POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Aerospaziale

Tesi di Laurea Magistrale

Nuove osservazioni sulla caratterizzazione spaziale di flussi turbolenti attraverso la teoria delle reti complesse

New insights into spatial characterization of turbulent flows: a complex network-based analysis

Relatore:

Ing. Stefania Scarsoglio

Candidato:

Giovanni Iacobello

Aprile 2016

Alla mia Famiglia,

ad Aurora,

ad Anna

e...

Sommario

Lo studio dei flussi turbolenti è stato da sempre oggetto di grande interesse nell'ambito della ricerca fluidodinamica. La turbolenza, infatti coinvolge molti aspetti della vita quotidiana, dai fenomeni naturali (come le correnti oceaniche e atmosferiche) alle applicazioni industriali (tra cui, ad esempio, flussi nei tubi, turbine, processi chimici e di combustione o nelle estremità alari di velivoli ad ala fissa e rotante). Nonostante sia largamente presente nelle applicazioni pratiche, tuttavia, la turbolenza rappresenta una delle più grandi sfide della ricerca scientifica; ad oggi, infatti, gli strumenti a disposizione — sia concettuali, sia sperimentali — non riescono a fornire un approccio generale alla risoluzione dei problemi riguardanti la turbolenza. Gli approcci più comuni sono quindi di tipo numerico e sperimentale, le cui simulazioni forniscono un enorme quantità di dati, i quali necessitano di essere opportunamente esaminati ed interpretati.

Con lo scopo di ottenere una maggiore e migliore comprensione delle dinamiche dei flussi turbolenti e della loro caratterizzazione spaziale, nuovi e alternativi approcci di carattere multidisciplinare e nuovi strumenti di analisi divengono dunque necessari per un'appropriata valutazione statistica della turbolenza. In tale contesto, la teoria delle reti complesse — combinando elementi della teoria dei grafi e della fisica statistica — fornisce un potente strumento per lo studio di sistemi complessi, tra cui i flussi turbolenti, i quali sono composti da un elevato numero di elementi che interagiscono dinamicamente nel tempo. Sebbene ampiamente usata con successo in molte applicazioni pratiche (dalla sociologia e l'economia alla biologia, informatica e alla viabilità dei mezzi di trasporto), negli ultimi anni relativamente pochi approcci alla turbolenza basati sulla teoria delle reti complesse sono stati proposti. La maggior parte di questi, comunque, si basa sulla trasformazione di ciascuna serie temporale — valutata in differenti posizioni del flusso — delle grandezze fisiche misurate (ad es. energia o temperatura) in una corrispettiva rete complessa. Nel presente lavoro, invece, un'unica rete viene costruita a partire dai dati spazio-temporali acquisiti.

Un nuovo metodo di analisi dei flussi turbolenti, basato sulle reti complesse, viene quindi qui proposto — in alternativa soprattutto ai classici metodi statistici — con lo scopo di offrire una migliore comprensione della caratterizzazione spaziale delle dinamiche della turbolenza. Inoltre, i risultati successivamente mostrati sono stati sottoposti per pubblicazione nella rivista scientifica *Europhysics Letters* (EPL) e sono stati accettati per essere esposti come presentazione orale all'ottava conferenza EPFDC (*European Postgraduate Fluid Dynamics Conference*).

Contents

Introduction

Turbulence is an important and widely investigated topic which involves several aspects of everyday life, from natural phenomena (as oceanic or atmospheric currents) to research and industrial applications (as flow through pumps, turbines, chemical reactors, and aircraft-wing tips). Although it is so common in practical applications, turbulence represents one of the greatest scientific research challenges and an unsolved problem because so far both conceptual and experimental tools fail to provide a general approach to the solution of problems in turbulence. Richard Feynman said [1]: "*Turbulence is the most important unsolved problem of classical physics*", but many other quotes pointed out the importance of turbulence as well its high complexity.

Nowadays, experimental and numerical simulations provide a great amount of detailed data but these need to be examined and interpreted properly. Therefore, in order to achieve a better description of the dynamics and spatial characterization of turbulence, alternative, interdisciplinary approaches and investigative tools become necessary for appropriate statistical analyses. In this context, complex network theory — by combining graph theory and statistical physics — provides a powerful tool to analyze complex systems, just as turbulent flows, consisting of a huge number of dynamically interacting elements.

Despite widely and successfully used in other applications (from sociology and economics to biology, informatics, transportation and drivability), in the last years only few network-based approaches to turbulence have been proposed, mainly to study two-phase flows, turbulent jets or fully developed turbulent flows. Most of them focuses on temporal data measured in different spatial locations and, by means of the visibility algorithm [2] or recurrence plots [3], convert each time series into a network. Differently to what carried out so far, in this work each temporal series is not transformed into a network but a single global network from spatio-temporal data is constructed.

With this work a new approach to analyze turbulent flows through the complex networks is then proposed, trying to provide a better comprehension of the spatial characterization of the dynamic of turbulence and to overcome the limitations of the classic statistical methods.

The results achieved in this work have been submitted for publication to *Europhysics Letters* (EPL) journal [4] and have been accepted for oral presentation at the next *8th European Postgraduate Fluid Dynamics Conference* (EPFDC)[5].

More in detail, the study of a forced isotropic turbulent field is carried out; the network is built starting from the evaluation of the correlation coefficients of the turbulent kinetic energy, getting a single framework from spatio-temporal data. The so

obtained network is studied both by using typical structural and topological metrics of complex network theory — such as centrality metrics, clustering and community detection — and by introducing new, *ad hoc*, metrics — as the average physical distance.

The thesis is thus organized as follows.

In Chapter 1 turbulence is discussed from a general point of view. First, the properties which characterize turbulent flows are briefly reported. Then, the main difficulties encountered in dealing with turbulence and the classical approaches to the problem are summarized.

Chapter 2 is about the complex network theory. This topic is introduced with an overview of applications and then it is explained in its structural properties. The main definitions and the metrics used in this work are then listed and expounded. Finally, two short reviews are presented: the first one reports the typical features and properties of real and spatial networks, while the second one is an outline of the generally used tools to analyze and visualize large complex networks.

Chapter 3 is dedicated to the description of the assumptions and of the operative steps which led to the network building. The first part contains a description of the turbulence database acquired and the pre-processing stages. The hypotheses and the procedures that have been used are then explained and illustrated. In the last part of the chapter, instead, a short overview about the recent development of the complex network approach to turbulence is provided.

Finally, Chapter 4 contains a critical analysis of all the results obtained as a consequence of the post-processing stage. After a general description of the main features of the network under consideration, the outcomes of the examination of complex network metrics are showed and widely discussed. Therefore the physical interpretation of results is reported. The chapter ends with two analyses of sensitivity of two parameters of the network building and a further study applied to a different turbulent region of the same fluid flow domain.

Concluding remarks, future developments and possible applications are discussed in the Conclusions section.

Chapter 1:

An Overview of Turbulent Flows

1.1 Basic concepts

An easy explanation of what is the turbulence or a turbulent flow is generally an hard task because of the lack of a clear definition. The best way to clarify these concepts is to show several examples of turbulent flows in our everyday surroundings. In reality, most flows occurring in industrial applications and in nature are turbulent. Typical examples in which at least a part of fluid is turbulent are: the smoke from a cigarette or a chimney; the flow around bluff bodies or the wakes of cars, ships, aircrafts and bicycles; the terrestrial atmospheric; the blood flow; most combustion processes; the river and oceanic currents and so on. The counterpart of a turbulent flow is said a *laminar* flow. Intuitively, a laminar flow is a *simple* and *regular* flow, instead a turbulent flow is a *chaotic*, *random*, and *whirling* flow (Fig. 1.1-1, Fig. 1.1-2).
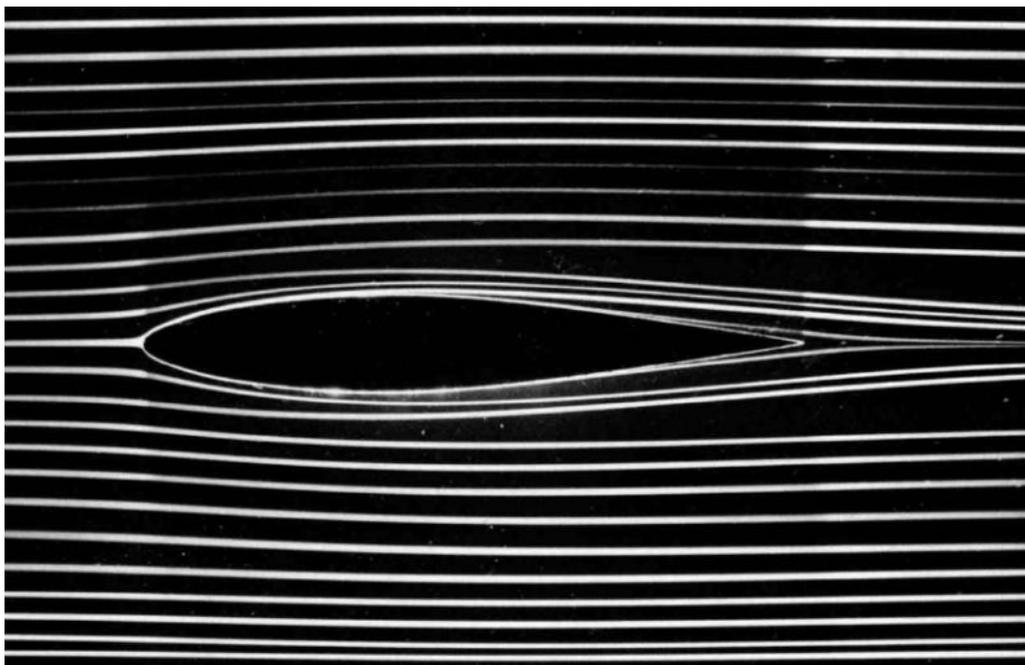
Fig. 1.1-1: Symmetric plane flow past an NACA 64A015 airfoil at zero incidence in a water tunnel; Re = 7000 [6]**.**
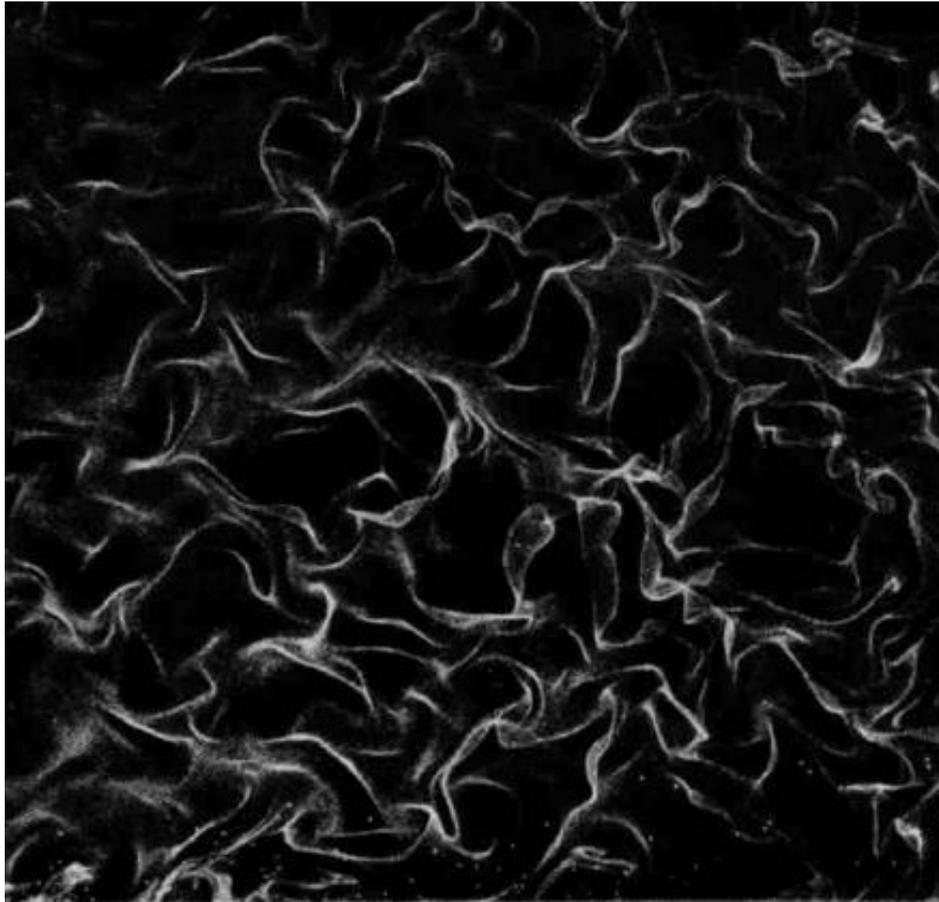
Fig. 1.1-2: Wrinkling of a fluid surface in isotropic turbulence [6] .

As the reader can easily imagine, thus, turbulence is the rule while laminar flows are a rare exception. The answer to the question "why is it so difficult to define precisely a turbulent flow?" probably comes directly from what said before, i.e. because of the great diversity of occurrence and its apparently random and chaotic nature. It is important to point out the meaning of "random" in this context. A variable is random if it does not have a unique value every time an experiment is repeated under the same set of conditions. However, the Navier-Stokes (NS) equations are valid both for laminar and turbulent flows and they present a deterministic nature. Therefore, the random character of the variables of a turbulent flow (visible for example in a time-history of a component of the velocity field, Fig. 1.1-3) does not come from the NS equations but from its "*context*", its "*environment*", i.e. the set of initial conditions that includes the initial and boundary conditions and material properties. In particular, two aspects play a key role:

1) *perturbations* in the set of initial conditions, that are almost unpredictable and — above all — they are unavoidable;

2) turbulent flow fields display an acute sensitivity to such perturbations; the sensitivity of a flow can be expressed as follows: two different flows with nearly the same set of initial conditions move rapidly apart with time, i.e. slight differences in initial conditions lead to large different behaviors later (state-space view).
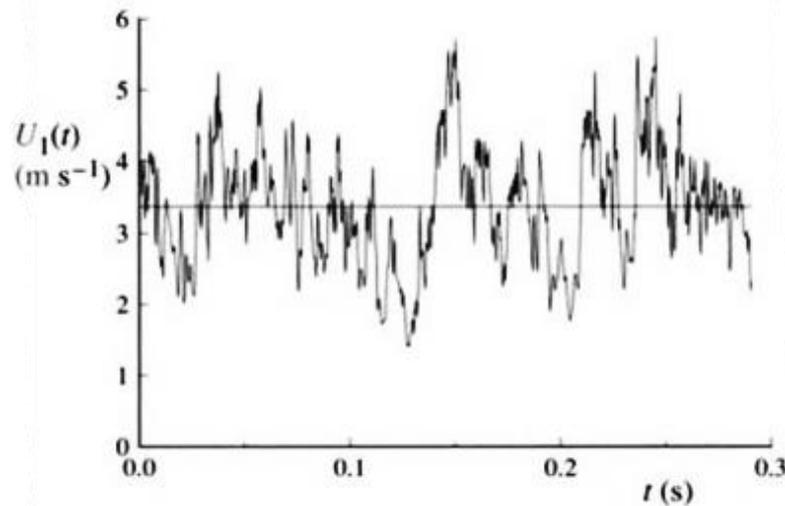
Fig. 1.1-3: The time history of the axial component of velocity $U_1(t)$ on the centerline of a turbulent jet [7].

The combination of these two factors explains the apparent random and chaotic nature of turbulence. It is worth the effort to underline that both the elements contribute to explain this aspect of a turbulent flow. The perturbations in initial and boundary conditions and material properties, indeed, are also present in laminar flows, but the extreme sensitiveness of turbulent flows is the trigger of such a behavior which is also referred as *deterministic chaos*. Turbulence theory, therefore, does not attempt to deal with all kinds of flows in a general way but it concentrates on families of flows with fairly simple boundary conditions (like boundary layers, jets and wakes). Nonlinearity of NS equations is crucial to the appearance of chaos and is also important in generating and maintaining turbulence. The theory of chaos was summarized very efficiently by Edward Lorenz as [8]: "*Chaos: When the present determines the future, but the approximate present does not approximately determine the future*". That is the case of turbulent flows.

1.2 General properties of turbulence

In order to overcome the limitation in the definition of turbulence, it is possible to specify a set of general properties common to almost all the turbulent flows. Furthermore, this procedure helps to give a deeper view of the physical nature of turbulence, rather than a simple definition. Below, the main characteristics of a turbulent flow are listed [9],[10].

- **IRREGULARITY or RANDOMNESS.**

   This property was discussed before and it concerns the large number of spatial degrees of freedom and the consequently impossibility to use a deterministic approach to turbulence problems.

- **DIFFUSIVITY**

   The diffusivity of turbulence is an important feature which causes rapid mixing and increased rates of mass transfer, momentum and heat. If a

flow pattern looks random but does not exhibit spreading of velocity fluctuations through the surrounding fluid, it is not turbulent. The contrails of a jet aircraft are a typical example where the flow is turbulent only when it was generated and then it extends for long distances with a nearly constant diameter.

- **LARGE REYNOLDS NUMBER**

  Turbulent flows always occur at high Reynolds numbers, *Re*. The Reynolds number is the ratio of the nonlinear inertial forces, responsible for the flow instability, to the linear dissipative damping, which converts kinetic energy into thermal energy. Turbulence often originates as an instability of laminar flows if the Reynolds number becomes large enough (the value depends on specific experiment). In general, a system in a state of equilibrium is said unstable if any perturbations from the equilibrium state are amplified by the system itself, usually leading the system to another completely different equilibrium state or to the so called "*divergence*". In a flow, instabilities are related to the interaction of viscous terms and nonlinear inertia terms in the equations of motion. With rising Reynolds number, the nonlinear convective term in the Navier-Stokes equations assumes increasing importance compared to the viscous term, and the tendency to instability, which is damped by viscosity, increases.

- **WIDE RANGE OF SCALES**

  The "*jiggle*" in velocity measurements at high *Re* (Fig. 1.1-3) reflects the existence of a continuum of different space and time scales of the flow. The large scales are evident in the overall fluctuations of a graph of velocity versus spatial position or time, whereas small ones are visible through the fine-scale oscillations. A graph of velocity as a function of position is furry when the large scales of a high Reynolds number flow are considered. Applying higher and higher magnifications, one eventually finds a scale at which the velocity is revealed to be a smooth function, defining the smallest scales of the flow. Smoothness reflects viscous action and thus the size of the smallest scales depends on the viscosity.

  The size of the large scales is typically fixed by the overall geometry of the flow whereas that of the smallest ones adjusts itself according to the viscosity. Quantitative measures of large scales in a turbulent flow can be made using the concept of temporal and spatial correlation of velocities. Correlation is a statistical parameter that reveals how velocities in two points or two times change together. Hence, scales defined using correlations give an idea of the size (in space and/or in time) of the largest turbulent structures. The smallest scale associated with turbulence, instead, mainly dictated by viscosity and by energy dissipation rate, is referred as the Kolmogorov length scale, η.

- **A CONTINUUM PHENOMENON**

  Turbulence is a continuum phenomenon, i.e. the smallest scales are ordinarily far larger than any molecular length scale. In other words, defining as l the size of the smallest turbulent scale and λ the mean free path, the Knudsen number of a turbulent flow is typically $Kn = \lambda/l \ll 1$. Exceptions may occur for gases and plasma if the turbulent Mach number is large enough.

- **THREE-DIMENSIONAL and ROTATIONAL**

  Turbulence is rotational and three-dimensional. In particular, turbulence is characterized by high levels of fluctuating vorticity that could not maintain itself if the velocity fluctuations were two dimensional. An important vorticity-maintenance mechanism is known as *vortex stretching* and it is an inviscid typical three-dimensional mechanism. In summary, turbulence exhibits high levels of three dimensional fluctuating vorticity.

- **DISSIPATION**

  Turbulent flows are always dissipative; viscous shear stresses work to increase the internal energy of the fluid at the expense of kinetic energy of the turbulence. Therefore, if there is no energy supply to maintain the flow, turbulence decays and eventually ceases to be active because the Reynolds number is no longer large enough.

- **A FLUID FLOW FEATURE**

  Turbulence is not a feature of fluids but of fluid flows. Most of the dynamics of turbulence is the same in all fluids, provided that the Reynolds number is large enough.

- **SMALL-SCALE RANDOM VORTICITY**

  Turbulent flows are rotational, i.e. they contain vorticity. This is defined as the curl of the velocity, $\overline{\omega} = \nabla \times \overline{U}$, thus involving its spatial derivatives, and represents rotation of small fluid particles about their centroid.

  Laminar flows can also posses vorticity but, a characteristic of high Reynolds number turbulence is that the vorticity has intense, small-scale, random variations in both space and time. Furthermore, the magnitude of these vorticity fluctuations is much larger than the mean vorticity and they are randomly orientated in direction.

  Qualitatively, *vortex convection* and *vortex stretching* — far from being fully understood — may be thought as the two inviscid mechanisms by which the intense, small-scale vorticity fluctuations of high *Re* turbulence are generated and maintained. Conversely, viscous diffusion tends to cause vorticity to spread, counteracting both the intensity amplification and scale-reducing effects of vortex stretching. In particular, since the viscous diffusive term contains second derivatives in space, it increases in importance as the scale considered decreases.

In conclusion, viscous diffusion of vorticity places both a lower limit (the Kolmogorov scale) on the size of vortical structures attainable by stretching, resulting in the velocity field being smooth when viewed at such scales, and an upper limit on the amplification of vorticity fluctuations.

- **LARGE SCALE INSENSITIVITY TO VISCOSITY**

    If the turbulent Reynolds number is "high enough", the dynamics of the large scales are essentially inviscid and hence insensitive to the precise value of the large Reynolds number. While the size of the smallest scales changes according to the fluid viscosity so as to dissipate energy — at an appropriate rate controlled by the large scales — , the smallest scales are thought to have little direct effect on the large ones, which interact mainly with scales immediately below them in the cascade.

    Hence, those properties of turbulence which are determined by the large scales should be largely unaffected by changes in the viscosity. This leads to the conjecture that, as instance, the mean energy dissipation rate, mean velocity and root-mean-squared velocity fluctuations of a turbulent flow approach limiting values as the Reynolds number tends to infinity (or at least that they ought to vary much more slowly with changes in the Reynolds number). Furthermore, although viscous dissipation of energy is dominated by the smallest scales, the average rate of energy dissipation reflects the mean rate of energy supplied from the large scales and it is consequently also believed to tend to a limit as the viscosity goes to zero.

1.3 Techniques for the study of turbulence

1.3.1. General considerations

Turbulent flows have been investigated for more than a century, but no general approach to the solution of problems in turbulence exists. The equations of motion have been analyzed in great detail, but it is still next to impossible to make accurate quantitative predictions without relying heavily on empirical data [9]. Many different techniques have been developed to overcome many different problems regarding turbulence and turbulent flows. As discussed earlier, at large enough Reynolds numbers there is a great separation of scales having different behaviors determined by different factors (overall geometry on one side, viscosity and energy dissipation rate on the other). This assumptions suggest different approaches on solving and prediction of different scales in turbulent flows. In general, studies on turbulent flows can be divided roughly into three categories [7]:

I) *Discovery.* Experimental or simulation studies, aimed at providing qualitative or quantitative information about particular flows.

II) *Modelling.* Theoretical or modelling studies, aimed at developing *tractable* mathematical models that can predict properties of turbulent flows.

III) *Control.* Studies aimed at manipulating or controlling the flow or the turbulence in a beneficial way, usually involving both experimental and theoretical components.

The objective of first category is to develop an understanding of the dominant physical processes of simple turbulent flows, and how they are related to the equations of motions. For studies in the second category, instead, the word *tractable* is crucial as better explained later. The studies concerning the second category can also be used in some studies of the third category.

Let us focus on second category listed before, in particular the meaning of *tractable* adjective. The NS equations describe in details fluid flows — laminar and turbulent — so that the amount of information deriving from a direct approach of solving NS equations is impossible to handle in practice. Thus, NS equations, despite their accuracy, do not provide a *tractable* approach for turbulent flows. The direct approach is called *Direct Numerical Simulation* (DNS); while it is intractable for high-Re flows (the computational cost for DNS grows with Re^3), it can be useful as a research tool for simple turbulent flows at moderate Reynolds number. For high-Re flows the natural alternative to DNS is the statistical approach. A model based on statistical variables (such as mean velocity field) can lead to a *tractable* set of equations, because statistical fields vary smoothly in space and time (Fig. 1.3-1). However, by introducing a model it is inevitable to lose some information — according to the characteristics of the model — or it is not possible to generalize the specific model considered for all families of turbulent flows. Hence, for each experiment involving a turbulent flow there will be an appropriate model that approximates, as best as it can, the real flow.
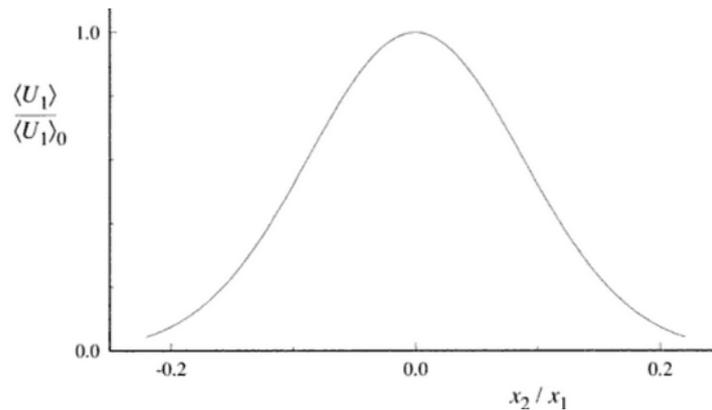
Fig. 1.3-1: The mean axial velocity $\langle U_1 \rangle$ profile in a turbulent jet with Re = 95500, normalized by its value on the centerline $\langle U_1 \rangle_0$; the cross-stream (radial) coordinate x_2 is normalized by the distance from the nozzle x_1 [7].

It is worth the effort to distinguish between the two definition of *turbulent-flow simulation* and *turbulence model*: in the former, equations are solved for time-dependent velocity field that represents the velocity field for one realization of the turbulent flow; in the latter, instead, equations are solved for some mean quantities. However, the word "model" is generally used to refer both to simulations and to turbulence models. Typical example of simulations are the DNS and the *Large-Eddy Simulation* (LES). In LES, equations are solved for a "*filtered*" velocity field which is representative of the large-scale turbulent motion — which are of most interest in several practical cases — while the small-scale motion is represented with simple models (because the smaller scales are responsible for fine grids in numerical simulations and make calculations expensive).

Since there is a broad range of different turbulent flows, it is useful and appropriate to have a broad range of models that vary in their attributes. The principal criteria that can be used to assess different models are [7]:

- Level of description, i.e. the amount of information determined with the model; DNS, as instance, has the highest level of description.

- Completeness. A model is said complete if constituent equations of the model are free from flow-dependent specifications (i.e. material properties, boundary and initial conditions).

- Cost and ease of use, i.e. the computational difficulty to perform a calculus in terms of CPU time and *flops*, memory, algorithm developing.

- Range of applicability. Not all models are applicable to all flows. However, a model is applicable to a flow if the model equations are well posed and can be solved, even if the solutions are not accurate. Computational requirements place limitations on the applicability of some models as in DNS, where the computational requirements rise steeply with Reynolds number.

- Accuracy. This can be determined by comparing model calculations with experimental measurements and it is a desirable attribute of any model. The accuracy is affected by errors of measurements and numerical errors, mainly coming from the impossibility to set the same initial and boundary conditions of the real flow experiment.

The suitability of a particular model for a particular turbulent-flow problem depends on a weighted combination of the criteria listed above.

1.3.2. Statistical approach to turbulence

The apparently random character of turbulent flows, as depicted in the previous sections, suggests a statistical approach. In particular, thank to the increasing computational power of modern computers, the relevant properties of turbulent flows can be calculated numerically. The detailed behavior of the flow in any realization is extremely sensitive to small changes in the initial or boundary conditions, which the experimenter cannot control to infinite precision. Hence, it would be necessary to repeat many times the same experiment — trying to maintain the same boundary and initial conditions — in order to collect as much information as possible. This type of problem is ideally suited to statistical methods [10]. An important step here must be considered, despite a lot of times that is implied. The statistical approach is usually based on averaged variables — above all, the velocity field — but the average operation is, by definition, expressed in terms of *statistical ensemble,* i.e. the mean value of a general variable can be obtained repeating the same experiment, with the same conditions, N times, N $\rightarrow \infty$. In practice, the measured averages are *temporal averages* so, in order to overcome the impossibility to repeat the same experiment a large number (at limit infinite) of times, it is used the *ergodic hypothesis.* According to this hypothesis, *the average of a variable over time and the average over the statistical ensemble are the same*, so it is possible to carry out an experiment once but for a long time (at limit infinite). For example, the hypothesis says that to throw one billion dice is statistically nearly the same that to throw only one dice one billion times. Usually, it is not easy to demonstrate the ergodic hypothesis so it is often assumed *a priori.* Furthermore, practically, the temporal averages can be made only in a finite interval of time, while by definition those have to be done in the time interval $[0, +\infty)$. A good approximation is obtained using a time interval greater than the order of magnitude of the characteristic time (*Eddy Turnover Time*) of turbulent structures considered.

From a historical point of view, the first statistical method to analyze turbulent flows was proposed in 1894 by Osborne Reynolds who assumed that turbulent flows can be described by ensemble averages, without considering the details of each flow realization [11]. He then decomposed the velocity field $\overline{U}$ into a mean contribution $\langle \overline{U} \rangle$ plus fluctuations $\overline{u}$ and rewrote the Navier–Stokes equations to predict the evolution of mean velocity, obtaining the so called *Reynolds-Averaged-Navier-Stokes equations* (RANS). However, to solve the Reynolds equations one should compute the second order moment of the velocity fluctuations, called the *Reynolds stress tensor*, which in fact depends on the third order moment which depends on the fourth order moment, and so on *ad infinitum.* This is the *closure problem*: there are more unknowns than equations and to solve the hierarchy of Reynolds equations the traditional strategy is to introduce another equation, or system of Reynolds equations, chosen from some *a priori* phenomenological hypotheses, to close the set of equations [11]. Several models have been developed to determine the Reynolds stresses in RANS (solved for the mean

velocity field) either via the *turbulent viscosity hypothesis* or more directly from modeled Reynolds-stress transport equations. Focusing on the former, the models based on that hypothesis are relatively simple but they show several limitations that should always be kept in mind. The turbulent viscosity hypothesis (introduced by Boussinesq in 1877) is mathematically analogous to the stress-rate-of-strain relation for a Newtonian fluid and is based on the analogy with molecular diffusion, which smoothes velocity gradients for scales smaller than the molecular mean free path [11]. According to the hypothesis, the deviatoric Reynolds stress $D_{R,ij}$ is proportional to the mean rate of strain $2\bar{S}_{ij}$ through a positive scalar coefficient v_T called *turbulent viscosity* (or *eddy viscosity*) [7]

$$D_{\mathrm{R,ij}} = -\rho\langle u_i u_j\rangle + \frac{2}{3}\rho K \delta_{ij} = \rho v_T \left(\frac{\partial\langle U_j\rangle}{\partial x_i} + \frac{\partial\langle U_i\rangle}{\partial x_j}\right) = 2\rho v_T \bar{S}_{ij} \ , \qquad (1.1)$$

where $-\rho\langle u_i u_j\rangle$ is the Reynolds stress tensor with density ρ, $\frac{2}{3}\rho K \delta_{ij}$ is its isotropic part, $K = \frac{1}{2}\langle u \cdot u\rangle$ is the turbulent kinetic energy and δ_{ij} is the Kronecker delta. The average operation is indicated with the notation $\langle \cdot \rangle$. Given the turbulent viscosity field $v_T(\bar{x}, t)$, the Reynolds stress term assumes the advantage of having the same form as the NS equations; however, for many flows the accuracy of the hypothesis is poor. The most important and common turbulent models are grouped, as follows, in order of increasing complexity [1]:

First Order Models

- **Algebraic (Zero Equations) Models:**
  - Uniform Turbulent Viscosity;
  - Mixing-Length.
- **One Equation Models:**
  - Prandtl's K-equation model;
  - Spalart-Allmaras model,
    - Detached Eddy Simulation (DES).
- **Two Equation Models:**
  - K-ε model;
  - K-ω model.

Second Order Models

- Algebraic Stress Models (ASM)
- Reynolds Stress Models (RSM)
- Nonlinear Eddy Viscosity Models

Probability-Density-Function of Velocity Methods (PDF)

Fig. 1.3-2: Turbulent models review.

First order models are based on the analogy between laminar and turbulent flows. They are also called *Eddy Viscosity Models* (EVM), as shown in the Boussinesq model. For instance, to close the Reynolds equations, Prandtl introduced a characteristic scale for the velocity fluctuations, called *mixing length*, which led him to rewrite the Reynolds stress tensor as a turbulent diffusion term. Following the hypothesis proposed by Boussinesq, Prandtl assumed that there exists a turbulent diffusion which regularizes the mean velocity gradients, for scales smaller than the mixing length [11]. Unfortunately this hypothesis is wrong because, contrary to molecular diffusion — which is decoupled from the large scale motions and can then be modeled by a linear operator (Laplacian), with an appropriate transport coefficient (viscosity) — turbulent motions interact nonlinearly at all scales and there is no spectral gap to decouple large scale motions from small scale motions. This is a major obstacle faced by all turbulence models and the closure problem remains still open [11].

The central concept of second order models, instead, is to make direct use of the governing equations for the second order moments (Reynolds stresses and turbulent fluxes) as an alternative of the questionable Boussinesq hypothesis. The motivation is to overcome the limitations of first order models in dealing with the isotropy of turbulence and the extra strains. The overshoot of this approach is the large number of partial differential equations induced, which involve many unknown or impossible to find correlations [1].

A fundamental contribute to turbulence was also given by A.N. Kolmogorov, who published three papers in 1941 on the statistical theory of fully developed turbulence, now referred to as *K41 theory*. He used spectral analysis to study the way in which NS equations in three dimensions distribute energy among the different scales of the flow, identified by their respective wave number. The K41 theory is based upon the *energy cascade* concept — introduced by Richardson in 1922 — and the three hypotheses stated by Kolmogorov himself. In his work, Richardson supposes that turbulence can be considered to be composed of *eddies* (swirling turbulent structures) of different sizes, where the largest are characterized by length-scale comparable with the flow macro-scale and can contain also smaller eddies. Furthermore, Richardson stated that large eddies are unstable and break up, transferring their energy to smaller eddies at a rate which is supposed to be constant, in a cascade-like way, until the local *Re* is small enough to make the motion stable and let the molecular viscosity dissipate the kinetic energy. Kolmogorov then advanced the following three hypotheses [7].

- **Local isotropy**: at sufficiently high *Re*, the small scale turbulent motions — contrarily to large scales — are statistically isotropic.

- **First similarity hypothesis**: in every turbulent flow at sufficiently high *Re*, the statistics of the small scale motions have a universal form that is uniquely determined by viscosity and energy dissipation rate. Just as the directional information (i.e. isotropy) of the large scales is lost as the energy passes down the cascade, Kolmogorov argued that all information about the geometry of large eddies is lost. Hence, the statistics of small scale motions are in a sense "universal". So, given the viscosity and the energy dissipation

rate there are unique length, velocity and time scales that can be formed; these are the *Kolmogorov scales*.

- **Second similarity hypothesis**: in every turbulent flow at sufficiently high *Re*, the statistics of the motion of scales ℓ in the range $\eta \ll \ell \ll \ell_0$, where η is the Kolmogorov length scale and ℓ_0 the lengthscale of largest eddies, have a universal form that is uniquely determined by the energy dissipation rate, independently of viscosity. This range of scales is named *inertial subrange*, while the range of smaller scales is the *dissipation range* (cf. Fig. 1.3-3 and Fig. 1.3-4).
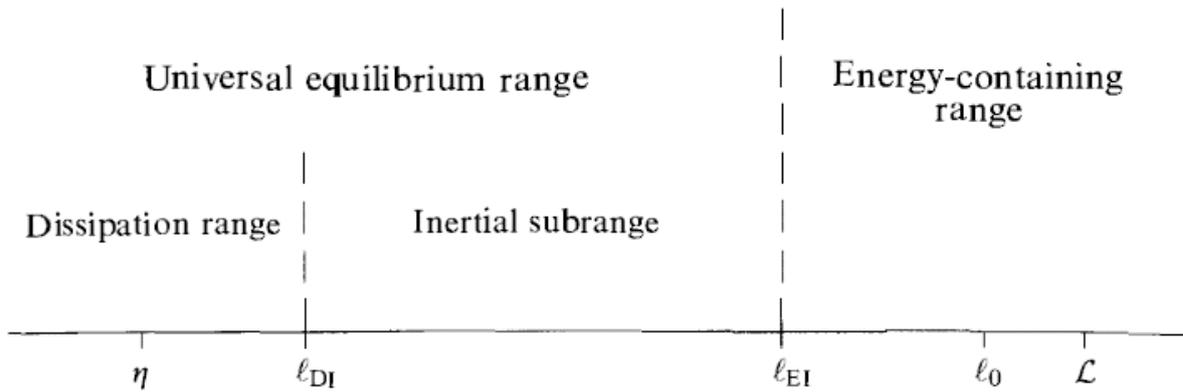
Fig. 1.3-3: Eddy sizes (on a logarithmic scale) at very high Re; flow scale $\mathcal{L}$; lengthscale of largest eddies ℓ_0; lengthscale ℓ_{EI} is the demarcation between the anisotropic large eddies ($\ell > \ell_{EI}$) and the isotropic small eddies ($\ell < \ell_{EI}$); lengthscale ℓ_{DI} separates the inertial subrange and the dissipation range; η, Kolmogorov scale [7].
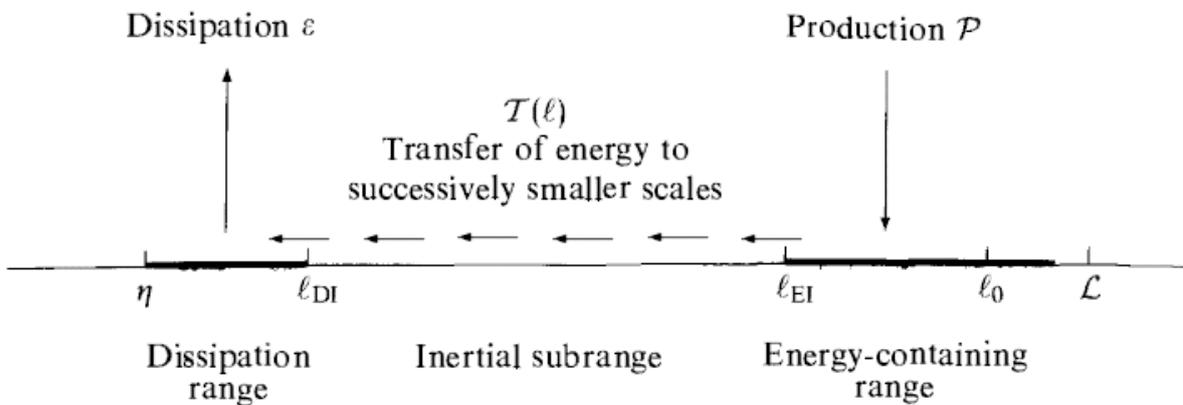
Fig. 1.3-4: A schematic diagram of the energy cascade at very high Reynolds number [7].

It remains to be determined how turbulent kinetic energy is distributed among eddies of different sizes. Kolmogorov found that, for statistically homogeneous and isotropic turbulent flows, the K41 model predicts that the energy spectrum scales with $k^{-5/3}$, where k is the modulus of the wave number, averaged over directions, corresponding to the inverse of the scale. For two-dimensional turbulence there is a statistical theory similar to Kolmogorov's theory developed by Batchelor and Kraichnan [11].

1.3.3. Tools for statistical analysis

In the previous section the importance of a statistical approach for the study of turbulent flows is repeatedly highlighted, but without specifying the tools typically used. There are many statistical tools that can be used to characterize a random field but, often, only few parameters are more meaningful or useful than others. In this context, five of the tools used for the analysis of turbulence are briefly introduced.

- **Probability Density Function (PDF).**

  PDF is a function that describes the relative probability for a random variable to assume a given value. The probability of a random variable falling within a range of values is given by the integral of the variable's PDF over the range considered. It is very useful because fully characterizes the random variable; if two or more random variables have the same PDF are said to be *statistically identical.*

- **Mean and Higher Order Moments.**

  The n-th moment of a real random variable U about a value c is defined as

$$\mu_n(U) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} (V - c)^n f(V) dV \ , \tag{1.2}$$

  where V can assume all possible values of U and $f(V)$ is the PDF of U. If $c = 0$ and $n = 1$, one obtains the *mean* $\langle U \rangle$ of the random variable U. If c is equal to the mean then μ_n is called *central moment*. For instance, the *variance* of the variable U is the second order central moment of U — the *standard deviation* is its square-root.

- **Structure Functions.**

  The second-order velocity structure function is the covariance of the difference in velocity between two points distant $\vec{r}$, at the time t [7]

$$D_{ij}(\vec{r}, \vec{x}, t) = \langle [U_i(\vec{x} + \vec{r}, t) - U_i(\vec{x}, t)] [U_j(\vec{x} + \vec{r}, t) - U_j(\vec{x}, t)] \rangle \ , \tag{1.3}$$

  with $i, j = 1,2,3$ the three component of the velocity field U. Supposing that the assumptions of Kolmogorov are satisfied (as high Reynolds number and homogeneity), D_{ij} can be rewritten as an isotropic function of $\vec{r}$ [7]

$$D_{ij}(\vec{r}, t) = D_{NN}(r, t)\delta_{ij} + [D_{LL}(r, t) - D_{NN}(r, t)] \frac{r_i r_j}{r^2} \ , \tag{1.4}$$

  where the scalar functions D_{LL} and D_{NN} are called, respectively, the *longitudinal* and *transverse structure functions.*

If the coordinate system is chosen so that $\vec{r} = r\hat{\imath}$, i.e. $\vec{r}$ is in the x direction then

$$D_{LL} = D_{11}; \quad D_{NN} = D_{22} = D_{33}; \quad D_{ij} = 0, \qquad \text{for } i \neq j \ .$$

Furthermore, as a consequence of homogeneity and isotropy, from the continuity equation it follows that

$$D_{NN}(\vec{r},t) = D_{LL}(r,t) + \frac{r}{2}\frac{\partial D_{LL}(r,t)}{\partial r} \ . \tag{1.5}$$

Thus, in locally isotropic, homogeneous turbulence, $D_{ij}(\vec{r},t)$ is determined by the single scalar function $D_{LL}(r,t)$.

This tool was used by Kolmogorov in his theory and is very useful to verify the three hypotheses and to compare them with experimental data.

- **Correlation Functions.**

  Generally, correlation functions between random variables are statistical indicators of dependencies as a function of distance in time or space. When one considers the correlation function between random variables representing the same quantity, then this quantity is often referred to as *autocorrelation function*; instead, when one considers correlation functions of different random variables then these are sometimes called *cross-correlation functions* to emphasize that different variables are being considered. A measure of the linear correlation between two variables A and B is the *Pearson product-moment correlation coefficient* or simply *correlation coefficient* [II]

  $$\rho(A,B) = \frac{cov(A,B)}{\sigma_A \sigma_B}, \qquad -1 \leq \rho \leq 1 \ , \tag{1.6}$$

  where $cov(A,B)$ is the *covariance* of the two variables A, B (a measure of how much A and B change together), while σ is the standard deviation.

  If $\rho = \pm 1$ the variables are totally linearly correlated (or anti-correlated) while if $\rho = 0$ they are uncorrelated. For samples A_i, B_i with $i = 1,2,\ldots,N$ scalar observations and with mean μ, the Pearson coefficient is

  $$\rho(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{A_i - \mu_A}{\sigma_A}\right)\left(\frac{B_i - \mu_B}{\sigma_B}\right) , \tag{1.7}$$
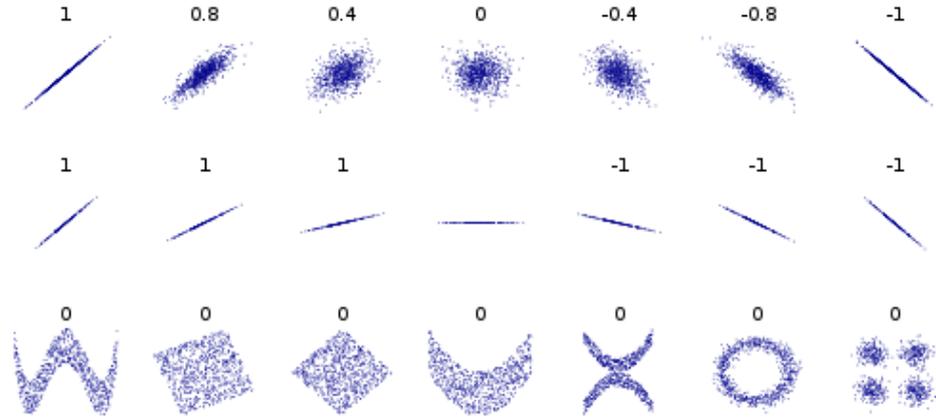
Fig. 1.3-5: Several sets of (x, y) points, with the correlation coefficient for each set. The correlation reflects the non-linearity and direction of a linear relationship (top row), but not the slope of that relationship (middle), nor many aspects of nonlinear relationships (bottom).

The correlation coefficient matrix of two random variables is the matrix of correlation coefficients for each pairwise variable combination,

$$R(A, B) = \begin{pmatrix} \rho(A, A) & \rho(A, B) \\ \rho(B, A) & \rho(B, B) \end{pmatrix} = \begin{pmatrix} 1 & \rho(A, B) \\ \rho(B, A) & 1 \end{pmatrix}, \tag{1.8}$$

which is symmetrical, being $\rho(A, A) = \rho(B, B) = 1$ and $\rho(A, B) = \rho(B, A)$. For correlations of $M > 2$ samples, $A_1, A_2, \ldots, A_M$, each with N scalar observations, the $M \times M$ matrix of correlation coefficients R with entries r_{ij} becomes

$$r_{ij} = \begin{cases} 1 \\ \rho(A_i, A_j) = \rho(A_j, A_i) \end{cases} if \quad \begin{matrix} i = j, \\ i \neq j. \end{matrix} \tag{1.9}$$

As instance, for a statistically stationary process arising in turbulent flows, using the velocity fluctuations $u(t) = U(t) - \langle U \rangle$, the autocorrelation function between the process at time t and $(t + s)$ is

$$\rho(s) = \frac{\langle u(t)u(t + s) \rangle}{\langle u(t)^2 \rangle},$$

and one expects that to diminish as the lag time s increases. In this manner, the *integral timescale* of the process can be defined as [7]

$$\bar{\tau} = \int_0^\infty \rho(s)ds.$$

The same procedure could be carried out for one-time, spatial correlation which is often referred as *two-point correlation* [7]

$$R_{ij}(\vec{r}, t) \equiv \langle u_i(\vec{x} + \vec{r}, t) u_j(\vec{x}, t) \rangle . \tag{1.10}$$

Just as with the structure function D_{ij}, as a consequence of isotropy, R_{ij} can be re-expressed in terms of two scalar functions $f(r, t)$ and $g(r, t)$ called respectively *longitudinal* and *transversal autocorrelation functions*

$$R_{ij}(\vec{r}, t) = u'^2 \left[g(r, t)\delta_{ij} + [f(r, t) - g(r, t)] \frac{r_i r_j}{r^2} \right] , \tag{1.11}$$

where $u'^2 \delta_{ij} = \langle u_i u_j \rangle = R_{ij}(0, t)$. The two autocorrelation functions $f(r, t)$ and $g(r, t)$ are related as the two structure function are related in the Eq. (1.5). Then, generally speaking, just as the integral timescale $\bar{\tau}$, an integral lengthscale can be defined as

$$L_{ijk} = \frac{\int_0^\infty R_{ij}(\vec{r}, t) dr_k}{R_{ij}(0, t)} . \tag{1.12}$$

Since there are nine different R_{ij} and three different directions, twenty-seven integral lengthscale can be defined. However, taking into account the isotropy assumption and the relation between $f(r, t)$ and $g(r, t)$, only one lengthscale is obtained, that is

$$L = \frac{\int_0^\infty f(r, t) dr}{f(0, t)} , \tag{1.13}$$

or, equivalently, considering $g(r, t)$ in place of $f(r, t)$.

Another important lengthscale in characterization of turbulent flows is the so called *Taylor microscale* [7]

$$\lambda_g(t) = \left[-\frac{1}{2} \frac{\partial^2 g}{\partial r^2}(0, t) \right]^{-\frac{1}{2}} , \tag{1.14}$$

and it can be seen as the characteristic dimension of the smallest dynamically significant eddies of the flow.

- **Spectral Analysis - Fourier Transform.**

   Another important tool is the spectral analysis, a technique of decomposing into simpler parts a complex signal through a sum (or integral) of many individual frequency components. A useful quantity, mentioned earlier especially about K41 theory, is the energy spectrum function $E(\kappa, t)$ such that [7]

$$\int_0^\infty E(\kappa, t)\, d\kappa = \frac{1}{2}\langle u_i u_i \rangle = \frac{1}{2} R_{ii}(0, t) \ ,$$

   and it represents the contribution to the turbulent kinetic energy from all modes with wavenumber κ.

1.4 Closing remarks

Because of the lack of a general theory, turbulence is primarily an experimental problem, i.e. we know much more about it from experience than from theory. A continuing challenge to research is to develop methodologies to calculate the flow and turbulence properties of practical relevance. There is a broad range of turbulent flow problems for which various levels of description and accuracy are required.

In general, the turbulence studies touch several natural aspects of physics, engineering, chemistry and other disciplines; in this view, a multidisciplinary approach suits the purpose of a deeper knowledge of the dynamics of turbulence. The present work embraces this idea, suggesting an innovative and strongly interdisciplinary technique for the characterization of turbulent flows that could overcome the limits or enhance the knowledge of the other methods.

Chapter 2:

The Complex Network Theory

2.1 An introduction to the complex networks

Network theory is a part of *graph theory*, which is the study of *graphs* — mathematical structures used to model pairwise relations between entities. A graph (or network) is then made up of *vertices* (or *nodes*, *points*), that are connected by *edges* (or *arcs*, *lines*). Thanks to their easy nature, graphs are a flexible tool to describe a lot of everyday life "systems". Well-known examples are the Internet, highways or subway systems (Fig. 2.1-1) and neural networks. However, graphs are used in many other applications, such as art and literature (Fig. 2.1-2), music (Fig. 2.1-3), sports (Fig. 2.1-4) or biology (Fig. 2.1-5).

Fig. 2.1-1: The latest re-design of the Paris metro map, produced by the design agency bdcconseil in 2003 [III].
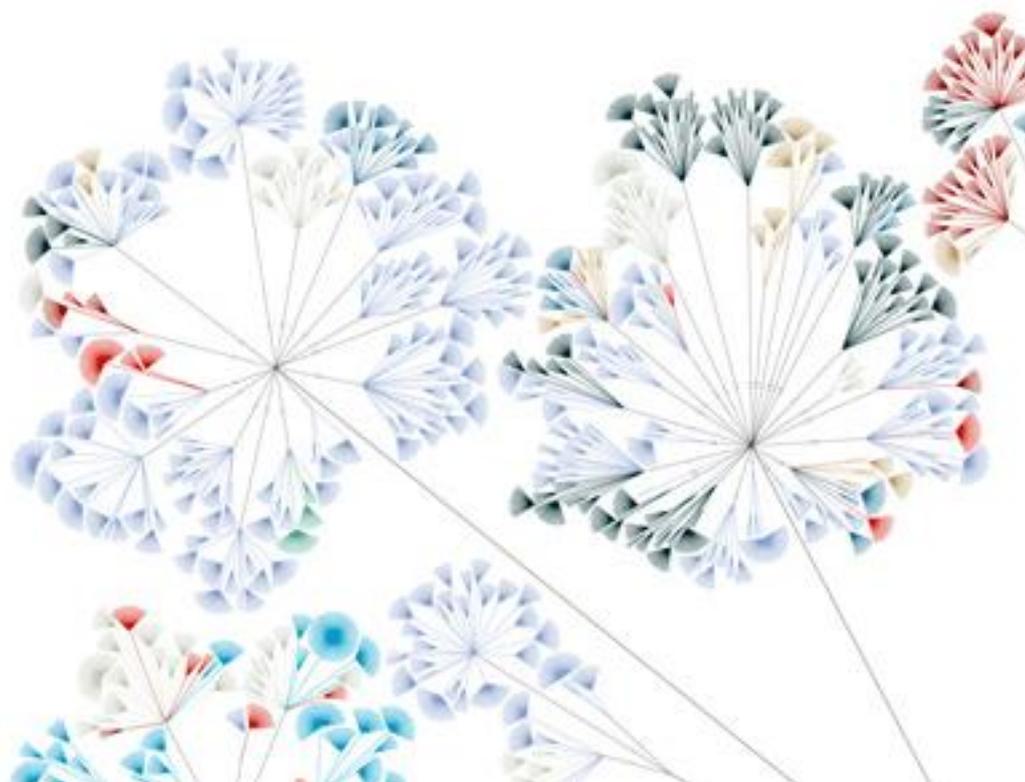
Fig. 2.1-2: Writing Without Words, by Stephanie Posavec; the image shows an example of visualization of modern classic. Each literary component was divided hierarchically into even smaller parts - Part, Chapters, Sections, Sentences, and ultimately Words, also divided with different colors [III].
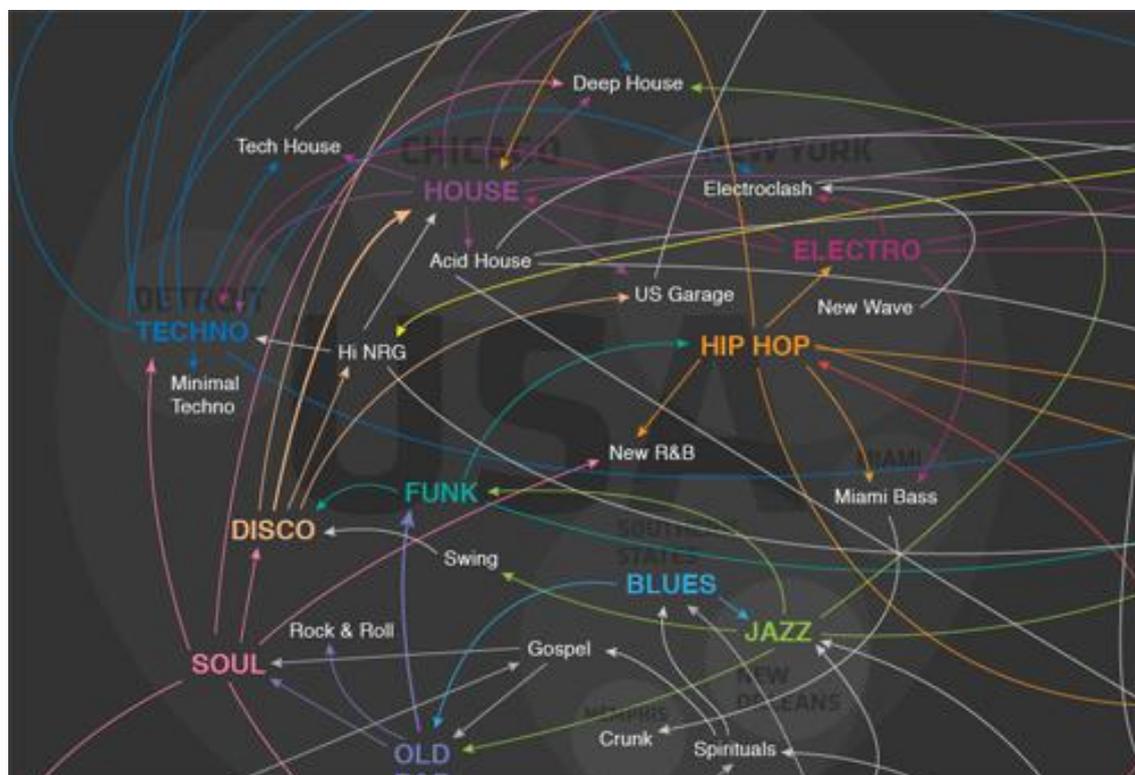
Fig. 2.1-3: A part of the map of cultural transmission of music across the world and across time, from 1800-2000s, and traces the geographic and temporal evolution of Western dance music [III].
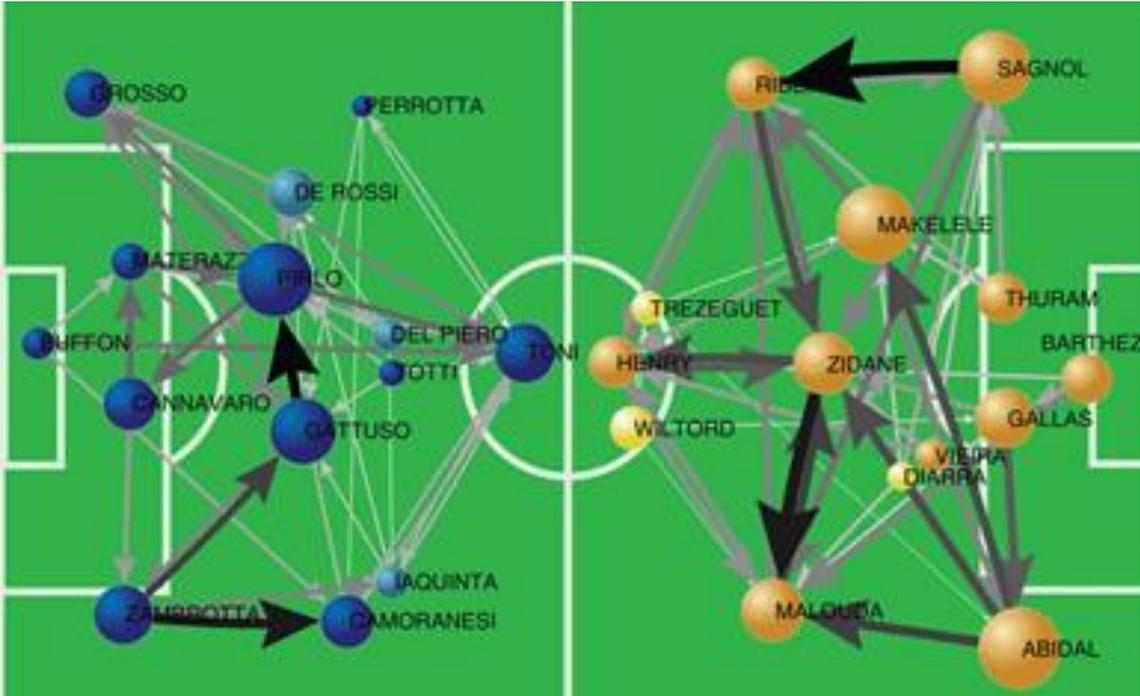
Fig. 2.1-4: The network shows the passes from every player to those three team-mates he passes to most frequently during FIFA World Cup Germany 2006. Strength of arcs displays the number of passes. Size of nodes displays the influence (flow betweenness) of a player [III].
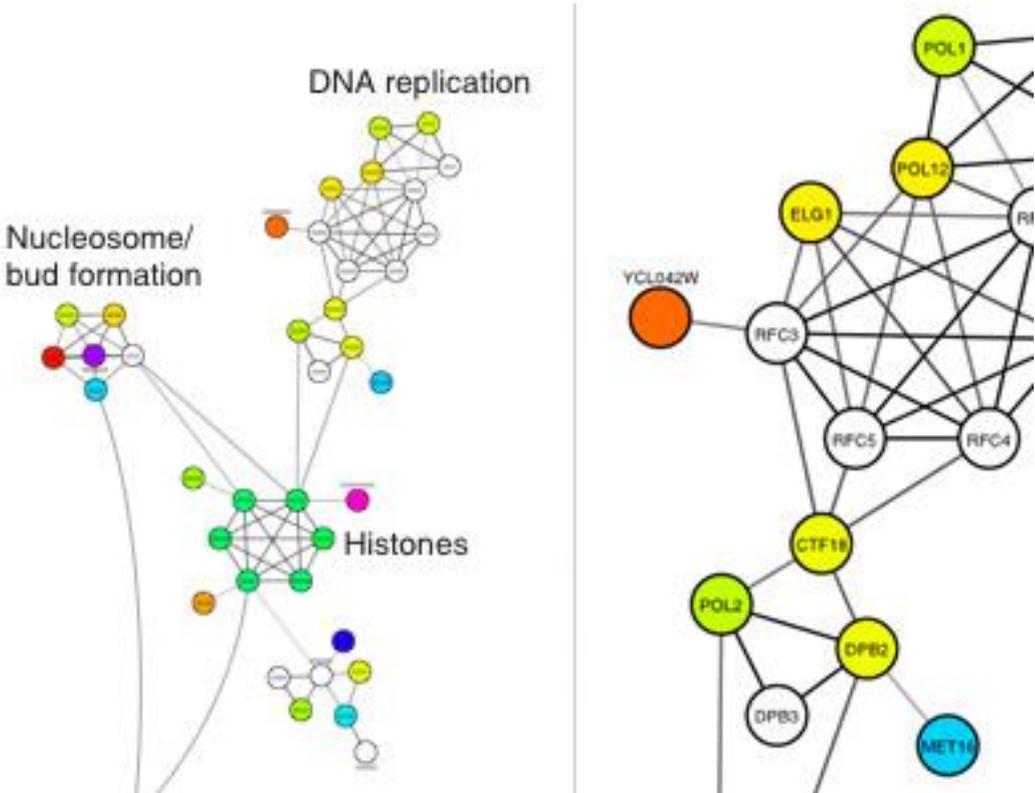
Fig. 2.1-5: The graph represents a temporal protein interaction network of the yeast mitotic cell cycle. Cell cycle proteins that are part of complexes or other physical interactions are shown within the circle. For the dynamic proteins, the time of peak expression is shown by the node color; static proteins are represented by white nodes [III].

In this context, a *complex network* is a graph with non-trivial topological features, typically occurring in networks modeled from real systems and involving thousands or millions of nodes. The network theory was developed hugely during all 20th century both in mathematical areas and in all its application fields. The study of complex networks, instead, is a very young area of scientific research, which took hold mainly in the last years of 1990s decade. Despite the relatively recent growth of interest in network theory, historically, the birth of the graph theory coincides with the solution of the so called *Seven Bridges of Königsberg problem* by Leonhard Euler in 1736 [12], which is deservedly referred as one of the father of graph theory.
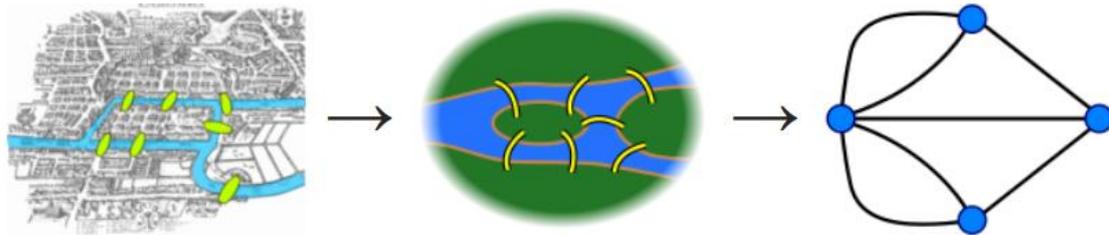
Fig. 2.1-6: Diagram of Seven Bridges of Königsberg with a graph representation [IV].

Focusing on complex network theory, this can be thought as combination of graph theory and statistical physics, endowing it with a multidisciplinary nature [13]. First developments can be traced back to percolation and random graphs works by Flory [14], Rapoport [15]-[17], and Erdős and Rényi [18]-[20]; however, the main reason for recent increase of interest in complex networks was the discovery that real networks have characteristics different from uniform random graphs. Networks derived from real data, indeed, may involve *community structures*, *power-law degree distributions*, *hubs* and other particular structural features [13]. Three particular developments have contributed to the onset of interest on complex networks [13]: Watts and Strogatz's investigation of *small-world* networks [21], Barabási and Albert's characterization of scale-free models [22] and Girvan and Newman's identification of the community structures [23]. These developments have been certainly supported by the increased computing powers and by the possibility to study the properties of a plenty of large databases of real networks [12].

2.1.1. Brief outline of applications

Graph theory was applied to several areas such as geometry and mathematics, theoretical computer science, game theory and city planning; many of these are optimization problems — which graph theory successfully resolved — such as [12] "what is the maximum flow per unit time from source to sink in a network of pipes", "how to color the regions of a map using the minimum number of colors so that neighboring regions receive different colors" — leading to the so called *four color theorem* — or "how to fill n jobs by n people with maximum total utility". Complex network theory, instead, has taken place in more practical areas such as sociology, telecommunications, biology, medicine (e.g. neural networks, genetic, cancer metastasis [24]), physics and more recently on engineering [25],[26], economy and earth

science (e.g. climate [27] and earthquake occurrence [28]). Other examples of network theory applications include transportation networks, phone call networks, the Internet and the World Wide Web, the actors' collaboration network in movie databases, scientific co-authorship[12]. Such a wide range of applications is favorite by statistical mechanics elements mixed with graph theory, conferring to complex network theory more flexibility and reliability in representing virtually real complex systems.

2.2 Complex network structure: classification and measurements

Recently, particular attention has been focused on the relationship between the structure and dynamics of complex networks, while relatively little attention has been given to the subject of network measurements. However, it is only by obtaining informative quantitative features of the networks topology that they can be characterized and analyzed. Therefore, network measurements are essential resources in many network investigations, from representation and characterization to classification and modeling [13]. However, both the classification and characterization of analyzable structures (natural or human-made) using complex network theory do not provide a rigorous procedure to choose the most appropriate measurements. The choice, hence, should reflect the specific interests and applications, often implying redundancy in evaluation of topological features.

2.2.1. Main definitions and notations

In general, there are four main types of complex networks (Fig. 2.2-1): *weighted digraphs* (directed graphs), *unweighted digraphs*, *weighted undirected graphs* (or simply weighted graph) and *unweighted graphs*. Starting from the concept of weighted digraph, all the other three types can be derived, applying the *threshold* and *symmetry* operations.
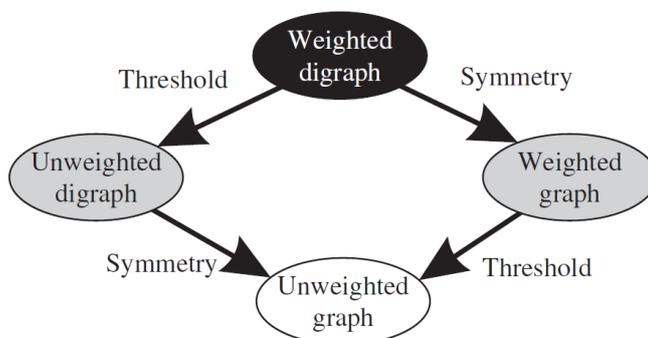
Fig. 2.2-1: The four main types of complex networks and their transformations [13].

A weighted directed graph, $G(N, M) = (\mathcal{V}, \mathcal{L})$, is defined by a set $\mathcal{V}$ of N vertices (or nodes) and a set $\mathcal{L}$ of M edges (or links). Each vertex can be identified by an integer value $i = 1, 2, \ldots, N$ while each edge can be identified by a pair (i, j) and by a scalar $\omega(i, j)$, that respectively represents a connection going from vertex i to vertex j — the link is said to be incident in nodes i and j — and its weight. Usually it is assumed that there are no *self-connections* (or *loops*) and *multiple connections*, i.e. each vertex has not a link with itself and each pair of nodes have no more than one link. Graphs with loops or duplicate connections are sometimes called *multigraphs* or *degenerate graphs* [13]; in the following it will be analyzed only non-degenerate graphs. In matrix notation, a weighted digraph can be completely represented by its *weight matrix W*, where each entries $w_{ij} = \omega(i, j)$.

Unweighted digraphs are directed graphs without a particular weight mapping and, in this case, usually, all set of links have $\omega(i, j) = cost. = 1$. Undirected graphs — both weighted and unweighted — instead, are graphs in which edges have no direction, i.e. the same connection exists from i to j and from j to i.
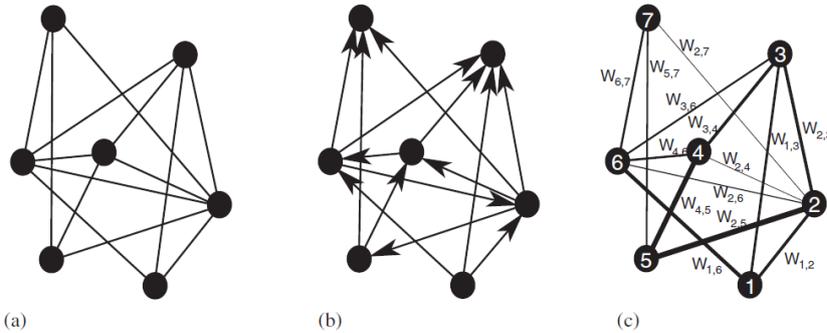
Fig. 2.2-2: Graphical representation of an undirected (a), a directed (b), and a weighted undirected (c) graph. In the weighted graph, the weights of the links are graphically represented by the link thicknesses [12].

The threshold operation permits to obtain from weighted graphs (directed or not) their unweighted counterpart. In particular, said T the threshold value, the operation of threshold is applied to the matrix W, yielding the matrix A in which the entries

$$a_{ij} = \begin{cases} 1, & if \ |w_{ij}| > T \\ 0, & otherwise. \end{cases}$$

The square matrix A is said *adjacency* (or *connectivity*) *matrix*. Moreover, a weighted digraph can be transformed into its undirected counterpart by applying the symmetry operation $W + W^T$, where W^T is the transpose of W [13].

For undirected graphs, two vertices i and j are said to be *adjacent* or *neighbors* if $a_{ij} \neq 0$; for digraphs, instead, if $a_{ij} \neq 0$ then node i is said a *predecessor* of j and node j is said a *successor* of i. The *neighborhood* of a vertex i, represented as $v(i)$, corresponds to the set of vertices adjacent to i, or in other words, the set of all nodes with which vertex i has a link. In many cases it is useful to consider only a part of the whole graph, in order to analyze local properties. A *subgraph $G'(N', M') = (\mathcal{V}', \mathcal{L}')$* of $G(N, M) = (\mathcal{V}, \mathcal{L})$ is a graph such $\mathcal{V}' \subseteq \mathcal{V}$ and $\mathcal{L}' \subseteq \mathcal{L}$ [12].

Another important concept is the one of reachability of two different nodes. In fact, in a general case, two nodes in a graph are not adjacent but they can be connected through a sequence of m edges which form a *walk*. Two vertices are said *connected* if there is at least one walk connecting them, otherwise they are said *unconnected*. If all pairs of vertices are connected, then also the graph is said connected; instead, if there is at least a pair of nodes unconnected, then the graph is said unconnected. Furthermore, a *trail* is a walk in which no edge is repeated, while a *path* is a walk in which no node is visited more than once [12]. Generally, two nodes could be connected by more than one path; however, the path of minimum length, called *shortest-path*, plays a key role in several analysis, especially in those where information transferring is relevant. With the notation $d(s,t)$ is denoted the *topological distance* between node s and t, i.e. the length of the shortest-path connecting vertices s and t. Finally, a *cycle*, C_k of length k is defined as a closed walk in which no edge is repeated (e.g. C_3 is a triangle, C_4 is a quadrilater and so on).

2.2.2. Centrality metrics

Indicators of centrality identify the most important vertices within a graph, providing, therefore, a ranking of *the most important vertices* in the network. Rankings based on centrality, however, have some limitations, such as "*how to choose the best metric*" or "*how to quantify the difference in importance between different levels of the ranking*". In the former, usually, the choice is not univocal but is dependent on the application and the network topology [29]. The latter, instead, is more subtle and it can be mitigated by applying *Freeman centralization* [30]. Moreover, the features which (correctly) identify the most important nodes in a network may be meaningless for the remaining vertices of the network. This behavior is a consequence of the heterogeneous topology of some complex networks and it explains why, as instance, only the first few results of a Google-image-search appear in a reasonable order. Therefore, Freeman centralization based on a specific centrality metric can provide some insight to the relative importance of nodes in a network.

A list of the common centrality indicators is reported below.

- **Degree Centrality, k.**

  The *degree* (or *connectivity*) k_i of a node i is the number of edges connected to that node, i.e. the cardinality of the neighborhood $v(i)$ of i. This is a simple local measure, since is based on the notion of neighborhood, following the idea that *an important node is involved in large number of interactions* [V].

  For directed networks there are two kinds of degree: the *out-degree*, k_i^{out}, equals to the cardinality of the set of successors of i, and the in-*degree*, k_i^{in}, equals to the cardinality of the set of predecessors of i,

$$k_i^{out} = \sum_{j=1}^{N} a_{ij}, \qquad k_i^{in} = \sum_{j=1}^{N} a_{ji} \ . \tag{2.1}$$

Hence, the total degree is $k_i = k_i^{out} + k_i^{in}$. For undirected networks $k_i^{out} = k_i^{in} = k_i$ (despite the use of the same notation, in this case k_i is equal to half of total degree of directed networks). The previous definitions can be also normalized, taking into account the size of the network, dividing by the total number of possible neighbors $(N-1)$,

$$k_i = \frac{1}{N-1} \sum_{j} a_{ij} \ . \tag{2.2}$$

In the following, except when explicitly specified, the normalized version of the degree centrality will be adopted, which permits a comparison between network of different size.

The *average degree* is the average of k_i for all vertices in the network

$$\langle k \rangle = \langle k^{out} \rangle = \langle k^{in} \rangle = \frac{1}{N} \sum_{i,j=1}^{N} a_{ij} \ . \tag{2.3}$$

The highest-degree nodes are often called "*hubs*". Information on how the degree is distributed among the nodes of a network can be obtained calculating its *degree distribution*, $P(k)$, defined as the probability that a node (chosen uniformly at random) has degree k, or equivalently, the fraction of vertices in a network with degree k. For digraphs, the two distributions $P(k^{in})$ and $P(k^{out})$ can be considered. Moreover, sometimes it can be advisable to calculate the *cumulative* degree distribution $P_{cum}(k)$, defined as [12]

$$P_{cum}(k) = \sum_{k'=k}^{\infty} P(k') \ . \tag{2.4}$$

- **Closeness Centrality, CC.**

  The closeness of a node is an index based on the distance of that node from all others, following the idea that *an important node is typically "close" to, and can communicate quickly with, the other nodes in the network* [V]. Despite several closeness-based metrics have been developed, some of these are commonly used. The closeness centrality can be defined as *the reciprocal of the total distance from a vertex s to all other vertices $t \in \mathcal{V}$* [31]

$$CC(s) = \left[\sum_{t \in \mathcal{V}} d(s,t)\right]^{-1} , \tag{2.5}$$

or as *the mean of shortest paths lengths from a node to others*, which is the normalized version of the (2.5). These definitions works well for connected graphs; in disconnected graphs there exists at least one pair of vertices with distance $d = \infty$, i.e. they are not connected. Therefore, for these cases it is preferred to use the *harmonic centrality* of a node s, defined as [32]

$$H(s) = \sum_{s \neq t \in \mathcal{V}} \frac{1}{d(s,t)} ,$$

with the convention $1/\infty = 0$. Unlike degree centrality, this is a global metric, since the computation involves all nodes in the network. So, vertices with a smaller total distance are more central and thus they are considered more important.

- **Betweenness Centrality, BC.**

  In networks — assuming that interactions follow the shortest paths between nodes pairs — it is possible to quantify the importance of a vertex or a edge in terms of its betweenness centrality [33], that is the number of paths in which a vertex or edge participates. The idea is that *an important node/link will lie on a high proportion of paths between other nodes in the network* [V]. Vertex betweenness centrality can be quantified as

$$BC(v) = \sum_{v \neq s \neq t \in \mathcal{V}} \frac{\sigma_{st}(v)}{\sigma_{st}} , \tag{2.6}$$

  where σ_{st} is the total number of shortest paths from node s to node t, while $\sigma_{st}(v)$ is the number of those paths that pass through node v. High values of BC suggest that a node can reach other nodes on relatively short paths, or that a vertex lies on a significant fraction of shortest paths connecting pairs of other vertices [29].

  The definition (2.6) can be normalized taking into account the total number of pairs of vertices not including v, that is $(N-1)(N-2)$ for digraphs and $(N-1)(N-2)/2$ for undirected ones.

  A related concept is *random-walk betweenness centrality* [34], which takes into account that the shortest paths might not be known or not be relevant. This metric, hence, considers random walks connecting all couples of nodes instead of the shortest paths.

Besides, calculating both closeness and betweenness centralities generally require large computational time for large networks (i.e. with a huge number of nodes/links), because of the expensive calculation of all the shortest paths in the graph. Betweenness centrality, indeed, typically requires a computational cost of the order $O(N^3)$ with the Floyd–Warshall algorithm [29]. However, faster and *ad hoc* algorithms have been developed in order to decrease the computational costs. As instance, to take advantage of the sparse nature of typical real-world graphs [13], the algorithms developed by Brande compute BC for all vertices in the graph with the order $O(N^2 log N + MN)$ for weighted graphs, and $O(MN)$ for unweighted graphs , where M is the total number of links. Finally, Bader and Madduri [29] proposed a parallel computing of centrality metrics that further reduce the computational cost in large, real networks.

- **Eigenvector Centrality.**

  A more sophisticated version of the degree centrality is the so-called *eigenvector centrality.* This metric is based on an authority classification, where *an important node is connected to important neighbors* [V]. In general, connections to vertices which are themselves influential will lend a node more influence than connections to less influential nodes. Denoting the centrality of a vertex i by x_i, this effect can be quantified as [35]

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{ij} x_j \ , \tag{2.7}$$

  where λ is a constant. Defining as $\bar{x} = (x_1, x_2, \ldots, x_N)$ the vector of centralities, the equation (2.7) can be re-written in matrix form

$$\lambda \bar{x} = A \cdot \bar{x} \ ,$$

  showing that $\bar{x}$ is an eigenvector of the adjacency matrix A, with eigenvalue λ. In general, there will be many λ for which an eigenvector exists. However, hypothesizing that $x_i \geq 0, \forall i = 1, \ldots, N$, it can be shown (using the Perron–Frobenius theorem) that λ must be the largest eigenvalue of A and $\bar{x}$ the corresponding eigenvector [35].

- **Freeman Centralization.**

  As just said, the centralization of any network is a measure of how central its most central node is, in relation to how central all the other nodes are [30]. Centralization indices depend on the centrality metric chosen, but, in general, they should have certain features in common: (i) they should index the degree to which the centrality of the most central point exceeds the centrality of all other points, and (ii) they should each be expressed as

a ratio of that excess to its maximum possible value for a graph containing the observed number of points. An acceptable index is [30]

$$C_X = \frac{\sum_{i=1}^{N}[C_X(p^*) - C_X(p_i)]}{max \sum_{i=1}^{N}[C_X(p*) - C_X(p_i)]} \ ,$$

(2.8)

where $C_X(p_i)$ is a centrality metric for a node i of the network, $C_X(p^*)$ is the largest value of $C_X(p_i)$ and the denominator is the maximum possible sum of differences in point centrality for a graph of N nodes. Being expressed as a ratio of values to their maximum, $0 \leq C_X \leq 1$, where $C_X = 0$ if and only if all $C_X(p_i)$ are equal, and $C_X = 1$ if and only if one point, i.e. p^*, completely dominates the network with respect to centrality.

Using as centrality metric the degree k, its centralization index is [30]

$$C_k = \frac{\sum_{i=1}^{N}[k_{max} - k_i]}{max \sum_{i=1}^{N}[k_{max} - k_i]} \ ,$$

(2.9)

where the denominator is equal to $(N-1)(N-2)$ if k is not normalized — this expression is obtained if the graph as *star* or a *wheel* —, while it is equal to $(N-2)$ if the degree is calculated with the Eq.(2.2).

2.2.3. Degree correlation indices

It is often useful to verify — especially for real networks — whether different vertices shows correlation between one of their metrics. The most natural approach is to consider the correlations between two connected nodes in terms of degree centrality. Degree correlations can be used to characterize networks and to validate the ability of network models to represent real network topologies. In addition, the degree correlations have strong influence on dynamic processes like instability and they are related to the network evolution process, therefore, should be taken into account in the development of new models [13]. In the following, five main correlation indices are briefly described; in general, however, a network is said to be *assortative* or *disassortative* if its vertices show a preference to attach to others vertices that are similar or different in some way. If there is no correlation, the network is said *uncorrelated* or *non-assortative* [12],[13],[36].

- **Joint Degree Distribution, $P(k, k')$.**
  The joint degree distribution (JDD) is the probability that an arbitrary edge of the network connects a vertex of degree k to a vertex of degree k' [13]. It can be quantified as a square matrix $P(k, k')$; if many nodes of similar degree are linked, then P will have large entries along and near its diagonal, while, if many node of dissimilar degree are linked, then P will have large entries far from the diagonal.

- **Conditional Probability, $P(k|k')$.**

  The conditional probability is defined as the probability that a link is incident to a pair of nodes of degrees k and k' [13]

  $$P(k'|k) = \frac{\langle k \rangle P(k, k')}{k \, P(k)} \ .$$

  <div align="right">(2.10)</div>

  That satisfies the normalization $\sum_{k'} P(k'|k) = 1$ and the degree detailed balance condition $[k \, P(k'|k)P(k)] = [k' \, P(k|k')P(k')]$ [12],[13]. While the degree distribution $P(k)$ completely determines the statistical properties of uncorrelated networks, for correlated ones (such as a large part of real networks) it is necessary to use the conditional probability.

- **k-nearest Neighbors, k_{nn}.**

  Because of finite size N of networks, the evaluation of the JDD and conditional probability in practice is difficult, providing extremely noisy results especially for fat-tailed distribution. This problem can be addressed by computing the *average degree of the nearest neighbors* of vertices with degree k, [12],[13]

  $$k_{nn}(k) = \sum_{k'} k' \, P(k'|k) \ ,$$
  $$k_{nn,i} = \frac{1}{k_i} \sum_{j \in v_i} k_j \ ,$$

  <div align="right">(2.11)</div>

  with v_i the set of first neighbors of vertex i. If there are no correlations then $k_{nn}(k)$ is independent of k, resulting $k_{nn} = \langle k^2 \rangle / \langle k \rangle$; conversely, correlated networks are said assortative (disassortative) if $k_{nn}(k)$ is an increasing (decreasing) function ok k [12],[36].

- **Pearson Degree Correlation Coefficient, r.**

  Another way to quantify the correlation in assortative networks is the Pearson degree correlation coefficient applied to each pair of linked nodes. Similarly to definition in Eq.(1.6), the result lies in the range $[-1, +1]$, where negative values indicate that vertices of dissimilar degree tend to be linked, while positive values indicate that vertices of similar degree tend to be linked. More details regarding quantitative expressions are well explained in [12],[13] and [36].

- **Rich-Club Metric, ϕ.**

    Degree correlation, especially Pearson correlation, is often incorrectly interpreted to be an indicator of *the tendency of high degree vertices to link to each other*, whereas it is an indicator *of the tendency of similar degree nodes to link to each other* (Fig. 2.2-3). For instance, in science, influential researchers of some areas tend to form collaborative groups and publish papers together [13]. This phenomenon is known as *Rich-club* and it can be measured by the Rich-club coefficient $\phi(k)$ of degree k, defined as [37]

    $$\phi(k) = \frac{E_{>k}}{N_{>k}(N_{>k} - 1)/2} \, ,\tag{2.12}$$

    where $E_{>k}$ and $N_{>k}$ are respectively the number of edges between vertices of degree greater than or equal to k, and the number of those vertices. When $\phi(k) \to 1$ for $k \to k_{max}$ then it is interpreted that high degree nodes of the network are well connected.

    A high Rich-club coefficient implies that the hubs are well connected, and global connectivity is resilient to any one hub being removed. However, one must notice that vertices with higher degree will be naturally more likely to be more densely connected than vertices with smaller degree simply due to the fact that they have more incident edges [38]. In these situations it can be useful to evaluate a normalized version of $\phi(k)$, dividing by the Rich-club coefficient of an equivalent random network with the same degree distribution $P(k)$.
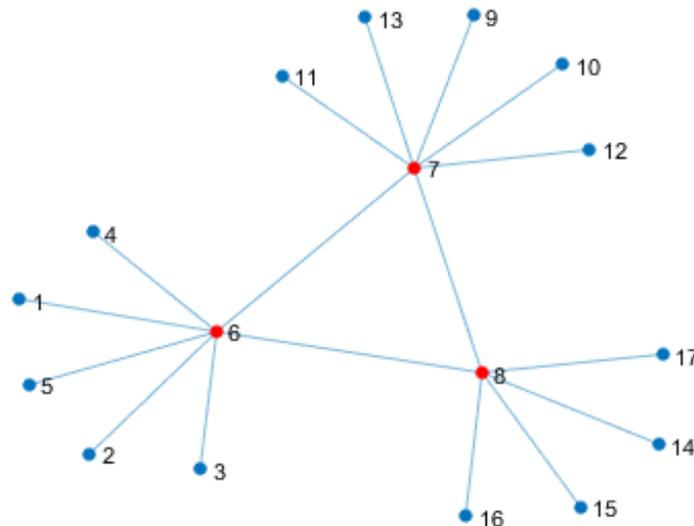
Fig. 2.2-3: An example of disassortative network that shows the Rich-club effect (nodes 6, 7 and 8).

2.2.4. Other measurements

Although centrality metrics play a fundamental role in the characterization of a large number of complex networks — especially the degree and the betweenness centralities — several other parameters are important and meaningful in network analyses. These measurements arise from the necessity to investigate network properties regarding the relative distance between vertices or the tendency to form sets of tightly connected vertices (clusters or communities).

- **Clustering Coefficient, C.**

    One way to characterize the presence of *triangles* (defined as *loops of order three* or *sets of three vertices with edges between each pair of vertices*) in a network is through the *clustering coefficient*, also known as *transitivity* [12],[13]. In other words, the clustering coefficient gives the probability that two randomly chosen neighbors of a node i are also neighbors. It can be defined a *local* and a *global* clustering coefficient. The local clustering coefficient is [27]

$$c_i = \frac{e(v_i)}{k_i(k_i - 1)/2} \, ,$$

(2.13)

   that is the fraction of links between the vertices $e(v_i)$ within the neighborhood v_i of i to the total number of possible links in v_i.

   A first global clustering coefficient is then

$$C = \frac{1}{N} \sum_{i=1}^{N} c_i \, .$$

(2.14)

   By definition, $0 \leq c_i \leq 1$ and $0 \leq C \leq 1$; however, if $k_i = 1$, $c_i \to \infty$, then some tools force that if $k_i = 1$, $c_i = 0$, other instead $c_i = 999999998$. In this work the former choice is adopted.

   A second definition for the global clustering coefficient is said *transitivity* and it is calculated for undirected, unweighted networks as [13]

$$C = \frac{3N_\Delta}{N_3} \, ,$$

(2.15)

   where N_Δ is the number of triangles in the network and N_3 is the number of connected triples. The factor three accounts for the fact that each triangle can be seen as consisting of three different *connected triples*, giving, also in this definition, $0 \leq C \leq 1$. A connected triple is a set of three vertices where each vertex can be reached from each other (directly or indirectly) [13].

Values calculated with the two definitions of C are different because the "average" in Eq. (2.15) gives the same weight to each triangle while Eq. (2.14) gives the same weight to each vertex, resulting in different values because vertices of higher degree are possibly involved in a larger number of triangles than vertices of smaller degree [13].

- **Average Physical Distance, L.**

  The *average physical distance* of a node is here introduced and defined as

$$L_i = \frac{1}{(N-1)k_i} \sum_{j \in v(i)} l(i,j) \ , \tag{2.16}$$

  where $l(i,j)$ is the physical distance between node i and another node j of its neighborhood, and k_i is the normalized degree centrality of i in Eq.(2.2). Since in this work all the networks are *spatial* networks (cf. section 2.3), then an alternative to the *average topological distance* — that is proportional to the inverse of the closeness centrality, Eq. (2.5) — is introduced. The calculation of L_i, then, let to avoid the high computational cost due to the calculation of all shortest-path lengths (such as in the betweenness and closeness centralities). The mean value of L_i is obtained averaging on all nodes N as

$$\langle L \rangle = \frac{\sum_i L_i}{N} \ .$$

  Later, L can be a very useful parameter to characterize the average size of the neighborhood of vertices.

- **Edge Density, ρ.**

  The *edge density* is defined here as

$$\rho = \frac{\# \ active \ links}{\# \ total \ possible \ links} \ , \tag{2.17}$$

  and it will be used to characterize the overall dimensions of the networks.

- **Modularity, Q.**

  Many real networks show the presence of groups whose vertices are more densely interconnected to one another than with the rest of the network. These groups are commonly referred as *communities* and their identification in large networks is useful because vertices belonging to the same community are more likely to share properties and dynamics [13]. Despite the importance of the concept of community, there is no

consensus about its definition. Communities are defined in a *strong sense*, i.e. if all vertices of a subgraph have more connections between them than with the rest of the network, and in a *weak sense* if the sum of all vertex degrees inside the subgraph is larger than outside it [13]. Newman and Girvan [39] proposed a measurement, called *modularity*, Q, to quantify the division of networks. Modularity is *the fraction of the edges that fall within the given groups minus the expected such fraction if edges were distributed at random*. If a network is split in c communities then a symmetric $c \times c$ matrix, E, can be defined where its elements along the main diagonal e_{ii} give the fraction of connections between vertices in the same community i while the other elements, e_{ij} ($i \neq j$) identify the fraction of connections between vertices in the different communities i and j [13]. Therefore, if the network is split into two communities, $c = 2$, and Q can be calculated as [13]

$$Q = \frac{1}{4M} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2M} \right) s_i s_j = Tr(E) - \|E^2\| \, , \qquad (2.18)$$

where $Tr(E)$ is the *trace* of matrix E, $s_{i,j} = \pm 1$ according to the nodes i and j belongs to community 1 or 2 (of course the indices $i, j = 1, \ldots, c$). If a particular division gives no more within-community edges than would be expected by random chance we will get $Q = 0$. Values other than zero indicate deviations from randomness, and in practice values greater than about $Q = 0.3$ appear to indicate significant community structure [40]. Furthermore, there exist several methods to find community structures in networks such as spectral methods, agglomerative and divisive methods or a method — proposed by Newman — based on maximization of the modularity. Optimizing Q theoretically results in the best possible partition of a given network; however, going through all possible iterations of the nodes into groups is impractical (graph partition problems fall under the category of *NP-hard* problems) so heuristic algorithms are used. An example of heuristic method based on modularity maximization to find communities is the Louvain method [41] that appears to run with a computational cost $O(N \, logN)$. Recently, Newman [42] proposed a method which reformulates the modularity concept in terms of the eigenvectors of the so called *modularity matrix*

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2M} \, .$$

Thus, Newman proposed a new definition of communities as *indivisible sub-graphs*, i.e. sub-graphs whose division would not increase the modularity. Currently, this method is believed to be one of the most precise, as it is able to find a division with the highest value of modularity for many networks [13].

2.3 Real and spatial networks

As said before, many systems in nature and in technology can be approximated as a complex network, since they present a large number of highly interconnected elements. Naturally, an approximation cannot provide every details of the real system modeled. Nevertheless, in many cases of practical interest, such an approximation provides a simple but still very informative representation of the entire system. In last decades, analyzing many topological features of a lot of real networks, it was found out that, despite the intrinsic differences, most of the real networks are characterized by the same topological properties, such as relatively small characteristic path lengths, high clustering coefficients, fat-tailed shapes in the degree distributions, degree correlations, and the presence of community structures [12],[13]. For instance, high clustering coefficients and power-law degree distributions are ubiquitous in most real networks [12],[13].

In regular hyper-cubic lattices in D dimensions, the mean number of vertices one has to pass by in order to reach an arbitrarily chosen node grows with the lattice size as $N^{1/D}$. Conversely, in most of the real networks, despite of their often large size, there is a relatively short path between any two nodes. This feature is known as the *small-world property* and is characterized by a small average-shortest-path length [12]. This property has been observed in a variety of real networks, including biological and technological ones. Differently from random graphs, the small-world property in real networks is often associated with the presence of clustering, denoted by high values of the clustering coefficient [12]. Watts and Strogatz [21], have proposed to define small-world networks as those networks having both a small value of average-shortest-path length, like random graphs (as instance, Erdős-Rényi random networks [18]-[20]), and a high clustering coefficient, like regular lattices [12],[13]. Additionally, it was found that most of the real networks — unlike regular lattices or random graphs — display power-law shaped degree distributions $P(k) \sim k^{-\gamma}$, with $2 < \gamma < 3$. Such networks have been named *scale-free networks*, because power-laws have the property of having the same functional form at all scales [12]; of course, this does not necessarily implies that such graphs are scale-free with respect to other measurable structural properties. In finite-size networks, fat-tailed degree distributions have natural cut-offs. When analyzing real networks, it may happen that the data have a rather strong intrinsic noise due to the finiteness of the sampling. Therefore, in these cases, to smooth the statistical fluctuations present in the tails of the distribution, it is sometimes suitable to use the cumulative degree distribution — cf. Eq.(2.4).

A particular class of networks is that of *spatial networks*, i.e. networks whose nodes occupy a precise position in Euclidean, real space (two or three-dimensional), and whose edges are real physical connections. There are a lot of examples of spatial networks in many areas, such as in neural networks, information/communication networks, electric power grids, transportation systems, nature and engineering applications networks [12],[25]-[28]. It is not surprising that the topology of spatial networks is constrained by their geographical embedding [12]. Some "conditioned" topological characteristics of such networks regard, for instance, the node degree,

since the number of edges that can be connected to a single node is limited by the physical space to connect them. The fact that long-range connections are constrained by the Euclidean distance has also important consequences on the small-world behavior.

Csányi and Szendrői [43] have worked out an alternative definition to characterize the small-world behavior, proving that networks with strong geographical constraints are not small worlds, suggesting an indication of the small-world scaling as an exponential relation between the number of nodes that can be reached from vertex i in at most r steps, and the steps r (instead of the average shortest path, which scales logarithmically with the network size N) [12]. Besides, many spatial networks show trivial clustering-degree correlations; many other real networks, however, show a hierarchical behavior where $C(k) \sim k^{-1}$. In networks with strong geographical constraints, hierarchy is absent because of the limitations imposed by the topology [12].

2.4 Large graph visualization and analysis tools

A large graph (equivalently a network) is typically classified as *large* if it is made up of a great number of nodes (usually $N > 10^4$) and links (usually $M > 10^6$). Since a lot of real complex networks are large graphs (such as in sociology, biology, and computer science [44]), it is crucial to have tools able to handle easily these networks, both in visualization and analysis. Some tools are free to download while other tools have appeared in research papers but are not available for public use. In general, there are few software suitable for all the aspects required by network analyses coming from different research areas; conversely, most of tools are *ad-hoc* software for specific areas. Some of these tools are: *igraph, Gephi, Cytoscape, Tulip, WiGis, CGV, VisANT, Pajek, In Situ Framework, Honeycomb, UCINET, Graphviz, InfiniteGraph, R, Mathematica, Matlab*. In this work were used mainly *Matlab*, *Tulip*, *Pajek* and *VisIt*, where the last one is an open source tool for visualizing and analyzing different types and very large data [VI]; as a check tool in many operations was also used *Gephi*.

Chapter 3:

Building the Turbulent Network

In the previous chapters a general overview about turbulence and complex network theory was presented, particularly introducing their main properties and measurements. In this chapter, instead, the set of hypotheses and procedures necessary to build the network, based on a turbulent flow, will be provided.

In Fig. 3-1 the *workflow* of the entire work is shown, in order to give a visual, easy summary of each stage.
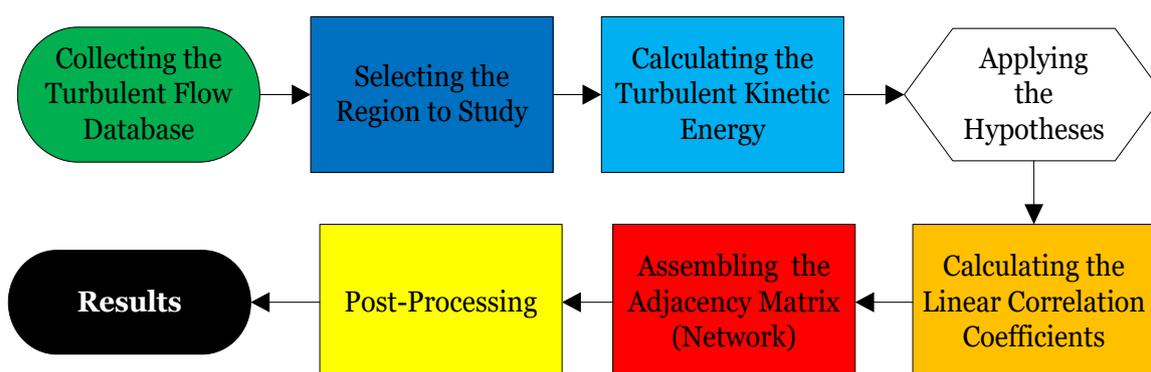
Fig. 3-1: Workflow diagram: a flowchart representation.

In the next sections each block of the workflow will be explained in detail, from the collecting of the turbulent flow database to the formation of the adjacency matrix, which represents entirely the network, for the established hypotheses. The last two blocks, i.e. the post-processing and the comparing of results, instead, will be treated in the next chapter.

3.1 Turbulence and networks: a novel approach

As mentioned previously, in Chapter 2, the study of complex networks is a young area of scientific research, which took hold mainly in the last two decades. Furthermore, complex network theory found a wide range of application to real world problems [24]-[28] revealing to be a very powerful tool to analyze the high complexity of real world systems.

In turbulence, instead, few and very recent network-based approaches have been proposed. The main works in this research area concern: the study of two-phase flows [25],[47]-[50], primarily with the aim to identify flow-patterns, which results an hard task due to the complexity of such flows; the study of turbulent jets in order to map stochastic processes [51] or discriminating various regions of the jet relative to its axis [26]; the study of reacting [52] and fully developed turbulent flows [53],[54].

In general four approaches of mapping time series to complex networks have been proposed [55]:

- Visibility graph (also visibility algorithm) [2];
- Recurrence networks (also recurrence plots) [3];
- Phase-space networks [56];
- Correlation networks, which uses linear correlation coefficients as the criterion to connect nodes [57].

Most of works concerning turbulence or turbulent flows focuses on the first two approaches, i.e. recurrence plots and visibility algorithm, while a minor part uses the phase-space networks and the correlation networks.

This work, instead, is placed in a novel perspective (both for the methodology — by using a correlation network —, as well as for the objective and purpose of the study) with respect to other works that deal with complex networks in turbulent flow.

3.2 Description of the turbulent flow database

The present work uses the forced isotropic *coarse* turbulence dataset from *Johns Hopkins Turbulence Databases* (JHTDB), which consists of a DNS of a forced isotropic turbulence, with a 1024^3 periodic spatial grid and 1024 timesteps available [45],[VII]. The dataset chosen is part of a bigger database, which includes also a DNS of a forced magneto-hydro-dynamic turbulence (MHD), a DNS of a channel flow and a DNS of a homogeneous buoyancy driven turbulence, comprising more than 230 Terabytes of data. A summary of the main characteristics of the turbulent field, with the respective formulas, are listed below.

Domain:	$(2\pi)^3$
Grid:	1024^3
Viscosity	$\nu = 1.85 \cdot 10^{-4}$
Simulation timestep	$\Delta t = 2 \cdot 10^{-4}$
Storing time interval	$\delta t = 0.002 = \Delta t/10$
Stored time	$t \in [0, 2.048]$

Table 3.1: Turbulence simulation parameters [VII].

Total kinetic energy	$E_{tot} = \langle \sum_k \frac{1}{2} u \cdot u^* \rangle_{time} = 0.695$
Dissipation	$\varepsilon = \langle \sum_k (\nu k^2 u \cdot u^*) \rangle_{time} = 0.0928$
R.m.s. velocity	$u' = \sqrt{2E_k/3} = 0.681$
Taylor microscale	$\lambda = \sqrt{15\nu u^2/\varepsilon} = 0.118$
Taylor-scale Reynolds	$Re_\lambda = u\lambda/\nu = 433$
Kolmogorov timescale	$\tau_\eta = \sqrt{\nu/\varepsilon} = 0.0446$
Kolmogorov lengthscale	$\eta = (\nu^3/\varepsilon)^{1/4} = 0.00287$
Integral scale	$L = \frac{\pi}{2u^2} \int \frac{E(k)}{k} dk = 1.376$
Large-eddy turnover-time	$T_L = L/u' = 2.02$
Distance between nodes	$d = \frac{\Delta d_{domain}}{\Delta d_{grid}} = \frac{2\pi - 0}{1024 - 1} = 6.142 \cdot 10^{-3}$

Table 3.2: Statistical characteristics of turbulence (averaged over the stored time $t \in [0, 2.048]$) [VII].

Focusing on the forced isotropic turbulence dataset, energy is injected by keeping constant the total energy in modes such that their wave-number magnitude $|k|$ is less or equal to 2. After the simulation has reached a statistical stationary state,

1024 frames of data — which includes the three components of the velocity vector and the pressure field— are generated and stored into the online database. The duration of the stored data is about one large-eddy turnover time T_L. The domain of the field is a $(2\pi)^3$ cube, i.e. a range of x, y and z coordinate of $[0, 2\pi]$, while the simulation time interval is $t \in [0, 2.048]$ with a timestep $\Delta t = 0.0002$. However, data are stored every 10 DNS timesteps in the *coarse* version, obtaining the just mentioned 1024 timesteps available. There is another dataset available to download, a finer one made for testing purposes, that acquired the data every single time step.

In Fig. 3.2-1 and Fig. 3.2-2, the radial kinetic energy spectrum of the dataset, the total kinetic energy and Re_λ, as function of time, are shown; in particular, the scaling $k^{-5/3}$ Kolmogorov law is well represented for the inertial subrange. This scaling also applies for smaller k values (i.e. large scales), because of the forcing that acts on bigger scales. The dashed lines represents data before storing the database, that is before reaching a statistically stationary state ($t \geq 0$).
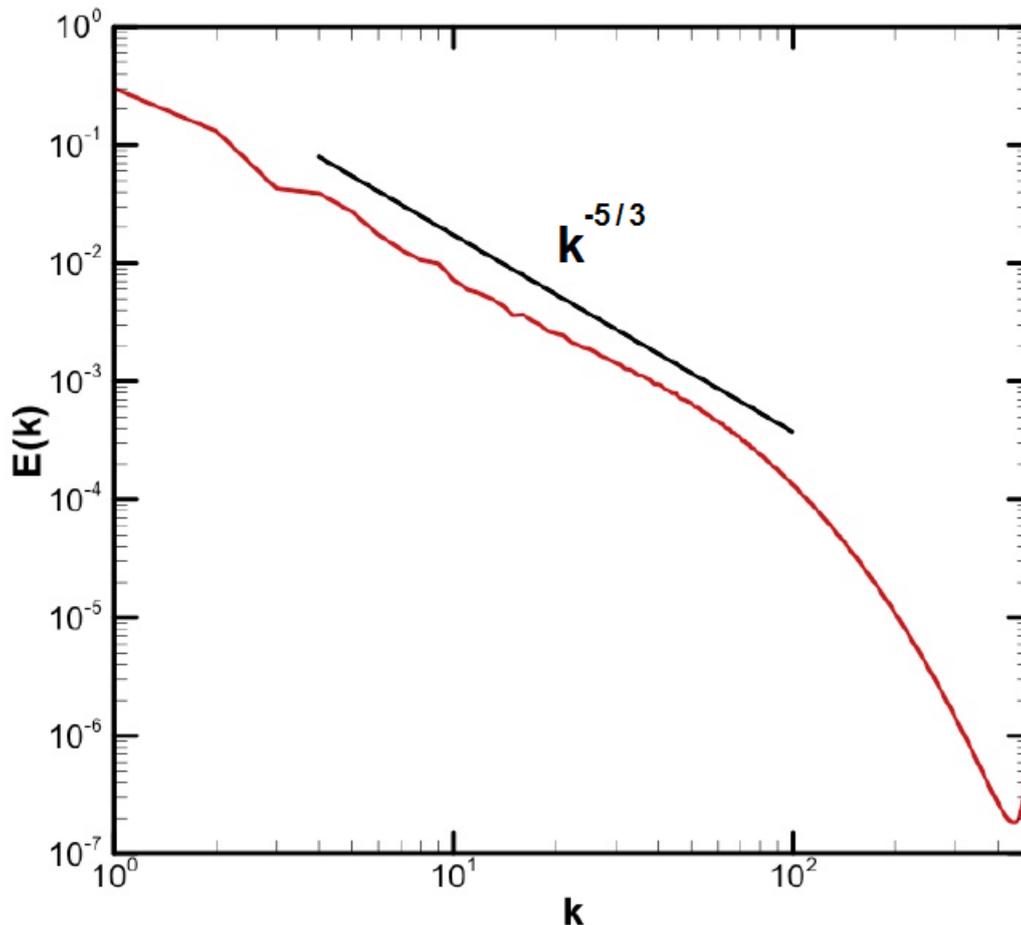
Fig. 3.2-1: Radial kinetic energy spectrum, averaged in time t = [0, 2.048] [VII].
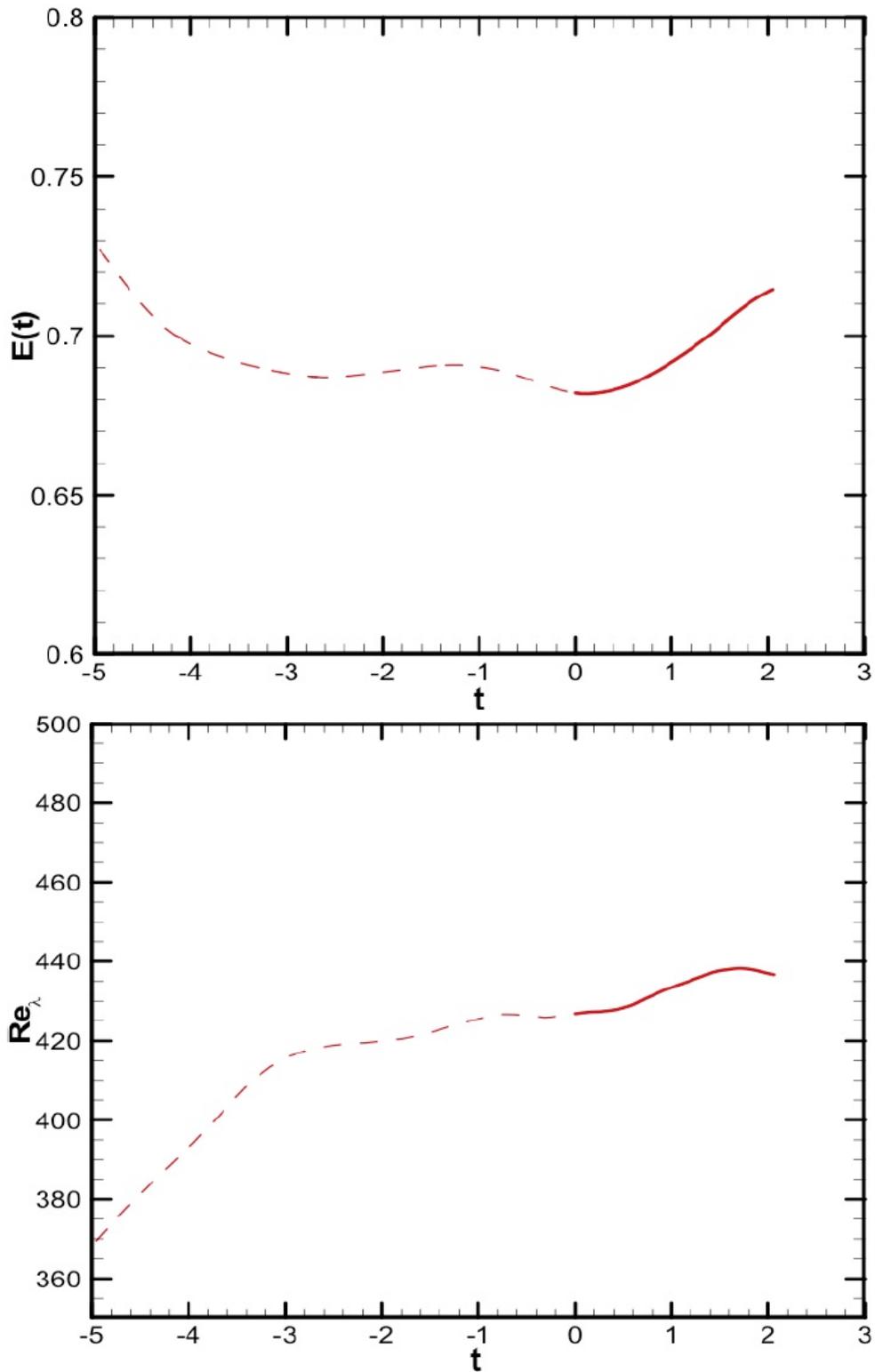
Fig. 3.2-2: Total kinetic energy (upper panel) and micro-scale Reynolds number (lower panel) as function of time. Data corresponding to the database is show using solid line between t=0 and 2.048 [VII].

3.2.1. Collecting the data: Pre-processing

In this part, the pre-processing operations — summarized in the Fig. 3.2-3 — are briefly explained.

Fig. 3.2-3: Pre-processing stage of the workflow diagram.

In particular, in order to acquire the data from the stored database, it was used the new Beta version of the *Cutout Service* provided by the website [VII], in which several options are available, such as the *authorization token* (Fig. 3.2-4), obtained thanks to a website operator. From an operative point of view, the user can firstly select a dataset from the list; in this case it was selected the *Isotropic Coarse* dataset. After, it can be chosen a data field among velocity, pressure (in format *hdf5*, *.h5* file extension) and vorticity fields (in format *.vtk*). Finally, the cutout can be performed specifying the time and space intervals. However, during the downloading process, due to some limitation of the database server, it was possible to download cutouts of limited dimensions.

Fig. 3.2-4: Screenshot of the new data-cutout service. [VII]

In this work, two parts of the domain were downloaded and from now these will be referred as *regions*. In detail, the first region — which will be referred as *region-1* from now — is a cube with 83 nodes for each side and all 1024 times, centered in the node

$(i_x, j_y, k_z) = (391, 391, 512)$, where (i_x, j_y, k_z) are the coordinates expressed in the nodes grid reference system. The second region — referred as *Region-2* in the following — is also a cube (with the same dimensions of the previous one) but it is centered in the node $(i_x, j_y, k_z) = (530, 673, 475)$. Region-1 is the objective of analysis of this work, while the region-2 is analyzed as a comparison with region-1.

Because of download limitations each file acquired corresponds to one of the 83 Z-section of the cube, i.e. 83 square grids of (83×83) nodes normal to the Z-axis, with 1024 times for each node. A *MatLab* script was then created to read the fields downloaded and re-assemble them in a $(83 \times 83 \times 83 \times 1024)$, $4D$-matrix which consists of the velocity or the vorticity vector fields time series.

It's worth the effort to specify from now that — in order to avoid notation confusions — only one reference system it is adopted, that is *the global reference system* with origin in the node $(i_O, j_O, k_O) = (0, 0, 0)$ and that corresponds to the physical point $(X_O, Y_O, Z_O) = (0, 0, 0)$. The set of all points of the region selected to analyze are expressed in node-coordinates, although for practical and computational uses they are converted in physical coordinate with the transformation relation

$$(X_N, Y_N, Z_N) = \frac{2\pi}{1023}(i_x, j_y, k_z) \; . \tag{3.1}$$

When the temporal series of the three components of the velocity field are stored, the turbulent kinetic energy scalar field, E, can be computed for each node identified by the set of coordinates (i, j, k) as

$$E = \frac{1}{2}(U^2 + V^2 + W^2) \; , \tag{3.2}$$

where U, V and W are the three components of the vector velocity field $\bar{U} = (U, V, W)$, respectively along (X, Y, Z) axes. Furthermore, the modulus of the vorticity field

$$|\bar{\omega}| = \sqrt{\omega_x^2 + \omega_y^2 + \omega_z^2} = |\nabla \times \bar{U}| \; , \tag{3.3}$$

is stored (provided directly from the on-line database) in order to show the time evolution of the regions selected.

3.3 Network-building hypotheses

At this step of the workflow, it is necessary to state the hypotheses that each point of the regions considered must fulfill in order to be classified as a node of the network.

First of all, in order to avoid directional preferences, it was decided to work in spherical symmetry. Therefore, for each cubical region selected, an internal sphere is obtained; the centers of the two spheres are the same two points indicated previously for the cube, i.e. $C = (391, 391, 512)$ and $C' = (530, 673, 475)$ respectively, while the radii are both chosen equal to $r_c = 0.24$ that is about twice the Taylor micro-scale λ (cf. Table 3.2). Hence, each point inside the sphere of radius r_c is a potential node of the network. The 4-D matrix of the time series of the energy field E, then, is re-shaped both in a $N_T \times N_S$ matrix and three $N_S \times 1$ coordinates vectors, where N_T is the number of temporal observations (equals to 1024 if all times are considered), while N_S is the total number of points selected inside the sphere of radius r_c.

For $r_c = 0.24$ it is found $N_S = 250023$. Each entry of the reshaped matrix corresponds to a temporal "measured" value of the energy of a point inside the sphere, while the other three vectors gives the (X, Y, Z)-coordinates of that point (Fig. 3.3-1).
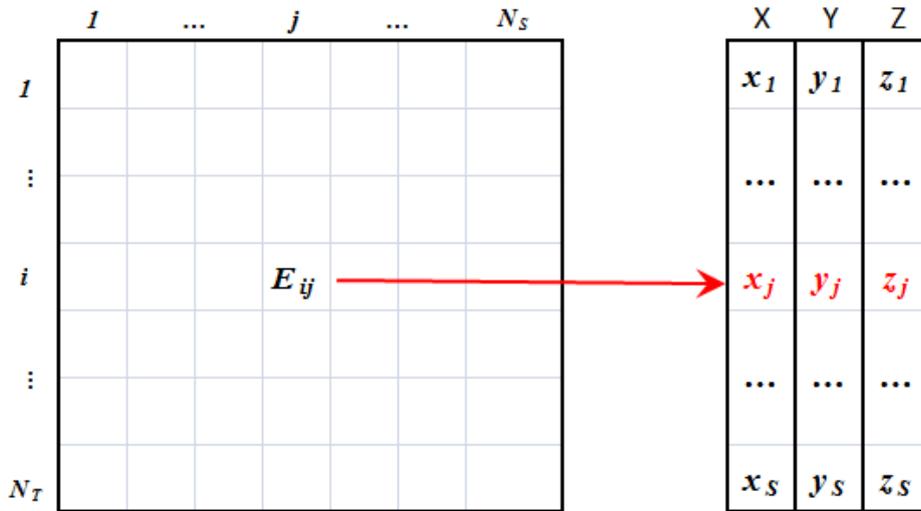
Fig. 3.3-1: Sketch of the matrix with temporal observations of the kinetic energy of the N_S nodes and their coordinates,.

After that, other three hypotheses are stated; specifically, a pair of distinct points $i, j = 1, \ldots, N_S$ are said to be two connected nodes of the network (i.e. they have a direct link) if and only if:

1) at least one between the two nodes lies inside the internal sphere with radius $r_i = 0.12 \cong \lambda$ and concentric to the sphere of radius r_c;

2) the physical distance l between the two nodes i and j is $l_{ij} \leq r_i$;

3) the linear correlation coefficient R_{ij} between the two nodes, computed for the N_T temporal observations, is $|R_{ij}| > \tau$, with τ a threshold value included in the range $[0, 1)$.

Since only different nodes are selected (that is $i \neq j$), the networks obtained will have not self-connections. The first two hypotheses are geometrical constraints while the third guarantees that the two nodes are sufficiently "linked", i.e. spatially correlated, over time. In particular, the hypothesis 1) reduces the region of study to a smaller sphere with radius nearly equals to the Taylor micro-scale which, as stated in Eq. (1.14), can be seen as the characteristic dimension of the smallest dynamically significant eddies of the flow. This choice is due to the fact that it is considered relevant what happens at scales of the order of λ or smaller, where the spatial correlation is high. The smaller sphere, then, can be seen as an influence sphere for the center node; however, to extend the possibility that each node inside the internal sphere has the same influence sphere, the hypothesis 2) is stated. The Fig. 3.3-2 shows a sketch of the boundaries of the regions and summarizes the geometrical restrictions.
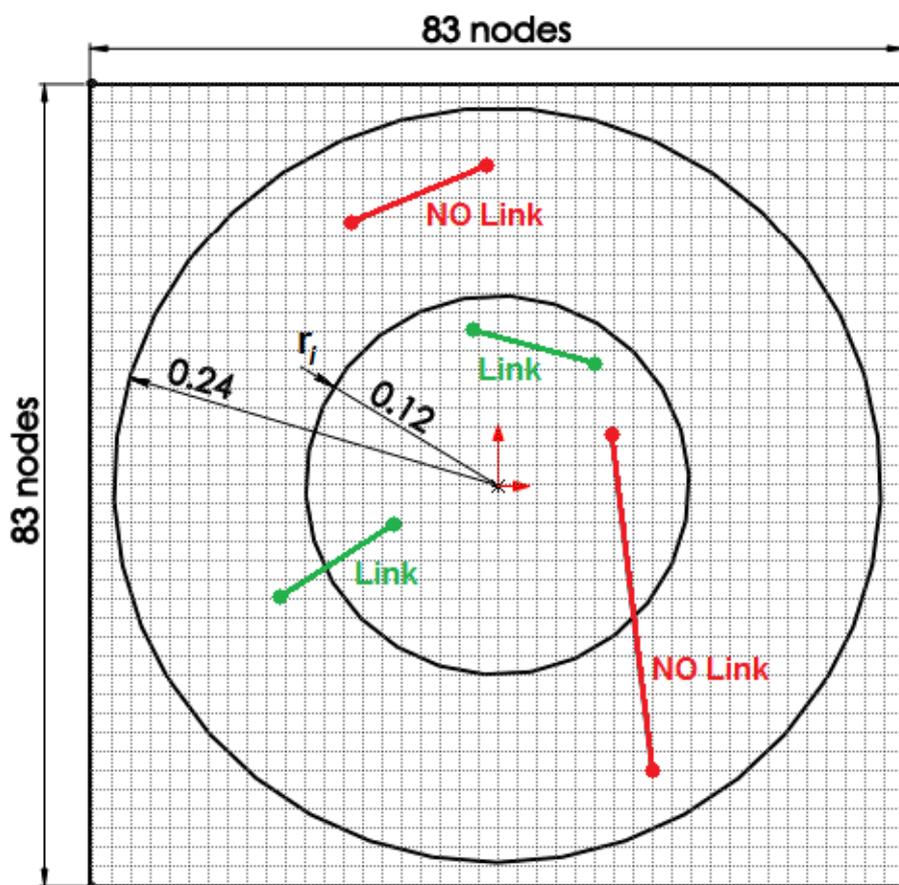
Fig. 3.3-2: Graphical representation of the domains (the first cubic, the others spherical) processed and a visual review of the possibility to active a link by a pair of nodes, as stated above in the three hypotheses.

In such a way, every node included in the inner sphere has the same potential number of links that are discerned, then, by the absolute value of the correlation coefficient $|R_{ij}| > \tau$. The selection of the threshold, τ, is a non-trivial aspect of building the network [46]; indeed, it should be set in order to take into account both to evidence strong spatial correlations and to provide a manageable number of nodes. In the following, to conclude, a part of the turbulent kinetic energy time history and of the modulus of the vorticity field for the region-1 are shown.
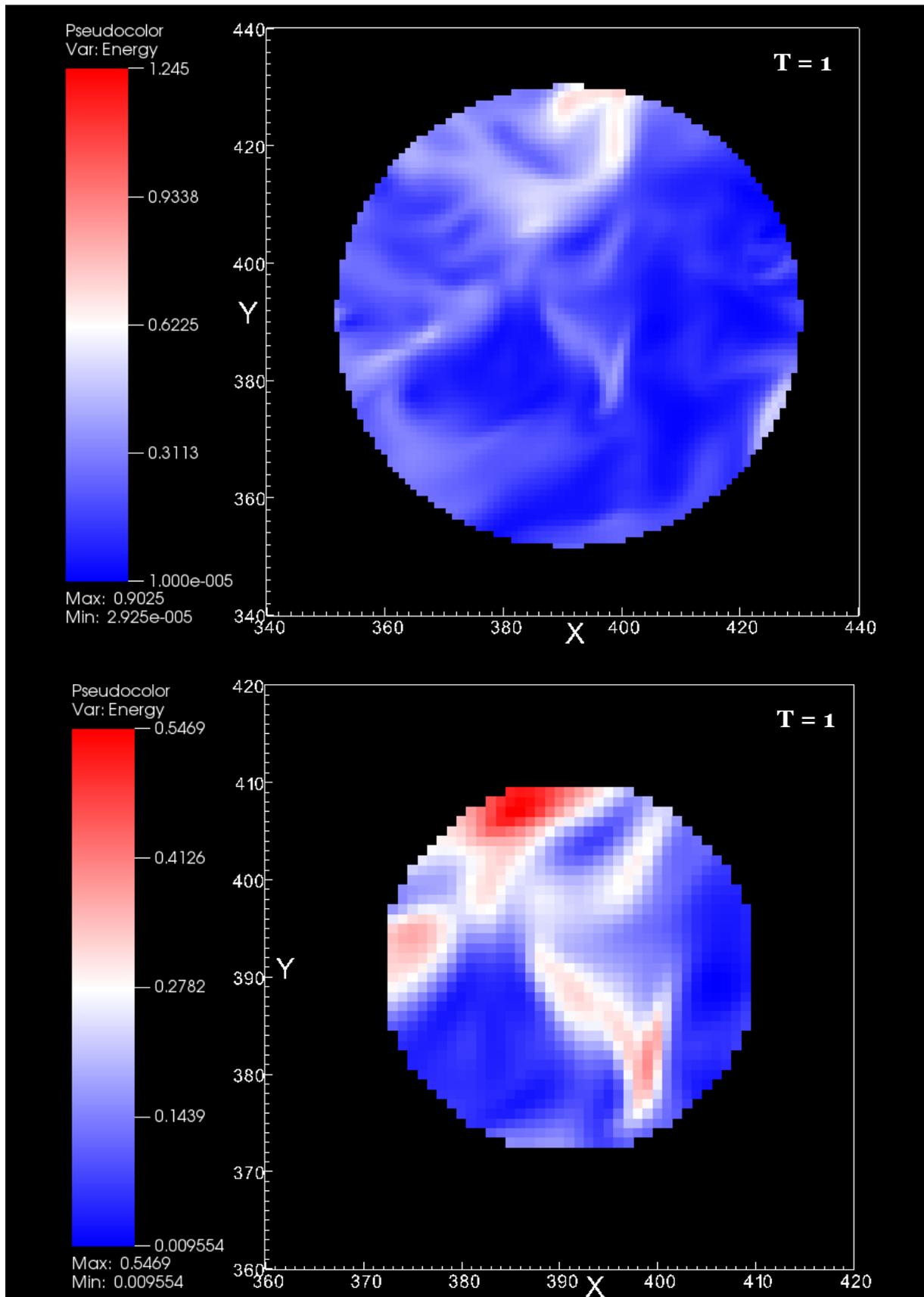
Fig. 3.3-3: Time evolution of the kinetic energy field in the section Z=512, T=1. Upper panel, section of the outer sphere (a unique scale with the other sections); lower panel, a zoom on the inner sphere (with specific color-scale).
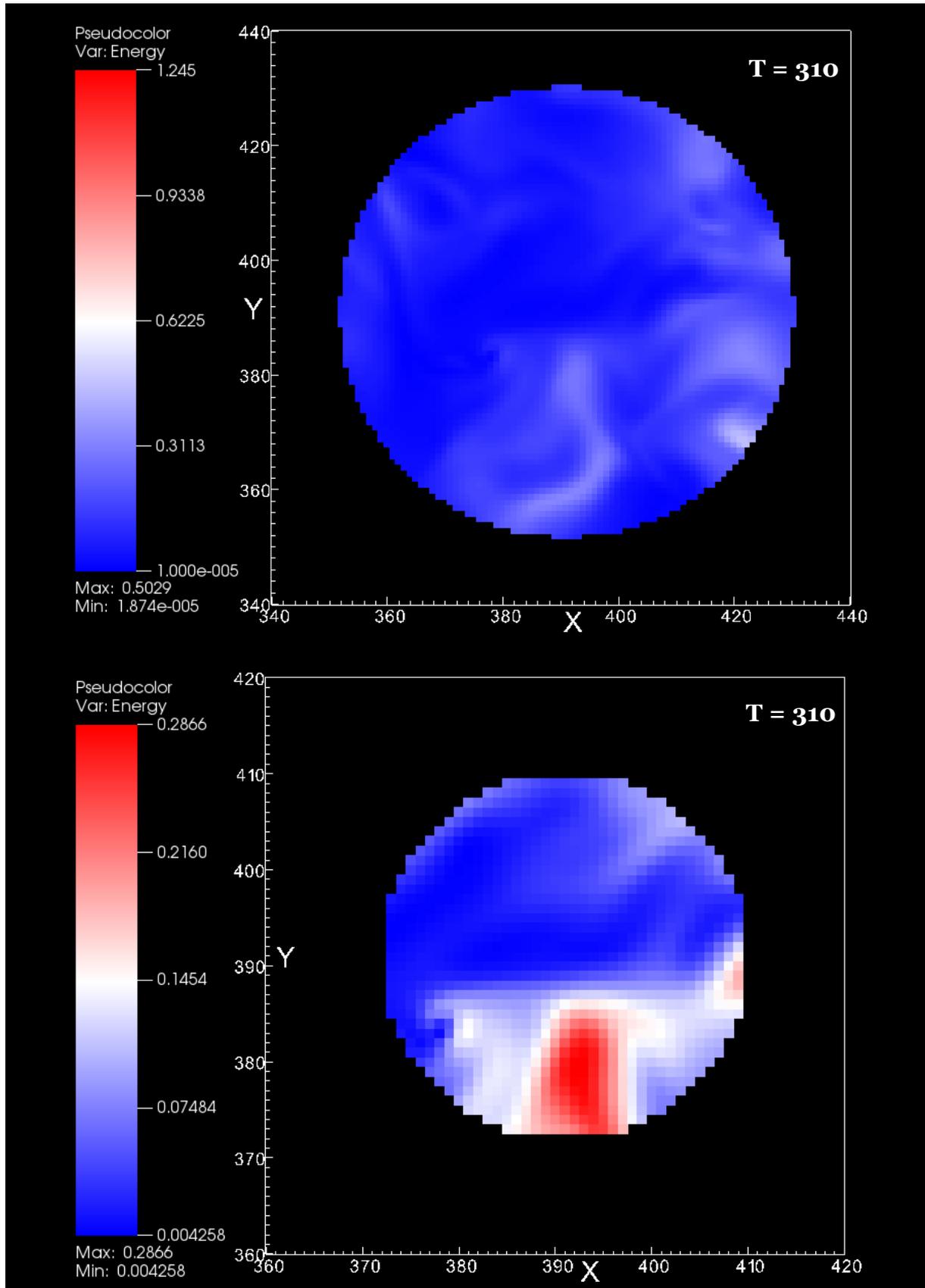
Fig. 3.3-4: Time evolution of the kinetic energy field in the section Z=512, T=310. Upper panel, section of the outer sphere (a unique scale with the other sections); lower panel, a zoom on the inner sphere (with specific color-scale).
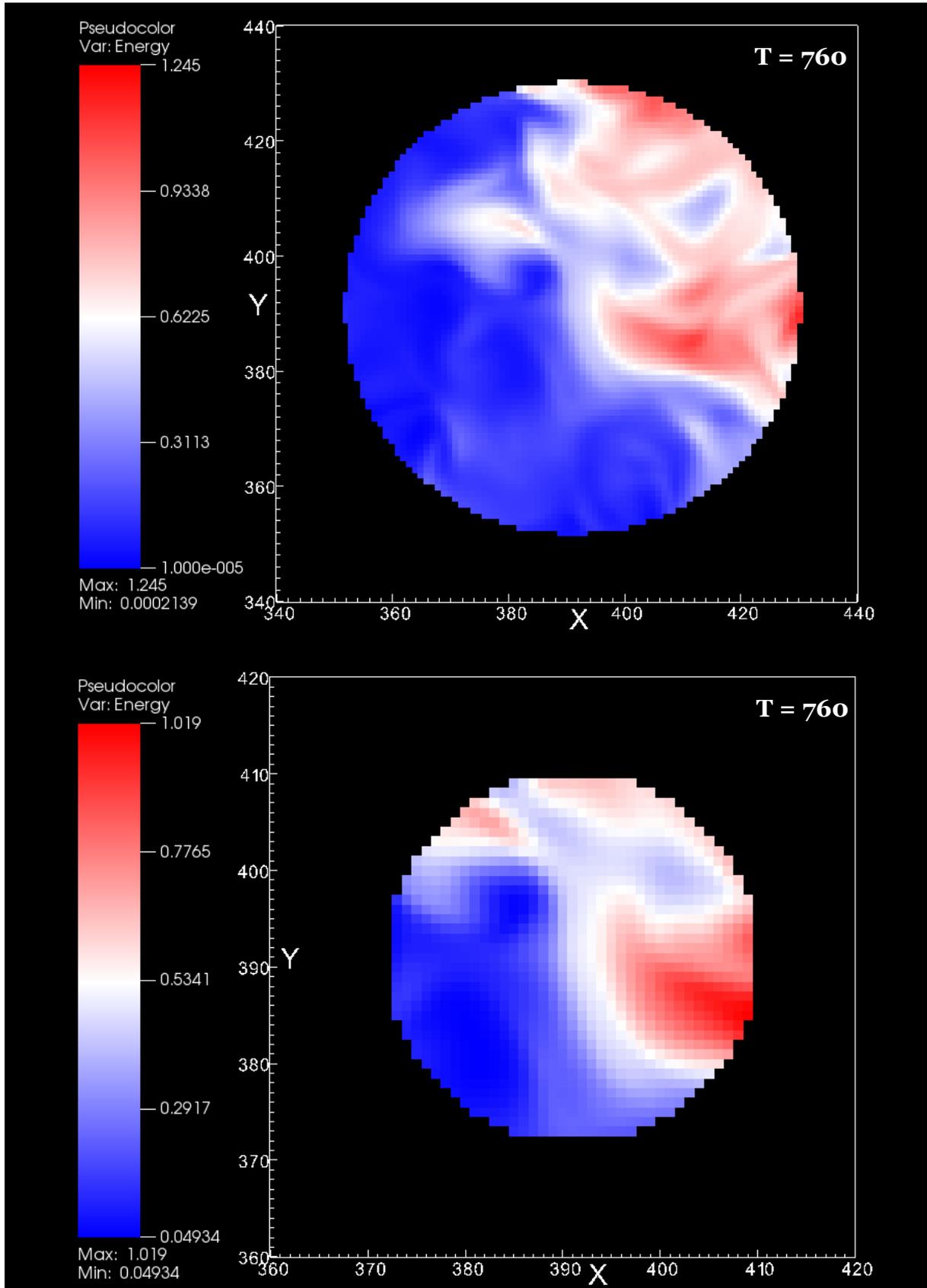
Fig. 3.3-5: Time evolution of the kinetic energy field in the section Z=512, T=760. Upper panel, section of the outer sphere (a unique scale with the other sections); lower panel, a zoom on the inner sphere (with specific color-scale).
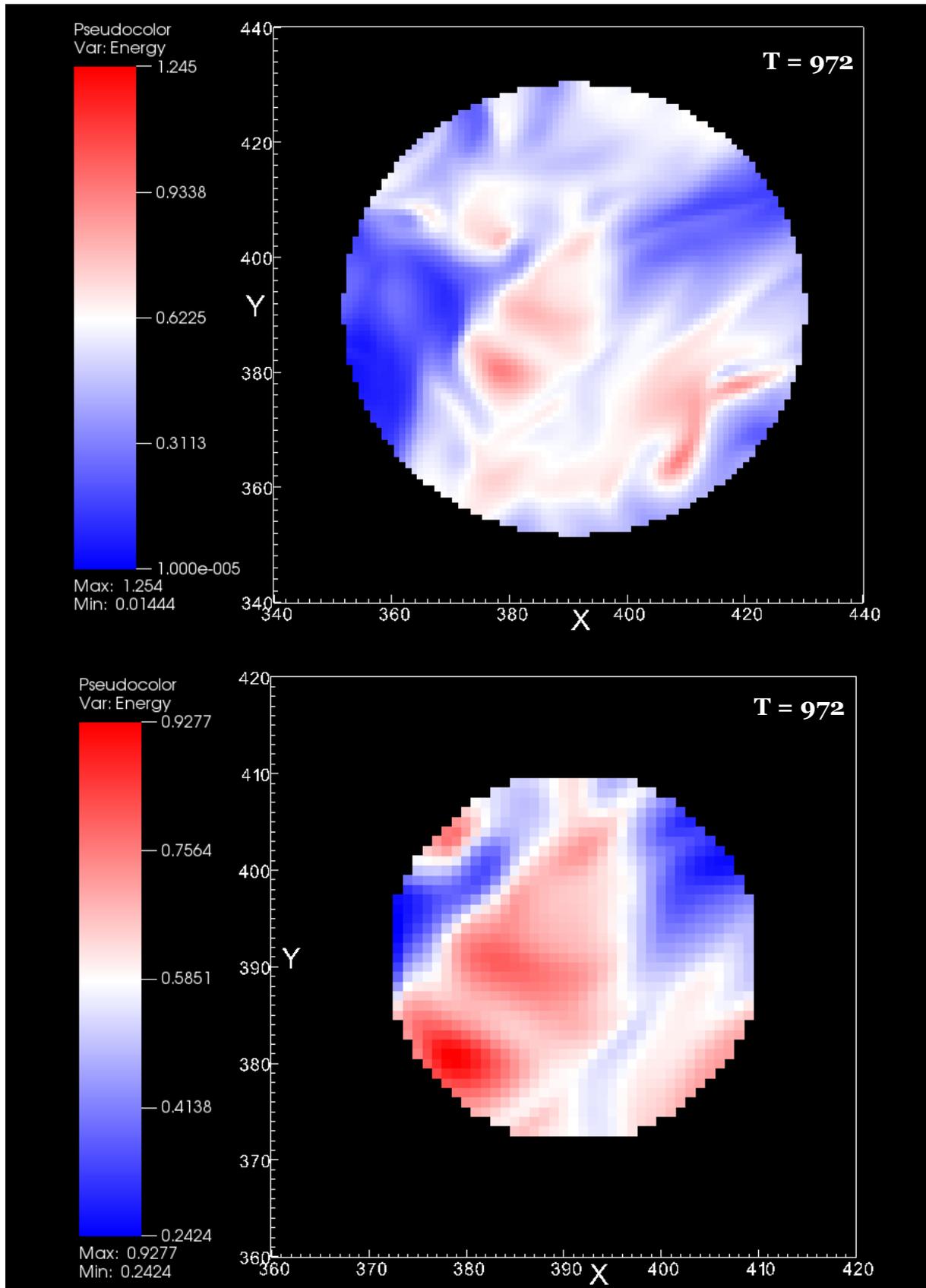
Fig. 3.3-6: Time evolution of the kinetic energy field in the section Z=512, T=972. Upper panel, section of the outer sphere (a unique scale with the other sections); lower panel, a zoom on the inner sphere (with specific color-scale).
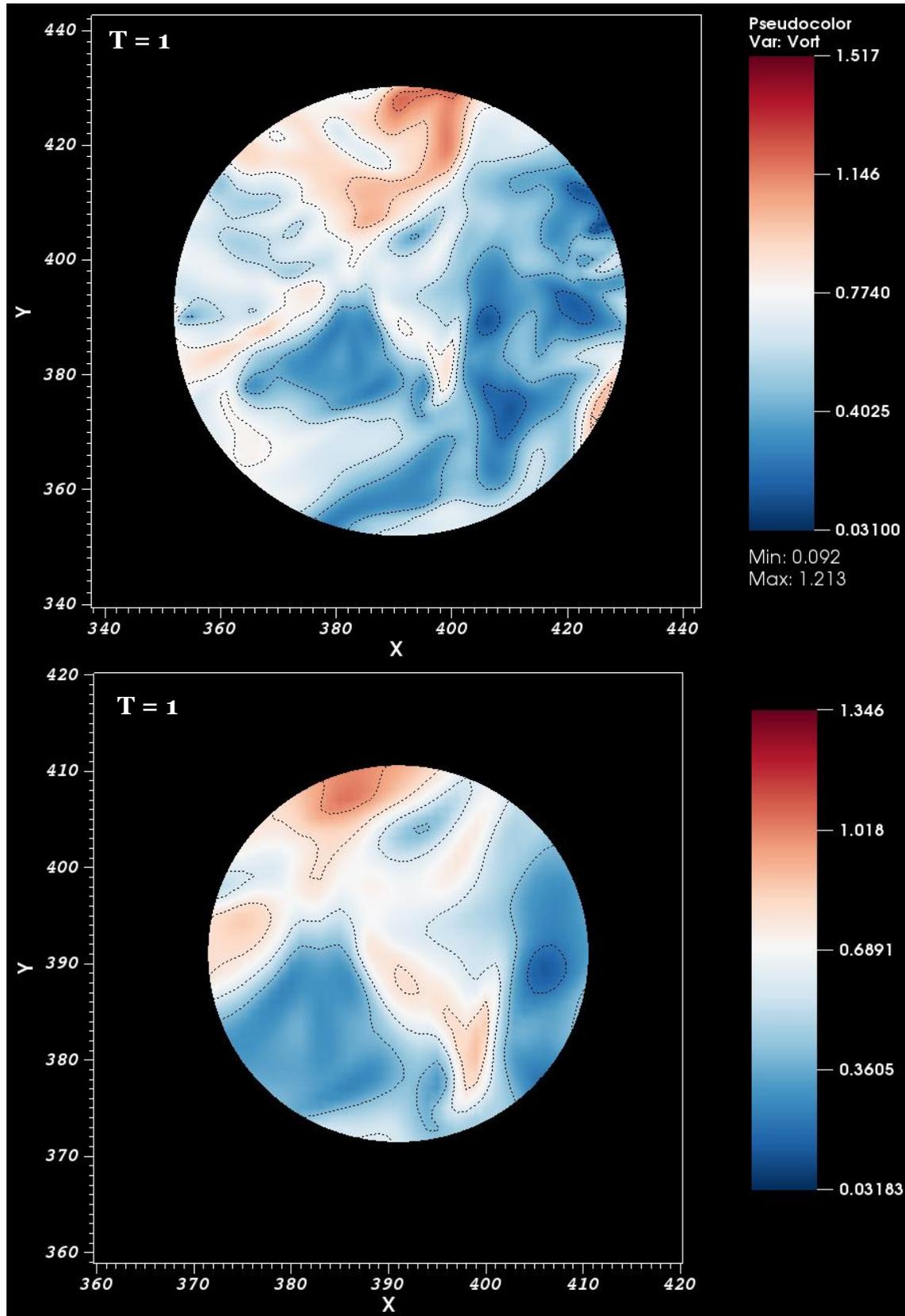
Fig. 3.3-7: Time evolution of the modulus of the vorticity field in the section Z=512, T=1. Upper panel, section of the outer sphere (a unique scale with the other sections); lower panel, a zoom on the inner sphere (with specific color-scale). Contour levels are highlighted with dotted lines.
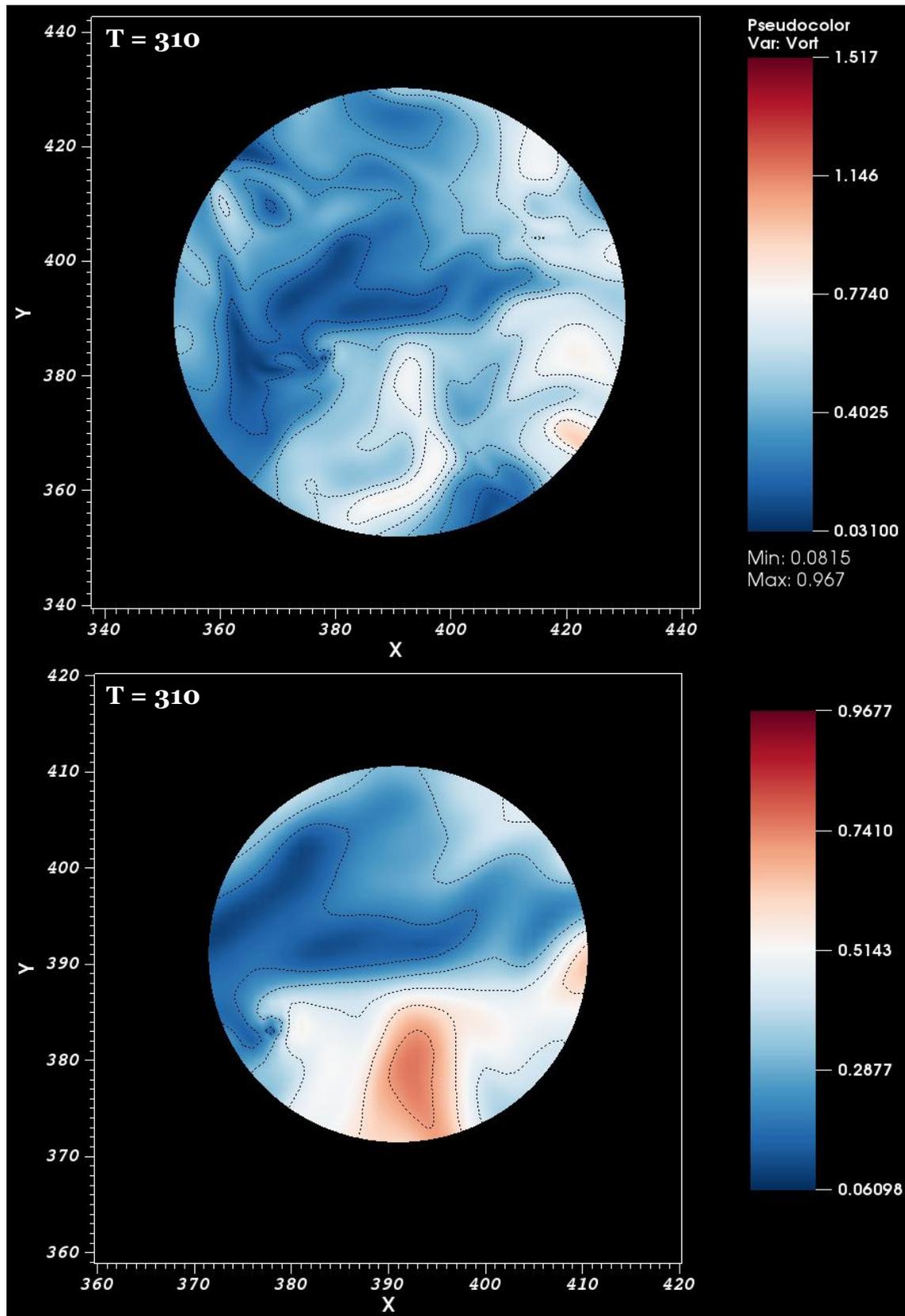
Fig. 3.3-8: Time evolution of the modulus of the vorticity field in the section Z=512, T=310. Upper panel, section of the outer sphere (a unique scale with the other sections); lower panel, a zoom on the inner sphere (with specific color-scale). Contour levels are highlighted with dotted lines.
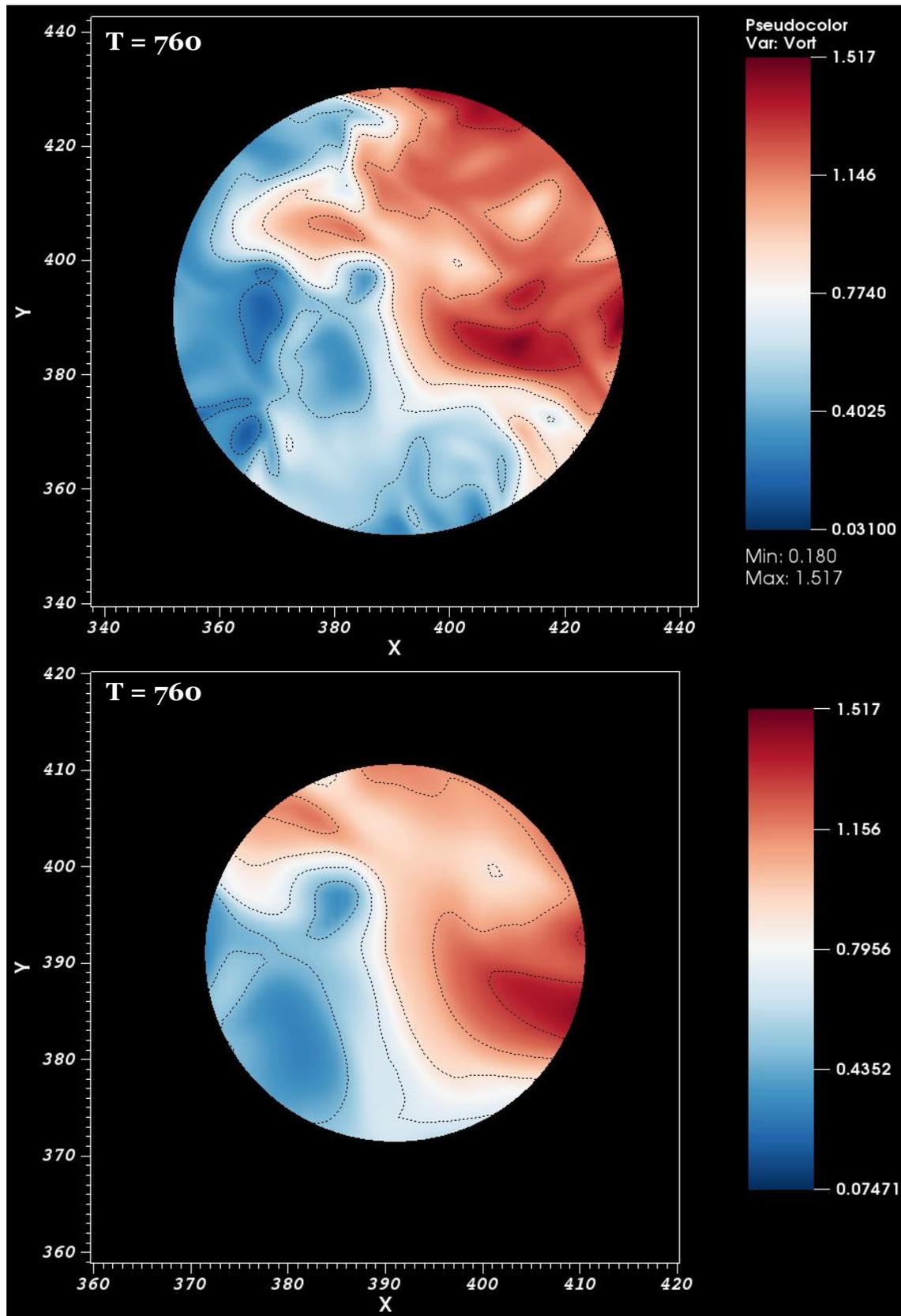
Fig. 3.3-9: Time evolution of the modulus of the vorticity field in the section Z=512, T=760. Upper panel, section of the outer sphere (a unique scale with the other sections); lower panel, a zoom on the inner sphere (with specific color-scale). Contour levels are highlighted with dotted lines.
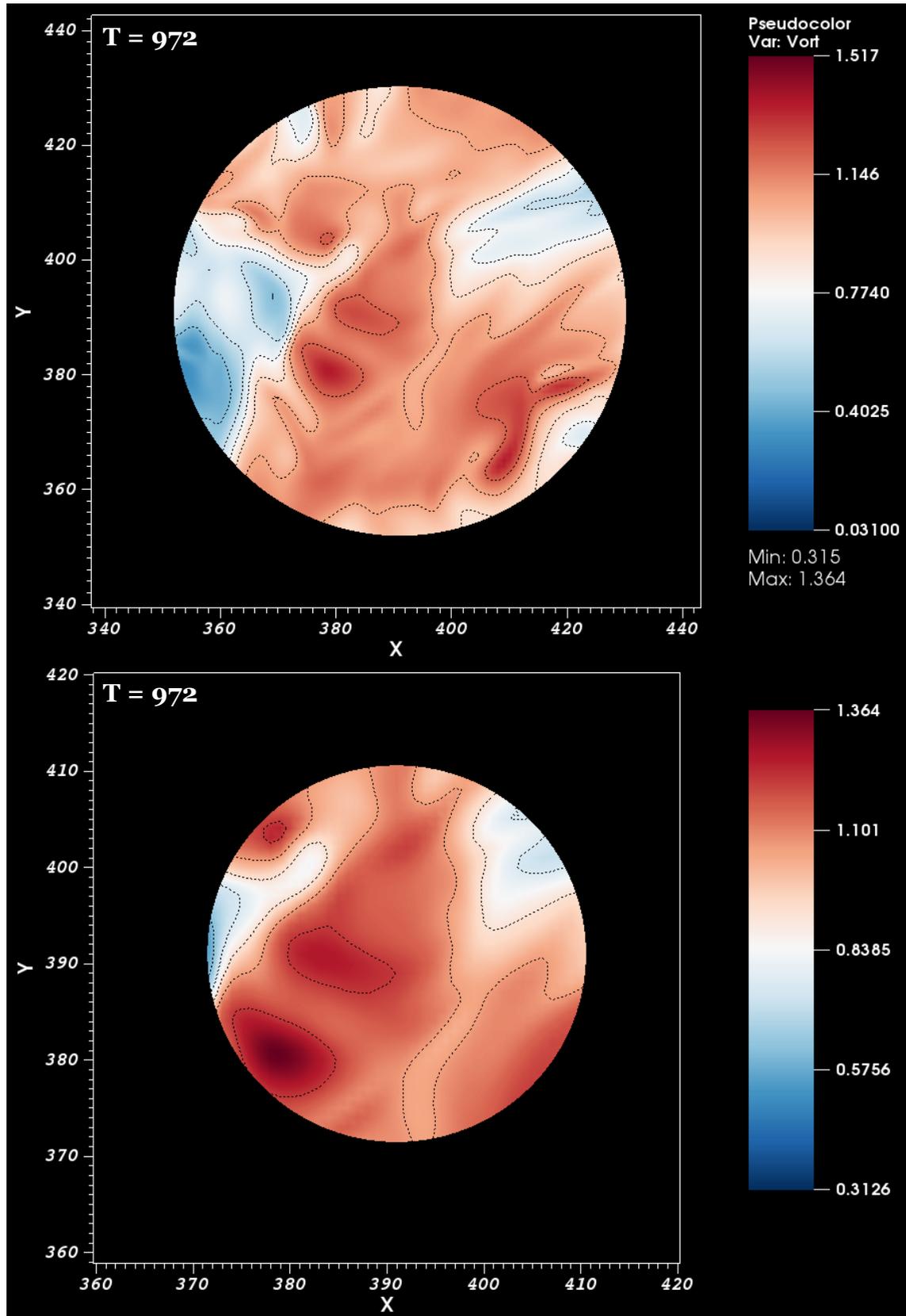
Fig. 3.3-10: Time evolution of the modulus of the vorticity field in the section Z=512, T=972. Upper panel, section of the outer sphere (a unique scale with the other sections); lower panel, a zoom on the inner sphere (with specific color-scale). Contour levels are highlighted with dotted lines.

3.4 The correlation and adjacency matrices

Once the hypotheses are established, the next step is to apply them to the region to study. So, this step will provide the set of procedures implemented which lead to the construction of the adjacency matrix (Fig. 3.4-1), that, in practice, is the representation of the complex network of the turbulent flow considered.
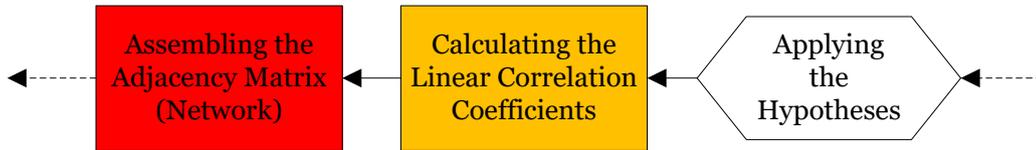
Fig. 3.4-1: Operative part of the workflow diagram.

Among the three main hypotheses stated above, the two geometrical ones are easy to apply once the three coordinates vector — called X, Y, Z in Fig. 3.3-1 — are stored. It is sufficient, indeed, to verify that the mutual physical distance between two points inside the sphere of radius r_c and the respective physical distances of this pair from the center of the spherical region selected are less or equal than $r_i = 0.12 \cong \lambda$. The third assumption, instead, requires a particular attention.

Despite the correlation coefficients R_{ij} between two points i, j can be calculated simply applying the Eq. (1.7) — repeating this operation for each pair of points —, acquiring and saving all the R_{ij} can be useful later, in the construction of the adjacency matrix and in the post-processing step. Nevertheless, since the total number of points inside the sphere of radius $r_c = 0.24$ is (as said before) $N_S = 250023$, the matrix R, calculated with the Eq.(1.8) and (1.9), with entries R_{ij}, will have dimensions $N_S \times N_S$ that is not storable or tractable in practice. Therefore, a few adaptations are necessary:

- Firstly, by definition, only half matrix have to be stored due to the symmetry (i.e. $R_{ij} = R_{ji}$).

- Then, if the threshold τ is know *a priori* and it is *large enough* (that means, indicatively, $\tau > 0.5$), it can be helpful to apply a filter excluding all the $R_{ij} < \tau'$, with τ' slightly smaller than τ (as instance if $\tau = 0.9$, it may be $\tau' = 0.8$).

- Moreover, since the order of magnitude of τ is, often, about $(10^{-1} \div 10^{-3})$, it is possible to round the correlation coefficients off to the nearest $N_{RO} \geq 1$ decimal digits. The rounding operation, however, should take into account the number of significant digits of τ; as a rule of thumb, it is sufficient to set N_{RO} as the number of significant digits of τ plus one. For instance, if $\tau = 0.85 \rightarrow N_{RO} = 3$.

Applying these reductions, the memory-costs significantly decrease. However, the size of the matrix R is still $N_S \times N_S$, preventing a single-file-saving. In order to overcome this obstacle, the correlation coefficients matrix is split in P parts, considerably smaller than R, that can be saved individually. For simplicity, the square

size N_P of all parts R_P are the same. It may happen, though, as in this case, that N_s is not easy to be divided. Indeed, the divisors of $N_s = N_F \times N_P = 250023$ are 3 and 83341, which represent different orders of magnitude; since the number of parts (files) P grows as

$$P = \frac{N_F(N_F + 1)}{2} \, ,$$ (3.4)

selecting an N_F too small (such as 3) the sub-matrices R_P will have dimensions still too big ($N_P \times N_P = 83341 \times 83341$, in this case) but with the advantage of having a small number of files P (6 in this case). On the other hand, selecting an N_F too big (such as 83341) the sub-matrices R_P will be minimal (3×3, in this case) but the number of files will be enormous (with $N_F = 83341$, $P \approx 3.47 \cdot 10^9$). Moreover, applying this method with other values of r_c, it may happen that N_s is a prime number. Hence, to attenuate all the complications listed above, the square size of R (that is N_s) is increased by n-zeros rows/columns, with $n \ll N_s$, such that the modified dimensions are

$$[R]_{mod} = [(N_s + n) \times (N_s + n)] \, .$$

The statement $n \ll N_s$ is crucial to avoid that the sizes of R grows too much. Every single file saved in this way will include the correlation coefficients of the sub-matrix R_P and a (4×1) vector with the indices of the first and last rows and columns of R_P (Fig. 3.4-2).
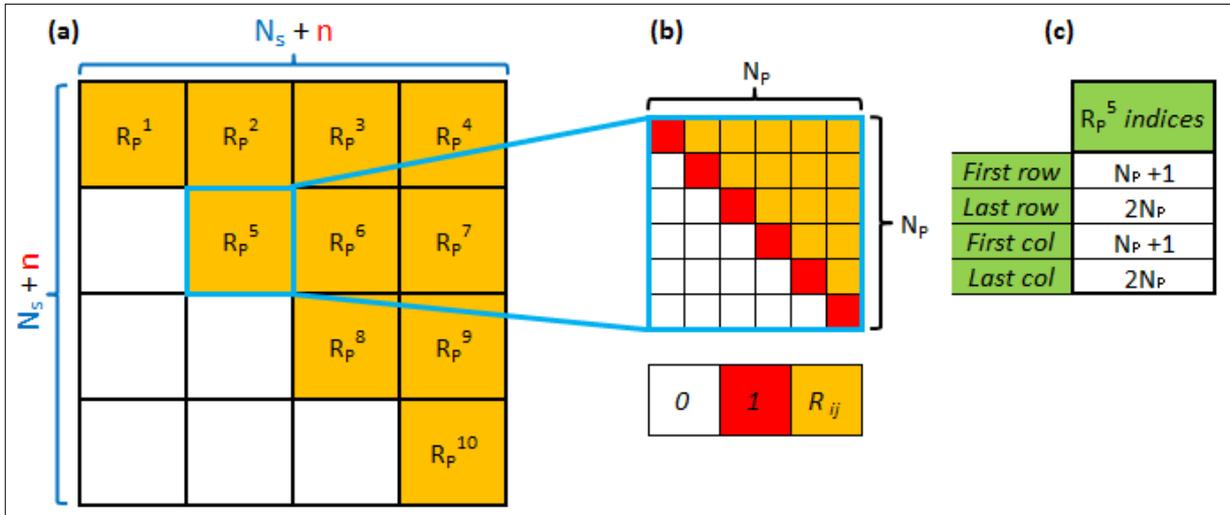
Fig. 3.4-2: A sketch example of a partitioning of the correlation coefficients matrix.

As instance, setting $n = 6$ the ultimate square size of R is $(n + N_S) = 250029$ and it can be easily split as $(n + N_S) = 39 \cdot 6411$ so that, setting $N_P = 6411$ and $N_F = 39$, it will lead to a number of saved files $P = 780$. This example gives an idea of a trade-off between a relatively small value of n (that affects the *Hard-Disk* storable capacity), N_P

(which affects the performances in terms of *RAM*) and *P* (that affects the manageability of the codes).

Once that all the correlation coefficients are stored, the three hypotheses stated in the previous section can be applied to the set of N_S points internal to the sphere of radius $r_c = 0.24$. So, the adjacency matrix, *A*, is obtained with entries $a_{ij} = 1$ if $i \neq j$ and if the pair of nodes fulfill the three hypotheses. According to the definition, hence, in the following each network analyzed will be undirected and without self-connections. The size of *A* is the same of *R* (i.e. $N_S \times N_S$) so it is also partitioned and stored in $P_A \leq P$ parts, because it may happen that in some parts R_P there are no correlation coefficients greater than the threshold τ. Nevertheless, frequently, the adjacency matrix of real networks is *sparse* [13], that is the number of non-zero entries (links) is a considerable small fraction all possible links. This property allows to re-assemble the P_A parts in one single *sparse-matrix*, which have the size $N_S \times N_S$ but only the non-zero entries are stored.

Chapter 4:

Results and Network Analyses

Since the adjacency matrix is assembled, each network built from the turbulent flow can be analyzed in its topological and structural properties, as defined in Chapter 3. In this chapter, then, a detailed review of the results coming from post-processing step will be showed and discussed.

```
┌─────────┐        ┌──────────────────┐        ┌──────────────────────┐
│ Results │ ◄───── │  Post-Processing │ ◄────── │  Assembling the      │
└─────────┘        └──────────────────┘         │  Adjacency Matrix    │
                                                 │  (Network)           │
                                                 └──────────────────────┘
```

Fig. 4-1: Analysis and post-processing stage of the workflow diagram.

After a brief overview of the general properties of the network, the centrality analysis is first carried out, in order to indentify the most important nodes or groups of nodes. Based on centrality, the analysis moved on correlation between nodes and their tendency to link each other (assortativity). Later, results of other measurements are showed, in particular the clustering analysis of the network, provided by clustering coefficient, and the grouping in communities, through modularity concept. In the last part of the chapter two sensitivity analyses are discussed, considering different threshold values τ and different temporal windows. Finally, another network, built on region-2, is then studied ad compared to the network constructed on region-1.

4.1 General properties of the network

First of all, it is reminded that the domain analyzed, called *region-1*, is composed by two concentric spheres, centered in the node $C = (391, 391, 512)$, with radii $r_c = 0.24$ and $r_i = 0.12$. Then, a threshold value $\tau = 0.9$ is chosen, representing very high degree of energy correlation; this value will be kept unchanged in the following, except in the sensitivity analysis. The resulting network is composed by $N = 128785$ nodes, $M = 80920781$ links and $N_i = 31343$ nodes inside the inner sphere. It should point out that this value is greater than the "ideal" value $N_{i,id} = \frac{4}{3}\pi\left(\frac{1023}{2\pi}r_i\right)^3 \cong 31241$, confirming the fullness of the internal sphere (no holes in the inner part of the

network); the two values are not the same because of numerical approximations in the computing of the distances from the center. The number of links between only internal nodes also diminish to $M_i = 38721023$, but the ratios $M/M_i > N/N_i$ saying that, despite the internal nodes are only about $N_i \sim N/4$, excluding the external nodes $N_{ext} = N - N_i$, the remaining links are about 50% of the total, confirming that internal nodes are strongly linked.
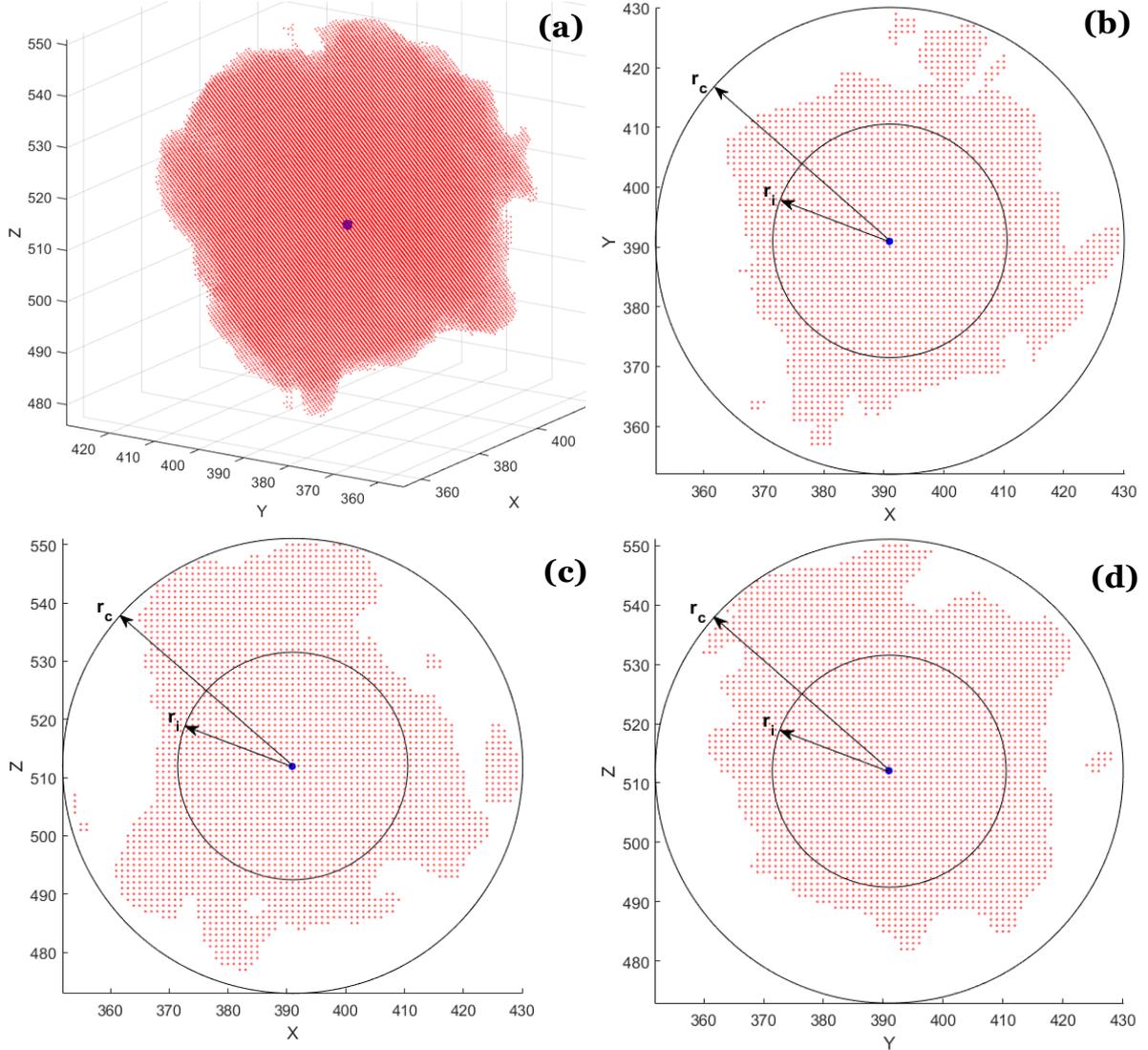
Fig. 4.1-1: Visualizations of nodes of the network; the center *C* is highlighted in blue. (a) a 3D view; (b) 2D section view on the plane Z=512; (c) 2D section view on the plane Y=391; (b) 2D section view on the plane X=391.

The edge density is evaluated, according to the definition in Eq. (2.17), as

$$\rho(\tau) = \frac{M}{\dfrac{N(N-1)}{2} - \dfrac{N_{ext}(N_{ext}-1)}{2}} = 2.282 \cdot 10^{-2} \ ,$$

where the numerator is the number of active links, while the number of total possible links is computed as the total number of links between all N nodes minus the total

number of links between only the external links. This definition, however, does not take into account the physical distance constraints imposed by the hypotheses. A second edge density, $\rho_2(\tau, l)$, is then introduced as a normalized, bi-dimensional, cumulative density function

$$\rho_2(\tau, l) = \frac{m(\tau, l)}{D(\tau = 0, l)} \, , \qquad (4.1)$$

where $l \in [d, r_i]$ is the physical distance between two nodes, $m(\tau, l)$ is the number of active links above the threshold τ and at a fixed l and $D(\tau = 0, l)$ is the total number of links that can be activated at the distance l. A graphical representation of $\rho_2(\tau, l)$ is reported in Fig. 4.1-2: high density values are found for small physical distances, confirming that at $\tau = 0.9$ short-term links are always active ($\rho_2 \to 1 \ if \ l \to d$).
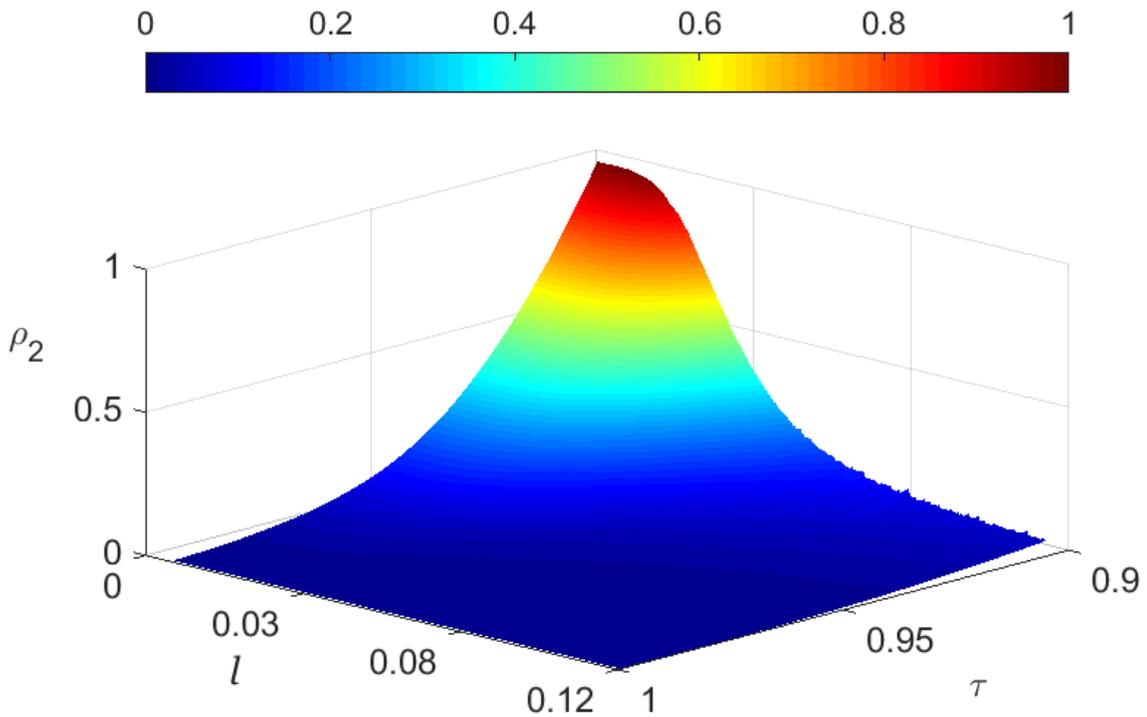
Fig. 4.1-2: Combined bidimensional edge density, $\rho_2(\tau, l)$.

In detail, $\rho_2(\tau, l)$ is defined as *cumulative* because $m(\tau, l)$ is calculated considering the number of links *above* the threshold τ; so, each section at fixed threshold τ in Fig. 4.1-2 represents the distribution of links as a function of the distance l for a hypothetical network built with that specified threshold value τ. The distribution of $m(\tau = 0.9, l)$ is reported in Fig. 4.1-3 as a histogram plot; it is clear the noisy behavior. Rather, a decreasing function of l was expected since long-range connections should be much less than shorter ones. The explanation of such a tendency is explained if one reminds that the domain analyzed is an equally spaced, discrete mesh so each node can activate only a limited number of links at a fixed distance (cf. Fig. 4.1-4 as an example).
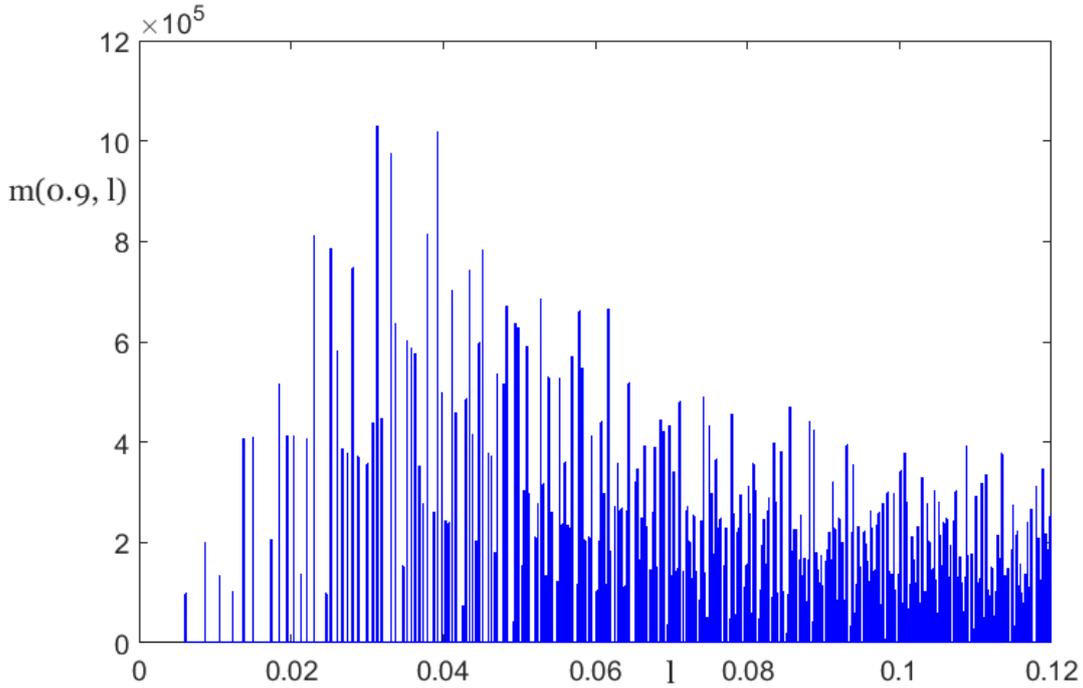
Fig. 4.1-3: Histogram plot of the number of active links at physical distance l, for $\tau = 0.9$.
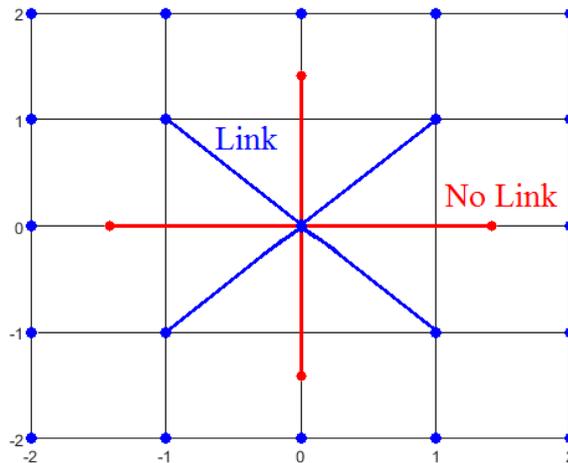
Fig. 4.1-4: Sketch of possible links (in blue) at distance $\sqrt{2}$ that can be activated in a uniformly spaced grid (blue nodes).

Therefore, the number of links that can be activated with nodes at a fixed distance depends on the possibility to find nodes on the grid; naturally, as l increases also the number of potential links grows. In Fig. 4.1-5 the links distribution of Fig. 4.1-3 is displayed after the normalization with the possibility to find a node at a fixed distance. In so doing, the new normalized distribution m' is quite smooth and, above all, decreases for increasing distances. However, the values of m' have no practical meanings because they represent neither the counted values for each distance (such as in Fig. 4.1-3) nor a probability function. The latter can be then evaluated normalizing with the number of links which can be activated at a fixed distance without threshold restrictions, that is the edge density ρ_2 (Fig. 4.1-6). Finally, the probability density function of the correlation coefficients as function of τ is reported in Fig. 4.1-7.
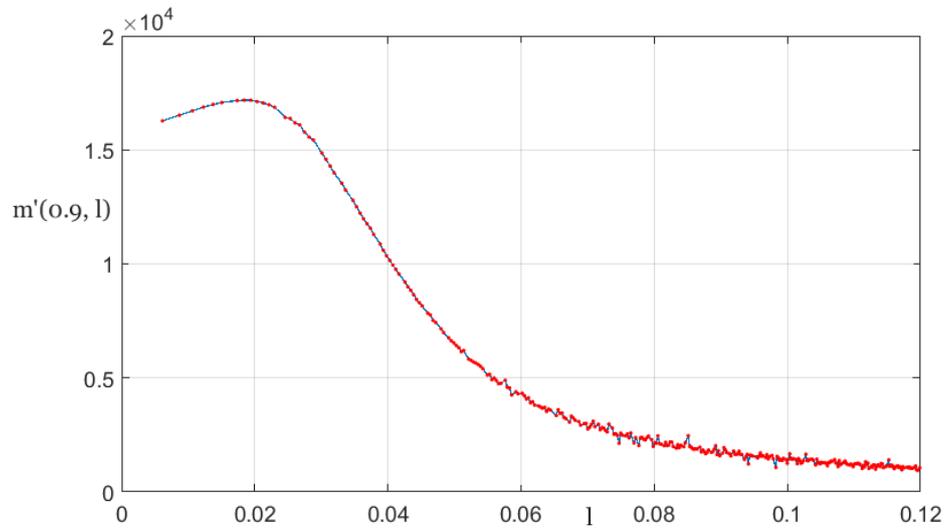
Fig. 4.1-5: Links distribution m(0.9,l) normalized with the possibility to find a node at a physical distance l.
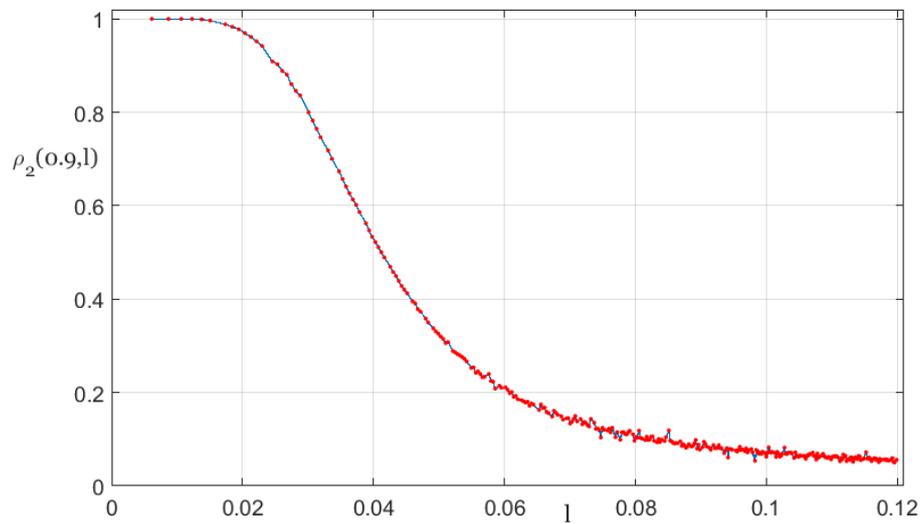
Fig. 4.1-6: The section $\tau = 0.9$ of the combined bidimensional edge density, $\rho_2(\tau, l)$.
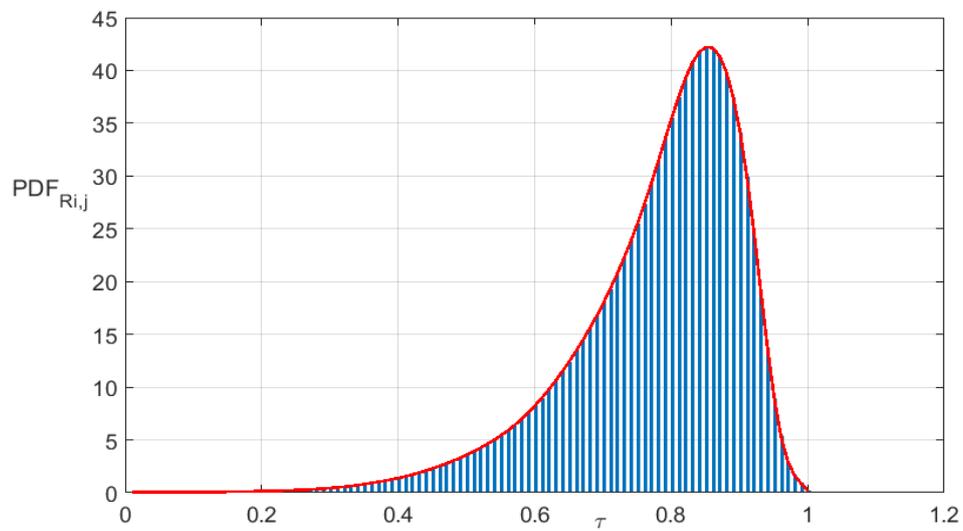
Fig. 4.1-7: The PDF of the correlation coefficients as function of τ; the probability is obtained dividing by 10^3.

4.2 Network centrality examination

The first centrality index analyzed is the degree centrality, k_i — Eq. (2.1), (2.2) and (2.3) — defined as the number of links incident to a node i, normalized with the total number of possible links (i.e. $N-1$) or as the normalized cardinality of the neighborhood of node i. The average degree of the network is $\langle k \rangle = 9.758 \cdot 10^{-3}$.

The Fig. 4.2-1 shows the degree centrality in a 3D perspective view of the entire network.
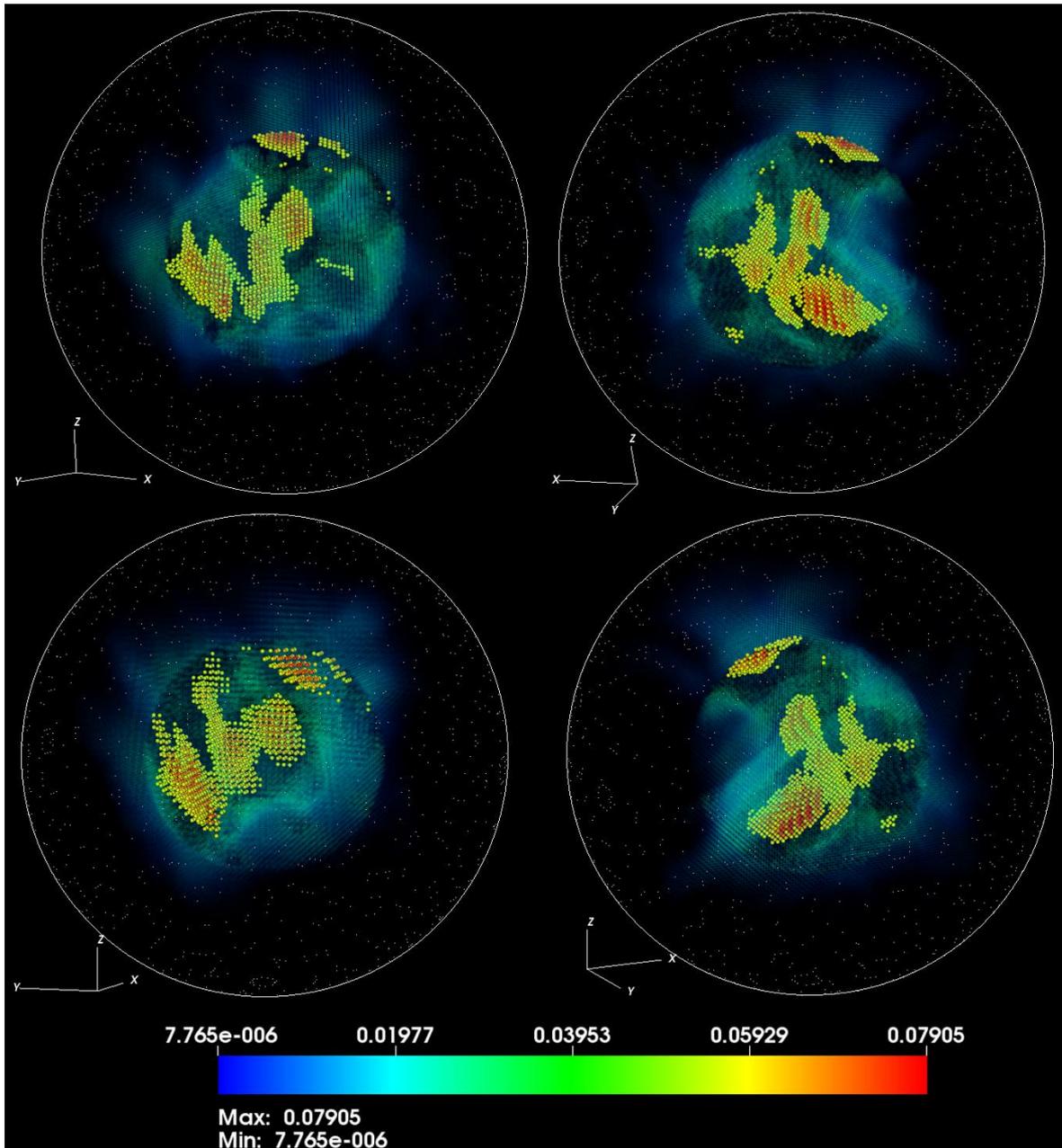
Fig. 4.2-1:Four 3D views of the degree centrality of the network in a unique scale of values. The white circle indicates the edges of external sphere. Nodes above the threshold $k_i \geq 0.7 k_{i,max}$ are highlighted.

In Fig. 4.2-2 is reported the degree centrality map in a 3D perspective view of the inner sphere, highlighting the nodes with high k_i values. From these views, two considerations arise: first, the N_{ext} nodes have low degree centrality values while the more central nodes in terms of degree are only in the inner sphere; then, the high-k_i nodes tends to "group each other", i.e. they are clearly distinguishable in roughly defined clusters inside the sphere of radius r_i. The former consideration is a consequence of the two geometrical hypotheses. Indeed, more importance is given to the vertices inside the inner sphere, while the N_{ext} nodes are considered only to make the influence sphere equal for all the N_i nodes. The latter, instead, is an unexpected result if one considers that it is analyzed an isotropic, homogeneous flow field. Moreover, Fig. 4.2-3 shows the k_i values in 2D section views normal to the X, Y, Z axes.
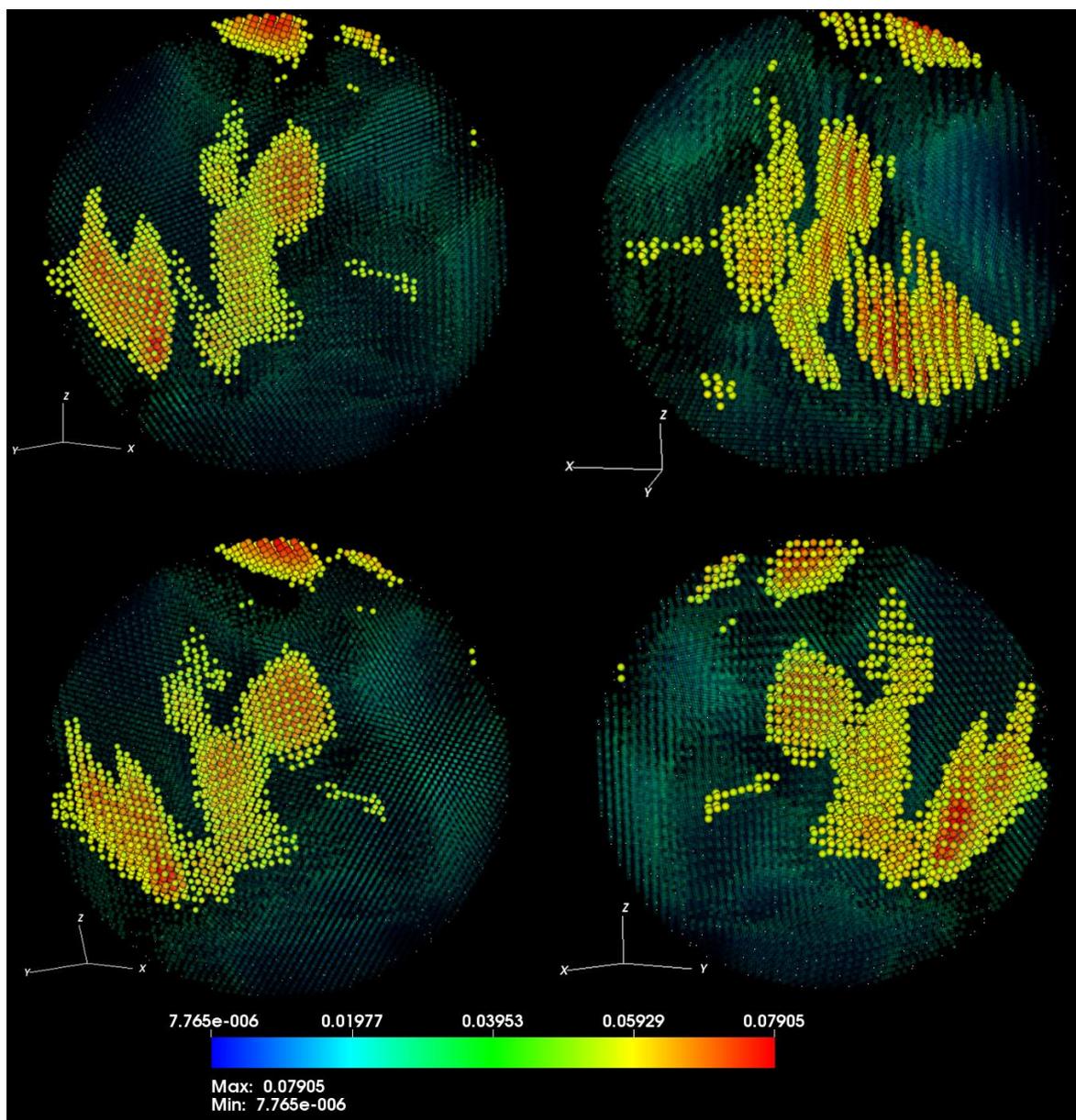
Fig. 4.2-2: A zoom to the inner spheres of Fig. 4.2-1. Nodes above the threshold $k_i \geq 0.7k_{i,max}$ are highlighted.
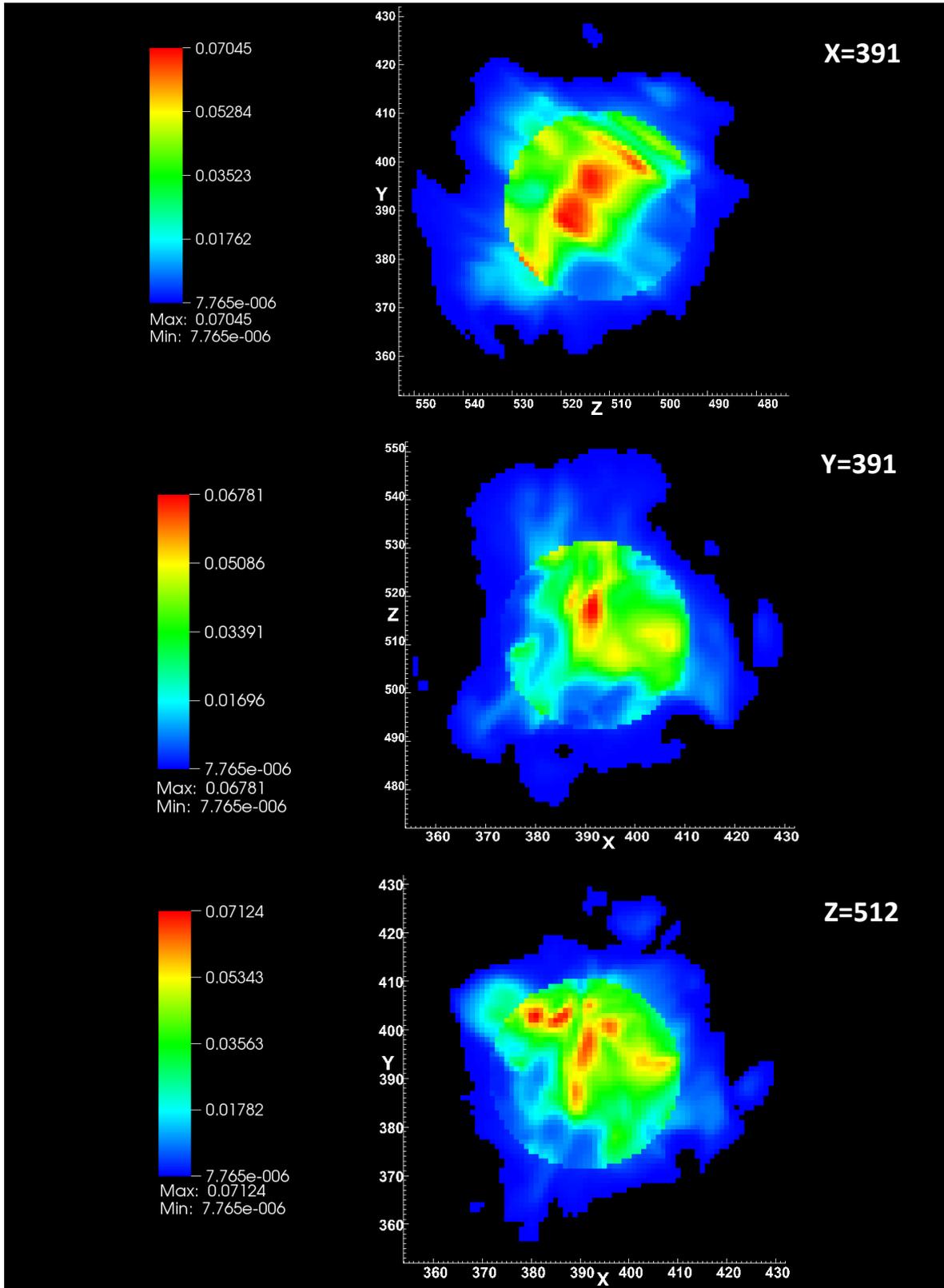
Fig. 4.2-3: Degree centrality k_i; 2D section views on the planes X=391, Y=391 and Z=512.

Then, the degree distribution $P(k)$ of the network, i.e. the fraction of vertices in a network with degree k, is discussed. The Fig. 4.2-4 shows such a distribution in log-log

plot and in which is well clear the power-law behavior and the noisy, fat-tail for high values of k. As suggested by Boccaletti et al. [12], the cumulative degree distribution is evaluated. In Fig. 4.2-5, indeed, the cumulative distribution is smooth and it is easy to find the exponent $\gamma_{cum} \cong 0.77$ of the power-law $P_{cum}(k) \sim k^{-\gamma_{cum}}$ in the range $k \in [0.003, 0.04]$ (which contains about the 60% of all values of k), with an high coefficient of determination $R^2 \cong 0.93$, meaning that the network is scale-free. Hence, the exponent of $P(k) \sim k^{-\gamma}$ can be calculated as $\gamma = 1 + \gamma_{cum} \cong 1.77$, which is very close to the real network scaling range $\gamma_{real} \in [2, 3]$, although typically found over the entire range of k.
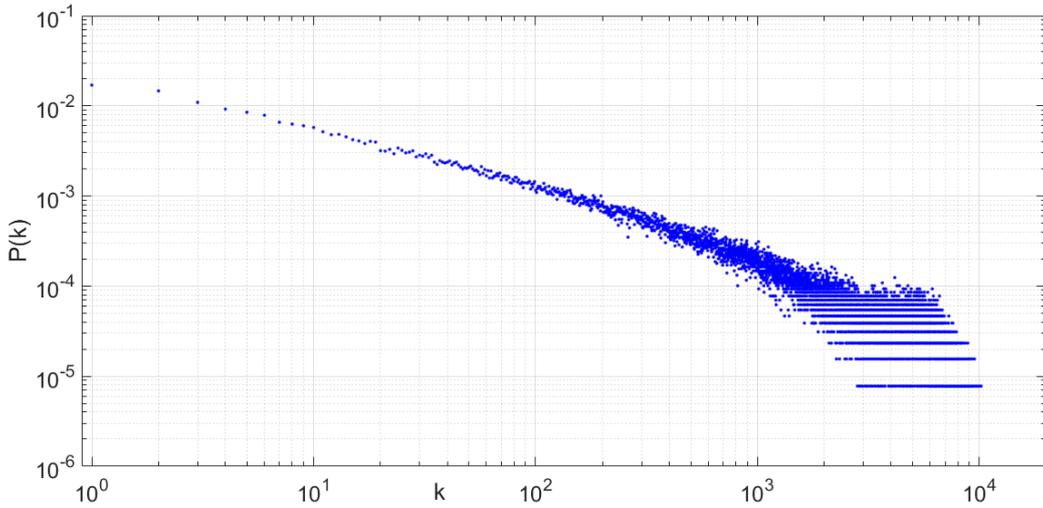
Fig. 4.2-4: Degree distribution P(k).
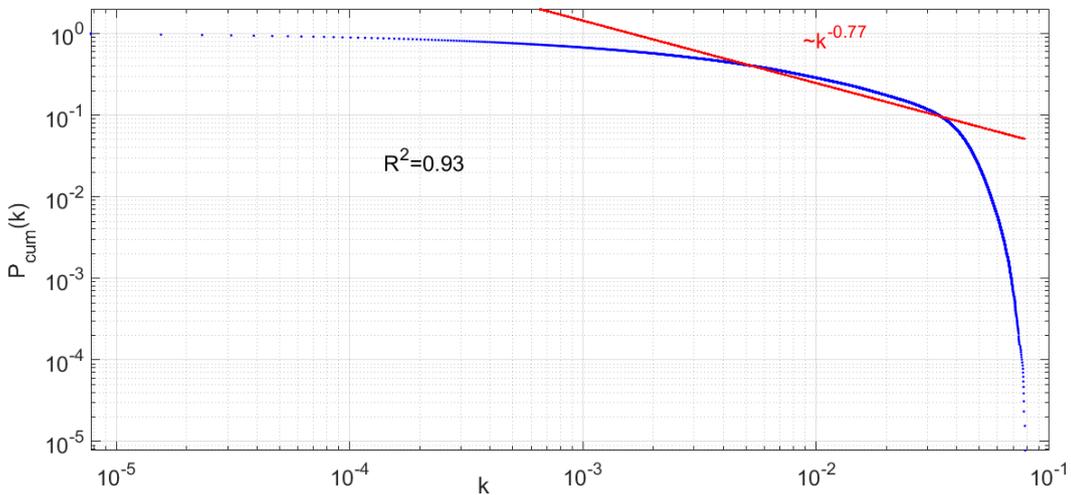
Fig. 4.2-5: Cumulative degree distribution (blue) and the power law behavior (red).

It is worth the effort to highlight that the existence of high-degree pattern well distinguishable was totally unexpected analyzing a statistically homogeneous and isotropic field. That is because, usually, coherent turbulent structures appear if there are spatial inhomogeneities, such as a wall or a body in the fluid flow. Therefore, since the energy forcing acts to larger scales while the analysis is based on the Taylor scale, it was expected a spotted or homogeneous spatial distribution.

The second centrality index analyzed is the eigenvector centrality defined in Eq. (2.7). The results illustrated both in Fig. 4.2-6 and in Fig. 4.2-7 confirm what previously found with the degree centrality analysis. Despite the eigenvector centrality index seems to exhibit a coarser centralization of the nodes, the analogy with the degree centrality is well clear both in a three-dimensional (cf. Fig. 4.2-2) and two-dimensional views. However, while the degree centrality values have a specific topological meaning — being the cardinality of the neighborhood of a node in the network — the values of ranking provided by eigenvector centrality have no particular topological meaning. Indeed, the same partition of the network would be obtained rescaling the eigenvector of the adjacency matrix by the same constant quantity.
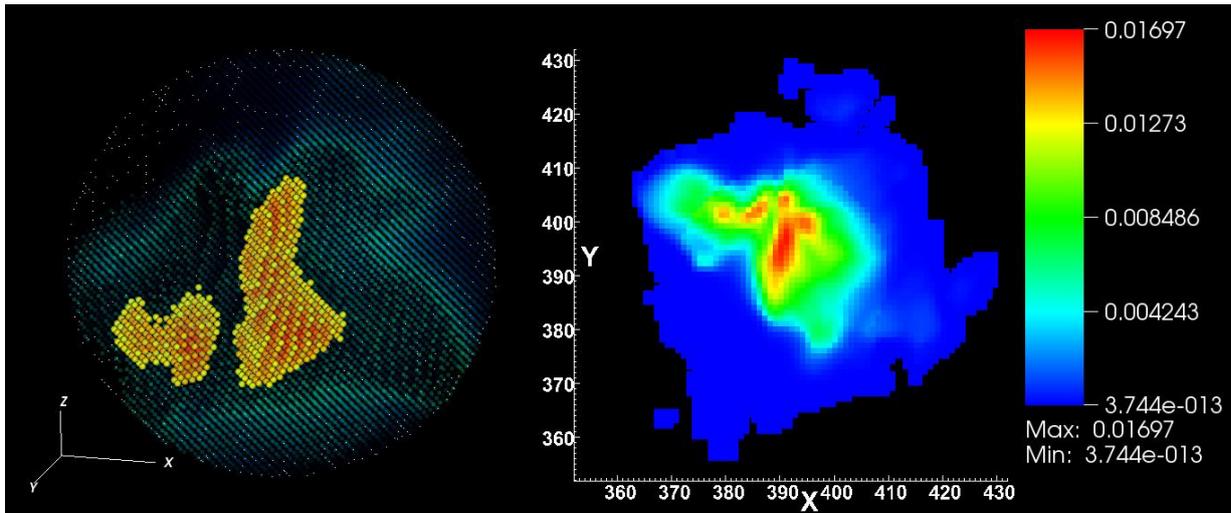
Fig. 4.2-6: Eigenvector centrality distribution. (left) 3D perspective; (right) 2D section on the Z=512 plane.
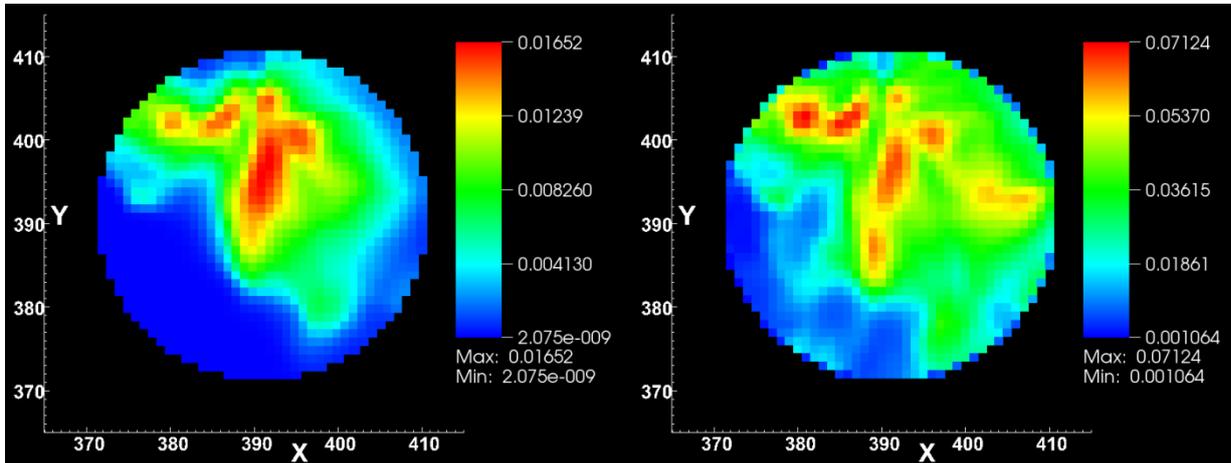
Fig. 4.2-7: Comparison of eigenvector (left) and degree (right) centralities in 2D sections on the Z=512 plane of the inner sphere. Scales of values are referred to the sections.

The third centrality index analyzed is the betweenness centrality, defined in the equation (2.6), limited to the only inner sphere. Unlike the degree and eigenvector centralities, the BC exhibits a spotted spatial distribution with significantly high gradient values; indeed, there are a few high values that are three to four orders of magnitude greater than the lower ones (Fig. 4.2-8).
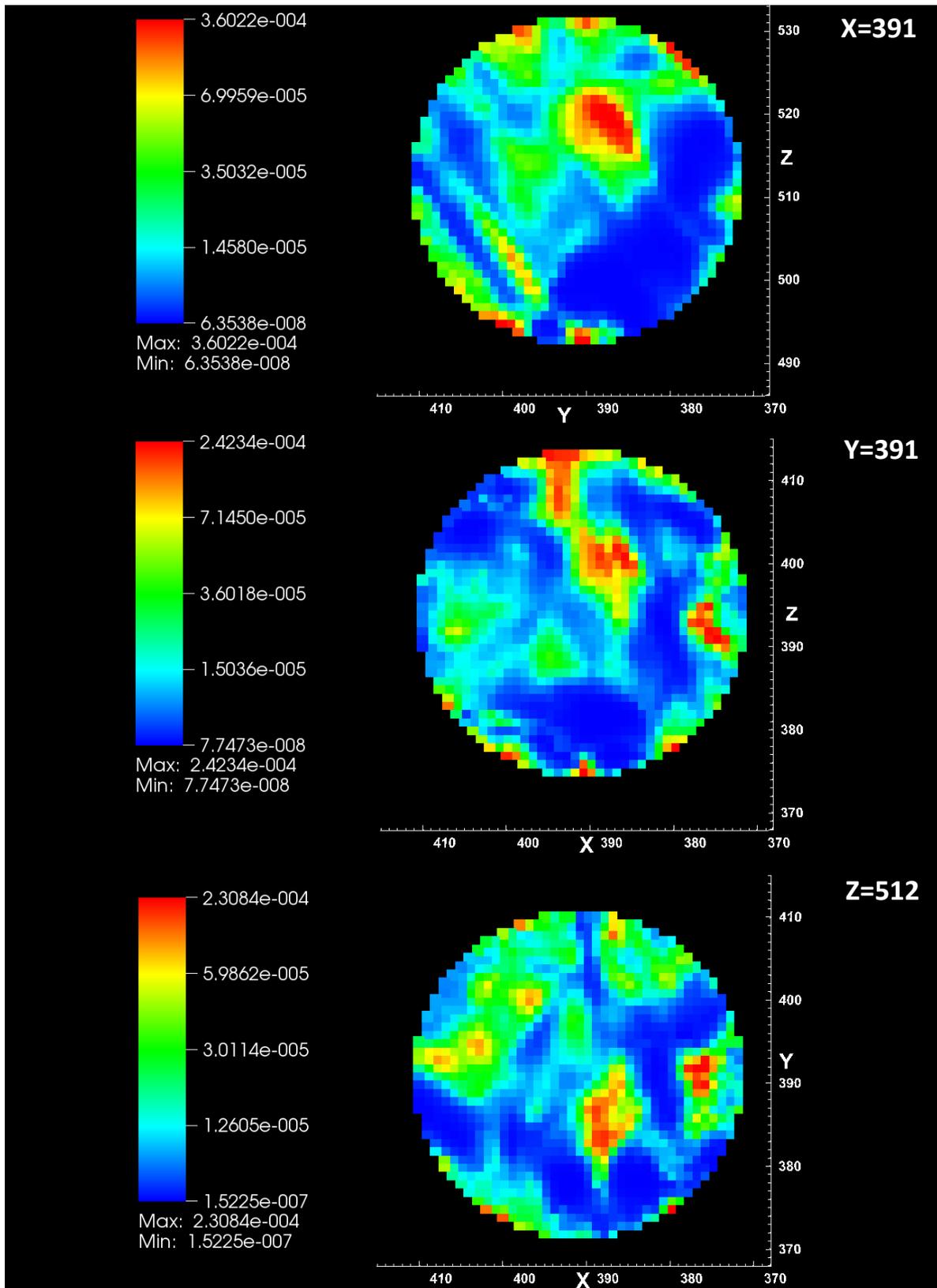
Fig. 4.2-8: Betweenness centrality distribution; 2D section views of the inner sphere on the planes X=391, Y=391 and Z=512. The values are in a non-linear color scale in order to highlight lower values and to mitigate the strong gradient.

Since no sources of inhomogeneity and anisotropy are present in the field, there are no preferential pathways based on shortest-paths; an alternative option could be, then, to evaluate the random betweenness centrality that includes contribution from many paths that are not optimal in any sense. As said in the previous chapter, high computation time is required in the evaluation of the shortest paths, due to the complexity and large size of the network. Therefore, considering the strong spatial constraints and spherical symmetry of network, the closeness centrality of Eq. (2.5) is not evaluated and then it is left to a future analysis.

For the sake of completeness, the degree-based Freeman centralization index defined in Eq. (2.9) is also calculated only for nodes N_i belonging to the inner sphere. Since the centralization measure is an index of how tightly the network is organized around its most central point, the resulting value $C_k = 0.184$ is as expected quite low because the spatial closeness of the nodes in network and the high link density make the network far to be a stair (or a wheel).

4.2.1. Assortativity analysis

Here the degree correlation of the network is briefly showed and discussed. Remembering that a network can be classified as

- assortative, if its vertices show a preference to attach to others vertices that are similar in some way, specifically in degree centrality,

- disassortative, if vertices show a preference to attach to others vertices that are different in degree centrality,

- uncorrelated or non-assortative, if there is no correlation between vertices,

the following five parameters are considered: the joint degree distribution; the k-nearest neighbors; the Pearson degree correlation coefficient; the conditional probability and the Rich-club metric. More precisely, only the first three give information about assortativity of the network, from the most detailed to the most summarizing; the fourth one is considered only if degree correlations between nodes is found. The Rich-club coefficient, instead, can be viewed as a specific index of how nodes link together.

The joint degree distribution, i.e. the probability that an arbitrary link connects two nodes of degree k and k', is shown in Fig. 4.2-9 as a pseudocolor (checkerboard) plot. Evidently, the plot is symmetric because the network is undirected, it is quite spread and display very low probability values. In particular, nodes with lower k_i tends to link with nodes having medium-high degree (upper and left parts of the figure), nodes with medium k_i tends to link roughly uniformly with all other nodes (middle part of the figure), while nodes with higher k_i tend to link mainly with nodes with high degree (lower and right parts of the figure). To summarize, the plot of Fig. 4.2-9 does not show a strong tendency of any degree values towards any other degree values with high probability; the network, therefore, seems to be almost uncorrelated. The k-nearest neighbors parameter, instead, gives a more condensed view on how nodes of specific degree k tend to be linked on average. Fig. 4.2-10 confirms the

prevision suggested by the joint degree distribution: indeed, the plot does not exhibits a monotonic behavior for $k \to k_{max}$, whereas the tendency is "fluctuating" and limited to a k_{nn} range about $k_{nn} \in [1000, 6000]$.

The Pearson degree correlation parameter summarizes what previously said in a scalar value. The calculated coefficient is, indeed, $r \cong 1.417 \cdot 10^{-3}$ which means that the assortativity is almost absent (very low value of r) but it tends to be positive.
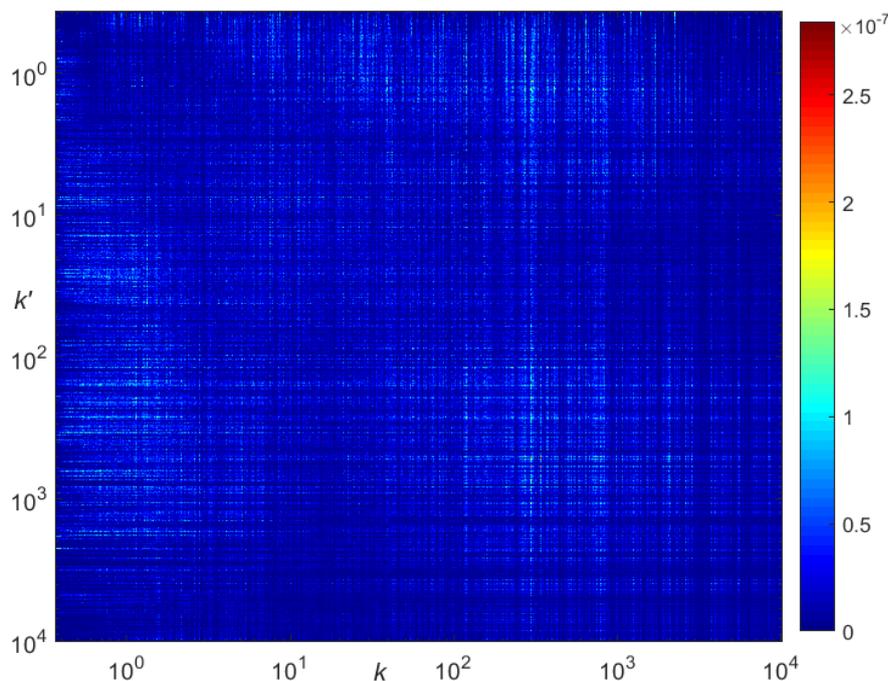
Fig. 4.2-9: The joint degree distribution as a pseudocolor (checkerboard) plot.
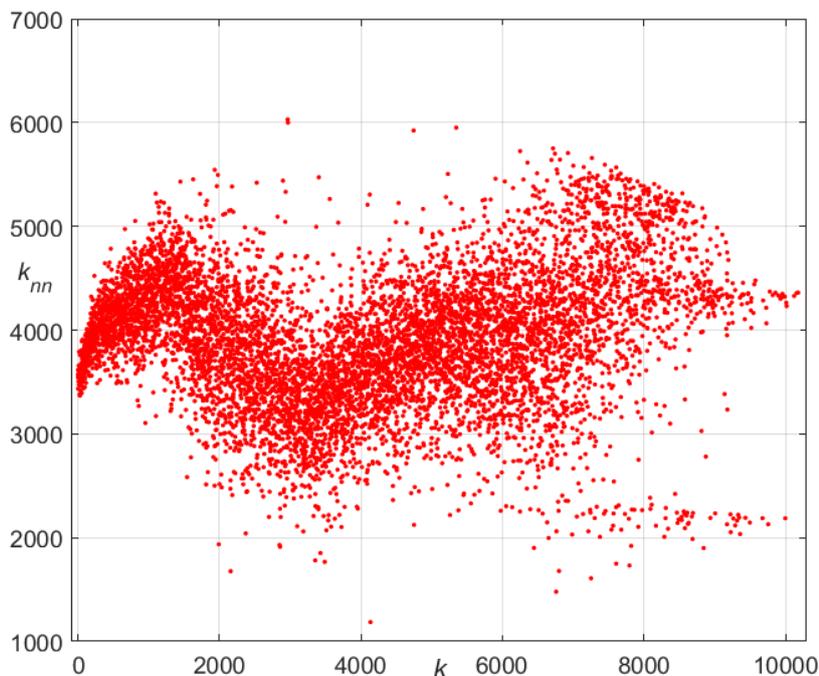
Fig. 4.2-10: The k-nearest neighbors distribution k$_{nn}$.

In conclusion, the network analyzed is non-assortative, though with a slight inclination of nodes of similar degree to be linked. The degree distribution $P(k)$, hence, fully determines the statistical properties of the network. However, for the sake of completeness, in Fig. 4.2-11 the conditional probability $P(k'|k)$, i.e. the probability that a link is incident to a pair of nodes of degrees k and k', is shown.
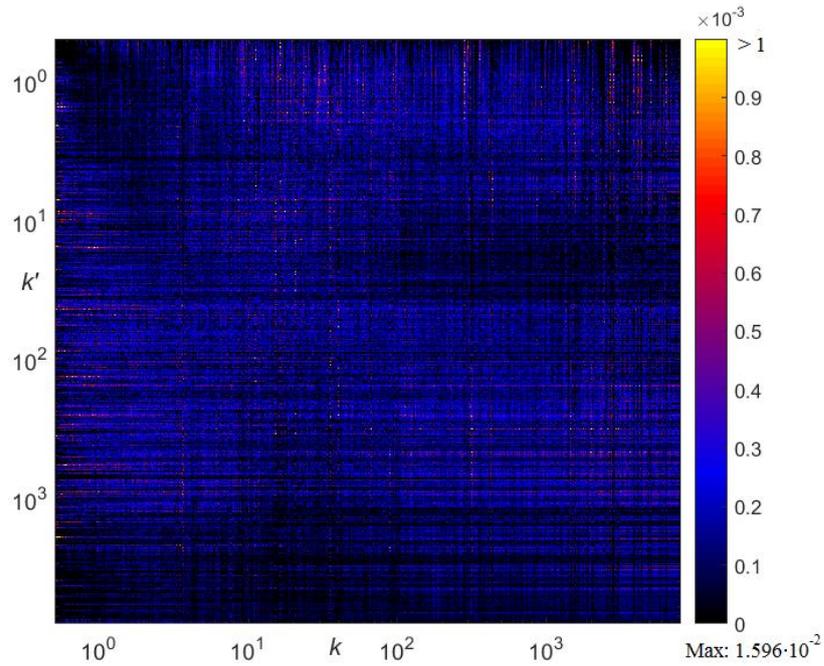
Fig. 4.2-11: The conditional probability P(k'|k) as a pseudocolor (checkerboard) plot.

Last parameter evaluated is the Rich-club coefficient $\phi(k)$, which is an indicator of the tendency of high degree vertices to link to each other. The plot in Fig. 4.2-12 reveals, despite some noise for high k, an increasing trend for $k \to k_{max}$. The network then can be said to show a Rich-club phenomenon, despite it is non-assortative.
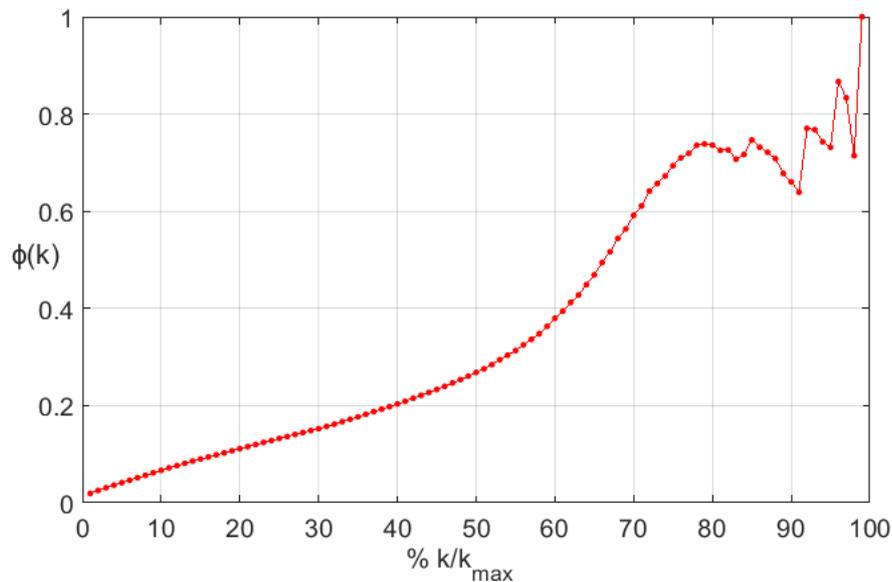
Fig. 4.2-12: The Rich-club coefficient $\phi(k)$; the degree k is normalized with the maximum value and reported in %.

4.3 Other structural properties

The clustering coefficients defined in equations (2.13) and (2.14) are here discussed. Fig. 4.3-1 shows the local clustering in the section Z = 512 and it is interesting to note that the highest and lowest values are located in the outer sphere. This behavior can be explained as follows: since the cardinality of the neighborhoods of external nodes is low (i.e. low degree centrality values), it is easy to find that nodes in small neighborhood are linked at all or not at all, without middle ways. So, the analysis is limited to the solely inner sphere.
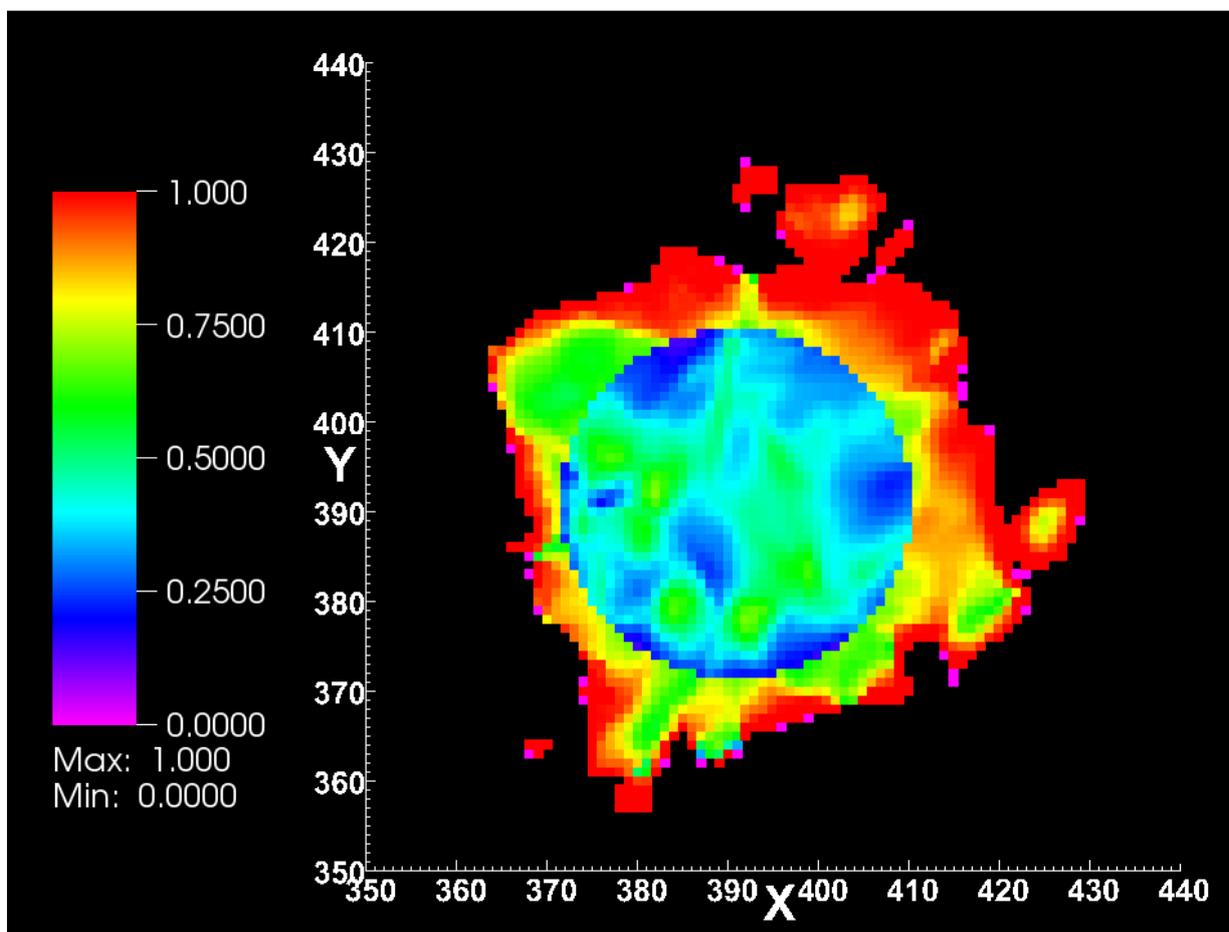
Fig. 4.3-1: Clustering coefficient c_i on the section Z=512.

As can be seen in the Fig. 4.3-2 in which the section Z = 512 is considered, the local clustering (bottom panel) is poorly related to the degree centrality (top panel); the same consideration is true also for the other sections of the domain (cf. Fig. 4.2-3 and Fig. 4.3-3). As previously said, many real networks have a hierarchical dependence between $C(k)$ and k, while in many others where there are strong spatial constrains, such as the network here analyzed, the clustering coefficient and the degree centrality are independent.
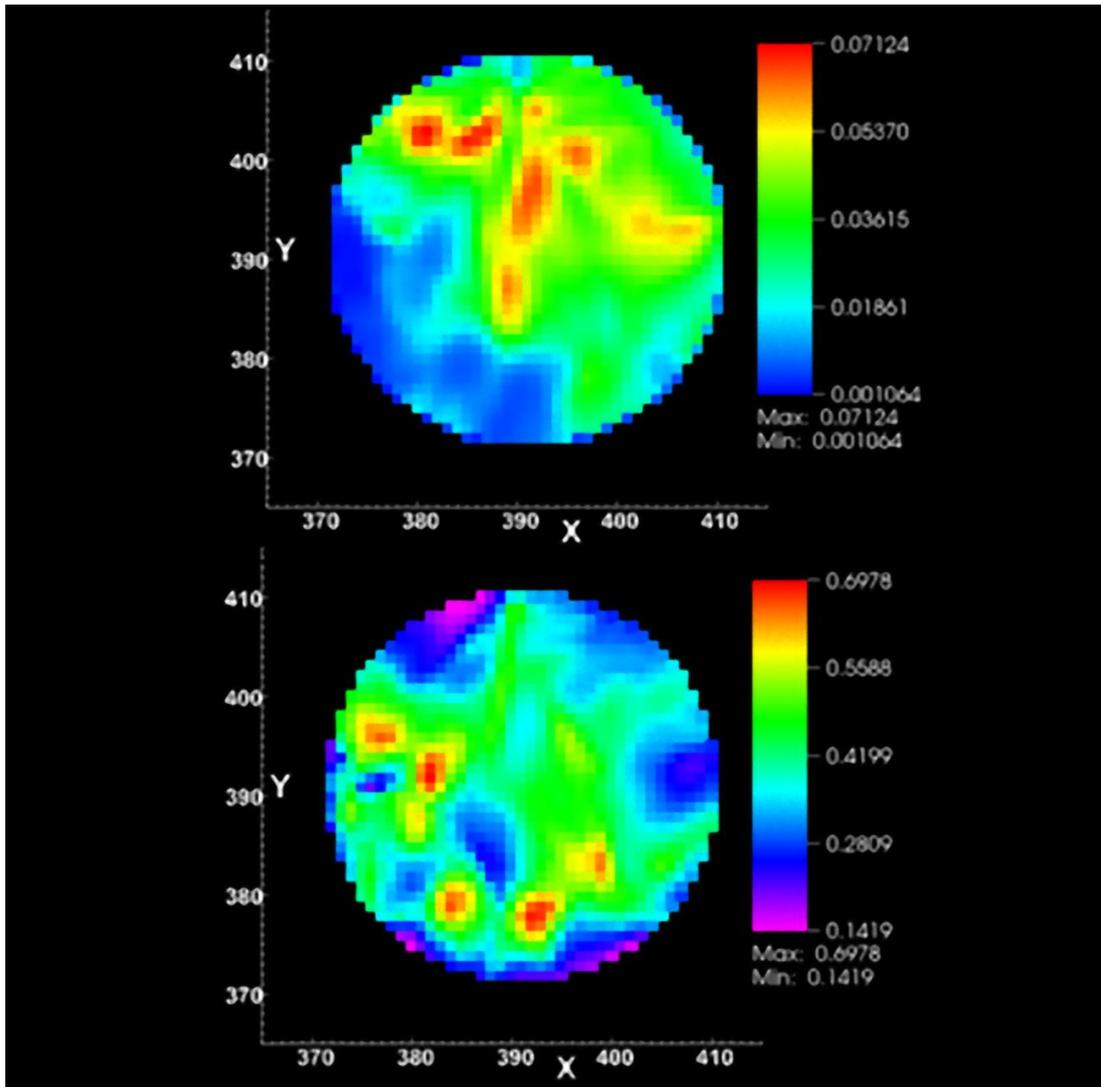
Fig. 4.3-2: Comparison of degree centrality (upper panel) and clustering coefficient (lower panel) in 2D sections on the Z=512 plane of the inner sphere. Scales of values are referred to the sections.

The global clustering coefficient of the entire network $C = 7.211 \cdot 10^{-1}$ is a quite high value, a typical behavior of real networks (in particular those that exhibit a small-world property). Evaluating the global clustering coefficient for the inner sphere network only, the value diminishes to $C_{int} = 5.828 \cdot 10^{-1}$ because many clusters of order three are removed since they are located externally from the inner sphere.
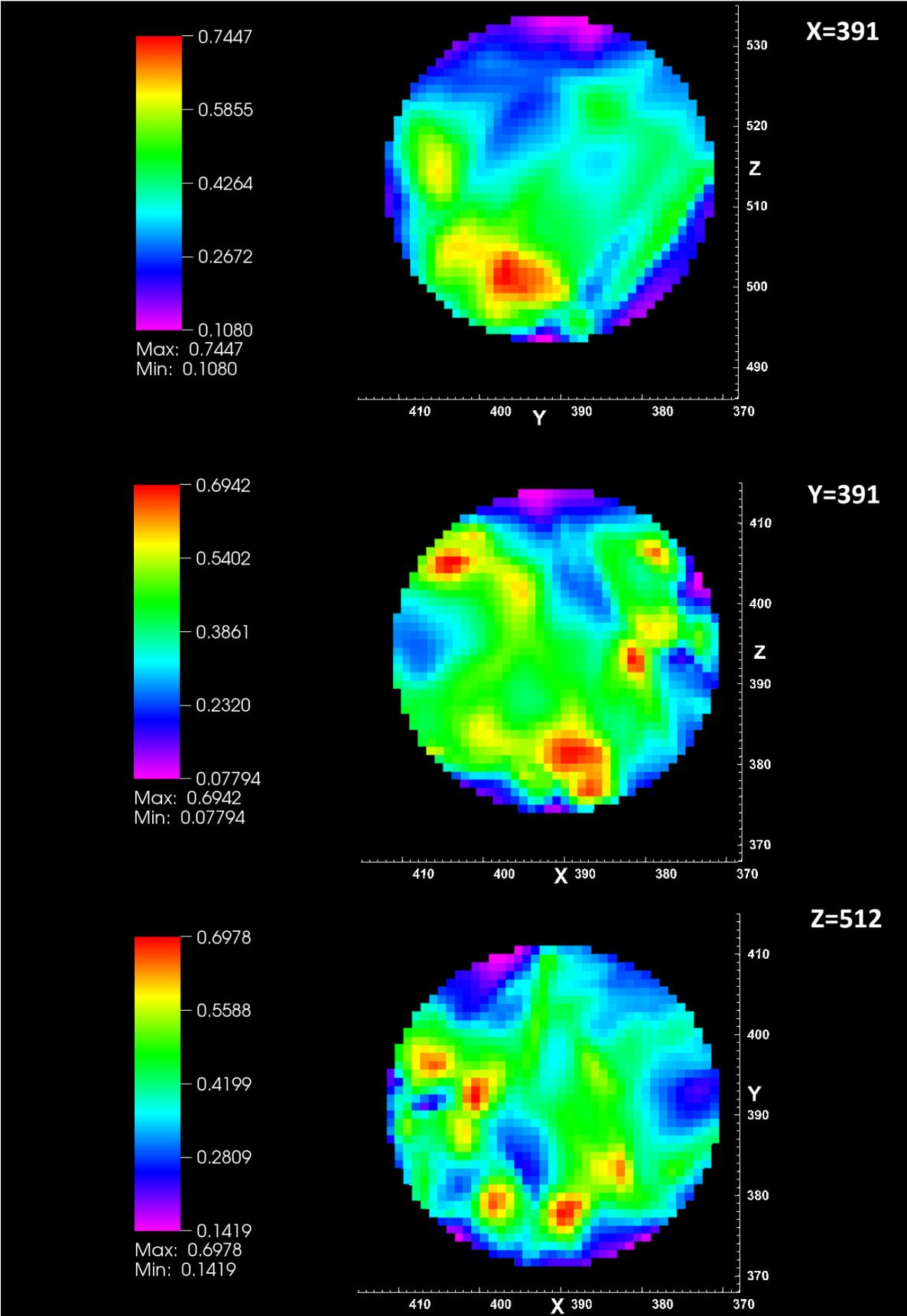
Fig. 4.3-3: Clustering coefficient c_i; 2D section views of the inner sphere on the planes X=391, Y=391 and Z=512.

To sum up, the local clustering, the degree, the eigenvector and the betweenness centralities of the network in the section Z = 512 are displayed in Fig. 4.3-4.
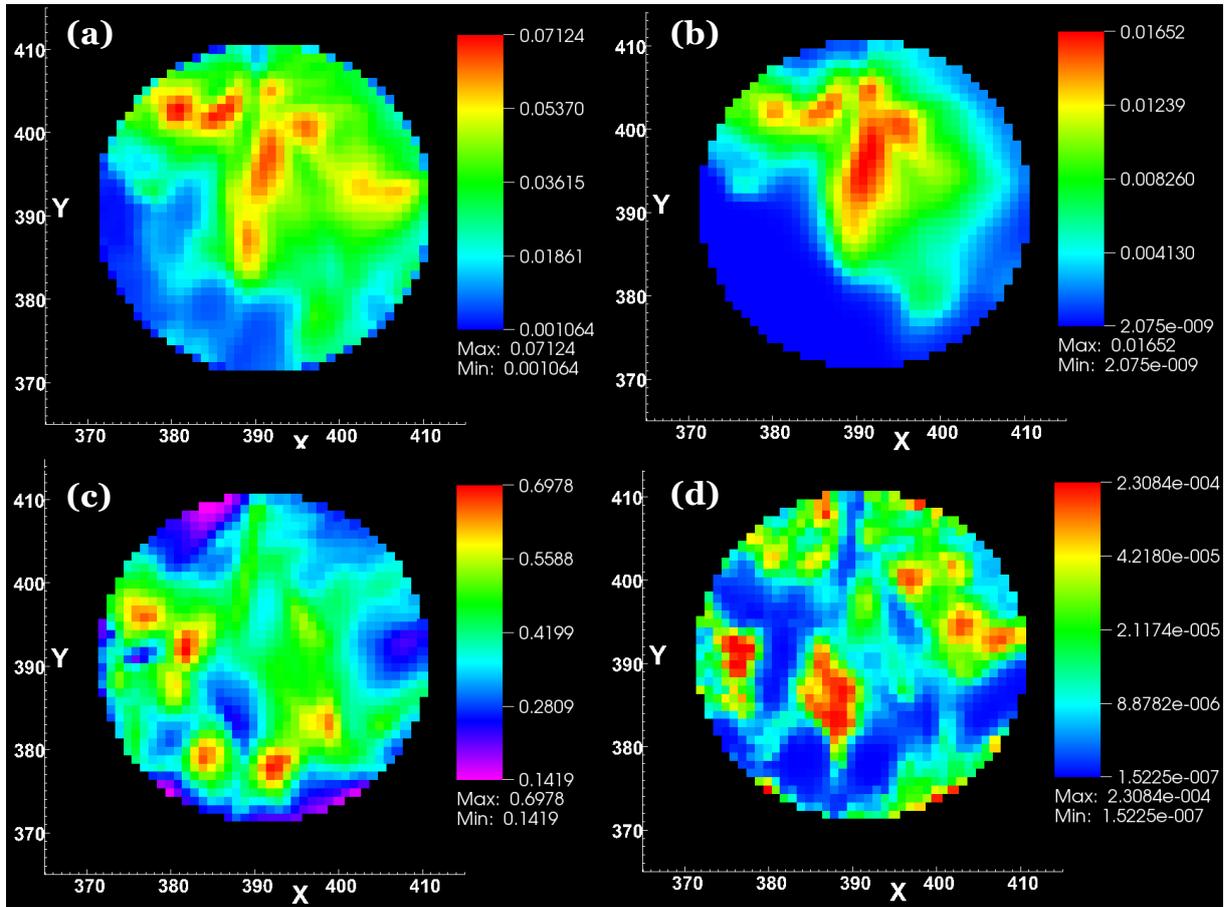
Fig. 4.3-4: (a) Degree centrality, (b) eigenvector centrality, (c) local clustering coefficient, (d) betweenness centrality (a non-linear color scale is used). Results are displayed on the section Z = 512 of the inner sphere.

The average physical distance defined in Eq. (2.16) is then analyzed. This index is introduced in this work mainly for two reasons: firstly — just as for betweenness and closeness centralities — in order to avoid the high computational cost in terms of time, due to the expensive calculation of all the shortest-paths of such a large network; secondly, an average topological distance loses significance in a network with high-spatial restrictions and spherical symmetries. The average physical distance L_i, instead, is a more appropriate index because it considers physical distances — which are more suitable for a spatial network — and, taking the averages only among nodes of the same neighborhood, provides a direct information about the order of magnitude of the size of the effective influence region of each node.

Fig. 4.3-5 shows L_i in the section Z = 512 and it is pretty clear that — just as in the local clustering — the minimum and maximum values of L_i are located externally from the inner sphere. Thus, since both the cardinality of the neighborhoods of the external nodes is low and their location is far from the center of the network, the average topological distance varies from high values (where dominate the farness) to low values (where dominate the small size of the neighborhood).
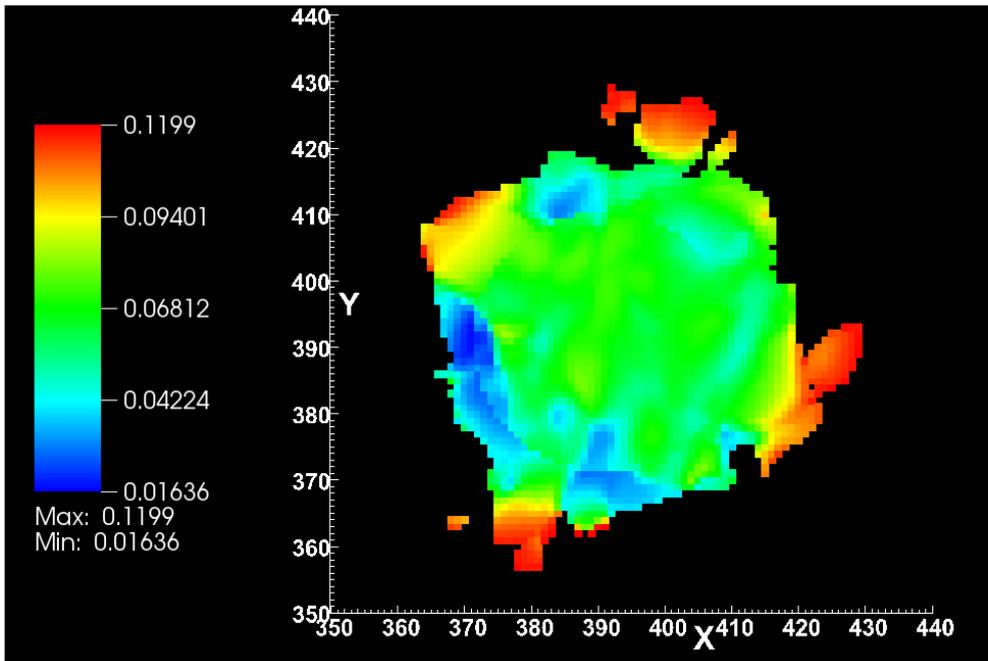
Fig. 4.3-5: Average physical distance L_i on the section Z=512.

Moreover, L_i is compared with the degree centrality in Fig. 4.3-6. Without overstating the analogy, it can be affirmed that the average physical distance is closely related with the degree centrality, so L_i gives the order of magnitude of the spatial patterns identified by k_i distribution. The mean value for the entire network is $\langle L \rangle = 7.703 \cdot 10^{-2}$ which corresponds to a grid distance of about 12.5 cells.
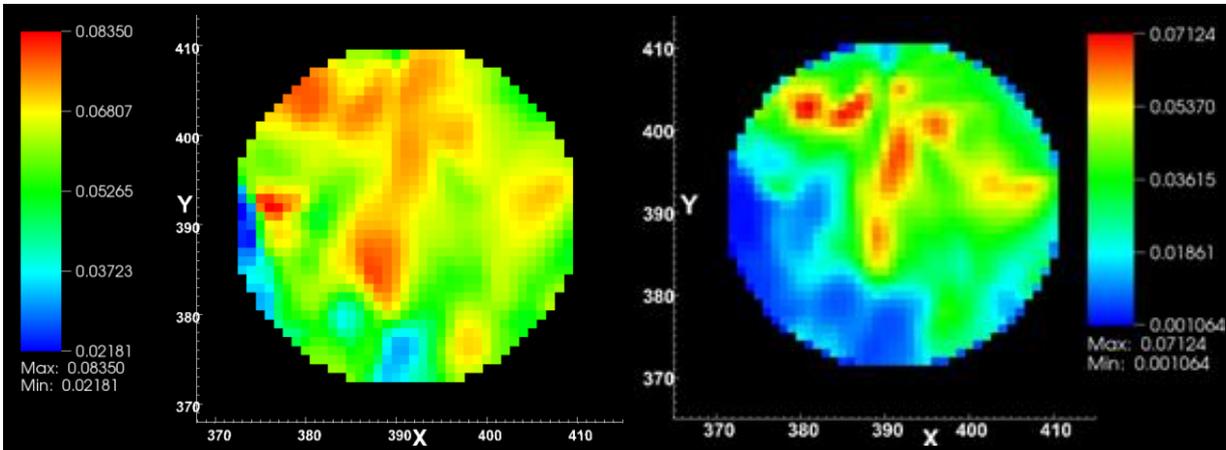
Fig. 4.3-6: Comparison of average physical distance (left) and degree centrality (right) in 2D sections on the Z=512 plane of the inner sphere. Scales of values are referred to the sections.

To complete the set of structural metrics, the tendency to form organized groups in the network, called communities, is discussed and carried out through the analysis of modularity Q defined in Eq. (2.18). As previously stated, the modularity is a parameter used to quantify the division of a network and a way to find the best partitioning in communities of the network is to maximize Q. It can also be remembered that $Q = 0$ should be found if links are distributed randomly in the graph and values of Q above 0.3 suggest strong division of the network [40].

Among all the algorithms and techniques used for this purpose, in this work, only two have been used, both following the idea to maximize Q:

- the Newman's spectral algorithm, based on the concept of modularity matrix, that provides a division of the network in "indivisible sub-networks" [42];

- the fast and heuristic Louvain algorithm.

Since the former is well documented as one of the best algorithm to find communities, it is here used for the partitioning of the network. The latter, instead, it is used only as a comparison, mainly to verify the order of magnitude of Q rather than to find effectively the network division. Table 4.1 summarizes the main results for the two algorithms selected. Both the algorithms provide a modularity value above 0.3, suggesting then a strong division of the network.

Algorithm	Network	Modularity, Q	# Communities, c
Newman	Internal [b]	0.31	28
Louvain[a]	Full-Internal	$0.47 \div 0.6$	$6 - 8$

Table 4.1: Community features. [a]The values are calculated considering the inner sphere only, with a resolution parameter between 0.8 and 1.1.

Considering only the Newman's algorithm, a deeper analysis can be carried out such as the modularity and the mean degree of each community or how well each community identifies the high-degree clusters just seen (cf. Fig. 4.2-1 and Fig. 4.2-2). The modularity and mean degree distributions, $q(c)$, $\langle k \rangle(c)$, over found communities c is reported in Fig. 4.3-7; the following relation holds $Q = \sum_c q(c)$.
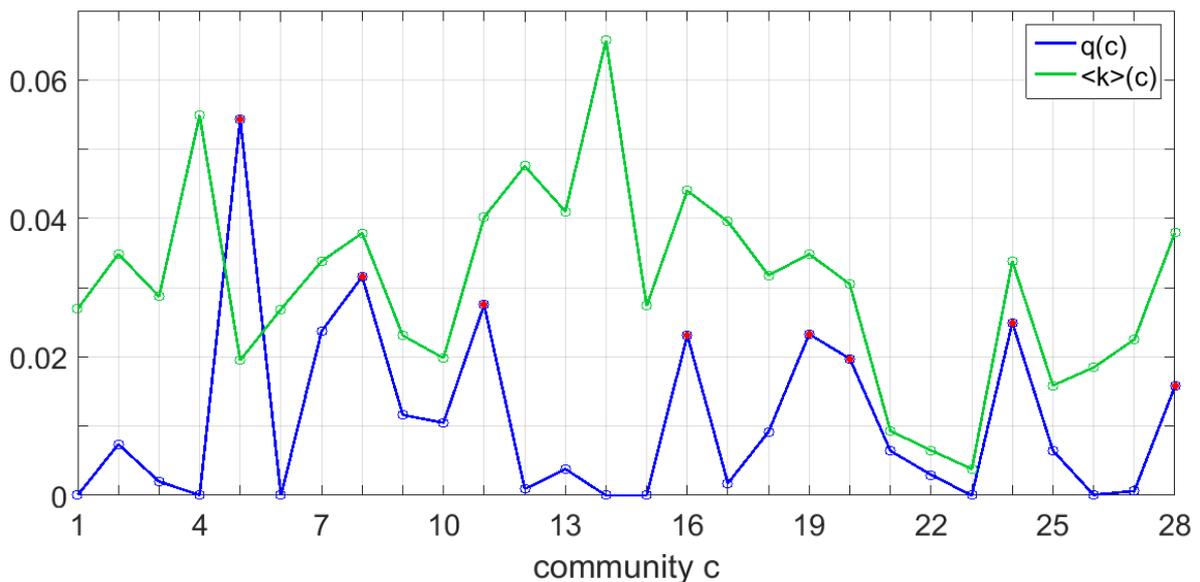
Fig. 4.3-7: In blue, modularity distribution over the twenty-eight communities, q(c); in green, the mean degree $\langle k \rangle$ of each community. Red spots highlight the eight more numerous communities.
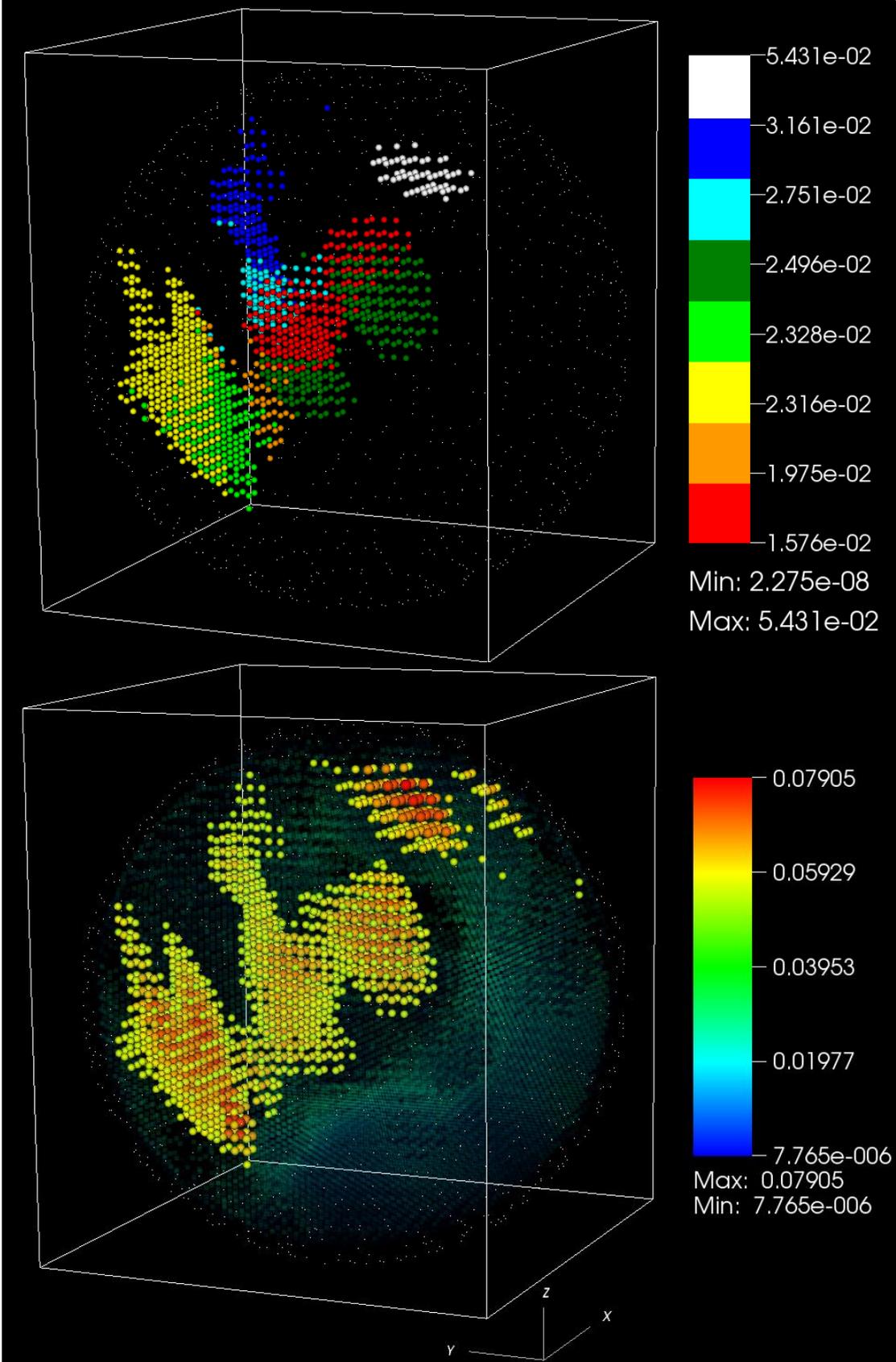
Fig. 4.3-8: A comparison of the eight more numerous communities (top), with their modularity value q(c), and the degree centrality in the same 3D perspective (bottom). In both cases nodes with $k_i \geq 0.7 k_{i,max}$ are filtered.

Modularity is not uniformly distributed over the communities, as eight modules have q values close to zero, while one community has the highest q value (0.055), which is about 18% of the total modularity value, Q. The analysis is then limited to high-degree nodes, in particular nodes with degree $k_i \geq 0.7k_{max}$. Fig. 4.3-8 compares clusters of high-k nodes (bottom panel) with nodes of high-k in the eight communities with highest $q(c)$ values and most numerous (top panel;cf. also red spots in Fig. 4.3-7).

The eight communities are chosen according to the relation $q(c) \geq 0.3\ q_{max}$ and avoiding those with a number of nodes with high-degree ($k_i \geq 0.7k_{max}$) less than 35 (which is about 10% of the higher cardinality). From the comparison of Fig. 4.3-8 it is clear that high modularity communities include a large number of high-k nodes; indeed, the fraction of nodes with degree above 70% of the maximum value inside the eight high-$q(c)$ communities is about 86.1%. Newman algorithm, hence, turn out to be a very useful tool to find communities in real, spatial networks.

4.4 Physical interpretation of results

In the light of results just discussed in the previous sections, the most meaningful parameter turns out to be the degree centrality, k, together with the eigenvector centrality, both direct measures of the importance of nodes in the network. In order to interpret the network results in terms of physical properties of the turbulent field, the highest degree centrality node and another one with very low degree centrality are considered, respectively called HDC and LDC nodes. Moreover, other two nodes, called A and B, at an intermediate physical distance $l_{ph} = 6.142 \cdot 10^{-2}$ (10 grid cells) from nodes HDC and LDC respectively are considered. Some features of HDC and LDC nodes are reported in Table 4.2, while for the nodes A and B is sufficient to specify their respective coordinates, $A = (395, 401, 508)$ and $B = (372, 377, 510)$.

	Node HDC	**Node LDC**
Coordinates	$(385, 401, 508)$	$(372, 387, 510)$
Degree centrality, k_i	$7.905 \cdot 10^{-2}$	$3.052 \cdot 10^{-3}$
Neighborhood cardinality, $\#v(i)$	10180	393
Average physical distance, L_i	$7.515 \cdot 10^{-2}$	$2.985 \cdot 10^{-2}$

Table 4.2: **Topological and spatial features of two nodes with high (HDC) and low (LDC) degree centrality.**

The neighborhoods of the HDC and LDC nodes are displayed in Fig. 4.4-1. The temporal series of the vorticity modulus $|\bar{\omega}|$ for the two pairs of nodes, $(HDC - A)$ and $(LDC - B)$ are then evaluated and showed in Fig. 4.4-2 and Fig. 4.4-3.
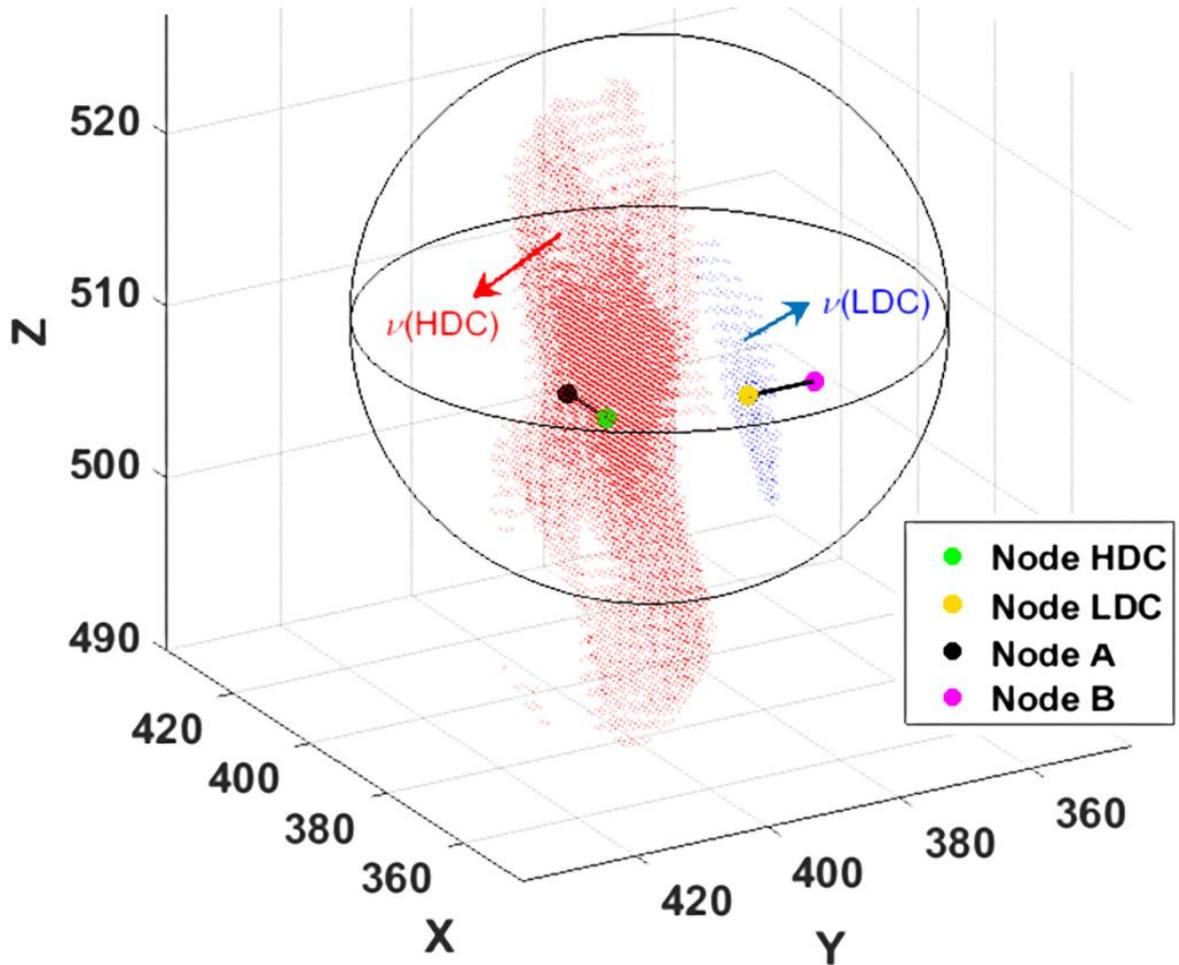
Fig. 4.4-1: Positions of HDC, A, LDC and B nodes and neighborhoods of HDC and LDC nodes are shown.
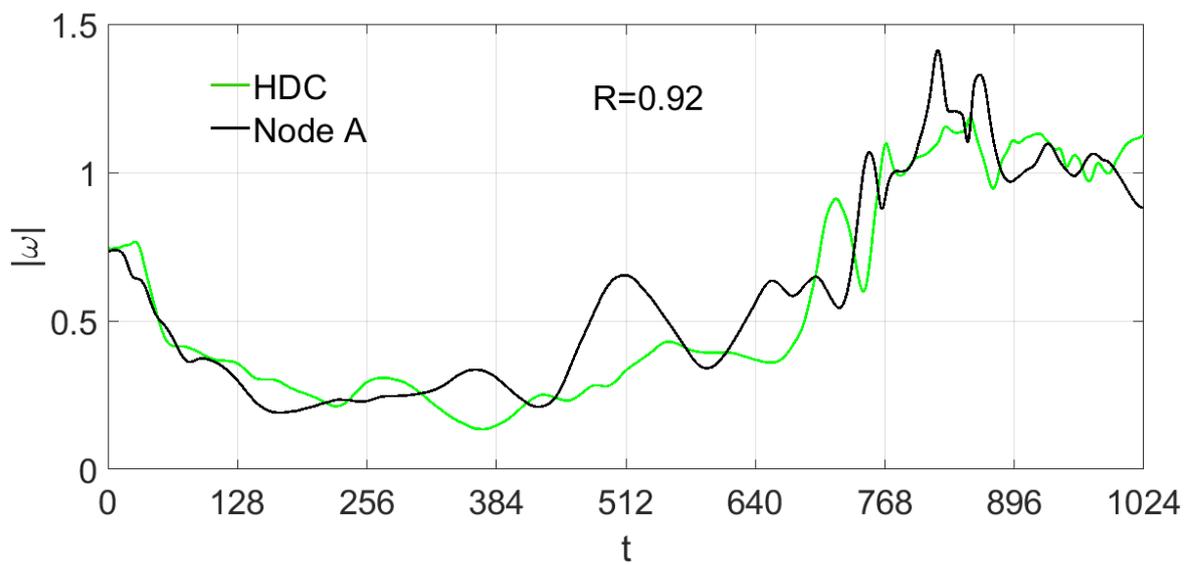
Fig. 4.4-2: Time series of the vorticity modulus $|\omega|$ are shown for the pair (HDC-A); the corresponding correlation coefficient, R, is also reported.
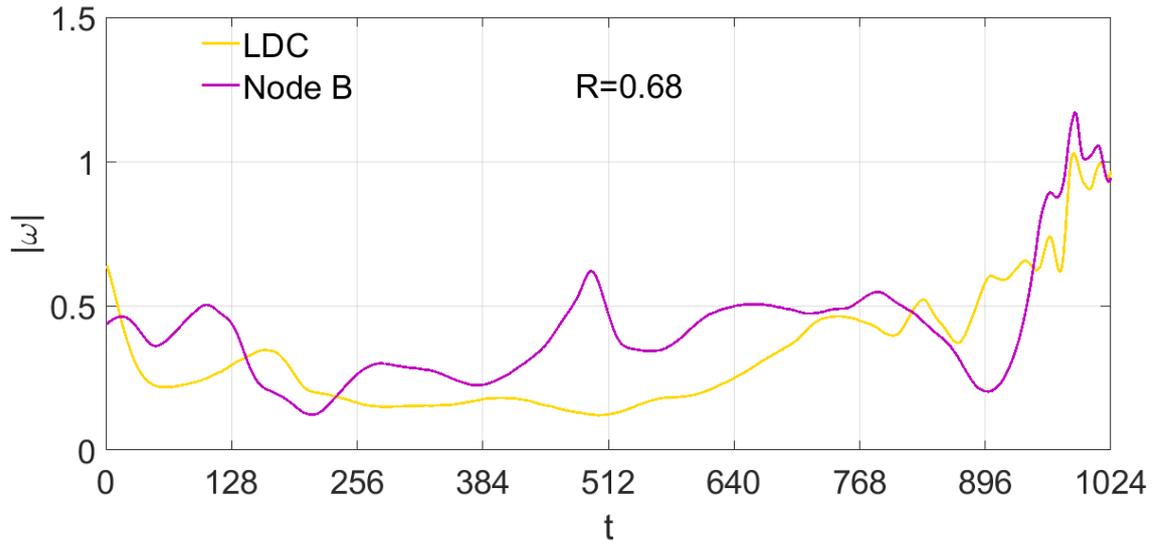
Fig. 4.4-3: Time series of the vorticity modulus $|\omega|$ are shown for the pair (LDC-B); the corresponding correlation coefficient, R, is also reported.

For each pair of nodes is also calculated the linear correlation coefficient R, reported in the figures: the couple $(HDC - A)$ exhibits a strong temporal correlation $R = 0.92$ for $|\bar{\omega}|$ while the couple $(LDC - B)$ has a much weaker temporal correlation $R = 0.68$. The behavior of the pairs $(HDC - A)$ and $(LDC - B)$ is also found for many other couples of nodes in the network. As instance, considering two nodes HDC' and LDC', with features reported in Table 4.3, and nodes $A' = (385, 401, 508)$ and $B' = (382, 393, 508)$, distant 10 grid cells along the X-axis from nodes HDC' and LDC', the linear correlation coefficients for the couples $(HDC' - A')$ and $(LDC' - B')$ are $R = 0.92$ and $R = 0.59$ respectively.

	Node HDC$'$	Node LDC$'$
Coordinates	$(375, 401, 508)$	$(372, 393, 508)$
Degree centrality, k_i	$5.641 \cdot 10^{-2}$	$5.536 \cdot 10^{-3}$
Neighborhood cardinality, $\#\nu(i)$	7265	713
Average physical distance, L_i	$7.539 \cdot 10^{-2}$	$3.263 \cdot 10^{-2}$

Table 4.3: Topological and spatial features of other two nodes with high (HDC') and low (LDC') degree centrality.

Thus, the behaviors for the two pairs just discussed are representative — respectively — of high and low degree centrality regions. It can be concluded that in general high degree centrality values indicate regions with the same instantaneous vorticity, that is, by definition, turbulent patterns coherently moving over the acquired time scale T_L. Besides, the average physical distance L_i, since is directly correlated with the degree centrality k_i (cf. Fig. 4.3-6), gives the order of magnitude of the spatial patterns identified by the k_i distribution, whose size ranges between the dissipative scale and the Taylor microscale.

4.5 Sensitivity analyses

The results previously illustrated and discussed have shown many interesting features. Indeed, they reveals a lot of typical properties of real and spatial networks (as a power law degree distribution) but also new insights into the spatial characterization of a turbulent flow (especially by degree centrality and by average physical distance). All the provided results are evaluated on a network built with a fixed threshold value $\tau = 0.9$ and considering the entire set of temporal series $t \in [1, 1024]$.

Two sensitivity analyses are here carried out, firstly considering different thresholds for a fixed set of energy time series, later examining different temporal windows, shorter than the complete one, at a fixed threshold τ.

4.5.1. Selecting a different threshold τ

The first sensitivity analysis consists in changing the threshold τ on the hypothesis 3) of the network building. Beside $\tau = 0.9$, networks for two different threshold values have been analyzed, specifically $\tau = 0.85$ and $\tau = 0.95$.

In Table 4.4 some general, topological and spatial features of the three networks are compared. As the threshold τ decreases, the network size grows both in N and M but the number links increases more rapidly than the number of nodes (Fig. 4.5-1); the opposite behavior holds if the threshold increase from $\tau = 0.9$ to $\tau = 0.95$.

In detail, while the number of nodes N_i inside the inner sphere does not vary as τ changes, the ratio M_i/M of internal links with respect to the total number increases when $\tau \to 1$. This behavior confirms again that nodes inside the inner sphere are strongly linked and this is emphasized when the threshold grows in value.

	$\tau = 0.85$	$\tau = 0.9$	$\tau = 0.95$
Nodes, N	172,713	128,785	75,874
Internal nodes, N_i	31,343	31,343	31,343
Links, M	243,355,115	80,920,781	11,061,400
Internal links, M_i	98,274,386	38,721,023	7,195,078
Edge density, $\rho(\tau)$	4.944%	2.282%	0.586%
Average degree, $\langle k \rangle$	$1.632 \cdot 10^{-2}$	$9.758 \cdot 10^{-3}$	$3.843 \cdot 10^{-3}$
Mean A.P.D., $\langle L \rangle$	$9.060 \cdot 10^{-2}$	$7.703 \cdot 10^{-2}$	$4.621 \cdot 10^{-2}$
L_{HDC}	$8.377 \cdot 10^{-2}$	$7.515 \cdot 10^{-2}$	$3.784 \cdot 10^{-2}$
L_{LDC}	$5.931 \cdot 10^{-2}$	$2.985 \cdot 10^{-2}$	$1.146 \cdot 10^{-2}$

Table 4.4: Main features of the networks with $\tau = 0.85; 0.9; 0.95$. HDC = (385, 401, 508) and LDC = (372, 387, 510).
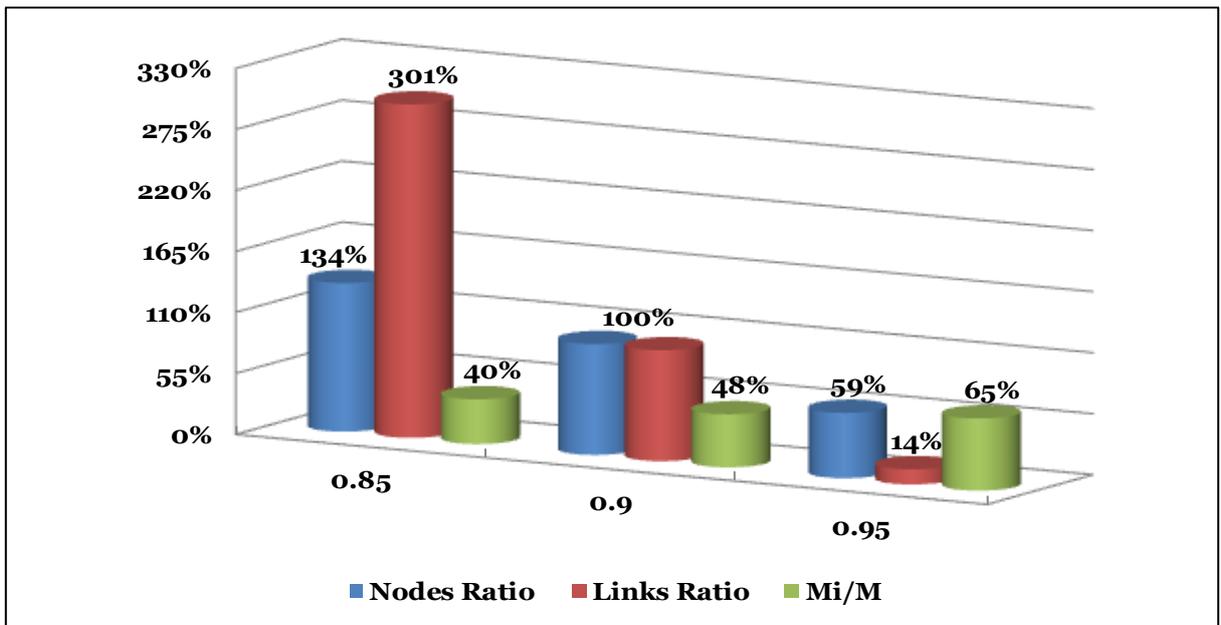
Fig. 4.5-1: Ratios of nodes (blue) and links (red) for networks at different τ to those at $\tau = 0.9$. (Green) ratio of internal and total links.

In Fig. 4.5-2 the degree centrality and the average physical distance on the plane $Z = 512$ are reported for the three values of the threshold τ. As the threshold is reduced, both the average degree centrality $\langle k \rangle$ and the mean average physical distance $\langle L \rangle$ decrease (Fig. 4.5-3), the former more rapidly than the latter; moreover, the regions with high values of k_i and L_i become more spatially expanded (Fig. 4.5-2). However, the order of magnitude of the spatial patterns — identified by high-k_i nodes and quantified by $\langle L \rangle$ for the entire network and by L_{HDC} and L_{LDC} for two representative nodes — does not substantially vary.
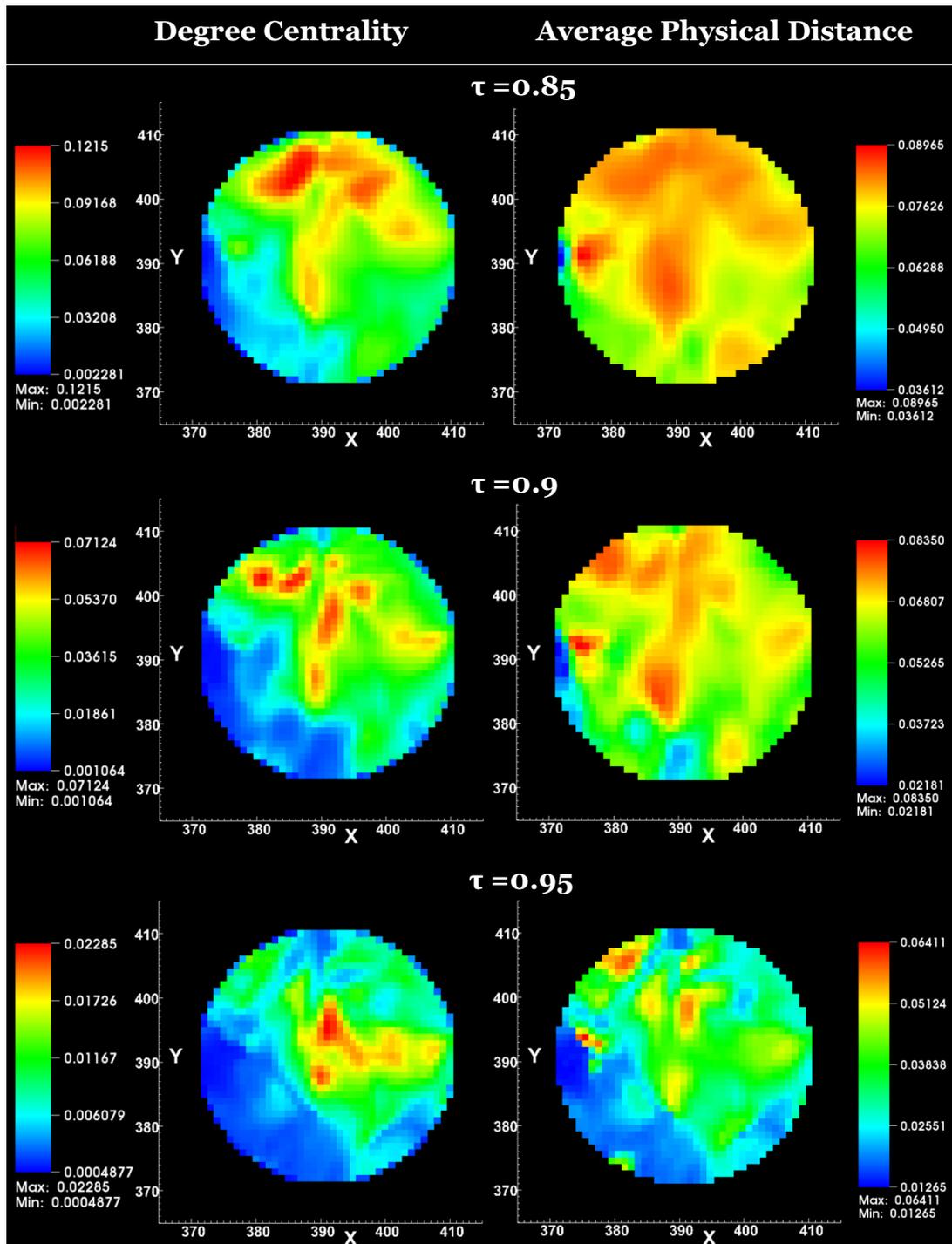
Fig. 4.5-2: Comparison of degree centrality (left) and average physical distance (right) in 2D sections on the Z=512 plane of the inner sphere, for the three threshold values of τ. Scales of values are referred to the sections.

In conclusion, despite the specific values assumed and the qualitative changes induced by the three threshold values, the spatial pattern detection is essentially independent from the choice of τ. If properly set, thus, the threshold value becomes a resolution parameter able to bring out spatial patterns (if τ is close to one too much information may be discarded, while a too low value of τ could spread the topological distributions, making the pattern detection much harder).
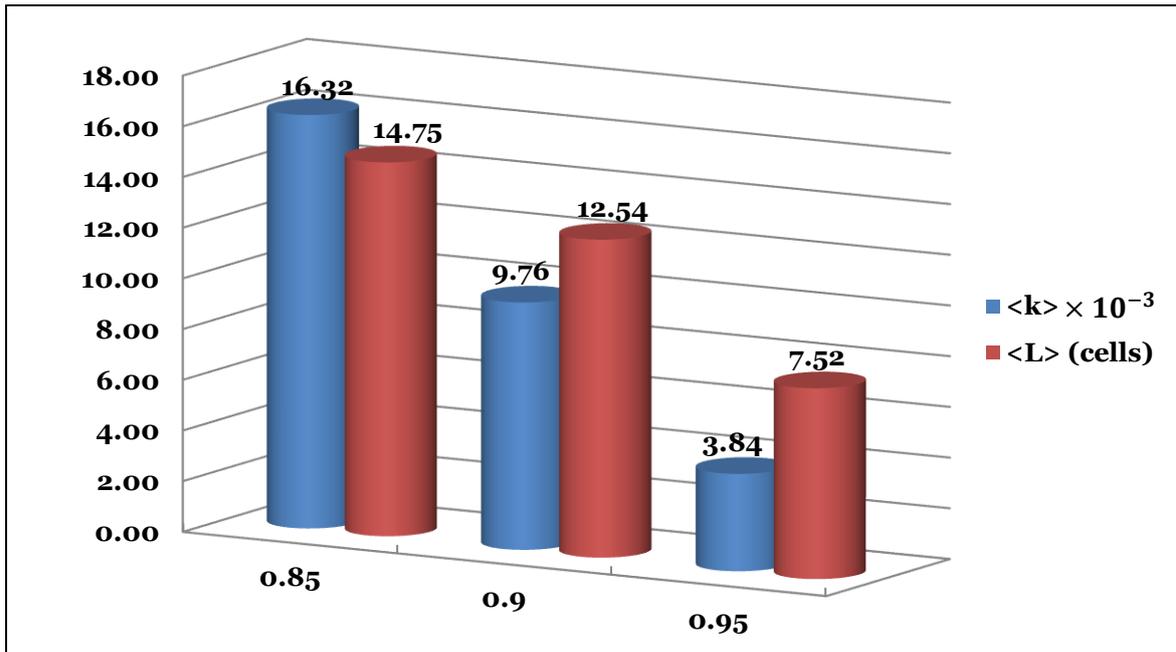
Fig. 4.5-3: Average degree centrality $\langle k \rangle$ and mean value of average physical distance $\langle L \rangle$ for the three networks built with the different threshold values $\tau = 0.85; 0.9; 0.95$.

4.5.2. Networks based on different temporal windows

A second sensitivity analysis is carried out considering different temporal windows from the complete one (TWC), and at fixed threshold $\tau = 0.9$. For the purpose, the following three slices of the complete time series (1-1024) are considered (Fig. 4.5-4):

- TW1, with time series $t \in [1, 600]$,
- TW2, with time series $t \in [213, 812]$,
- TW3, with time series $t \in [425, 1024]$,

which partially overlap and have all the same length (600 acquired times). In Table 4.5 and Fig. 4.5-5, the main features of the three networks are reported and compared with the complete temporal window (TWC).
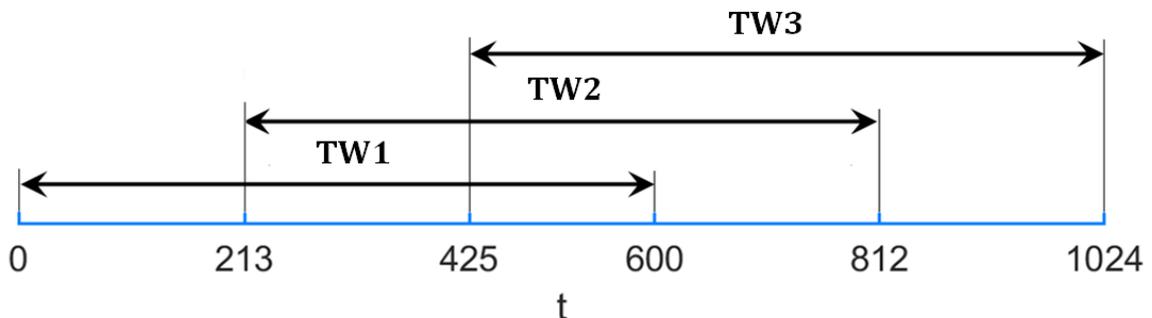
Fig. 4.5-4: Sketch of the three temporal windows analyzed in comparison with the complete one (TWC), $t \in [1, 1024]$.

	TW1	TW2	TW3	TWC
Nodes, N	82,641	123,593	157,420	128,785
Internal nodes, N_i	31,343	31,343	31,343	31,343
Links, M	9,316,475	87934279	108,262,150	80,920,781
Internal links, M_i	3,337,583	38,758,376	41,645,762	38,721,023
Edge density, $\rho(\tau)$	0.410%	2.600%	2.437%	2.282%
Average degree, $\langle k \rangle$	$2.728 \cdot 10^{-3}$	$1.151 \cdot 10^{-2}$	$8.738 \cdot 10^{-3}$	$9.758 \cdot 10^{-3}$
Max degree, $max(k_i)$	$1.086 \cdot 10^{-1}$	$1.082 \cdot 10^{-1}$	$8.923 \cdot 10^{-2}$	$7.905 \cdot 10^{-2}$
Min degree, $min(k_i)$	$1.210 \cdot 10^{-5}$	$8.091 \cdot 10^{-6}$	$6.352 \cdot 10^{-6}$	$7.765 \cdot 10^{-6}$
Mean A.P.D., $\langle L \rangle$	$6.166 \cdot 10^{-2}$	$8.056 \cdot 10^{-2}$	$8.663 \cdot 10^{-2}$	$7.703 \cdot 10^{-2}$

Table 4.5: Main topological and spatial features of the networks built on different temporal windows; last column reports the values of the complete temporal window as a comparison.
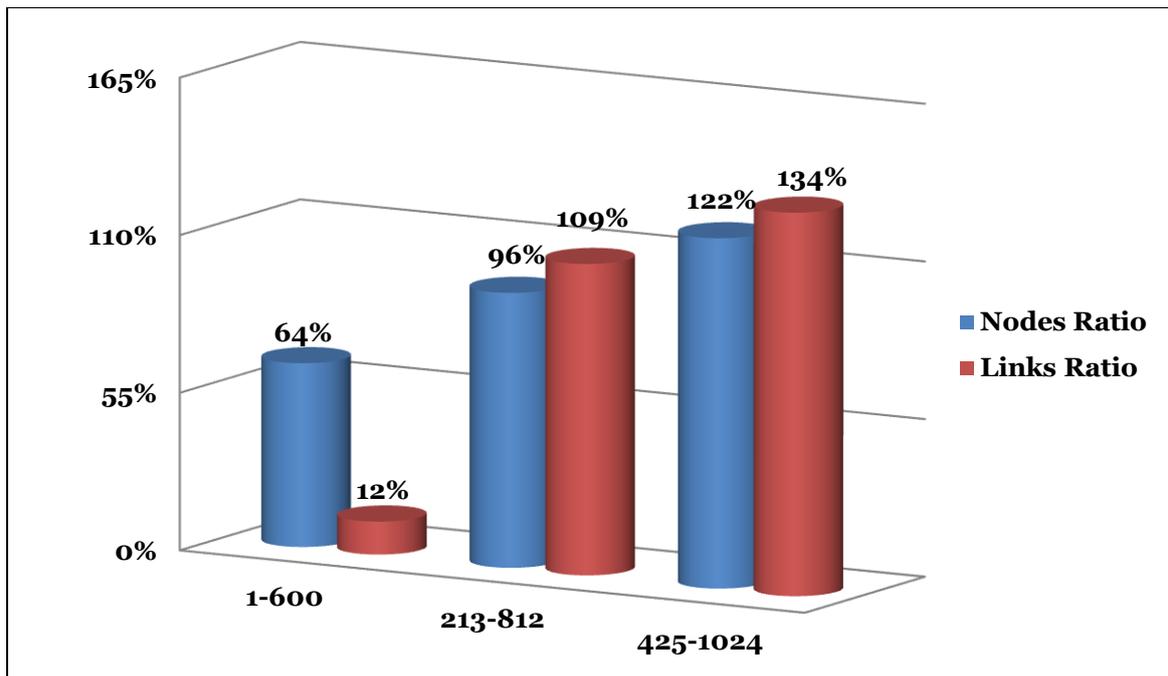
Fig. 4.5-5: Ratios of nodes (blue) and links (red) for networks built on the three different temporal windows to those of the complete temporal windows $[1, 1024]$.

The degree centrality spatial distributions are displayed both in a three-dimensional (Fig. 4.5-6) and bi-dimensional (Fig. 4.5-7) section views. As can be clearly observed, the sub-networks visibly change their features passing from the TW1 to the TW3. Both the number of nodes and links grow but at different rate (Fig. 4.5-5); indeed, M increases of two order of magnitude from TW1 to TW3, while N only doubles.
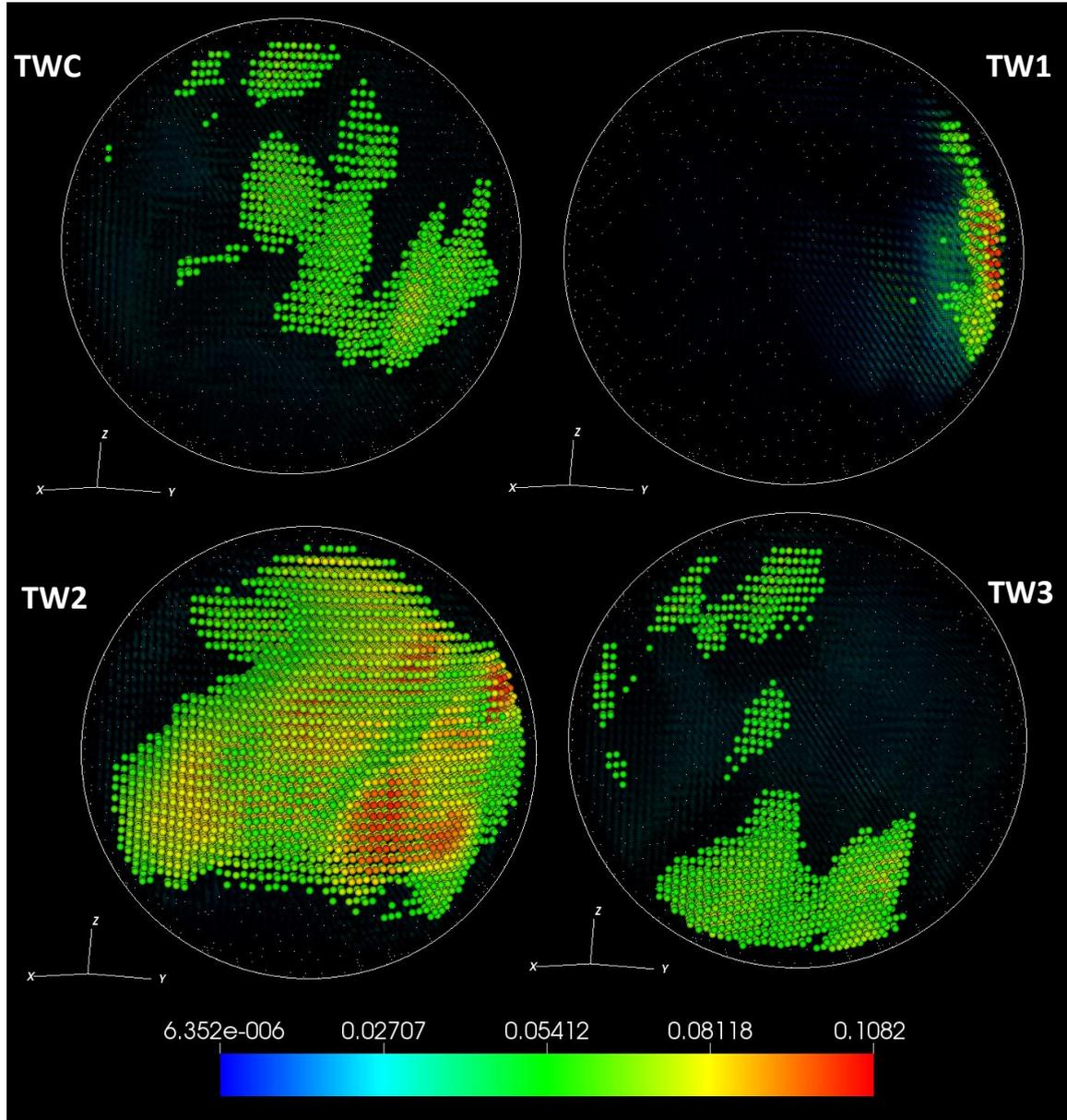
Fig. 4.5-6: Degree centrality in a 3D perspective for the networks with $\tau = 0.9$ and based on different time series. TWC: complete series $[1, 1024]$; TW1: first temporal series $[1, 600]$; TW2: second temporal series $[213, 812]$; TW3: third temporal series $[425, 1024]$.

Then, despite the last sub-network shows just more nodes and links than the TW2 sub-network, the degree centrality distributions look much different if the nodes with k_i above the 50% of the maximum value (in a unique scale) are highlighted, as reported in Fig. 4.5-6. This behavior is reflected on the mean values of k_i (Fig. 4.5-8): the TW2 sub-network, indeed, has the highest value of $\langle k \rangle$ because there is a lower number of nodes than the TW3 sub-network but they have higher degree values. To summarize, the temporal evolution of the three sub-networks suggests that spatial patterns can emerge and persist as long as the statistical stationarity of the field is preserved. The complete window network, instead, summarizes a framework which is the result of all the spatiotemporal connection variations over an integral timescale, T_L.
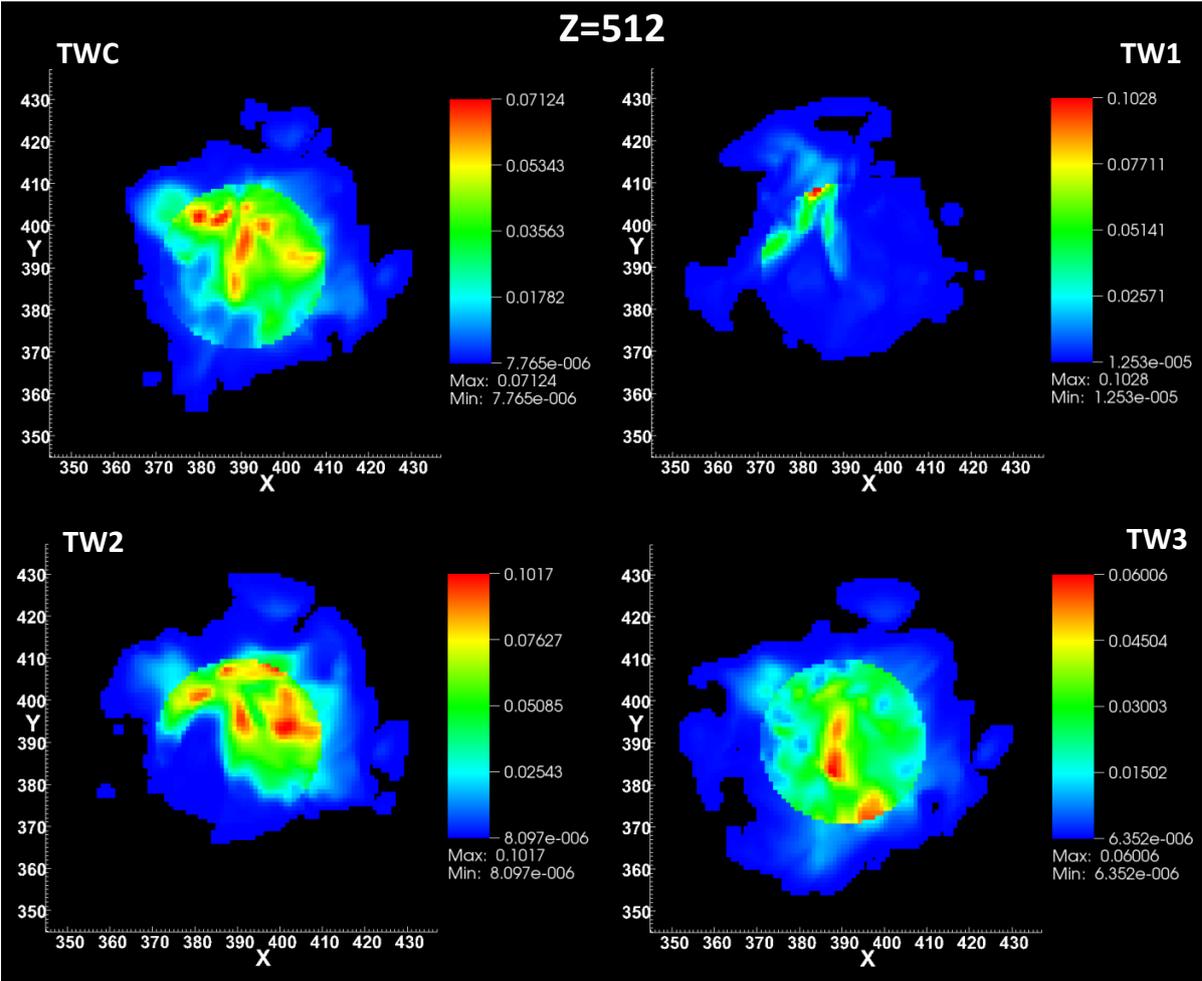
Fig. 4.5-7: Degree centrality of networks based on the four temporal windows in 2D section views on the plane Z=512.
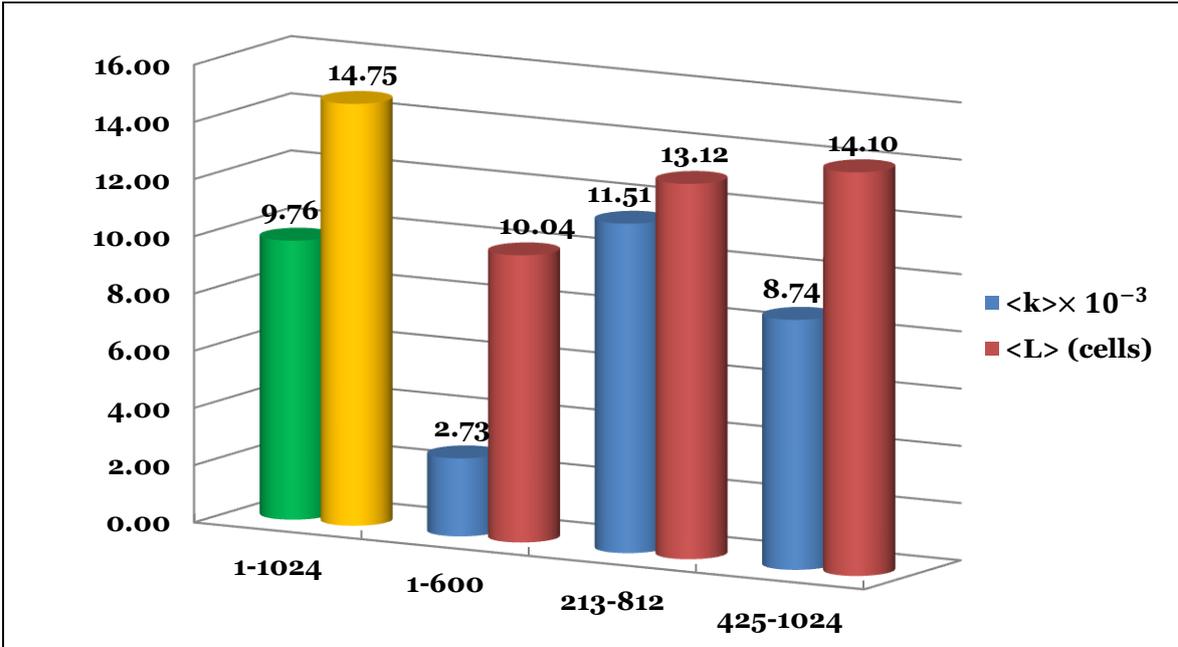
Fig. 4.5-8: A comparison of average degree centrality $\langle k \rangle$ and mean value of average physical distance $\langle L \rangle$ between the three sub-networks built on smaller time series and the complete one. Threshold $\tau = 0.9$.

4.6 Network on different regions: a comparison

As a final point, another network based on a different sphere of radius $r_c = 0.24$ has been studied. This part of domain, previously called region-2, is centered in $C' = (530, 673, 475)$, whose physical distance from node $C = (391, 391, 512)$ is about 1.93, which is largely exceeding the integral scale, $L = 1.376$.

The network is built with the same three hypotheses used in region-1 analysis, with a fixed threshold value $\tau = 0.9$ and considering the complete kinetic energy temporal series, $t \in [1, 1024]$, reported in eight panels in Fig. 4.6-1 and Fig. 4.6-2.
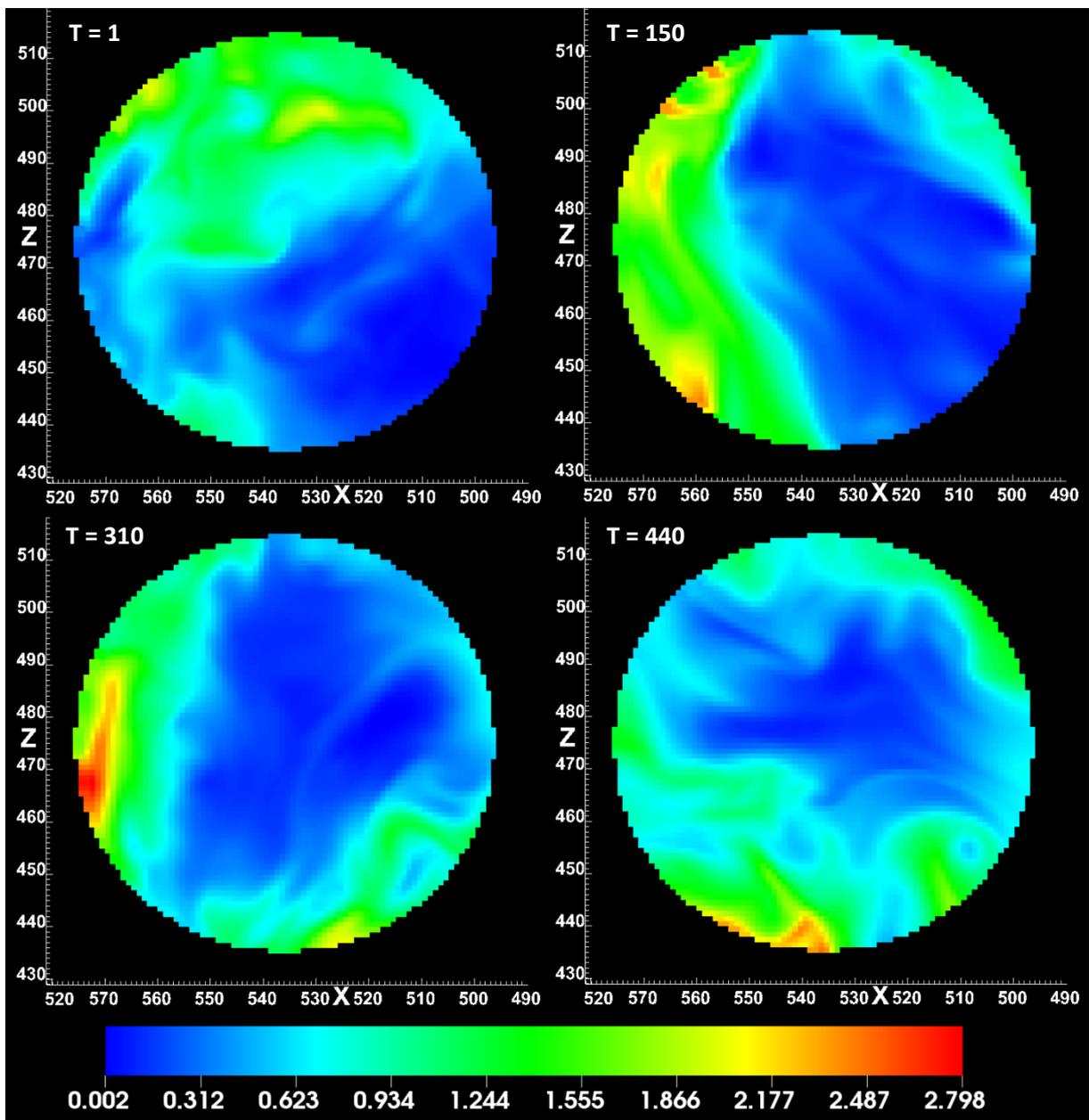
Fig. 4.6-1: Time evolution of the kinetic energy field in the section Y=673 of the outer sphere, T=1, 150, 310, 440.
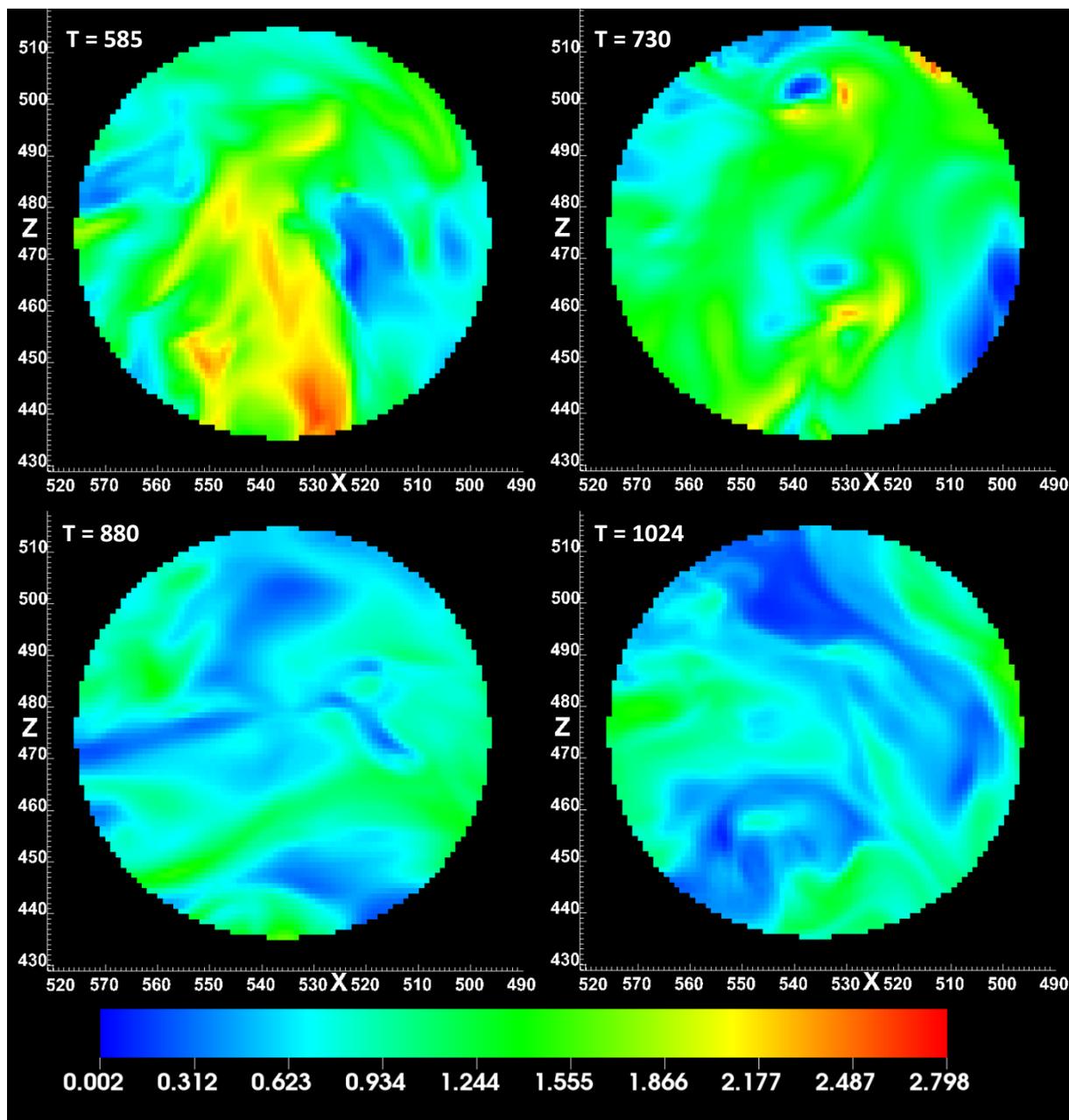
Fig. 4.6-2: Time evolution of the kinetic energy field in the section Y=673 of the outer sphere, T=585, 730, 880, 1024.

In Table 4.6 and Fig. 4.6-3, the main results are reported and compared.

	Region-1	Region-2
Nodes, N	128785	74432
Internal nodes, N_i	31343	31343
Links, M	80920781	38799854
Internal links, M_i	38721023	26525492
Edge density, $\rho(\tau)$	2.282%	2.107%
Average degree, $\langle k \rangle$	$9.758 \cdot 10^{-3}$	$1.401 \cdot 10^{-2}$
Average neighborhood cardinality, $\langle \#v \rangle = (N-1)\langle k \rangle$	1257	1043
Mean A.P.D., $\langle L \rangle$	$7.703 \cdot 10^{-2}$	$4.853 \cdot 10^{-2}$
Clustering Coefficient, C	$7.211 \cdot 10^{-1}$	$6.969 \cdot 10^{-1}$

Table 4.6: Main topological and spatial features of networks of different regions.
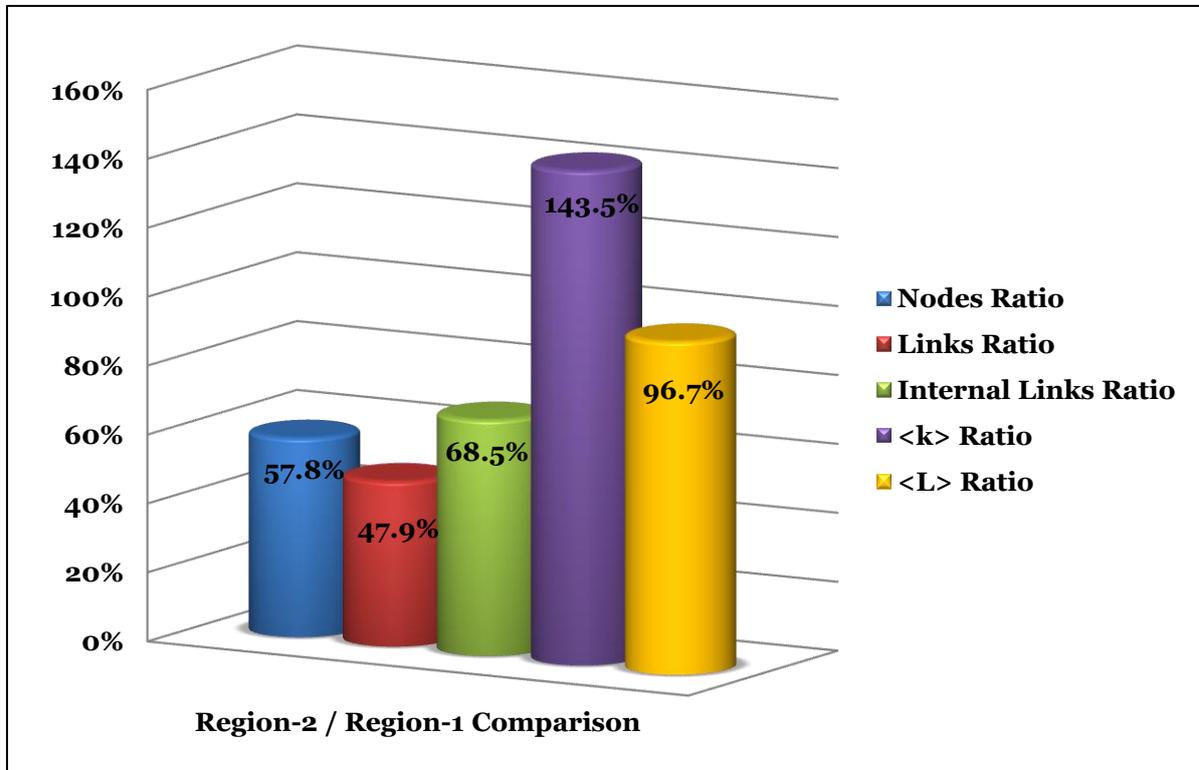
Fig. 4.6-3: : Ratios of nodes, total and internal links, average degree $\langle k \rangle$ and mean value of average physical distance $\langle L \rangle$ for network of region-2 to those of region-1.

As can be seen, both the number of nodes and links are lower in region-2 but, generally speaking, results are quite similar to those of region-1. The average degree, the mean value of the average physical distance and the global clustering coefficient are, indeed, different in values but their order of magnitude does not change significantly.

In addition, the network built on second region also displays a power law behavior for the cumulative degree distribution with a power exponent nearly the same of region-1 (Fig. 4.6-4) and with the same degree of approximation (the coefficient of determination $R^2 \cong 0.93$ is the same in both cases).
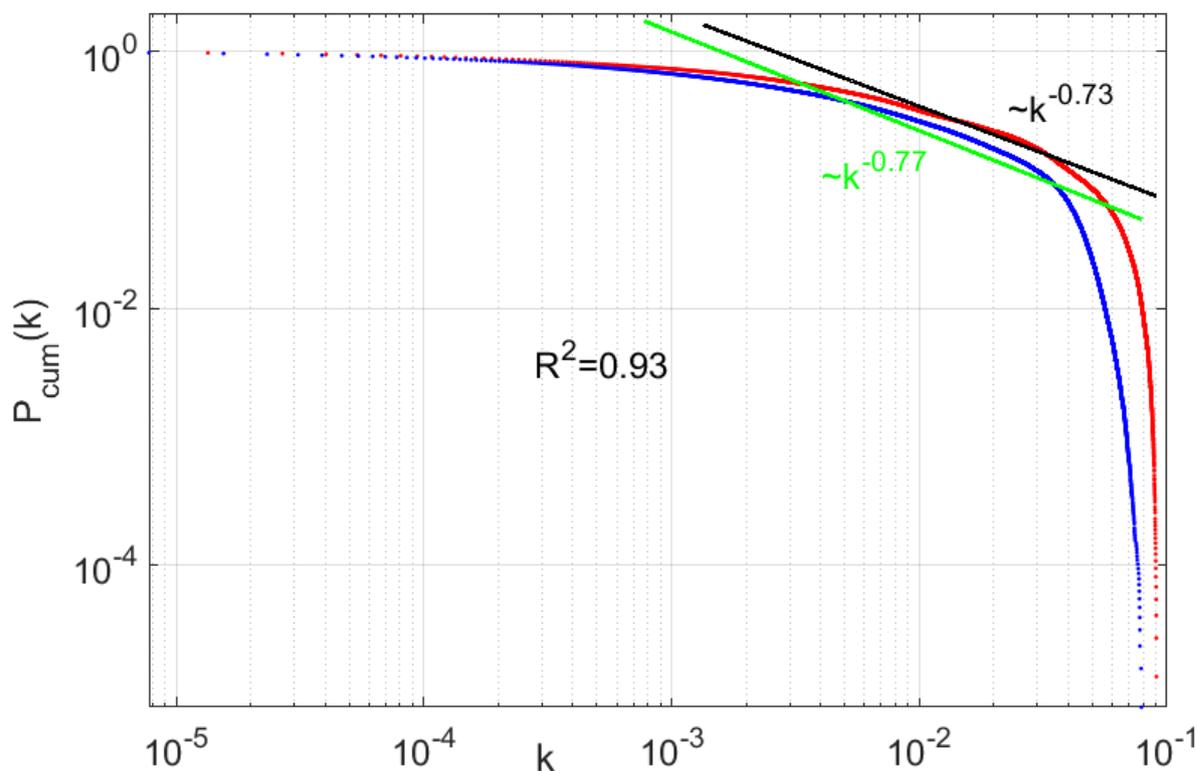
Fig. 4.6-4: Cumulative degree distributions (blue, red) and the power law behaviors (green, black) for region-1 and region-2, respectively. The exponents are almost equal with the same coefficient of determination.

Fig. 4.6-5 shows two 3D views of the k_i map for the region-2. Despite the spatial distribution differences with the region-1 and the specific values of the topological features, it is quite evident the presence of clusters with high degree with nearly the same mean value of average topological physical distance.
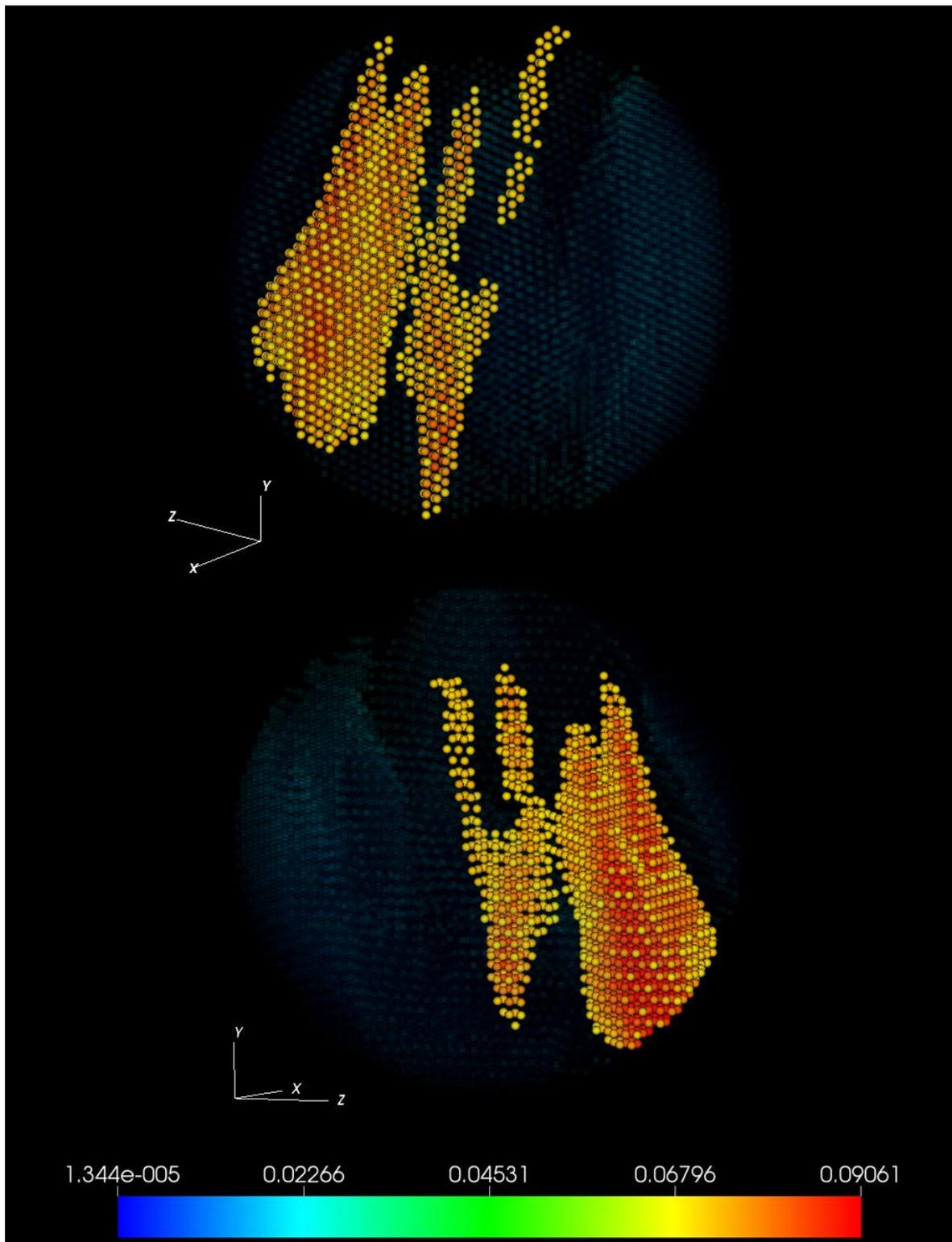
Fig. 4.6-5: Two 3D visualizations of the inner spheres of region-2. Nodes above the threshold $k_i \geq 0.07$ are highlighted.

Fig. 4.6-6 shows the degree and eigenvector centralities, the average physical distance and the local clustering coefficient in the 2D section $Z = 475$.
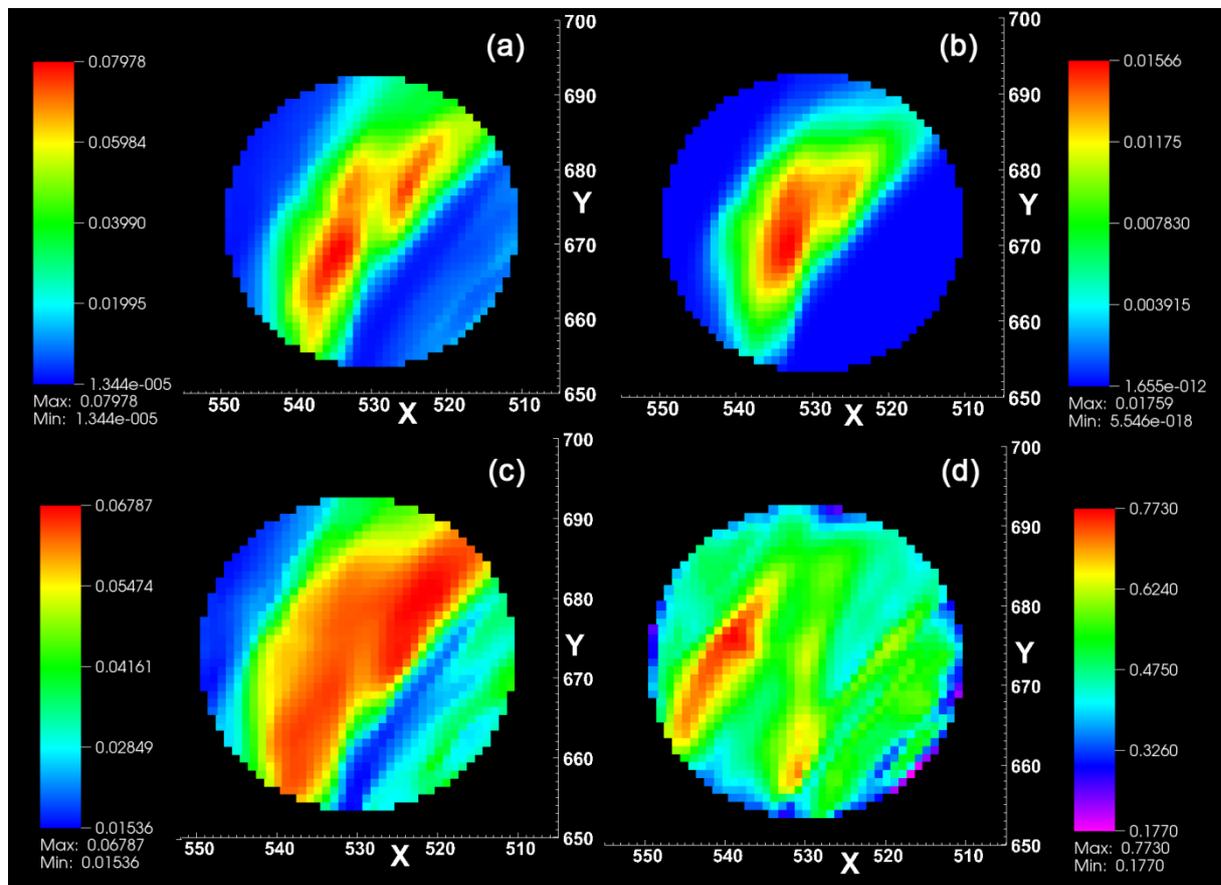
Fig. 4.6-6: Region-2: (a) degree centrality, (b) eigenvector centrality, (c) average physical distance, (d) local clustering coefficient. Results are displayed on the section $Z = 475$ of the inner sphere.

Therefore, the presence of spatial patterns with different size and intensity (though, quite similar as orders of magnitude) is not limited to the chosen domain portion but can involve the whole turbulent field, despite it is statistically homogeneous and isotropic.

Conclusions

In the present work, the complex network tools have been applied to study a forced isotropic turbulent field. It was expected that the analysis of a spherical, randomly selected sub-region of the domain (with size of the order of the Taylor microscale) does not provide particular, organized patterns but spotted or homogeneous spatial distributions of the fluid dynamics parameters evaluated. Indeed, both the features of the chosen turbulent flow (which is statistically homogeneous and isotropic at smaller scales with a energy forcing at larger ones) and the relatively short simulation time interval acquired (that is just only one integral timescale) do not suggest — *a priori* — a plausible reason to find a network metric that may highlight coherent patterns in the turbulent flow.

The evaluation of the spatial distributions of the degree centrality, k_i, and the eigenvector centrality, however, reveals that spatial coherent patterns exist and that they are identified by high values (indicatively above of the 70% of the maximum) of both the centrality indices. A new metric, named as average physical distance, L_i, was then introduced to estimate the size of the neighborhoods of the nodes in the network. The spatial distribution of L_i (which varies from small scales up to the Taylor microscale) follows, as expected, quite closely the degree centrality distribution; the average physical distance provides, hence, a useful measure of the size of such spatial patterns.

Conversely, the betweenness centrality and the clustering coefficient showed spotted distributions, in particular with high gradient values in the former. The clustering coefficient and the degree centrality are found to be independent presumably because of the strong spatial constraints of the network; that is a typical behavior of spatial, real network as reported in many other works.

The community detection is the other important topological feature of the network that has been studied. Two algorithms were used, both based on the maximization of the modularity: (i) the Newman's method, which applies a recursive, spectral algorithm to the so called modularity matrix of smaller and smaller sub-networks; (ii) the Louvain method that is much faster than the Newman's one but less accurate, since a heuristic algorithm is used. Both the methods provided modularity values $Q > 0.3$, meaning there is a pronounced division of the network in smaller, self-organized structures. Moreover, the network division obtained with the Newman's method indicates that nodes with high degree centrality values also belong to community with higher values of modularity. Therefore, these results confirm the elevate capability of the Newman's method to find communities quite accurately.

Besides, other features of the network were briefly analyzed. The power law scaling of the degree distribution $P(k)$ and of the cumulative degree distribution $P_{cum}(k)$

indicates that the network is scale-free in a limited range of degree centrality values. An assortativity study was also carried out revealing — through the evaluation of the joint degree distribution, the k-nearest neighbors parameter and the Pearson degree correlation parameter — that the network is non-assortative but it exhibits a Rich-club phenomenon.

In order to complete the study, a physical interpretation of results was provided. Two pairs of nodes were selected: the first couple includes one node with a high-k_i value (HDC) while the second couple includes a node with a low-k_i value (LDC) and the other two nodes were selected to be at a fixed distance from HDC and LDC nodes. The linear correlation coefficients of the time series of the vorticity modulus was then calculated for each pair: while the couple with HDC node exhibits a high temporal correlation, the couple with LDC node instead shows lower values of correlation. This behavior is representative of high and low degree regions, meaning that, in general, high values of k_i indicate regions with the same instantaneous vorticity, i.e. turbulent patterns coherently moving over the integral timescale T_L.

Additionally, the results of the two sensitivity analyses carried out indicate that — despite the specific values assumed and the qualitative changes induced by the threshold value τ and by changing the temporal window — the spatial pattern detection is essentially independent from the choice of τ and such patterns (as well as all the sub-networks) exhibit a temporal evolution over the integral timescale.

Furthermore, despite the differences in general, spatial and topological features (such as the number of nodes and links), the structural analysis of the network built on the region-2 shows the same behavior of the first region. Indeed, both the degree and eigenvector centralities highlight the presence of coherent patterns over the integral timescale, while the average physical distance quantifies the size of such patterns. The presence of spatial patterns with different size and intensity is not limited then to the chosen sub-domain but can involve the whole turbulent field.

In the wake of the great development of the complex networks in practical applications that took place in the last years — in particular those regarding turbulent flows —, the procedure illustrated in this work turns out to be a useful tool to identify, in a systematic way, spatial patterns on turbulent regions. In this context, the metrics and the structural properties provided by network theory are a suitable tool to investigate the spatial characterization of turbulent flows, since a huge amount of detailed information can be synthesized in a single framework.

However, the ability of the large complex networks to condensate in a single framework the essential information is a double-edged sword. On one side, indeed, a multipoint analysis can be carried out with few suitable metrics; on the other side, though, the huge amount of data that needs to be handled, both in the pre and post-processing stages, requires high computational capabilities and costs (which are not even known *a priori*). The first great operative limitation encountered was the evaluation and the assembly of the matrix with the correlation coefficients, since its size scales as the square of the number of nodes in the network, and that is reflected on

the adjacency matrix. The partitioning as used in this work is a solution but, naturally, higher computational performance are suggested as first improvement.

Moreover, beyond the computational point of view, the choice of which software should be used to build the network and to visualize the results is not trivial. There exists, as a matter of fact, many tools (in some cases developed *ad hoc* for specific applications) capable to handle huge amount of data or to deal with large complex networks. Nonetheless, each new implementation of the network theory to a different research area or a particular subject (such as the study of turbulent flows) involve particular requirements often satisfied by a combination of more than one software. Anyhow, this choice is subordinated or in conjunction with the available computational capacity, as already mentioned.

Future investigations can involve several aspects of the work. First of all, different types of flows may be explored, such as flows with strong spatial, thermal or density inhomogeneities; these can be also studied through other approaches, as the visibility algorithm or recurrence networks. Secondly, the hypotheses underlying the network building may be slightly modified, especially in the size of spheres considered. In addition, further parameters can be evaluated, as alternative to those unsuitable for the purpose, such as the random betweenness centrality instead of betweenness centrality defined with the shortest paths. A deeper attention may be also dedicated to the sensitivity analysis based on different temporal windows: sub-networks built with shorter time intervals could be a starting point to develop a dynamic study of the complex network considered.

In conclusion, the proposed approach can suggest new insights into the spatial characterization of turbulent flows and the application to a wide range of fluid systems seems to be promising and is worth additional future investigation.

Bibliography

[1] Feynman, R. P., Leighton, R. B., Sands, M. (2013). *The Feynman Lectures on Physics, Desktop Edition Volume I* (Vol. 1). Basic books.

[2] Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuno, J. C. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, *105*(13), 4972-4975.

[3] Donner, R. V., Small, M., Donges, J. F., Marwan, N., Zou, Y., Xiang, R., Kurths, J. (2011). Recurrence-based time series analysis by means of complex network methods. *International Journal of Bifurcation and Chaos*, *21*(04), 1019-1046.

[4] Scarsoglio, S., Iacobello, G., Ridolfi, L. (2016). Complex networks unveiling spatial patterns in turbulence. *EPL (Europhysics Letters)*, Under Review.

[5] Iacobello, G., Scarsoglio, S., Ridolfi, L. (2016). New insights into spatial characterization of turbulent flows: a complex network-based analysis. *8th European Postgraduate Fluid Dynamics Conference* (EPFDC)

[6] Dyke, Milton Van. (1982) An Album of Fluid Motion. *Stanford, CA: Parabolic. Print.*

[7] Pope, S. B. (2000) Turbulent Flows. *Cambridge: Cambridge UP. Print.*

[8] Danforth, Christopher M. (2013). Chaos in an Atmosphere Hanging on a Wall. *Mathematics of Planet Earth 2013*. Retrieved 4 April 2013.

[9] Tennekes, H., Lumley, J. L. (1972). A first course in turbulence. *MIT press*.

[10] Mathiew J., Scott J. (2000). An Introduction to Turbulent Flow. *Cambridge University Press*

[11] Farge, M., Kevlahan, N. K. R., Perrier, V., Schneider, K. (1999). Turbulence analysis, modelling and computing using wavelets. In *Wavelets in Physics* (Vol. 1, p. 117).

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

[13] Costa, L. D. F., Rodrigues, F. A., Travieso, G., Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, *56*(1), 167-242.

[14] Flory, P. J. (1941). Molecular size distribution in three dimensional polymers. i. gelation1. *Journal of the American Chemical Society*, *63*(11), 3083-3090.

[15] Rapoport, A. (1951). Nets with distance bias. *The bulletin of mathematical biophysics*, 13(2), 85-91.

[16] Rapoport, A. (1953). Spread of information through a population with socio-structural bias: I. Assumption of transitivity. *The bulletin of mathematical biophysics*, 15(4), 523-533.

[17] Rapoport, A. (1957). Contribution to the theory of random and biased nets. *The bulletin of mathematical biophysics*, 19(4), 257-277.

[18] Erdős, P., Rényi, A. (1959). On random graphs I. *Publ. Math. Debrecen*, *6*, 290-297.

[19] Erdős, P., Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, *5*, 17-61.

[20] Erdős, P., Rényi, A. (1961). On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, *12*(1-2), 261-267.

[21] Watts, D. J., Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, *393*(6684), 440-442.

[22] Barabási, A. L., Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509-512.

[23] Girvan, M., Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, *99*(12), 7821-7826.

[24] Chen, L. L., Blumm, N., Christakis, N. A., Barabasi, A. L., Deisboeck, T. S. (2009). Cancer metastasis networks and the prediction of progression patterns. *British Journal of Cancer*, *101*(5), 749-758.

[25] Gao, Z., Jin, N. (2009). Flow-pattern identification and nonlinear dynamics of gas-liquid two-phase flow in complex networks. *Physical Review E*, *79*(6), 066303.

[26] Charakopoulos, A. K., Karakasidis, T. E., Papanicolaou, P. N., Liakopoulos, A. (2014). The application of complex network time series analysis in turbulent heated jets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *24*(2), 024408.

[27] Scarsoglio, S., Laio, F., Ridolfi, L. (2013). Climate dynamics: a network-based approach for the analysis of global precipitation. *PloS one*, *8*(8), e71129.

[28] Abe, S., Suzuki, N. (2004). Scale-free network of earthquakes. *EPL (Europhysics Letters)*, *65*(4), 581.

[29] Bader, D., Madduri, K. (2006, August). Parallel algorithms for evaluating centrality indices in real-world networks. In *Parallel Processing, 2006. ICPP 2006. International Conference on* (pp. 539-550). IEEE.

[30] Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215-239.

[31] Alex Bavelas. Communication patterns in task-oriented groups. *Journal of the acoustical society of America*, 22(6):725–730, 1950.

[32] Rochat, Y. (2009). Closeness centrality extended to unconnected graphs: The harmonic centrality index. In *ASNA* (No. EPFL-CONF-200525).

[33] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[34] Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Social networks*, *27*(1), 39-54.

[35] Newman, M. E. (2008). The mathematics of networks. *The new palgrave encyclopedia of economics*, 2(2008), 1-12.

[36] Newman, M. E. (2002). Assortative mixing in networks. *Physical review letters*, *89*(20), 208701.

[37] Colizza, V., Flammini, A., Serrano, M. A., Vespignani, A. (2006). Detecting rich-club ordering in complex networks. *Nature physics*, *2*(2), 110-115.

[38] McAuley, J. J., da Fontoura Costa, L., Caetano, T. S. (2007). Rich-club phenomenon across complex network hierarchies. *Applied Physics Letters*, *91*(8), 084103.

[39] Newman, M. E., Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, *69*(2), 026113.

[40] Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical review E*, *69*(6), 066133.

[41] Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

[42] Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, *103*(23), 8577-8582.

[43] Csányi, G., Szendrői, B. (2004). Fractal–small-world dichotomy in real-world networks. *Physical Review E*, *70*(1), 016122.

[44] Jarukasemratana, S., Murata, T. (2013). Recent Large Graph Visualization Tools: A Review. *Information and Media Technologies*, *8*(4), 944-960.

[45] Perlman, E., Burns, R., Li, Y., Meneveau, C. (2007, November). Data exploration of turbulence simulations using a database cluster. In *Proceedings of the 2007 ACM/IEEE conference on Supercomputing* (p. 23). ACM.

[46] Donges, J. F., Zou, Y., Marwan, N., Kurths, J. (2009). Complex networks in climate dynamics. *The European Physical Journal Special Topics*, *174*(1), 157

[47] Gao, Z. K., Jin, N. D. (2012). Characterization of chaotic dynamic behavior in the gas–liquid slug flow using directed weighted complex network analysis. *Physica A: Statistical Mechanics and its Applications*, *391*(10), 3005-3016.

[48] Gao, Z. K., Fang, P. C., Ding, M. S., Yang, D., Jin, N. D. (2015). Complex networks from experimental horizontal oil–water flows: Community structure detection versus flow pattern discrimination. *Physics Letters A*, *379*(8), 790-797.

[49] Gao, Z. K., Yang, Y. X., Fang, P. C., Zou, Y., Xia, C. Y., Du, M. (2015). Multiscale complex network for analyzing experimental multivariate time series. *EPL (Europhysics Letters)*, *109*(3), 30005.

[50] Gao, Z. K., Fang, P. C., Ding, M. S., Jin, N. D. (2015). Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow. *Experimental Thermal and Fluid Science*, *60*, 157-164.

[51] Shirazi, A. H., Jafari, G. R., Davoudi, J., Peinke, J., Tabar, M. R. R., Sahimi, M. (2009). Mapping stochastic processes onto complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2009*(07), P07046.

[52] Murugesan, M., Sujith, R. I. (2015). Combustion noise is scale-free: transition from scale-free to order at the onset of thermoacoustic instability. *Journal of Fluid Mechanics*, *772*, 225-245.

[53] Liu, C., Zhou, W. X., Yuan, W. K. (2010). Statistical properties of visibility graph of energy dissipation rates in three-dimensional fully developed turbulence. *Physica A: Statistical Mechanics and its Applications*, *389*(13), 2675-2681.

[54] Manshour, P., Tabar, M. R. R., Peinke, J. (2015). Fully developed turbulence in the view of horizontal visibility graphs. *Journal of Statistical Mechanics: Theory and Experiment*, *2015*(8), P08031.

[55] Fan, X., Wang, L., Xu, H., Li, S., Tian, L. (2015). Characterizing air quality data from complex network perspective. *Environmental Science and Pollution Research*, 1-11.

[56] Xu, X., Zhang, J., Small, M. (2008). Superfamily phenomena and motifs of networks induced from time series. *Proceedings of the National Academy of Sciences*, *105*(50), 19601-19605.

[57] Yang, Y., Yang, H. (2008). Complex network-based time series analysis. *Physica A: Statistical Mechanics and its Applications*, *387*(5), 1381-1386.

List of Links

[I] TURBULENCE MODELING FOR BEGINNERS, Tony Saad, University of Tennessee Space Institute

www.cfd-online.com/W/images/3/31/Turbulence_Modeling_For_Beginners.pdf

[II] Mathworks. (2015). *Documentation-Matlab-Mathematics-Elementary Math-Descriptive Statistics-corrcoef* (R2015b).

http://it.mathworks.com/help/matlab/ref/corrcoef.html

[III] Visual Complexity, A unified resource space for the visualization of complex networks.

http://www.visualcomplexity.com/vc/

[IV] Seven Bridges of Königsberg, Visual Representation

https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

[V] An Introductory Course on Network Analysis, An introduction to Centrality measures.

https://sites.google.com/site/networkanalysisacourse/

[VI] VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data.

https://wci.llnl.gov/simulation/computer-codes/visit

[VII] Johns Hopkins Turbulence Database. Database Access and Documentation.

http://turbulence.pha.jhu.edu