2015-11-17

Topics in Complexity: From Physical to Life Science Systems

Pedro D. Manrique Charry
University of Miami, pmanriq@gmail.com

Recommended Citation

Manrique Charry, Pedro D., "Topics in Complexity: From Physical to Life Science Systems" (2015). *Open Access Dissertations*. 1525.
https://scholarlyrepository.miami.edu/oa_dissertations/1525

UNIVERSITY OF MIAMI

TOPICS IN COMPLEXITY: FROM PHYSICAL TO LIFE SCIENCE SYSTEMS

By

Pedro David Manrique Charry

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2015

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

TOPICS IN COMPLEXITY: FROM PHYSICAL TO LIFE SCIENCE SYSTEMS

Pedro David Manrique Charry

Approved:

Neil F. Johnson Ph.D. Professor of Physics	Joshua L. Cohn Ph.D. Professor of Physics
Chaoming Song Ph.D. Assistant Professor of Physics	Dean of the Graduate School
Stefan Wuchty Ph.D. Associate Professor of Computer Science	

MANRIQUE CHARRY, PEDRO DAVID (Ph.D., Physics)
Topics in Complexity: From Physical to Life Science Systems (December 2015)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Neil F. Johnson.
No. of pages in text. (158)

Complexity seeks to unwrap the mechanisms responsible for collective phenomena across the physical, biological, chemical, economic and social sciences. This thesis investigates real-world complex dynamical systems ranging from the quantum/natural domain to the social domain. The following novel understandings are developed concerning these systems' out-of-equilibrium and nonlinear behavior. Standard quantum techniques show divergent outcomes when a quantum system comprising more than one subunit is far from thermodynamic equilibrium. Abnormal photon inter-arrival times help fulfill the metabolic needs of a terrestrial photosynthetic bacterium. Spatial correlations within incident light can act as a driving mechanism for an organisms adaptation toward more ordered structures. The group dynamics of non-identical objects, whose assembly rules depend on mutual heterogeneity, yield rich transition dynamics between isolation and cohesion, with the cohesion regime reproducing a particular universal pattern commonly found in many real-world systems. Analyses of covert networks reveal collective gender superiority in the connectivity that provides benefits for system robustness and survival. Nodal migration in a network generates complex contagion profiles that lie beyond traditional approaches and yet resemble many modern-day outbreaks.

To God and my beloved family for their selfless love,

understanding, support and encouragement

Acknowledgments

I would like to thank my supervisor Dr. Neil F. Johnson for his guidance and encouragement during these years of study. Through his mentorship I was able to accomplish several goals and establish new challenges for the future. In addition, I extend my gratitude to Professor Luis Quiroga, Professor Ferney Rodríguez, Professor Stefan Wuchty, Professor Chaoming Song, Professor Joshua Cohn and Professor Pak Ming Hui for their collaborative work on several topics of this dissertation.

I thank the faculty and staff members of the department of Physics at the University of Miami for their active contribution during my years of formation.

PEDRO MANRIQUE CHARRY

University of Miami

December 2015

Contents

List of Figures

List of Tables

List of Publications

1. N.F. Johnson, P. Manrique and P.M. Hui. *Modeling Insurgent Dynamics Including Heterogeneity A Statistical Physics Approach*, J. Stat. Phys. **151**, 395 (2013)

2. N.F. Johnson, P. Medina, G. Zhao, D.S. Messinger, J. Horgan, P. Gill, J.C. Bohorquez, W. Mattson, D. Gangi, H. Qi, P. Manrique, N. Velasquez, A. Morgenstern, E. Restrepo, N. Johnson, M. Spagat and R. Zarama. *Simple Mathematical Law Benchmarks Human Confrontations*. Sci. Rep. **3**, 3463 (2013)

3. N. Johnson, G. Zhao, F. Caycedo, P. Manrique, H. Qi, F. Rodriguez and L. Quiroga. *Extreme Alien Light Allows Survival of Terrestrial Bacteria* Sci. Rep. **3** 2198, (2013)

4. A.P. Morgenstern, N. Velásquez, P. Manrique, H. Qi, N. Johnson and N.F Johnson. *Modeling Political Conflict, Violence, and Wars: A Survey*. American Journal of Physics **81**, 805 (2013)

5. P. Manrique, H. Qi, A. Morgenstern, N. Velásquez, T.C. Lu and N. Johnson. *Context Matters: Improving the Uses of Big Data for Forecasting Civil Unrest* IEEE Intl. Conf. on Intelligence and Security Informatics. 169-172 (Seattle, WA, 2013)

6. P.D. Manrique, A. De Mendoza, F. Caycedo-Soler, F. Rodríguez, L. Quiroga and N.F. Johnson. *Survivability of Photosynthetic Bacteria in Non-Terrestrial Light*. Journal of Astrobiology & Outreach. **3** 124, (2015)

7. P.D. Manrique, F. Rodríguez, L. Quiroga, and N.F. Johnson. *Nonequilibrium Quantum Systems: Divergence between Global and Local Descriptions*. Advances in Condensed Matter Physics. **2015**, 615727, (2015)

List of Preprints

1. P. Manrique, A. De Mendoza, F. Caycedo-Soler, F. Rodríguez, L. Quiroga, N. Johnson. *Thermal Photon Spatial Coherence Favors Photosynthetic Life*
2. P. Manrique, Z. Cao, A. Gabriel, J. Horgan, P. Gill, H. Qi, E.M. Restrepo, D. Johnson, S. Wuchty, C. Song and N. Johnson. *Female Minority Emerges as Dominant in Networks under Extreme Pressure*
3. P.D. Manrique, P.M. Hui, N.F. Johnson. *Internal Character Dictates Transition Dynamics Between Isolation and Cohesive Grouping*
4. P.D. Manrique, H.Qi, M.Zheng, C.Xu, P.M. Hui & N.F. Johnson. *Anomalous Contagion and Renormalization in Dynamical Networks*
5. H. Qi, P.D. Manrique, M. Zheng, C.Xu, P.M. Hui & N.F. Johnson. *Common Space Contagion: ISIS and Beyond*
6. H. Qi, N. Velásquez, Y. Vorobyeva, A. Gabriel, P. Manrique, A. Morgenstern, D. Johnson, C. Song, S. Wuchty, E.M. Restrepo & N. Johnson. *Latent Escalation of Self-Organized Cyber Aggregates Predicts Sudden Uprisings*

Chapter 1

Introduction

Complexity analyzes phenomena emerging from a collection of interacting objects [1; 2; 3; 4; 5; 6; 7]. Large everyday life phenomena such as market crashes [2; 8], cancer tumors [9], mass mobilization [10; 11], among others enter into the focus of complexity. There is no unified theory of complexity. However, emerging phenomena rise from the build-up of individual 'bricks' that are somehow connected to each other i.e., they are network-structured. Hence, graph theory as well as agent-based models become essential tools to study complex systems [12; 13; 14]. Real-world networks (e.g. protein networks, internet) show complex, adaptive and self-organized behavior [15; 16]. Two common perspectives are used in order to study these systems: topology and dynamics. Several studies focus on only one perspective, and how one affects the other, arguing that dynamics and growth occur on different time scales. However, new studies point to the evolution of the network to be an interplay between topology and dynamics. Assembling mechanisms based on topological constrains (e.g. stability) determine the ultimate progression of the system [17; 18]. The mechanisms behind the interaction among individuals or *agents* are key to construct models that quantify and predict the evolution of the system and hence explore how they influence it at the macro-level. Likewise, exploring connections between systems that previously were considered unrelated. For instance, it has been reported that the severity of violent actions by insurgent groups [19; 20; 21], the size distribution of pockets in superconducting coherence in fragmented materials [22], the size distribution of neuronal avalanches [23] and the size distribution of herds of agreeing traders [24], all share the emergent feature of a 5/2 power-law size distribution of aggregates. Though these systems are fundamentally different from each other at the single-particle level, they have shown to exhibit this universal pattern.

Complexity covers a wide variety of systems. However, certain commonalities have been identified as essential ingredients of a complex system [1]: (i) they are typically non-Markovian, i.e., memory effects are relevant to their dynamics. (ii) They are open, meaning that the dynamics is affected by the interaction with an external environment. (iii) The system tends to fluctuate between states of order and disorder. (iv) They are far from equilibrium. Therefore, fluctuations could give rise to collective effects that would not be predicted only by the knowledge of the system at the single-particle level. (v) The system is capable of adaptation. These ingredients are essential in, for instance, biology. Living organisms have developed over the course of millions of years of evolution, the capacity of adapting to highly fluctuating environmental conditions. This is the case of photosynthetic single cell organisms whose apparatuses adapt to different light sources in order to successfully satisfy their metabolic needs [25; 26]. They manage to optimize light absorption and energy funneling on low light intensity and protect the membrane from photo-damage by boosting dissipation mechanisms on high light intensity. These highly non-trivial properties have been the cause of heated debates in the physical, biological, and chemical communities for decades. Currently, a detailed atom-based computational description of a photosynthetic unit is in progress at the University of Illinois in Urbana-Champagne [27; 28; 29; 30]. The system is comprised of 100 million atoms that account for lipids, proteins, cofactors and background water where the full photosynthetic vesicle is embedded. The supercomputer performs nearly 27 quadrillion calculations per second with the goal of simulating $1\mu s$ of molecular processes. Given the size of the system, only the transferring of data could take days to complete. Though such detailed description could indeed help in the understanding of some of the processes involved in the photosynthetic apparatus, the method is computationally very expensive and fails to obtain a big picture of the full photosynthetic process. The reason is that these processes involve several competing time-scales that range from 1fs to 1s [31; 32; 25; 33; 34]. The path to understand the adaptation to different environmental conditions through this traditional reductionist approach is still dark. Complexity provides an alternative route to deal with this problem. Caycedo-Soler, et al. [35; 36; 37] have been able to explain this adaptation by quantifying a dynamical interplay between energy transfer kinetics and energy processing cycling. They used a stochastic method to describe the collective processing of light energy packets (photons) by the network of light-harvesting complexes. In this way, complexity provides a holistic description of the system that captures the underlying physics responsible for the chromatic adaptation. More on this in chapter 3.

In this thesis, we investigate real-world complex dynamical systems from the quantum/natural domain to the social domain. The goal is to unveil the mechanisms behind their typical non-linear, out-of-equilibrium behavior. The systems are fundamentally different and hence the methodology developed and employed, tends to depend on the specific system to be studied. However, it does show certain commonalities in terms of the focus on extreme behavior. In this way, this thesis pushes the field forward in different areas and as a whole since many of the techniques used are common across the chapters.

In chapter 2, a simple but highly non trivial problem is presented. The quantum process of thermalization of qubits is studied when the system is far from thermodynamic equilibrium. The ultrafast timescales (fs-ps) challenge traditional physical assumptions that the system can be considered to be isolated [38; 39]. Given that the most fundamental process for Life on Earth (i.e. photosynthesis) processes excitations on ultrafast scales, has raised the question of whether we truly understand its dynamics [40; 41; 42]. By employing two popular approaches, we show that these methods yield to contrasting outcomes when the number of qubits is greater than one. This divergence highlights the need to establish rigorous ranges of applicability for such methods in modeling nanoscale transfer phenomena.

Chapters 3 and 4 are devoted to structural and physical aspects of bacterial photosynthesis. Harvesting of sunlight underpins Life on Earth as well as driving novel energy device design[26; 31; 43; 44; 45; 25; 35; 36]. Although several experiments suggest that both photosynthetic excitation energy transport and charge separation [31; 46; 47; 48; 49; 50] may benefit from the quantum nature of their dynamics, the effects of spatial coherences in the incident light have been largely ignored. Here it is shown that spatial correlations in the incident light likely play an important role in trapping light and adding robustness, as well as providing a driving mechanism for an organism's adaptation toward more ordered structures. Our theory is grounded on empirical inputs, while its output is validated against testable predictions. The results suggest a key role for the spatio-temporal correlation between photons, a fundamental property of the quantum world, in understanding early Life and in improving the design of artificial photosynthetic systems.

Chapter 5 looks at the effects introduced by individual heterogeneity to a simple but highly non-trivial fission/fusion grouping model. Dynamical grouping underlies myriad collective phenomena across the physical, biological, chemical, economic and social sciences [51; 52; 53; 54; 55; 56; 57; 58]. Whether the underlying N objects are particles, people or proteins, the issue of whether they evolve as isolated individuals or aggregates has significant consequences at the macro-level [57; 58; 59; 60; 61; 62; 63]. A serious shortcoming is that the underlying objects (e.g. people, cells, animals etc.) are typically not identical, and it is this heterogeneity that typically dictates their interactions and ultimately their collective behavior. An outstanding question is therefore how this diversity in individual characters affects the dynamics of groups [64; 65; 66; 67]. And how can this individual-level heterogeneity be reconciled with the emergent universality observed across many diverse real-world phenomena? This chapter attempts to address these questions by adding a simple, continuous 'character' variable x_i to each object i, and then allowing objects' characters to influence how they interact with each other. Analytical and numerical calculations reveal that different critical points arise for different heterogeneity-dependent grouping mechanisms. These critical points move in opposite directions as the population's diversity decreases. Regardless of the grouping mechanism, cohesion regime follows a particular universal pattern that is commonly found in real-world systems. Previous work [68; 69; 67], including in the absence of character, suggests that our main conclusions will hold for a variety of model generalizations.

In chapter 6, a study of covert networks under pressure is presented. Off-line networks of Provisional Irish Republican Army (PIRA)[70; 71] and on-line networks of Islamic State (ISIS) are quantitatively studied and compared with basic elements of graph theory and agent based modeling. These networks, in addition to being the only operational networks under extreme pressure for which detailed longitudinal information is known concerning the links and nodes over time, are also the most successful examples for each genre [71; 72]. Our results point to a collective female superiority in the network's connectivity that affords benefits for system robustness and survival [62; 73], despite of women being out numbered by men. This result challenges the common perception that women naturally play more minor roles to men as environments become more dangerous and aggressive [74; 75; 76; 77]. Female centered clusters are associated to an increment in the lifetime of neighboring actors as well as the group's impact as a whole. A fission/fusion model which captures the non-monotonic evolution in the connectivity, suggests that women spread a team-oriented culture which offers benefits to group's secrecy and autonomy. Findings are tested against null models discarding effects due to noise. Our results propose that the effective disruption of a extreme network could lie on female-centric approaches where the interconnectivity of few members is struck as opposed to the majority of men.

Chapter 7 looks at contagion in dynamical networks. Modern-day human contagion phenomena [78; 79], whether online [80], offline [81], social [82] or disease-based [83; 84; 85; 86; 87; 88; 89]. Unfortunately such modern-day outbreak profiles often lie well beyond the predictions of standard infection models [82; 84; 90; 91]. Here we show that the simple human tendency to move through online and offline spaces, such as social sites or communities, generates novel contagion dynamics that capture outbreak profiles ranging from online pro-ISIS activity to the spread of financial rumors [92] and mass street protests [10; 11]. Generally speaking, our results reveal a highly nonlinear dependence on human throughput that offers a less invasive control method for future outbreaks than traditional approaches [78]. Our findings establish the counterintuitive result that by increasing throughput of a common online or off-line space, an outbreak's severity can actually be decreased. Even if the number of nodes remains static on average, the mobility of individuals entering and leaving the common place, impacts the severity of the outbreak. Extension to multiple common places is subsequently expanded where commonalities with electric circuits are explored.

Chapter 2

Non-equilibrium quantum systems

The rapid recent development of quantum control, quantum engineering and information processing at the nanoscale in biological, chemical and condensed matter systems, has led to a crucial need to improve our understanding of *open* quantum systems. The typical physics assumptions of an isolated system, particularly under non equilibrium conditions, cannot hold for systems probed on the quantum scale at optical pico- and femtosecond scales [39]. The accurate description of excitation dynamics at these ultrafast scales is essential for the understanding of fundamental processes for Life on Earth such as photosyhthesis [40; 42; 41]. For such reasons, theoretical physicists have begun to develop theoretical and experimental tools to study the dynamical behavior of an open system interacting with its environment. The key lies in identifying an accurate way of removing the environmental degrees of freedom, and hence obtain a closed equation of motion for the reduced system of interest [38; 93; 94; 95; 96].

Quantum master equations have been tested for exactly solvable interacting harmonic systems in thermal equilibrium and nonequilibrium providing reliable results [97]. However, many systems of interest (e.g. optically probed semiconductor quantum dots, or biological photosynthetic processes) show fluctuations which are far from equilibrium. Moreover, recent evidence suggests that excitation energy transfer in biological systems, particularly photosynthetic membranes, might involve some level of quantum coherence [98; 99]. In this chapter, we provide analytic results for the steady state of an open quantum system interacting with two reservoirs at different temperatures following local and global approximations. This mimics the situation in photosynthetic membrane reaction centers – and elsewhere in natural and artificial systems – in which transfer occurs between few-level molecular complexes that in turn may be coupled to reservoirs with different effective temperatures. In both cases we use the same assumptions and the same procedure – the only difference is the election of the free Hamiltonian and, therefore, the basis set. The local approach is commonly used to model incoherent transfer phenomena in small quantum systems [42], while the global approach is a more rigorous way to calculate the quantum properties such as coherence and entanglement [100; 101]. For the simple case of an interacting qubit dimer where one of the qubits is coupled with a thermal bath, the approximations are equivalent only for the case of zero bath temperature [102]. Similar studies have been performed for two interacting quantum harmonic oscillators under non equilibrium thermal conditions, where the local approach is found to violate the second law of thermodynamics for the non-symmetrical case [103] (non identical systems). This discrepancy represents a serious challenge for modeling of quantum systems. We consider an open system comprising interacting subunits (qubits), which could provide a simple representation of interacting two-level systems in a photosynthetic membrane reaction center. The question of how to solve the problem can then be addressed in two ways: *(i)* Diagonalize the Hamiltonian of the open system and solve the problem in terms of the diagonal basis set (i.e. *global* approach). *(ii)* Use the direct product of the individual basis set of the interacting subunits (i.e. *local* approach). We here apply both these methods in parallel, and show explicitly how each formulation leads to a different result when the number of interacting subunits is greater than one.

This chapter is divided in four parts. In the second part the methods are presented for an arbitrary system. Analytic expressions for populations and heat current are derived and applied to a two level system. In the third part, the quantum system is extended to an interacting qubit dimer where the quantities are calculated. The fourth part is devoted to analysis and conclusions.

2.1 Formalism

Consider a quantum system under non-equilibrium thermal conditions. Each reservoir is modeled as an infinite collection of harmonic oscillators in thermal equilibrium at temperature given by $\beta_j = 1/k_B T_j$, $j = 1, 2$. We assume that the coupling strength between the reservoirs and the central sub system is weak, hence we can express the total density operator as a direct product of the reduced density operators of the open system $\hat{\rho}_s$ and the reservoirs $\hat{\rho}_1$ and $\hat{\rho}_2$: $\hat{\rho} = \hat{\rho}_s \otimes \hat{\rho}_1 \otimes \hat{\rho}_2$. The Hamiltonian of the whole system is then

$$\hat{H} = \hat{Q} + \hat{R}_1 + \hat{R}_2 + \hat{S}_1 + \hat{S}_2 \tag{2.1}$$

where $\hat{Q}$ is the Hamiltonian of the open system and the terms $\hat{R}_j$ and $\hat{S}_j$ are the reservoir Hamiltonian and the interaction term associated with the reservoir j, respectively. The usual route to this problem is to solve directly the second-order expansion of the Liouville von Neumann equation. This is achieved by obtaining the reduced density operator for the central subsystem through a partial trace over the reservoirs degrees of freedom. Within the interaction picture representation, the dynamics for the whole system is given by

$$\frac{d\hat{\rho}}{dt} = -i[\hat{H}_I(t), \hat{\rho}(0)] - \int_0^t dt_1 [\hat{H}_I(t), [\hat{H}_I(t_1), \hat{\rho}(t_1)]] \tag{2.2}$$

The Hamiltonian in the interaction picture representation is defined as $\hat{H}_I(t) = \hat{U}_0^\dagger \hat{H}_I \hat{U}_0$, where $\hat{U}_0$ is the evolution unitary operator of the free system. The free system is considered as a subsystem whose solutions are well known. For the present work, we use the two methods to describe the free system: global and local.

2.1.1 Global approach

For this global approach, it is useful to express the interaction term in the form

$$\hat{S}_j = \sum_\mu \hat{V}_{j,\mu} \hat{f}_{j,\mu} = \sum_\mu \hat{V}_{j,\mu}^\dagger \hat{f}_{j,\mu}^\dagger \tag{2.3}$$

where the operators $\hat{V}_{j,\mu}$ act over the open system's degrees of freedom while the operators $\hat{f}_{j,\mu}$ act on the reservoir j. The operators $\hat{V}_{j,\mu}$ are chosen in such a way that they follow the following commutation relationship:

$$[\hat{Q}, \hat{V}_{j,\mu}] = \omega_\mu \hat{V}_{j,\mu}. \tag{2.4}$$

This decomposition is always possible [38; 97; 104]. For instance, for an operator $\hat{V}_\mu = |i\rangle\langle i|\hat{S}|k\rangle\langle k|$, where $|i\rangle$ and $|k\rangle$ are eigenstates of $\hat{Q}$, it results in $\omega_\mu = \langle i|\hat{Q}|i\rangle - \langle k|\hat{Q}|k\rangle$. The dynamics, as was mentioned above, is governed by the Liouville von Neumann equation of motion $d\hat{\rho}/dt = -i[\hat{H}, \hat{\rho}]$. By using the Born-Markov approximation, the dynamics of the open system in terms of its reduced density operator $\hat{\rho}_s$ is given by:

$$\frac{d}{dt}\hat{\rho}_s = -i[\hat{Q}, \hat{\rho}_s] - \sum_{j=1}^{2} \mathcal{L}_j(\hat{\rho}_s) \tag{2.5}$$

where $\mathcal{L}_j$ is the Lindbland super-operator associated to the reservoir j [100]

$$\mathcal{L}_j(\hat{\rho}_s) = \sum_{\mu,\nu} J^{(j)}_{\mu,\nu}(\omega_\nu) \left\{ [\hat{V}_{j,\mu}, [\hat{V}^{\dagger}_{j,\nu}, \hat{\rho}_s]] - (1 - e^{\beta_j \omega_\nu})[\hat{V}_{j,\mu}, \hat{V}^{\dagger}_{j,\nu}\hat{\rho}_s] \right\} \tag{2.6}$$

where the indices μ and ν run over the complete range of operators and $J^{(j)}_{\mu,\nu}(\omega_\nu)$ is the spectral density is defined by [97; 100; 104]

$$J^{(j)}_{\mu,\nu}(\omega_\nu) = \int_0^\infty d\tau e^{i\omega_\nu \tau} \mathrm{Tr}_{R_j} \left\{ \rho_j \bar{f}^{\dagger}_{j,\nu}(\tau) \hat{f}_{j,\mu} \right\} \tag{2.7}$$

where $\bar{f}_{j,\nu}(\tau) = e^{-i\hat{R}_j \tau} \hat{f}_{j,\nu} e^{i\hat{R}_j \tau}$ is the interaction picture representation of the reservoir operator $\hat{f}_{j,\nu}$. For each selection of the open system, a new set of operators $\{\hat{V}_{j,\mu}\}$ is found.

2.1.2 Local approach

Similarly, the problem can be addressed by using a free Hamiltonian which is formed by summing all zero-point Hamiltonians of each subunit. For a simple subunit such as the qubit, the Hilbert space is spanned by two states with energy gap of ϵ, and the Hamiltonian can be written in terms of the 2×2 pauli matrices as $\hat{Q} = (1/2)\sigma^z \epsilon$. For instance, for the case where the open system is composed by a linear chain of N qubits and considering an *XX*-like interaction between them, the Hamiltonian of the open system can be written as $\hat{Q} = \hat{Q}_0 + \hat{Q}_I$, where $\hat{Q}_0 = \sum_{q=1}^{N} \frac{1}{2}\epsilon_q \sigma^z_q$ is the contribution of each sub system and $\hat{Q}_I = \sum_{i=1}^{N-1} K_i \hat{\sigma}^+_i \hat{\sigma}^-_{i+1} + \mathrm{h.c.}$ describes the inter-qubit interaction. Consequently, the free Hamiltonian $\hat{H}_0$ and interaction Hamiltonian $\hat{H}_I$ can be identified as

$$\hat{H}_0 = \hat{Q}_0 + \hat{R}_1 + \hat{R}_2 \tag{2.8}$$

$$\hat{H}_I = \hat{Q}_I + \hat{S}_1 + \hat{S}_2\,. \tag{2.9}$$

Hence in the interaction picture representation, the Hamiltonian can be written as

$$\begin{aligned}\hat{H}_I(t) &= \bar{Q}_I(t) + \bar{S}_1(t) + \bar{S}_2(t) &(2.10)\\ \bar{Q}_I(t) &= \sum_{i=1}^{N-1} K_i \hat{\sigma}_i^+ \hat{\sigma}_{i+1}^- e^{i(\epsilon_i - \epsilon_{i+1})t} + \text{h.c.} &(2.11)\\ \bar{S}_j(t) &= \sum_k g_k^{(j)} \hat{a}_{k,j} \sigma_\lambda^+ e^{i\left(\epsilon_\lambda - \epsilon_{j,k}\right)t} + \text{h.c.} &(2.12)\end{aligned}$$

where $\hat{a}_{k,j}^\dagger$ creates an excitation in mode k of reservoir j with a coupling strength of $g_k^{(j)}$ and energy $\epsilon_{j,k}$. The subindex λ labels the subunit interacting with the reservoir j, i.e. $\lambda = 1$ when $j = 1$ and $\lambda = N$ for $j = 2$. Note that for the energy associated to the central system we use only one subindex, e.g. ϵ_λ, while for the energy associated with one mode of the bath we use two, e.g. $\epsilon_{j,k}$. We use the commutation relations for Pauli operators ($[\sigma^+, \sigma^-] = \sigma^z$ and $[\sigma^z, \sigma^\pm] = \pm 2\sigma^\pm$), and for bosonic operators ($[\hat{a}_k, \hat{a}_{k'}^\dagger] = \delta_{k,k'}$). As a result of this transformation, we can use equation 2.2 to solve the problem. Specifically, we take the partial trace over the reservoirs and use the Born-Markov approximation. Furthermore, we take the continuous limit for the reservoir energies and the wide band limit on the interaction with the central subsystem. This procedure yields to a delta function in the energies ($\delta(\epsilon_\lambda - \epsilon_{j,k})$) that collapses all the energy spectra of reservoit j into the energy of the subunit λ. Under these conditions, the quantum optical Master equation for a chain of qubits whose endpoints ($\lambda = 1$ and $\lambda = N$) interact with bosonic reservoirs, is given by

$$\begin{aligned}\frac{d\hat{\rho}_c}{dt} &= -i[\bar{Q}_I(t), \hat{\rho}_c] + \sum_{j=1}^{2} J^{(j)}(\epsilon_\lambda) e^{\beta_j \epsilon_\lambda} \left(\hat{\sigma}_\lambda^- \hat{\rho}_c \hat{\sigma}_\lambda^+ - \frac{1}{2}\{\hat{\sigma}_\lambda^+ \hat{\sigma}_\lambda^-, \hat{\rho}_c\} \right)\\ &\quad + J^{(j)}(\epsilon_\lambda) \left(\hat{\sigma}_\lambda^+ \hat{\rho}_c \hat{\sigma}_\lambda^- - \frac{1}{2}\{\hat{\sigma}_\lambda^- \hat{\sigma}_\lambda^+, \hat{\rho}_c\} \right) \qquad (2.13)\end{aligned}$$

where $J^{(j)}$ denotes the spectral density function associated with the interaction between the qubit λ and the reservoir j given in terms of the spontaneous emission rate γ_j and the

Bose-Einstein distribution $\bar{N}_j(\epsilon) = (e^{\beta_j \epsilon} - 1)^{-1}$

$$J^{(j)}(\epsilon) = \gamma_j \bar{N}_j(\epsilon). \tag{2.14}$$

For simplicity, we consider $\gamma_j = 1$.

2.1.3 Observable: thermal energy

As a test observable, we use the steady state thermal energy transferred from the reservoir to the quantum system given the non-zero thermal bias. The steady-state heat flux is defined as the trace of the product of the Liouvillian super-operator with the subsystem Hamiltonian:

$$\mathcal{Q}_j = \mathrm{Tr}\{\hat{Q}\mathcal{L}_j\} \tag{2.15}$$

The resultant expression for the heat flux in the global approach can be written in the following compact form:

$$\mathcal{Q}_j = \sum_{\mu,\nu} \omega_\mu J^{(j)}_{\mu,\nu}(\omega_\nu) \left\{ \left\langle \hat{V}^\dagger_{j,\nu} \hat{V}_{j,\mu} \right\rangle - e^{\beta_j \omega_\nu} \left\langle \hat{V}_{j,\mu} \hat{V}^\dagger_{j,\nu} \right\rangle \right\}. \tag{2.16}$$

In a similar way, the result for the heat flux in the local approach is found to be

$$\mathcal{Q}_j = \epsilon_\lambda J^{(j)}(\epsilon_\lambda) \left\{ \langle \hat{\sigma}^-_\lambda \hat{\sigma}^+_\lambda \rangle - e^{\beta_j \epsilon_\lambda} \langle \hat{\sigma}^+_\lambda \hat{\sigma}^-_\lambda \rangle \right\}. \tag{2.17}$$

As expected, the formula (2.17) gives a zero flux value when the reservoirs are set at the same temperature. Furthermore, it can be seen that the global and local approaches lead to the same result when applied to a single qubit under non-equilibrium thermal conditions where the system operators are the same, $\hat{V}^{(\dagger)}_{j,\mu} = \hat{\sigma}^{+(-)}$. The steady-state population of the excited state n of one qubit system with energy gap ϵ is found to be

$$n = \langle \hat{\sigma}^+ \hat{\sigma}^- \rangle = \frac{1}{2} e(\epsilon) \tag{2.18}$$

where e_j is a simple universal function defined as

$$e(\omega_j) = \frac{\bar{N}_1(\omega_j) + \bar{N}_2(\omega_j)}{1 + \bar{N}_1(\omega_j) + \bar{N}_2(\omega_j)} \leq 1. \tag{2.19}$$

Furthermore, the heat flux for this system is found to be

$$\begin{aligned} \mathcal{Q}_1 &= \epsilon J^{(1)}(\epsilon) J^{(2)}(\epsilon) \frac{e^{\beta_2 \epsilon} - e^{\beta_1 \epsilon}}{J^{(1)}(\epsilon)(1 + e^{\beta_1 \epsilon}) + J^{(2)}(\epsilon)(1 + e^{\beta_2 \epsilon})}. && (2.20) \\ &= \frac{\epsilon}{2} (1 - e(\epsilon)) \left(\bar{N}_1(\epsilon) - \bar{N}_2(\epsilon) \right) && (2.21) \end{aligned}$$

As can be seen, the formula 2.21 leads to a zero value when the thermal bias is set as zero. In addition, note the heat flux is always positive for positive bias ($\bar{N}_1 > \bar{N}_2$) and negative, otherwise.

2.2 Divergence for a dimer

We consider a dimer composed of two interacting qubits, where each quit is connected to a different reservoir in thermal equilibrium at temperature T_j $(j = 1, 2)$. Figure 2.1 shows schematically the system to be studied. The Hamiltonian of the qubit dimer is

$$\hat{Q} = \frac{1}{2}\epsilon_1 \hat{\sigma}_1^z + \frac{1}{2}\epsilon_2 \hat{\sigma}_2^z + K(\hat{\sigma}_1^+ \hat{\sigma}_2^- + \hat{\sigma}_1^- \hat{\sigma}_2^+) \tag{2.22}$$

where $\hat{\sigma}_1 = \hat{\sigma} \otimes I$ and $\hat{\sigma}_2 = I \otimes \hat{\sigma}$, with I as the 2×2 identity matrix. In addition, we assume that the qubit labeled as j interacts with the bath labeled as j only, for $j = 1, 2$. Hence the interaction Hamiltonian between the dimer and the reservoirs can be written as

$$\hat{S}_j = \sum_k g_{k,j} \sigma_j^+ \hat{a}_{k,j} + \text{h.c.} \tag{2.23}$$

The operators $\hat{\sigma}_j^{\pm}$ do not commute with $\hat{Q}$, so for the global approach it is necessary to find a transformation that allow us to write the interaction Hamiltonian in the form of Equation

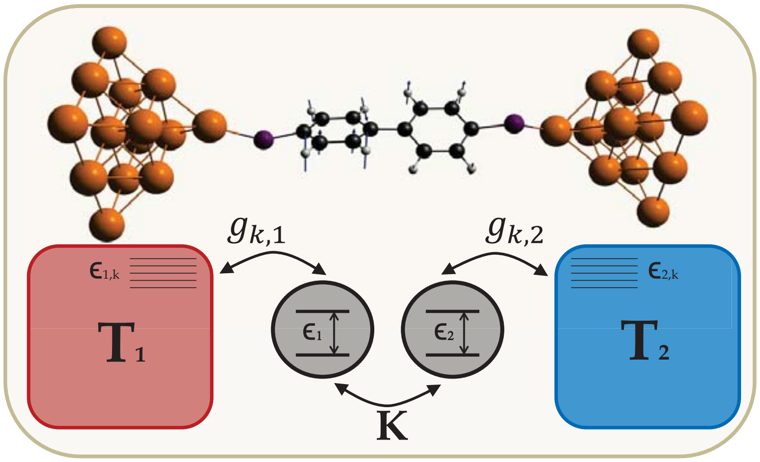

Figure 2.1: Extended quantum system interacting with two thermal reservoirs.

(2.3), so that the condition (2.4) is fulfilled. The eigenstates and eigenenergies of $\hat{Q}$ are $|s_1\rangle = |0,0\rangle$ ($E_1 = -\frac{\epsilon_1+\epsilon_2}{2}$), $|s_2\rangle = |1,1\rangle$ ($E_2 = \frac{\epsilon_1+\epsilon_2}{2}$), $|s_3\rangle = c_{3,1}|1,0\rangle + c_{3,2}|0,1\rangle$ ($E_3 = \alpha$), and $|s_4\rangle = c_{4,1}|1,0\rangle + c_{4,2}|0,1\rangle$ ($E_4 = -\alpha$) where the amplitudes and constants are given by:

$$\begin{aligned} c_{3,1} &= \frac{K}{\sqrt{2\alpha^2 - \alpha\Delta\epsilon}} \quad , \quad c_{3,2} = \frac{\alpha - \Delta\epsilon/2}{\sqrt{2\alpha^2 - \alpha\Delta\epsilon}} \\ c_{4,1} &= \frac{K}{\sqrt{2\alpha^2 + \alpha\Delta\epsilon}} \quad , \quad c_{4,2} = -\frac{\alpha + \Delta\epsilon/2}{\sqrt{2\alpha^2 + \alpha\Delta\epsilon}} \end{aligned} \tag{2.24}$$

$$\alpha = \sqrt{K^2 + \frac{\Delta\epsilon^2}{4}} \quad , \quad \Delta\epsilon = \epsilon_1 - \epsilon_2 \; . \tag{2.25}$$

The transformation of the coupling operators from the individual qubits basis set into the dimer diagonal basis set, can be calculated as

$$\hat{\sigma}_j = \sum_{p=1}^{4}\sum_{q=1}^{4} |s_p\rangle \langle s_p| \hat{\sigma}_j |s_q\rangle \langle s_q| \tag{2.26}$$

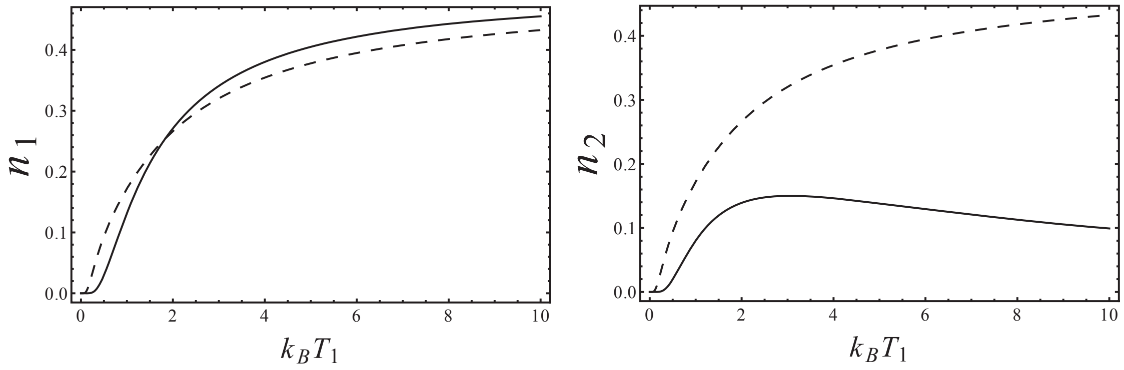

Figure 2.2: Steady state populations n_1 (left) and n_2 (right) as a function of the temperature of the reservoir 1 for both approximations. The solid lines represent the local approach while the dashed lines represent the global approach. In all cases, the temperature of reservoir 2 is set as zero, $K = 1$ and $\epsilon = 1.5$.

With this transformation, the condition (2.4) is fulfilled $\left[\hat{Q}, |s_p\rangle\langle s_q|\right] = (E_p - E_q)|s_p\rangle\langle s_q|$.

In this way, the operators can be found to be

$$\begin{aligned}
\hat{V}_{j,1} &= [c_{3,2}\delta_{j,1} + c_{3,1}\delta_{j,2}]\,|s_2\rangle\langle s_3|\,, \quad \omega_1 = E_2 - E_3 \\
\hat{V}_{j,2} &= [c_{4,2}\delta_{j,1} + c_{4,1}\delta_{j,2}]\,|s_2\rangle\langle s_4|\,, \quad \omega_2 = E_2 - E_4 \\
\hat{V}_{j,3} &= [c_{3,1}\delta_{j,1} + c_{3,2}\delta_{j,2}]\,|s_3\rangle\langle s_1|\,, \quad \omega_3 = E_3 - E_1 \\
\hat{V}_{j,4} &= [c_{4,1}\delta_{j,1} + c_{4,2}\delta_{j,2}]\,|s_4\rangle\langle s_1|\,, \quad \omega_4 = E_4 - E_1
\end{aligned} \tag{2.27}$$

For simplicity we consider the symmetric case where the energy gap of each qubit is the same: $\epsilon_1 = \epsilon_2 = \epsilon$. In particular, we have that $\omega_1 = \omega_4 = |\epsilon - K|$, $\omega_2 = \omega_3 = \epsilon + K$ and the amplitudes are $c_{3,1} = c_{3,2} = c_{4,1} = 1/\sqrt{2}$, and $c_{4,2} = -1/\sqrt{2}$. The steady-state populations can be calculated for each of the qubits

$$n_1 = \left\langle \hat{\sigma}_1^+ \hat{\sigma}_1^- \right\rangle = \rho_{22}^{(G)} + \frac{1}{2}\left(\rho_{33}^{(G)} + \rho_{44}^{(G)} + 2\mathrm{Re}\left\{\rho_{34}^{(G)}\right\}\right) \tag{2.28}$$

$$n_2 = \left\langle \hat{\sigma}_2^+ \hat{\sigma}_2^- \right\rangle = \rho_{22}^{(G)} + \frac{1}{2}\left(\rho_{33}^{(G)} + \rho_{44}^{(G)} - 2\mathrm{Re}\left\{\rho_{34}^{(G)}\right\}\right) \tag{2.29}$$

where the matrix elements $\rho_{ij}^{(G)} = \langle s_i|\hat{\rho}_s|s_j\rangle$ are calculated in the diagonal basis. For the weak-coupling case ($\epsilon > K$), the populations are found to be ($K = 1$):

$$\begin{aligned} \rho_{11}^{(G)} &= \left(1 - \frac{e_1}{2}\right)\left(1 - \frac{e_2}{2}\right), \quad \rho_{22}^{(G)} = \frac{e_1}{2}\frac{e_2}{2} \\ \rho_{33}^{(G)} &= \left(1 - \frac{e_1}{2}\right)\frac{e_2}{2}, \quad \rho_{44}^{(G)} = \frac{e_1}{2}\left(1 - \frac{e_2}{2}\right) \end{aligned} \tag{2.30}$$

where $e_j = e(\omega_j)$. For the steady-state, the matrix element $\rho_{34}^{(G)}$ is zero, therefore the populations take the simple form of $n_{1,2} = (e_1 + e_2)/4$. In a similar way the populations for the local approach can be derived by solving equation (13) for $N = 2$ leading to $n_{1(2)} = \rho_{11}^{(L)} + \rho_{22(33)}^{(L)}$, where the matrix elements $\rho_{ij}^{(L)}$ are calculated in the local basis, i.e. $\rho_{11}^{(L)} = \langle 11|\hat{\rho}_s|11\rangle$, $\rho_{22}^{(L)} = \langle 10|\hat{\rho}_s|10\rangle$, and $\rho_{33}^{(L)} = \langle 01|\hat{\rho}_s|01\rangle$. The following results follow:

$$n_1 = \frac{2K^2 e(\epsilon) + (1 + 2\bar{N}_2(\epsilon))\bar{N}_1(\epsilon)}{4K^2 + (1 + 2\bar{N}_1(\epsilon))(1 + 2\bar{N}_2(\epsilon))} \tag{2.31}$$

$$n_2 = \frac{2K^2 e(\epsilon) + (1 + 2\bar{N}_1(\epsilon))\bar{N}_2(\epsilon)}{4K^2 + (1 + 2\bar{N}_1(\epsilon))(1 + 2\bar{N}_2(\epsilon))} \tag{2.32}$$

In Figure 2.2 we can see how the populations change as the temperature of the reservoir 1 changes while the temperature of reservoir 2 is set to be zero. The two approaches for n_1 converge when $k_B T_1 >> \epsilon$. On the other hand, the population n_2 decays in the local approach as $k_B T_1 >> \epsilon$ while the global approach predicts an asymptotic growth to the mixed state of $1/2$. This divergence in predictions suggests that one of the approaches is not correct.

Another way to see this discrepancy is by looking at the heat flux. Using the system operators (2.27), the steady state heat flux (2.16) for reservoir 1 is:

$$\mathcal{Q}_1 = \sum_{i=1}^{2} \omega_i J^{(1)}(\omega_i) J^{(2)}(\omega_i) \frac{e^{\beta_2 \omega_i} - e^{\beta_1 \omega_i}}{(1 + e^{\beta_1 \omega_i}) J^{(1)}(\omega_i) + (1 + e^{\beta_2 \omega_i}) J^{(2)}(\omega_i)} \tag{2.33}$$

$$= \frac{\omega_1}{2}(1 - e_1)(\bar{N}_1(\omega_1) - \bar{N}_2(\omega_1)) - (1 \leftrightarrow 2) \tag{2.34}$$

The heat flux is expressed as a sum over the energy channels ω_1 and ω_2 that depend on the inter-qubit coupling K and qubit energy gap ϵ. The flux is always positive for positive thermal bias. On the other hand, the steady state heat flux from the reservoir 1 to the qubit 1, calculated in the local approach for the symmetric case, can be found to be

$$\begin{aligned} \mathcal{Q}_1 &= \frac{\epsilon J^{(1)}(\epsilon)J^{(2)}(\epsilon)(e^{\beta_2\epsilon}-e^{\beta_1\epsilon})}{J^{(1)}(\epsilon)(1+e^{\beta_1\epsilon})+J^{(2)}(\epsilon)(1+e^{\beta_2\epsilon})}\frac{4K^2}{4K^2+J^{(1)}(\epsilon)J^{(2)}(\epsilon)(1+e^{\beta_1\epsilon})(1+e^{\beta_2\epsilon})} \\ &= \left[\frac{\epsilon}{2}\left(1-e(\epsilon)\right)\left(\bar{N}_1(\epsilon)-\bar{N}_2(\epsilon)\right)\right]\frac{4K^2}{4K^2+J^{(1)}(\epsilon)J^{(2)}(\epsilon)(1+e^{\beta_1\epsilon})(1+e^{\beta_2\epsilon})} \end{aligned} \tag{2.35}$$

The local approach for the dimer leads to an expression for the heat flux expression equal to the one for the monomer weighted by a positive function that depends on the interqubit coupling K and the reservoirs' temperature. This function ensures that the flux tends to zero as the inter-qubit coupling decreases and remains finite for large K. This is reasonable since the qubits are weakly coupled (i.e. $K \to 0$), and therefore the heat transferred should decrease. However it also shapes the flux in such a way that an optimal value is found, i.e., the heat flux exhibits a maximum with the temperature and then decays to zero as the thermal bias grows. This behavior is not found for either the monomer or the dimer in the global approach.

As an illustration, Figure 2.3 shows the observable $\mathcal{Q}_1$ for both approximations and three different values of the qubit energy gap ϵ, as a function of the temperature difference. To simplify the presentation, we have set the temperature of the second reservoir to be close to zero and have depicted the flux as a function of $k_B T_1$. We can clearly see the maximum of $\mathcal{Q}_1$ for a specific value of the energy $k_B T_1$ in the local approach, while the result for the global approach grows asymptotically to ϵ as the thermal bias increases.

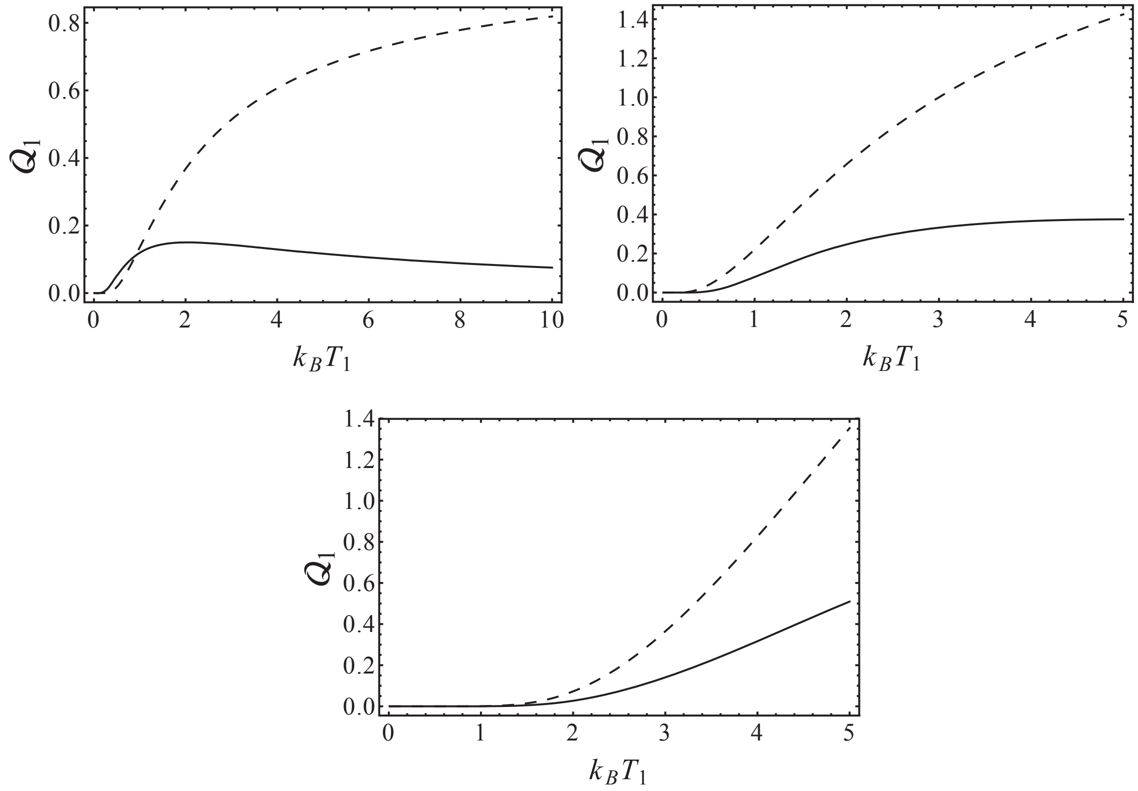

Figure 2.3: Dimer heat flux $\mathcal{Q}_1$ as a function of the temperature of reservoir 1 in both approximaions. The solid lines represent the local approach while the dashed lines represent the global approach. The steady state heat flux is calculated for three different values of the qubit energy gap ϵ: Left top $\epsilon = 1.001K$, Right top $\epsilon = 2.5K$, Bottom $\epsilon = 10K$. In all the cases we have set the temperature of the reservoir 2 as zero and $K = 1$.

The dimer system thus provides the simplest non-trivial physical scenario where a quantitative comparison of quantum non-equilibrium thermal quantities can be performed: those obtained within a rigorous global approach and results coming from a less rigorous, and hence restricted in validity, local approach. Furthermore, our results as tested in simple systems show that some physical magnitudes calculated within the local framework could be misleading. This work suggests various future avenues of research, one of which concern the systematic analysis of the scaling of the divergence 'distance' between different approach results as the system size (complexity) and the separation from equilibrium increase.

2.3 Summary of chapter 2

We have shown that even in the symmetrical case, the problem of an open quantum system interacting with two reservoirs at different temperatures leads – within local and global approximations – to different results when the number of interacting subunits in the open system is greater than one. The formulations are equivalent and identical for the monomer. By contrast, the results for an interacting dimer open several urgent questions about the range of applicability of the underlying approximations and hence methods. In the global approach, the populations for both qubit are identical in the steady state. On the contrary, the local approach predicts that the population of the qubit interacting with the cold reservoir is always smaller than the population of the other qubit. Second, the local approach predicts a maximum in the heat flux as a function of the temperature gradient, followed by a gradual decay to zero as the thermal bias grows. By contrast, the global approach predicts a saturation of the flux as the bias increases. Finally, we note that the outcome of the local approach in the strong inter-qubit coupling limit concludes that the dimer can be modeled as a single qubit which resembles the properties of a classical system. Future work with larger numbers of qubits will elucidate the differences in these approaches, as will a careful comparison to future experiments which are able to distinguish between the divergence in their predictions.

Chapter 3

Generalities and modeling of bacterial photosynthesis

Photosynthesis, the remarkable process in which light energy is transformed into chemical energy, is responsible for most of the metabolic processes underlying Life on Earth [26; 31; 43; 105; 45; 106; 107; 108; 109; 34; 44; 110]. Photosynthetic organisms are conventionally grouped into two classes: oxygenic and anoxygenic organisms. Higher plants, algae and cyanobacteria are characterized to use oxygenic photosynthesis which produces oxygen through the reduction of carbon dioxide and the oxidation of water molecules[25]. On the other hand, anoxygenic photosynthesis does not produce oxygen since water is not used for carbon dioxide reduction [31]. Despite of the many hypotheses about the origin of photosynthesis, there is little certainty about its earliest origin. Analyses of meteorites reveal a rich organic composition [111; 112; 113] which has led to speculation that some life forms could have resulted from the chemical development of organic material provided from outer space [114]. This raises the question as to whether Earth-like life forms might have developed elsewhere – and as a corollary, whether existing Earth-like lifeforms would survive near more alien light sources, where we define an ’alien’ light source as one with temporal correlations that are qualitatively different from those experienced on the Earth due to the Sun.

In this chapter, the reader will find an overview of the biological structure of photosynthetic purple bacteria together with the physical processes involved in bacterial photosynthesis. Findings from atomic force microscopy imaging connect the illumination conditions with structural aspects of the photosynthetic apparatus such as antennae stoichiometry and architecture. In order to quantify these remarkable findings, a detailed dynamical model of the photosynthetic process is introduced and explained. The model considers the processes underlying bacterial photosynthesis such as light absorption, excitation transfer, dissipation and reaction center cycling. A simple analytic approach is proposed in order to verify the accuracy of the model and it is shown that it captures some of the structural variations connected to the light intensity conditions. Finally, an anlysis on how variations on the statistics of the temporal photon arrival potentially influence the well functioning and hence survival of the photosynthetic organism, is presented.

3.1 Overview of bacterial photosynthesis

Bacterial photosynthesis is arguably the oldest form of photosynthetic life [115; 111]. Its structural simplicity, as compared with higher organisms like algae or plants[31], makes it an ideal focus for research. In particular, purple bacteria is known to sustain their metabolism in anaerobic environments such as deep waters (e.g. lagoons, streams and ponds), hence green and far red (¿750nm) components of sunlight spectra are used as its energy source. There are several studies on bacterial photosynthesis that go into a deep description of the structure, organization and dynamics of the photosynthetic apparatus [31; 25]. The building blocks of the photosynthetic apparatus are the chlorophyll pigments which are present in all light driven organisms. They are responsible for the absorption,

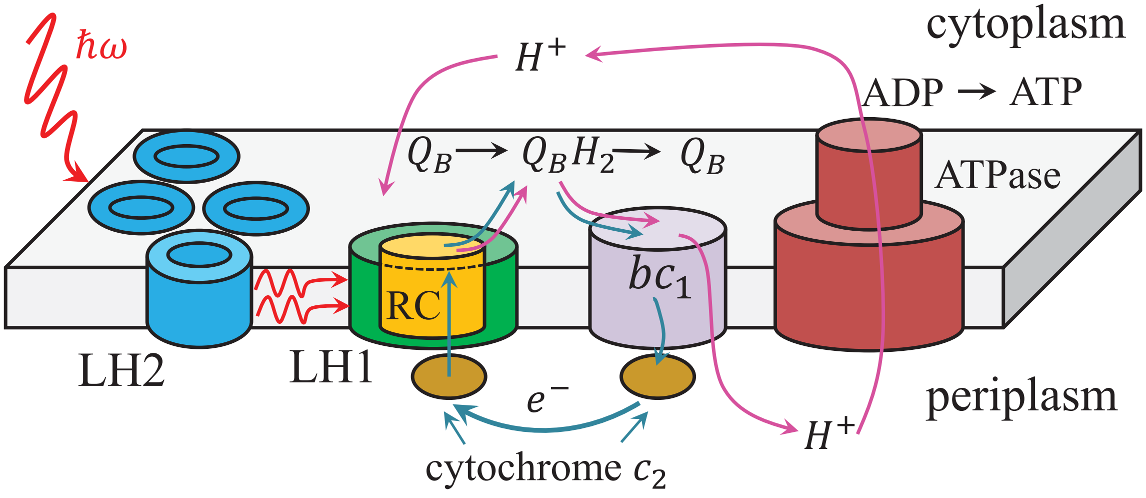

Figure 3.1: Schematic representation of photosynthetic apparatus of purple bacteria. The antennae complexes LH2 (blue ring) and LH1 (green ring) absorb incoming photons at a rate γ_A. They transfer the electronic excitation (red arrows) to the reaction center (yellow cylinder) where a series of charge separations produce quinol (Q_BH_2) from quinone with two cytoplasmatic photons (pink arrows). Quinol molecules migrate to bc_1 complex (purple cylinder) where it is reverted to quinone by a proton gradient that expels protons (pink arrows) to ATPase (Red cylinder) where ATP is synthesized. Meanwhile, remaining electrons and quinone at bc_1 are taken back to the reaction center through the cytochrome c_2 (green arrows) so the RC neutrality is reestablished and the cycle starts from the beginning.

transfer and early stages of light energy processing. In bacteria they are known as bacteriochlorophills (BChls) which are found throughout the chromatophore vesicle that lies in the bacteria cellular membrane.

The components of the chromatophore vesicle and the basic photosynthetic cycle are shown illustrative in Fig.3.1. Incoming photons with energy $\hbar\omega$ are absorbed by antennae complexes LH2 and LH1 creating an electronic excitation by pigment molecules such as BChls and Carotenoids (Car). The excitation is subsequently transfered to the reaction center (RC) where a charge separation process is initiated when charge carriers are available. The electron in the RC reduces a quinone (Q_B) which uses a cytoplasmic proton to form a hydrogen atom that is added to it. The resulting hole from the charge separation is filled with a periplasmic electron provided by a protein called cytochorme c_2 reestablishing the neutrality. A second excitation is therefore received at the RC to repeat the process forming quinol ($Q_B H_2$) with the reception of the second hydrogen atom. Subsequently, quinol detaches from the RC and diffuses to a protein complex called cytochrome bc_1. The quinol is reverted to quinone at bc_1 where the electrons are later moved back to the RC through c_2 to fill the holes resulting from the charge separation. The resulting proton gradient is used by the protein complex ATP synthase (ATPase) in the last stage of the process, to phosphorilate adenosine diphosphate (ADP) to adenosine triphophate (ATP) which is a very stable form of energy used for all living organisms.

3.1.1 Light harvesting complex 2

Pigments molecules, BChls and Car, absorb maximally in the far red and green, respectively. They are embedded in protein helices called α and β which form a ring-like structure known as Light Harvesting Complex (LHC). Most species of purple bacteria contain two types of LHC, named Light Harvesting 1 (LH1) and Light Harvesting 2 (LH2). LH2 complexes contains 9 pairs of BChls in the inner ring and 9 chromophores in the outer ring, that absorb maximally at 850nm (B850) and 800nm (B800) respectively (see Fig.3.2(a)), together with 9 Car that absorb maximally at 500nm. LH2 complexes are found largely $(100 - 400)$ on the chromathophore increasing the vesicle's absorption cross section. This is particularly crucial for purple bacteria, given that they live in environments where the light intensity is low (Intensity $\sim$1-10W/m^2). An excitation is easily transferred from Car to BChls within 100 femptoseconds (fs) given that energetically it is inexpensive [25; 116].

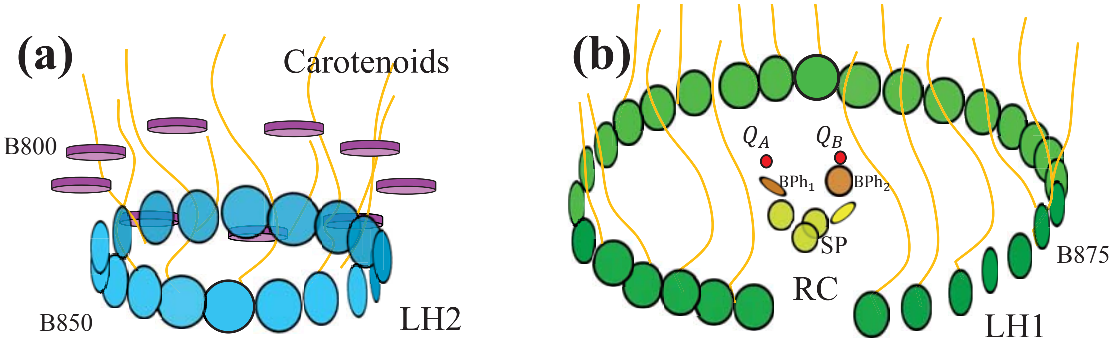

Figure 3.2: Pigment arrangement for (a) light harvesting complex 2 and (b) light harvesting complex 1 and reaction center core complex. Bactriochlorophyls are represented by colored circles while carotenoids by thick curved lines. Organic pigments B850 (blue circles) and B800 (purple circles), are embedded on a circular arrangement of 9 helix protein dimer α-β (not shown). Analogously for pigments B875 (green circles) with 16 dimers. Special pair dimer (yellow circles) at the reaction center is responsible for charge separation and release of an electron to the pigment bacteriopheophytin (brown circle) which transfer the electron to a quinone molecule (red circles). Pigment molecules at the reaction center are bound to protein subunits (not shown).

The typical fluorescence lifetime of light-driven molecules is 1 nanosecond (ns). Therefore the excitation needs to be transfered to neighboring complexes on smaller timescales in order to avoid exceptional loss.

3.1.2 Light harvesting 1 - Reaction center core complex

The core complex LH1-RC (Fig.3.2(b)) is responsible for receiving excitations from neighbor LH2, photon absorption as well as of initiate charge separation process. LH1 complexes contain 32 BChls arranged on 16 bi-chromophore units that absorb maximally at 875 nm (B875). Similar to LH2 complex, LH1 forms a ring-like structure but with a larger radius which surrounds the RC. Species like *Rhodospirillum (Rsp.) viridis* and *Rsp. photometricum* display monomeric core LH1-RC while species like *Rhodobacter (Rb.) sphaeroides* and *Rb. blasticus* dimeric, where the helix protein PufX induces its dimerization. The RC is composed by four BChls molecules, two bacteriophaeophytin molecules, and quinones A and B. Excitation from the LH1 is transferred to the RC that produce ionization of BChls dimer known as the special pair. An electron is transferred to the bacteriophaeophytin molecule which moves it to the quinone A and subsequently to quinone B. As mentioned above, a second electron repeats the cycle to create quinol which detaches from the RC and a new quinone molecules replaces it while electrons from cytochrome c_2 fill the holes from the charge separation, and the cycle repeats from the beginning.

3.2 Physical processes involved in bacterial photosynthesis

3.2.1 Photon absorption

Pigment molecules BChls and Car absorb incoming photons at a rate proportional to its cross section. The absorption rate for a specific complex can be calculated as the product of the spectral density of the incident light as a function of its wavelength $n_\gamma(\lambda)$ and the complex cross-section. Francke and Amezc [117] reported a cross-section for BClhs effectively measured in vivo of 2.32 Å^2 at the wavelength of maximum absorption[32]. Since each protein pair α-β holds 3 BClhs, and a specific number of protein pairs α-β are found on LHC complexes, the effective cross-section for the full ring σ_{LH1} and σ_{LH2} can be calculated. The photon spectrum normalized to the total light intensity of a representative environment for *Rb. sphaeroides* of 18W/m^2, is used to estimate the rate of absorption. For instance the absorption rate for LH1, that counts with 16 protein pair α-β, can be estimated as [32]

$$\gamma_1 = \int d\lambda n_\gamma(\lambda)\sigma_{LH1}(\lambda) = 18s^{-1} \tag{3.1}$$

Analogously for LH2 the absorption rate can be estimated as 10s^{-1}. If the spectrum is normalized to 1W/m^2, the absorption rate for LH1 and LH2 complexes is $\gamma_1 = 1$s^{-1} and $\gamma_2 = 0.55$s^{-1}. Therefore, for vesicles containing N_1 LH1 complexes and N_2 LH2 complexes at a light intensity of I, the total absorption rate (γ_A) can be calculated as [35]

$$\gamma_A = I(\gamma_1 N_1 + \gamma_2 N_2). \tag{3.2}$$

3.2.2 Excitation transfer

General theory

Excitation transfer between an excited donor and unexcited acceptor chromophores occurs via Coulomb interaction between corresponding electrons [25]. The wave functions associated with the initial and final states within the Born-Oppenheimer approximation can be expressed as a product of the electronic wave function and the vibrational wave function

$$\Psi_{D^*}\Psi_A = \psi_{D^*}\psi_A\chi(w_D^*)\chi(w_A) \tag{3.3}$$

$$\Psi_D\Psi_{A^*} = \psi_D\psi_{A^*}\chi(w_D)\chi(w_A^*), \tag{3.4}$$

where $w_{D(D^*)}$ and $w_{A(A^*)}$ describe the continuum of vibrational and bath states associated with the ground (excited) states for the donor and acceptor, respectively. For an electronic coupling V_{DA} so that the inter and intra molecular relaxation processes are faster than the excitation transfer, and using the Condon approximation [25], the transfer rate can be written as

$$k_{DA} = \frac{2\pi}{\hbar}|U_{DA}|^2 J_{DA} \tag{3.5}$$

where U_{DA} depends only on the electronic part while J_{DA} depends only on the vibrational states. The latter is defined as

$$J_{DA} = \int dE G_D(E) G_A(E) \tag{3.6}$$

where $G_D(E)$ and $G_A(E)$ are known as the Frank-Condon weighted and thermally averaged combined density of states [25]. The former is related with the Coulomb interaction as

$$U_{DA} = \langle \phi_{D^*}\phi_A | V_{DA} | \phi_D \phi_{A^*} \rangle. \tag{3.7}$$

The interaction energy in the donor-acceptor pair is formally given by

$$V_{DA} = \frac{1}{2} \sum_{\substack{m,n,p,q \\ \in I_D \cup I_A}} \sum_{\sigma,\sigma'} \langle \phi_m \phi_n | V | \phi_p \phi_q \rangle c^\dagger_{m\sigma} c^\dagger_{p\sigma'} c_{q\sigma'} c_{n\sigma} \tag{3.8}$$

where $c^\dagger_{m\sigma}$ and $c_{n\sigma}$ are the fermion creator and annihilation operators associated to electrons with spins σ and σ' in the mutually orthogonal basis of atomic orbitals ϕ_m and ϕ_n and $I_{A(D)}$ is the set of atomic orbitals associated to the donor (acceptor) molecule. The bracket $\langle \phi_m \phi_n | V | \phi_p \phi_q \rangle$ is the Coulomb integral given by

$$\langle \phi_m \phi_n | V | \phi_p \phi_q \rangle = \int\int d\vec{r}_1 d\vec{r}_2 \phi^*_m(\vec{r}_1) \phi_p(\vec{r}_2) \frac{e^2}{|\vec{r}_1 - \vec{r}_2|} \phi^*_n(\vec{r}_1) \phi_q(\vec{r}_2) \tag{3.9}$$

Intra-molecular contributions arise where m, n, p, q are either all donor or all acceptor states. On the other hand, intermolecular contributions occur through two types of mechanisms. If ϕ_m and ϕ_p are states from the donor, and ϕ_n and ϕ_q are states of the acceptor, the mechanism is direct Coulomb interaction. It is related to the de-excitation of an electron in the donor molecule D, which immediately excites an electron on the acceptor A. By contrast, if ϕ_m and ϕ_n describe states of the donor and ϕ_p and ϕ_q states of the acceptor, the mechanism is called exchange interaction. In this case, excitation is transfered when the excited electron of the donor is exchanged by a non-excited electron of the acceptor. Using anti-commutation relationships of fermions the interaction can be separated in two contributions for each transfer mechanism:

$$V_{DA} = \sum_{\substack{i,j \in I_D \\ \sigma}} \sum_{\substack{k,l \in I_A \\ \sigma}} \left[\langle \phi_i \phi_j | V | \phi_k \phi_l \rangle c^\dagger_{i\sigma} c_{j\sigma} c^\dagger_{k\sigma'} c_{l\sigma'} - \langle \phi_i \phi_l | V | \phi_k \phi_j \rangle c^\dagger_{i\sigma} c_{j\sigma} c^\dagger_{k\sigma'} c_{l\sigma'} \right] \tag{3.10}$$

Inter complex transfer rates

An effective Hamiltonian description agrees with the spectrum of B850 where individual BChl Q_y excitations are taken into account [25]. The state $|j\rangle$ describes the $2N$ BChl system where the j-th molecule is excited, while the rest are in the ground state. This formulation allows the construction of the completely delocalized exciton state $|\tilde{n}\rangle$ as a linear combination of individual excitations $|j\rangle$:

$$|\tilde{n}\rangle = \sum_{j=1}^{2N} a_{\tilde{n},j}|j\rangle \tag{3.11}$$

The effective Hamiltonian presents diagonal elements $\langle j|H|j\rangle = \epsilon$ corresponding to the excitation energy of the Q_y state. The nearest neighbors matrix elements $\langle j|h|j+1\rangle$ for pigments within the same α-β protein dimer is denoted as ν_1 and for neighbors of different dimers ν_2. For non-nearest neighbors pigments j and k, a dipole-dipole approximation is used

$$\langle j|H|k\rangle = C\left(\frac{\vec{d}_j \cdot \vec{d}_k}{r_{jk}^3} - \frac{3(\vec{r}_{jk} \cdot \vec{d}j)(\vec{r}_{jk} \cdot \vec{d}k)}{r_{jk}^5}\right). \tag{3.12}$$

The parameters that reproduce the exact spectrum for B850 hexadecamer of LH2 from *Rs. molischianum* are $\epsilon = 13242\text{cm}^{-1}$, $\nu_1 = 790\text{cm}^{-1}$, $\nu_2 = 369\text{cm}^{-1}$ and $C = 505644\text{Å}^3\text{cm}^{-1}$ [25]. Within the time-scale of inter-complex energy transfer (a few picoseconds), exciton states thermalize. Therefore they become a statistical ensemble characterized by the density operator

$$\rho = \frac{1}{\sum_{\tilde{m}} e^{-E_{\tilde{m}}/k_B T}} \sum_{\tilde{m}} e^{-E_{\tilde{m}}/k_B T} |\tilde{m}\rangle\langle\tilde{m}|. \tag{3.13}$$

from / to	LH1	LH2	RC
LH1	20ps	15.5ps	15.8ps (30-50 ps)
LH2	7.7ps (3.3ps)	10ps	-
RC	8.1ps (8ps)	-	-

Table 3.1: Theoretical estimation and experimental measurements (in brackets) of inter complex transfer times. Dashes indicate that the value is not available.

Hence, the net inter-complex interaction can be calculated to be

$$\begin{aligned} U_{DA} &= \mathrm{Tr}\{\rho V_{DA}\} \\ &= \frac{1}{\sum_{\tilde{m}} e^{-E_{\tilde{m}}/k_B T}} \sum_{\tilde{n},\tilde{m}} e^{-E_{\tilde{m}}/k_B T} \langle \tilde{m}|V_{DA}|\tilde{n}\rangle. \end{aligned} \tag{3.14}$$

Finally, the matrix elements $\langle \tilde{m}|V_{DA}|\tilde{n}\rangle$ associated with interactions of exciton states from different complexes can be calculated from the individual chromophore inter-complex interaction as

$$\langle \tilde{m}|V_{DA}|\tilde{n}\rangle = \sum_{i,j} a_{n,i} a^*_{m,j} \langle i|H|j\rangle. \tag{3.15}$$

Hence, equations (3.5), (3.15) and (3.14) allow to estimate transfer times from different complexes. Table 3.1 shows estimated transfer times which show to be in good agreement with pump probe spectroscopy experiments (in brackets) [118; 33; 119; 120].

3.2.3 Dissipation

In chromophores there are two main mechanisms of dissipation: fluorescence and internal conversion. As mentioned above, fluorescence is presented on a time scale of hundred of picoseconds to a few nanoseconds and is the result of spontaneous emission due to interaction with the electromagnetic vacuum [121]. Internal conversion is the decay of the electronic excited state to the ground stated due to overlap with the molecule's vibrational

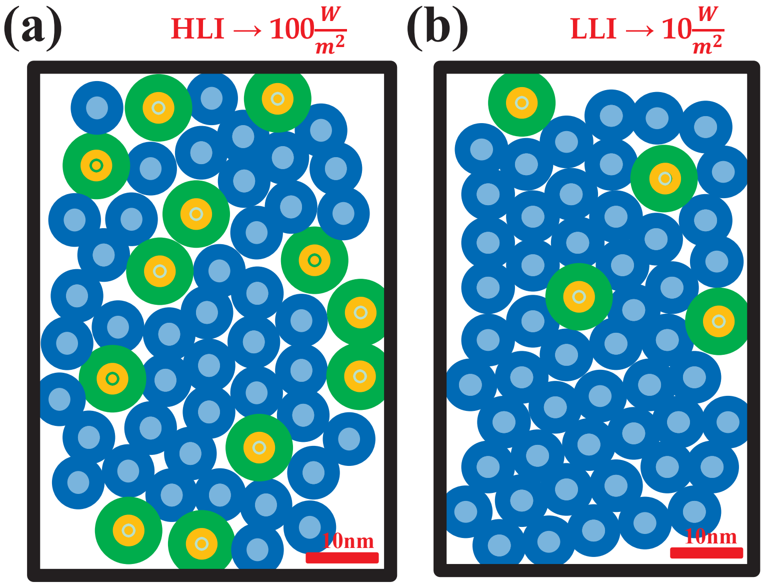

Figure 3.3: Representation of real-world membranes from AFM images grown at different light intensity conditions. (a) High Light Intensity (100W/m^2) and (b) Low Ligh Intensity (10Wm^2). Blue rings represent LH2 complexes while green rings represent LH1 complexes. Yellow dots represent the reaction center.

degrees of freedom. Calculations have estimate a dissipation rate of 1ns^{-1} including both mechanisms [25].

3.3 Evidence from atomic force microscopy

During the last decade, experimentalists have been devoted considerable effort to study the membrane structure through Atomic Force Microscopy (AFM) imaging, revealing a rich organization of the photosynthetic apparatus. Two remarkable findings are of particular importance for this document. The first one is related to the differences found in the ratio of LH2 to core complexes (i.e. stoichiometry, s_0) when membrane are grown under different light intensity conditions. The second is related to the organization of the antennae complexes within the membrane where several architectures are found for different species and growth environmental conditions.

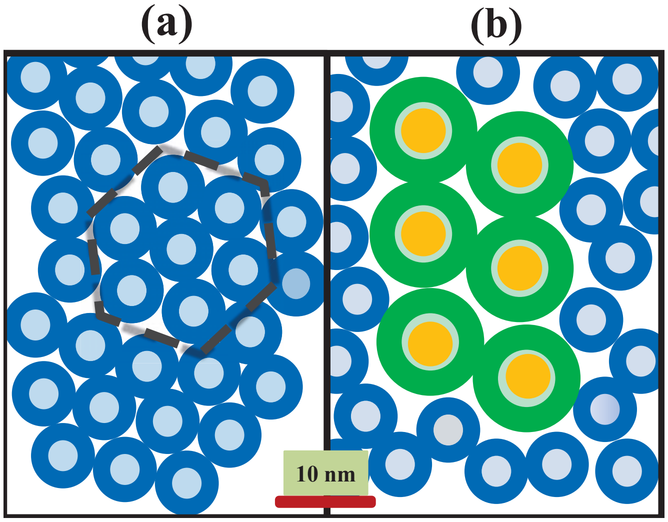

Figure 3.4: Representation of real-world membranes from AFM images.(a) Low-light adapted native membrane of *Rsp. photometricum* displaying LH2 hexagonal packing from [124] (b) *Rb. sphaeroides* from [125] showing high core-core clustering.

3.3.1 Chromatic adaptation

In 2005, it was reported for *Rsp. Photometricum* an interesting relationship between the membrane's stoichiometry and the light intensity during the growing stage [26]. Under high-light intensity (HLI) ($I \approx 100$W/m^2), membranes grow with a ratio of LH2 to core complex of $s_0 \approx 3.5$-4, while membranes under low-light intensity (LLI) conditions ($I \approx 10$W/m^2) express a ratio of $s_0 \approx 7$-9. Fig.4.2(a) and Fig.4.2(b) show a representation of real membranes for HLI and LLI, respectively. Further experiments have confirmed this link in other types of bacterial systems such as *Rps. Palustris* [122] and *Rb. sphaeroides* [123]. This chromatic adaptation is of particular importance to the well functioning and survival of the organism. It optimizes the capture of photons under low-light conditions by expressing a large amount of LH2 antennae complexes, while preventing damaging effects due to over exposure for high-light intensity conditions.

3.3.2 Antennae organization

Additional details regarding the organization of antennae complexes are shown in Figure 3.4. In reduced light intensity conditions, it is found that LH2-only regions are formed on a hexagonally packed domain (Fig.3.4(a)). This is considered to be a response arising from the chromatic adaptation to low light [124]. A particular case of honeycomb crystal architectures of identical harvest found in *R. rubrum* is reported in Ref.[126]. Variations from a perfect hexagonal configuration are found among species and sample preparation methods. In addition, in species like *Rb. blasitcus* and *Rb. sphaeroides*, where dimerization of core complexes is found[127], are observed to exhibit a linear alignment of core-core dimers surrounded by LH2 antennae (Fig.3.4(b)). Interestingly, membranes displaying core-core clustering are also found on species that do not display dimerization, that is the case of high-light adapted *Rps. photometricum*[44]. Qualitative explanations are proposed to support these findings. For instance, it is expected that core-core clustering may assist in the transfer and processing of excitations to the RC by reducing the effective path and hence avoiding dissipation [128]. In chapter 4 it will be shown that this statement is quantitatively correct and spatial correlations play a fundamental role in the degree of enhancement of the photosynthetic efficiency.

3.4 Photosynthetic membrane model

3.4.1 Stochastic approach

The proposed model uses a stochastic approach to the classical rate equations for a large number of LHC ($\approx$400). The advantage provided by a stochastic method is the level of

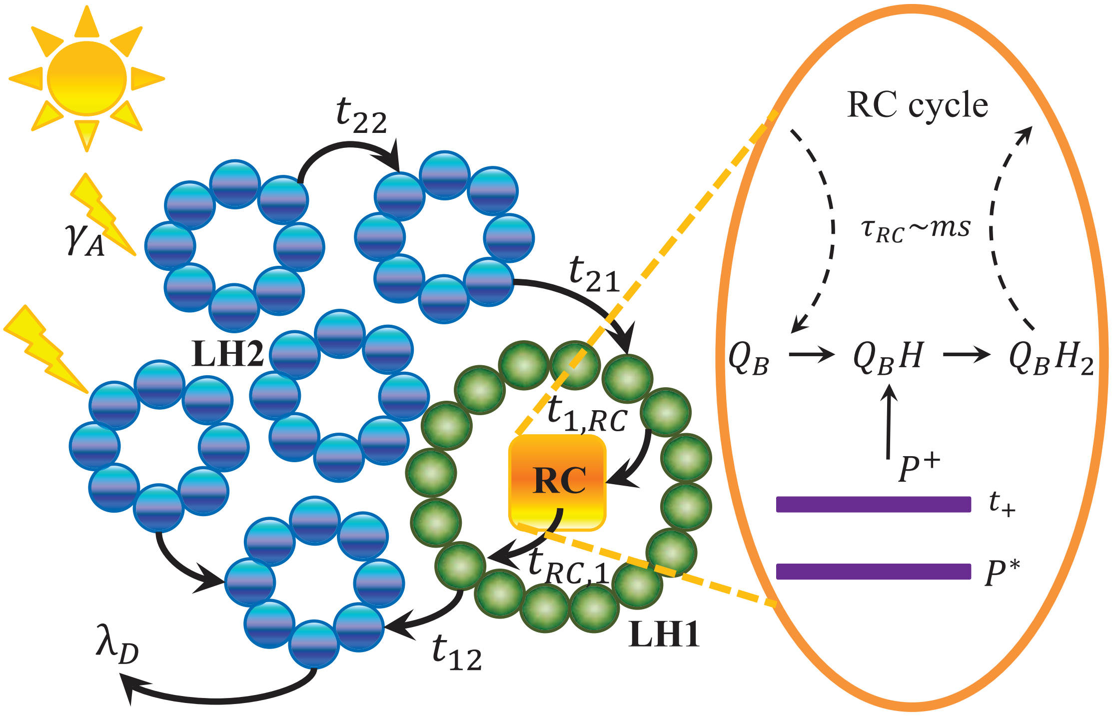

Figure 3.5: Schematic representation of photosynthetic process. Radiation from a thermal source (e.g. our Sun) is absorbed by LHC complexes 1 (blue rings) or 2 (green rings) at a combined rate γ_A. The resulting excitation is transfered from complex i to complex j at a specific rate t_{ij} or it is dissipated at a rate λ_D. At the reaction center, the double excitation of the special pair P yields to the formation of quinol (Q_BH_2) molecule in a cycle that lasts a few milliseconds.

detail that can be achieved despite of the large number of complexes typically present on a photosynthetic vesicle. Accounting for two states (occupied and unoccupied) for each complex, the dimension of the matrix to be solved grows as $2^N \times 2^N$, being N the total number of complexes. Therefore, a master equation approach is not practical for a realistic vesicle. Moreover, a stochastic description provides detailed time series of every step of the process that become indicators to further investigate the system.

Figure 3.5 shows the schematics of the model. It accounts for photon absorption, photoexcitation transfer and RC cycling for a given architecture and light statistics. At each time-step ($\delta t \approx 0.025$ps), incoming photons following a specific arrival statistic, are being absorbed by the antenna complexes LH1 and LH2 with absorption rate $\gamma_A = I(\gamma_1 N_1 + \gamma_2 N_2)$, where $\gamma_{1(2)}$and $N_{1(2)}$ are the absorption rates [32] and number of LH1(2) complexes, respectively. This means that on a timestep corresponding to absorption, a random complex of type i is chosen to absorb the photon with probability

$$p_i = \frac{\sum_k \sigma_k^{LHi}}{\sigma_T}, \quad i = 1, 2, \tag{3.16}$$

where σ_T is the total cross-section of the membrane. At the same time, present photoexcitations diffuse throughout the membrane in search for an open RC according to transfer rates resulting from experiments [118; 33; 119; 120]. Some of these excitations will be dissipated through fluorescence or internal conversion at constant rate λ_D [32]. The excitation transfer from the complex i to any of the $\{j\}$ possibilities (transfer to other neighboring complex or dissipation) occur with a probability

$$P_{i \to \{j\}} = \delta t \sum_j \gamma_{ij} \tag{3.17}$$

The exact acceptor j is determined if a generated pseudo-random number a lies in the interval

$$\frac{\gamma_{ij}\delta t}{P_{i\to\{j\}}} \leq a < \frac{\gamma_{i(j+1)}\delta t}{P_{i\to\{j\}}}, \tag{3.18}$$

and the sum is performed including the dissipation path of constant rate λ_D. Meanwhile, closed RCs are processing the excitations that have already been received. Once an open RC has received two photo-excitations, it is set closed and no other photo-excitation is allowed to enter. After a time τ_{RC} has elapsed from the moment in which the second photo-excitation has entered, the RC is set open and the cycle starts from the beginning. This open/close mechanism accounts for the time where two electrons produce quinol ($Q_B H_2$) before it undocks and a new quinone (Q_B) substitutes it. This process lasts a few milliseconds and has shown to be key in order to explain the chromatic adaptation of the membrane stoichiometry (ratio of LH2 to LH1) under different light intensity conditions [35; 36].

Transfer rate measures from pump-probe experiments agree with generalized Förster calculated rates [25], assuming intra-complex delocalization. LH2→LH2 transfer has been given as $t_{22} = 10$ps [25], while LH2→ LH1 transfer has been measured for *R. Sphaeroides* as $t_{21} = 3.3$ps [118]. Back-transfer LH1→ LH2 is approximately $t_{12} = 15.5$ps while the LH1→ LH1 mean transfer time t_{11} has been calculated using a generalized Förster interaction [106] as 20 ps. Second and third lowest exciton lying states cause LH1→ RC transfer due to ring symmetry breaking [33], consistent with a transfer time of $35 - 37$ ps found experimentally at 77 K [120; 119]. As proposed by Ref. [129], increased spectral overlap at room temperature improves the transfer time to $t_{1,RC} = 25$ ps. The back-transfer from an RC's fully populated lowest exciton state to higher-lying LH1 states occurs in a calculated time of $t_{RC,1} = 8.1$ ps [33], which is close to the experimentally measured 7-9 ps estimated from decay kinetics after RC excitation [130]. The subsequent passage through the RC complex depends on whether a charge carriers is available (i.e. the RC is in an open state), to occur within $t_+ = 3$ps.

3.4.2 Analytic approach

An analytical description of the model is used to verify the validity of its results. The analytic model considers N_E photo-excitations created at the membrane at a rate γ_A. In addition, photo-excitations leave the system by either dissipation at rate λ_D or by entering into a RC at rate λ_C. On the other hand, the number of closed RCs per unit time is $\lambda_C N_E/2$ and the number of open RCs per time unit is $(N_1 - N_O)/\tau_{RC}$. Therefore the number of photo-excitations and the number of open RCs is connected via the following pair of differential equations:

$$\frac{dN_E}{dt} = -\left[\lambda_C(N_O) + \lambda_D\right] N_E + \gamma_A \tag{3.19}$$

$$\frac{dN_O}{dt} = \frac{1}{\tau_{RC}}\left(N_1 - N_O\right) - \frac{1}{2}\lambda_C(N_O)N_E \tag{3.20}$$

The rate of entering to a RC λ_C depends on the number of open RCs N_O. When there is no RC available $\lambda_C = 0$, and it takes the maximum value (λ_C^0) when all RCs are open. We can assume a linear relation between λ_C and N_O to be a good approximation due to the fast photo-excitation hopping compared to the cycling time, i.e., $\lambda_C \approx \lambda_C^0 N_O/N_1$. Moreover, the efficiency is given by $\eta = \lambda_C N_E/\gamma_A$, the steady state solution to Eqs. (1) and (2) is

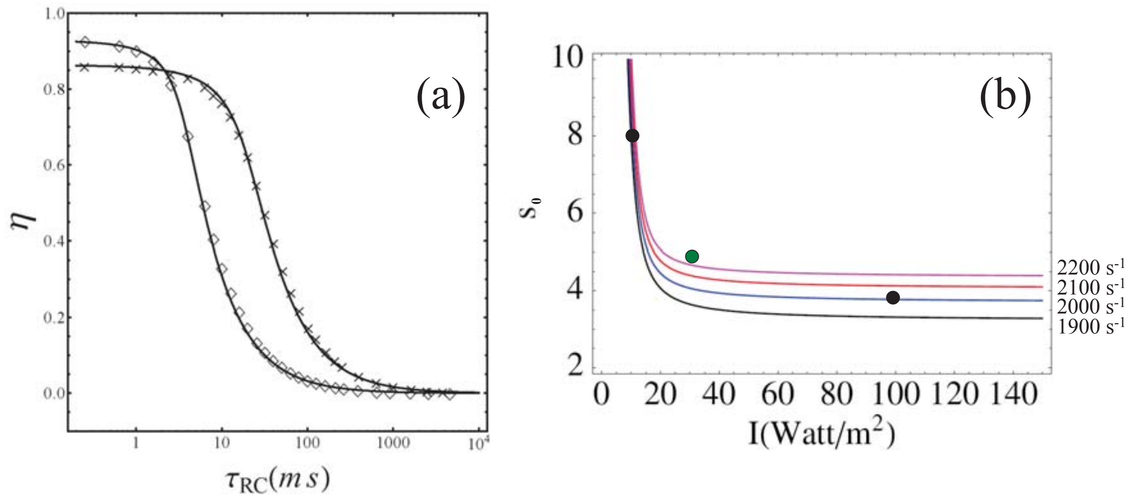

Figure 3.6: (a) Photosynthetic efficiency from stochastic simulation (symbols) and analytic approximation (lines) as a function of the RC closing time τ_{RC}. Diamonds (and corresponding line) are calculated for light intensity of 100W/m^2 while crosses (and line) are for light intensity of 10W/m^2. Model prediction (curves) for stoichiometry s_0 as a function of light intensity for several values of quinol production rate W [35; 36]. Dots are experimental results for reference [26] (black) and [131] (green).

[35]:

$$\begin{aligned}\eta &= \frac{1}{2\gamma_A\lambda_C^0\tau}\Big\{2N_1(\lambda_C^0+\gamma_D)+\gamma_A\lambda_C^0\tau \\ &\quad -\sqrt{4n_1^2(\lambda_C^0+\gamma_D)^2+4N_1\gamma_A\lambda_C^0(\gamma_D-\lambda_C^0)\tau+(\gamma_A\lambda_C^0\tau)^2}\Big\}\end{aligned} \tag{3.21}$$

Figure 3.6(a) compares results from the analytic model and numerical simulation. It illustrates the photosynthetic efficiency as a function of the closure time τ_{RC} for high and low light intensity conditions. The agreement between the formulations is noteworthy. As expected, for short cycling times ($\tau_{RC} < 5$ms) membranes grown on high light intensity experience a larger efficiency than LLIM given the larger number of RC's. However, for larger cycling times, the situation is overturned. For $\tau_{RC} > 5$ms, LLIM experience a larger efficiency regardless of their reduced number of RC's. This behavior talks about the mechanism that better benefits the organism for a given light intensity environment. For high light intensity, the system exploits the dissipation mechanism in order to avoid overexposure and potential damage, which results in a lower photosynthetic efficiency. On the other hand, for low light intensity environments, the system develops more LH2 antennae in order to capture most of the incoming light and hence increasing overall efficiency.

The good agreement is used to explore additional implications related to the chromatic adaptation. Under the hypothesis that adaptations occur so that the system meet metabolic needs, an expression for the quinol production rate W is derived in terms of the stoichiometry and the light intensity. The result is: [35; 36]

$$\begin{aligned} 2W(s_0, I) &= \frac{\gamma_A(s_0, I)}{2} + \frac{1}{B(s_0)}\left(1 + \frac{\lambda_D}{\lambda_c^0}\right) \\ &+ \sqrt{\left(\frac{\gamma_A(s_0, I)}{2} + \frac{1}{B(s_0)}\left(1 + \frac{\lambda_D}{\lambda_c^0}\right)\right)^2 + \frac{\gamma_A(s_0, I)}{2B(s_0)}} \end{aligned} \tag{3.22}$$

where $B(s_0) = \tau_{RC}(s_0)(A_1 + s_0 A_2)$. Figure 3.6(b) compares contour curves of W as a function of light intensity and stoichiometry with real-world membranes grown under different light intensities and displaying different ratios of LH2 to LH1 complexes [26; 131]. The agreement is remarkable and predicts that the vesicle could express extreme values of stoichiometry for very low light intensity conditions.

3.5 Implications of photon arrival statistics

The validity of the model allows us to investigate potential implication of statistical variables on the overall dynamics of the system. Let us consider the statistical properties of photon arrival to the photosynthetic membrane (input) together with those of the quinol production in the RC (output). Following Ref. [37; 132] the focus is on two statistical properties to characterize the output of quinol production $\phi(t)$ from the reaction centers. The specific focus is the burstiness B and memory M of this time-series $\gamma(t)$ [133]. A

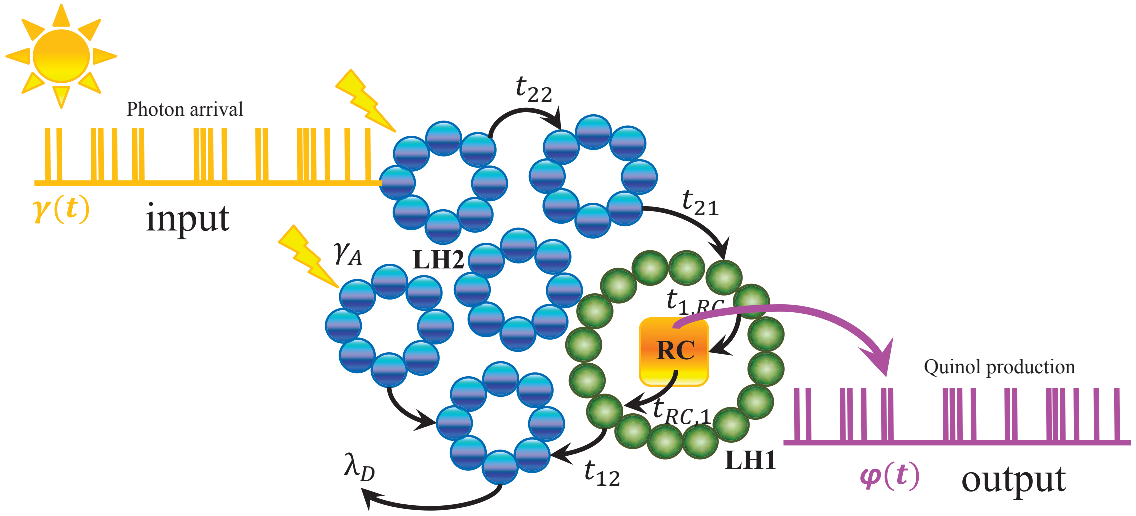

Figure 3.7: Schematic diagram showing the process of photosynthetic harvesting of incident light (e.g. sunlight or other alien light source) which is a priori temporally correlated. Our results analyze how these correlations in the incident light affect the ultrafast physics up to, but not including, the slow chemical processes following reaction center conversion. Excitations (excitons) are produced ($\gamma(t)$) and migrate across a large network containing light-harvesting complexes. Excitations are then processed at reaction centers (RC) producing quinol output ($\phi(t)$) for chemical metabolism.

Poisson process corresponds to $(M, B) \approx (0, 0)$ which is the approximate result expected for photosynthetic organisms which live (and hence survive) on Earth. Values of B and M that deviate significantly from $(0, 0)$ correspond to a quinol production that is very different from that experienced by the organism on Earth, and hence would likely kill it [37]. Recently, there have been observations of extreme value statistics in phenomena such as optical Rogue waves and coherent anti-Stokes Raman scattering in silicon [134]. These sources emerge from physical processes allowed throughout the universe and they are not restricted to a BM value close to zero, hence they open new roads of research.

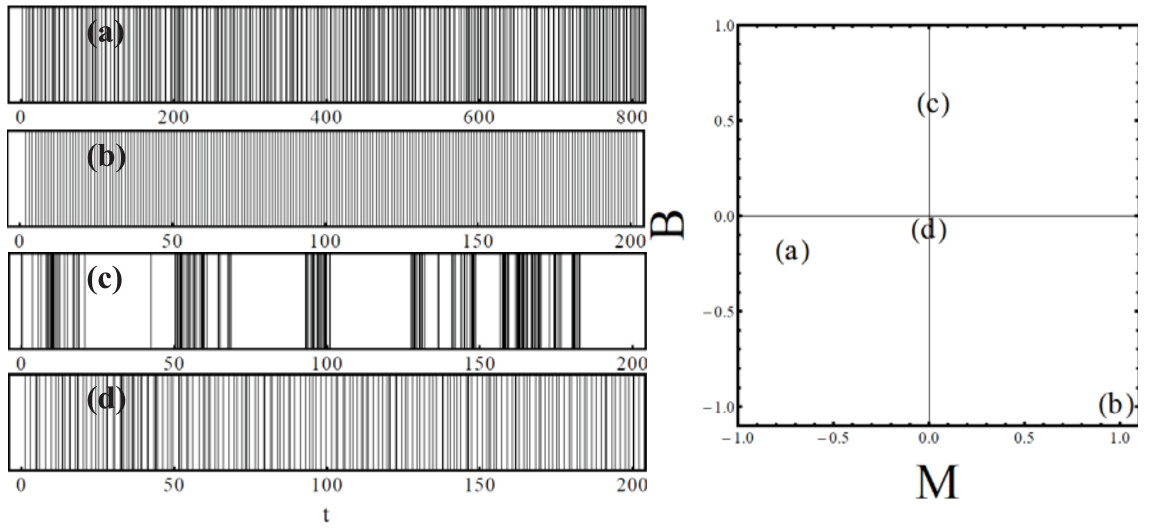

Figure 3.8: Time series characterized by different BM measures. (a) Long inter-event times are followed by short and vice-versa. (b) Periodic series. (c) Two state model. In active state the system generates events at a higher rate than the inactive state. The system changes from active to inactive randomly. (d) Poisson process.

3.5.1 Burstiness and memory

The temporal characteristics of both the input photon arrival and the quinol output are quantified using the two statistical measures B and M, which were introduced in Ref. [133] and employed in references [37; 132]. A random process is characterized by a time independent probability for events (e.g. photon absorption) to occur. The inter-event time Δt between consecutive events obeys a Poisson distribution. The burstiness B measures how far a distribution is from that emerging out of a Poisson process, i.e.,

$$B \equiv \frac{(\sigma_{\Delta t} - m_{\Delta t})}{(\sigma_{\Delta t} + m_{\Delta t})} \tag{3.23}$$

where $m_{\Delta t}$ and $\sigma_{\Delta t}$ are the mean and standard deviation of the inter event time series, respectively. For a Poisson process the mean is equal to the standard deviation, therefore a Poisson process presents zero burstiness B. Periodic signals present small standard deviation that yields to a burstiness value of -1. The memory M between consecutive intervals is defined as:

$$M \equiv \frac{1}{(n_{\Delta t}-1)} \sum_{i=1}^{n_t-1} \frac{(\Delta t_i - m_1)(\Delta t_{i+1} - m_2)}{\sigma_1 \sigma_2}, \tag{3.24}$$

where $n_{\Delta t}$ is the number of intervals and $m_{1(2)}$ and $\sigma_{1(2)}$ are the mean and standard deviation of the $\Delta t_i(\Delta t_{i+1})$'s respectively $(i = 1, ..., n_{\Delta t-1})$ [133]. As an illustration, Fig.3.8 illustrates four different artificial time series that yields to different values of B and M measures. Series (a) is constructed so that a short inter-event time is followed by a long one and subsequently followed by a short one again. This series generates both B and M measures negative. Series (b) is periodic, therefore presents M near to one while B close to -1. Series (c) corresponds to a two state model proposed on Ref.[133] where the ground state excecutes independent events with long inter-event times while the excited state performs correlated events with higher frequency. The system changes from ground (excited) state to excited (ground) state with an independent (correlated) probability. The resulting series is therefore bursty (positive B) but memoryless. Finally, series (d) represents a Poisson process whose values of B and M lie near to zero.

3.5.2 Potential survivability under extreme light

Although some inputs may have values of B and M that are very different from zero, and hence very unlike the experience on Earth, the quinol production output may end up lying very close to $(B, M) \approx (0, 0)$. This arises from the non-linearity of the membrane processing of excitations (Fig. 3.7) which signals potential survivability of the corresponding organism under these extreme alien light conditions.

Figure 3.9 shows an updated version of our preliminary results from Ref. [37]. It identifies the regions in the $B - M$ space that fulfill the requirement of survivability. As an

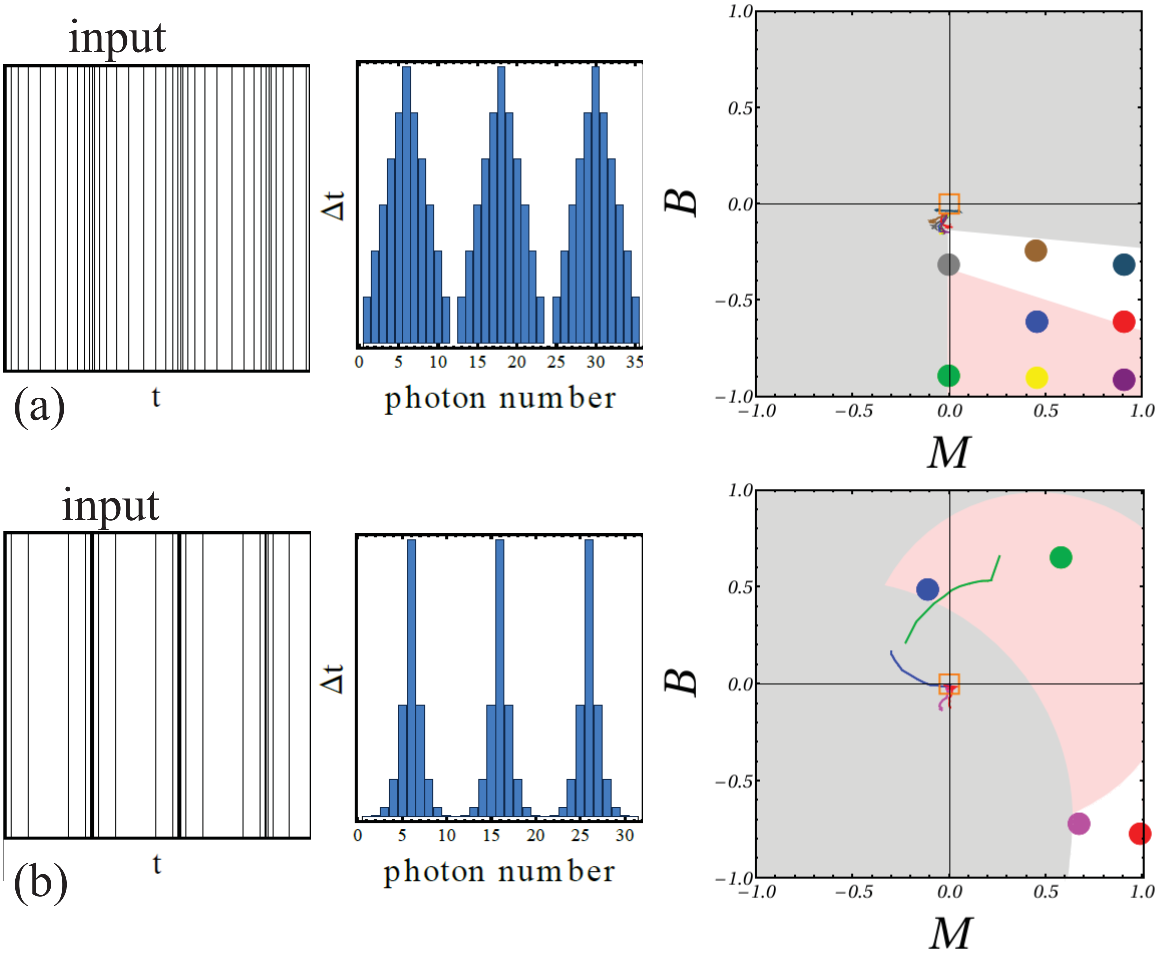

Figure 3.9: BM measurements for two types of photon arrival time series: bunched (a) and power law (b). The large colored circles are the BM values for the input while the trajectories represent the quinol production output for different values of RC closed time τ_{RC}. The gray regions are forbidden for the specific input time series. The red region shows the values where the output for the organism is bursty and likely to be too toxic. The white region produces output trajectories close to zero within the margin of 0.05 (orange square), and therefore correspond to survival in principle, i.e. these are the regions of survivability.

illustration, we have chosen the bunched (a) and power law (b) input for the photons being absorbed (see Fig.3.7). The circles represent $B - M$ values for the incident photon absorption time series (input), while the trajectory of the same color illustrates the quinol production output for different values of closed time τ. We assume that all (or a constant fraction) of the arriving photons get absorbed and hence that the time series of absorbed photons is statistically equivalent to that of the incident light. The gray region represents the forbidden values of B and M for this type of time series. The red region illustrates the region in the $B - M$ plane where the output is bursty and therefore potentially damaging for the metabolism of the photosynthetic organism. The white region shows the region where the quinol production output resembles the one on Earth and therefore the survival of the bacteria. The time series are generated in such a way that they preserve the average intensity regardless of their inter-event statistics, hence preserving the temperature of the system.

Our results in Fig.3.9 confirm that even though certain inputs of incident light may be very unlike that experienced by photosynthetic bacteria on Earth, their metabolic quinol production output may end up lying very close to that required for survival. The reason lies in the non-linearity of the membrane processing of excitations (Fig. 3.7). This means that they could potentially survive under these extreme alien light conditions. We hope our results encourage the experimental study of bacteria under such unusual light sources as characterized by their temporal correlations.

3.6 Summary of chapter 3

In this chapter the problem of light harvesting in bacterial organisms has been introduced. An overview of the photosynthetic process was presented where details about the structure, functionality and environmental effects were explained. Subsequently, a detailed microscopic model for bacterial photosynthesis was bestowed. It accounts for light absorption, excitation transfer, dissipation and RC cycling for large architectures and arbitrary light intensity and statistics. The model is able to explain the chromatic adaptation of photosynthetic vesicles under different light intensity conditions as an interplay of excitation transfer and RC cycling. The functionality of the model allows to exploration of potential implications of abnormal photon arrival statistics. Using the statistical measures of burstiness and memory, extreme light sources different from our Sun, showed not to be harmful for the photosynthetic organism. This opens the possibility for bacterial life to survive in habitats with illumination conditions very different from Earth's.

Chapter 4

Thermal photon spatial coherence favors photosynthetic life

The nature and possible uses of the quantum mechanical correlations within light, continue to generate heated debates. While the photoelectric effect establishes that light transfers energy to materials through the capture of individual energy packets (photons), experiments probing Bell inequalities and Hanbury-Brown-Twiss effects [135; 136] show that photons can also exhibit highly non-trivial spatial correlations – even within natural sunlight[137; 138; 139]. A fundamental open question concerns the extent to which Nature might exploit such correlations[140; 141]. The study of coherent effects in biologically relevant systems has attracted increasing attention due to the observation of long-lasting oscillatory signals measured with time-resolved techniques [46; 47; 48; 142; 50; 49]. This has been followed by a wave of theoretical work concentrated on these systems' exciton and vibrational properties, [143; 144; 145], and recently on the role of electronic coherence in the process of light absorption [146].

Spatial correlations of thermal light have attracted attention for their nontrivial consequences such as sub-wavelength interference [139], ghost imaging [147] and photon-number correlations [147; 148]. In this chapter we explore the potential implications that spatial correlations may have in bacterial photosynthesis and the way these organisms might have exploited them. With this in mind, it is considered the simplest possible setup for a photosynthetic system: sunlight incident on N antennae complexes that can absorb a photon and transfer the energy to neighboring complexes. Sunlight is described as thermal light and on occasions it be could referred to as radiation from a blackbody emitter [149]. The first two sections are devoted to the basic theory of optical coherence where the main definitions and general formalisms are presented and limiting cases described. Subsequently, the formalism is applied to primitive clusters of molecular antennae where it is found that specific motif architectures, (similar to the ones found on real-world membrane), profit from spatial coherence by maximizing the probability of photon absorption. Further analysis on arbitrary detection time demonstrates that the findings are robust for times which are comparable and even larger than the coherence time associated to the incoming radiation. Finally, the full photosynthetic model introduced in the previous chapter is implemented to account for spatial correlations on absorbing photons. It is found that the clustering of core complexes LH1-RC provides greater sensibility to spatial correlations than disordered membranes which increases the overall photosynthetic efficiency when compared with non correlated light.

4.1 Generalities and definitions in optical coherence theory

The theory of optical coherence is formulated to describe the properties of fluctuating electromagnetic fields [149; 121]. Optical coherence phenomena are manifestations of the correlation between these fluctuations. Every electromagnetic field in nature presents fluctuations; being phenomena of interference the simplest manifestation of correlations in

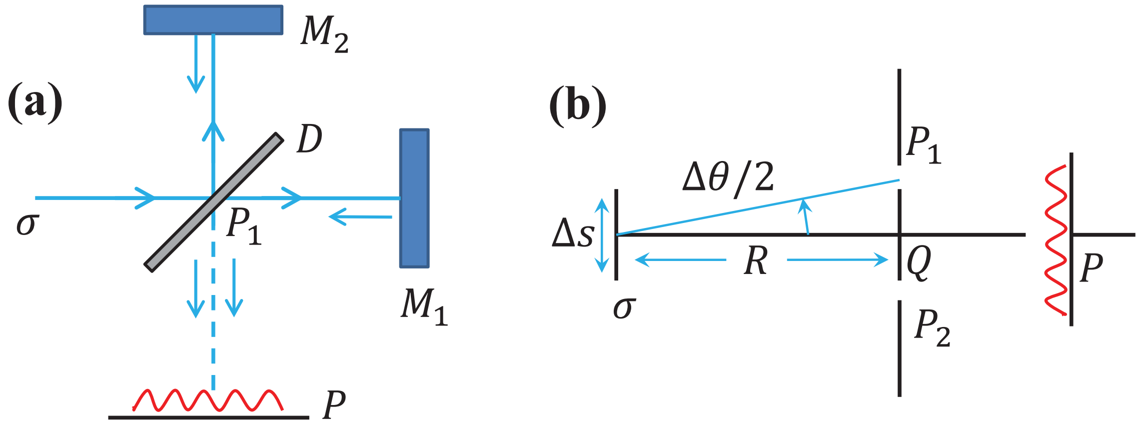

Figure 4.1: (a) Michelson interferometer simplified setup. The source is identified as σ, D is the beam divider, M_1 and M_2 are mirrors and P is the observation screen. (b) Young's interference experiment. A source σ with dimension Δs. A plane Q located at a distance R from the source and has two pinholes P_1 and P_2. If the pinholes are close enough to the symmetry axis, fringes will be formed at the screen P.

fluctuating optical fields. For a thermal field, fluctuations originated from a beam at some space-time coordinate is the result of a large number of contributions which are independent from each other, and their superposition can be described on statistical terms. Even well established lasers present some degree of random fluctuation, given the presence of spontaneous emission phenomena. Although optical fluctuations are not evident, appropriate experiments are able to determine their existence and potential involvement on different optical phenomena [150] (e.g. light absorption on photosynthesis).

4.1.1 Coherence time and area of coherence

Figure 4.1(a) shows a Michelson interferometer where a light beam from a quasi monochromatic source sigma is divided into two beams by a beam splitter denoted as D. The beam is said to be quasi-monochromatic given that the bandwidth $\Delta\nu$ is much smaller than the mean frequency $\bar{\nu}$. The beams are then reflected on the mirrors M_1 and M_2 and reunited on the screen P. If the path difference $\Delta l = c\Delta t$ (with c being the speed of light in vacuum) is small enough, interference fringes are formed on P. Experiments show that the formation of fringes is restricted by the relation $\Delta t \Delta \bar{\nu} \leq 1$. Therefore, the formation or absence of fringes in the screen P is a manifestation of the coherence or lack of coherence of the incident field, respectively. The maximum value for the time delay $\Delta t = 1/\Delta\nu \equiv \tau_c$ is known as the *coherence time* of the beam. Similarly, the maximum path difference $\Delta l = c\tau_c \equiv l_c$ is called the *longitudinal coherence length.*

Figure 4.1(b) illustrates another of interference experiment; the Young's double slit experiment. Consider a thermal source σ of dimension Δs radiating quasi-monochromatic light through the pinholes $P1$ and $P2$ of the plane Q. The plane is located at a distance R from the source and an angle of $\Delta\theta$ characterizes the separation between P_1 and P_2 with respect to the source. If the distance between the pinholes P_1 and P_2 is much smaller than R, fringes are observed at the screen at the axial point P. This phenomena is said to be a manifestation of the spatial coherence between two light beams at P from the pinholes P_1 and P_2. It is found experimentally that if R is large enough, fringes are formed at P if $\Delta\theta\Delta s \leq \bar{\lambda}$, where $\bar{\lambda}$ is the mean wavelength of the incoming field. This implies that the region on the plane Q where the pinholes P_1 and P_2 must be located cannot exceed an area ΔA given by $\Delta A \sim (R\Delta\theta)^2$. This region is known as the *coherence area* associated to the source [149].

The area of coherence associated to sunlight is estimated to be $\Delta A \approx 3.67 \times 10^3 \mu m^2$, by Wolf et.al. [151]. This is compatible with the estimate of its coherence length $r_c \sim \hbar c/k_B T \sim 400nm$ for blackbody radiation at the Sun's surface temperature $T \simeq 5700K$. Photosynthetic membranes lie within this length-scale range, therefore they are expected to experience the effects of spatial coherence. The striking question is how they cope with, adapt to and even profit from, such correlations in the incident light.

4.1.2 Coherence function and degree of coherence

Let us now consider that the pinholes P_1 and P_2 of Figure 4.1(b) are located at the coordinate $\vec{r}_1$ and $\vec{r}_2$, respectively. The fluctuating field $V(\vec{r}, t)$ (e.g. a cartesian component of the electric field) at a point $\vec{r}$ on the screen P can be expressed as a superposition of the fields originated in the pinholes as [149]

$$V(\vec{r}, t) = K_1 V(\vec{r}_1, t - t_1) + K_2 V(\vec{r}_2, t - t_2) \tag{4.1}$$

where t_i is the time needed for the light to go from the pinhole P_i to the screen P, and K_i is a constant factor that depends on the geometrical properties of the pinhole i. The instantaneous field intensity at the point P is defined as

$$I(\vec{r}, t) = V^*(\vec{r}, t) V(\vec{r}, t). \tag{4.2}$$

The ensemble average intensity associated to the field in the point P, under the assumption of ergodicity and stationarity is given by

$$\begin{aligned} \langle I(\vec{r}, t)\rangle &= |K_1|^2 \langle I(\vec{r}_1, t)\rangle + |K_2|^2 \langle I(\vec{r}_2, t)\rangle \\ &\quad + 2\mathrm{Re}\{K_1^* K_2^* \Gamma(\vec{r}_1, \vec{r}_2, t_1 - t_2)\} \end{aligned} \tag{4.3}$$

where the function $\Gamma(\vec{r}_1, \vec{r}_2, t_1, t_2)$ is the cross-correlation function associated to the random process $V(\vec{r}_1, t)$ and $V(\vec{r}_2, t)$ defined as

$$\Gamma(\vec{r}_1, \vec{r}_2; t_1, t_2) = \langle V^*(\vec{r}_1, t_1) V(\vec{r}_2, t_2)\rangle \tag{4.4}$$

The cross correlation function is called the mutual coherence function and it constitutes the central quantity on the elementary theory of optical coherence. It is convenient to define the normalized cross correlation function

$$\begin{aligned} g(\vec{r}_1, \vec{r}_2, t_1 - t_2) &= \frac{\Gamma(\vec{r}_1, \vec{r}_2, t_1 - t_2)}{[\langle \Gamma(\vec{r}_1, \vec{r}_1, 0) \rangle]^{1/2} [\langle \Gamma(\vec{r}_2, \vec{r}_2, 0) \rangle]^{1/2}} \\ &= \frac{\Gamma(\vec{r}_1, \vec{r}_2, t_1 - t_2)}{[\langle I(\vec{r}_1, 0) \rangle]^{1/2} [\langle I(\vec{r}_2, 0) \rangle]^{1/2}}. \end{aligned} \tag{4.5}$$

This quantity is called the second order degree of coherence associated to the field vibrations at the points $P_1(\vec{r}_1)$ and $P_2(\vec{r}_2)$. For any set of values of $\vec{r}_1$, $\vec{r}_2$ and $t_1 - t_2$, $|g(\vec{r}_1, \vec{r}_2, t_1 - t_2)|$ is defined only within the interval $[0, 1]$.

4.1.3 Functional formalism for photoelectric counting distribution

Investigations of the statistical behavior of fluctuating light beams are done through photo counting statistical techniques [152; 153; 154]. The photo count distribution can be obtained from the momentum generating function. Consider an incident radiative random field over a set of N detectors located at single points $\vec{r}_1, \vec{r}_2, ..., \vec{r}_N$ on a detection time interval τ. The multidimensional generating function associated to an incident field can be written as [152]:

$$G(s_1, s_2, ..., s_N; \tau) = \left\langle \prod_{k=1}^{N} \exp\left(-s_k J_k\right) \right\rangle \tag{4.6}$$

where s_k is the expanssive parameter associated to the random function J_k which is given by

$$J_k = \alpha_k \int_{-\tau/2}^{\tau/2} |V(\vec{r}_k, t)|^2 dt, \tag{4.7}$$

α_k is proportional to the quantum efficiency of the kth detector and the incident field on a point $\vec{r}_k$ is $V(\vec{r}_k, t)$. The generating function $G(s_1, s_2, ..., s_N; \tau) = G(\{s_k\}; \tau)$ is connected to the joint probability $P(n_1, n_2, ..., n_N)$ that n_k photo-excitations are registered by the k-th detector ($k = 1, ..., N$) through the relation [152; 153]

$$P(n_1, n_2, ..., n_N) = \left\{ \prod_{k=1}^{N} \frac{(-1)^{n_k}}{n_k!} \frac{\partial^{n_k}}{\partial s_k^{n_k}} \right\} G\left(\{s_k\}\right)\big|_{\{s_k=1\}}. \tag{4.8}$$

The average ensemble in equation (4.6) is calculated over the set of random variables $V_k = V(\vec{r}_k, t)$. The probability functional $W[V_1^*, V_2^*, ..., V_N^*, V_1, V_2, ..., V_N] = W[V^*, V]$ is therefore $2N$ dimensional, and the multidimensional generating function is given by

$$G(s_1, s_2, ..., s_N; \tau) = \int W[V^*, V] \exp\left(-\sum_{l=1}^{N} s_l \alpha_l \int_{\tau/2}^{\tau/2} V_l^* V_l dt \right) \prod_{l=1}^{N} dV_l^* dV_l. \tag{4.9}$$

It is convenient to work with the characteristic function Φ, which is known for some fields and under some conditions, it represents the $2N$ dimensional Fourier transform of W. The general form is

$$\Phi[\zeta, \zeta^*] = \exp\left\{ i \int \zeta(\vec{r}, t) V^*(\vec{r}, t) d\vec{r} dt + i \int \zeta(\vec{r}, t)^* V(\vec{r}, t) d\vec{r} dt \right\} \tag{4.10}$$

The function $\zeta(\vec{r}, t)$ is then a complex vector field whose Fourier expansion contains positive frequency amplitudes. For a set of N point detectors, we choose $\zeta(\vec{r}, t)$ to be

$$\zeta(\vec{r}, t) = \sum_{k=1}^{N} \delta(\vec{r} - \vec{r}_k) \zeta_k(t), \tag{4.11}$$

where the function $\zeta_k(t)$ vanishes outside of the interval $(-\tau/2, \tau/2)$. The probability functional W can therefore be written in terms of its inverse Fourier transform as

$$\begin{aligned} W[V^*, V] \quad &= \quad \int \Phi[\zeta, \zeta^*] \exp\left(-i \sum_k (\zeta_k, V_k^*) - i \sum_k (\zeta_k^*, V_k) \right) \\ &\qquad \times \prod_{l=1}^{N} d\left(\frac{\zeta_l}{2\pi}\right) d\left(\frac{\zeta_l^*}{2\pi}\right) \end{aligned} \tag{4.12}$$

where we have used the notation for scalar products

$$(V_l^*, V_l) = \int_{-\tau/2}^{\tau/2} V_l^*(t) V_l(t) dt. \tag{4.13}$$

Finally, after performing the integration over V_l^* and V_l, the final form of the multidimensional generating function is

$$\begin{aligned} G(s_1, ..., s_N; \tau) &= \int \Phi[\sqrt{\alpha_1 s_1}\zeta_1, ..., \sqrt{\alpha_N s_N}\zeta_N, \sqrt{\alpha_1 s_1}\zeta_1^*, ..., \sqrt{\alpha_N s_N}\zeta_N^*] \\ &\times \exp\left(-\sum_{k=1}^{N}(\zeta_k, \zeta_k^*)\right) \prod_{l=1}^{N} d\left(\frac{\zeta_l}{2\pi}\right) d\left(\frac{\zeta_l^*}{2\pi}\right). \end{aligned} \tag{4.14}$$

For a stationary thermal field, the characteristic functional is given by [152]

$$\Phi[\zeta, \zeta^*] = \exp\left\{-\sum_{k,l=1}^{N} \int_{-\tau/2}^{\tau/2} \zeta_k(t_1)\Gamma_{k,l}(t_1 - t_2)\zeta_l^*(t_2) dt_1 dt_2\right\} \tag{4.15}$$

where $\Gamma_{k,l}(t_1 - t_2) = \langle V_k^*(t_1) V_l(t_2)\rangle$ is the mutual coherence function. It is convenient to define the kernel

$$D_{k,l}(t_1 - t_2) = \sqrt{s_k \alpha_k}\Gamma_{k,l}(t_1 - t_2)\sqrt{s_l \alpha_l} + \delta_{k,l}\delta(t_1 - t_2) \tag{4.16}$$

and the generating function takes the following form

$$\begin{aligned} G(s_1, ..., s_N; \tau) &= \int \exp\left\{-\frac{1}{2}\sum_{k,l=1}^{N} \int_{-\tau/2}^{\tau/2}\int_{-\tau/2}^{\tau/2} \zeta_k(t_1) D_{k,l}(t_1 - t_2)\zeta_l^*(t_2) dt_1 dt_2\right\} \\ &\times \prod_{l=1}^{N} d^2\left(\frac{\zeta_l}{\sqrt{2\pi}}\right) \end{aligned} \tag{4.17}$$

The integration results on the inverse of the determinant of $D_{k,l}$ that corresponds to the product of the inverse of the eigenvalues of the following set of integral equations

$$\sum_{l=1}^{N} \int_{-\tau/2}^{\tau/2} \sqrt{s_k \alpha_k}\Gamma_{k,l}^*(t_1 - t_2)\sqrt{s_l \alpha_l}\phi_l^{(\mu)}(t_2) dt_2 + \phi_k^{\mu}(t_1) = \lambda_\mu \phi_k^{(\mu)}(t_1), \quad k = 1, ..., N \tag{4.18}$$

And the final form for the generating function associated to thermal light can therefore be written as

$$G(s_1, ..., s_N; \tau) = \prod_{\mu} \lambda_{\mu}^{-1}. \tag{4.19}$$

4.2 Fully coherent limit and incoherent limit

The generating function can be calculated explicitly if we consider photo-excitations arising in a time interval (τ) that is shorter than the coherence time of the light (τ_c). However, our main findings are robust to detection times that are a factor of 10 or more higher than τ_c. Under this condition on τ the integral equation (4.18) can be solved explicitly given that the functions do not change considerably within the interval $[-\tau/2, \tau/2]$. The generating function can be expressed as:

$$G(\{s_k\}; \tau) = |\Delta_N|^{-1} \tag{4.20}$$

with $|\Delta_N|$ being the determinant of the $N \times N$ matrix

$$(\Delta_N)_{i,j} = \delta_{i,j} + \sqrt{\langle n_i \rangle s_i} g_{i,j} \sqrt{\langle n_j \rangle s_j}. \tag{4.21}$$

The second-order degree of coherence associated to thermal light between detectors located at $\vec{r_i}$ and $\vec{r_j}$ is $g_{i,j}$ and $\langle n_i \rangle$ is the average number of photo-counts at the i-th-detector. The latter is directly connected with the mean intensity through the relation $\langle n_i \rangle = \alpha_i \langle I_i \rangle \tau$, where α_i is the quantum efficiency of the i-th-detector. It is known from *Rsp. photometricum* purple bacteria [107], which are among the first examples of photosynthetic membranes formed on Earth, that a typical $\simeq 200$ nm wide photo-receptive harvesting vesicle will be entirely contained inside the coherence area of sunlight.

The probability of detection of n photons regardless of the specific counting record of any individual detector:

$$P(n) = \sum_{\{n_i\}} \delta \left(n - \sum_{i=1}^{N} n_i \right) P(n_1, n_2, ..., n_N) \tag{4.22}$$

which can be formally obtained from Eq.(4.21)) by putting $\langle n_i \rangle \, s_i \to \langle n \rangle s$, yielding to

$$P(n) = \frac{(-1)^n}{n!} \frac{\partial^n}{\partial s^n} \{\det \left| \hat{1} + \langle n \rangle \; s \, \hat{g} \right| \}_{s=1}^{-1} \tag{4.23}$$

where $\hat{1}$ and $\hat{g}$ are the identity and the second-order degree of coherence $N \times N$ matrices, respectively. Two limiting cases illustrate the important impact of spatially correlated photons.

4.2.1 Full spatial coherence

This limit is characterized by situations where the area contained by the N antennae is much smaller than the coherence area of the thermal light ($g_{i,j} = 1$ for any pair of detectors). For this limit we obtain:

$$\begin{aligned} P(n) &= \frac{(-1)^n}{n!} \frac{\partial^n}{\partial s^n} \left[\frac{1}{1 + N \, \langle n \rangle \; s} \right]_{s=1} \\ &= \frac{\tilde{n}^n}{(1+\tilde{n})^{n+1}} \end{aligned} \tag{4.24}$$

which corresponds to the Bose-Einstein distribution (BED) for an average photon number $\tilde{n} = N \langle n \rangle$, a single parameter distribution $P_{\tilde{n}}(n)$.

4.2.2 The totally uncorrelated limit

At this limit the second order degree of coherence yields to the $N \times N$ identity matrix ($g_{i,j} = 0$ for any pair of detectors, $i \neq j$), and corresponds to situations where the distance between any pair of detectors is larger than the spatial coherence length of the thermal light. The result is

$$\begin{aligned} P(n) &= \frac{(-1)^n}{n!}\frac{\partial^n}{\partial s^n}\left[\frac{1}{(1+\langle n\rangle\ s)^N}\right]_{s=1} \\ &= \binom{N+n-1}{n}\left(\frac{1}{1+\langle n\rangle}\right)^N\left(\frac{\langle n\rangle}{1+\langle n\rangle}\right)^n \end{aligned} \tag{4.25}$$

which is the so-called negative binomial or Pólya distribution (a two parameter distribution $P_{N,\langle n\rangle}(n)$) with $\binom{a}{b}$ denoting a binomial coefficient. Notice that in the $N = 1$ case Eq.(4.25) reduces to Eq.(4.24), i.e. a BED with $\tilde{n} = \langle n\rangle$, as it should be since a single point detector is always within the thermal coherence area for any finite coherence length. On the other hand, Eq.(4.25) can be rewritten as

$$P_{N,\tilde{n}}(n) = \frac{\tilde{n}^n}{n!}\frac{\Gamma(N+n)}{\Gamma(N)(N+\tilde{n})^n}\left(1+\frac{\tilde{n}}{N}\right)^{-N} \tag{4.26}$$

which in the limits $N \to \infty$ and $\langle n\rangle \to 0$ with $\tilde{n} = N\langle n\rangle = constant$, reduces to the Poisson distribution

$$P_{\tilde{n}}(n) = \lim_{N\to\infty} P_{N,\tilde{n}}(n) = \frac{\tilde{n}^n}{n!}e^{-\tilde{n}} \tag{4.27}$$

again, a single parameter distribution. A Poisson distribution of detected photons in photosynthetic membranes has been always taken for granted. However, this assumption must be seen as highly questionable when thermal light is used and the area spanned by the set of point detectors is smaller than the coherence area of the radiation field (specially for counting times smaller than or of the order of τ_c for which biochemistry individual processes in the membranes occur). According to Ref. [155] the mean photon number $\langle n\rangle$

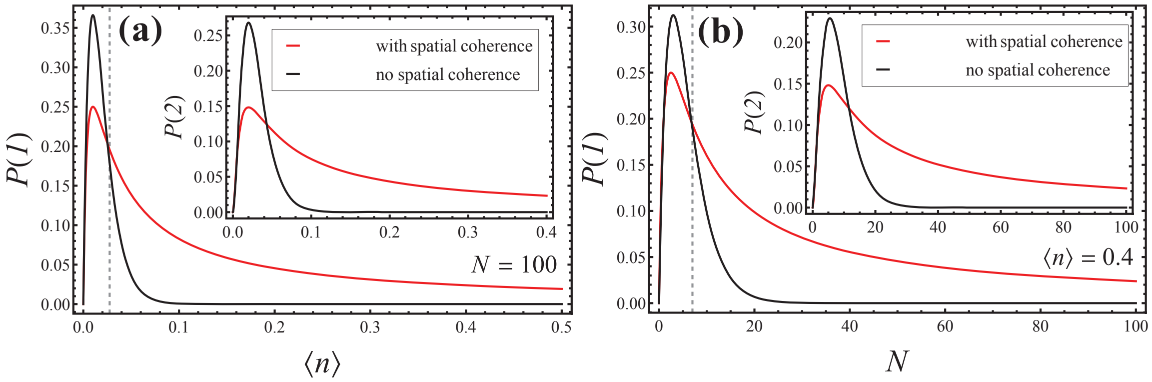

Figure 4.2: Probability of detection of one photon $P(1)$ as a function of (a) intensity $\langle n \rangle$ for a cluster of $N = 100$ molecular antennae, and (b) number of molecular antennae (N) for a fixed intensity ($\langle n \rangle = 0.4$). Red lines correspond to full spatial coherences, i.e. $g_{i,j} = 1$. Black lines correspond to no spatial coherence, i.e. $g_{i,j} = 0$ with $i \neq j$. Dashed lines indicate the crossing point of the fully coherent and incoherent description at (a) $\langle n \rangle = 0.03$ and (b) $N = 7$. Insets show the probability for detecting pairs, $P(2)$.

in a coherence volume of thermal light (blackbody radiation) is of the order of 1, showing that thermal light is on the borderline between classical and quantum regimes. The fact that the probability distribution with full spatial coherence is greater than that of no spatial coherence (see Fig. 4.2) across essentially the full range of $\tilde{n}$, proves that the coherence of sunlight will increase the photon detection.

Figure 4.2 shows explicitly how the incident light's spatial correlations impact the probability of photon detection by the N molecular antennae. Different light intensities are particular to the specific environment of the photosynthetic organism. For example, purple bacteria *Rsp. photometricum* initiates specific light intensity adaptations when subject to growth intensity variations [35; 36; 26] from 10 to 100 W/m^2 in laboratory. However, the light intensities in natural environment can decrease as low as 1 W/m^2. For small light intensities ($\langle n \rangle > 0.03$ in Fig. 4.2(a)) and for molecular antennae that are clustered within the coherence radiation area as in realistic photosynthetic membranes ($N > 7$ in Fig. 2(b)), the probabilities for 1 and 2-photon detection are consistently higher for spatially coherent light than for fully incoherent light. These probabilities are also fairly insensitive to fluctuations in $\langle n \rangle$ and N and hence offer robustness protection to the underlying membrane against atmospheric variations and the recurrent photo-bleaching of pigments.

4.3 Primitive clusters of molecular antennae

To analyze the effect of spatial coherence on simple antennae architecture motifs, the expression for the second order degree of coherence associated to thermal light as registered by detectors located at $\vec{r}_i$ and $\vec{r}_j$, $g_{i,j}$, is given by [154]

$$g_{i,j} = \frac{2J_1(u_{i,j})}{u_{i,j}} \tag{4.28}$$

where $J_1(k)$ is the Bessel function of the first kind, $u_{i,j} = \kappa|\vec{r}_i - \vec{r}_j|$ is the effective ratio between the detectors' separation and the spatial coherence length, where κ depends on the average wavelength, the size of the source and the source-detector separation. Figure 4.3 illustrates the photo-count statistics for a configuration of five detectors and three values of light intensity. The configuration consists of three molecular antennae in an equilateral triangle and a fourth one at the center. The probabilities of detecting $n = 1, 2, 3$ and 4 photons are calculated as a function of the position of the fifth detector.

The configuration of photo-detectors (Fig. 4.3) that maximizes the probability of absorbing a specific number of photons, is remarkably similar to the honeycomb architecture found in several species of purple bacteria [156; 122; 124; 126], completely analogous to the hexagonal unit cell 'honeycomb' configuration. The number of photons n, whose probability is maximized, is directly related to the light intensity. Importantly, for low intensity ($\langle n \rangle = 0.5$), the honeycomb architecture maximizes the probability of detecting one

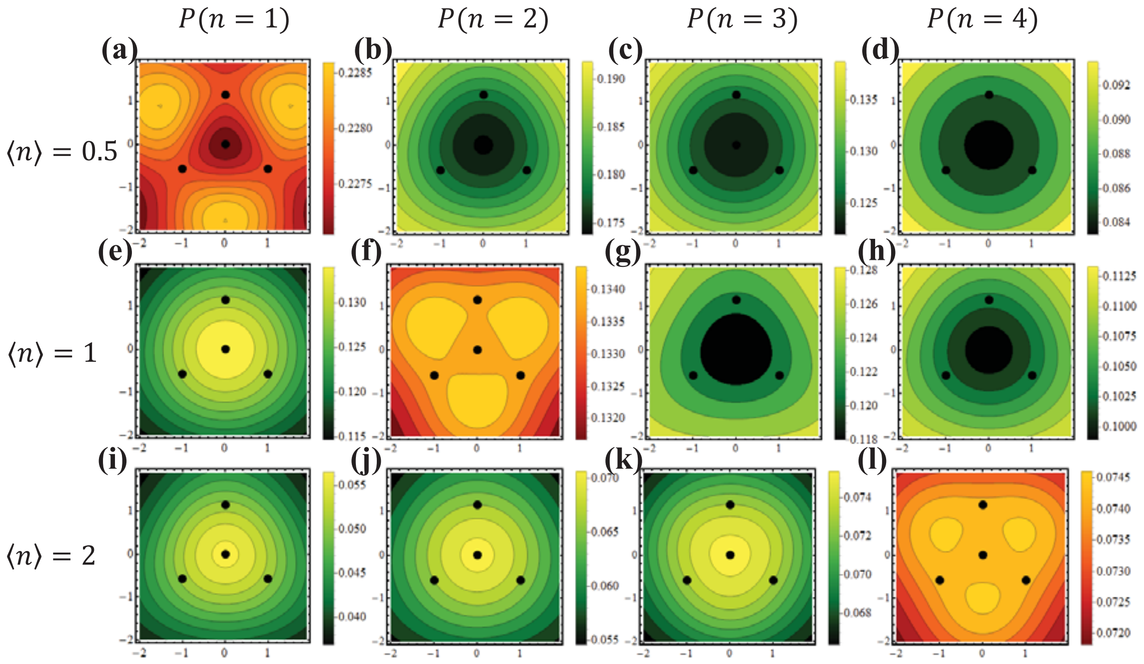

Figure 4.3: Results from numerical evaluation of the full joint probability for $N = 5$ molecular antennae. The value of $P(n)$ is plotted as a function of the position of the 5-th detector in the x-y plane, with the remaining 4 detectors fixed in space (large black dots). Results are shown for spatially coherent light at low intensity (a-d), medium intensity (e-h) and high intensity (i-l). Similar results follow for different geometries (see SI). The dimensionless distance between nearest neighbor detectors is fixed to $u = 1$ (see Eq.(4.28)) and all distances in the $x - y$ plane are measured in terms of this inter-detector distance.

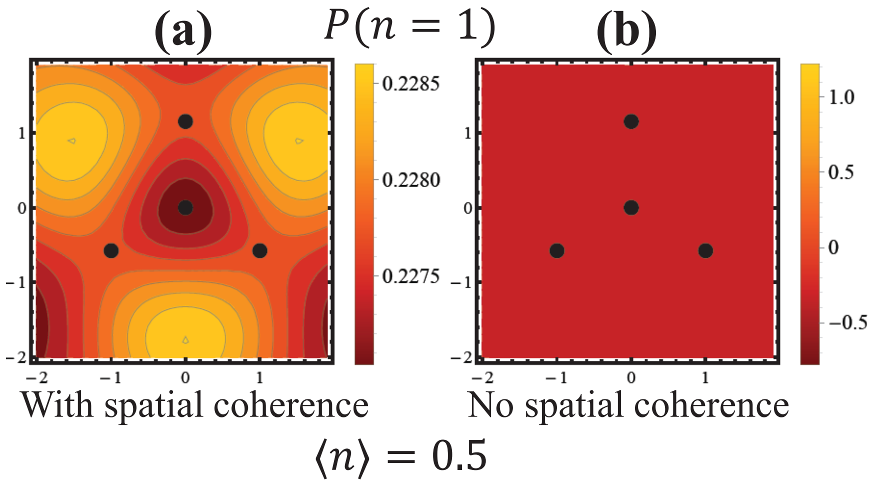

Figure 4.4: Spatial coherence effects on a system presented in Fig.4.3. Probability of detecting $n = 1$ photon is shown as a function of the position of the fifth detector for radiation whose intensity is proportional to $\langle n \rangle = 0.5$. (a) With spatial coherence. (b) No spatial coherence.

photon (Fig 4.3.[a]). Similarly, for medium ($\langle n \rangle = 1$) and high ($\langle n \rangle = 2$) light intensity, the honeycomb architecture maximizes the probability of detecting two (Fig 4.3.[f]) and four photons (Fig 4.3.[l]) respectively. This suggests that for a given light intensity, the organism positions antennae detectors such that the spatial coherence can better support the metabolic needs by either maximizing absorption or protecting the membrane from damage due to overexposure. For example in environments where the light intensity is high, an organism could express a non-symmetric architecture as a defense mechanism from the light – while for low light intensity environments, the system could express a symmetric architecture (e.g. honeycomb) in order to maximize the absorption of incoming photons. Intermediate configurations could maximize the absorption probability of a specific number of photons given the metabolic needs of the organism. These statements are consistent with the observed chromatic adaptation of the antennae complexes to the light intensity during the growing stage [26; 122]. However, our findings suggest that the photosynthetic apparatus adapts to both the incident light's intensity *and* correlations. By contrast, in the case of no spatial coherence, there is no preferred position for the fifth detector and hence the light offers no guiding organizational principle as shown in Fig. 4.4.

4.4 Arbitrary detection times

The previous results were calculated under the assumption of detection times τ shorter than the coherence time of light τ_c. Here we extend the conclusions obtained previously for systems where the underlying assumption does not necessarily apply. To that end we consider that the radiation generated by the superposition of beams preserve the spectral distribution of the original source. This property is known as cross spectral purity [152; 153] and allows to further simplify the set integral equations (4.18). The coherence function can be separated into a product of the spatial contribution and the temporal contribution.

$$\begin{aligned} g_{i,j}(t-t') &= g_{i,j}(0)g(t-t') \qquad (4.29) \\ g_{i,j}(0) &= \langle I_i\rangle^{-1/2}\langle I_j\rangle^{-1/2}\Gamma_{i,j}(0). \qquad (4.30) \end{aligned}$$

Using this condition and introducing the matrix $B_{k,l} = \sqrt{s_k\alpha_k\langle I_k\rangle}g_{k,l}(0)\sqrt{s_l\alpha_l\langle I_l\rangle}$, the equation (4.18) yields to:

$$\sum_{l=1}^{L} B_{k,l}\int_{-\tau/2}^{\tau/2} g(t_1-t_2)\phi_l^{(\mu)}(t_2)dt_2 + \phi_k^{(\mu)}(t_1) = \lambda_\mu\phi_k^{(\mu)}(t_1). \qquad (4.31)$$

A unitary transformation to the matrix $B_{l,k}$ brings it to its diagonal form where each diagonal element constitutes an eigenvalue b_k of $B_{l,k}$

$$\sum_{m,n=1}^{N} U_{k,m}B_{m,n}U_{n,l}^{-1} = b_k\delta_{k,l} \qquad (4.32)$$

Multiplying by $U_{k,m}$ to the left of (4.31), introducing an identity operator $\sum_{n=1}^{N} U_{n,l}^{-1} U_{n,l}$ to the left of $B_{k,l}$, and summing over m, the equation (4.31) results in

$$b_k \int_{-\tau/2}^{\tau/2} g(t_1 - t_2)\psi_k^{(\mu)}(t_2)dt_2 + \psi_k^{(\mu)}(t_1) = \lambda_\mu \psi_k^{(\mu)}(t_1) \tag{4.33}$$

where

$$\psi_k^{(\mu)}(t) = \sum_{l=1}^{N} U_{k,l}\psi_l^{(\mu)}(t), \quad k = 1, ..., N \tag{4.34}$$

The temporal dependence of (4.33) is 1-fold, therefore the eigenvalues can be expressed as $\lambda_\mu = 1 + \omega_\nu b_k$, where ω_ν's are the eigenvalues of the 1-fold temporal integral equation

$$\int_{-\tau/2}^{\tau/2} g(t_1 - t_2) f_\nu(t_2)dt_2 = \omega_\nu f_\nu(t_1), \tag{4.35}$$

and the eigenfunctions $f_\nu(t)$ are proportional to $\psi_k(t)$ by a time independent vector ψ_k. The generating function (4.19) is therefore written as a double product

$$G(s_1, ..., s_N; \tau) = \prod_\nu \prod_{k=1}^{N} (1 + \omega_\nu b_k)^{-1} \tag{4.36}$$

Notice that in the case where the detection time τ is much smaller than the coherence time of incoming light τ_c, the integral equation (4.35) yields to a single eigenvalue $\omega_\nu = \tau$, hence the generating function is fully determined by the spatial contribution of the kernel which is equivalent to Eq.(4.20).

$$G(s_1, ..., s_N; \tau) = \prod_{k=1}^{N} (1 + \tau b_k)^{-1} = [\det|\hat{1} + s\langle n\rangle \hat{\gamma}|]^{-1}, \tag{4.37}$$

where $\hat{1}$ and $\hat{\gamma}$ are the $N \times N$ identity and second order degree of coherence matrix, respectively. Let us now consider a particular case where the thermal source is characterized by a Lorentzian of line width $1/\tau_c$ whose kernel is given by

$$g(t, t') = \exp[-|t - t'|/\tau_c]. \tag{4.38}$$

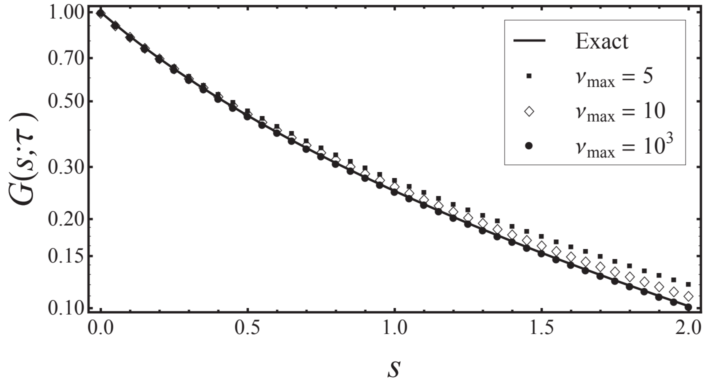

Figure 4.5: Comparison between numerical approximation (symbols) and exact result (solid line) for the generating function for the case of a single detector, ($N = 1$). The parameter ν_{max} is the number of terms evaluated in the product over ν of the equation (4.36). Parameters: $\beta = 1$ and $\langle n \rangle = 2$.

For this kernel, it can be shown that the eigenvalues ω_ν of the equation (4.35) are given by [157]

$$\omega_\nu = \frac{2\beta\tau}{\beta^2 + 4\chi_\nu^2}, \quad \beta \equiv \tau/\tau_c, \tag{4.39}$$

where χ_ν are the roots of the transcendental equations $\chi_\nu \tan\chi_\nu = \beta/2$ and $\chi_\nu \cot\chi_\nu = -\beta/2$. For the simple case of one detector ($N = 1$), the generating function takes the compact form [152; 153]

$$G(s;\beta) = \frac{e^\beta}{\cosh z + \sinh z(\frac{\beta}{2z} + \frac{z}{2\beta})}, \tag{4.40}$$

where $z = (2\beta \langle n \rangle s + \beta^2)^{1/2}$ and $\langle n \rangle = \alpha \langle I \rangle \tau$ is the mean number of detections on a time interval τ. In this way the generating function is fully determined by the parameters β and $\langle n \rangle$. For an arbitrary configuration of detectors, a numerical calculation is more suitable if we are looking for finite and nonzero values of β. For the limit of large β, the roots χ_ν can be approximated to integers multiples of $\pi/2$, and the generating function can be calculated explicitly as

$$G(s;\beta) = \prod_{k=1}^{N} \frac{\beta \sinh\beta \operatorname{csch}\sqrt{\beta^2 + 2\beta\tau b_k}}{\sqrt{\beta^2 + 2\beta\tau b_k}}, \tag{4.41}$$

$P(n)$	Exact	$\nu_{max} = 5$	$\nu_{max} = 10$	$\nu_{max} = 10^3$
$P(0)$	0.248552	0.271604	0.259346	0.248652
$P(1)$	0.260658	0.260819	0.260961	0.260663
$P(2)$	0.187893	0.181122	0.184725	0.187863
$P(3)$	0.119955	0.113937	0.11707	0.119928

Table 4.1: Probability values associated with the detection of n photons by one detector. Values calculated by an exact calculation and numeric approximations for different number of terms on the product of equation (4.36). Parameters: $\beta = 1$, $\langle n \rangle = 2$ and $N = 1$.

and taking the limit for $\beta \to \infty$, the expression takes the well known Poisson generating function

$$G(s;\beta) = \prod_{k=1}^{N} e^{-\tau b_k} = e^{-s\langle n \rangle N}. \tag{4.42}$$

At this limit, the probability distribution only depends on the number of detectors N independent from their location in space. For finite non-zero values of β it is required to construct the set χ_ν and numerically calculate the generating function.

On the other hand, for the case where the coherence time of light and the detection time are comparable ($\beta \approx 1$) we need to use the formal expression given by Eq.(4.36) with the appropriate coefficients ω_ν given by Eq.(4.39). Since the analytic solution is known for the case of $N = 1$, we could estimate the number of zeros of the transcendental equation to accurately calculate ω_ν. Figure 4.5 compares the exact generating $G(s;\tau)$ and numerical approximations for the particular case of $\beta = 1$ and $\langle n \rangle = 2$. The number of coefficients required to obtain a good estimate for the generating function is 1000. Larger values of β require larger number of coefficients. Table 4.1 shows the numerical values of the probability of detecting $n = 0, 1, 2, 3$ photons from the different approximations.

Figure 4.6 summarizes the finding of the effects of spatial coherence on photo-count statistics for different values of detection time β. For simplicity, it is presented the curve of the increment on the probability of detecting 1 photon by the configuration if the fifth detector lies on a line at $y = 1$ as defined in Fig.4.6(a). Notably, the triple maximum effect observed previously which supports a honeycomb architecture as optimal, is preserved for finite values of β which gives robustness to the original result. However, the increment in β results on a reduction on $\Delta P(1)$ as shown in Fig.4.6(b). This is consistent with the result for the limit of $\beta \rightarrow \infty$ where the Poisson limit was found. In addition, it provides a temporal measure on which coherent effects are relevant for the system only if it is fully contained on the area of coherence of light. The small effects on the probability should not be considered negligible since it is known that even small increments on overall measures (e.g. photosynthetic efficiency) could be determined in order to keep the metabolic cycle. These results are consistent with architectures found in real species that are particularly grown under low light intensity conditions (extreme environments).

4.5 Full architecture model

Having established the impact of light's spatial coherence for highly primitive geometric motifs, we examine results for actual architectures in purple bacteria (e.g. Fig. 3.4(b) corresponding to *Rsp. Photometricum*). We consider a simple model for photo-bunching by spatially correlating the absorption of two consecutive photons by the photosynthetic membrane. The maximum transverse distance between antenna complexes where two consecutive absorptions occur, is defined as r. The site is determined stochastically with specific probabilities weighted by the complex's cross section based in turn on an individual pigment's cross-section which yields to 2.32 Å^2. When the neighborhood's radius r is large enough to cover the whole membrane, we recover the limit of random arrival.

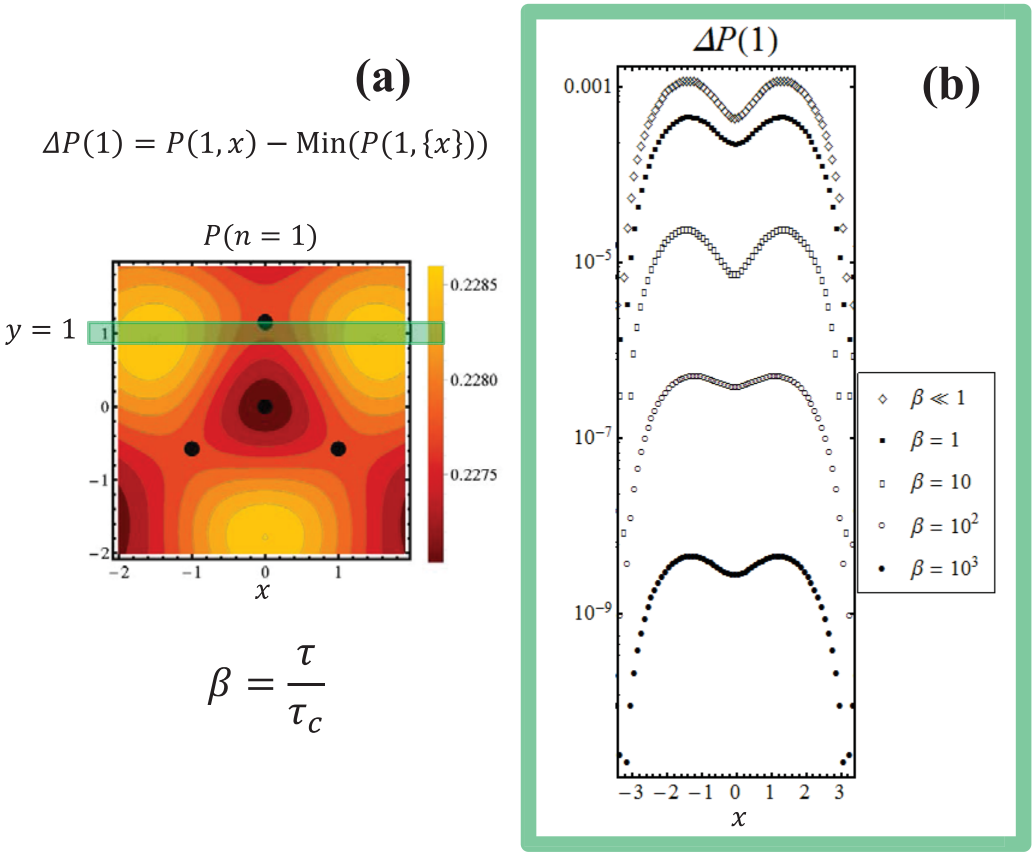

Figure 4.6: Variation in the photo-detection probability $P(n = 1)$ ($\Delta P(1)$ as defined in (a)) for a configuration of four fixed detectors arranged on an equilateral triangle (black dots) as a function of the location of the fifth detector, for different values of $\beta \ll 1$, $\beta = 1, 10, 10^2, 10^3$. The curves (b) illustrate the green path highlighted in the contour plot (a) for $y = 1$ as a function of x.

The validity of our detailed theoretical model of the membrane's excitation dynamics [35; 36; 25], which inputs empirical excitation transfer times (see chapter 2), is verified by its demonstrated prediction of structural preferences in purple bacteria as a function of light intensity [35; 36] together with the close agreement between its output and an analytical version based on coupled Master Equations. Recent empirical work also demonstrated its applicability in other primitive bacterial systems such as *Rps. Palustris* [122] and *Rb. sphaeroides* [123]. The functional relevance of the correlated absorption can be quantified in terms of the excitations that successfully initiate charge separation in the RC. After such an event, the RC is then is closed for a few milliseconds before an available charge carrier establishes its neutrality within a time τ_{RC}. We calculate the relative efficiency of the photosynthetic process as a function of the RC closure time τ_{RC} and correlation distance r for two architectures with the same number of complexes ($N = 400$) and stoichiometry ($N_{LH2}/N_{LH1} \approx 8.09$), but different levels of core-core clustering. The efficiency (η) is defined as the ratio of photo-excitations that produce charge separation at an RC, to the total number of absorbed photons. Similarly, we define the random arrival efficiency η_{RA} to be the value when spatial correlations are not present. This value is equivalent to the photosynthetic efficiency in the limit when the correlation parameter r is much greater than the membrane length. Finally, the relative efficiency is defined as the ratio of the membrane efficiency to the random arrival efficiency, η/η_{RA}. The clustering is quantified by the mean number of RC-LH1 neighbors for a given RC-LH1 core complex. Consequently, high clustering relates to high numbers of RC-LH1 complex neighbors while low clustering means low numbers of core complex neighbors. Our calculations consider two contrasting levels of organization: (i) High Core-Core (HCC) clustering with mean number of RC-

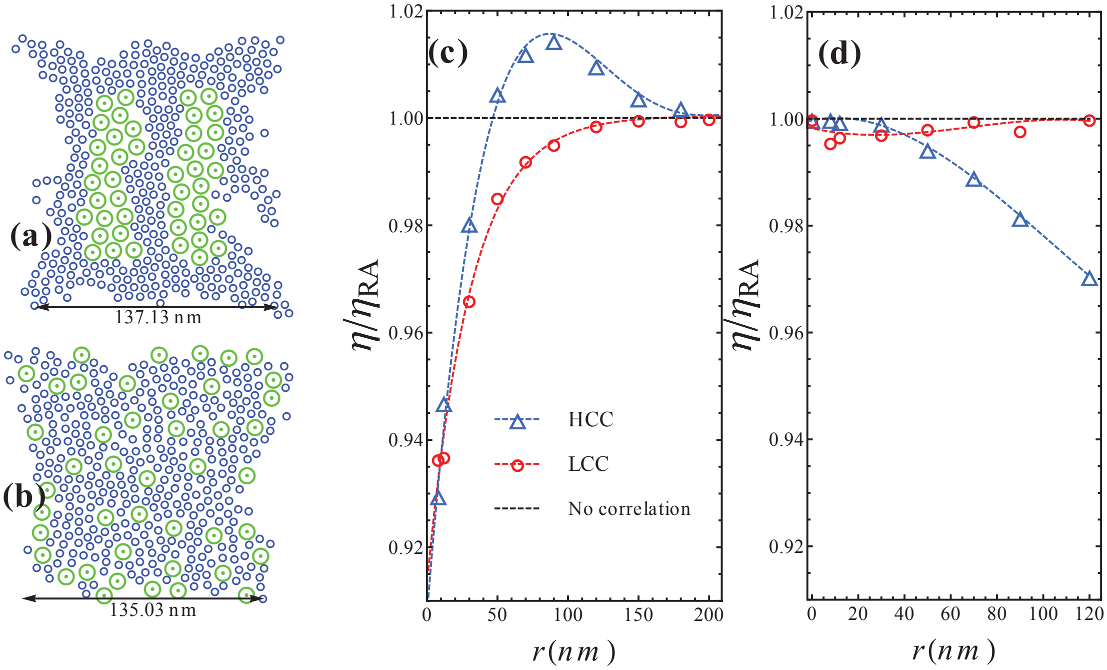

Figure 4.7: Architectures of the antenna complexes (a) HCC and (b) LCC. They are obtained as local minima in a large-scale Monte Carlo energy minimization, and are hence realistic as (locally stable) minimum energy structures. Blue rings represent LH2 antennae and green rings are core RC-LH1 complexes. (c) Photosynthetic relative efficiency as a function of the correlation parameter r for architectures HCC (triangles) and LCC (circles). Dashed black line illustrates the result without light correlation. Symbols represent the result from our model while lines correspond to the fitting of our simulation points. (d) Analogous to (c) but for negative correlation. The RC closure time is $\tau_{RC} = 12.5$ms.

LH1 neighbors as 3.4, which corresponds to the maximally core-core clustering observed in *Rsp. photometricum* [44]; (ii) Low Core-Core (LCC) clustering which is characterized by an average of 0.31 RC-LH1 neighbors for each core complex.

Figure 4.7(c) shows the striking result that for high core-core (HCC) clustering, the structure benefits from the specific range of correlation distances within the incident sunlight. The left panel illustrates the LHC network for our two sample architectures: HCC (Fig.4.7(a)) and LCC (Fig.4.7(b)) while the right panel displays their photosynthetic relative efficiencies from our model. By contrast, there is no such peak enhancement for the membrane with LCC clustering. Our simulation uses an RC closure time of 12.5ms, in agreement with the average time for quinone release [158]. However we have shown that values up to 50ms preserve this peak enhancement. This novel finding suggests that the organism may group the core-complexes to make the most of the spatial correlations in the incoming light. In particular, our simulations reveal that the aggregation of core complexes reduces the lifetime of excitations that produce charge separation on the RC, which indicates a reduction in the number of complexes visited before ionization takes place. Therefore, LH1-RC core aggregation helps reduce the mean path of excitations and hence their probability of being dissipated. This could indeed be the case for *Rb. blasticus* and *Rb. sphaeroides*[125; 124] where the presence of the helix protein PufX induces dimerization of core-complex [127], as well as *Rsp. photometricum*[44] and *R. rubrum* [126]. We have checked that other incident light inputs with different inter-photon arrival time statistics (e.g. Poisson, power-law and periodic) but the same light intensity, show minimal difference – hence the effect is purely spatial nature. For negative spatial correlation in the incident light (i.e. opposite of natural sunlight), the peak disappears as shown in Fig.4.7(d). As mentioned above, the increment in efficiency showed on Fig. 4.7(c) should not be considered as negligible. It has been demonstrated that on these organisms large structural variations occurred as a result of 'small' metabolic benefit. For instance, *Rsp. acidophilia* which, under low illumination condition, is known to gradually replace complexes LH2 by LH3 whose absorption maximum is shifted to 820nm. This modification increases the funneling toward LH1 and hence increasing its efficiency. However, this expensive structural adaptation yields to only 3-4$\%$ of increment [106].

These results also speak to open questions about the purpose of membrane organization [125; 156; 159], showing that it allows the system to harvest not only the light's energy, but also to profit from its spatial correlations. While other energetic and metabolic factors may ultimately dictate a given membrane's architecture in a given environment, this work shows for the first time that spatial correlations in sunlight potentially play a key role, and should not be ignored.

4.6 Summary of chapter 4

This chapter provides quantitative evidence that (i) the organization of the antennae complexes within a photosynthetic membrane complements the spatial correlations naturally present in sunlight; (ii) either an ordered or disordered antennae structure is favored depending on the availability of light and its metabolic needs, yielding high-efficiency photon absorption for low light-intensity environments and protection for high light-intensity environments; (iii) the macro-molecular aggregation of harvesting structures displays a high sensitivity to spatial correlations which then affects the exciton migration and subsequent change separation at the reaction center (RC); (iv) high core-core clustering membranes yield an enhancement of the photosynthetic efficiency that is not observed for membranes with low core-core clustering; (v) the complementarity between the spatial correlations in the incident light and the evolved biomaterial structure not only enhances photon capture and efficiency, but can also guide the organization (Fig 3.4.[a]) and types of antennae (Fig 3.4.[b]) to better serve the organism's metabolic needs. Previous work has discarded spatial correlations in the incident light as negligible and hence overlooked these five consequences.

Chapter 5

Character-driven transition in dynamical grouping

The understanding of emerging collective phenomena is a challenge encompassing disciplines such as physics, chemistry, economics and social sciences [51; 52; 53; 54; 55; 56; 57; 58]. Whether the individual objects are particles, proteins or people, the impact at the macro-scale is determined by their evolution as aggregates or isolated individuals [57; 58; 160; 60; 161; 162; 63]. Phenomena of super-radiance results from the coherent coupling of two-level systems with a bosonic model [163]; several neuro-degenerative diseases are driven by protein aggregation [164]; large market movements are driven by traders' herding [24; 165; 166]; insurgencies are driven by illicit human grouping [167; 168; 169; 170]; collective neuronal avalanches is an aspect of brain activity [23].

Despite the innate heterogeneity of these real-world systems, an emergent universality has been reported when looking at the size distribution of aggregates [19; 20; 167]. Specifically, a power-law with exponent of $5/2$ is documented for: (i) the size distribution of pockets of superconductive coherence in fragmented materials [22]; (ii) the size distribution of neuronal avalanches, once the probability of avalanche initiation from a random neuron is corrected (i.e. $k \cdot k^{-5/2} = k^{-3/2}$[23]); (iii) the size distribution of herds of agreeing traders and the price movements that they inflict [24]; (iv) the size distribution of insurgent groups and therefore the severity of attacks inflicted over a civilian population [167; 170]. The challenge is to reconcile the observed universality with the innate diversity of the individuals.

This chapter attempts to address this challenge by discussing a model of aggregation and fragmentation of heterogeneous individuals. An internal variable called ’character’ is introduced to account for individual heterogeneity among agents. The mechanism for breaking and forming groups is proposed be character-driven. Two contrasting mechanisms are presented where either similarity or dissimilarity between individuals is favored in order to form and break groups. It will be demonstrated (through simulations and analytic approach) that the introduction of the heterogeneity preserves the $5/2$-exponent size distribution for the different character-driven mechanisms. In addition, the group mechanism and the diversity distribution dictate a dynamical transition between isolation and group cohesive regime. In particular, different critical points arise for the various group mechanisms (favoring similarity or dissimilarity) which move in opposite direction as the population’s diversity decreases. Finally, a combined infective-based mechanism is developed and it is shown that it correctly predicts the non-monotonic time evolution in the connectivity of a real-world system.

5.1 Model description

Consider a system composed by N individual objects. Each object can establish a non-directed link to any of the $N-1$ remaining individuals. Connected objects form a cluster. Over time, clusters can grow by establishing links with other clusters as well as fragment into individual pieces (Fig.5.1). A parameter p is introduced to be the probability that a

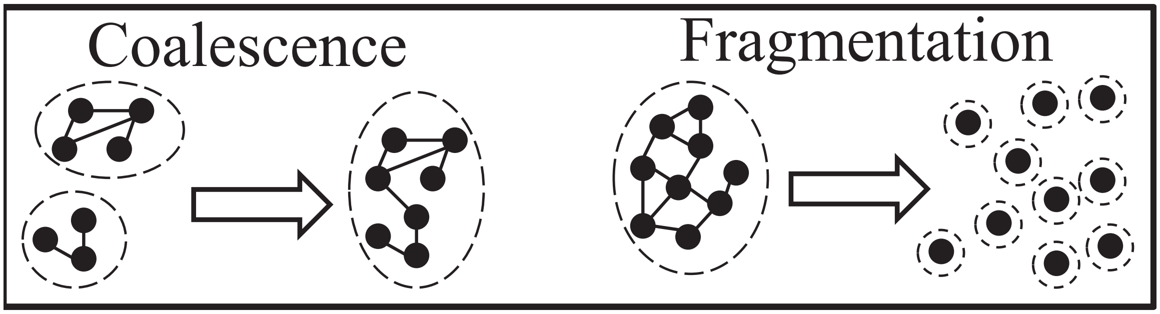

Figure 5.1: Processes carried out by the model. Coalescence: two objects establish a link and forming a new larger cluster. Fragmentation: a cluster shatters into individual pieces.

time step corresponds to one that an object considers to form a link with another object, furthermore $1-p$ is the probability that a time step corresponds to one that an already established link is chosen so that the whole cluster where the link belongs to may fragment. We assume that each agent carries a unique and fixed character value x which is assigned randomly at the beginning of the simulation. For simplicity, we consider the characters $\{x_i\}$ to be real numbers uniformly distributed between 0 and 1, but in general they can follow any distribution $q(x)$. We define the similarity $S_{i,j}$ between objects i and j is defined as $1-|x_i-x_j|$. In this way, objects with character values alike have S_{ij} near unity while objects with character values dissimilar (or complementary) produce S_{ij} near zero. Character-based group formation is studied. For this reason the probabilities of coalescence (c) and fragmentation (f) depend directly on the similarity. We study four types of formation mechanisms that favor similarity, complementarity, intermediate and character free models. The initial conditions are set so that all the N objects are isolated. A flow diagram applicable to all the model variations is presented in Fig.5.2.

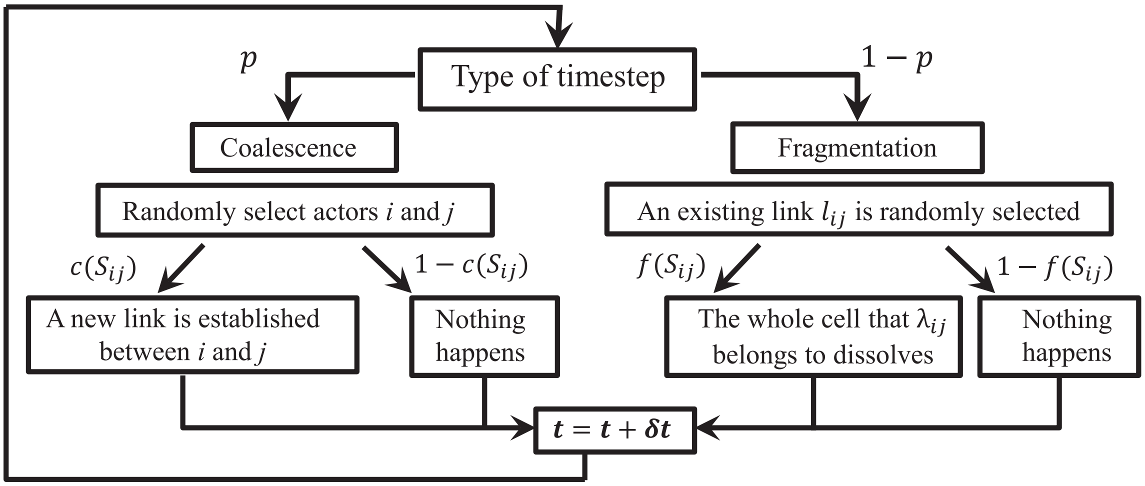

Figure 5.2: Flow diagram describing the formation and fragmentation process of our model including heterogeneity among the abjects.

5.1.1 Mechanism favoring similarity (M1)

For timesteps in which clusters coalesce: An object i is picked randomly from the collection of N objects. The object i can be part of a cluster or can be isolated. Similarly, a second object j is picked randomly from the collection. With a probability $c = S_{ij}$, a new link is established between object i and object j and hence a new cluster is created. The size of the new cluster is the sum of the sizes of the clusters that object i and object j belong to. If the objects were isolated, the new cluster's size is two. A link between object i and object j will not be formed with a probability of $1 - S_{ij}$. In turn, this mechanism promotes connection between objects with similar characters.

For timesteps in which clusters fragment: An existing link is randomly chosen. The end objects of that link contain character values x_i and x_j. With a probability of $f = 1 - S_{ij}$ the whole clusters where the link belongs to fragments. With a probability of S_{ij} the cluster will not fragment. Here, the similarity serves as a measure of the link's strength. A connection between actors with different character values is weak and hence easy to break.

5.1.2 Mechanism favoring complementarity (M2)

For timesteps in which clusters coalesce: An object i is picked randomly from the collection of N objects. The object i can be part of a cluster or can be isolated. In the same way, a second object j is picked randomly from the collection. With a probability $c = 1 - S_{ij}$, a new link is established between object i and object j and hence a new cluster is created. The size of the new cluster is the sum of the sizes of the clusters to which object i and object j belong. If the objects were isolated, the new cluster's size is two. A link between object i and object j will not be formed with a probability of S_{ij}. This mechanism therefore promotes connection between objects with dissimilar (or complementary) characters.

For timesteps in which clusters fragment: An existing link is randomly chosen. The end objects of that link contain character values x_i and x_j. With a probability of $f = S_{ij}$ the whole cluster to which the link belongs fragments. With a probability of $1 - S_{ij}$ the cluster will not fragment. Now, the similarity serves as a measure of the likelihood of the link to be broken.

5.1.3 Additional mechanisms M3 & M4

Mechanism M3 is proposed to be an intermediate set of grouping rules between M1 and M2. Clusters are formed favoring similarity among objects, $c = S_{ij}$, while the fragmentation process supports complementarity. Therefore, a connection with dissimilar character values is strong, $f = S_{ij}$. Finally, we consider a character-free mechanism (M4) where the objects coalesce and fragment independently from their character. The probability of coalescence and fragmentation is in all cases equal to one ($f = c = 1$).

5.2 Analytic treatment

5.2.1 Mean field probabilities for coalescence and fragmentation

The analytic description is a mean field approximation. Let F be the probability that two randomly chosen objects coalesce, while Q is the probability of fragmenting a cluster containing an arbitrarily chosen link. At first approximation, the quantities F and Q are constants.

Consider the model under M1 rules. For a uniform distribution of x_i, the probability density function (PDF) of $y = S_{ij}$, $k(y)$ is $k(y) = 2y$ with $y \in [0, 1]$. The probability F that two objects will be connected is

$$F = \int_0^1 k(y) y dy = \frac{2}{3}. \tag{5.1}$$

Similarly, the PDF of y associated with links is $g(y) = 3y^2$, hence the probability Q that a randomly selected link breaks is

$$Q = \int_0^1 g(y)(1 - y) dy = \frac{1}{4} \tag{5.2}$$

On the other hand, for a group under M2 rules, let us define $z = 1 - S_{ij}$. The PDF $k(z) = 2(1 - z)$ while the PDF for links is $g(z) = 6(1 - z)z$, and the probabilities F and Q are given by

$$\begin{aligned} F &= \int_0^1 k(z) z dz = \frac{1}{3} && (5.3) \\ Q &= \int_0^1 g(z)(1 - z) dy = \frac{1}{2}. && (5.4) \end{aligned}$$

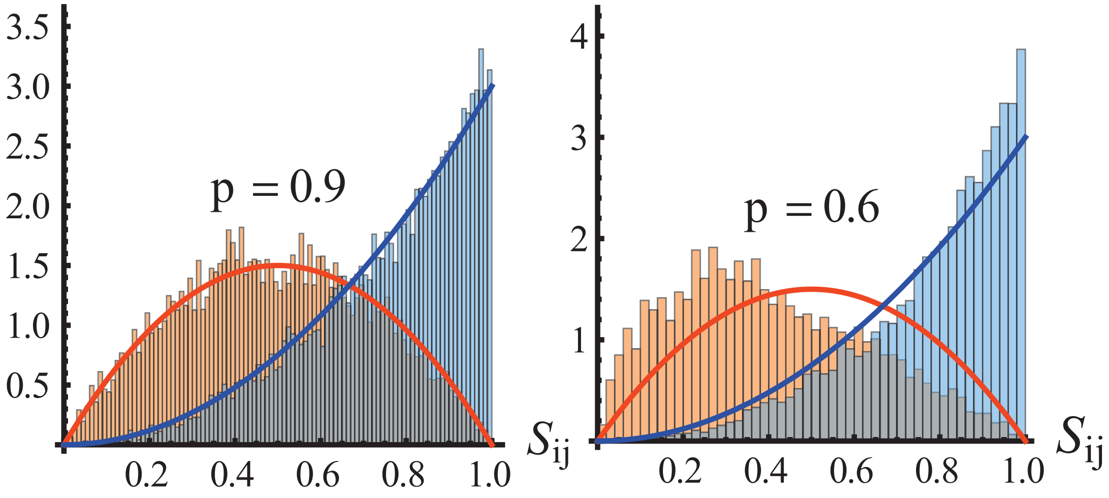

Figure 5.3: Distribution of links with similarity $g(S_{ij})$ for M1 (e.g. kinship, shown in blue) and M2 (e.g. team, shown in orange). Charts represent a snapshop of the numerical simulation in the steady state for $N = 10^4$ individuals and solid curves are functions $g(S_{ij})$ from the mean field approximation.

	c	f	p_c (mean field theory)	p_c(numerical)	F	Q
M1 (e.g. kinship)	S_{ij}	$1-S_{ij}$	$3/11$	0.10	$2/3$	$1/4$
M2 (e.g. team)	$1-S_{ij}$	S_{ij}	$3/5$	0.51	$1/3$	$1/2$
M3 intermediate	S_{ij}	S_{ij}	$9/17$	0.49	$2/3$	$3/4$
M4 character-free	1	1	$1/2$	0.50	1	1

Table 5.1: Values associated with different grouping mechanisms.

Figure 5.3 illustrates the distribution of similarity for the group mechanisms M1 (blue) and M2 (orange). It reflects the tendency of M1 and M2 to favor similarity and dissimilarity, respectively. Solid lines are PDF, $g(y)$, associated with link formation from our mean field approach. The agreement at first approximation is noteworthy, specially for large values of p. The differences are due to neglect of higher order correlations.

5.2.2 Distribution of size of clusters n_k

The coupled differential equation for the number of clusters size k, n_k, can be constructed as follow: the two master equations are:

For $k \geq 2$,

$$\begin{aligned}\frac{\partial n_k}{\partial t} &= -P(\text{size } k \text{ group fragments}) + P(\text{smaller groups combine to size } k) \\ &\quad -P(\text{size } k \text{ groups joins other groups}); \end{aligned} \tag{5.5}$$

while for $k = 1$,

$$\frac{\partial n_1}{\partial t} = -P(\text{single actor joins other group}) + P(\text{big group fragments}). \tag{5.6}$$

Mathematically, the equation for $k \geq 2$ becomes

$$\begin{aligned}\frac{\partial n_k}{\partial t} &= -(1-p)Q\frac{(k-1)n_k}{\sum_{r=2}^{\infty}(r-1)n_r} - 2pF\frac{kn_k}{N^2}\sum_{r=1}^{\infty} rn_r \\ &+ pF\frac{1}{N^2}\sum_{r=1}^{k} rn_r(k-r)n_{k-r}. \end{aligned} \tag{5.7}$$

The equation for $k = 1$ becomes

$$\frac{\partial n_1}{\partial t} = -2pF\frac{n_1}{N^2}\sum_{r=1}^{\infty} rn_r + (1-p)Q\frac{\sum_{k=2}^{\infty} k(k-1)n_k}{\sum_{r=2}^{\infty}(r-1)n_r}. \tag{5.8}$$

In the steady state, the left-hand side of each of these equations becomes zero. Using $\sum_{r=1}^{\infty} rn_r = N$ and $(k-1)n_k \approx kn_k$, we get from Eq.(5.7) that

$$\begin{aligned} kn_k &= \frac{pF(N-n_1)}{N[Q(1-p)N + 2pF(N-n_1)]}\sum_{r=1}^{k} rn_r(k-r)n_{k-r} \\ &\equiv \gamma\sum_{r=1}^{k} rn_r(k-r)n_{k-r} \end{aligned} \tag{5.9}$$

Let $g(w) = \sum_{r=2}^{\infty} rn_r e^{-wr}$, hence

$$\begin{aligned} g(w)^2 =&(2n_2e^{-2w} + 3n_3e^{-3w} + 4n_4e^{-4w} + \cdots)(2n_2e^{-2w} + 3n_3e^{-3w} + 4n_4e^{-4w} + \cdots) \\ =&(3n_1n_3 + 4n_2^2 + 3n_1n_3)e^{-4w} + (4n_1n_4 + 6n_2n_3 + 6n_2n_3 + 4n_1n_4)e^{-5w} + \cdots \\ &- 6n_1n_3e^{-4w} - 8n_1n_4e^{-5w} - \cdots \\ =&\sum_{r=1}^{4} rn_r(4-r)n_{4-r}e^{-4w} + \sum_{r=1}^{5} rn_r(5-r)n_{5-r}e^{-5w} + \cdots - 6n_1n_3e^{-4w} - \cdots \\ =&\frac{1}{\gamma}[2n_2e^{-2w} + 3n_3e^{-3w} + 4n_4e^{-4w} + 5n_4e^{-5w} + \cdots - 2n_2e^{-2w} - 3n_3e^{-3w}] \\ &- 2n_1e^{-w}(2n_2e^{-2w} + 3n_3e^{-3w} + 4n_4e^{-4w} + \cdots - 2n_2e^{-2w}) \\ =&\frac{1}{\gamma}[g(w) - 2n_2e^{-2w} - 3n_3e^{-3w}] - 2n_1e^{-w}[g(w) - 2n_2e^{-2w}] \end{aligned} \tag{5.10}$$

From Eq.(5.9), $2n_2 = \gamma n_1^2$ and $3n_3 = 4\gamma n_1 n_2$, so Eq.(5.10) becomes

$$g(w)^2 - \left[\frac{1}{\gamma} - 2n_1e^{-w}\right] g(w) + n_1^2e^{-2w} = 0 \tag{5.11}$$

For $w = 0$, $g(0) = \sum_{r=2}^{\infty} rn_r = N - n_1$, hence we get $g(0) = \gamma N^2$ and

$$n_1 = N - g(0) = N(1 - \gamma N). \tag{5.12}$$

Note that γ is itself a function of n_1. Equation (5.12) can be solved for n_1. The solutions are

$$n_1 = N \quad \text{or} \quad n_1 = \frac{pF + (1-p)Q}{2pF} N\ . \tag{5.13}$$

The smaller root is kept as the physical solution. Under the condition that $\frac{pF+(1-p)Q}{2pF} \geq 1$, i.e.,

$$p \leq \frac{Q}{F+Q}, \tag{5.14}$$

n_1 becomes N, indicating that there are only isolated actors in the system. Note that the model undergoes a transition from a regime of isolated objects ($n_1 = N$) to a regime of cluster formation $n_1 = N[pF + (1-p)Q]/2pF$. Since the mean field probabilities F and Q are taken to be constant, the critical value can be determined to be $p_c = Q/(F+Q)$. Consequently, for $p \leq p_c$ system is on a isolation regime and for $p > p_c$ the system is on a group cohesion regime.

Consider now the case of any general value of w. The quadratic equation Eq.(5.11) can be solved to give

$$g(w) = \frac{1}{2\gamma} - n_1 e^{-w} \pm \frac{1}{2\gamma}\sqrt{1 - 4\gamma n_1 e^{-w}} \tag{5.15}$$

The solution with the negative sign in the square root is kept. Using the property that $\sqrt{1+x} = \sum_{k=0}^{\infty} \frac{(\frac{1}{2})!}{k!(\frac{1}{2}-k)!} x^k$, we have

$$\begin{aligned} g(w) &= -\frac{1}{2\gamma}\sum_{k=2}^{\infty} \frac{(\frac{1}{2})!}{k!(\frac{1}{2}-k)!}(-4\gamma n_1 e^{-w})^k \\ &= \sum_{k=2}^{\infty} k n_k e^{-kw} \, . \end{aligned} \tag{5.16}$$

Therefore, the group size distribution for $k \geq 2$ is

$$n_k = -\frac{1}{2\gamma k}\frac{(\frac{1}{2})!}{k!(\frac{1}{2}-k)!}(-4\gamma n_1)^k \tag{5.17}$$

This is the full (exact) form of the group distribution $\{n_s\}$. Note that this expression also holds for n_1, i.e., plugging in $k = 1$ gives the previous n_1. Using the property of the gamma function,

$$(\frac{1}{2} - n)! = \Gamma\left(\frac{1}{2} - n + 1\right) = \frac{(-1)^{n-1} 2^{n-1}\sqrt{\pi}}{(2n-3)!!}. \tag{5.18}$$

For $n = 0$, $(\frac{1}{2})! = \frac{\sqrt{\pi}}{2}$. The $(2n-3)!!$ term can be expressed as

$$(2n-3)!! = \frac{(2n-2)!}{2^{n-1}(n-1)!}. \tag{5.19}$$

Therefore,

$$\frac{1}{k}\frac{(\frac{1}{2})!}{k!(\frac{1}{2}-k)!}=\frac{2(2k-2)!}{(-1)^{k-1}(k!)^2 4^k}. \tag{5.20}$$

Substituting into Eq.(5.17) and using the Stirling approximation

$$\ln(z!)\approx\frac{1}{2}\ln 2\pi+(z+\frac{1}{2})\ln z-z\ , \tag{5.21}$$

we have for large k that $\ln n_k$ is given approximately by the expression

$$\begin{aligned} & k\ln(\gamma n_1)+(2k-\frac{3}{2})\ln(2k-2)-(2k-2) \\ & -2(k+\frac{1}{2})\ln k+2k-\frac{1}{2}\ln 2\pi-\ln\gamma\ . \end{aligned} \tag{5.22}$$

Taking $2k-2\approx 2k$, we have

$$\begin{aligned} \ln n_k &\approx k\ln(\gamma n_1)+(2k-\frac{3}{2})\ln(2k)-2k \\ &- 2(k+\frac{1}{2})\ln k+2k-\frac{1}{2}\ln 2\pi-\ln\gamma \\ &= k\ln(\gamma n_1)-\frac{5}{2}\ln k+2k\ln 2-\frac{3}{2}\ln 2-\frac{1}{2}\ln 2\pi-\ln\gamma \\ &= k\ln(4\gamma n_1)-\frac{5}{2}\ln k-2\ln 2-\frac{1}{2}\ln\pi-\ln\gamma\ . \end{aligned} \tag{5.23}$$

Therefore

$$n_k\sim\frac{1}{4\gamma\sqrt{\pi}}(4\gamma n_1)^k k^{-5/2}=\frac{N}{2\sqrt{\pi}}\frac{p(1-p_c)}{p-p_c}\left[1-\left(\frac{p_c(1-p)}{p(1-p_c)}\right)\right]^k k^{-5/2}. \tag{5.24}$$

This approximate result for the distribution of clusters of size k presents the power-law behavior with exponent $-5/2$ and exponential cutoff, which is consistent with the empirical results. Therefore, introducing the heterogeneity among the objects retains the empirical observed feature. Another key point is that the result is independent of the formation mechanism. Groups that are created favoring similarity, dissimilarity, intermediate or character-

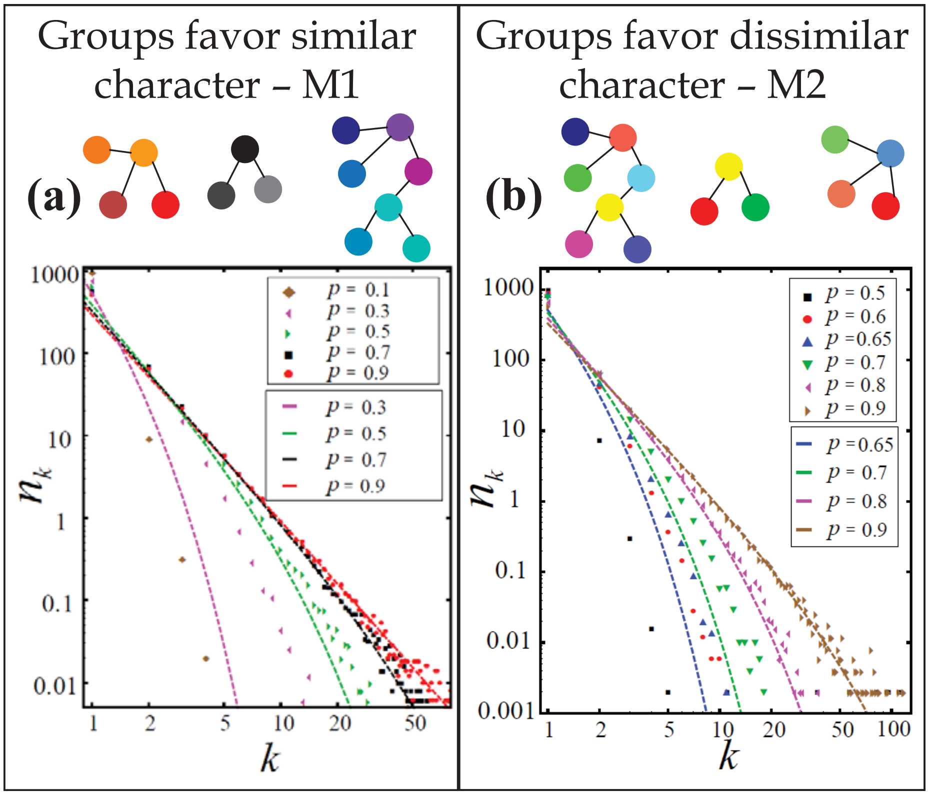

Figure 5.4: (a) Groups favoring similar character (e.e kin) illustrated by similar colors. Underneath, group size distribution n_k showing simulations (symbols) and analytical (lines) results for different p values. (b) Same as (a) but now groups favoring diverse characters (e.g. team).

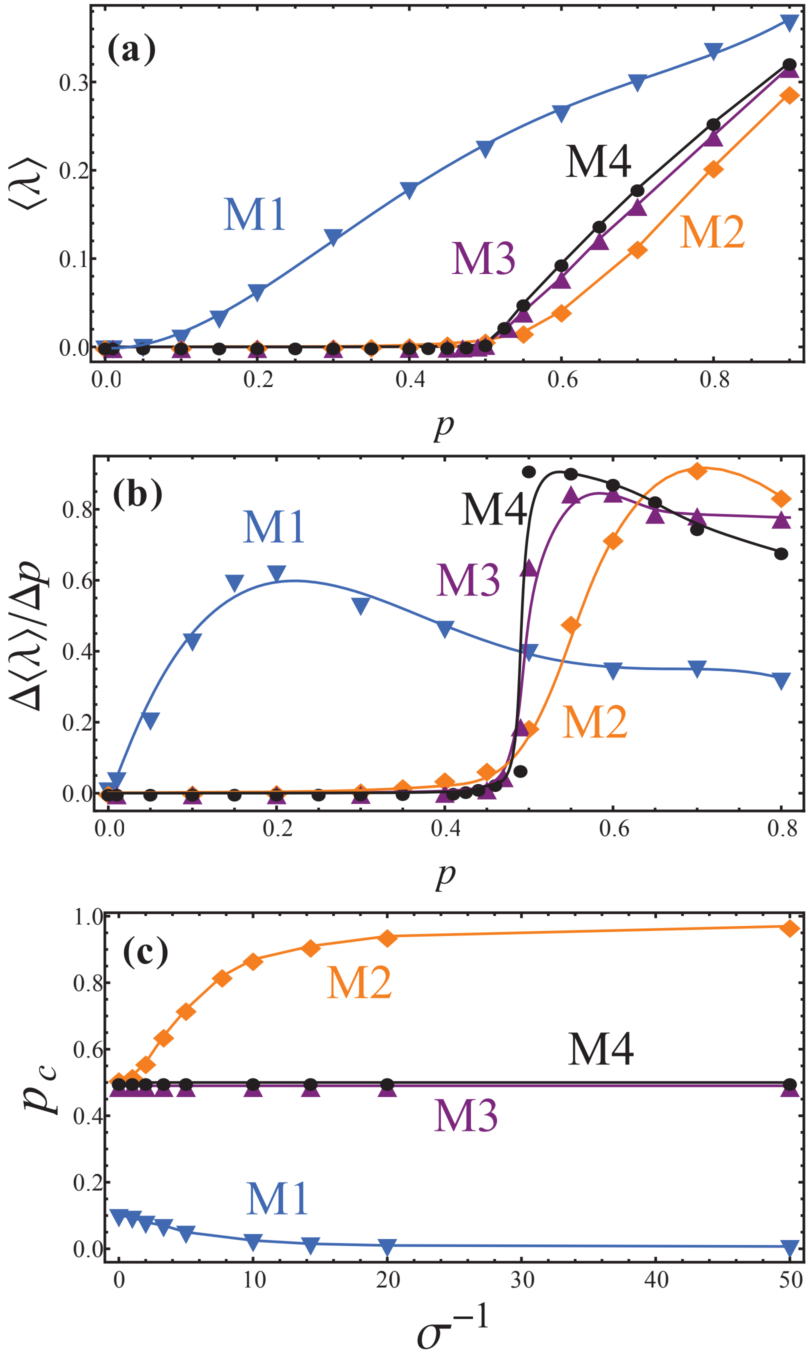

Figure 5.5: (a) show $\langle\lambda\rangle$ versus p for $N = 10^4$ objects. Blue points: grouping mechanism M1 favoring similar characters (e.g. kinship). Orange points: grouping mechanism M2 favoring diverse characters (e.g. team). Purple points: M3 intermediate between M1 and M2. Black points: M4 character-free. (b) Rate of change. Symbols on (a) and (b) are calculated numerically for an uniform distribution of character values $q(x)$ while the lines correspond to the fitting of our simulation points. (c) p_c for M1 (bottom, blue) and M2 (top, orange), M3 and M4 (horizontal) versus inverse standard deviation (σ^{-1}) of gaussian character distribution $q(x)$ centered on $\mu = 0.5$.

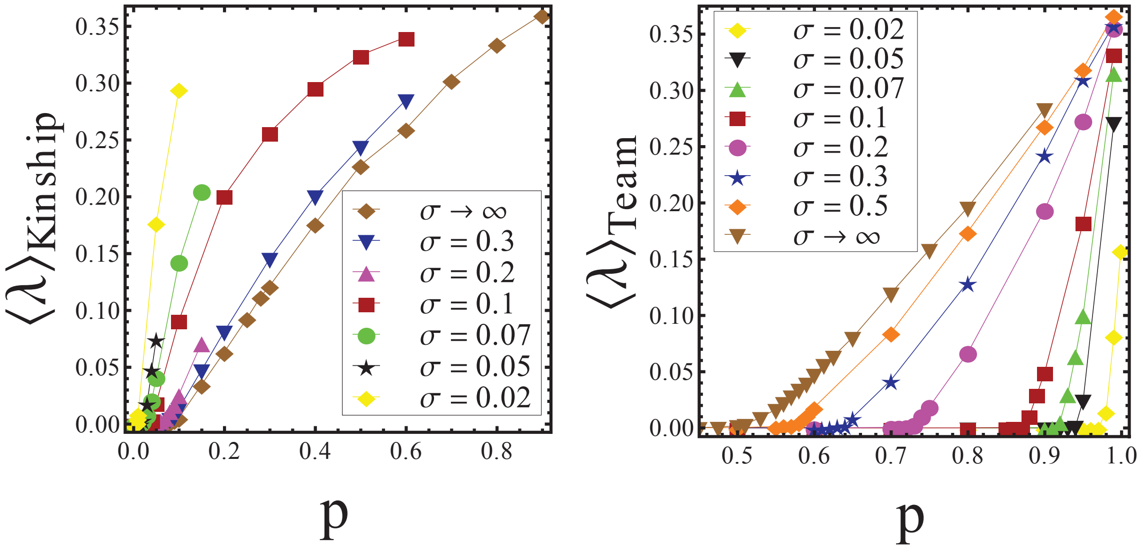

Figure 5.6: $\langle\lambda\rangle$ as a function of p for a gaussian character distribution $q(x)$ centered at $\mu = 0.5$ and different values of standard deviation σ, the limit $\sigma \to \infty$ is equivalent to the uniform character distribution $q(x)$. Left: M1 (e.g. kinship). Right: M2 (e.g. team). Simulations calculated for $N = 10^5$ objects.

free tendencies are prompted to be described by our fission-fusion model. Figure 5.4 contrasts numerical simulations and analytic results from equation 5.24, for several values of p and formation mechanisms M1 and M2. As seen in figure 5.3, small differences arise in the neighborhood of the transition point p_c, the agreement for large p is noteworthy. Numerical results are averages from simulations with $N = 10^4$ objects. Each simulation is comprised of 10^5 timesteps and data collected in the steady state.

5.3 Transition from isolation regime to group cohesion regime

As mentioned on the previous section, the model undergoes a transition from a regime where all objects are isolated to a regime of group cohesion. The group cohesion regime is characterized by a common macroscopic order; the cluster size distribution n_k. Since the critical point p_c only depends on the mean field probabilities F and Q, all the mechanisms M1-M4 experience these transition. In our simulations, the transition behavior is captured

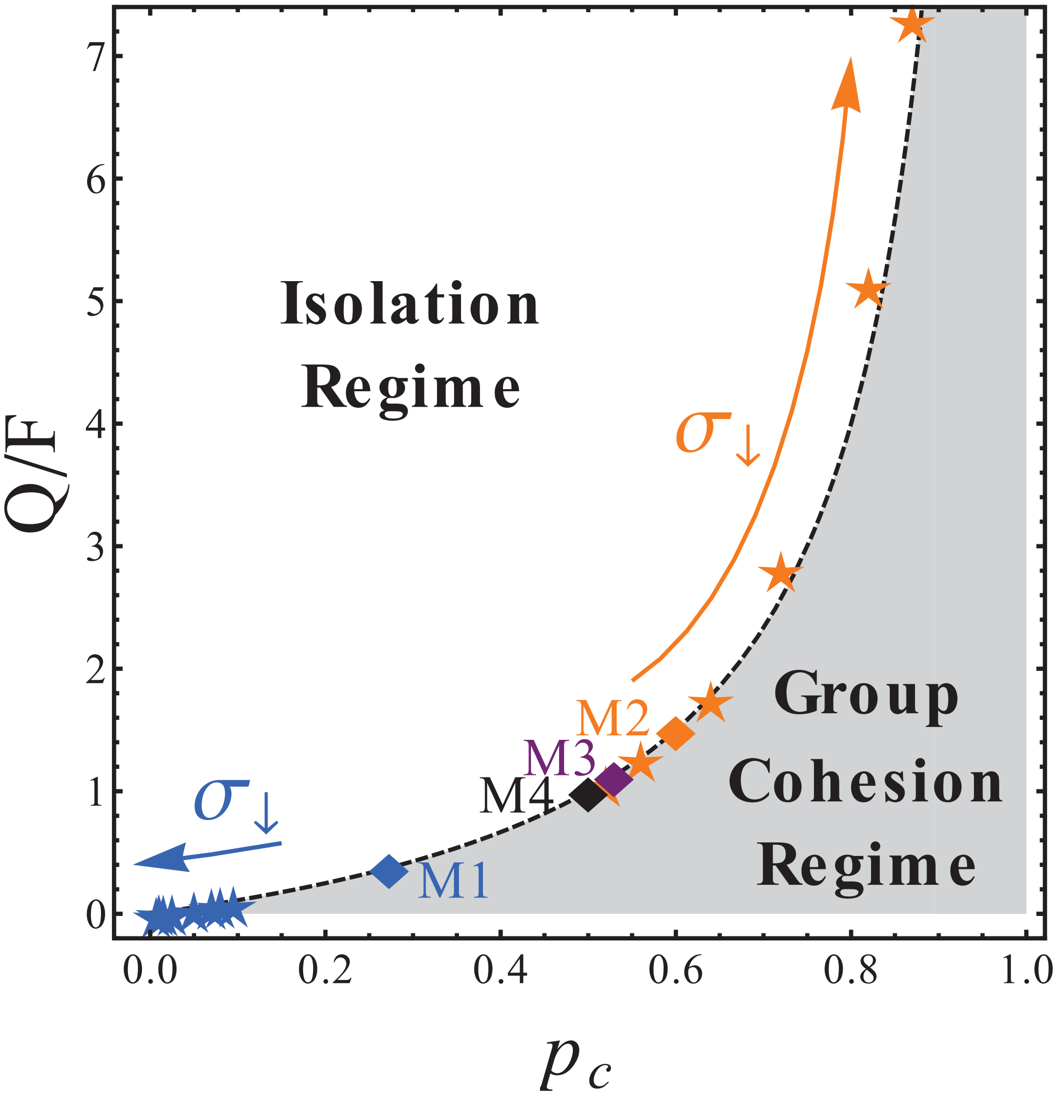

Figure 5.7: Regime diagram illustrating the parameter range corresponding to isolation and group cohesion. Curved regime boundary is our mean-field analytical result $p \equiv p_c = Q(F + Q)^{-1}$. Diamonds show p_c for uniform character distribution $q(x)$. Stars show numerical results for gaussian $q(x)$ from Fig. 2(c). M1 blue, M2 orange. Arrows illustrate that the critical point p_c for M1 moves on the opposite direction of M2 as the diversity (σ) is reduced.

by calculating the average number of links per object $\langle\lambda\rangle$ as a function of p. Figure 5.5(a) shows that even for a uniform character distribution $q(x)$, rich behavior emerges. As p increases, $\langle\lambda\rangle$ increases from zero, indicating that groups spontaneously form from a population of individuals. Figure 5.5.(b) shows the corresponding rate of change. Values of p_c from mean field approach and numerical simulation are summarized in Table 5.1. The value of p_c and the shape of the curve depend on the particular mechanism, where M1 is less abrupt and requires low values of p, while $M2$ requires higher values of p to create groups. This suggests that for a population with high diversity, team formation needs to be encouraged by externally incrementing the probability $p > p_c$, while for kinship formation arise naturally for very small p.

This implications could be seen more clearly by decreasing the diversity among individuals. Consider a character distribution $q(x)$ that follow a gaussian distribution centered at $\mu = 0.5$ and arbitrary standard deviation σ. Figure 5.5.(c) and 5.6 show that p_c's indeed shift in opposite directions, (M1 opposed to M2) as the heterogeneity of the underlying population is reduced. Therefore, kinship groups require an even lower p to form as the population becomes more homogeneous. By contrast, teams require an even higher p to form, and eventually as the homogeneity increases, the population will be comprised by mostly isolated objects.

Figure 5.7 shows the corresponding regime diagram. It marks the regions where the model yields to isolation of individuals or group cohesion. The boundary is the analytic curve $Q/F = p_c(1 - p_c)^{-1}$ from the mean field approach. Diamonds indicate the mechanisms M1-M4 for a uniform distribution of clusters $q(x)$. Interestingly, for a gaussian distribution with mean $\mu = 1/2$ and variable standard deviation σ denoted as stars into the diagram, lies remarkably close to the analytic curve. This fact provides further support to the mean field approach. In addition, the transition point p_c for M1 and M2, shifts in opposite directions as the diversity decreases, as indicated by arrows.

5.4 Team culture spreading over kinship

In this section we explore a variation on the model mechanism through contagion. Up until now we have considered the system to evolve, using only one mechanism. However, real-world organizations can undergo changes over time (e.g. to adapt to their changing environmental conditions). Provisional Irish Republican Army (PIRA), which is the best-known insurgency network to date [70], is a a good example of this sort of structure reorientation. PIRA operated in Northern Ireland from 1969 to 1998 and constitutes the most successful and innovative organization in history in terms of inventing, developing, planting and detonating improvised explosive devices (IEDs). PIRA started as a formal homogeneous army. However it experienced a bottom-up transition from the end of 1970s to beginning of 1980s, where the group operated as self-organized clusters of teams. The result of this transition was an increase in the productivity measured as number of attacks whose civilian casualties where negligible. A deeper discussion regarding other aspects of the group will take place in the next chapter.

We therefore test our model with the available data from PIRA. For that purpose, the model is modified as follows: In the initial state (i.e. birth of PIRA) N agents are unconnected since, by definition, PIRA was not yet involved in any IED attacks. In each time step ($\delta t \sim 1$ day) a new connection may be created or an old cluster may be fragmented with equal probability. $N-1$ actors initially follow M1 (i.e. susceptible) while one individual

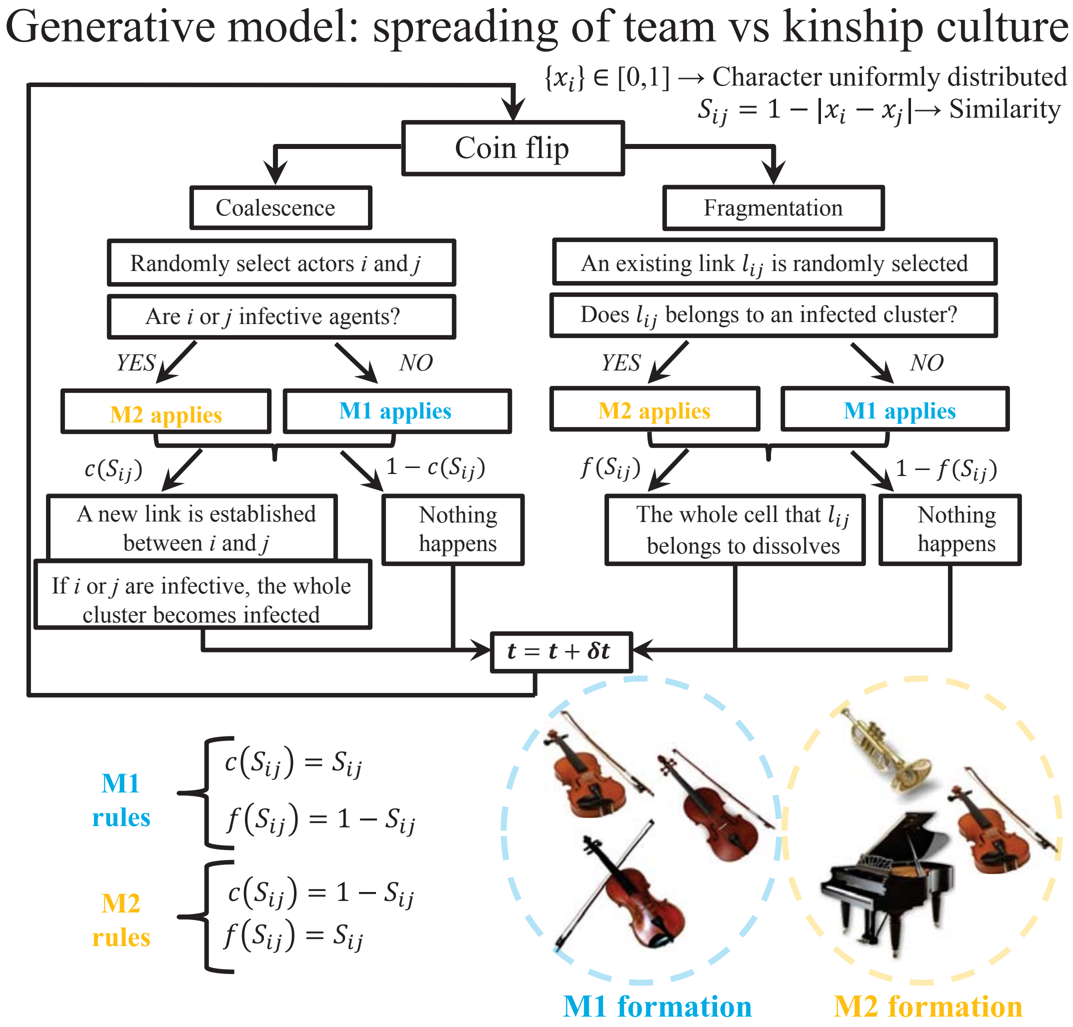

Figure 5.8: Flow diagram of generative multi-agent model.

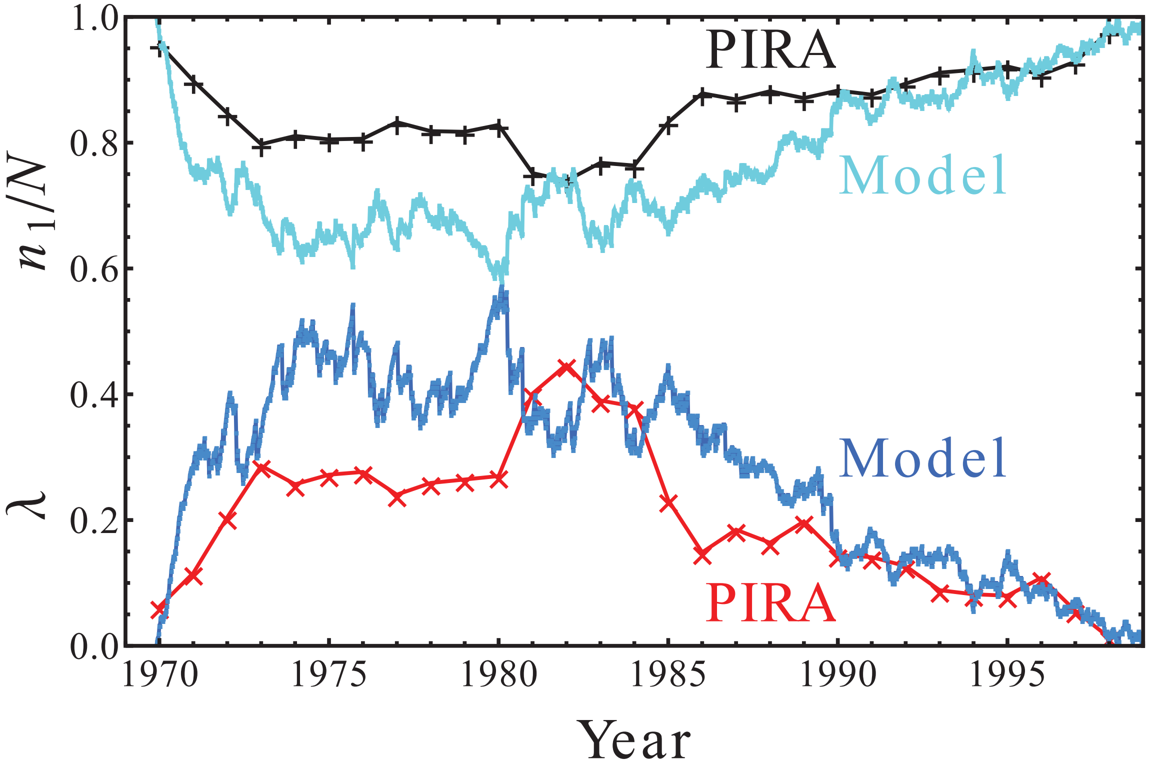

Figure 5.9: Our model's prediction vs. actual PIRA temporal variation for (top two curves) the fraction of isolated individuals (n_1/N), and (bottom two curves) the ratio between the total number of links in the network and the total number of individuals λ. PIRA data from Ref. [70]. The model considers equally formation and fragmentation time steps ($p = 0.5$). At $t = 0$, most of the actors follow M1 rules and only one object follows M2. However, this object spreads M2 rules as it joins other groups. The spreading overturns the formation rules from M1 to M2, just like the bottom-up transition from homogeneous to team-like structure experienced by PIRA. A timestep in the model corresponds to a day in real data

is a team oriented (M2) actor (i.e. infected). An actor is further labeled by its character x_i that was sampled from a uniform distribution $q(x)$. Susceptible (infected) agents follow M1 (M2) formation rules. In each time step, we randomly choose either the coalescence or fragmentation scenario with equal probability ($p = 0.5$). In the case of coalescence, a pair of agents i and j are randomly picked. If one of the agents is infected, M2 grouping mechanism apply, preferably linking actors with low similarity. In turn, M1 group mechanism links actors with high similarity. Essentially, infective agents can spread the team culture to the clusters where they join, which is not the case for susceptible agents. This uses the fact that team structure improved PIRA's efficacy and therefore got strengthened over time from ground local level. In the case of fragmentation, an existing link λ_{ij} between agents i and j is randomly picked. If λ_{ij} belongs to an infected cluster, M2 mechanism applies. Otherwise, M1 group mechanism applies, with a high probability of dissolving all links in the underlying cluster if similarity is low. Figure 5.8 presents the flow diagram associated to the underlying process.

Figure 5.9 compares the outcome of the model with the measure for the average number per links per actor λ as well as the fraction of isolated actors n_1/N associated with PIRA, over time. Notably, our combined model rightfully predicts the non-monotonic behavior of network connectivity with a detailed precision of a day and, this being the sole model parameter. In addition, it is worth mentioning that the features are not fully related (i.e. $\lambda + n_1/N \neq$const) and still our model is able to capture the dynamics. Finally, individual formation mechanism fails to produce these features. Unfortunately the data is insufficient to have detailed dynamics. The collected data are aggregated over time in the resolution of a year which limits our study.

5.5 Summary of chapter 5

In summary, introducing an internal character variable to individuals in a population yields rich transition behavior. The critical points depend on the character-driven mechanism and the distribution of character values (population's diversity) within the community. In addition, both simulation and analytic approximation preserve the group size universality despite large differences in the underlying objects and the grouping mechanisms. Finally, a simple combination of group mechanism was proposed and it showed to be in a good agreement with the non-monotonic time evolution of the connectivity of a real-world system.

Chapter 6

Network analysis in covert organizations

The quantitative understanding of terrorist organizations constitutes a challenge for today's scientists and government leaders [19; 20; 171]. The lack of high resolution data on covert organizations restricts the thoroughness of the studies and leaves several questions unanswered. New efforts on acquiring detailed data have been pursued by government organizations and universities in a multi-disciplinary effort to unfold the mechanisms that make illicit organization successful [172; 72]. Recently, the International Center for the Study of Terrorism (ICST) conducted a 42-month project where 28 years of Improvised Explosive Device (IED) activity in Northern Ireland [70] was investigated. Detailed open source data were collected for a two-year period with specific information on Provisional Irish Republican Army (PIRA) IED attacks, militant structure and strategies against counter-terrorist actions. This effort constitutes the best available data to date [70]. PIRA is a paramilitary organization whose main goal was to separate Northern Ireland from the United Kingdom and reunite Ireland. Despite the efforts of the British government to control trade of materials for IED, PIRA managed to find new machinery and components to continue its illicit activity. Nowadays, PIRA is known to be responsible of the greatest innovation in the construction, development and planting of IED.

PIRA was created in 1969 as an offshoot from the IRA and operated until 1998 where peace and political power were finally attained. Within this 29 year period, five different phases of the conflict have been identified [70]: (*i*) Between 1969 and 1976 indiscriminant violent acts were performed by both sides (PIRA and British government). Throughout the conflict, civilians represented the greatest casualties. (*ii*) From 1977 to 1980 the network underwent a structural transition toward a decentralized group of clusters which facilitated the secrecy of their activities. (*iii*) Between 1981 and 1989 PIRA gained political strength. In addition, the number of attacks increased steadily while attacks become more strategic,defining targets and causing less collateral casualties. (*iv*) From 1990 and 1994 negotiations opened the pathway toward peace. Finally, from 1995 and 1998 the cease of fire occurred together with the gain of political power by the IRA.

In this chapter a quantitative study of the networks associated with PIRA's IED attacks is presented. This study attempts to unfold key information about the network dynamical structure that other non quantitative approaches fail to unveil. Basic graph theory, statistics, agent based models along with other quantitative tools are utilized. The findings reveal that female minorities emerge with stronger collective network connectivity which provides benefits for system robustness and survival. These findings are further supported by analogous ongoing studies of recently aquired data from the current illicit organization known as Islamic State (IS).

6.1 PIRA's structural transition in the late 1970's and early 1980's

The transition carried out in the late 1970s' and early 1980s' brought important contributions to the ongoing IED activity of PIRA. While the activity during the of 1970s was

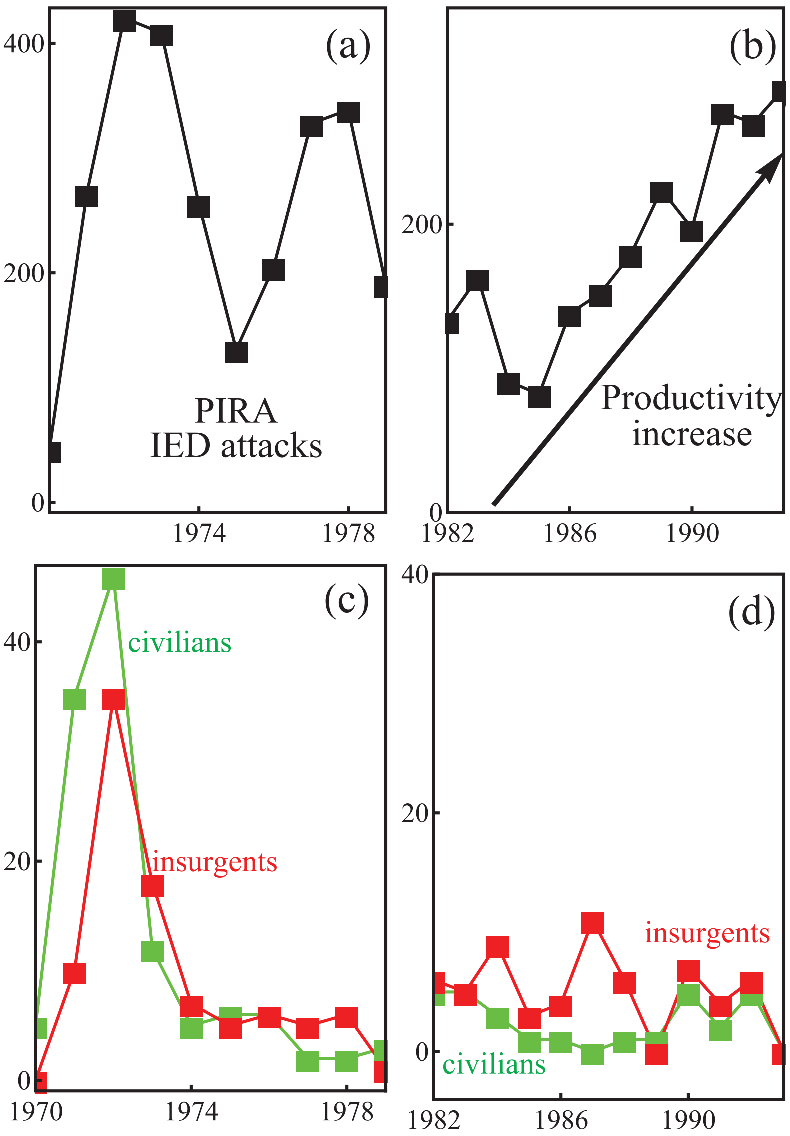

Figure 6.1: Number of PIRA IED attacks before (a) and after structural transition (b). Civil (green) and insurgent (red) casualties resulting from from British government activity, before (c) and after structural transition (d).

very intense, the IED attacks were sparse and highly fluctuating in both accuracy and frequency (Fig.6.1(a)). In addition, retaliation from the British government during that period (Fig.6.1(c)) was similarly intense which resulted on the largest number of casualties of both, civilians and insurgents, for the whole length of the conflict. By contrast, from 1982 onward, the trend of IED events shows a remarkable steady increase over time which speaks about a growth in the group's productivity within that term (Fig.6.1(b)). Moreover, as shown in Fig.6.1(d), effective activity from the British government decreased enormously when compared with the previous period. Also, government retaliation affected both, civilians and insurgents in a similar way. This talks about the level of uncertainty on government activities, and high secrecy of PIRA's actions.

An interesting question is: What are the key factors that contributed to this increment of productivity? Here, an initial answer is put forward by looking at the available data from PIRA's activities and members. Networks regarding PIRA's IED activity are aggregated over a one year period. Any two actors that were associated to an IED event have a link for that particular year. Individuals were identified from published news outlets such Irish Times and LexisNexis. First, global quantities are analyzed followed by specific network calculations on different aspects of member features. Agent-based models will subsequently provide insights about the transition on group's connectivity by reproducing the time evolution of basic network features as well as collective connectivity. Comparison with other networks will be presented at the end.

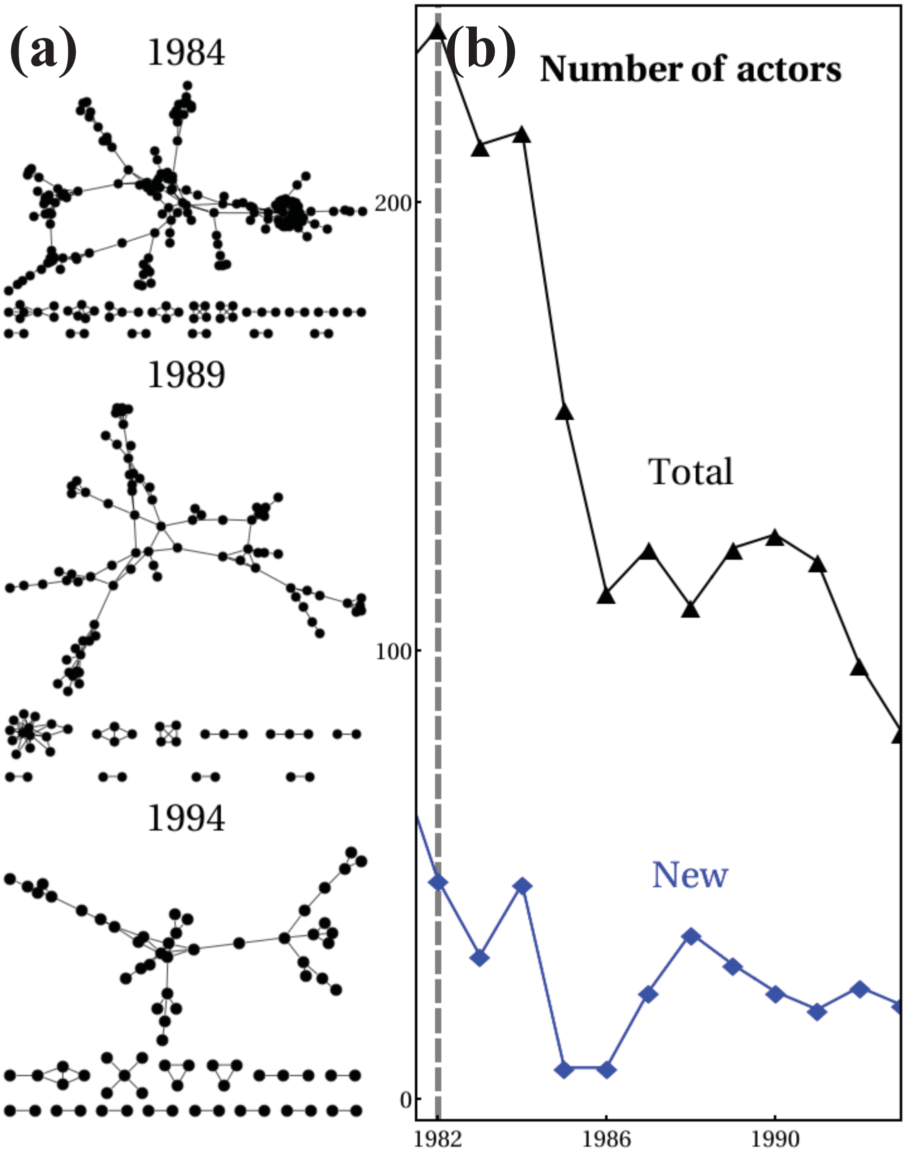

Figure 6.2: (a) Illustrative temporal snapshots of the PIRA network after self-organized restructuring at the end of 1970s. (b) Number of total (black) and new (blue) actors over time.

6.1.1 Global quantities

During the whole active period, PIRA always operated as a small group counting with 1240 members aggregated over time but with only, on average, 140 active members every year. This fact favored secrecy, given that individuals could be targeted by security forces and imprisoned, or killed by opposing paramilitary groups. Threats could also be extended to their families and children. Therefore, the risk was enormous and to keeping the number of members small was crucial [72; 71]. Figure 6.2(a) illustrates temporal snapshots of the PIRA network after the self-organized transition at the end of the 1970s. In addition to the large cluster, there are several small operational units that are evidence of the group's significant autonomy [72; 71; 173; 174]. Fig.6.2(b) shows the number of actors (total and new) that took active role in the group's activities. The decreasing trend in total members rejects a potential claim pointing to an increase in PIRA membership in order to explain the increase in PIRA productivity in Fig.6.1(g). Additionally, the recruitment (i.e. new actors) remains roughly stable in time which is not a clear indicator of increased productivity.

Every stage of PIRA's illicit activities was carefully prepared [70; 72; 173]. Specialized roles were assigned to particular individuals in order to trade arsenal and ammunition materials, manufacture explosives, select target and planting IED. Available information on PIRA actors allows for a comparison based on gender. Figure 6.3 shows the number of actors broken down by both, role (a) and gender (b). Fig.6.3(a) Illustrates that even though the role population fluctuates in time, the tendency where bomber dominates in number followed by gunrunner, gunman, bomb maker, officer and other is preserved. In addition, as shown in Fig.6.3(b) the network is dominated by men where only approximately 6%

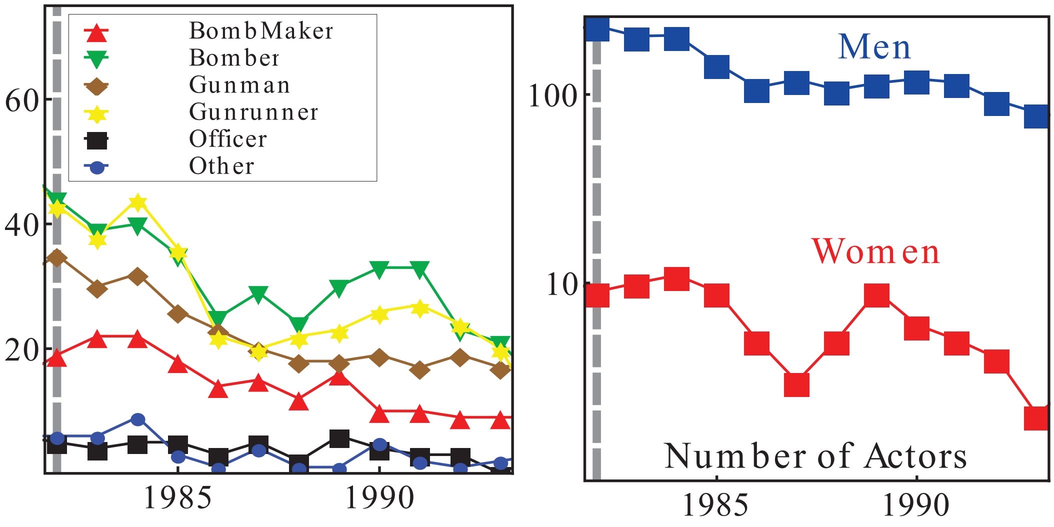

Figure 6.3: Number of PIRA actors over time per role (left) and per gender (right).

of its members are women. This general information does not provide any insight of the dynamical transition nor the increase in productivity.

6.1.2 Centrality measures

Degree and Betweenness

Given that there is no clear association of the number of actors, we look at the role of the node within the network by calculating centrality measures. In graph theory, centrality measures quantify the importance of a node or group of nodes. Since there are many possible way in which a node can be significant, there are several centrality measures for networks. *Degree centrality* measures the number of links that a particular node has with neighboring nodes. This is a basic indicator and is often used as the first step in network analysis. The degree centrality C_D associated to the node i is formally defined as [175]:

$$C_D(i) = \sum_{j \neq i}^{N} A_{ij} \tag{6.1}$$

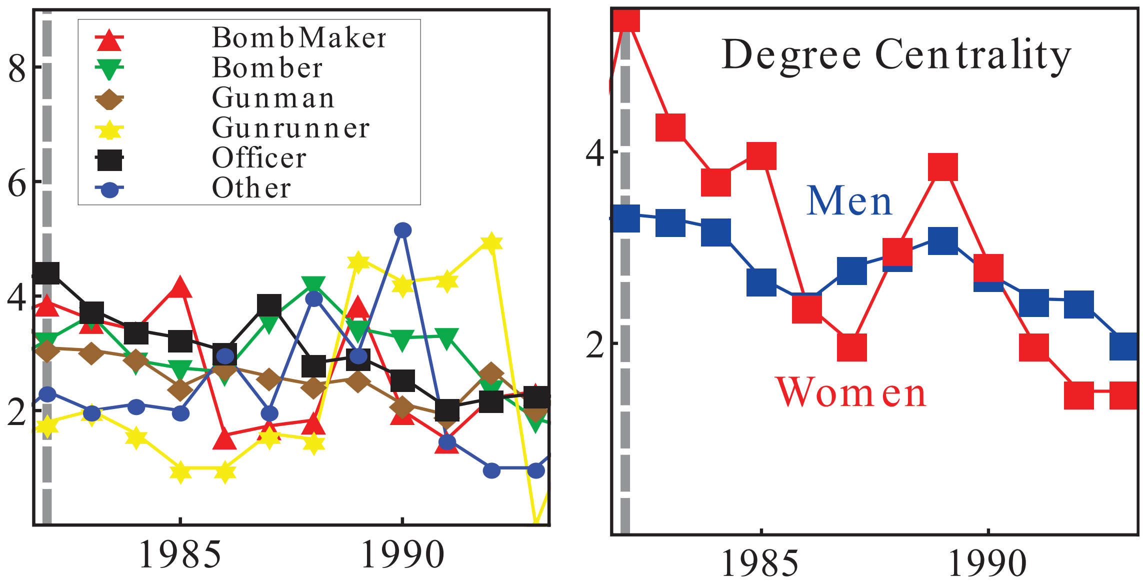

Figure 6.4: Degree centrality broken down by actor's role (left) and actor's gender (right)

where N is the total number of nodes and A_{ij} is the matrix element of the adjacency matrix A whose elements are defined as $x_{ij} = 1$ if node j is connected to node i, and $x_{ij} = 0$ otherwise. A high degree centrality speaks of the popularity or leadership that the underlying node has. It is also connected to the likelihood of neighboring nodes to be influenced by the high degree node. For instance, a scientific paper that has been cited several times presents a high degree centrality (more exactly a high in-degree since citation is a directed network). A straightforward extension of degree centrality is *eigenvector centrality* which measures the importance of a node by looking at the importance of its neighbors. Figure 6.4 shows the degree centrality

A crucial measure in covert networks is related to the capacity of a node to serve as bridge for communication, flow of resources or ideas, and brokerage. This measure is called *betweenness centrality* and measures the fraction of shortest (i.e. geodesic) paths

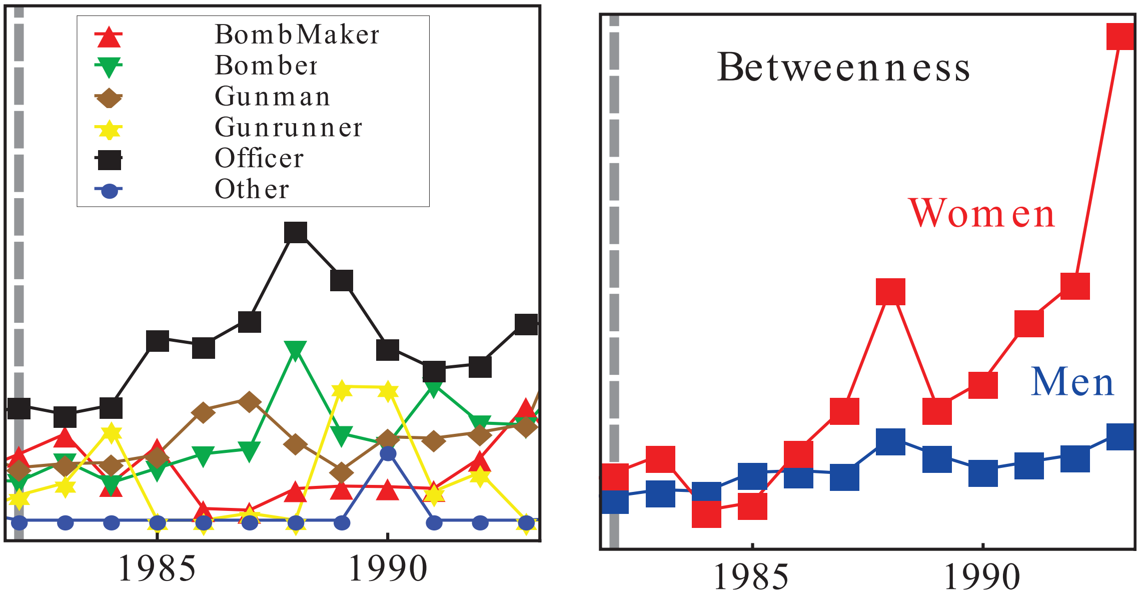

Figure 6.5: Betweenness centrality broken down by actor's role (left) and actor's gender (right)

that connect any two nodes that pass through a specific node. A particular actor with high betweenness centrality improves the channels of communication. Hence, its removal could cause the disruption of the network. The betweenness centrality C_B associated to the node i is formally defined as

$$C_B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{6.2}$$

where σ_{lm} is the total number of shortest paths from node l to node m and $\sigma_{lm}(i)$ is the number of these paths that pass through node i [175]. Betweenness is particularly useful for the analysis of innovation networks where secrecy and fast communication are crucial [172; 176; 177]. Figure 6.4 and 6.5 show results for degree and betweenness centrality broken down by actor's role and gender, respectively. For betweenness, it is calculated with the common normalization where the value is further divided by $(N-1)(N-2)/2$, where N is the size of the giant component of the underlying network. While degree centrality shows a rather noisy behavior with a not so well defined trend, betweenness centrality shows: (i) officers to carry a higher betweenness than other roles and (ii) a remarkable steady increment in women's betweenness centrality when compare to men's during the same period when PIRA's productivity in IED attacks increased (see Fig.6.1(b)). Although it could be argued that the increment in women's betweenness is connected to them being officers, the acquired data dispels that connection since (1) only one woman is identified as officer which is only present on one particular year (1992) and with negligible betweenness centrality and (2) women with high betweenness were not assigned to particular role, rather they are found to occupy several different roles within the network.

One could argue that the small number of women explains why they happened to present a high betweenness during that period, i.e. they were somehow exceptional and therefore more likely to be central. That argument was refuted by subject matter experts John Horgan and Paul Gills, who worked first-hand on the data collection project [70]. A random list of PIRA women was given to them with the request to rank them in terms of relative importance. Most women with high betweenness did not rank high on the expert's list. These are three examples of women with high betweenness centrality and did not appear on the expert's list. Only name initials are provided since some of them are still alive:

- MC: Age of recruitment 18. Occupation gunman. No university education. No notable other education. not married. No children. A simple Google search of her name, with or without PIRA added to the search, shows no hits connected to her

- DR: Age of recruitment 29. Occupation unknown (Other). No university education. No notable other education. not married. No children. A simple Google search of her name, with or without PIRA added to the search, shows no hits connected to her

- HR: Age of recruitment 70. Occupation housing meetings (Other). No university education. No notable other education. not married. No children. A simple Google search of her name, with or without PIRA added to the search, shows no hits connected to her
- FM: Age of recruitment unknown. Occupation unknown (Other). No university education. No notable other education. not married. No children. A simple Google search of her name, with or without PIRA added to the search, shows no hits connected to her

By contrast, several women with low betweenness centrality ranked high on the expert's list and appear on Google searches as former PIRA volunteers. Therefore, the hypothesis that women occupying a particular role or 'exceptional' (i.e. notorious to experts or web-search) are responsible for high betweenness centrality can be rejected.

Null hypothesis and fission/fusion model

In order to dispel the possibility that the results of high betweenness for women happened by chance, a null model is proposed. While keeping the network architecture and the total number of men and women intact, the gender of the nodes is randomly assigned. For each realization, the normalized betweenness for men and women is calculated and the procedure is repeated 10000 times. Then, the average and standard deviation is calculated and compared with the original dataset. The results are shown in Figure 6.6 where it can be seen that the null approach lies far from the original result. To further check this result a p value is calculated. It represents the probability that the test point is at least as extreme as the one generated by the null model. For this calculation it has been assumed that the null model

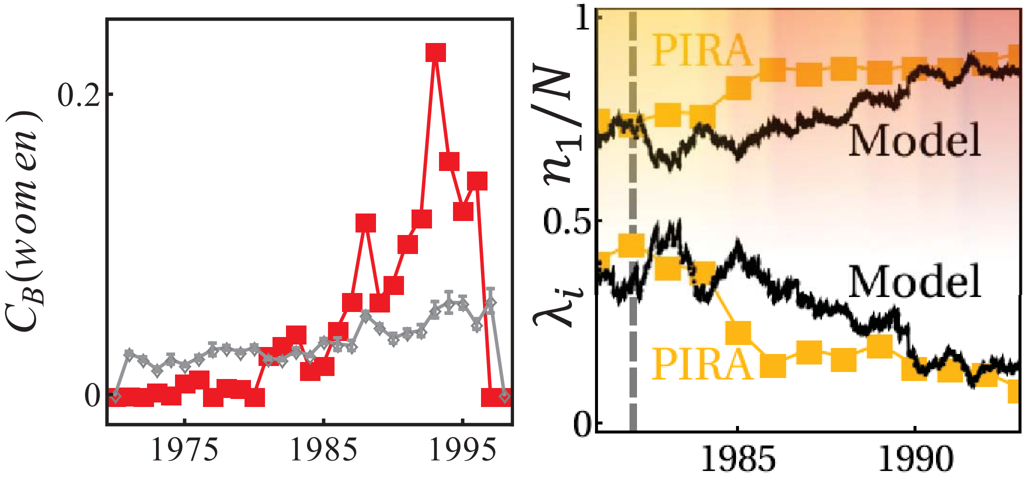

Figure 6.6: Left: contrast of betweenness centrality of PIRA women (Red) with null model (gray) and the corresponding averaged results obtained by randomizing the position of women in the PIRA network 10,000 times. The corresponding standard deviations are small while the mean expected values are distant to the observed values, allowing us to reject the hypothesis that the increase of women's centrality was a random process. Right: Generative fission/fusion model introduced in chapter 4 comparing network features such as number of isolated actors n_1 and mean degree λ_i. Shadow on top is the betweenness centrality of the seed of infection.

obeys a normal distribution (Normal p value). Calculations for p values for each year yield to numbers smaller than 0.05. By convention, this indicates that the null hypothesis can be rejected and hence the high centrality for women did not happen by chance.

We further analyzed the fission/fusion model introduced in Chapter 4 which reproduces the non monotonic temporal evolution of connectivity features such as fraction of isolated actors n_1/N and mean degree λ. The results of betweenness motivates the formulation of the hypothesis that women introduced a team-oriented type of connection different from the homogeneous kinship-like of men. In addition, it is assumed that men actors are susceptible to women's team culture. Once a connection is established, the team-like culture is preserved. These assumptions are go in line with the model's predictability of a high betweenness centrality of the seed which spreads team-oriented group formation. Figure

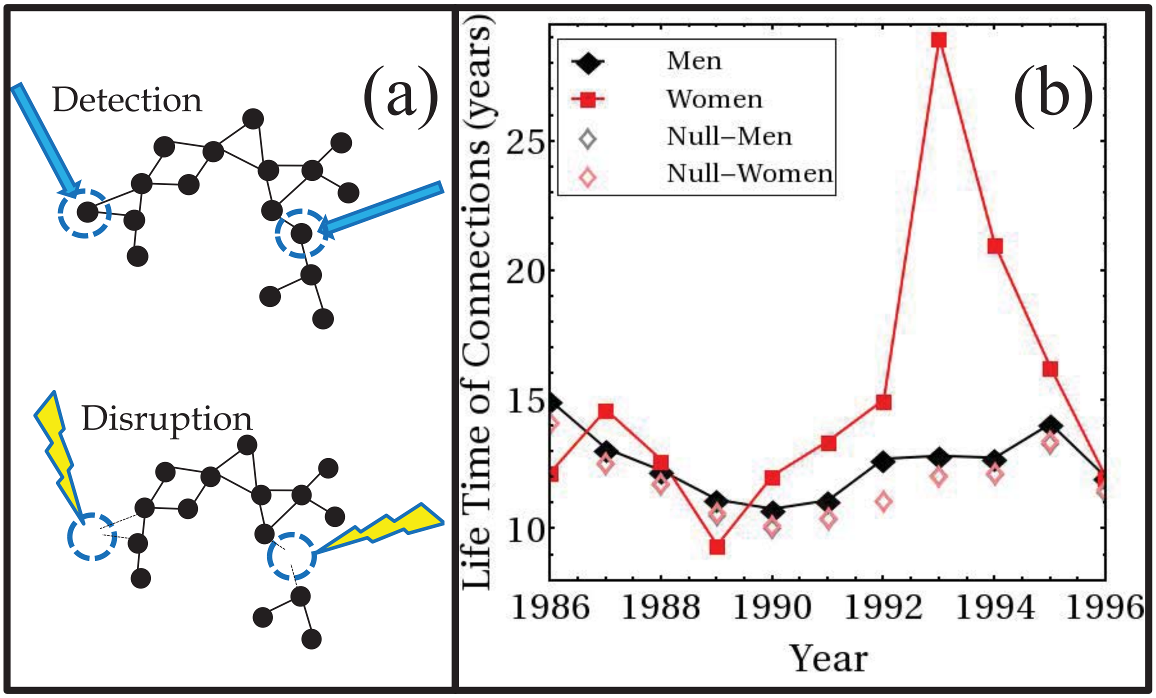

Figure 6.7: (a) Main mechanism for death of an individual is due to individual targeting by an opponent. (b) Statistical association whereby actors directly connected to women have a longer lifetime on average, i.e. they survive longer. This is in contrast to the null model results as shown, where the gender of the actor was scrambled and the average (lighter symbols) over 5000 realizations were calculated

6.6 illustrates the results of the model where betweenness of the seed is represented by the shadow, in good agreement with the findings of PIRA. Therefore, the results suggest an innate ability of women to mediate teamwork. This allows them to, on average, assume more central positions than men in the underlying network. The importance of a team culture compared to kinship is consistent with studies of online covert networks in online gaming [178], which also allows gender anonymity. The simulations also support a bottom-up transition which otherwise would be costly by manipulating $\sim N^2$ links. This is highly unlikely for an organization who wants to remain under the radar.

Analysis at a local-level points to a benefit that self-organized positioning of women, on average, produce to its nearest neighbors. The main mechanism for death of an individual in an extreme network under pressure is being targeted by an opponent. Figure 6.7 shows that women's connection have superior longevity to men's which is consistent with women being better embedded in the network. In addition, the results of the null model where the actor's gender is scrambled over 5000 realizations, demonstrates that it did not occur by chance.

6.2 Comparison with other networks

6.2.1 Academia and industry

The innovative skill developed by PIRA in which women appear to be in more central positions than men could be contrasted with other gender studies. In particular, we look at the work of Prof. Kjersten Whittington at Reed college which studied academic and industrial networks, environments where innovation is essential. For university and biotechnology firm data, we utilized the Boston subset of 89 inventors at top tier US universities and inventors from 482 biotechnology firms [179]. Covering a period from 1980 to 2000, actors were connected if they filed for a patent together. The percentage of women in the university and biotech-industry dataset was roughly 27% and 34% in a given year. A focus on gender leads to Fig. 6.8, with women emerging as significantly more centrally located than men in the covert PIRA environment. Determining the ratio of the mean betweeness centrality b of women and men, $r = bw/bm$, we observe that the women's average betweeness centrality eventually rises to more than four times the men's corresponding value. In the two non-covert innovation environments, the ratio fluctuates around $b_w/b_m \sim 1$ for industry, while men occupy more central positions than women in the university setting

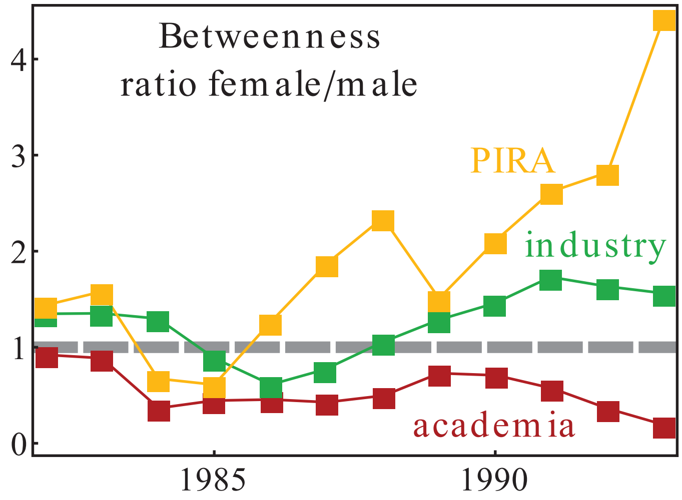

Figure 6.8: Ratio of betweenness centrality of women b_w and men b_m, where $r = b_w/b_m$ (vertical axis), for PIRA (yellow), as compared to industry (green) and academia (red) over the same period.

$(b_w/b_m < 1)$.

Indeed the academic and industrial environment are very different from that of an illicit group. Academia is highly hierarchical and patent production is usually linked to individual faculty contributions, which suggest that men are more centrally located than women (as seen in Fig.6.8). By contrast, biotechnological companies are less individual-oriented and the innovative success is attributed to the company. This suggests a team-oriented approach which we have hypothesized to be driven in higher likelihood by women, and is interestingly supported by a higher women's betweenness as shown in Fig.6.8. Though, there are indeed several other sociological factors that influence the positioning of individuals on the networks, we note that for these examples (academia and industry) the secrecy, level of aggression and physical danger are far smaller than in PIRA. Hence, we suggest that women are likely to emerge with superior network connectivity in extreme environments that offer benefits such as survival. In addition to PIRA, the next example provides support to this claim.

6.2.2 Online support network of Islamic State

There has been an unexpected online support for Islamic State in social media.Though several media sites such as Facebook turn down any pro-ISIS activity quickly, other social websites act slower. This is the case of VKontakte (www.vk.com) where our team within the complexity initiative has been able to monitor the evolution of pro-ISIS groups until they are detected and shut down (Fig.6.10(a)). The network is comprised by $> 10^5$ individuals, which are free to provide gender information, that share pro-ISIS material such as videos, prayers, PDF files, etc. Given that the network is bipartite (follower $\rightarrow$ group) we analyze the unweighted projection over the followers, i.e., followers of a particular group form a fully connected cluster among them.

The summary of the findings for online pro-ISIS network is presented in Fig.6.9. Daily snapshots of the network re shown in Fig.6.9(a-c). We looked at betweenness centrality (Fig.6.9(d)) over time where women, despite of the fluctuations, are found to be on consistently higher betweenness position than men's (showed by orange rings). We attributed fluctuations to the dynamics of followers entering and leaving groups as well as the sudden death of a particular group due to a shut down. In addition, by looking at the aggregated network over two months, women's degree centrality (vertical dashed line) is also higher than men's and it distances for more than 4 standard deviations from a null model (histogram) where nodes' gender are scrambled over 10^5 realization (Fig6.9(e)).The ratio of women to men followers is higher than PIRA (~ 0.7) hence the potential claim that the

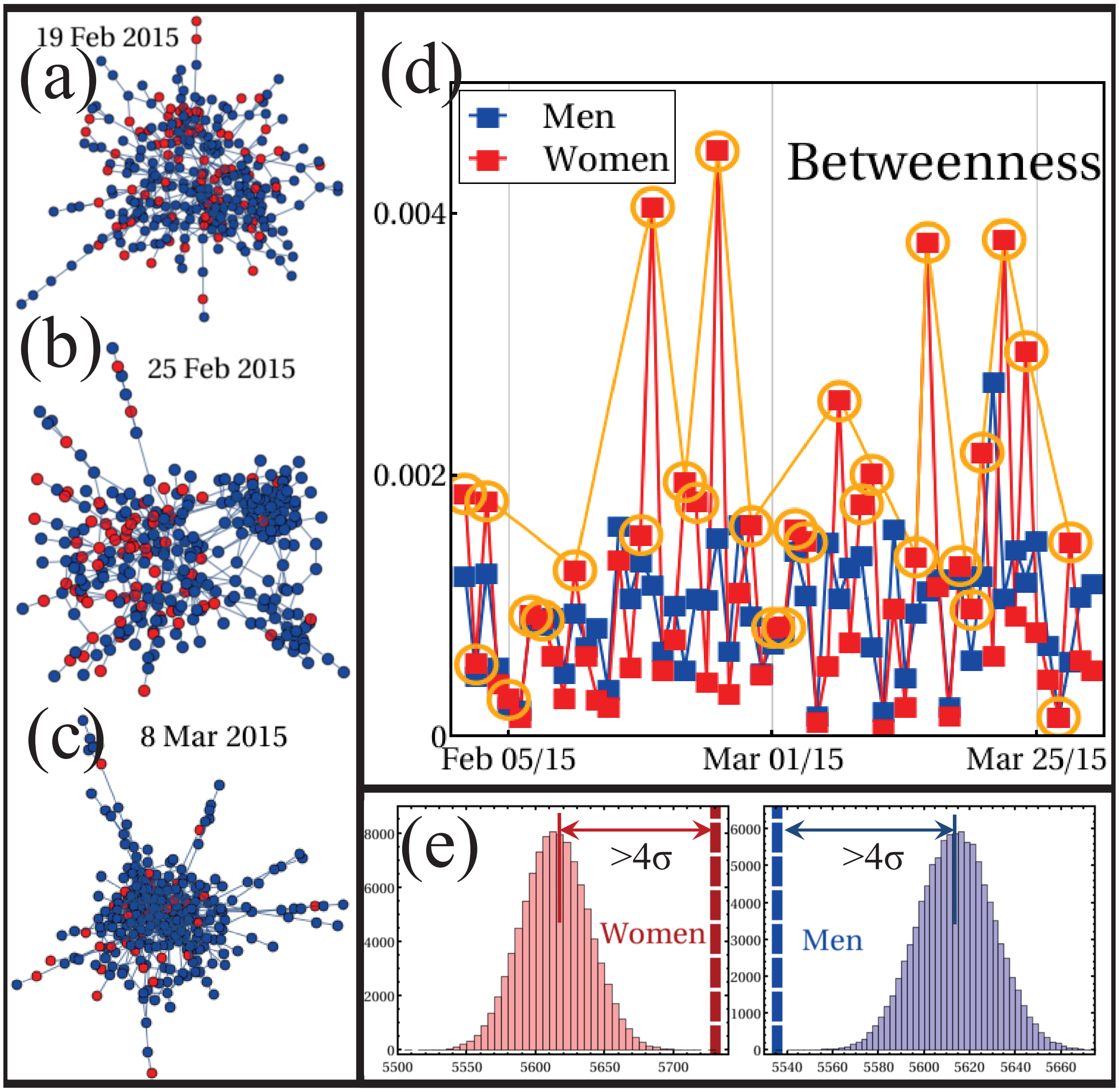

Figure 6.9: (a-c) Illustrative temporal snapshots of a subset of the global online network containing $> 10^5$ pro-ISIS followers. Followers (circles) aggregate spontaneously around online pages such as organizational accounts (squares), e.g. on Vkontake (www.vk.com) as shown. (d) Betweenness centrality (BC) over time for women and men in follower network. Orange rings show days when betweenness centrality of women b_w is superior to that of men b_m. Orange lines connect the subset of these days where signal exceeds noise (¿σ). (e) Left: Degree centrality of female followers (vertical line) in time-averaged follower network is more than 4σ (i.e. Z¿4) larger than the distribution of null model results in which the gender of all nodes is repeatedly scrambled. Right: Degree centrality of male followers (vertical line) is more than 4σ (i.e. Z¿4) smaller than the null model result. For subset of followers without declared gender, which likely contains a similar number of women and men, the degree centrality is statistically indistinguishable from the null model result, as expected.

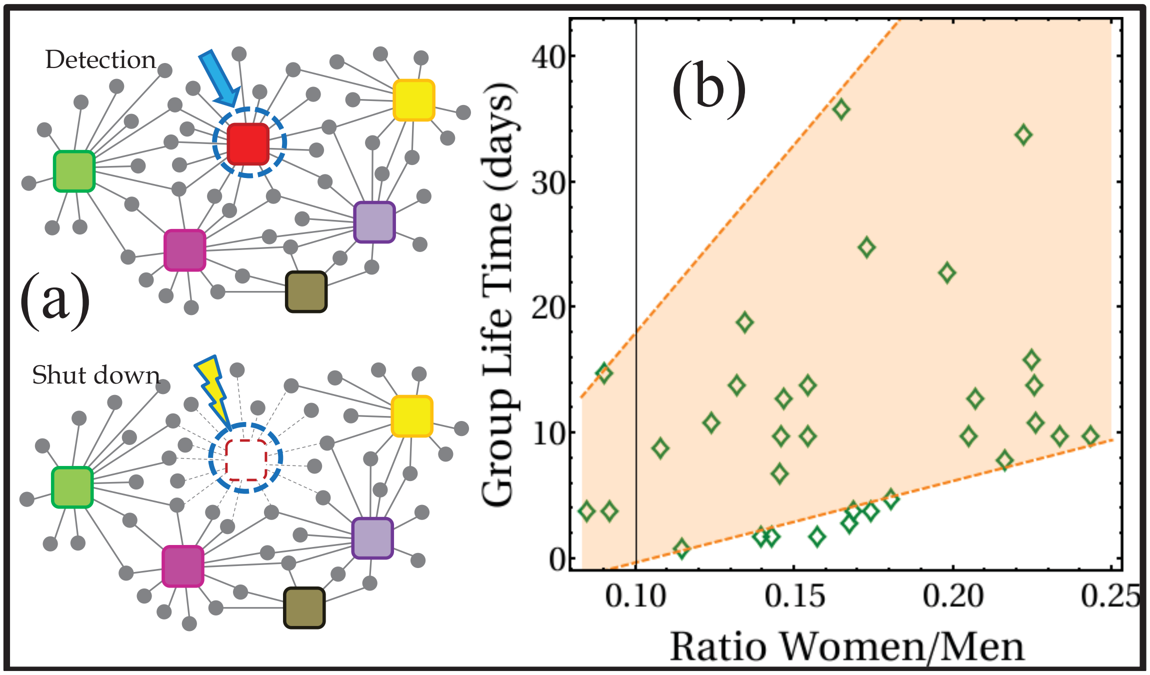

Figure 6.10: (a) Main mechanism for death of a group in an extreme network under external pressure (e.g. ISIS or PIRA) is fragmentation due to being detected and hence shut down by an opponent. (b) As the ratio of women to men increases, there is an association with an increase in the group lifetime in ISIS, i.e. group survives longer

fewer females makes them exceptional can be dispelled. The network's size and density (41 thousand of followers creating nearly 130 million links) makes it nearly impossible for an up-to-bottom coordination to manipulate the structure.

A global-level advantage afforded by women's position was found when looking at the lifetime of a group. Online pro-ISIS groups where the ratio of women-to-men is high, tend to survive longer which can be associated to the local-level advantage found in PIRA where women's connections tend to survive longer than those of men's. The results are presented in Figure 6.10(b).

6.3 Summary of chapter 6

We analyze different dynamical networks (off-line and on-line) using several quantitative tools. Global features, centrality measures and agent-based models were used to analyze the internal dynamics and outcome (i.e. IED activity) of PIRA network. The findings point to a transition to a self-organized network where a female minority collectively emerged with higher connectivity than men. A fission/fusion model is able to capture the non monotonic behavior of some network features by gradually introducing a team-oriented culture with an infection-driven mechanism over a more homogeneous kinship culture. The model captures an increasing betweenness centrality of the seed which supports the hypothesis that woman are more likely to embrace and spread a team-oriented behavior. The overall results revealed that, although men dominate numerically, women emerge with higher connectivity than men when the environment is extreme. This collective feature is associated with benefits to the network in their local and global level in terms of longevity. The results were contrasted with null models which rejected the hypothesis that the positioning of individuals within the network happened by chance. These findings suggest that female-centric approaches can be adopted for disrupting extreme networks, by focusing on the interconnectivity of these few women as opposed to the majority of men.

Chapter 7

Anomalous contagion in dynamical networks

Despite the high diversity of epidemiological models [78; 82; 180; 181], there is little quantitative understanding on how the dynamics of individuals affect the overall features (e.g. severity, duration, etc.) of outbreaks. Unexpected rise on activities such as online support for social and military causes (e.g. ISIS), on-street giant protests at the level of a country, sudden changes on global currency trade prices, along with others, exhibit profiles that hardly resemble traditional approaches (e.g. SIR model)[182]. This creates a need to explore additional mechanisms which accounts for unexplored aspects of modern-day human contagion.

The digital era opened a new level of human interaction which does not require individuals to be at the same physical location to reach out. Chat rooms, websites, cell phones, among others, allow a temporary connection among individuals and hence transmit information (e.g. rumors, ideologies) that could result into outbreaks of collective activity. Online social media has become an excellent vehicle for promoting, planning and scheduling large mass events. Individuals entering and leaving online sites continuously throughout the day, create communities whose nodes change frequently in time [78; 79; 183; 184]. During the online period, nodes are exposed, increasing the chance for infection. However, there is the off-line period in which the individual is exempt from external influences. The on-line/off-line dynamics becomes a key mechanism that provides explanation to modern-world infection patterns.

In this chapter a new model of contagion is presented. It accounts for the dynamics of individuals on a common space where contagion is possible and an outside region where they are unexposed to the infection. The addition of the dynamics to the contagion model is able to reproduce specific infection profiles' features that traditional approaches (e.g. SIR) fail to explain. In particular, the results reveal a non-linear dependence of dynamical features (e.g. mobility) with extensive observables such as duration and severity. Interestingly, the model predicts that the severity of an infection could be reduced by increasing the mobility of the individuals which challenges traditional understanding of epidemiological processes. The model is tested against real-wold data from different domains such as online pro-ISIS groups, on-street civil unrest protests in Latin America and global online currency trading. Data from pro-ISIS online communities is the result of efforts of the complexity initiative at University of Miami, which both manual and computer-based techniques were used. First, online groups were identified manually through the VK (http://www.vk.com) search options such as hashtags with key explicit supporting words (e.g. ISIS, islamicstate, khilafah, caliphate, fisyria, etc.) and/or publishing ISIS-related propaganda. Subsequently, information of the followers was used in order to capture the extent of the network and hence the extent of the infection spreading. Later, the resulting data was cross-checked manually by subject matter experts which excluded non related groups and followers that were not explicitly supporters of ISIS. The process was then automatized and performed on a daily basis at the same time of the day. The data set counts for ~ 40000 individuals at any one time with ~ 130 million links in less than two months. Data for civil unrest in Latin America is discussed in Ref [185] while data of currency exchange is discussed in [82].

The chapter is divides as follows: Section 1 is devoted to the basic description of the model. An initial comparison with the different real-world data sets is then presented. In section 2 extensive features of the infection profile (e.g. severity and duration) are contrasted with those of the data sets highlighting the lack of explanatory power of traditional approaches such as the SIR model. In section 3 the model is expanded to two connected communities expanding the boundaries of the single community model. Finally, a comparison based on electric circuit is developed while looking for similarities and generalizations for multiple connected communities. A summary of the chapter is presented at the end.

7.1 Single CSC model

The Common Space Contagion (CSC) model is a generalization of a standard SIR process to account for objects entering and leaving a dynamical group (common space) where they could become infected. The SIR model describes the spreading of an infection over a population over time. SIR stands for *Susceptible* (S), *Infective* (I) and *Recovered* (R), which are the three possible states that an object (agent) within a population can have. The mechanism of infection is through direct contact between a susceptible agent and an infective agent with a probability q_i per timestep. Once an agent is infected, there is a probability q_r to recover per timestep and then it becomes immune i.e., it can no longer get infected. It is convenient to define an infection's contact rate as $\lambda = q_i/q_r$. Every object that is outside of the group has a probability per unit time p_j to enter into it. Similarly, objects inside the group have a probability per unit time of p_l to leave the common space (see Figure 7.1[a]). The equation of motion that describes the population dynamics within the common space, where N_g is the number of objects inside of the group at a given time t, can be written as

$$\frac{dN_g}{dt} = -p_l N_g + p_j(N - N_g). \tag{7.1}$$

In the steady state ($dN_g/dt = 0$), the mean number of agents within the common space $\langle N_g \rangle$ is given by

$$\langle N_g \rangle \;=\; N\frac{p_j}{p_j + p_l} = N\gamma_s \tag{7.2}$$

In addition, the mobility μ, which is the sum of agents joining and leaving the space in a unit time can be calculated as

$$\begin{aligned} \mu &= (N - \langle N - g \rangle)p_j + \langle N_g \rangle p_l = N[(1-\gamma_s)p_j + \gamma_s p_l] \\ &= N\frac{2p_l p_j}{p_l + p_j} = N\gamma_m \end{aligned} \tag{7.3}$$

where $\gamma_m = \mu/N$ is the mobility per unit object. The single CSC model integrates the SIR model with the dynamical component. Consider N agents (initially set as S) joining and leaving a common space with probability p_j and p_l, respectively. The agents outside the group are isolated, while the agents inside are able to interact with one another (e.g. fully connected network). Once the system of susceptible attains the steady state, an agent is randomly selected within the group and it becomes infected. Consequently, the infected agent is able to poison other agents within the group with a probability q_i as well as get recovered with a probability q_r. Under these circumstances, only within the common space, the infection can be transmitted to other individuals, while the recovery process takes place everywhere. The dynamics of the model is studied by a discrete (simulation) and continuous (differential equations) approach.

7.1.1 Simulation and dynamical equations

The simulation is a discrete time description of the model. It considers that, at each time step, all the agents undergo an SIR process, followed by the dynamics of leaving and joining. The number of susceptible, infected and recovered agents is traced over time. When the number of infective agents drops to zero, the epidemic ends. On the other hand, the differential equations consider a continuous time and continuous population approach of the system. We have six variables since we describe the populations inside and outside the common space. We use $S(t)$, $I(t)$ and $R(t)$ to be the number of susceptible, infected and recovered agents in the whole system, respectively and $s_g(t)$, $i_g(t)$ and $r_g(t)$, to be the respective quantities inside the common space. The differential equation for each of these quantities is given by:

$$\begin{aligned}
\frac{dS}{dt} &= -q_i s_g i_g \\
\frac{dI}{dt} &= q_i s_g i_g - q_r I \\
\frac{dR}{dt} &= q_r I \\
\frac{ds_g}{dt} &= -q_i s_g i_g - p_l(s_g - q_i s_g i_g) + p_j(S - s_g) \\
\frac{di_g}{dt} &= q_i s_g i_g - q_r i_g - p_l(i_g + q_i s_g i_g - q_r i_g) + (1 - q_r) p_j (I - i_g) \\
\frac{dr_g}{dt} &= q_r i_g - p_l(r_g + q_r i_g) + p_j((R - r_g) + q_r(I - i_g)) \qquad (7.4)
\end{aligned}$$

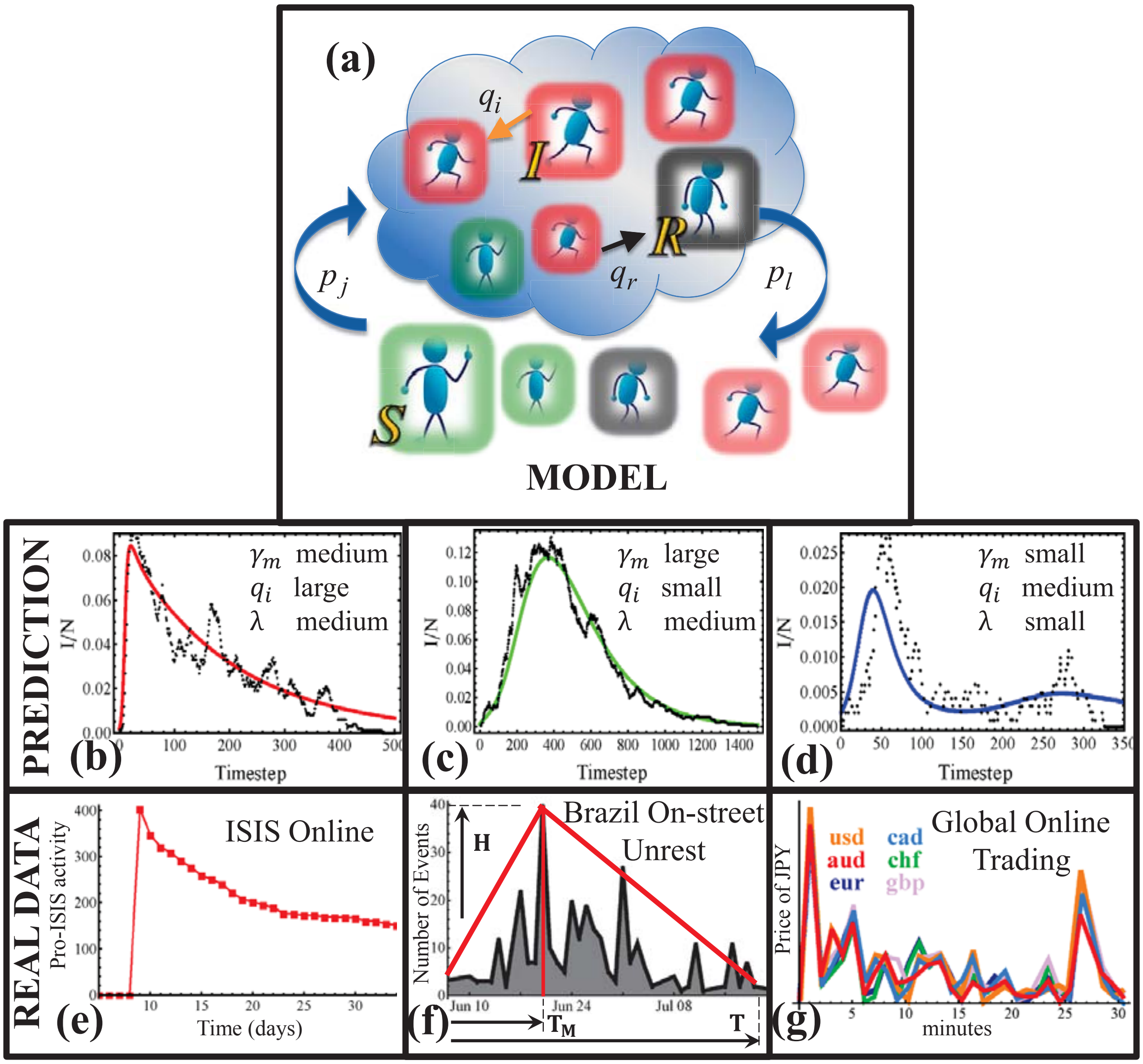

Figure 7.1: Outbreak profiles in real-world systems. (a) our common space theoretical model. Common space (blue cloud) can be an online site or chatroom, a network community or group, a physical place etc. and the transmission can range from information or rhetoric to a real virus. An individual outside (or inside) the common space has a probability p_j (or p_l) to enter (or leave) at each timestep. Infected individuals (i.e. activated: red) inside the common space have a probability q_i to infect other susceptibles (green) inside the common space. Infecteds both inside and outside recover (black) with probability q_r. $\lambda = q_i/q_r$. (b)-(d) illustrates the qualitatively distinct outbreak profiles predicted by our model, with the corresponding parameter regime. Black line is simulation, colored line is from integrating the coupled differential equations (see SI). (e)-(g) shows how these theoretical profiles capture various modern-day outbreak profiles: (e) pro-ISIS activity online 2014 matches (b); (f) protests on-street in Brazil 2013 matches (c); (g) global online currency trading during transmission of rumour of re-evaluation of Yuan, matches (d). Profile shows the variation of all major currency exchange rates[92]. For (e)-(g), similar profiles appear repeatedly in our datasets (see Fig. 2) confirming that the $I(t)$ variation is a reproducible signal, e.g. for (g) an almost identical profile occurred several months later when the same rumour circulated again[82; 92].

Figures 7.1(b-d) illustrates the dynamics of the infective population for three set of microscopic parameters for a total number of 1000 individuals. The smooth colored lines result from the numerical integration of the differential equations (7.4), while the black curves represent a typical time discrete simulation. As shown, the results of simulations are in good agreement with the differential equations. The infection dynamics reveals qualitatively different behavior depending on the parameters. The red curve (Fig.7.1(b)) undergoes a rapid increase followed by a slow decay. The green curve goes through a gradual increase and a slow decay (Fig.7.1(c)), and the blue curve (Fig.7.1(d)) experiences a revival after the gradual rise and decay. Figures 7.1(e-g) shows typical infection profiles from real-world systems. As shown, they do not exhibit an exponential decay as in SIR type of models together with features such as slow increase and revivals. Interestingly, the sensitivity of the CSC model serves to reproduce these features by modifying the infection and mobility parameters. Figure 7.1[g] (match of Fig.7.1(b)) illustrates the pro-ISIS online activity of a typical group that explicitly express support for and circulates pro-ISIS related documents (e.g. news, propaganda). Even though not all the members are active (infected), the actual number of infected is proportional to the group's size where all the members present the same probability of becoming active. Figure 7.1(f) (matching Fig.7.1(c)) shows off-line activity in Brazil starting in mid-June of 2013. The vertical axis quantifies the number of on-street protests taking place in several cities and supporting a wide variety of civil-unrest causes [185], where the peak of infection occurs nearly at the center of the burst. Fig.7.1(g) shows online global currency trading rates through the course of rumor spreading of Yuan re-evaluation[82]. Time for peak, duration and revival are in good agreement with the parameter set of the CSC model as shown in Fig.7.1(d).

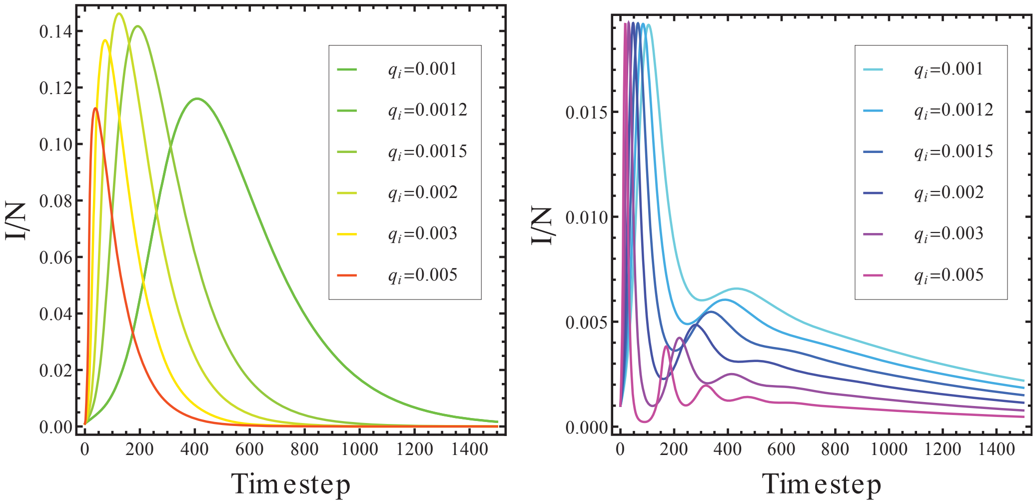

Figure 7.2: Infection profiles form the numerical integration of the equations (7.4) for different values of infective probability q_i. Left: $\gamma_m = 0.018$ and $\lambda = 0.1$. Right: $\gamma_m = 0.0018$ and $\lambda = 0.022$.

Additional observations of the model's infection profiles allow us to understand how the microscopic parameters affect the dynamics. For instance, the infective probability q_i is directly related with the growth rate of the infected population. This is supported by Figure 7.2 which shows the dynamics of the infective population for different values of infection probability q_i, and for two sets of the parameters γ_m and λ. The outcome confirms the previous affirmation about the parameter q_i. Interestingly, the high growth rate does not imply a high infection peak. In fact, as shown in the left panel of Fig.7.2, there is an optimal value of q_i that results in a highest infected peak. Another key point is that the total duration of the infection gets longer for a small infection probability as well as its time to peak. Equally important, the oscillatory behavior is exclusive for small mobility γ_m and small infection contact rate λ. This is depicted in the right hand side of Fig.7.2, where the frequency of the oscillation increases with the infection probability.

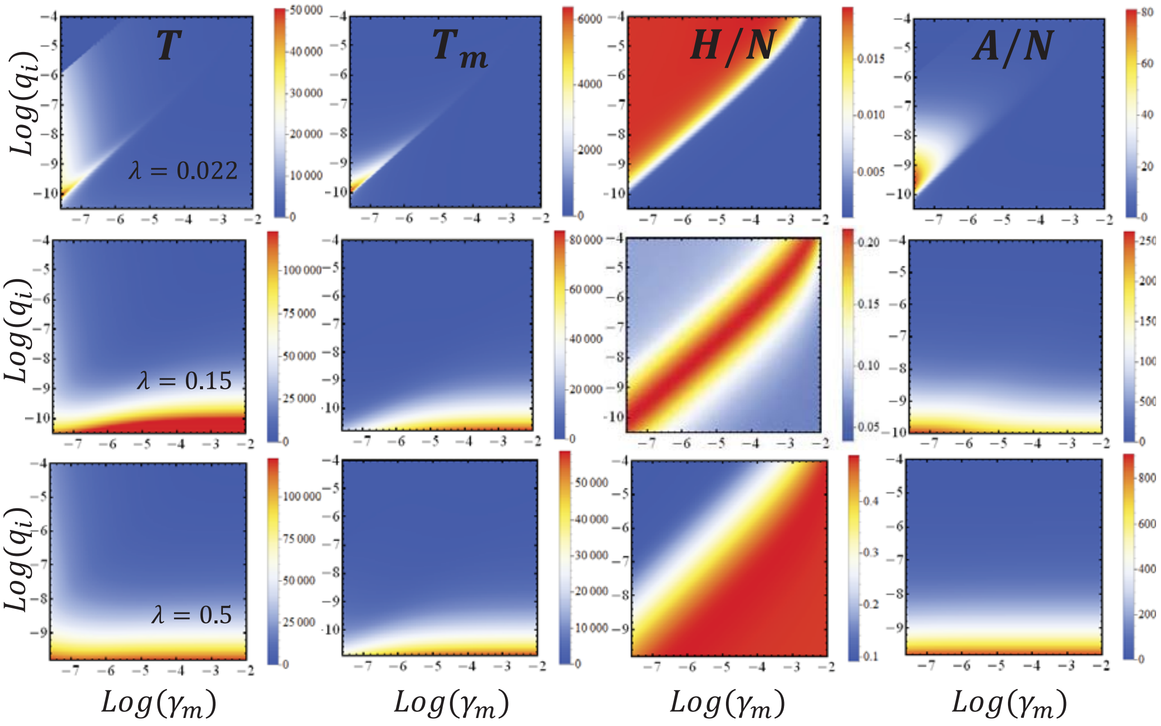

Figure 7.3: Extensive quantities of the infection profiles (from left to right: T, T_m, H/N and A/N) as a function of γ_m and q_i. For $\gamma_s = 0.1$ and four values of λ (from top to bottom: 0.022, 0.15, and 0.5).

7.2 Profile features: duration, time to peak and severity

We characterize the profile differences by looking at *extensive* features of the infection profiles. This becomes particularly useful when comparing with real complex systems where the information of the microscopic parameters is usually unknown. We set T to be the duration of the epidemic, H to be the peak of infection i.e., the moment where the number of infected is maximum, T_m to be the time where the maximum (i.e. H) is attained, and the area below the curve of the infection dynamics, denoted as A. Fig.7.1(f) illustrates these macroscopic quantities associated to an infection profile. Figure 7.3 shows the behavior of the extensive quantities by integrating numerically Eqs.7.4. Profile features are shown as a function of the mobility γ_m and infection probability q_i for six different values of the infection's contact rate λ. The relationship between the duration, time to peak, and area becomes evident by showing the similar qualitative results for a given value of λ. These are some key points to highlight: (1) As λ grows, the times (duration and time to peak) and area become independent from mobility. (2) As q_i increases, the times and area become smaller. (3) By increasing the parameter λ the maximum height grows. (4) The highest severity value of H shows a linearity with γ_m and q_i (i.e., $q_i = e^3\gamma_m$). (5) The regions on the q_i-γ_m space where the maximum height is located change from low mobility and high infection probability, for small λ, to the region of low infection probability and high mobility, for large λ. The transition between these two limits can be seen to occur around $\lambda = 0.15$. (6) For small values of λ, the times and area follow the linearity on their maximum value with γ_m and q_i. Before the transition point at $\lambda = 0.1$, the linearity is lost.

7.2.1 Effects of the initial seed

As mentioned above, the initial conditions consider an infected individual on the common space where the spreading process initiates. However, in real systems some infections are controlled or naturally dissipated before a large-scale spreading is reached. The numeric simulation accounts for these types of situations. Figure 7.4 illustrates the distribution of infection's duration for 1000 different realizations and different initial conditions which varies in the number of initially infected objects (seed s). Each run leads to a slightly different dynamics whose mean values are well captures by the equation of motion. For small values of s, the probability of having a short outbreak ($T \approx 0$) is very high in comparison with realizations for larger values of s. For this illustration, the recovery probability is selected to be approximately 50 times greater than the infection probability. Hence, the

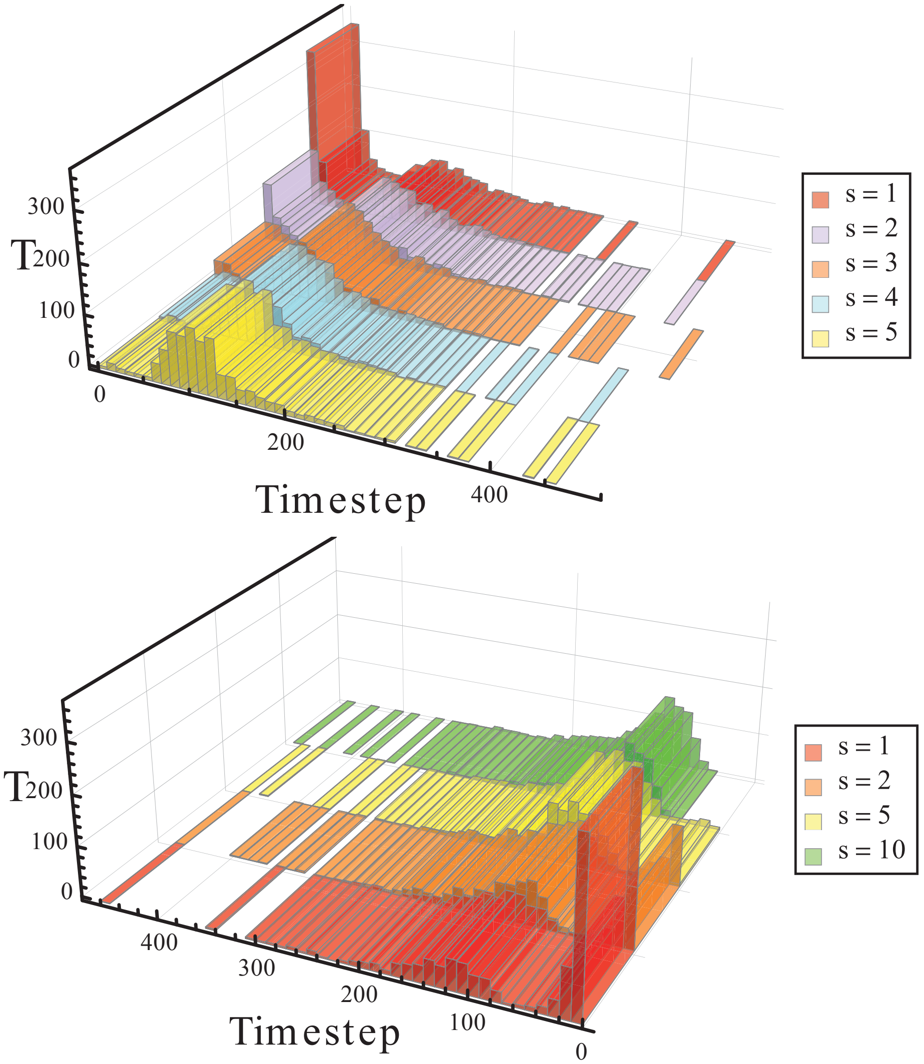

Figure 7.4: Distribution of duration of infection for different values of initial seed s. Parameters: $\lambda = 0.022$, $q_i = 0.002$, $\gamma_m = 0.0018$ and $N = 10^3$.

distribution for small s shows a large probability of a short infection, i.e., for most of the runs, the few infected agents get recovered faster than they can spread the infection. On the contrary, as s increases, the probability of short durations decreases and the distribution gets populated on a duration that is similar for all the values of s. Interestingly, the point where the distribution is maximum (after the short duration peak for small s) is only slightly shifted to shorter times as s grows. It becomes more evident in the right hand side of figure 7.4 by looking at the difference between $s = 5$ and $s = 10$. This time shift can be explained as the time interval where the $s - 1$ agents got infected from the first infected agent. The duration that resembles the result from the differential equations which is around $T \approx 400$ (see blue curve in figure 7.1), has a very low probability for all s values. In turn, for this selection of parameters, the simulation is far from the mean field approach i.e., differential equation result.

7.2.2 Contrast simulation and differential equations

As an illustration, figure 7.5 depicts the result for the extensive quantities of the infection profile as a function of mobility contrasting the results from the differential equations (solid curve) with the mean value from the simulations (dotted curve). The results show that for small values of mobility, the duration predicted by the differential equations is greater than the mean from simulation, in agreement with the previous finding. Interestingly, the affirmation is valid for the area and time to peak but it is false for maximum height. The latter displays, for small γ_m, a good agreement between the simulations and the equations. In contrast, as the mobility is increased the previous affirmations are no longer accurate for the duration and maximum height. For instance, the simulation result grows with a smaller

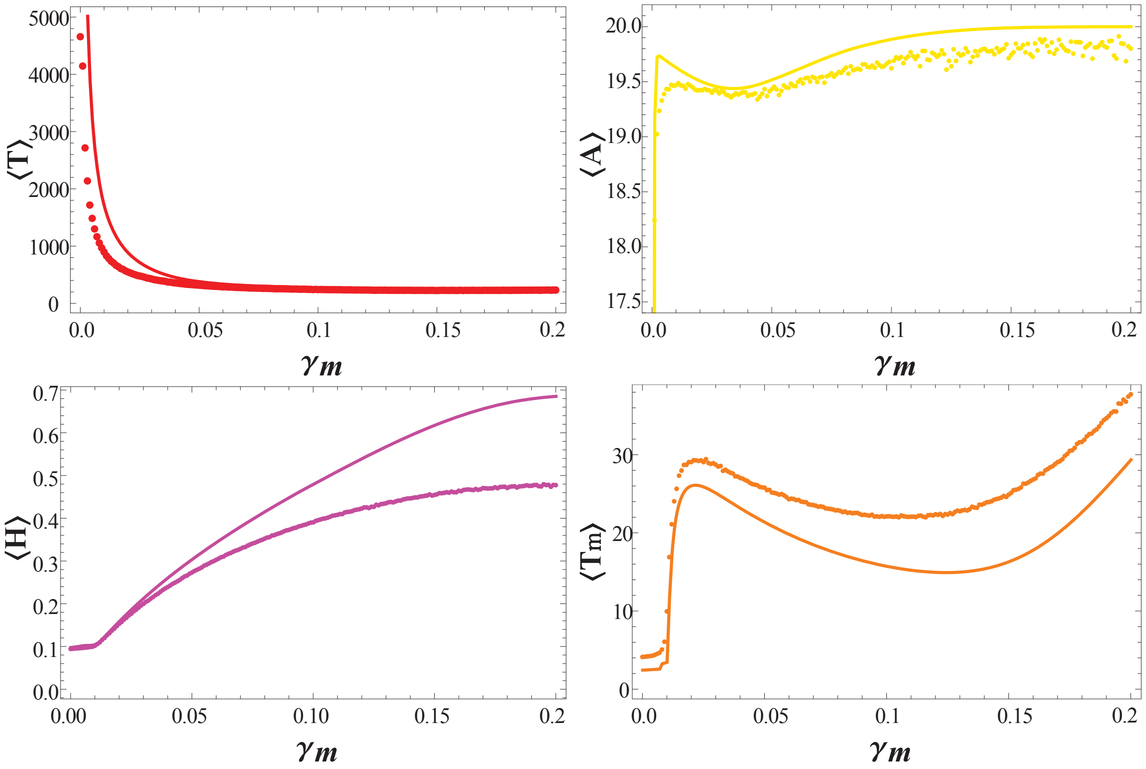

Figure 7.5: Extensive quantities as a function of mobility from numerical integration of differential equations (solid curve) and mean values of 10^4 simulations (dotted curve). The quantities are duration (top left), area (top right), severity (bottom left) and time-to-peak (bottom right). Parameters: $\lambda = 0.1$, $q_i = 0.005$, $\gamma_s = 0.1$ and $N = 10^4$.

rate than the differential equation for the maximum height while the agreement between the results for duration grows as the mobility is increased.

7.2.3 Comparison with empirical outbreaks

The mobility γ_m introduces non linear effects to the infection profiles different from commonly used zero-mobility or well-mixed approximations. These effects allow to remarkably shape different modern-day contagion phenomena as shown briefly on Fig.7.1. A larger illustration of this statement is shown on Fig.7.6 where modern-day outbreaks are compared with theoretical results. Triangles represent the empirical data of online pro-ISIS activity. While circles depict civil unrest outbreaks in Latin America between 2012 and 2014 [10]. The black thick curve is the result from standard SIR model which shows a limited range of applicability. By contrast, the results from the model (colored curves) cover a much wider range than SIR, which reproduces most of the empirical outbreaks. Each curve is generated by a different value of mobility which increases the diversity in the infection profiles.

7.3 Multiple CSC model

Here, a simple extension of the model to two communities is presented. Inspired by electric circuits, two variations are proposed: series and parallel. The parallel model (Fig.7.7(a)) considers that after entering to the common space, there are two available but mutually exclusive sub spaces where agents can migrate. The series model (figure 7.7(b)) considers that each agent must go across two consecutive sub spaces before leaving the group. The motivation behind this extension (besides academic purposes) is to find broader infection profiles that single communities fail to produce. Moreover, it is interesting to explore poten-

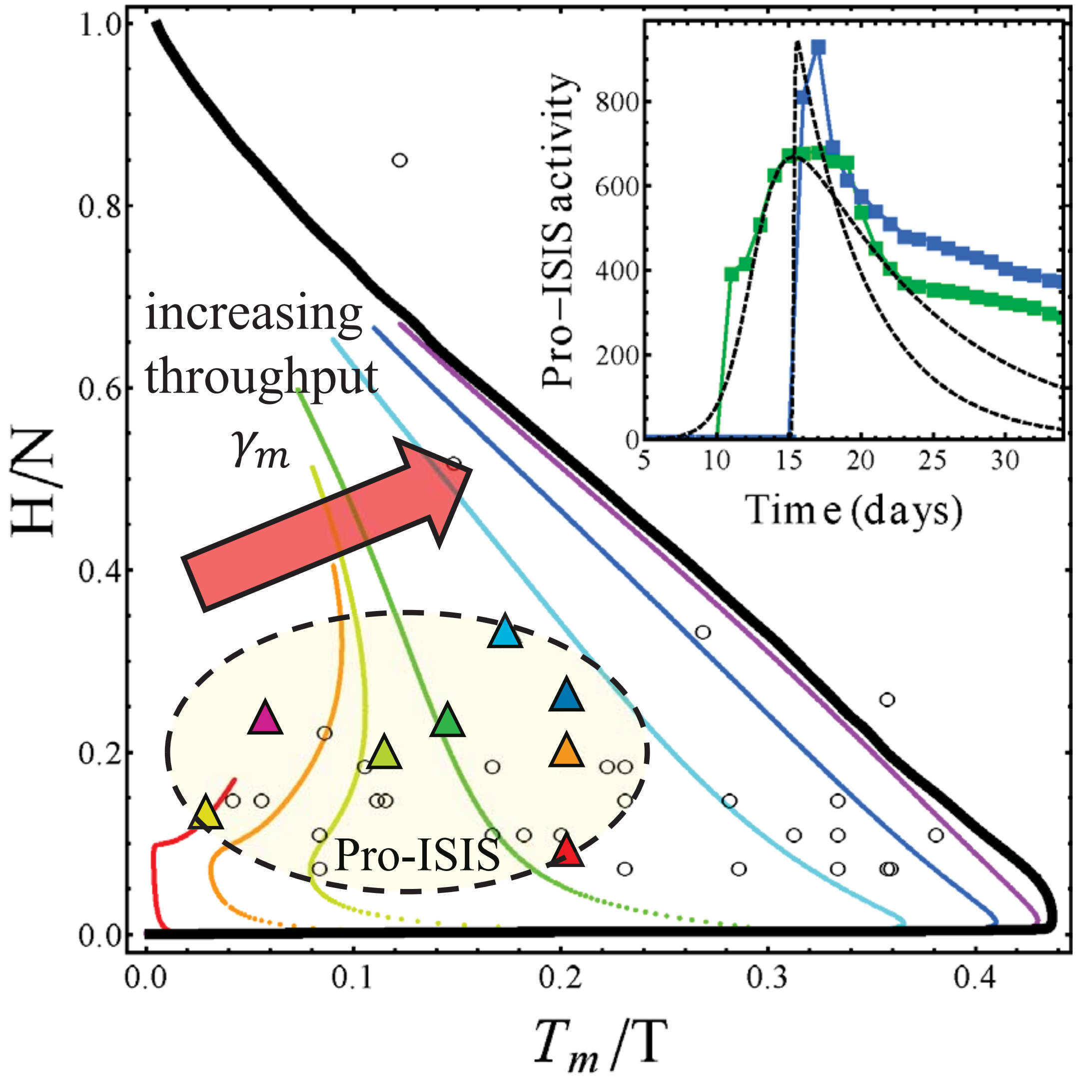

Figure 7.6: Empirical and theoretical results as a function of severity (H), duration (T) and time to peak (T_m). Colored triangles are empirical data from pro-ISIS online groups. Circles are civil unrest outbreaks across Latin America between 2012-2014. Theoretical results (colored lines) obtained from integrating the six coupled differential equations in our model, for different values of throughput γ_m (red line, $\gamma_m = 0.001$; orange line, $\gamma_m = 0.005$; yellow line, $\gamma_m = 0.01$; green line, $\gamma_m = 0.02$; cyan line, $\gamma_m = 0.05$; blue line, $\gamma_m = 0.1$; purple line, $\gamma_m = 0.2$). The same quantities are also calculated for the standard SIR model (black line). The same population size ($N = 1000$) is being used in all cases, and the same infection probability ($q_i = 0.002$). Each trajectory starts near the origin for $\lambda \equiv q_i/q_r = 10^{-3}$ and grows until $\lambda = 1$ in steps of $\delta\lambda = 10^{-3}$. Inset: Two typical empirical profiles for online outbreaks of pro-ISIS activity, identified as **club81567093** (blue) and **interes.publics** (green). Black line shows that even a best-fit standard SIR curve with the benefit of freely varying parameters, fails to capture the overall $I(t)$ profile

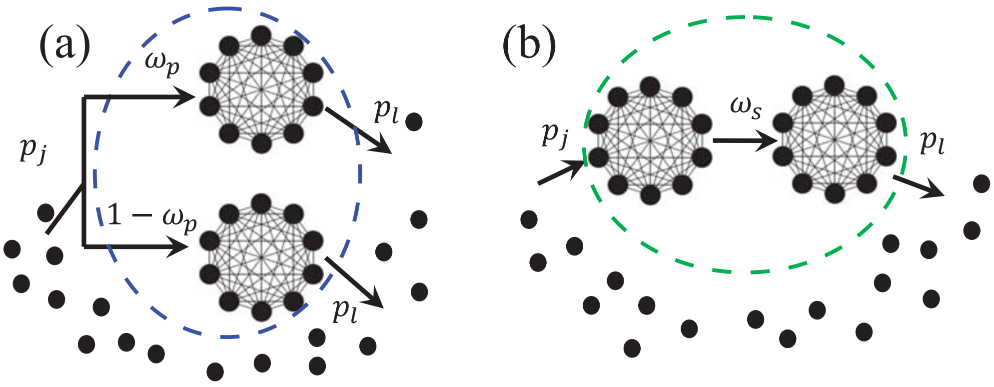

Figure 7.7: Multiple common space contagion model in parallel (a) and in series (b)

tial equivalences between the extensive profile quantities with the well-known formulas for electric resistors and capacitors in series and parallel. Furthermore, in real-world systems the model could capture insights about the influence of different sources on a susceptible population over time.

7.3.1 Parallel model

As before, it is considered N agents joining and leaving a common space with probability p_j and p_l, respectively. The common space is divided into two separate sub spaces arranged in parallel. This feature adds to the dynamics an extra probability of ω_p to migrate to the first subspace and a probability of $1-\omega_p$ to migrate to the second one. The probability to leave the common space, from each subspace is the same. First, we find the equilibrium population for each common space by considering the master equation of the system. Let N_i be the population of the sub space i and N_{out} be the population outside the common space. From this, the constrain of the system is $N = N_{out} + N_1 + N_2$.

$$\frac{dN_1}{dt} = -p_l s_1 + p_j \omega_p N_{out} \tag{7.5}$$

$$\frac{dN_2}{dt} = -p_l N_2 + p_j (1-\omega_p) N_{out} \tag{7.6}$$

The equilibrium solutions are given by solving equation 7.5 and 7.6 for $dN_1/dt = 0$ and $dN_2/dt = 0$, respectively. The equilibrium values can be written as $\langle N_i \rangle = \gamma_{s,i} N$, for $i = 1, 2$, and the parameter $\gamma_{s,i}$ is given by

$$\begin{aligned}\gamma_{s,1} &= \frac{p_j}{p_j + p_l}\omega_p \\ \gamma_{s,2} &= \frac{p_j}{p_j + p_l}\left(1 - \omega_p\right)\end{aligned} \tag{7.7}$$

Inside the common space, the agents undergo an SIR process with contagion types to be person to person, broadcast, or a combination of them with probability q_i to get infected. The infected agents outside the common space may get recovered with a probability q_r. The dynamical equations for the parallel model undergoing person to person type of contagion are given by

$$\begin{aligned}\frac{dS}{dt} &= -q_i\left(s_1 i_1 + s_2 i_2\right) \\ \frac{dI}{dt} &= q_i\left(s_1 i_1 + s_2 i_2\right) - q_r I \\ \frac{dR}{dt} &= q_r I\end{aligned} \tag{7.8}$$

where the capital letters S, I and R represent the total number of susceptible, infected and recovered agents, respectively. The lowercase letters with sub index $k = 1, 2$ represent the agents inside the first and second sub space, respectively. The dynamical equations for the agents inside the common space are

$$\begin{aligned}\frac{ds_k}{dt} &= -q_i s_k i_k - p_l(s_k - q_i s_k i_k) + p_j \omega_k \left(S - s_1 - s_2\right) \\ \frac{di_k}{dt} &= q_i s_k i_k - q_r i_k - p_l\left(i_k + q_i s_k i_k - q_r i_k\right) + (1 - q_r) p_j \omega_k \left(I - i_1 - i_2\right) \\ \frac{dr_k}{dt} &= q_r i_k - p_l\left(r_k + q_r i_k\right) + p_j \omega_k \left((R - r_1 - r_2) + q_r(I - i_1 - i_2)\right)\end{aligned} \tag{7.9}$$

for $k = 1, 2$ and where $\omega_1 = \omega_p$ and $\omega_2 = 1 - \omega_p$.

Infection profiles for parallel model

In figure 7.8 we find the numerical solution of the differential equations (7.8) for the total infection dynamics as a function of the parameter ω_p, and for three sets of microscopic parameters. First of all, we can see that at the boundaries, that is, when $\omega_p \rightarrow 1$ or $\omega_p \rightarrow 0$, we recover the profile of the single common space contagion model. Secondly, it is interesting to find that for green and blue parameters, there is a drop on the infection profile as ω_p approaches the value of $1/2$ from either left or right. For red we have a slow reduction of the peak followed by a revival that reaches its maximum at $\omega_p = 1/2$. This behavior can be understood by looking at the infection rate value q_i for each configuration. If the infection rate is small, as well as the number of susceptible agents within the sub space, the number of newly infected decreases. In addition, the point where the distribution of the agents makes that the population is the smallest at both subspaces simultaneously is for $\omega_p = 1/2$. The configuration that presents the highest infection rate q_i is red and that for green and blue is one fifth and less than one half of red. Moreover, the rate of decaying of the infection peak with ω_p is slower at blue than green, given that the infection rate for blue is twice than the infection rate in green. Another reason however, for the rapid decay in green is the value of the mobility. Since the SIR process takes place in full at the common space, a slow mobility may help on the infection spreading. The green configuration experiences the highest mobility value, which makes shorter the effective time for a susceptible agent to get infected and which combining it with the low infection rate q_i, reduces the infection profile for small number of agents on both sub spaces. Therefore, the essential

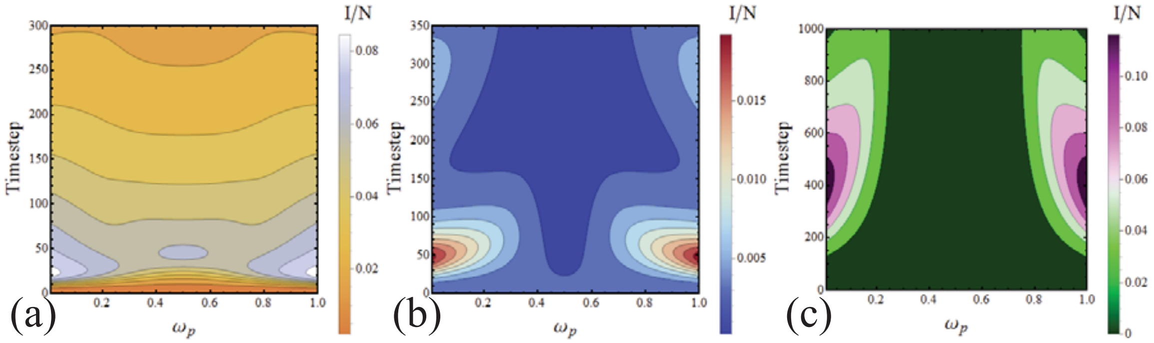

Figure 7.8: Parallel Burst Profiles. Dynamical behavior of the infected population as a function of the parameter ω_p for the three sets of microscopic parameters: Red ($q_i = 0.005$, $\lambda = 0.1$, $\gamma_m = 0.009$), Green ($q_i = 0.001$, $\lambda = 0.1$, $\gamma_m = 0.018$) and Blue ($q_i = 0.002$, $\lambda = 0.022$, $\gamma_m = 0.0018$), respectively.

ingredients to have an infective population for small number of agents on both sub spaces ($\omega_p \rightarrow 1/2$) are a small mobility and a large infection rate value.

7.3.2 Series model

Let us now consider the spaces to be arranged in series. In this case, the journey of each agent consists of two main steps. Generally, any agent outside the common space joins the space 1 with a probability p_j. However, the agent has to pass through space 2 in order to leave the common space. For each agent in space 1 there is a probability ω_s to join space 2. Finally, any agent in the space 2 is able to leave the common space with a probability of p_l. Hence, the differential equation for the population on each subspace is given by

$$\frac{dN_1}{dt} = -\omega_s N_1 + p_j N_{out} \tag{7.10}$$

$$\frac{dN_2}{dt} = -p_l N_2 + \omega_s N_1. \tag{7.11}$$

As in the parallel case, the steady state solutions for the populations can be calculated by setting $dN_j/dt = 0$ and solving for the populations. The results can be written as a function of the parameter γ_s of the single common space model as

$$\begin{aligned}\langle N_1\rangle &= N\gamma_s\left(\frac{\omega_s}{p_l}+\frac{p_j}{p_l+p_j}\right)^{-1}\\ \langle N_2\rangle &= N\gamma_s\left(1+\frac{p_l}{\omega_s}\frac{p_j}{p_j+p_l}\right)^{-1}.\end{aligned}\tag{7.12}$$

The equilibrium populations vary with the relationship between the parameters ω_s, p_l and p_j. For instance, for $\omega_s \ll p_l$, the equilibrium population of subspace 1 tends to $N/(1+\kappa)$, where κ is the ratio of ω_s to p_j, while the equilibrium population for subspace 2 vanishes. On the other hand, for $\omega_s \gg p_l$, the population of subspace 1 decays to zero while for subspace 2 we recover the equilibrium population for the single common space contagion model. Finally, the populations are identical only for $\omega_s = p_l$. Inside the common space, the agents go through an SIR process, and outside infected agents may get recovered. The dynamical equation for the total quantities S, R and I are the same as the parallel model

$$\begin{aligned}\frac{dS}{dt} &= -q_i\left(s_1i_1+s_2i_2\right)\\ \frac{dI}{dt} &= q_i\left(s_1i_1+s_2i_2\right)-q_rI\\ \frac{dR}{dt} &= q_rI.\end{aligned}\tag{7.13}$$

Additionally, for the first common space the quantities are given by

$$\begin{aligned}\frac{ds_1}{dt} &= -q_is_1i_1-\omega_s\left(s_1-q_is_1i_1\right)+p_j\left(S-s_1-s_2\right)\\ \frac{di_1}{dt} &= q_is_1i_1-q_ri_1-\omega_s\left(i_1+q_is_1i_1-q_ri_1\right)+\left(1-q_r\right)p_j\left(I-i_1-i_2\right)\\ \frac{dr_1}{dt} &= q_ri_1-\omega_s\left(r_1+q_ri_1\right)+p_j\left(\left(R-r_1-r_2\right)+q_r\left(I-i_1-i_2\right)\right),\end{aligned}\tag{7.14}$$

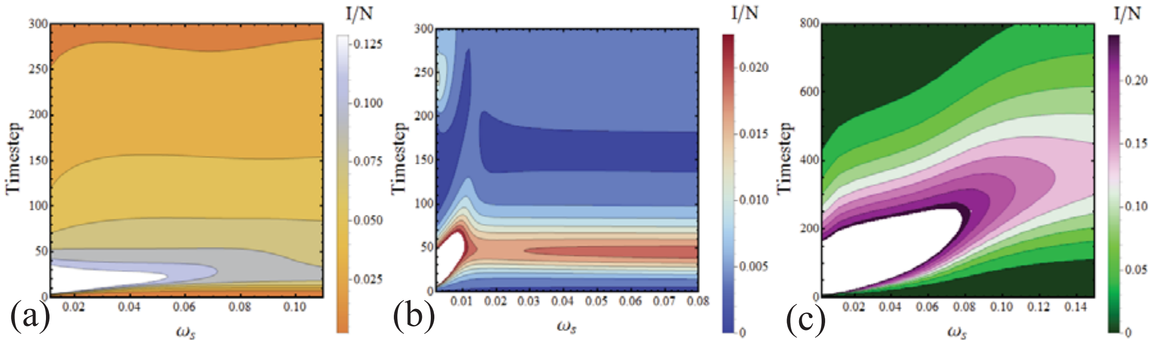

Figure 7.9: Series Burst Profiles. Dynamical behavior of the infected population as a function of the parameter ω_s for the three sets of microscopic parameters: Red ($q_i = 0.005$, $\lambda = 0.1$, $\gamma_m = 0.009$), Green ($q_i = 0.001$, $\lambda = 0.1$, $\gamma_m = 0.018$) and Blue ($q_i = 0.002$, $\lambda = 0.022$, $\gamma_m = 0.0018$), respectively.

while for the second space we have

$$\begin{aligned}
\frac{ds_2}{dt} &= -q_i s_2 i_2 - p_l\left(s_2 - q_i s_2 i_2\right) + \omega_s(s_1 - q_i s_1 i_1) \\
\frac{di_2}{dt} &= q_i s_2 i_2 - q_r i_2 - p_l\left(i_2 + q_i s_2 i_2 - q_r i_2\right) + \omega_s\left(i_1 + q_i s_1 i_1 - q_r i_1\right) \\
\frac{dr_2}{dt} &= q_r i_2 - p_l\left(r_2 + q_r i_2\right) + \omega_s\left(r_1 + q_r i_1\right). \qquad (7.15)
\end{aligned}$$

Infection profiles for series model

Figure 7.9 depicts the burst profiles for the series model and a function of the parameter ω_s. We find that the profile experiences the greatest variations for small ω_s, where the peak of infection is notably greater than for larger values of ω_s. This feature is independent of the other parameters sets. Moreover, this is in agreement with the equations (7.12), where for $\omega_s \ll p_l$ the population of agents on the first sub space grows to the limit of the total population N for $\omega_s = 0$. In this way, the large number of agents located at the sub space 1 will be more prompted to get infected and consequently to infect other actors. On the other hand, as ω_s grows, the population on the second sub space approaches to the single

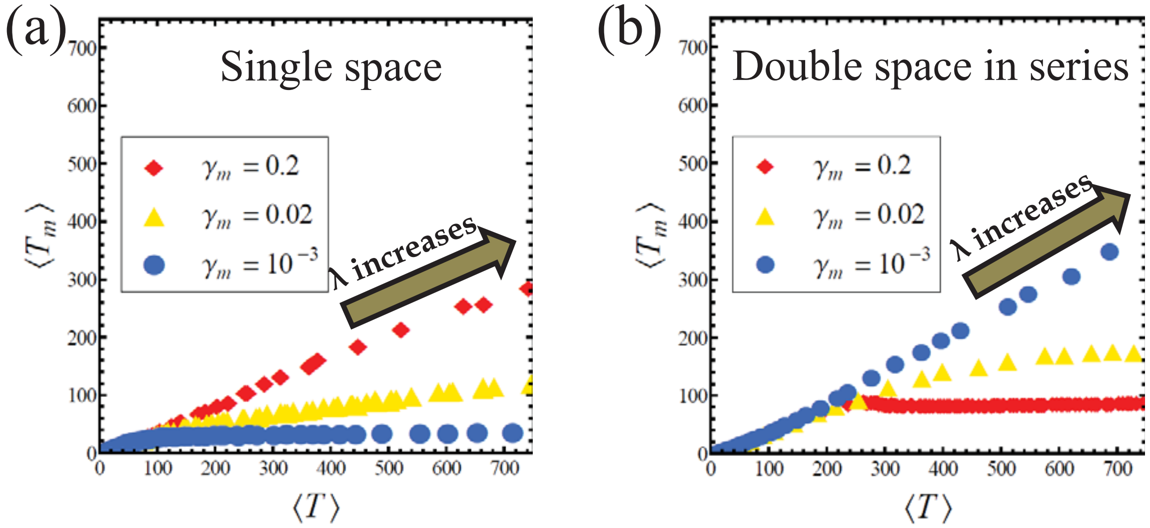

Figure 7.10: Non-linearity of the outbreak time-to-peak T_m as a function of duration T, for different values of mobility γ_m. Each dot represents an average value over 1000 realizations and changes as λ grows in steps of 10^{-2}. (a) Single common space model with person-to-person contagion mechanism. (b) Series model where first sub space undergoes person-to-person contagion mechanism while the second one broadcast.

CSC model while the population of the first sub space decays to zero. The series infection profile therefore, reproduces the single CSC model under the condition of $\omega_s \gg p_l$.

7.4 Model variations and predictability power

7.4.1 Contagion mechanisms

Variations of contagion mechanisms preserve a strong nonlinear dependence on mobility. For instance, consider that every individual within the common space has a constant probability to get infected, independently of the number of infected individuals present in the common space. This mechanism of contagion is known as $broadcast$, and shows contrasting behavior when compared with person-to-person mechanisms or when combined with person-to-person on a multiple space system. Figure 7.10 compares single and series multiple space by looking at the duration and time-to-peak as a function of λ and mobility γ_m.

The contagion mechanism for the single space and the first sub space of the series model is person-to-person, while for the second sub space is broadcast. The motivation behind studying this particular combination for the series model is two-fold. (1) It produces a largest ratio T_m/T when compared with the single space or other type of series combination. A ratio $T_m/T > 0.5$ has been observed on real-world outbreaks such as civil unrest. (2) The multi-stage model (i.e. series) resembles multiple source of information that may influence individuals such as radio news stations (broadcast) and online forum (person-to-person). Notably, a series combination as showed in Fig.7.10(b), on average, preserves the ratio T_m/T for small values of mobility at a value ≈ 0.55. By contrast, for the single model the ratio is preserved for large values of mobility but it is not larger than 0.45.

Another key point is to note that these values are averages over 10^3 realizations. Therefore, larger values of T_m/T can be obtain by the combined series model on a case-by-case basis. This is shown on Fig.7.11 where a wider range of civil unrest outbreaks are positioned on a T_m vs T plane. The gray region delimits the outbreaks that could be obtained by the single model where the ratio T_m/T is near zero (red curve). Outbreaks in the boundary of the gray and white region, (i.e. $T_m/T \approx 0.5$) can be described by either the single model or the averaged combined series model (green curve). Finally, for outbreaks whose peak of severity is located near the end of the burst, a combined model is better suited (black simulation). The latter case is still problematic since no model combination can produce, on average, a profile with these characteristics.

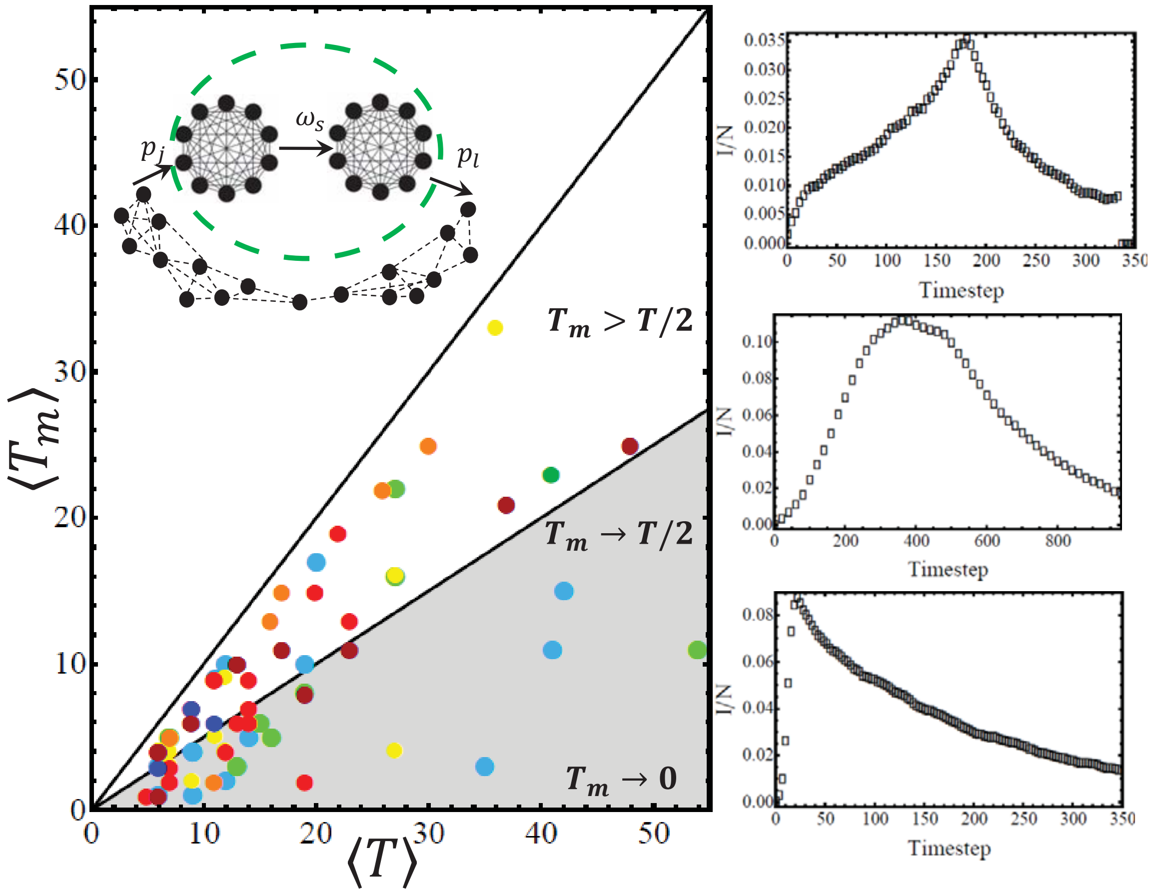

Figure 7.11: Real-world outbreaks exhibit extreme profiles (e.g. ratio T_m/T of time-to-peak to duration) represented by colored circles. Though these lie beyond standard SIR model and single model, they can be explained using series combined model. Illustrative infection profiles are shown.

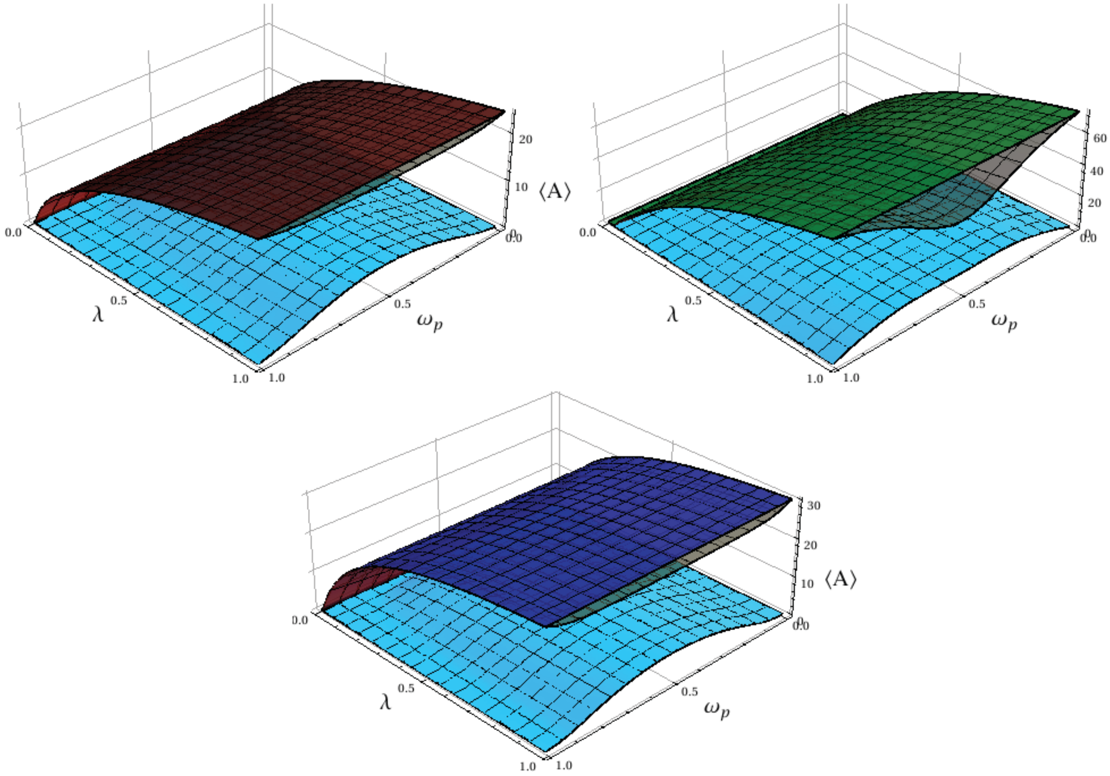

Figure 7.12: Comparison between *parallel* CSC model and single CSC model by using the area below the I/N curve as a function of ω_p and the infection rate λ. The area of each sub space is denoted by a_i. The light blue plane represents the quantity $a_1a_2/(a_1+a_2)$ while the gray plane is a_1+a_2. The colored planes (red, green and blue) are the result from the single CSC model for these type of parameters.

7.5 Possible Correspondence with RC circuits

We address this by calculating the extensive quantity a for each sub space independently and comparing it with the single CSC model. Given the architecture of the multiple CSC model, an interesting question is how we can relate the results of the model with an electric circuit. Is this quantity resistor-like or capacitor-like and under what conditions? We address the question by calculating the extensive quantity $\mathcal{Q}$ for each sub space independently and comparing it with the single CSC model. Given the previous results, we know how to recover the single CSC model while working with the double CSC model in series or in parallel. For instance, the parallel model for $\omega_p \to 0$ or $\omega_p \to 1$ is equivalent to the single CSC model. Therefore, we can conclude that any extensive quantity behaves like a capacitor, since for these limits one of the extensive quantities values is equal to zero and the remainder reproduces the single CSC model. In the same way, we may conclude that the series model behaves like resistors for the cases where $\omega_s \gg p_l$. For the other values of ω_p and ω_s, the comparison is made with the extensive quantity associated to the common space only, i.e., infective agents outside the common space will not be included.

Figure 7.12 compares the single CSC model and the parallel two spaces CSC model by using as an extensive quantity the area below the I/N curve for three sets of parameters as a function of the infection rate λ and the parameter ω_p. The light blue plane represents the sum of the individual areas associated with the parallel CSC model. The Blue plane is the outcome by adding the inverses of the individual areas and taking the inverse. And the upper colored planes represent the single CSC model result. The results reveal that there is no direct relation between the areas through a circuit analogy other than the trivial

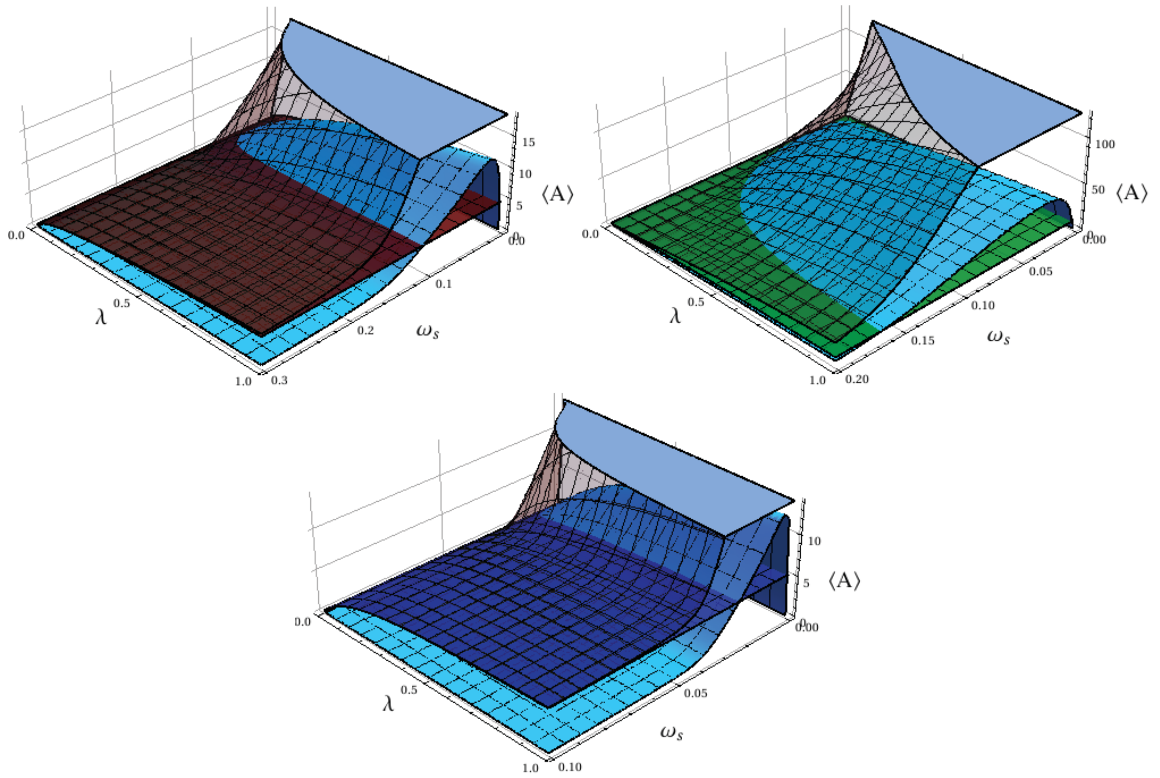

Figure 7.13: Comparison between *series* CSC model and single CSC model by using the area below the I/N curve as a function of ω_s and the infection rate λ. The area of each sub space is denoted by a_i. The light blue plane represents the quantity $a_1 a_2/(a_1 + a_2)$ while the gray plane is $a_1 + a_2$. The colored planes (red, green and blue) are the result from the single CSC model for these type of parameters.

result when the parallel model reproduces the single model ($\omega_p \to 0, 1$). Other values of ω_p generate different results independently of the value of λ.

On the other hand, the outcome from the series CSC model reveals a richer behavior. Figure (7.13) depicts the area relations as a function of ω_s and λ. In addition to the known result for large ω_s, we find coincidences between the capacitance relation of the areas and the single CSC model for specific values of λ and ω_s. The coincidences correspond to the curve where the light blue plane (capacitance relation between areas) and the colored ω_s independent plane (single CSC model result) intersect. In other words, the single CSC model can be mapped by a series CSC model where the areas are modeled as capacitors and the

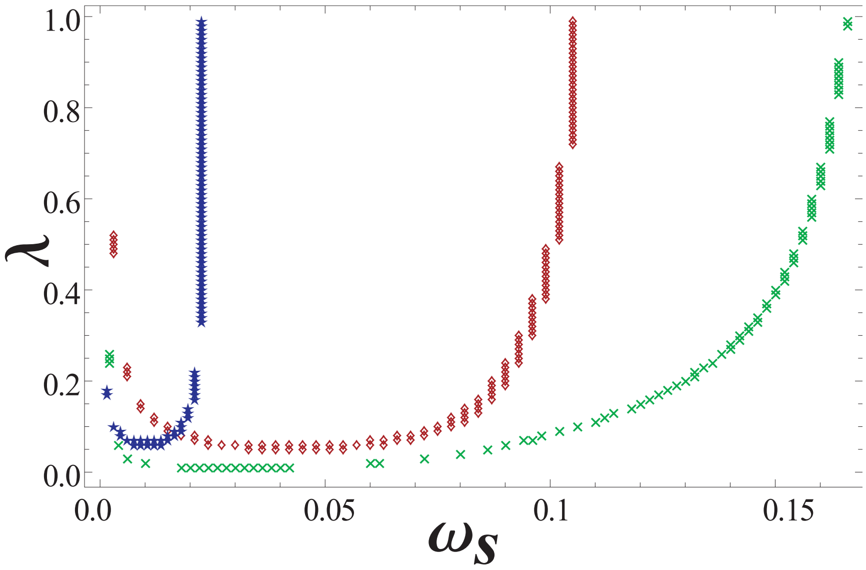

Figure 7.14: Curves on the $\lambda - \omega_s$ plane where the area of the series model show a capacitor-like behavior.

values of λ and ω_s lie in the curve where the planes meet. Figure 7.14 illustrates the curves on the $\lambda - \omega_s$ plane where the areas of the series CSC model behaves like capacitors for the three sets of parameters studied before. The curves reveal an exponential-like behavior between the quantities λ and ω_s.

7.6 Summary of chapter 7

An agent-based model of contagion including the human tendency of moving through on-line and off-line spaces was presented. The dynamical component allows greater flexibility to capture a wider variety of extensive features of modern-day outbreaks that other traditional approaches (e.g. SIR) lack. In particular, civil-unrest activity in Latin America, global on-line trading and pro-ISIS online support lie on the range generated by the model. Results reveal a high non-linear dependence on mobility that yield to the counter-intuitive prediction that by increasing the flow of individuals through a region of contagion (on-line or off-line), the infection's severity could be decreased. Extension to multiple regions of contagion, together with variations on the contagion mechanisms, are explored and wider ranges of profile features have been found.

Chapter 8

Summary and perspectives

The results provided in this thesis constitute a series of steps forward into the understanding and modeling of complex dynamical systems. While traditional approaches focus on asymptotic behavior, quasi-static transitions and nearly infinite number of particles, many real-world systems lie beyond the scope of these assumptions. On the other hand, reductionist approaches aim to describe the microscopic details at a fundamental level, and hence might fail in capturing the big picture of the system. This thesis aims for a midway. We use a holistic approach in order to capture the underlying mechanisms that generate emerging out-of-equilibrium behavior. This approach has allowed us to analyze additional implications that may lie beyond the empirical data available.

For example, the stochastic model for bacterial photosynthesis bestowed in chapter 3 has been verified by its demonstrated prediction of structural preferences [35; 36]. This empirical ground opens the gate to explore potential implications. For example, the impact of abnormal photon arrival times and hence, examine conditions for survival in environments different from Earth's. In addition, our approach was able to provide quantitative understanding of the complementarity between the macro molecular aggregation of light harvesting structures in primitive photosynthetic organisms, and spatial correlations naturally present in sunlight. Either an ordered or disordered antennae assembly is favored depending on light availability and the organisms' metabolic needs, yielding high-efficiency photon absorption for low light-intensity environments and protection for high light-intensity environments.

Following this line of thought. Having identified the grouping mechanisms that produce the emergent feature of a 5/2 power-law size distribution of aggregates, a step forward has been taken by eliminating the serious shortcoming which assumes that the underlying objects are identical. We have demonstrated that the introduction of objects' heterogeneity, where it typically dictates their grouping rules and hence their collective behavior, preserves the universally observed feature while revealing rich transition behavior. Additionally, the implementation of a character-driven fission-fusion mechanism, mimics connectivity features and produces sparse networks which are quantitatively similar to those observed empirically. These commonalities give us insights about the orientation-type of grouping which drives the evolution of a dynamical adaptive network.

We analyzed additional quantitative measures to study contrasting real-world networks from several social domains. Interestingly, for networks where the environment becomes more aggressive and dangerous, we have found a remarkable emerging superiority of a female minority in terms of network connectivity. This trend is statistically associated with benefits to the network in the local and global level, in aspects such as longevity. Our findings raise more general questions about who the important members of a network actually are in terms of an organization's development. As a result, we suggest to investigate whether the contributions of female members would be better appreciated if measured in terms of collective connectivity, as opposed to measures favoring individual connectivity.

Finally, we looked at the problem of contagion in dynamical networks. We developed a simple but highly non-trivial model that, by introducing a modern-day tendency of passing through a common place of interaction (e.g. social media), leads to a wide variety of infection profiles that resemble those found in several real-world systems. In addition, the results revealed the anti-intuitive outcome which states that the severity of an epidemic could in fact decrease if the mobility of individuals entering and leaving the contagion region increases. Additional developments of the model, accounting for multiple places of interaction were advanced. These expansions have shown to increase the applicability to real-world outbreaks when compared to the single common space model.

There are several possible avenues of research. A vision for the future accounts for the following:

1. Integrate the quantum dimer with the stochastic model to study the influence of quantum dynamic effects (e.g. entanglement) in the photosynthetic reaction center and how they impact the overall efficiency of the photosynthetic process.
2. Explore the usage of stochastic methods to charge migration modeling on particular types of novel light-driven organic materials, and examine how spatial coherence could guide the engineering on their light detectors.

3. Study the robustness of the 5/2 distribution of aggregates by a generalization of our heterogeneous fission/fusion model to account for a variable number of individuals.

4. Investigate commonalities of the isolation/cohesion transition with the defect formation equivalent to the Kibble-Zurek effect [186] and explore potential phenomenological usage on coupled quantum systems (e.g. Dicke model).

5. Examine potential physical applications of the common space contagion model. In particular, the two-state model SIS where infected agents relax to a susceptible state instead of a recovered state. This model can be implemented on two-level systems where, for instance, particles can change their state due to active interaction ($S+I \rightarrow I+I$) on a subset, or due to passive relaxation ($I \rightarrow S$).

6. Explore generalizations of multiple common spaces that can be re-normalized to a single ’dressed’ common space and mapped into an effective electric circuit.

Bibliography

[1] N. Johnson. *Simply Complexity*. Oneworld Publications, Oxford, 2007.

[2] N. Johnson, P. Jefferies, and P.M. Hui. *Financial Market Complexity*. Oxford University Press, Oxford, 2003.

[3] R. Lewis. *Complex Systems: Science at the Edge of Chaos*. Phoenix Press, Phoenix, 2001.

[4] A.L. Barabasi. *Linked: How Everything is Connected to Everything Else and What It Means for Business, Science and Everyday Life*. Plume Books, 2003.

[5] R. Sole and B. Goodwin. *Signs of Life: How Complexity Prevades Biology*. Basic Books, 2002.

[6] J. Gribbin. *Deep Simplicity: Chaos Complexity and the Emergence of Life*. Penguin Books, 2005.

[7] S. Strogatz. *Sync: The Emerging Science of Spontaneous Order*. Penguin Books, 2004.

[8] P. Jefferies, M.L. Hart, P.M. Hui, and N. Johnson. From markets games to real-world markets. *Eur. Phys. J. B*, 20:493–501, 2001.

[9] S.C. Choe, G. Zhao, D. Zhao, Z. andRosenblatt, H.M. Cho, S.U. Shin, and N. Johnson. Model for in vivo progression of tumors based on co-evolving cell population and vasculature. *Sci. Rep.*, 1(31), 2011.

[10] P. Manrique, H. Qi, A. Morgernstern, N. Velásquez, and N. Johnson. Context matters: Improving the uses of big data for forecasting civil unrest. *IEEE Intl. Conf. on Intelligence and Security Informatics.*, pages 169–172, 2013.

[11] A. Morgenstern, N. Velasquez, P. Manrique, H. Qi, N. Johnson, and N.F. Johnson. Modeling political conflict, violence and wars. *American Journal of Physics*, 81(805), 2013.

[12] D.J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton Studies in Complexity, 2003.

[13] J. Scott. *Social Network Analysis*. SAGE publications, London, 2013.

[14] N. Johnson et al. Multi-agent complex systems and many-body physics. *Europhys. Lett.*, 74(5), 2006.

[15] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47, 2002.

[16] R.V. Solé et al. *Complexity*, 8(20), 2002.

[17] G. Caldarelli and D. Garlaschelli. Self-organization and complex networks. *arXiv:0806.1655*, 2008.

[18] J.I. Perotti et al. Emergent self-organized complex network topology out of stability constraints. *Phys.Rev.Let*, 103:108701, 2009.

[19] M. Lim, R. Metzler, and Y. Bar-Yam. Global pattern formation and ethnic/cultural violence. *Sience*, 317:1540–1544, 2007.

[20] A. Clauset, M. Young, and K.S. Gleditsch. On the frequency of severe terrorist events. *J. Conflict Resolut.*, 51:58–87, 2007.

[21] N. Johnson, et al. Universal patterns underlying ongoing wars and terrorism. *arXiv:physics/0605035*, 2006.

[22] M. Fratini, et al. Scale-free structural organization of oxygen interstitials in la_2cuo_{4+y}. *Nature*, 466(841), 2010.

[23] D.R. Chialvo. Emergent complex neural dynamics. *Nature Physics*, 6(744), 2010.

[24] X. Gabaix, et al. Institutional investors and stock market volatility. *Quaterly Journal of Economics*, 121(461), 2006.

[25] X. Hu, T. Ritz, A. Damjanovic, F. Autenrieth, and K. Schulten. Photosynthetic apparatus of purple bacteria. *Quarterly Reviews of Biophysics*, 35(1), 2002.

[26] S. Scheuring and J.N. Sturgis. Chromatic adaptation of photosynthetic membranes. *Science*, 309(484), 2005.

[27] D.E. Chandler et. al.,. Light harvesting by lamellar chromatophores in *rhodospirillum photometricum*. *Biophys. J.*, 106:2503–2510, 2014.

[28] M.L. Cartron et. al.,. Integration of energy and electron transfer processes in the photosynthetic membrane of *rhodobacter sphaeroides*. *Biochimica et Biophysica Acta.*, (1837):1769–1780, 2014.

[29] J. Strümpfer and K. Schulten. Open quantum dynamics calculations with the hierarchy equations of motion on parallel computers. *J. Chem. Phys.*, 137(065101), 2012.

[30] J. Strümpfer and K. Schulten. Open quantum dynamics calculations with the hierarchy equations of motion on parallel computers. *Journal of Chemical Theory and Computation*, 8:2808–2816, 2012.

[31] R. Blankenship. *Molecular Mechanisms of Photosynthesis*. Wiley-Blackwell, 2002.

[32] T. Geyer and V. Helms. Reconstruction of a kinetic model of the chromatophore vesicles from *rhodobacter sphaeroides*. *Biophys. J.*, 91(927), 2006.

[33] A. Damjanović, T. Ritz, and K. Schulten. Excitation energy trapping by the reaction center of rhodobacter sphaeroides. *Int. J. Quantum Chem.*, 77(139), 2000.

[34] G.R. Flemming and R. van Grondelle. Femtosecond spectroscopy of photosynthetic light-harvesting systems. *Curr. Opin. Struct. Biol.*, 7(738), 1997.

[35] F. Caycedo-Soler, F.J. Rodriguez, L. Quiroga, and N.F. Johnson. Light-harvesting mechanism of bacteria exploits a critical interplay between the dynamics of transport and trapping. *Physical Review Letters*, 104(158302), 2010.

[36] F. Caycedo-Soler, F.J. Rodriguez, L. Quiroga, and N.F. Johnson. Interplay between excitation kinetics and reaction-center dynamics in purple bacteria. *New Journal of Physics*, 12(095008), 2010.

[37] N. Johnson, G. Zhao, F. Caycedo, P. Manrique, H. Qi, F. Rodriguez, and L. Quiroga. Extreme alien light allows survival of terrestrial bacteria. *Scientific Reports*, 3(298), 2013.

[38] H.P. Breuer and F. Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, 2006.

[39] G.M. Moy. J.J. Hope and C.M. Savage. Born and markov approximation for atom lasers. *Phys. Rev. A*, 59(667), 1999.

[40] A. Olaya-Castro, C. Fan Lee, F. Fassioli Olsen, and N. Johnson. Efficiency of energy transfer in a light-harvesting system under quantum coherence. *Phys. Rev. B*, 78(085115), 2008.

[41] F. Caycedo-Soler. F. Rodriguez. L. Quiroga and N. Johnson. Light-harvesting mechanism of bacteria exploits a critical interplay between the dynamics of transport and trapping. *Phys. Rev. Lett*, 104(158302), 2010.

[42] L. Campos Venuti and P. Zanardi. Excitation transfer through open quantum networks: a few basic mecanisms. *Phys. Rev. B*, 84(134206), 2011.

[43] D. Abbott, P.C. Davies, and A.K. Pati. *Quantum Aspects of Life*. Imperial College Press, 2008.

[44] S. Scheuring, J.L. Rigaud, and J.N. Sturgis. Variable lh2 stoichiometry and core clustering in native membranes of *Rhodospirillum photometricum*. *EMBO J.*, 23(4127), 2004.

[45] G. Scholes, G. Flemming, A. Olaya-Castro, and R. van Grondelle. Lessons from nature about solar light harvesting. *Nature Chemistry*, 3(763774), 2011.

[46] G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mancal, Y. C. Cheng, R. E. Blankenship, and G. R. Fleming. Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature*, 446:782–786, 2007.

[47] E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer, and G. D. Scholes. Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature*, 463:644–647, 2010.

[48] G. Panitchayangkoon, D. Hayes, K. A. Fransted, J. R. Caram, E. Harel, J. Wen, R. E. Blankenship, and G. S. Engel. Long-lived quantum coherence in photosynthetic complexes at physiological temperature. *Proc. Natl. Acad. Sci. USA*, 107:12766–12770, 2010.

[49] F. D. Fuller, J. Pan, A. Gelzinis, V. Butkus, S. S. Senlik, D. E. Wilcox, C. F. Yocum, L. Valkunas, D. Abramavicius, and J. P. Ogilvie. Vibronic coherence in oxygenic photosynthesis. *Nat. Chem.*, 6:706–711, 2014.

[50] E. Romero, R. Augulis, V. I. Novoderezhkin, M. Ferretti, J. Thieme, D. Zigmantas, and R. van Grondelle. Quantum coherence in photosynthesis for efficient solar-energy conversion. *Nature Phys.*, 10:676–682, 2014.

[51] M. Anghel, Z. Totoczkai, K.E. Bassler, and G. Korniss. Competition-driven network dynamics: Emergence of a scale-free leadership structure and collective efficiency. *Phys. Rev. Lett*, 92(058701), 2004.

[52] A. Soulier and T. Halpin-Healy. The dynamics of multidimensional secession: Fixed points and ideological condensation. *Phys. Rev. Lett*, 90(258103), 2003.

[53] B. Goncalves and N. Perra. *Social Phenomena: Data Analytics and Modeling*. Springer, Berlin, 2015.

[54] G. Palla, A.L. Barabasi, and T. Vicsekm. Quantifying social group evolution. *Nature*, 446:664–667, 2007.

[55] P.L. Krapivsky, S. Redner, and E. Ben-Naim. *A Kinetic View of Statistical Physics*. Cambridge University Press, Cambridge, 2010.

[56] E. Estrada. Communicability in temporal networks. *Phys. Rev. E*, 88(042811), 2013.

[57] C. Song, S. Havlin, and H. Makse. Origins of fractality in the growth of complex networks. *Nature Phys.*, 2(275), 2006.

[58] G. Caldarelli. *Scale-free networks: Complex webs in nature and technology*. Oxford University Press, Oxford, 2007.

[59] A.L. Barabasi and T. Halpin-Healy. *Fractal Concepts in Surface Growth*. Cambridge University Press, Cambridge, 1995.

[60] F. Radicchi and S. Fortunato. Explosive percolation: A numerical analysis. *Phys. Rev. E*, 81(036110), 2010.

[61] J. Nagler, A. Levina, and M. Timme. Impact of single links in competitive percolation. *Nature Physics*, 7(265), 2011.

[62] D.J. Watts and P.S. Doods. Influential networks and public opinion formation. *Journal of Consumer Research*, 34(441), 20057.

[63] I.D. Couzin, et al. Effective leadership and decision-making in animal groups on the move. *Nature*, 433(513), 2005.

[64] S. Wuchty, et al. The increasing dominance of teams in production of knowledge. *Science*, 316(1036), 2007.

[65] D.R. Forsyth. *Group Dynamics*. Cengage, New York, 2013.

[66] D. Centola et al. Homophily, cultural drift, and the co-evolution of cultural groups. *J. Confl. Resolut.*, 51(905), 2007.

[67] A. Wyld and Rodgers G.J. Models for random graphs with variable strength edges. *Physica A*, 374(491), 2007.

[68] B. Ruszczycki et al. Relating the microscopic rules in coalescence-fragmentation models to the cluster-size distribution. *Eur. Phys. J. B*, 72(289), 2009.

[69] N. Johnson, P. Manrique, and Pac Ming Hui. Modeling insurgent dynamics including heterogeneity. *J. Stat. Phys.*, 151(395), 2013.

[70] J. Horgan, et al. *From Bomb to Bomb-maker*. Office of Naval Research, Arlington, 2013.

[71] A. Moloney. *A Secret History of the IRA*. Norton, New York, 2003.

[72] V. Asal, P. Gill, R.K. Rethemeyer, and J. Horgan. Killing range: Explaining lethality variance within a terrorist organization. *Journal of Conflict Resolution*, 59:401–427, 2015.

[73] B. Uzzi and S. Dunlap. How to build you network. *Harvard Business Review*, 83(53), 2005.

[74] C. Shipman and K Kay. *The Confidence Code*. Harper Collins Publishers, 2014.

[75] H. Ibarra. Paving an alternative route: Gender differences in managerial networks. *Social Psychology Quarterly*, 60(91), 1997.

[76] G. Moore. Structural determinants of men's and women's personal networks. *American Sociological Review*, 55(726), 1990.

[77] A.H. Eagly and A. Mladinic. Are people prejudice against women? some answers from research on attitudes, gender stereotypes, and judgments of competence. *European Review of Social Psychology*, 5(1), 1994.

[78] E. Estrada. *The structure of Complex Networks*. Oxford University Press, Oxford, 2012.

[79] M.C. Gonzalez et al. Understanding individual human mobility patterns. *Nature*, 453(779), 2008.

[80] L. Feng, Y. Hu, V. Li, H.E. Stanley, S. Havlin, and L.A. Braunstein. Competing for attention in social media under information overload conditions. *PLoS ONE*, 10(e0126090), 2015.

[81] M.V. Tomasello et al. The role of endogenous and exogenous mechanisms in the formation of r&d network. *Sci. Rep.*, 4(5679), 2014.

[82] Z. Zhao, J.P. Calderon, C. Xu, G. Zhao, D. Fenn, D. Sornette, R. Crane, P.M. Hui, and N.F. Johnson. Effect of social group dynamics on contagion. *Phys. Rev. E*, 81(056107), 2010.

[83] L.D. Valdez et al. Predicting the extinction of ebola spreading in liberia due to mitigation strategies. *Sci. Rep.*, 5(12172), 2015.

[84] D.J. Watts et al. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences*, 102(11157), 2005.

[85] M. Karsai et al. Time varying networks and the weakness of strong ties. *Sci. Rep.*, 4(4001), 2014.

[86] L.G. Alvarez Zuzek et al. Epidemic model with isolation in multilayer networks. *Sci. Rep.*, 5(12151), 2015.

[87] N. Ferguson. Capturing human behaviour. *Nature*, 446(733), 2007.

[88] S. Lui and N. Baronchelli, A. Perra. Contagion dynamics in time-varying metapopulation networks. *PRE*, 87(032805), 2013.

[89] J.P. Onnela and F. Reed-Tsochas. The spontaneous emergence of social influence in online systems. *Proceedings of the National Academy of Sciences*, 43(18375), 2010.

[90] D. Sornette et al. Endogenous versus exogenous shocks in complex networks: An empirical test using book sale rankings. *Phys. Rev. Lett.*, 93(22), 2004.

[91] F. Deschâtres and D. Sornette. The dynamics of book sales: Endogenous versus exogenous shocks in complex networks. *Phys. Rev. E*, 72(016112), 2005.

[92] M. McDonald, O. Suleman, S. Williams, S. Howison, and N.F. Johnson. Impact of unexpected events, shocking news, and rumors on foreign exchange market dynamics. *Phys. Rev. E*, 77(046110), 2008.

[93] H.P. Breuer. Non-markovian generalization of the lindbland theory of open quantum systems. *Phys. Rev. A*, 75(022103), 2007.

[94] U. Weiss. *Quantum Dissivative Systems*. World Scientific, 1999.

[95] A.J. Leggett. S. Chakravarty. A.T. Dorsey. M. Fisher. A. Gard and W. Zwerger. Dynamics of the dissipative two-state system. *Rev. Mod. Phys*, 59(1), 1987.

[96] H.P. Breuer. Non-markovian quantum dynamics and the method of correlated projection superoperators. *Lecture Notes in Physics*, 787:125–139, 2010.

[97] A. Rivas. A.D.K. Plato. S.F. Huelga and M.B. Plenio. Markovian master equations: A critical study. *New Journal of Physics*, 12(113032), 2010.

[98] Y.C. Cheng and G.R. Fleming. Dynamics of light harvesting in photosynthesis. *Annu. Rev. Phys. Chem*, 60:241–262, 2009.

[99] A. Chin. J. Prior. R. Rosenbach. F. Caycedo-Soler. S. Huelga and M.B. Plenio. The role of non-equilibrium vibrational structures in electronic coherence and recoherence in pigment-protein complexes. *Nature Physics*, 9:13–18, 2013.

[100] L. Quiroga. F. Rodríguez. M. Ramírez and R. París. Nonequilibrium thermal entanglement. *Phys. Rev. A*, 75(032308), 2007.

[101] J.C. Castillo. F.J. Rodriguez and L. Quiroga. Enhanced violation of a leggett-garg inequality under nonequilibrium thermal conditions. *Phys. Rev. A*, 88(022104), 2013.

[102] G.L. Deçordi and V. Vidiella-Barranco. Coherence and entanglement in a two-qubit system coupled to a finite temperature reservoir: A comparative study. *arXiv:1406.0528v1 [quant-ph]*, 2014.

[103] A. Levy and R. Kosloff. The local approach to quantum transport may violate the second law of thermodynamics. *Europhys. Lett.*, 107(20004), 2014.

[104] M. Goldman. Formal theory of spin-lattice relaxation. *Journal of Magnetic Resonance*, 149:160–187, 2001.

[105] H. Van Amerongen, L. Valkunas, and R. van Grondelle. *Photosynthetic Excitons*. World Scientific, 2000.

[106] T. Ritz, S. Park, and K.J. Schulten. Kinetics of excitation migration and trapping in the photosynthetic unit of purple bacteria. *Phys. Chem. B*, 105(8459), 2001.

[107] S. Scheuring, J.N. Sturgis, V. Prima, A. Bernadac, D. Levi, and J.L. Rigaud. Watching the photosynthetic apparatus in native membranes. *Proc. Natl. Acad. Sci.*, 101(11293), 2004.

[108] R.S. Knox. *Primary Processes of Photosynthesis*. Elsevier-North Holland, 1977.

[109] T. Pullerits and V. Sundström. Photosynthetic light-harvesting pigment-protein complexes: Toward understanding how and why. *Acc. Chem. Res.*, 29(381), 1996.

[110] R.J. Cogdell and C. Mullineaux. Photosynthetic light harvesting - introduction. *Photosynthesis Research*, 95(117), 2008.

[111] M.F. Hohmann-Marriott and R.E. Blankenship. Evolution of photosynthesis. *Annu. Rev. Plant Biol.*, 62:515–548, 2011.

[112] E. Anders. Pre-biotic organic matter from comets and asteroids. *Nature*, 342:255–257, 1989.

[113] J.R. Cronin et al. *Meteorites and the Early Solar System.* Univ. Ariz. Press, Tuscon, 1988.

[114] A.I. Oparin. *Origins of Life: The Central Concepts*. Jones & Barlett, Boston, 1994.

[115] J. Xiong, W.M. Fischer, K. Inoue, M. Nakahara, and C.E. Bauer. Molecular evidence for the early evolution of photosynthesis. *Science*, 289(1724), 2000.

[116] Y. C. Cheng and R. J. Silbey. Coherence in the B800 ring of purple bacteria LH2. *Phys. Rev. Lett.*, 96:028103, 2006.

[117] C. Francke and J. Amesz. The size of the photosynthetic unit in purple bacteria. *Photosynth. Res.*, 46:347–352, 1995.

[118] S. Hess, M. Chachisvilis, K. Timpmann, M.R. Jones, C.N. Fowler, G.J. Hunter, and V. Sundström. Temporally and spectrally resolved subpicosecond energy transfer within the peripheral antenna complex (lh2) and from lh2 to the core antenna complex in photosynthetic purple bacteria. *Proc. Natl. Acad. Sci. USA*, 92(12333), 1995.

[119] K. J. Visscher, H. Bergström, V. Sunsdtröm, C. N. Hunter, and R. van Grondelle. Temperature dependence of energy transfer from the long wavelength antenna bchl-896 to the reaction center in *rhodospirillum rubrum, rhodobacter sphaeroides* (w.t. and m21 mutant) from 77 to 177k, studied by picosecond absorption spectroscopy. *Photosynth. Res.*, 22(211), 1989.

[120] H. Bergström, R. van Grondelle, and V. Sunsdström. Characterization of excitation energy trapping in photosynthetic purple bacteria at 77 k. *FEBS Lett.*, 250(503), 1989.

[121] M.O. Scully and M.S. Zubairy. *Quantum Optics*. Cambridge University Press, Cambridge, 1997.

[122] S. Scheuring, R.P. Goncalves, V. Prima, and J.N. Sturgis. The photosynthetic apparatus of *rhodopseudomonas palustris*: Structures and organization. *Journal of Molecular Biology*, 358:83–96, 2006.

[123] P.G. Adams and C.N. Hunter. Adaptation of intracytoplasmic membranes to altered light intensity in *rhodobacter sphaeroides*. *Biochimica et Biophysica Acta*, 1817:1616–1627, 2012.

[124] J. Sturgis and R. Niedermanm. Atomic force microscopy reveals multiple patterns of antenna organization in purple bacteria: implications for energy transduction mechanisms and membrane modeling. *Photosynth Res.*, 95:269–278, 2008.

[125] S. Bahatyrova, R.N. Frese, C. Alistair Siebert, J.D. Olsen, K.O. van der Werf, R. van Grondelle, R.A. Niederman, P.A. Bullough, C. Otto, and C.N. Hunter. The native architecture of a photosynthetic membrane. *Nature*, 430:1058–1062, 2004.

[126] S. Karrasch, P. A. Bullough, and R. Ghosh. The 8.5 $\AA$ projection map of the light-harvesting complex 1 from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits. *EMBO J.*, 14:631–638, 1995.

[127] S. Sheuring. Afm studies of the supramolecular assembly of bacterial photosynthetic core-complexes. *Current Opinion in Chemical Biology*, 10:387–393, 2006.

[128] M. Sener et al. Structural model and excitonic properties of the dimeric rclh1pufx complex from *rhodobacter sphaeroides*. *Chemical Phys.*, 357:188–197, 2009.

[129] R. van Grondelle, J.P. Dekker, T. Gillbro, and V. Sundström. Energy transfer and trapping in photosynthesis. *Biochim. Biophys. Acta*, 1187(1), 1994.

[130] K. Timpmann, F. G. Zhang, A. Freiberg, and V. Sundström. Detrapping of excitation energy from the reaction centre in the photosynthetic purple bacterium rhodospirillum rubrum. *Biochim. Biophys. Acta*, 1183(185), 1993.

[131] L. Liu et al. Quinone pathways in entire photosynthetic chromatophores of rhodospirillum photometricum. *J. Mol. Biol.*, 393(27), 2009.

[132] P.D. Manrique et al. Survivability of photosynthetic bacteria in non-terrestrial light. *Journal of Astrobiology*, 13(124), 2015.

[133] K.I. Goh and A.L. Barabasi. Burstiness and memory in complex systems. *Europhys. Lett.*, 81:48002, 2008.

[134] D. Borlaug, S. Fathpour, and B. Jalali. Extreme value statistics in silicon photonics. *IEEE Photonics Journal*, 1:33–39, 2009.

[135] R. Hanbury Brown and R.Q. Twiss. Interferometry of the intensity fluctuations in light. i. basic theory: The correlation between photons in coherent beams of radiation. *Proc. R. Soc.*, A242(300), 1957.

[136] R. Hanbury Brown and R.Q. Twiss. Interferometry of the intensity fluctuations in light ii. an experimental test of the theory for partially coherent light. *Proc. R. Soc.*, A243(291), 1958.

[137] S. Ragy and G. Adesso. Nature of light correlations in ghost imaging. *Scientific Reports*, 2(651), 2012.

[138] S. Ragy and G. Adesso. Unveiling the hanbury brown and twiss effect through rényi entropy correlations. *Physica Scripta*, T153, 2013.

[139] D.Z. Cao, G.J. Ge, and K. Wang. Two-photon subwavelength lithography with thermal light. *Appl. Phys. Lett.*, 97(051105), 2010.

[140] N. Sim, et al. Measurement of photon statistics with live photoreceptor cells. *Phys. Rev. Lett.*, 109(113601), 2012.

[141] P. Brumer, et al. Molecular response in one-photon absorption via natural thermal light vs. pulsed laser excitation. *Proceedings of the National Academy of Sciences*, 109(19575), 2012.

[142] R. Hildner, D. Brinks, J. B. Nieder, R. J. Cogdell, and N. F. Van Hulst. Quantum coherent energy transfer over varying pathways in single light-harvesting complexes. *Science*, 340:1448–1451, 2013.

[143] N. Christensson, H. F. Kauffmann, T. Pullerits, and T. Mančal. Origin of long-lived coherences in light-harvesting complexes. *J. Phys. Chem. B*, 116:7449–7454, 2012.

[144] V. Tiwari, W. K. Peters, and D. M. Jonas. Electronic resonance with anticorrelated pigment vibrations drives photosynthetic energy transfer outside the adiabatic framework. *PNAS*, 110:1203–1208, 2013.

[145] A. W. Chin, J. Prior, R. Rosenbach, F. Caycedo-Soler, S. F. Huelga, and M. B. Plenio. The role of non-equilibrium vibrational structures in electronic coherence and recoherence in pigment protein complexes. *Nature Phys.*, 9:113–118, 2013.

[146] C. A. Schroeder, F. Caycedo-Soler, S. F. Huelga, and M. B Plenio. Optical signatures of quantum delocalization over extended domains in photosynthetic membranes. *J. Phys. Chem. A*, 2015.

[147] Y. Zhou, J. Simon, J. Liu, and Y. Shih. Third-order correlation function and ghost imaging of chaotic thermal light in the photon counting regime. *Phys. Rev. A*, 81(043831), 2010.

[148] H. Chen, T. Peng, and Y. Shih. 100% correlation of chaotic thermal light. *Phys. Rev. A*, 88(023808), 2013.

[149] L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, Cambridge, 1995.

[150] L. J. Wang, et al. Propagation of thermal light through a dispersive medium. *JOSA*, B6(964), 1989.

[151] G.S. Agarwal, G. Gbur, and E. Wolf. Coherence properties of sunlight. *Opt. Lett.*, 29(5), 2004.

[152] A. Zardecki. Application of the functional formalism to investigation of the photoelectric counting distribution. *Can. J. Phys.*, 49(1724), 1971.

[153] G. Bedard. N-fold joint photon counting distributions associated with the photodetection of gaussian light. *Physical Review*, 161(5), 1967.

[154] C.D. Cantrell, et al. Effect of spatial coherence on the photoelectric counting statistics of gaussian light. *Phys. Rev. A*, 7(2063), 1973.

[155] L. Mandel. Photon occupation numbers in black body radiation. *JOSA*, 69(1038), 1979.

[156] S. Scheuring and J.N. Sturgis. Dynamics and diffusion in photosynthetic membranes from *rhodospirillum photometricum*. *Biophys. J.*, 91, 2006.

[157] D. Slepian. Estimation of signal parameters in the presence of noise. *Trans. IRE, Professional Group on Information Theory*, PGIT-3(82), 1954.

[158] F. Milano, A. Agostiano, F. Mavelli, and M. Trotta. Kinetics of the quinone binding reaction at the q_b site of reaction centers from the purple bacteria rhodobacter sphaeroides reconstituted in liposomes. *Eur. J. Biochem.*, 270:4595–4605, 2003.

[159] J. Kikuchi, Williamson, K. M. Shimada, and T. Asakura. Structure and dynamics of photosynthetic membrane-bound proteins in *rhodobacter sphaeroides*, studied with solid-state nmr spectroscopy. *Photosynthesis Research*, 63:259–267, 2000.

[160] A.L. Barabasi and H.E. Stanley. *Fractal Concepts in Surface Growth*. Cambridge University Press, Cambridge, 1995.

[161] J. Nagler, A. Levina, and M. Timme. Impact of single links in competitive percolation. *Nature Phys.*, 7(265), 2011.

[162] P.S. Dodds, D.J. Watts, and C.F. Sabel. Information exchange and the robustness of organizational networks. *Proc. Natl. Acad. Sci. USA*, 100(12516), 2003.

[163] O.L. Acevedo, et al. New dynamical scaling universality for quantum networks across adiabatic quantum phase transitions. *Phys. Rev. Lett.*, 112(030403), 2014.

[164] Z. Zhao, et al. Modeling the self-assembly dynamics of macromolecular protein aggregates underlying neurodegenerative disorders. *J Comput. Theor. Nanosci.*, 6(1), 2009.

[165] V. Eguíluz and M. Zimmermann. Transmission of information and herd behavior: An application to financial markets. *Phys. Rev. Lett.*, 85(5659), 2000.

[166] R. D'hulst and Rodgers G.J. Hamming distance and history distribution in the minority game. *Intern. Jour. of Theor. and Appl. Fin.*, 3(609), 2000.

[167] J.C. Bohorquez, et al. Common ecology quantifies human insurgency. *Nature*, 462(911), 2009.

[168] J. Robb. *Brave New War*. Wiley, New York, 2007.

[169] D. Kilcullen. *The Accidental Guerrilla*. Oxford University Press, Oxford, 2009.

[170] N.F. Johnson, et al. Simple mathematical law benchmarks human confrontations. *Sci. Rep.*, 3(3463), 2013.

[171] S. Galam and A. Mauger. On reducing terrorism power: a hint from physics. *Physica A*, 323:695–704, 2003.

[172] L.M. Gerdes. Mapping dark networks: A data transformation method to study clandestine organizations. *Network Science*, 2:213–253, 2014.

[173] P. Gill, J. Lee, K. Rethemeyer, J. Horgan, and V. Asal. Lethal connections: The determinants of network connections in the provisional irish republican army, 19701998. *International Interactions: Empirical and Theoretical Research in International Relations*, 40:52–78, 2014.

[174] M. Bloom, P. Gill, and J. Horgan. Tiocfaidh ár mná: Women in the provisional irish republican army. *Behavioral Sciences of Terrorism and Political Aggression*, 4(60), 2012.

[175] M.E.J. Newman. *Networks, An Introduction.* Oxford University Press, Oxford, 2012.

[176] C.G. Morselli and K. Petit. *Social Networks*, 29:143–153, 2007.

[177] W.E. Baker and R.R. Faulkner. The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. *American Sociological Review*, 58:837–860, 1993.

[178] N. Johnson et al. Human group formation in online guilds and offline gangs driven by a common team dynamic. *Phys. Rev E*, 79(066117), 2009.

[179] K.B. Whittington. *Working Paper*, 2012.

[180] L.D. Valdez et al. Predicting the extinction of ebola spreading in liberia due to mitigation strategies. *Sci. Rep.*, 5(12172), 2015.

[181] J.F. Garcia et al. Spread of pathogens in the patient transfer network of us hospitals. *e-print*, (arXiv:1504.08343), 2015.

[182] J.M. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, 2008.

[183] L. Feng et al. Competing for attention in social media under information overload conditions. *PLOS ONE*, 10(e0126090), 2015.

[184] M.V. Tomaselio et al. The role of endogenous and exogenous mechanisms in the formation of r&d networks. *Sci. Rep.*, 4(5679), 2014.

[185] N. Ramakrishnan et al. ’beating the news’ with embers: Forecasting civil unrest using open source indicators. *arXiv:*, (1402.7035), 2013.

[186] P.M. Chesler, A.M. García-García, and H. Liu. Defect formation beyond kibble-zurek mechanism and holography. *Phys. Rev. X*, 5(021015), 2015.