COMPLEXITY IN PHYSICAL, LIVING AND MATHEMATICAL SYSTEMS

Memoria que presenta para optar al grado de Doctor

LUCAS LACASA SAIZ DE ARCE
LICENCIADO EN FÍSICA

DIRECTOR DE TESIS:
BARTOLO LUQUE SERRANO, DOCTOR EN CIENCIAS FÍSICAS

MADRID, 2009
Tribunal nombrado por el Magfco. Y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día de de 200

Presidente: ____________________________________________________________
Secretario: ____________________________________________________________
Vocal: ________________________________________________________________
Vocal: ________________________________________________________________
Vocal: ________________________________________________________________
Suplente: ______________________________________________________________
Suplente: ______________________________________________________________

Realizado el cado de defensa y lectura de Tesis el día de de 200

En la E:T.S. I / Facultad

EL PRESIDENTE LOS VOCALES

EL SECRETARIO
A Isabel Saiz de Arce, mi madre.
# Table of Contents

Table of Contents v

Agradecimientos viii

Resumen en español ix

Abstract xxi

1 Complex behavior in living systems 1
   1.1 Sociophysics: How hierarchy takes place in a collective? 4
      1.1.1 Section summary 4
      1.1.2 Introduction 4
      1.1.3 The Bonabeau model 5
      1.1.4 Mean field approximation in the Bonabeau model 7
      1.1.5 Additive relaxation 15
      1.1.6 Mean field in the Stauffer version 17
      1.1.7 Conclusions 21
   1.2 Ecology: How should we design ecological reserves to allocate maximal biodiversity? 22
      1.2.1 Section summary 22
      1.2.2 Introduction 22
      1.2.3 Maximizing biodiversity: two reserves 24
      1.2.4 Generalization to \( r \) reserves 30
      1.2.5 Relaxing the neutral assumption 34
      1.2.6 Discussion 35

Bibliography 37
## 2 Complexity in mathematical systems

2.1 Phase transition and computational complexity in a number theoretic model ................................................. 43
2.1.1 Section summary .................................................. 43
2.1.2 Introduction ..................................................... 43
2.1.3 The model ......................................................... 45
2.1.4 Phase transition .................................................. 46
2.1.5 Computational complexity ...................................... 56
2.1.6 Conclusions ...................................................... 62

2.2 Scale-free topology induces Self-Organized Criticality .......................................................... 63
2.2.1 Section summary .................................................. 63
2.2.2 Introduction ..................................................... 63
2.2.3 the division model: simplest SOC system ......................... 64
2.2.4 Analytical developments: solving the model ..................... 68
2.2.5 Discussion: a new mechanism for the onset of SOC? .......... 71
2.2.6 Future work and open questions .................................. 72

2.3 Patterns in primes .................................................. 73
2.3.1 Section summary .................................................. 73
2.3.2 The first digit frequencies of primes and Riemann zeta zeros tend to uniformity following a size-dependent generalized Benford’s law ......................................................... 73
2.3.3 Explanation of the patterns ..................................... 80
2.3.4 Final remarks and applications .................................. 85
2.3.5 Appendix: Some mathematical digressions and statistical tests ......................................................... 86

### Bibliography

## 3 New methods in Complex systems analysis

3.1 Self-overlap as a method of stability analysis in Ising models .................................................. 108
3.1.1 Section summary .................................................. 108
3.1.2 Introduction ..................................................... 108
3.1.3 Ising model: damage spreading vs. self-overlap method ........ 112
3.1.4 Master equation, transition probabilities, and mean field ........ 115
3.1.5 Thermodynamics: magnetization .................................. 118
3.1.6 Stability .......................................................... 121
3.1.7 Conclusion ...................................................... 124
3.1.8 Appendix: Detailed analysis of the stability ................... 125

3.2 The visibility graph, a mapping between time series and complex networks ................................................. 128
3.2.1 Section summary ........................................... 128
3.2.2 The mapping ............................................... 128
3.2.3 The visibility graph inherits series structure ............. 130
3.2.4 Some remarks ............................................. 138
3.3 The visibility graph: a new Hurst exponent estimator ........ 140
  3.3.1 Section summary ......................................... 140
  3.3.2 Fractional Brownian motion ............................... 140
  3.3.3 The visibility graph as a new Hurst exponent estimator: ana-
    lytical developments and numerical simulations ............ 141
  3.3.4 Application to gait cycle series .......................... 146
  3.3.5 Conclusions ............................................. 147

Bibliography 149

4 Conclusions, perspectives and published papers 154
  4.1 Summary of conclusions and perspectives ...................... 154
  4.2 Publications derived from this thesis ........................ 159
    4.2.1 Chapter 1 ............................................ 159
    4.2.2 Chapter 2 ............................................ 160
    4.2.3 Chapter 3 ............................................ 160
Agradecimientos

Agradezco y dedico...

A mis padres, Isabel y Jaime, por la oportunidad de descubrir lo que me gusta y dedicarme a ello.
A mis hermanas, Alejandra y Miranda, con todo mi cariño. A la abuela Teresa.
A Tolo, mi director de tesis, con admiración: gracias por ser como eres y compartirlo conmigo. Gracias por tu confianza y tu apoyo, por todas las risas y las sonrisas.
A los colegas en la ciencia: Juancar, Octavi, Susa, Damián, Ugo, Jordi B., Jordi L., Fer, Javi G., por vuestra generosidad.
A la gente del departamento de Matemática Aplicada de Aeronáuticos: Javito, Euse, Pablo, Jose, Mario, Marta, Mariola, Ignacio, Vicky, Juan, por vuestra hospitalidad: han sido cuatro años maravillosos.
A los colegas de siempre: Manolo, Punki, Mica, Nes, Iván, Andy, Alex, Alfon, Navarro, Carlos bicho, Afri, Greg y todos vosotros que sabéis quiénes sois, por estar siempre ahí. Al dúo Frottington. A los que han costeado esta tesis con sus impuestos. A los Cinnamon.
A Ana, por muchas razones, pero sobretodo por amor.
Resumen en español

To those who do not know mathematics it is difficult to get across a real feeling as to the beauty, the deepest beauty, of nature ... If you want to learn about nature, to appreciate nature, it is necessary to understand the language that she speaks in.

-Richard Feynman

La imaginación es más importante que el conocimiento.

-Albert Einstein

La presente tesis doctoral está redactada en inglés. Son numerosos los motivos de esta elección y, para no extenderme diré que, para mal o para bien, en la actualidad el lenguaje de la ciencia es el inglés. En ese sentido creo que la mejor forma de que esta tesis acaso no acabe como tantas otras, desconocidas u olvidadas por inaccesibles, es redactarla en el lenguaje que pueda maximizar el número de lectores potenciales. Sin embargo, para que cualquier lector interesado pueda conocer, al menos de forma superficial, algunos de los resultados más importantes de esta tesis sin necesidad de entender otro idioma, presentaré en estas páginas un extenso resumen. ¿Resumen de qué? Una primera opción sería contextualizar y resumir las contribuciones que vienen desarrolladas en los diferentes capítulos. A modo de breve sinopsis, podríamos decir: la sección adecuada para un lector cansado o cuanto menos, ocupado. No duden que de esto habrá, y de hecho la primera parte de este resumen está dedicada a esta cuestión. Pero he preferido evitar escribir una disertación sesuda y árida, para resumir en cambio de forma somera algunos de los temas y problemáticas abordados,
junto con algunos resultados generales. Entiendo que aquellas personas que estén interesados en los entresijos y en ciertas técnicas particulares del trabajo, podrán seguir los desarrollos en los capítulos 1, 2 y 3. Para una lectura diagonal, recomiendo al lector ir saltando de capítulo en capítulo leyendo las introducciones, que relacionan las diferentes problemáticas específicas y describen las contribuciones originales. A su vez, los resultados más importantes de cada capítulo están recogidos en los Section Summary.

Si que aprovecharé estas líneas, y este idioma, para resumir lo que han sido para mí estos últimos cuatro años. Una época plagada de vivencias, aprendizajes, encuentros y desencuentros, que seguramente merece a modo de vista atrás mucho más que una pobre y desangelada lista con los resultados científicos más notables.

**Problemáticas y resultados importantes**

Esta tesis doctoral se engloba dentro de la Ciencia de la Complejidad o análisis de Sistemas Complejos. Sistema complejo es aquel sistema compuesto por muchos elementos en interacción, cuyo comportamiento global no puede inferirse exclusivamente del conocimiento de los comportamientos individuales. Se dice que los sistemas complejos evidencian comportamiento emergente. Un fluido, compuesto por millones de moléculas en constante interacción, es un sistema complejo: podemos conocer las propiedades físico-químicas de las moléculas de agua, pero sólo a partir de esta información, ¿cómo podríamos entender que al enfriarse el agua se congela y cambia su aspecto de forma tan dramática? Este fenómeno se presenta a una escala global del sistema y se antoja necesario tener en cuenta la totalidad del mismo. Una sociedad o un colectivo, sistema compuesto por un gran número de individuos en interacción, es también un sistema complejo: podemos intentar estudiar al individuo de forma aislada. Pero sólo a partir de conceptos individuales (psicología, fisiología, etcétera), ¿cómo podríamos predecir la difusión de las modas o de los rumores en una sociedad? ¿Cómo podemos dar cuenta de diversos fenómenos como el pánico social o el comportamiento de manada, que tienen carácter colectivo? El cerebro, compuesto por miles
de millones de neuronas en interacción constante, es de nuevo un sistema complejo: podemos conocer las reglas bioquímicas que actúan en la transmisión de información entre neuronas. Pero a partir de esta conocimiento, ¿cómo podríamos inferir el concepto de conciencia?

Todos estos ejemplos son sistemas complejos que evidencian comportamiento emergente. Cuando un sistema manifiesta tal comportamiento, un análisis reduccionista del problema (romper el sistema en sistemas más sencillos, entender lo básico, y deducir el comportamiento de lo agregado como simple suma o agregación de comportamientos básicos) no suele funcionar. Es necesario estudiar el sistema de forma global, tener una cierta visión general del cuadro. Por el carácter de dicho enfoque, suele ser necesario aplicar tanto técnicas estadísticas en los desarrollos analíticos como el uso de simulaciones por ordenador.

De forma complementaria a este punto de vista holístico, se dice que los sistemas complejos presentan universalidad. A grandes rasgos, esto significa que dos sistemas complejos diferentes en apariencia pueden evidenciar en esencia el mismo comportamiento emergente. Para muestra, un botón. Cuando reducimos la temperatura de un metal por debajo de un valor denominado temperatura de Curie, este se vuelve repentinamente un imán: de forma espontánea cobra una imanación permanente (de forma rigurosa, son los materiales ferromagnéticos los que evidencian este comportamiento, entre los cuales se encuentran muchos metales como el hierro o el níquel). El cambio brusco de las propiedades cualitativas de un material se denomina transición de fase (en particular, la imanación espontánea de los metales por debajo de la temperatura de Curie es una transición de fase de segundo orden). El metal puede por tanto encontrarse en un estado o fase de imanación, también denominada fase ferromagnética o fase ordenada o, por el contrario, encontrarse en una fase de no imanación, conocida como paramagnética o desordenada. El parámetro que hace pasar al sistema de una fase a otra se conoce como parámetro de control, y en este caso no es otro que la temperatura. ¿Cómo y por qué se produce este cambio cualitativo en el metal? Este material es parecido a un retículo formado por partículas
cuyo espín se mueve aleatoriamente en cualquier dirección. Cuando todos los espines se alinean en una misma dirección, se genera una componente neta magnética que es lo que da lugar al fenómeno de imanación. La cuestión es que cada espín está sometido en todo momento a esencialmente dos tipos de perturbaciones: por un lado, se ve afectado por los espines que le rodean, de tal modo que tiende a alinear su dirección con la de sus vecinos (esto a su vez tiende a reducir la energía total del material, en acuerdo con el principio de mínima energía). Por otro lado, cada espín sufre fluctuaciones por ruido térmico, que tienden a romper las correlaciones anteriores para aumentar la entropía o desorden interno del material. Esta competencia entre la tendencia a disminuir la energía interna (interacción entre vecinos) y aumentar la entropía (ruido térmico) es la que caracteriza la fase en la que se encuentra el material. Técnicamente se dice que la fase estable es la que tiene menor energía libre, siendo esta última un compromiso entre sendas cantidades anteriores. Por encima de la temperatura de Curie, las fluctuaciones térmicas son más fuertes que la tendencia de los espines contiguos a alinearse, de tal forma que por azar las direcciones de los espines se anulan entre sí: la fase de menor energía libre es la desordenada o paramagnética. Por debajo de la temperatura de Curie, la interacción entre vecinos es suficientemente fuerte para que los espines se alineen y que el metal se imane: estamos en la fase ordenada. La transición entre esas dos fases es además brusca, y el punto que los separa, denominado punto crítico, tiene unas características bien definidas que no dependen de las particularidades del sistema sino de propiedades generales como su dimensionalidad, sus simetrías, etc. Nótese que el modelo del metal que acabamos de describir se conoce como modelo de Ising y es un paradigma dentro de la física estadística.

El segundo sistema complejo que vamos a utilizar para describir el concepto de universalidad es una sociedad, entendida como una colección de individuos que interactúan entre sí. A priori, podríamos pensar que no existe ninguna similitud entre una sociedad y un metal. Ahora, centrámonos en el fenómeno colectivo de formación de opinión. ¿Cómo se genera consenso alrededor de una cuestión en una sociedad?
Simplificando, podemos asumir que la sociedad está formada por una colección de individuos extendidos en una región espacial que interactúan entre sí. Cada individuo tiene una opinión acerca de una cuestión pública (al igual que cada partícula de metal tiene espín en una orientación). Dicha opinión puede variar por un lado cuando el individuo discute con sus vecinos (eventualmente, un individuo puede cambiar de opinión si se ve influido por sus vecinos). Este mecanismo es semejante a la minimización de la energía interna en el metal. Por otro lado, la opinión de un individuo también se ve influida por muchos otros elementos sociales espurios que podemos catalogar como ruido social (entre otros: los medios de comunicación, mecanismos publicitarios de diferente índole, así como intereses particulares, estados de ánimo, etcétera). De forma general, diremos que una sociedad tiene consenso sobre un tema cuando todos (o la gran mayoría) opinan del mismo modo: esta sería la fase ordenada y el fenómeno de consenso en la sociedad sería análogo al fenómeno de imanación en el metal. La fase desordenada sería aquella en la que no existiera consenso: no hay correlación entre las opiniones de cada individuo. Sorprendentemente, la formación de consenso se puede modelizar como una transición de fase ferromagnética muy parecida a la que sufre un metal al disminuir la temperatura por debajo del punto de Curie. En el caso social, el parámetro que distingue la fase estable es el nivel de ruido social. Matemáticamente, el consenso social y la imanación de los metales son dos fenómenos colectivos cualitativamente iguales. Este es uno de los ingredientes de la llamada universalidad de los sistemas complejos. Diversos fenómenos emergentes cualitativamente parecidos tienen lugar en sistemas totalmente dispares. La razón de fondo: al igual que en el metal y en la sociedad, el responsable de la transición de fase son las interacciones locales, es decir, los parámetros y la forma con respecto a los cuales los elementos individuales se comportan respecto a sus vecinos y al resto. Los detalles individuales (el tipo de partícula, el tipo de ciudadano) son poco importantes. De ahí que el comportamiento emergente sea ubicuo: es robusto frente a cambios en los detalles del sistema. ¿Interaccionamos con nuestro entorno de la misma forma que interaccionan dos moléculas de agua? En cierta manera, sí y es posiblemente ésta, la
razón de que encontremos los mismos patrones a lo largo y ancho de la naturaleza.

El trabajo desarrollado en esta tesis doctoral no pretende ser una disertación acerca de un problema en particular o un sistema complejo concreto. Representa un conjunto de diferentes aportaciones puntuales con una misma temática: la descripción de fenómenos emergentes universales en sistemas complejos dispares y la descripción en cada caso del paso de lo simple (comportamiento individual, interacción) a lo complejo (comportamiento emergente) en estos sistemas. Las técnicas empleadas pertenecen al análisis de sistemas multicomponentes, entre las que se encuentran la dinámica no lineal (teoría de sistemas dinámicos, procesos estocásticos), la teoría de grafos (redes complejas) o la física estadística. A continuación se resumen los problemas abordados:

En el primer capítulo abordamos el estudio de sistemas complejos vivos. En particular, describimos en una primera parte la presencia de transiciones de fase en sistemas sociales, donde el cambio cualitativo que se percibe no es un cambio brusco de volumen (como le pasa al agua al congelarse) sino un cambio brusco en el nivel de estratificación social. De esta forma damos con un posible mecanismo generador de jerarquía (clases sociales) en una sociedad inicialmente igualitaria. Mediante tanto técnicas analíticas (estabilidad de sistemas dinámicos, teoría de campo medio) como numéricas (simulaciones de Monte Carlo, modelos de agentes) logramos caracterizar la susodicha transición así como otros patrones estructurales en la generación de jerarquía social.

En una segunda parte, abordamos otro tipo de sistemas complejos vivos: los ecosistemas. Concretamente, estudiamos el problema teórico del diseño óptimo de reservas ecológicas. Asumiendo que tenemos \( r \) reservas con áreas a priori diferentes, calculamos la distribución del área de las mismas bajo la que se maximiza la biodiversidad.

En el capítulo 2 abordamos modelos matemáticos que evidencian complejidad sistémica, en un intento por estudiar la emergencia de complejidad en sistemas lo más simple
posible. En una primera parte, analizamos un generador estocástico de números primos que evidencia una transición de fase, donde lo que cambia de forma brusca es la capacidad del algoritmo para generar primos. Estudiamos qué ingredientes del sistema son los que dan cuenta del fenómeno colectivo y encontramos relaciones interesantes entre teoría de números, física estadística y la teoría de la complejidad computacional.

En una segunda parte, proponemos un nuevo sistema basado en la división entre números enteros que se auto-organiza en un estado crítico. Relacionamos la dinámica crítica auto-organizada de este simple modelo con su estructura y deducimos que es la topología de la red de interacciones subyacente la que induce criticalidad en la dinámica del sistema. Esto constituye por un lado una nueva relación entre las redes libres de escala y la criticalidad auto-organizada, y por otro un ejemplo más de la interacción entre estructura y dinámica en un sistema complejo.

Finalmente, en la tercera parte de este capítulo presentamos un nuevo patrón estadístico tanto en la distribución de números primos como en la distribución de los ceros no triviales de la función zeta de Riemann.

En el tercer capítulo se presentan dos nuevos métodos para el análisis de Sistemas Complejos. En primer lugar, presentamos el método de autosolapamiento, que permite estudiar la estabilidad y propiedades termodinámicas de sistemas cooperativos genéricos. Aplicamos este método al modelo de Ising bidimensional para calcular tanto analítica como numéricamente propiedades termodinámicas en la transición ferromagnética.

En segundo lugar, introducimos un algoritmo que mapea series temporales a grafos, el grafo de visibilidad, mediante el que se abre un nuevo campo de exploración en el análisis de series temporales. Presentamos diferentes propiedades del método y lo aplicamos para calcular el exponente de Hurst en series fractales mediante técnicas provenientes de la teoría de redes. Finalmente en la cuarta sección presentamos un resumen de las conclusiones más importantes.
Generalmente, toda tesis pretende ser un manuscrito autocontenido, en el que tanto los conceptos como las técnicas empleadas son contextualizados y explicados. El campo de los Sistemas Complejos constituye una notable excepción a esta regla: debido a su horizontalidad y a su naturaleza interdisciplinar, se antoja difícil autocontener en un único manuscrito todos y cada uno de los conceptos y técnicas empleados. En la medida de lo posible hemos intentado contextualizar las problemáticas abordadas en cada caso, y por ello cada capítulo contiene una introducción particular. Las técnicas tanto físicas como matemáticas han sido muy variadas y por lo tanto es posible que el lector no encuentre en este manuscrito las definiciones y propiedades de todas ellas. Sin embargo hemos aportado un número elevado de referencias bibliográficas específicas en cada capítulo, e invitamos encarecidamente al lector a hacer uso de las mismas para cualquier duda que pueda surgirle en el plano técnico.

**Resumen de todo lo demás**

El primer contacto que tuve con la Ciencia de la Complejidad fue pocos días antes de inscribirme en el programa de doctorado *Física de Sistemas Complejos* que ofertaban, de modo interuniversitario, las Universidades Complutense, Politécnica y Carlos III de Madrid y la UNED. Si bien mi primera intención al terminar la carrera de Física fue la de especializarme en Física Fundamental (Relatividad General, Mecánica Cuántica), me sedujó más un programa plagado de cursos con nombres de lo más atractivo como *Teoría del Caos, Geometría Fractal, Fenómenos críticos* o *Redes Complejas*. La mayoría de los temas que se contemplaban apenas los conocía de oídas, y, por supuesto, no porque me lo hubieran contado (que no enseñado) en la facultad. Todos esos temas nuevos constituirían un mundo por explorar y comprender: ¿cuál es la relación entre lo simple y lo complejo? La universalidad de patrones, encontrar orden dentro del desorden, simetrías, la belleza: un auténtico rompecabezas aún sin ordenar. Para ser sincero, creo que este punto de vista platónico lo fui descubriendo poco a poco, con el tiempo y con discusiones de lo más iluminadores con quien además de
mi director de tesis se convirtió en uno de mis grandes amigos, Bartolo Luque. En un primer momento, yo tenía en mente problemas más concretos: la impredicibilidad del caos, sus efectos en la dinámica bursátil o en el la meteorología, en definitiva, las corrientes habituales. Sin embargo ahora puedo decir que hay un aspecto quizá más fundamental. Para explicarlo, usemos otro ejemplo: ¿Conocen el teorema de Noether? En pocas palabras, dice que por cada simetría que tiene un sistema, una cantidad del mismo se conserva. Así, un sistema con simetría traslacional (invariante bajo traslaciones) conserva el momento lineal, uno con simetría rotacional (invariante bajo rotaciones) conserva el momento angular y uno con simetría temporal (invariante bajo cambios del tipo $t \rightarrow -t$, es decir reversión temporal) conserva la energía. Este teorema es uno de los más importantes en la Física Fundamental, y es uno de los más bellos de aquellos que me contaron en la facultad. Es bello porque te dice cosas sobre la esencia última, sobre la base de cómo se comporta la naturaleza. Y es de lo más lógico. Un sistema invariante bajo reversión temporal es como una película que vista hacia delante es igual que vista hacia atrás. Pensemos en un péndulo oscilando. Si grabamos su movimiento durante un pequeño período de tiempo, veremos que presenta invariancia bajo reversión temporal: la película enseña lo mismo vista hacia delante que vista hacia atrás. Ahora bien, si esperamos el tiempo suficiente, debido al rozamiento principalmente, el péndulo se irá frenando poco a poco hasta que, después de un rato, se parará. La película claramente enseñará en este caso dos cosas diferentes si la vemos hacia delante o hacia atrás. ¿Qué ha cambiado? El rozamiento ha hecho frenar al péndulo: el movimiento del péndulo ha perdido energía (de hecho, parte de la energía del péndulo se ha disipado y convertido en calor). Por tanto, podemos concluir que si un sistema tiene invariancia bajo reversión temporal, la energía no puede disiparse. ¡Justo lo que dice el teorema de Noether! No puede resultar más sencillo y elegante. Uso este ejemplo porque me parece que ahonda en la esencia de la física: entender la naturaleza a partir de sus simetrías, de sus reglas básicas de funcionamiento, me parece algo bello. Este tipo de enfoques, dentro del mundo de los Sistemas Complejos, ¡son hoy en día problemas abiertos! ¿Cuáles son
las relaciones básicas, las propiedades últimas, que se combinan para generar complejidad en la naturaleza? ¿Cuál es la relación entre los ladrillos, lo simple, con la construcción vista como un todo?

Pero estaba explicando mis primeros contactos con los Sistemas Complejos, estoy divagando. En ocasiones tiendo a ello, y a dispersarme, también, de vez en cuando. Esto es herencia directa de mi director de tesis: el gusto por la variedad. ¿Por eso en mi tesis toco temas tan dispares como la auto-organización social, la búsqueda de patrones en números primos o la manera de caracterizar series fractales? Tal vez. O tal vez, todas sean caras de una misma moneda o tengan en esencia las mismas simetrías, la misma relación entre los ladrillos básicos y la construcción vista como un todo. Lo que quiero decir es que los Sistemas Complejos, que es el campo en donde he desarrollado esta tesis doctoral, son aquellos sistemas compuestos de muchos elementos que interaccionan entre sí (habitualmente de forma no lineal) para dar lugar a comportamientos globales que no son inferibles o entendibles únicamente a partir del conocimiento individual de los elementos. Ese comportamiento emergente es lo que distingue un sistema complicado, donde el comportamiento global se puede entender como una suma de comportamientos individuales, de un sistema complejo. Sistemas complejos son el cerebro, las sociedades, la bolsa, internet, las redes ecológicas, el clima y un larguísimo etcétera. Todos ellos formados por elementos individuales, todos ellos con relaciones básicas, simetrías, cantidades conservadas. Todos ellos exhibiendo comportamientos aparentemente impredecibles, pero con un orden oculto por descubrir. Esta tesis se encuentra a medio camino entre la descripción de sistemas que desarrollan complejidad y un buceo entre la relación entre lo simple y lo complejo.

Y sí, de nuevo me he escapado del hilo. Contaba mis primeras impresiones. Empecé a trabajar con Bartolo a los pocos meses de empezar los cursos de doctorado, a finales del año 2004. En aquellos días aprendía sobre Sistemas Dinámicos, Caos
y Transiciones de fase pero también sobre Sociología y auto-organización social. Intentábamos comprender si sería posible que en una sociedad formada por personas inicialmente iguales, se pudieran desarrollar clases sociales, jerarquía, de una forma puramente estocástica, por simple amplificación de fluctuaciones aleatorias. E intentábamos entenderlo desde las matemáticas: estudiar si era posible que ese tipo de configuraciones fueran estables o no. Acabamos escribiendo un artículo en donde presentábamos diversos resultados, tanto analíticos como numéricos, sobre varios modelos de formación de jerarquía. La respuesta fue que sí, este tipo de configuraciones eran posibles y estables, o dicho de otro modo, la igualdad social aparecía como un régimen matemáticamente inestable. También descubrimos, de paso, que uno de los modelos al uso en formación de jerarquía, era incorrecto conceptualmente. En los años posteriores de tesis cambiaron varias cosas. Empecé a descubrir el mundo de la investigación: congresos, cursos, viajes, estancias. Compartir lo aprendido y mejor aún, lo que no, con otros compañeros. Conocí Argentina, México, Turquía, Chile. Sitios increíbles y gentes maravillosas. Me da la impresión de que esta es la única faceta de la ciencia que mis amigos envidian: el poder viajar y conocer otros países y culturas ‘con el dinero de sus impuestos’. Creo que tienen razón en parte, porque poder viajar a todos esos sitios fue alucinante.

Estudié bastante, amplié mi perspectiva de los Sistemas Complejos. Dejé de centrarme en la complejidad en sistemas vivos (sociales, ecológicos) para estudiar otro tipo de sistemas más abstractos, como sistemas puramente numéricos que sorprendentemente también evidencian comportamientos emergentes. Aprendí sobre conceptos bastante dispares, como Complejidad Computacional, Teoría Local de Bifurcaciones, Teoría de Números, Redes Complejas. Entender cómo funciona el paso de lo simple a lo complejo, de lo individual a lo colectivo. Comprendí que la técnica es necesaria para resolver problemas, pero que la imaginación es la que formula las preguntas más agudas. Otra herencia de Bartolo. El fruto de todo este trabajo se vio plasmado en varias publicaciones internacionales en diferentes contextos: Ecología
teórica, Sistemas Complejos y Teoría de Números, Mecánica Estadística, Redes Complexas, Análisis de Series Temporales, Criticalidad Auto-organizada y en múltiples combinaciones de los mismos. Estos trabajos versan sobre temas como transiciones de fase algorítmicas en sistemas de números, procesos de tipo crítico auto-organizado embebidos en redes, mapeos entre series temporales y redes complejas o patrones en la distribución de números primos. El presente manuscrito compendia la síntesis de algunos de estos trabajos. Debido a la aparente disparidad de contenidos, en su día me supuso un buen dolor de cabeza el pensar cómo podría hilar los temas, para que no pareciese una simple colección de artículos inconexos. Sin embargo, creo sinceramente que una misma idea sobrevuela y auna todos ellos, la misma que llevo repitiendo en estas líneas: estudiar el paso de lo simple a lo complejo, estudiar cómo ese paso es universal, estudiar, en diferentes contextos, las relaciones básicas en sistemas de muchos cuerpos, la formación de complejidad.

Recordamos al lector que el resto de la tesis está redactada en inglés.
Abstract

From time immemorial, man has desired to comprehend the complexity of nature in terms of as few elementary concepts as possible.

-Abdus Salam

This thesis is in a first place a compendium of different works addressing the emergence of analogous complex behavior in areas of different garment, such as Social systems, Ecology, Networks, Stochastic Algorithms or Mathematical sequences. It also proposes in a second place some new methods and methodologies for Complex systems analysis. Some material is based on seven published papers and two preprints under review whose references can be found in the Concluding section. The work is divided in three general chapters, each of these having a specific introduction as well as a specific reference section.

Chapter 1 studies the onset of complex behavior in living systems. Concretely, the first part of this chapter addresses the problem of hierarchy formation in societies as a collective induced phenomenon. Two different models of hierarchy formation are studied, namely the Bonabeau model and a modified version of the former one introduced by Stauffer. These agent based models present a phase transition distinguishing a phase where the system lacks hierarchy (the so called egalitarian regime) and a phase where hierarchy grows up sharply. Applying stability analysis techniques, we introduce new analytical developments and derive results in mean field approximation that agree with the numerics.
In the second part of this chapter the task of designing ecological reserves with maximal biodiversity is studied from a probabilistic viewpoint. We analytically find the size distribution that maximizes biodiversity among a set of \( r \) reserves for a neutral case of uniform species colonization probability, and provide numerical simulations of the non neutral case.

Chapter 2 embodies the study of different mathematical systems that evidence complex behavior. In the first part we present a stochastic algorithm that generates primes. An algorithmic phase transition takes place distinguishing the ability of the algorithm to generate primes. Both Monte Carlo simulations and analytical developments are provided in order to characterize the dynamics of the system and explain what mechanisms are the responsible for the onset of the phase transition. Some connections between Computational complexity theory, Statistical physics, Number theory and Network theory are also outlined.

In the same spirit, the second part of the chapter focuses on a simple algorithm whose dynamics evidence Self-Organized Criticality (SOC). We prove that this dynamical behavior is directly related to the underlying network of interactions of the system. Specifically, we find that the underlying network’s scale free topology induces criticality in the system’s dynamics. We claim that this is a general mechanism for the onset of SOC.

Finally, in the third part of this chapter we present an as yet unnoticed statistical pattern in both the prime number distribution and the Riemann zeta zero distribution. We prove that the nature of this pattern is a consequence of the prime number theorem.

In Chapter 3 we gather some new methods and tools for Complex System analysis. In the first part we present the Self-Overlap method (SO), a method for the analysis of generic cooperative systems, and compare it with the well known Damage Spreading method (DM). We claim that SO is computationally faster than DS, analytically
simpler and lacks some of the ambiguities that DS evidences. We justify our claim by analyzing the thermodynamics and stability of the well known 2D Ising model through SO.

In the second part of the chapter we introduce the Visibility graph, an algorithm that maps time series into networks and stands as a brand new tool for time series analysis. We present the method and some of its properties, explaining how Network theory can be used to describe the properties of time series. Finally, in the third part of the chapter we show that the Visibility algorithm stands as a new method to estimate the Hurst exponent of fractal series (namely fractional Brownian motion and $f^{-\beta}$ noises). Both numerical simulations and analytical developments are outlined, as well as some applications to real time series.

We finally provide a concluding chapter that gathers the particular conclusions of each chapter as well as a list of the publications derived from this thesis.

SOME KEYWORDS

Some of the technical concepts of this thesis include:


-Stochastic processes: random walk, first return time, brownian motion, fractional Brownian motion, Hurst exponent, time series, extreme events, fractal series, noise.

-Complex network theory: degree distribution, scale free network, small world network.

-Dynamical systems: Bifurcation theory, Stability analysis, fixed points, chaos.

-Computational complexity: NP class, easy-hard-easy pattern, decision problems.

-Probability and statistics: probability distribution, goodness-of-fit tests, Benford’s law, Lagrange multipliers.
-Number theory: divisor function, counting function, Riemann zeta zero, Cramer model, natural density, logarithmic integral, prime number theorem, primitive set.

Some of the specific problems tackled in this thesis belong to:
Sociophysics, Theoretical ecology, Computer science (Computational complexity theory, algorithmics), Number theory (primitive set theory, prime number distribution), Complex Networks theory, Time series analysis, Statistical physics.
Chapter 1

Complex behavior in living systems

The molecules are like individuals, [...] and the properties of gases only remain unaltered because the number of these molecules which on the average have a given state is constant [...]. This opens a broad perspective if we do not only think of mechanical objects. Let’s consider to apply this method to the statistics of living beings, society, sociology and so forth.

-Ludwig Boltzmann

The whole is more than the sum of the parts.

-Aristotle

Living societies are composed by a plethora of heterogeneous individuals. These can be either people forming a human society, animals forming a collective, plants and insects forming an ecosystem, and so on. Individuals interact locally with other individuals according to specific behavioral rules, which usually are of a nonlinear nature: while sometimes the effect of an individual act is proportional to the action itself, typically you never know with accuracy which will be the consequences of your acts. Additionally, some external type of
driving or forcing mechanisms are also present in these systems, affecting each individual in several and heterogeneous manners. Some examples of these mechanisms include the effect of mass media in societies or the climatic aspects as regards to ecosystems. Some noise is also inevitably present in terms of stochastic deviations from the design rules. Social systems are thus an archetypical case of a complex system. In this context, physicists have shown a rapidly growing interest for a statistical physics modeling of social phenomena. The same has occurred in ecology and economics, for instance, where the same basic principles also hold. Indeed, recently multidisciplinary teams composed by theoretical physicists, mathematicians, sociologists, economists, biologists, social psychologists, etc, have started to study the development of social collective phenomena, such as the onset of consensus, the formation of hierarchy, or the population dynamics in ecosystems from a collective phenomenon perspective. Typically, these issues are analyzed within mathematical models that aim to capture the basic ingredients of the complex social phenomenon. The basic techniques include the analysis of coupled differential equations (ordinary, partial and/or stochastic) describing the evolution and interaction of individuals, and the numerical simulation of agent models (Agent based models, cellular automata, etc). These models are oversimplified versions of reality, however they are still able to describe the qualitative properties of large scale social behavior. This is reminiscent of the universality concept: in most situations, these properties do not depend on the microscopic details of the system, and only higher level features, such as the nature of the interaction rules, symmetries, conservation laws, are the responsible of the aggregated behavior. At the end, while the human being is probably the most complex system by far, when people aggregate and we look at how we behave... we’re like gases.

In this first chapter we will tackle two particular collective phenomena taking place in living systems, namely (i) the process of hierarchy formation, in particular the mechanism of stochastic hierarchy formation in many-body systems and (ii) the maximization of biodiversity
in complex ecosystems and its relation to the optimum design of species reserves.
1.1 Sociophysics: How hierarchy takes place in a collective?

1.1.1 Section summary

What basic processes generate hierarchical stratification in a collective? The Bonabeau model provides us a simple mechanism based on randomness. In this model, agents interact and self-organization is reached through both winner/looser effects and a relaxation process. A phase transition between stationary egalitarian and hierarchic states has been found in previous works. In this first section we will introduce a discrete version of the Bonabeau model. Applying a mean field approximation, we are able to reproduce previous numerical results and find a phase transition that distinguishes egalitarian from hierarchic regimes. Moreover, our discrete version of the model allows us to characterize the complex structure of the hierarchic phase. In the same philosophy, we then proceed study a more recent version of the Bonabeau model, introduced by Stauffer et al. some years ago. Several previous works described numerically the presence of a similar phase transition in this later version. We find surprising results in this context that can be interpreted as the non-existence of any phase transition in this version of Bonabeau model, but just a modification of the fixed point structure. We conclude that this later version is not a well-defined model of hierarchy formation.

1.1.2 Introduction

It is of common in sociological works to describe how global behavior appears, in many levels of social activities [1]. Before that, it is more fundamental to understand in which way citizens gather [2]; since in a little collective every single seems to display the same status, in big societies diversity appears [3]. Hierarchic dominance and hierarchic stratification has been studied with several different approaches [2, 4, 5]. As long as these matters can be
considered as many-body dynamical systems, they have attracted the attention of physicists in the latest years. The emergent area of Sociophysics involves those social complex systems, dealing with many different social situations with a statistical physics approach [1, 3, 8]. In this terms, a simple and fruitful model of diversity generation has been proposed by Bonabeau et al. [6, 7]. Related to this model, some modified versions have been proposed, as the Stauffer et al. [10] version, or the one from Ben-Naim and Redner [16] (this later one has been solved analytically).

Our purpose here is double: first, we pretend to introduce analytical developments that may provide rigor and completeness to the Bonabeau model. Using a mean field approximation in a discrete version of the Bonabeau model, we will be able to reproduce analytically some of the numerical results found by Bonabeau. We will also point out a non trivial structure in the hierarchy generation path. After this, we will apply the same scheme to Stauffer version [9, 10], a widely used model of hierarchy generation [10, 9, 13, 14, 11, 15], in order to obtain analytical evidences of its numerical behavior. To our surprise, we must finally conclude that this later version of the Bonabeau is ambiguous and ill defined, and in any case does not stand as a model of hierarchy generation by amplification of stochastic fluctuations.

1.1.3 The Bonabeau model

The Bonabeau model [6, 7] has been proposed as a simple model showing self-organization to explain hierarchic dominance in Ethology. With subtle modifications it has been reallocated in Sociophysics area as a model of social stratification [9, 10, 11, 16]. It starts from a community composed of \( N \) agents, randomly distributed over a regular lattice \( L \times L \), that is to say, with a population density \( \rho = \frac{N}{L \times L} \). Each agent \( i = 1, 2, ..., N \) is characterized by a time dependent variable \( h_i(t) \), the agent fitness, that from now on we will name status. Initially all agents share the same status \( h_i(t = 0) = 0 \): the so-called egalitarian situation. System rules are:
(1) *Competition with feedback*: an agent $i$ chosen randomly moves in a four nearest neighbor regular lattice. If the target site is empty, the agent takes the place. If it is already occupied by an agent $j$, a fight takes place. The attacker agent $i$ will defeat agent $j$ with some probability:

$$ P_{ij}(t) = \frac{1}{1 + \exp(\eta(h_j(t) - h_i(t)))}. $$

(1.1.1)

Where $\eta > 0$ is a free parameter. If $i$ wins, he exchanges positions with $j$. Otherwise, positions are maintained. After each combat, status $h_i(t)$ are updated, increasing by 1 the winner’s status and decreasing by $F$ the looser’s status. Note that $F$ is a parameter of the system that weights the defeats such that $F \geq 1$. The case $F = 1$ will be the symmetric case from now on. In the asymmetric case, $F > 1$, the effect of loosing will be more significative for the individual status than the fact of winning [9].

(2) *Relaxation*: a time step is defined as $N$ movement processes (with or without combat) in order to have a parallel updating. After each time step all agents update their status by a relaxation factor $(1 - \mu)$, such that $0 < \mu < 1$; this effect is interpreted as a fading memory of agents.

Notice that competition rule (1) is a feedback mechanism: status differences $h_j(t) - h_i(t)$ drive the future winning/loosing probabilities of agents $i$ and $j$. If agent $i$ wins/looses it’s winning/loosing probability increases afterwards. This mechanism amplifies agent inhomogeneity. On the other hand, relaxation rule (2) drives the agent status $h_i(t)$ to equalize: status differences are absorbed and toned town. The balance between both mechanisms can eventually generate asymptotic stability on $h_i(t)$. Common sense would lead us to expect low fights when the agent’s density is low, so that the relaxation mechanism would overcome and egalitarian situation would prevail ($h_i = h_j = 0, \forall i, j$). But if the system possesses high agent’s density, the rate of fights would increase, and competition mechanism would prevail, leading the system to an effective inhomogeneity. The interplay between both mechanisms
consequently depends on the density \( \rho \). Simulations ran by Bonabeau et al. [6, 7] show how this compromise between both effects brings about a phase transition at a critical density, between a global egalitarian state for low densities and a hierarchical state for high densities.

A natural measure for the status diversification is the standard deviation of its stationary distribution \( \{h_i^*\}_{i=1,\ldots,N} \). However, in [9] another measure is proposed: the standard deviation of stationary probability distribution \( \{P_{ij}\}_{i=1,\ldots,N} \), defined as:

\[
\sigma = \left( \langle P_{ij}^2 \rangle - \langle P_{ij} \rangle^2 \right)^{\frac{1}{2}}.
\]

(1.1.2)

This choice turns out to be more suitable since it is a bounded parameter: \( \sigma \in [0,1] \), and works as an order parameter of the system. For densities lower than the critical, the egalitarian state imposes that every \( h_i \) have the same value and therefore all \( P_{ij} \) too, thus \( \sigma = 0 \). Above the critical density, status diversify and their associated probabilities are consequently different: this leads to a non zero value of \( \sigma \).

1.1.4 Mean field approximation in the Bonabeau model

In order to tackle analytically tackle the system, we will obviate spatial correlations, re-defining \( \rho \) as the probability of two agents to combat. In the spatial correlated model this is equivalent to a random mixing of the agent’s positions every time step. Therefore, at each time step, an agent \( i \) will possess:

1. Probability \( 1 - \rho \) of no combat. In that case, the agent will only suffer relaxation.
2. Probability \( \rho \) of leading a combat (with probability \( 1/(N-1) \) the attacked agent will be \( j \)). In that case: agent \( i \) will increase, in average, its status by one with probability \( P_{ij}(t) \), and will decrease by \( F \) its status with probability \( 1 - P_{ij}(t) \). Relaxation will also be applied in this case.

In this mean field approximation, the dynamics of the model can be described by an \( N \)
Figure 1.1: Equality-hierarchy phase transition with control parameter $\rho$ and order parameter $\sigma$, in the mean field model of two automata ($N = 2$). From left to right we represent the transition derived from iteration of the two automata equation system 1.1.4, for different values ($F; \mu; \eta$): (2; 0.1; 1) diamonds, (1; 0.1; 1) squares, (1; 0.1; 0.5) left triangles, and (1; 0.3; 1) circles. The continuous lines are just guides for eye.

Equation system ($i = 1, \ldots, N$) of the following shape:

$$h_i(t+1) = (1-\rho)(1-\mu)h_i(t) + \frac{\rho(1-\mu)}{N-1} \sum_{j=1; j \neq i}^{N} \left\{ P_{ij}(t)(h_i(t)+1) + (1-P_{ij}(t))(h_i(t)-F) \right\}.$$  

(1.1.3)

Let us start the analysis with the simplest version $N = 2$. Having in mind that for two agents $P_{12}(t) = 1 - P_{21}(t)$, the system (1.1.3) simply reduces to:

$$h_1(t+1) = (1-\mu)h_1(t) + \rho(1-\mu)\{P_{12}(t)(1 + F) - F\}$$

$$h_2(t+1) = (1-\mu)h_2(t) + \rho(1-\mu)\{1 - P_{12}(t)(1 + F)\}.$$  

(1.1.4)

In the egalitarian phase (below the critical density), the fixed point $(h_1^*, h_2^*)$ of the two-automata system 1.1.4 will have stationary status of the same value, say $h_1^* = h_2^*$. In order to find the fixed point of the system we can define the mean status of the system as:
Figure 1.2: The two automata system switches from having one fixed point $h^* = h_2^* - h_1^* = 0$ for densities $\rho < \rho_c$, to three fixed points when $\rho > \rho_c$: $h^* = \{0, +a, -a\}$ (first one unstable and the rest stable), equivalent to the three fixed points of the two automata system, say $\{(0,0), (+a/2, -a/2), (-a/2, +a/2)\}$.

$\langle h(t) \rangle = (h_1(t) + h_2(t))/2$. The system turns into:

$$\langle h(t+1) \rangle = (1 - \mu)\langle h(t) \rangle + \rho(1 - \mu)(1 - F),$$

with a fixed point:

$$\langle h^* \rangle = \frac{\rho(1 - \mu)(1 - F)}{2\mu}, \quad (1.1.5)$$

always stable (this can be interpreted as the system’s energy, which is conserved).

Notice that in the egalitarian phase we have $h_1^* = h_2^* = \langle h^* \rangle$. In order to study the stability of $((\langle h^* \rangle, \langle h^* \rangle))$ we proceed by computing the Jacobian matrix of the system, evaluated in that fixed point:

$$J = (1 - \mu) \begin{pmatrix} 1 - A & A \\ A & 1 - A \end{pmatrix}, \quad (1.1.6)$$

where:

$$A = -\frac{\rho \sigma (1 + F)}{4}. \quad (1.1.7)$$
Figure 1.3: Bifurcation of stationary values $h_1^*$ and $h_2^*$ of the reference case (symmetric case $F = 1$ with $\eta = 1$ and $\mu = 0.1$). At the egalitarian zone we have: $h_1^* = h_2^* = \langle h^* \rangle = 0$. At $\rho = \rho_c$ a bifurcation occurs, from where, due to symmetry, $h_1^* = -h_2^* \neq 0$.

By stability analysis we conclude that egalitarian phase is stable as long as:

$$\rho < \rho_c = \frac{2\mu}{\eta(1 - \mu)(1 + F)}.$$  \hfill (1.1.8)

In figure 1.1 we plot the stationary values of the order parameter $\sigma$ as a function of the control parameter $\rho$, for the 2-automata system described by equations 1.1.4. Each of the curves correspond to specific values of the system’s parameters $F$, $\eta$ and $\mu$. The critical density $\rho_c$, expressed here as a bifurcation point, corresponds with the values of equation 1.1.8. For the symmetric case ($F = 1$) with $\eta = 1$ and $\mu = 0.1$ (squares) we obtain a critical density $\rho_c \approx 0.11$. We’ll take this particular case as the reference case from now on.

In order to get a deeper understanding of the two-automata system dynamics, let us apply the following change of variables: $h^* = h_2^* - h_1^*$. The fixed points of the system correspond to the solutions of

$$h^* = \frac{\rho(1 - \mu)(1 + F)}{\mu} \left(1 - \frac{2}{1 + \exp(\eta h^*)}\right).$$
Figure 1.4: Bifurcation of stationary values of $h_1^*$ and $h_2^*$ for values $F = 2$, $\mu = 0.1$ and $\eta = 1.0$.

Notice that for each value of $\rho$ the system has two fixed points for symmetry reasons: $(h_1^*, -h_2^*), (h_2^*, -h_1^*)$. In figure 1.2 we represent a case below transition ($\rho < \rho_c$), where we find a single solution and consequently a single degenerated fixed point. Above transition $\rho_c$ we find three solutions that correspond to three different fixed points. This feature is evidenced in figure 1.3 which plots the stable values of $h_1^*$ and $h_2^*$. In the egalitarian phase ($\rho < \rho_c$) we have: $h_1^* = h_2^* = \langle h^* \rangle = 0$ (single degenerated fixed point), at $\rho = \rho_c$ a bifurcation takes place, and in the hierarchic phase ($\rho > \rho_c$) we have $h_1^* = -h_2^* \neq 0$ (two stable and symmetric fixed points). The third fixed point of the hierarchic phase is unstable, as it should.

In the successive figures we can observe how both the stable fixed points and the bifurcation point deviate from those found in the reference case when the values of parameters $F$, $\eta$ and $\mu$ are modified. Concretely: (1) Increasing the asymmetry $F$ (figure 1.4) decreases $\rho_c$
and grows up inequality. (2) Increasing the relaxation $\mu$ (figure 1.5) increases $\rho_c$ and diminish inequality. (3) A decrease of $\eta$ (figure 1.6) increases $\rho_c$ and has a null effect on inequality.

Once the reference case has been analyzed, we will apply the same method to the general system of $N$ agents 1.1.3. We may define the mean status of the system as: $\langle h(t) \rangle = \frac{1}{N} \sum_{i=1}^{N} h_i(t)$. Now taking advantage on the fact that $P_{ij} + P_{ji} = 1$ and thus $\sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} P_{ij} = \frac{N(N-1)}{2}$, we can define an average status evolution equation:

$$\langle h(t+1) \rangle = (1-\mu)\langle h(t) \rangle - \frac{\rho(1-\mu)(F-1)}{2},$$

whose fixed point is independent of the number of automata and agrees with the very first result given at (1.1.5) in the case of two-automata (the system’s energy). Again we get that $\langle h_1^*, h_2^*, \ldots, h_N^* \rangle$ with $h_1^* = h_2^* = \cdots = h_N^* = \langle h^* \rangle$ is a fixed point of the system whose stability determines the phase (egalitarian/hierarchic). The Jacobian matrix of the
Figure 1.6: Bifurcation of stationary values of \( h_1^* \) and \( h_2^* \) for values \( F = 1, \mu = 0.1 \) and \( \eta = 0.5 \).

The linearized system, evaluated at the fixed point, is:

\[
J = (1 - \mu) \begin{pmatrix}
1 - A & \frac{A}{N-1} & \ldots & \frac{A}{N-1} \\
\frac{A}{N-1} & 1 - A & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
\frac{A}{N-1} & \ldots & \ldots & 1 - A
\end{pmatrix},
\]

a circulating matrix [12] where \( A = -\frac{\rho \eta (1+F)}{4} \) and its eigenvalues being: \( \lambda = (1 - \mu)((1 - A) - \frac{A}{N-1}) \) with multiplicity \( N - 1 \), and \( \lambda = (1 - \mu) \) with multiplicity 1. The stability of the egalitarian phase is therefore assured for

\[
\rho < \rho_c = \frac{4\mu(N-1)}{\eta(1-\mu)N(1+F)},
\]

while for \( \rho > \rho_c \) equality turns an unstable state. This result is on agreement with the particular case of two-automata \( (N = 2) \), and for \( N >> 1 \) we have:

\[
\rho_c = \frac{4\mu}{\eta(1-\mu)(1+F)}.
\]

(1.1.9)
Figure 1.7: Parameter space: delimitation, for $N \gg 1$, of the regions where transitions can be whether achieved or not achieved.

Notice that as far as $0 \leq \rho_c \leq 1$, the phase transition has a physical meaning only when the system’s parameters obey the following inequality

$$\mu < \frac{\eta(1 + F)}{4 + \eta(1 + F)}.$$  \hspace{1cm} (1.1.10)

In figure (1.7) we represent for $N \gg 1$ the parameter space, where we can distinguish the zone where the egalitarian-hierarchical transition is allowed according to equation 1.1.10.
1.1.5 Additive relaxation

Bonabeau et al. in their seminal paper [6], proposed an additive relaxation mechanism as the effect of fading memory, as an alternative to the multiplicative relaxation developed above. That additive relaxation updates the status as it follows:

\[ h_i \rightarrow h_i - \mu \tanh(h_i). \] (1.1.11)

They developed a mean field approximation accordingly, based on stochastic differential equations, and found the egalitarian-hierarchical phase transition. We can of course apply, in the discrete model that we propose here, this additive mechanism of relaxation. In the two-automata system, mean field equations would reduce to:

\[ h_1(t + 1) = h_1(t) + \rho (P_{12}(t)(1 + F) - F) - \mu \tanh(h_1) \]
\[ h_2(t + 1) = h_2(t) + \rho (1 - P_{12}(t)(1 + F)) - \mu \tanh(h_2). \]

Following our previous steps, it’s quite easy to deduce that the fixed point \( h^*_1 = h^*_2 \) is stable as long as \( \rho < \rho_c = \frac{2 \mu}{\eta(1 + F)} \). This result is on agreement with those of the continuum model proposed by Bonabeau et al. [6].

Comparing additive to multiplicative relaxation in the Bonabeau model, we must conclude that additive works somehow worse than multiplicative: status differences grow excessively in the former case, leading to eventually computational divergences (due to the exponential burst). This divergence can be overcome by introducing a new parameter \( Q \geq 0 \) (instead of increasing by one the winner status, we increase it by \( Q \), so that we can tune both winning and losing effects \( Q, F \)). In figure 1.8-a we plot the stationary values \( h^*_i \) of this latter model for \( N = 10, F = 0.7, Q = 0.7, \mu = 0.0001 \) and \( \eta = 0.001 \). As we can see, while status values are finite, they reach values of five order of magnitude, what we think is not realistic. Moreover, the hierarchy phase in the multiplicative relaxation model (figures 1.8-b-f) develops much more complexity than the additive one. An increase of density, above critical one, stills generates hierarchy, as common sense would have dictated us. This fact is
Figure 1.8: Bifurcations of stationary values of fixed point components (system 1.1.3) depending on $\rho$, for (a) the symmetric case of reference, in the additive relaxation model, when $N = 10$, (b) for the symmetric case of reference with multiplicative relaxation when $N = 3$, (c) $N = 4$, (d) $N = 6$, (e) $N = 8$ and (f) $N = 10$. 
traduced by a periodic fixed point coordinates $h_i^*$ splitting, even after the phase transition. This hierarchy growing is not trivial. In the additive relaxation model the hierarchical structure is simple, it doesn’t change with $\rho$ at hierarchical phase. Instead of that, there is a fixed point coordinates $h_i^*$ splitting at $\rho_c$, and dynamical evolution in hierarchical phase is poor.

In any case, we think that the criterium for choosing whether the additive or multiplicative relaxation process should be of a biological garment, not mathematical.

### 1.1.6 Mean field in the Stauffer version

![Graph](image)

Figure 1.9: $\sigma$ vs $\rho$ in the 2 automata system. Initial conditions for each $\rho$ are taken randomly. This leads to stationary values of $\sigma$ being zero (egalitarian zone) or non zero (hierarchical zone) depending of those chosen initial conditions.

The model developed by Stauffer et al. comes from Bonabeau’s seminal idea. It was firstly introduced to carry out the supposed lack of transition of the previous model, discussed numerically in [10, 9, 11]. In this new version, the free parameter $\eta$ is now exchanged with the order parameter $\sigma$, such that
Figure 1.10: Crossover from one stable fixed point of the system at $\rho < \rho_c$ (triangles) to two stable fixed points -egalitarian/hierarchic- (circles) and an unstable fixed point as the delimiting branch between both domains of attraction.

\[ P_{ij}(t) = \frac{1}{1 + \exp(\sigma(t)(h_j(t) - h_i(t)))}. \]  

(1.1.12)

This modification somehow introduces a dynamical feedback to the system: probability of winning/loosing is directly related to the global inequality of the system, therefore, depending on the climate’s aggressiveness of the system, agents will behave more or less aggressive themselves.

Just as in the case of Bonabeau’s model, redefining $\rho$ as the probability of two agents combat we may use a mean field approximation to reproduce analytically the phase transition that has been found numerically in previous works.

In the case of a two-automata system, we develop the mean field equations having in mind that:

1. Each automaton updates at each time step with the same dynamics that the Bonabeau model. The probability calculation is different though ($\eta \rightarrow \sigma$).
At each time step, variable $\sigma$ is updated: the order parameter is now a dynamical parameter of the system and therefore evolves with it.

We once again will redefine $h = h_2 - h_1$. With this change of variable, we reduce the dimension of the system, from a three equation system $(h_1, h_2, \sigma)$ to a two equation system $(h, \sigma)$. This can be done without lack of generality since $h_1$, $h_2$ and $\sigma$ are related. The mean field equations are therefore:

\[
h(t + 1) = (1 - \mu)h(t) + \rho(1 - \mu)(1 + F) \cdot \left(1 - \frac{2}{1 + \exp(\sigma(t)h(t))}\right)
\]

\[
\sigma(t + 1) = \left|\frac{1}{1 + \exp(\sigma(t)h(t))} - \frac{1}{2}\right|.
\]

Now, we should expect the same qualitative results concerning the stability analysis of this system than those found in Bonabeau’s model, that is to say, loss of stability of the egalitarian regime (i.e. the fixed point $h^* = h_2^* - h_1^* = 0, \sigma^* = 0$) at some critical density $\rho_c$.

The Jacobian matrix of the linearized system is

\[
J = \begin{pmatrix}
1 - \mu + AB\sigma^* & 2ABh^*\\
-\sigma^*A & -h^*A
\end{pmatrix},
\]

with

$A = \exp(\sigma^*h^*)/(1 + \exp(\sigma^*h^*))^2$,

$B = \rho(1 - \mu)(1 + F)$.

Evaluating $J$ in the egalitarian fixed point $(h^* = 0; \sigma^* = 0)$, we find that eigenvalues of $J$ are:

\[\lambda_1 = 0 < 1, \lambda_2 = 1 - \mu < 1, \forall \rho\]

We thus conclude that the egalitarian phase is stable for all densities. How come a phase transition can then take place, as claimed in many previous works [10, 9, 13, 14, 11, 3, 15]? How come hierarchical situation can be achieved starting from equality, if the egalitarian
The answer is set on the simulation methods that have been applied so far. Figure (1.9) shows the stationary values of the order parameter $\sigma$, that characterizes the status inhomogeneity, as a function of the control parameter $\rho$. For each $\rho$, we have taken different random initial conditions for status and $\sigma$ (these can be both zero or non-zero) and have iterated the evolution equations until reaching a stationary state. The figure shows that $\sigma = 0$ for all initial conditions, below a certain density. Above, we have stationary values of $\sigma$ being zero in some cases (we remain in the egalitarian zone) and non-zero in others (hierarchical zone).

We can deduce that the system has a single stable fixed point below the critical density $\rho_c$ which is actually $(h^* = 0; \sigma^* = 0)$, and two stable fixed points above $(h^* = 0; \sigma^* = 0, h^* \neq 0; \sigma \neq 0)$. The egalitarian zone is therefore always stable $(\forall \rho)$. At $\rho_c$, a saddle-node bifurcation is taking place, and brings about the hierarchic branch. Notice that the stability scheme is totally different from what we founded in the Bonabeau model: while in that model equality-hierarchy transition was generated across Pitchfork bifurcation, due to loss of stability of the egalitarian regime, in the Stauffer version the egalitarian regime is always stable, but here at $\rho_c$ a saddle-node bifurcation takes place. The stable branch of this bifurcation is related to the hierarchic regime, and the unstable branch (not drawn in figure 1.9) plays the role of frontier between the two domains of attraction. Depending what initial conditions we give to $(h, \sigma)$, (i.e. depending in which domain of attraction we start), the system, above $\rho_c$, will evolve towards the egalitarian domain or the hierarchic one.

In figure 1.10 we can understand how this stability is developed. Below $\rho_c$ the system has only one fixed point (triangles), indeed stable due to (13): every set of initial conditions $h, \sigma$ will evolve towards $h^* = 0, \sigma^* = 0$ (egalitarian zone). Above $\rho_c$ the system has three fixed points (circles), moreover, from (13) and figure 1.9 we know that upper and lower fixed point are stable and characterize both egalitarian and hierarchic zones, this leads to an unstable fixed point between them which constitutes the frontier. If the initial conditions
belong to the egalitarian basin, the system will evolve towards an asymptotic egalitarian state. On the contrary, if the initial conditions belong to the hierarchic basin, the system will evolve to an asymptotic hierarchical state.

If in the model simulations, we set egalitarian initial conditions, the result would be the “absence of transition”. But if we fix some other initial conditions, this could lead us to interpret the results as “existence of transition”. In a recent paper by Stauffer [3] they say “for the first ten Monte Carlo steps per site, $\sigma = 1$ to allow a buildup of hierarchies”. This fact probably allows initial fluctuations to grow in such a way that the system falls into the hierarchic basin.

1.1.7 Conclusions

The Bonabeau model has been criticized [10, 9, 11] in the last years. In this work we revisit Bonabeau model in order to obtain analytical evidences that give clear proof of the phase transition that the system shows. We obtain, in the model with multiplicative relaxation, a high complex structure of the hierarchical regime, a fact that we think deserves an in-depth investigation.

The Stauffer version, which is an alternative to Bonabeau model, proposed by Stauffer et al. [10, 9, 11], is the base of recent works, basically numerical [13, 14, 15]. Here we have tackled this version with the same philosophy applied in the Bonabeau model. Surprisingly, this one doesn’t show a phase transition in rigor, as far as there is no sudden growth of hierarchy if we start from equality.
1.2 Ecology: How should we design ecological reserves to allocate maximal biodiversity?

1.2.1 Section summary

In this second part of the chapter we will focus on complex ecosystems. We will develop a probabilistic approach to optimum reserve design based on the species-area relationship. Specifically, we will focus on the distribution of areas among a set of reserves maximizing biodiversity. We begin by presenting analytic solutions for the neutral case in which all species have the same colonization probability. The optimum size distribution is determined by the local-to-regional species richness ratio $k$. There is a critical $k_t$ ratio defined by the number of reserves raised to the scaling exponent of the species-area relationship. Below $k_t$, a uniform area distribution across reserves maximizes biodiversity. Beyond $k_t$, biodiversity is maximized by allocating a certain area to one reserve and uniformly allocating the remaining area to the other reserves. We then proceed by numerically exploring the robustness of our analytic results when departing from the neutral assumption of identical colonization probabilities across species.

1.2.2 Introduction

The theory of island biogeography predicts the number of species in an island as a balance between colonization and extinction events [30]. The number of species $s$ (hereafter biodiversity) of an island of area $A$ can be described by the following power-law relationship:

$$s = cA^z,$$  \hspace{1cm}(1.2.1)

where $c$ is a fitted constant and the scaling exponent $z$ has values in the range 0.2-0.4 [24]. Several explanations for the above species-area relationship have been proposed, including species abundance distributions [21], population dynamics [22], and the interplay between
a skewed species abundance distribution and intraspecific spatial aggregation [23]. The small range of empirical z-values has recently been derived from the specific form of the canonical lognormal species abundance distribution [17], which served to unify the species-area relationship with two other power laws in ecology: species frequency versus species length, and maximal body size versus area [17].

The theory of island biogeography has been used to generate simple rules of thumb in conservation biology. One classical example is the problem of choosing between one large or two small reserves. Higgs & Usher [19] used the species-area relationship and elegantly showed that the answer depends on the species overlap, that is, the fraction of common species contained in both smaller reserves. Thus, it is better to have two reserves for low overlaps, whereas one reserve maximizes biodiversity if the overlap is larger than a specific threshold.

Here we extend the one versus two reserves approach [19] for the case of multiple reserves. Given a set of r reserves, we ask the following questions: (i) what is the size distribution among these reserves that maximizes biodiversity? and (ii) how does this solution depend on the total protected area and regional diversity?

Our analytic approximation assumes neutrality. MacArthur and Wilson [30] assumed that all species are equivalent in the sense of having the same extinction and colonization rates (see also [22] for an important generalization at the individual level). However, research in island biogeography since the decade of the 1980s has unequivocally shown that species are distributed non-randomly across reserves. Specifically, due to different colonization (and/or extinction) rates, some species are more widespread than others. The observed pattern is nested, in which species inhabiting small reserves form perfect subsets of the species inhabiting larger reserves [29, 28, 27, 26, 32]. To assess to what extent these non-random patterns of species distribution affect our analytic results, we end up by analyzing numerically an extension of our model. We thus ask: (iii) how robust are our analytic results when non-neutral, species-specific colonization rates are incorporated? Our analytical
approach differs from alternative approaches in reserve design such as site-selection algorithms [20, 34, 25, 31, 33] that analyze real systems and predict the optimum set of reserves given some finite budget. Our work presents an idealized system that, although necessarily simplistic, is able to predict general and robust rules of thumb based on a few ubiquitous general laws such as the species-area relationship.

1.2.3 Maximizing biodiversity: two reserves

Let's start by illustrating the case of two reserves. Although this reproduces Higgs & Usher [19], it will be important for our generalization to \( r \) reserves in the next section. Higgs & Usher assumed a fixed area distribution between both reserves and derived the critical species overlap dictating whether it is better to have a large reserve or two small ones. Our approach in here is slightly different: we assume that we have 2 reserves (\( r \) in the following section) and are able to tune the area distribution. That is, having in mind that the total area \( A \) satisfies \( A = A_1 + A_2 \), we can determine to our convenience \( p \) satisfying \( A_1 = pA \) and \( A_2 = (1 - p)A \). Let us assume that \( n \) is the regional number of species (i.e., the total number of species in the nearby continent). Each one of these species has a probability of colonizing any of the above reserves. The number of species \( s_1 \) in reserve 1 will be:

\[
 s_1 = cA_1 = cA^2p^2. 
\]  

(1.2.2)

and similarly, the second reserve will host \( s_2 \) species given by:

\[
 s_2 = cA_2 = cA^2(1 - p)^2. 
\]  

(1.2.3)

The problem is then to calculate the value of \( p \) maximizing biodiversity, i.e., the total number of species in both reserves.

In a realistic scenario there are species with high colonization rates (these ones will likely appear in both reserves), and species with low colonization rates (we will hardly see any of
these). Let us assume the following probability distribution of reserve colonization across the $n$ species in the pool:

$$P(x) \propto x^{-\gamma},$$  \hspace{1cm} (1.2.4)

with $x = 1, 2, ..., n$.

Notice that the above probability distribution would produce a nested pattern as found in island biogeography [29, 28, 27]. For example, only the species with the highest colonization probability would be found in the far distant reserve, while this and the other species would be found in the closest reserve. That is, species in remote reserves form well-defined subsets of the species found in close reserves.

To be able to derive analytical results, we start by assuming that every species has the same colonization rate. This corresponds to the limiting case $\gamma = 0$, that is, a uniform colonization probability distribution. This neutral scenario will provide the minimum overlap between species in the two reserves. In the last section we will relax this neutral assumption.

Let’s take a number $s_1$ of different species randomly from the $n$ species pool to occupy the first reserve. For the second reserve we must choose randomly $s_2$ different species from the pool. We can now imagine that the pool has been divided in two urns: the first with $s_1$ species and the second with $n - s_1$ different species. We will compute the probability $q_m$ that, after taking $s_2$ random species, $m$ of them were actually present in the first urn. $q_m$ is thus the probability of having an overlap of $m$ common species between the two reserves.

The $s_2$ species group will be constituted by $m$ species from the urn with $s_1$ species and $s_2 - m$ from the urn with $n - s_1$ species. There are $\binom{s_1}{m}$ different, even possibilities of choosing $m$ species from the first urn. Similarly, there are $\binom{n-s_1}{s_2-m}$ different, even ways of choosing $s_2 - m$ species from the second urn. Having in mind that every choice is independent, that we assume a uniform probability distribution of colonization, and that the total number of choices is $\binom{n}{s_2}$, the probability $q_m$ of having $m$ common species is given by the hypergeometric distribution:
\[ q_m = \frac{\binom{s_1}{m}}{\binom{n}{s_2-m}}, \]  

(1.2.5)

where, if \( s_2 \geq s_1 \), \( m = 0, 1, 2, \ldots, s_1 \); and if \( s_2 \leq s_1 \), \( m = 0, 1, 2, \ldots, s_2 \).

Figure 1.11: Relative biodiversity \( B(p, k) \) versus relative reserve size \( p \) between two reserves for different values of the local-to-regional species richness ratio \( k = 0.1, 0.2, \ldots, 0.9 \) from top to bottom. Dots represent numerical simulations (average over 100 realizations, where the regional pool is \( n = 10000 \) species), and lines depict the theoretical equation (3.1.35). \( B(p, k) > 1 \) indicating that it is always better to choose two reserves to maximize biodiversity.

The mean species overlap between both reserves is determined by the mean of the hypergeometric distribution:

\[ < q > = \frac{s_1s_2}{n}, \]  

(1.2.6)

We are interested in maximizing biodiversity. Therefore, we need to maximize the following function [19]:

\[ F(p, s, n) = s_1 + s_2 - < q > = s_1 + s_2 - \frac{s_1s_2}{n}. \]  

(1.2.7)

Taking into account the species-area relationship (1.2.1, 1.2.2, 1.2.3), biodiversity is given by:
Figure 1.12: Relative biodiversity $B(p,k)$ versus relative reserve size $p$ between two reserves for different values of the local-to-regional species richness ratio $k = 0.91, 0.93, 0.95, 0.96, 0.97,$ and $0.98$ from top to bottom. Dots represent numerical simulations (average over 100 realizations, where the regional pool is $n = 10000$ species), and lines depict the theoretical equation (3.1.35). When $k > k_c \approx 0.947$, choosing one or two reserves will depend on $p$: for low to moderate values of $p$, $B(p,k) < 1$ indicating that the best option is now choosing only one reserve.

$$F(p,s,n) = s[p^n + (1-p)^n] - \frac{s^2}{n}p^n(1-p)^n.$$  \hspace{1cm} (1.2.8)

Let us define the ratio $k = s/n$, where once more $s$ is the number of species supported by a single reserve of total area $A$ (1.2.1), and $n$ is the regional species pool. $k$ is thus a local-to-regional species richness ratio; small $k$-values indicate rich continents, diverse taxons, and/or a small protected area. If we now divide equation (1.2.8) by $s$, we can define an index of relative biodiversity $B(p,k)$:

$$B(p,k) \equiv \frac{F(p,k)}{s} = p^n + (1-p)^n - kp^n(1-p)^n.$$  \hspace{1cm} (1.2.9)

The solution $B(p,k) = 1$ defines a critical line in such a way that for $B(p,k) > 1$, having two small reserves maximizes biodiversity, whereas if $B(p,k) < 1$, having only one reserve
is the best option. Note that, as long as the species pool \( n \) is larger than \( s \), \( 0 < k = s/n \leq 1 \) so as a fact of symmetry, we only have to consider the situation \( 0.5 \leq p \leq 1 \).

The behavior of \( B(p,k) \) for several values of \( k \) is plotted in figs. 1.11 and 1.12. Hereafter we assume without lack of generality \( z = 0.3 \). Note that for values of \( k \) between 0.1 and 0.9 (fig.1.11), the relative biodiversity \( B(p,k) \) is always larger than 1. This means that regardless of the reserve size distribution \( p \), it is always better to have two small reserves than a big one.

Above some critical value \( k_c = 0.94655... \), choosing one or two reserves depends strongly on the size distribution \( p \) (see fig.1.12). For low \( p \)-values, one reserve is better \( (B(p,k) < 1) \), but after a large enough \( p \)-value, two reserves maximize biodiversity as before \( (B(p,k) > 1) \). \( k_c \) can be derived easily by solving \( B(p,k)|_{p=1/2} = 1 \). The above results are summarized in figure fig.1.13, where the isocline \( B(p,k) = 1 \) is plotted in the space \( p - k \). Points \((p,k)\) below the critical line indicate situations in which two reserves maximize biodiversity.

Now, \( k \) does not only determine whether one or two reserves maximize biodiversity through

![Figure 1.13](image-url)
the critical $k_c$ value explored above. Within the domain of two reserves, there is another critical $k$ value ($k_t$) that determines the optimum size allocation between the two reserves. Note in fig.1.11) that for every value of $k \leq 0.8$, relative biodiversity reaches its maximum when $p = 0.5$, that is, for two reserves of the same size. However, for $k \geq 0.91$, $p = 0.5$ still represents an extrema of the biodiversity index, but has changed from maximum to minimum (fig. 1.12). The maximum relative biodiversity is now associated to higher values of $p$. All these conclusions can be derived in detail from the extrema analysis of $B(p, k)$. In order to find directional extrema $(p, k)^*$ of $B(p, k)$, we fix $k$. This converts $B(p, k)$ into a parametric function of $k$, say $B_k$. We then solve:

$$\frac{\partial B_k(p)}{\partial p} = 0. \quad (1.2.10)$$

A first solution of this equation is $p = 0.5 \ \forall k$. Now we tackle the second derivative, which gives information both on the function’s convexity and on the nature of the extrema. Now

![Figure 1.14: Extrema $(k, p)$ of the relative biodiversity function $B(p, k)$. Note that at the threshold $k_t \approx 0.862$ an extrema bifurcation takes place. Below this threshold, $p = 0.5$ is a maximum of $B(p, k)$. Above it, $p = 0.5$ converts into a minimum of $B(p, k)$, and a new maximum of $B(p, k)$ appears for $p > 0.5$. This maximum strongly depends on $p$.](image-url)
we can evaluate for which value of \( k = k_t \), the size allocation \( p = 0.5 \) changes from maximum to minimum. That is:

\[
\frac{\partial^2 B_k}{\partial p^2} \bigg|_{p=1/2} = 0.
\]

(1.2.11)

The solution to this equation is:

\[
k_t = (1 - z)^2 z,
\]

(1.2.12)

that in our case \((z = 0.3)\) is \( k_t \simeq 0.862 \). This is the threshold that distinguishes the domain where \( p = 0.5 \) represents either a maximum or a minimum of biodiversity. In fig. 1.14 we represent the extrema \((p, k)\) of \( B(p, k) \). We can clearly observe the extrema bifurcation: under \( k_t \), \( p = 0.5 \) (two reserves of the same size) maximizes the function. This extrema turns into a minimum above \( k_t \), and a new maximum appears with \( p > 0.5 \) (favoring an asymmetric distribution of reserves).

### 1.2.4 Generalization to \( r \) reserves

The problem can be generalized from 2 reserves to a generic number \( r \). The argument is as follows:

First, suppose again that we can determine the area distribution of the reserves, so that the area of the \( i \)-th reserve will be \( A_i = p_i A \) (with \( A = \sum_{i=1}^{m} p_i A \)). Then, for each reserve \( i \), we have:

\[
s_i = cA^2 p_i^2 = sp_i^2,
\]

(1.2.13)

where again, \( s = cA^2 \).

If we have only one reserve, the function that we should maximize will be, trivially, the constant function \( F_1 = s_1 \). We have just seen that in the case of two reserves, we could divide the pool in two urns, one with \( s_1 \) species and the other one with \( n - s_1 \). This fact leads us to maximize the function \( F_2 = F(p, s, n) = s_1 + s_2 - \frac{s_1 s_2}{n} \) giving the total number of different species in both reserves.
In the case of three reserves, we can repeat the process of dividing the pool in two urns: now the first urn will contain $F_2$ different species and the other one $n - F_2$. Reasoning as before, we would obtain a new function $F_3 = F_2 + s_3 - \frac{F_2 s_3}{n}$. We can generalize for $r$ reserves through the following recurrence equation:

$$F_r = F_{r-1} + s_r - \frac{F_{r-1} s_r}{n}.$$  \hspace{1cm} (1.2.14)

It is easy to demonstrate by induction that:

$$F_r = n \left\{ 1 - \prod_{i=1}^{r} \left( 1 - \frac{s_i}{n} \right) \right\}.$$  \hspace{1cm} (1.2.15)

Using the species-area relationship (1.2.1), defining again $k = s/n$ and dividing it by $s$, we find a generalized expression for the relative biodiversity:

$$B_r(\{p_i\}, k) \equiv \frac{F_r}{s} = \frac{1}{k} \left\{ 1 - \prod_{i=1}^{r} \left( 1 - kp_i \right) \right\}.$$  \hspace{1cm} (1.2.16)

Thus, the problem now becomes a search of the area distribution $\{p_i\}$ that maximizes $B_r(\{p_i\}, k)$. This corresponds to minimizing the following function:

$$G_r(\{x_i\}) = \prod_{i=1}^{r} (1 - x_i),$$  \hspace{1cm} (1.2.17)

for $i = 1, \ldots, r$, where we have defined the variables $x_i$ such that $x_i = k^{1/z} p_i$.

We make use of the Lagrange multipliers method to perform this task. As long as the logarithmic operator is a monotonically increasing function, the minimum of $G_r$ will coincide with the minimum of $\log(G_r)$. Applying this transformation:

$$\log G_r(\{x_i\}) = \sum_{i=1}^{r} \log (1 - x_i),$$  \hspace{1cm} (1.2.18)

will be the function to minimize. Note that:
\[
\sum_{i=1}^{r} x_i = \sum_{i=1}^{r} k^{1/z} p_i = k^{1/z} \sum_{i=1}^{r} p_i = k^{1/z},
\]
so that we can write the Lagrangian associated to (1.2.18) as:

\[
\mathcal{L} = \sum_{i=1}^{r} \log (1 - x_i^z) - \lambda \left[ \sum_{i=1}^{r} x_i - k^{1/z} \right].
\]  

(1.2.20)

Solving the system and undoing the changes to \(x_i\) we get:

\[
\lambda = \frac{p_1^{z-1}}{kp_1^z - 1} = \cdots = \frac{p_i^{z-1}}{kp_i^z - 1} = \cdots = \frac{p_r^{z-1}}{kp_r^z - 1}.
\]  

(1.2.21)

A trivial solution is the uniform distribution:

\[
p_1 = \cdots = p_i = \cdots = p_r = \frac{1}{r}.
\]  

(1.2.22)

A second solution is:

\[
p_1 = p, \quad p_i = \frac{1 - p}{r - 1}, \quad i = 2, \ldots, r.
\]  

(1.2.23)

Note that when \(r = 2\) we get our previous results. In fact, if we fix \(r = 2\) in (1.2.21), we get equation (1.2.10) as expected (the solution of Lagrange multipliers gives us the extrema).

Again, if we set \(r > 2\), we have that the uniform distribution (1.2.22) acts as a maximum until a critical value \(k_t\) is reached, from which it acts as a minimum, letting the distribution (1.2.23) act as the maximum.

From now on we will focus on the uniform case, where (1.2.22) maximizes biodiversity. Starting from equation (1.2.16) and assuming a uniform distribution (1.2.22) of reserve sizes, the relative biodiversity will be:

\[
B_r(p, k) = \frac{1}{k} \left\{ 1 - \left( 1 - \frac{k}{r^z} \right)^r \right\}.
\]  

(1.2.24)
As in the case of \( r = 2 \), we have to check that this distribution maximizes biodiversity until some threshold \( k_t \) (that is, that this distribution, being an extrema of \( B_r(p, k) \), changes from maximum to minimum). For this, we have to solve \( H_r = (h_{ij})_{r \times r} \), the Hessian of \( B_r(p, k) \), fixing \( k \) and assuming \( p = 1/r \). Thus, the diagonal terms of the Hessian will be:

\[
h_{ii} = \frac{z(z - 1)}{r^2 - 2} \left( 1 - \frac{1}{r^2} \right)^{r-1} \equiv a, \tag{1.2.25}
\]

and the non-diagonal terms:

\[
h_{ij} = -\frac{z^2 k}{r^2 z - 2} \left( 1 - \frac{1}{r^2} \right)^{r-2} \equiv b. \tag{1.2.26}
\]

Note that when we set \( r = 2 \), the conditions under which \( p = 1/r \) represents a maximum of biodiversity are \( h_{11} < 0 \) and \( |H_2| > 0 \). Solving this set of inequalities, we find again the expected solution \( k < k_t = (1 - z)2^z \).

In the general case \( r > 2 \), we proceed as follows:

The first condition is \( h_{ii} < 0 \), which is satisfied trivially \( \forall r \). The second condition is that the determinant of the Hessian changes from positive to negative at some value \( k_t \). That is, we need to find \( k_t \) that satisfies \( |H_r| = 0 \). To solve the determinant of an \( r \)-order matrix is in general a tough problem. However, due to the fact that the determinant is an algebraic invariant, we just have to diagonalize the Hessian, and ask when any eigenvalue becomes null. As a fact of symmetry, we find that the Hessian has the following shape:

\[
H_r = \begin{pmatrix}
a & b & b & \ldots & b \\
b & a & b & \ldots & b \\
b & b & a & \ldots & \ldots \\
\vdots & \vdots & \vdots & a & b \\
b & \ldots & \ldots & b & a
\end{pmatrix},
\]

which is a circulant matrix \( r \times r \) with \( r \) eigenvalues:
\( \lambda_1 = a - b \), with multiplicity \( \sigma(\lambda_1) = r - 1 \)
\( \lambda_2 = a + (j - 1)b \) with multiplicity \( \sigma(\lambda_2) = 1 \).

Hence, \( |H_r| = 0 \) provides two solutions depending on whether \( a = b \) or \( a = (1 - r)b \).

The first possibility gives us a mathematical solution with \( k_t > 1 \), which has no physical meaning. The second possibility gives us the relation:

\[
k_t = \frac{r^z(z - 1)}{z(2 - r) - 1},
\]

which is on good agreement with the case \( r = 2 \) and is the general solution of the problem.

We can conclude that in the case of \( r \) reserves, the size distribution \( p = 1/r \) maximizes biodiversity as long as the local-to-regional species richness ratio \( k \) is lower than the critical value \( k_t \). Beyond this threshold and as a fact of consistency, the size distribution that will maximize biodiversity will be the other extreme found in (1.2.23).

### 1.2.5 Relaxing the neutral assumption

Up to here we have assumed neutrality, i.e., that all species have the same colonization probability. This allowed analytic tractability. In order to see how robust previous results are in the face of relaxing neutrality, we will now present numerical results for the general case with a more realistic colonization probability distribution. Finding an analytical expression of the distribution overlap similar to equation (1.2.5) is a difficult problem when the colonization probability distribution is no longer uniform, but a power-law (equation (1.2.4)). However, we are only interested in the mean of that distribution, i.e., the mean overlap. We can assume, for a fixed \( k \), the following ansatz for the mean of that distribution:

\[
< q > = \frac{s_1 s_2}{n} P(\gamma) 2^{-\gamma},
\]

where \( P(\gamma) \) is a polynomial whose coefficients will have to be estimated through fitting. In fig. 1.15 we compare some numerical results with this ansatz for the case \( k = 0.9 \). Note that the agreement is quite good. We find as the best fitting for \( P(\gamma) \) a second order polynomial
Figure 1.15: Similar to Fig. 1 but for $r$ reserves and different colonization probability distributions described by values of $\gamma$ in equation (1.2.4). Squares represent Monte Carlo simulations and lines represent the ansatz (1.2.28). $k = 0.9$, and from top to bottom, $\gamma = 0$ (corresponding to the uniform probability distribution), 0.1, 0.25, 0.5, and 0.75. As noted, departing from neutrality ($\gamma = 0$) does not affect largely the analytic solution.

of the following shape: $P(\gamma) \approx 1.0 + 0.7\gamma + 0.41\gamma^2$. Unfortunately, we have not found a general simple ansatz so that this polynomial must be fitted for each value of $k$.

The numerical results shown in fig.1.15 clearly illustrate that for values of $\gamma < 1$, the species-specific colonization probabilities reduce relative biodiversity by less than 3%.

### 1.2.6 Discussion

We have developed a probabilistic framework to optimum reserve design. It dictates the optimum size allocation among a set of $r$ reserves. We have found that a simple variable $k$ depending on the area allocated to reserves and the regional species richness is a key determinant of the best size distribution. For high regional species richness and low reserve areas, a uniform area distribution maximizes biodiversity. For low regional species richness and high reserve areas, the optimum size allocation consists of allocating a certain area to
one reserve and uniformly distributing the remaining area among the remaining reserves. Recent research has linked the species-area relationship with two other independently derived power laws in ecology [17], namely species frequency versus species length, and maximum body size versus area. Here we add to this work by showing yet another relationship of the species-area exponent $z$. Interestingly enough, the critical value $k_t$ separating the two optimum reserve size allocation is determined by the number of reserves raised to the power-law exponent of the species-area relationship (see equation (1.2.27)). This connection between identical variables sets up the possibility of extending some of the current findings in the context of other ecological laws. For example, the commonly observed value of the exponent $z$ is related to the underlying lognormal species abundance distribution [17, 23], and thus one could explore how species abundance distributions may affect optimum reserve design. Exponent $z$ also depends on habitat and scale [23], so despite the spatially implicit assumptions of our model, such details could be incorporated through $z$.

Our analytical solutions depend only on the underlying species-area relationship, which although seems to be a good descriptor of real distributions if (i) individuals cluster in space and (ii) if abundance distribution is similar to Preston’s lognormal, it is independent on specific details of these properties [23]. This suggests that our approach is also independent on details, something which is reminiscent of the universality of complex (eco)systems.

The numerical solutions in the previous section allow us to relax the neutrality assumption. Our analytic results are robust for moderate departures from neutrality. This implies that specific complexities in the colonization rates across species would probably affect only quantitatively but not qualitatively our analytic results. This suggest the value of simple, yet general analytic predictions, which despite their simplicity can be used to provide general rules of thumb.
Bibliography


Chapter 2

Complexity in mathematical systems

*God may not play dice with the universe, but something strange is going on with the prime numbers.*

-Paul Erdos

*Nature uses as little as possible of anything.*

-Johannes Kepler

In the preceding chapter we have undertaken, by means of particular techniques coming from Statistical physics, Dynamical Systems theory and Probability theory, the task of identifying which are the key ingredients that enable collective phenomena in some particular living systems. The underlying philosophy assumes that the onset of complex behavior in multicomponent systems only depends in few aspects of such system, *e.g.* local interactions, while much of the details remain redundant and are consequently not incorporated in
the mathematical modeling. This ‘detail independence’ is a cornerstone of the universality of complex behavior. We have thus assumed that few ingredients of the system takes the major contribution of the emergence of complex structures. Now, whether this is correct or not, it is straightforward that living systems, where complexity pervades, are fully ‘intoxicated’ by the former redundant details: every living system is composed by a plethora of interconnected complex subsystems interacting at different scales. Identifying the basic ingredients that bring about the complex behavior turns out to be at least a challenging task in living systems.

In this second chapter we will be still analyzing the development of organized complexity in particular systems. However, here we make a turn and focus on the connections and mutual implications between two apparent separate fields: Complex systems and Number theory. In particular, we analyze how complex behavior, such as critical phenomena (phase transitions, self-organized criticality) or pattern formation develops on systems only made by numbers that interact with each other according to their particular number theoretical properties. We will try to unveil how the particular (local) relations between numbers, i.e. the properties of integers, are enough for the emergence of (global) complex behavior. We claim that in order to unveil the main ingredients that bring about the development of organized complexity, one should focus on rather simple models, or models where redundancy and uncertainty is reduced as much as possible. Loosely speaking, our fancy bet is that few basic internal symmetries in a system are the responsible for the onset of global complexity, and these symmetries are also present in numbers, as the most basic (i.e. less redundant) image of reality. On the other hand, we have seen, somewhat unexpectedly, number theory being applied by physicists to solve physical problems and, perhaps even more unexpectedly, techniques developed by physicists applied to problems in number theory. This chapter is also embodied in this line. Some key questions of this chapter are: Can cooperative and collective behavior develop in systems only made by numbers that interact according to their basic mathematical properties? Can we find examples of organized complexity in purely
mathematical systems? Which are the connections between the mathematical properties of these systems and the complex signatures that they eventually develop?

In what follows we will focus on three different aspects of number theory and complexity, namely: the onset of phase transitions and Self-organized criticality in systems made of numbers (sections 1 and 2), and their relation with the mathematical aspects of the models (combinatorics, computational complexity theory, complex networks theory, primitive set theory). Finally we will advance in section 3 an as yet unnoticed pattern of statistical regularity in both the prime number and the Riemann zeta zero sequences.
2.1 Phase transition and computational complexity in a number theoretic model

2.1.1 Section summary

In this section we introduce a prime number generator in the form of a stochastic algorithm. The character of such algorithm gives rise to a continuous transition which distinguishes a phase where the algorithm is able to reduce the whole system of numbers into primes and a phase where the system reaches a frozen state with low prime density. After presenting some phenomenological features characterizing the transition as well as some analytical developments, we proceed by redefining the model as a search problem, fitting it in the hallmark of computational complexity theory. We suggest that the system belongs to the class \( \text{NP} \). The computational cost is maximal around the transition threshold, as common in many algorithmic phase transitions, revealing the presence of an easy-hard-easy pattern in the system. We finally relate the nature of the phase transition to an average-case classification of the problem.

2.1.2 Introduction

Computer science and physics, although different disciplines in essence, have been closely linked since the birth of the first one. More recently, computer science has met together with statistical physics in the so called combinatorial problems and their relation to phase transitions and computational complexity (see [1] for a compendium of recent works). For instance, Erdős and Renyi, in their pioneering work on graph theory [2], found the existence of zero-one laws in their study of cluster generation. These laws have a clear interpretation in terms of phase transitions, which appear extensively in many physical systems. Computer science community has recently detected this behavior in the context of algorithmic problems [3, 4, 5, 6, 7, 8, 9, 10, 11]. The so called threshold phenomenon [1] distinguishes
zones in the phase space of an algorithm where the problem is, computationally speaking, either tractable or intractable. It is straightforward that these three phenomena can be understood as a unique concept, and in this sense to build bridges between each other seems an appealing idea.

Related to the concept of a phase transition is the task of classifying combinatorial problems. The theory of computational complexity distinguishes problems which are tractable, that is to say, solvable in polynomial time by an efficient algorithm, from those which are not. The so-called \( NP \) class gathers problems that can be solved in polynomial time by a non-deterministic Turing machine [4]. This class generally includes many hard or eventually intractable problems, although this classification is denoted \( \text{worst-case} \), that is to say, a rather pessimistic one, since the situations that involve long computations can be eventually rare. In the last years, numerical evidences suggest the presence of the threshold phenomenon in \( NP \) problems. These phase transitions may in turn characterize the \( \text{average-case} \) complexity of the associated problems, as pointed out recently [6].

Here we discuss a stochastic algorithm [12, 13, 14] that generates prime numbers by means of a stochastic integer decomposition. In section 2.1.3 we will describe the model, which stands as a stochastic prime number generator. In section 2.1.4 we will characterize the phase transition present in the system. Concretely, we will firstly outline the breaking of symmetry responsible for the order-disorder transition. An analytical approach to the system will also be considered at this point, in terms of an annealed approximation. In section 2.1.5, we will reinterpret the model as a search problem, and will analyze it from the context of computational complexity theory. We will show that the system belongs to the \( NP \) class in a worst-case classification. We point out an easy-hard-easy pattern in the algorithm, as common in many \( NP \) problems, related in turn to the critical slowing down near the the transition point. According to [6], we will finally relate the nature of the phase transition with the average-case complexity of the problem. We will show that while the
problem is in $NP$ and is consequently classified as a ‘hard’ computational problem, the resource usage only grows polynomially when the system’s size is increased, on average. In section 2.1.6 we will point out some conclusions on this topic.

2.1.3 The model

Suppose [12] that we got a pool of positive integers $\{2, 3, ..., M\}$, from which we randomly extract a certain number $N$ of them (this will constitute the system under study). Note that the chosen numbers can be repeated, and that the integer 1 is not taken into account. Now, given two numbers $n_i$ and $n_j$ taken from the system of $N$ numbers, the algorithm division rules are the following:

- Rule 1: if $n_i = n_j$ there is no division, and the numbers are not modified.

- Rule 2: If the numbers are different (say $n_i > n_j$ without loss of generality), a division will take place only if $n_j$ is a divisor of $n_i$, i.e. if $n_i \mod n_j = 0$. The algorithm’s outcome is then schematized as

$$n_i \oplus n_j \mapsto n_k \oplus n_j,$$

where $n_k = \frac{n_i}{n_j}$.  

- Rule 3: if $n_i > n_j$ but $n_i \mod n_j \neq 0$, no division takes place.

The result of a division will be the extinction of $n_i$ and the introduction of a smaller one, $n_k$.

The algorithm goes as follows: after randomly extracting from the pool $\{2, 3, ..., M\}$ a set of $N$ numbers, we pick at random two numbers $n_i$ and $n_j$ from the set. We then apply the division rules. In order to have a parallel updating, we will establish $N$ repetitions of this process ($N$ Monte Carlo steps) as a time step. Note that the algorithm rules tend to reduce numbers, hence this dynamic when iterated may generate prime numbers in the system.
We say that the system has reached stationarity when no more divisions can be achieved, whether because every number has become a prime or because rule 2 cannot be satisfied in any case -frozen state-. The algorithm then stops.

2.1.4 Phase transition

Preliminary insight

Figure 2.1: Numerical simulation of the steady values of $r$ versus $N$, for a pool size $M = 2^{14}$. Each run is averaged over $2 \cdot 10^4$ realizations in order to avoid fluctuations. Note that the system exhibits a phase transition which distinguishes a phase where every element of the system becomes a prime in the steady state and a phase with low prime density.

As stated in the previous section, this algorithm clearly tends to generate primes as far as possible: when the algorithm stops, one may expect the system to have a large number of primes or at least have a frozen state of non-divisible pairs. A first indicator that can evaluate properly this feature is the unit percentage or ratio of primes $r$, that a given system of $N$ numbers reaches at stationarity [14]. In figure 2.1 we present the results of Monte Carlo simulations calculating, as a function of $N$ and for a concrete pool size $M = 2^{14}$,
the steady values of $r$. Every simulation is averaged over $2 \cdot 10^4$ realizations in order to avoid fluctuations. We can clearly distinguish in figure 2.1 two phases, a first one where $r$ is small and a second one where the prime number concentration reaches the unity. This

![Disordered phase](image1)

![Ordered phase](image2)

Figure 2.2: The left figure stands for the steady state distribution (averaged over $2 \cdot 10^4$ realizations) of the $N$ elements, for $N = 10$ and $M = 10^4$ (phase with low $r$): this one is a uniform distribution $U(2,M)$ (note that the distribution is not normalized). The right figure stands for the same plot for $N = 110$ and $M = 10^4$ (phase where $r$ reaches the unity): this one is a power law $P(x) \sim 1/x$.

is the portrait of a phase transition, where $N$ would stand as the control parameter and $r$ as the order parameter. In the phase with small $r$, the average steady state distribution of the $N$ elements is plotted in the left side of figure (2.2): the distribution is uniform (note that the vertical scale is zoomed in such a way that if we scale it between $[0,1]$ we would see a horizontal line), which is related to an homogeneous state. In this regime, every number has the same probability to appear in the steady state. In the other hand, the average steady state distribution of the $N$ numbers in the phase of high $r$ is plotted in the right side of figure (2.2): the distribution is now a power law, which is related to a biased, inhomogeneous state. In this regime, the probability of having -in the steady state- a composite number is practically null, and the probability of having the prime $x$ is in turn proportional to $1/x$ [16]. The breaking of this symmetry between steady distributions leads us to assume an order-disorder phase transition, the phase with small proportion of
primes being the disordered phase and the ordered phase being the one where $r$ tends to one.

A second feature worth investigating is the size dependence of the transition. In figure 2.3 we plot $r$ versus $N$, for a set of different pool sizes $M$. Note that the qualitative behavior is size invariant, however the transition point increases with $M$. This size dependence will be considered later in the text.

As a third preliminary insight, we shall study the temporal evolution of the system. In figure 2.4 we plot, for a given pool size $M = 10^4$, the cumulated number of divisions that a system of $N$ numbers needs to make in order to reach stationarity. According to this, in figure 2.5 we plot, for the same $(N, M)$, the evolving value $r(t)$. In the disordered phase we can see that the system is rapidly frozen: the algorithm is not efficient in producing primes, and $r$ is asymptotically small. In the ordered phase the system needs more time to reach stationarity: this is due to the fact that the algorithm is producing many primes, as the evolving value of $r$ reaches the unity. It is however in a neighborhood of the transition
where the system takes the higher times to reach the steady state: the system is producing many divisions, but not that many primes. This fact can be related to a critical slowing down phenomenon, and is studied further in the text.

It is worth noting in figures 2.1 and 2.3 that in the disordered phase the order parameter doesn’t vanish, as it should. This is due to the fact that in a pool of $M$ numbers, following the prime number theorem, one finds on average $M/\log(M)$ primes [15]. Thus, there is always a residual contribution to the ratio $1/\log(M)$ not related to the system’s dynamics which only becomes relevant for small values of $N$, when the algorithm is not able to produce primes.

**New order parameter**

Let us now see how this phase transition can be understood as a dynamical process embedded in a network having integer numbers as the nodes. Consider two numbers of that network, say $a$ and $b$ ($a > b$). These numbers are connected ($a \rightarrow b$) if they are exactly
Figure 2.5: Ratio $r$ as a function of the time steps for the same configurations as for figure 2.4.

divisible, that is to say, if $a/b = c$ with $c$ being an integer. The topology of similar networks has been studied in [17, 18, 19], concretely in [19] it is shown that this network exhibits scale-free topology [33]: the degree distribution is $P(k) \sim k^{-\lambda}$ with $\lambda = 2$. In our system, fixing $N$ is equivalent to selecting a random subset of nodes in this network. If $a$ and $b$ are selected they eventually can give $a/b = c$; in terms of the network this means that the path between nodes $a$ and $b$ is traveled thanks to the ‘catalytic’ presence of $c$. We may say that our network is indeed a catalytic one [21, 22] where there are no cycles as attractors but two different stationary phases: (i) for large values of $N$ all resulting paths sink into primes numbers, and (ii) if $N$ is small only a few paths are traveled and no primes are reached. Notice that in this network representation, primes are the only nodes that have input links but no output links (by definition, a prime number is only divisible by the unit and by itself, acting as an absorbing node of the dynamics). When the temporal evolution of this algorithm is explored for small values of $N$, we have observed in figures 2.4, 2.5 that the steady state is reached very fast. As a consequence, there are only few traveled paths over the network and since $N$ is small the probability of catalysis is small as well, hence the paths ending in prime nodes are not traveled. We say in this case that the system
freezes in a disordered state. In contrast when \( N \) is large enough, many divisions take place and the network is traveled at large. Under these circumstances, an arbitrary node may be catalyzed by a large \( N - 1 \) quantity of numbers, its probability of reaction being high. Thus, in average all numbers can follow network paths towards the prime nodes: we say that the system reaches an ordered state.

In the light of the preceding arguments, it is meaningful to define a new order parameter \( P \) as the probability that the system has for a given \((N, M)\) to reduce every number from \( N \) into primes, that is to say, to reach an ordered state. In practice, \( P \) is calculated in the following way: given \((N, M)\), for each realization we check, once stationarity has been reached, whether the whole set of elements are primes or not, and we subsequently count the fraction of runs in which all the remaining numbers are prime in the steady state.

In figure 2.6 we plot \( P \) versus \( N \), for different pool sizes \( M \). The phase transition that the system exhibits has now a clear meaning; when \( P = 0 \), the probability that the system has to be able to reduce the whole system into primes is null (disordered state), and viceversa when \( P \neq 0 \).

In each case, \( N_c(M) \), the critical value separating the phases \( P = 0 \) and \( P \neq 0 \), can now be defined. Observe in figure 2.6 that \( N_c \) increases with the pool size \( M \). In order to describe this size dependence, we need to find some analytical argument by means of which define a system’s characteristic size. As we will see in a few lines, this one won’t be \( M \) as one would expect in a first moment.

**Annealed approximation**

The system under hands shows highly complex dynamics: correlations take place between the \( N \) numbers of the system at each time step in a non trivial way. Find an analytical solution to the problem is thus completely out of the focus of this work. However, an annealed approximation can still be performed. The main idea is to obviate these two-time correlations, assuming that at each time step, the \( N \) elements are randomly generated. This
way, we can calculate, given $N$ and $M$, the probability $q$ that at a single time step, no pair of numbers among $N$ are divisible. Thus, $1 - q$ will be the probability that there exist at least one reacting pair. Note that $1 - q$ will somehow play the role of the order parameter $P$, in this oversimplified system.

In a first step, we can calculate the probability $p(M)$ that two numbers randomly chosen from the pool $M$ are divisible:

$$p(M) = \frac{2}{(M - 1)^2} \sum_{x=2}^{[M/2]} \left\lfloor \frac{M - x}{x} \right\rfloor \approx \frac{2 \log M}{M},$$

(2.1.1)

where the floor brackets stand for the integer part function. Obviously, $1 - p(M)$ is the probability that two numbers randomly chosen are not divisible in any case. Now, in a system composed by $N$ numbers, we can make $N(N - 1)/2$ distinct pairs. However, these pairs are not independent in the present case, so that probability $q(N,M)$ isn't simply $(1 - p(M))^{N(N-1)/2}$. Correlations between pairs must be somehow taken into account. At
this point, we can make the following ansatz:

\[ q(N, M) \approx \left( 1 - \frac{2 \log M}{M} \right)^{N^{1/\alpha}} \quad (2.1.2) \]

where \( \alpha \) characterizes the degree of independence of the pairs. The relation \( 1 - q(N, M) \) versus \( N \) is plotted in figure 2.7 for different values of the pool size \( M \). Note that for a given \( M \), the behavior of \( 1 - q(N, M) \) is qualitatively similar to \( P \), the order parameter in the real system.

For convenience, in this annealed approximation we will define a threshold \( N_c \) as the one for which \( q(N_c, M) = 0.5 \). This value is the one for which half of the configurations reach an ordered state. This procedure is usual for instance in percolation processes, since the choice of the percolation threshold, related to the definition of a spanning cluster, is somewhat arbitrary in finite size systems [23]. Taking logarithms in equation (2.1.2) and expanding
up to first order, we easily find an scaling relation between $N_c$ and $M$, that reads

$$N_c \sim \left( \frac{M}{\log M} \right)^\alpha.$$  \hfill (2.1.3)

This relation firstly suggests that the system’s characteristic size is not $M$, as one would expect in a first moment, but $M/\log(M)$. In figure 2.8 we plot, in log-log, the scaling between $N_c$ and the characteristic size $M/\log(M)$ that can be extracted from figure 2.7. The best fitting provides a relation of the shape (2.1.3) where $\alpha = 0.48 \pm 0.01$ (note that the scaling is quite good, what gives consistency to the leading order approximations assumed in equation 2.1.3).

Data collapse and some remarks

The annealed approximation introduced in the preceding section suggests that the characteristic size of the system is not $M$ as one would expect but rather $M/\log(M)$. This is quite reasonable if we have in mind that the number of primes that a pool of $M$ integers
has is on average $M/\log(M)$ [15]: the quantity of primes doesn’t grow linearly with $M$.

In order to test if this scaling also applies to the prime number generator, in figure 2.9 we represent (in log-log) the values of $N_c$ (obtained numerically from the values where $P(N,M)$ becomes non-null for the first time) as a function of $M/\log(M)$. We find the same scaling relation as for the annealed system (equation 2.1.3), but with a different value for $\alpha = 0.59 \pm 0.05$ due to the obviation of correlations.

In figure 2.10 we have collapsed all curves $P(N,M)$ from figure 2.6 according to the preceding finite-size scaling relations. Note at this point that $N$ is an extensive variable, and in order to find the transition point in the thermodynamic limit, it is meaningful to define a reduced control parameter $n = \frac{N}{M/\log(M)}$, which is now an intensive variable. In the thermodynamic limit, we find $n_c = 0$: the transition exists only for finite sizes. This kind of transitions have already been noticed in other systems [?]. While in rigor the phase transition is meaningful only in the thermodynamic limit (what would lead us to assume that the

Figure 2.9: Scaling of the critical point $N_c$ versus the characteristic system’s size $M/\log(M)$ in the prime number generator, for pool size $M = \{2^{10} - 2^{18}\}$. The plot is log-log: the slope of the curve provides an exponent $\alpha = 0.59$. 

In order to test if this scaling also applies to the prime number generator, in figure 2.9 we represent (in log-log) the values of $N_c$ (obtained numerically from the values where $P(N,M)$ becomes non-null for the first time) as a function of $M/\log(M)$. We find the same scaling relation as for the annealed system (equation 2.1.3), but with a different value for $\alpha = 0.59 \pm 0.05$ due to the obviation of correlations.

In figure 2.10 we have collapsed all curves $P(N,M)$ from figure 2.6 according to the preceding finite-size scaling relations. Note at this point that $N$ is an extensive variable, and in order to find the transition point in the thermodynamic limit, it is meaningful to define a reduced control parameter $n = \frac{N}{M/\log(M)}$, which is now an intensive variable. In the thermodynamic limit, we find $n_c = 0$: the transition exists only for finite sizes. This kind of transitions have already been noticed in other systems [?]. While in rigor the phase transition is meaningful only in the thermodynamic limit (what would lead us to assume that the
Figure 2.10: Data collapse of curves \((N, P)\) for different values of \(M\), assuming the scaling relation 2.1.3. The collapse is very good, the scaling relation seems to be consistent.

former transition is only a finite size effect), for every finite size we actually can distinguish two phases and for practical purposes we can still be speaking of a pseudo-critical transition.

2.1.5 Computational complexity

As pointed out in [7], phase transitions quite similar to the former one as percolation processes for instance can be easily related to search problems. In the case under study we can redefine the system as a decision problem in the following terms: one could ask when does the clause *every number of the system is prime when the algorithm reaches stationarity* is satisfied. It is clear that through this focus, the prime number generator can be understood as a SAT-like problem [1], as long as there is an evident parallelism between the satisfiability of the preceding clause and our order parameter \(P\). Thereby, in order to study the system from the focus of computational complexity theory, we must address the following questions: which is the algorithmic complexity of the system? and how is related
the observed phase transition to the problem’s tractability?

**Worst-case classification**

The algorithm under study is related to both primality test and integer decomposition problems. Although primality was believed to belong to the so-called \( NP \) problems [25] (solvable in non-deterministic polynomial time), it has recently been shown to be in \( P \) [26]: there exists at least an efficient deterministic algorithm that tests if a number is prime in polynomial time. The integer decomposition problem is in turn a harder problem, and to find an algorithm that would factorize numbers in polynomial time is an unsolved problem of computer science. Furthermore, exploring the computational complexity of the problem under hands could eventually shed light into these aspects.

For that task, let us determine in a first place how does the search space grows when we increase \( N \). In a given time step, the search space corresponds to the set of configurations that must be checked in order to solve the decision problem: this is nothing but the number of different pairs that can be formed using \( N \) numbers. Applying basic combinatorics, the set of different configurations \( G \) for \( N \) elements and \( N/2 \) pairs is:

\[
G(N) = \frac{N!}{2^{N/2} \cdot (N/2)!} = (N - 1)!!.
\]  

(2.1.4)

We get that the search space increases with \( N \) as \( (N - 1)!! \). On the other hand, note that the decision problem is rapidly checked (in polynomial time) if we provide a candidate set of \( N \) numbers to the algorithm. These two features lead us to assume that the problem under hands belongs, in a worst-case classification [1], to the \( NP \) complexity class.

This is not surprising: the preceding sections suggested that the process is embedded in a (dynamical) scale-free network, and most likely non-planar [27]. Now, it has been shown that non-planarity in this kind of problems usually leaves to NP-completeness [28] (for instance, the Ising model in two-dimensions is, when the underlying network topology is non-planar, in \( NP \)).
Easy-hard-easy pattern

Figure 2.11: Characteristic time $\tau$ as defined in the text versus $N$, for different pool sizes, from left to right: $M = 2^{10}, 2^{11}, 2^{12}, 2^{13}, 2^{14}$. Every simulation is averaged over $2 \cdot 10^4$ realizations. Note that for each curve and within the finite size effects $\tau(N)$ reaches a maximum in a neighborhood of its transition point (this can be easily explored in figure 2.6).

An ingredient which is quite universal in the algorithmic phase transitions is the so called easy-hard-easy pattern [1]: in both phases, the computational cost of the algorithm (the time that the algorithm requires to find a solution, that is, to reach stationarity) is relatively small. However, in a neighborhood of the transition, this computational time reaches a peaked maximum. In terms of search or decision problems, this fact has a clear interpretation: the problem is relatively easy to solve as long as the input is clearly in one phase or the other, but not in between. In the system under study, the algorithm is fast in reaching an absorbing state of low concentration of primes for small $N$ because the probability of having positive divisions is small. On the other hand, the algorithm is also relatively fast in reaching an absorbing state of high concentration of primes for high $N$, because the system
Figure 2.12: Data collapse of $\tau$ for the curves of figure 2.11. The goodness of the collapse validates the scaling relations.

has enough “catalytic candidates” at each time step to be able to reduce them, the probability of having positive divisions is high. In the transition’s vicinity, the system is critical. Divisions can be achieved, however, the system needs to make an exhaustive search of the configuration space in order to find these divisions: the algorithm requires in this region much more time to reach stationarity.

Note that this easy-hard-easy pattern is related, in second order phase transitions, to the phenomenon of critical slowing down, where the relaxation time in the critical region diverges [1].

We have already seen in figure 2.4 and 2.5 that the system reaches the steady state in a different manner, depending on which phase is located the process. More properly, when $N << N_c$ (disordered phase), the system rapidly freezes, without practically achieving any reaction. When $N >> N_c$ (ordered phase), the system takes more time to reach the steady state, but it is in the regime $N \sim N_c$ where this time is maximal. In order to
be able to properly compare these three regimes, let us define a characteristic time in the system $\tau$ as the number of average time steps that the algorithm needs to take in order to reach stationarity. Remember that we defined a time step $t$ as $N$ Monte Carlo steps ($N$ operations). Thus, normalizing it over the set of numbers considered, it is straightforward to define a characteristic time as:

$$\tau(N) = \frac{t}{N}. \quad (2.1.5)$$

Note that $\tau$ can be understood as a measure of the algorithm’s time complexity [4]. In figure 2.11 we plot $\tau$ versus $N$ for a set of different pools $M = 2^{10} \ldots 2^{14}$ (simulations are averaged over $2 \cdot 10^4$ realizations). Note that given a pool size $M$, $\tau$ reaches a maximum in a neighborhood of its transition point $N_c(M)$, as can be checked according to figure 2.6. As expected, the system exhibits an easy-hard-easy pattern, as long as the characteristic time $\tau$ required by the algorithm to solve the problem has a clear maximum in a neighborhood of the phase transition. Moreover, the location of the maximum shifts with the system’s size according to the critical point scaling found in equation 2.1.3. In the other hand, this maximum also scales as:

$$\tau_{\text{max}} \left( \frac{M}{\log(M)} \right) \sim \left( \frac{M}{\log(M)} \right)^{\delta}, \quad (2.1.6)$$

where the best fitting provides $\delta = 0.13 \pm 0.1$. Note that in the thermodynamic limit, the characteristic time would diverge in the neighborhood of the transition. It is straightforward to relate this parameter with the relaxation time of a physical phase transition. According to these relations, we can collapse the curves $\tau(N, M)$ of figure 2.11 into a single universal one. In figure 2.12 this collapse is provided: the goodness of the former one supports the validity of the scaling relations.

**Average-case classification**

While Computational Complexity Theory has generally focused on worst-case classification, one may readily generalize this study to different complexity definitions. Average-case
analysis is understood as the classification of algorithms according to which is the resource usage needed to solve them on average (for instance, the execution time needed to solve an algorithm often depends on the algorithm’s inputs and consequently one can define an average execution time). This classification seems more relevant for practical purposes than the worst-case one. As a matter of fact, although NP-complete problems are generally thought of as being computationally intractable, some are indeed easy on average (however some remain complete in the average case, indicating that they are still difficult on randomly generated instances).

The system under study has been interpreted in terms of a search problem, belonging to the $NP$ class in a worst-case classification. Now, an average-case behavior, which is likely to be more useful in order to classify combinatorial problems, turns out to be tough to describe. In [6], Monasson et al. showed that there where $NP$ problems exhibit phase transitions (related to dramatic changes in the computational hardness of the problem), the order of the phase transition is in turn related to the average-case complexity of the problem. More specifically, that second order phase transitions are related to a polynomial growing of the resource requirements, instead of exponential growing, associated to first order phase transitions.

It has been shown that the system seems to exhibit a second order phase transition and an easy-hard-easy pattern. Following Monasson et al. [6], while our prime generator is likely to belong to the $NP$ class, its average-case complexity class would only be polynomial. This means that as the pool size $M$ grows, the execution time that the algorithm needs in order to solve the problem (to reach the steady state) increases polynomially on average (on average means in the ensemble of all possible initial configurations, i.e. considering many realizations with an initial random configuration). We may argue that one of the reasons of this hardness reduction is that the algorithm doesn’t realize a direct search but on the contrary this search is stochastic: the search space is not exhaustively explored. Thereby, the average behavior of the system and thus the average decision problem can be easily
solved by the algorithm, in detriment of the probable character of this solution.

2.1.6 Conclusions

In this first part of the chapter a (stochastic) algorithmic model which stands for a prime number generator has been considered. This model exhibits for every finite size a phase transition which distinguishes a phase where the algorithm has the ability to reduce every element into a prime, and a phase where the system is rapidly frozen. Analytical and numerical evidences suggest that the transition is continuous. On a second part, the model has been reinterpreted as a search problem. As long as the model searches paths to reduce integers into primes, the combinatorial problem is related to primality test and decomposition problem. It has been shown that this model belongs to the \( NP \) class in a worst-case classification, moreover, an easy-hard-easy pattern has been found, as common in many algorithmic phase transitions. According to the fact that the transition is continuous, and based on previous works, it has been put into relevance that the average-case complexity may be only polynomial. This hardness reduction is in turn related to the fact that the algorithm only yields probable states.
2.2 Scale-free topology induces Self-Organized Criticality

*Self-organized criticality is a new way of viewing nature... perpetually out-of-balance, but organized in a poised state.*

–Per Bak

2.2.1 Section summary

In this second part of the chapter we will present a general mechanism by which simple dynamics running on networks become self-organized critical for scale free topologies. We illustrate this mechanism with a simple arithmetic model of division between integers, the division model. This is the simplest self-organized critical model advanced so far, and in this sense it may help to elucidate the mechanism of self-organization to criticality. Its simplicity allows analytical tractability, characterizing several scaling relations. Furthermore, its mathematical nature brings about interesting connections between statistical physics and number theoretical concepts. We show how this model can be understood as a self-organized stochastic process embedded on a network, where the onset of criticality is induced by the topology.

2.2.2 Introduction

In the late 80s Bak, Tang and Wiesenfeld (BTW) [29, 30] introduced the concept of Self-Organized Criticality (SOC) as a mechanism explaining how multicomponent systems can evolve naturally into barely stable self-organized critical structures without external “tuning” of parameters. This single contribution sparked an enormous theoretical and experimental research interest in many areas of physics and interdisciplinary science, and many
natural phenomena were claimed to exhibit SOC [32, 31, 33]. However, there was not a
general accepted definition of what SOC exactly is, and the conditions under which it is
expected to arise. In order to disengage the mechanism of self-organization to criticality one
should likely focus on rather ‘simple’ models, and in this sense Flyvbjerg recently introduced
the “simplest SOC model” along with a workable definition of the phenomenon [34, 35],
namely ‘a driven, dissipative system consisting of a medium through which disturbances can
propagate causing a modification of the medium, such that eventually, the disturbances are
critical, and the medium is modified no more—in the statistical sense’.

On the other hand, in the last years it has been realized that the dynamics of processes
taking place on networks evidence a strong dependence on the network’s topology [76, 71].
Concretely, there exist a current interest on the possible relations between SOC behavior
and scale-free networks [71], characterized by power law degree distributions $P(k) \sim k^{-\gamma}$,
and how self-organized critical states can emerge when coupling topology and dynamics
[38, 39, 40, 41].

Here we introduce a rather simple and general mechanism by which the onset of criticality
in the dynamics of self-organized systems is induced by the scale-free topology of the under-
lying network of interactions. To illustrate this mechanism we present a simple model, the
division model from now on, based uniquely in the division between integers. We show that
this model compliances with Flyvbjerg’s definition of SOC and to our knowledge, constitutes
the simplest SOC model advanced so far that is also analytically solvable. Interestingly,
this model establishes connections between statistical physics and number theory (see [42]
for a complete bibliography on this topic).

2.2.3 the division model: simplest SOC system

In number theory, a primitive set of $N$ integers is the one for which none of the set elements
divide exactly any other element [43, 44, 58]. Consider an ordered set of $M - 1$ integers
$\{2, 3, 4, \ldots, M\}$ (notice that zero and one are excluded, and that integers are not repeated),
that we will name as the pool from now on. Suppose that we have extracted $N$ elements from
the pool to form a primitive set. The division model proceeds then by drawing integers at
random from the remaining elements of the pool and introducing them in the set. Suppose
that at time $t$ the primitive set contains $N(t)$ elements. The algorithm updating rules are
the following:

(R1) Perturbation: an integer $a$ is drawn from the pool at random and introduced in the
primitive set.

(R2) Dissipation: if $a$ divides and/or is divided by say $s$ elements of the primitive set,
then we say that an instantaneous division-avalanche of size $s$ takes place, and these latter
elements are returned to the pool, such that the set remains primitive but with a new size

$$N(t+1) = N(t) + 1 - s.$$ 

This process is then iterated, and we expect the primitive set to vary in size and composition
accordingly. The system is driven and dissipative since integers are constantly introduced
and removed from it, its size temporal evolution being characterized by $N(t)$.

In order to unveil the dynamics undergoing in the model, we have performed several Monte
Carlo simulations for different values of the pool size $M$. In the upper part of figure 2.13 we
have represented for illustration purposes a concrete realization of $N(t)$ for $M = 10^4$ and
$N(0) = 0$. Note that after a transient, $N(t)$ self-organizes around an average stable value $N_c$, fluctuating around it. In the inner part of the bottom figure 2.13, we have plotted in
log-log the power spectrum of $N(t)$: the system evidences $f^{-\beta}$ noise, with $\beta = 1.80 \pm 0.01$.
The former fluctuations are indeed related to the fact that at each time step a new integer
extracted from the pool enters the primitive set (external driving R1). Eventually (accord-
ing to rule R2), a division-avalanche can propagate and cause a modification in the size
and composition of the primitive set. These avalanches constitute the disturbances of the
system. In figure 2.14 (up) we have represented an example of the avalanche’s size evolution
in time. In the same figure (bottom) we show the probability $P(s)$ that a division-avalanche
Figure 2.13: Upper figure: Single realization of the division model showing the time evolution of the primitive set size $N(t)$ for a pool size $M = 10^4$ and $N(0) = 0$. Notice that after a transient, $N(t)$ self-organizes around an average stable value $N_c$, fluctuating around it. Bottom: (black dots) Scaling behavior of the average stable value $N_c$ as a function of the system’s characteristic size $M/\log M$. The best fitting provides $N_c \sim (M/\log M)^\gamma$, with $\gamma = 1.05 \pm 0.01$. (squares) Scaling of $N_c$ as predicted by equation 2.2.8. Inner figure: plot in log-log of the power spectrum of $N(t)$, showing $f^{-\beta}$ noise with $\beta = 1.80 \pm 0.01$ (this latter value is the average of $10^5$ realizations of $N(t)$ for 4096 time steps after the transient and $M = 10^4$).

of size $s$ takes place, for different pool sizes $M$. These latter distributions are power laws $P(s) \sim s^{-\tau} \exp(s/s_0)$ with $\tau = 2.0 \pm 0.1$: disturbances are thus critical. Observe that the power law relation suffers from a crossover to exponential decay at a cut-off value $s_0$ due to finite size effects (pool is finite), and that the location of these cut-offs scales with the system’s characteristic size $s_0 \sim (M/\log M)^\omega$ with $\omega = 1.066 \pm 0.003$, what is typically characteristic of a finite size critical state [31] (the characteristic size $M/\log M$ will be explained later in the text). We can conclude that according to Flyvbjerg’s definition [34], the division model exhibits SOC. Division-avalanches drive the system to different marginally
Figure 2.14: Upper figure: Single realization of the division model showing the time distribution of division-avalanches. Bottom figure: Probability distribution $P(s)$ that a division-avalanche of size $s$ takes place in the system, for different pool sizes $M = 2^{10}$ (triangles), $M = 2^{11}$ (inverted triangles), $M = 2^{12}$ (diamonds) and $M = 2^{13}$ (circles). In every case we find $P(s) \sim s^{-\tau} \exp(s/s_0)$ with $\tau = 2.0 \pm 0.1$. Note that the power law relation evidences an exponential cut-off due to finite size effects at particular values of $s_0$. Inner figure: Scaling of the cut-off value $s_0$ as a function of the system’s characteristic size $M/\log M$, with an exponent $\omega = 1.066 \pm 0.003$.

stable states, that are nothing but primitive sets of different sizes and composition. Accordingly, for a given pool $[2, M]$, these time fluctuations generate a stochastic search in the configuration space of primitive sets.
2.2.4 Analytical developments: solving the model

In what follows we discuss analytical insights of the problem. Consider the divisor function \([48]\) that provides the number of divisors of \(n\), excluding integers 1 and \(n\):

\[
d(n) = \sum_{k=2}^{n-1} \left( \left\lfloor \frac{n}{k} \right\rfloor - \left\lfloor \frac{n-1}{k} \right\rfloor \right),
\]

where \(\lfloor \cdot \rfloor\) stands for the integer part function. The average number of divisors of a given integer in the pool \([2, M]\) is then:

\[
\frac{1}{M-1} \sum_{n=3}^{M} d(n) = \frac{1}{M-1} \sum_{k=2}^{M} \left\lfloor \frac{M}{k} \right\rfloor \approx \sum_{k=2}^{M} \frac{1}{k} \approx \log M + 2(\gamma - 1) + O\left(\frac{1}{\sqrt{M}}\right).
\]

Accordingly, the mean probability that two numbers \(a\) and \(b\) taken at random from \([2, M]\) are divisible is approximately \(P = Pr(a|b) + Pr(b|a) \approx 2 \log M/M\). Moreover, if we assume that the \(N\) elements of the primitive set are uncorrelated, the probability that a new integer generates a division-avalanche of size \(s\) is on average \((2 \log M/M)N\). We can consequently build a mean field equation for the system’s evolution, describing that at each time step an integer is introduced in the primitive set and a division-avalanche of mean size \((2 \log M/M)N\) takes place:

\[
N(t + 1) = N(t) + 1 - \left(\frac{2 \log M}{M}\right)N(t),
\]

whose fixed point \(N_c = M/(2 \log M)\), the stable value around which the system self-organizes, scales with the system’s size as

\[
N_c(M) \sim \frac{M}{\log M}.
\]

Hitherto, we can conclude that the system’s characteristic size is not \(M\) (pool size) as one should expect in the first place, but \(M/\log M\). This scaling behavior has already been noticed in other number-theoretic models evidencing collective phenomena \([46, 47]\) such as
the prime number generator analyzed in the previous section. In figure 2.13 we have plotted (black dots) the values of $N_c$ as a function of the characteristic size $M/\log M$ provided by Monte Carlo simulations of the model for different pool sizes $M = 2^8, 2^9, ..., 2^{15}$ ($N_c$ has been estimated averaging $N(t)$ in the steady state). Note that the scaling relation 2.2.4 holds, however the exact numerical values $N_c(M)$ are underestimated by eq.2.2.3. This is reasonable since we have assumed that the primitive set elements are uncorrelated, what is obviously not the case: observe for instance that any prime number $p \geq \lfloor M/2 \rfloor$ introduced in the primitive set will remain there forever. Fortunately this drawback of our mean field approximation can be improved by considering the function $D(n)$ that defines the exact number of divisors that a given integer $n \in [2, M]$ has, i.e. the amount of numbers in the pool that divide or are divided by $n$:

$$D(n) = d(n) + \left\lfloor \frac{M}{n} \right\rfloor - 1. \quad (2.2.5)$$

Define $p_n(t)$ as the probability that the integer $n$ belongs at time $t$ to the primitive set. Then, we have

$$p_n(t+1) = \left(1 - \frac{D(n)}{M - N(t)}\right)p_n(t) + \frac{1}{M - N(t)}(1 - p_n(t)), \quad (2.2.6)$$

that leads to a stationary survival probability in the primitive set:

$$p_n^* = \frac{1}{1 + D(n)}. \quad (2.2.7)$$

In Fig.2.15 (right) we depict the stationary survival probability of integer $n$ (black dots) obtained through numerical simulations for a system with $M = 50$, while squares represent the values of $p_n^*$ as obtained from the eq.2.2.7. Note that there exists a remarkable agreement. We now proceed to estimate the critical size values $N_c(M)$ as:

$$N_c(M) \approx \sum_{n=2}^{M} p_n^* = \sum_{n=2}^{M} \frac{1}{1 + D(n)}. \quad (2.2.8)$$

In the bottom of figure 2.13 we have represented (squares) the values of $N_c(M)$ predicted by eq.2.2.8, showing good agreement with the numerics (black dots).
Figure 2.15: Left: Histogram of the amount of integers in \([2, 10^6]\) that have \(D\) divisors. The histogram have been smoothed (binned) to reduce scatter. The best fitting provides a power law \(P(D) \sim D^{-\tau}\) with \(\tau = 2.01 \pm 0.01\), in agreement with \(P(s)\) (see the text). Right: (black dots) Stationary survival probability of integer \(n\) in a primitive set for a pool size \(M = 50\), obtained from Monte Carlo simulations of the model over \(10^6\) time steps (a preliminary transient of \(10^4\) time steps was discarded). (squares) Theoretical prediction of these survival probabilities according to equation 2.2.7.

Finally, previous calculations point out that system’s fluctuations, i.e. division-avalanches distribution \(P(s)\) is proportional to the percentage of integers having \(s\) divisors. In order to prove this conjecture, in figure 2.15 (left) we have plotted a histogram describing the amount of integers having a given number of divisors, obtained from computation of \(D(n)\) for \(M = 10^6\). The tail of this histogram follows a power law with exponent \(\tau = 2.0\).

This can be proved analytically as it follows: the responsible for the tail of the preceding histogram are those numbers that divide many others, i.e. rather small ones \((n \ll M)\). A small number \(n\) divides typically \(D(n) \simeq \lfloor \frac{M}{n}\rfloor\). Now, how many ‘small numbers’ have \(D(n)\) divisors? The answer is \(n, n+1, \ldots, n+z\) where

\[
\left\lfloor \frac{M}{n} \right\rfloor = \left\lfloor \frac{M}{n-1} \right\rfloor = \ldots = \left\lfloor \frac{M}{n-z} \right\rfloor.
\]

(2.2.9)

The maximum value of \(z\) fulfills \(\frac{M}{n-z} - \frac{M}{n} = 1\), that is \(z \simeq n^2/M\). The frequency of \(D(n)\) is thus \(fr(D(n)) = n^2/M\), but since \(s \equiv D(n) \simeq M/n\), we get \(fr(s) \sim Ms^{-2}\), and finally
normalizing, \( P(s) \sim s^{-2} \).

### 2.2.5 Discussion: a new mechanism for the onset of SOC?

Coming back to the Flyvbjerg’s definition of SOC, which is the medium in the division model? Observe that the process can be understood as embedded in a network, where nodes are integers, and two nodes are linked if they are exactly divisible. The primitive set hence constitutes a subset of this network, that is dynamically modified according to the algorithm’s rules. The degree of node \( n \) is \( D(n) \), and consequently the degree distribution is \( P(k) \sim k^{-2} \): the network is scale-free. Hence the SOC behavior, which arises due to the divisibility properties of integers, can be understood as a sort of anti-percolation process taking place in this scale-free network. Observe that the division model is a particular case of a more general class of self-organized models: a network with \( M \) nodes having two possible states (on/off) where the following dynamics runs: (R1) perturbation: at each time step a node in the state off is randomly chosen and switched on, (R2) dissipation: the \( s \) neighbors of the perturbed node that were in the state on in that time step are switched off, and we say that an instantaneous avalanche of size \( s \) has taken place. \( N(t) \) measures the number of nodes in the state on as a function of time. Its evolution follows a mean field equation that generalizes eq. 2.2.3:

\[
N(t+1) = N(t) + 1 - \frac{\langle k \rangle}{M} N(t),
\]

(2.2.10)

where \( \langle k \rangle \) is the network’s mean degree. Accordingly, in every case \( N(t) \) will self-organize around an average value \( N_c(M) \). Within regular or random networks, fluctuations (avalanches) around \( N_c(M) \) will follow a Binomial or Poisson distribution respectively. However, when the network is scale free with degree distribution \( P(k) \sim k^{-\gamma} \), fluctuations will follow a power law distribution \( P(s) \sim s^{-\tau} \) with \( \tau = \gamma \), and the dynamics will consequently be SOC. In this sense, we claim that scale-free topology induces criticality.
2.2.6 Future work and open questions

Some questions concerning this new mechanism can be depicted, namely: which is the relation between the specific topology of scale-free networks and the power spectrum of the system’s dynamics? Which physical or natural systems evidence this behavior?

With regard to the division model, the bridge between statistical physics and number theory should also be investigated in depth. This includes possible generalizations of this model to other related sets such as $k$-primitive sets [49], where every number divides or is divided by at least $k$ others ($k$ acting as a threshold parameter), to relatively primitive sets [50] and to cross-primitive sets [44] (where this will introduce coupled SOC models). From the computational viewpoint [1], properties of the model as a primitive set generator should also be studied. Of special interest is the task of determining the maximal size of a $k$-primitive set [44, 49], something that can be studied within the division model through extreme value theory [32].
2.3 Patterns in primes

2.3.1 Section summary

Prime numbers seem to distribute among the natural numbers with no other law than that of chance, however its global distribution presents a quite remarkable smoothness. Such interplay between randomness and regularity has motivated scientists of all ages to search for local and global patterns in this distribution that eventually could shed light into the ultimate nature of primes. In this last part of the chapter we show that a generalization of the well known first-digit Benford’s law, which addresses the rate of appearance of a given leading digit $d$ in data sets, describes with astonishing precision the statistical distribution of leading digits in the prime numbers sequence. Moreover, a reciprocal version of this pattern also takes place in the sequence of the nontrivial Riemann zeta zeros. We prove that the prime number theorem is, in the last analysis, the responsible of these patterns. Some new relations concerning the prime numbers distribution are also deduced, including a new approximation to the counting function $\pi(n)$. We finally point out several applications of the generalized Benford’s law formalism, from its relation with random matrix theory to stock market analysis or fraud detection.

2.3.2 The first digit frequencies of primes and Riemann zeta zeros tend to uniformity following a size-dependent generalized Benford’s law

The individual location of prime numbers within the integers seems to be random, however its global distribution exhibits a remarkable regularity [51]. Certainly, this tenseness between local randomness and global order has lead the distribution of primes to be, since antiquity, a fascinating problem for mathematicians [52] and more recently for physicists
The Prime Number Theorem, that addresses the global smoothness of the counting function \( \pi(n) \) providing the number of primes less or equal to integer \( n \), was the first hint of such regularity \[89\]. Some other prime patterns have been advanced so far, from the visual Ulam spiral \[56\] to the arithmetic progression of primes \[57\], while some others remain conjectures, like the global gap distribution between primes or the twin primes distribution \[89\], enhancing the mysterious interplay between apparent randomness and hidden regularity. There are however many open problems to be solved, and the prime number distribution is yet to be understood \[58, 59, 83\]. For instance, there exist deep connections between the prime number sequence and the nontrivial zeros of the Riemann zeta function \[54, 61\]. The celebrated Riemann Hypothesis, one of the most important open problem in mathematics, states that the nontrivial zeros of the complex-valued Riemann zeta function \( \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \) are all complex numbers with real part \( 1/2 \), the location of these being intimately connected with the prime number distribution \[61, 62\].

Here we address the statistics of the first significant or leading digit of both the sequences of primes and the sequence of Riemann nontrivial zeta zeros. The leading digit of a number stands for its non-zero leftmost digit. For instance, the leading digits of the prime 7703 and the zeta zero 21.022... are 7 and 2 respectively. The most celebrated leading digit distribution is the so called Benford’s law, after physicist Frank Benford \[63\] who empirically found that in many disparate natural data sets and mathematical sequences, the leading digit \( d \) wasn’t uniformly distributed as might be expected, but instead had a biased probability of appearance

\[
P(d) = \log_{10}(1 + 1/d),
\]  

(2.3.1)

where \( d = 1, 2, \ldots, 9 \). While this empirical law was indeed firstly discovered by astronomer Simon Newcomb in 1881 \[64\], it is popularly known as the Benford’s law or alternatively as the Law of Anomalous Numbers. Several disparate data sets such as stock prices, freezing
points of chemical compounds or physical constants exhibit this pattern at least empirically. While originally being only a curious pattern [85], practical implications began to emerge in the 1960s in the design of efficient computers [66]. In recent years goodness-of-fit test against Benford’s law has been used to detect possible fraudulent financial data, by analyzing the deviations of accounting data, corporation incomes, tax returns or scientific experimental data to theoretical Benford predictions [88]. Indeed, digit pattern analysis can produce valuable findings not revealed by a mere glance, as is the case of recent election results [68, 88].

Many mathematical sequences such as \((n^n)_{n \in \mathbb{N}}\) and \((n!)_{n \in \mathbb{N}}\) [63], binomial arrays \(\binom{n}{k}\) [84], geometric sequences or sequences generated by recurrence relations [85, 69] to cite a few are proved to be Benford. One may thus wonder if this is the case for the primes. In figure 2.16 we have plotted the leading digit \(d\) rate of appearance for the prime numbers placed in the interval \([1, N]\) (red bars), for different sizes \(N\). Note that intervals \([1, N]\) have been chosen such that \(N = 10^D, D \in \mathbb{N}\) in order to assure an unbiased sample where all possible first digits are equiprobable a priori (see the appendix section for further details). Benford’s law states that the first digit of a series data extracted at random is 1 with a frequency of 0.301\%, and is 9 only about 0.6\%. Note in figure 2.16 that primes seem however to approximate uniformity in its first digit. Indeed, the more we increase the interval under study, the more we approach uniformity. As a matter of fact, Diaconis [84] proved that primes are not Benford distributed as long as their first significant digit is asymptotically uniformly distributed.

Several mathematical insights of the Benford’s law have been also advanced so far [71, 72, 85, 69], and Hill [92] proved in 1995 a Central Limit-like Theorem which states that random entries picked from random distributions form a sequence whose first digit distribution converges to the Benford’s law, explaining thereby its ubiquity. This law has been for a long time practically the only distribution that could explain the presence of skewed first digit
frequencies in generic data sets. Recently Pietronero et al. [93] proposed a generalization of Benford’s law based in multiplicative processes. It is well known that stochastic process with probability density $1/x$ generates data which are Benford, therefore series generated by power law distributions $P(x) \sim x^{-\alpha}$ with $\alpha \neq 1$, would have a first digit distribution that follow a so-called Generalized Benford’s law (GBL):

$$P(d) = C \int_d^{d+1} x^{-\alpha} dx = \frac{1}{10^{1-\alpha} - 1} \left[ (d + 1)^{1-\alpha} - d^{1-\alpha} \right],$$  

(2.3.2)

where the prefactor is fixed for normalization to hold and $\alpha$ is the exponent of the original power law distribution (for $\alpha = 1$, the GBL reduces to the Benford’s law).

Although Diaconis showed that the leading digit of primes distributes uniformly in the infinite limit, there exist a clear bias from uniformity for finite sets (see figure 2.16). In this figure we have also plotted (grey bars) the fitting to a GBL. Note that in each of the four intervals, there is a particular value of exponent $\alpha$ for which an excellent agreement holds (see the appendix for fitting methods and statistical tests). Interestingly, the fitting parameter $\alpha$ decreases as the interval, hence the number of primes, increases in a very particular way. In the left part of figure 2.17 we have plotted this size dependence, showing that a functional relation between $\alpha$ and $N$ takes place:

$$\alpha(N) = \frac{1}{\log N - a},$$  

(2.3.3)

where $a = 1.10 \pm 0.05$ for large values of $N$. Notice that $\lim_{N \to \infty} \alpha(N) = 0$, and in this situation this size-dependent GBL reduces to the uniform distribution, in consistency with previous theory [84]. Despite the local randomness of the prime numbers sequence, its first digit distribution seems to converge smoothly to uniformity in a very precise trend: as a GBL with a size dependent exponent $\alpha(N)$.

At this point an extension of the GBL to include, not only the first significative digit, but the first $k$ significative ones can be done [92]. Given a number $n$, we can consider its $k$
Figure 2.16: Leading digit histogram of the prime number sequence. Each plot represents, for the set of prime numbers comprised in the interval \([1, N]\), the relative frequency of the leading digit \(d\) (red bars). Sample sizes are: 5761455 primes for \(N = 10^8\), 50847534 primes for \(N = 10^9\), 455052511 primes for \(N = 10^{10}\) and 4118054813 primes for \(N = 10^{11}\). Grey bars represent the fitting to a Generalized Benford distribution (eq. 2.3.2) with a given exponent \(\alpha(N)\).

First significative digits \(d_1, d_2, \ldots, d_k\) through its decimal representation: \(D = \sum_{i=1}^{k} d_i 10^{k-i}\), where \(d_1 \in \{1, \ldots, 9\}\) and \(d_i \in \{0, 1, \ldots, 9\}\) for \(i \geq 2\). Hence, the extended GBL providing the probability of starting with number \(D\) is

\[
P(d_1, d_2, \ldots, d_k) = P(D) = \frac{1}{(10^k)^{1-\alpha} - 10^{k-1}} \left[ (D + 1)^{1-\alpha} - D^{1-\alpha} \right]. \tag{2.3.4}
\]

Figure 2.18 represents the fitting of the 4118054813 primes appearing in the interval \([1, 10^{11}]\) to an extended GBL for \(k = 2, 3, 4\) and 5: interestingly, the pattern still holds.
Figure 2.17: Size dependent parameter $\alpha(N)$. Left: Red dots represent the exponent $\alpha(N)$ for which the first significant digit of prime number sequence fits a Generalized Benford Law in the interval $[1, N]$. The black line corresponds to the fitting, using a least squares method, $\alpha(N) = 1/(\log N - 1.10)$. Right: same analysis as for the left figure, but for the Riemann nontrivial zeta zeros sequence. The best fitting is $\alpha(N) = 1/(\log N - 2.92)$.

Once the pattern has been put forward in the case of the prime number sequence, we may wonder if a similar behavior holds for the sequence of nontrivial Riemann zeta zeros (zeros sequence from now on). This sequence is composed by the imaginary part of the nontrivial zeros (actually only those with positive imaginary part are taken into account by symmetry reasons) of $\zeta(s)$. While this sequence is not Benford distributed in the light of a theorem by Rademacher-Hlawka [75] that proves that it is asymptotically uniform, will it follow a size-dependent GBL as in the case of the primes?

In figure 2.19 we have plotted, in the interval $[1, N]$ and for different values of $N$, the relative frequencies of leading digit $d$ in the zeros sequence (blue bars), and in grey bars a fitting to a GBL with density $x^\alpha$, i.e.:

$$P(d) = C \int_d^{d+1} x^\alpha dx = \frac{1}{10^{1+\alpha} - 1} \left[ (d + 1)^{1+\alpha} - d^{1+\alpha} \right]$$  \hspace{0.5cm} (2.3.5)

(this reciprocity is clarified later in the text). Note that a very good agreement holds again for particular size-dependent values of $\alpha$, and the same functional relation as equation 2.3.3.
Figure 2.18: Extension of GBL to the $k$ first significant digits. In this figure we represent
the fitting of an extended GBL following eq. 2.3.4 (black line) to the first two significant
digits relative frequencies (up-left), first three significant digits relative frequencies (up-
right), first four significant digits relative frequencies (down-left) and first five significant
digits relative frequencies (down-right) of the 4118054813 primes appearing in the interval
$[1, 10^{11}]$ (red dots).

holds with $a = 2.92 \pm 0.05$. As in the case of the primes, this size dependent GBL tends to
uniformity for $N \to \infty$, as it should [75]. Moreover, the extended version of equation 2.3.5
for the $k$ first significative digits is

$$P(d_1, d_2, \ldots, d_k) = P(D)$$
$$= \frac{1}{(10^k)^{1+\alpha} - 10^k-1} \left[ (D + 1)^{1+\alpha} - D^{1+\alpha} \right], \quad (2.3.6)$$

and also holds in this case (see figure 2.20).
Figure 2.19: Leading digit histogram of the nontrivial Riemann zeta zeros sequence. Each plot represents, for the sequence of Riemann zeta zeros comprised in the interval \([1, N]\), the observed relative frequency of leading digit \(d\) (blue bars). Sample sizes are: 10142 zeros for \(N = 10^4\), 138069 zeros for \(N = 10^5\), 1747146 zeros for \(N = 10^6\) and 21136126 zeros for \(N = 10^7\). Grey bars represent the fitting to a GBL following equation 2.3.5 with a given exponent \(\alpha(N)\).

2.3.3 Explanation of the patterns

Why do these two sequences exhibit this unexpected pattern in the leading digit distribution? What is the responsible for it to take place? While the prime number distribution is deterministic in the sense that precise rules determine whether an integer is prime or not, its apparent local randomness has suggested several stochastic interpretations. In particular, Cramér [89, 90] defined the following model: assume that we have a sequence of urns \(U(n)\) where \(n = 1, 2, ...\) and put black and white balls in each urn such that the probability
Figure 2.20: Extension of GBL to the $k$ first significant digits. In this figure we represent the fitting of an extended GBL following eq. 2.3.6 (black line) to the first two significant digits relative frequencies (up), first three significant digits relative frequencies (down-left), and first four significant digits relative frequencies (down-right) of the 21136126 zeros appearing in the interval $[1, 10^7]$ (blue dots).

of drawing a white ball in the $k^{th}$-urn goes like $1/\log k$. Then, in order to generate a sequence of pseudo-random prime numbers we only need to draw a ball from each urn: if the drawing from the $k^{th}$-urn is white, then $k$ will be labeled as a pseudo-random prime. The prime number sequence can indeed be understood as a concrete realization of this stochastic process, where the chance of a given integer $x$ to be prime is $1/\log x$. We have repeated all statistical tests within the stochastic Cramér model, and have found that a statistical sample of pseudo-random prime numbers in $[1, 10^{11}]$ is also GBL distributed and reproduce all statistical analysis previously found in the actual primes (see the appendix for an in-depth analysis). This result strongly suggests that a density $1/\log x$, which is nothing but the mean local primes density by virtue of the prime number theorem, is likely to be the responsible for the GBL pattern.

Recently, it has been shown that disparate distributions such as the Lognormal, the Weibull or the Exponential distribution can generate standard Benford behavior [91] for particular
values of their parameters. In this sense, a similar phenomenon could be taking place with
GBL: can different distributions generate GBL behavior? One should switch the emphasis
from the examination of data sets that obey GBL to probability distributions that do so,
others than power laws. We have confirmed (appendix) that there exists statistical confor-
mance between the prime number cumulative distribution \( \pi(n) \) (conveniently normalized
in \([1, N]\)) and a GBL with exponent \( \alpha(N) \). The same holds for the Eulerean logarithmic
integral \( \text{Li}(n) \), which constitutes an asymptotic approximation to \( \pi(n) \):

\[
\pi(N) \sim \text{Li}(N) = \int_2^N \frac{1}{\log x} \, dx
\]  

(one of the formulations of the Riemann hypothesis actually states that \( |\text{Li}(n) - \pi(n)| < c \sqrt{n \log n} \), for some constant \( c \) [61]).

We can indeed prove (see the appendix) that any cumulative distribution function \( F_T(z) \)
that evidences GBL behavior must fulfill

\[
\sum_{d=0}^{n} \{ F_T(v 10^d) - F_T(10^d) \} = z,
\]

where \( v \equiv [(10^{1-\alpha} - 1)z + 1]^{\frac{1}{1-\alpha}} \) and \( z \in [0, 1] \). This latter relation is trivially fulfilled
by power law densities \( x^{-\alpha} \). In particular, eq. 2.3.8 reduces for the prime cumulative
distribution \( \pi(n) \) to

\[
\sum_{d=0}^{D} \left\{ \pi(v \cdot 10^d) - \pi(10^d) \right\} = \pi(10^{D+1})z,
\]

and

\[
\sum_{d=0}^{D} \left\{ \text{Li}(v \cdot 10^d) - \text{Li}(10^d) \right\} = \text{Li}(10^{D+1})z,
\]

for \( \text{Li}(n) \). Equations 2.3.40 and 2.3.41 are numerically fulfilled (see the appendix).

Hitherto, we have provided statistical arguments that indicate that other distributions than
\( x^{-\alpha} \) such as \( 1/\log x \) can generate GBL behavior. In what follows we provide analytical
arguments that support this fact.
Table 2.1: Up to integer $N$, values of the prime counting function $\pi(N)$, the approximation given by the logarithmic integral $\text{Li}(N)$, $N/\log N$, the counting function $L(N)$ defined in equation 2.3.9 and the ratio $L(N)/\pi(N)$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\pi(N)$</th>
<th>$\text{Li}(N)$</th>
<th>$N/\log N$</th>
<th>$L(N)$</th>
<th>$L(N)/\pi(N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^2$</td>
<td>25</td>
<td>30</td>
<td>22</td>
<td>29</td>
<td>0.85533</td>
</tr>
<tr>
<td>$10^3$</td>
<td>168</td>
<td>178</td>
<td>145</td>
<td>172</td>
<td>0.97595</td>
</tr>
<tr>
<td>$10^4$</td>
<td>1229</td>
<td>1246</td>
<td>1086</td>
<td>1228</td>
<td>1.00081</td>
</tr>
<tr>
<td>$10^5$</td>
<td>9592</td>
<td>9630</td>
<td>8686</td>
<td>9558</td>
<td>1.00352</td>
</tr>
<tr>
<td>$10^6$</td>
<td>78492</td>
<td>78628</td>
<td>72382</td>
<td>78280</td>
<td>1.00278</td>
</tr>
<tr>
<td>$10^7$</td>
<td>664579</td>
<td>664918</td>
<td>620421</td>
<td>662958</td>
<td>1.00244</td>
</tr>
<tr>
<td>$10^8$</td>
<td>5761455</td>
<td>5762209</td>
<td>5428681</td>
<td>5749998</td>
<td>1.00199</td>
</tr>
<tr>
<td>$10^9$</td>
<td>50847534</td>
<td>50849235</td>
<td>48254942</td>
<td>50767815</td>
<td>1.00157</td>
</tr>
<tr>
<td>$10^{10}$</td>
<td>455052511</td>
<td>455055615</td>
<td>434294482</td>
<td>454484882</td>
<td>1.00125</td>
</tr>
<tr>
<td>$10^{20}$</td>
<td>2220819602560918840</td>
<td></td>
<td></td>
<td></td>
<td>1.00027</td>
</tr>
</tbody>
</table>

Li$(N)$ possesses the following asymptotic expansion

$$\text{Li}(N) = \frac{N}{\log N} \left(1 + \frac{1}{\log N} + \frac{2}{\log^2 N} + O\left(\frac{1}{\log^3 N}\right)\right). \quad (2.3.11)$$

Now, a sequence whose first significant digit follows a GBL has indeed a density that goes as $x^{-\alpha}$. One can consequently derive from this latter density a function $L(N)$ that provides the number of primes appearing in the interval $[1, N]$ as it follows:

$$L(N) = c\alpha(N) \int_2^N x^{-\alpha(N)} \, dx \quad (2.3.12)$$

where the prefactor is fixed for $L(N)$ to fulfill the prime number theorem and consequently

$$\lim_{N \to \infty} \frac{L(N)}{N/\log N} = 1 \quad (2.3.13)$$

(see table 2.1 for a numerical exploration of this new approximation to $\pi(N)$). Now, we
can asymptotically expand $L(N)$ as it follows

$$L(N) = \frac{\alpha(N)e}{1 - \alpha(N)} N^{1 - \alpha(N)}$$

$$= \frac{N}{\log N - (a + 1)} \cdot \exp \left( -\frac{a}{\log N - a} \right)$$

$$\leq \frac{N}{\log N} \left\{ 1 + \frac{a + 1}{\log N} + \frac{(a + 1)^2}{\log^2 N} + O \left( \frac{1}{\log^3 N} \right) \right\} \cdot$$

$$\leq \frac{N}{\log N} \left\{ 1 - \frac{a}{\log N - a} + \frac{a^2}{(\log N - a)^2} + O \left( \frac{1}{(\log N - a)^3} \right) \right\}$$

$$= \frac{N}{\log N} \left\{ 1 + \frac{1}{\log N} + \frac{1 + a - a^2/2}{\log^2 N} + O \left( \frac{1}{\log^3 N} \right) \right\}.$$  (2.3.14)

Comparing equations 2.3.11 and 2.3.14, we conclude that $\text{Li}(N)$ and $L(N)$ are compatible cumulative distributions within an error

$$E(N) = \frac{N}{\log N} \left\{ \frac{2}{\log^2 N} - \frac{1 + a - a^2/2}{\log^2 N} + O \left( \frac{1}{\log^3 N} \right) \right\}.$$  (2.3.15)

that is indeed minimum for $a = 1$, in consistency with our previous results. Hence, within that error we can conclude that primes obey a GBL with $\alpha(N)$ following equation 2.3.3.

What about the Riemann zeros? Von Mangoldt proved [61] that on average, the number of nontrivial zeros $R(N)$ up to $N$ (zeros counting function) is

$$R(N) = \frac{N}{2\pi} \log \left( \frac{N}{2\pi} \right) - \frac{N}{2\pi} + O(\log N).$$  (2.3.16)

$R(N)$ is nothing but the cumulative distribution of the zeros (up to normalization), which fulfills

$$R(N) \approx \frac{1}{2\pi} \int_2^N \log \left( \frac{x}{2\pi} \right) dx.$$  (2.3.17)

The nontrivial Riemann zeros average density is thus $\log(x/2\pi)$, which is nothing but the reciprocal of the prime numbers average density (see eq. 2.3.7). Comparing equations 2.3.17 and 2.3.7 one can straightforwardly deduce a power law approximation to the cumulative distribution of the non trivial zeros similar to equation 2.3.12:

$$R(N) \sim \frac{1}{2\pi e\alpha(N/2\pi)} \int_2^N \left( \frac{x}{2\pi} \right)^{\alpha(N/2\pi)} dx.$$  (2.3.18)
We conclude that zeros are also GBL for $\alpha(N)$ satisfying the following change of scale

$$\alpha\left(\frac{N}{2\pi}\right) = \frac{1}{\log\left(\frac{N}{2\pi}\right) - a} = \frac{1}{\log N - (\log(2\pi) + a)}. \quad (2.3.19)$$

Hence, since $a = 1$ (equation 2.3.15) one should expect for the constant $a$ associated to the zeros sequence the following value: $\log(2\pi) + 1 \approx 2.83$, in good agreement with our previous numerical analysis.

### 2.3.4 Final remarks and applications

To conclude, we have unveiled a statistical pattern in the prime numbers and the nontrivial Riemann zeta zeros sequences that has surprisingly gone unnoticed until now. Several applications and future work can be depicted: first, since the Riemann zeros seem to have the same statistical properties as the eigenvalues of a concrete type of random matrices called the Gaussian Unitary Ensemble [78], the relation between GBL and random matrix theory should be investigated. Second, this finding may also apply to several other sequences that, while not being strictly Benford distributed, can be GBL, and in this sense much work recently developed for Benford distributions [79] could be readily generalized. Finally, it has not escaped our notice that several applications recently advanced in the context of Benford’s law, such as fraud detection or stock market analysis [88], could eventually be generalized to the wider context of GBL formalism. This generalization also extends to stochastic sieve theory [80], dynamical systems that follow Benford’s law [81, 69] and their relation to stochastic multiplicative processes [82].
2.3.5 Appendix: Some mathematical digressions and statistical tests

How to pick an integer \textit{at random}?

**Visualizing the Generalized Benford law pattern in prime numbers as a biased random walk** In order the pattern already captured in figure 2.16 to become more evident,

we have built the following 2D random walk

\begin{align*}
  x(t+1) &= x(t) + \xi_x \\
  y(t+1) &= y(t) + \xi_y,
\end{align*}

(2.3.20)
where $x$ and $y$ are cartesian variables with $x(0) = y(0) = 0$, and both $\xi_x$ and $\xi_y$ are discrete random variables that take values $\in \{0, -1, 1\}$ depending on the first digit $d$ of the numbers randomly chosen at each time step, according to the rules depicted in figure 2.21. Thereby, in each iteration we peak at random a positive integer (grey random walk) or a prime (red random walk) from the interval $[1, 10^6]$, and depending on its first significative digit $d$, the random walker moves accordingly (for instance if we peak prime 13, we have $d = 1$ and the random walker rules provide $\xi_x = 1$ and $\xi_y = 1$: the random walker moves up-right).

We have plotted the results of this 2D Random Walk in figure 2.21 for random picking of integers (grey random walk) and for random picking of primes (red random walk). Note that while the grey random walk seems to be a typical uncorrelated Brownian motion (enhancing the fact that the first digit distribution of the integers is uniformly distributed), the red random walk is clearly biased: this is indeed a visual characterization of the pattern. Observe that if the interval in which we randomly peak either the integers or the primes wasn’t of the shape $[1, 10^D]$, there would be a systematic bias present in the pool and consequently both integer and prime random walks would be biased: it comes thus necessary to define the intervals under study in that way.

**Natural density** If primes were for instance Benford distributed, one should expect that if we pick a prime at random, this one should start by number 1 around 30% of the time. But what does the sentence ‘Pick a prime at random’ stand for? Notice that in the previous experiment (the 2D biased Random Walk) we have drawn whether integers or primes at random from the pool $[1, 10^6]$. All over this work, the intervals $[1, N]$ have been chosen such that $N = 10^D$, $D \in \mathbb{N}$. This choice isn’t arbitrary, much on the contrary, it relies on the fact that whenever studying infinite integer sequences, the results strongly depend on the interval under study. For instance, everyone will agree that intuitively the set of positive integers $\mathbb{N}$ is an infinite sequence whose first digit is uniformly distributed: there exist as
many naturals starting by one as naturals starting by nine. However there exist subtle difficulties at this point that come from the fact that the first digit natural density is not well defined. Since there exist infinite integers in \( \mathbb{N} \) and consequently it is not straightforward to quantify the quote ‘pick an integer at random’ in a way in which satisfies the laws of probability, in order to check if integers have a uniform distributed first significant digit, we have to consider finite intervals \([1, N]\). Hereafter, notice that uniformity \textit{a priori} is only respected when \( N = 10^D \). For instance, if we choose the interval to be \([1, 2000]\), a random drawing from this interval will be a number starting by 1 with high probability, as there are obviously more numbers starting by one in that interval. If we increase the interval to say \([1, 3000]\), then the probability of drawing a number starting by 1 or 2 will be larger than any other. We can easily come to the conclusion that the first digit density will oscillate repeatedly by decades as \( N \) increases without reaching convergence, and it is thereby said that the set of positive integers with leading digit \( d \) (\( d = 1, 2, ..., 9 \)) does not possess a natural density among the integers. Note that the same phenomenon is likely to take place for the primes (see Chris Caldwell’s The Prime Pages [83] for an introductory discussion in natural density and Benford’s law for prime numbers and references therein).

In order to overcome this subtle point, one can: (i) choose intervals of the shape \([1, 10^D]\), where every leading digit has equal probability a priori of being picked. According to this situation, positive integers \( \mathbb{N} \) have a uniform first digit distribution, and in this sense Diaconis [84] showed that primes do not obey Benford’s law as their first digit distribution is asymptotically uniform. Or (ii) use average and summability methods such as the Cesaro or the logarithm matrix method \( \ell \) [85] in order to define a proper first digit density that holds in the infinite limit. Some authors have shown that in this case, both the primes and the integers are \textit{weak} Benford sequences [85, 86, 87].

As we are dealing with finite subsets and in order to check if a pattern \textit{really} takes place
for the primes, in this work we have chosen intervals of the shape \([1, 10^D]\) to assure that samples are unbiased and that all first digits are equiprobable \textit{a priori}.

Statistical methods

\textbf{Method of moments} In order to estimate the best fitting between a GBL with parameter \(\alpha\) and a data set, we have employed the method-of-moments. If GBL fits the empirical data, then both distributions have the same first moments, and the following relation holds:

\[ \sum_{d=1}^{9} dP(d) = \sum_{d=1}^{9} dP_e(d) \]  

(2.3.21)

where \(P(d)\) and \(P_e(d)\) are the observed normalized frequencies and GB expected probabilities for digit \(d\), respectively. Using a Newton-Raphson method and iterating equation 2.3.21 until convergence, we have calculated \(\alpha\) for each sample \([1, N]\).

\textbf{Statistical tests} Typically, chi-square goodness-of-fit test has been used in association with Benford’s Law [88]. Our null hypothesis here is that the sequence of primes follow a GBL. The test statistic is:

\[ \chi^2 = M\sum_{d=1}^{9} \frac{(P(d) - P_e(d))^2}{P_e(d)} \]

(2.3.22)

where \(M\) denotes the number of primes in \([1, N]\). Since we are computing parameter \(\alpha(N)\) using the mean of the distribution, the test statistic follows a \(\chi^2\) distribution with \(9 - 2 = 7\) degrees of freedom, so the null hypothesis is rejected if \(\chi^2 > \chi_{a,7}^2\), where \(a\) is the level of significance. The critical values for the 10%, 5%, and 1% are 12.02, 14.07, and 18.47 respectively. As we can see in table 2.2, despite the excellent visual agreement (figure 1 in the main text), the \(\chi^2\) statistic goes up with sample size and consequently the null hypothesis can’t be rejected only for relatively small sample sizes \(N < 10^9\). As a matter of fact, chi-square statistic suffers from the excess power problem on the basis that it is
size sensitive: for huge data sets, even quite small differences are statistically significant [88]. A second alternative is to use the standard $Z$-statistics to test significant differences. However, this test is also size dependent, and hence registers the same problems as $\chi^2$ for large samples. Due to this facts, Nigrini [88] recommends for Benford analysis a distance measure test called Mean Absolute Deviation (MAD). This test computes the average of the nine absolute differences between the empirical proportions of a digit and the ones expected by the GBL. That is:

$$\text{MAD} = \frac{1}{9} \sum_{d=1}^{9} |P(d) - P^e(d)|$$  (2.3.23)

This test overcomes the excess power problem of $\chi^2$ as long as it is not influenced by the size of the data set. While MAD lacks of cut-off level, Nigrini [88] suggests that the guidelines for measuring conformity of the first digits to Benford Law to be: MAD between 0 and $0.4 \cdot 10^{-2}$ imply close conformity, from $0.4 \cdot 10^{-2}$ to $0.8 \cdot 10^{-2}$ acceptable conformity, from $0.8 \cdot 10^{-2}$ to $0.12 \cdot 10^{-1}$ marginally acceptable conformity, and finally, greater than $0.12 \cdot 10^{-1}$, nonconformity. Under these cut-off levels we can not reject the hypothesis that the first digit frequency of the prime numbers sequence follows a GBL. In addition, the Maximum Absolute Deviation $m$ defined as the largest term of MAD is also showed in each case.

As a final approach to testing for a similarity between the two histograms, we can check the correlation between the empirical and theoretical proportions by the simple regression correlation coefficient $r$ in a scatterplot. As we can see in table 2.2 the empirical data are highly correlated with a Generalized Benford distribution.

The same statistical tests have been performed for the case of the Riemann non trivial zeta zeros sequence (table 2.3), with similar results.
Table 2.2: Table gathering the values of the following statistics: $\chi^2$, Maximum Absolute Deviation ($m$), Mean Absolute Deviation (MAD) and correlation coefficient ($r$) between the observed first significant digit frequency of the set of $M$ primes in $[1, N]$ and the expected Generalized Benford distribution (equation 2.3.2 with an exponent $\alpha(N)$ given by equation 2.3.3 with $a = 1.1$). While $\chi^2$-test rejects the hypothesis for very large samples due to its size sensitivity, every other test cannot reject it, enhancing the goodness-of-fit between the data and the GB distribution.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$M = #$ primes</th>
<th>$\chi^2$</th>
<th>$m$</th>
<th>MAD</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^4$</td>
<td>1229</td>
<td>0.45</td>
<td>$0.32 \cdot 10^{-2}$</td>
<td>$0.19 \cdot 10^{-2}$</td>
<td>0.96965</td>
</tr>
<tr>
<td>$10^5$</td>
<td>9592</td>
<td>0.62</td>
<td>$0.21 \cdot 10^{-2}$</td>
<td>$0.65 \cdot 10^{-3}$</td>
<td>0.99053</td>
</tr>
<tr>
<td>$10^6$</td>
<td>78498</td>
<td>0.61</td>
<td>$0.50 \cdot 10^{-3}$</td>
<td>$0.26 \cdot 10^{-3}$</td>
<td>0.99826</td>
</tr>
<tr>
<td>$10^7$</td>
<td>664579</td>
<td>0.77</td>
<td>$0.17 \cdot 10^{-3}$</td>
<td>$0.11 \cdot 10^{-3}$</td>
<td>0.99964</td>
</tr>
<tr>
<td>$10^8$</td>
<td>5761455</td>
<td>2.2</td>
<td>$0.15 \cdot 10^{-3}$</td>
<td>$0.56 \cdot 10^{-4}$</td>
<td>0.99984</td>
</tr>
<tr>
<td>$10^9$</td>
<td>50847534</td>
<td>11.0</td>
<td>$0.11 \cdot 10^{-3}$</td>
<td>$0.42 \cdot 10^{-4}$</td>
<td>0.99988</td>
</tr>
<tr>
<td>$10^{10}$</td>
<td>455052511</td>
<td>61.2</td>
<td>$0.90 \cdot 10^{-4}$</td>
<td>$0.33 \cdot 10^{-4}$</td>
<td>0.99991</td>
</tr>
<tr>
<td>$10^{11}$</td>
<td>4118054813</td>
<td>358.5</td>
<td>$0.74 \cdot 10^{-4}$</td>
<td>$0.27 \cdot 10^{-4}$</td>
<td>0.99993</td>
</tr>
</tbody>
</table>

Table 2.3: Table gathering the values of the following statistics: $\chi^2$, Maximum Absolute Deviation ($m$), Mean Absolute Deviation (MAD) and correlation coefficient ($r$) between the observed first significant digit frequency in the $M$ zeros in $[0, N]$ and the expected Generalized Benford distribution (equation 2.3.5 with an exponent $\alpha(N)$ given by equation 2.3.3 with $a = 2.92$). While $\chi^2$-test rejects the hypothesis for very large samples due to its size sensitivity, every other test cannot reject it, enhancing the goodness-of-fit between the data and the GB distribution.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$M = #$ zeros</th>
<th>$\chi^2$</th>
<th>$m$</th>
<th>MAD</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^4$</td>
<td>649</td>
<td>0.14</td>
<td>$0.32 \cdot 10^{-2}$</td>
<td>$0.13 \cdot 10^{-2}$</td>
<td>0.99701</td>
</tr>
<tr>
<td>$10^4$</td>
<td>10142</td>
<td>0.23</td>
<td>$0.11 \cdot 10^{-2}$</td>
<td>$0.41 \cdot 10^{-3}$</td>
<td>0.99943</td>
</tr>
<tr>
<td>$10^5$</td>
<td>138069</td>
<td>0.75</td>
<td>$0.54 \cdot 10^{-3}$</td>
<td>$0.20 \cdot 10^{-3}$</td>
<td>0.99974</td>
</tr>
<tr>
<td>$10^6$</td>
<td>1747146</td>
<td>3.6</td>
<td>$0.34 \cdot 10^{-3}$</td>
<td>$0.13 \cdot 10^{-3}$</td>
<td>0.99983</td>
</tr>
<tr>
<td>$10^7$</td>
<td>21136126</td>
<td>20.3</td>
<td>$0.23 \cdot 10^{-3}$</td>
<td>$0.86 \cdot 10^{-4}$</td>
<td>0.99988</td>
</tr>
</tbody>
</table>
**Cramér’s model**

The prime number distribution is deterministic in the sense that primes are determined by precise arithmetic rules. However, its apparent local randomness has suggested several stochastic interpretations. Concretely, Cramér [89, 90] defined the following model: assume that we have a sequence of urns $U(n)$ where $n = 1, 2, ...$ and put black and white balls in each urn such that the probability of drawing a white ball in the $k^{th}$-urn goes like $1/\log k$. Then, in order to generate a sequence of pseudo-random prime numbers we only need to draw a ball from each urn: if the drawing from the $k^{th}$-urn is white, then $k$ will be labeled as a pseudo-random prime. The prime number sequence can indeed be understood as a concrete realization of this stochastic process. With such model, Cramér studied amongst others the distribution of gaps between primes and the distribution of twin primes as far as statistically speaking, these distributions should be similar to the pseudo-random ones generated by his model. Quoting Cramér: ‘With respect to the ordinary prime numbers, it is well known that, roughly speaking, we may say that the chance that a given integer $n$ should be a prime is approximately $1/\log n$. This suggests that by considering the following series of independent trials we should obtain sequences of integers presenting a certain analogy with the sequence of ordinary prime numbers $p_n$’.

In this work we have simulated a Cramér process, in order to obtain a sample of pseudo-random primes in $[1, 10^{11}]$. Then, the same statistics performed for the prime number sequence have been realized in this sample. Results are summarized in table 2.4. We can observe that the Cramér’s model reproduces the same behavior, namely: (i) The first digit distribution of the pseudo-random prime sequence follows a GBL with a size-dependent exponent that follows equation 3 (main text). (ii) The number of pseudo-primes found in each decade matches statistically speaking to the actual number of primes. (iii) The $\chi^2$-test evidences the same problems of power for large data sets. Having in mind that the sample elements in this model are independent (what is not the case in the actual prime sequence),
Table 2.4: Table gathering the values of the following statistics: $\chi^2$, Maximum Absolute Deviation ($m$), Mean Absolute Deviation (MAD) and correlation coefficient ($r$) between the observed first significant digit frequency in the Cramér model for $M$ pseudo-random primes in $[1,N]$ and the expected Generalized Benford distribution (equation 2.3.2) with an exponent $\alpha(N)$ given by equation 2.3.3 with $a = 1.1$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$M = #$ pseudo-random primes</th>
<th>$\chi^2$</th>
<th>$m$</th>
<th>MAD</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^4$</td>
<td>1189</td>
<td>1.20</td>
<td>0.17·$10^{-1}$</td>
<td>0.92·$10^{-2}$</td>
<td>0.639577</td>
</tr>
<tr>
<td>$10^5$</td>
<td>9673</td>
<td>0.43</td>
<td>0.33·$10^{-2}$</td>
<td>0.21·$10^{-2}$</td>
<td>0.969031</td>
</tr>
<tr>
<td>$10^6$</td>
<td>78693</td>
<td>0.39</td>
<td>0.59·$10^{-3}$</td>
<td>0.14·$10^{-2}$</td>
<td>0.990322</td>
</tr>
<tr>
<td>$10^7$</td>
<td>664894</td>
<td>0.09</td>
<td>0.23·$10^{-3}$</td>
<td>0.99·$10^{-4}$</td>
<td>0.999626</td>
</tr>
<tr>
<td>$10^8$</td>
<td>5762288</td>
<td>0.24</td>
<td>0.15·$10^{-3}$</td>
<td>0.53·$10^{-4}$</td>
<td>0.999855</td>
</tr>
<tr>
<td>$10^9$</td>
<td>50850064</td>
<td>1.23</td>
<td>0.11·$10^{-3}$</td>
<td>0.42·$10^{-4}$</td>
<td>0.999892</td>
</tr>
<tr>
<td>$10^{10}$</td>
<td>455062569</td>
<td>6.84</td>
<td>0.90·$10^{-4}$</td>
<td>0.33·$10^{-4}$</td>
<td>0.999914</td>
</tr>
<tr>
<td>$10^{11}$</td>
<td>4118136330</td>
<td>41.0</td>
<td>0.73·$10^{-4}$</td>
<td>0.27·$10^{-4}$</td>
<td>0.999937</td>
</tr>
</tbody>
</table>

we can conclude that the rejection of the null hypothesis by the $\chi^2$-test for huge data sets is not related to a lack of data independence but much likely from the test’s size sensitivity.

(iv) The rest of statistical analysis is similar to the one previously performed in the prime number sequence.

**Statistical conformance of prime number distribution to GBL**

The analysis of our numerical experiments with the Cramér’s model is suggesting that the probability density $1/\log n$ is likely the responsible for the GBL pattern in the prime number sequence. Note that this may be not that surprising, since in the case of standard Benford law, Leemis et al. [91] have recently shown that different standard distributions such as the Log normal, the Weibull, the Gamma or the Exponential distribution, for particular values of its parameters, generate data sets Benford distributed. Accordingly, one can switch the emphasis from the examination of data sets that obey GBL to probability distributions that do so others than power laws.
\( \chi^2 \)-test for conformance between distributions. The prime counting function \( \pi(N) \) provides the number of primes in the interval \([1, N]\) and up to normalization, stands as the cumulative distribution function of primes. While \( \pi(N) \) is a stepped function, a nice asymptotic approximation is the offset logarithmic integral:

\[
\pi(N) \sim \int_{2}^{N} \frac{1}{\log x} \, dx = \text{Li}(N),
\]

(2.3.24)

where we can interpret \( 1/\log x \) as an average prime density and the lower bound of the integral is set to be 2 for singularity reasons. Following Leemis et al. \([91]\), we can calculate a chi-square goodness-of-fit of the conformance between the first digit distribution generated by \( \text{Li}(N) \) and a GBL with exponent \( \alpha(N) \). The test statistic is in this case:

\[
c = \sum_{d=1}^{9} \frac{[\Pr(Y = d) - \Pr(X = d)]^2}{\Pr(X = d)},
\]

(2.3.25)

where \( \Pr(X) \) is the first digit probability (equation 2 in the main text) for a GBL associated to a probability distribution with exponent \( \alpha(N) \) and \( \Pr(Y) \) is the tested probability. In table 2.5 we have computed, fixed the interval \([1, N]\), the chi-square statistic \( c \) for two different scenarios, namely the normalized logarithmic integral \( \text{Li}(n)/\text{Li}(N) \) and the normalized prime counting function \( \pi(n)/\pi(N) \), with \( n \in [1, N] \). Note that this test does not evidence the excess power problem as far as it is size independent. In both cases there is a remarkable good agreement and we cannot reject the hypothesis that primes are size-dependent GBL.

Conditions for conformance to GBL. In \([92]\) Hill wondered about which common distributions (or mixtures thereof) satisfy Benford’s law. In \([91]\) Leemis et al. tackled this problem and quantified the agreement to Benford’s law of several standard distributions. They concluded that the ubiquity of Benford behavior could be related to the fact that many distributions follow Benford’s law for particular values of their parameters. Here, following the philosophy of that paper \([91]\), we will develop a mathematical framework that provide conditions for conformance to a GBL.
<table>
<thead>
<tr>
<th>(N)</th>
<th>(c) for (\pi(x)/\pi(N))</th>
<th>(c) for (\text{Li}(x)/\text{Li}(N))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10^4)</td>
<td>(0.59 \cdot 10^{-2})</td>
<td>(0.58 \cdot 10^{-2})</td>
</tr>
<tr>
<td>(10^4)</td>
<td>(0.86 \cdot 10^{-3})</td>
<td>(0.57 \cdot 10^{-3})</td>
</tr>
<tr>
<td>(10^5)</td>
<td>(0.12 \cdot 10^{-3})</td>
<td>(0.13 \cdot 10^{-3})</td>
</tr>
<tr>
<td>(10^6)</td>
<td>(0.57 \cdot 10^{-4})</td>
<td>(0.61 \cdot 10^{-4})</td>
</tr>
<tr>
<td>(10^7)</td>
<td>(0.32 \cdot 10^{-4})</td>
<td>(0.33 \cdot 10^{-4})</td>
</tr>
<tr>
<td>(10^8)</td>
<td>(0.17 \cdot 10^{-4})</td>
<td>(0.17 \cdot 10^{-4})</td>
</tr>
</tbody>
</table>

Table 2.5: Chi-square goodness-of-fit test \(c\) of the conformance between primes cumulative distributions \((\pi(x)/\pi(N) \text{ and } \text{Li}(x)/\text{Li}(N))\) and a GBL with exponent \(\alpha(N)\) (equation 2.3.3) in the interval \([1, N]\). The null hypothesis, prime number distribution obeys GBL, cannot be rejected.

The probability density function of a discrete GB random variable \(Y\) is:

\[
f_Y(y) = \Pr(Y = y) = \frac{1}{10^{1-\alpha} - 1}[(y + 1)^{1-\alpha} - y^{1-\alpha}], \ y = 1, 2, ..., 9. \tag{2.3.26}
\]

The associated cumulative distribution function is therefore:

\[
F_Y(y) = \Pr(Y \leq y) = \frac{1}{10^{1-\alpha} - 1}[(y + 1)^{1-\alpha} - 1], \ y = 1, 2, ..., 9. \tag{2.3.27}
\]

How can we prove that a random variable \(T\) extracted from a probability density \(f_T(t) = \Pr(t)\) has an associated (discrete) random variable \(Y\) that follows equation 2.3.26? We can readily find a relation between both random variables. Suppose without loss of generality that the random variable \(T\) is defined in the interval \([1, 10^{D+1}]\). Let the discrete random variable \(D\) fulfill:

\[
10^D \leq T < 10^{D+1} \tag{2.3.28}
\]

This definition allows us to express the first significative digit \(Y\) in terms of \(D\) and \(T\):

\[
Y = \lfloor T \cdot 10^{-D} \rfloor, \tag{2.3.29}
\]

where from now on the floor brackets stand for the integer part function. Now, let \(U\) be a random variable uniformly distributed in \((0, 1)\), \(U \sim U(0, 1)\). Then, inverting the
cumulative distribution function 2.3.27 we come to:

\[ Y = \lfloor (10^{1-\alpha} - 1) \cdot U + 1 \rfloor^{1-\alpha} \]  
(2.3.30)

This latter relation is useful to generate a discrete GB random variable \( Y \) from a uniformly distributed one \( U(0,1) \). Note also that for \( \alpha = 0 \), we have \( Y = \lfloor 9 \cdot U + 1 \rfloor \), that is, a first digit distribution which is uniform \( \Pr(Y = y) = 1/9, \ y = 1, 2, ..., 9 \), as expected. Hence, every discrete random variable \( Y \) that distributes as a GB should fulfill equation 2.3.30, and consequently if a random variable \( T \) has an associated random variable \( Y \), the following identity should hold:

\[ \lfloor T \cdot 10^{-D} \rfloor = \lfloor (10^{1-\alpha} - 1) \cdot U + 1 \rfloor^{1-\alpha} \]  
(2.3.31)

and then,

\[ Z = \left( \frac{T 10^{-D} 10^{1-\alpha} - 1}{10^{1-\alpha} - 1} \right) \sim U(0,1). \]  
(2.3.32)

In other words, in order the random variable \( T \) to generate a GB, the random variable \( Z \) defined in the preceding transformation should distribute as \( U(0,1) \). The cumulative distribution function of \( Z \) is thus given by:

\[ F_Z(z) = \sum_{d=0}^{n} \left\{ \Pr(10^d \leq T < 10^{d+1}) \cdot \Pr \left( \left( \frac{T 10^{-D} (10^{1-\alpha} - 1)}{10^{1-\alpha} - 1} \right) \leq z \middle| 10^d \leq T < 10^{d+1} \right) \right\} = z, \]  
(2.3.33)

that in terms of the cumulative distribution function of \( T \) becomes

\[ \sum_{d=0}^{n} \left\{ F_T(v 10^d) - F_T(10^d) \right\} = z, \]  
(2.3.34)

where \( v \equiv [(10^{1-\alpha} - 1)z + 1]^{1-\alpha} \).

We may take now the power law density \( x^{-\alpha} \) proposed by Pietronero et al. [93] in order to show that this distribution exactly generates Generalized Benford behavior:

\[ f_T(t) = \Pr(t) = \frac{1 - \alpha}{10^{D+1}(1-\alpha) - 1} t^{-\alpha}, \ t \in [1, 10^{D+1}) \]  
(2.3.35)
Its cumulative distribution function will be:

\[ F_T(t) = \frac{t^{1-\alpha} - 1}{10^{(D+1)(1-\alpha)} - 1}, \]  
(2.3.36)

and thereby equation 2.3.34 reduces to:

\[ \sum_{d=0}^{D} \{ F_T(v10^d) - F_T(10^d) \} = \frac{z(10^{1-\alpha} - 1)}{10^{(D+1)(1-\alpha)} - 1} \sum_{d=0}^{D} (10^{1-\alpha})^d = z, \]  
(2.3.37)

as expected.

**GBL holds for prime number distribution.** While the preceding development is in itself interesting in order to check for the conformance to GBL of several distributions, we will restrict our analysis to the prime number cumulative distribution function conveniently normalized in the interval \([1, 10^D]\):

\[ F_T(t) = \frac{\pi(t)}{\pi(10^{D+1})}, \ t \in [1, 10^D+1) \]  
(2.3.38)

Note that previous analysis showed that

\[ \alpha(10^{D+1}) = \frac{1}{\ln(10^{D+1}) - 1}. \]  
(2.3.39)

Since \(\pi(t)\) is a stepped function that does not possess a closed form, the relation 2.3.34 cannot be analytically checked. However a numerical exploration can indicate into which extent primes are conformal with GBL. Note that equation 2.3.34 reduces in this case to

\[ \sum_{d=0}^{D} \left\{ \pi(v \cdot 10^d) - \pi(10^d) \right\} = \pi(10^{D+1})z \]  
(2.3.40)

where \(v \equiv [(10^{1-\alpha(10^{D+1})} - 1)z + 1]^{\frac{1}{1-\alpha(10^{D+1})}}\) and \(z \in [0, 1]\). Firstly, this latter relation is trivially fulfilled for the extremal values \(z = 0\) and \(z = 1\). For other values \(z \in (0, 1)\), we have numerically tested this equation for different values of \(D\), and have found that it is fulfilled with negligible error (we have performed a scatterplot of equation 2.3.40 and have found a correlation coefficient \(r = 1\)).
The same numerical analysis has been performed for the Eulerean logarithmic integral. In this case the equation that must be fulfilled is

$$\sum_{d=0}^{D} \left\{ \text{Li}(v \cdot 10^d) - \text{Li}(10^d) \right\} = \text{Li}(10^{D+1})z,$$

(2.3.41)

with similar remarkable results provided that we fix \( \text{Li}(1) \equiv 0 \) for singularity reasons.
Bibliography


[16] This is something quite intuitive if we take into account that there are typically $N/2$ multiples of two, $N/3$ multiples of three, and so on: the probability that the prime $x$ appears in a random composite number is in average $1/x$.


Chapter 3

New methods in Complex systems analysis

You can know the name of a bird in all the languages of the world, but when you’re finished, you’ll know absolutely nothing whatever about the bird... So let’s look at the bird and see what it’s doing – that’s what counts. I learned very early the difference between knowing the name of something and knowing something.
- Richard Feynman

Hitherto, we have analyzed several particular systems where complexity develops. These systems have been chosen from a wide set of areas and complexity topics: Sociology and social self-organization, Ecology and biodiversity, Pure mathematics and patterns, Computer Science, algorithmic phase transitions and computational complexity, and so on. In this sense, this thesis firstly pretends to give a broad overview of how complexity settles down with the same signatures in systems of different garment. The techniques that have
been employed in each case are however quite standard so far. We have made use of Dynamical Systems theory (including Bifurcation theory) for the stability analysis of evolution equations. We have made use of Network theory for the characterization of the underlying network of interactions in several systems, and Monte Carlo methods for numerical simulations of such systems. Techniques coming from Statistical physics (finite size scaling, mean field and annealed approximation) and Stochastic processes have been also employed at several stages of the system’s description. Finally, Probability, Statistics, and Combinatorics concepts pervade. Pretty complete, but at the end, nothing new. In this chapter we pretend to bridge the gap, by introducing and/or analyzing in detail some new tools and techniques for the description and characterization of different aspects of complexity. Concretely we will introduce two methods: the Self-overlap method and the Visibility graph. The former is a method for the study of discrete cooperative systems which is closely related to the well known Damage Spreading method, and can be used specifically for the stability analysis of Ising-like models. The latter is an algorithm that maps time series into graphs, and stands as a brand new tool for time series analysis. In what follows we will introduce both methods and present some applications of potential interest.
3.1 Self-overlap as a method of stability analysis in Ising models

3.1.1 Section summary

The damage spreading method (DS) provided a useful tool to obtain analytical results of the thermodynamics and stability of the $2D$ Ising model –amongst many others–, but it suffered both from ambiguities in its results and from large computational costs. In this first part of the chapter devoted to ‘new methods for complex systems analysis’, we propose an alternative to DS, the so called self-overlap method. This method is based on the study of correlation functions measured at subsequent time steps as the system evolves towards its equilibrium. In order to show the goodness of the method, we tackle the paradigmatic $2D$ Ising model and compare the results of both DS versus the self-overlap method. Concretely, applying markovian and mean field approximations we obtain both analytical and numerical results on the thermodynamics that agree with the expected behavior. We also provide some analytical results on the stability of the system. Since only a single replica of the system needs to be studied, this method would seem to be free from the ambiguities that afflicted DS. It also seems to be numerically more efficient and analytically simpler.

3.1.2 Introduction

The damage spreading (DS) method [1] is a remarkable tool amongst the many ones developed in recent years in the effort to understand the dynamics of cooperative systems. Very roughly speaking the goal of the method is to study the stability of a cooperative system under a small perturbation: if perturbations die off after some time then the system must be in a stable, ordered state; if small perturbations always get amplified however then the system must be in a disordered, chaotic state. By studying how far away the final states are from the initial ones given the initial perturbation one can get information about the
system such as, for example, the Lyapunov exponents.

Of course there is a key aspect that differentiates cooperative systems from classical dynamical systems. Namely, that in the former case given the complexity of the systems under study we almost never have at our disposal a detailed analytical solution to the equations of motion in order to study the system’s stability under perturbations. Here is where the DS method comes in: its operational side amounts to an algorithm designed to study how small perturbations spread within the system by working in detail how each of the system’s components react to the changes. The method has been applied to many different dynamical systems such as Ising systems [2, 3, 4, 5, 6, 9, 10, 11], Kauffman networks [12, 13, 14], spin glasses [15, 16], cellular automata [17] amongst others, yielding in many cases useful information about their evolution and stability. Succinctly speaking, the algorithm analyzes the evolution of two almost identical states of the system. The damage (difference between the two initial states) is specified as part of the initial conditions. That is, on one side we have a specified state of the system, and on the other we have a replica that only differs in a small perturbation (the damage) from this original state. One then fixes the stochastic evolution to be the same for each replica (in a Monte Carlo simulation the method imposes the same random numbers at each step of time on both copies for instance). As we let the two copies evolve, the method analyzes their distance (Hamming distance) as a function of time. Useful information about the system can then be extracted from this, not only numerically but in some cases also analytically.

However, as was shown in [7], [8] and [9] (and references therein) DS has been shown to be ill-defined in the sense that different—and equally legitimate—algorithmic implementations of the same physical system’s dynamics can yield different DS properties. This ambiguity stems from the fact that while the transfer matrix for the evolution of a single system is completely determined by the one-point correlation functions [9], the simultaneous evolution of two
replicas however is governed by a joint transfer matrix determined by two-point correlation functions. For example, Glauber and both standard and uncorrelated heat bath (HB) algorithms satisfy detailed balance with respect to the same Hamiltonian. It follows that these three different update rules generate the same equilibrium ensemble and are therefore equally legitimate to mimic the evolution in time of an Ising system coupled to a thermal reservoir. Accordingly, the one-point correlation functions for the three cases coincide and the corresponding transfer matrices for single systems are identical. On the other hand the two-point functions for HB and Glauber dynamics are different; hence damage evolves differently in either case (see [9] and references for a extended quantitative version of this argument). As long as the results depend on the algorithm being implemented, one can not assert that the results obtained from a given DS analysis are conclusive and unambiguous. This handicap is a major motivation in order to search for an alternative method of stability analysis. Our goal in this work will be to propose a different approach to study the stability of cooperative systems. By relying heavily on the above mentioned fact that the evolution of a single system is determined only by the one-point correlation functions we will try to eliminate some of ambiguities found in the DS method.

In order to be specific, as a test case we will focus on the study of a well-known type of system: Ising models. In [2] Votja tackled the $2D$ Glauber-Ising model via DS. He obtained results on the thermodynamics (magnetization, ferromagnetic transition) and stability (regular vs. chaotic behavior) of the model both analytically and numerically. Due to the nature of the method however (at every step of time we must keep account of the two replicas), there is obvious room for improving the computational efficiency. This is also the case in the analytical realm, where accounting for the way in which at each step of time the differences between the two copies may increase inevitably leads to lengthy computations (as was shown in [2, 3, 4, 5, 6]). This on itself constitutes a second motivation in order to search for alternative methods.
The method of analysis that we will propose here, so-called the self-overlap method (SO) [18, 19, 20], has already been successfully used in the study of the stability and critical points of random Boolean networks, a system that is multi-component albeit deterministic. In this work we will show that the method can also be successfully applied to a stochastic system such as a spin network. We will obtain analytical and numerical results on the 2D Glauber-Ising model that exactly match those yielded by DS. However, contrary to DS, SO proceeds by handling only one replica and analyzing its own evolution in time –using basically one-point correlation functions at subsequent time steps for that task. The computational costs are thus lower in SO than in DS. As we will see the analytical calculations also become much simpler while yielding the same results. Furthermore, and what is more important, the SO method is free from the ambiguities that afflicted DS due to its use of two replicas. This comes as a direct consequence of the already mentioned fact that on a single replica it does not matter whether one uses Glauber or HB dynamics since they possess the same one-point correlation functions.

We will follow here the development applied by Vojta in [2], comparing in each case the results obtained using DS and our results (using SO). This work is organized as it follows: in section 3.3.3 we quickly introduce both the 2D Glauber-Ising model and SO. We then apply the method to the 2D Glauber-Ising model in section 3.3.4, obtaining a system of equations (master equation) that describe the dynamical evolution of the system. We discuss then how to apply a mean-field approximation to the system, and compare it with the methodology used by Vojta [2]. In section 3.3.5 we obtain an analytical expression for the magnetization of the system in both ferromagnetic/paramagnetic phases similar to that obtained by Vojta [2]. Numerical results are provided at this point in order to validate the mean field approximation assumed in the analytical development. Finally, in section 3.3.6 we provide some analytical and numerical results on the stability of the model, showing that
the system is chaotic (disordered) in the paramagnetic phase. Conclusions are presented in section 3.3.7.

3.1.3 Ising model: damage spreading vs. self-overlap method

Glauber Ising model

We will work with a kinetic Ising model, a lattice of $N$ spins, $s_i \in \{+1, -1\}$, that follows Glauber dynamics. That is, at every time step a lattice site $i$ is chosen at random. If the spin value of site $i$ at time $t$ is given by $s_i(t)$, at time $t+1$ it will be given by:

$$s_i(t+1) = \text{sgn}\left[\phi(h_i(t)) - \frac{1}{2} + s_i(t)\left(\xi_i(t) - \frac{1}{2}\right)\right],$$

(3.1.1)

where $\xi_i(t)$ is a random number such that $\xi_i(t) \in [0, 1)$. The transition probability $\phi(h_i)$ is given by the usual Glauber expression:

$$\phi(h_i(t)) = \frac{e^{h_i(t)/T}}{e^{h_i(t)/T} + e^{-h_i(t)/T}}$$

(3.1.2)

where $T$ denotes the temperature and $h_i(t)$ is the local field seen by spin $i$ at time $t$:

$$h_i(t) = \sum_{j=\text{n.n.}} J_{ij} s_j(t) + h_0.$$  

(3.1.3)

In this expression $h_0$ represents an external magnetic field, and the sum in the interaction term applies only to the nearest neighbors (three for example in an hexagonal lattice). Without loss of generality from now on we will take $h_0 = 0$ and $J_{ij} = 1$.

Damage spreading and self-overlap

As stated above we will use the SO method to study the dynamics of the system. This procedure was introduced by Luque and Ferrera [18] and its underlying philosophy is similar to that of the DS method used by Vojta to study the thermodynamics of phase transitions in spin systems. The main difference between the two procedures lies in that, while damage
spreading uses two copies of a system with slightly different initial conditions (the damage) and computes the evolution of these differences, the self-overlap method uses the difference between successive temporal states of a single system as the system evolves towards equilibrium. For instance, in DS the damage $D(t)$ at time $t$ is defined as:

$$D(t) = \frac{1}{2N} \sum_{i=1}^{N} |s_i^{(1)}(t) - s_i^{(2)}(t)|$$

and measures the (averaged) Hamming distance between the states of the two replicas at that time (i.e., the proportion of sites for which the spin state differs between the system (1) and the damaged replica (2)). In SO however the self-overlap $a(t)$ at time $t$ is defined as one minus the averaged Hamming distance between the states of a spin site at time $t - 1$ and at time $t$:

$$a(t) = 1 - \frac{1}{2N} \sum_{i=1}^{N} |s_i(t) - s_i(t-1)|.$$

In order to describe the time evolution of the system it is useful to define the “up state self-overlap” $a_{++}(t)$ at time $t$ as the average number of spin sites that had $s_i = +1$ both at time $t - 1$ and at time $t$. We also define $a_{--}(t)$, $a_{+-}(t)$ and $a_{-+}(t)$ in a completely similar fashion. By normalization we must then have:

$$a_{++}(t) + a_{--}(t) + a_{+-}(t) + a_{-+}(t) = 1.$$

Since the sites that remain in the same state at times $t - 1$ and $t$ drop from the sum in the definition (3.1.5) we also must have

$$a(t) = 1 - a_{+-}(t) - a_{-+}(t) = a_{++}(t) + a_{--}(t).$$

Once the equilibrium as been reached the relation $a_{+-} = a_{-+}$ must be satisfied, where we have dropped the time dependence to indicate equilibrium values. Then trivially

$$a_{+-} = a_{-+} = \frac{1-a}{2}.$$

At this point it is interesting to note that the self-overlap functions can be understood in terms of autocorrelation functions, more precisely, two-time autocorrelation functions. For
instance, in equation (3.1.5), one can rewrite $|s_i(t) - s_i(t - 1)|$ as $\frac{(1 - s_i(t)s_i(t - 1))}{2}$, which is a shifted autoresponse function measured at subsequent time steps. In a similar way, the rest of self-overlap functions can be written as linear combinations of the basis of autoresponse functions.

Autocorrelation functions have been widely used as efficient tools in order to measure spatial or temporal correlations in physical and biological systems (repeated patterns, relaxation, frustration, etc). Their applications range from investigations in transport properties of fluids [21] or the analysis of climatological models [22] to studies of decoherence in quantum systems [23], to cite but a few. Autocorrelation functions are the center of interest in theoretical studies of the relaxation of non-equilibrium systems. In this sense, much work has been recently done in order to characterize dynamical scaling and other invariant behavior in the ageing regimes of Ising-like systems [24, 25, 26, 27, 28]. In our case it would be fair to say that the self-overlap functions are really measurements of autocorrelations under a different garment. To dwell on a deeper review of the existing literature on autoresponse functions would go beyond the scope of this work however. We would like to emphasize nonetheless that what is new here is: i) the fact that this particular combination of self-correlation functions measured at subsequent time steps manages to capture the essence of the (same-site) temporal correlations in systems that undergo order/disorder phase transitions, and ii) this is then combined with a philosophy inspired by DS, namely: an evolution equation towards the equilibrium state for the correlations, and a mean field approximation directly extracted from DS in order to be able to solve this equation. Once the evolution equation and the mean field approximation are in place the self-overlaps will allow us to study the stability of the different states accessible to the system, and hence the phase transition itself.
3.1.4 Master equation, transition probabilities, and mean field

Master equation

Generally speaking, the self-overlap method would proceed by solving some master evolution equation for the \( a \)'s in order to obtain their equilibrium values, much in the vein of the damage spread method. We begin by defining the probability of finding a spin site in the \(+ (\cdot -)\) state at time \( t \), \( P_+(t) \) \( (P_-(t)) \)

\[
P_{\pm}(t) = \frac{n_{\pm}(t)}{N},
\]

where \( n_{\pm}(t) \) is the number of sites with spin up (down) at time \( t \). Obviously \( P_+(t) + P_-(t) = 1 \). By the definition of \( a_{++}(t), a_{--}(t) \) it follows that

\[
P_+(t) = a_{-+}(t) + a_{++}(t) = \frac{1 + a_{++}(t) - a_{--}(t)}{2}
\]

and analogously for the down states

\[
P_-(t) = a_{+-}(t) + a_{--}(t) = \frac{1 + a_{--}(t) - a_{++}(t)}{2}.
\]

As noted above in the limit \( t \to \infty \) the \( a \)'s ought to reach their equilibrium values and one can drop the \( t \) dependence.

Of particular interest to us will be the transition probabilities from one state to another, i.e., the elements of the transition matrix of our Markov process. Let \( W_{++}(t) \) \( (W_{--}(t)) \) be the average probability of changing from the \(+ (\cdot -)\) state at time \( t \) to the \(+ (\cdot -)\) state at time \( t+1 \), where the precise meaning of this average will be made clear shortly. In an mean field approximation we will then have

\[
a_{++}(t) = W_{++}(t - 1)P_+(t - 1),
\]

\[
a_{--}(t) = W_{--}(t - 1)P_-(t - 1)
\]

(3.1.12)
and analogously with $W_{+-}, W_{-+}$. Note that these $W$’s will then be the elements of an average Markov matrix for the evolution of the system. Combining (3.1.10), (3.1.11) and (3.1.12) together it is easy to arrive to a couple of mean field evolution equations for $a_{++}(t)$ and $a_{--}(t)$, namely

$$\frac{d}{dt}a_{++}(t) = -a_{++}(t)W_{+-}(t) + a_{++}(t)W_{++}(t),$$

$$\frac{d}{dt}a_{--}(t) = -a_{--}(t)W_{-+}(t) + a_{--}(t)W_{--}(t).$$

(3.1.13)

These two equations are of course nothing but the reaction-diffusion equations for the $a$’s that common sense would have dictated us to begin with. We now proceed to evaluate a mean field approximation for the $W$’s so that we may solve (3.1.13).

**Mean-field approximation**

To begin with, note that in a system that follows Glauber dynamics the transition probability at site $i$ for a given local field $h_i$ is given by (3.1.2) above. This means that

$$W_{++}(h_i) = \phi(h_i), \quad W_{+-}(h_i) = 1 - \phi(h_i),$$

$$W_{-+}(h_i) = \phi(h_i), \quad W_{--}(h_i) = 1 - \phi(h_i).$$

(3.1.14)

That is, as is well known for a given local field $h_i$ the probability that the spin at site $i$ will be in the $+$ state at time $t + 1$ is always $\phi(h_i)$, whereas the probability that its state be $-$ will be $1 - \phi(h_i)$, regardless of the initial state of the site. Thus, finding average values for the $W$’s is equivalent to finding an average $\bar{\phi}$. The mean field approximation that we will use closely follows the spirit of the effective-field approximation used by Vojta [2]. This consists basically in averaging over all the possible configurations that can surround a given site, where in the average each configuration is weighted by its probability of taking place. Thus, with three nearest neighbors per site, the
transition probabilities can take the values (remember that we are taking $J_{ij} = 1$)

$$
\begin{align*}
\phi_0{+++} &= \phi(3) = \frac{e^{3/T}}{2 \cosh (3/T)}, \\
\phi_1{+++} &= \phi(1) = \frac{e^{1/T}}{2 \cosh (1/T)}, \\
\phi_2{+-} &= \phi(-1) = \frac{e^{-1/T}}{2 \cosh (-1/T)}, \\
\phi_3{--} &= \phi(-3) = \frac{e^{-3/T}}{2 \cosh (-3/T)}. \\
\end{align*}
\tag{3.1.15}
$$

Note that the calculations are much simpler than those needed in DS [2]. The probability associated to each configuration will be

$$
\begin{align*}
P(\phi_0) &= P_+^3, \\
P(\phi_1) &= 3P_+^2(1 - P_+), \\
P(\phi_2) &= 3P_+(1 - P_+)^2, \\
P(\phi_3) &= (1 - P_+)^3, \\
\end{align*}
\tag{3.1.16}
$$

where to simplify the notation we have dropped the time dependence, although in this case one must be aware that we are not dealing with equilibrium values (this will be the case for the next several equations). Using equations (3.1.10) and (3.1.11), we can now write after some trivial manipulations

$$
\bar{\phi} = \sum_{k=0}^{3} P(\phi_k)\phi_k = 
\begin{align*}
&\frac{1}{2} + \frac{3}{8}(a_{++} - a_{--}) \left[ \tanh \left( \frac{3}{T} \right) + \tanh \left( \frac{1}{T} \right) \right] \\
&+ \frac{1}{8}(a_{++} - a_{--})^3 \left[ \tanh \left( \frac{3}{T} \right) - 3 \tanh \left( \frac{1}{T} \right) \right]. \\
\end{align*}
\tag{3.1.17}
$$

Using the relations between the $a$’s and applying the mean field to the right hand side of the differential equations (3.1.13) we can rewrite them as

$$
\frac{d}{dt}a_{++} = -a_{++}(1 - \bar{\phi}) + \left( \frac{1 - a_{++} - a_{--}}{2} \right) \bar{\phi} \\
\tag{3.1.18}
$$
\[
\frac{d}{dt} a_{--} = -a_{--} \overline{\phi} + \left( \frac{1 - a_{++} - a_{--}}{2} \right) (1 - \overline{\phi}),
\] (3.1.19)

which by (3.1.17) is now a system of equations depending only on \(a_{++}\) and \(a_{--}\). Note that it is easy to generalize the mean field approximation to the case of \(n\) nearest neighbors (that is, for a given topology):

\[
\overline{\phi} = \sum_{k=0}^{n} \binom{n}{k} P^{n-k} P^k \frac{1}{1 + \exp \left( \frac{2n-4k}{T} \right)}. \tag{3.1.20}
\]

This would be much harder to do using DS, if at all possible.

### 3.1.5 Thermodynamics: magnetization

At this point we are going to link the self-overlaps to the average magnetization per spin, \(m\). With \(P_+(t), P_-(t)\) as defined above

\[
P_{\pm}(t) = \frac{n_{\pm}(t)}{N}, \tag{3.1.21}
\]

we must then obviously have for the average magnetization \(m\)

\[
m(t) = P_+(t) - P_-(t), \tag{3.1.22}
\]

or, since \(P_+(t) + P_-(t) = 1\),

\[
P_+(t) = \frac{1 + m(t)}{2}, \quad P_-(t) = \frac{1 - m(t)}{2}. \tag{3.1.23}
\]

By the definition of \(a_{++}(t), a_{--}(t)\) it follows then

\[
a_{--}(t) + a_{++}(t) = P_+(t) = \frac{1 + m(t)}{2}, \tag{3.1.24}
\]

and analogously with \(a_{--}(t), a_{++}(t)\) and \(P_-(t)\).

Since

\[
m(t) = a_{++}(t) - a_{--}(t), \tag{3.1.25}
\]

the system of equations (3.1.17,3.1.18,3.1.19) can be rewritten as
\[
\frac{d}{dt} m = \frac{m}{2} \left\{ -1 + \frac{3}{4} \left[ \tanh \left( \frac{1}{T} \right) + \tanh \left( \frac{3}{T} \right) \right] \right\} + \frac{m^3}{8} \left\{ \tanh \left( \frac{3}{T} \right) - 3 \tanh \left( \frac{1}{T} \right) \right\}.
\]

(3.1.26)

Within the limits of our approximation this equation describes the evolution towards equilibrium of the magnetization \( m \) for the case of \( n = 3 \) nearest neighbors. Setting \( \frac{dm}{dt} = 0 \) one can obtain an expression for the temperature dependence of its equilibrium value \( m(T) \), and from it one can extract the transition temperature for the ferro-paramagnetic transition —this was the approach originally followed by Vojta [2].

Equation (3.1.26) yields a critical temperature \( T_c \approx 2.104 \) above which the magnetization is zero. When \( T < T_c \), we have:

\[
m = \pm \sqrt{\frac{-1 + \frac{3}{4} \left[ \tanh \left( \frac{1}{T} \right) + \tanh \left( \frac{3}{T} \right) \right] \frac{3}{4} \tanh \left( \frac{1}{T} \right) - 1}{\frac{3}{4} \tanh \left( \frac{1}{T} \right) - 1}}.
\]

(3.1.27)

Both results completely coincide with those in [2]. Note however that the calculations involved here have been considerably simpler —again basically due to the fact that in SO we only consider one replica of the system, which results in a considerable reduction in the number of configurations that need to be taken into account.

In our Monte Carlo simulations, the procedure to measure the (equilibrium) self-overlap goes as follows: let us suppose that we generate a random initial condition for the \( N \) spin lattice. Then we let it evolve towards equilibrium by applying the Glauber dynamics with 4 neighbors (square lattice). Once equilibrium has been reached we compute the states of the system for a sufficiently large number of time steps. We have used in all cases \( 10,000 \times N \) time steps for a square lattice of \( N = 100 \times 100 \) spins (that is, defining a system time step \( t \) as \( N \) steps of the simulation, we use \( t = 10000 \) system time steps). If we then count the number of times that a spin site is in the “up” state, \(+\), both at time \( t \) and \( t - 1 \) and
average over all sites and time steps, this will give us the equilibrium value of the up state self-overlap $a_{++}$. Repeating this procedure with the down state, $-\text{-}$, will then obviously give us the “down state self-overlap”, $a_{--}$, and so on. Each value of the simulation is averaged over 100 realizations.

In figure (3.1) we plot the average equilibrium magnetization vs. temperature in order to visualize how our mean-field approximation performs —we note here that we are basically interested in the thermodynamic limit of infinite lattice size and that we are removing the inherent degeneracy of the system by plotting only positive magnetization. First, note that our Monte Carlo simulations in a square lattice (squares) are in fair agreement with the Onsager (infinite size) solution —dashed line— except in the proximity of the phase transition, where finite size effects are relevant and difficult to suppress. Comparing then the Monte Carlo simulations and the mean-field solution with $n = 4$ neighbors we can see that qualitatively speaking they provide the same results, with the mean field typically overestimating the critical temperature. We stress here however that the purpose of this work was not so much to present a mean field technique able to reproduce the exact results, but rather to introduce a new technique able to exactly reproduce previously known mean field results while at a much lower cost. For illustrative purposes and to allow comparison with the results obtained by Vojta we also show in figure 3.1 the mean-field result for $n = 3$ neighbors (hexagonal lattice), which underestimates the $n = 4$ critical temperature $T_c$.

In figure (3.2) we plot the equilibrium values $a^*_{++}, a^*_{--}$ vs. temperature, following the same methodology of figure (3.1): we compare our Monte Carlo simulations (circles) with the numerical resolution of the mean field equations (note that again, the mean field with $n = 3$ underestimates the quantitative behavior and the one with $n = 4$ overestimates it). As we can see in the figure, the self-overlap $a = a_{++} + a_{--}$ acts as an order parameter.
Figure 3.1: Magnetization of the system versus temperature in the case of: (squares) Monte Carlo simulation of a $100 \times 100$ spin square lattice (the solid line here is just a guide for the eye), with 10,000 system steps and averaged over 100 realizations; (dashed-dot line) mean field approximation for $n = 3$ neighbors; (solid line) mean field approximation for $n = 4$ neighbors; (dashed line) Onsager solution. Note that the mean field approximation recovers the expected behavior, that is, null magnetization above $T_c$, non null magnetization below $T_c$, which tends to a constant maximum value at $T = 0$. The difference lies on the quantitative value of $T_c$ in each case, overestimated by the mean field in the case of $n = 4$ neighbors.

### 3.1.6 Stability

In equation (3.1.26), the stability of the fixed point $m^* = 0$ (paramagnetic phase) is related to the sign of the eigenvalue:

$$\lambda(T) = 1/2(-1 + 3/4[\tanh(1/T) + \tanh(3/T)]).$$

(3.1.28)

Note that (3.1.26) falls into the normal form of a pitchfork bifurcation at $T = T_c$ where the fixed point is not hyperbolic and the Hartman-Grobman theorem [29] does not apply. For $T > T_c$, $m^* = 0$ is stable, and below it, it becomes unstable. The fact that we have a pitchfork bifurcation at $T_c$ implies that in the ferromagnetic phase (i.e., below $T_c$) two other stable fixed points must appear. They are indeed $\pm m^*$, where $m^*$ is now given by (3.1.27).
Figure 3.2: Stationary values of $a_{++}$ and $a_{--}$ in the case of: mean-field approximation with $n = 3$ first neighbors (dashed line), mean-field approximation with $n = 4$ first neighbors (solid line) and Monte Carlo simulation of a $100 \times 100$ spin square lattice (here the solid line is just a guide for the eye), with 10,000 system steps and averaged over 100 realizations (circles). Note that at $T_c$ a pitchfork bifurcation takes place in the three cases. The bifurcation value is underestimated by the mean field approximation in the case of $n = 3$ neighbors and overestimated in the case of $n = 4$ neighbors. Below the critical temperature $a_{++} \to 1$ while $a_{--} \to 0$ for a system that chooses the $m = +1$ vacuum, whereas the opposite is true if the system goes to $m = -1$. Above $T_c$ the system tends to $(a_{++},a_{--})=(1/4,1/4)$. Note that although the critical temperature is only predicted qualitatively, the stationary values for $(a_{++},a_{--})$ yielded by our simple model exactly match the Onsager predictions.

Taking into account the relation between $m$ and $a$, with a little algebra we arrive at

$$a^* = m^*(1 - \phi^*) + \phi^*. \hspace{1cm} (3.1.29)$$

Hence, the fixed point $m^* = 0$ leads to $\phi^* = 1/2$ (according to (3.1.17)) and $a^* = 1/2$, which are thus stable at $T > T_c$. Note that $a = 1/2$ is the minimal self-overlap that the system can show.

We can define at this point a Hamming-like distance between successive temporal states (a self-distance), as:

$$d(t) = 1 - a(t). \hspace{1cm} (3.1.30)$$
Figure 3.3: Values of the temperature dependent eigenvalue (3.1.28) of $J$: when it is negative, $(a_+, a_-) = (1/4, 1/4)$ is stable, thus the self-distance $d = 1/2$ is the attractor of the system (chaotic phase). When the eigenvalue is positive, the value $(1/4, 1/4)$ is a saddle point and thus an unstable fixed point. At $T_c \approx 2.104$ the eigenvalue is null, thus the fixed point is not hyperbolic -a bifurcation takes place.-

The fixed point $a^* = 1/2$ implies that we must have a fixed point for $d$ at $d^* = 1/2$ which, since it is taking place at the minimal self-overlap, is equivalent to the maximal self-distance of the system (total disorder). Following Wolf’s method as in the case of random Boolean networks [19], this self-distance would enable us to determine a Lyapunov exponent of the system. However, one can simply apply the Hartman-Grobman theorem directly [29]. Near the fixed points the self-distance of our system can be expressed in terms of $d(t) \sim \exp(\lambda t)$, where $\lambda$ is given by (3.1.28). This eigenvalue can also be understood as a Lyapunov exponent. Note that nevertheless it would not be a standard Lyapunov exponent: when $d(t)$ tends to its fixed point, the system is actually tending to the maximal disorder, thus $\lambda < 0$ means chaos.

Summing up, in the paramagnetic phase, $m^* = 0$ is stable, thus $d^* = 1/2$ is stable too: the system tends exponentially to the maximal disorder and the phase is chaotic.
Figure (3.3) is a plot of equation (3.1.28). Note that when $T > T_c$ (paramagnetic phase) an increase of the temperature leads to an increase of chaos, with the self-distance of the system tending faster to the attractor $d^* = 1/2$.

In the ferromagnetic phase however the stable stationary value of $d$ is

$$d^*(T) = 1 - m^*(T)(1 - \bar{\phi}^*) - \bar{\phi}^*,$$

(3.1.31)

with $m^*$ given by (3.1.27) and $\bar{\phi}^*$ the fixed point value of the mean field. The self-distance tends to zero for low $T$, and thus the system is in a frozen state (order). When we increase the temperature the self-distance also increases up to the maximum value $d = 1/2$, which is reached at $T_c$ (no correlation). These results agree with those found in the paramagnetic phase. We can conclude therefore that our approach correctly reproduces an ordered behavior in the ferromagnetic phase and disordered (chaotic) behavior in the paramagnetic phase. In Appendix (A) we perform a more detailed analysis of the stability of the system that confirms this conclusion.

### 3.1.7 Conclusion

In this first part of the chapter we have introduced the self-overlap method by using it to study both analytically and numerically the 2D Ising model. Since the properties of this model are obviously well known our main concern was to show that SO is an unambiguous method (with respect to changes in the algorithm implementation) that correctly reproduces the standard results while being very advantageous from both the numerical and the analytical point of view. The SO method could thus constitute a rather simple and efficient method of stability analysis in this kind of multicomponent systems (Ising-like models, spin glasses, CA, Kauffman networks, etc). Many other physically relevant quantities in these systems (measures of complexity, information theory measures such as the mutual information, and so on) can be studied and measured by applying SO, something that we think...
deserves further investigation. Wherever damage spreading was supposed to have been useful and the equilibrium state of the system is ergodic, we think that self-overlap ought to work too and do so in a non ambiguous manner. Moreover, it should also be more efficient numerically speaking, and simpler from the analytical viewpoint.

3.1.8 Appendix: Detailed analysis of the stability

We undertake here a deeper study on the stability of the system. For this task we go back to the evolution equations (3.1.18, 3.1.19), which constitute a nonlinear differential system. The fixed points of this system are obtained from equating (3.1.18,3.1.19) to zero (reducing the differential system to a linear system). This yields a total of three fixed points, namely:

$$(a_{++}(T)^*, a_{--}(T)^*) = \left( \frac{\phi}{2} (1 + m^*) - \frac{\phi}{2} (1 + m^*) + \frac{\phi}{2} \right),$$

$$$(a_{++}(T)^*, a_{--}(T)^*) = \left( m^* \frac{\phi}{2} - 1 + \frac{\phi}{2} \frac{\phi}{2} (1 + m^*) \right),$$

(3.1.32)

when $T < T_c$ (where $m^*$ is given by (3.1.27)), and $(a_{++}(T)^*, a_{--}(T)^*) = (1/4, 1/4) \forall T$ (this solution is obviously related to the fixed point $m^* = 0$).

We can write $\bar{\phi}$ as

$${\bar{\phi}} = \frac{1}{2} + \frac{1}{2} A(T)(a_{++} - a_{--}) + \frac{1}{8} B(T)(a_{++} - a_{--})^3,$$  

(3.1.33)

where

$$A(T) = \frac{3}{4} [\tanh(3/T) + \tanh(1/T)],$$

(3.1.34)

and

$$B(T) = [\tanh(3/T) - 3 \tanh(1/T)].$$

(3.1.35)

Let’s start with the stability analysis of the fixed point $(a_{++}^*, a_{--}^*) = (1/4, 1/4)$. This solution is independent of $T$ and for $T > T_c$ is the only fixed point (note that in this case $\bar{\phi}$ takes the value 1/2 independently of the number $n$ of neighbors as it can be proved after some trivial algebra). Computing the jacobian $J$ at this fixed point, we come to:
with eigenvalues $\lambda_1 = -1$, and $\lambda_2 = 1/2(A(T) - 1)$. We will distinguish then three situations: when $A(T) < 1$, $(1/4, 1/4)$ is an hyperbolic (indeed stable) fixed point (which is obviously related to the fact that $m^* = 0$ is stable when $T > T_c$). When $A(T) > 1$ the fixed point is again hyperbolic, but now it is unstable (a saddle point). In these two situations we can apply the developed formalism, due to the Hartman-Grobman theorem [29]. Hence, $A(T) < 1 \iff T > 2/\ln(2^{2/3} + 1) \approx 2.104$ (and viceversa for $A(T) > 1$).

We thus get that when $T > T_c$ (that is, in the paramagnetic phase), the stationary solution $(1/4, 1/4)$ is stable. In the ferromagnetic phase however ($T < T_c$) this fixed point becomes unstable.

At this point we can introduce the self-distance defined in (3.1.30). The stability of the $(1/4, 1/4)$ solution directly implies that $d$ will have a stable value of $1/2$ in the paramagnetic phase, whilst this value will become unstable in the ferromagnetic phase. Since in the paramagnetic phase $(1/4, 1/4)$ is the only fixed point the self-distance necessarily goes to the attractor (stable fixed point) $d^* = 1/2$, indeed exponentially due to the Hartman-Grobman theorem, and the phase is thus chaotic. However in the ferromagnetic phase $(1/4, 1/4)$ is unstable: orbits with initial conditions arbitrarily close from this fixed point will separate from it exponentially, correlations will take place and the phase will become ordered.

When $A(T) = 1$, applying Peixoto’s theorem [29], we can conclude that $(1/4, 1/4)$ is a bifurcation point (lack of structural stability), that is, $T_c$ constitutes a bifurcation value. What kind of bifurcation is taking place?. It is easy to see that the linearized system
has a symmetry of the type $a_{++} - a_{--}$. Using this symmetry, the system of equations (3.1.18,3.1.19) can be transformed into (3.1.26). This equation falls into the normal form of a codimension one bifurcation, a pitchfork bifurcation (indeed, subcritical). This means that two branches of equilibria appear for $T < T_c$ associated with values of $m \neq 0$, either positive (positive branch) or negative. Undoing the change of variables we get that below $T_c$ we must have, for a given $T$, two extra stationary points –other than $(1/4,1/4)$– of the shape $[(a,b),(b,a)]$. These fixed points correspond obviously to (3.1.32). Moreover, since as the Poincaré index is a topological invariant these two new fixed points are both stable in the ferromagnetic phase (in the paramagnetic phase the global index is +1 because the fixed point $(1/4,1/4)$ is a sink, whereas in the ferromagnetic phase $(1/4,1/4)$ is a saddle point with index $-1$, so the other two fixed points must have index +1). Depending on the initial conditions, the system will evolve to a fixed point of the shape $(a,b)$ or to $(b,a)$. In other words, the Ising model will give us either positive or negative magnetization in the ferromagnetic phase, depending on the initial condition. If the system starts at $T > T_c$, where the magnetization is zero, and we lower its temperature below the critical one, fluctuations will take the system either to the upper or to the lower branch indistinctively.

In figure (3.2) we plot together the stationary values $(a_{++}^*, a_{--}^*)$ of the differential system (3.1.18,3.1.19) for both $n = 3$ and $n = 4$ nearest neighbors and the results from our Monte-Carlo simulation (again, a square lattice of $100 \times 100$ spins, where we ran 10,000 system steps after reaching equilibrium, and averaging over 100 realizations). We can see that the results are qualitatively similar, that is, the stationary value $(1/4,1/4)$ is stable above the Curie temperature and unstable below it. As expected, at $T_c$ a pitchfork bifurcation takes place and when $T < T_c$ the system has two stable fixed points, i.e. $(a,b)$ and $(b,a)$ for each $T$. 
3.2 The visibility graph, a mapping between time series and complex networks

3.2.1 Section summary

In this second part of the chapter we proceed by introducing a simple and fast computational method, the visibility algorithm, that converts a time series into a graph. The constructed graph inherits several properties of the series in its structure. Thereby, periodic series convert into regular graphs, and random series do so into random graphs. Moreover, fractal series convert into scale-free networks, enhancing the fact that power law degree distributions are related to fractality, something highly discussed recently. Some remarkable examples and analytical tools are outlined in order to test the method’s reliability. Many different measures, recently developed in the complex network theory, could by means of this new approach characterize time series from a new point of view.

3.2.2 The mapping

Here we introduce a new tool for time series analysis: the visibility algorithm. This algorithm maps a time series into a network. The main idea is to study into which extend the techniques and focus of graph theory are useful as a way to characterize time series. As will be shown below, this network inherits several properties of the time series, and its study reveals non trivial information about the series itself.

For illustrative purposes, in figure (3.4) we present a scheme of the visibility algorithm. In the upper zone we plot the first twenty values of a periodic series using vertical bars (the data values are displayed above the plot). Considering this as a landscape, we link every bar (every point of the time series) with all those that can be seen from the top of the considered one (gray lines), obtaining the associated graph (shown in the lower part of the
Figure 3.4: Example of a time series (20 data) and the associated graph derived from the visibility algorithm. In the graph, every node corresponds, in the same order, to a series data. The visibility rays between the data define the links connecting nodes in the graph.

In this graph, every node corresponds, in the same order, to a series data, and two nodes are connected if there exists visibility between the corresponding data, that is to say, if there is a straight line that connects the series data, provided that this “visibility line” does not intersect any intermediate data height.

More formally, we can establish the following visibility criterium: two arbitrary data \((t_a, y_a)\) and \((t_b, y_b)\) will have visibility, and consequently will become two connected nodes of the associated graph, if any other data \((t_c, y_c)\) placed between them fulfills:

\[
y_c < y_b + (y_a - y_b) \frac{t_b - t_c}{t_b - t_a}.
\]  

We can easily check that by means of the present algorithm, the associated graph extracted from a time series is always:

(i) connected: each node sees at least its nearest neighbors (left and right).

(ii) undirected: the way the algorithm is built up, there is no direction defined in the links.
(iii) invariant under affine transformations of the series data: the visibility criterium is invariant under rescaling of both horizontal and vertical axis, as well as under horizontal and vertical translations (see figure 3.5).

In a recent work [30], Zhang & Small (ZS) introduced another mapping between time series and complex networks. While their philosophy is similar to ours (to encode the time series in a graph in order to characterize the series using graph theory), there exist fundamental differences between both methods, mainly in what refers to the range of applicability (ZS only focus on pseudoperiodic time series, associating each series cycle to a node and defining links between nodes via temporal correlation measures, while the visibility graph can be applied to every kind of time series) and the graph connectedness (in ZS the giant component is assured only ad hoc, meanwhile the visibility graph is always connected by definition).

3.2.3 The visibility graph inherits series structure

The key question is to know whether the associated graph inherits some structure of the time series, and consequently if the process which generated the time series can be characterized using graph theory. In a first step we will consider periodic series. As a matter of fact, the example plotted in figure 3.4 is nothing but a periodic series with period 4. The associated visibility graph is regular, as long as it is constructed by periodic repetition of a pattern. The degree distribution of this graph is formed by a finite number of peaks related to the series period, much in the vein of the Fourier Power Spectrum of a time series. Generally speaking, all periodic time series are mapped into regular graphs, the discrete degree distribution being the fingerprint of the time series periods. In the case of periodic time series, its regularity seems therefore to be conserved or inherited structurally in the graph by means of the visibility map.
Figure 3.5: The visibility graph of a time series remains invariant under several transformation of the time series: a) original time series with visibility links b) translation of the data c) vertical rescaling d) horizontal rescaling e) addition of a linear trend to the data. As can be seen in the bottom figure, in all these cases the visibility graph remains invariant.

As an opposite to periodic series, in a second step we will tackle a series $R(t)$ of $10^6$ data extracted from a uniform distribution in $[0, 1]$. Although one would expect in a first moment a Poisson degree distribution in this case (as for uncorrelated random graphs [31]), a random time series has indeed some correlation, since it is an ordered set. In fact, let $k_t$ be the connectivity of the node associated to the data $t$. If $k_t$ is large (related to the fact that the data has a large value and that consequently it has large visibility), one would expect that $k_{t+1}$ would be relatively small, since the time series is random and two consecutive data with a large value are not likely to occur. It is indeed due to these 'unlikely' large values (the hubs) that the tail of the degree distribution deviates from the Poisson distribution. Two large values in the series data can be understood as two rare events in a random process. The time distribution of these events is indeed exponential [32, ?], therefore we should expect the tail of the degree distribution in this case to be exponential instead of Poissonian, as
long as the form of this tail is related to the hub’s distribution. More concretely, given a
data with a large value (a rare event), the probability $q(t)$ that another large value appears
after $t$ time steps is given by:

$$q(t) = \int_0^t \lambda \exp(-\lambda x) dx = 1 - \exp(-\lambda t).$$  \hspace{1cm} (3.2.2)

Assuming now that a node associated to a large value has enough visibility to be connected
to every node until a new rare event happens, the degree distribution of these rare events
is:

$$p(k) = (1 - q(k - 1)) q(k) = (1 - \exp(-\lambda k)) \exp(-\lambda (k - 1)),$$  \hspace{1cm} (3.2.3)

where $1/\lambda$ stands for the mean of the distribution. In the left side of figure 3.6 we depict the
first 250 values of $R(t)$. In the right side we plot the degree distribution $P(k)$ of its visibility
graph. The tail of this distribution fits quite well an exponential distribution, as expected
(the plot is is semi-log, and therefore a straight line denotes an exponential relation). Note
at this point that time series extracted randomly from other distributions than uniform
have also been addressed. In every case the algorithm captures the random nature of the
series, and the particular shape of the degree distribution of the visibility graph is related to the particular random process.

Hitherto, ordered (periodic) series convert into regular graphs, and random series convert into exponential random graphs: order and disorder structure in the time series seem to be inherited in the topology of the visibility graph. Thus, the question arises: What kind of visibility graph is obtained from a fractal time series? This question is in itself interesting at the present time. Recently, the relationship between self-similar and scale-free networks [33, 38, 39, 40, 41] has been intensively discussed [34, 35, 36, 37]. While complex networks [38] usually exhibit the Small-World property [42] and cannot be consequently size-invariant, it has been recently shown [34] that applying fitted box-covering methods and renormalization procedures, some real networks actually exhibit self-similarity. So, whereas self-similarity seems to imply scale-freeness, the opposite is not true in general.

In order to explore these issues in more detail, the following two fractal series will be considered: the well-known Brownian motion $B(t)$ and the Conway series [43]. While the Brownian motion represents a well-known case of self-affinity (indeed, the following relation holds: $B(t) = a^{1/2}B(t/a)$), the Conway series $a(n) - n/2$ is the recursively generated fractal series from:

$$
\begin{align*}
a(1) &= a(2) = 1 \\
a(n) &= a(a(n - 1)) + a(n - a(n - 1)); \quad n > 2.
\end{align*}
$$

(3.2.4)

In figure 3.7 we have plotted the behavior of these series, the degree distribution $P(k)$ of their respective visibility graphs and their mean path length $L(N)$ as a function of the series length. First, both series have visibility graphs with degree distributions that correspond to power laws of the shape $P(k) \sim k^{-\alpha}$, where we get different exponents in each case: this result enhances the fact that in the context of the visibility algorithm, power law degree
distributions (that is, scale free networks [38, 39, 40, 41]) arise naturally from fractal series. Moreover, this relation seems to be robust as long as the preceding examples show different kinds of fractality: while $B(t)$ stands for a stochastic self-affine fractal, the Conway series is a deterministic series recursively generated.

Figure 3.7: Upper part, from left to right: First 4000 data from a Brownian series of $10^6$ data. In the middle, the degree distribution of the visibility graph associated to the Brownian motion. This one is a power law $P(k) \sim k^{-\alpha}$ with $\alpha = 2.00 \pm 0.01$. In the right part of the figure we plot the mean path length of this network as a function of the network size $N$. The best fitting provides a logarithmic scaling $L(N) = 1.21 + 0.51 \log(N)$. This network shows Small-World effect in addition to being scale-free. Bottom part, from left to right: First $10^5$ data from a Conway series of $4 \cdot 10^6$ data. In the middle, the degree distribution of the visibility graph associated to the Conway series. This one is a power law $P(k) \sim k^{-\alpha}$ with $\alpha = 1.2 \pm 0.1$. The mean path length as a function of the size $N$ is depicted in the right part of the figure. The best fitting provides a power law scaling $L(N) = 0.76N^{0.38}$. Then, this network is scale-invariant.
On the other hand, while the Brownian visibility graph seems to evidence the Small-World effect (right top figure 3.7) since \( L(N) \sim \log(N) \), the Conway series shows in turn a self-similar relation (right bottom figure 3.7) of the shape \( L(N) \sim N^\beta \). This fact can be explained in terms of the so called hub repulsion phenomenon [36]: visibility graphs associated to stochastic fractals such as the Brownian motion or generic noise series do not evidence repulsion between hubs (in these series it is straightforward that the data with highest values would stand for the hubs, and these data would have visibility between each other), and consequently won’t be fractal networks following Song et al. [36]. On the other hand, the Conway series actually evidence hub repulsion: this series is concave-shaped and consequently the highest data won’t in any case stand for the hubs; the latter ones would be located much likely in the monotonic regions of the series, which are indeed hidden from each other (effective repulsion) across the series local maxima. The Conway visibility graph is thus fractal.

Since a fractal series is characterized by its Hurst exponent, we may argue that the visibility graph can actually distinguish different types of fractality, something that will be explored in detail in further work. Note at this point that some other fractal series have been also studied (Q series [44], Stern series [45], Thue-Morse series [46], etc) with similar results. Moreover, observe that if the series under study increases its length, the resulting visibility graph can be interpreted in terms of a model of network growth and may eventually shed light into the fractal network formation problem.

In order to cast light into the relation between fractal series and power law distributions, in the left part of figure 3.8 we present a deterministic fractal series generated by iteration of a simple pattern of three points. The series starts (step 0) with three points \((A,B,C)\) of coordinates \((0,1)\), \((1,1/3)\) and \((2,1/3)\) respectively. In step \(p\), we introduce \(2^{p+1}\) new points with height \(3^{-p-1}\) and distanced \(3^{-p}\). The series form a self-similar set: applying an isotropic zoom of horizontal scale \(3^p\) and vertical scale \(3^p\) to the pattern of order \(p\), we
recover the original pattern.

Figure 3.8: Left: Fractal series obtained by iteration of the original pattern (points A,B,C) with \( p = 10 \) steps. Right: Values of \( K_r \) (circles) and \( K_l \) (squares) as a function of the fractal size, related to equations (3.2.5,3.2.7). Note that the plot is log-linear: the relation is thus exponential. The straight lines correspond to the approximations deduced in equations (3.2.6,3.2.8).

Note that this time series is not data uniformly spaced as the previous examples. However its usefulness is set on the fact that it is simple enough to handle it analytically, that is to say, to find the degree distribution of its visibility graph. The main idea is to find a recurrence behavior in the way that a given node increases its connectivity when the fractal step (hence the fractal size) is increased [47]. Then we calculate how many nodes (self-similar to it) appear in each step, and from both relations we obtain the degree distribution.

First, from a quick visual exploration of left figure 3.8 one comes to the conclusion that nodes A and B have ‘typically’ the same degree. In the other hand, the degree of node C can be decomposed in two terms, the left degree (due to visibility of nodes placed at the left hand side of C) and right degree. The degree of A and B is the same as the right degree of C (statistically speaking, A and B increase their connectivity as the fractal size increases much in the way of the right part of C). Thereby, the degree of C contains the
whole information of the system. We will quote $K_r(C, n)$ the right degree of node $C$ in an $n$-step fractal (respectively, $K_l(C, n)$ stands for the left degree).

Applying the visibility criterium, one can geometrically find that

$$K_r(C, n) = \sum_{m=1}^{n} \frac{1}{m} \sum_{d|m} \mu(d) \cdot 2^{m/d}, \quad (3.2.5)$$

where $\mu$ is the Moebius function. Note that this summation agrees with the number of irreducible polynomials of degree at most $n$ over the Galois field GF(2) [48], something which deserves an in-depth investigation. This expression can be approximated by

$$K_r(C, n) \sim 2^{4n/5}. \quad (3.2.6)$$

On the other hand, there is a recurrence for the left degree that reads

$$K_l(C, n) = 2K_l(C, n-1) + 1, \quad (3.2.7)$$

whose leading term is

$$K_l(C, n) \sim 2^n. \quad (3.2.8)$$

The degree of node $C$ is $K(C) = K_r(C, n) + K_l(C, n)$. In figure 3.8 (right) we plot the values of $K_r$ (circles) and $K_l$ (squares) as a function of the fractal size (the number of iterations $n$). Exact values are plotted as the outer circles and squares, while the inner dots represent equations (3.2.5,3.2.7). Note that both formulas reproduce the exact data. The straight lines correspond to the approximation equations (3.2.6) and (3.2.8).

Now, in a generic step $p$, $2^p$ nodes which are self-similar to $C$ appear (by construction). Those nodes will have a generic degree $K(C, n-p) = 2^4(n-p) + 2^{n-p}$ that, for large values of $n-p$, can be approximated to $K(C, n-p) \simeq 2^{n-p}$. Defining $f(K)$ as the degree distribution, we get that $f(K(C, n-p)) = 2^p$, and with the change of variable $u \equiv 2^{n-p}$, we finally have

$$f(u) \sim u^{-1}, \quad (3.2.9)$$
that is, the degree distribution related to the C-nodes is a power law.

Although this simple example doesn’t provide a general explanation of why series fractality is traduced in power law degree distributions, we may use the preceding arguments as a general methodology for those deterministic fractal series generated by iteration.

### 3.2.4 Some remarks

Once the visibility method has been presented, some remarks can be stated:

Note that typically two series that only differ by an affine transformation will have the same visibility graph; in this sense the algorithm absorbs the affine transformation. Furthermore, it is straightforward to see that that some information regarding the time series is inevitably lost in the mapping from the fact that the network structure is completely determined in the (binary) adjacency matrix. For instance, two periodic series with the same period as $T_1 = \{.., 3, 1, 3, 1, ..\}$ and $T_2 = \{.., 3, 2, 3, 2, ..\}$ would have the same visibility graph albeit being quantitatively different. While the spirit of the visibility graph is to focus on time series structural properties (periodicity, fractality, etc), the method can be trivially generalized using weighted networks (where the adjacency matrix isn’t binary and the weights determine the slope of the visibility line between two data) if we eventually need to quantitatively distinguish time series like $T_1$ and $T_2$ for instance.

While in this introduction we have only tackled undirected graphs, note that one could also extract a directed graph (related to the temporal axis direction) in such a way that for a given node one should distinguish two different connectivities: an ingoing degree $k_{in}$, related to how many nodes see a given node $i$, and an outgoing degree $k_{out}$, that is the number nodes that node $i$ sees. In that situation, if the direct visibility graph extracted from a given time series is not invariant under time reversion (that is, if $P(k_{in}) \neq P(k_{out})$),
one could assert that the process that generated the series is not conservative. In a first approximation we have studied the undirected version and the directed one will be eventually addressed in further work.

There are some direct applications of the method that can be put forward. The relation between the exponent of the degree distributions and the Hurst exponent of the series will be addressed in next section. In particular, it turns out that the method presented here constitutes a reliable tool to estimate Hurst exponents, as far as a functional relation between the Hurst exponent of a fractal series and the degree distribution of its visibility graph holds. Note that the estimation of Hurst exponents is an issue of major importance in data analysis that is yet to be completely solved (see for instance [50]). Moreover, the ability of the algorithm to detect not only the difference between random and chaotic series but also the spatial location of inverse bifurcations in chaotic dynamical systems is another fundamental issue that will also be at the core of further investigations [49]. Finally, the visibility graph characterizes non trivial time series and in that sense, the method may be relevant in specific problems of different garments, such as human behavior time series recently put forward [51].

In summary, a brand new algorithm that converts time series into graphs is presented. The structure of the time series is conserved in the graph topology: periodic series convert into regular graphs, random series into random graphs and fractal series into scale-free graphs. Such characterization goes beyond, since different graph topologies arise from apparently similar fractal series. In fact, the method captures the hub repulsion phenomenon associated to fractal networks [36] and thus distinguishes scale free visibility graphs evidencing Small-World effect from those showing scale invariance.

With the visibility algorithm a natural bridge between complex networks theory and time series analysis is now built.
3.3 The visibility graph: a new Hurst exponent estimator

3.3.1 Section summary

Fractional Brownian motion (fBm) has been used as a theoretical framework to study real time series appearing in diverse scientific fields. Because its intrinsic non-stationarity and long range dependence, its characterization via the Hurst parameter $H$ requires sophisticated techniques that often yield ambiguous results. In this third part of the chapter we show that fBm series map into a scale free visibility graph whose degree distribution is a function of $H$. Concretely, it is shown that the exponent of the power law degree distribution depends linearly on $H$. This also applies to fractional Gaussian noises (fGn) and $f^{-\beta}$ noises. Taking advantage of these facts, we propose a brand new methodology to quantify long range dependence in these series. Its reliability is confirmed with extensive numerical simulations and analytical developments. Finally, we illustrate this method analyzing the persistent behavior of human gait dynamics.

3.3.2 Fractional Brownian motion

Self-similar processes such as fractional Brownian motion (fBm) [52] are currently used to model fractal phenomena of different nature, ranging from Physics or Biology to Economics or Engineering. To cite a few, fBm has been used in models of electronic delocalization [53], as a theoretical framework to analyze turbulence data [54], to describe geologic properties [55], to quantify correlations in DNA base sequences [56], to characterize physiological signals such as human heartbeat [57] or gait dynamics [58], to model economic data [59] or to describe network traffic [62, 60, 61]. Fractional Brownian motion $B_H(t)$ is a non-stationary random process with stationary self-similar increments (fractional Gaussian noise) that can
be characterized by the so called Hurst exponent, $0 < H < 1$. The one-step memory Brownian motion is obtained for $H = \frac{1}{2}$, whereas time series with $H > \frac{1}{2}$ shows persistence and anti-persistence if $H < \frac{1}{2}$.

While different fBm generators and estimators have been introduced in the last years, the community lacks consensus on which method is best suited for each case. This drawback comes from the fact that fBm formalism is exact in the infinite limit, i.e. when the whole infinite series of data is considered. However, in practice, real time series are finite. Accordingly, long range correlations are partially broken in finite series, and local dynamics corresponding to a particular temporal window are overestimated. The practical simulation and the estimation from real (finite) time series is consequently a major issue that is, hitherto, still open. An overview of different methodologies and comparisons can be found in [62, 63, 64, 65, 66, 67, 68, 69] and references therein. Most of the preceding methods operate either on the time domain (e.g. Aggregate Variance Method, Higuchi’s Method, Detrended Fluctuation Analysis, Range Scaled Analysis, etc) or in the frequency or wavelet domain (Periodogram Method, Whittle Estimator, Wavelet Method).

### 3.3.3 The visibility graph as a new Hurst exponent estimator: analytical developments and numerical simulations

Here we introduce an alternative and radically different method, the Visibility Algorithm, based in graph theoretical techniques. In the preceding section this new tool for analyzing time series has been presented. A preliminary analysis has shown that series structure is inherited in the visibility graph. Accordingly, periodic series map into regular graphs, random series into random graphs and fractal series into scale free graphs [71]. In particular, it has been shown that the visibility graph obtained from the well-known Brownian motion
has got both the scale-free and the small world properties [70]. In this third part of the chapter we show that the visibility graphs derived from generic fBm series are also scale free. This robustness goes further, and we prove that a linear relation between the exponent $\gamma$ of the power law degree distribution in the visibility graph and the Hurst exponent $H$ of the associated fBm series exists. Therefore, the visibility algorithm provides an alternative method to compute the Hurst exponent and then, to characterize fBm processes. This also applies to fractional gaussian noise (fGn) [52] which are nothing but the increments of a fBm, and generic $f^{-\beta}$ noises, enhancing the visibility graph as a method to detect long range dependence in time series.

In fig.3.9 we have depicted in log-log the degree distribution of the visibility graph associated with three artificial fBm series of $10^5$ data, namely (i) triangles: extracted from a fBm series of $10^5$ data with $H = 0.3$, (ii) squares: extracted from a fBm series of $10^5$ data with $H = 0.5$, (iii) circles: extracted from a fBm series of $10^5$ data with $H = 0.8$. Note that distributions are not normalized. The three visibility graphs are scale-free since their degree distributions follow a power law $P(k) \sim k^{-\gamma}$ with decreasing exponents $\gamma_{0.3} > \gamma_{0.5} > \gamma_{0.8}$.

Figure 3.9: Degree distribution of three visibility graphs, namely (i) triangles: extracted from a fBm series of $10^5$ data with $H = 0.3$, (ii) squares: extracted from a fBm series of $10^5$ data with $H = 0.5$, (iii) circles: extracted from a fBm series of $10^5$ data with $H = 0.8$. Note that distributions are not normalized. The three visibility graphs are scale-free since their degree distributions follow a power law $P(k) \sim k^{-\gamma}$ with decreasing exponents $\gamma_{0.3} > \gamma_{0.5} > \gamma_{0.8}$.
$H = 0.3$ (triangles), a memoryless Brownian motion with $H = 0.5$ (squares) and a persistent fBm with $H = 0.8$ (circles). As can be seen, these distributions follow a power law $P(k) \sim k^{-\gamma}$ with decreasing exponents $\gamma_{0.3} > \gamma_{0.5} > \gamma_{0.8}$.

Figure 3.10: (Black dots) Numerical estimation of exponent $\gamma$ of the visibility graph associated with a fBm series with exponent $H$. In each case $\gamma$ is averaged over 10 realizations of a fBm series of $10^4$ data, in order to avoid non-stationary biases (the error bars are included in the dot size). The straight line corresponds to the numerical fitting $\gamma(H) = a - bH$, where $a = 3.1 \pm 0.1$ and $b = 2.0 \pm 0.1$.

In order to compare $\gamma$ and $H$ appropriately, we have calculated the exponent of different scale free visibility graphs associated with fBm artificial series of $10^4$ data with $0 < H < 1$ generated by a wavelet based algorithm [74]. Note at this point that some bias is inevitably present since artificial series generators are obviously not exact, and consequently the nominal Hurst exponents have an associated error [72]. For each value of the Hurst parameter we have thus averaged the results over 10 realizations of the fBm process. We have estimated exponent $\gamma$ in each case through Maximum Likelihood Estimation (MLE) [76]:

$$\gamma = 1 + n \left[ \sum_{i=1}^{n} \log \frac{x_i}{x_{\min}} \right]^{-1},$$

(3.3.1)

where $n$ is total number of values taken into account, $x_i, i = 1, \ldots, n$ are the measured values and $x_{\min}$ corresponds to the smallest value of $x$ for which the power law behavior holds. In
fig.3.10 we have represented the relation between $\gamma$ and $H$ (black circles). As can be seen, a linear relation holds (the straight line represents the fitting $\gamma = 3.1 - 2H$).

That fBm yields scale free visibility graphs is not that surprising. The most highly connected nodes (hubs) are the responsible for the heavy tailed degree distributions. Within fBm series, hubs are related to extreme values in the series, since a data with a very large value has typically a large connectivity. In order to calculate the tail of the distribution we consequently need to focus on the hubs, and thus calculate the probability that an extreme value has a degree $k$. Suppose that at time $t$ the series reaches an extreme value (a hub) $B_H(t) = h$. The probability of this hub to have degree $T$ is

$$P(T) \sim P_{fr}(T)r(T),$$

(3.3.2)

where $P_{fr}(T)$ provides the probability that after $T$ time steps, the series returns to the same extreme value, i.e. $B(t + T) = h$ (and consequently the visibility in $t$ gets truncated in $t + T$), and $r(T)$ is the ratio of nodes between $t$ and $t + T$ that $t$ may see. $P_{fr}(T)$ is
nothing but the first return time distribution, which is known to scale as \( P_{fr}(T) \sim T^{H-2} \) for fBm series [73]. On the other hand, the ratio of visible nodes between two extreme values is related to the roughness of the series in that basin, that is, to the way that a series of \( T \) time steps folds. This roughness is encoded in the series standard deviation [52], such that intuitively, we have \( r(T) \sim T^H/T = T^{H-1} \) (this fact has been confirmed numerically). Finally, notice that in this context \( T \equiv k \), so eq.3.3.2 converts into

\[
P(k) \sim k^{H-2}k^{H-1} = k^{2H-3},
\]

what provides a linear relation between the exponent of the visibility graph degree distribution and the Hurst exponent of the associated fBm series:

\[
\gamma(H) = 3 - 2H,
\] (3.3.4)

in good agreement with our previous numerical results. A scatter plot of the theoretical (eq.3.3.4) versus the empirical estimation of \( \gamma(H) \) provides statistical conformance with a correlation coefficient \( c = 0.99 \).

To check further the consistency of the visibility algorithm an estimation of the power spectra is performed. It is well known that fBm has a power spectra that behaves as \( 1/f^\beta \), where the exponent \( \beta \) is related to the Hurst exponent of an fBm process through the well known relation [75]

\[
\beta(H) = 1 + 2H.
\] (3.3.5)

Now according to eqs.3.3.4 and 3.3.5, the degree distribution of the visibility graph corresponding to a time series with \( f^{-\beta} \) noise should be again power law \( P(k) \sim k^{-\gamma} \) where

\[
\gamma(\beta) = 4 - \beta.
\] (3.3.6)

In fig.3.11 we depict (triangles) the empirical values of \( \gamma \) corresponding to \( f^{-\beta} \) artificial series of \( 10^6 \) data with \( \beta \) ranging from 1.2 to 2.8 in steps of size 0.1. For each value of \( \beta \) we have again averaged the results over 10 realizations and estimated \( \beta \) through MLE
The straight line corresponds to the theoretical prediction eq.3.3.6, showing good agreement with the numerics. In this case, a scatter plot confronting theoretical versus empirical estimation of $\gamma(\beta)$ also provides statistical conformance between them, up to $c = 0.99$.

Finally, observe that eq.3.3.5 holds for fBm processes, while for the increments of an fBm process, known as a fractional Gaussian noise (fGn), the relation between $\beta$ and $H$ turns to be [75]:

$$\beta(H) = -1 + 2H,$$  \hspace{1cm} (3.3.7)

where $H$ is the Hurst exponent of the associated fBm process. We consequently can deduce that the relation between $\gamma$ and $H$ for a fGn (where fGn is a series composed by the increments of a fBm) is

$$\gamma(H) = 5 - 2H.$$  \hspace{1cm} (3.3.8)

Notice that eq.3.3.8 can also be deduced applying the same heuristic arguments as for eq.3.3.4 with the change $H \rightarrow H - 1$.

### 3.3.4 Application to gait cycle series

In order to illustrate this latter case, we finally address a realistic and striking dynamics where long range dependence has been recently described. Gait cycle (the stride interval in human walking rhythm) is a physiological signal that has been shown to display fractal dynamics and long range correlations in healthy young adults [77, 78]. In the upper part of fig.3.12 we have plotted to series describing the fluctuations of walk rhythm of a young healthy person, for slow pace (bottom series of 3304 points) and fast pace (up series of 3595 points) respectively (data available in www.physionet.org/physiobank/database/umwdb/[79]). In the bottom part we have represented the degree distribution of their visibility graphs. These ones are again power laws with exponents $\gamma = 3.03 \pm 0.05$ for fast pace and $\gamma = 3.19 \pm 0.05$ for slow pace (derived through MLE). According to eq.3.3.6, the visibility
algorithm predicts that gait dynamics evidence $f^{-\beta}$ behavior with $\beta = 1$ for fast pace, and $\beta = 0.8$ for slow pace, in perfect agreement with previous analysis using Detrended Fluctuation Analysis [77, 78]. These series record the fluctuations of walk rhythm (that is, the increments), so according to eq.3.3.8, the Hurst exponent is $H = 1$ for fast pace and $H = 0.9$ for slow pace, that is to say, dynamics evidences long range dependence (persistence) [77, 78].

3.3.5 Conclusions

As a summary, the visibility graph is an algorithm that map a time series into a graph. In so doing, classic methods of complex network analysis can be applied to characterize
time series from a brand new viewpoint. In this section we have pointed out how graph theory techniques can provide an alternative method to quantify long range dependence and fractality in time series. We have reported analytical and numerical evidences showing that the visibility graph associated to a stochastic fractal series with Hurst exponent $H$ is a scale free graph, whose degree distribution follows a power law $P(k) \sim k^{-\gamma}$ such that $H(\gamma) = \frac{3-\gamma}{2}$ for fBm and $H(\gamma) = \frac{5-\gamma}{2}$ for fGn. We have also proved that the power spectra $f^{-\beta}$ is related to the degree distribution such that $\beta(\gamma) = 4 - \gamma$.

The reliability of this methodology has been confirmed with extensive simulations of artificial fractal series and real (small) series concerning gait dynamics. To our knowledge, this is the first method for estimation of long range dependence in time series based in graph theory techniques advanced so far. Some questions concerning its accuracy, flexibility and computational efficiency will be at the core of further investigations. In any case, we do not pretend in this work to compare its accuracy with other estimators, but to propose an alternative and simple method based in completely different techniques with potentially broad applications.
Bibliography


[49] To be published.


Chapter 4

Conclusions, perspectives and published papers

I’m on this business to have fun.
-Bartolo Luque

Philosophers say a great deal about what is absolutely necessary for science, and it is always, so far as one can see, rather naive, and probably wrong.
-Richard Feynman

4.1 Summary of conclusions and perspectives

The ubiquitous heavy-handed concluding summary should be omitted; a talk should tell such a good story that a summary is uncalled for. Imagine ‘War and Peace’ ending with a summary.
While I definitively agree with professor Mermin, I understand that the present thesis tackles different topics of sufficiently different garment to merit a concluding summary.

In Chapter 1 we have focused on two particular problems concerning many body systems modeling in the context of Life Sciences. The main findings are:

- Social stratification is a collective induced phenomenon and is ubiquitous not only in human behavior but also in Ethology. The mechanisms of hierarchy formation can be many in principle. Now, can hierarchy appear in societies only by means of a stochastic amplification mechanism? That is, do homogeneous societies tend to be stratified? Is social equality an unstable state? The Bonabeau model provides an affirmative answer. It shows sudden hierarchy formation beginning from equality due to stochastic amplification of the agents perturbations at some critical density. In the first section of chapter 1 we have provided analytical evidences of the phase transition between egalitarian and hierarchical states that the Bonabeau model evidences, and have pointed out that the hierarchical regime evidences a rich structure. Furthermore, we have also tackled analytically a modified version of this model due to Stauffer, which has been tackled numerically in recent years. To our surprise, we have found that this latter model is ill-defined since no hierarchy formation takes place if the initial state is homogeneous (equality is always stable in this model).

- In the second section of Chapter 1 we have developed a probabilistic framework to optimum reserve design. It dictates the optimum size allocation among a set of $r$ reserves. We have found analytically that a simple variable $k$ depending on the area allocated to reserves and the regional species richness is a key determinant of the best size distribution. For high regional species richness and low reserve areas, a uniform
area distribution maximizes biodiversity. For low regional species richness and high reserve areas, the optimum size allocation consists of allocating a certain area to one reserve and uniformly distributing the remaining area among the remaining reserves.

In Chapter 2 we have focused on purely mathematical systems in order to investigate if complex behavior can emerge in what we could call laboratory conditions. The main findings are

- First section: In this first part of the chapter a (stochastic) algorithmic model which stands for a prime number generator has been considered. This model exhibits for every finite size a phase transition which distinguishes a phase where the algorithm has the ability to reduce every element into a prime, and a phase where the system is rapidly frozen. Analytical and numerical evidences suggest that the transition is continuous. On a second part, the model has been reinterpreted as a search problem. As long as the model searches paths to reduce integers into primes, the combinatorial problem is related to primality test and decomposition problem. It has been shown that this model belongs to the $NP$ class in a worst-case classification, moreover, an easy-hard-easy pattern has been found, as common in many algorithmic phase transitions. According to the fact that the transition is continuous, and based on previous works, it has been put into relevance that the average-case complexity may be only polynomial. This hardness reduction is in turn related to the fact that the algorithm only yields probable states.

- Second section: In this second part of the chapter we have presented a new general mechanism by which simple dynamics running on networks become self-organized critical for scale free topologies. We have illustrated this mechanism within a simple arithmetic model of division between integers, the division model. This is the simplest self-organized critical model advanced so far, and in this sense it may help to elucidate the mechanism of self-organization to criticality. Its simplicity allows analytical
tractability, characterizing several scaling relations. Furthermore, its mathematical
nature brings about interesting connections between statistical physics and number
theoretical concepts. We have shown how this model can be understood as a self-
organized stochastic process embedded on a network, where the onset of criticality is
induced by the topology.

• Third section: In the third part of the chapter we have unveiled a statistical pattern
in the prime numbers and the nontrivial Riemann zeta zeros sequences that has
surprisingly gone unnoticed until now. Several applications and future work have
been pointed out: first, since the Riemann zeros seem to have the same statistical
properties as the eigenvalues of a concrete type of random matrices called the Gaussian
Unitary Ensemble, the relation between GBL and random matrix theory should be
investigated. Second, this finding may also apply to several other sequences that,
while not being strictly Benford distributed, can be GBL, and in this sense much work
recently developed for Benford distributions could be readily generalized. Finally, it
has not escaped our notice that several applications recently advanced in the context
of Benford’s law, such as fraud detection or stock market analysis, could eventually be
generalized to the wider context of GBL formalism. This generalization also extends
to stochastic sieve theory, dynamical systems that follow Benford’s law and their
relation to stochastic multiplicative processes.

Chapter 3 has been devoted for the introduction of some new methods for Complex System’s
analysis. The main conclusions are.

• First section: In this first part of the chapter we have introduced the self-overlap
method (SO), a method for the stability analysis of Ising-like systems. This method
pretends to be an alternative to the well known Damage Spreading (DS). We have
tackled as a test case the thermodynamics of the 2D Ising model. The results point
out that SO correctly reproduces the standard results provided by DS while being
very advantageous from both the numerical and the analytical point of view, and also being free from the ambiguities that afflicted DS. The SO method could thus constitute a rather simple and efficient method of stability analysis in this kind of multicomponent systems (Ising-like models, spin glasses, CA, Kauffman networks, etc).

- Second section section: In a second part we have presented a simple and fast computational method, the \textit{visibility algorithm}, that converts a time series into a graph. The constructed graph inherits several properties of the series in its structure. Thereby, periodic series convert into regular graphs, and random series do so into random graphs. Moreover, fractal series convert into scale-free networks, enhancing the fact that power law degree distributions are related to fractality, something highly discussed recently. Some remarkable examples and analytical tools are outlined in order to test the method’s reliability. Many different measures, recently developed in the complex network theory, could by means of this new approach characterize time series from a new point of view.

- Third section: In the last part of the chapter we have pointed out how the visibility algorithm constitutes an alternative method to quantify long range dependence and fractality in time series. We have reported analytical and numerical evidences showing that the visibility graph associated to a generic fractal series with Hurst exponent $H$ is a scale free graph, whose degree distribution follows a power law $P(k) \sim k^{-\gamma}$ such that $H(\gamma) = \frac{3-\gamma}{2}$ for fBm and $H(\gamma) = \frac{5-\gamma}{2}$ for fGn. We have also proved that the power spectra $f^{-\beta}$ is related to the degree distribution such that $\beta(\gamma) = 4 - \gamma$. The reliability of this methodology has been confirmed with extensive simulations of artificial fractal series and real (small) series concerning gait dynamics. To our knowledge, this is the first method for estimation of long range dependence in time series based in graph theory techniques advanced so far. Some questions concerning
its accuracy, flexibility and computational efficiency should be at the core of further investigations.

Finally, we can draw some general conclusions. First, that even extremely simple systems, such as the division model for instance, can develop complexity. This finding strengthens the assumption that the emergence of complex behavior usually depends on few aspects of the systems (interactions) and eventually pushes our knowledge of Complexity a little step forward. Second, some interesting connections between apparent separate fields have been presented. Such is the case of the relations between Statistical Physics and Number theory (Chapter 2), or the mapping between Time Series Analysis and Graph theory (Chapter 3). Further work should eventually confirm and evaluate the extent of these latter connections. I can lastly affirm that I had such an enormous fun working on this business, despite the fact that this PhD has been completely realized without any particular fellowship (what is dramatically quite common in Spain).

4.2 Publications derived from this thesis

In what follows we depict the publications derived from this work. We have skipped proceeding papers and communications published in conferences or elsewhere and have restricted ourselfed to publications in indexed journals with impact factor (according to the Journal of Citation Report (JCR)).

4.2.1 Chapter 1


4.2.2 Chapter 2


4.2.3 Chapter 3

