POLITECNICO DI TORINO

Corso di Laurea in Ingegneria Aerospaziale

Tesi di Laurea

Studio di un flusso turbolento isotropo mediante la teoria delle reti complesse

Study of an isotropic turbulent flow with the complex network theory

Relatore:

Ing. Stefania Scarsoglio

Candidato:

Davide Prino

Luglio 2016

Sommario

I fenomeni turbolenti sono largamente diffusi in natura e nei contesti tecnologici, essi non sono una prerogativa del solo campo aerospaziale, ma riguardano i più svariati rami dell'ingegneria. Per fare qualche esempio basti pensare al flusso d'acqua che esce dal nostro rubinetto di casa, alle correnti atmosferiche e oceaniche o al flusso d'aria che attraversa le pale di una turbina.

La prima esperienza scientifica sulla turbolenza risale alla fine del diaciannovesimo secolo e fu condotta da Osborne Reynolds. Essa consisteva nell'osservazione visiva dei moti di un filetto fluido all'interno di un tubo, ciò è realizzabile utilizzando un tubo trasparente collegato ad un serbatoio e servendosi di un colorante per distinguere un certo filetto fluido dal resto del liquido. Reynolds osservò che i fenomeni turbolenti si presentavano alle alte velocità, infatti la turbolenza è caratterizzata da un alto numero di Reynolds. A partire da questo pionieristico contributo sono stati fatti grandissimi passi in avanti nello studio della materia, attualmente si usano sia simulazioni sperimentali che numeriche.

Data la natura caotica del fenomeno l'approccio naturale per lo studio di questa materia è di tipo statistico, tuttavia questi tipi di studio possono rivelarsi estremamente complessi. Nel seguente lavoro si presenta un approccio differente: utilizzando il visibility algorithm si trasforma una serie temporale (in questo caso l'evoluzione temporale dell'energia cinetica) in un grafo. Successivamente, attraverso lo studio di alcune caratteristiche topologiche di questo grafo, si cerca di arrivare ad una interpretazione fisica del fenomeno. Questo lavoro si basa sull'ipotesi che alcune caratteristiche fisiche del sistema si trasmettono al grafo. Il grande vantaggio di uno studio di questo tipo è che si possono ricavare alcune informazioni sulla fisica del problema utilizzando un algoritmo di semplice implementazione.

Nel capitolo 1 si descrive in generale il fenomeno della turbolenza, descrivendone le principali caratteristiche e dando particolare importanza alla turbolenza omogenea e isotropa.

Nel capitolo 2 si riporta una introduzione sulla teoria dei grafi, si danno le principali definizioni e notazione e si descrivono gli indici che verranno poi utilizzati nella trattazione.

Nel capitolo 3 si parla del visibility algorithm e si fornisce una possibile implementazione con Matlab, cercando di sottolineare i passaggi critici (il codice completo è riportato in appendice A).

Nel capitolo 4 si descrive il database dal quale si attingono i dati per il seguente studio, si tratta del John Hopkins Turbulence Databases. Esso è costituito di più parti, ma viene descritta più nel dettaglio esclusivamente la parte che concerne una simulazone numerica di un flusso turbolento omogeneo ed isotropo. Si forniscono infine le coordinate dei punti che saranno successivamente presi in considerazione.

Nel capitolo 5 si studiano le caratteristiche dei grafi ricavati dai due punti descritti nel capitolo precedente, infine si comparano i risultati ottenuti per i due diversi punti e si cerca di dare una interpretazione fisica ai dati ottenuti.

Contents

Introduction

Turbulent phenomena are largely diffused both in nature and technological context, they are not a prerogative of the aerospace area, but they concern so many branches of engineering. Some examples can be: water flow that goes out from a water tap, atmospheric and oceanic currents and air flow through the blades of a turbine.

The first scientific experience on turbulence was made by Osborne Reynolds in the nineteenth century. Using glass pipes connected to a reservoir, he observed the flow pattern in the pipes at various speeds of water by injecting a thin steam of dye in the main stream. He found out that at low velocities the dye filament travelled straight and parallels to the walls of the tube, instead at high velocities it diffuses over the whole depth and loses its identity. The former is the laminar condition, whereas the latter is the turbulent condition. In fact, turbulence is characterised by high Reynolds number. Since this pioneering contribution, considerable developments were done in understanding the nature of turbulence, nowadays physical and numerical simulation are largely used in the study of this subject.

Due to the chaotic nature of the phenomenon, a statistical approach is the natural method for studying turbulence, but these studies can be extremely difficult. In this work another kind of approach is presented: using the visibility algorithm a time series (in this case the temporal evolution of the kinetic energy) is transformed in a graph, then, studying some topological features of the graph, a physical interpretation of the phenomenon is given. This work is based on the hypothesis that some physical characteristics of the system are transferred to the graph. The great advantage of this kind of study is simplicity of visibility algorithm.

Chapter 1 is an introduction on turbulence: its main properties are described and particular importance is given to the isotropic and homogeneous turbulence.

Chapter 2 deals with the graph theory: at first the main definitions and notations are given, then the indices useful for the forward analysis are described.

In chapter 3 the visibility algorithm is defined and a possible implementation with Matlab is presented, giving particular emphasis to the crucial points (complete Matlab code is reported in appendix A).

In chapter 4 the John Hopkins Turbulence Databases are described, they are composed by many parts, but only the numerical simulation of an isotropic and homogeneous turbulent flow is described in detail. Then the coordinates of the two points subsequently studied are given.

In chapter 5 the characteristics of the two graphs obtained from the two points described in the previous chapter are studied. Eventually obtained results are compared and a physical interpretation is given to these data.

Chapter 1

Turbulent Flows

When the Reynolds number exceeds a certain limit, disturbances are amplified. When amplification is large, disturbances break down into chaotic motion. Repeated breakdown of disturbances leads to completely chaotic sustained motion, commonly known as turbulence.
It is difficult to give a precise definition of turbulence. Hinze defined turbulence as follows: "Turbulent fluid motion is an irregular condition of flow in which various quantities show a random variation with time and space coordinates so that statistically distinct averages can be discerned"[2].
The lateral movement of fluid particles in the case of laminar flow is due to molecular diffusion, so it is very small. Instead, in turbulent flows, this movement can be caused also by the presence of an eddy, which is a large group of fluid particles characterised by high value of vorticity. During its life an eddy can change its shape, stretch and rotate or brake into two or more eddies.
If one were to focus attention at a point in flow field, passage of small and large eddies through this point would induce velocity fluctuations of small magnitude and large frequency, and large magnitude and small frequency.
The turbulent flow is described by the Navier-Stokes equations, but the deterministic solution of this problem is very hard to find due to the presence of non linearity, so it is usually studied with statistical and computational methods or, like in the present work, with the complex networks theory.

1.1 Characteristics

Turbulence is a chaotic phenomenon, which has some important characteristics:

- The presence of fluctuations both spatial and temporal, which make the flow three-dimensional and not steady;
- Three-dimensional behaviour even if the initial conditions are two-dimensional. Transition from laminar to turbulent flow starts with small two-dimensional vibrations, that grow and become vortices, which warp and move due to mutual induction;
- The non linear motion become more important for high Reynolds number, in fact the convective term, which is non linear, assumes more relevance and the viscous term is not able to dull the fluctuations. Due to this characteristic a small perturbation in the initial conditions can involve a big one in the solution, which become bigger and bigger while time is going on.
- The presence of three-dimensional fluctuations of vorticity;
- This phenomenon has a wide range of scales. The biggest eddies have the dimension of the overall turbulent region and they contain the major part of the kinetics energy, while the dimension of the smallest ones depends on the Reynolds number, the higher is Re the

smaller are the eddies, in fact the ratio between the two scales has the same order of magnitude of the Reynolds number. Thanks to the vortex stretching there is a transfer of energy from the bigger scales to the smaller scales, until the gradient of velocity becomes so high (because of the conservation of the angular momentum) that the viscous stresses cause the dissipation of kinetics energy;

- The diffusivity is very strong because fluctuations cause a rapid mixing of mass, momentum and heat, higher than in the laminar case, in which the diffusivity is only due to molecular agitation.
- The dissipation is strong too, because of the viscous stresses in the small eddies.

To sum up turbulence is a chaotic flow characterized by three-dimensional vorticity, non linearity and high values of diffusivity and dissipation.

1.2 Kolmogorov's hypotheses

Kolmogorov formulated three hypotheses:

1. First Kolmogorov's hypothesis (local isotropy): for high enough Reynolds number the turbulent movement of the small scales ($l \ll l_0$, where l_0 is the length scale of largest eddies) are statistically isotropic;
2. Second Kolmogorov's hypothesis (first hypothesis of similarity): for high enough Reynolds number the statistical physical quantities of small scales ($l \ll l_{EI}$) have a universal form defined by viscosity (ν) and energy traded between the big and the small scales (ε);
3. Third Kolmogorov's hypothesis (second hypothesis of similarity): for high enough Reynolds number the statistical physical quantities of the scale l with $l_{DI}<l<l_{EI}$ have a universal form defined only by ε.

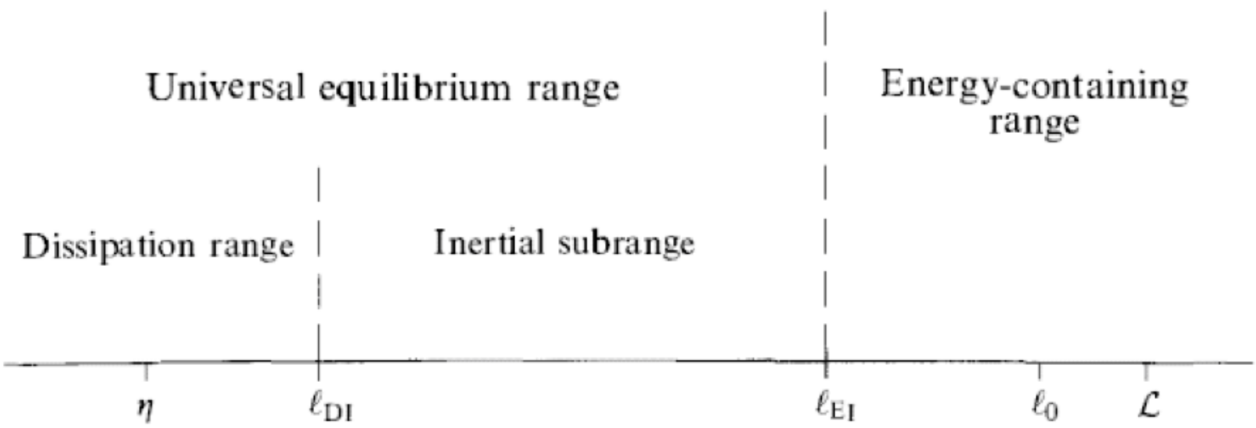

Figure 1.1: The representation of the ranges of the Kolmogorov's hypotheses

Generally, the largest eddies are anisotropic, but they lose this feature during the trade of energy between the big and the small scales, this phenomenon happens in the inertial range ($l_{DI} \ll l \ll l_{EI}$). They lose also information about their geometry, so the motion of the small scales is universal, which means that it is similar in every flows that have high Re. If ν and ε are known, dimensional scale, scale of velocity and temporal scale (Kolmogorov's scales) can be defined, so that the dimensionless quantities can be built. These quantities are statistically the same for every turbulent flow with high Re. Peculiarity of inertial range are intermediate eddies, that are big enough not to be influenced by viscosity, but small enough to be in the range of universality.

1.3 homogeneous and isotropic turbulence

The homogeneous and isotropic turbulence is a flow in which there are no interactions with solid objects and there is no a mean flow of velocity imposed from the outside.
A turbulent flow is homogeneous if the velocity field doesn't change due to a translation of the reference system and it is isotropic if it doesn't change due to a rotation of the reference system.
It is physically realized letting a fluid pass, with a constant flow, through a grid. If Reynolds number is high enough, beyond the grid the flow becomes turbulent and it can evolve in a free space, generating the homogeneous and isotropic turbulence. Feeding the turbulence for a long time requires a source of energy because of the strong dissipation of the turbulence. The production of energy is concentrated in the big scales and the dissipation in the small ones. In this case the source is represented by the flow through the grid.

Chapter 2

Graph theory

Many real systems from very different fields, for example chemical systems, social interacting species, the World Wide Web, are composed by a large number of highly interconnected units. They can be represented by a graph, where the units are drawn as points (nodes) and the interactions between them as edges. The analysis of many different networks has produced an important result: real networks have a series of unifying principles and statistical properties in common, for example the degree distribution can exhibits a power law, they have relatively short path between any two nodes (small-world property) and the presence of a large number of short cycles or specific motifs. The present work is based on the hypothesis that some physical behaviours are reflected in the topology of the network, so the analysis of the graph can be a useful alternative to a more classical statistical study for knowing more about turbulence.

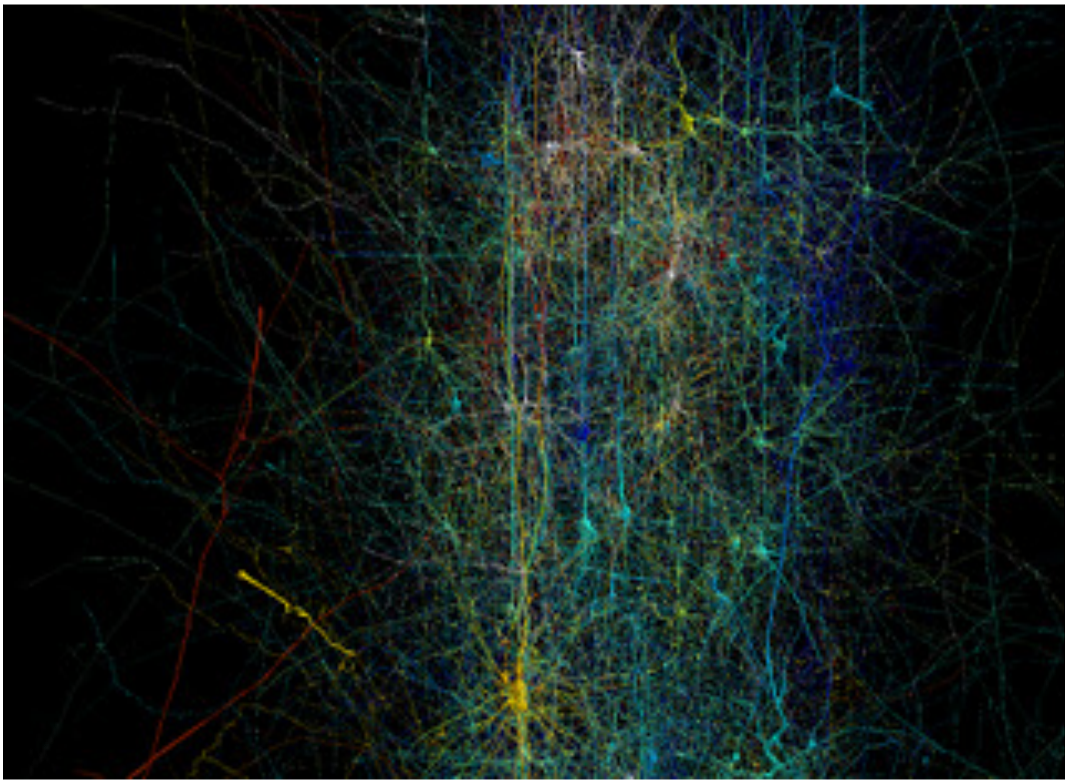

Figure 2.1: a neural network made of 100 neurons

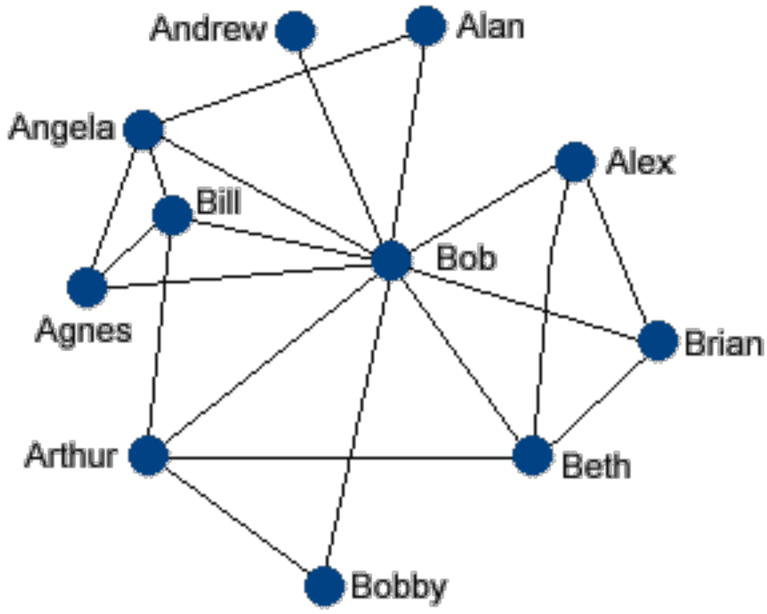

Figura 2.2: An example of social network.

2.1 Definitions and notations

An unweighted (i.e. all the links have the same values, there is no priority in correlations) and undirected (i.e. the links don't have direction) network can be regarded as a graph G=(N,E), where N is a set of nodes and E a set of edges.
The graph can't contain loops, i.e. links from a node to itself, or multiple edges, i.e. more than one connection between two nodes.
Two nodes joined by a link are addressed as adjacent or neighbouring. An important concept is the reachability of two different nodes, a walk from node i to node j is a sequence of adjacent nodes that begins with i and ends with j. Its length is the number of edges in the sequence. A path is a walk in which no node is visited more than once, the shortest path between two nodes is an important characteristic of the network.
A graph is "connected" if there is a path for each pair of nodes.
A subgraph G'=(N', E') of G is a graph such that N' is contained in N and E' in E. If G' contains all the edges of G that join two nodes in N', G' is called "subgraph induced by N'". The subgraph of the neighbours of a given node i (G_i) is the subgraph induced by the set of the nodes adjacent to i (N_i).
It can be used a matricial representation of a graph, it consists in a N x N matrix (the adjacency matrix) whose entry a_{ij} assumes the value 1 if the nodes i and j are joined, else $a_{ij} = 0$. For undirected graph it is a symmetrical matrix, and all the terms on the diagonal, a_{ii}, are zero, because of the absence of loops.

Another important matrix for the characterization of the network is matrix D, in which the entry d_{ij} is the length of the shortest path from node i to node j. It is symmetrical for undirected graph.

2.2 Metrics and indices [4]

The **degree** k_i of a node i is the number of edges incident with the node.

$$k_i = \sum_{j \in N} a_{ij} \quad (2.1)$$

The **average degree** of a network is:

$$\langle k_i \rangle = \frac{1}{N} \sum_i k_i = \frac{1}{N} \sum_{ij} a_{ij} \quad (2.2)$$

The **degree distribution** P(k) is the probability that a node chosen uniformly at random has degree k, so it is the fraction of nodes in the graph having degree k.
The **shortest path** is the length of the minimum walk between two nodes, without visiting the same node more than once. It plays an important role in the transport of information and communication within a network, since the transfer of information is faster if nodes are connected with short path.
The **diameter** is the maximum value of shortest paths:

$$D = \max_{ij} d_{ij} \quad (2.3)$$

The **average shortest path** (or characteristic path length) is the average number of edges along the shortest path for all possible nodes in the network:

$$\langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \quad (2.4)$$

This definition diverges if there are disconnected points in the network. This causes an issue, and a way to avoid the divergence is to introduce the efficiency E, which is an indicator of the traffic capacity of a network.

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (2.5)$$

The **betweenness centrality** b_i is a measure of the relevance of a node, in fact the communication between two non-adjacent nodes depends on the paths connecting them, so the importance of a given node can be obtained counting the number of shortest path going through it and calculating b_i (the more the path passing through the node, the more importance is given to the node itself).

$$b_i = \sum_{j \neq k} \frac{n_{jk}(i)}{n_{jk}} \quad (2.6)$$

where n_{jk} is the number of shortest paths connecting j and k, while $n_{jk}(i)$ is the number of shortest path connecting j and k and passing through i.
The **clustering coefficient** (or transitivity) measures the probability that two nodes, which are neighbour of a given node i, are linked together. This means the presence of triangles.

$$c_i = \frac{2e_i}{k_i(k_i-1)} = \frac{\sum_{j,m} a_{ij}a_{jm}a_{mi}}{k_i(k_i-1)} \qquad (2.7)$$

Where e_i is the number of edges in G_i (the subgraph induced by the node i) and $k_i(k_i-1)/2$ is the maximum number of edges that G_i can have. a_{ij} and a_{mi} verify that nodes j and m are both neighbours of i, then a_{jm}=1 if j and m are joined by an edge.
It can be useful to calculate the **average clustering coefficient**.

$$C = \frac{1}{N}\sum_N c_i \qquad (2.8)$$

The **modularity** of a network is a measure of the structure of a complex network for detecting communities. A high value of modularity indicates a strong division into groups. Nodes belonging to the same community have a lot of connections between them, this implies a faster rate of transmission of information.

Chapter 3

Visibility algorithm

3.1 General properties

The visibility algorithm is a simple computational method to convert time series into graphs. Every point of the time series becomes a node of the associated graph and two nodes (t_a, y_a) and (t_b, y_b) are connected if any other data (t_c, y_c) between them fulfil the following propriety:

$$y_c < y_b + (y_a - y_b)\frac{t_b - t_c}{t_b - t_a} \qquad (3.1)$$

$(y_a - y_b)/(t_b - t_a)$ is the gradient of the straight line that joins the two data a and b. So a and b are linked if there are not any data between them with a value above this line, (the visibility line does not intersect any intermediate data height).

In figure 3.1 is represented the time series y=sin(t), where t is a vector with ten values between 0 and 2π, with the same gap between two adjacent values. On this figure the edges that fulfil the visibility algorithm are also drawn. The edges on the right of the fourth node are of different colours, in order to create less confusion. Then, in figure 3.2, is reported the graph of the series made with Gephi. This is an open-source and free software for visualization and exploration of all kinds of graphs and networks. It can be downloaded at the link below[1].

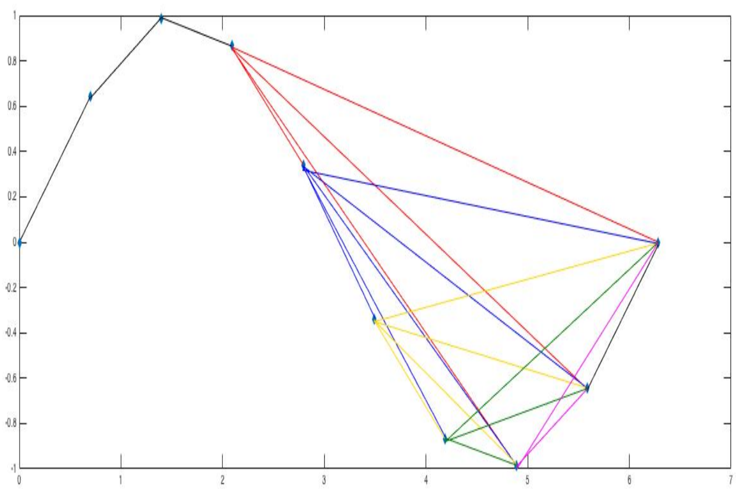

Figure 3.1: the time series y=sin(t), with links drawn in order to fulfil the visibility algorithm.

[1] https://gephi.org

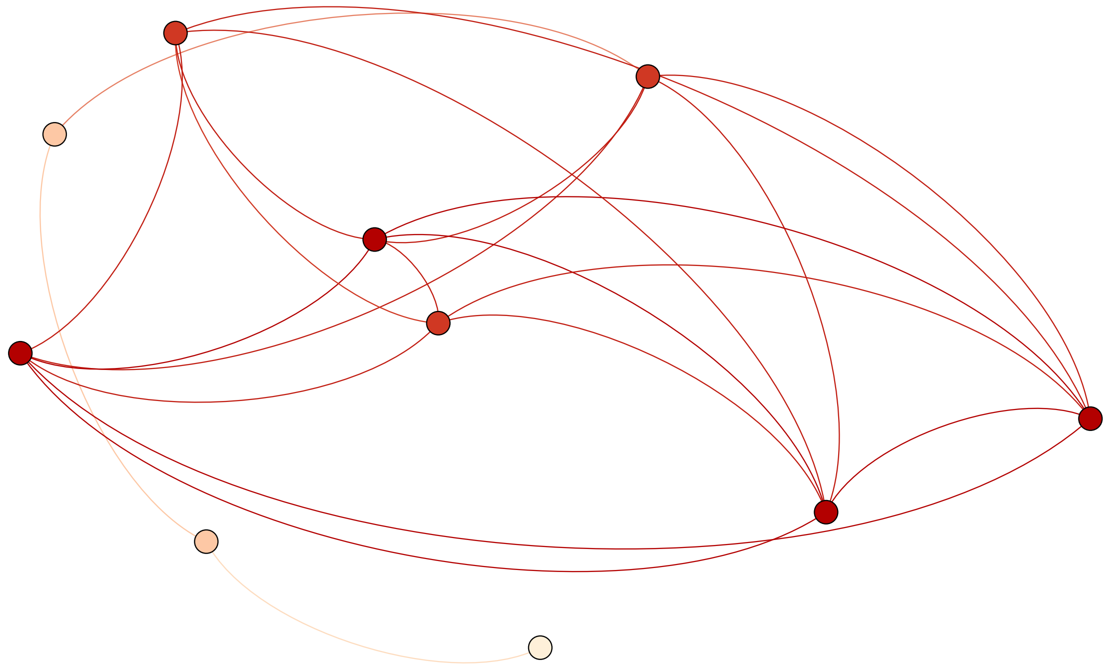

Figure 3.2: the graph of the series y=sin(t), made with Gephi.

3.2 Properties and implementation

In this work the implementation of the visibility algorithm to build adjacency matrix A is done with Matlab, (macro is reported in appendix A). This program also calculates some important indices: average degree, degree distribution, average path length, diameter and clustering coefficient.
To build the adjacency matrix is important to consider some properties of the visibility algorithm, that make the calculation faster.
First of all, the graph obtained with this method is undirected, so the adjacency matrix is symmetric, it can be built only the triangular matrix (in this case the superior triangular matrix is built). Then there are no loops, so the values on the diagonal are all zeroes. The graph is always connected, in fact each node sees at least its nearest neighbours, left and right, this means that all the terms of the adjacency matrix on the left or on the right of the diagonal are 1. For these reasons the "for cycle" that sounds out the line of the matrix starts from i=1 and ends with i=(N-1), where N is the number of nodes and the size of the matrix, in fact A is initialized to zero and the only term of the last line that needs to be calculated is the one on the diagonal, which is 0. Instead the "for loop" that sounds out the columns starts with j=i+2, the terms on the diagonal and its nearest terms are known a priori.
To verify if a pair of nodes are joined the introduction of an auxiliary variable c is needed, at first it assumes the value 1, that means that the two nodes are supposed to be connected, then a third "for cycle" controls if there is no visibility between the nodes, if so c is put equal to zero and the

cycle is interrupted. The reason of the choice of initializing c to 1 is the fact that two nodes can be supposed to be neighbours unless the contrary in proved, but is not true the vice versa.
At the end of all of these iterations the adjacency matrix is calculated.
The next step of the macro is to build matrix D that contains the length of the path between all pairs of nodes. This can be done referring to the fact that the entry (i, j) of matrix A^k is equal to the number of walks of length k from node i to node j. Initially D is posed equal to A, in fact for every pair of nodes joined by an edge the length of the path is 1, then a "for loop" from k=2 to k=(N-1), which is the maximum length possible for a path, analyzes one by one the matrices A^k to find all the entries that fulfil the following properties:

- $D_{i,j} = 0$, if it is not zero it means that a path between the nodes i and j was already found;
- $A_{i,j}^k \neq 0$, if this is true there is at least a walk of length k between the nodes i and j;

if both the properties are respected it can be said that $D_{i,j} = k$.
It is important to proceed increasing k, and not decreasing it, because it must be found the length of the path, which is the shortest walk that joins a pair of nodes.
Another issue of the program is to calculate the average path length and the diameter of the graph, which are respectively the mean and the maximum value of terms of matrix D.
The degree of the node i is calculated summing all the entries of a line or a column of matrix A, the average degree is simply the mean value of the degree of all the nodes.
The maximum degree is also calculated because it is useful for the calculation of the degree distribution, which is a vector whose number of components is equal to this value. The first "for cycle" defines the value i of the degree, while the second one sounds up the vector that contains the degree of all the nodes to find the number of nodes which has degree equal to i.
The local clustering coefficient c_i is calculated using the equation (2.7). It is important to pay attention to the fact that if the degree of the node i is 1, c_i diverges, but if a node has only one neighbour its c_i is 0, this is why Matlab code uses an "if" command to control if the degree is equal to 1.
Eventually all the graphics are plotted: time series, degree distribution, degree of each node and local clustering coefficient. The degree distribution is managed using functions "histogram" and "ksdensity", that evaluate the density estimate at rispectively 20 and 50 points covering the range of the data.

Chapter 4

JHTDB database

4.1 Description of the database

The data used in the present work are taken from the *Johns Hopkins Turbulence Databases* (JHTDB). It contains four kind of turbulence fields: a homogeneous and isotropic turbulence, a forced magneto-hydro-dynamic turbulence (MHD), a channel flow and a homogeneous buoyancy driven turbulence. All these flows are obtained from a direct numeric simulation (DNS); this means that the Navier-Stokes equations are solved numerically without the use of a turbulence model. All the temporal and spatial scales are solved, from the integral scales to the dissipative scales, starting from appropriate initial conditions and boundary conditions. This method is the simplest approach to the turbulence problem, but it has some limitations. Due to the complexity of the Navier-Stokes equations it can not be used to solve every kind of flow. Then it needs very high calculation power, making the simulation very expensive, but, in the last years, thanks to the development of technology which leads to producing powerful computers, it is largely used in research.
From JHTDB database the velocity components, the pressure and the magnetic field can be downloaded.

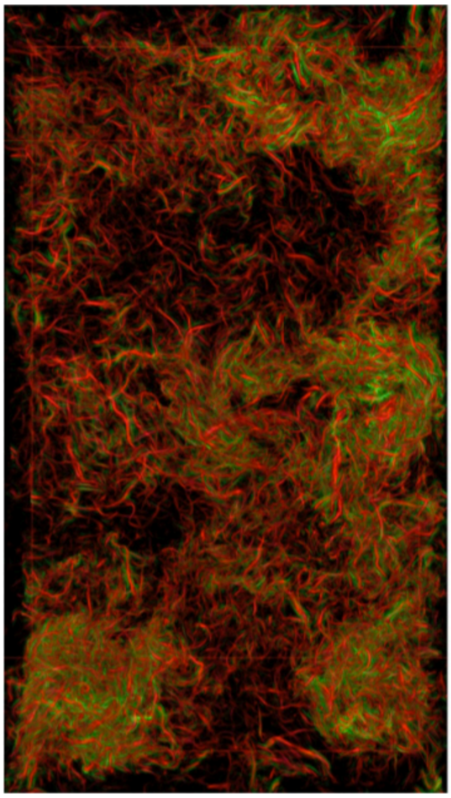

Figure 4.1: a representation of vorticity of the forced turbulent flow, take from the JHTDB web site

The domain of the field is a cube 2π x 2π x 2π and the velocity or pressure values are reported on a spatial grid with 1024^3 nodes. The viscosity is ν = 0.000185. The timestep is Δt = 0.0002, but the storing time interval is 0.002, finally the simulation time interval is t ∈ [0,2048], obtaining the 1024 timesteps available, the total time of the simulation is about one large eddy turnover time.
The main characteristics of the isotropic field are reported in the table below:

Distance between nodes	$d = \frac{2P}{1024} = 6.142\times10^{-3}$
Total kinetic energy	$E_{tot} = \langle \sum_k \left(\frac{1}{2}\boldsymbol{u}\cdot\boldsymbol{u}^*\right)\rangle = 0.695$
Dissipation	$\langle \sum_k \nu k^2 \boldsymbol{u}\cdot\boldsymbol{u}^*\rangle = 0.0928$
Taylor micro-scale	$\lambda = \sqrt{\frac{15\nu u^2}{\varepsilon}} = 0.118$
Taylor-scale Reynolds number	$Re_\lambda = \frac{u\lambda}{\nu} = 433$
Kolmogorov lengthscale	$\eta = \left(\frac{\nu^3}{\varepsilon}\right)^{1/4} = 0.00287$
Kolmogorov timescale	$\tau_\eta = \left(\frac{\nu}{\varepsilon}\right)^{1/2} = 0.0446$
Integral scale	$L = \frac{\pi}{2u^2}\int \frac{E(k)}{k}dk = 1.376$
Large eddy turnover time	$T_L = {}^{L}/_{u'} = 2.02$

Table 4.1: characteristic of the turbulent flow

To obtain data from the database an authorization token is needed, obtainable asking the permission to the database administrator, this is useful for knowing the number of users of the different services available.
On the web site there is a simple interface from which the user can chose one of the four kind of flow available, the field (pressure or velocity), the starting coordinates (temporal and spatial) and the size of the cutout.
Files downloaded from the database are saved in Hierarchical Data Format (extension .h5).

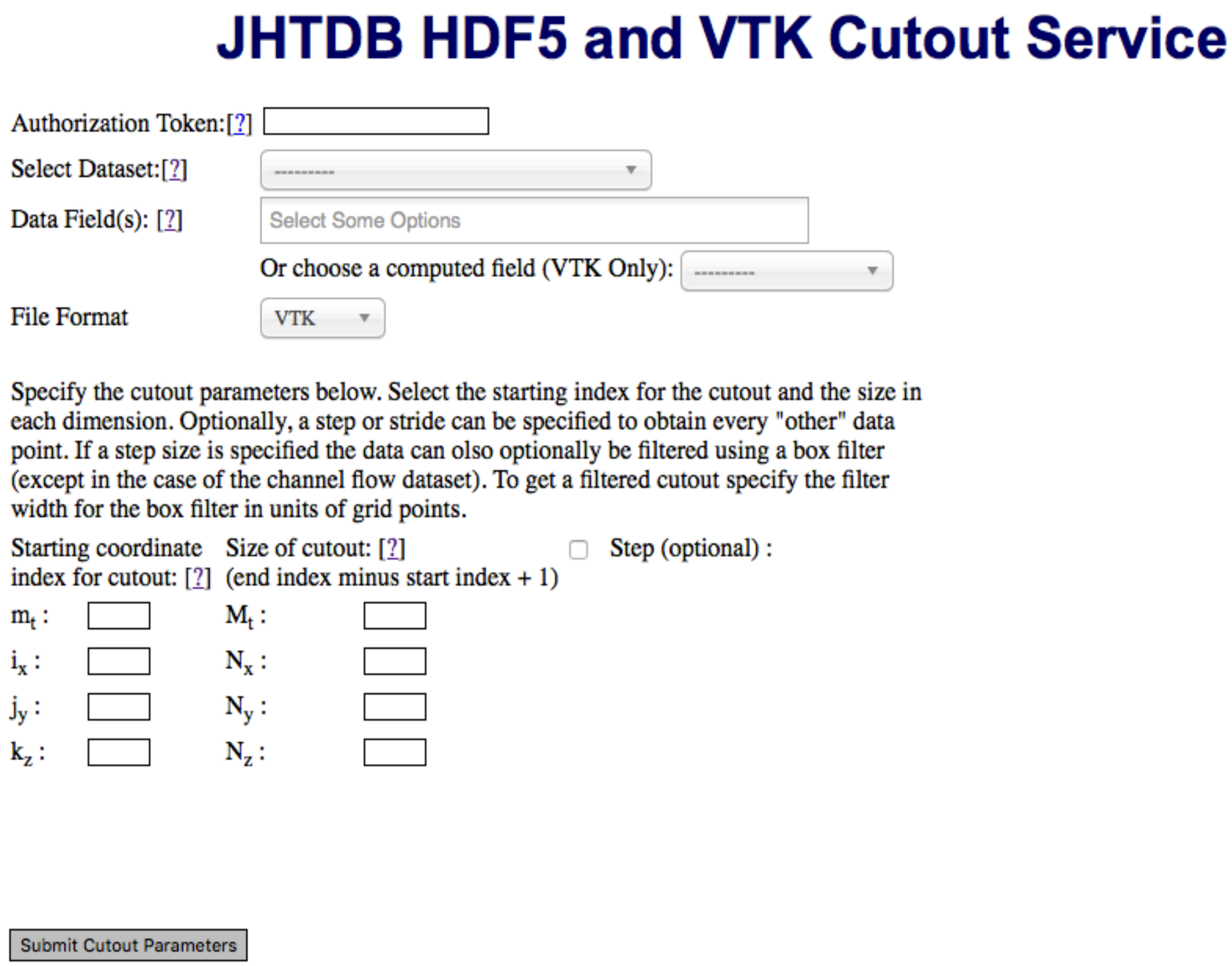

Figure 4.2: the interface of data-cutout service

4.2 Obtaining data

In this work the time series of two points of a homogeneous and isotropic turbulence are analyzed. From the database are taken all the 1024 time values of the velocity, then the kinetic energy is calculated:

$$K = \frac{1}{2}(u^2 + v^2 + w^2) \quad (4.1)$$

where u, v, w are the three components of the velocity vector.

The choice of the two points is based on another work[2], in which a cutout of the isotropic turbulence of the JHTDB database is analyzed with the complex network theory, using a procedure that converts the spatio-temporal data in a spatial network. In particular, this work focus on the study of two points of the network: a node with high degree centrality value (HDC) and a node with low

[2] “Nuove osservazioni sulla caratterizzazione spaziale di flussi turbolenti attraverso la teoria delle reti complesse” by Giovanni Iacobello, magistral thesis.

degree centrality value (LDC). The coordinates are (385, 401, 508) and (372, 387, 510) respectively for HDC and LDC node. High values of k_i indicate regions with similar instantaneous vorticity, this means that there are turbulent patterns coherently moving over the acquired time scale T_L. The domain of this study is a sphere with radius equal to Taylor micro-scale, that is the characteristic dimension of the smallest dynamically significant eddies of the flow. It is the intermediate length scale at which fluid viscosity significantly affects the dynamic of turbulent eddies of the flow. Length scales which are larger than Taylor micro-scale, the integral range, are not strongly influenced by viscosity, while below the Taylor micro-scale there is the dissipation range.

Chapter 5

Network analysis and physical interpretation

In this chapter the two networks built with the visibility algorithm applied to time series of the kinetic energy of the HDC and the LDC nodes are analyzed, using the indices described in chapter 2. Distinct features of a time series can be mapped onto networks with distinct topological features, so the resulting networks inherit characteristics of the time series and this series can be investigated from a complex network perspective. Thanks to this assumption a physical interpretation can be given to the results obtained.

5.1 HDC network analysis

Starting from the HDC node, below is reported the graphic of the time series.

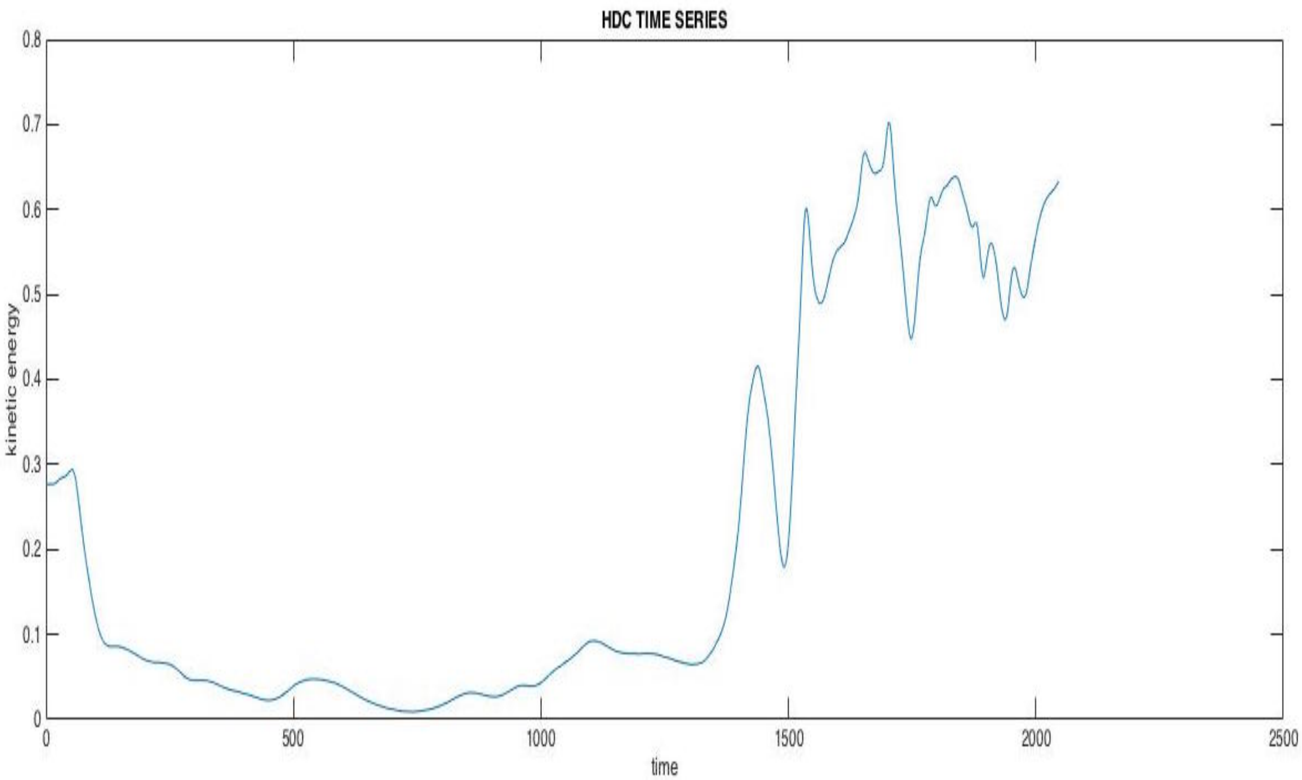

Figure 5.1: the graphic of the HDC energy time series

The time series is converted in a graph with the visibility algorithm, then the graph is drawn with Gephi (figure 5.2). Nodes change their colour from white to red due to their degree (red means high degree).

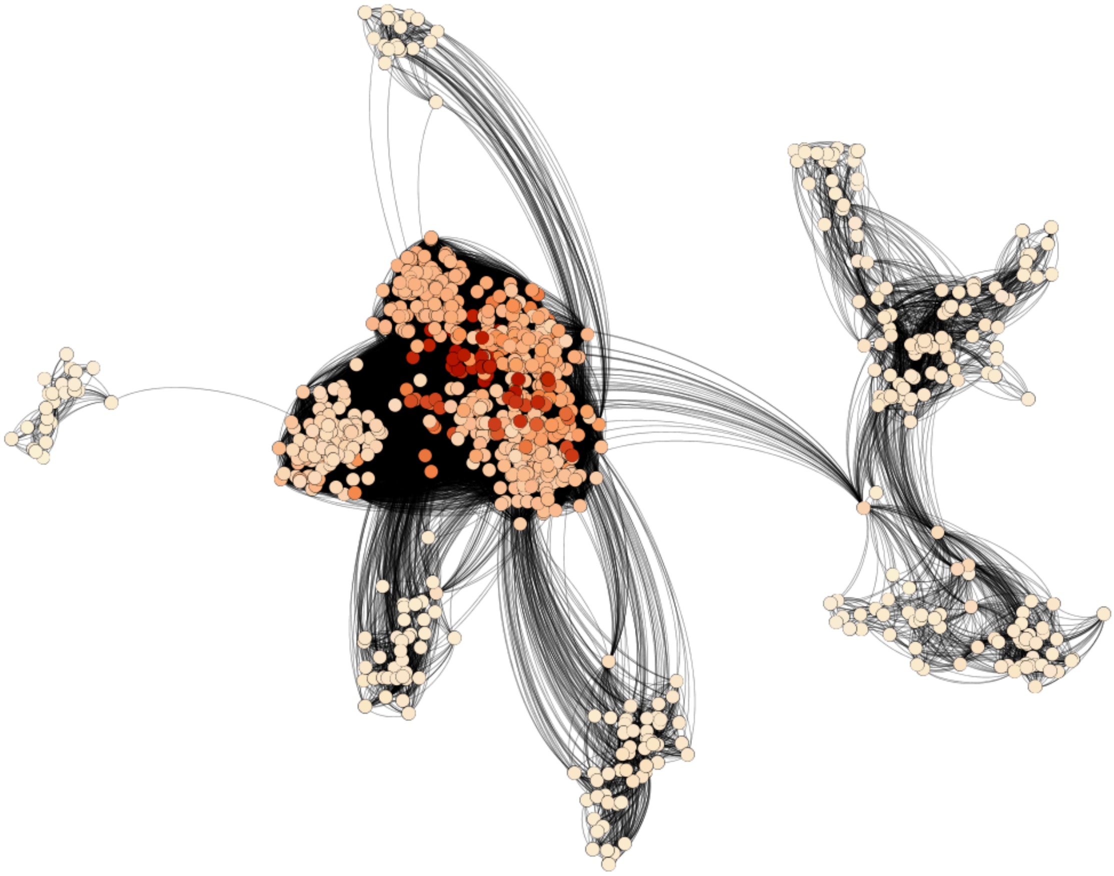

Figure 5.2: graph of the time series of HDC point

The average degree of the graph is <k> = 159.176
Degree distribution is reported in figure 5.3. It is evident that the majority of nodes have a degree minor than 250, while there are only a few nodes with very high degree (more than 600).

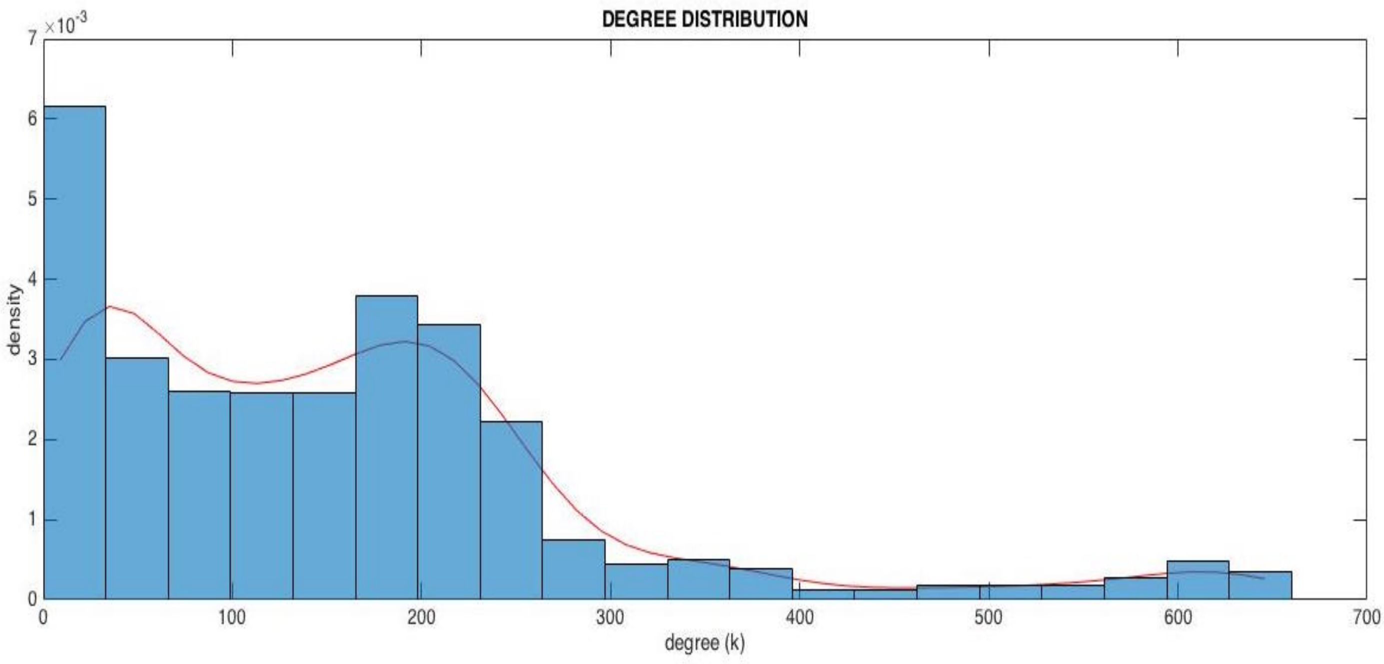

Figure 5.3: degree distribution P(k) of HDC point

In figure 5.4 is represented the degree associated to every node of the graph. It is important to remember that the visibility algorithm converts every point of the time series into a node and there is a gap of 2 between two consecutive times stored. The number associated to a node is the value of time related to the node itself (at time 2 corresponds node 2, at time 4 the node 4, etcetera...).

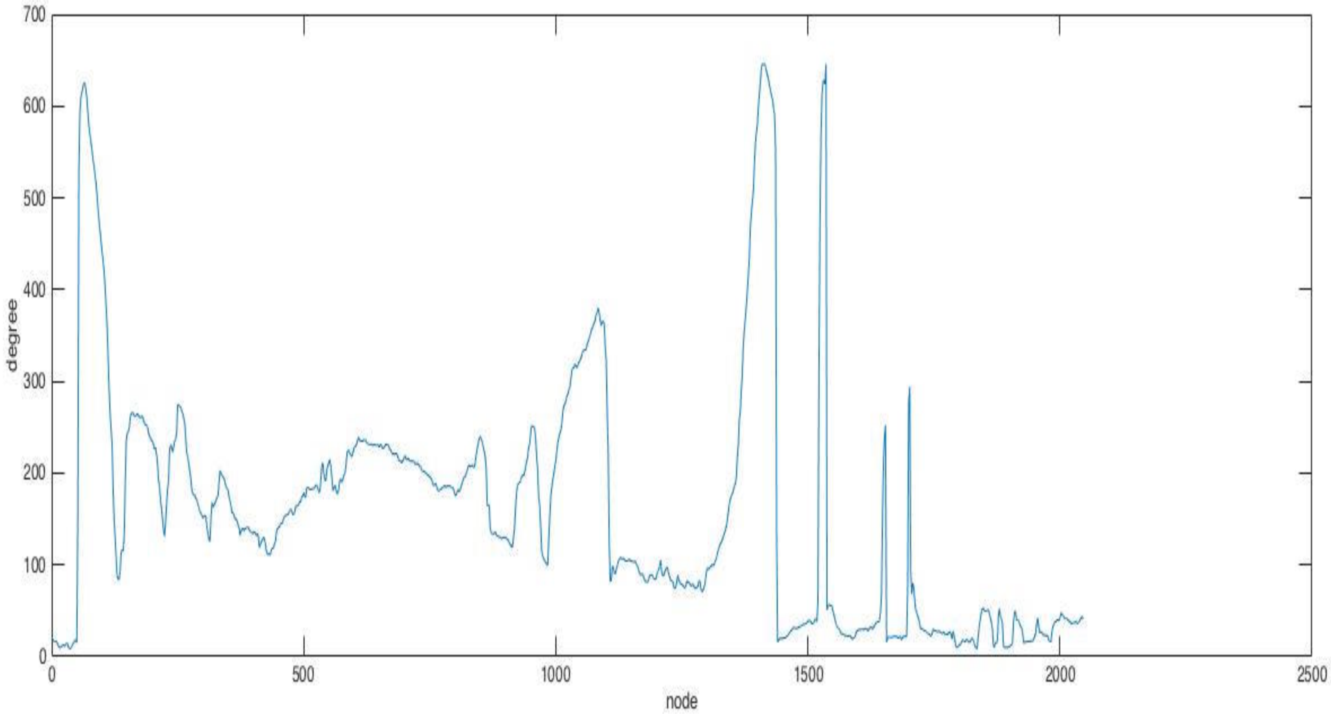

Figure 5.4: a graphic where is reported the value of the degree for every node of the graph

Comparing this graphic with figure 5.1 it can be seen that the peaks are situated at the same points. This is not surprising, considering the fact that a peak of the time series has a large visibility on the other nodes and covers the nodes next to it, in fact after a peak on this graphic there is a strong decrease of degree.

The diameter and the average shortest path are respectively D = 7 and <d> = 2.602, it is not necessary to introduce the efficiency, due to the fact that the graph does not have disconnected points, so <d> is not divergent.

The average clustering coefficient is <c> = 0.642 and the local clustering coefficients are reported in figure 5.5, they have very variable values. There is an alternation of regions with high values of c_i and regions with low values. A low value means that the nodes are more independent one from each other, while the presence of a large number of triangles involves having a strong transitivity. The strong variability of this index probably involves a strong division in cluster of the graph.

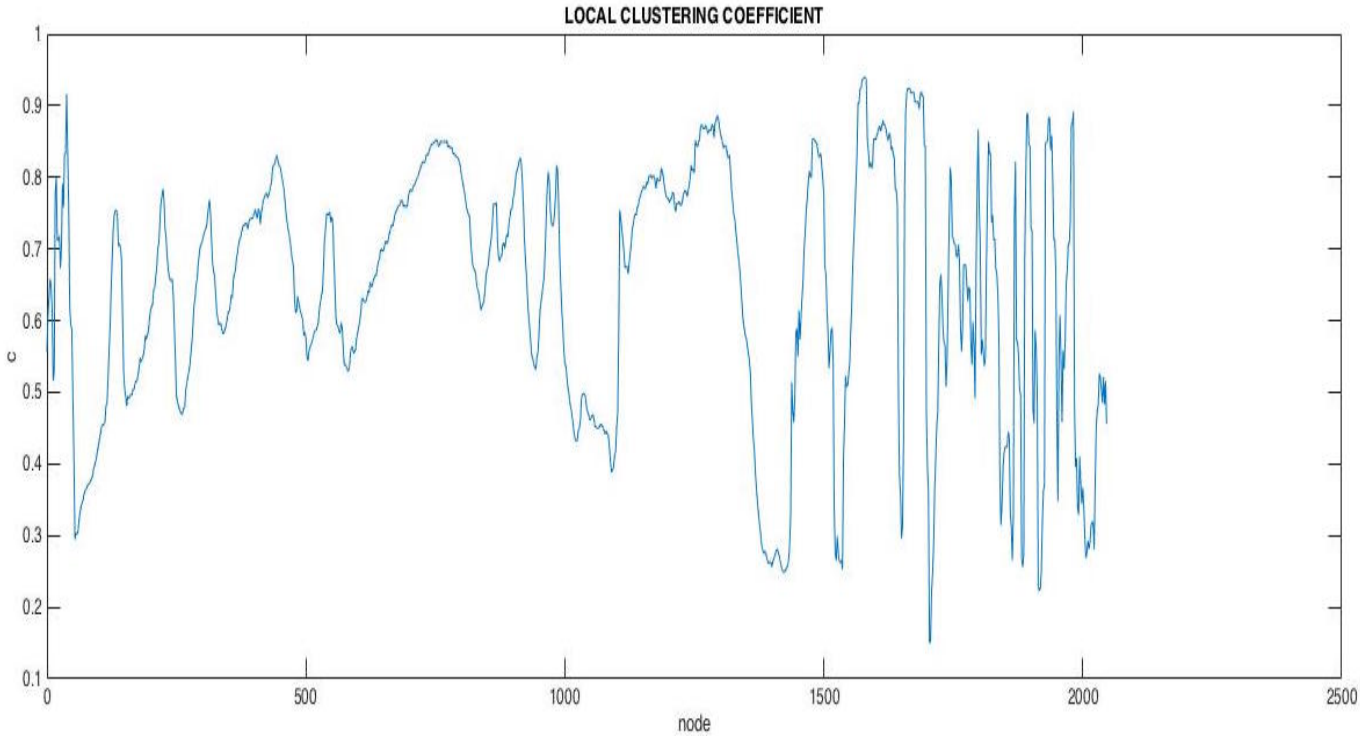

Figure 5.5: a graphic where is reported the local clustering coefficient for every node of the graph

In order to have a better knowledge of the division in communities of the graph, the modularity is studied with Gephi: Q = 0.321

There are 9 communities. In the graphic below is reported how many nodes each community contains.

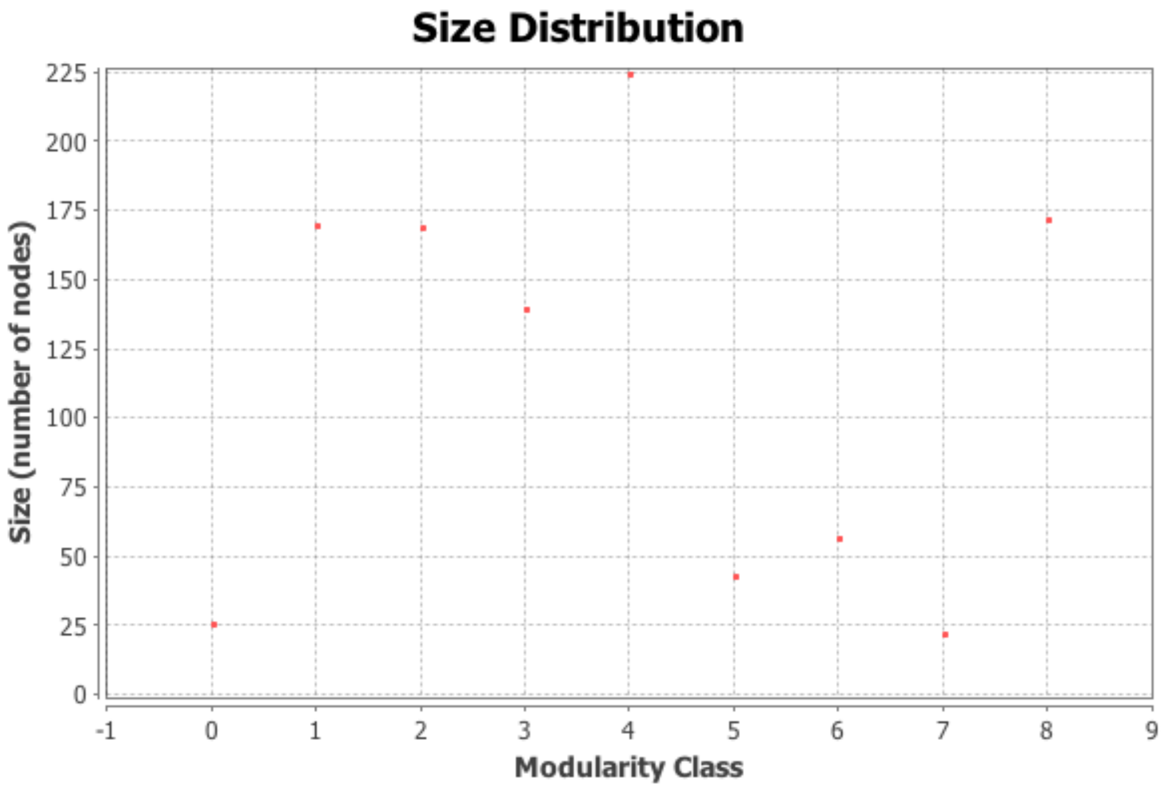

Figure 5.6: a graphic of the size of every community of the graph

5.2 LDC network analysis

The time series and the correlated graph are reported below.

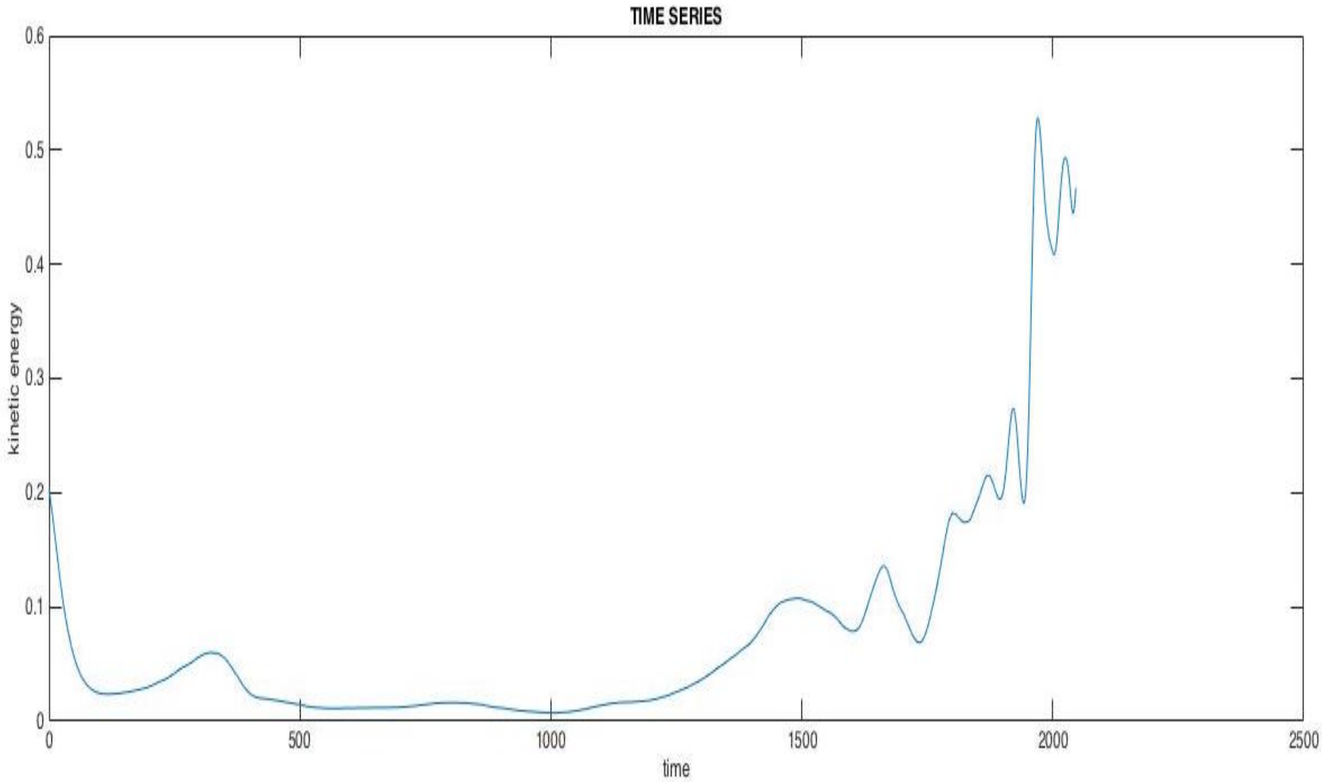

Figure 5.7: the graphic of the LDC energy time series

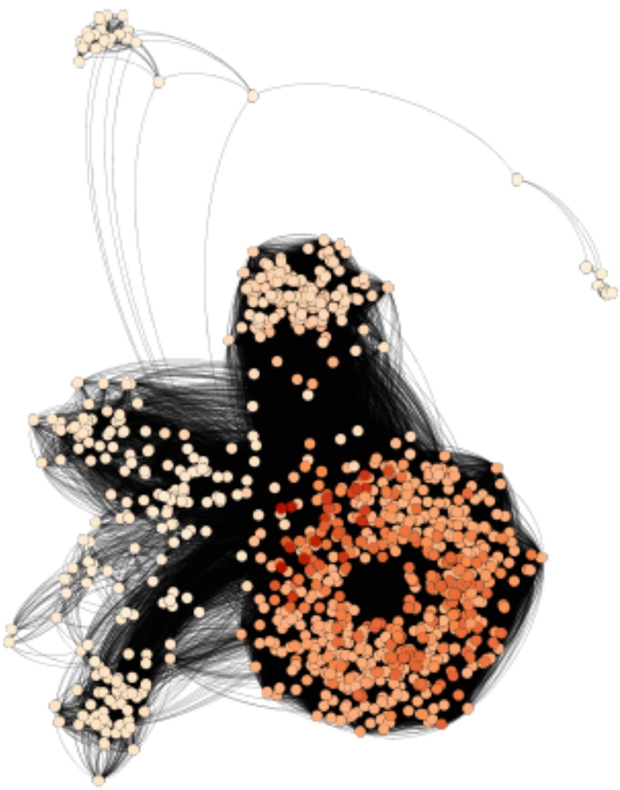

Figure 5.8: graph of the time series of LDC point

The average degree of the graph is <k> = 277.16
In figure 5.9 is reported the degree distribution. There are a lot of nodes with a value of degree less than 200. There is also a large group of nodes with degree near 300. Very few nodes have a very high degree (more than 800).

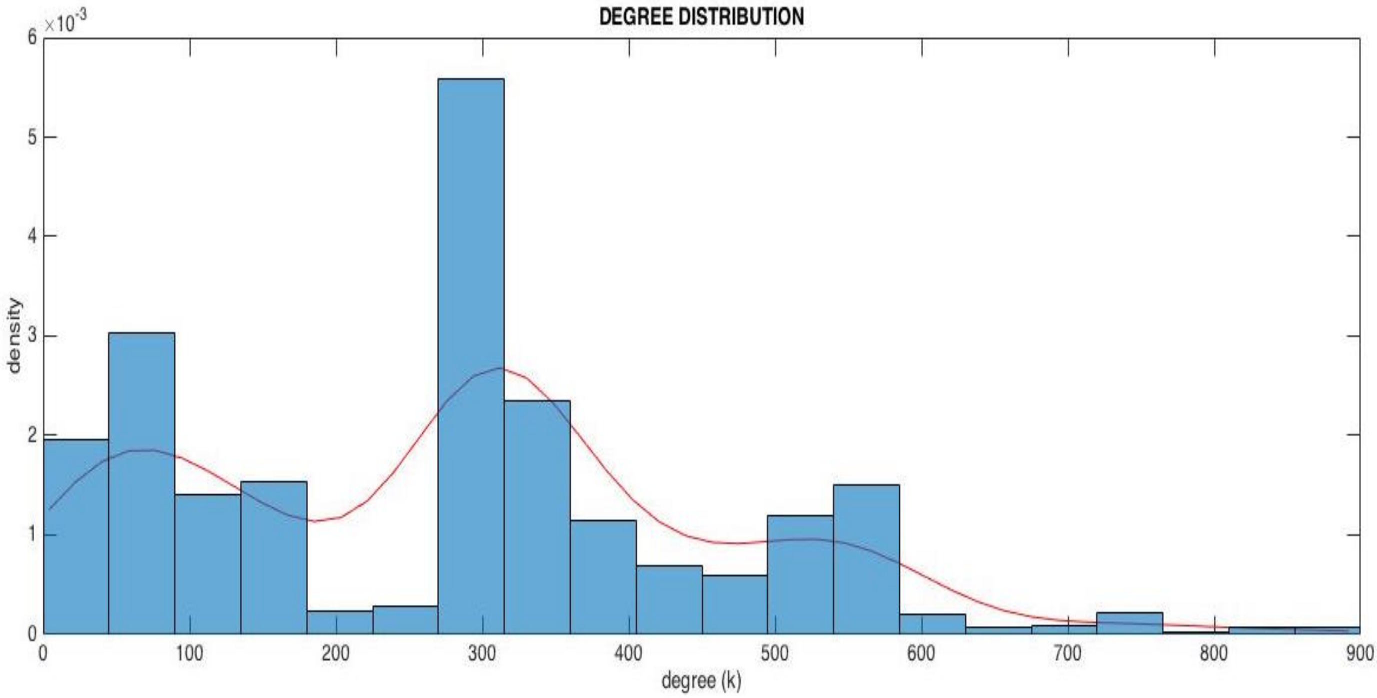

Figure 5.9: degree distribution P(k) of LDC point

The next graphic associates at every node its own degree. Considerations on this figure are the same of the ones made in the previous chapter.

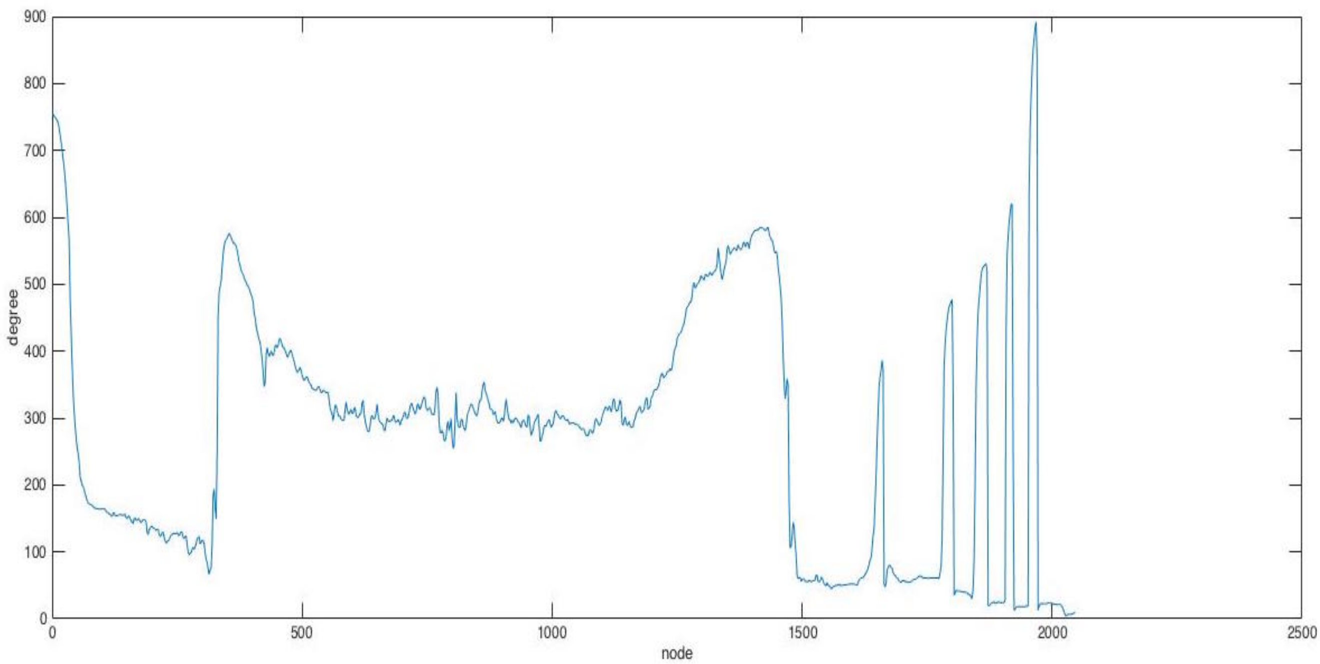

Figure 5.10: a graphic where is reported the value of the degree for every node of the graph

The diameter and the average shortest path are respectively D = 7 and <d> = 1.875
The average clustering coefficient is <c> = 0.687 and the local clustering coefficients are reported in figure 5.11, the graphic has quite a continuous form in the first part, then, in the last part, the values are more variable.

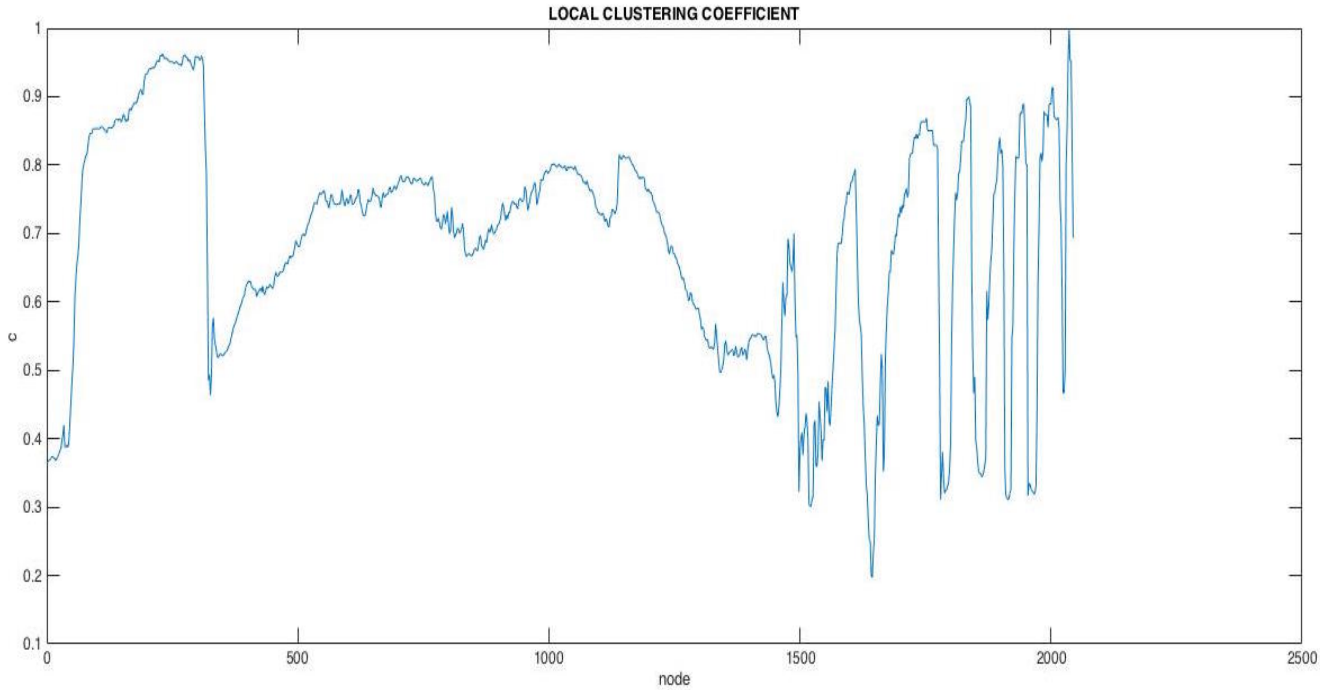

Figure 5.11: a graphic where is reported the local clustering coefficient for every node of the graph

The modularity is Q = 0.26
There are 5 communities, in the graphic below is reported how many nodes each community contains.

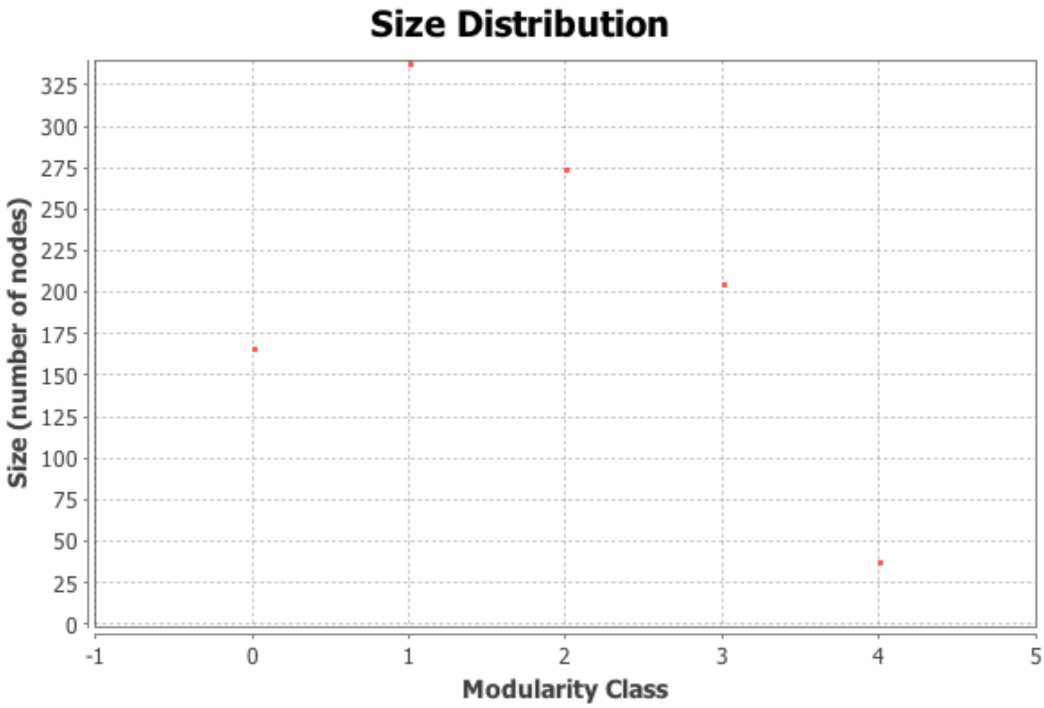

Figure 5.12: a graphic of the size of every community of the graph

5.3 Comparison and physical interpretation

In order to give a physical interpretation to the results discussed in the previous two paragraphs, it is important to compare the data obtained for the two points. In the table below are reported the main topological features of the two networks.

	HDC	**LDC**
Average degree	159.176	277.16
Diameter	7	7
Average path length	2.602	1.875
Clustering coefficient	0.642	0.687
Modularity	0.321	0.26
Number of communities	9	5

The average degree of HDC point is lower, this means that there are fewer connections in its graph. A probable cause of this behaviour may be the presence of more perturbations in the motion of the fluid particles, which involves in rapid fluctuations of the kinetic energy. This decreases the visibility between nodes. Perturbations may be caused by the continuous formation and breakdown of vortical structures.

Both the graphs have some nodes with a very high degree, as seen in chapter 5.1 these nodes correspond to peaks in the time series, which have a high value of kinetic energy. At these points there is a strong influence of fluid particles with a large value of kinetic energy on particles with lower value.

The diameter is equal for the two points; it does not give information about the difference between HDC and LDC point. Due to the conceptual similarity of the diameter and the average path length (they are both related to the path length between nodes) also the second index is not taken into account.

The clustering coefficient of HDC point is lower, so its graph has a smaller number of triangles and the nodes are more independent. Also this fact can be explained with the presence of vortexes: perturbations make the closure of triangles harder.

A similar conclusion can be obtained studying the modularity, which is higher for the HDC point. A high value of Q means a strong division in communities, in fact the HDC graph has more groups with smaller population. The presence of vortexes causes often perturbations of the system dynamics and as a results the successive states lose connectivity.

In the following figure are reported the graphics of local clustering coefficients for the two points:

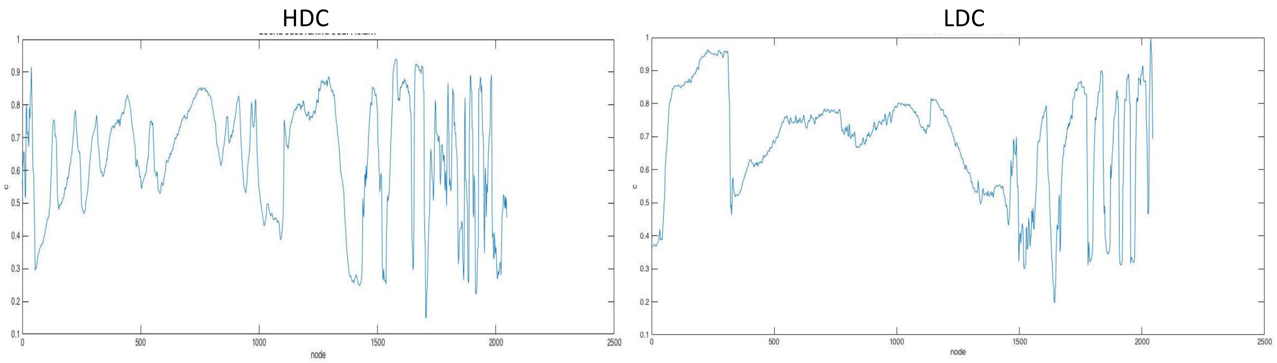

Figure 5.13: on the left the graphic of the local clustering coefficient of the HDC point, on the right the same graphic of the LDC point

The form of the last part of the graphic (characterized by the presence of a lot of peaks) calculated in the LDC point is similar to the form of the other graphic. Probably during the last times of the simulation vortexes are comparing in the LDC point. This hypothesis is supported by the presence of many peaks in the graphic of the degree in this step of time, that can be interpreted like perturbations.

In conclusion, the HDC point is characterized by the presence of many fluctuations of its energy, maybe caused by a continuous formation and breakdown of vortical structures. This means a low connectivity of the network. While LDC point has lower energy, but more constant, which increases connectivity.

Conclusions

In the present work two points of a forced isotropic turbulent field have been studied with the complex network tools. At first the visibility algorithm has been applied in order to transform the time series of kinetic energy into a graph, then its topology has been analyzed. The aim of the study was to find out correlations between the physical characteristics of the turbulent field and the indices used for studying the topology of the network.
Vortical structures promote in fast fluctuations of the kinetic energy, and this in turn decreases the visibility between the nodes of the graph.
The results obtained suggest that, to find out a point which has been much influenced by the presence of vortical structures during the time of the simulation, we have to search for point with precise characteristics:

- low average degree value;
- low clustering coefficient value;
- high modularity;
- the presence of many communities.

The diameter and the average path length are not much useful for this aim.
Eventually, the local clustering coefficient, due to its feature to be “local”, can give information about the temporal evolution of the phenomenon, for example it can suggests the birth of new vortexes at a certain time.
Thus some physical behaviours can be analysed using the visibility algorithm and the complex networks theory. This method can be useful because of its simplicity, since we can get information about the system without using the statistical approach.

Appendix A

Implementation of visibility algorithm

```
diametro=0;
path_medio=0;
C=0;
grado=zeros(1,N);
grado_medio=0;

%% MATRICE ADIACENZA
A=zeros(N);
for i=1:(N-1)
    A(i,i+1)=1;
    for j=(i+2):N
        c=1;
        for m=(i+1):(j-1)
            if X(m)>=X(i)+(X(j)-X(i))*((m-i)/(j-i))
                c=0;
                break;
            end
        end
        A(i,j)=c;
    end
end
for i=2:N
    for j=1:(i-1)
        A(i,j)=A(j,i);
    end
end

%% MATRICE DELLE LUNGHEZZE
D=A;
for k=2:(N-1)
    B=A^k;
    for i=1:(N-1)
        for j=(i+2):N
            if D(i,j)==0 && B(i,j)>0
                D(i,j)=k;
            end
        end
    end
end
for i=2:N
    for j=1:(i-1)
        D(i,j)=D(j,i);
    end
end

%% DIAMETRO E PATH MEDIO
somma=0;
for i=1:N
    for j=1:N
        somma=somma+D(i,j);
        if D(i,j)>diametro
            diametro=D(i,j);
        end
```

```
    end
end
path_medio=somma/(N*(N-1));


%% GRADO DEI NODI

K=zeros(1,N);
for i=1:N
    somma=0;
    for j=1:N
        somma=somma+A(i,j);
    end
    K(i)=somma;
end
grado(1,:)=K;
somma=0;
grado_max=0;
for i=1:N
    somma=somma+K(i);
    if K(i)>grado_max
        grado_max=K(i);
    end
end
grado_medio=somma/N;

%% DEGREE DISTRIBUTION
P=zeros(1,grado_max);
for i=1:grado_max
    for j=1:N
        if K(j)==i
            P(i)=P(i)+1;
        end
    end
end

%% CLUSTERING COEFFICENT
c=zeros(1,N);

for i=1:N
    somma=0;
    for j=1:N
        for m=1:N
        somma=somma+A(i,j)*A(j,m)*A(m,i);
        end
    end
    if K(i)==1
        c(i)=0;
    else
    c(i)=somma/(K(i)*(K(i)-1));
    end
end
somma=0;
for i=1:N
    somma=somma+c(i);
end
C=somma/N;

%% GRAFICI

% serie di tempo
time=0:2:2046;
```

```
figure
plot(time,X)
xlabel('time')
ylabel('kinetic energy')
title('TIME SERIES')

% degree distribution
degree=1:1:grado_max;
figure
plot(degree,P);
xlabel('degree (k)')
ylabel('number of nodes')
title('degree distribution')

grado_min=K(1);
for i=2:N
    if K(i)<grado_min
        grado_min=K(i);
    end
end
k_vettore=linspace(1,grado_max,50);
[g]=ksdensity(K,k_vettore);
figure
plot(k_vettore,g,'r')
xlabel('degree (k)')
ylabel('density')
title('DEGREE DISTRIBUTION')
hold on
pdf_k=histogram(K,20,'Normalization','pdf');

% grado dei nodi
figure
plot(time,K)
xlabel('nodes')
ylabel('degree')

% local clustering coefficient
figure
plot(time,c)
xlabel('nodes')
ylabel('c')
```

Bibliography

[1] Renzo Arina, 2015, *Fondamenti di aerodinamica*, Levrotto&Bella.

[2] R. J. Garde, 1994, *Turbulent Flow*, John Wiley & Sons.

[3] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J. C. Nuno, 2008, *From time series to complex networks: The visibility graph*, Proceeding of the National Academy of Sciences, vol. 105 no. 13, 4972-4975.

[4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. U. Hwang, 2006, *Complex networks: Structure and dynamics*, Physics reports, 424, 175-308.

[5] A. K. Charakopoulos, T. E. Karakasidis, P. N. Papanicolaou, A. Liakopoulos, 2014, *The application of complex network time series analysis in turbulent heated jets*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 24, 024408.

[6] Giovanni Iacobello, 2016, *Nuove osservazioni sulla caratterizzazione spaziale di flussi turbolenti attraverso la teoria delle reti complesse*, magistral thesis, Politecnico di Torino.