NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SCHOOL OF APPLIED MATHEMATICAL
AND PHYSICAL SCIENCES

MATHEMATICAL MODELLING IN MODERN
TECHNOLOGIES AND FINANCE

Master Thesis

Analysis of Rough Surfaces with Network Theory

Nikoleta V. Karasmani

Supervisor: Antonios Symvonis, Professor

March 2012, Athens

Acknowledgements

I would like to thank Mr V. Constantoudis for his support, encouragement and valuable guidance. I thank Mr A. Symvonis and Mr. C. Siettos for their assistance. I would also like to thank Kostas for the support (technical and ethical) and his incredible patience. I am grateful to my friends: Natassa for her practical help, Nikoleta, Vaso, Giannis and Chara for their continuous interest. Finally, I would like to thank my family for their support, faith and interest in me.

Abstract

The motivation for studying the roughness of surfaces has two aspects: the technological aspect is that surface roughness affects several surface properties (adhesion, catalysis, friction, conductivity) and is of increased importance in nanotechnology since the surface to volume ratio increases on nanoscale. Moreover, the theoretical aspect is that a rough surface is a complex system (an hierarchical synergy of order and randomness) whose evolution is a non equilibrium process. However, the characterization of this synergy in rough surfaces has open issues. At this point the need for a new approach for studying rough surfaces emerges. Complex network theory, which provides us with a new perspective for understanding a complex system from the relations between the elements in a global way, can constitute this new approach.

The aim of this thesis is to investigate the structure of rough surfaces through complex network theory. First we propose a method for constructing a complex network from a rough surface. From random to fully periodic as well as fractal self–affine and mounded rough surfaces are generated and studied. Thereafter, the network's statistical measures, such as the mean degree, the degree distribution, the clustering coefficient and the average shortest path length are examined. In some cases, a comparison of the constructed networks with other existing network models, such as the ring lattice and the random graph, is investigated. The main findings of the thesis are:

1. The mean degree $\langle k \rangle$ in the cases of white noise, square and triangular pulse can be analytically estimated.

2. The peak of the degree distribution $P(k)$ is sensitive to the surface's periodicity. In fully periodic rough surfaces we can have either a type I peak—coming from points in the same height zone—, either a type II peak —coming from points in different height zones—, or both types of peaks. In a white noise we have a Poisson degree distribution and in fractal self–affine, $P(k)$ is sensitive to any change of the rms parameter. Finally, in mounded rough surfaces, when we have deviations from full periodicity both type I and II peaks are affected.

3. The clustering coefficient C for networks produced by white noise or triangular pulse approximates the vale $\frac{3}{4}$. A recursive proof that $C \sim 0.75$ is given. In general, our constructed networks are highly clustered.

4. The average shortest path length L is also long and in some cases diverges.

Committee:
Antonios Symvonis, Professor, N.T.U.A.
Constantinos Siettos, Assistant Professor, N.T.U.A.
Vassilios Constantoudis, PhD, N.C.S.R. "Demokritos"

Contents

Chapter 1

Introduction and Motivation

1.1 Characterization of Rough Surfaces

1.1.1 Preliminaries

Roughness is a property of surfaces that can be observed very often in nature, either on macroscopic level (terrain morphology like mountains, lakes, rivers and canyons) or on microscopic scale (metallic surfaces are assumed to be smooth but, from a closer look, height fluctuations are revealed).

In modern scientific researches as well as engineering studies, rough surfaces have attracted more and more attention in a wide range of fields including, for example, acoustics, remote sensing and geophysics. Interface roughness is one of the key features in many important thin film technologies. The roughness of the interface directly controls many physical and chemical properties of the films and the reduction of this property is desirable for many applications in microelectronic and optoelectronic devices.

Surface roughness is a measure of the texture of a surface. It is quantified by the vertical deviations of a real surface from its ideal form. If these deviations are large, the surface is rough; if they are small the surface is smooth. A classification of rough surfaces (RS) according to their statistical properties can give us two categories: deterministic and stochastic RS.

A deterministic surface is defined in the way that the surface height fluctuation can be expressed as a deterministic function of surface position. Once the form of a height function is given, the property of the whole surface is determined. In the statistical point of view, surface heights between two arbitrary points on a deterministic surface are totally correlated.

A random rough surface can be described mathematically as $h = h(\vec{r})$, where h is the surface height of a rough surface with respect to a smooth reference surface defined by a mean surface height and $\vec{r}$ is the position vector on the surface. The height h is a single-valued function of the position vector $\vec{r}$. We can assume that the height fluctuation is a random field with respect to the position. For a complete description of a random field one needs to know the n-dimensional joint distribution function $p_n(h_1, h_2, ..., h_n; \vec{r_1}, \vec{r_2}, ..., \vec{r_n})$, where $\vec{r_1}, \vec{r_2}, ..., \vec{r_n}$ are a set of different positions on the surface and $h_1, h_2, ..., h_n$ are the corresponding random variables. A random field is called homogeneous if all the joint probability distribution functions remain the same when the set of locations $\vec{r_1}, \vec{r_2}, ..., \vec{r_n}$ is translated (but not rotated) in the parameter space. This implies that all of the probabilities depend only on the relative, not the absolute locations of the points $\vec{r_1}, \vec{r_2}, ..., \vec{r_n}$. The field is isotropic if the joint probability density functions are invariant after the constellation of points $\vec{r_1}, \vec{r_2}, ..., \vec{r_n}$ is rotated in the parameter space.

1.1.2 First-Order Statistics

One of the main characteristics of a random rough surface is the *height distribution function*, $p(h)$. The meaning of $p(h)$ is that the probability of a surface height between h and $h + dh$ at any point on the surface is $p(h)dh$. The distribution $p(h)$ is a non-negative function of h and is normalized such that

$$\int_{-\infty}^{+\infty} p(h)dh = 1.$$

The height distribution function provides a complete specification of the random variable $h(\vec{r})$ at a position $\vec{r}$.

To describe a particular property of a random variable h, it is often much more convenient to use simpler numerical statistics determined by the distribution function. The nth-order moment of a variable h, defined as

$$m_n = E\{h^n\} = \int_{-\infty}^{+\infty} h^n p(h)dh,$$

is the most important numerical statistic. More generally, one can define the nth-order central moment

$$v_n = E\{(h - \bar{h})^n\} = \int_{-\infty}^{+\infty} (h - \bar{h})^n p(h)dh,$$

where $E\{...\}$ is an average operator, which takes the ensemble average of the expression insides the braces and $\bar{h} = m_1$ is the average surface height or the 1st-order moment. Central moments form one set of values by which the properties of a probability distribution can be usefully characterized. Central moments are used in preference to ordinary moments because then the values higher order quantities relate only to the spread and shape of the distribution, rather than to its location. For a realistic rough surface, usually we take $\bar{h} = 0$ and then $m_n = v_n$. The assumption that $\bar{h} = 0$ holds at any position on the surface ensures that the random field $h(\vec{r})$ is homogeneous.

The 2nd-order moment of the variable h is one of the most important physical parameters people use to describe the surface roughness and it is denoted as

$$w^2 = m_2 = \int_{-\infty}^{+\infty} h^2 p(h)dh,$$

where w is called the root-mean-square (RMS) roughness or the interface width. Here it also equals to the the standard deviation since $\bar{h} = 0$. The w describes the fluctuations of surface heights around an average height. Fig. 1.1 shows three different rough surfaces with different RMS roughness values. The larger the value of w, the rougher the surface, under the condition that the other roughness parameters are the same.

The 1st-order moment $\bar{h}$ and the 2nd-order moment w^2 are sufficient to characterize surface roughness for most purposes. However, high order moments can give more information about the surface height distribution and sometimes one needs higher-order moments to differentiate surfaces in more detail. For example, Fig. 1.2 shows two different surfaces with the same RMS roughness value w. One has Gaussian height distribution, in which the surface height fluctuates symmetrically around the average surface height. The other one has an exponential distribution which possesses many protrusions on the surface. To describe this difference, higher-order moments would be needed.

The most important higher order moments are the 3rd and 4th-order moments. The 3rd-order moment defines the skewness of surface height:

$$\gamma_3 = \frac{m_3}{w^3} = \frac{1}{w^3} \int_{-\infty}^{+\infty} h^3 p(h)dh.$$

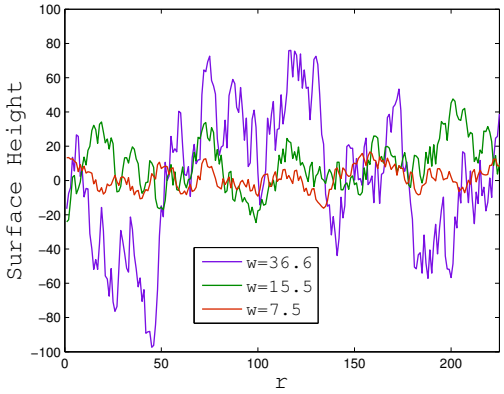


Figure 1.1: Three different rough surfaces with different rms roughness values.

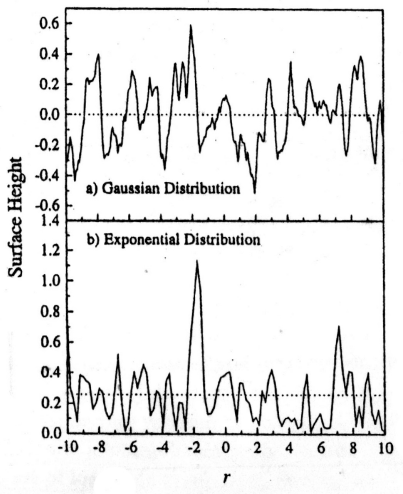


Figure 1.2: Two different surfaces with the same RMS roughness value w but different height distributions $h(r)$. The dotted lines in the figure indicate the average surfaces height positions.

The skewness is dimensionless in contrast to the RMS roughness which is in units of length. Skewness is a measure of the symmetry of a distribution about a mean surface level. The sign of the skewness, positive or negative, tells that the farther points are proportionately above or below the mean surface level, respectively. In other words, a positive skewness represents a distribution with an asymmetric tail extending out toward more positive height with respect to the mean surface level, while a negative skewness represents a distribution whose tail extends out toward more negative height with respect to the mean surface level, as shown in Fig. 1.3. For example, a distribution of bumps on surface would have a positive skewness, while a distribution of holes will have a negative skewness. For a symmetric distribution like Gaussian, the skewness is zero.

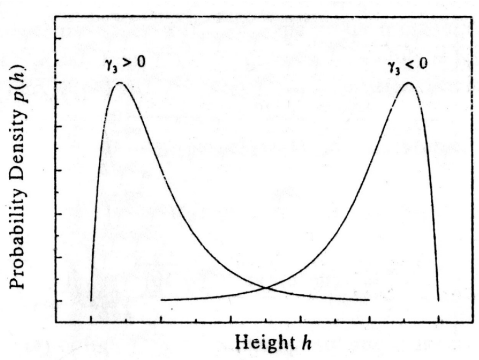


Figure 1.3: The surface height distribution function with positive ($\gamma_3 > 0$) skewness and negative ($\gamma_3 < 0$) skewness.

The 4th-order moment defines the kurtosis of surface height:

$$\gamma_4 = \frac{m_4}{w^4} = \frac{1}{w^4} \int_{-\infty}^{+\infty} h^4 p(h) dh.$$

Kurtosis is also a dimensionless quantity. It is a measure of the sharpness of the height distribution function. If most of the surfaces features are concentrated close to the mean surface level, the kurtosis will be less than that of the height distribution containing a larger portion of the surface features lying farther from the mean surface level. In addition, kurtosis describes the randomness of the surface profile relative to that of a perfectly random surface (Gaussian distribution) that has kurtosis 3.0. For $\gamma_4 < 3$, the distribution is platykurtic (mild peak) and for $\gamma_4 > 3$ the distribution is leptokurtic (sharp peak). Fig. 1.4 shows typical shapes of these two different distributions and a Gaussian distribution.

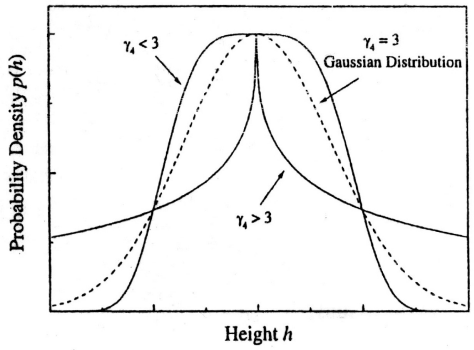


Figure 1.4: The surface height distribution functions with different kurtosis values γ_4.

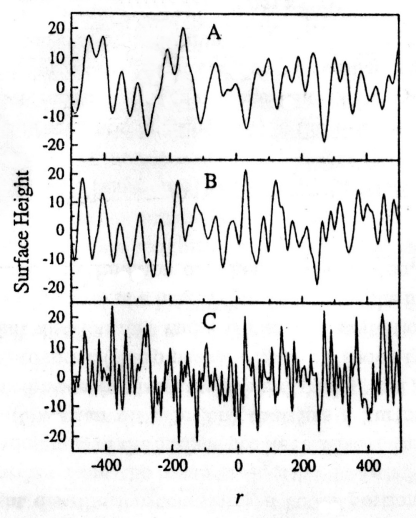


Figure 1.5: Surface profiles with the same rms w but different lateral correlation lengths ξ.

1.1.3 Second-Order Statistics

The first-order statistics or the height distribution function only describes the statistical properties of random variables of a random field at individual positions. It cannot reflect the connection between random variables at different positions. Different rough surfaces can have the same height distribution $p(h)$ and RMS roughness w (interface width). For example, all the three sampled surface profiles shown in Fig. 1.5 have the same Gaussian height distribution and the same interface width but they look totally different because the changes in heights occur in different length scales along the surfaces. In other words, the height fluctuation frequencies are different. Intuitively, one can say that profile C is much rougher than profile A. In order to differentiate the spatial difference, one needs to know the connection of a random field $h(r)$ at two different positions r_1 and r_2. To do that, the joint distribution probability density function $p_j(h_1, h_2; r_1, r_2)$ of $[h(r_1), h(r_2)]$ is introduced and it satisfies:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_j(h_1, h_2; r_1, r_2)dh_1dh_2 = 1,$$

$$\int_{-\infty}^{+\infty} p_j(h_1, h_2; r_1, r_2)dh_1 = p(h_2),$$

$$\int_{-\infty}^{+\infty} p_j(h_1, h_2; r_1, r_2)dh_2 = p(h_1),$$

where $p(h_1)$ and $p(h_2)$ are called the marginal distributions of $p_j(h_1, h_2; r_1, r_2)$ and for an homogeneous random field we have $p(h_1) = p(h_2) = p(h)$. In general, $p_j(h_1, h_2; r_1, r_2)$ is related not only to the height distribution, but also to the correlation of heights between two separated positions.

One special case for the joint distribution function is that h_1 and h_2 are independent of each other. In this case, $p_j(h_1, h_2; r_1, r_2)$ can be simply expressed as

$$p_j(h_1, h_2; r_1, r_2) = p(h_1)p(h_2).$$

The most important statistical characteristic of a joint distribution $p_j(h_1, h_2; r_1, r_2)$ is the auto-covariance function, $G(r_1, r_2)$, defined as

$$G(r_1, r_2) = E\{h(r_1)h(r_2)\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h_1h_2p_j(h_1, h_2; r_1, r_2)dh_1dh_2,$$

or the auto-correlation function:

$$R(r_1, r_2) = \frac{G(r_1, r_2)}{w^2}.$$

Both $G(r_1, r_2)$ and $R(r_1, r_2)$ reflect the extension of a correlation of heights at two positions and depend on positions r_1 and r_2. Sometimes $R(r_1, r_2)$ is called the auto-correlation coefficient and is a dimensionless function. For the homogeneous, isotropic rough surfaces, $G(r_1, r_2)$ and $R(r_1, r_2)$ depend only on the distance between two positions r_1 and r_2:

$$G(r_1, r_2) = G(|r_1 - r_2|) = G(\rho),$$

and

$$R(r_1, r_2) = R(\rho),$$

where $\rho = |r_1 - r_2|$. The quantity ρ is the translation. It is obvious that

$$G(0) = G(r, r) = E\{h(r)h(r)\} = w^2.$$

That means that the value of an auto-covariance function $G(\rho)$ at $\rho = 0$ is equal to the variance of a surface height. Since $G(\rho)$ and $R(\rho)$ have the same property, the difference between them is only a factor w^2. The auto-correlation function $R(\rho)$ has the following properties:

1. $R(0) = 1$.
2. $R(-\rho) = R(\rho)$, i.e., $R(\rho)$ is an even function.
3. $|R(\rho)| \leq R(0)$, i.e., $R(0)$ is a maximum of $R(\rho)$.
4. Given that $h(r)$ and $h(r+\rho)$ are independent of each other for $\rho \to \infty$, and also $E\{h(r)\} = 0$, then $\lim_{\rho \to +\infty} R(\rho) = 0$.

For a truly random rough surface, $R(\rho)$ usually decays to 0 with the increase of ρ. The shape of this decay depends on the type of random surfaces and the rate of decay depends on the distance over which two points become uncorrelated. The correlation length (or the lateral correlation length) ξ of an auto-correlation function is usually defined as the value of the lag length at which the auto-correlation function drops to $1/e$ of its value at zero lag:

$$R(\xi) = 1/e.$$

The correlation length ξ defines a representative lateral dimension of a rough surface. If the distance between two surface points is within ξ, the heights at these two points can be considered correlated. However, if the separation of two surface points is much larger than ξ, then we say that the heights at these two points are independent of one another.

Fig. 1.6 shows the auto-correlation functions for the three different surfaces in Fig. 1.5. We can see that surface A has the largest correlation length while surface C has the smallest correlation length. Therefore surface C looks rougher.

For a random rough surface, the information of the correlation functional form is very important because it represents a good description of the surface morphology. Based on the discussion of roughness parameters so far, one tends to think that for surface roughness characterization, two parameters, the RMS roughness w and the lateral correlation length ξ, would be sufficient. However, real surfaces are not so simple. For example, the two sampling surface profiles shown in Fig. 1.7 both have a Gaussian height distribution, the same RMS roughness w and the lateral correlation length ξ, but the forms of $R(\rho)$ are different: profile A has a Gaussian correlation and profile B has an exponential correlation. We can tell by visual inception that profile B is

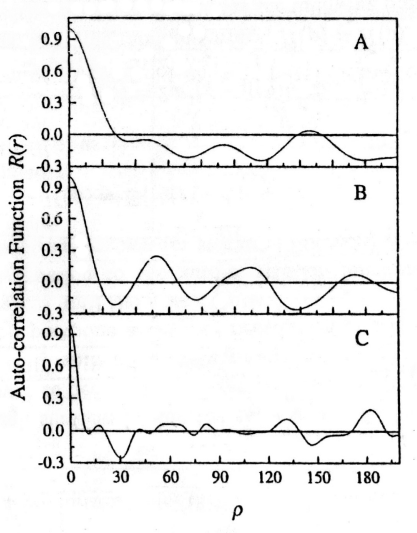


Figure 1.6: The auto-correlation functions of the surface profiles shown in Fig. 1.5.

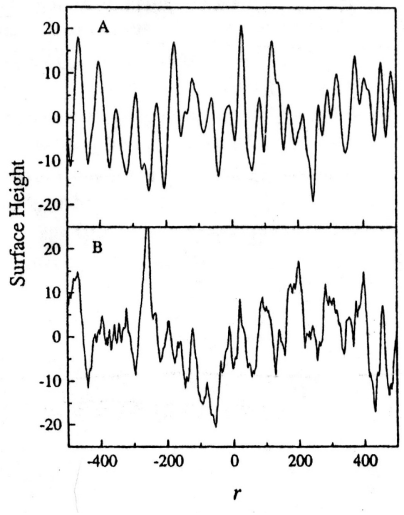


Figure 1.7: Examples of two surfaces having the same w and ξ but which look very different.

rougher than profile A. Therefore, the two parameters w and ξ alone are not enough to give a full description of surface roughness. One needs at least one additional parameter to describe this difference. This parameter is the roughness exponent (or Hurst exponent) a.

An equivalent function to the auto-correlation function $R(\rho)$ is the height-height correlation function (or structure function) $H(\rho)$, defined as

$$H(\rho) = E\{[h(r) - h(r+\rho)]^2\}.$$

It can be related to $R(\rho)$ as

$$H(\rho) = 2w^2[1 - R(\rho)].$$

For a homogeneous, isotropic random rough surface, $H(-\rho)$ has the following properties:

1. $H(-\rho) = H(\rho)$, i.e., it is also an even function with respect to ρ.
2. $H(0) = 0$.
3. $H(\rho \to +\infty) = 2w^2$.

1.1.4 Self-Affine Surfaces

The fractal is a very useful concept for describing rough surfaces. The idea of fractal geometry is closely associated with the property of invariance under a change of scale. The simplest fractal object is a self-similar object, which is invariant under similarity transformations. A self-similar object looks the same (or statistically the same) when the space it occupies is stretched uniformly by a factor ε. This property is called scale invariance. Another very useful concept for describing surface morphology is the self-affine fractal. A self-affine object looks the same after an affine transformation: if a small piece of the object is stretched with different ratios in different directions, then the enlarged object recovers (or statistically recovers) the original object.

A rough surface can be described by a single-valued self-affine function, which has the property:

$$h(x_1, ..., x_n) = \varepsilon_1^{-\alpha_1} ... \varepsilon_n^{-\alpha_n} h(\varepsilon_1 x_1, ..., \varepsilon_n x_n), \tag{1.1}$$

where h is the surface height and α_i is called the roughness exponent or Hurst exponent. Typically, there is only one characteristic roughness exponent and Eq. (1.1) has a simpler form, $h(x) = \varepsilon^{-\alpha} h(\varepsilon x)$. For example, for a single variable x, Eq. (1.1) demonstrates the fact that the function h is invariant under the following rescaling: shrink the variable x along the x-axis by a factor of $1/\ \varepsilon$; rescale the value of the function by a different factor $\varepsilon^{-\alpha}$.

Therefore, a self-affine surface is a class of fractal objects that can be described by a roughness exponent, which is related to the fractal dimension of the surface. The height-height correlation function of such a surface has the form

$$H(r) = 2w^2 f\left(\frac{r}{\xi}\right),$$

where $f(x)$ is a scaling function, having the following properties:

$$f(x) = \begin{cases} x^{2a}, & \text{for } x \ll 1, \\ 1, & \text{for } x \gg 1. \end{cases}$$

The parameter α is called the roughness exponent ($0 \leq \alpha \leq 1$) and describes how wiggly the surface is. The asymptotic behavior of the height-height correlation function can be written as

$$H(r) = \begin{cases} (mr)^{2\alpha}, & \text{for } r \ll \xi, \\ 2w^2, & \text{for } r \gg \xi, \end{cases}$$

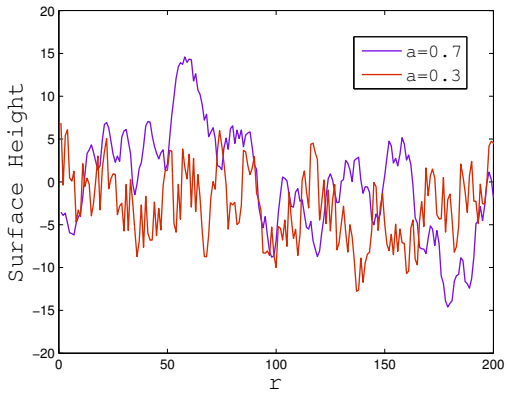


Figure 1.8: Rough surface profiles for different values of α. These two surfaces have the same values of w and ξ, but different values of the roughness exponent α.

where $m = w^{1/\alpha}/\xi$ is the local slope and it characterizes the short-range properties of the surface. Surfaces with different α values are shown in Fig. 1.8. One can see that a larger value of $\alpha(= 0.7)$ corresponds to a locally smooth surface structure while a smaller value of $\alpha(= 0.3)$ corresponds to a more jagged local surface morphology. The roughness exponent is directly related to the local surface fractal dimension D_s by $d + 1 - D_s$, where $d + 1$ is the tool boxdimension of the embedded space.

The three parameters w, α and ξ are independent from each other and completely characterize a self-affine surface.

1.2 Characterization of Complex Networks

1.2.1 Definitions and Denotations

Graph theory is the natural framework for the exact mathematical treatment of complex networks and, formally, a complex network can be represented as a graph. There are four main types of complex networks which include *weighted digraphs* (*directed graphs*), *unweighted digraphs*, *weighted graphs* and *unweighted graphs*. The operation of *symmetry* can be used to transform a digraph into a graph and the operation of *thresholding* can be applied to transform a weighted graph into its unweighted counterpart. Strarting from the concept of weighted digraph, all the other three types can be derived.

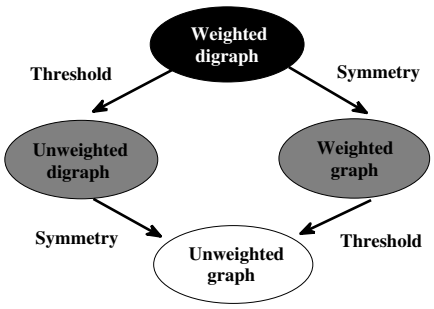


Figure 1.9: The four main types of complex networks and their transformations.

A *weighted directed graph* $G^W = (\mathcal{N}, \mathcal{L}, \mathcal{W})$ consists of a set $\mathcal{N} = \{n_1, n_2, ..., n_N\}$ of *nodes* (or *vertices*, or *points*), where $\mathcal{N} \neq 0$, a set of *links* $\mathcal{L} = \{l_1, l_2, ..., l_K\}$ (or *edges*, or *lines*), which is a set of ordered pairs of elements of $\mathcal{N}$, and a set of *weights* (or *values*) $\mathcal{W} = \{w_1, w_2, ..., w_K\}$ that are real numbers attached to the links. The number of elements in $\mathcal{N}$ and $\mathcal{L}$ are denoted by N and K respectively. A node is usually referred to by its order i in the set $\mathcal{N}$. Each of the edges is defined by a couple of nodes (i, j) or l_{ij}, that represents a connection going from vertex i to vertex j.

A weighted digraph can be completely represented in terms of its *weights matrix* $\mathcal{W}$, a $N \times N$ matrix whose entry w_{ij} expresses the weight of the connection from vertex i to vertex j and $w_{ij} = 0$ if the nodes i and j are not connected ($w_{ii} = 0\ \forall i$, unless explicitly mentioned).

The operation of thresholding can be applied to a weighted digraph to produce an unweighted counterpart. This operation, henceforth represented as $\delta T(\mathcal{W})$, is applied to each element of the matrix $\mathcal{W}$, yielding the matrix $\mathcal{A} = \delta T(\mathcal{W})$. The elements of the matrix $\mathcal{A}$ are computed comparing the corresponding elements of $\mathcal{W}$ with a specified threshold T: in case $|w_{ij}| > T$ we have $a_{ij} = 1$, otherwise $a_{ij} = 0$. The resulting matrix $\mathcal{A}$ can be understood as the *adjacency matrix* of the unweighted digraph obtained as a result of the thresholding operation. Any weighted digraph can be transformed into a graph by using the symmetry operation $\sigma(\mathcal{W}) = \mathcal{W} + \mathcal{W}^T$, where $\mathcal{W}^T$ is the transpose of $\mathcal{W}$.

For undirected graphs, the diagonal of the adjacency matrix $\mathcal{A}$ also contains zeros and $\mathcal{A}$ is a symmetric matrix. Two vertices i and j are said to be *adjacent* or *neighbours* if $a_{ij} \neq 0$. The link (i, j) is said to be *incident* in nodes i and j, or to join the two nodes; the two nodes i and j are referred to as the *end-nodes* of link (i, j). The concept of adjacency can also be used in digraphs by considering *predecessors* and *successors* as adjacent vertices: if $a_{ij} \neq 0$ then i is a predecessor of j and j is a successor of i. The neighbourhood of a vertex i, henceforth represented as $v(i)$, corresponds to the set of vertices adjacent to i.

Examples of an undirected and of a directed graph, both with $N = 7$ but with different number of edges, are shown in Fig. 1.10a and 1.10b respectively.

(a) (b)

Figure 1.10: Graphical representation of (a) an undirected graph with $N = 7$ nodes and $K = 11$ edges and (b) a directed graph with $N = 7$ and $K = 8$.

Multigraph is called a graph that contains either *loops* – links from a node to itself – either *multiple edges* – couples of nodes connected by more than one link.

For a graph G of size N, the number of edges K is at least 0 and at most $N(N-1)/2$ (when all the nodes are pairwise adjacent). G is said to be *sparse* if $K \ll N^2$ and *dense* if $K = \mathcal{O}(N^2)$. A graph $G_{N,K}$ is said a *complete* N-graph if $K = \binom{N}{2} = N(N-1)/2$ and is denoted by K_N. For example K_3 is a complete graph and is called a triangle.

A *subgraph* $G' = (\mathcal{N}', \mathcal{L}')$ of $G = (\mathcal{N}, \mathcal{L})$ is a graph such that $\mathcal{N}' \subseteq \mathcal{N}$ and $\mathcal{L}' \subseteq \mathcal{L}$. If G' contains all links of G that joint two nodes in $\mathcal{N}'$, then G' is said to be a subgraph induced by $\mathcal{N}'$ and is denoted as $G' = G[\mathcal{N}']$. The subgraph of the neighbours of a given node i is denoted as G_i. G_i is also defined as the subgraph induced by $\mathcal{N}_i$, the set of nodes adjacent to i, i.e. $G_i = G[\mathcal{N}_i]$.

(a) (b)

Figure 1.11: (a) Graph G, (b) subgraph G' of G, $G' \equiv G_4$

A central concept in graph theory is that of reachability of two different nodes of a graph. In fact, two nodes that are not adjacent may nevertheless be reachable from one to the other. A *walk* from node i to node j is an alternating sequence of nodes and edges (a sequence of adjacent nodes) that begins with i and ends with j. The length of the walk is defined as the number of edges in the sequence (in Fig. 1.11a the sequence $0-(0,1)-1-(1,2)-2$ is a walk from node 0 to node 2 of length 2). A *trail* is a walk in which no edge is repeated. A *path* is walk in which no node is visited more than once. The walk of minimal length between two nodes is known as *shortest path* or *geodesic*. A *cycle* is a closed walk, of at least three nodes, in which no edge is repeated. A cycle of length k is usually said a k-cycle and denoted as C_k. For example C_3 is a triangle ($C_3 = K_3$), C_4 is called a quadrilater, C_5 a pentagon and so on.

A graph is said to be *connected* if, for every pair of distinct nodes i and j, there is a path from i to j, otherwise it is said *unconnected* or *disconnected*.

1.2.2 Node Degree and other Topological Properties

Node degree k_i

The simplest local characteristic of a vertex is its *degree*. In undirected graphs, the *degree* (or *connectivity*) k_i of a node i is the number of edges incident with the node, i.e. the cardinality of the set $v(i)$, and is defined in terms of the adjacency matrix $\mathcal{A}$ as:

$$k_i = \sum_{j \in \mathcal{N}} a_{ij} = \sum_{j \in \mathcal{N}} a_{ji}. \tag{1.2}$$

If the graph is directed, the degree of the node has two components: the *out-degree*, k_i^{out}, equal to the number of outgoing edges (i.e. the cardinality of the set of successors) and the *in-degree*, k_i^{in}, corresponding to the number of incoming edges (i.e. the cardinality of the set of predecessors):

$$k_i^{out} = \sum_j a_{ij} \tag{1.3}$$

$$k_i^{in} = \sum_j a_{ji} \tag{1.4}$$

The total degree is then defined as $k_i = k_i^{out} + k_i^{in}$.

In a weighted graph, the natural generalization of the degree k_i of a node i is the *node strength* (or *node weight*, or *node weighted connectivity*) s_i, defined as:

$$s_i = \sum_{j \in \mathcal{N}} w_{ij}. \tag{1.5}$$

The strengths integrate the information on the number (degree) and the weights of links incident in a node. When the weights are independent on the topology, the strength of the vertices of degree k is $s(k) \simeq \langle w \rangle k$, where $\langle w \rangle$ is the average weight.

Degree distribution $P(k)$

The most basic topological characterization of a graph G can be obtained in terms of the *degree distribution* $P(k)$ (also denoted as P_k or p_k), defined as the probability that a node chosen uniformly at random has degree k or, equivalently, as the fraction of nodes in the graph having degree k. A plot of P_k for any given network can be formed by making an histogram of the degrees of vertices.

In the case of directed networks one needs to consider two distributions: $P(k^{in})$ and $P(k^{out})$. It can also be considered the joint in-degree and out-degree distribution $P(k^{in}, k^{out})$,which gives the probability of finding a vertex with in-degree k^{in} and out-degree k^{out}.

Information on how the degree is distributed among the nodes of an undirected network can be obtained either by a plot of $P(K)$ or by the calculation of the moments of the distribution. The n-moment of $P(k)$ is defined as:

$$\langle k^n \rangle = \sum_k k^n P(k) \tag{1.6}$$

The first moment, $\langle k \rangle$, is the mean degree of G. The second moment, $\langle k^2 \rangle$, measures the fluctuations of the connectivity distribution.

The degree distribution completely determines the statistical properties of uncorrelated networks. However, a large number of real networks are correlated in the sense that the probability that a node of degree k is connected to another node of degree k' depends on k. In these cases,

it is necessary to introduce the *conditional* probability $P(k'|k)$, being defined as the probability that a link from a node of degree k points to a node of degree k'. $P(k'|k)$ satisfies the normalization $\sum_{k'} P(k'|k) = 1$ and the degree detailed balance condition $kP(k'|k)P(k) = k'P(k|k')P(k')$.

Topological properties

The original formulation, proposed by Watts and Strogatz [1998] is considered here. Let a generic graph G with N nodes and K edges. G is assumed to be:

1. Unweighted. The edges are not assigned any a priori weight and therefore are all equal.
2. Simple. This means that either a couple of nodes is connected by a direct edge or it is not: multiple edges between the same couple of nodes are not allowed.
3. Sparse. This property means that $K \ll N(N-1)/2$, i.e. only a few of the total possible number of edges $N(N-1)/2$ exist.
4. Connected. K must be small enough to satisfy property 3, but on the other side it must be large enough to assure that there exist at least one path connecting any couple of nodes. For a random graph this property is satisfied if $K \gg Nln(N)$.

All the information necessary to describe such a graph are therefore contained in a single matrix $\{a_{ij}\}$, the adjacency matrix. In order to quantify the structural properties of G, two quantities will be evaluated: the characteristic path length L and the clustering coefficient C.

The characteristic path length L: Shortest paths or geodesics play an important role in the transport and communication within a network. Suppose one needs to send a data packet from one computer to another through the Internet: the geodesic provides an optimal path way, since one would achieve a fast transfer and save system resources. For such a reason, shortest paths have played an important role in the characterization of the internal structure of a graph.

It is useful to represent all the shortest path lengths of a graph G as a matrix $\mathcal{D}$ in which the entry d_{ij} is the minimum number of edges traversed to get from a vertex i to another vertex j. By definition $d_{ij} \geq 1$, where $d_{ij} = 1$ if there exists a direct edge between i and j. The maximum value of d_{ij} is called the diameter, $diam(G)$.

In general the geodesic between two vertices may not be unique: there may be two or more shortest paths (sharing or not sharing similar vertices) with the same length. The matrix D can be extracted from the adjacency matrix, $\{a_{ij}\}$: there is a huge number of different algorithms in the literature from the standard breadth-first search algorithm, to more sophisticated algorithms as the Floyd-Warshall or the Dijkstra algorithm.

The characteristic path length or average shortest path length L of graph G is defined as the average of the geodesic lengths over all couples of nodes:

$$L = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N},\ i \neq j} d_{ij} \tag{1.7}$$

and measures the typical separation between two vertices in the graph (a global property).

The assumption that G is connected is crucial in the calculation of L. It implies that there exists at least one path connecting any couple of vertices with a finite number of steps, d_{ij} finite $\forall i \neq j$, and therefore it assures that also L is a finite number.

A problem with this definition is that L diverges if the assumption of connectedness is removed, i.e. if there are disconnected components in the graph. To avoid this problem, only couples of nodes belonging to the largest connected component are included in the sum. This

skips the divergence, but introduces a distortion for networks with many unconnected pairs of vertices, which will show a small value of average distance, expected only for networks with a high number of connections.

The clustering coefficient C: One way to characterize the presence of loops of order three is through the clustering coefficient. Two different clustering coefficients are frequently used.

The first, also known as transitivity, is a typical property of social networks: if vertex A is connected to vertex B and vertex B to vertex C, then there is a heightened probability that vertex A will also be connected to vertex C. In other words, the friend of your friend is likely also to be your friend. In terms of network topology, transitivity means the presence of a high number of triangles[1] in the network and it can be quantified as the fraction of connected triples[2] of nodes which also form triangles of interactions:

$$T = \frac{3 \times \#\text{ of triangles in } G}{\#\text{ of connected triples of vertices in } G} = \frac{3N_\Delta}{N_3}, \tag{1.8}$$

where $N_\Delta = \sum_{k>j>i} a_{ij}a_{ik}a_{jk}$ and $N_3 = \sum_{k>j>i} (a_{ij}a_{ik} + a_{ji}a_{jk} + a_{ki}a_{kj})$. The factor 3 in the numerator accounts for the fact that each triangle contributes to three connected triples and ensures that T lies in the range $0 \leq T \leq 1$, with $T = 1$ for a complete graph, K_N.

An alternative way is to use the clustering coefficient C, a measure introduced by Watts and Strogatz. This quantity gives the average cliquishness of a typical neighbourhood of G (a local property), and is defined as follows. First of all a quantity c_i, the local clustering coefficient of node i, is defined as:

$$c_i = \frac{N_\Delta(i)}{N_3(i)}, \tag{1.9}$$

where $N_\Delta(i)$ is the number of triangles involving vertex i and $N_3(i)$ is the number of connected triples having i as the central vertex:

$$N_\Delta(i) = \sum_{k>j} a_{ij}a_{ik}a_{jk}, \tag{1.10}$$

$$N_3(i) = \sum_{k>j} a_{ij}a_{ik}. \tag{1.11}$$

If k_i is the number of neighbours of vertex i, then $N_3(i) = k_i(k_i-1)/2$ stands for the maximum possible number of edges in $G(i)$—the subgraph of neighbours of i. $N_\Delta(i)$ counts the actual number of edges between neighbours of i, which can be represented as e_i. Now Eq.(1.9) can be rewritten as:

$$c_i = \frac{2e_i}{k_i(k_i-1)} \tag{1.12}$$

The quantity c_i denotes the fraction of these allowable edges that actually exist and expresses how likely $a_{jk} = 1$ for two neighbours j and k of node i.

For vertices with degree 0 or 1, for which both numerator and denominator are zero, we put $c_i = 0$.

The clustering coefficient C of the whole graph G is then defined as the average of c_i over all the vertices i of G:

[1] A triangle is a set of three vertices with edges between each pair of vertices.

[2] A connected triple is a set of three vertices where each vertex can be reached from each other (directly or indirectly), i.e. two vertices must be adjacent to another vertex—the central vertex.

$$C = \langle c \rangle = \frac{1}{N} \sum_{i \in \mathcal{N}} c_i, \tag{1.13}$$

By definition, $0 \leq c_i \leq 1$ and $0 \leq C \leq 1$.

This definition reverses the order of operations of taking the ratio of triangles to triples and of averaging over vertices—one here calculates the mean of the ratio, rather than the ratio of the means. It tends to weight the contributions of low-degree vertices more heavily, because such vertices have a small denominator in Eq.(1.9) and hence can give quite different results from Eq.(1.8). The first definition (1.8) is easier to calculate analytically while definition (1.9) is easily calculated on a computer and has found wide use in numerical studies and data analysis.

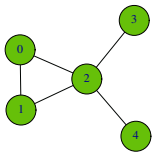


Figure 1.12: Illustration of the definition of the clustering coefficient C, Eq.(1.8): This network has one triangle and eight connected triples, and therefore has a clustering coefficient of $3 \times \frac{1}{8} = \frac{3}{8}$. The individual vertices have local clustering coefficients, Eq.(1.9), of $1, 1, \frac{1}{6}, 0$ and 0, for a mean value, Eq.(1.13), of $C = \frac{13}{30}$.

The efficiency E: The approach of Watts and Strogatz can be used when the only information retained of a real network is about the existence or the absence of a link, and nothing is known about the physical length of the link, the weight associated to the link, and multiple edges between the same couple of nodes are not allowed. Moreover the assumption of connectedness is necessary because otherwise the quantity L would diverge.

Latora and Manchiori proposed a more general set-up to investigate real networks. They considered a generic graph G, weighted and possibly even non-connected and non-sparse graph. This weighted graph needs two matrices to be described:

- the adjacency matrix $\{a_{ij}\}$, defined the same as for the unweighted graph,
- the matrix of physical distances $\{l_{ij}\}$, a matrix of the weights associated to each link. The number l_{ij} can be the space distance between the two vertices or the strength of their possible interaction. l_{ij} is supposed to be known even if in the graph there is no edge between i and j.

l_{ij} can be the geographical distance between stations in transportation systems or the time taken to exchange a packet of information between routers in the Internet. In the particular case of an unweighted graph $l_{ij} = 1 \ \forall \ i \neq j$. The shortest path length d_{ij} between two points i and j is the smallest sum of the physical distances throughout all the possible paths in the graph from i to j. The matrix d_{ij} is therefore calculated by using the information contained both in matrix $\{a_{ij}\}$ and in matrix $\{l_{ij}\}$. We have $d_{ij} \geq l_{ij} \ \forall \ i, j$, the equality being valid when there is an edge between i and j.

Let us now suppose that every vertex sends information along the network, through its edges. The efficiency ϵ_{ij} in the communication between vertices i and j can then be defined to be inversely proportional to the shortest distance: $\epsilon_{ij} = 1/d_{ij} \ \forall i, j$. When there is no path in the graph between i and j, $d_{ij} = +\infty$ and, consistently, $\epsilon_{ij} = 0$. The average efficiency of G can be defined as:

$$E = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N},\ i \neq j} \epsilon_{ij} = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N},\ i \neq j} \frac{1}{d_{ij}} \ . \tag{1.14}$$

Formula (1.14) gives a value of E that can vary in the range $[0, \infty]$. E can be normalized in the interval $[0, 1]$ by considering the ideal case G^{ideal} in which the graph G has all the $N(N-1)/2$ possible edges. In such a case the information is propagated in the most efficient way since $d_{ij} = l_{ij}\ \forall i, j$, and E assumes its maximum value $E(G^{ideal}) = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N},\ i \neq j} \frac{1}{l_{ij}}$. The efficiency $E(G)$ is divided by $E(G^{ideal})$ and therefore $0 \leq E(G) \leq 1$. Though the maximum value $E = 1$ is typically reached only in complete graphs.

This formalism, proposed by Latora and Manchiori, has the advantage that a single measure, the efficiency E, can constitute an alternative approach for the measures: characteristic path length L and clustering coefficient C.

In fact, on one side, the quantity defined in Eq.(1.14) can be evaluated as it is for the whole graph G to characterize the global efficiency of G:

$$E_{glob} = \frac{E(G)}{E(G^{ideal})} \tag{1.15}$$

E_{glob} plays a role similar to the inverse of the characteristic path length L as it is the harmonic mean of geodesic lengths, d_{ij}. This measurement quantifies the efficiency of the network in sending information between vertices, assuming that the efficiency for sending information between two nodes i and j is proportional to the reciprocal of their distance. In the particular case o a disconnected graph the difference between the two quantities is evident because $L = +\infty$ while E_{glob} is a finite number.

On the other side the same measure E, since it is defined also for a disconnected graph, can characterize the local properties of G by evaluating for each vertex i the efficiency of G_i:

$$E_{loc} = \frac{1}{N} \sum_{i \in G} \frac{E(G_i)}{E(G_i^{ideal})} \tag{1.16}$$

Here, for each vertex i, the normalization factor $E(G_i^{ideal})$ is the efficiency of the ideal case in which the graph G_i has all the $k_i(k_i - 1)/2$ possible edges. E_{loc} is an average of the local efficiency and plays a role similar to the clustering coefficient C. Since $i \notin G_i$, the local efficiency E_{loc} tells how much the system is fault tolerant, thus how efficient is the communication between the first neighbours of i when i is removed.

1.2.3 Topology of real networks

During the last decade, the grown availability of large databases, the optimized rating of computing facilities, as well as the development of powerful and reliable data analysis tools have constituted a better and better machinery to explore the topological properties of networked systems from the real world such as communication, social and biological systems. This study has revealed that, despite the inherent differences, most of the real networks are characterized by the same topological properties, as for instance relatively small characteristic path lengths, high clustering coefficients, fat-tailed shapes in the degree distributions and degree correlations. All these features make real networks radically different from regular lattices and random graphs, the standard models studied in mathematical graph theory. This has led to a large attention towards the understanding of the evolution mechanisms that have shaped the topology of a network, and to the design of new models retaining the most significant properties empirically observed.

The small-world property

The study of several dynamical processes over real networks has pointed out the existence of shortcuts, i.e. bridging links that connect different areas of the networks, thus speeding up the communication among otherwise distant nodes. This feature is known as the *small-world property* and is mathematically characterized by an average shortest path length L, defined as in Eq.(1.7), that depends at most logarithmically on the network size N. In regular lattices the characteristic path length L grows linearly with N while in random graphs L grows only logarithmically with N.

The small-world property has been observed in social networks, where everyone in the world can be reached through a short chain of social acquaintances. This concept originated from the famous experiment made by Milgram in 1967, who found that two US citizens chosen at random were connected by an average of six acquaintances (six-degrees of separation).

The small-world property in real networks is also associated with the presence of clustering, denoted by high values of the clustering coefficient, defined as in Eq.(1.13). Regular lattices are also high-clustered networks while random graphs are considered as poorly-clustered.

Scale-free degree distributions

The usual case in science until a few years ago was that of homogeneous networks. Homogeneity in the interaction structure means that almost all nodes are topologically equivalent, like in regular lattices or in random graphs. In these latter ones, for instance, each of the $N(N-1)/2$ possible links is present with equal probability, and thus the degree distribution is binomial or Poisson in the limit of large graph size. However real-world networks are mostly found to be very unlike the random graph in their degree distributions. Scientists found that most of the real networks display power law shaped degree distribution $P(k) \sim Ak^{-\gamma}$, with exponents varying in the range $2 < \gamma < 3$. Far from having a Poisson distribution, the degrees of the vertices in most networks are highly right-skewed, meaning that their distribution has a long right tail of values that are far above the mean.

Measuring this tail is somewhat tricky. Although in theory one just has to construct an histogram of the degrees, in practice one rarely has enough measurements to get good statistics in the tail, and direct histograms are thus usually rather noisy. There are two accepted ways to get around this problem. One is to construct an histogram in which the bin sizes increase exponentially with degree. For example, the first few bins might cover degree ranges 1, 2-3, 4-7, 8-15, and so on. The number of samples in each bin is then divided by the width of the bin to normalize the measurement. This method of constructing an histogram is often used when the histogram is to be plotted with a logarithmic degree scale, so that the widths of the bins will appear even. Because the bins get wider as we get out into the tail, the problems with statistics are reduced, although they are still present to some extent as long as $P(k)$ falls off faster than k^{-1}, which it must if the distribution is to be integrable.

An alternative way of presenting degree data is to make a plot of the cumulative distribution function

$$P_k = \sum_{k'=k}^{\infty} p_{k'}, \tag{1.17}$$

which is the probability that the degree is greater than or equal to k. Such a plot has the advantage that all the original data are represented. When we make a conventional histogram by binning, any differences between the values of data points that fall in the same bin are lost. The cumulative distribution function does not suffer from this problem. The cumulative distribution also reduces the noise in the tail: the statistical fluctuations generally present in the tails of the distributions are smoothed.

1.2.4 Network models

Random graphs

The random graph developed by Solomonoff and Rapoport and independently by ERDdS and R&WI [1959] can be considered the most basic model of complex networks. The term random graph refers to the disordered nature of the arrangement of links between different nodes. Erdös and Rényi studied, by means of probabilistic methods, the properties of graphs as a function of the increasing number of random connections. In their 1959 paper, they introduced a model to generate random graphs with N nodes and K edges which are called *Erdös and Rényi (ER) random graphs.*

There are two closely related variants of the Erdös and Rényi random graph model:

- the $G^{ER}_{N,K}$ model: starting with N disconnected nodes, the network is constructed by connecting couples of randomly selected nodes, prohibiting multiple and self connections, until the number of edges equals K. In this model, a graph is chosen uniformly at random from the collection of all graphs which have N nodes and K edges. For example, in the $G^{ER}_{3,2}$ model, each of the three possible graphs with three vertices and two edges are included with probability 1/3.

- the $G^{ER}_{N,p}$ model: This alternative model is constructed by connecting nodes randomly. Each edge is included in the graph with probability $0 < p < 1$ independent from every other edge. All graphs with N nodes and K links have equal probability of $p^K(1-p)^{\frac{N(N-1)}{2}-K}$. The parameter p in this model can be thought of as a weighting function; as p increases from 0 to 1, the model becomes more and more likely to include graphs with more edges and less and less likely to include graphs with fewer edges. In particular, the case $p = 0.5$ corresponds to the case where all $2^{\binom{N}{2}}$ graphs with N vertices are chosen with equal probability.

The two models coincide in the limit of large N. The limit $N \to \infty$ is taken at fixed $\langle k \rangle$, which corresponds to fixing $2K/N$ in the first model and $p(N-1)$ in the second one. Although the first model seems to be more pertinent to applications, analytical calculations are easier and usually are performed in the second model.

The limit with a fixed $\langle k \rangle$ as $N \to \infty$ corresponds to a sparse network. In a sparse net, the mean number of connections of a vertex is much less than the number of connections of a vertex in a fully connected graph. The important feature of a network is its giant connected component. This is a set of mutually reachable vertices containing a finite fraction of vertices of a large network. Without the giant connected component, a net is only a set of small separated clusters. It turns out that in the classical random graphs, the giant connected component exists if the mean number of connections of a vertex exceeds one. So, this characteristic point of a network—the point of "the birth of the giant connected component"—is just in the range of extremely low degrees, $k \sim 1 \ll N$.

The probability that a node i has $k = k_i$ edges is the binomial distribution $P(k_i = k) = C^k_{N-1} p^k (1-p)^{N-1-k}$, where p^k is the probability for the existence of k edges, $(1-p)^{N-1-k}$ is the probability for the absence of the remaining $N-1-k$ edges and $C^k_{N-1} = \binom{N-1}{k}$ is the number of different ways of selecting the end points of the k edges. Since all the nodes in a random graph are statistically equivalent, each of them has the same distribution and the probability that a node chosen uniformly at random has degree k has the same form as $P(k_i = k)$. For large N and fixed $\langle k \rangle$, the degree distribution is well approximated by a Poisson distribution:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \ . \tag{1.18}$$

For this reason, ER graphs are sometimes called *Poisson random graphs*. ER random graphs are, by definition, uncorrelated graphs, since the edges are connected to nodes regardless of their degree.

Small-world networks

Watts and Strogatz proposed a network model to describe various real networks between order and randomness. Both regular and random networks failed to capture two important characteristics of the real networks, the high clustering coefficient and the short average path length: lattices have high clustering but long characteristic path length; random graphs have short characteristic path length but no clustering at all.
The network model they proposed, the *small-world network*, is a method to construct graphs, denoted as $G^{WS}_{N,K}$, having both a short average path length and a high clustering coefficient. The model is based on a rewiring procedure of the edges implemented with probability p. To construct a small-world network, one starts with a regular lattice of N nodes in which each node is connected to m nearest neighbours in each direction, totalizing $2m$ connections, where $N \gg m \gg log(N) \gg 1$. The total number of edges is $K = mN$. Next, for every node, each link connected to a clockwise neighbour is rewired to a randomly chosen node with probability p and preserved with probability $1 - p$.
For various values of p we have the following cases:

- if $p \to 0$, the case of regular lattice, $L_{regular} \sim N/2m \gg 1$ and $C_{regular} \sim 3/4$,
- if $p \to 1$, the case where all edges are rewired randomly so that the model produces a random graph with the constraint that each node has a minimum connectivity $k_{min} = m$, $L_{random} \sim ln(N)/ln(m)$ and $C_{random} \sim m/n \ll 1$,
- if $0 < p < 1$ the procedure generates graphs where both short distances and a large number of loops are present. Then, L is almost as small as as L_{random} yet $C \gg C_{random}$.

In the efficiency-based formalism, such a definition corresponds to networks having a high value of global efficiency E_{glob} and a high value of local efficiency E_{loc} i.e. to networks extremely efficient in exchanging information both at a global and at a local scale.

Scale-free networks

Barabási and Albert showed, in their 1999 paper, that the degree distribution of many real systems is characterized by an uneven (inhomogeneous) degree distribution. More specifically, the degree distribution has been found to follow a power law for large k. These networks are called scale-free networks. A characteristic of this kind of networks is the existence of hubs, i.e. a few nodes linked to many other nodes, and a large number of poorly connected elements..

The Barabási-Albert (BA) network model is based on two basic rules:

1. *Growth*: network models discussed thus far assume that we start with a fixed number N of vertices that are then randomly connected or rewired, without modifying N. In contrast, most real-world networks describe open systems that *grow* by the continuous addition of new nodes. Starting from a small nucleus of nodes, the number of nodes increases throughout the lifetime of the network by the subsequent addition of new nodes. For example, the World Wide Web grows exponentially in time by the addition of new web pages, and the research literature constantly grows by the publication of new papers.
2. *Preferential attachment*: network models discussed so far assume that the probability that two nodes are connected (or their connection is rewired) is independent of the nodes degree,

i.e., new edges are placed randomly. Most real networks, however, exhibit preferential attachment, such that the likelihood of connecting to a node depends on the nodes degree. For example, a web page will more likely include hyperlinks to popular documents with already high degrees, because such highly connected documents are easy to find and thus well known, or a new manuscript is more likely to cite well-known and thus much-cited publications than less-cited and consequently less-known papers.

More precisely, an undirected graph $G_{N,K}^{BA}$ is constructed as follows: starting with m_0 isolated nodes, at each time step $t = 1, 2, 3, ..., N - m_0$ a new node j with $m \leq m_0$ links is added to the network. The probability that a link will connect j to an existing node i is linearly proportional to the actual degree of i:

$$\prod_{j \to i} = \frac{k_i}{\sum_u k_u}. \tag{1.19}$$

Because every new node has m links, the network at time t will have $N = m_0 + t$ nodes and $K = mt$ links, corresponding to an average degree $\langle k \rangle = 2m$ for large times.

1.3 Motivation and Objective of Thesis

The measurement and characterization of a rough surface is a long standing problem. Traditionally, surface metrology was seen as important owing to its strong association with other disciplines such as control of machine tools and manufacturing processes, quality of optical components, tribology and surface engineering in general. For example, the control of surface texture allows automobile engines to have reduced running-in times and to be more fuel efficient with reduced emissions. It allows orthopaedic implants to last longer through optimized surface topography. It also enables bearings in machines such as hard disk drives to run more efficiently and wear less. In addition, when optical components have smoother surfaces they scatter less light and have better optical qualities. The usefulness of surface metrology is not in doubt: it has been viewed as a means of obtaining a fingerprint of the surface to characterize and understand how the unique surface topography plays a dominant role in the functional performance of components in a system [Jiang et al., 2007].

A paradigmatic shift took place when fractal theory and scaling concepts were introduced based on the assumption that most surface in nature and laboratory exhibit the self-similar (or more exactly self-affine) symmetry [Peter, 1984]. However, with the advent of Scanning Probe Microscopies and the digital measurement of a large amount of surfaces on micro and nanoscale, it was realized that the fractal description of roughness is mostly limited to a specific range of scales [Malcai et al., 1998]. This limitation can be crucial when surface structures with interesting properties like periodicity appear at scales outside this range. Furthermore, conventional tools for periodicity detection and evaluation suffer from their own limitations. For example, Fourier transform due to the kernel of harmonic waves it implies is not only sensitive to the degree of periodicity but also to the shape of the repeated structure. The above limitations motivate us to seek alternative methods for the characterization of surface roughness.

The key idea here is to transform a surface to a network and thereafter to examine network's structure for different values of the surface's parameters. The study of network's statistical measures, such as the mean degree, the degree distribution, the clustering coefficient and the average shortest path length as well the comparison of the constructed networks with other existing graphs is the objective of this work.

Chapter 2

Methodology

2.1 Previous Litterature

Similar attempts have been made in time series analysis:

Zhang and Small [2006] first introduced a transformation from pseudoperiodic time series, such as the human electrocardiogram (ECG), human speech, laser output and annual sunspot numbers, to complex networks. A pseudoperiodic time series is divided into disjoint cycles according to the local minima or maxima, and each cycle is considered as a basic node of the network. Two nodes are deemed to be connected if the phase space distance or the linear correlation coefficient between the corresponding cycles is less than a predetermined threshold. A complementary approach was suggested by Xu et al. [2008] who studied the structural shape of the direct neighbourhood of the phase space trajectory by a motif classification.

Lacasa et al. [2008] have proposed an alternative algorithm, the so-called visibility algorithm, to characterize periodic, random and fractal time series. Their scheme is based on a similar philosophy with that of Zhang et al. [2008]: successive scalar time series points are mapped to nodes and two nodes are connected if there is a straight line that connects the series data, with the assumption that this "visibility line" does not intersect any intermediate data. The structure of the time series is conserved in the graph topology: periodic, random and fractal series convert into regular, random and scale-free networks respectively.

Subsequently, many researchers considered the visibility algorithm in their scientific work. Yang et al. [2009] investigated six important exchange rate series and found that some of them convert into scale-free networks while others into exponential networks. Ni et al. [2009] investigated the visibility graphs extracted from fractional Brownian motions and multifractal random walks and emphasized in results for the degree distributions of the constructed networks.

Marwan et al. [2009] proposed a more natural and simple approach. They linked the recurrence matrix, calculated from time series, with the adjacency matrix of an undirected, unweighted network and studied the direct application of the corresponding network measures to the recurrence matrix.

Gao and Jin [2009] introduced a reliable and effective method for constructing complex networks from a time series based on phase space reconstruction. Each vector point of the reconstructed phase space represents a single node of the network and each edge is determined by the phase space distance.

Shirazi et al. [2009] proposed a general method in order to map stochastic processes onto complex networks with distinct geometrical properties. Considering a discrete stationary process $x(t)$, each bin of the probability distribution function of $x(t)$ is represented by a node in the equivalent network of the series. Then nodes i and j are linked if as the time increases the value of $x(t)$ in bin i changes to that in bin j in one time step. The networks they constructed were both directed and weighted.

So far, previous work can be classified into three categories. Thus, the transformation from time series to networks has been done through:

1. The visibility criterion with all the points as nodes and links between points with visibility (Lacasa et al. [2008], Yang et al. [2009], Luque et al. [2009]).

2. The phase space reconstruction with the phase space points as nodes and links between the neighbouring points (Xu et al. [2008], Gao and Jin [2009], Marwan et al. [2009]).

3. A partition of time series values into bins with the bins of the partition as nodes and links between bins with one step time correlations (Shirazi et al. [2009]).

2.2 Height Similarity Method (HS method)

In the following section, the height similarity method that maps a rough surface into a network is presented.

Formally, let the N points of the generated rough profile be the nodes of the expected graph. One way to represent mathematically a graph is via its adjacency matrix $A_{N\times N}$, where the element $a_{ij} = 1$ if there is a link between the nodes i, j and $a_{ij} = 0$ otherwise.

For the links of the graph, the following criterion—the height similarity criterion—is established:

$$\text{if } \; p(i,j) = e^{-|y_i - y_j|} > p_{thres} \;\; \text{then } \; a_{ij} = 1, \; \text{otherwise } \; a_{ij} = 0, \;\; \forall \; i \neq j,$$

where y_i, y_j are the heights of the points i, j respectively and p_{thres} is a threshold that can take values into the interval $(0, 1)$. It is assumed that the constructed graph has no self connections which means that the diagonal elements of its adjacency matrix are equal to zero.

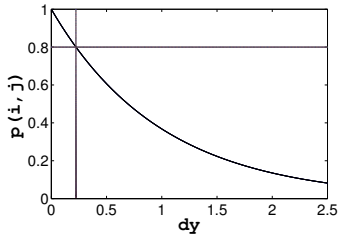


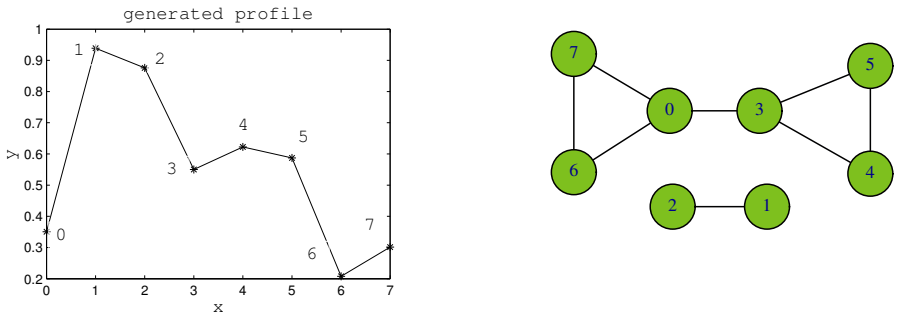
Figure 2.1: The height similarity criterion for $p_{thres} = 0.8$: $p(i,j) = e^{-|y_i - y_j|} > 0.8$ $\Rightarrow \; dy = |y_i - y_j| < -ln(0.8)$.

The criterion above focuses on the correlations of the points according to their heights. Particularly, the distance for each pair of points according to the height–axis is mapped to the value $p(i, j)$. If this value is greater than a threshold that we choose then there is a link between the points i and j. For example, let $p_{thres} = 0.8$. Then the height similarity criterion is: $p(i,j) = e^{-|y_i - y_j|} > 0.8 \Rightarrow \; dy = |y_i - y_j| < -ln(0.8)$. Thus, the two nodes i and j are considered to be connected if their height difference, dy, is less than the value $-ln(0.8)$. The associated graph extracted from a rough surface is always undirected, without multiple connections.

In Fig. 2.1 the adopted criterion is illustrated while in Fig. 2.2 a scheme of the computational method is presented.

Comments:

1. Intuitively, for each point i under examination the HS method "constructs" a *height–similarity zone* around it. Any other point that belongs to this zone is considered to be



(a) (b)

Figure 2.2: (a) A profile composed of $N = 8$ random points on the interval $(0, 1)$ and $p_{thres} = 0.8$ and (b) the respective graph.

connected to the initial point. Thus, the node degree, k_i, of each point i is equal to the number of the points that belong to the HS zone except the point i itself. When the value of p_{thres} is increased the HS zone is shrinking, thus the degree k_i decreases.

2. The extrema of the profile define two zones in which the points contained are named *boundary points*. Particularly, these two zones are defined by the ranges $[h(mim), h(min)+\frac{-\ln p_{thres}}{2})]$ and $[(h(max)-\frac{-\ln p_{thres}}{2}, h(max)]$. The boundary points in the transformed network have a low node degree since their HS zone has smaller range than the HS zones of the rest points.

3. The degree of each node, k_i, is composed of the local and non–local contributions:

$$k_i = k_{i,local} + k_{i,nonlocal}, \quad \text{for } i = 1, ..., N.$$

The local degree $k_{i,local}$ includes the points of the HS zone which are neighbours to the i-th point and depends on the local slope $\left(k_{i,local} \sim \frac{1}{\frac{dy}{dx}\Big|_{x_i}} \right)$. On the other hand, $k_{i,nonlocal}$ comes from the repetitions of the profile and is roughly proportional to their number (Fig. 2.3).

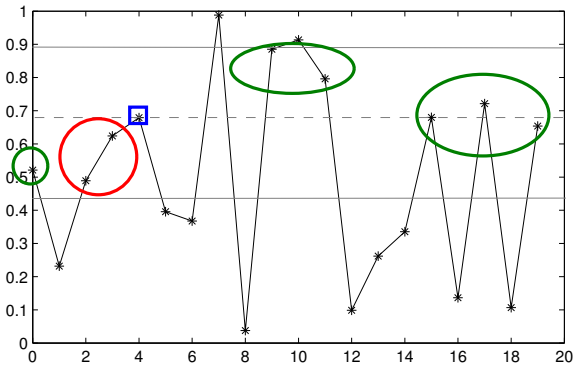


Figure 2.3: The point in the blue square is examined: the points in the red circle are the local contributions while those in green circles are the non-local contributions to the degree of the node under examination.

Chapter 3

Results

The HS method have been applied to different cases of 1D rough surfaces. For different values of the surfaces' parameters, the effect on the statistical measures of the corresponding networks have been examined. In Table 3.1 an overview of all cases is presented.

Type of Rough Surfaces	Input: Surface parameters (plus $p_{thres},\ N$)	Output: Network characteristics
white noise	A	$\langle k \rangle,\ P(k)$
fully periodic	A	C
fractal self–affine	$rms,\ \xi,\ \alpha$	L
mounded (almost periodic)	$A_p,\ A_h,\ A_w$	*Graph footprint*

Table 3.1: An overview of all the studied cases.

The clustering coefficient C for cases where the points of a surface are uniformly distributed on an interval[1], is expected to be high and specifically to be of order 3/4 ($C \sim 0.75$). A recursive proof of our intuition is given below.

Let the height–similarity zone of a point i. Suppose that i has node degree $k_i = k$. If we divide the HS zone vertically into n strips then each of these strips will contain $\frac{k}{n}$ points. The clustering coefficient c_i is recursively estimated as follows:

In the 1st upper strip, there are contained $\frac{k}{n}$ points. Each point is connected with all the other points in the same strip, i.e. $\frac{k}{n}$ points, as well as with points of different strips but in the same HS zone, i.e. $\frac{k}{2}$ points[2]. Hence we have:

$$
\begin{aligned}
1st \quad & \frac{k}{n} \longrightarrow \frac{k}{2} + \frac{k}{n} \\
2nd \quad & \frac{k}{n} \longrightarrow \frac{k}{2} + 2\frac{k}{n} \\
jth \quad & \frac{k}{n} \longrightarrow \frac{k}{2} + j\frac{k}{n} \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\frac{n}{2}th \quad & \frac{k}{n} \longrightarrow \frac{k}{2} + \frac{n}{2}\frac{k}{n} = k
\end{aligned}
$$

[1]White noise and triangular pulse are such cases.

[2]These last points belong in the range from the 2nd to the $\left(\frac{n}{2}+1\right)$th strip.

At this point we can calculate the actual number of edges between neighbours of i, denoted as e_i:

$$2e_i = 2\sum_{j=1}^{n/2} \frac{k}{n}\left(\frac{k}{2} + j\frac{k}{n}\right) = 2\frac{k}{n}\left[\sum_{j=1}^{n/2}\frac{k}{2} + \sum_{j=1}^{n/2} j\frac{k}{n}\right] = 2\frac{k}{n}\left[\frac{k}{2}\frac{n}{2} + \frac{k}{n}\frac{n}{4}\left(\frac{n}{2}+1\right)\right] = \frac{k^2}{2}+\frac{k^2}{2}\left(\frac{1}{2}+\frac{1}{n}\right)$$

.

For $n \to \infty$, we get $2e_i = \frac{3}{4}k^2$ and finally $c_i = \frac{2e_i}{k(k-1)} \simeq 0.75$.

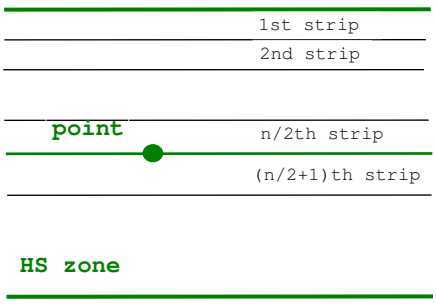


Figure 3.1: A schematic representation of the proof that $C \simeq 0.75$ in a profile where the points are uniformly distributed.

Another property that the constructed networks inherit from the HS method is that they are highly–clustered nets with large average shortest path length. For example, let *point 1* and *point 2* and their HS zones, *1* and *2* respectively. These two zones intersect and the points belonging to the intersection are connected to each other as well as to our initial points (Fig. 3.2).

A path between two points at different height zones has to cross over different HS zones. The number of these zones depends on the height distance of the points as well as the threshold of the HS method. Let $dh = |h(p_1) - h(p_2)|$ be the height distance of two points p_1 and p_2. The shortest path length between p_1 and p_2 is given by the formula: $l(p_1 \to p_2) = \frac{dh}{-ln(p_{thres})}$.

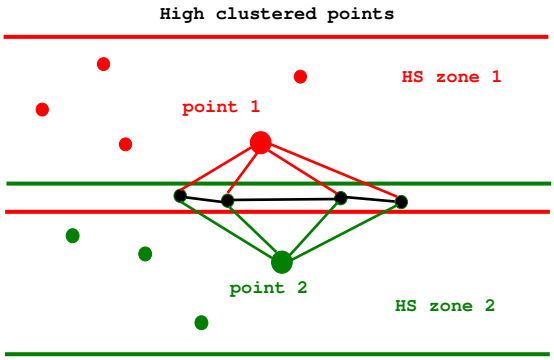
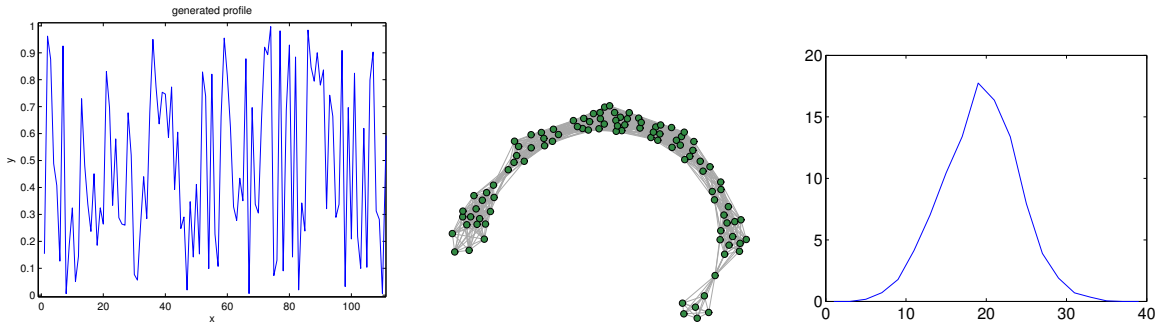


Figure 3.2: Description of the method's nature: highly clustered points.

3.1 Extreme Cases

3.1.1 White noise

In a first step, a white noise is considered, i.e a 1-D surface profile composed of N random points uniformly distributed on an interval with amplitude A (Fig. 3.3).



(a) (b) (c)

Figure 3.3: (a) A white noise composed of $N = 100$ points distributed on the height interval $(0, 1)$, (b) the corresponding graph with $p_{thres} = 0.9,\ C = 0.7799,\ L = 3.9503, \langle k \rangle = 19.7492$ and (c) the degree distribution $P(k)$.

In white noise, the contribution of the local node degree approximates zero, $k_{i,local} \sim 0$, and $k_i = k_{i,nonlocal}$. The associated graph possesses a Poisson degree distribution and the average degree's analytical formula is: $\langle k \rangle = \dfrac{-2 \ln p_{thres}}{A} N$. In Fig. 3.4, 3.5 the effects of A and of p_{thres} on the degree distribution $P(k)$ (as well as in the clustering coefficient C and the average shortest path length L) are illustrated:

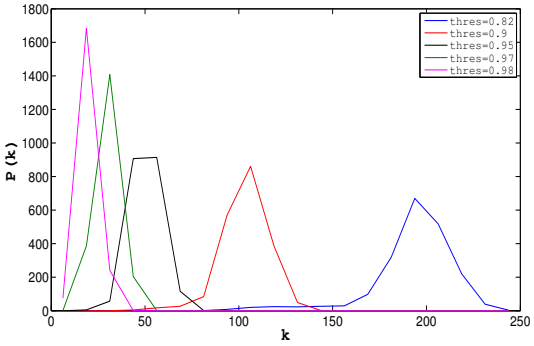
- when p_{thres} increases, fewer points are connected to each other and that results in more and more points to have a low degree. So the peak of $P(k)$ is becoming higher and is shifted toward smaller degrees. For smaller values of p_{thres} an enhancement of the left tail of $P(k)$ is observed which reflects the broadening of the height-similarity zone. Furthermore, clustering coefficient reduces, although the decrement is negligible. Regarding L, increasingly small values of p_{thres} make the presence of shortcuts more and more intense.
- when A increases, the points are sparsely distributed in space and that results in fewer connections between them with similar consequences for $P(k)$, C and L as above.

An interesting outcome is that the graphs created from white noise by the HS method have a Poisson degree distribution. This allows us to make a comparison between our *HS graphs* and Erdös–Rényi random graphs. Moreover, since the points are uniformly distributed on an interval of amplitude A, ring lattice's statistical measures can also be compared with "white noise network". For the two comparisons above, we will take extreme values regarding the values of parameters A and p_{thres}:

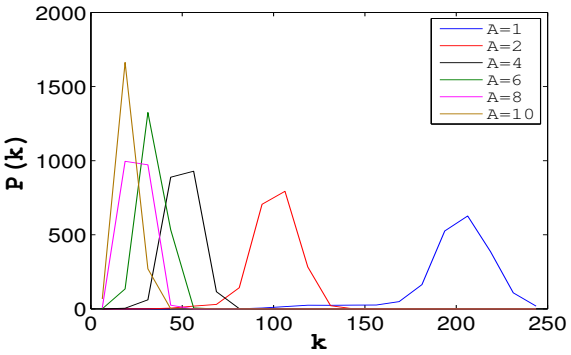
1. For $N = 2000$, $A = 1$, $p_{thres} = 0.82$ we get $\langle k \rangle = 714.9295$, $C = 0.8040$ and $L = 1.6524$.

   In a random graph with $\langle k \rangle_{random} = \langle k \rangle$ the probability p of connecting each pair of nodes is $p = \dfrac{\langle k \rangle}{N-1} = 0.3576$. Hence, $C_{random} = p = 0.3576$ and $L_{random} = \dfrac{\ln N}{\ln \langle k \rangle} = 1.1565$.

   In a ring lattice, $C_{ring} = \dfrac{3(K-2)}{4(K-1)} = 0.7489$ and $L_{ring} = \dfrac{N}{2K} = 1.4035$.



(a)



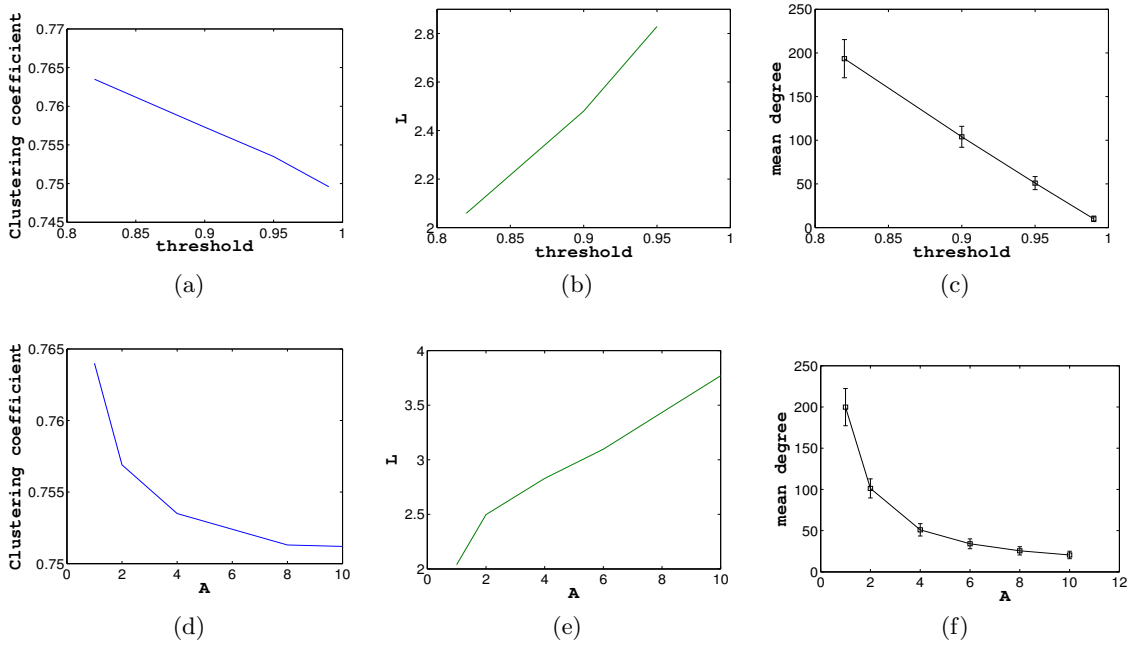
(b)

Figure 3.4: Degree distributions for white noise. Effects of A, p_{thres} in $P(k)$ for $N = 2000$, $runs = 100$ (a) $A = 4$ (b) $p_{thres} = 0.95$.

2. For $N = 2000$, $A = 10$, $p_{thres} = 0.95$ we get $\langle k \rangle = 20.4555$, $C = 0.7515$ and $L = 3.7749$.

   In the random graph we have $C_{random} = 0.0102$ and $L_{random} = 2.5181$ while in the ring lattice we take $C_{ring} = 0.7115$ and $L_{ring} = 46.0756$.

In both cases, we observe that $C_{random} \ll C \sim C_{ring}$ and $L_{random} < L \sim L_{ring}$. Moreover, the degree distribution of white noise networks and random graphs have a Poison degree distribution while $P(k)_{ring}$ is a delta function.



(a) (b) (c)

(d) (e) (f)

Figure 3.5: Clustering coefficient, average shortest path length and mean degree of white noise. In the calculations above the standard parameters are: $N = 2000$, $runs = 100$, (a)–(c) $A = 4$, (d)–(f) $p_{thres} = 0.95$.

3.1.2 Fully periodic 1-D surfaces

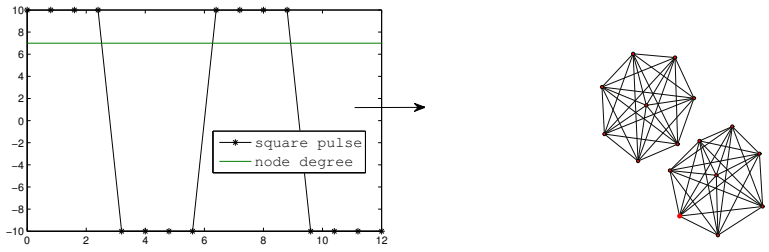
Square pulse:

For a square pulse of N points, the method gives a graph of two disconnected components, the one for the $N/2$ points in the upper zone of the pulse and the other for the $N/2$ points of the bottom (Fig. 3.6).

Since the constructed graph is disconnected, the shortest average path length, L, diverges. Each component of the graph constitutes a complete graph therefore for each of the subgraphs the shortest path length is $l_i = 1$, $\forall i \in N$ and the clustering coefficient C takes its larger value, i.e. $C = 1$. The degree of each node is $k_i = \dfrac{N}{2} - 1$ (self connections are not permitted) and the degree distribution is a Kronecker delta function,

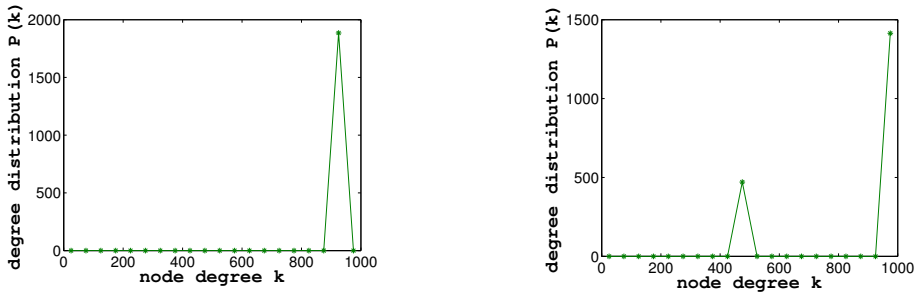
$$P(k) = \begin{cases} N, & \text{for } k = \frac{N}{2} - 1 \\ 0, & \text{otherwise} \end{cases}$$

The peak of the degree distribution is at high node degree as there are many points in the same height zone. From so on, this kind of peak will be denoted as *type I peak* (Fig. 3.7a).



(a) (b)

Figure 3.6: (a) A square pulse with $N = 16,\ A = 10,\ p_{thres} = 0.9$ and (b) the respective disconnected graph with $k_i = 7$ and $c_i = 1,\ \forall i \in N$, L diverges for the whole graph while for each subgraph $L = 1$.



(a) (b)

Figure 3.7: Degree distributions for square pulse. In the calculations above, the standard parameters are $N = 1885,\ A = 1,\ p_{thres} = 0.95$ (a) dc=0.5 (b) dc=0.75.

For a square pulse with duty cycle [3] (dc) not equal to 0.5 the degree distribution is:

$$P(k) = \begin{cases} N*(dc), & \text{for } k = N*(dc)-1 \\ N*(1-dc), & \text{for } k = N*(1-dc)-1 \\ 0, & \text{otherwise} \end{cases}$$

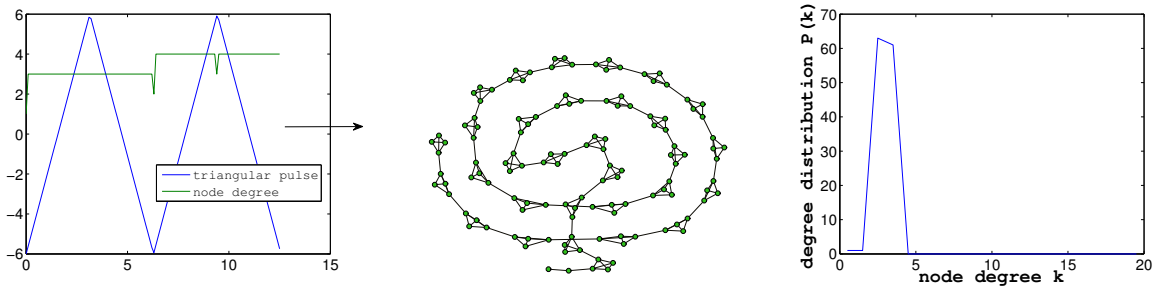
In any change of the parameters w and p_{thres}, the two groups[4] of points in the square pulse remain in the same height zones.

Triangular pulse:

The node degree of each node in a triangular pulse is $k_i = 2m(k_{i,local}+1)-1$, where $k_{i,local} \sim \frac{T}{2A}$ and m is the number of iterations. The degree distribution $P(k)$ is a Kronecker delta function,

$$P(k) = \begin{cases} N, & \text{for } k = k_i \\ 0, & \text{otherwise} \end{cases}$$

The peak of $P(k)$ is at low degree as there are many points in different height zones but with similar node degree. These points are coming from the linear slopes of the triangular pulse. From so on, this kind of peak will be denoted as *type II peak*.



(a) (b) (c)

Figure 3.8: (a) A triangular pulse with $N = 126,\ A = 6,\ p_{thres} = 0.8,\ m = 2$ (b) the respective graph with $\langle k \rangle = 3.4603,\ C = 0.7421,\ L = 21.6667$ (c) the degree distribution $P(k)$.

There two ways to decrease the network's node degree. The first way is by increasing the amplitude of the profile. Then, since our period is fixed, the same points form a triangular pulse with a steeper slope. The other way is by increasing the threshold thereafter by decreasing the range of the height–similar zones.

The results describing the clustering coefficient and the average shortest path length are expected. When we decrease the node degree then C is getting smaller while L becomes larger.

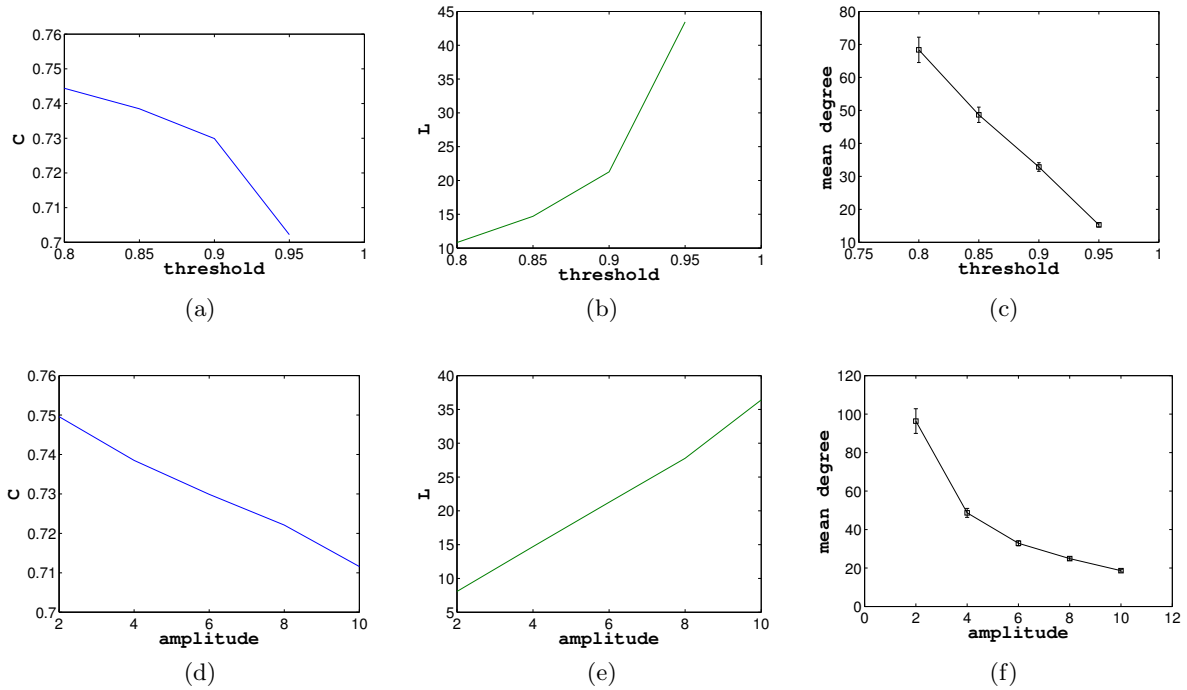
At this point, a comparison between networks coming from the triangular pulse and the ring lattice can be done: a ring lattice is a counterpart of this highly ordered lattice-type network. The nodes of a triangular pulse (which are uniformly distributed) can be considered as the nodes of a ring lattice. In Table 3.2 the arithmetical results for C and L are compared with the analytical results for C_{ring} and L_{ring}. We observe that $C \simeq C_{ring}$ while $L < L_{ring}$.

[3] The duty cycle or duty ratio is the percent of period in which the signal is positive and takes values in the space $(0, 1)$.

[4] The one group gives the positive and the other the negative signal.

$N = 1885,\ m = 3,\ p_{thres} = 0.9$					
A	$\langle k \rangle$	C	C_{ring}	L	L_{ring}
2	96.3756	0.7496	0.7421	8.0768	9.7794
6	32.8467	0.7299	0.7264	21.2704	28.6939
8	24.9103	0.7221	0.7186	27.7726	37.8358
10	18.6164	0.7116	0.7074	36.4060	50.6274

Table 3.2: Comparison between a triangular network and a ring lattice. The clustering coefficient and the average shortest path length are examined.



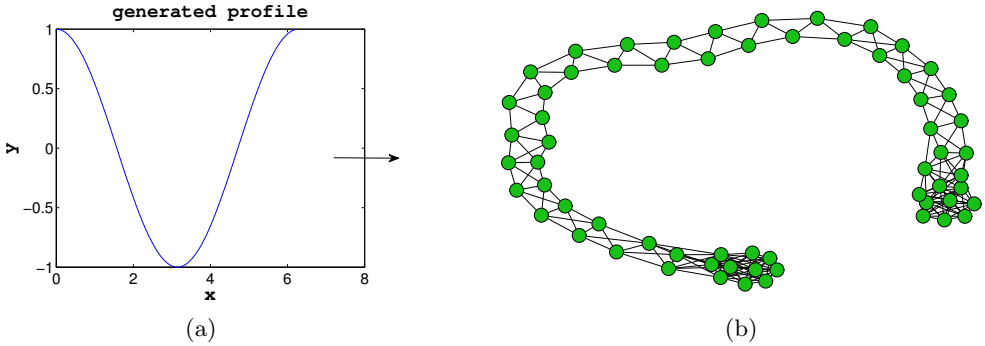
(a) (b) (c)

(d) (e) (f)

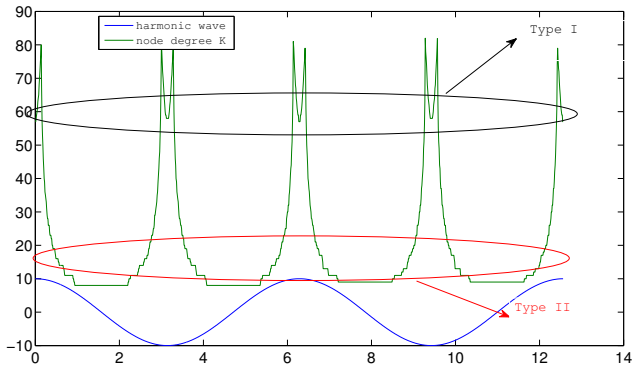
Figure 3.9: Clustering coefficient, average shortest path length and mean degree of triangular pulse. In the calculations above the standard parameters are $N = 1885,\ m = 3,\ A = 6,\ p_{thres} = 0.9$ (a)–(c) $p_{thres} = 0.8,\ 0.85,\ 0.9,\ 0.95$ (d)–(f) $A = 2,\ 4,\ 6,\ 8,\ 10$.

Harmonic wave:

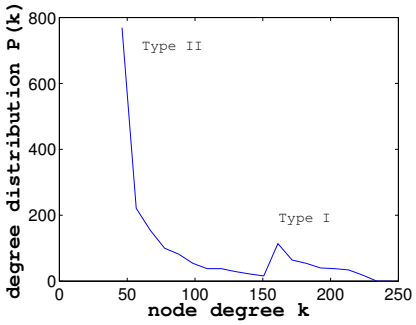
Let an harmonic wave $y(x) = A \cdot \cos(\omega x + \phi)$. In the degree distribution of the generated graph, two peaks make their appearance, the type I and type II peaks. The type I peak is a low peak at high degree corresponding to the points at same height zone, i.e. to the maxima and minima of the profile while the type II peak is a high peak at low degree coming from the almost linear slopes of the wave (Fig. 3.10c, 3.10d).



(a) (b)



(c)



(d)

Figure 3.10: (a) Harmonic wave $y(x) = \cos x,\ N = 63,\ p_{thres} = 0.9$ and (b) the respective graph with $\langle k \rangle = 6.4762,\ C = 0.6657,\ L = 9.4936$ (c) degree per node (d) degree distribution $P(k)$ of an harmonic wave $y(x) = 3 \cdot \cos x,\ N = 1885,\ p_{thres} = 0.9$.

In Fig. 3.11 the effect of the amplitude is presented: when the amplitude of the harmonic wave is increasing, the slope of the profile becomes steeper. This has effect on a smaller node degree for both the maxima, minima and the points belonging to the almost linear slopes of the profile. One can observe an overall shift of the degree distribution $P(k)$ towards lower degrees.

$N = 1885,\ p_{thres} = 0.9$			
A	$\langle k \rangle$	C	L
2	120.1634	0.7754	9.7957
6	46.8456	0.7458	27.5572
10	29.9416	0.7258	46.1404

(a)

$N = 1885,\ A = 6$			
p_{thres}	$\langle k \rangle$	C	L
0.8	89.6032	0.7670	13.3004
0.85	68.3119	0.7589	17.8745
0.95	24.4594	0.7110	58.0457

(b)

Table 3.3: Arithmetical results for the mean degree $\langle k \rangle$, the clustering coefficient C and the average shortest path length L of the network induced from an harmonic wave. Effect of the (a) amplitude A, (b) p_{thres}

Moreover as we increase A and p_{thres}, the clustering coefficient becomes smaller while the average shortest path length gets bigger. In Table the values of our measures are presented.

3.2 Fractal self-affine surfaces

A self-affine fractal surface is a class of fractal objects that can be described by the parameters:

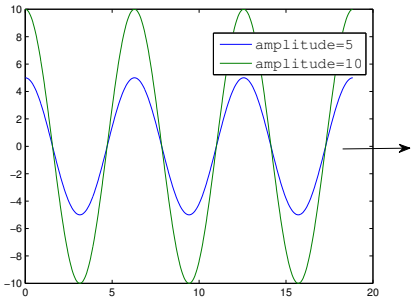
1. Rms value (root mean square) or interface width: it describes the fluctuations of surface heights around an average surface height. 99% of surface points are included in a zone of width $6rms$.
   As rms increases, the vertical roughness fluctuations (height fluctuations) around $\bar{h} = 0$ become more intense (Fig. 3.13a). In the respective network, the degree of many points becomes lower because the same number of surface points are distributed in a broader range (many points in different height zones). Thus the degree distribution $P(k)$ shifts to lower $k's$ and an increment of the peak is observed. Contrary, as rms becomes smaller, many points are in the same height zone and their node degree increases (Fig. 3.13d).

   Regarding clustering coefficient C, large values of rms make the graph gradually sparse and contribute to a significant decrement of C (Fig. 3.14a).

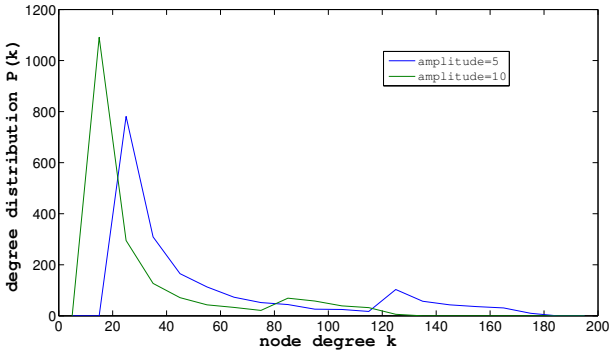
2. Roughness exponent α: it describes how wiggly the surface is. A larger value of α corresponds to a locally smooth surface structure while a smaller value of α corresponds to a more jagged local surface morphology (Fig. 3.13b).
   As α increases, the contribution of the local degree, k_{local}, becomes bigger. In the degree distribution, an enrichment of high $k's$ as well as a lowering of the peak are appearing (Fig. 3.13e). The increment of α has small effect at the mean degree $\langle k \rangle$ (Fig. 3.14e).

   Regarding C, for $\alpha \in [0.2, 0.5]$ a small decrement of C's value is observed. For $\alpha \in [0.5, 0.9]$ there is no effect in C (Fig. 3.14b).

3. Correlation length ξ: it defines a representative lateral dimension of a rough surface. If the distance, r, between two surface points is within ξ ($r < \xi$), the heights at these two points can be considered correlated. However, if the separation of two surface points is much larger than ξ, then the heights at these two points are independent of one another. Between two surfaces, the one with the smallest ξ looks rougher (Fig. 3.13c).
   As ξ increases, similarly with α, a lowering of the peak and an enhancement of high $k's$ is observed in the degree distribution diagram (Fig. 3.13f). The effect of ξ in the mean degree $\langle k \rangle$ is also negligible (Fig. 3.14f).



(a)

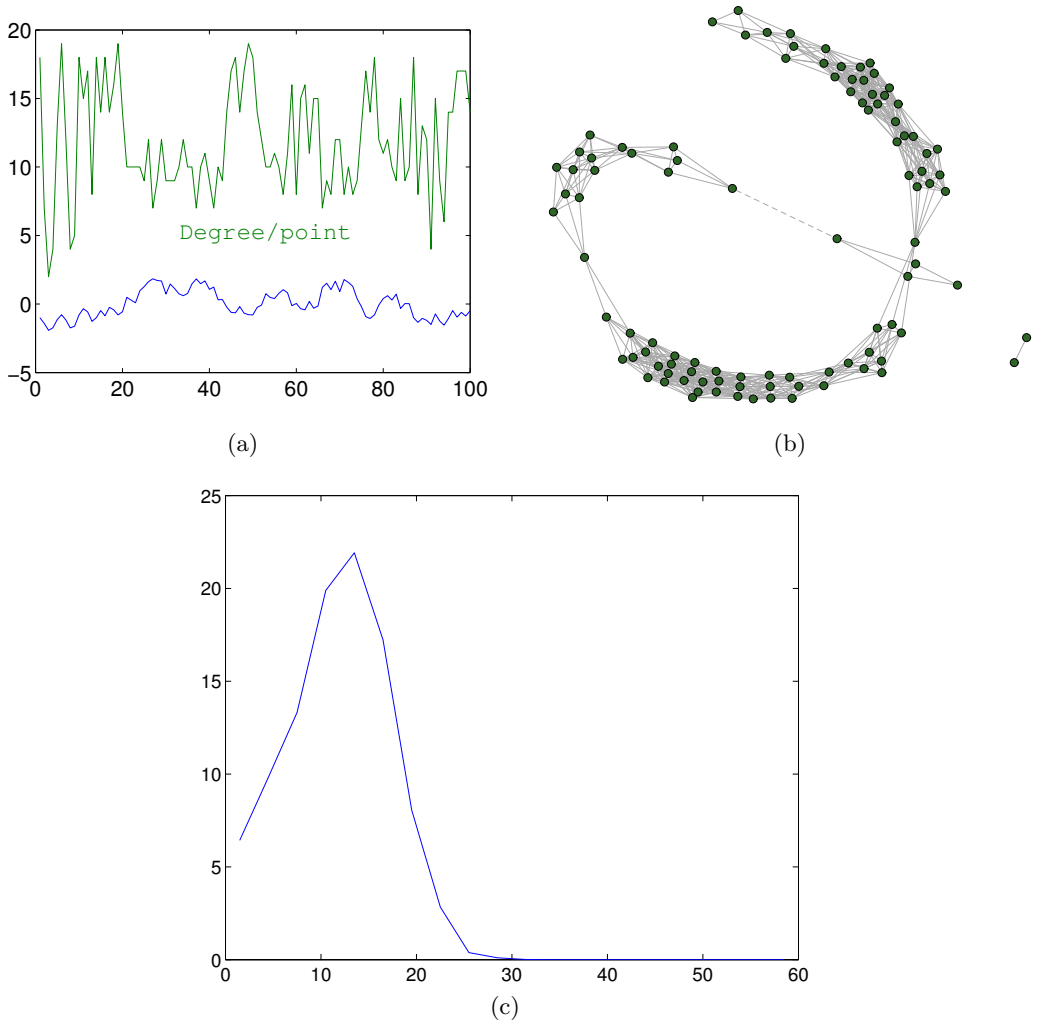


(b)

Figure 3.11: (a) Two harmonic waves $y_1(x) = 5 \cdot \cos x,\ y_2(x) = 10 \cdot \cos x,\ N = 1885,\ p_{thres} = 0.9$ and (b) effect of the amplitude to the degree distribution $P(k)$: $\langle k_1 \rangle = 54.8881,\ \langle k_2 \rangle = 29.9416$.

Regarding C, it takes its minimum value for $\xi = 10$ and its maximum value for $\xi = 100$, although the effect of ξ to C is very small.

The peak of the degree distribution $P(k)$ is a type I peak which comes from the points in the same (or similar) zones. Finally, the average shortest path length L diverges for $N = 2000$, $p_{thres} = 0.8 - 0.95$, $runs = 100$ and different values of rms, ξ, and α. This means tha the constructed networks are disconnected.



(a) (b)

(c)

Figure 3.12: (a) Degree per node and (b) a graph of a fractal self–affine surface with $N = 100$, $p_{thres} = 0.8$, $rms = 1$, $\xi = 10$, $\alpha = 0.5$, $\langle k \rangle = 12.1234$, $C = 0.7465$, L *diverges*, (c) degree distribution $P(k)$.

3.3 Mounded Surfaces

In this section, mounded 1-D surfaces are examined. The mounded profiles considered here are composed by Gaussian humps distributed on x-axis. Gaussian humps can be described by the position, the height and the width of the Gaussian peak. (Fig. 3.15)

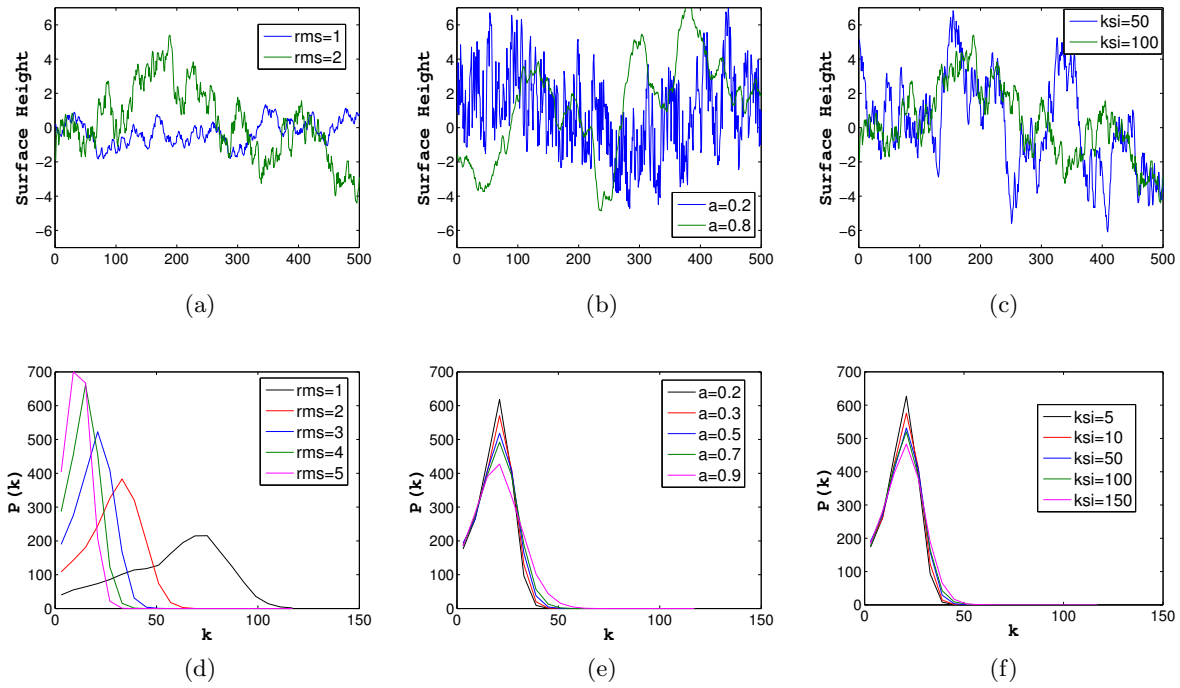


Figure 3.13: Profiles and degree distributions of fractal self–affine surfaces. In the calculations above the standard parameters are: $N = 2000$, $p_{thres} = 0.95$, $runs = 100$, $rms = 3$, $\xi = 50$, $\alpha = 0.5$.

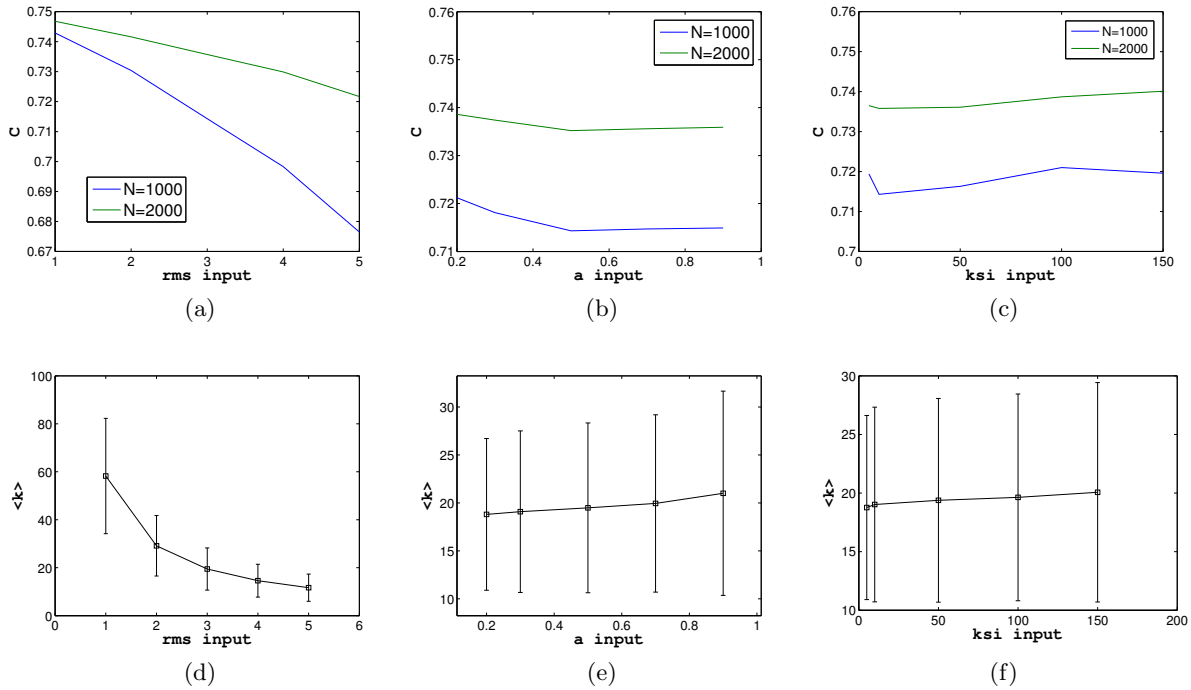


Figure 3.14: Clustering coefficient and mean degree of fractal self–affine surfaces. In the calculations above the standard parameters were: $N = 2000$, $p_{thres} = 0.95$, $runs = 100$, $rms = 3$, $\xi = 10$, $\alpha = 0.5$. (a)–(c) Effect of the parameters at the clustering coefficient C. For $N = 2000$ the decline rate of C is smaller than for $N = 1000$. (d)–(f) Effect of the parameters at the mean degree $\langle k \rangle$.

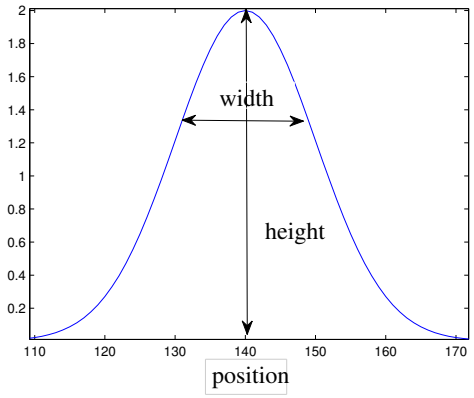


Figure 3.15: The three parameters for describing Gaussian peaks: position, width and height.

3.3.1 Full Periodicity

The first case to be considered is when the mound profile is fully periodic, i.e. the profile is composed by identical Gaussian humps which are equally distributed on x-axis. Equivalently, there are no deviations from a mean position, a mean height and a mean width.

In the degree distribution $P(k)$ two peaks make their appearance: the type I peak coming from the minima of the profile and the type II peak coming from the almost linear slopes and the maxima of the Gaussian motif. Regarding minima, as the distance between two successive positions of Gaussian peaks increases, these points become gradually more. (Fig. 3.16) A question that arises is how the degree distribution $P(k)$ is affected by the deviations from full periodicity.

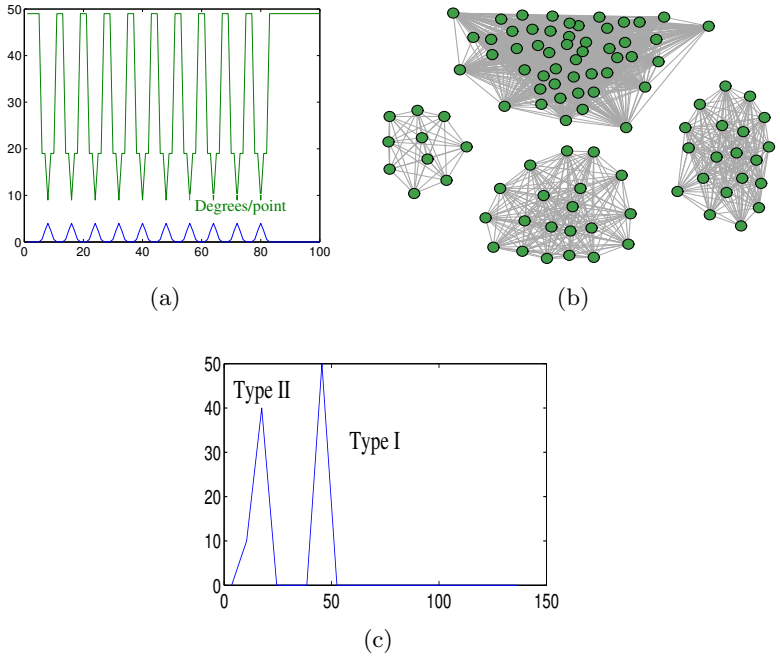


Figure 3.16: In the calculations above the standard parameters are: $N = 100$, $runs = 100$, $p_{thres} = 0.9$, $distance = 8$, $height = 4$, $width = 1$. (a) Degree per node, (b) network with $\langle k \rangle = 33$, $C = 1$, $L = 1$ for each subgraph and (c) degree distribution $P(k)$ of a fully periodic mounded surface.

3.3.2 Deviations from Full Periodicity

There can be three ways of deviation from periodicity, so many as the number of parameters that describe the Gaussian humps.

1. Deviations due to position fluctuations:

   The peak position vector is given by the formula: $d_i(dev) = d_i(per) + A_p \times \eta_i,\ \forall$ peak i, where $d_i(per)$ is the peak position vector in a fully-periodic mound profile, A_p is the deviation of the peak distances and η_i is the Gaussian white noise. (Fig. 3.17a)

   Different values of the parameter A_p affect the type I and type II peaks of the degree distribution. When deviation from the position, A_p, increases, the number of minima, i.e. points which are on same height zones, increases too. For $A_p = 20$, the type I peak is shifted towards larger node degree values. Moreover, the right tail of this peak denotes an enhancement of high k's. The mean node degree grows and the clustering coefficient becomes smaller although the effect at C is negligible.

2. Deviations due to height fluctuations

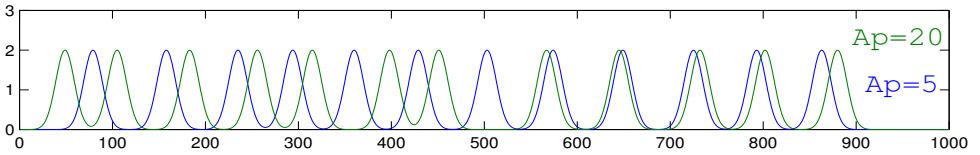
   The peak height vector is given by the formula: $h_i(dev) = h(per) + A_h \times \eta_i,\ \forall$ peak i, where $h_i(per)$ is the peak height vector in a fully-periodic mound profile and A_h is the deviation of the peak heights. (Fig. 3.17b)

   Different values of the parameter A_h affect both type I and type II peaks of the degree distribution. As A_h becomes larger, Gaussian humps have both steeper and smoother inclines than the inclines for small values of A_h. The points belonging to the steep slopes of the mounded surface are responsible for the type II peak's shift to lower k's. Points distributed in smoother slopes have effect at the type I peak: an enhancement of high k's as well as a peak's shift to larger k's is observed. For $A_h \in [0, 0.4]$ we have a significantly lowering of the type II peak while for $A_h \in [0.4, 1]$, the type I peak is affected. Regarding the clustering coefficient, for $A_h \in [0, 0.4]$ a significant decline is observed. This decline is coming from the steep inclines of the Gaussian humps.

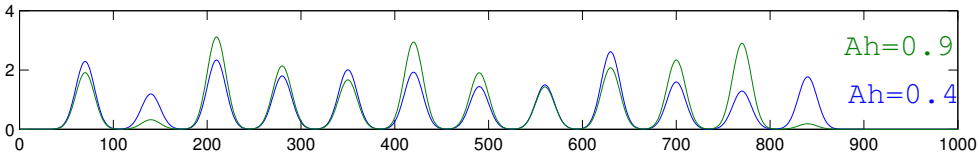
3. Deviations due to width fluctuations

   The peak width vector is given by the formula: $w_i(dev) = w(per) + A_h \times \eta_i,\ \forall$ peak i, where $w_i(per)$ is the peak width vector in a fully-periodic mound profile and A_w is the deviation of the peak widths. (Fig. 3.17c)

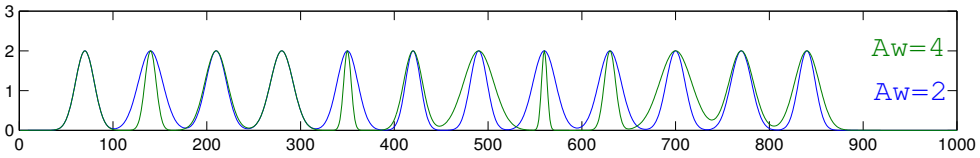
   Different values of the parameter A_w affect both type I and type II peaks of the degree distribution. All the significant results concerning the degree distribution and the clustering coefficient are observed for $A_w \in [0, 2]$. As the width deviations increase, we have an enrichment of small k's in the type I peak which tends to be smoother. In both cases a lowering of the peaks is observed.



(a)



(b)



(c)

Figure 3.17: Profiles of mounded surfaces with deviations from full periodicity. The standard parameters in the calculations above are: $N = 1000$, *mean distance* $= 70$, *mean height* $= 2$, *mean width* $= 10$

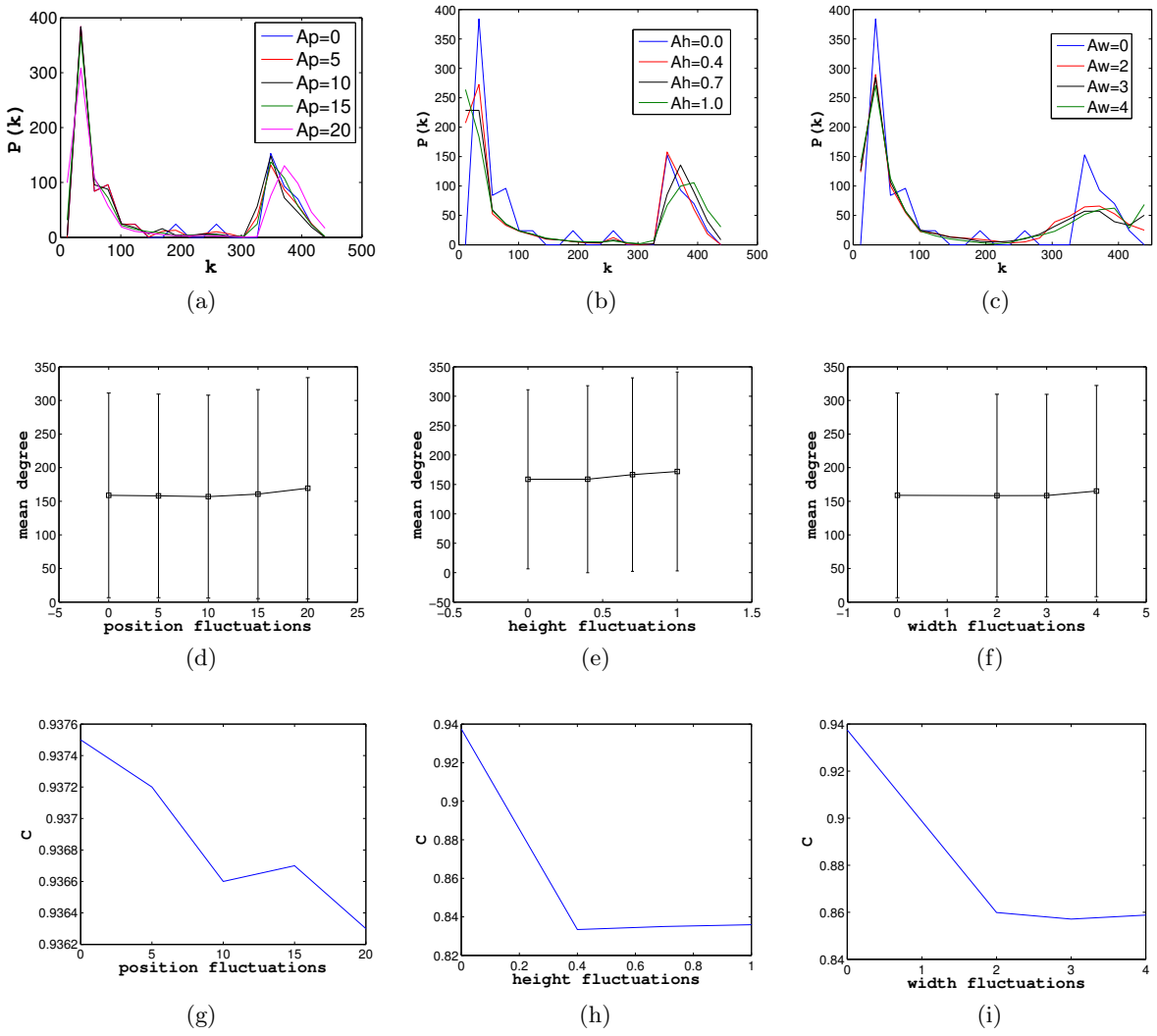


Figure 3.18: Degree distribution, mean degree and clustering coefficient of mounded profiles with deviations from full periodicity. In the calculations above the standard parameters were: $N = 1000$, $p_{thres} = 0.95$, $runs = 100$, $d(per) = 70$, $h(per) = 2$, $w(per) = 10$, (a)–(c) node degree distribution, (d)–(f) mean degree, (g)–(i) effect in clustering coefficient.

Chapter 4

Summary and Conclusions

In this work we have introduced the *height similarity method* which transforms a rough surface into a network. The nodes of the network are the points of the surface's profile. Concerning the links of the network, the method takes into account correlations regarding the height distances of surface's points: distant points in the spacial dimension, although belonging in the same height–similar zone, are considered to be connected.

The HS method have been applied to different cases of rough surfaces: from random to fully periodic as well as to fractal self–affine and mounded surfaces. For different values of the surfaces' parameters, the effect on the corresponding networks have been examined. The statistical measures that we studied are the node degree $\langle k \rangle$, the degree distribution $P(k)$, the clustering coefficient C and the average shortest path length L. Network footprints for all the types of surfaces have been illustrated in order to get a more intuitive picture of the networks' structure. Our results can be summarized in Tables 4.1, 4.2, 4.3.

The main findings of the thesis are:

1. The mean degree $\langle k \rangle$ in the cases of white noise, square and triangular pulse can be analytically estimated.

2. The peak of the degree distribution $P(k)$ is sensitive to the surface's periodicity. In fully periodic rough surfaces we can have either a type I peak—coming from points in the same height zone—, either a type II peak —coming from points in different height zones, or both types of peaks. In a white noise we have a Poisson degree distribution and in fractal self–affine, $P(k)$ is a type I peak sensitive to any change of the rms parameter. Finally, in mounded rough surfaces, when we have deviations from full periodicity both type I and II peaks are affected.

3. The clustering coefficient C for networks produced by white noise or triangular pulse approximates the vale $\frac{3}{4}$. A recursive proof that $C \sim 0.75$ is given. In general, our constructed networks are highly clustered.

4. The average shortest path length is also long and in some cases diverges.

Things for future consideration

1. Estimation of more network statistical measures (efficiency E in cases where L diverges)

2. Extension of the method to include the spatial distance, transformation to geographical models

3. Examination of real experimental surfaces, comparison with the conventional methods for roughness analysis

Surfaces	$\langle k \rangle$	$P(k)$	C	L
White noise (wn)	$\frac{-2\ln(p_{thres})N}{A}$	Poisson	$\gg C_{random}$, $\sim 3/4$, $\sim C_{ring}$	$> L_{random}$
Square pulse (sq.p)	$\frac{N}{2} - 1$	delta function, type II peak	1	diverges
Triangular pulse (tr.p)	$2m(k_{i,local}+1)-1$	delta function, type II peak	$\sim C_{ring}$	$< L_{ring}$
Harmonic wave (hw)/Mounded fully periodic (mfp)		type I, II peaks		
$A,\ p_{thres} \nearrow \quad \Rightarrow \quad \langle k \rangle \searrow,\ C \searrow\ (s.e.),\ L \nearrow$ for wn, tr.p, (n.e.) for sq.p				

Table 4.1: Results for random and fully periodic rough surfaces.

	Fractal self–affine surfaces		
	$rms \nearrow$	$\xi \nearrow$	$\alpha \nearrow$
$\langle k \rangle$	$\searrow$	$\nearrow$ (s.e.)	$\nearrow$ (s.e.), $k_{local} \nearrow$
$P(k)$ (type I peak)	shifts to lower k's, increment of the peak	enhancement of high k's, lowering of the peak	
C	$\searrow$	$\searrow$ (s.e.), minimum for $\xi = 10$, maximum for $\xi = 100$	$\searrow$ (s.e.) for $\alpha \in [0.2, 0.5]$, (n.e) for $\alpha \in [0.5, 0.9]$
L	diverges, for $N = 1000,\ 2000,\ p_{thres} = 0.8 - 0.95,\ runs = 100,$ $rms \in [1, 5],\ \xi \in [5, 150],\ \alpha \in [0.2, 0.9]$		

Table 4.2: Results for fractal self–affine surfaces.

Mounded surfaces			
Deviations from full periodicity	position fluctuations $A_p \nearrow$	height fluctuations $A_h \nearrow$	width fluctuations $A_w \nearrow$
$\langle k \rangle$	$\nearrow$		
$P(k)$	for $A_p = 20$, type I: shifts to large k's, enhancement of high k's, type II: lowering of the peak	type I: shifts to large k's, enhancement of high k's, lowering of the peak for $A_h \in [0.4, 1]$, type II: shifts to low k's, lowering of the peak	for $A_w \in [0, 2]$ type I: enhancement of low k's, lowering of the peak, type II: lowering of the peak
C	$\searrow$ (s.e.)	$\searrow$, for $A_h \in [0, 0.4]$	$\searrow$, for $A_w \in [0, 2]$
L	diverges, for $N = 1000,\ p_{thres} = 0.95,\ runs = 100$ $d(per) = 70,\ h(per) = 2,\ w(per) = 10$ $A_p = [5, 20],\ A_h = [0, 1],\ A_w = [0, 4]$		

Table 4.3: Results for mounded surfaces with deviations from full periodicity.

Bibliography

R. Albert and A. Barabsi. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

A. Barabsi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439): 509–512, Oct. 1999. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.286.5439.509. URL http://www.sciencemag.org/content/286/5439/509.

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(45):175–308, Feb. 2006. ISSN 0370-1573. doi: 10.1016/j.physrep.2005.10.009. URL http://www.sciencedirect.com/science/article/pii/S037015730500462X.

L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas. Characterization of complex networks: A survey of measurements. *arXiv:cond-mat/0505185*, May 2005. doi: 10.1080/00018730601170527. URL http://arxiv.org/abs/cond-mat/0505185. Advances in Physics, Volume 56, pages 167 - 242, Issue 1 (2007).

R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths. Ambiguities in recurrence-based complex network representations of time series. *Physical Review E*, 81(1), Jan. 2010. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.81.015101. URL http://link.aps.org/doi/10.1103/PhysRevE.81.015101.

S. Dorogovtsev and J. Mendes. The shortest path to complex networks. *Arxiv preprint cond-mat/0404593*, 2004.

P. ERDdS and A. R&WI. On random graphs i. *Publ. Math. Debrecen*, 6:290297, 1959.

Z. Gao and N. Jin. Complex network from time series based on phase space reconstruction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 19(3):033137, 2009. ISSN 10541500. doi: 10.1063/1.3227736. URL http://link.aip.org/link/CHAOEH/v19/i3/p033137/s1&Agg=doi.

X. Jiang, P. J. Scott, D. J. Whitehouse, and L. Blunt. Paradigm shifts in surface metrology. part i. historical philosophy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 463(2085):2049–2070, Sept. 2007. ISSN 1364-5021, 1471-2946. doi: 10.1098/rspa.2007.1874. URL http://rspa.royalsocietypublishing.org/content/463/2085/2049.

L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. Nuo. From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences*, 105(13):4972, 2008.

V. Latora and M. Marchiori. Efficient behavior of Small-World networks. *Physical Review Letters*, 87(19), Oct. 2001. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.87.198701. URL http://link.aps.org/doi/10.1103/PhysRevLett.87.198701.

V. Latora and M. Marchiori. Economic small-world behavior in weighted networks. *The European Physical Journal B - Condensed Matter*, 32(2):249–263, Mar. 2003. ISSN 1434-6028, 1434-6036. doi: 10.1140/epjb/e2003-00095-5. URL http://www.springerlink.com/Index/10.1140/epjb/e2003-00095-5.

B. Luque, L. Lacasa, F. Ballesteros, and J. Luque. Horizontal visibility graphs: exact results for random time series. *Physical Review E*, 80(4):046103, 2009.

O. Malcai, D. A. Lidar, O. Biham, and D. Avnir. Scaling range and cutoffs in empirical fractals. *arXiv:cond-mat/9801127*, Jan. 1998. doi: 10.1103/PhysRevE.56.2817. URL http://arxiv.org/abs/cond-mat/9801127. Phys. Rev. E, vol. 56, pp. 2817-2828 (1997).

N. Marwan, J. Donges, Y. Zou, R. Donner, and J. Kurths. Complex network approach for recurrence analysis of time series. *Physics Letters A*, 373(46):42464254, 2009.

M. Newman. The structure and function of complex networks. *SIAM review*, page 167256, 2003.

X. Ni, Z. Jiang, and W. Zhou. Degree distributions of the visibility graphs mapped from fractional brownian motions and multifractal random walks. *Physics Letters A*, 373(42): 3822–3826, Oct. 2009. ISSN 0375-9601. doi: 10.1016/j.physleta.2009.08.041. URL http://www.sciencedirect.com/science/article/pii/S0375960109010536.

P. Peter. Fractal dimension as working tool for surface-roughness problems. *Applications of Surface Science*, 18(12):146–164, June 1984. ISSN 0378-5963. doi: 10.1016/0378-5963(84)90042-4. URL http://www.sciencedirect.com/science/article/pii/0378596384900424.

D. Shi, S. X. Zhu, and L. Liu. Clustering coefficients of growing networks. *Physica A: Statistical Mechanics and its Applications*, 381:515 – 524, 2007. ISSN 0378-4371. doi: DOI:10.1016/j.physa.2007.03.054. URL http://www.sciencedirect.com/science/article/pii/S0378437107003342.

A. Shirazi, G. Reza Jafari, J. Davoudi, J. Peinke, M. Reza Rahimi Tabar, and M. Sahimi. Mapping stochastic processes onto complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009:P07046, 2009.

D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684): 440–442, June 1998. ISSN 0028-0836. URL http://dx.doi.org/10.1038/30918.

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

Y. Yang, J. Wang, H. Yang, and J. Mang. Visibility graph approach to exchange rate series. *Physica A: Statistical Mechanics and its Applications*, 388(20):4431–4437, Oct. 2009. ISSN 0378-4371. doi: 10.1016/j.physa.2009.07.016. URL http://www.sciencedirect.com/science/article/pii/S037843710900569X.

J. Zhang and M. Small. Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters*, 96(23):238701, June 2006. doi: 10.1103/PhysRevLett.96.238701. URL http://link.aps.org/doi/10.1103/PhysRevLett.96.238701.

J. Zhang, J. Sun, X. Luo, K. Zhang, T. Nakamura, and M. Small. Characterizing pseudoperiodic time series through the complex network approach. *Physica D: Nonlinear Phenomena*, 237 (22):28562865, 2008.