Characterizing the Solar Activity Using the Visibility Graph Method

Thesis
submitted to the
University Of Chile
in partial fulfillment of the requirements
for the degree of
Master of Sciences with mention in Physics
Faculty of Sciences

by

Tomás Zurita Valencia

April, 2024

Thesis Advisor: **Dr. Víctor Muñoz**

FACULTY OF SCIENCES

UNIVERSITY OF CHILE

APPROVAL REPORT

MASTER THESIS

The Graduate School of the Faculty of Sciences is informed that the Master's Thesis presented by the candidate

Tomás Zurita Valencia

has been approved by the Thesis Evaluation Committee as a requirement for the Master's degree, in the Private Thesis Defense examination given on _____, 2024.

Thesis Advisor

Dr. Víctor Muñoz _____

Thesis Evaluation Committee

Dra. Denisse Pastén _____

Dr. Víctor Pinto _____

Dr. Rodrigo Soto _____

Biography

I was born in Valparaíso on February 18, 1998. I am the firstborn son of Julio and Nilda, and the elder brother of Gonzalo. Both my parents are from Bolivia, they settled in Chile shortly before I was born. I grew up in Quilpué, where I completed my primary and secondary education at Colegio Coeducacional de Quilpué. After finishing high school I decided to study Physics at the Universidad de Chile. At that moment, I learned there are two bachelor's degree programs in Physics at the university, but I chose to study at the Faculty of Sciences. Then I moved to Santiago. After finishing my BSc in Physics, I decided to continue my studies by pursuing a graduate program at the same university. I decided to enter the MSc in Physics program and am now carrying it out through this thesis.

Acknowledgments

I would like to thank the friends I have met with whom I have shared most of my time during this process, for your company and the many conversations, laughs and hugs shared. You were crucial in this path and I would not have made it without you and your help. Special mention to Nicolás and Sebastián, we began together our undergraduate studies in 2016 and we are almost finishing our graduate studies now, I am grateful to you for giving me the space and trust to be your friend. Thank you to Alejandro for these past years in the office and for making that space a happier place.

To my advisor, professor Víctor Muñoz, I would like to thank you for guiding me in the turbulent process of becoming a scientist. I learned a lot from you. Thank you for your patience, your help, your advice and for believing in me. I would also like to thank professors Pablo Moya and Denisse Pastén for their advice, guidance and orientation.

Finally, I would like to thank four special people who are a pillar in my life, my parents, my brother and my Bele. Thank you to my parents, for your love, encouragement and guidance in my whole life, you have always been a constant support. Thank you to my brother, for allowing me to be your brother, for all the laughs and good moments, for your tolerance, for your love and for being you. Thank you to my girlfriend Bele, for your love and company, for being there encouraging me and listening to me when I thought I could not do it or when I did not have confidence in myself. Thank you to these four special people for that unconditional love and support.

Contents

Abstract

In this thesis, the Sun and its behavior are studied by means of complex networks. The complex network was built using the Visibility Graph algorithm. This method maps time series into graphs in which every element of the time series is considered as a node and a visibility criterion is defined in order to connect them. Using this method, we construct complex networks for magnetic field and sunspots time series encompassing four solar cycles, and various measures such as degree, clustering coefficient, mean path length, betweenness centrality, eigenvector centrality and decay exponents were calculated. In order to study the system in several time scales, we perform both a global, where the network contains information on the four solar cycles, and a local analysis, involving moving windows. Some metrics correlate with solar activity, while others do not. Interestingly, those metric which seem to respond to varying levels of solar activity in the global analysis, also do in the moving windows analysis. We study these correlations with different methods such as quantile-quantile plot, correlation matrix and student's t-test. Our results suggest that complex networks can provide a useful way to follow solar activity, and reveal new features on solar cycles.

Resumen

En esta tesis se estudia el Sol y su comportamiento mediante el uso de redes complejas. La red compleja se construyó utilizando el algoritmo Visibility Graph. Este método convierte las series temporales en grafos en los que cada elemento de la serie temporal se considera un nodo y se define un criterio de visibilidad para conectarlos. Con este método, construimos redes complejas para las series temporales de campos magnéticos y manchas solares que abarcan cuatro ciclos solares, y diversas medidas como el grado, el coeficiente de *clustering*, el camino medio, medidas de centralidad como *betweenness* y *eigenvector* y los exponentes de decaimiento. Para estudiar el sistema en varias escalas temporales, realizamos tanto un análisis global donde la red contiene información sobre los cuatro ciclos solares, y un análisis local, con ventanas móviles. Algunas métricas se correlacionan con la actividad solar, mientras que otras no. Curiosamente, las métricas que parecen responder a distintos niveles de actividad solar en el análisis global, también lo hacen en el análisis de ventanas móviles. Estudiamos estas correlaciones con distintos métodos, como el gráfico cuantil-cuantil, la matriz de correlación y el método estadístico *student's t test*. Nuestros resultados sugieren que las redes complejas pueden proporcionar una forma útil de seguir la actividad solar y revelar nuevas características de los ciclos solares.

Chapter 1

Introduction

Various measures of complexity can provide relevant ways to study the dynamics of magnetized plasma and, in particular, complex networks have been largely used to study a vast number of physical systems [1, 2], as their graph representation has been found to be helpful to characterize and model their phenomenology. Complementing these studies, mathematical tools from statistical physics have also proven to be particularly suitable for studying and understanding complex networks [3].

These works show that the underlying phenomenology in various systems can be inferred from their complex behavior, thus suggesting the great potential of complex networks to tackle problems in a variety of fields, such as economy [4–8], biology [9, 10], or in the study of geophysical problems such as earthquakes, magnetic storms or atmospheric flows [11–15], which prove the versatility of the method and its robustness.

The Sun is a particularly interesting system to study from the point of view of complexity. The interaction of particles and magnetic fields in the Sun's plasma, leads to a nonlinear dynamics which, in turns, leads to varying levels of solar activity, as manifested in the evolution of sunspots on the Sun's photosphere, velocity and turbulence levels of the solar wind, events such as solar flare or coronal mass ejections,

etc. Since the Sun is our closest star, it is essential to understand its behavior and the impact of solar activity on our planet, especially the impact of its magnetic activity and its effects on the Earth through the Earth's magnetic field and solar wind coupling [16, 17], which may lead to intense geomagnetic storms that may affect human communications and spacecrafts in periods of high solar activity [18]. In fact, the study of geomagnetic storms is quite relevant because they can have significant impact in a wide range of technological instruments such as damages and disruptions to satellites and communication systems [19, 20]. Their relation and occurrence with solar activity has been studied showing that descending phase of the solar cycle correlates with the following maximum phase [21] of the next solar cycle.

Various complexity analyses have been carried out to study this rich behavior. For instance, fractal and multifractal features have been identified in the Sun's photosphere, which have been shown to correlate with the evolution of solar activity [22], and have been proposed to be related to the emergence of solar flares [23–25].

Other works have focused on the chaotic and persistent features of the sunspots time series [26, 27]. Also, self-organized critical models have been proposed to represent the Sun's flare activity [28] and its power-law statistics [29].

In this work, we intend to follow a different approach, based on complex networks. Various recent works have carried out complex network analyses to study the Sun's activity, focusing on its major features: sunspots. For instance, in Ref. [30], the spatiotemporal patterns of sunspots are mapped into a complex network, showing that some topological measures of the network correlate with the solar cycle, while others anticorrelate, or remain essentially invariant. This is consistent with the fact that different measures inform about different features of the network topol-

ogy, so that some measures vary in response to the changes in sunspots number and location, whereas others point at complex properties which remain invariant along the solar cycle.

The previous work maps the spatiotemporal evolution of the sunspots distribution into a complex network. Nodes represent their location, and links represent their time sequence. However, various works have shown that valuable information about complex systems can be extracted by focusing on the time domain, by mapping time series into complex networks.

This was introduced by Lacasa et al. [31], and thanks to this and other works, it has been established that the resulting complex network has topological properties that reflect properties of the original time series [32, 33]. Thus, the Visibility Graph method (see details in Section 3.3) becomes an interesting tool, allowing, through the study of complex networks, to infer properties of the underlying dynamics. In the context of space physics, Suyal et al. [34] applied it to analyze the solar wind, a turbulent plasma whose origin is the upper atmosphere of the Sun and which leads to dynamic phenomena throughout the heliosphere on various temporal and spatial scales. In the following years, several authors have further explored the use of VG to various issues related to space and astrophysical physics, such as the analysis of reversibility in the turbulent states of solar wind simulations [35], the analysis of high-energy emission mechanisms of blazars [36], characterization of sunspot time series [37], statistical studies of solar flares [38], discrimination between types of variable pulsating stars [39], among others.

In particular, the work of Zou et al. [37] is interesting, since the VG analysis provides a complexity perspective to the analysis of the number of sunspots, which has been the traditional indicator of solar activity for centuries. There, the authors

perform a global analysis, constructing the VG from the complete time series of the number of sunspots, from the mid-nineteenth century to the first decade of the twentieth century.

Nonetheless, since the solar magnetic activity is not constant, which manifests itself, *e.g.*, in 11-year cycles [40], it is also relevant to study the complex properties of the sunspot configuration as a function of time. For example, it has been shown how the fractal dimension of the solar photosphere correlates with solar activity [22], and more recently, it has been studied how complex networks constructed from the spatio-temporal configuration of sunspots, also present various metrics that correlate or anti-correlate with the solar cycle [30].

Considering these results, we propose to carry out a VG study of the sunspot time series, using moving time windows to establish whether the complexity of this time series and its evolution provide information about variations in solar activity, complementing similar results based on fractal dimensions and complex networks [22, 30].

We also notice that both works just mentioned are based on image analysis of solar magnetograms. However, these images are actually a representation of the magnitude of the solar magnetic field, so, as a first approach to consider the physical information contained in the magnetic field itself, we will analyze, in this work, the time series of the average solar magnetic field.

Given a complex network, a large variety of measures could be calculated in order to characterize its topological structure. In Ref. [37], the VG analysis is focused on the degree distribution. However, other measures may provide additional insight, or may turn out to be less useful, depending on the specific system studied. For instance, Muñoz et al. [30] have shown that some metrics are correlated with so-

lar activity (degree centrality), or anti-correlated (eigenvector centrality), or remain constant throughout the variations in solar activity (clustering coefficient) [1]. This is a clear example that the complex network contains nontrivial information from a system, since metrics such as degree centrality are expected to be sensitive to variations due to their explicit dependence on the number of connections, while others, more elaborate metrics such as clustering coefficient, that quantify the grouping between neighbors, do not exhibit major variations throughout the cycle. All this suggests that the topology of the complex network contains nontrivial information about the physical state of the system, which is an important motivation for this work.

Thus, following Ref. [30], in this work, besides the degree distribution, we will focus on the clustering coefficient and various centrality measures (which are a measure of importance within the network), in order to examine the complex network from multiple perspectives. Furthermore, we will not only consider their average values, but also their distributions, by means of their respective critical exponents and Gini coefficients.

Thus, the interest and objective of this project lies in characterizing nonlinear dynamical processes (in this case, the evolution of solar activity) through the complexity parameters that the system itself can provide.

[1] These metrics and other are a measure of the properties of the network. They will be discussed later in Section 3

Chapter 2

Solar Activity

The Sun, being a magnetically active star, undergoes intense eruptions capable of influencing and distorting Earth's magnetosphere, leading to significant disruptions. These disturbances extend to both Earth and the technological facilities in orbit. Over the course of decades, continuous observation of the Sun has offered a remarkably detailed perspective on its structure and day–to–day changes in the life of a star. Recent high-resolution observations from both Earth and space have significantly enhanced our comprehension of the Sun's interior and atmosphere, providing deeper insights into its structure and evolution. All the behavior of the Sun is driven by its magnetic field, which induces the manifestations we observe, such as sunspots, solar flares, coronal mass ejections, among others. All these processes are part of what we know as solar activity.

2.1 Solar Magnetic Field

Getting a handle on what drives the magnetic system is crucial for understanding the nature of interplanetary space throughout the solar system. The Sun's magnetic field is responsible for everything from the solar explosions that cause space weather phenomena on Earth, such as auroras, to the interplanetary magnetic field

and radiation. Therefore, the Sun's magnetic field is fundamental to driving all the phenomena that define solar activity. It is generated by the motion of conductive plasma inside the Sun. When charged particles move, they naturally create magnetic fields, which in turn have an additional effect on how the particles move. The plasma in the Sun, therefore, sets up a complicated system of cause and effect in which plasma flows inside the Sun create the Sun's magnetic fields. This system is known as the solar dynamo. Within the Sun, it is widely agreed upon that a conductive fluid resides in the interior, moving amidst a turbulent environment. This fluid's motion generates electric currents, altering the original magnetic field and enhancing its strength. Consequently, this process becomes self-perpetuating, ensuring its continuity. This cyclical renewal of the Sun's extensive magnetic field forms the basis for all phenomena grouped under the term "solar activity".

We can observe the shape of the magnetic fields above the Sun's surface because they guide the motion of the plasma and we can measure its strength and direction using an instrument called magnetograph, which produces solar magnetograms. In figure 2.1 we can observe the Marshall Space Flight Center magnetograph from NASA, which has a gold plated H-alpha telescope mounted on the side of the magnetograph telescope. This H-alpha (Hydrogen-alpha) filter is a combination of several filters that are designed to provide additional safety to the viewer and remove all out of band transmission from the etalon or Fabry-Pérot interferometer [41] (two flat and parallel optical surfaces separated by a gap of air or solid material), allowing only the transmission at 656.28 nm to pass. The magnetograph works by measuring the polarization of light at various wavelength positions within a solar spectral line. Circular polarization in the opposite sense on either side of a magnetically sensitive spectral line gives a measure of the longitudinal magnetic field (the strength of the

Figure 2.1: Photograph of MSFC (Marshall Space Flight Center) magnetograph. James Smith (retired), former chief observer, is shown besides the instrument. Photograph taken from NASA repository https://magnetograph.msfc.nasa.gov

field directed toward and away from the instrument). Linear polarization provides information on the strength and direction of the magnetic field transverse to the line of sight.

A solar magnetogram is an image that shows the strength, polarity, and location of the magnetic fields on the Sun. In a grayscale magnetogram, regions of strong magnetic field (active regions) are shown as white (positive polarity, magnetic field lines coming towards the observer) and black (negative polarity, magnetic field lines pointing away from the observer) areas, while gray areas indicate regions of weak magnetic fields. A complete understanding of the Sun's magnetic field, including knowing exactly how it is generated and its structure deep inside the star is not yet available. What we know is that the solar magnetic system is known to drive the approximately 11-year activity cycle on the Sun. The Sun's magnetic field grows

more complicated over time until it peaks at solar maximum, some 11 years after the previous solar maximum (the same interval applies for solar minimum). From that moment and with every eruption, the Sun's magnetic field smooths out slightly until it reaches its simplest state. At that point, the Sun experiences what is known as solar minimum, when solar explosions are least frequent. Then, the process is perpetuated over time, considering that the energy released during eruptions is not enough to mitigate the increase in complexity in the magnetic field until the new solar maximum is reached.

2.2 Sunspots

Sunspots are phenomena of the Sun's photosphere (the visible surface of the Sun) that appear as temporary spots that are darker than the surroundings. They are regions of reduced surface temperature caused by concentrations of magnetic flux that inhibit convection. In other words, these magnetic fields keep some of the heat within the Sun from reaching the surface. Sunspots appear within active regions, usually in pairs of opposite magnetic polarity. Their number varies according to the approximately 11-year solar cycle. Sunspots appear in a wide variety of shapes and forms. The darkest area of a sunspot (also the first to be observed) is called the umbra. As the sunspot matures (becomes more intense), a less dark, outlying area of well-defined fibril-like structure develops around the umbra, called penumbra (see Figure 2.2).

In the umbra, the large magnetic field and its almost vertical inclination inhibits the convective energy transport, lowering its temperature to $T = 3500$–5000 K and making it appear darker than the quiet photosphere, where $T_{\text{pho}} = 6000$–6500 K. In the umbra, the average magnetic field is orientated vertically with respect to

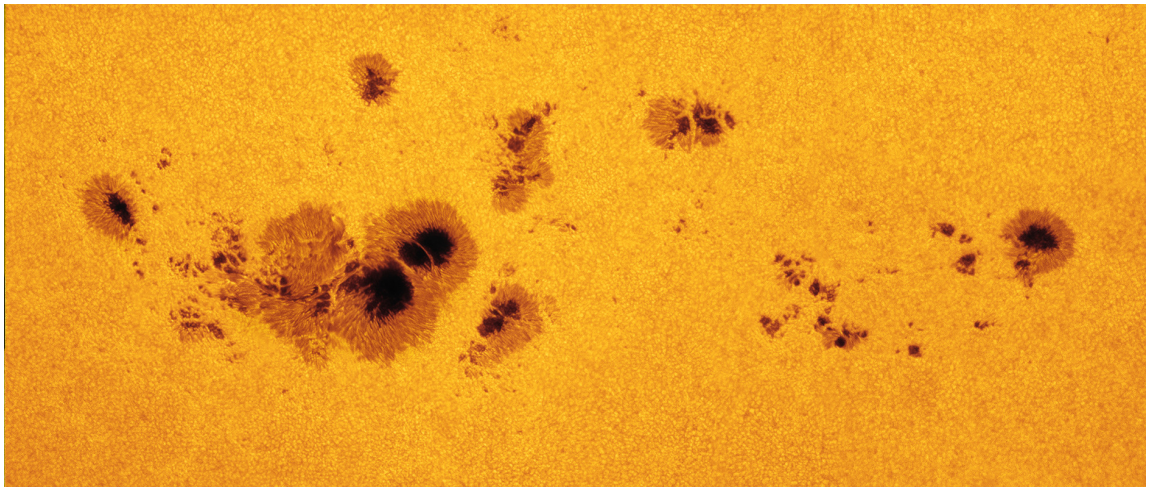


Figure 2.2: Image of a sunspot where the dark region corresponds to the umbra and the lighter, more diffuse area corresponds to the penumbra.

the solar surface, becoming slightly inclined as we approach the umbral–penumbral boundary. The magnetic field strength ranges from $B = 2000$–3500 G, achieving the largest values for big sunspots. Except for some oscillatory phenomena, the umbra is commonly found to be at rest. If many phenomena observed in the umbra remain unknown, the penumbra overtakes its darker brother by a large amount. Even the question of why sunspots have a penumbra and how does it form remains unanswered.

Sunspots can grow from an individual unipolar spot into more organized bipolar spot groups; or even evolve into immense, very complex sunspot groups with mixed magnetic polarities throughout the group, as we can see in the magnetogram shown in Figure 2.3. The largest sunspot groups can cover large swaths of the Sun's surface and be many times the size of Earth. Individual sunspots or groups of sunspots may last anywhere from a few days to a few months, but eventually decay. Sunspots expand and contract as they move across the surface of the Sun, with diameters ranging from 16 km to 160 000 km. Larger sunspots can be visible from Earth without the aid of a telescope. They may travel at relative speeds, or proper motions, of a

few hundred meters per second when they first emerge. Indicating intense magnetic activity, sunspots accompany other active region phenomena such as coronal loops, prominences, and reconnection events. Most solar flares and coronal mass ejections originate in these magnetically active regions around visible sunspot groupings. A new solar cycle is considered to have begun when sunspot groups emerge at higher latitudes with the magnetic polarities of the leading spots opposite that of the previous cycle.

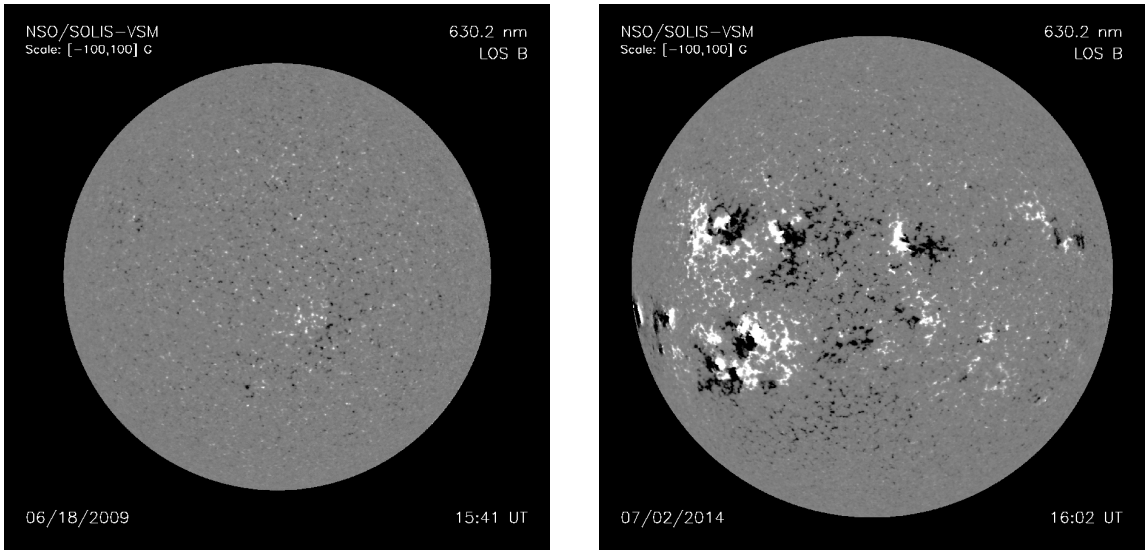


Figure 2.3: Examples of magnetograms taken during solar minimum (left) and solar maximum (right). The complexity of the magnetic morphology during periods of high level of solar activity during solar maximum is clear. Images taken from the National Solar Observatory https://nso.edu/data/nisp-data/magnetograms/

There is a model that postulates various characteristics of sunspots and their active regions stem from the emergence of sizable magnetic flux tubes across the solar surface [42]. As a result of the Sun's convection, plasma rises, allowing solar magnetic field lines to surface, curve, and re-enter the solar photosphere. This model suggests that near the solar surface, these flux tubes fragment into smaller ones, with

the larger segments forming visible sunspots. Essentially, the observable sunspot represents the point where this flux tube intersects the solar surface (see Figure 2.4). Within an electrically conductive atmosphere, a horizontal magnetic flux tube floats and naturally ascends. This magnetic buoyancy is potent enough to occasionally draw flux from the broader solar toroidal field to the photosphere [43].

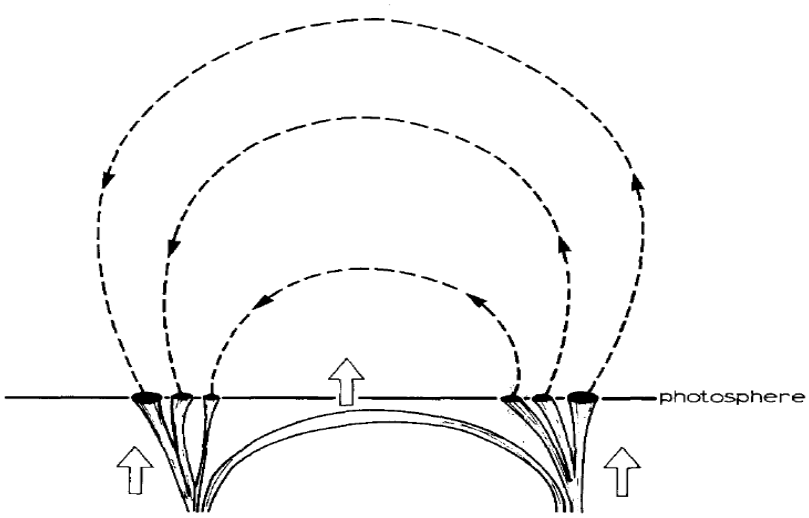


Figure 2.4: Emergence of flux bundle and coalescence of spots. For simplicity a small number of tubes constituting the bundle was chosen [44].

Chapter 3

Method

To study solar activity, we use time series data of the number of sunspots observed on the solar surface [45], and Sun's global magnetic field [46]. Both series will be considered between 1975 and 2015 with a one-day resolution, comprising Solar Cycles 21, 22, and 23, and the beginning of Solar Cycle 24. Data are shown in Figure 4.1. Three solar cycles are chosen in order to have a relevant sample of solar activity in the last years. Since the Wilcox Solar Observatory (WSO) project started collecting data in 1975, while the sunspot data date back to 1818, we chose to collect the characteristic parameters of solar activity since 1975. We set as day zero the measurements of 16 May 1975, ending in day 14,185, corresponding to 16 December 2015.

Using these data, the method consists of constructing complex networks from the time series, using two variations of the visibility algorithms as connection criteria, aiming to extract statistical properties of the system. Thus, for each time series, two complex networks will be constructed, so that we will have a total of four networks.

3.1 Complex Networks

Complex networks have been studied extensively due to their relevance to many real-world systems such as the worldwide web, the internet, energy landscapes and biological and social systems [47]. They have become, in recent decades, a very popular strategy to study complex systems, since they allow to abstract in a very general way systems composed of a large number of components somehow interrelated. Then, we will present some basic concepts about complex networks, which will be useful in the development of this work.

A complex network is a graph consisting in a set of nodes and their connections with non-trivial statistical and topological properties [48]. By representing an abstraction of some system, the definition of nodes and connections must consider both the properties of that system and the type of study to be performed. Therefore, different complex networks can be obtained from the same system, depending on the construction method and what is the question of interest about that system. If the connections of a network have an assigned direction, we are in the presence of a *directed network*. We can find examples of this type of networks in works that considered connections in the temporal direction such as earthquakes [49], sunspots [30] and non-thermal plasmas [35]. If the connections have no privileged direction, then we have an *undirected network*. In this work, we will consider this type of connections. The usual way to picture a graph is by drawing a dot for each node and joining two dots by a line if the two corresponding nodes are connected by a link. How these dots and lines are drawn is irrelevant, and the only thing that matters is which pair of nodes form a link and which ones do not. Examples of directed and undirected graphs are shown in Figure 3.1. As the size of a network increases,

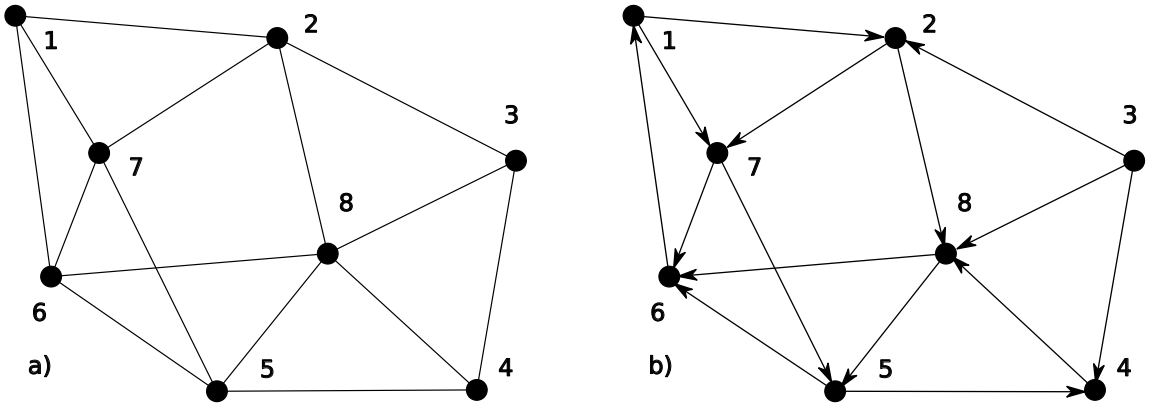


Figure 3.1: Graphical representation of an undirected (a) and directed (b) graph. In this example we consider $N = 8$ nodes and $K = 15$ connections or links. In the directed graph, the arrows indicate the direction of each connection.

it becomes impossible to analyze just visually, and the development of statistical measures of connectivity becomes necessary [50].

A central concept in graph theory is that of reachability of two different nodes of a graph. In fact, nodes that are not adjacent can still be linked through a reachable path. A *walk* from node i to j is an alternating sequence of nodes and edges (a sequence of adjacent nodes) that begins in i and ends in j. A *path* is a walk in which no node is visited more than once. The length of the walk is defined as the number of edges in the sequence. The walk of minimal length between two nodes is known as *shortest path* or *geodesic*. A graph is said to be connected if for every pair of nodes (i, j), there is a path from i to j. Otherwise, it is said to be disconnected. It is often useful to consider the matrix representation of a graph given by the adjacency matrix $\mathcal{A}$. This matrix is an $N \times N$ square matrix whose element a_{ij} is equal to 1 when there is a connection between i and j, and zero otherwise. The diagonal of $\mathcal{A}$ contains zeros. This is thus a symmetric matrix for undirected graphs. The adjacency matrix for the network shown in Figure 3.1 (the undirected case) can be

written as shown in equation (3.1). We must note that this is the adjacency matrix for the node labeling that we have chosen, which is arbitrary. Therefore, depending on the labeling chosen for the nodes, there will be a corresponding adjacency matrix, which is a permutation of rows and columns of other possible labelings, without affecting the topology of the network.

$$\mathcal{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} . \tag{3.1}$$

We build the networks using the Visibility Graph method (described in detail in Section 3.3) on the magnetic field and sunspots time series. After the networks are built, various measures must be calculated in order to study their possible correlation with solar activity along the 3 solar cycles previously mentioned. In particular, we considered node degree as a measure of connections, clustering coefficient as a measure of groupability among nodes, mean path length between nodes as an indicator of network performance in terms of connectivity, and two centrality measures, namely, betweenness and eigenvector centrality, as a measure of importance within the network.

3.2 Metrics

As mentioned in Section 3.1, we calculate several metrics to characterize the networks, which are quantities containing information about the network structure and dynamics. Usually, one single metric is not enough to characterize the network, so several of them are necessary to describe it adequately. Here we define the metrics that we will use in the rest of the thesis, which are various ways to describe the connectivity within the network. The node degree is the number of connections that a node has [51]. We will use a normalized degree, to make it independent of the network size. If ν is an arbitrary node, and it has n_ν connections then its normalized degree is

$$g(\nu) = \frac{n_\nu}{n-1}\,, \tag{3.2}$$

where n is the total number of nodes in the network and $n-1$ denotes the maximum number of possible connections for a node ν.

Additionally, the degree or connectivity $g(\nu)$ of a node ν in terms of the adjacency matrix $\mathcal{A}$ is the number of edges incident with the node, and is defined as

$$g(\nu) = \sum_{j \in \mathcal{N}} a_{\nu j}\,. \tag{3.3}$$

For directed graphs, the degree of a node has components, the number of outgoing links $g_{\text{out}}(\nu) = \sum_j a_{\nu j}$ and the number of ingoing links $g_{\text{in}}(\nu) = \sum_j a_{j\nu}$. The total degree is then defined as $g(\nu) = g_{\text{out}}(\nu) + g_{\text{in}}(\nu)$.

The clustering coefficient of a node ν is the fraction of possible triangles which

contain that node, and is defined as

$$c(\nu) = \frac{2T(\nu)}{g(\nu)(g(\nu)-1)}\,, \tag{3.4}$$

where $T(\nu)$ is the number of triangles containing node ν. This metric quantifies how clustered the nodes are within the network. It shows the tendency for two neighboring nodes of a node ν to be connected to each other. We can think of this metric as the probability that pairs of neighboring nodes of a node ν are connected. If this node ν has $g(\nu)$ neighbors, then there are at most $g(\nu)(g(\nu)-1)/2$ connections among them, which occurs when every neighbor of ν is connected to all of its neighbors. Then, $c(\nu)$ is the fraction of these allowed connections that actually occur [52].

The clustering coefficient of the graph is then given by the average $c(\nu)$ over all the nodes

$$C = \langle c \rangle = \sum_{i \in V} c(\nu)\,. \tag{3.5}$$

The mean path length corresponds to the average number of steps along the shortest paths for all possible pairs of network nodes. It then may be considered as a measure of the efficiency of information or mass transport on a network. It is defined as

$$l = \sum_{s,t \in V} \frac{d(s,t)}{n(n-1)}\,, \tag{3.6}$$

where V is the set of nodes and $d(s,t)$ is the minimum distance from node s (*source* node) to node t (*target* node).

The centrality metrics measure the relevance of a node within the network. Betweenness centrality is the sum of the fraction of all-pairs shortest paths that pass

through a node ν [53–56],

$$b(\nu) = \sum_{s,t \in V} \frac{\sigma(s,t|\nu)}{\sigma(s,t)}\,, \tag{3.7}$$

where V is the set of nodes, $\sigma(s,t|\nu)$ is the number of paths passing through some node ν other than s,t and $\sigma(s,t)$ is the number of shortest paths between s and t. When $s = t$, then $\sigma(s,t) = 1$ and if $\nu = s$ or $\nu = t$, then $\sigma(s,t|\nu) = 0$ [57]. This metric quantifies the importance of a node within the network in the sense of a bridge of information, which connects other nodes. This metric is one of the standard measures of node centrality, originally introduced to quantify the importance of an individual in a social network [53, 54, 58].

If a node of high betweenness centrality is removed, it should have a large effect on the connectivity between any two nodes of the network, and thus betweenness centrality is expected to have some relationship with node degree. In fact, betweenness-degree correlations have been studied in [59–61]. We will further discuss this correlation in Chapter 5.

Eigenvector centrality computes the centrality for a node based on the centrality of its neighbors. The intuition is that a node is important if it is connected to important nodes. Following this argument, let x_i be the centrality of node i and we assume x_i is proportional to the sum of the centralities of its neighbors (x_j for $j \neq i$). Then we have the equation

$$x_i = \frac{1}{\lambda} \sum_j a_{ij} x_j\,, \tag{3.8}$$

where a_{ij} is an element of the adjacency matrix that indicates the connection (or not) between node i and j, and λ is a proportionality constant. The above-mentioned

equation can be expressed in matrix form, obtaining

$$\mathcal{A}\mathbf{x} = \lambda\mathbf{x}\,, \tag{3.9}$$

which is an eigenvalue equation for $\mathbf{x}$.

Then, the problem of computing node importance is equivalent to solving the eigenvalue problem for the adjacency matrix. In principle, there can be many solutions for this problem, but when computing eigenvector centrality, the eigenvector associated with the largest eigenvalue is used because it represents the most dominant and stable influence within the network.

Finally, we will calculate the degree probability distribution (degree distribution) of a node having degree k, $P(k)$, which corresponds to the fraction of nodes with k connections n_k over the total amount of nodes n, that is

$$P(k) = \frac{n_k}{n}\,. \tag{3.10}$$

The probability distribution satisfies that

$$\sum_{k=1}^{n} P(k) = 1\,. \tag{3.11}$$

We consider these abovementioned metrics to analyze information about underlying processes (given by the degree distribution) of the solar time series, to study the behavior of the solar cycles (by clustering coefficient), and to analyze the dynamics in the resulting networks (by betweenness and eigenvector centrality).

To clarify these concepts, we consider a network example, as shown in Figure 3.2. This netowrk consists of seven nodes and eleven connections. We will calculate the metrics abovementioned for this network and we will show the results in Figures 3.3-3.6.

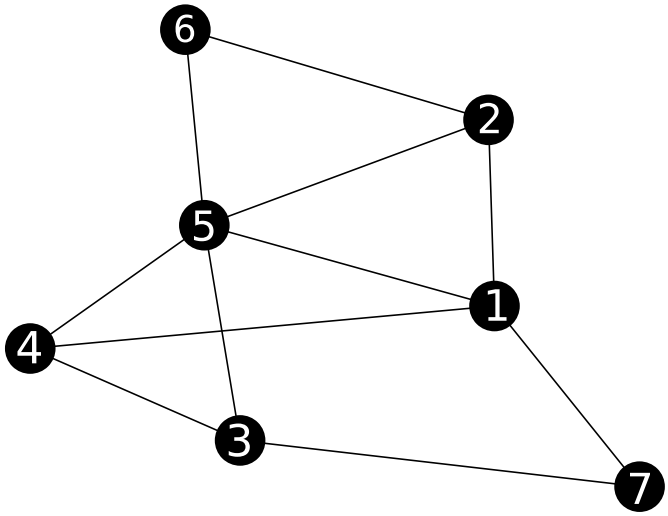


Figure 3.2: Example of an illustrative network with 7 nodes and 11 connections.

The colorbar indicates values of the corresponding metric, from darker (small values of the metric) to lighter colors (large values of the metric). As we can see in Figure 3.3, each node has an assigned color depending on the value of its metric, in this case, its degree centrality. We can clearly see in this case that node 5 is the node with the maximum value of degree because it has more connections than the others. In figure 3.4, we can see that node 5 is the node with the largest value of betweenness, because it connects more times other pair of nodes in the network (acts as bridge). For instance, connects these pair of nodes $(2, 4)$, $(3, 6)$ and $(3, 2)$. In figure 3.5, we can see that node 6 has the largest value of clustering in the network (and also the maximum possible value, 1), in this case it is not because it is grouped with many

other nodes, but instead the connections it has are well-grouped, *i. e.*, it is connected to two other nodes and forms only triangle, matching the necessary criteria to be well-grouped, according to the definition. Finally, in Figure 3.6 we observe once again that node 5 is the one with the largest value of eigenvector centrality, and we can graphically see this by noticing it is connected to other important nodes in the network.

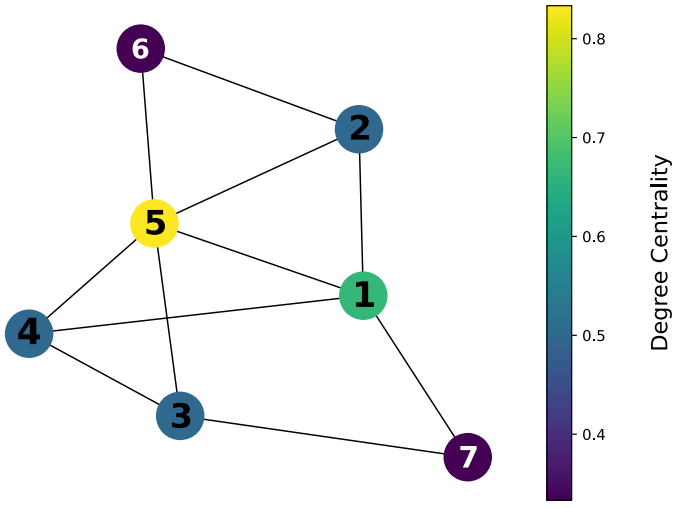


Figure 3.3: Degree centrality for the network shown in Figure 3.2, where the colorbar indicates values of this metric.

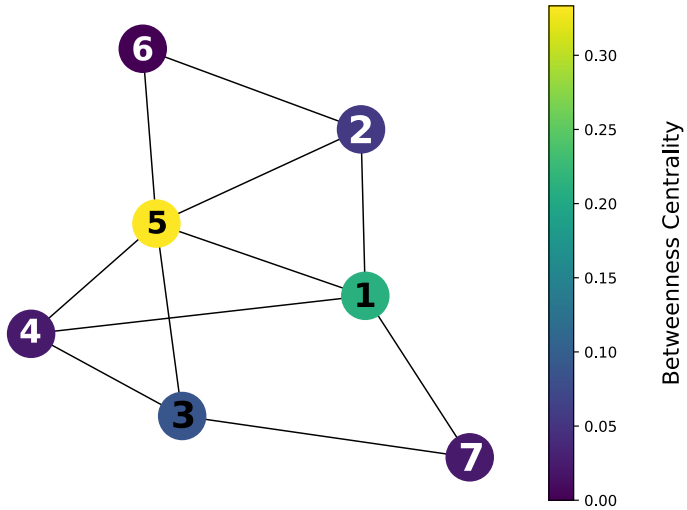


Figure 3.4: Betweenness centrality for the network shown in Figure 3.2, where the colorbar indicates values of this metric.

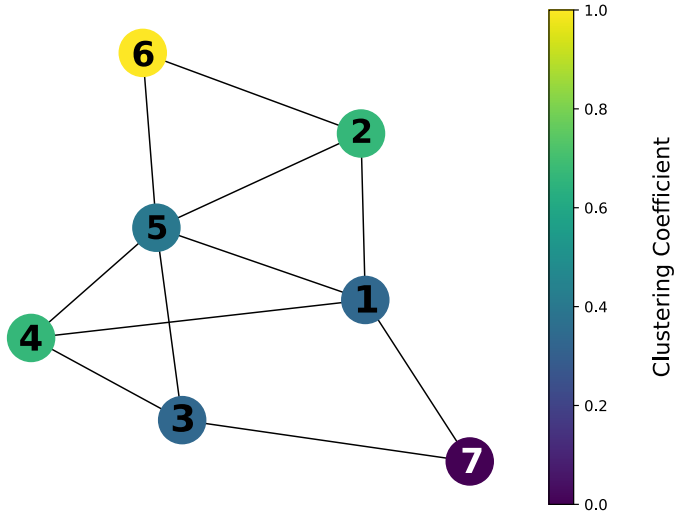


Figure 3.5: Clustering coefficient for the network shown in Figure 3.2, where the colorbar indicates values of this metric.

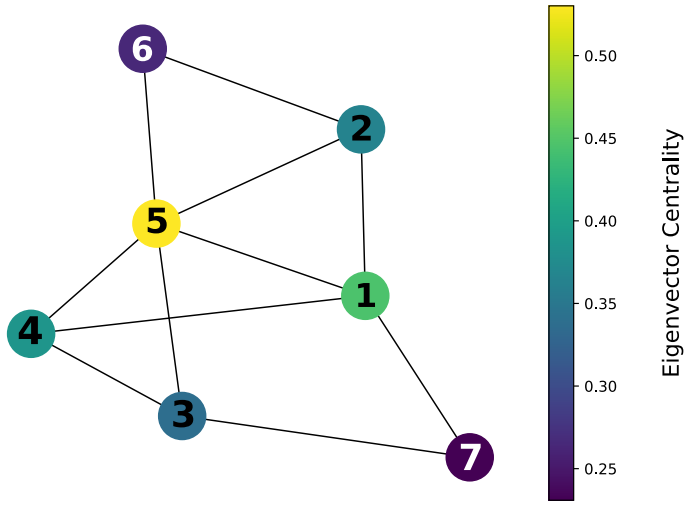


Figure 3.6: Eigenvector centrality for the network shown in Figure 3.2, where the colorbar indicates values of this metric.

3.3 Connection Criteria

Visibility Graph

We use the Visibility Graph (VG) algorithm (or also known as Natural Visibility Graph), whose statistical properties have been studied in several publications [62, 63]. The definition of VG for time series comes from the concept of visibility between nodes. Each element of the time series can be identified by a time t and its respective associated value $x(t)$, which represents some physical quantity. Therefore, a node in the network is defined by the point $(t, x(t))$. Two nodes are connected if they "see" each other, *i.e.* if there is a straight line connecting them without being interrupted by other intermediate nodes. Formally, given a data series X_N, two arbitrary nodes x_a and x_b are connected if, for every node x_c between them, then [63, 64]

$$x_c \leq x_b + (x_a - x_b)\frac{t_b - t_c}{t_b - t_a}\,. \tag{3.12}$$

Therefore, the edges of the network take into account the temporal information explicitly. **The key question is to know whether the associated graph inherits some structure of the time series, and consequently whether the process that generated the time series may be characterized by using graph theory**. Most other methods for constructing complex networks from time series data are dependent on the choice of some parameters, *e.g.*, the threshold ϵ of recurrence networks [65]. In contrast, the visibility graph is not affected by these elections since its construction relies purely on the geometric or topological relationship between the points, specifically on the visibility criterion. For the current discussion, we prefer the simplicity of the visibility graph. The visibility graph method becomes especially intriguing in the case of specific stochastic processes in which the statistical characteristics of the resultant network can be directly correlated with the fractal characteristics of the time series [63, 66–69]. A graphical example of this algorithm can be seen in Figure 3.7.

Horizontal Visibility Graph

A variant of VG known as Horizontal Visibility Graph (HVG) consists of restricting the visibility between nodes to a horizontal line (a simplification of the visibility algorithm). If $\{X_i\}_{\{i=1,2,\ldots,N\}}$ is a time series of size N, two nodes i and j will be visible if for all nodes n such that $i < n < j$, then [62, 70, 71]

$$X_i\,,\, X_j > X_n\,. \tag{3.13}$$

Therefore, horizontal visibility for two nodes occurs if there is no other node greater in magnitude between them. We must note that the geometric criterion

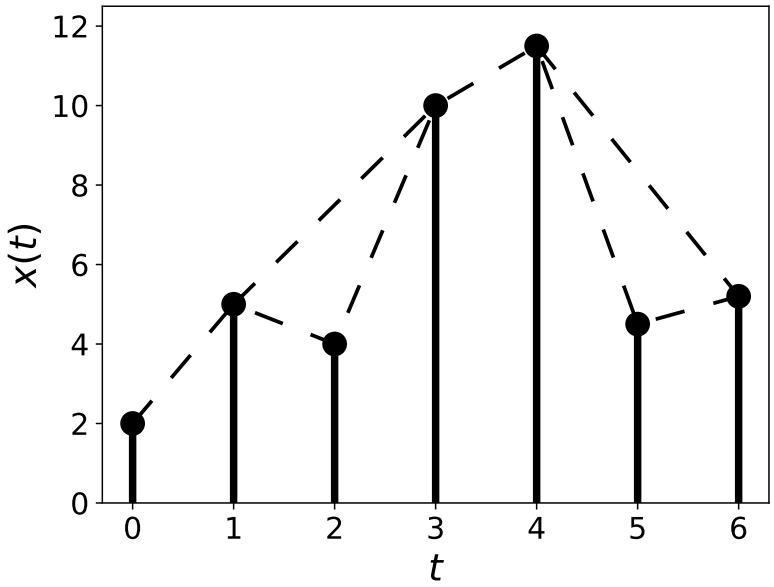


Figure 3.7: Example of a time series. The visibility rays between the data define the links connecting nodes in the graph.

defined for the HVG is more *visibility restrictive* than its analogous for the general case. That is to say, the nodes within the HVG will have *less visibility* than their counterparts within the VG. As a matter of fact, notice that given a time series, its corresponding horizontal visibility graph is always a subgraph of its associated visibility graph. Accordingly, as in the former case, the HVG is always connected, each node sees at least its nearest neighbors (adjacent nodes). This method is well fitted to distinguish different degrees of chaos from a sequence of uncorrelated random variables. A graphical example of this method can be seen in Figure 3.8.

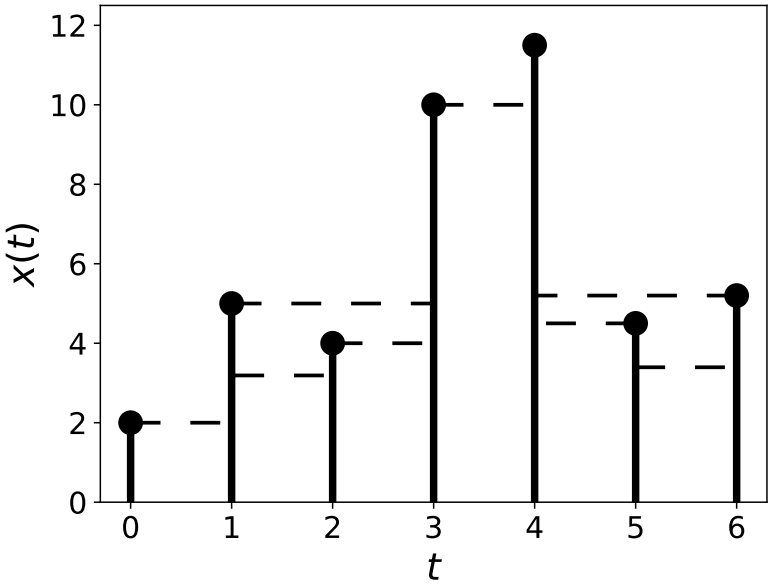


Figure 3.8: Example of a time series. The horizontal visibility rays between the data define the links connecting nodes in the graph.

Chapter 4

Solar activity results

Construction of the complex network from the time series involves not only the decision on what will be regarded as a node, and what will be the criterion to connect two nodes, but also the length of the time window within which data will be considered. Figure 4.1 shows that solar activity has variations on various timescales. Thus, in order to obtain a better perspective of the solar activity, we carry out two analyses: a global analysis, considering the complete time series, and a local one, using time windows.

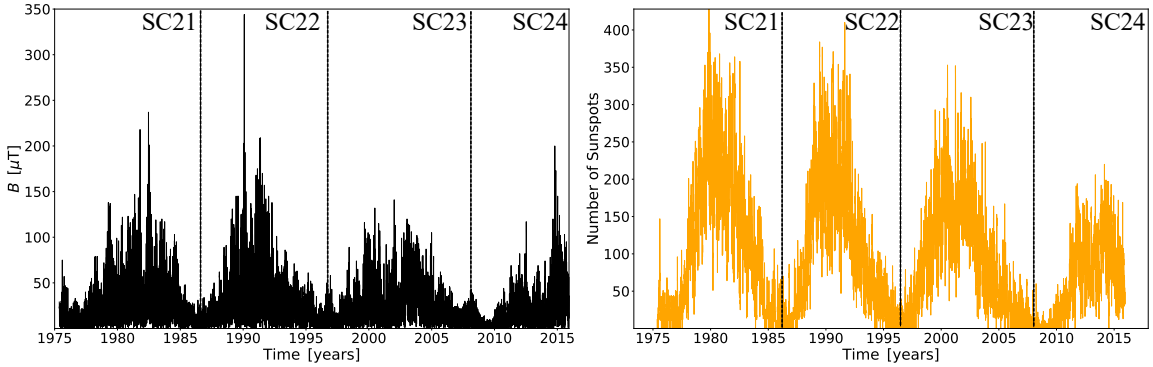


Figure 4.1: Time series used in this work. The solar cycles are indicated in the figure (from solar cycle 21 to 24) and separated with vertical dashed lines. **Left** panel: mean magnetic field on the surface of the Sun. **Right** panel: number of sunspots.

4.1 Global Analysis

We first consider the global analysis. For this case, we take time series of sunspots and magnetic field from 1975 to 2015, comprising Solar Cycles 21, 22, 23 and the beginning of Solar Cycle 24 (see Figure 4.1). We then build a complex network for each time series using two connection criteria, visibility graph and horizontal visibility graph. For these networks we compute the metrics described in Section 3.

The results are shown in two consecutive panels, where the left panel represents the results for the networks built with the visibility graph method, while the right panel represents the results for the networks built with the horizontal visibility graph. In order to distinguish the time series, we use purple for magnetic field and green for sunspots. Figure 4.2 shows the resulting degree for each node, normalized to the network size. Thus, the ordinate axis represents the fraction of nodes that each node is connected to.

We first observe that the number of connections is larger for the networks built by the VG method (note the different scales in the vertical axes), which is expected, since the HVG method restricts visibility to a horizontal line, and therefore, less connections can be established. We also notice that there is no particular dependence of this metric with the solar cycles. There are some prominent values for the VG graph (time $t \sim 5500$), which match high values of the magnetic field (ascending phase of Solar Cycle 22, Figure 4.1), but as a general rule, no correlation is observed. Given the definition of the VG, one would expect that maxima in the time series would "see" more data in the rest of the time series, as they would tend to be unobstructed by intermediate points, thus leading to maximum degree. However, except for the very large maximum noticed above, this does not hold in general, due

Table 4.1: Results for average clustering for both networks, subscripts ss and mf refers to sunspots and magnetic field, respectively.

	VG	HVG
C_{ss}	0.721	0.448
C_{mf}	0.739	0.458

to the rapid fluctuations. This result suggests that the degree is too simple a metric to study these time series, thus justifying the need for more elaborate metrics.

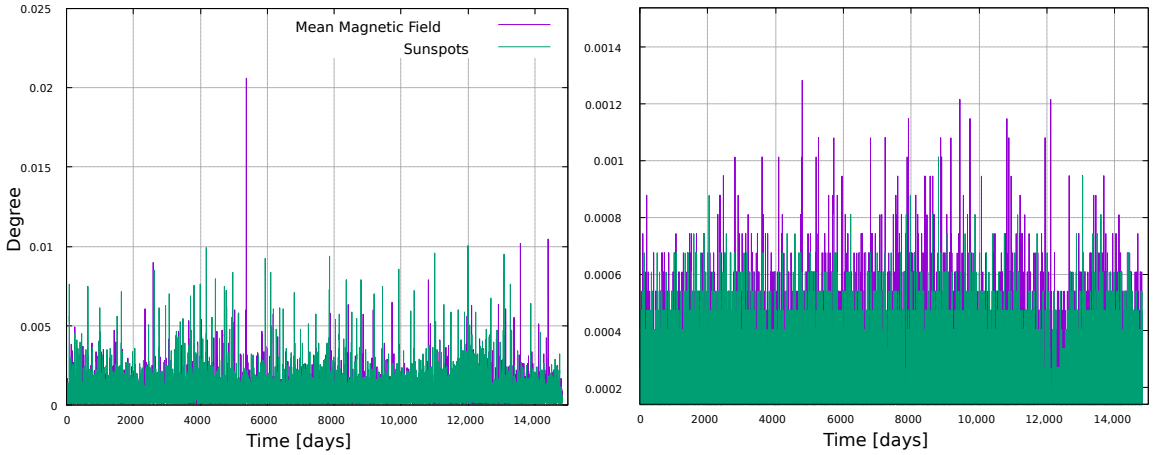


Figure 4.2: Degree for every node in the network, normalized by the size of the network, for the sunspots time series (green line) and the mean magnetic field time series (purple line). **Left** panel: VG; **right** panel: HVG.

No particular dependence on the solar cycle is observed for the clustering coefficient and the shortest path length per node, which is why we have not shown the corresponding plots. For the clustering coefficient one obtains average values for both VG and HVG, which are shown in Table 4.1.

These results indicate the presence of well-defined clusters within the time series, corresponding to solar cycles, for the VG networks, based on a large value for the clustering for both time series (≈ 0.7). The solar cycle shows that most of the tem-

poral points of the decreasing phase of one cycle are connected to those points of the increasing phase of the next solar cycle (see Figure 4.1). Therefore, the network is clustered into communities, each of which mainly consists of the temporal information of two subsequent solar cycles. When the sunspot number reaches a stronger but more infrequent extreme maximum, we have inter-community connections, since they have better visibility contact with more neighbors than other time points, forming hubs in the graph. The inter-community connections extend over several consecutive solar cycles encompassing the temporal cycle-to-cycle information.

In the case of VG networks, the results obtained for both mean magnetic field and sunspots mean paths are $l_{mf} \approx 6.36$ and $l_{ss} \approx 5.73$, respectively. Both values are much smaller than the network size ($N = 14\,185$), indicating that, although the network is large, nodes are close to one another on average, separated by at most 6 nodes.

The short distance between nodes notices above can be due, for instance, to a large number of connections between nodes, or to the presence of some highly important nodes, acting as bridges that connect different parts of the network. This can be quantified with the concept of betweenness centrality (BC) which, as other centrality measures, provides a way to assess the importance of nodes in the network. In this case, how important a node is to establish connection between nodes (see Equation (3.7)).

The results obtained for this metric can be seen in Figure 4.3. For the VG network (left panel), we observe three zones where a few nodes have BC values much larger than other nodes in the network. For the HVG case (right panel), these zones can also be seen, with better statistics, as more nodes have large BC values. These results hold for both time series (sunspots and mean magnetic field). The most interesting

feature of these results is the clear match between BC and variations in solar activity (Figure 4.1). As mentioned when Figure 4.2 was discussed, one would expect the highest points in the time series to be very well connected, as they should be able to "see" more nodes. However, this is not captured by the number of connections itself, as shown in Figure 4.2. This is unlike BC, where Figure 4.3 shows higher values of sunspots and mean magnetic field do not have, on average, more connections that the rest of the data, but they do play an important role in connecting nodes. In this sense, it is also interesting to note that the important nodes for the VG method are very few, where the nodes with large values of the BC belong to a narrow zone around solar maxima; whereas for the HVG, the BC has a wider distribution, following the sunspots and mean magnetic field time series in a smoother way. This behavior can be explained with the visibility criteria used by the two methods, in which for VG is more "relaxed" than the HVG, and therefore, the former method tends to create denser networks where nodes have more connections. This means that there are multiple alternative paths between nodes, whereas for HVG there will be fewer alternative paths, increasing their betwenness centrality.

We have computed a further centrality measure, namely the Eigenvector Centrality (EC). The results are shown in Figure 4.4.

Several features are interesting to observe, which differ from the previous plots. First, this metric clearly exhibits different results for each time series, as maximum values for the sunspots time series do not occur for the same nodes as for the magnetic field time series. This highlights the nontriviality and nonlinearity of the metrics, and supports the fact that it is interesting to calculate several metrics for a given network, as they may reveal different features. This is specially noticeable as both time series in Figure 4.1 show a similar behavior: sequences of maxima and minima which

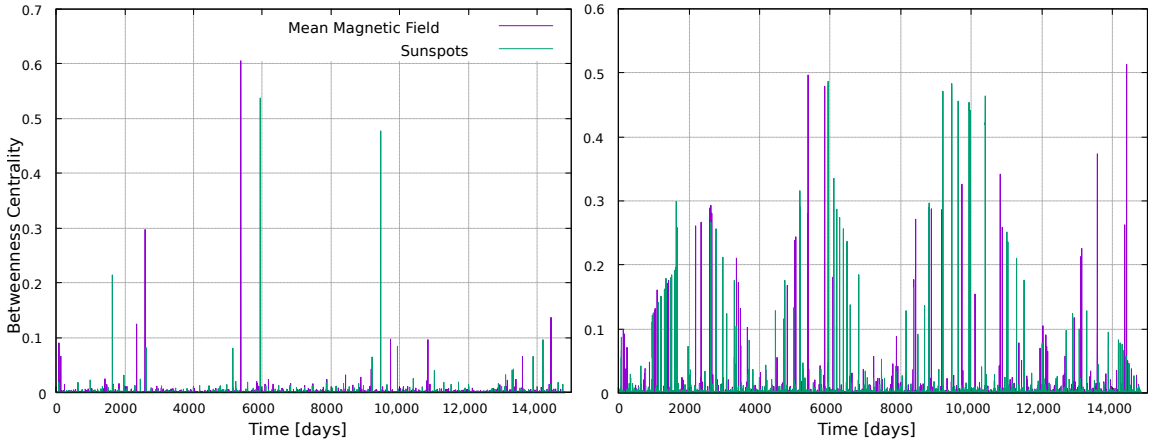


Figure 4.3: Betweenness Centrality for every node in the network, normalized by the size of the network. **Left** panel: VG method; **right** panel: HVG method.

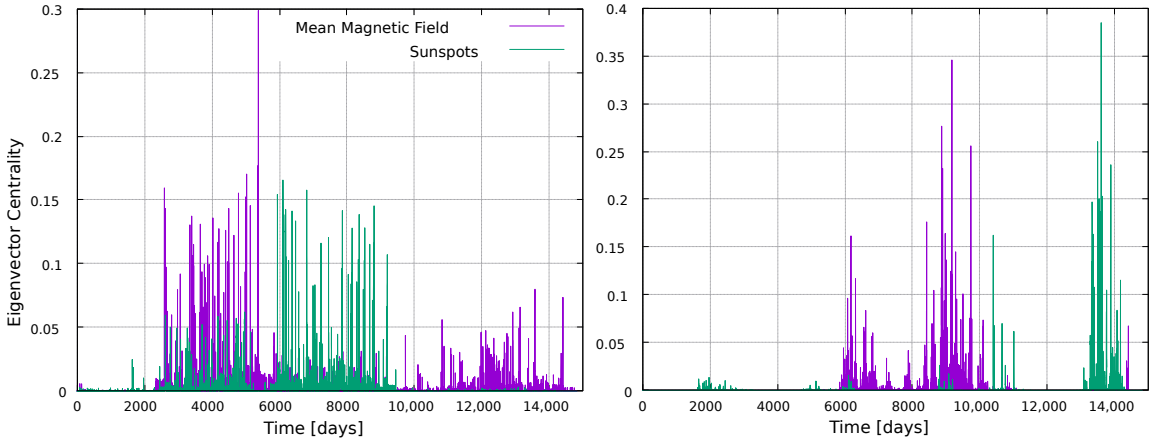


Figure 4.4: Eigenvector Centrality for every node in the network, normalized by the size of the network. **Left** panel: VG method; **right** panel: HVG method.

clearly mark all the solar cycles in the dataset, at essentially the same time. However, there are two different datasets being analyzed here, and this is clearly evident when considering to the eigenvalue centrality, not the metrics previously discussed.

Another interesting fact is that, in the VG case, maximum values of the eigenvector centrality tend to occur in between solar maxima, suggesting an anticorrelation

with the solar cycle. Notice, for instance, the EC maxima for the magnetic field time series, between the 21st and 22nd solar cycle, and the maxima for the sunspots time series, between the 22nd and 23rd solar cycle. However, there are three intercycle time windows in the data set, but only two noticeable maxima of the EC for both the magnetic field and the sunspots time series. Since the EC is related to the importance of neighboring nodes, it is possible that the analysis is affected by boundary effects, as no data exist before and after the selected time window. However, it is worth noticing that several papers have been devoted to the prediction of features of the next solar cycle [72–74], such as its intensity. Since the EC for the VG seems to be most sensitive during the intercycle period, with different behaviors for each time series (*e.g.*, the existence of maxima for the magnetic field series at day $\sim$13 000, while no important maxima occur for the sunspots series) it would be interesting to explore to what extent the EC could provide useful information on the next solar maximum before it is actually reached.

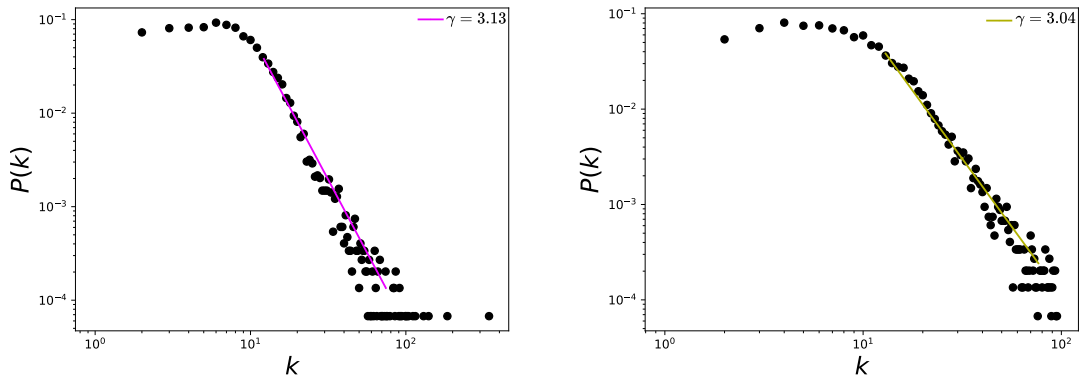
As for the HVG method, Figure 4.4 also shows that EC maxima do not occur simultaneously with sunspots maxima. Rather, they seem to cluster during the ascending or descending phase of cycles.

The analysis so far has focused on the value of network measures per node. However, the distribution of values $P(k)$ may also have information, as it can provide insight about the physical processes underlying the network formation [48, 75].

For instance, the VG method typically leads to power-law distributions, $P(k) \sim k^{-\gamma}$, and the HVG method typically leads to exponential degree distributions, $P(k) \sim \exp(-\gamma k)$, and it has been suggested that its decay exponent γ is related to the type of randomness [76]. Specifically, it has been suggested that a threshold value $\gamma_{\text{th}} = \ln(3/2) \approx 0.405$ exists, such that $\gamma < \gamma_{\text{th}}$ corresponds to a

chaotic process, whereas $\gamma > \gamma_{\text{th}}$ corresponds to a correlated stochastic process.

In figure 4.5 we can observe the expected power-law behavior for the networks built using the VG method. We calculate its respective γ by computing the slope of the linear fit of the log-log distribution at the tail.



(a) (b)

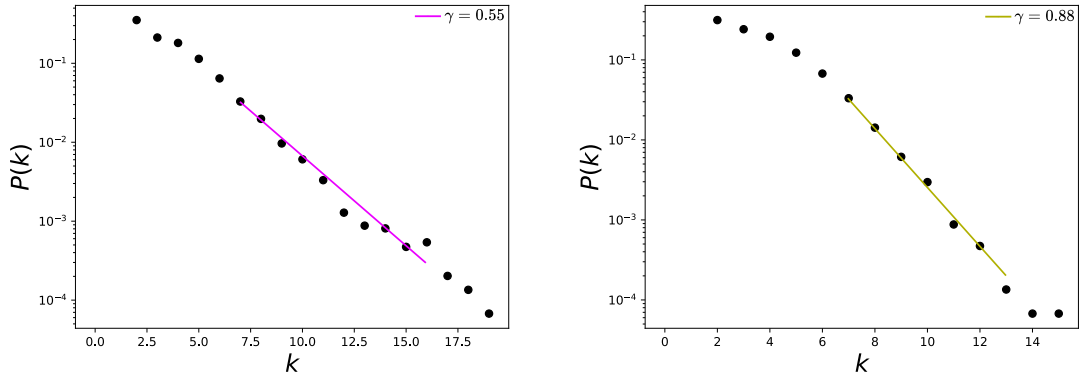
Figure 4.5: Log-log plot of the degree distributions. **Left** panel: magnetic field time series; **right** panel: sunspots time series.

Table 4.2: Values for γ for the two networks built by the VG method.

	Magnetic Field	Sunspots
γ_{vg}	3.13 ± 0.08	3.04 ± 0.05

We can also observe in Figure 4.6 that networks for both time series follow an exponential distribution $P(k)$, as expected. The value of γ in this case is also given by the slope of the linear fit, but of the semi-log distribution, considering only the tail [77], where a linear relation for $\ln P(k)$ and k holds.

The estimated values for γ_{vg} are 3.13 and 3.04 for the magnetic field and sunspots networks, respectively. The estimated values of γ_{hvg} are 0.55 and 0.88 for the magnetic field and sunspots networks, respectively, which suggests an underlying corre-



(a) (b)

Figure 4.6: Semi-log plot of the degree distributions. **Left** panel: magnetic field time series; **right** panel: sunspots time series.

Table 4.3: Values for γ for the two networks built by the HVG method.

	Magnetic Field	Sunspots
γ_{hvg}	0.55 ± 0.03	0.88 ± 0.05

lated stochastic process [76].

4.2 Local Analysis

Now, we employ moving time windows to follow the evolution of the network measures along the solar cycle. Two window sizes were chosen: 1-year windows, with a 1-month overlap; and 11-years windows, with a 1-year overlap. This leads to 493 windows of 1-year width, and 30 windows of 11-years width. We then plot results by associating, to each window, the time corresponding to its center.

The same metrics as in the global analysis were calculated. Results for the degree are shown in Figures 4.7 and 4.8. In this analysis, we present the results in two consecutive panels, where the left panel represents the magnetic field networks and the right panel represents the sunspots networks. The line color indicates the method used to build the networks, in black we have the VG method and in magenta we have the HVG. As expected, the VG method leads to larger number of connections than the HVG.

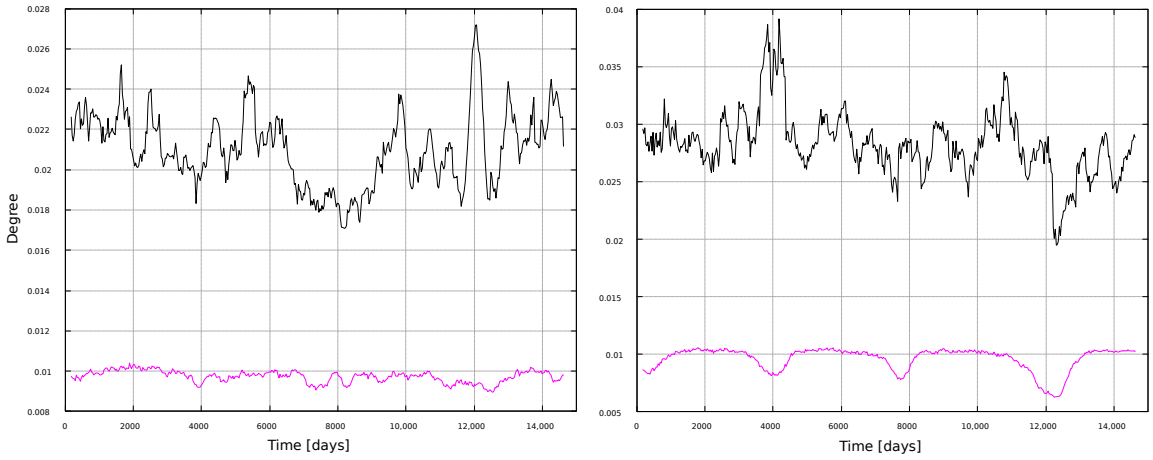


Figure 4.7: Degree for 1-year windows. **Left** panel: magnetic field networks; **right** panel: sunspots networks. Line color indicates the type of graph: VG (black line) and HVG (magenta line).

In general, results are consistent with the global results in Figure 4.2: the degree

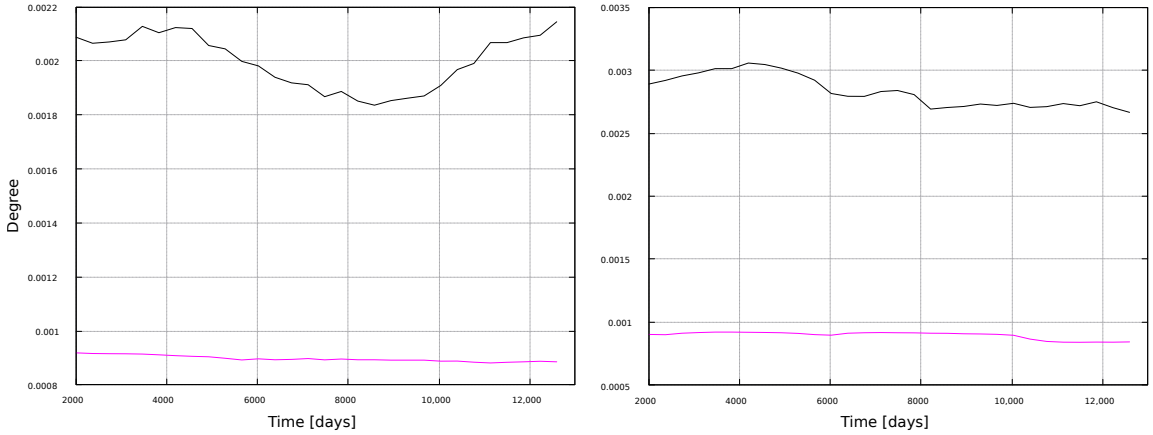


Figure 4.8: Degree for 11-year windows. **Left** panel: magnetic field networks; **right** panel: sunspots networks. Line color indicates the type of graph: VG (black line) and HVG (magenta line).

does not correlate with the solar cycle, regardless of the timescale of observation. The only exception is the HVG analysis for the sunspots time series, with 1-year windows (Figure 4.7), where clear minima close to solar minima can be found. This shows the nontriviality of the metrics involved. Although the degree is the simplest metric, it does not exhibit a clear correlation to the solar cycle, either for the complete time series of using a moving windows analysis, unless a specific strategy and time series is involved: an HVG for 1-year windows using the sunspots time series.

Figures 4.9 and 4.10 show the corresponding results for the clustering coefficient.

Larger values are obtained for the VG method, for both window types. Furthermore, results do not show clear correlations with solar activity, but notably, the HVG method has the same kind of oscillating behavior as for the degree (Figure 4.7), but more pronounced (notice that both measures are normalized, so that their maximum possible value is 1). Interpretations of this behavior will be discussed later, in Section 7.

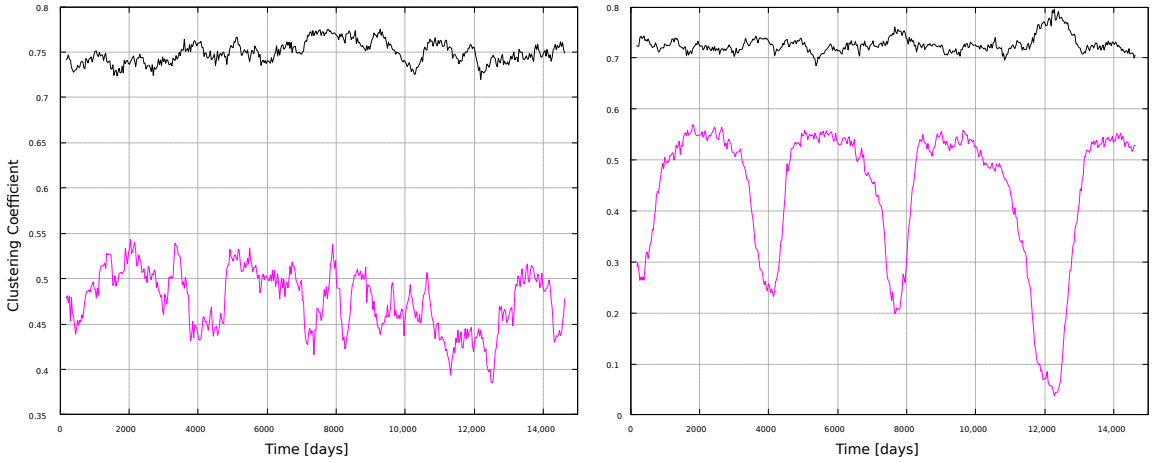


Figure 4.9: Clustering coefficient for 1-year windows. **Left** panel: magnetic field networks; **right** panel: sunspots networks. Line color indicates the type of graph: VG (black line) and HVG (magenta line).

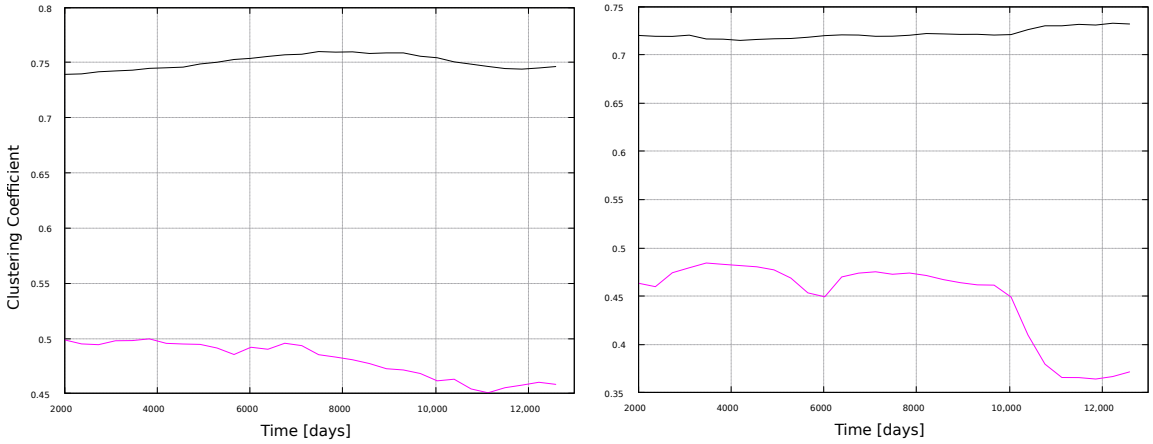


Figure 4.10: Clustering coefficient for 11-year windows. **Left** panel: magnetic field networks; **right** panel: sunspots networks. Line color indicates the type of graph: VG (black line) and HVG (magenta line).

Betweenness centrality results are shown in Figures 4.11 and 4.12. We already noticed, in the global analysis (Sec. 4.1), that BC was an interesting metric, due to its apparent sensitivity to the solar cycle (Figure 4.3). This is found here for the wider windows as well, Figure 4.12, showing peaks associated to maxima in solar activity (Figure 4.1). Thus, BC correlates well with solar activity, but if large timescales are

studied (full time series in Figure 4.3, 11 years in Figure 4.12), and if the HVG is used. If shorter, 1-year windows are taken, or if the VG method is used, then the BC does not convey information on solar activity.

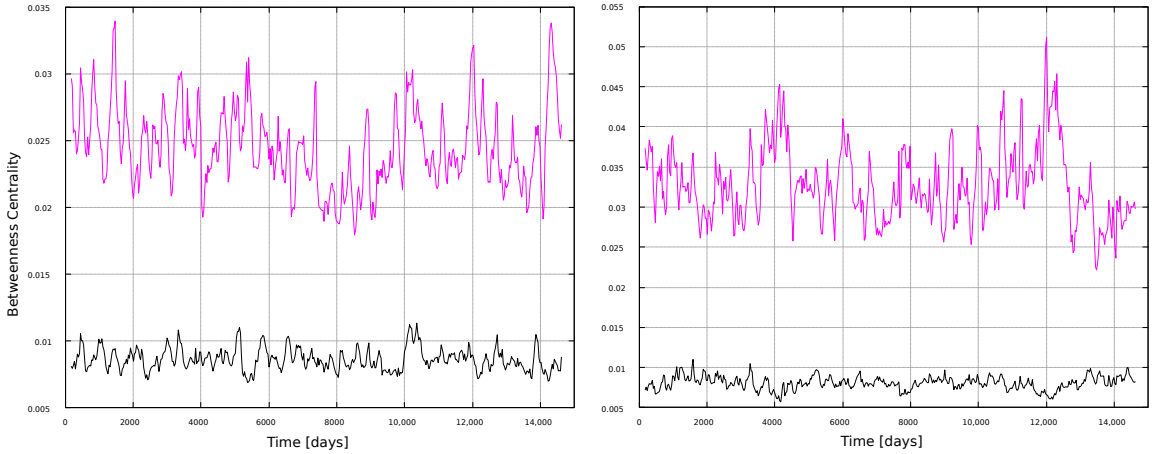


Figure 4.11: Betweenness centrality for 1-year windows. **Left** panel: magnetic field networks; **right** panel: sunspots networks. Line color indicates the type of graph: VG (black line) and HVG (magenta line).

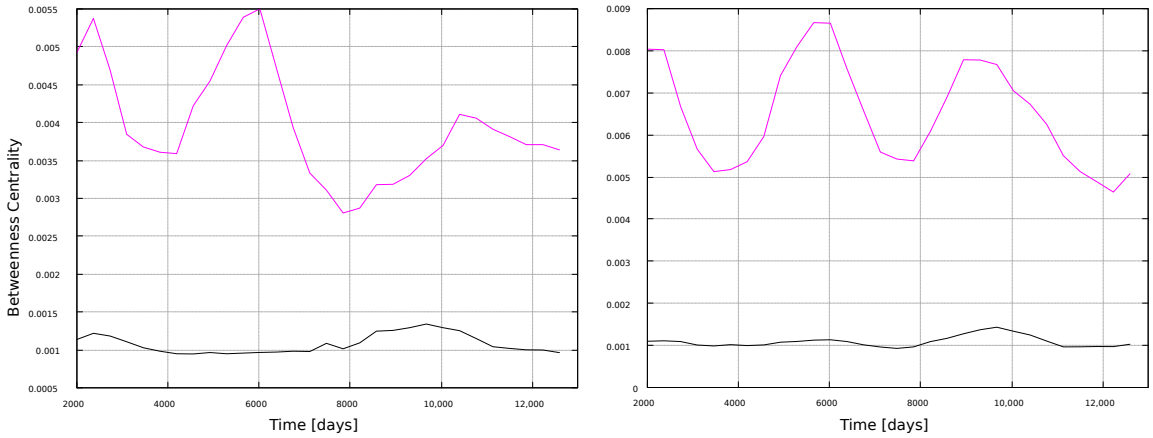


Figure 4.12: Betweenness centrality for 11-year windows. **Left** panel: magnetic field networks; **right** panel: sunspots networks. Line color indicates the type of graph: VG (black line) and HVG (magenta line).

Finally, we compute the eigenvector centrality, shown in Figures 4.13 and 4.14. Unlike Figure 4.4, this measure does not show interesting results for the local analysis,

regardless of the moving window width, thus highlighting again the nontriviality of the results, as the usefulness of the network approach to follow solar activity depends both on the metric and the timescale observed.

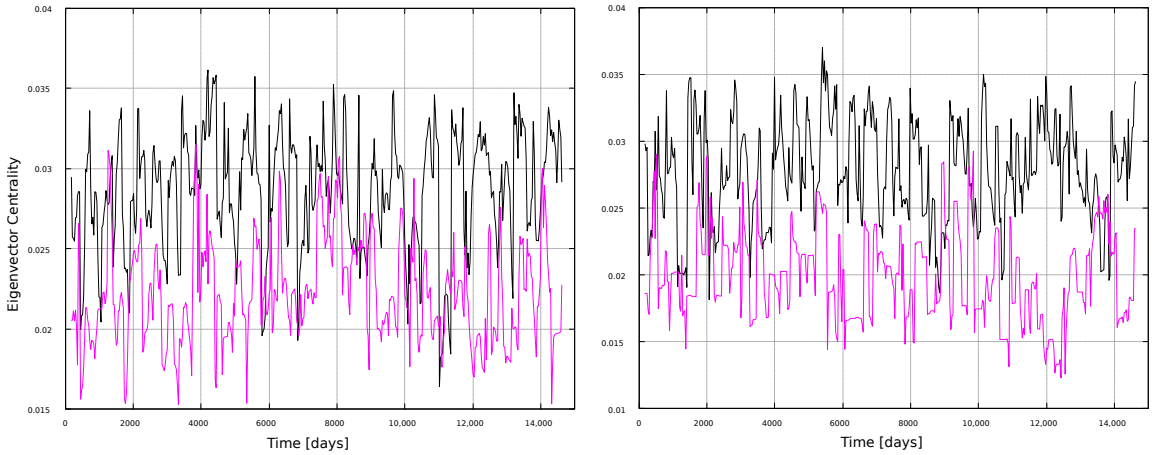


Figure 4.13: Eigenvector centrality for 1-year windows. **Left** panel: magnetic field networks; **right** panel: sunspots networks. Line color indicates the type of graph: VG (black line) and HVG (magenta line).

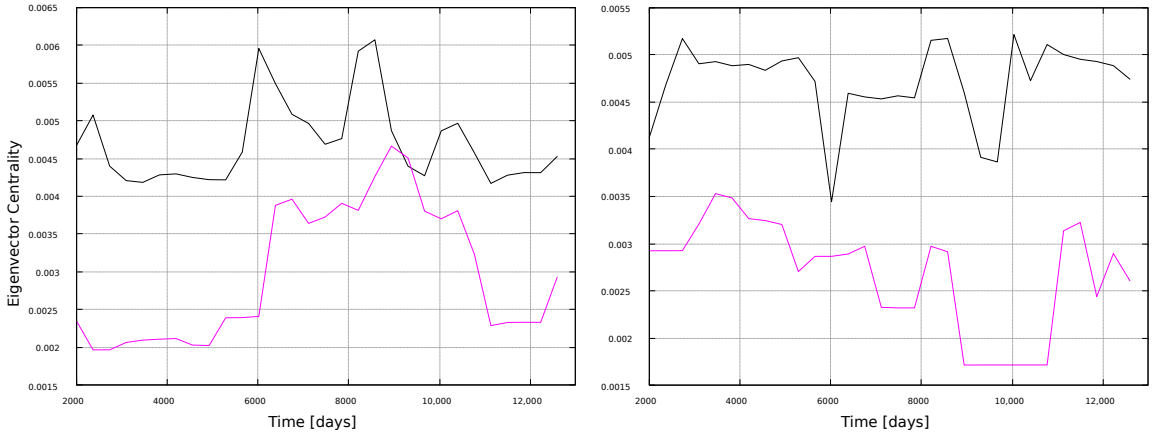


Figure 4.14: Eigenvector centrality for 11-year windows. **Left** panel: magnetic field networks; **right** panel: sunspots networks. Line color indicates the type of graph: VG (black line) and HVG (magenta line).

Regarding the degree distribution for the HVG method, all networks, for all time windows, exhibit an exponential topology, as in the case of the global networks,

consistent with previous results for the HVG [76]. The degree distributions $P(k)$ of every window are shown in Figures 4.15 and 4.16, for 1-year and 11-year windows, respectively.

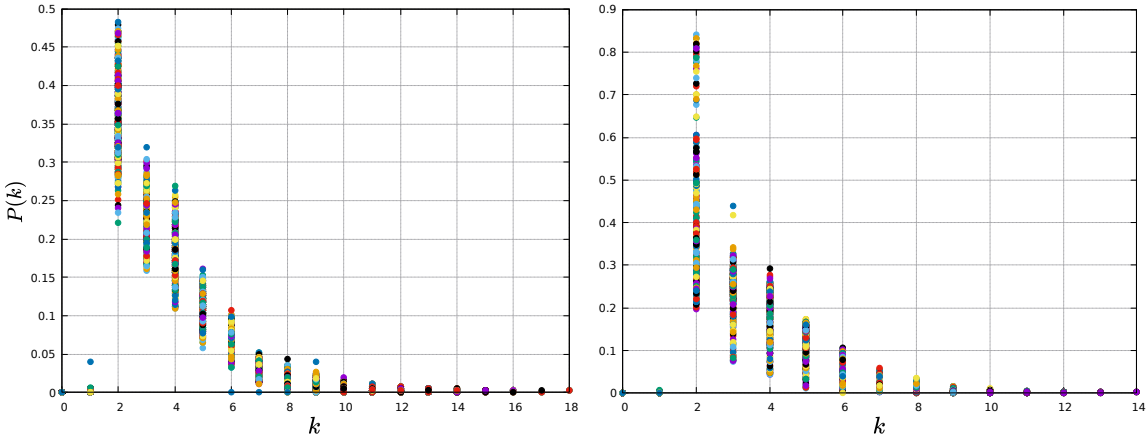


Figure 4.15: Degree distribution for each 1-year window. **Left** panel: magnetic field network; **right** panel: sunspots network. Each color represents a given window.

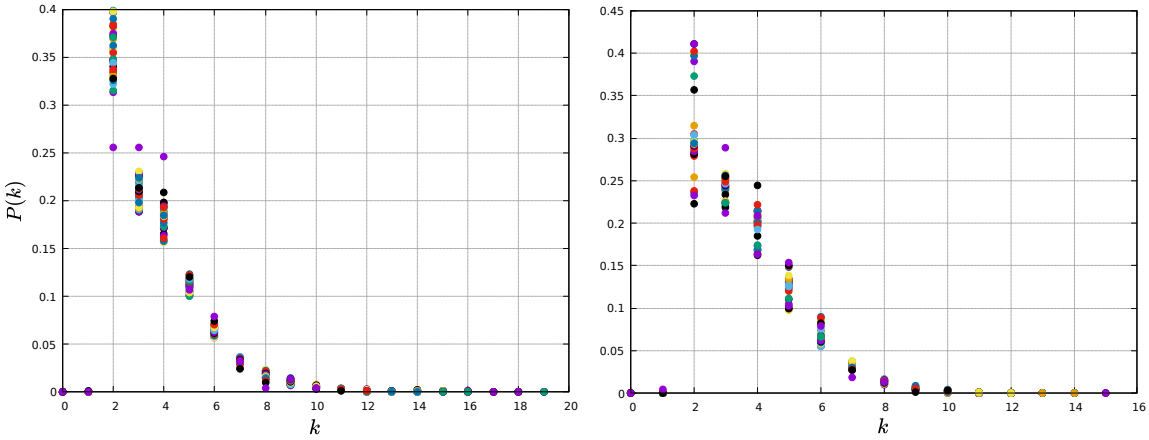


Figure 4.16: Degree distribution for each 11-year window. **Left** panel: magnetic field network; **right** panel: sunspots network. Each color represents a given window.

The decay exponent γ for each window is shown in Figure 4.17. For both window types, 1-year and 11-year windows, similar values $\gamma \sim 0.6$ are found along the solar

cycle. As mentioned before, this suggests correlated stochastic processes for every window, regardless of its length [76].

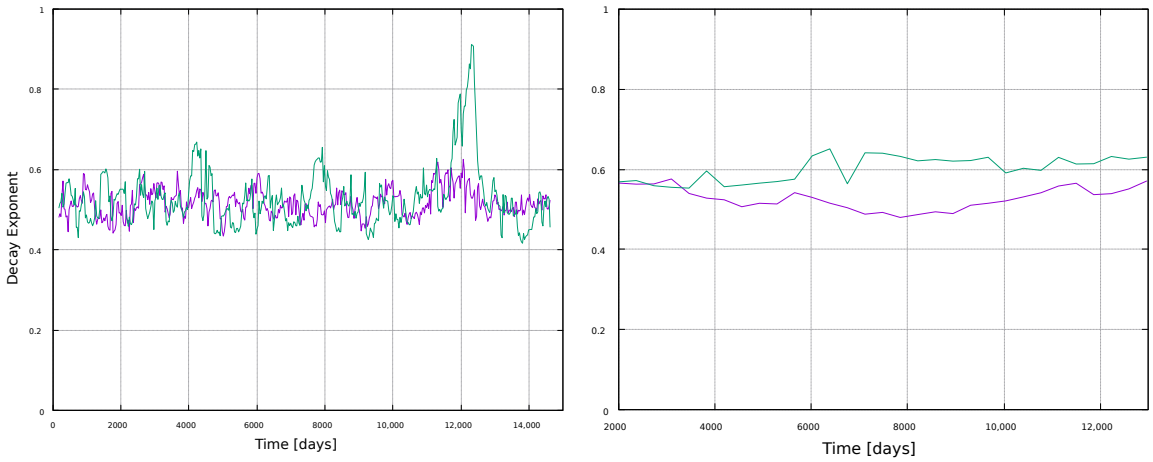


Figure 4.17: Evolution of the decay exponent for the degree distribution obtained from the HVG. **Left** panel: 1-year windows; **right** panel: 11-year windows. Line color indicate the time series used: magnetic field (purple line) and sunspots (green line).

4.3 Betweenness Centrality and Mean Magnetic Field

In the spirit of understanding the apparent sensitivity of the betweenness centrality with respect to the solar activity, we will focus the results only in this metric and the mean magnetic field. We will use the Savitsky-Golay filter [78] on the time series in order to obtain smoother curves without losing the original shape of the series, such as peaks and valleys. In Figure 4.18 we can observe the results for a smoothened magnetic field using the Savitsky-Golay filter and the betweenness centrality.

Additionally, we need to further analyze the relation between the BC metric and the solar cycle, because we only have a qualitative relation between these two variables. In order to analyze this behavior we employed a statistical method which quan-

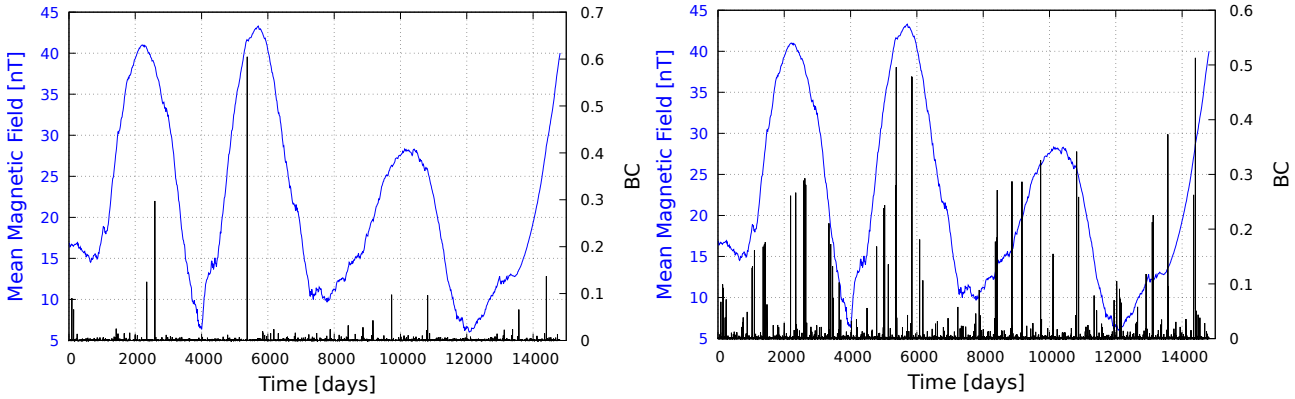


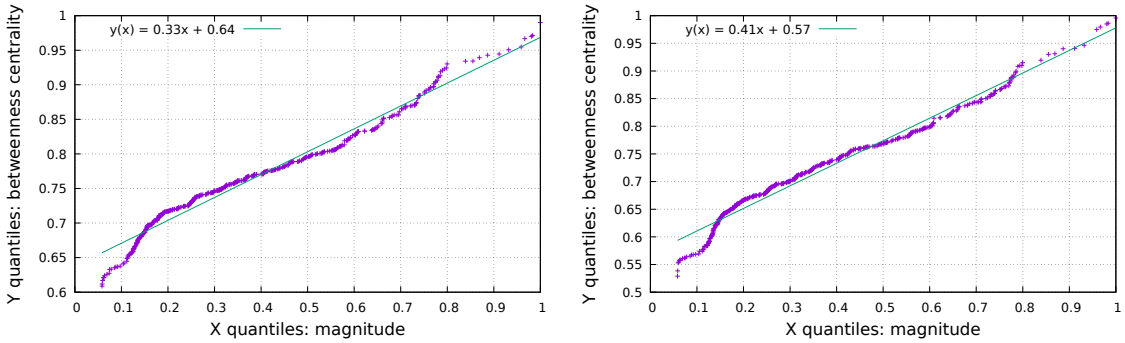
Figure 4.18: Betweenness centrality and mean magnetic field time series. **Left** panel: BC computed with the VG analysis; **right** panel: BC computed with the HVG analysis.

tifies correlations between variables, namely, the quantile-quantile plot [79]. This is an empirical method which determines if two data sets come from populations with common distribution. As the name indicates, it is a plot of the quantiles of the first data set against the quantiles of the second data set. By a quantile, we mean the fraction of points below a given value. If the two data sets being compared are similar, the Q-Q plot will approximately lie on a line. The advantages of the Q-Q plot are that the sample sizes do not need to be equal and we can test simultaneously many distributional aspects. This method has been used to study the time evolution for intraplate earthquakes [80] suggesting strong correlation between seismic events magnitudes and the fractal dimension of the events.

We carry out a Q-Q analysis to compare the betweenness centrality and the mean magnetic field datasets, for the moving windows strategy. That is, the datasets to be compared correspond to the average betweenness centrality for a certain time window, and the mean magnetic field during the same window.

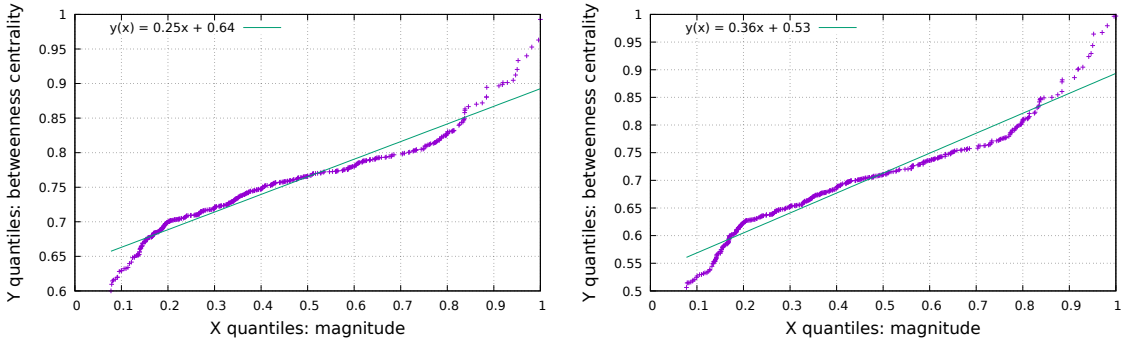
Results for various window sizes, with various overlaps, are shown in Figs. 4.19–4.25. It can be seen that, for all window sizes, an approximately linear behavior

is found, showing that both time series are indeed correlated across all timescales examined. This correlation opens many possibilities for solar studies. For example, temporal calculations of the betweenness centrality or other sensitive metrics may be used to forecast solar activity, a possibility which is also suggested by previous results based on spatiotemporal networks [30].



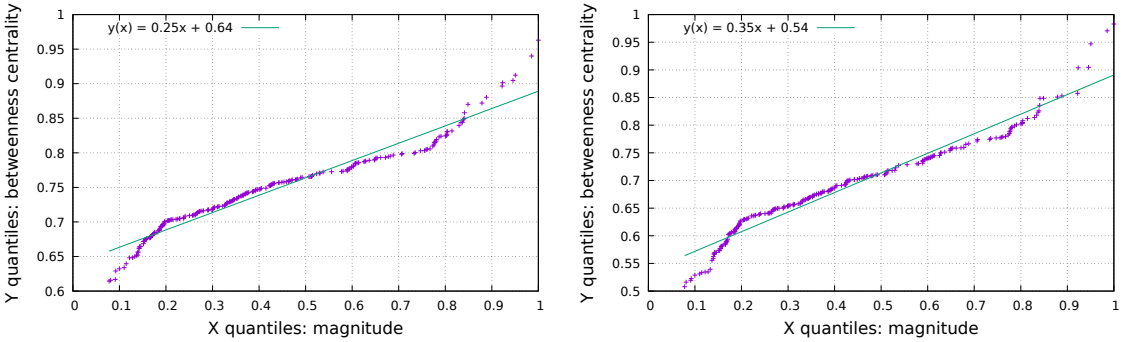
(a) VG analysis (b) HVG analysis

Figure 4.19: 1-year windows with 1-month overlap.



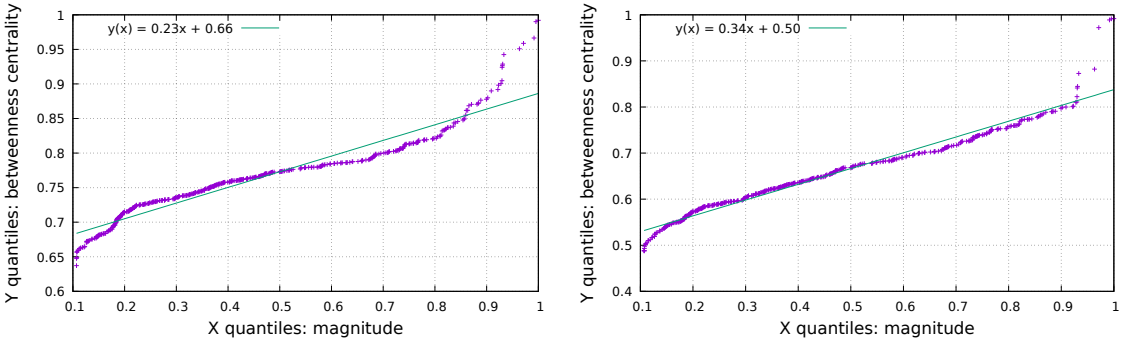
(a) VG analysis (b) HVG analysis

Figure 4.20: 548-days windows with 1-month overlap.



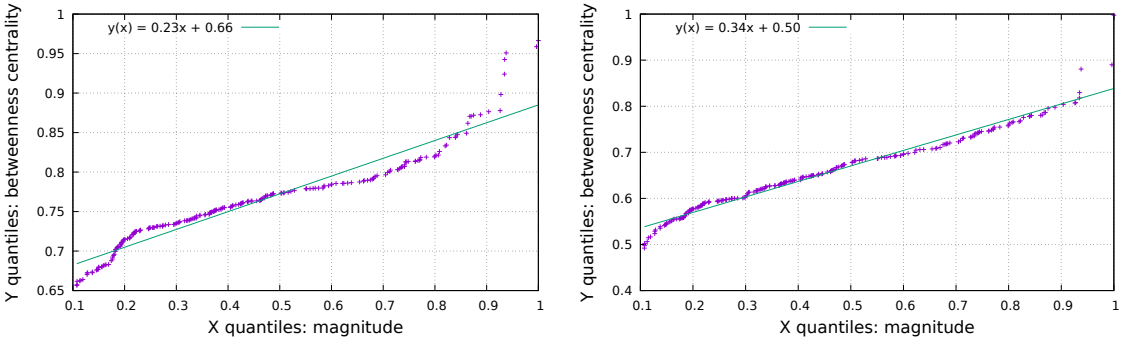
(a) VG analysis (b) HVG analysis

Figure 4.21: 548-days windows with 2-months overlap.



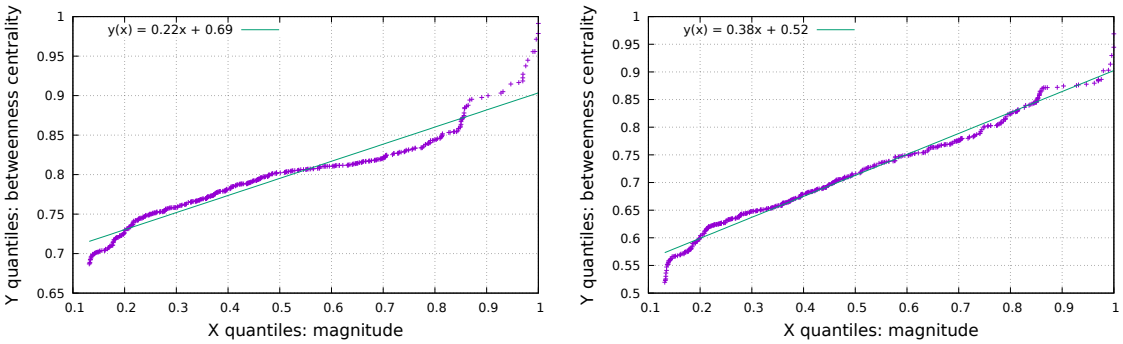
(a) VG analysis (b) HVG analysis

Figure 4.22: 730-days windows with 1-month overlap.



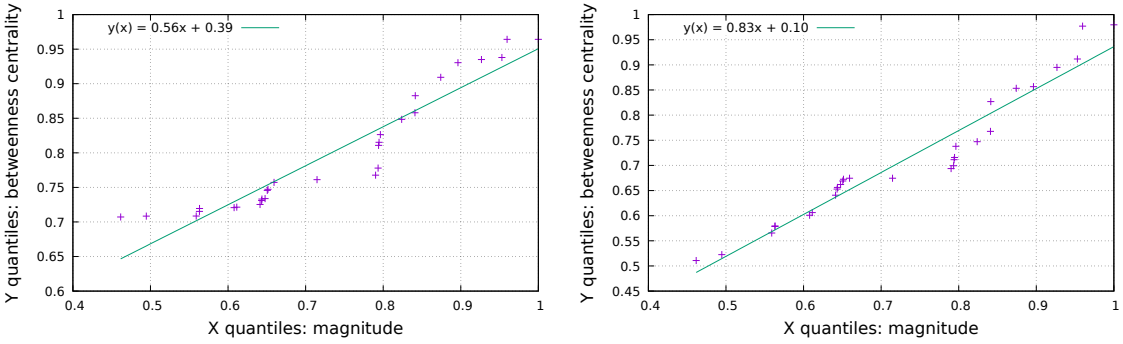
(a) VG analysis (b) HVG analysis

Figure 4.23: 730-days windows with 2-months overlap.



(a) VG analysis (b) HVG analysis

Figure 4.24: 1000-days windows with 1-month overlap.



(a) VG analysis (b) HVG analysis

Figure 4.25: 11-years windows with 1-year overlap.

Chapter 5

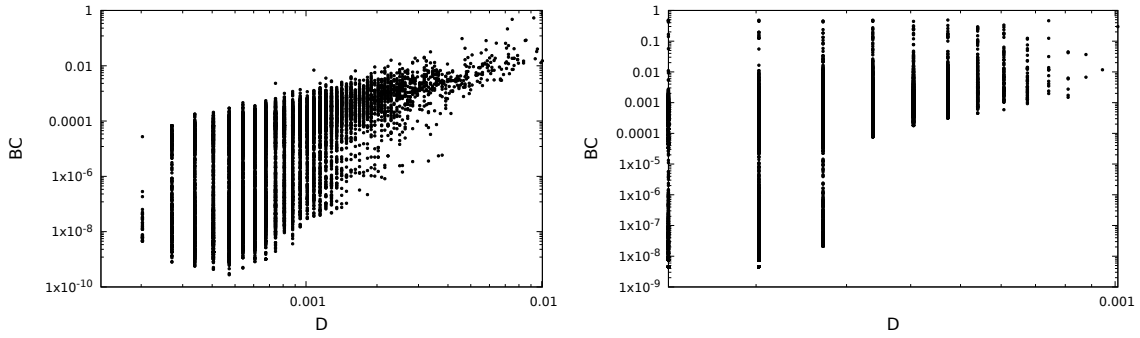
Correlations in the Networks

For the two networks described in Chapters 3 and 4, we seek to find the relation among metrics in order to obtain universal parameters that help to characterize the solar activity.

For example, both the degree and betweenness centrality, to some extent, quantify how important a node is within the network and we expect to obtain some correlation between these two metrics. The latter is based on the assumption that a node with high degree should also have high betweenness. We have found that the centrality measures (betweenness centrality and eigenvector centrality) and the clustering coefficient correlate with the degree as in [81], where real-world networks were studied.

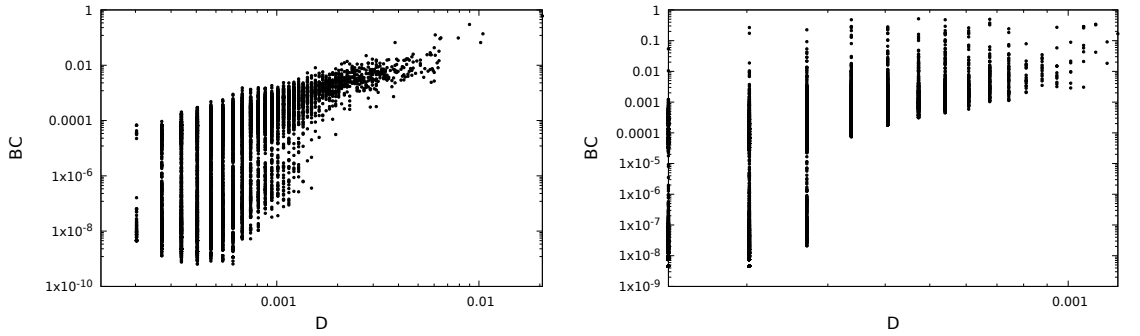
The calculation of the betweenness is based on the global information on paths connecting all pair of nodes, while the degree, by definition, is a quantity that depends only on local information. As we said before, it is expected for the two metrics to be correlated if we understand them as a measure of importance in the network. The results are shown in the scatter plot (logarithmic scale) of node betweenness (BC) versus node degree (D) in Figures 5.1 and 5.2, for the global network. Both the magnetic field and the sunspots networks show clear signs of correlation between the

degree and the betweenness.



(a) VG analysis (b) HVG analysis

Figure 5.1: Node betweenness centrality versus node degree for the sunspots network. Both graphs are in log-log scale.

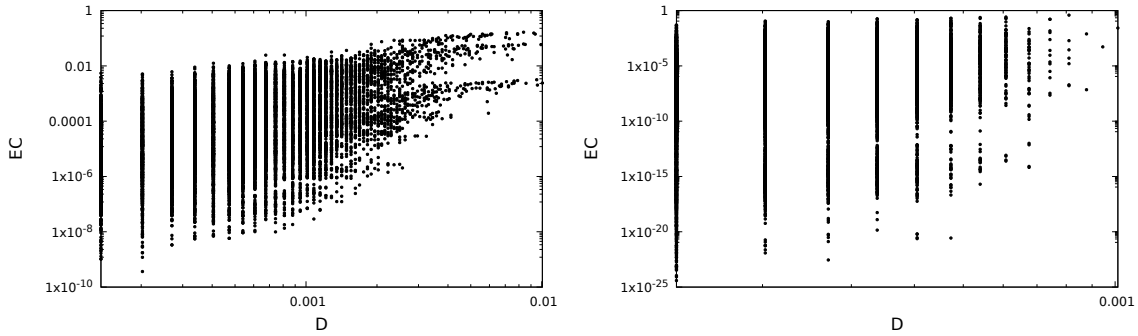


(a) VG analysis (b) HVG analysis

Figure 5.2: Node betweenness centrality versus node degree for the magnetic field network. Both graphs are in log-log scale.

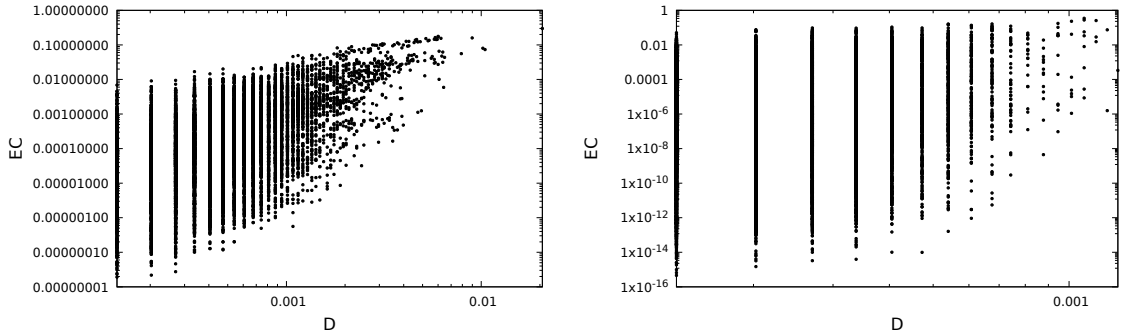
A similar case occurs in Figures 5.3 and 5.4 for the eigenvector centrality. That is, on average, higher values of the eigenvector centrality are found for nodes with higher degree.

The results shown in Figures 5.5 and 5.6 correspond to the scatter plot between the clustering coefficient and the degree. It is interesting to notice not only a clear correlation between the variables, but also a power-law behavior, suggesting that



(a) VG analysis (b) HVG analysis

Figure 5.3: Eigenvector degree versus degree for the sunspots network. Both graphs are in log-log scale.

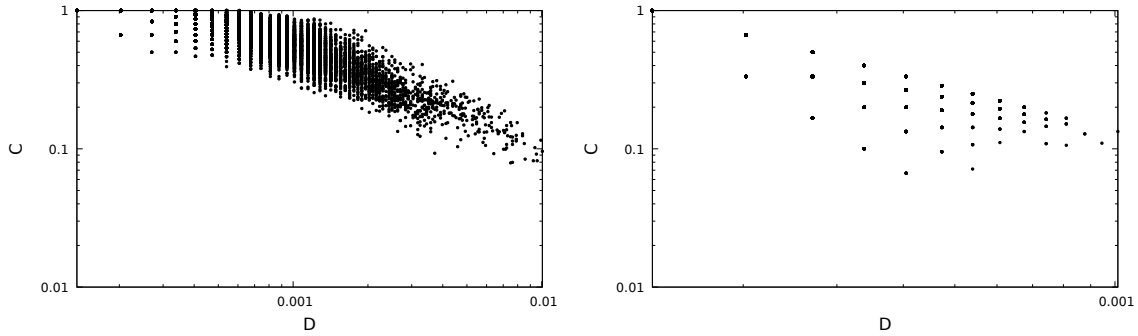


(a) VG analysis (b) HVG analysis

Figure 5.4: Eigenvector centrality versus degree for the magnetic field network. Both graphs are in log-log scale.

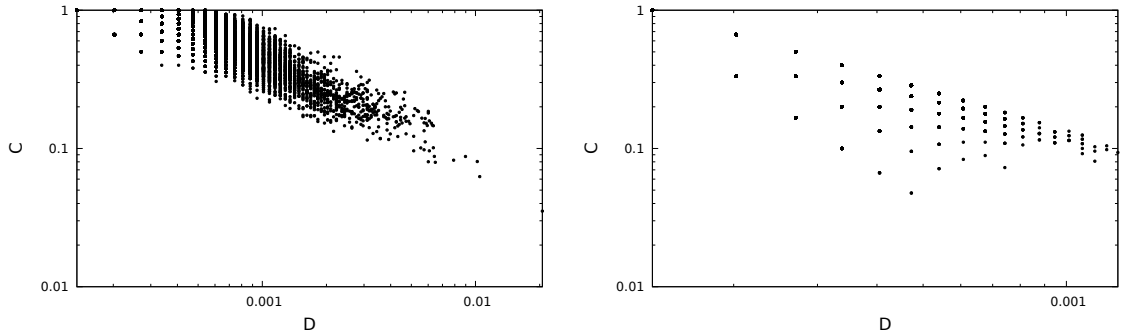
the node clustering coefficient decays as $c(\nu) \approx k(\nu)^{-\delta}$, with δ the characteristic exponent of the decay.

In order to further understand if there are correlations or not, we compute the linear correlation coefficient $\rho(i, j)$ (or usually abbreviated as r for no good reason except that there are two r's in "correlation") between any two metrics i and j defined in Section 3.2. This correlation coefficient is computed using a statistical definition named "standard units", which is a way of putting different kinds of observations on



(a) VG analysis (b) HVG analysis

Figure 5.5: Clustering coefficient versus degree for the sunspots network.



(a) VG analysis (b) HVG analysis

Figure 5.6: Clustering coefficient versus degree for the magnetic field network.

the same scale following the next idea: replace a datum by the number of standard deviations it is above the mean of the data, if a datum is above the mean it is positive and if it is below it is negative. The definition of the conversion can be expressed as

$$\frac{\text{original datum} - \text{mean of original data}}{\text{standard deviation of original data}}.$$

The objective of this conversion is to obtain a new set of data with mean zero and the standard deviation is one. The matrix ρ is called the correlation matrix. The absolute value $0 \leq |\rho(i,j)| \leq 1$ characterizes the strength of the correlation between the corresponding metrics i and j. If $|\rho(i,j)|$ is close to zero, the two metrics are

almost uncorrelated whereas a $|\rho(i,j)|$ close to 1 implies a strong correlation. The sign of $\rho(i,j)$ is not further explored (positive or negative correlation), because it is the strength that indicates to which extent a metric can be predicted from the other.

The results are shown in the following figures: VG analysis for magnetic field network (Fig. 5.7) and sunspots network (Fig. 5.8), and HVG analysis for magnetic field (Fig. 5.9) and sunspots networks (Fig. 5.10). The colors are indicated in the color bar at the right of every figure, where lighter colors are larger values for $|\rho(i,j)|$, and darker ones are lower values. In the figures we analyze the four metrics: degree, clustering coefficient, betweenness centrality and eigenvector centrality, and the magnetic field time series. Each row or column represents one of these data.

In the case of VG, as we can see in the Figures 5.7 and 5.8, there is a strong correlation between the degree and the clustering coefficient for both networks, there is also a good correlation for the eigenvector centrality and the degree for the magnetic field network, and for the sunspots network this correlation is weaker, but still the correlation value is greater than 0.5, closer to 1 than 0. The other metrics show lower values of correlation, indicating an uncorrelated rather than correlated behavior, with the particular case of betweenness centrality and degree, which is just in the limit of 0.5, indicating neither correlated nor uncorrelated behavior.

On the other hand, for both networks constructed by the HVG method, we can observe that there are only uncorrelated behaviors for the metrics, indicating that this case is not good enough for the purpose of predicting metrics from one another.

The previous plots show correlations within the network, but we should also analyze the correlations between the network metrics and the time series, particularly with the magnetic field times series which is indeed the physical variable of interest.

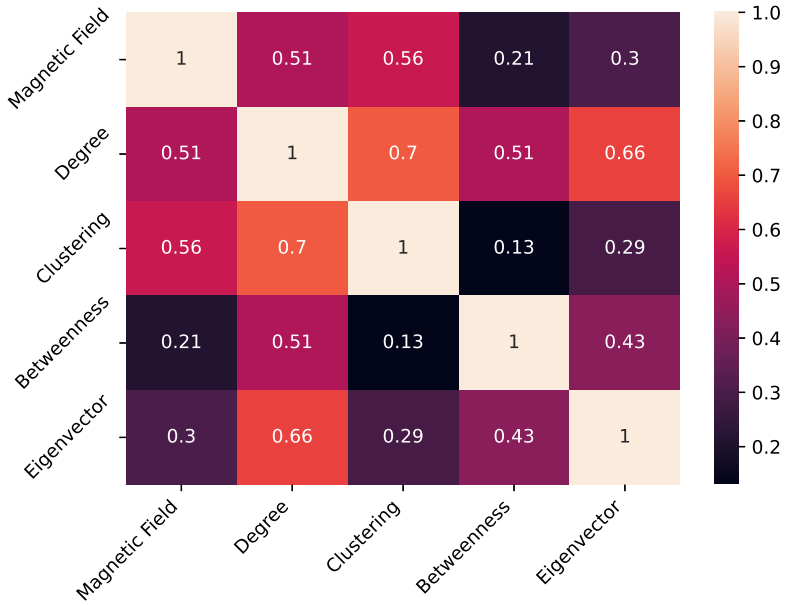


Figure 5.7: Correlation matrix for the metrics of the magnetic field network constructed by using the VG method.

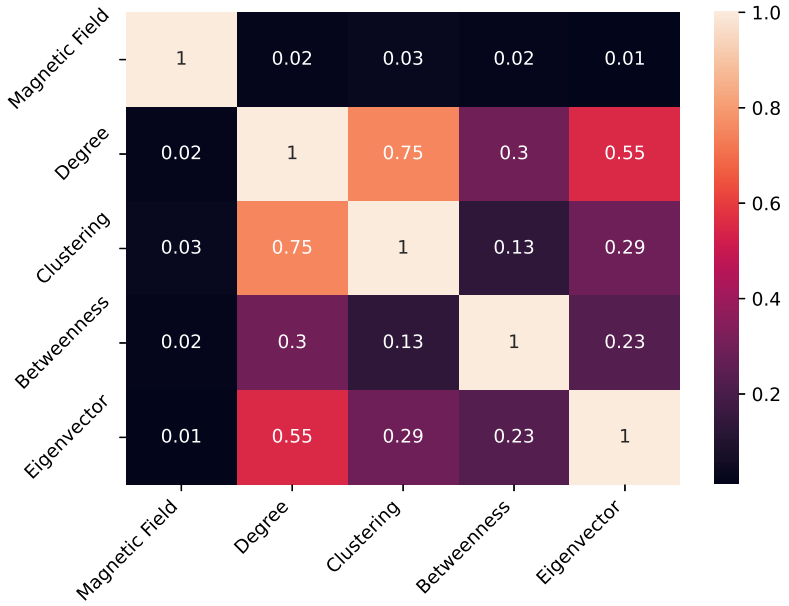


Figure 5.8: Correlation matrix for the metrics of the sunspots network constructed by using the VG method.

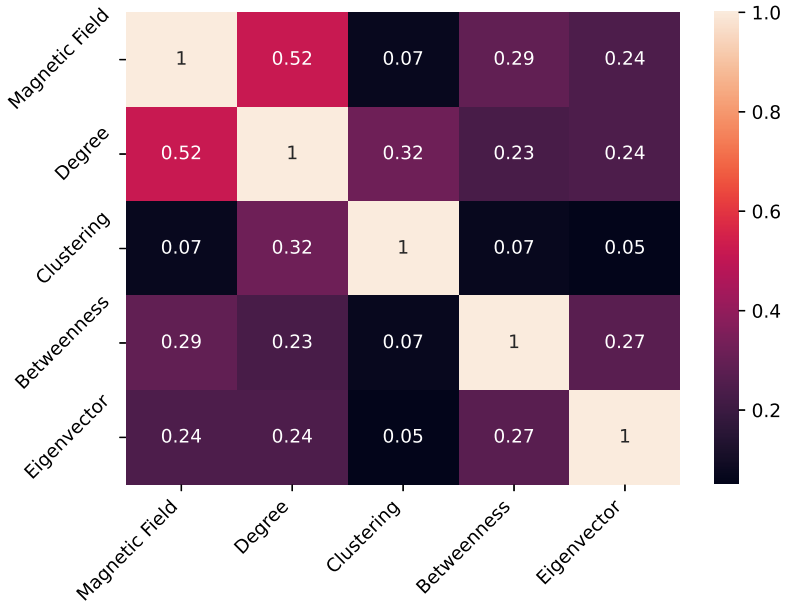


Figure 5.9: Correlation matrix for the metrics of the magnetic field network constructed by using the HVG method.

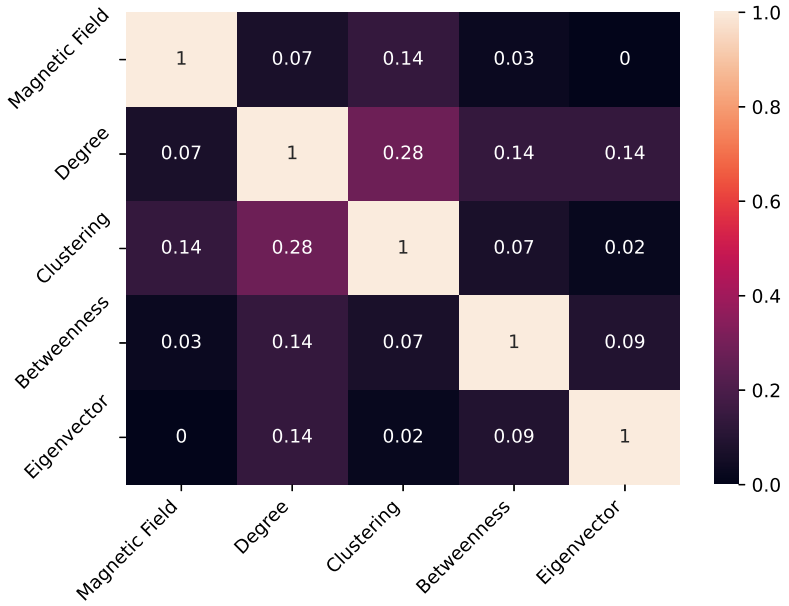


Figure 5.10: Correlation matrix for the metrics of the sunspots network constructed by using the HVG method.

For this purpose, we have computed in Chapter 4 the quantile-quantile plot for several time windows and for the particular case of betweenness centrality and magnitude of magnetic field.

As we can see from Figures 5.7, 5.8, 5.9 and 5.10, there is no sign of strong correlation between the magnetic field and any metric, except for the clustering coefficient in the VG network (Fig. 5.7), where the maximum value for correlation is obtained, but still it is not strong. In principle this result for the magnetic field is not so promising as we would have expected, considering the evident visual correlation for betweenness centrality and magnetic field time series. In this sense, we have to keep in mind that the correlation matrix is used to analyze or measure lineal relationships between variables. In this case, the results are saying that we have no linear corellation between the magnetic field and the betweenness centrality or to some extent, the correlation coefficient may not accurately capture the strength of the relationship.

Following this idea, we have to further explore other correlation methods because, as we have said before, in this work we are interested in studying possible correlations of the metrics and the magnetic field and for this reason we need to further quantify the significance of correlations. Another test to measure the statistical significance of a correlation between time series is the Student t-test, which is a measure of the significance of a difference of means between two distributions X_a and X_b. In particular, for this case we use the Welch's t-test or the unequal-variance t-test.

If N_a and N_b correspond to the total number of data of the X_a and X_b time series, respectively, and $\mathrm{Var}(X)$ is the variance of the time series X, then a quantity called the t value is computed, which measures how many standard errors the sample means are apart, that is, the correlation is determined by analyzing the disparity in

their averages.

Two statistical values are relevant in the test, namely the t value and the degrees of freedom df:

$$t = \frac{\overline{X}_a - \overline{X}_b}{\sqrt{\dfrac{\mathrm{Var}(X_a)}{N_a} + \dfrac{\mathrm{Var}(X_b)}{N_b}}}\,, \tag{5.1}$$

and

$$df = \frac{\left(\dfrac{\mathrm{Var}(X_a)}{N_a} + \dfrac{\mathrm{Var}(X_b)}{N_b}\right)^2}{\dfrac{(\mathrm{Var}(X_a)/N_a)^2}{N_a - 1} + \dfrac{(\mathrm{Var}(X_b)/N_b)^2}{N_b - 1}}\,, \tag{5.2}$$

which represent the number of independent values in a dataset that are free to vary.

Given the expressions for the t value and the degrees of freedom, the definition of the p value, for a two-tailed test, is given by [82]

$$p = 2\left[1 - F(|t|, df)\right]\,, \tag{5.3}$$

where $F(x, df)$ is the cumulative distribution function (CDF) for the variable x, with df degrees of freedom. Multiplying by 2 accounts for the two-tailed nature of the test (deviations in both directions or tails from the null hypothesis).

The hypothesis will be satisfied if the p-value is $p < 0.05$ (which means that there is a 5% chance of obtaining a significant result by chance alone), meaning that the difference of the means of the time series is statistically significative.

We have applied this analysis for betweenness centrality (obtained through VG and HVG), and the magnetic field time series. We have obtained p values close to zero ($p < 10^{-5}$) and t values of 184.23 and 180.72, for VG and HVG betweenness centrality, respectively.

Chapter 6

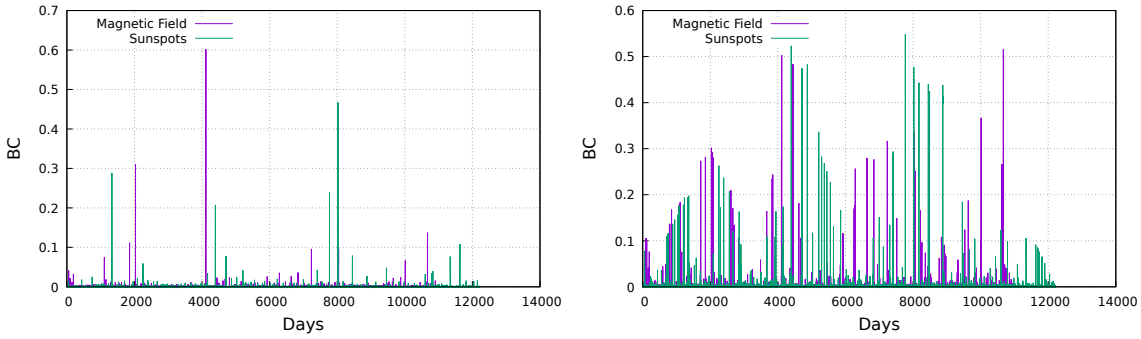
Method Robustness

In the examination of network vulnerabilities, the method by which nodes are chosen for removal follows an open selection process. It is indeed possible to enhance the extent of damage when a specific number of nodes (or edges) are eliminated. Nevertheless, achieving this requires a comprehensive understanding of the entire network's structure. In this chapter we propose an attack to the networks, in which we will remove data from the time series using two different methods. The first method, namely *Periodic Removal*, consists of removing an amount N of data from the time series separated by a certain gap T. The second method, *Random Removal*, consists in removing N random data from the time series. As more nodes are removed, the structure of the network should change, and therefore, different values and distributions for the metrics would be obtained. For the two removal methods, the same analysis for the networks as in Chapter 4 was carried on.

6.1 Periodic Removal

This method consists of creating gaps in the time series. This can simulate, for instance, the existence of observational gaps in the observation of a physical phenomenon. For example, if the time series corresponds to the light intensity received from a star, as measured by a telescope on the Earth's surface, there will be periodic gaps due to Earth's rotation [39].

We focus on the betweenness centrality, from the global networks, since it exhibits the best correlations with the magnetic field time series as shown in Chapters 4 and 5. Figures 6.1–6.4 show the results for both the VG and HVG analysis, for various choices of the windows and gap lengths: removal of 30 nodes every 365 days (Fig. 6.1), 100 nodes every 50 days (Fig. 6.2), 200 nodes every 75 days (Fig. 6.3), and 365 nodes every 30 days (Fig. 6.4). We chose these window sizes and gaps to test over different possibilites, without a preferential selection.

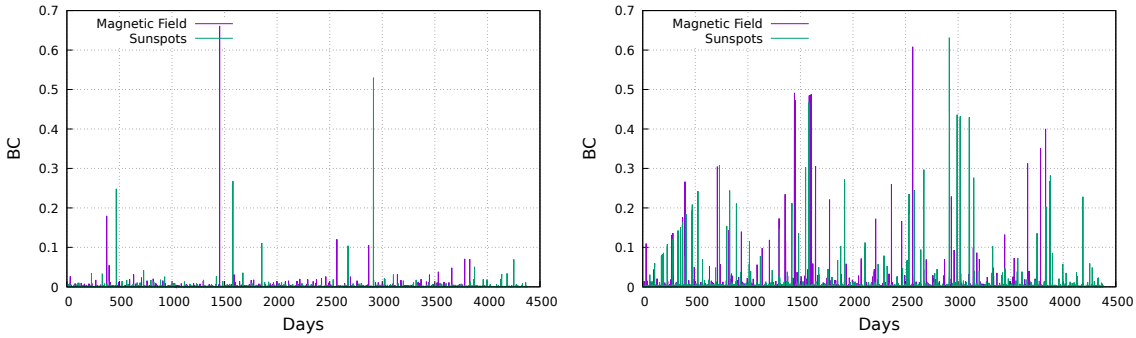


(a) VG analysis

(b) HVG analysis

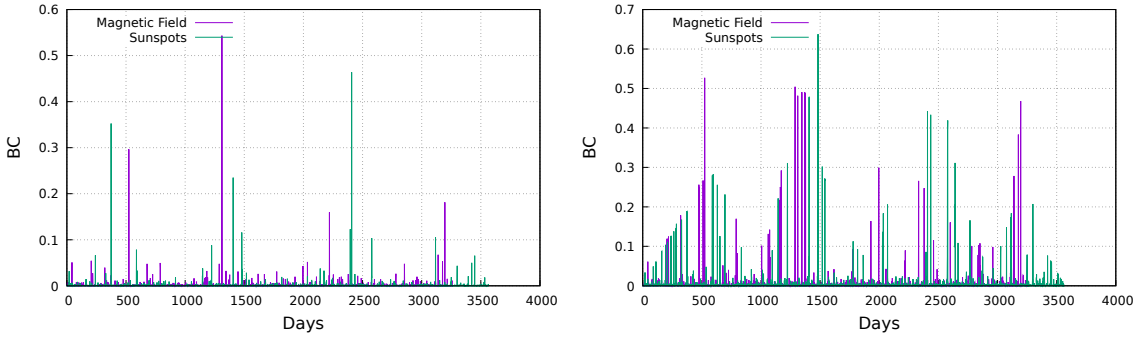
Figure 6.1: Removing 30 nodes every 365 days.

It can be observed that, despite data removal, the same trend is observed: betweenness centrality maxima and minima follow solar activity maxima and minima. The peaks in BC are sharper for the VG graphs, while the HVG graphs show a slower



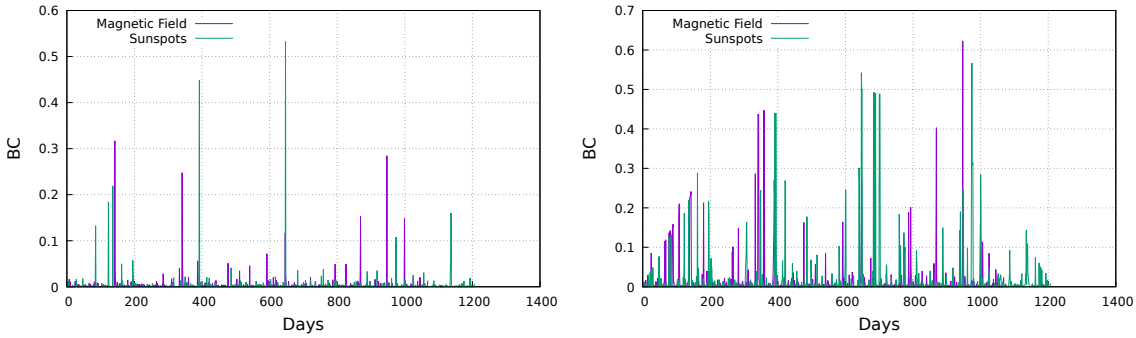
(a) VG analysis (b) HVG analysis

Figure 6.2: Removing 100 nodes every 50 days.



(a) VG analysis (b) HVG analysis

Figure 6.3: Removing 200 nodes every 75 days.

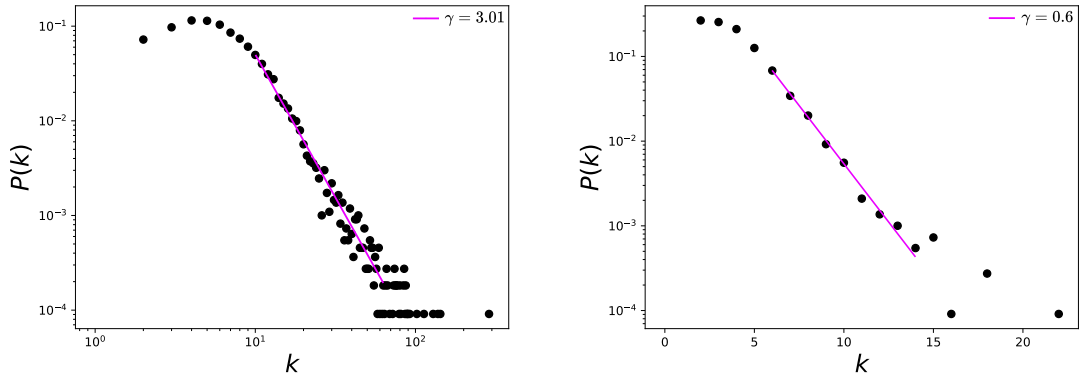


(a) VG analysis (b) HVG analysis

Figure 6.4: Removing 365 nodes every 30 days.

variation, which is consistent with previous plots, but the overall trend is the same. In all cases, the result is essentially not affected by the existence of gaps. Similar results have been observed for light curves of pulsating variable stars [39] , suggesting that the VG/HVG graph is a useful tool to study systems where observational gaps are unavoidable.

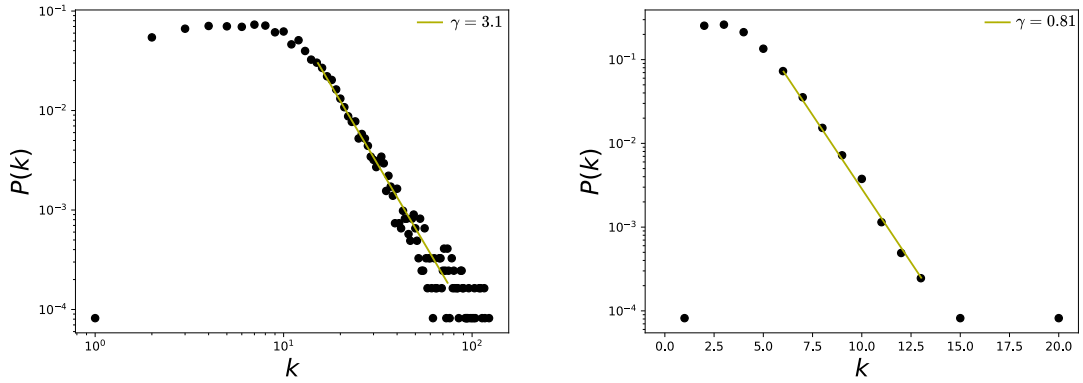
As noted in Section 4, degree distributions follow a power-law for the VG graph, and an exponential decay for the HVG graph (see, *e.g.*, Figures 4.5 and 4.6). This is also observed for the time series with gaps, except that, as the number of deleted data increases, the statistics worsens. This can be seen in Figs. 6.5–6.12.



(a) VG analysis (b) HVG analysis

Figure 6.5: Degree distribution analysis associated to the magnetic field networks built by removing 30 nodes every 365 days. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.

A variation of the decay exponent is observed with the different gaps introduced, but the functional form is the same, suggesting again the robustness of this approach, as the network topology is not significantly modified to missing data.



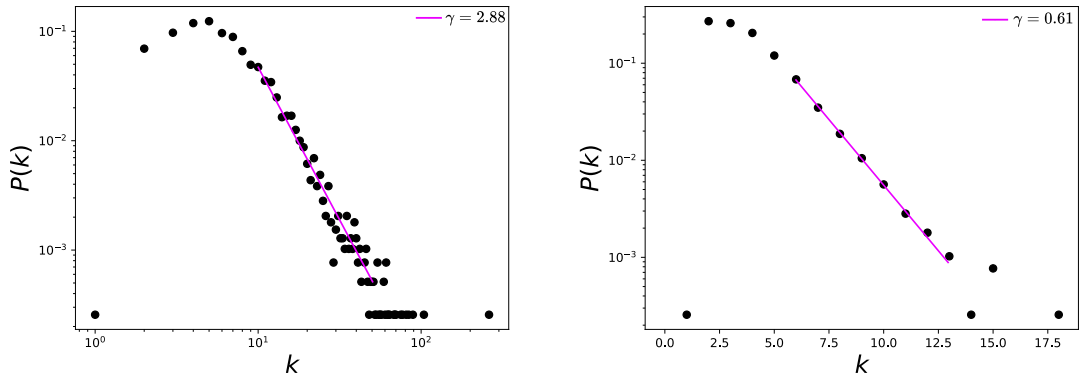
(a) VG analysis

(b) HVG analysis

Figure 6.6: Degree distribution analysis associated to the sunspots networks built by removing 30 nodes every 365 days. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.

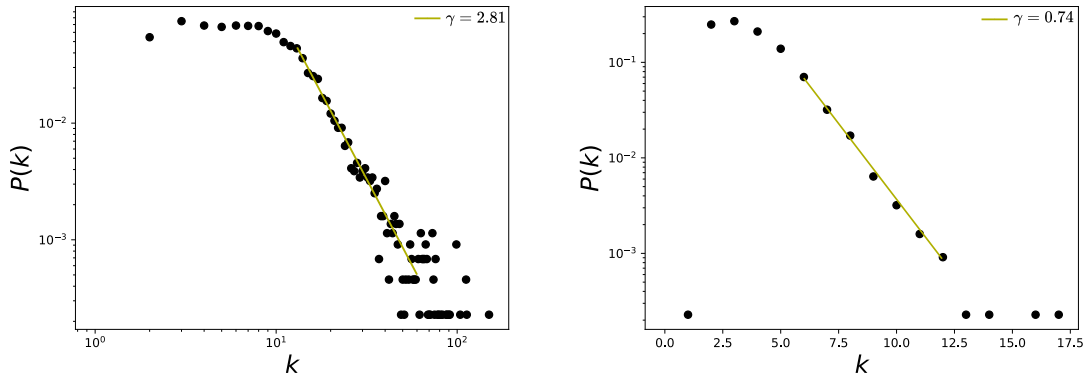
Table 6.1: γ values for the different removal intervals selected

		γ_{vg}	γ_{hvg}
Removing 30 nodes every 365 days	Mean magnetic field	3.01	0.6
	Sunspots	3.1	0.81
Removing 100 nodes every 50 days	Mean magnetic field	2.88	0.61
	Sunspots	2.81	0.74
Removing 200 nodes every 75 days	Mean magnetic field	3.06	0.7
	Sunspots	2.96	0.74
Removing 365 nodes every 30 days	Mean magnetic field	3.04	0.8
	Sunspots	2.49	0.61



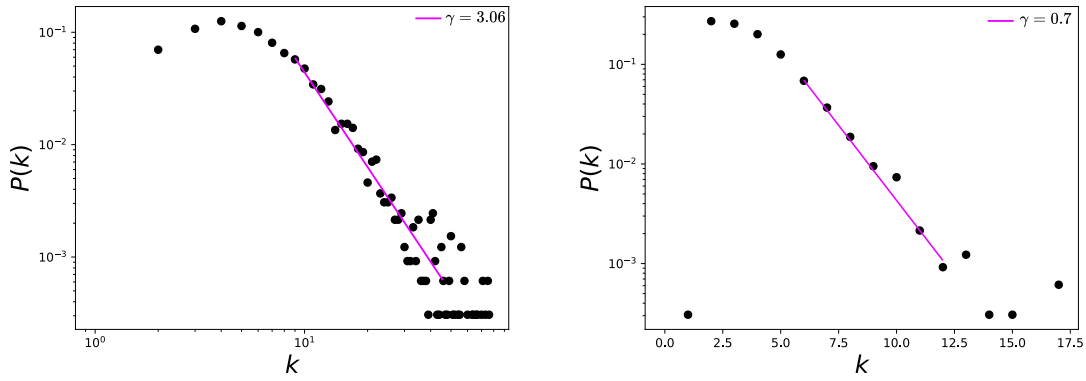
(a) VG analysis (b) HVG analysis

Figure 6.7: Degree distribution analysis associated to the magnetic field networks built by removing 100 nodes every 50 days. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



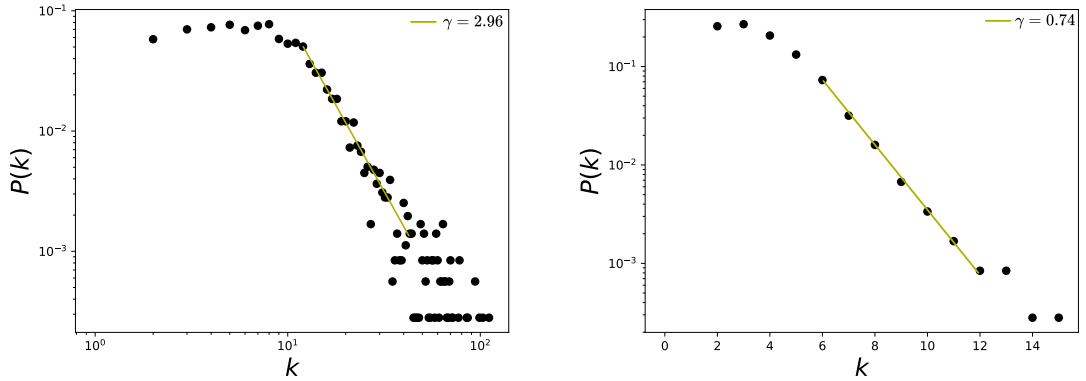
(a) VG analysis (b) HVG analysis

Figure 6.8: Degree distribution analysis associated to the sunspots networks built by removing 100 nodes every 50 days. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



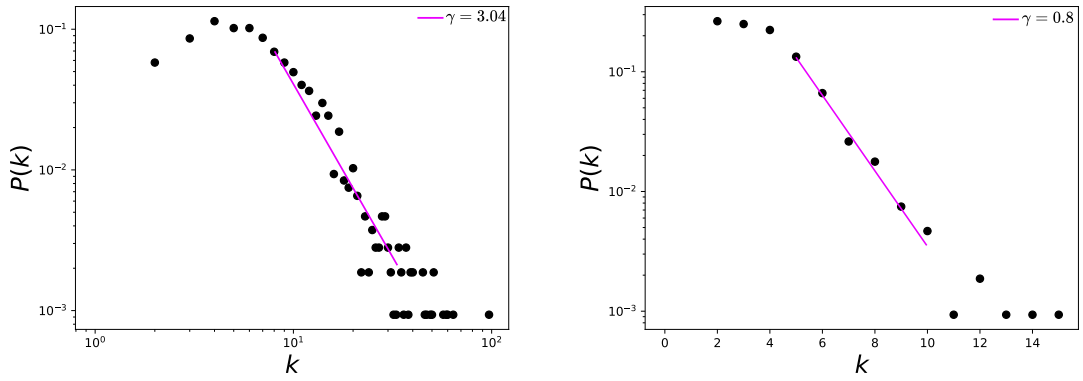
(a) VG analysis (b) HVG analysis

Figure 6.9: Degree distribution analysis associated to the magnetic field networks built by removing 200 nodes every 75 days. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



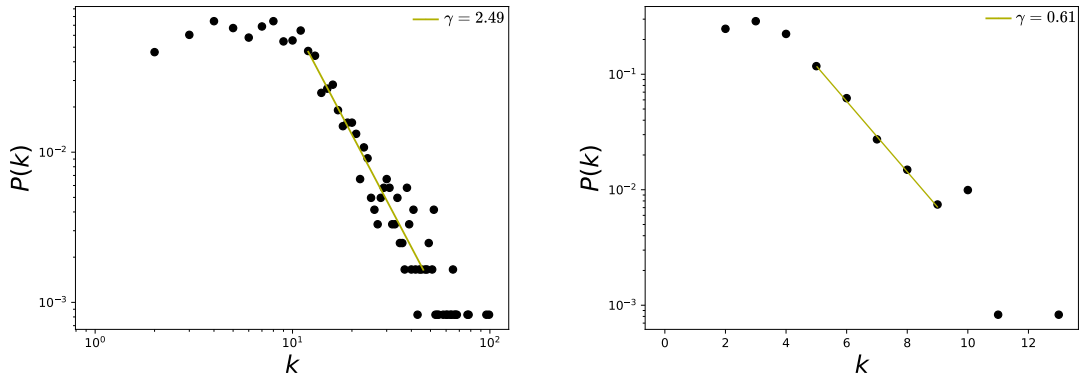
(a) VG analysis (b) HVG analysis

Figure 6.10: Degree distribution analysis associated to the sunspots networks built by removing 200 nodes every 75 days. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



(a) VG analysis (b) HVG analysis

Figure 6.11: Degree distribution analysis associated to the magnetic field networks built by removing 365 nodes every 30 days. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



(a) VG analysis (b) HVG analysis

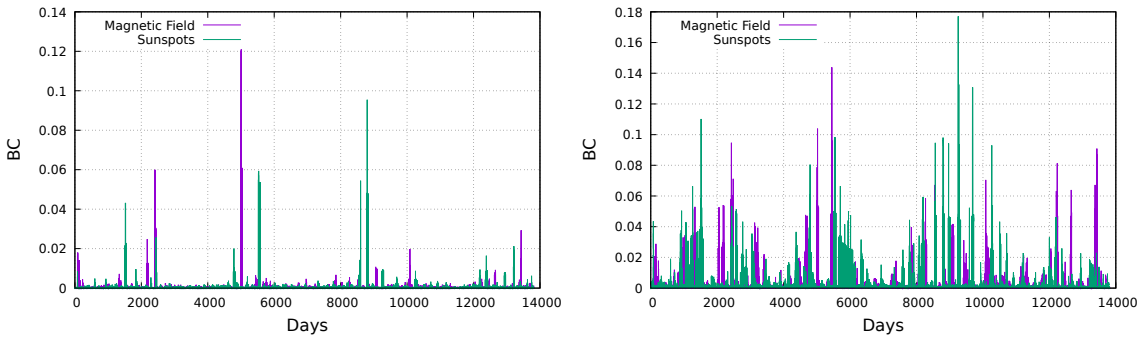
Figure 6.12: Degree distribution analysis associated to the sunspots networks built by removing 365 nodes every 30 days. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.

6.2 Random Removal

This method also intends to test the robustness of the network analysis and recreate common conditions in space observational data, where missing data appears often. In this case, we consider the case where data is missing at random places in the time series. As in the previous section, Sec. 6.1, we only show results for the betweenness centrality, due to its better sensitivity to solar activity (as observed in Chapter 4).

In Figures 6.13, 6.14, and 6.15 the results are shown for the betweenness centrality after removing 1000 (6.75% of the total data), 5000 (33.75% of the total data), and 10000 (67.5% of the total data) points at random locations in the series, using uniform randomness. The procedure consisted in removing n random data, then rearrange the new time series from 0 to $N - n$ (with the total number of data in the original time series) and compute the metrics, following the same steps as in Section 3. We repeat this process 10 times for each n, the take an average of the metrics for every corresponding node of the resulting networks.

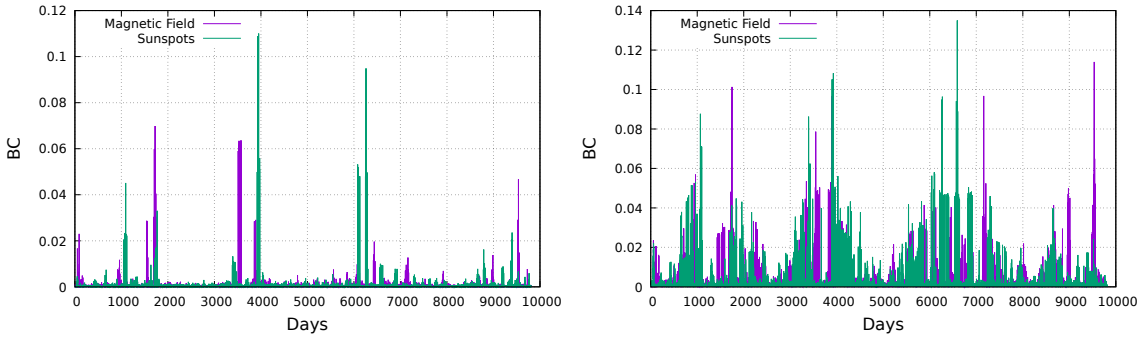
The plots suggest the method is robust against missing data, because we can observe the same 4 zones detected, each one corresponding to a different solar cycle.



(a) VG analysis

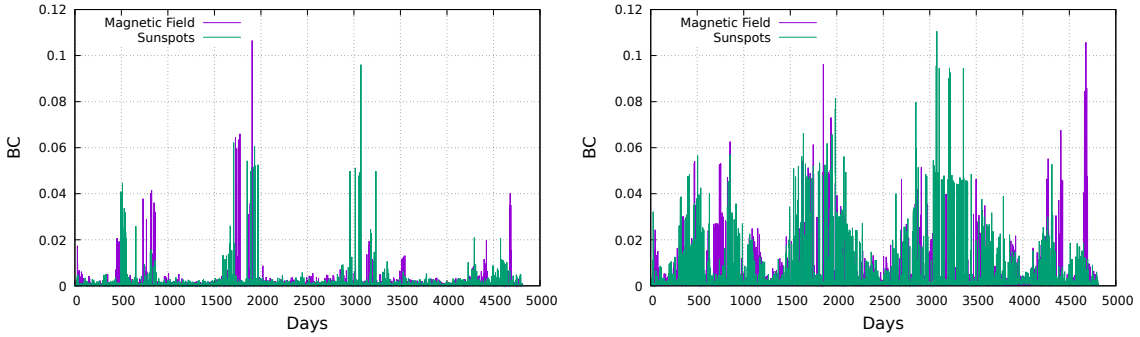
(b) HVG analysis

Figure 6.13: Randomly removing $N = 1000$ data of the time series.



(a) VG analysis (b) HVG analysis

Figure 6.14: Randomly removing $N = 5000$ data of the time series.



(a) VG analysis (b) HVG analysis

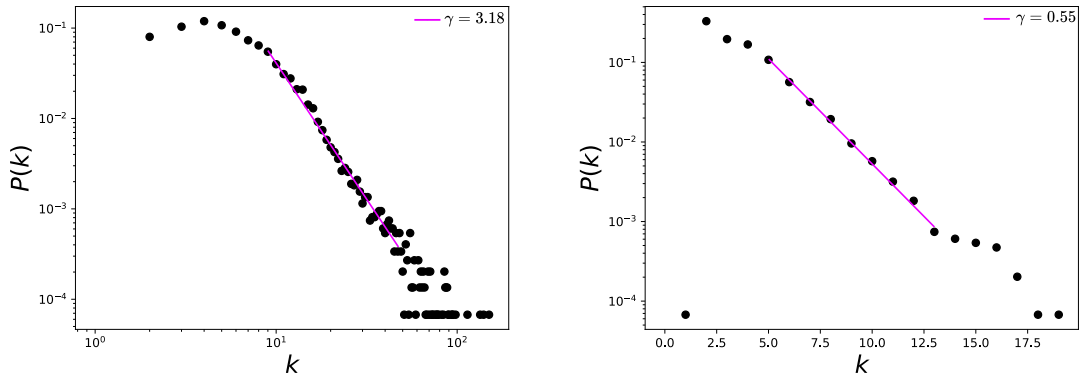
Figure 6.15: Randomly removing $N = 10000$ data of the time series.

Analogous to the previous section, we computed the degree distribution for the modified networks. This is shown in Figs. 6.16–6.21 and the γ exponents are shown in Table 6.2.

We obtained expected behaviors for each network, power-law for networks constructed using the visibility graph and exponential behaviors for networks constructed with the horizontal visibility graph. The decay exponents are affected by data removal, as shown in Table 6.2, but the overall topology is not changed by the removal.

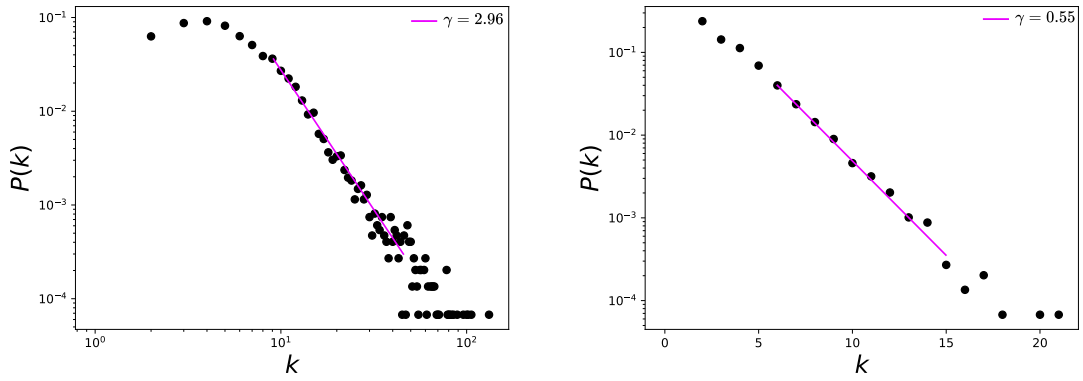
Table 6.2: γ values for the different removal intervals selected

		γ_{vg}	γ_{hvg}
Removing 1000 random data	Mean magnetic field	3.18	0.55
	Sunspots	3.05	0.85
Removing 5000 random data	Mean magnetic field	2.96	0.55
	Sunspots	3.0	0.74
Removing 10000 random data	Mean magnetic field	2.74	0.47
	Sunspots	2.64	0.59



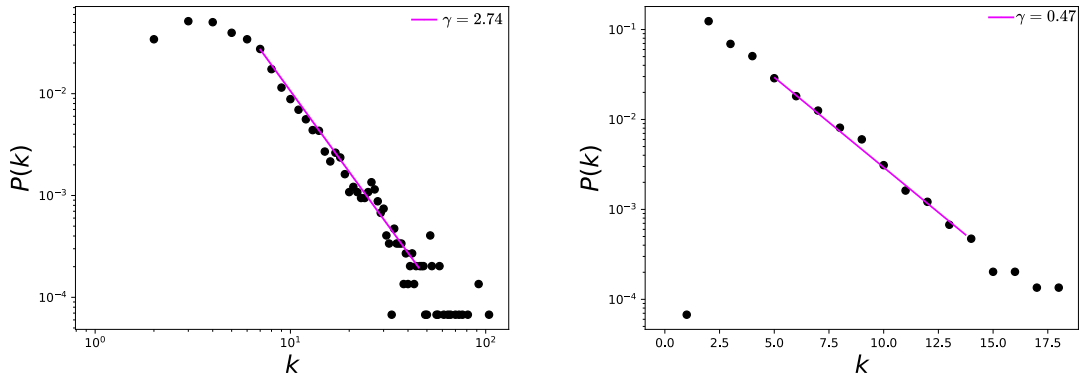
(a) VG analysis (b) HVG analysis

Figure 6.16: Degree distribution analysis associated to the magnetic field networks built by removing 1000 random data. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



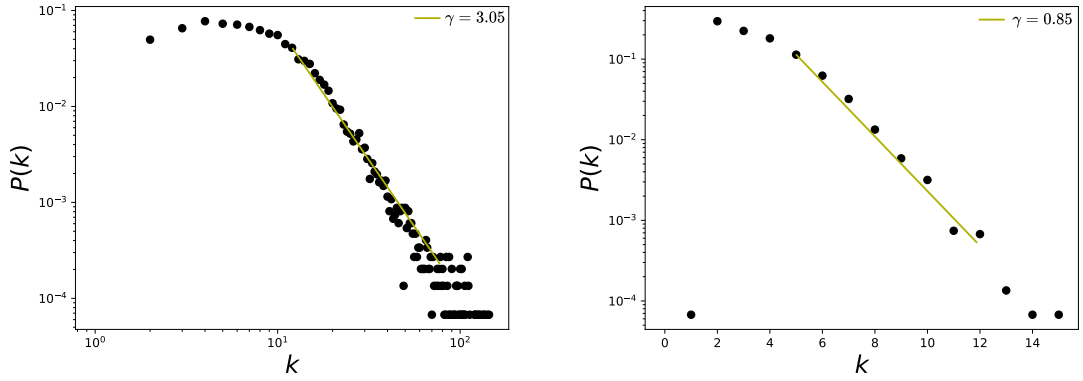
(a) VG analysis (b) HVG analysis

Figure 6.17: Degree distribution analysis associated to the magnetic field networks built by removing 5000 random data. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



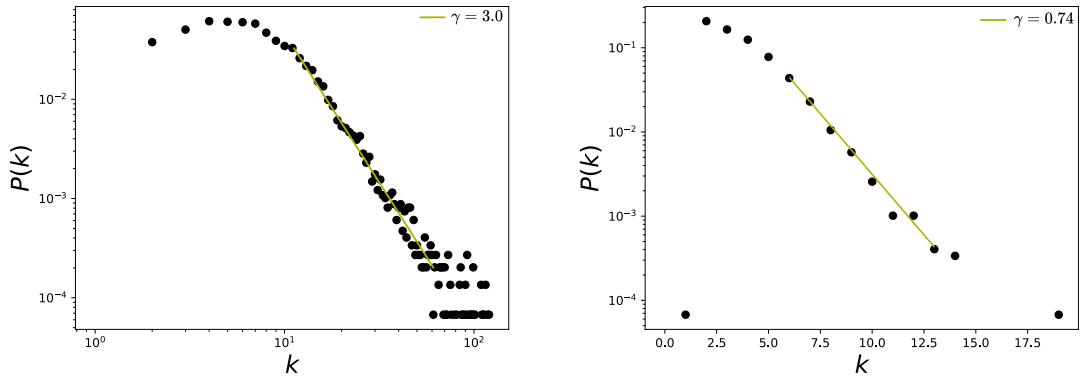
(a) VG analysis (b) HVG analysis

Figure 6.18: Degree distribution analysis associated to the magnetic field networks built by removing 10000 random data. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



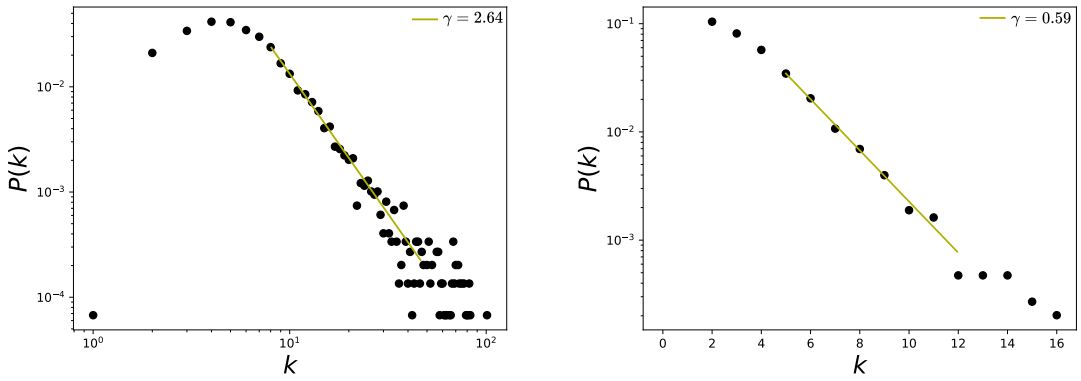
(a) VG analysis (b) HVG analysis

Figure 6.19: Degree distribution analysis associated to the sunspots networks built by removing 1000 random data. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



(a) VG analysis (b) HVG analysis

Figure 6.20: Degree distribution analysis associated to the sunspots networks built by removing 5000 random data. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.



(a) VG analysis

(b) HVG analysis

Figure 6.21: Degree distribution analysis associated to the sunspots networks built by removing 10000 random data. **Left** panel: Shows the power-law behavior for VG method; **right** panel: shows exponential behavior for HVG method.

Chapter 7

Discussion

In this work, we have studied and characterized solar activity using a complex network approach. By means of the visibility algorithms mentioned in Chapter 3, time series related to the Sun's dynamics, are mapped into a complex network. We have also tested the robustness of this method (see Chapter 6) and obtained consistent results with the ones shown in Chapter 4, confirming empirically that the method is particularly robust to time series that are not necessarily complete, either by observational periodic gaps or random missing data.

Various network metrics are calculated, which are related to node connectivity, edge density, distance between nodes, and node relative importance. In general, larger values of the degree are found for the VG as compared with the HVG (Figure 4.2). This is an expected result since the HVG has a limited visibility, restricted to horizontal lines, and therefore less connections can be established.

For the global analysis, using the full time series, the most interesting metrics were the centrality measures. From Figure 4.3, we can observe sensitivity of the betweenness centrality to the solar cycle both for VG and HVG (even more clear in Figure 4.18). This is a nontrivial result, because larger values of the time series would be expected to have more connections, because they should be more "visible"

to other nodes. However, the degree itself does not capture variations in solar activity, whereas the betweenness centrality, which is a more elaborate measure, clearly does.

The eigenvector centrality also shows a dependence on the solar activity, but of a different kind. First, behavior is different for both time series (magnetic field and sunspots), thus this is the only metric, among those studied here, that distinguishes the physical quantity being observed. Besides, for the VG, maxima tend to lie close to solar minima, whereas for the HVG they tend to lie in the ascending or descending phases of the cycle. It is also interesting to observe the small values obtained for SC24, with the VG method, representing almost non-influence in the network, consistent with the substantially lower activity of this cycle with respect to other recent solar cycles. The HVG, on the other hand, yields different results to the VG ones. Considering that EC tends to show maxima outside solar maxima, and that it distinguishes between sunspots and magnetic field times series, it should be interesting to study to what extent this measure is able to provide information on the next solar maximum, before it is actually reached. We plan to examine this in more detail in the future.

The results for the local analysis are, in general, consistent with the global analysis. Figures 4.7 and 4.8 show the expected result that the VG yields larger values for the degree than the HVG. It is also interesting to notice that the HVG degree shows a slight trend to decrease during its evolution, for the 1-year windows. However, one should take into account that values are normalized to the interval $[0, 1]$, and that the obtained values are very small $\sim 10^{-2}$, thus the degree could be regarded as essentially constant, regardless of the size of the time windows. However, a similar and clearer trend is observed for the HVG, if other metrics are considered.

Figure 4.9 shows that, whereas the degree is different for VG and HVG, the clus-

tering coefficient for the magnetic field and sunspots time series has about the same value, $\sim$0.75, for the VG method. On the other hand, the HVG method is able to pick variations associated to the solar cycle in the sunspots network. This sensitivity, though, is not present for the larger timescales, when the 11-year windows are used.

We have also observed interesting variations in the BC, for both the VG and HVG methods, with the larger scale time windows (11 years), as seen in Figure 4.12. This is consistent with the behavior found for the BC for the global analysis. One should consider, anyway, that calculated values are normalized to 1, and thus the variations shown in those figures are very small, of the order of 10^{-2}. In this sense, the behavior of the BC for the global analysis is much stronger, but the subtle variations in the local analysis may also be interesting, specially because they are consistent with the local analysis for the degree and the clustering coefficients, which did not exhibit any special dependence on solar activity in the global analysis.

In general, most curves shown in Figures 4.7–4.14 are featureless, with a few of them, as discussed above, showing noticeable variations which are consistent with the solar cycles. This is worth pointing out, because, although the sunspots and magnetic field time series clearly show variations in solar activity along solar cycles, and despite the interesting capabilities of the VG approach to identify statistical features in time series, it is interesting to point out when the VG can be most useful to study solar activity, and when it does not provide useful information.

The degree distributions are found to show an exponential behavior at the tail, as seen in Figure 4.6. The fast decay shows that on average, most nodes are connected to only a few nodes (degree probability is different from zero for $k < 6$). However, the mean path length is very small compared with the size of the network, suggesting a small-world behavior. Basically, the information within magnetic field and sunspots

networks is efficiently transferred toward the entirety of the system, locally and globally [83]. These results are preserved when the analysis is carried out in moving windows, as shown in Figures 4.15 and 4.16. For this latter analysis, we also observe an essentially constant value of the decay exponent despite variations in solar activity, as shown in Figure 4.17.

Despite simple metrics like the degree may not exhibit strong dependence with solar activity, more elaborate ones like the clustering coefficient and centrality measures may show clear variations with the solar cycle. The centrality measures are particularly interesting, due to the strong dependence of the BC for the global analysis, and the distinction between the magnetic field and sunspots time series that the EC displays. Further analysis should be carried out to determine to what extent these findings may contribute to characterize future solar cycles in advance, but our findings highlight the nontriviality of the information extracted by each metric, as results depend on the algorithm used, and the time scale examined, complementing other, recent works, on complex network analyses for solar activity [84–86]. In particular, we have previously observed that observing with different network metrics the same time series (sunspots number), various results can be found, with some metrics correlated, others anti-correlated, and other being essentially constant along the solar cycle [30]. The present work also complements these results. Our findings also show that different time series, although they may be related to the same underlying physics (solar dynamics), are not equivalent for the VG algorithm, which is consistent with the fact that one cannot expect a single technique to provide all the possible information on a given phenomenon. Besides, the correlation of certain metrics, for some timescales, with solar activity, opens the question of to what extent this correlation may be used to either characterize solar cycles, or inform us about

the dynamo process driving sunspots emergence and magnetic field variability along the solar cycle. In this sense, we have used several methods for correlations, including the ones presented in Chapter 5 and the quantile-quantile plot shown in Chapter 4, which is indeed another correlation analysis. These results have shown interesting results for the metrics and specially for the magnetic activity measured in terms of the betweenness centrality, given the fact that from both the quantile-quantile plot and the student t-test analysis support the qualitative correlation observed in graphs, and further exploration of this metric combined with prediction algorithms, could break new ground in the context and challenge of solar activity prediction. Therefore, our future interest would be the analysis of additional solar cycles, in order to understand in detail why some metrics perform better and their connections to physical features, beyond the results presented here.

Publications

The article derived from this thesis is listed below:

1. Zurita-Valencia, T.; Muñoz, V. Characterizing the Solar Activity Using the Visibility Graph Method. Entropy 2023, 25, 342. https://doi.org/10.3390/e25020342

Bibliography

[1] E. Cervi, A. Cammi, and E. Zio, "A new approach for nuclear reactor analysis based on complex network theory," *Progress in Nuclear Energy*, vol. 112, pp. 96–106, 2019. [Online]. Available: https://doi.org/10.1016/j.pnucene.2018.12.008

[2] R. Ding, N. Ujang, H. B. Hamid, M. S. A. Manan, R. Li, S. S. M. Albadareen, A. Nochian, and J. Wu, "Application of complex networks theory in urban traffic network researches," *Networks and Spatial Economics*, vol. 19, no. 4, pp. 1281–1317, 2019. [Online]. Available: https://doi.org/10.1007/s11067-019-09466-5

[3] A. Ghavasieh and M. De Domenico, "Statistical physics of network structure and information dynamics," *Journal of Physics: Complexity*, vol. 3, no. 1, p. 011001, 2022. [Online]. Available: https://doi.org/10.1088/2632-072X/ac457a

[4] J. Garcia-Algarra, G. G. Bengoechea, and M. L. Mouronte-López, "Reducing trade inequality: a network-based assessment," *Complexity*, vol. 2020, 2020. [Online]. Available: https://doi.org/10.1155/2020/1593215

[5] E. Zhuang, M. Small, and G. Feng, "Time series analysis of the developed financial markets integration using visibility graphs," *Physica A: Statistical Mechanics and its Applications*, vol. 410, pp. 483–495, 2014. [Online]. Available: https://doi.org/10.1016/j.physa.2014.05.058

[6] M.-C. Qian, Z.-Q. Jiang, and W.-X. Zhou, "Universal and nonuniversal allometric scaling behaviors in the visibility graphs of world stock market indices," *Journal of Physics A: Mathematical and Theoretical*, vol. 43, no. 33, p. 335002, 2010. [Online]. Available: https://doi.org/10.1088/1751-8113/43/33/335002

[7] G. Corso, L. Lucena, and Z. Thomé, "The small-world of economy: a speculative proposal," *Physica A: Statistical Mechanics and its Applications*, vol. 324, no. 1-2, pp. 430–436, 2003. [Online]. Available: https://doi.org/10.1016/S0378-4371(02)01883-6

[8] L. Bargigli, A. Lionetto, S. Viaggiu *et al.*, "A statistical equilibrium representation of markets as complex networks," *Journal of Statistical Physics*, vol. 165, pp. 351–70, 2013.

[9] C. Liu, X. Wu, R. Niu, X. Wu, and R. Fan, "A new SAIR model on complex networks for analysing the 2019 novel coronavirus (COVID-19)," *Nonlinear Dynamics*, vol. 101, no. 3, pp. 1777–1787, 2020. [Online]. Available: https://doi.org/10.1007/s11071-020-05704-5

[10] L. Li, J. Zhang, C. Liu, H.-T. Zhang, Y. Wang, and Z. Wang, "Analysis of transmission dynamics for zika virus on networks," *Applied Mathematics and Computation*, vol. 347, pp. 566–577, 2019. [Online]. Available: https://doi.org/10.1016/j.amc.2018.11.042

[11] L. Telesca, D. Pastén, and V. Muñoz, "Analysis of Time Dynamical Features in Intraplate Versus Interplate Seismicity: The Case Study of Iquique Area

(Chile)," *Pure and Applied Geophysics*, vol. 177, no. 10, pp. 4755–4773, 2020. [Online]. Available: https://doi.org/10.1007/s00024-020-02554-5

[12] S. Abe, D. Pastén, V. Muñoz, and N. Suzuki, "Universalities of earthquake-network characteristics," *Chinese Science Bulletin*, vol. 56, no. 34, pp. 3697–3701, 2011. [Online]. Available: https://doi.org/10.1007/s11434-011-4767-6

[13] D. Pastén, F. Torres, B. Toledo, V. Muñoz, J. Rogan, and J. A. Valdivia, "Time-based network analysis before and after the mw 8.3 Illapel earthquake 2015 Chile," in *The Chile-2015 (Illapel) Earthquake and Tsunami*. Springer, 2017, pp. 123–131. [Online]. Available: https://doi.org/10.1007/s00024-016-1335-7

[14] L. Orr, S. Chapman, J. Gjerloev, and W. Guo, "Network community structure of substorms using SuperMAG magnetometers," *Nature communications*, vol. 12, no. 1, pp. 1–10, 2021. [Online]. Available: https://doi.org/10.1038/s41467-021-22112-4

[15] E. Ser-Giacomi, V. Rossi, C. López, and E. Hernandez-Garcia, "Flow networks: A characterization of geophysical fluid transport," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 3, p. 036404, 2015. [Online]. Available: https://doi.org/10.1063/1.4908231

[16] T. Andriyas and S. Andriyas, "Periodicities in solar wind-magnetosphere coupling functions and geomagnetic activity during the past solar cycles," *Astrophysics and Space Science*, vol. 362, no. 9, pp. 1–14, 2017. [Online]. Available: https://doi.org/10.1007/s10509-017-3141-9

[17] Y. Singh *et al.*, "Study of the influence of magnetic fluctuations and solar plasma density on the solar wind–magnetosphere coupling," *Journal of*

atmospheric and solar-terrestrial physics, vol. 75, pp. 15–21, 2012. [Online]. Available: https://doi.org/10.1016/j.jastp.2011.05.005

[18] M. Moldwin, *An introduction to space weather*. Cambridge University Press, 2022. [Online]. Available: https://doi.org/10.1017/CBO9780511801365

[19] S. C. Chapman, R. B. Horne, and N. W. Watkins, "Using the index over the last 14 solar cycles to characterize extreme geomagnetic activity," *Geophysical Research Letters*, vol. 47, no. 3, p. e2019GL086524, 2020. [Online]. Available: https://doi.org/10.1029/2019GL086524

[20] G. Wrenn, "Chronology of 'killer'electrons: Solar cycles 22 and 23," *Journal of atmospheric and solar-terrestrial physics*, vol. 71, no. 10-11, pp. 1210–1218, 2009. [Online]. Available: https://doi.org/10.1016/j.jastp.2008.08.002

[21] P. I. Reyes, V. A. Pinto, and P. S. Moya, "Geomagnetic storm occurrence and their relation with solar cycle phases," *Space Weather*, vol. 19, no. 9, p. e2021SW002766, 2021. [Online]. Available: https://doi.org/10.1029/2021SW002766

[22] M. Domínguez, V. Muñoz, and J. A. Valdivia, "Temporal evolution of fractality in the Earth's magnetosphere and the solar photosphere," *Journal of Geophysical Research: Space Physics*, vol. 119, no. 5, pp. 3585–3603, 2014. [Online]. Available: https://doi.org/10.1002/2013JA019433

[23] M. J. Aschwanden and P. D. Aschwanden, "Solar flare geometries. i. the area fractal dimension," *The Astrophysical Journal*, vol. 674, no. 1, p. 530, 2008. [Online]. Available: https://doi.org/10.1086/524371

[24] ——, "Solar flare geometries. ii. the volume fractal dimension," *The Astrophysical Journal*, vol. 674, no. 1, p. 544, 2008. [Online]. Available: https://doi.org/10.1086/524370

[25] M. K. Georgoulis, "Are solar active regions with major flares more fractal, multifractal, or turbulent than others?" *Solar Physics*, vol. 276, no. 1, pp. 161–181, 2012. [Online]. Available: https://doi.org/10.1007/s11207-010-9705-2

[26] C. Price, D. Prichard, and E. Hogenson, "Do the sunspot numbers form a "chaotic" set?" *Journal of Geophysical Research: Space Physics*, vol. 97, no. A12, pp. 19 113–19 120, 1992. [Online]. Available: https://doi.org/10.1029/92JA01459

[27] A. Ruzmaikin, J. Feynman, and P. Robinson, "Long-term persistence of solar activity," *Solar physics*, vol. 149, pp. 395–403, 1994. [Online]. Available: https://doi.org/10.1007/BF00690625

[28] E. T. Lu and R. J. Hamilton, "Avalanches and the distribution of solar flares," *The astrophysical journal*, vol. 380, pp. L89–L92, 1991. [Online]. Available: https://doi.org/10.1086/186180

[29] V. Carbone, R. Cavazzana, V. Antoni, L. Sorriso-Valvo, E. Spada, G. Regnoli, P. Giuliani, N. Vianello, F. Lepreti, R. Bruno *et al.*, "To what extent can dynamical models describe statistical features of turbulent flows?" *EPL (Europhysics Letters)*, vol. 58, no. 3, p. 349, 2002. [Online]. Available: https://doi.org/10.1209/epl/i2002-00645-y

[30] V. Muñoz and E. Flández, "Complex Network Study of Solar Magnetograms,"

Entropy, vol. 24, no. 6, p. 753, 2022. [Online]. Available: https://doi.org/10.3390/e24060753

[31] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, "From time series to complex networks: The visibility graph," *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 4972–4975, 2008. [Online]. Available: https://doi.org/10.1073/pnas.0709247105

[32] R. Zhang, Y. Zou, J. Zhou, Z.-K. Gao, and S. Guan, "Visibility graph analysis for re-sampled time series from auto-regressive stochastic processes," *Communications in Nonlinear Science and Numerical Simulation*, vol. 42, pp. 396–403, 2017. [Online]. Available: https://doi.org/10.1016/j.cnsns.2016.04.031

[33] L. Lacasa and R. Flanagan, "Time reversibility from visibility graphs of nonstationary processes," *Physical Review E*, vol. 92, no. 2, p. 022817, 2015. [Online]. Available: https://doi.org/10.1103/PhysRevE.92.022817

[34] V. Suyal, A. Prasad, and H. P. Singh, "Visibility-graph analysis of the solar wind velocity," *Solar Physics*, vol. 289, no. 1, pp. 379–389, 2014. [Online]. Available: https://doi.org/10.1007/s11207-013-0332-6

[35] B. Acosta-Tripailao, D. Pastén, and P. S. Moya, "Applying the horizontal visibility graph method to study irreversibility of electromagnetic turbulence in non-thermal plasmas," *Entropy*, vol. 23, no. 4, p. 470, 2021. [Online]. Available: https://doi.org/10.3390/e23040470

[36] B. Acosta-Tripailao, W. Max-Moerbeck, D. Pastén, and P. S. Moya, "Assigning degrees of stochasticity to blazar light curves in the radio band using

complex networks," *Entropy*, vol. 24, no. 8, 2022. [Online]. Available: https://doi.org/10.3390/e24081063

[37] Y. Zou, M. Small, Z. Liu, and J. Kurths, "Complex network approach to characterize the statistical features of the sunspot series," *New Journal of Physics*, vol. 16, no. 1, p. 013051, 2014. [Online]. Available: https://doi.org/10.1088/1367-2630/16/1/013051

[38] A. Najafi, A. H. Darooneh, A. Gheibi, and N. Farhang, "Solar Flare Modified Complex Network," *The Astrophysical Journal*, vol. 894, no. 1, p. 66, 2020. [Online]. Available: https://doi.org/10.3847/1538-4357/ab8301

[39] V. Muñoz and N. E. Garcés, "Analysis of pulsating variable stars using the visibility graph algorithm," *Plos one*, vol. 16, no. 11, p. e0259735, 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0259735

[40] K. Labitzke, J. Austin, N. Butchart, J. Knight, M. Takahashi, M. Nakamoto, T. Nagashima, J. Haigh, and V. Williams, "The global signal of the 11-year solar cycle in the stratosphere: observations and models," *Journal of atmospheric and solar-terrestrial physics*, vol. 64, no. 2, pp. 203–210, 2002. [Online]. Available: https://doi.org/10.1016/S1364-6826(01)00084-0

[41] M. Vaughan, *The Fabry-Perot interferometer: history, theory, practice and applications*. Routledge, 2017. [Online]. Available: https://doi.org/10.1201/9780203736715

[42] M. Schüssler, "The formation of sunspots and starspots," *Astronomische Nachrichten*, vol. 323, no. 3-4, pp. 377–382, 2002.

[43] E. N. Parker, "The formation of sunspots from the solar toroidal field." *Astrophysical Journal, vol. 121, p. 491*, vol. 121, p. 491, 1955. [Online]. Available: https://doi.org/10.1086/146010

[44] C. Zwaan, "On the appearance of magnetic flux in the solar photosphere," *Solar Physics*, vol. 60, pp. 213–240, 1978. [Online]. Available: https://doi.org/10.1007/BF00156523

[45] "Sunspot data from the World Data Center SILSO, Royal Observatory of Belgium, Brussels," https://www.sidc.be/silso/datafiles.

[46] "The Wilcox Solar Observatory (WSO project)," wso.stanford.edu.

[47] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, vol. 433, no. 7024, pp. 392–395, 2005. [Online]. Available: https://doi.org/10.1038/nature03248

[48] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002. [Online]. Available: https://doi.org/10.1103/RevModPhys.74.47

[49] D. Pastén, F. Torres, B. A. Toledo, V. Muñoz, J. Rogan, and J. A. Valdivia, "Non-universal critical exponents in earthquake complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 445–452, 2018. [Online]. Available: https://doi.org/10.1016/j.physa.2017.09.064

[50] M. E. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003. [Online]. Available: https://doi.org/10.1137/S003614450342480

[51] M. Newman, *Networks*. Oxford university press, 2018. [Online]. Available: https://doi.org/10.1093/oso/9780198805090.001.0001

[52] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998. [Online]. Available: https://doi.org/10.1038/30918

[53] S. Wasserman and K. Faust, "Social network analysis: Methods and applications," 1994. [Online]. Available: https://doi.org/10.1017/CBO9780511815478

[54] J. Scott, *Social network analysis*. Sage Publications, 1992. [Online]. Available: https://doi.org/10.4135/9781529716597

[55] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977. [Online]. Available: https://doi.org/10.2307/3033543

[56] L. C. Freeman *et al.*, "Centrality in social networks: Conceptual clarification," *Social network: critical concepts in sociology. Londres: Routledge*, vol. 1, pp. 238–263, 2002. [Online]. Available: https://doi.org/10.1016/0378-8733(78)90021-7

[57] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Social networks*, vol. 30, no. 2, pp. 136–145, 2008. [Online]. Available: https://doi.org/10.1016/j.socnet.2007.11.001

[58] V. Latora and M. Marchiori, "A measure of centrality based on the network efficiency," *New Journal of Physics*, 2007. [Online]. Available: https://doi.org/10.1088/1367-2630/9/6/188

[59] K.-I. Goh, E. Oh, B. Kahng, and D. Kim, "Betweenness centrality correlation in social networks," *Physical Review E*, vol. 67, no. 1, p. 017101, 2003. [Online]. Available: https://doi.org/10.1103/PhysRevE.67.017101

[60] R. Guimera and L. A. N. Amaral, "Modeling the world-wide airport network," *The European Physical Journal B*, vol. 38, pp. 381–385, 2004. [Online]. Available: https://doi.org/10.1140/epjb/e2004-00131-0

[61] R. Guimera, S. Mossa, A. Turtschi, and L. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 22, pp. 7794–7799, 2005. [Online]. Available: https://doi.org/10.1073/pnas.0407994102

[62] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, "Horizontal visibility graphs: Exact results for random time series," *Physical Review E*, vol. 80, no. 4, p. 046103, 2009. [Online]. Available: https://doi.org/10.1103/PhysRevE.80.046103

[63] L. Lacasa, B. Luque, J. Luque, and J. C. Nuno, "The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion," *EPL (Europhysics Letters)*, vol. 86, no. 3, p. 30001, 2009. [Online]. Available: https://doi.org/10.1209/0295-5075/86/30001

[64] L. Lacasa, A. Nunez, É. Roldán, J. M. Parrondo, and B. Luque, "Time series irreversibility: a visibility graph approach," *The European Physical Journal B*, vol. 85, no. 6, pp. 1–11, 2012. [Online]. Available: https://doi.org/10.1140/epjb/e2012-20809-8

[65] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, "Recurrence networks—a novel paradigm for nonlinear time series analysis," *New Journal of Physics*, vol. 12, no. 3, p. 033025, 2010. [Online]. Available: https://doi.org/10.1088/1367-2630/12/3/033025

[66] J. Elsner, T. Jagger, and E. Fogarty, "Visibility network of united states hurricanes," *Geophysical Research Letters*, vol. 36, no. 16, 2009. [Online]. Available: https://doi.org/10.1029/2009GL039129

[67] A. M. Nuñez, L. Lacasa, J. P. Gomez, and B. Luque, "Visibility algorithms: A short review," *New frontiers in graph theory*, pp. 119–152, 2012. [Online]. Available: https://doi.org/10.5772/34810

[68] R. V. Donner and J. F. Donges, "Visibility graph analysis of geophysical time series: Potentials and possible pitfalls," *Acta Geophysica*, vol. 60, pp. 589–623, 2012. [Online]. Available: https://doi.org/10.2478/s11600-012-0032-x

[69] J. F. Donges, R. V. Donner, and J. Kurths, "Testing time series irreversibility using complex network methods," *Europhysics Letters*, vol. 102, no. 1, p. 10004, 2013. [Online]. Available: https://doi.org/10.1209/0295-5075/102/10004

[70] A. Nunez, L. Lacasa, E. Valero, J. P. Gómez, and B. Luque, "Detecting series periodicity with horizontal visibility graphs," *International Journal of Bifurcation and Chaos*, vol. 22, no. 07, p. 1250160, 2012. [Online]. Available: https://doi.org/10.1142/S021812741250160X

[71] M. G. Ravetti, L. C. Carpi, B. A. Gonçalves, A. C. Frery, and O. A. Rosso, "Distinguishing noise from chaos: objective versus subjective criteria using

horizontal visibility graph," *PloS one*, vol. 9, no. 9, p. e108004, 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0108004

[72] K. Schatten, D. J. Myers, and S. Sofia, "Solar activity forecast for solar cycle 23," *Geophysical research letters*, vol. 23, no. 6, pp. 605–608, 1996. [Online]. Available: https://doi.org/10.1029/96GL00451

[73] D. H. Hathaway, "Solar cycle forecasting," pp. 401–412, 2008. [Online]. Available: https://doi.org/10.1007/s11214-008-9430-4

[74] K. Petrovay, "Solar cycle prediction," *Living Reviews in Solar Physics*, vol. 17, no. 1, pp. 1–93, 2020. [Online]. Available: https://doi.org/10.1007/s41116-020-0022-z

[75] P. Erdös and A. Rényi, "On random graphs i," *Publicationes Mathematicae Debrecen*, vol. 6, p. 290, 1959.

[76] L. Lacasa and R. Toral, "Description of stochastic and chaotic series using visibility graphs," *Physical Review E*, vol. 82, no. 3, p. 036120, 2010. [Online]. Available: https://doi.org/10.1103/PhysRevE.82.036120

[77] L. Telesca and M. Lovallo, "Analysis of seismic sequences by using the method of visibility graph," *EPL (Europhysics Letters)*, vol. 97, no. 5, p. 50002, 2012. [Online]. Available: https://doi.org/10.1209/0295-5075/97/50002

[78] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964. [Online]. Available: https://doi.org/10.1021/ac60214a047

[79] R. Gnanadesikan and M. B. Wilk, "Probability plotting methods for the analysis of data," *Biometrika*, vol. 55, no. 1, pp. 1–17, 1968. [Online]. Available: https://doi.org/10.2307/2334448

[80] D. Pastén and C. Pavez-Orrego, "Multifractal time evolution for intraplate earthquakes recorded in southern norway during 1980–2021," *Chaos, Solitons & Fractals*, vol. 167, p. 113000, 2023. [Online]. Available: https://doi.org/10.1016/j.chaos.2022.113000

[81] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Physical review E*, vol. 65, no. 5, p. 056109, 2002. [Online]. Available: https://doi.org/10.1103/PhysRevE.65.056109

[82] D. Freedman, R. Pisani, and R. Purves, *Statistics*. Norton, 2007.

[83] M. Marchiori and V. Latora, "Harmony in the small-world," *Physica A: Statistical Mechanics and its Applications*, vol. 285, no. 3-4, pp. 539–546, 2000. [Online]. Available: https://doi.org/10.1016/S0378-4371(00)00311-3

[84] Y. Zou, R. V. Donner, N. Marwan, M. Small, and J. Kurths, "Long-term changes in the north–south asymmetry of solar activity: a nonlinear dynamics characterization using visibility graphs," *Nonlinear Processes in Geophysics*, vol. 21, no. 6, pp. 1113–1126, 2014. [Online]. Available: https://doi.org/10.5194/npg-21-1113-2014

[85] F. Daei, H. Safari, and N. Dadashi, "Complex network for solar active regions," *The Astrophysical Journal*, vol. 845, no. 1, p. 36, 2017. [Online]. Available: https://doi.org/10.3847/1538-4357/aa7ddf

[86] N. Lotfi, M. Javaherian, B. Kaki, A. H. Darooneh, and H. Safari, "Ultraviolet solar flare signatures in the framework of complex network," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 4, p. 043124, 2020. [Online]. Available: https://doi.org/10.1063/1.5129433

List of Equations

List of Figures