Christian André Andresen

Properties of fracture networks and other network systems

Thesis for the degree of Philosophiae Doctor

Trondheim, November 2008

Norwegian University of Science and Technology
Faculty of Natural Sciences and Technology
Department of Physics

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Natural Sciences and Technology
Department of Physics

© Christian André Andresen

ISBN 978-82-471-1376-9 (printed ver.)
ISBN 978-82-471-1377-6 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2009:5

Printed by NTNU-trykk

Abstract

This thesis is submitted as partial fulfillment of the requirements for the degree of doctor philosophiae and is a result of four years of research at the Department of Physics, the Norwegian University of Science and Technology.

The work presented represents applications of concepts from statistical physics and modern network theory. The systems studied vary and include fractured geological media, crumpled paper, political polling results and pure network theory. In the introductory chapters I familiarise the reader with the main concepts that are applied in the papers at the end. A total of five papers are included.

Paper I [1] : The network formed by ridges in a straightened sheet of crumpled paper is studied using a laser profilometer. Square sheets of paper were crumpled into balls, unfolded and their height profiles measured. From these profiles the ridges formed during the crumpling were extracted and viewed as networks. Nodes were defined as intersections between ridges, and links as the various ridges connecting the nodes. Several network and spatial properties have been investigated such as the ridge length distribution, the degree distribution and the facet area distribution. The tail of the ridge length distribution was found to follow a power-law whereas the shorter ridges followed a log-normal distribution. The degree distribution was found to have an exponentially decaying tail, and the degree correlation was found to be disassortative. The facets created by the ridges and the Voronoi diagram formed by the nodes have also been investigated.

Paper II [2] : In this paper we present a generalised version of the classical cluster coefficient, which can also be applied to networks with directed links. This generalisation takes into account more than the immediate nearest neighbours, giving more detailed information about the network structure than the classical version. The introduced concept is compared to earlier generalisation attempts, and it is applied to the directed protein interaction network of the *S. cerevisiae* yeast cell.

Paper III [3] : This paper describes the application of statistical methods to political polling data in order to look for correlations and memory effects. We

propose measures for quantifying the political memory using the correlation function and scaling analysis. These methods reveal time correlations and self-affine scaling properties. We have applied these measures to polling data from Norway. Power-law dependencies have been found between correlation measures and party size, and different scaling behaviour has been found for large and small parties.

Paper IV [4] : Fracture data from eight granite outcrops from the Laxemar and Simpevarp areas in south-east Sweden have been investigated as complex networks. For each outcrop a network was generated by viewing each fracture as a node, and linking it to any other fractures that it intersects. We show that the networks have complex behaviour with a broad degree distribution. The clustering coefficients and efficiencies are compared with values for rewired and random versions of the networks, and on this basis we have classified the networks as small-world networks. The degree-degree correlations of the networks indicates that they are disassortative.

Paper V [5] : In this paper we present a modern complex network (Graph theory) analysis of a discrete fracture network (DFN) model in two dimensions aimed at mimicking geological fracture networks. The main feature of the DFN model is the power-law fracture length distribution and the fractally distributed fracture centres. Several key network properties have been analysed such as degree distribution, degree correlations, clustering and efficiency. It has been shown that the degree distribution follows a power-law, $p(k) \propto k^{-\alpha_k}$ where α_k is dependent on the fracture length power-law exponent α_l, and insensitive to the fractal dimension D_2 of the fracture centre distribution. The nodes of the network are more likely to link to nodes of similar degree and the networks are therefore assortative. Both the clustering and the efficiency of the network have been shown to depend on the fracture length distribution, with an increase of both clustering and efficiency for higher α_l. Finally the networks are classified as "small-world" networks, giving important implications for the transport and stability properties of the networks.

Acknowledgements

I hear and I forget
I see and I remember
I do and I understand
Confucius (551–479 BC)

I do not think that any candidate that starts a PhD has an understanding of what he or she has started. I have experienced the last four years working towards my PhD as both extremely rewarding and extremely frustrating. It took a considerable amount of time before I realised that doing a PhD is just as much an education of me as it is a production of the papers. In retrospect, I'm grateful for the experience, and I feel that I have developed an enthusiasm for physics combined with the needed skepticism in addition to the technical skills that come with the work.

On the top of my thank-you list is my supervisor Alex Hansen. I would like to thank him for giving me the opportunity to do a PhD, and his support over the last four years. As all PhD-students I have had times where the problems seemed too hard and the motivation was suffering. To have such an inspirational and engaged supervisor has sometimes, honestly, been the only reason why I have carried on. Not a single time have I left a meeting with Alex and not felt more motivated and privileged to work on a PhD. I would also like to thank Alex for the opportunity to travel frequently to various events and meet the scientific community. Among the most rewarding trips was my stay in Strasbourg, the two trips to the SigmaPhi conference on Crete, the 2007 APS March-meeting in Denver, the Computational summer-school at the *Forschung Sentrum Jülich* and the conference on mathematics in geophysics on Svalbard. In total I have had the pleassure to give more than a dozen talks on various topics at different conferences and meetings.

I would also like to thank my fellow PhD-students in the Complex group that have provided a unique social and academic environment over the last few years. Especially I would like to thank Henning F. Hansen for the great effort he put in

to our common papers, and for this wizardry in the kitchen producing unspeakable pleasures for our cake lunches. Also Jan Øystein Bakke, Thomas Ramstad and Henrik Tollefsen have contributed to the work environment at the Complex group.

In addition to the Comlex group I would like to thank the other PhD-students that have helped supply me with endless amounts of coffee and discussions. These include Frantz Stabo-Eeg, Jan Petter Morten, Bjarte G. B. Solheim, Kjetil Børkje, Martin S. Grønsleth, Alex Klein-Paste, Per Kristian Hove, Zbigniew Rozynek, Asle H. Vaskinn, Eskil K. Dahl, Hilde Sand, Justin Wells, Lars Erik Walle, Steinar Kragset, Jo Smiseth, Eivind Smørgrav and Stein Olav Skrøvseth. Without whom my days at the office would have been bleak indeed.

Two of my papers would not have been possible without the kind help from Jean Schmittbuhl, who let me stay at his group in Strasbourg and use his lab there, and Philippe Davy that welcomed me in Rennes and kindly provided us with outcrop data and insights. I thank you both.

The work presented in this thesis was funded by the VISTA foundation that is a collaboration between Statoil (now StatoilHydro) and The Norwegian Academy for Science and Letters. I kindly thank them for their support. During my time as a PhD-student I have also had the opportunity to teach 25% of my time for the Department of Physics at NTNU. This provided me with an extra year for my studies.

I would also like to thank my family for their interest in my work and supportive conversations. Finally I would like to thank Marte Pernille Hatlo for being the most important person in my life. Her help, understanding and support is invaluable, and all of my efforts are futile and pointless without her.

Christian André Andresen

Trondheim, Norway
November 2008

List of papers

Paper I, Reference [1]: C. A. Andresen, A. Hansen, and J. Schmittbuhl,
Ridge Network in Crumpled Paper,
Phys. Rev. E **76**, 026108 (2007).

Paper II, Reference [2]: H. F. Hansen, C. A. Andresen, and A. Hansen,
A quantitative measure for path structures of complex networks,
Europhysics Letters **78**, 48005 (2007).

Paper III, Reference [3]: C. A. Andresen, H. F. Hansen, A. Hansen,
G. L. Vasconcelos, and J. S. Andrade Jr,
Correlations between political party size and voter memory:
A statistical analysis of opinion polls,
International Journal of Modern Physics C **19**(11), 1-11, 2008.

Paper IV, Reference [4]: C. A. Andresen, R. Le Goc, P. Davy, and A. Hansen,
Network analysis of outcrop data,
Submitted to Journal of Geophysical Research - Solid Earth.

Paper V, Reference [5]: C. A. Andresen, and A. Hansen,
Network properties of Discrete Fracture Network Model, To be submitted.

My contribution to the papers

Paper I: *Ridge Network in Crumpled Paper.* I performed the experiment, analysed the data, made the figures and wrote the paper all in collaboration with Alex Hansen and Jean Schmittbuhl.

Paper II: *A quantitative measure for path structures of complex networks.* In close collaboration with Henning Frydenlund Hansen I developed the idea, performed the calculations and made the figures. We wrote the paper in close collaboration with Alex Hansen.

Paper III: *Correlations between political party size and voter memory: A statistical analysis of opinion polls.* In close collaboration with Henning Frydenlund Hansen I collected the data, performed the calculations and made the figures. Moreover, we wrote the paper in collaboration with Alex Hansen, Giovani L. Vasconcelos and José Soares de Andrade Jr.

Paper IV: *Network analysis of outcrop data.* The linearised fracture data was supplied by Romain Le Goc and Philippe Davy. I performed the network analysis, made the figures and wrote the paper in collaboration with Alex Hansen and Philippe Davy.

Paper V: *Network properties of Discrete Fracture Network Model.* I implemented the model, performed the analysis, made the figures and wrote the paper all in collaboration with Alex Hansen.

Contents

1 Introduction **1**

2 Concepts from Statistical Physics **3**
2.1 Scale-free distributions and Power-laws . . . 3
2.2 Fractals . . . 7
2.3 Self-similarity and self-affinity . . . 10
2.3.1 Box-counting . . . 12
2.3.2 Power spectrum density . . . 12
2.3.3 Detrended fluctuation analysis . . . 13
2.3.4 Averaged wavelet coefficients . . . 13

3 Networks **15**
3.1 Basic network concepts . . . 16
3.1.1 Degree and strength distributions . . . 18
3.1.2 Diameter, Clustering and Efficiency . . . 19
3.2 Small-World networks . . . 20
3.3 Randomised and rewired networks . . . 23
3.4 Degree-degree correlations . . . 23
3.5 Ridge network in crumpled paper . . . 24
3.6 Path structures . . . 25

4 Geological Media **27**
4.1 The complexity of geological media . . . 27
4.2 Fracture modelling, brief history . . . 30
4.3 A discrete fracture network model . . . 32
4.4 Comparison between model and real-world results . . . 35
4.5 Thoughts for future work . . . 37

5 3D fracture project **39**
5.1 Purpose of experiment . . . 39
5.2 Choice of material and sample preparation . . . 40
5.3 Fracture production and AE results . . . 42
5.4 Missing fractures . . . 43

1 Introduction

> Science is facts;
> just as houses are made of stones,
> so is science made of facts;
> but a pile of stones is not a house,
> and a collection of facts is not necessarily science.
>
> *Henri Poincaré* (1854 – 1912)

The following chapters will provide the reader with an introduction to the theory and the setting that have motivated the papers attached at the end. This introduction is not meant to be a full review of any of the topics that are discussed, but to give readers not familiar with these topics the essentials needed to understand the papers. I have described not only the technical aspects, but also the setting and motivation for the work. In some cases I have summarised the main results from our papers and put these into context.

In all the chapters I have provided examples and references that I hope will be of interest to all readers, both novice and expert. For any reader who wishes to gain a deeper understanding of any of the fields I have provided references to some insightful review articles and books at the beginning of each chapter.

The papers presented at the end of this thesis span a range of topics loosely connected. The first paper [1] was the result of a training project where we wanted to investigate the ridge structure of crumpled paper. Crumpled paper is a much used analogy for many thin plate deformation problems, and the network formed by the ridges produced during crumpling was investigated. We proposed a novel power-law ridge length distribution, and confirmed the self-affine[1] nature of the height profile of the unfolded crumpled sheets.

The second paper [2] describes a new measure for the path-structure in complex networks. This work was motivated by the fact that the currently established

[1] See Section 2.3.

measures are either very local (i.e. the clustering coefficient) or very global (i.e. efficiency). Our measure for describing the path-structure of complex networks describes quantitatively the neighbourhood around each node including more than just the nearest neighbours. This gives new and very descriptive information concerning the structure of the network. The traditional clustering coefficient can be viewed as a special case of our novel measure.

Our third paper [3] introduces novel applications of concepts from statistical physics to analyse political polling data. Here we show how the auto-correlation function and scaling analysis can describe memory effects in the voter base for political parties. We have found a qualitative difference in the behaviour of small and large parties, and a power-law correspondence between party size and various correlation measures.

The fourth paper [4] presents a novel network analysis of the fracture traces from a selection of outcrops. A network definition is introduced and applied to the outcrop traces. The networks are shown to have a broad degree-distribution and large global and local connectivity. They are classified as *Small-World* networks giving important implications for their transport properties.

The fifth and final paper [5] applies the above mentioned network analysis to fractures generated by a two-dimensional discrete fracture network model. Using parameters motivated by real-world outcrop studies we have shown that the model reproduces both qualitatively and quantitatively many of the measures found for real-world outcrops. This includes the *Small-World* network classification and degree distribution. The model however shows a qualitatively different behaviour in the degree-degree behaviour.

2 Statistical concepts

> There are three kinds of lies:
> lies, damned lies and statistics.
>
> *Mark Twain* (1835 – 1910)

In this chapter I will introduce some of the statistical concepts that have been central to the papers in this thesis. As mentioned in the Introduction this is in no way an attempt to give an exhaustive record of the concepts discussed, but an introduction aimed at giving a reader not familiar with the field of statistical physics a brief overview and maybe inspire him/her to pursue the references. A reader familiar with the field could turn to the next chapter.

First, I will discuss the concept of scale-free distributions and their relation to power-laws. Thereafter I will briefly discuss fractal systems, and introduce some of the most central measures used to describe them. Finally, I will introduce the concept of roughness, and describe some of the standard methods of detecting the roughness exponent.

2.1 Scale-free distributions and Power-laws

Consider a sizeable population, perhaps the adult inhabitants of Norway, and think of the height and weight of these individuals. With little problem you would be able to estimate the average height and weight of this population well within a factor of two. If you were to estimate the average wealth or income of the same population it would not be as easy to come within a factor of two of the real value. There are no 100 m high or 1000 kg heavy persons to significantly influence the average of the weight and height distributions[1], but there are people with income and wealth many orders of magnitude larger than the median value found

[1] In fact even one person 100 m high or 1000 kg heavy would not significantly influence the average values.

in the population. These rare, but not negligibly few, large values significantly influences the distribution. Mathematically one finds that the height and weight distributions follows a normal or gaussian distribution, whereas the income and wealth distributions are power-laws [2].

In Figure 2.1 I have plotted two qualitatively different distributions. In the left hand panel is the height distribution of Norwegian soldiers for the last decade, and in the middle panel the wealth distribution for Norwegian tax-payers for a recent time interval[3]. Note that the wealth distribution is not very well represented in the middle plot and is much better represented with logarithmic axis as in the right hand panel.

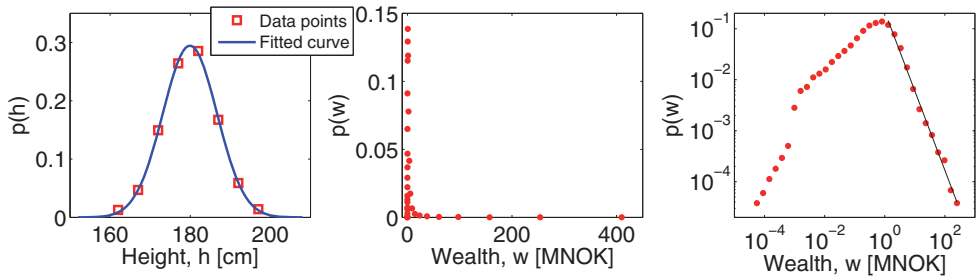


Figure 2.1: *a) Height distribution of Norwegian soldiers for the period 1998 – 2006. b) Wealth distribution of Norwegian tax-payers for the period 1986 – 2003. c) same as in plot 'b)' but with logarithmic axis.*

The values in the high end of the wealth distribution show a linear behaviour in the log-log plot, and this is the trademark of a power-law. A distribution with this behaviour is said to have a *fat tail*. Statistical properties of systems exhibiting such distributions are often completely controlled by the few but large data points in their tail. This motivates the saying *It's all in the tail*. A *power-law* distribution of a quantity x is given by

$$p(x) = Cx^{-\alpha} \tag{2.1}$$

where α is known as the power-law exponent and C is a normalising constant.

[2]These distributions are not power-laws over the whole range of values, but only in the large value end of the range known as the *tail*.

[3]Soldier height data from Statistics Norway (Statistisk sentral byrå) [6], and wealth distribution data from Hansen [7].

Note that there is no parameter which sets the typical or characteristic scale in Eq. 2.1 as there is in the normal, log-normal, poissonian or exponential distributions. That is because the function is scale-free, there is no typical scale attached to the distribution itself. Any scale information will have to be introduced via the bounds of the function. The power-law is the only function that obeys the relation $ap(x) = p(bx)$ for constant a and b. Much has been written about power-laws and a nice review is presented by Newman in Ref. [8]. Many systems have been shown to have properties following such a law for the whole or part of their distribution. These properties include the size of cities, length of rivers, number of books sold, frequency of words, fracture lengths and e-mail lengths to name some. A more comprehensive list of systems displaying scaling behaviour is given by Wiesenfeld [9].

The moments of a power-law are either divergent or convergent in the limit of infinite systems depending on the exponent α. Of special interest is the first order momentum or average value of the probability distribution $p(x)$ for some x in the range $x_{min} \leq x \leq x_{max}$ given by

$$\langle x \rangle = \int_{-\infty}^{\infty} xp(x)dx = C \int_{x_{min}}^{x_{max}} x^{1-\alpha}dx = C(x_{max}^{2-\alpha} - x_{min}^{2-\alpha}). \tag{2.2}$$

If we assume $\alpha > 0$ so that larger x are less likely, then $< x >$ is dominated by the x_{max} term for $\alpha < 2$ and by the x_{min} term for $\alpha > 2$. For most real systems we do not have the whole probability distribution, but only a limited interval. For $\alpha > 0$ we will then sample the low range portion of the distribution much better than the high range portion and therefore get a much more accurate value for x_{min} than for x_{max}. If $0 \leq \alpha \leq 2$ the value of $\langle x \rangle$ will then be completely dominated by the largest x in our selection, and therefore vary from selection to selection. Speaking of an average value for an ensemble in this range is therefore meaningless even though the average value always exists for a finite set of values.

In the limit of infinitely large sets of values the mean will allways diverge for $\alpha < 2$. For higher moments of x given by $\langle x^n \rangle = \int x^n p(x) \mathrm{d}x$, these will converge only if $\alpha > n + 1$.

Various methods of measuring the exponent α are also interesting. Often there are much noise in the data and a linear fit to the plot of $\log(p(x))$ vs $\log(x)$ is hard or associated with much uncertainty. One way of achieving a more accurate estimate is to use logarithmic binning. By doing this you increase the binning size as x increases in order to compensate for the lower frequency of samples in this range. This will lower the fluctuations in the tail of the distribution. Care must

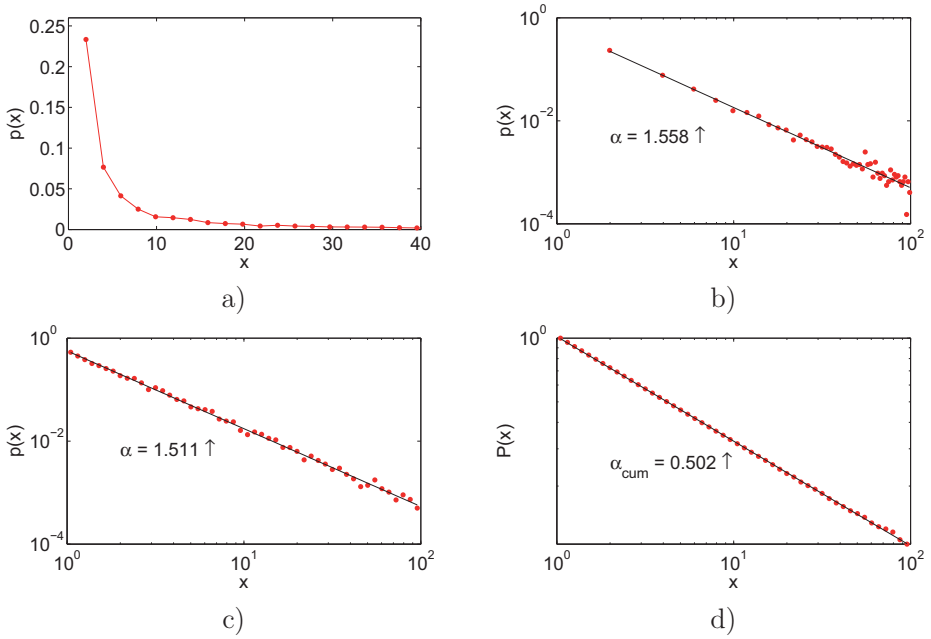


Figure 2.2: *a) Probability distribution for 10^5 realisations of a random variable x in the interval $1 \leq x \leq 100$ distributed according to a power-law with exponent $\alpha = 1.5$, equal binning size. b) The same distribution plotted with logarithmic axis. c) Same probability function binned with logarithmic binning size. c) Cumulative distribution with logarithmic bin sizes.*

be taken to account for the increasing binning size while plotting the histogram. Another method is to look at the cumulative distribution $P(x)$ given by

$$P(x) = \int_x^\infty p(x')dx' = \frac{C}{\alpha - 1} x^{-(\alpha-1)} = C' x^{\alpha_{cum}}. \tag{2.3}$$

Although much information is lost while constructing the cumulative distribution[4] it may help to form a smoother tail that is easier to analyse. Note that the slope of the linear fit changes to $\alpha_{cum} = \alpha - 1$. The various methods and their effect are visualised in Figure 2.2 for 10^5 realisations of a random variable distributed according to a power-law with exponent $\alpha = 1.5$.

Figure 2.2 shows that the accuracy of the estimated exponent generally increases

[4]This is detailed information about the behaviour of the probability distribution itself.

when using logarithmic binning and looking at the cumulative distribution. However care should be taken as these methods are in effect averaging methods and thus conceal information about the distribution that could be important. Another method for extracting the exponent is given by Newman in [8] where all n samples x_i over a given minimum x_{min} are taken into account

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln\left(\frac{x_i}{x_{min}}\right)\right]^{-1}. \tag{2.4}$$

For the distribution presented in Figure 2.2 this gives an exponent $\alpha = 1.67$ which is worse than any of the linear fittings.

2.2 Fractals

The term fractals was introduced by B. B. Mandelbrot. The publication of his books *Fractals: Form, Chance and Dimension* [10] (1977) and *The Fractal Geometry of Nature* [11] (1982) helped establish the field. A vast literature has emerged exploring this novel way of looking at Nature. Mandelbrot himself defined a fractal as [12],

A fractal is a shape made of parts similar to the whole in some way

A part of the whole can be called similar because it is identical to the whole, as in *deterministic fractals*, or statistically similar [5] as in *stochastic fractals*. Although a part of a line is similar to the whole it is not a fractal. In order to qualify the shape must have a non-integer fractal dimension as described below. This implies that pure fractals have structure on all scales, and are then perceived as rough or fragmented. In Nature there are many examples of *quasi-fractals* displaying fractal properties on a limited range of scales. These systems include clouds, snow flakes, crystals, coastlines, river networks and mountain ranges. The abundance of fractals in nature also includes fracture networks as we'll see in Chapter 4.

A fractal system is described[6] by its *fractal dimension*, D, that describes how the fractal spreads in the space that it is embedded in. As noted above a entity is only fractal if it has a non-integer fractal dimension. Various definitions for the fractal dimension have arisen, but the Hausdorff-Besicovitch dimension[7] is often

[5] Having the same statistical properties.
[6] Generally only true for mono-fractals.
[7] Called the Hausdorff dimension for short.

taken to be the most important [13] of these. The definition is based on a measure M_d on a set S of points in space. The set is covered by a test function[8] $h(\delta)$ of linear size δ and assosiated with a geometrical factor $\gamma(d)$ such that $h(\delta) = \gamma(d)\delta^d$. The measure is then given by the sum of test functions needed to cover the whole set, $M_d = \sum h(\delta)$. Generally as $\delta \to 0$ the measure M_d either goes to zero or infinity depending on d, known as the dimension of the measure. The critical dimension where the measure changes from zero to infinity is defined as the Hausdorff dimension D,

$$M_d = \sum \gamma(d)\delta^d = \gamma(d)N(\delta)\delta^d \quad \overset{\delta\to 0}{\to} \quad \begin{cases} 0, & d > D \\ \infty, & d < D. \end{cases} \tag{2.5}$$

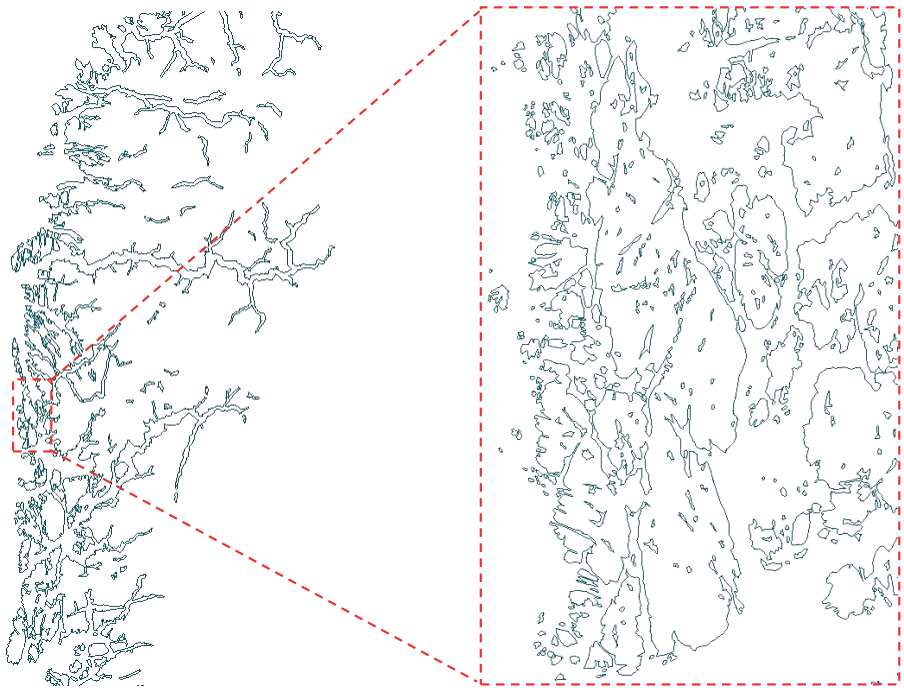
Here, if all the test functions are taken to be of the same size, $N(\delta)$ is the number of such funcions needed of linear size δ. This definition is very general and can be applied to any set. It should also be noted that it is a *local* definition that allows for a variation of the fractal dimension between various regions of the set. For a full mathematical description see Falconer [13]. Although it is mathematically convenient it is not computationally easy to implement and compute, and therefore the *box-counting dimension* is often preferred. For a nice discussion on fractal measures see the book *Fractals* by Feder [12].

The box-counting dimension is found by counting how many boxes $N(\delta)$ of size δ is needed to cover the set for various values of δ. For fractal shapes the two variables N and δ are realted by

$$N(\delta) \propto \frac{1}{\delta^D}, \tag{2.6}$$

where D is the box-counting dimension. D is easily found by a linear fit to the plot of $\log(N(\delta))$ vs $\log(\delta)$. This method has been refined and various implementation schemes have been developed. The measure is also easily generalised into a multi-fractal measure [14–16]. Note that the two definitions given in Eq. 2.5 and 2.6 do not necessarily give the same dimension. A nice example of a fractal shape is the coastline of Norway [12]. This is illustrated in Figure 2.3. The shape of the coast has been analysed with the box-counting technique and a fractal dimension of $D = 1.52$ has been found. A plot of $N(\delta)$ vs δ plotted with logarithmic axis can be seen in Figure 2.4.

[8] This test function may be a line, square, circle, ball, disk, cube ect, and the geometric factor associated with the test function is the pre factor of the length/area/volume occupied by the test function i.e. $\gamma = \pi/4$ for discs, $\pi/6$ for spheres, and 1 for cubes.



a) b)

Figure 2.3: *Outline of the coast of Norway at two different scales. Image a) is about 300 x 150 m^2 and image b) 30 x 15 m^2. Could you tell which was which scale if the topography of Norway was unknown to you? Map data: Statens kartverk (www.statkart.no)*

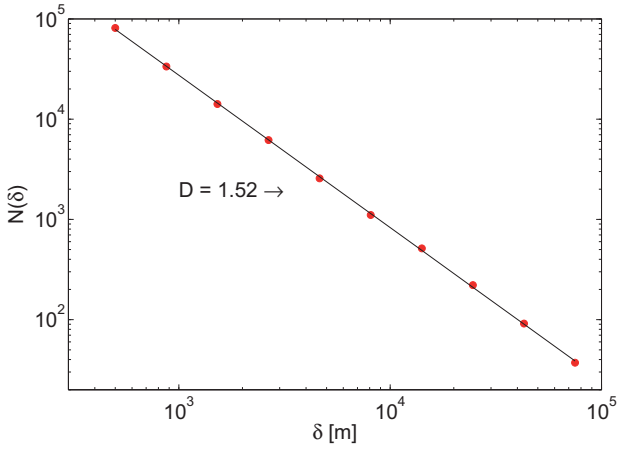


Figure 2.4: *Plot of number of boxes needed $N(\delta)$ vs box size δ for the coast of Norway, along with its linear fit in logarithmic axis.*

2.3 Self-similarity and self-affinity

The concept of fractallity is closely connected to the concepts of self-similarity and self-affinity. Imagine that you zoom in on a graph, $h(x)$, with the same factor λ in all directions, $h(x) \to \lambda h(\lambda x)$, and you find that the overall structure of the graph does not change, then the graph is *self-similar*. This is isotropic scaling, and the images in Figure 2.3 was scaled this way. The distinction between deterministic (exact replica) and stochastic (statistically equal) similarity applies as it does to fractallity. Now imagine that you would have to zoom with a different ratio in one of the directions in order to produce a similar graph. This is anisotropic scaling and shapes displaying this property are called *self-affine* shapes. Mathematically this can be expressed as [17]

$$h(x) = \lambda^{-H} h(\lambda x) \tag{2.7}$$

where λ is the scaling ratio, and H is the *Hurst exponent*[9]. The simplest self-affine structure is the discrete *random walk* where for each time step the walker takes one step up or down with equal probability. The random walk has no memory and it has $H = 1/2$. In Figure 2.3 I have plotted a random walk with 10^5 steps, I have

[9] Also known as the *roughness exponent*.

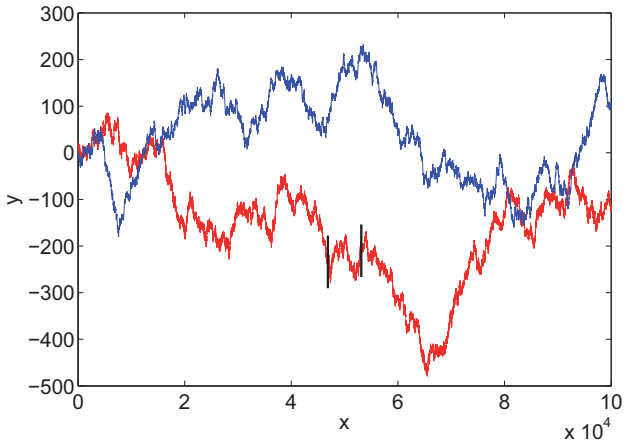


Figure 2.5: *Random walk with $N = 10^5$ steps. The red curve is the full random walk, and the blue curve is a scaled version of the section between the two black lines crossing the red curve. The blue curve has been scaled with a ratio $\lambda = 16$ in the x-direction, and $\lambda^{0.5} = 4$ in the y-direction.*

also plotted a scaled version of the middle portion of the walk. It is not possible to tell from their statistical properties which of the two graphs is a magnification of the other, and they look similar to the naked eye.

The Hurst exponent describes the auto-correlation of the graph in question. For a graph to be called self-affine it must have $0 \leq H \leq 1$. If $H > 1$ the graph is not asymptotically flat, and grows without bounds, and if $H < 0$ then the variance $\langle z^2 \rangle$ approaches a constant larger than zero, and the graph is known as a *fractional noise* [18]. For H in the interval $0 \leq H < 1/2$ the graph is anti-correlated and mean reverting, and for $1/2 < H \leq 1$ the graph is positively correlated and trend reinforcing. Examples of the three different behaviours can be seen in Figure 2.6 where I have plotted three different profiles with $H = 0.1$, 0.5 and 0.9. As the figure shows the profile with the lowest H seems to be rougher and spikier than the one with larger H.

Many different methods have been developed to determine the Hurst exponent of profiles and surfaces. Various studies [19, 20] have estimated the accuracy of these methods, and there have been shown to be strengths and weaknesses with all of them. In general it is preferable to use more than one method when estimating the Hurst exponent. The discrepancy between different methods gives a good indication of the accuracy with which the Hurst exponent has been estimated.

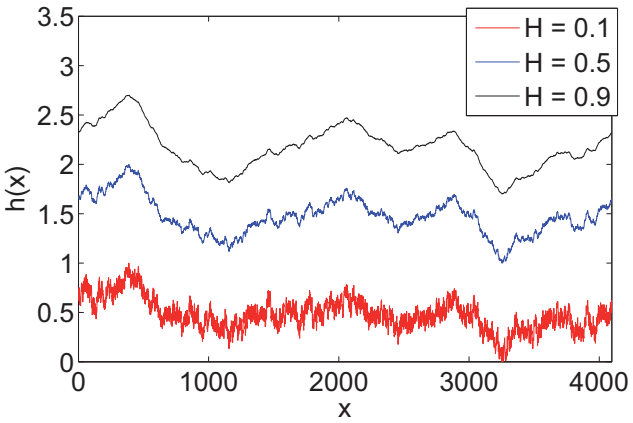


Figure 2.6: *Plot of three different profiles with different Hurst exponent.*

Below I will give a short introduction to some of the most important methods.

2.3.1 Box-counting

The box-counting method can measure the box dimension D_B of a profile $h(x)$. In general it is necessary to use rectangular and not square boxes to measure the fractal properties of a self-affine profile. For box sizes small compared to the variation of the profile the method will measure a local box dimension given by $D_B = 2 - H$ for a one-dimensional profile in two-dimensional space [12, 20]. Note that for larger scales when the box-sizes are large compared to the variations of $h(x)$ the method will report a fractal dimension of 1 since the function no longer is fractal at these scales. For such large scales the relation $D_B = 2 - H$ no longer holds.

2.3.2 Power spectrum density

The *Power Spectrum Density* (PSD) method is a Fourier based method where the spectral density of the profile is calculated. For a one-dimensional self-affine profile in two-dimensional space the power spectrum, $S(f)$, is given by [13]

$$S(f) \propto f^{-1-2H}, \tag{2.8}$$

where f is the frequency. The PSD method can be generalised to higher dimensions analysing shapes such as two-dimensional surfaces in three dimensions without having to make one-dimensional cross-sections of the surface. The method is reported [19] to perform well for relatively large system sizes under a range of conditions.

2.3.3 Detrended fluctuation analysis

The *Detrended Fluctuation Analysis* method [21] (DFA) measures the fluctuations, $F(s)$, of a signal around a polynomial trend for a range of scales s. The order of the polynomial used varies, and the method is often referred to as DFAn where n describes the order used. The order should be high enough so that any global trends in the signal are removed, and the analysis becomes insensitive to the choice of order. For a stationary signal with self-affine properties the fluctuations will scale as

$$F_2(s) \propto s^H, \tag{2.9}$$

where the subscript 2 indicates that the second moment of the fluctuations are considered. The method is easily generalised to a multifractal formulation (MF-DFAn) where other moments of the fluctuations are considered, and has been reported to give robust results for many systems with varying noise.

2.3.4 Averaged wavelet coefficients

The *Average Wavelet Coefficient* [22] (AWC) method is based on the wavelet transform of the function $h(x)$. The arithmetic mean of all the wavelet coefficients $W[h](a)$ for each scale a is calculated, and it has been found that for self-affine functions there is a scaling relation

$$W[h](\lambda a) \simeq \lambda^{(1/2)+H} W[h](a), \tag{2.10}$$

where λ is a rescaling factor. The method has been reported to perform well for small and sparse data sets, where other methods are plagued by noisy signals.

3 Networks

No man is an island, entire of itself...
John Donne (1572 – 1631)

Over the last few decades we have seen a tremendous development in our understanding of networks and network structures [23–27]. This development is largely motivated by the emergence of more detailed complex network[1] data and the numerical capabilities to study them. Network theory originates from the mathematical discipline of *graph theory* first discussed in 1735 by Leonard Euler [28] as he looked at the bridge structure between various islands in the city of Köningsberg. This study captures the essence of network theory as it describe the interactions between a set of connected entities.

In the following chapter I will introduce the basic concepts used to describe networks, and discuss various measures that have been constructed to classify their type and quantify their interactions. These measures enable the discussion of the Small-World concept, and this concept is central in the following chapters. Again this is not an exhaustive record, nor a reflection of the status of the research front in the field. At the end of this chapter I discuss the motivation for and context of the two first papers attached at the end of this thesis.

The chapter provides a foundation for further chapters, and puts our article on path structures in complex networks [2] into context.

[1] Specially concerning the Internet.

3.1 Basic network concepts

A *network*[2] is a set of entities named *nodes*[3] that are strongly or loosely connected in some sense via connections called *links*[4]. The nodes and links may be of the same type or there might be different types of each in the network as illustrated in Figure 3.1.

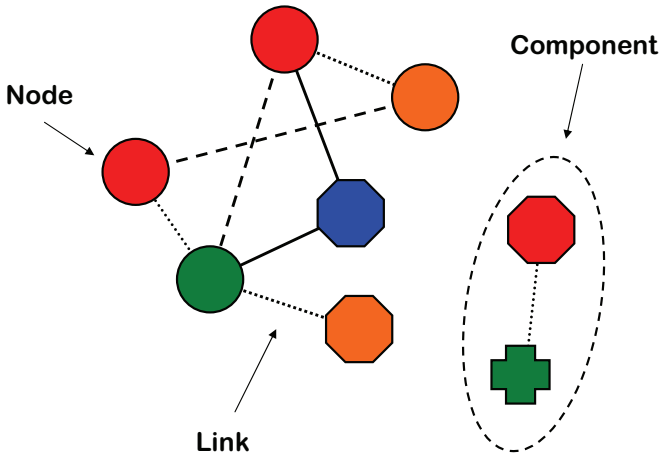


Figure 3.1: *Example network showing various types of nodes and links, and the concept of components.*

Two nodes that share a link are said to be connected, and a *component* of a network is a set of nodes between which you can traverse[5] either directly or via other nodes. The network in Figure 3.1 has two components where one is larger than the other. Components that are considerably larger than the other components in the network are called *giant components*, and in many studies the nodes not belonging to this component are not considered, in order to simplify the analysis. Networks where it is possible to traverse from all nodes to all other nodes have only one component and are called *fully connected*. The most common network with different types of nodes are the *bipartite* networks that have two types of nodes, and links only between unequal nodes.

[2] Also known as a *graph*.
[3] Also referred to as *vertices* or *actors*.
[4] Also referred to as *edges* or *bonds*.
[5] You can traverse from a node to all its connected nodes.

The links in a network may also have various properties, the most common of which are directionality and weight. Networks where the links have directionality are called *directed networks*, and networks where the links have weight are called *weighted networks*.

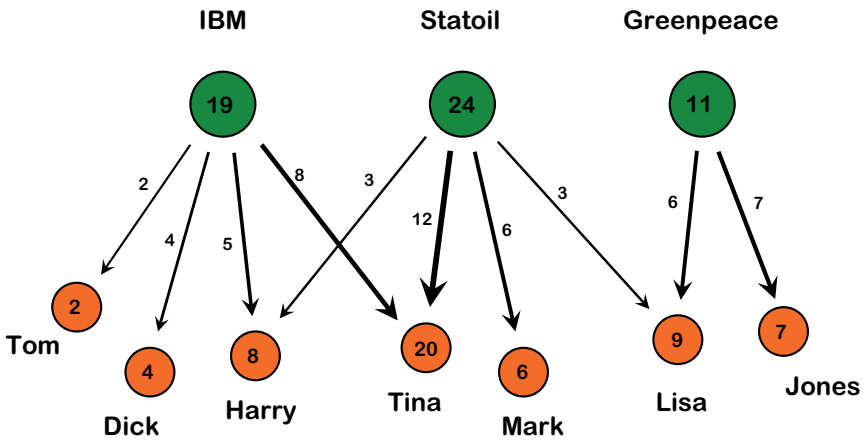


Figure 3.2: *Example network showing a weighted and directed bipartite network whith two types of nodes; employers and employees. The number listed at the links are the weight of the links, and the number listed for the nodes are the strength of the nodes.*

An example of a weighted, directed bipartite network is a network of employers and employees connected by their salary payments. We assume that no employer receives salary and that no employee pays wages. In Figure 3.2 I have displayed an example network to illustrate some basic concepts. This is a weighted, directed bipartite network of 3 employers and 7 employees. The arrows indicates payments from an employer to an employee, and the numbers associated with each arrow indicates the size of the payment that we shall call the *weight*, w, of the link.

One of the most fundamental properties of any node is the *degree* [6], k, of the node. The degree is the number of nodes that a given node is connected to. For our example network in Figure 3.2 the degree of the nodees *IBM* and *Statoil* is both four since they are connected to four other nodes, whereas the degree of the node *Tina* is two. For weighted networks nodes may be characterised by their *strength*, s, given [25] by

[6]This number is known as the *coordination number* in some disciplines.

$$s_i = \sum_{j \in NN} w_{ij}, \tag{3.1}$$

where w_{ij} is the weight of the link between nodes i and j, and j runs over all the nearest neighbours NN of node i. An unweighted network can be seen as a weighted network where all the weighs $w_{ij} = 1$ for all ij, then the strength is equal to the degree for all nodes.

3.1.1 Degree and strength distributions

The distribution of degrees $p(k)$ and/or strengths $p(s)$ is very descriptive of the complexity of a given network. A regular network, as for example a triangular lattice, will have a very narrow *degree distribution* $p(k) = \delta(k - 6)$ as each node[7] has six neighbours and therefore $k_i = 6$ for all i. Purely random unweighted and undirected networks where each pair of nodes are liked with a given probability p have a degree distribution given [24] by

$$p(k) = C^k_{N-1} p^k (1 - p)^{N-1-k}, \tag{3.2}$$

where N is the number of nodes in the network and $C^k_{N-1} = \binom{N-1}{k}$ is the binomial coefficient. In the limit of large N this approaches the Poisson distribution $p(k) = e^{-\langle k \rangle} \langle k \rangle^k / k!$ which is a well confined distribution with a pronounced peak. On the other hand many natural complex networks show a broad degree distribution that span many orders of magnitude. A special case of such networks having a degree distribution given by a power-law $p(k) \propto k^{-\beta}$ are called *scale-free networks*. Such a distribution indicates that the presence of nodes with a very high degree or strength is not neglibly small and that the presence of such nodes affects the behaviour of the network. Nodes that have a degree or strength considerably larger than the average or median node in the network are called *hubs*. Many systems, as diverse as international airport networks, citation networks, reference networks protein interaction networks and power-grid networks, show scale-free behaviour in their degree or strength distribution with a typical power-law exponent in the range $2 < \beta < 3$ [24, 25]. As we shall see in the following sections a broad degree distribution is connected to the concept of Small-World networks.

[7] Here a node is perceived as the intersection of links in the lattice.

3.1.2 Diameter, Clustering and Efficiency

In addition to the degree distribution it is also interesting to see how well the nodes interact both locally and globally in the network. Various measures for quantifying the interaction capabilities of networks have been introduced, including motives [29–31], community structures [32, 33] and weighted networks [34, 35], however the most prominent ones are the *clustering coefficient* for the local scale and the *diameter*, *characteristic path length* and the *efficiency* for the global scale. By local and global scale I mean the range of nodes included in the measure in terms of distance from a given node. Distance from a node is measured by the number of links that you have to traverse to move from one node to another, and is therefore a topological measure.

As mentioned above the clustering coefficient, C, is a measure of how well the nodes interact on a very local neighbour to neighbour scale. The measure is defined [25] as the mean of the local clustering coefficient, C_i, for each node

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \qquad\qquad C_i = \frac{2E_{\mathrm{nn},i}}{k_i(k_i-1)}, \tag{3.3}$$

where N is the number of nodes in the network, k_i is the degree of node i and $E_{\mathrm{nn},i}$ is the number of links between the nearest neighbours of node i. The local clustering coefficient can be seen as a measure of how many of your friends are friends with each other ($E_{\mathrm{nn},i}$) divided by how many could have been friends with each other ($k_i(k_i-1)/2$), and both C and C_i falls within the range $0 \leq C_{(i)} \leq 1$. A high C indicates a high degree of local interaction, and therefore the spreading of any quantity (abstract as information, or concrete as oil) is easy on this scale.

To discuss at the global measures I need to introduce the concept of a *geodesic path*, d_{ij}, which is the shortest path in terms of links that have to be transversed in order to go from node i to node j. Note that for undirected networks $d_{ij} = d_{ji}$ but for directed networks we generally have $d_{ij} \neq d_{ji}$. The *diameter*, D, of a network is the largest geodesic path in the network [25]

$$D = \max_i(\min_j(d_{ij})), \tag{3.4}$$

and hence describes how well the network is connected globally, and how easy or difficult it is to cross the network. If there is a small part of the network that has a three-structure[8] this will dominate the diameter measure and the real global

[8] A structure where there are no loops.

connectivity of the network will not be correctly reflected. To better probe the global connectivity the *characteristic path length*, L, was introduced [25]. The characteristic path length measures the arithmetic mean geodesic path

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij}, \tag{3.5}$$

and is a more robust measure than D. The *Small-World property* [9] is the property that the characteristic path length grows at most logarithmically with the number of nodes N in the network. This concept is illustrated in the famous *six degrees of separation* [36] expression stating that you are only six handshakes away from anybody in the world even though there are more than six billion people in the world.

For networks with more than one component there is however a problem. If a network is not fully connected there will be at least one $d_{ij} = \infty$, and hence $D = L = \infty$ for the entire network. As many interesting and complex networks have more than one component (and no giant component to look at) a better global measure is the *efficiency*, E, of the network. This measure describes the global transport properties of the network and is given [25] by the harmonic mean of the geodesic paths

$$E = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} \frac{1}{d_{ij}}. \tag{3.6}$$

The efficiency also lies in the interval $0 \leq E \leq 1$, and a high E indicates good global transport properties in the network.

3.2 Small-World networks

In their groundbreaking article in *Nature* in 1998 Watts and Strogatz [37] introduced the concept of *Small-World networks*. They defined such networks as networks with both large local connectivity, as measured by the clustering coefficient, and large global connectivity as measured by the characteristic path length.

[9] Note the difference between the *Small-World property* and the *Small-World class* of networks. In order to classify a network as belonging to the Small-World class it needs to have the Small-World property ensuring good global connectivity AND high clustering ensuring good local connectivity. This is described in more detail in Section 3.2.

They illustrated this by a model where a given number of nodes, N, are arranged in a circle, and every nodes was connected to the k nearest neighbours on both sides. A replication of their figure is shown in Figure 3.3. This model produces a regular network where each node is connected to $2k$ other nodes close by. This makes the networks very connected on a local scale, and gives it a large clustering coefficient. On the global scale however the network is not very well connected since you have to transverse a large number of links to move from one side of the circle to the other.

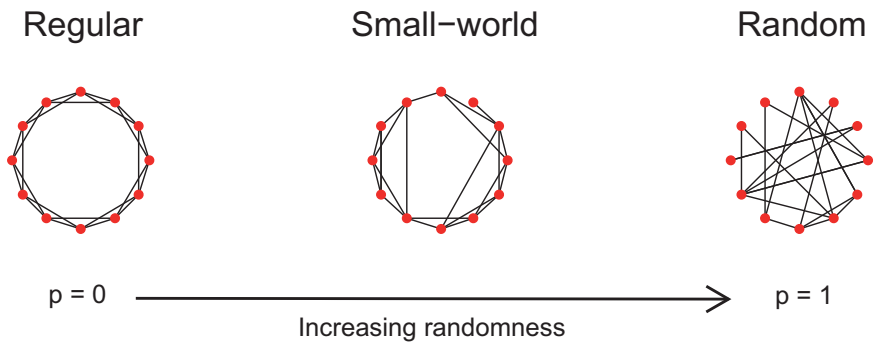


Figure 3.3: *Replication of figure plotted by Watts and Strogatz in their Nature article where they introduced the concept of Small-World networks. Here shown for $N = 12$ nodes connected to the $k = 2$ nearest neighbours.*

In their model Watts and Strogatz rerouted each link with a probability p. By rerouting a link they made it point to a randomly chosen node in the network, taking care not to duplicate any existing links. This produced long range links in the network that dramatically lowered the number of links needed to traverse form one side of the network to the other. For $p = 0$ the regular network is kept unchanged. For $p = 1$ all the links in the network are pointing to random nodes and the network is therefore a random network. Such networks are very well connected globally since there are many links across the network. However they are not well connected locally. Therefore they have a low clustering coefficient and a low characteristic path length.

For an interesting class on networks having $0 < p < 1$ a sufficiently large number of long range links is introduced to lower the characteristic path length significantly *and* still keep enough of the original links so that the clustering coefficient remains large. The function $C(p)$ and $L(p)$ vs p is reproduced in Figure 3.4. These intermediate values for p make the model produce Small-World networks.

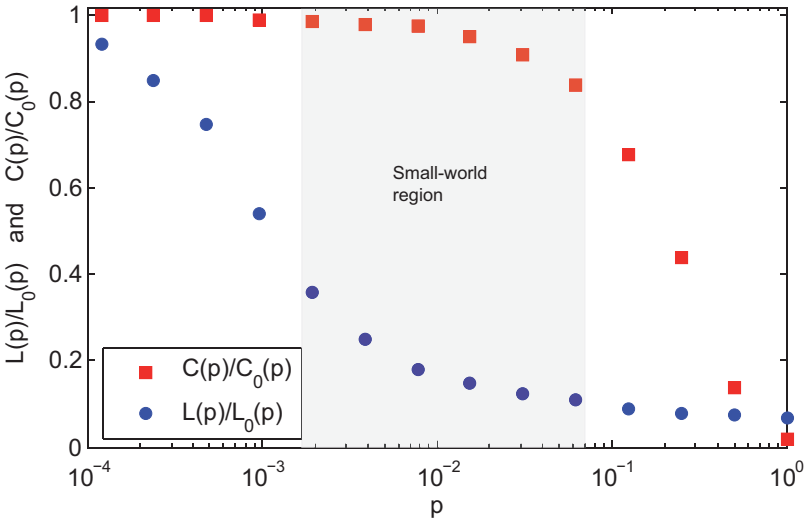


Figure 3.4: *Replication of figure plotted by Watts and Strogatz in their Nature article where they introduced the concept of Small-World networks. Both the clustering coefficient $C(p)$ and the characteristic path length $L(p)$ have been normalised by their values for the purely regular case $C_0(p) = C(p = 0)$ and $L_0(p) = L(p = 0)$ respectively. Note that for the shaded area in the middle of the figure the network have simultaneously a large C and a small L. The data is averaged over 100 realisations of networks with $N = 10^5$ nodes.*

Much has been written on the properties of Small-World networks [23, 26, 37–39], and they have proven to have remarkably good transport properties due to the combination of *both* good local connectivity *and* good global connectivity. They have also been shown to be surprisingly resilient to random removal of links or nodes for the same reasons. If a link or node were to be removed there would very likely be an alternative route with comparable transport properties. However a large group of Small-World networks are also scale-free networks where the hubs bind the network together and facilitate a majority of the transport. A targeted attack [40] disabling a portion of these hubs would seriously reduce the global connectivity, and these networks are therefore very vulnerable for such attacks.

For globally well connected networks one would expect both a low characteristic path length, since it is easy to traverse from one side of the network to the other, and a high efficiency since transport in such a network is easy. As pointed out above many interesting networks have more than one component causing the char-

acteristic path length to diverge. However for reasonably well behaved networks a small L is comparable to a large E. Note that the minimum value of d_{ij} is one, and therefore the harmonic measure E has no divergent terms. In such cases it is therefore possible to probe the local connectivity by looking at the clustering coefficient, and determine the global connectivity by looking at the efficiency.

3.3 Randomised and rewired networks

In order to determine whether a given value for C or E is *large* or *small* it is necessary to relate the values for the given network to a comparable network, as no value is large or small by itself. Two useful versions that any network can be compared with is the *randomised* and *rewired* versions of itself.

For the randomised version all links in the networks are removed and then distributed randomly among the nodes taking care not to produce duplicated links. This changes the degree distribution of the network to that of a random network given in Eq. 3.2 and removes any degree-degree correlations. Such a network is expected to have a very low clustering coefficient since the probability that two nodes connected to a given node are connected to each other by chance is small in a large network. On the other hand they will also have a low diameter and characteristic path length due to the large number of long range links[10].

In the rewired case two pairs of connected nodes are selected at random and their links are interchanged so that two new pairs of connected nodes are produced. Care must be taken not to select pairs that are already interconnected, or duplicate present links. This process is repeated until all the links have been rewired. This procedure preserves the degree-distribution of the network but removes any degree-degree correlations that may be present.

3.4 Degree-degree correlations

In addition to the measures mentioned above the *degree-degree correlation* of the network is an interesting property. *Do nodes predominantly link to other nodes of the same degree or to nodes of higher/lower degree?* I will discuss two methods for studying such correlations.

[10]Here assuming that there are only one component in the network or there is a giant component that is investigated. This is a reasonable assumption when the average degree $\bar{k} > 1$.

The first method was proposed by Pastor-Satorras *et al.* [41, 42] and builds on calculating the average degree $\langle k_{\mathrm{nn}} \rangle$ for the nearest neighbours of all nodes of degree k. Plotting $\langle k_{\mathrm{nn}} \rangle$ vs. k reveals any correlations in the linking. A flat curve indicates no special structure in the degree-degree correlations, whereas an increasing or decreasing curve indicates an underlying structure.

The second method was proposed by Maslov and Sneppen [43,44] and they propose to plot a two-dimensional correlation matrix $C(k_1, k_2)$ given by

$$C(k_1, k_2) = \frac{P(k_1, k_2)}{P_r(k_1, k_2)}, \tag{3.7}$$

where $P(k_1, k_2)$ is the probability that a node of degree k_1 is linked to a node of degree k_2, and $P_r(k_1, k_2)$ is the same probability of a rewired version of the same network. A value $C(k_1, k_2) > 1$ indicates an over-representation of links between nodes of degree k_1 and k_2, whereas a value $C(k_1, k_2) < 1$ indicates an under-representation. Networks where small-degree nodes tend to link to high-degree nodes are called disassortative networks and are found to be abundant in biological and technical networks. On the other hand networks where nodes tend to link to other nodes of similar degree are called assortative networks and are abundant in social networks where hubs tend to link to other hubs. The origin of the difference in behaviour is still not fully understood [45].

In order to look at the statistical significance of the correlations found in $C(k_1, k_2)$ Maslov and Sneppen introduced the matrix $Z(k_1, k_2)$ given by

$$Z(k_1, k_2) = \frac{P(k_1, k_2) - P_r(k_1, k_2)}{\sigma_r(k_1, k_2)}, \tag{3.8}$$

where $\sigma_r(k_1, k_2)$ is the standard deviation of $P_r(k_1, k_2)$ for an ensemble of rewired networks. A value of $Z(k_1, k_2) \sim 0$ indicates that the difference between the actual and the rewired probability is small compared to the fluctuation in the rewired probability, and therefore the result is not statistical robust. A value $|Z(k_1, k_2)| \gg 0$ indicates that the difference is large compared to the fluctuations, and the result is robust for the region.

3.5 Ridge network in crumpled paper

A surprisingly rich literature [46–50] has emerged on studies dealing with crumpled paper. This is an interesting problem largely because it is an easily investigated

version of the more general thin plate deformation problem. The paper *Ridge network in crumpled paper* [1] describes a crumpling experiment that we have conducted where we extracted the ridge network that formed when sheets of paper where crumpled. The ridge networks were investigated using the above outlined modern network theory, and a rich behaviour were found.

3.6 Path structures

In our paper *A quantitative measure for path structures of complex networks* [2] we try to bridge the gap between the very local nearest neighbour measure of the clustering coefficient and the global measures of diameter, characteristic path length and efficiency. We ask the question *How many paths of length n from node i are there to nodes which also can be reached with a path of length m?* In order to quantify this we introduce the measure $C^r(m,n)$ given by

$$C^r(m,n) = \frac{1}{N}\sum_{i=1}^{N} C_i^r(m,n) \tag{3.9}$$

where

$$C_i^r(m,n) = \frac{p_i(m,n)}{\prod_{j=0}^{n-1}(N_i^r - j)} \tag{3.10}$$

where N_i^r is the number of nodes in the sub-network[11] of radius r around node i and $p_i(m,n)$ is the actual number of paths of length n to nodes that can also be reached by a m-path from node i. Here the product $\prod_{j=0}^{n-1}(N_i^r - j)$ is the maximum possible number of such paths. This measure reproduces the classical clustering coefficient in Eq. 3.3 for the case $C^1(1,2)$, and it is also bounded by the unit interval $0 \leq C^r(m,n) \leq 1$. The strength of the measure is that it takes into account more than just the nearest neighbours of the node, and gives detailed information on the more long-ranged structure of the networks.

[11] A sub-network of radius r around node i is the set of nodes and links that can be reached by traversing r links.

4 Geological Media

> Clouds are not spheres, mountains are not cones,
> coastlines are not circles, and bark is not smooth,
> nor does lightning travel in a straight line.
>
> *Benoit Mandelbrot* (1942 –)

The field of *geology* dates back centuries if not millennia [51]. Throughout time humans have always strived to understand their surroundings, and geological media such as rock and sand have always been present. In this chapter I will introduce the central concepts and observations that form the foundation of our articles concerning fracture networks and summarise some of our key findings. I also outline the implications of the complexity of geological media with regard to modelling and upscaling.

4.1 The complexity of geological media

In Mandelbrot's groundbreaking book *The Fractal Geometry of Nature* [11] he opens by observing that Nature exhibits many scale-free properties. He describes coastlines, river networks, mountain ranges and clouds that span many orders of magnitude, and that have the same statistical properties over the entire range. The picture in Figure 4.1 depicts two typical geological sedimentary layer structures. There are two lens covers included for scale reference. If you turn to Figure 4.2 you will see that one of the lens covers is a giant replica, and the scale of this is illustrated by the person in the figure. It is impossible to tell the scale of the lens covers from the surroundings because the geological features in the picture are scale-free.

The same scaling invariance applies to fracture systems, and a nice review of many detailed studies is written by Bonnet *et al.* [52]. Bonnet reports the findings of about 40 different *outcrop* studies. Their main scale-free properties are the

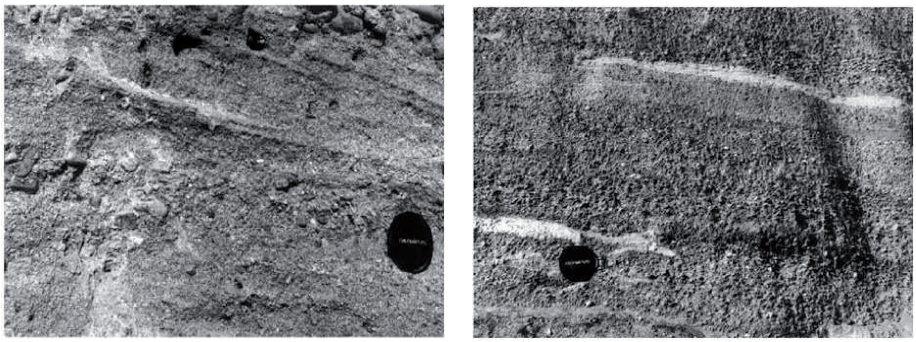


Figure 4.1: *Two pictures of typical geological sedimentary structures. Lens covers included for scale reference. Photo credit: S. W. Wheatcraft and S. W. Tyler.*

fractal placement of fracture centres and the power-law distribution of fracture lengths. The confirmation of such distributions and the determination of their cut-off lengths and exponent/dimension are highly non-trivial, and the results are not always very robust. However, there is overwhelming evidence that general fracture systems display these properties. For the case of the fractal dimension this is mostly measured by the box-counting method and the reported dimensions lie in the interval between 1.0 and 2.0[1]. In Figure 4.3 I have reproduced a figure from Bonnet [52] showing the spread of fractal dimensions between the different studies. For the power-law length distribution the various studies apply various numerical methods[2] for determining the exponents, and again I have reproduced Bonnets distribution of the exponents reported in Figure 4.3.

This lack of a typical scale leads to problems when one tries to model the behaviour of large geological formations [53–55]. In a numerical simulation one can only include a limited amount of features and one therefore has the need to upscale various properties in the model. It is not possible to measure properties such as the permeability for large volumes, such as for example a reservoir, and upscaling is therefore necessary from laboratory scale values. This is a highly nontrivial task [56, 57].

Another complicating factor concerning geological media is the lack of data. This specially applies to the characterisation of fractures and fracture systems in three

[1] This interval is netural because the fracture centers are distributed in a plane filling more of space than a line (with dimension 1) and less than the plane itself haveing dimension 2.

[2] as described in Chapter 2.

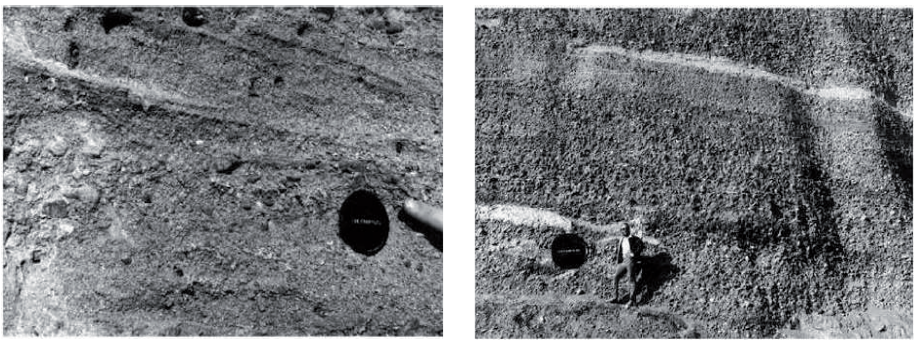


Figure 4.2: *The same two pictures as in Figure 4.1 but now with a finger and a person acting as real scale references. The scale-free nature of the sedimentary structures makes it impossible to se that the lens cover in the right panel is a dummy giant reproduction. Photo credit: S. W. Wheatcraft and S. W. Tyler.*

dimensions. It is currently not practically possible to record the three-dimensional fracture network in detail of any substantial volume. Although the size of a fracture may be large in the in-plane direction, the aperture of the fracture is mostly well below the resolution of seismic methods and the fractures are therefore not detected during seismic surveys. Some experiments have been conducted yielding detailed three-dimensional information on small systems [58–60]. However these are mostly laboratory scale experiments and yield too little data to lend them self to statistical analysis. Therefore the vast majority of studied fracture systems are either two-dimensional (outcrops) or one-dimensional (bore-holes). Due to the fractal nature of the fractures extrapolation from one- or two-dimensions distributions to three-dimensional distributions is not trivial [61–63].

Understanding the structure of geological media, and how single- or multi-phase flow behaves in such media, is crucial to many different fields. The extraction of hydrocarbons from reservoirs has been an important field, but also the preservation of ground-water supplies, the building of nuclear waste repositories and civil engineering are fields where this understanding is vital. The communication between the fields is unfortunately not always the best, leading to a slow exchange of ideas and the production of a waste literature published in a variety of journals. This makes it more complicated to define a research front in the field, and to familiarise oneself with the literature.

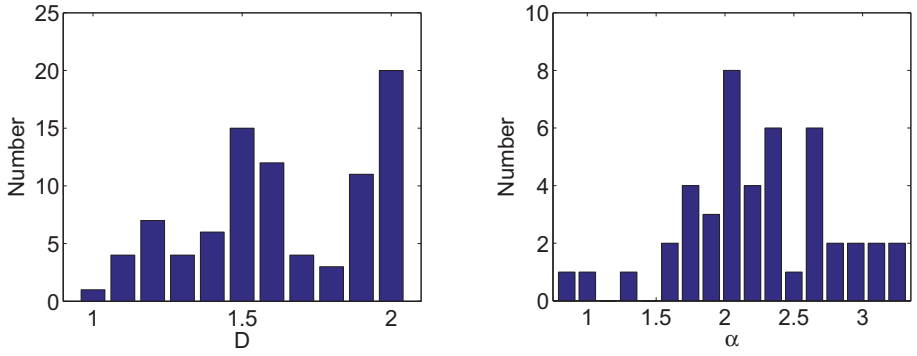


Figure 4.3: *Reproduction of figures from a review by Bonnet et al. [52]. a) Distribution of fractal dimension for fracture networks reported by various studies b) Distribution of power-law fracture length distributions reported by various studies.*

4.2 Fracture modelling, brief history

The origin of modelling of fractures in geological media must be seen in context with the modelling of porous media and percolation processes. A nice review of the early history of this development is given by Sahimi [64]. First porous media was modelled as a macroscopically homogenous system and various percolation and permeability properties were studied. Later the need to model the porous media as macroscopically heterogeneous manifested itself, as this was clearly the case for naturally occurring geological media. Fractures are one of the significant sources of macroscopic heterogeneity in geological media.

The first discrete fracture models were so called *double-porosity models* [65,66]. In these models fractures are regularly spaced and fully connected, and the *matrix*[3] is a set of disconnected regularly shaped blocks. All transport in the system goes through the fracture system, having one porosity, and this system is generally supplied with material from the matrix having a different porosity. One such model is the *sugar-cube model* [66] that is illustrated in Figure 4.4.

These models grew more sophisticated, including three or more degrees of porosity forming *multi-porosity models* [68, 69]. However the models contained many adjustable parameters with unclear physical meaning, and they proved unable to

[3]The material (the rock) that actually makes up the fractured geological formation is called the *matrix*.

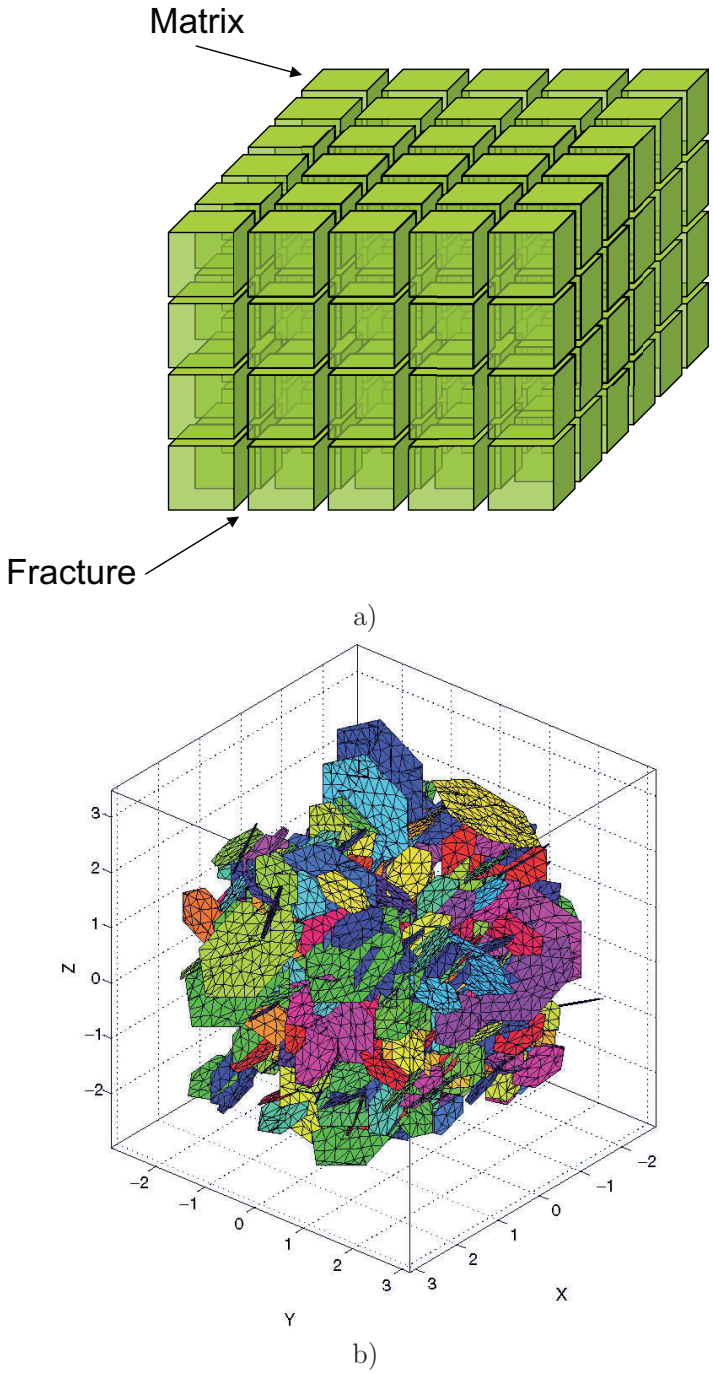


Figure 4.4: *a) Representation of fractures in an early sugar-cube model, b) Representation of fracture in a modern DFN model. Figure form Sisavath et al. [67].*

represent the unordered fractal fracture networks found in nature. Also *double-permeability* and *multi-permeability* models have been developed where there is flow between fractures, fractures and matrix and within the matrix [70, 71].

A model where each fracture is placed individually with correlated or uncorrelated position, size, orientation and other properties to build a network of fractures is more capable of incorporating the fractal nature of the fractures. This type of model was introduced in the context of electrical analogy models to study the flow in such networks [72, 73]. Also these models were sophisticated and more realistic properties for the various distributions were introduced [67, 74, 75]. Such models where each fracture is treated as a separate entity are called *discrete fracture networks* (DFN).

4.3 A discrete fracture network model

In our paper [5] we have implemented a DFN model where the main properties are the fractal placement of the fracture centres and the power-law distribution of the fracture lengths. This model is described in detail by Darcel *et al.* [63, 76, 77] who have developed the model in both two and three dimensions. They have also studied various percolation and stereological properties of the model. The fractallity of the fracture centres is imposed through a hierarchal multiplicative process. This process produces a multifractal probability field for the placement of the centres where the second-order dimension, D_2, also called the correlation dimension [78], is controlled. The generation of a power-law length distribution with a given exponent, α_l, is straight-forward using a pseudo-random number generator. In our case we have generated two-dimensional fracture networks at the percolation threshold, where there is a continuous path from one side of the system to the other. We have investigated the model for the parameter regime $1.5 \leq D_2 \leq 2.0$ and $2.0 \leq \alpha_l \leq 3.0$. This range of values was motivated by studies such as those described by Bonnet *et al.* [52]. A small D_2 gives a heterogeneous fracture center distribution with large voids between densely populated regions, whereas a large D_2 gives a more homogeneous and even fracture centre distribution. For the fracture lengths a large α_l gives a larger portion of smaller fractures, and in the limit of $\alpha_l \to \infty$ all the fractures have the same length. A smaller α_l gives, on the other hand, a larger portion of longer fractures. Examples of model realisations are given in Figure 4.5 for different choices of model parameters.

In order to analyse the network properties of the fracture networks generated we have developed a network representation of the fractures. We define each fracture as a node in the network and link all nodes where the corresponding fractures

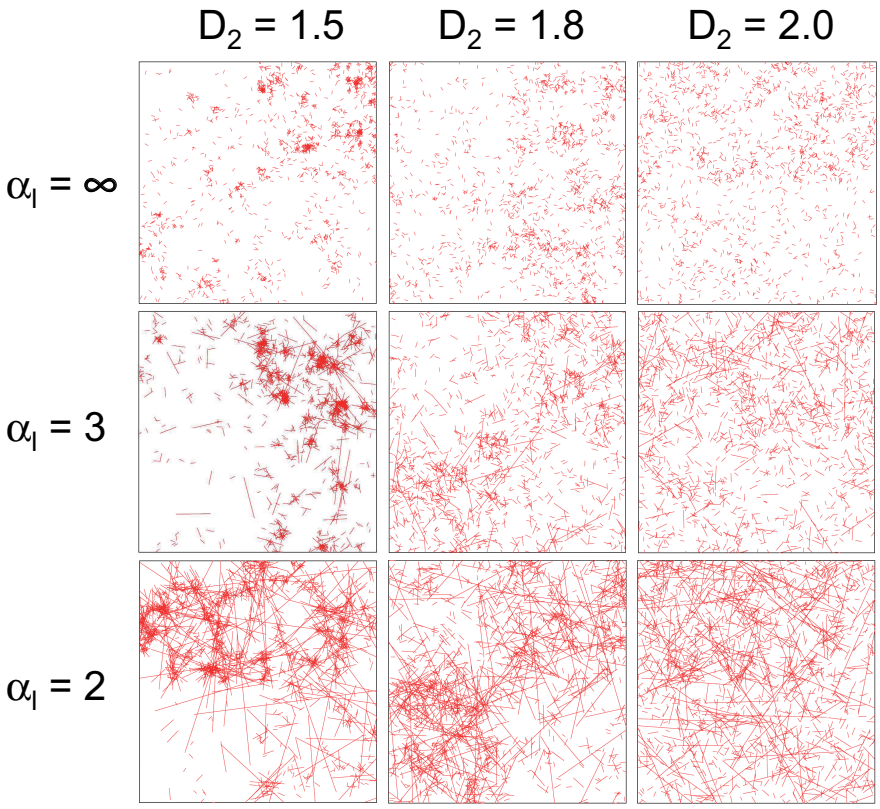


Figure 4.5: *Illustration of fracture sets produced by different choices of the model parameters. The number of fractures produced in each case is the same.*

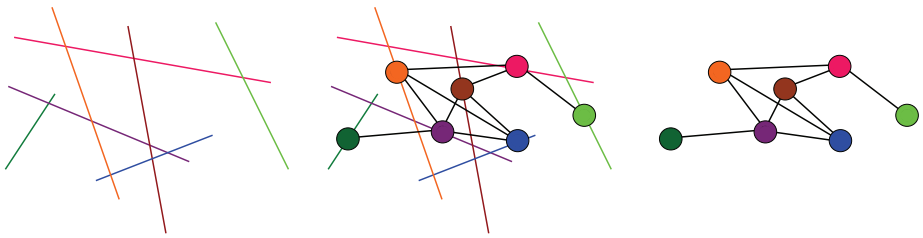


Figure 4.6: *Illustration of the applied networks definition. Each fracture is considered a node, and all nodes representing fractures that intersect each other are linked. This gives a topological network representation of the fracture network. a) Generated linear fractures, b) Fractures with nodes and links indicated, c) Pure network representation of the fractures.*

intersect in real space. This procedure is illustrated in Figure 4.6, and enables us to investigate the networks using the tools from modern network theory discussed in Chapter 3. The procedure is also illustrated for a real-world set of outcrop fractures in Figure 4.8.

The results of the analysis show that the networks formed have a broad degree distribution that can be fitted with a power-law in the tail. The power-law exponent seems to be insensitive to variations in D_2 but decreases monotonically with increasing α_l. The clustering coefficients and efficiencies also seems insensitive to the fractal dimension and increases monotonically with the fracture length distribution exponent. The efficiencies and clustering coefficients for various model parameters are plotted in Figure 4.7. Compared to rewired and random versions[4] of the same networks the model produces networks with large clustering coefficients and efficiencies, and they are therefore small-world networks. The model also forms assortative networks *i.e.* nodes of a given degree tend to link to nodes of similar degree and not link to nodes with degrees that differ significantly form its own. Figure 4.9 displays a typical degree-degree correlation matrix discussed in Section 3.4 for the networks produced by the model.

[4]This comparison is described in detail in our fifth paper [5].

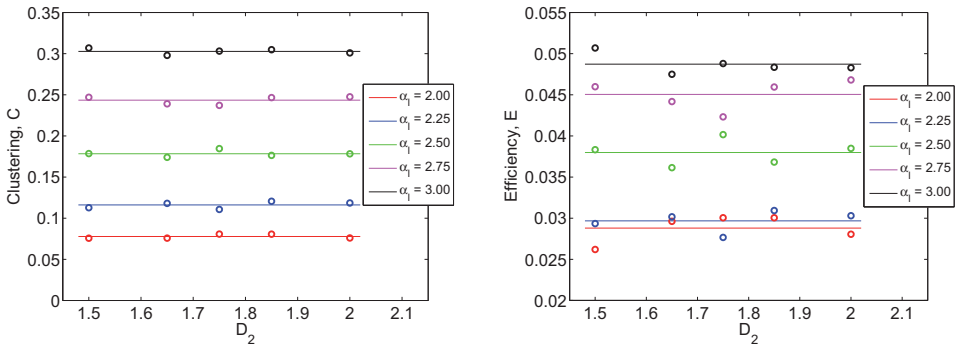


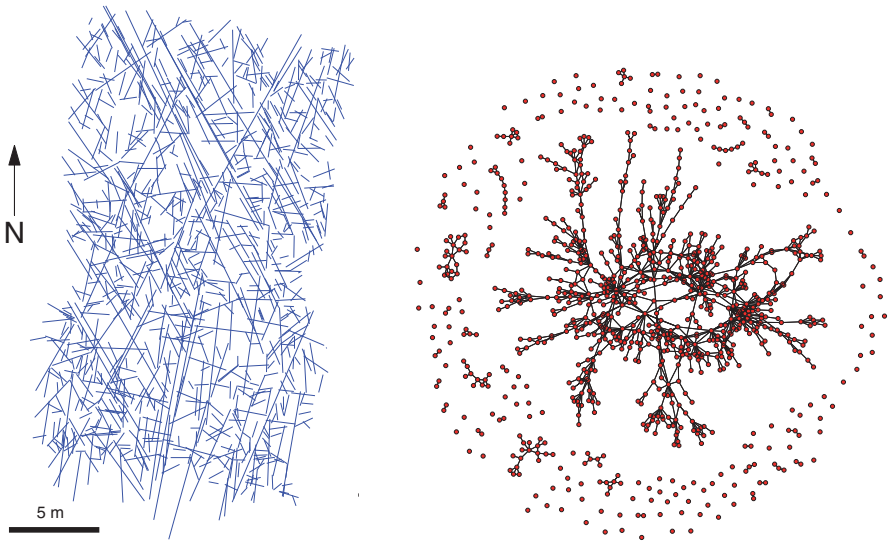
Figure 4.7: *a) Average values of the clustering coefficient for various values of the model parameters. b) Average efficiency values for the same networks.*

4.4 Comparison between model and real-world results

We have studied outcrop data obtained by *Svensk Kärnbränslehantering AB* (SKB), and applied the above outlined network analysis to these data. Figure 4.8 shows the fracture traces from a real-world outcrop study and the corresponding topological network representation.

The outcrop networks have a broad degree distribution, $p(k)$, comparable to that produced by the DFN model. The outcrop distribution can be fitted by a power-law for the tail of the distribution. The best fit gives an average exponent $\alpha_l = 2.3$ for the realation $p(k) \propto k^{-\alpha_k}$. This is somewhat higher than the model results described above that range from 0.86 to 2.2. Note that the data for the outcrop degree distribution were not abundant, and the estimate exponent for the tail of the distribution is connected with a large uncertainty.

The outcrops were found to have a clustering coefficient C, in the range $0.088 < C < 0.24$ with an average value of 0.18. This agrees well with the model results shown in Figure 4.7 that range from 0.079 to 0.31 and have an average value of 0.19. The efficiencies, E, found in the outcrops lie in the interval $0.0045 < E < 0.14$ with an average value of 0.065. This average is somewhat higher than what is found in the model results ranging from 0.026 to 0.051 with an average of 0.038. A higher efficiency in the outcrop networks is not surprising as the model networks are generated at the percolation threshold and the outcrop networks are believed



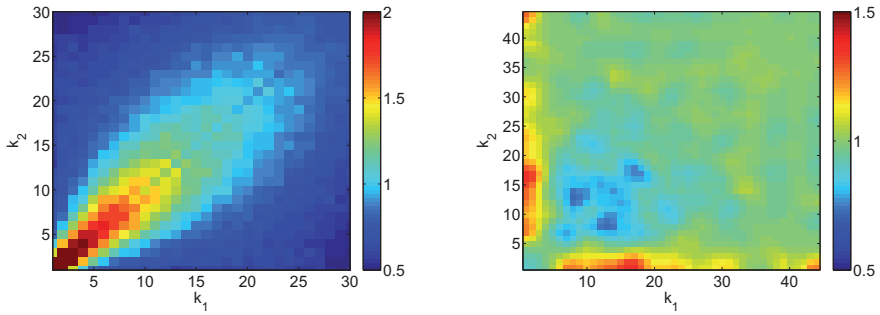
a) b)

Figure 4.8: *a) Outcrop fracture traces in real space. b) Topological network representation for the fractures in a).*

to have a fracture density somewhat above the percolation threshold. A higher fracture density will generally give a larger efficiency.

The outcrop fractures forms dissassortative networks i.e. low degree nodes tend to link to high degree nodes and vice versa, and equal degree nodes tend not to link to each other. This is in contrast to the model networks that form assortative networks. This difference in the degree-degree correlation matrix can be seen in Figure 4.9.

Maybe the most important finding is that the outcrop fractures form small-world networks. This means that they form very robust and efficient transport networks that are not vulnerable to the random removal of a few nodes or links. This is very beneficial if one wants to extract hydrocarbons from a fractured reservoir. However it is a major problem if one wants to confine a pollutant such as in a nuclear waste disposal or while trying to preserve a underground fresh water reserve.



a) b)

Figure 4.9: *a) Degree-degree correlation matrix for a typical model network. b) Average degree-degree correlation matrix for the outcrop networks studied.*

4.5 Thoughts for future work

All of the above presented results are for two-dimensional problems. A natural extension of this work is to look at three-dimensional systems where fractures are represented by (rough) two-dimensional sheets. This will add much complexities to the problem since the interactions between the fractures become more complicated. The intersection between two lines is just a point, however the intersection between two sheets is a profile of some sort. This change in behaviour/dimensionallity motivates the assignment of weights[5] to the links in the network since the lines will have different properties depending on the way that the fractures intersect.

Calculating the permeability of fracture networks is currently mainly done by methods like finite element or finite difference as in [79, 80]. However it may be possible to build (weighted) networks where one can calculate the permeability directly. A better understanding of the permeability of single (rough) fractures and fracture joints is a prerequisite for this.

Carbonate reservoirs are presumably the type of geological formation where this type of analysis will be of greatest interest due to the low permeability of such formations and the subsequent need for the fracture system to facilitate transport. In contrast to granite, from which the real world outcrops described above are formed, carbonate reservoirs are generally sectioned into sedimentary layers. These layers influences the fractures in the media. Some fractures terminate at layer-boundaries, some are inter-layer fractures and some layer boundaries can be

[5] See Section 3.1.

perceived as fractures. This heterogeneity has to be dealt with if an analogue study in three dimensions is to be attempted for such a formation.

The hierarchal method for generating fractally distributed fracture centres in the DFN model discussed above is not very flexible. Only one of the moments are tunable, and this gives little control over the multi-fractal behaviour of the method. Described by the f-alpha formalism the method produces a parabolic f-alpha function. A more general method, producing other f-alpha functions, would be interesting. Changing the multi-fractal behaviour of the fracture centres may influence the structure of the fracture network.

5 3D fracture project

I haven't failed,
I've found 10,000 ways that don't work.
Thomas Alva Edison (1854 – 1912)

As described in Section 4.1 there have been few reported findings concerning the full characterisation of true three-dimensional fracture networks. The detailed mapping of an *in situ* fracture network requires either a method that can reveal the internal structure of the rock such as seismic, CT or MRI, or a detailed dissection of the volume to be investigated. Current seismic methods does not have the resolution to reveal features smaller than the decimeter scale, and geological surveys suggest that most fracture networks are dominated by fractures with an aperture considerably smaller than that. For technics such as CT or MRI only laboratory scale samples can be investigated, and the reliability of the results are questionable largely due to poor statisitcs. Some work has been done on dissection and nuclear traces [58, 81–84]. However, no-one has reported production of fracture networks of sufficient scale and quality needed for a network analysis similar to what is outlined in Chapter 4. In this chapter I will discuss a project that we started in order to produce fracture networks suitable for such analysis. Unfortunately the project did not yield the results that we hoped for.

5.1 Purpose of experiment

The purpose of the project was to generate and characterise three-dimensional fracture networks in such detail that the location and extension of each individual fracture could be determined. By doing this in a statistically reproducible manner we could build fracture networks that could be analysed using tools from modern network theory. A secondary target was to measure the acoustic activity in the samples during the fracturing process. This would have given us insights into the dynamics of the fracturing process.

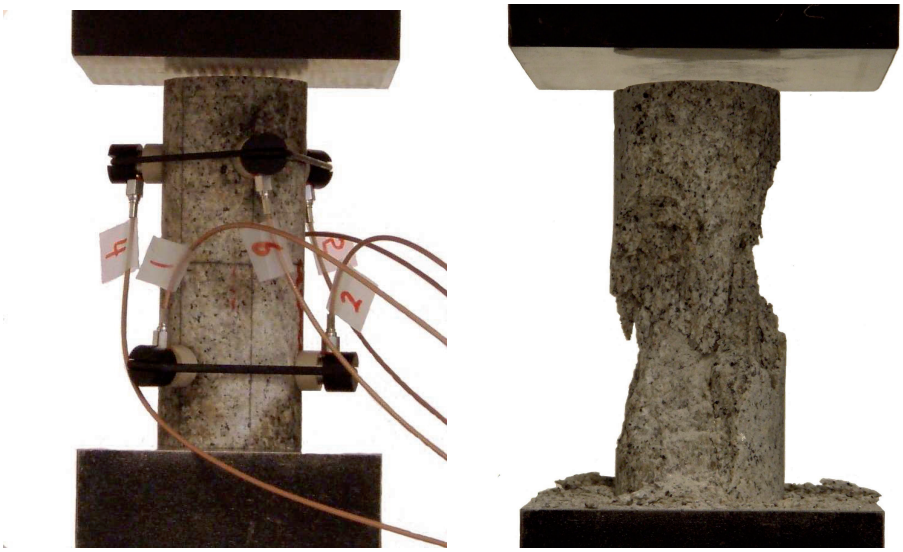
In order to fracture the samples we would apply stress very close to their yielding-stress, and monitor the activity with via *acoustic-emission* (AE) recordings. Thereafter we would cut the samples in thin layers and record the fractures by photographing them layer by layer. From this a full three-dimensional profile of the fracture networks could have been built.

For this experiment a *triaxial loading* configuration or a controllable confining pressure would have been desirable. The yielding stress of a rock is greatly increased when a confining pressure is applied, and this situation is closer to what we perceive as a natural fracturing process for geological media. However, the availability of such equipment is sparse, and the experimental complications involved are considerable. An *uniaxial loading* was therefore applied, with the extension to a more complex loading scheme to be considered in future work.

5.2 Choice of material and sample preparation

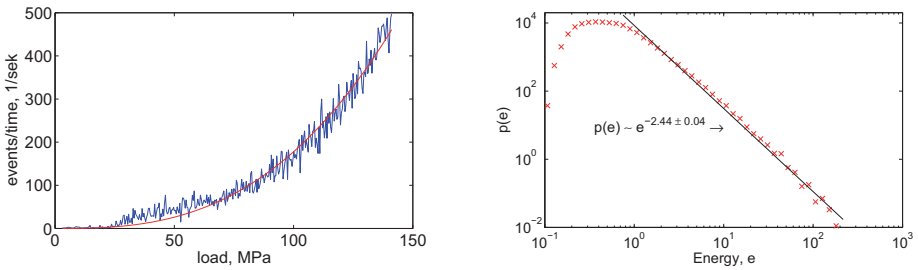
The choice of material was not trivial. Earlier works along similar lines have included *Carrara* marble [81], sandstone [83] and other natural rocks [82]. We also considered artificial samples such as aluminium-alloys, ceramics, ice and solidified granular matter. However all the artificial materials proved to have severe deficiencies such as being too brittle, too ductile or too porous. Another reason for choosing a natural rock was that the sample would be more realistic and closer to the geological systems that we wanted to imitate. Among the natural rock types sandstone was a good candidate, but the porous and coarse-grained structure of the rock would make it difficult to record any fracture pattern via dissection. We therefore concentrated on more brittle rocks, such as marble and granite. Granite was chosen because of its availability and an anticipation that the fractures would be easier to photograph in a granite sample than a marble sample. A picture of a granite sample with acoustic-emission microphones attached is given in Figure 5.1.

The granite samples used were cylinders of 50 mm in diameter and 125 mm long, having a density of 2.64 kg/dm^3. We used six microphones for each sample to record the acoustic activity while applying stress. These were placed in two planes normal to the cylinder axis equally distributed around the circumference of the sample. The position of the microphones can be seen in Figure 5.1.



a) b)

Figure 5.1: *a) A granite sample with acoustic-emission microphones attached mounted in a hydraulic press. b) A granite sample after the yielding-stress has been exceeded.*



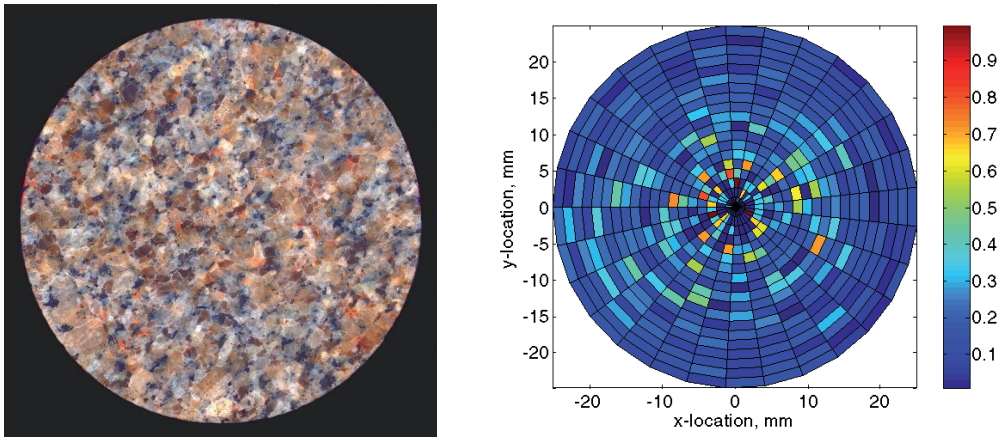
a) b)

Figure 5.2: *a) Plot of the AE activity as a function of the loading stress fitted with an exponential function. b) Histogram of the estimated AE event energy distribution plotted with logarithmic axis and fitted with a power-law function for the high-value tail.*

5.3 Fracture production and AE results

In order to estimate the yield-stress for the granite samples three samples were loaded until they yielded. A loading rate of 1.5 kN/s was used, and the three samples yielded 196.6 MPa, 169.7 MPa and 178.4 MPa respectively. A spread in the yielding-stress is expected due to heterogeneities in the samples and differences in the initial (micro) fracture population. We observed a $\sim 15\%$ difference between the highest and lowest yielding stress. This complicated the process of fracture production because the yielding stress of a given sample could not be known beforehand. It would therefore be very hard to load a sample very close to this threshold without destroying the sample. The recorded activity from the AE measurements gave a good indication of the level of fracturing within the rock. From this activity and visible signals such as sputtering of fragments from the sample, the loading was stopped very close to what was perceived as the yielding threshold of the samples.

Three samples were loaded with AE microphones attached. They were loaded up to 158.2 MPa, 194.9 MPa and 188.8 MPa respectively, and for each sample nearly $2{\cdot}10^5$ AE events were recorded. The AE equipment was calibrated and thresholded to only detect acoustic signals that were above the noise generated by the loading equipment. For each AE event, a hit number was assigned, and the energy of the event was estimated from the magnitude recorded at the various microphones. The location of the event was estimated by triangulation. The frequency of events increased exponentially with loading stress as shown in Figure 5.2.



a) b)

Figure 5.3: *a) Picture of a cross-section of a granite sample after loading. The picture is taken along the cylindrical axis. b) Histogram of the radial AE activity within a typical sample. The values are the number of AE hits in the various sections integrated along the cylindrical axis and normalised.*

A large number of papers have been published on AE measurements describing the deformation, fracturing and break-down of various materials [64,85–87]. In Figure 5.3 a histogram of the spatial distribution of the AE activity of a typical sample is shown. The histogram is a top-view plot, and displays the radial distribution of the activity. The histogram of the estimated event energies also plotted in Figure 5.2, and shows the tell-tale power-law tail that has been reported repeatedly throughout the literature. Note that no corrections of the event energy due to dispersion because of different travel lengths to the microphone were made [88].

5.4 Missing fractures

The main object of the project was to detect the fractures induced in the samples. The AE measurements showed that there was considerable activity in the samples during loading, and this should have produced fractures. A sample was therefore cut normal to its cylindrical axis at the middle of the sample, and the surface polished. A microscope was used to photograph the surface with a pixel-resolution of 1 μm. A full view of the sample surface can be seen in Figure 5.3. To our

surprise no fractures were visible in the photographs. It seems that the fractures were closed and not detectable.

In order to attempt to visualise the fracture we applied various fluorescent tracers. Both florescent molecules and nano-spheres were tested. These were diluted in a solution wettable to the rock. The tracers absorb florescent light at a given wavelength and emit at a different wavelength. This enabled us to illuminate the sample at one wavelength and filter this out while examining the sample. Neither of the two traces were able to penetrate into the fractures, and no fracture were recorded.

Due to the inability to record the fractures we abandoned the project. The fracture location obtained from the AE emissions could have given interesting information on the dynamic fractal distribution of fractures. However the accuracy needed for such an analysis is very high, and the triangulation was not able to achieve an accuracy better than $\sim$ 5mm. Therefore this part of the project was also abandoned.

Although the project did not yield the desired results it gave insights into the problem of recording three-dimensional fracture systems. The usefulness of such data for the understanding of both single fractures and fracture systems is so great that several labs are currently developing methods and experimental setups to record such data. Hopefully this project can help to make these efforts more successful.

Bibliography

[1] C. A. Andresen, A. Hansen, and J. Schmittbuhl. Ridge network in crumpled paper. *Physical Review E*, 76(026108), 2007.

[2] H. F. Hansen, C. A. Andresen, and A. Hansen. A quantitative measure for path structures of complex networks. *Europhysics Letters*, 78(48005), 2007.

[3] C. A. Andresen, H. F. Hansen, A. Hansen, G. L. Vasconcelos, and J. S. Andrade Jr. Correlations between political party size and voter memory. *International Journal of Modern Physics C*, 19(11):1–11, 2008.

[4] C. A. Andresen, R. Le Goc, P. Davy, and A. Hansen. Network analysis of outcrop data. Submitted to *Journal of Geophysical Research - Solid Earth.*, 2008.

[5] C. A. Andresen and A. Hansen. Network properties of discrete fracture network model. To be submitted, 2008.

[6] Statistisk sentral byrå, November 2008. www.ssb.no.

[7] H. F. Hansen, L. E. Walle, and A. Hansen. Wealth and income distributions in norway: A case study. To be publiched, 2008.

[8] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351, 2004.

[9] K. Wiesenfeld. Scaling laws. *The American Journal of Physics*, 69(9):938–942, 2001.

[10] B. B. Mandelbrot. *Fractals: Form, Chance and Dimension.* W. H. Freeman and Co., New York, USA, 1977.

[11] B. B. Mandelbrot. *The Fractal Geometry of Nature.* W. H. Freeman and Co., New York, USA, 1982.

[12] J. Feder. *Fractals.* Plenum Press, New York, USA, 1988.

[13] K. Falconer. *Fractal Geometry.* John Wiley and Sons, New York, USA, 2001.

[14] M. A. Alber and J. Peinke. Improved multifractal box-counting algorithm, virtual phase transitions, and negative dimensions. *Physical Review E*, 57(5):5489–5493, 1998.

[15] L. V. Meisel, M. A. Johnson, and P. J. Cote. Box-counting multifractal analysis. *Physical Review A*, 45(10):6989–6996, 1992.

[16] L. V. Meisel and M. A. Johnson. Convergence of numerical box-counting and correlation integral multifractal analysis techniques. *Pattern Recognition*, 30(9):1565–1570, 1997.

[17] A.L. Barábasi and H. E. Stanley. *Fractal Concepts in Surface Growth*. Cambridge Uni. Press, New York, USA, 1995.

[18] T. Hida. *Brownian Motion*. Springer, New York, USA, 1980.

[19] J. Ø. Bakke and A. Hansen. Accuracy of roughness exponent measurement methodes. *Physical Review E*, 76(031136), 2007.

[20] J. Schmittbuhl and J.P. Vilotte. Reliability of self-affine measurements. *Physical Review E*, 51(1):131–147, 1995.

[21] J. W. Kantelhardt *et al.* Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316:87–114, 2002.

[22] I. S. Simonsen, A. Hansen, and O. M. Nes. Determination of the Hurst exponent by use of wavelet transforms. *Physical Review E*, 58(3):2779–2787, 1998.

[23] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Review Modern Physics*, 74:47–97, 2002.

[24] S. Boccaletti. Complex networks: Structure and dynamics. *Physics Reports*, 424:175–308, 2006.

[25] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[26] S. H. Strogatz. Exploring complex networks. *Nature*, 410:268–275, 2001.

[27] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo. Topology of the fittest transport network. *Physical Review Letters*, 84(20):4745–4748, 2000.

[28] G. L. Alexanserson. Euler and Königsberg's bridges: A historical view. *Bulletin of the American Mathematical Society*, 43:567–573, 2006.

[29] R. Milo *et al.* Network motifs: Simple building blocks of complex networks. *Science*, 298(824), 2002.

[30] R. Milo *et al.* Superfamilies of evolved and designed networks. *Science*, 303(1538), 2004.

[31] S. S. Shen-Orr *et al.* Network motifs in the transcriptional regulation network of *Escherichia Coli*. *Nature Genetics*, 31(64), 2002.

[32] J. P. Bagrow. Local method for detecting communities. *Physical Review E*, 72(046108), 2005.

[33] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(026113), 2004.

[34] M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 70(056131), 2004.

[35] M. A. Serrano, M. Boguna, and R. Pastor-Satorras. Correlations in weighted networks. *Physical Review E*, 74(055101), 2006.

[36] A.L. Barábasi. *Linked*. Penguin Group, New York, USA, 2002.

[37] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Physical Review E*, 393:440–442, 1998.

[38] L. A. N. Amaral *et al.* Classes of small-world networks. *Proceedings of the National Academy of Sciences, USA*, 97(21):11149–11152, 2000.

[39] L. Rong Z. Zhang and F. Comellas. Evolving small-world networks with geographical attachment preference. *Journal of Physica A*, 39(13):3253–3261, 2006.

[40] R. Albert, H. Jeong, and A.L.Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

[41] R. Pastor-Satorras, A. Vázqez, and A. Vespignani. Dynamical and correlation properties of the internet. *Physical Review Letters*, 87(258701), 2001.

[42] A. Vazquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the internet. *Physical Review E*, 65(066130), 2002.

[43] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296:910–913, 2002.

[44] S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: Correlation profile of the internet. *Physica A*, 333:529–540, 2004.

[45] Virtual Round Table on ten leading questinos for network research. Virtual round table on ten leading questinos for network research. *The European Physical Journal B*, 38:143–145, 2004.

[46] E. Sultan and A. Boudaoud. Statistics of crumpled paper. *Physical Review Letters*, 95(136103), 2006.

[47] G. A. Vliegenthart and G. Gompper. Forced crumpling of self-avoiding elastic sheets. *Nature*, 5:216–221, 2006.

[48] P. A. Houle and J. P. Sethna. Acoustic emissions from crumpling paper. *Physical Review E*, 54(1):278–283, 1996.

[49] D. L. Blair and A. Kudrolli. Geometry of crumpled paper. *Physical Review Letters*, 94(166107), 2005.

[50] A. J. Wood (Editor). Witten's lectures on crumpling. *Physica A*, 313:83–109, 2002.

[51] G. Gohau, A. V. Carozzi, and M. Carozzi. *A history of geology*. Rutgers University Press, New Brunswick, USA, 1990.

[52] E. Bonnet *et al.* Scaling of fracture systems in geological media. *Reviews of Geophysics*, 39(3):347–383, 2001.

[53] O. M. Gurpinar and C. A. Kossack. Realistic numerical models for fractured reservoirs. *Society of Petroleum Engineers Journal*, 5(4):485–491, 2000.

[54] G. Yielding, T. Needham, and H. Jones. Sampling of fault populations using sub-surface data: A review. *Journal of Structual Geology*, 18(2-3):135–146, 1996.

[55] S. P. Neuman. Trends, prospects and challenges in quantifying flow and transport through fractured rocks. *Hydrogeological Journal*, 13:124–147, 2005.

[56] M. G. Gerritsen and L. J. Durlofsky. Modeling fluid flow in oil reservoirs. *Annual review of fluid mechanics*, 37:211–238, 2005.

[57] G. Bloschl and M. Sivapalan. Scale issues in hydrological modelling - A review. *Hydrological Processes*, 9(3-4):251–290, 1995.

[58] L. S. Gertsch. Three-dimensional fracture network models from laboratory-scale rock samples. *Int. J. Rock Mech. Min. Sci. & Geomech. Abstr.*, 32(1):85–91, 1995.

[59] L. J. Pyrak Nolte, C. D. Montemagno, and D. D. Nolte. Volumetric imaging of aperture distributions in connected fracture networks. *Geophysical Research Letters*, 24(18), 1997.

[60] A. Hodgkins, T. J. Marrow, P. Mummery, B. Marsden, and A. Fok. X-ray tomography observation of crack propagation in nuclear graphite. *Material Science and Technology*, 22(9):1045–1051, 2006.

[61] B. Berkowitz and P. M. Adler. Stereological analysis of fracture network structure in geological formations. *Journal of Geophysical Research*, 103(B7):15339–15360, 1998.

[62] P. Davy, C. Darcel, O. Bour, R. Munier, and J. R. de Dreuzy. A note on the angular correlation applied to fracture intensity profiles along a drill core. *Journal of Geophysical Research*, 111(B11408), 2006.

[63] C. Darcel, O. Bour, and P. Davy. Stereological analysis of fractal fracture networks. *Journal of Geophysical Research*, 108(B9), 2003.

[64] M. Sahimi. Flow phenomena in rocks - From continuum models to fractals, percolation, cellular-automata, and simulated annealing. *Reviews of Modern Physics*, 65(4):1393–1534, 1993.

[65] G. I. Barenblatt and I. P. Zheltov. Fundamental equations for the filtration of homogeneous fluids through fissured rocks. *Doklady Akademii Nauk. SSSR*, 132(3):545–548, 1960.

[66] J. E. Warren and P. J. Root. The behavior of naturally fractured reservoirs. *Society of Petroleum Engineers Journal*, 3(3):245–255, 1963.

[67] S. Sisavath, V. Mourzenko, P. Genthon, J. F. Thovert, and P. M. Adler. Geometry, percolation and transport properties of fracture networks derived from line data. *Geophysical Journal International*, 157(2):917–934, 2004.

[68] P. J. Closmann. Aquifier model for fissured reservoirs. *Society of Petroleum Engineers Journal*, 15(5):385–398, 1975.

[69] D. Abdassah and I. Ershagi. Triple-porosity systems for representing naturally fractured reservoirs. *SPE Formation Evaluation Journal*, 1(2):113–127, 1986.

[70] Z.-X. Chen. Transient flow of slightly compressible fluids through double-porosity, double-permeability systems: A state-of-the-art review. *Transport in Porous Media*, 4(2):147–184, 1989.

[71] Y. Zhang and D. Tong. The pressure transient analysis of deformation of fractal medium. *Journal of Hydrodynamics, Ser. B*, 20(3):306–313, 2006.

[72] R. W. Parsons. Directional permeability of heterogeneous anisotropic porous media - Discussion. *Society of Petroleum Engineers Journal*, 4(4):363–364, 1964.

[73] *Proceedings of symposium on percolation through fissured rocks*. International society for rock mechanics and international association of engineering geology, 1972. J. A. Caldwell.

[74] P. M. Adler and J.-F. Thovert. *Fractures and fracture networks*. Kluwer academic publishers, Boston, USA, 1999.

[75] V. V. Mourzenko, J. F. Thovert, and P. M. Adler. Percolation of three-dimensional fracture networks with power-law size distribution. *Physical Review E*, 72(036103), 2005.

[76] C. Darcel, O. Bour, P. Davy, and J. R. de Dreuzy. Connectivity properties of two-dimensional fracture networks with stochastic fractal correlation. *Water Resources Research*, 39(10), 2003.

[77] O. Bour, P. Davy, C. Darcel, and N. Odling. A statistical scaling model for fracture network geometry, with validation on a multiscale mapping of a joint network (Hornelen Basin, Norway). *Journal of Geophysical Research*, 107(B6), 2002.

[78] H. G. E. Hentschel and I. Proccacia. The infinite number of generalised dimensions of fractal and strange attractors. *Physica D*, 8:435–444, 1983.

[79] L. Moreno, J. Crawford, and I. Neretnieks. Modelling radionuclide transport for time varying flow in a chanel network. *Journal of Contaminant Hydrology*, 86:215–238, 2006.

[80] B. Berkowitz. Characterizing flow and transport in fractured geological media: A review. *Advances in Water Resources*, 25:861–884, 2002.

[81] S. Peacock, C. McCann, J. Sothcott, and T. R. Astin. Seismic velocities in fractured rocks: An experimental verification of Hudson's theory. *Geophysical Prospecting*, 42:27–80, 1994.

[82] B. Frieg, W. R. Alexander, H. Dollinger, C. Buhlerand P. Haag, A. Mori, and K. Ota. In situ resin impregnation for investigating radionuclide retardation in fractured repository host rocks. *Journal of Contaminant Hydrology*, 35(1-3):115–130, 1998.

[83] S. Baraka-Lokmane, R. Liedl, and G. Teutsch. Comparison of measured and modelled hydraulic conductivities of fractured sandstone cores. *Pure and Applied Geophysics*, 160:909–927, 2003.

[84] M. K. Fisher, C. A. Wright B. M. Davidson, A. K. Goodwin, E. O. Fielder, W. S. Buckler, and N. P. Steinsberger. Integrating fracture-mapping technologies to improve stimulations in the Barnett Shale. *Society of Petroleum Engineers, Production and Facilities*, 20(2):85–93, 2005.

[85] J. P. Sethna, K. A. Dahmen, and C. R. Myers. Crackling noise. *Nature*, 410, 2001.

[86] S. Zapperi, P. Ray, H. E. Stanley, and A. Vespignani. First-order transition in the breakdown of disordered media. *Physical Review Letters*, 78(8):1408–1411, 1997.

[87] S. P. Gross, J. Fineberg, M. Marder, W. D. McCormick, and H. L. Swinney. Acoustic emissions from rapidly moving cracks. *Physical Review Letters*, 71(19):3162–3165, 1993.

[88] D. Lockner. The role of acoustic-emission in the study of rock fracture. *Int. J. Rock Mechanics Minimg Sci. Geomech. Abstr.*, 30(7):883–899, 1993.

Index

acoustic emission, 40
actor, 16
average wavelet coefficient, 13

bonds, 16
box-counting dimension, 8

characteristic path length, 20
clustering coefficient, 19
component, 16
coordination number, 17

degree, 17
degree distribution, 18
degree-degree correlation, 23
deterministic fractal, 7
detrended fluctuation analysis, 13
diameter, 19
discrete fracture networks, 32
double-permeability models, 32
double-porosity model, 30

edges, 16
efficiency, 19, 20

fat tail, 4
fractal dimension, 7
fractional noise, 11
fully connected, 16

geodesic path, 19
geology, 27
giant components, 16
graph theory, 15

Hausdorff-Besicovitch dimension, 7
hub, 18
Hurst exponent, 10

links, 16

Mandelbrot B. B., 7
matrix, 30
multi-porosity models, 30

network, 16
nodes, 16

outcrop, 27

power spectrum density, 12
power-law, 4

quasi-fractals, 7

random walk, 10
randomised network, 23
rewired network, 23

scale-free networks, 18
self-affine, 10
self-similarity, 10
Small-World networks, 20
Small-World property, 20
stochastic fractals, 7
strength, 17
sugar-cube model, 30

triaxial loading, 40

uniaxial loading, 40

vertices, 16

weigh, 17
weighted networks, 17

C. A. Andresen, A. Hansen and J. Schmittbuhl,
Ridge Network in Crumpled Paper,
Phys. Rev. E **76**, 026108 (2007).

PHYSICAL REVIEW E **76**, 026108 (2007)

Ridge network in crumpled paper

Christian André Andresen* and Alex Hansen†
Department of Physics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

Jean Schmittbuhl‡
Institut de Physique du Globe Strasbourg, 5 rue René Descartes, 67084 Strasbourg, France
(Received 20 February 2007; revised manuscript received 1 June 2007; published 20 August 2007)

The network formed by ridges in a straightened sheet of crumpled paper is studied using a laser profilometer. Square sheets of paper were crumpled into balls, unfolded, and their height profile measured. From these profiles the imposed ridges were extracted as networks. Nodes were defined as intersections between ridges, and links as the various ridges connecting the nodes. Many network and spatial properties have been investigated. The tail of the ridge length distribution was found to follow a power law, whereas the shorter ridges followed a log-normal distribution. The degree distribution was found to have an exponentially decaying tail, and the degree correlation was found to be disassortative. The facets created by the ridges and the Voronoi diagram formed by the nodes have also been investigated.

DOI: 10.1103/PhysRevE.76.026108 PACS number(s): 89.75.Hc, 83.60.−a, 89.75.Fb, 42.62.−b

I. INTRODUCTION

The crumpling of paper is an everyday occurrence, yet it is a surprisingly rich and complex process. Paper is an elastic, flexible, and heterogeneous material, and many authors have tried to describe its crumpling properties analytically [1–3], numerically [4,5], and experimentally [6–11]. The crumpling process of paper is also interesting because it is a special case of the thin plate deformation problem that is central in describing processes that occur, for example, in car crashes and tank failures [12]. Earlier studies have tried to describe the ridge network of crumpled paper [4,6,10] and some results have been found; however, much is still unclear. This work aims at describing the ridge network formed during the common hand-crumpling process of ordinary printing paper. The application of modern network theory [13–15] has been specially emphasized.

This paper is organized as follows. In Sec. II the experimental procedure is described, and in Sec. III the ridge detection method is presented. The results are discussed in Sec. IV; in particular the ridge length and the degree distribution are discussed. Also, the degree-degree correlation, the clustering, and the surface roughness are investigated, in addition to the facet distribution and the angular ridge distribution. Finally the main conclusions are summarized in Sec. V.

II. EXPERIMENTAL PROCEDURE

Ordinary printing paper was used for all the experiments, and some of the properties of the paper are given in Table I. All the samples were cut into square sheets of 21×21 cm^2, and crumpled by hand into small balls. The diameters of the various balls produced are given in Table I. The hand-crumpling procedure has been applied before [6,8,10], and is practical because it is easy to conduct and produces a compact result. Unfortunately, the process in not repeatable and poorly controlled. Several test crumplings were conducted before the measured samples were crumpled in order to reduce the variance between the samples. Earlier studies [8] on acoustic emissions from crumpling of various materials have indicated that the emission spectra show a surprisingly low sensitivity to the crumpling method. This may indicate that the outcome of the crumpling is not highly sensitive to the details of the process. Balankin *et al.* [10] discuss the scaling behavior of the crumpling process for different paper thicknesses. They conclude that the impact of the variation of the applied confinement force F on the ball radius R is small since there is only a weak dependence $R\propto F^{-0.25}$. For these reasons, no special precautions, such as dents or initial folding, were taken to increase repeatability. After crumpling the samples, they were carefully unfolded, taking care not to tear the paper, introduce new ridges, or remove some of the original ridges. When the paper ball was unfolded the paper was stretched to a size of 20×20 cm^2, and fastened to an aluminum plate. This ensured that the vertical heights of the

TABLE I. List of samples investigated. x and y steps are the numbers of measured points in the x and y directions, respectively. Thickness is the thickness of the paper, and ball diameter is the diameter of the ball produced during the crumpling process. All samples were originally 21×21 cm^2 and thereafter unfolding and stretching to 20×20 cm^2 producing a maximum height of 12 mm. An area of 18×18 cm^2 in the center of the samples was measured.

Sample	x step	y step	Thickness (μm)	Weight (g/m^2)	Ball diameter (mm)
1	900	900	51±5	49.0±1.0	26±2
2	1800	1800	51±5	50.0±1.0	27±2
3	1000	1000	95±5	80.0±0.5	35±2
4	900	900	95±5	80.0±0.5	32±2
5	3600	3600	100±2	83.0±0.5	33±2
6	900	900	220±5	175.0±1.0	43±2

*Christian.Andresen@ntnu.no
†Alex.Hansen@ntnu.no
‡Jean.Schmittbuhl@eost.u-strasbg.fr

 ©2007 The American Physical Society

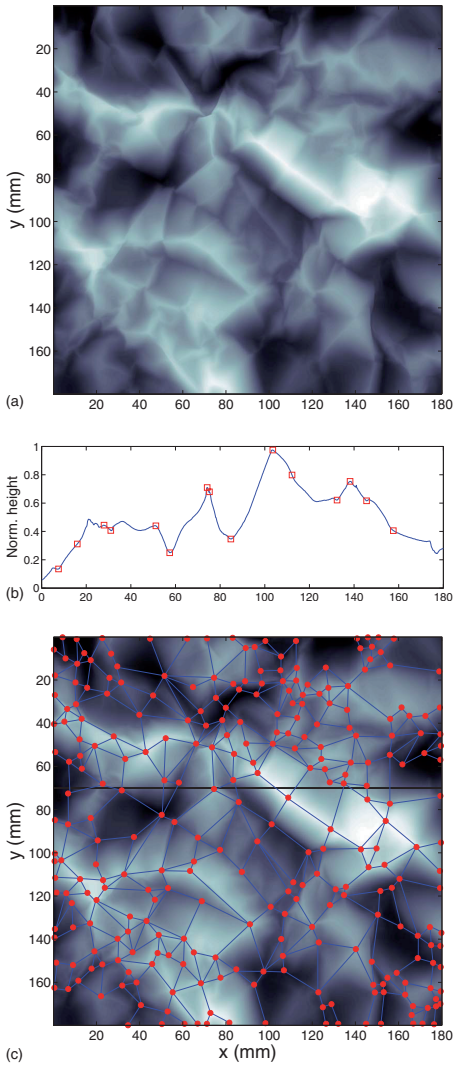


FIG. 1. (Color online) Top: Gray-scale height plot of sample 6 as function of x and y positions. Gray-scale indicates elevation (lighter is higher); ridges are clearly visible. Middle: A single one-dimensional profile from sample 6 (marked as a black line in the bottom plot). The points of the profile that give rise to ridges are marked by squares. Bottom: Network extracted from sample 6 superimposed on the gray-scale plot from the top figure.

samples were no more than 12 mm (the maximum range for the instrument used).

The full $(2+1)$-dimensional height mapping was measured profile by profile using a laser profilometer over an area of 18×18 cm^2 in the center of the samples. The height of each point was measured using a laser giving a voltage output linearly proportional to the distance between the probe and the paper surface. The voltage output was converted to a floating point length measure using a 16-bit analog-to-digital converter. The laser diameter used was 30 μm; however, accuracy considerably smaller than this could be achieved. Each profile was acquired by sliding the sample under the probe while measuring. Multiple profiles were acquired by stepping the probe normal to the sliding direction. A typical one-dimensional height profile and a complete $(2+1)$-dimensional map are shown in Fig. 1. The number of points per profile was kept equal to the number of profiles, resulting in a square grid of measurements or "pixels." The number of points used for the various samples is given in Table I. The in-plane accuracy of each point was no larger than 10 μm for any sample, and in the out-of-plane direction it was 0.5 μm for all samples.

III. RIDGE DETECTION

A ridge stands out as a line of high curvature in an otherwise smooth landscape. The curvature of any point in a height profile $\xi(\vec{x})$ can be calculated as the field $\nabla^2\xi(\vec{x})$, where $\vec{x}=(x,y)$ are the planar coordinates. In the present case, it proved necessary to smooth the height profile $\xi(\vec{x})$ with a short-range Gaussian filter before calculating $\nabla^2\xi(\vec{x})$ in order to filter away small-scale features. A range of different filters was tested, and the result did not seem sensitive to the details of the filter. The main effect of the filtering was the removal of single isolated high-curvature pixels, or small groups of such, and a narrowing of the ridge lines. After filtering, the curvature field was calculated and thresholded so that all points over a given value were considered to have a unit value and all other points to have a zero value. Any isolated points above the threshold were filtered away. From the remaining points, lines were detected as ridges. It is throughout this paper assumed that all ridges are straight lines. It proved difficult to automate the ridge extraction process from the thresholded field; finally this step had to be done manually. Some statistics of the produced networks are listed in Table II. Figure 1 shows an example of a full ridge network. In the middle plot of Fig. 1, a single one-dimensional profile is given, and all points along this profile giving rise to ridges are marked. It can be seen from this figure that not all sections of the profile that have high curvature give rise to a ridge, while some smooth sections do give rise to a ridge. This may stem from the directionality of the ridges relative to the profile shown. Ridges crossing the profile at a small angle may seem smooth, but small local dents crossing close to orthogonally may seem large.

Nodes are defined as intersections between ridges, and a ridge therefore extends only from one node to another. All the links are regarded as undirected since a paper ridge does not have any preferred direction. The networks formed are fully connected and have therefore only one component.

IV. NETWORK PROPERTIES

The different paper thicknesses used in the experiments showed a clear trend that thinner paper crumple more than thick paper, and therefore produce more nodes and links (see Tables I and II). Apart from the scale of the network created,

TABLE II. List of extracted networks with their number of nodes, number of links, clustering coefficient C, the clustering coefficient for the corresponding planar Delaunay network, C_D, the clustering coefficient for a nonplanar randomized network with the same degree distribution, C_R, and the maximum node degree for the network. The clustering coefficients for the random networks were calculated using an average over 1000 samples after each sample had 10 000 random rewirings.

Sample	Nodes	Links	C	C_D	C_R	Maximum Degree
1	503	890	0.182	0.4371	0.0045	8
2	1211	2238	0.190	0.4315	0.0020	9
3	190	293	0.138	0.4458	0.0095	6
4	350	580	0.162	0.4394	0.0064	8
5	929	1829	0.231	0.4326	0.0029	10
6	286	501	0.199	0.4384	0.0083	8

no significant differences in the various distributions referred to below were detectable. As a consequence, most distributions are averaged over all samples after each of them have been normalized appropriately. The lack of change in the behavior due to sample thickness may arise from the small amount of data available, and no correspondence between paper thickness and other properties can be excluded. From a scaling point of view, a qualitative change of behavior is not expected since a large and thick sheet of paper is equivalent to a thin and small sheet. Note that the needed confinement also varies with the paper thickness, and all our experiments are conducted at approximately the same confinement. Sultan and Boudaoud [11] discuss two regimes for the crumpling process, depending on the confinement of the sample. The transition confinement is partly dependent on the paper thickness. Our experiments are as mentioned conducted at approximately constant confinement (although it is poorly controlled), and it might therefore be that due to the varying paper thickness our samples lie in different regimes. However, the uniform behavior of the samples indicates that they are all in the same regime. Also the number of self-contacts is very large for all the samples, and this indicates that they are all in the highly confined regime.

It can be seen from Table I that samples 3 and 4 both have the same paper thickness, although they have significantly different numbers of links and nodes. This is most likely due to the difference in confinement. Sample 3 had a larger ball radius than sample 4, and was therefore less confined, and also has fewer nodes and links than sample 4.

A. Ridge length

The length of a ridge between nodes a and b is defined as the spatial length from node a to b, following the assumption that all ridges are straight lines. Previous works have reported log-normal, Γ, and exponential functions [3,6,10] to give good fits for this distribution. However, we find that, whereas the small-scale part of the distribution is well fitted by a log-normal function, the tail of the distribution is not well fitted by any of the above-mentioned functions. The

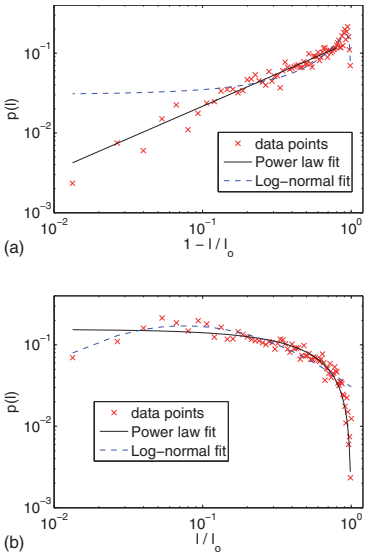


FIG. 2. (Color online) (a) Plot of the average noncumulative ridge length distribution $p(l)$ as a function of $1-l/l_0$, where l is the ridge length and l_0 is the maximum ridge length for any given sample. The data are fitted by a log-normal distribution and a power law $p(l) \propto (1-l/l_0)^{\beta}$ with $\beta=0.81$. (b) Plot of the average noncumulative ridge length distribution $p(l)$ as a function of normalized ridge length l/l_0.

large-scale part of the distribution is better fitted by a power-law function $p(l) \propto (1-l/l_0)^{\beta}$ where l is the ridge length and l_0 is the maximum ridge length for a given sample. Both fits can be seen in Fig. 2. We have found the tail to be best fitted by an exponent $\beta=0.81$. To compare the fits of the different functions they are plotted in Fig. 3 divided by the original distribution in order to emphasize any discrepancies.

The underlying reason for the shift in behavior may stem from the fact that the distribution of short ridges is dominated by remnants of originally long ridges. These ridges have been intersected by "younger" ridges crossing them after their formation. As outlined by Blair and Kudrolli [6], this random sectioning of ridges will give rise to a log-normal length distribution. The larger ridges, on the other hand, have not been so heavily sectioned by younger ridges. They are therefore not expected to follow the log-normal distribution of the shorter ridges. Instead, we detect a power-law dependency of the distribution of the difference between the longest ridge l_0 and the ridge length. It is reasonable to assume that larger samples will produce larger maximum ridges, and therefore l_0 is a sample-size-dependent quantity. Why this difference should exhibit a scale-free behavior is not clear.

B. Degree distribution

The degree of a node is defined as the number of ridges meeting at that node. The distribution has been found to have

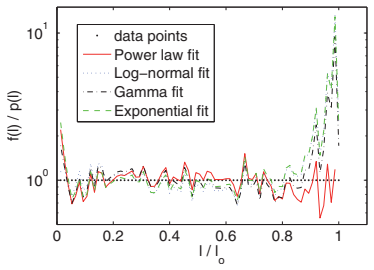


FIG. 3. (Color online) For comparison, the results shown in Fig. 2 from fitting the ridge length distribution $p(l)$ with Γ, log-normal, and exponential functions divided by the data themselves are shown, together with the same plot for the power-law fit as a function of normalized ridge length l/l_0.

a maximum probability at a median degree and produce a Gaussian-like form that is plotted in Fig. 4. The tail of the distribution is well fitted by a log-normal function of the same form as in Eq. (4). This is in strong contrast to many naturally occurring networks, which show a power-law tail, giving a larger portion of high-degree nodes than can be seen in the acquired samples.

C. Degree-degree correlation

The correlation between the degree of connected nodes has been studied using the procedure developed by Maslov and Sneppen [15]. They have defined a correlation measure

$$C(k_1,k_2) = \frac{P(k_1,k_2)}{P_R(k_1,k_2)}, \tag{1}$$

where $P(k_1,k_2)$ is the probability that a node of degree k_1 is linked to a node of degree k_2. $P_R(k_1,k_2)$ is the same average probability for a set of randomized networks. The randomized networks are assumed to have the same numbers of nodes and links, and the same degree distribution as the original network. A value $C(k_1,k_2)>1$ indicates that there is an over-representation of links between nodes with degree k_1 and k_2, whereas $C(k_1,k_2)<1$ indicates an under-representation. In order to look at the statistical signification of the correlation, Maslov and Sneppen introduced another correlation measure,

$$Z(k_1,k_2) = \frac{P(k_1,k_2) - P_R(k_1,k_2)}{\sigma_R(k_1,k_2)}, \tag{2}$$

where $\sigma_R(k_1,k_2)$ is the standard deviation of the samples used to generate $P_R(k_1,k_2)$. For $P(k_1,k_2)$ only the sample data are available. If a given coupling $P(k_1,k_2)$ is over-represented [that is, $P(k_1,k_2)>P_R(k_1,k_2)$] then $Z(k_1,k_2)>0$ and if it is under-represented $Z(k_1,k_2)<0$. If the standard deviation is small, the corresponding correlation coefficients are large, thus emphasizing statistically significant results. In all results presented here, 1000 randomized versions of the various samples were used to produce $P_R(k_1,k_2)$ and $\sigma_R(k_1,k_2)$. Each randomization used 10 000 rewirings of the original network.

Figure 5 shows the $C(k_1,k_2)$ matrix for all the samples. There is a tendency of small-degree nodes not to link to other small-degree nodes, but rather link to large-degree nodes. Links between large-degree nodes are also under-represented. This type of network is known as a disassortative network. Figure 6 shows the $Z(k_1,k_2)$ matrices for the same samples, and the same trends as in Fig. 5 can be observed. There is a clear trend in nearly all examined networks [16] that technical and biological networks such as the Internet and various protein interaction networks are disassortative, and that social networks such as acquaintance networks are assortative. The underlying reason for this is still not fully understood.

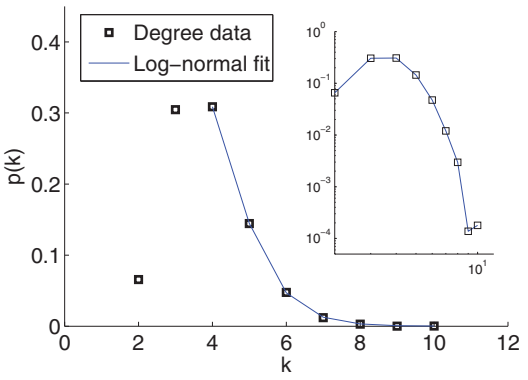


FIG. 4. (Color online) Plot of the degree distribution $p(k)$ as a function of node degree k with a fitted log-normal tail. The inset shows the same data plotted on log-log scale. This shows that the distribution does not have a power-law-distributed tail.

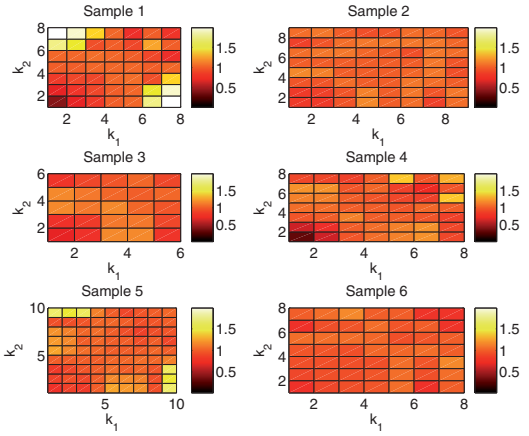


FIG. 5. (Color online) Correlation matrix $C(k_1,k_2) = P(k_1,k_2)/P_R(k_1,k_2)$ for each sample. The plots indicate that the networks are disassortative.

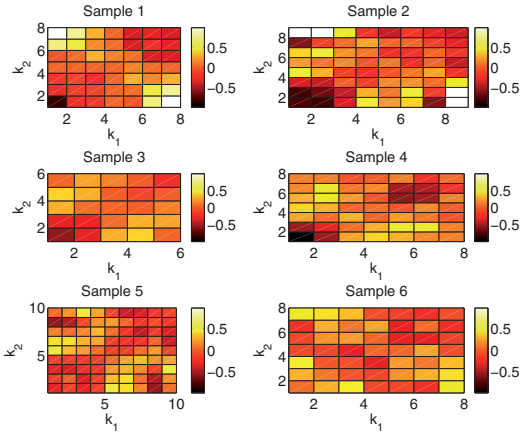


FIG. 6. (Color online) Z matrix $Z(k_1,k_2)=[P(k_1,k_2)-P_R(k_1,k_2)]/\sigma_R(k_1,k_2)$ for each sample. As in the $C(k_1,k_2)$ case of Fig. 5, the plot indicates a disassortative trend.

D. Clustering

The cluster coefficients C for all the samples are given in Table II, and they are all in the range 0.13–0.23. The definition used here is the standard

$$C = \frac{1}{N}\sum_{i=1}^{i=N} C_i, \tag{3a}$$

$$C_i = \frac{2E_{\mathrm{NN}}}{k_i(k_i-1)}, \tag{3b}$$

where N is the number of nodes in the network, E_{NN} is the number of links between nearest neighbors of node i, and k_i is the degree of node i [14]. A network embedded in two-dimensional Euclidean space with no crossing links is called a *planar network*, and has been described by West [17]. Generating a planar randomized network for comparing the cluster coefficients is very hard since no links can cross and the rewiring therefore must be local. However, the clustering can be compared with the Delaunay network [18] for the same spatial layout of nodes. For a given spatial node configuration and degree distribution, the Delaunay network gives the maximum possible clustering coefficient. The clustering coefficient for a Delaunay network made from nodes randomly distributed in the plane and with a number of nodes comparable to our samples is 0.44. Delaunay networks are closely linked to Voronoi diagrams, and both are described below. The cluster coefficient for a nonplanar random network, where the links can cross, having the same number of nodes and links and the same degree distribution is in the order of 0.001. The ridge networks have a much higher clustering than the nonplanar networks. This is expected because any node in a planar network has a low chance of being linked to a faraway node. This will generally increase the local clustering [17]. On the other hand the clustering is significantly lower than in the Delaunay case. This indicates that the ridge network does not form highly interconnected cliques.

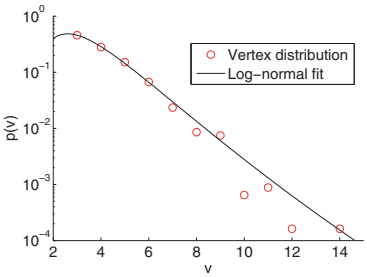


FIG. 7. (Color online) Average vertex distribution $p(v)$ as a function of the number of vertices v for the facets formed by the ridges. The data are fitted to a log-normal function with $\sigma=0.42$ and $\mu=1.13$.

V. GEOMETRICAL PROPERTIES

Various geometrical properties of crumpled thin sheets have been investigated in the past [6,19,20]. Here we discuss the size distribution of facets formed by the ridges and of the Voronoi sections formed by the location of the nodes. The angular distribution of the ridges and the three-cone structures is also investigated.

A. Facets

The nodes and links of the network form facets (also called domains) of various sizes and shapes. A facet is defined as an area of the crumpled paper confined by a closed loop of ridges that is simply connected, meaning that it contains no internal facets. The nodes bordering the facets are the corners or vertices of the facet. The distribution of facet areas and number of vertices for each sample have been calculated. The vertex distribution for all the samples was averaged, giving each sample equal weight. The number of facets with three, four, five, and six vertices was 46%, 28%, 15%, and 8%, respectively, and the number of facets with more than six vertices was 4%. The maximum number of vertices was 14. In Fig. 7 the distribution of the facet vertex number can be seen; the data are fitted with a log-normal function

$$p(a) = \frac{1}{\sqrt{2\pi}a\sigma} e^{-[\ln(a)-\mu]^2/(2\sigma)^2}, \tag{4}$$

where a is the vertex number, μ is the logarithm of the average number of vertices per facet, and σ is the standard deviation. The best fit was achieved with $\sigma=0.42$ and $\mu=1.13$.

The areas of the facets have also been investigated. The binned distribution of areas was normalized by the maximum area for each sample, and the average over all samples calculated. The resulting distribution can be seen in Fig. 8 together with a log-normal fit. The best fit parameters were $\sigma=1.17$ and $\mu=2.16$ in arbitrary units.

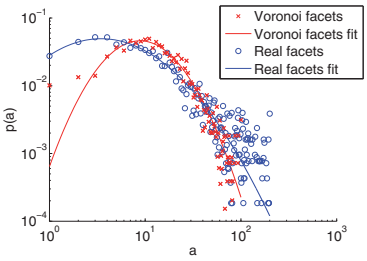


FIG. 8. (Color online) Average area distribution $p(a)$ as a function of the facet area a for the facets formed by the ridges and the Voronoi regions. Both sets of data are fitted by a log-normal function with $\sigma=1.17$ and $\mu=2.61$ for the facets and $\sigma=0.74$ and $\mu=2.73$ for the Voronoi regions, both in the same arbitrary units of area.

B. Voronoi networks

Given a set of nodes in space (or the plane) the Voronoi diagram [18] is a sectioning into areas around each node where each section contains all the points that are closest to the node in its interior. This partitions space (the plane) into sections filling the whole space (plane). The Delaunay network is a network where each node is linked to all the other nodes that it shares a Voronoi section border with. A visualization of this is given in Fig. 9, where the Voronoi diagrams for four of the samples are plotted. The gray scale of a given Voronoi section reflects the size of the section. Smaller sections have a lighter shade and larger sections have a darker shade. It can be seen that the sections are grouped according to size, making regions of the whole diagram that contains mainly large or small sections. The distributions of the areas of the various Voronoi sections have been calculated and fitted with a log-normal function. As in the facet case, each sample has been normalized by its maximum area. The distribution follows the same general shape as the facet distribution, and they can both be seen in Fig. 8.

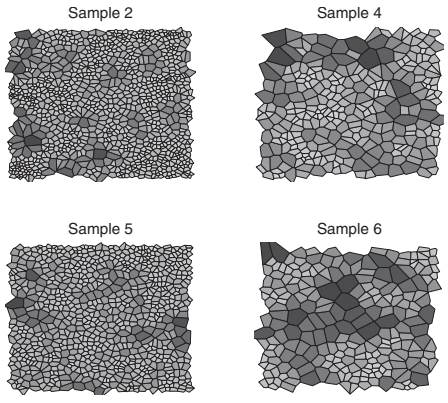


FIG. 9. Voronoi diagrams for four samples. The gray shading of the various sections represents the area of the sections. Lighter areas are smaller. There is a clear trend for sections of small (large) size to group with other small- (large-)sized sections.

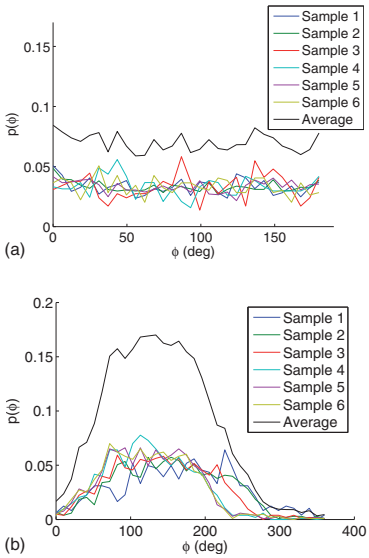


FIG. 10. (Color online) (a) Distribution of ridge angles, $p(\phi)$, as a function of angle ϕ for all the samples and the average (elevated for clarity). No trend is visible in the plot. (b) Distribution of separation angles, $p(\phi)$, as a function of opening angle ϕ in three ridge nodes. The average is elevated for clarity.

C. Angular distribution

The angular distribution of the ridges relative to the border of the sample has been studied in order to detect any preferred ridge direction or ordering among the ridges with regard to direction. No such preferred direction or ordering was found, and the distribution of ridge angles was reasonably uniform, both for each sample and for the average. A plot of the binned ridge angle distribution can be seen in Fig. 10.

The distribution of angles between ridges in a three-ridge cone (a node where three ridges meet, and hence form a conelike structure) has earlier been investigated both analytically and experimentally [6,19,20]. It has been reported that there are indications of preferred opening angles for such cones in the regions about 20°, 60°, and 110°, although all acquired distributions have been broad. All $k=3$ nodes have been investigated and the ridge separation angles show a broad distribution with a maximum in the range between 100° and 150°. There are no significant peaks in the distribution and this indicates a random ordering. However, 32% of all the angles lies in the interval between 90° and 150°. This suggests that the ridges tend to span out, trying to separate themselves from each other. Recall that 120° is the angle at which they are evenly separated. A plot of the distribution for all the samples and their average can be seen in Fig. 10.

VI. ROUGHNESS

The roughness of crumpled paper surfaces has been investigated before [4,6,10]. These investigations have reported self-affine behavior; this means that the surface is statistically characterized by

$$h(x) = \lambda^{-H} h(\lambda x), \tag{5}$$

where $h(x)$ is the height of the profile at position x, λ is a rescaling factor, and H is the Hurst exponent. We have investigated the one-dimensional profiles produced by the profilometer using the average wavelet coefficient (AWC) method [21], the power spectrum density (PSD) method [22], and the bridge method [23]. The results from all the methods indicate that the crumpled paper forms a self-affine surface. Earlier works have reported a small-scale region with a Hurst exponent $H_S \sim 1.0$ and a large-scale region with $H_L \sim 0.7$ [6,10] and $H_L \sim 0.8$ [4]. Our results follow the same trend in that there is a crossover scale between two scaling regimes. However, we found the small-scale exponent to be $H_S = 1.25 \pm 0.05$, indicating that the surface is asymptotically nonflat at these scales. Unfortunately, the data did not give a robust value for H_L because the sample size was too small compared to the crossover scale. The data did, however, indicate that $H_L < 1.0$ and in the range reported above. In Fig. 11, results from the PSD and AWC methods can be seen.

VII. CONCLUSION

The main points reported above are that the tail of the ridge length distribution is found to be well reproduced by a power-law distribution, and that the short ridges follow a log-normal distribution as reported earlier. The degree distribution has been shown not to have a power-law tail, but rather an exponential decay, and the networks have been found to be disassortative. The facet area distribution, the corresponding Voronoi diagram area distribution, and the Delaunay vertex distribution have all been found to fit log-normal distributions.

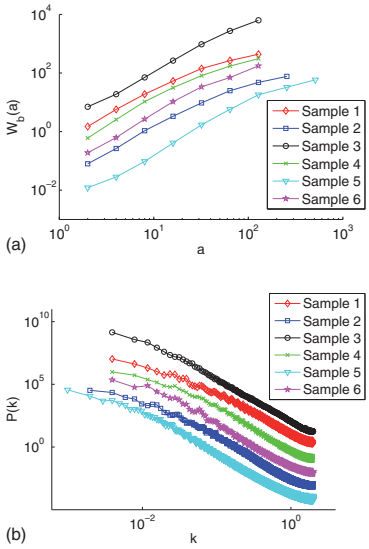


FIG. 11. (Color online) (a) Results from the AWC analysis showing the average wavelet coefficients $W_b(a)$ as a function of scale a; the scale represents measured points. (b) Results from the PSD analysis showing the power spectrum density $P(k)$ as a function of spatial frequency k. In both (a) and (b) the data sets are vertically shifted for clarity.

ACKNOWLEDGMENTS

This work have been supported by VISTA, a research cooperation between the Norwegian Academy of Science and Letters and Statoil. We would also like to thank J. Ø. Bakke and R. Toussaint for fruitful discussions, and the referees for valuable input.

[1] M. Ben Amar and Y. Pomeau, Proc. R. Soc. London, Ser. A **453**, 729 (1997).
[2] A. E. Lobkovsky and T. A. Witten, Phys. Rev. E **55**, 1577 (1997).
[3] A. J. WoodPhysica A **313**, 83 (2002).
[4] F. Plouraboué and S. Roux, Physica A **227**, 173 (1996).
[5] G. A. Vliegenthart and G. Gompper, Nat. Mater. **5**, 216 (2006).
[6] D. L. Blair and A. Kudrolli, Phys. Rev. Lett. **94**, 166107 (2005).
[7] L. Boué, M. Adda-Bedia, A. Boudaoud, D. Cassani, Y. Couder, A. Eddi, and M. Trejo, Phys. Rev. Lett. **97**, 166104 (2006).
[8] P. A. Houle and J. P. Sethna, Phys. Rev. E **54**, 278 (1996).
[9] J. P. Sethna, K. A. Dahmen, and C. R. Myers, Nature (London) **410**, 242 (2001).
[10] A. S. Balankin, O. S. Huerta, R. Cortes Montes de Oca, D. S. Ochoa, J. Martinez Trinidad, and M. A. Mendoza, Phys. Rev. E **74**, 061602 (2006).
[11] E. Sultan and A. Boudaoud, Phys. Rev. Lett. **96**, 136103 (2006).
[12] A. E. Lobkovsky, S. Gentges, H. Li, D. Morse, and T. A. Witten, Science **270**, 1482 (1995).
[13] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[14] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[15] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[16] Eur. Phys. J. B **38**, 143 (2004), special issue on ten leading questions for network research.
[17] D. B. West, *Introduction to Graph Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[18] F. Aurenhammer, ACM Comput. Surv. **23**, 345 (1991).
[19] E. Cerda and L. Mahadevan, Phys. Rev. Lett. **80**, 2358 (1998).

[20] S. Chaïeb, F. Melo, and J. C. Geminard, Phys. Rev. Lett. **80**, 2354 (1998).
[21] I. Simonsen, A. Hansen, and O. M. Nes, Phys. Rev. E **58**, 2779 (1998).
[22] A. R. Mehrabi, H. Rassamdana, and M. Sahimi, Phys. Rev. E **56**, 712 (1997).
[23] T. Engøy, K. J. Måløy, A. Hansen, and S. Roux, Phys. Rev. Lett. **73**, 834 (1994).

H. F. Hansen, C. A. Andresen, and A. Hansen,
A quantitative measure for path structures of complex networks,
Euro. Phys. Lett. **78**(4), 48005 (2007).

A Letters Journal Exploring the Frontiers of Physics

May 2007

EPL, **78** (2007) 48005
doi: 10.1209/0295-5075/78/48005

www.epljournal.org

A quantitative measure for path structures of complex networks

H. F. Hansen, C. A. Andresen and A. Hansen

Department of Physics, Norwegian University of Science and Technology - N-7491 Trondheim, Norway

received 15 Feburary 2007; accepted in final form 10 April 2007
published online 8 May 2007

PACS 89.75.Hc – Networks and genealogical trees
PACS 89.75.Fb – Structures and organization in complex systems
PACS 05.90.+m – Other topics in statistical physics, thermodynamics, and nonlinear dynamical systems

Abstract – In this paper we present a generalised version of the classical cluster coefficient, which also can be applied to networks with directed links. This generalisation takes into account more than the immediate nearest neighbours, giving more detailed information about the network structure than the classical version. The introduced concept is compared to earlier generalisation attempts, and it is applied to a directed version of the protein interaction network of the yeast cell *S. cerevisiae* and networks generated by the growing preferential attachment model of Barabási and Albert. Finally, we give some ideas on how our concept is related to modularity and community structures.

Copyright © EPLA, 2007

Introduction. – Over the past few years, a wide range of concepts and measures for complex networks have been proposed and investigated. However, complex networks are still most often described by three basic concepts. Perhaps the most important measure of complex, real-world networks is the degree distribution, $p(k)$ —the probability that a randomly selected node is connected to k other nodes [1]. For a surprisingly wide range of complex networks the degree distribution shows a scale-free character, described by a power law, $p(k) \sim k^{-\gamma}$, [2–8]. The second main characteristic is the small-world phenomenon, describing the fact that there is often a relatively short path between any two nodes in most networks. The maximum of the shortest paths between any two nodes in the network, referred to as the diameter, is often observed to grow logarithmically or slower with the network size, N [9]. Finally, the third main characteristic for complex networks is the cluster coefficient, which is related to the formation of cliques of linked nodes. The clustering in most real networks is observed to be considerably larger than the clustering in random networks [9].

More recently, there has been a considerable interest in understanding the structural properties of complex networks. This includes the study of local patterns, so-called, motifs [10–13], overrepresented compared to what is seen in random networks, and community structures [14–17], which can be described as modules of nodes interconnected by a relative small number of links, but where the nodes in each module are relatively highly linked to each other. Another field of interest has been networks with weighted links [18–22], and much attention has also been given to study the dynamical features of evolving networks [1,23].

As mentioned above, clustering is a central and well-studied concept in network theory. The classical cluster coefficient only concerns the formation of triplets and triangles in a given network. Many new ideas for a generalisation of the concept of clustering have been proposed, however most of them only apply to a specific type of networks or only look at a certain aspect of clustering. Some generalisations have successfully been applied to weighted networks where a cluster coefficient without a degree-correlation bias has been proposed [24]. Other examples include loop structures [25,26] and higher-order cluster coefficients taking into account more than the nearest neighbours of a node [27–29].

However, none of the above-mentioned attempts seems to satisfactorily produce a robust generalisation of the concept of clustering. The aim of this article is to present such a generalised measure that can be applied to all types of networks and give a broad range of information. We would like to introduce a quantitative measure for path structures in complex networks in order to better be able to describe and quantify the underlying structure and dynamics of networks. This measure, which contrary to the classical cluster concept includes long-range path structures, is motivated by the hope of giving insight to improve the understanding of clustering mechanisms.

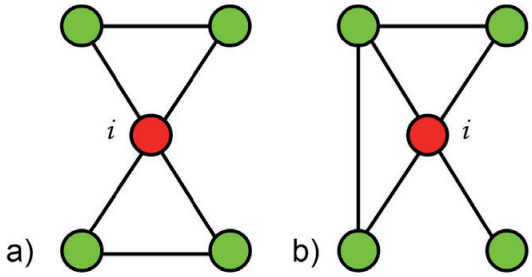


Fig. 1: (Colour on-line) Two small networks with the same number of links and nodes where the local clustering coefficient for the central node (marked red) is $C_i = 1/3$ even though the surrounding structure of the node is different.

The structure of this paper is as follows. First a motivation is given, then the new measure is introduced and compared to earlier generalisations. Then the concept is applied to a real-world protein interaction network and networks generated by the growing preferential attachment model of Barabási and Albert [30,31], showing the features of our idea. Finally we give some ideas about how our concept is related to modularity and community structures.

Motivation. – The classical cluster coefficient for a network with N nodes is defined [23] by

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i, \tag{1}$$

where

$$C_i = \frac{2E_{i,\mathrm{nn}}}{k_i(k_i-1)}, \tag{2}$$

where C_i is the local cluster coefficient for a given node i, k_i is the degree of node i and $E_{i,\mathrm{nn}}$ is the number of links between the nearest neighbours of node i. The sum in eq. (1) is restricted to nodes with degree larger than 1. As can be seen from the above definition, this concept only takes into account the immediate nearest neighbours of node i and the links between these nodes. From the networks displayed in fig. 1, it is clear that this is not always descriptive for the situation. The local structure for the two small networks are different when looking at the central node. However, the cluster coefficient given by eq. (1) is still the same for both cases.

The classical cluster coefficient is not defined for networks with directed links and this excludes a large number of networks. Another weakness in the classical definition can be seen when looking at regular networks such as those displayed in fig. 2. For the triangular network all nodes in the core of the network have a local clustering $C_i = 2/5$; however, for both the square and the hexagonal networks the cluster coefficient is zero. This in spite of the fact that these networks show clustering compared to random versions of the same networks. In

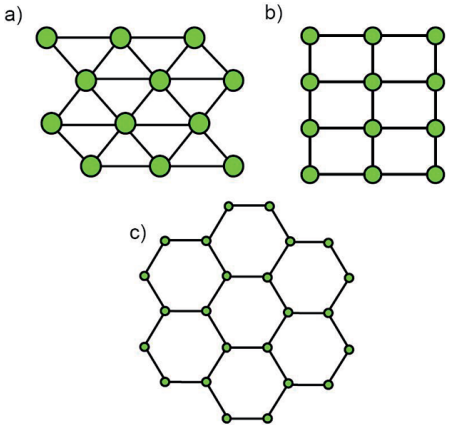


Fig. 2: (Colour on-line) A selection of regular networks: a) triangular network where all nodes not at the edge have a local clustering coefficient $C_i = 2/5$; b) square regular network where all nodes have a local $C_i = 0$; and c) hexagonal regular network where all local $C_i = 0$.

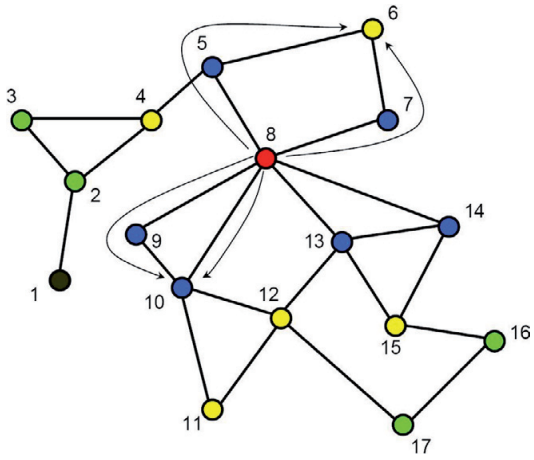


Fig. 3: (Colour on-line) Example network used for illustrating the definition and the generalised clustering concept. Note the colouring scheme that illustrates the radius from the central red node numbered 8. $k_{8,1} = 6$, $k_{8,2} = 5$, $k_{8,3} = 4$ and $k_{8,4} = 1$, resulting in $N_8^1 = 6$, $N_8^2 = 11$, $N_8^3 = 15$ and $N_8^4 = 16$.

fig. 3 an example network is given, and this network will be used for explaining the new concept and comparing it with the classical definition. Our concept reproduces the old definition as a special case. However, it also offers considerably more information about the clustering of a network, including the regular networks mentioned above.

Description. – We define a local sub-network characterised by a radius r, containing N_i^r nodes, around node i. We denote the number of node i's nearest neighbours, $k_{i,1}$,

node i's next nearest neighbours $k_{i,2}$, etc. For $r=1$ the sub-network only consists of node i's nearest neighbours, $N_i^1 = k_{i,1}$, for $r=2$ node i's next nearest neighbours are also included, $N_i^2 = k_{i,1} + k_{i,2}$, etc. This layering is illustrated in fig. 3 for the example network with node 8 as the central node. We find $N_8^1 = 6$, $N_8^2 = 11$, $N_8^3 = 15$ and $N_8^4 = 16$.

We relate the idea of clustering to paths in the network. A unique path is defined as a sequence of linked nodes where each node is visited only once. In fig. 3 node 10 can be reached directly with a path of length 1 from node 8, but it is also possible to go via node 9 to create a path of length 2. Other paths from node 8 to node 10 are also possible. In the same way, node 6 can be reached via two paths of length 2 from node 8, etc. As we have mentioned, this idea also makes is possible to look at clustering in networks with directed links. In networks with undirected links, any link is interpreted as two directed links going in opposite directions.

The question is: *How many paths of length n from node i are there to nodes which also can be reached with a path of length m*? The sub-network around node i consists of N_i^r nodes, not including node i itself. The potential number of paths of length 2 from node i is $N_i^r(N_i^r - 1)$, corresponding to $k_i(k_i - 1)$ for the classical cluster coefficient in eq. (2). The potential number of paths of length 3 is $N_i^r(N_i^r - 1) \times (N_i^r - 2)$, etc. Hence, the maximum number of potential paths of length n from node i is given by $\Pi_{j=0}^{n-1}(N_i^r - j)$, with $N_i^r \geqslant n$. The generalised cluster coefficient for node i, within radius r, is then given by

$$C^r(m,n) = \frac{1}{N}\sum_{i=1}^{N} C_i^r(m,n), \tag{3}$$

where

$$C_i^r(m,n) = \frac{p_i(m,n)}{\prod_{j=0}^{n-1}(N_i^r - j)}, \tag{4}$$

where $p_i(m,n)$ is the actual number of paths of length n to nodes that can also be reached by a m-path from node i. $C^r(m,n)$, as the classical cluster coefficient, is bounded by the unit interval, *i.e.*, $0 \leqslant C^r(m,n) \leqslant 1$. The sum in eq. (3) is restricted to nodes with $N_i^r \geqslant n$, giving a non-zero denominator. Note that the classical cluster coefficient is reproduced with $r=1$ and $(m,n) = (1,2)$, *i.e.*, $C^1(1,2)$. The definition generates symmetric results, *i.e.*, $C^r(m,n) = C^r(n,m)$. For the special case where $n=m$ paths of equal length is investigated, and the measure then describes the abundance of same length paths to nodes in the sub-network. As mentioned above, the central node of the two networks of fig. 1 has the same classical local cluster coefficient, $C_i^1(1,2) = 1/3$. However, the difference in the networks is reflected in the difference in the coefficients $C_i^1(1,3)$, which is 0 for network a and $1/12$ for network b.

Special cases of our generalised cluster coefficient grasp the same ideas as previously mentioned attempts to broaden the clustering concept. Both Bianconi and Capocci [25] and Fronczak *et al.* [27] are inspired by the growing network model of Barabási and Albert [30,31]. Bianconi and Capocci are focusing on the appearance of loops of length h, defined as a closed path of h links where each node is visited only once. This can, with undirected links, be related to our model by focusing on the cluster coefficient $C^r(\alpha, h-\alpha)$ with $0 < \alpha < h$ and $1 \leqslant r \leqslant \max(\alpha, h-\alpha)$. Fronczak *et al.* put forward an idea of a cluster coefficient of order x, involving the probability that there is a distance of length x between two neighbours of a given node. In our representation, this should be related to the coefficient $C^r(1, x+1)$. Jiang and Claramunt [28] are suggesting a k-cluster coefficient, involving the number of links among k neighbours of a given node. This could, only differed by an n-dependent factor, be related to $\sum_{n=1}^{n=k} C^k(n, n+1)$. In [29], Newman is focusing on so-called *Ego-centered networks*, illustrated by the concept of friends of friends, their friends, etc. Our coefficients $C^n(n, n+1)$ and $C^n(n,n)$ grasp much of this idea. Caldarelli, Pastor-Satorras and Vespignani [26], inspired by grid-like structures, are looking at so-called *quadrilaterals* —different cycles of length 4. Their *primary* and *secondary quadrilateral* seem to be grasping the same idea as $C^1(1,3)$ and $C^2(2,2)$, respectively.

Results. – We start by applying the general clustering concept to a protein interaction network from the *S. cerevisiae* yeast cell. This example was chosen because it belongs to a well-known class of biological networks that have been studied for some time. The network is treated as a set of directed links and therefore illustrates the application of our concept to this type of networks. The network data is made publicly available by Jeong *et al.* [32] on their web pages [33]. The network originally contained 1870 nodes; however, 412 of these did not belong to the giant component. The next largest component contained seven nodes, and we have therefore removed all nodes not belonging to the giant component. This left us with a network of 1458 nodes and 3941 directed links. The network has a scale free degree distribution [32], giving it a broad range of degrees. This class of networks have been shown to often display interesting properties such as the small-world property and high clustering [1]. The network is shown in fig. 4.

The results from the generalised cluster coefficient is in the form of a set of $n \times m$ matrices $C^r(m,n)$, one for each radius investigated. In table 1 these results are listed for radius $r=1$ up to $m=n=5$. The same type of results have been obtained for $r=3$ for later visualisation of clustering trends in the network, shown in fig. 5. For coefficients $C^r(m,n)$ with $r > \max(m,n)$ no new information is gained compared to results for lower r, with the exception of $C^r(r,r)$, because no new paths are included. Note that the number of nodes in the sub-networks created for each node during the calculation of $C_i^r(m,n)$ is much larger for $r=3$ than for $r=1$. In the

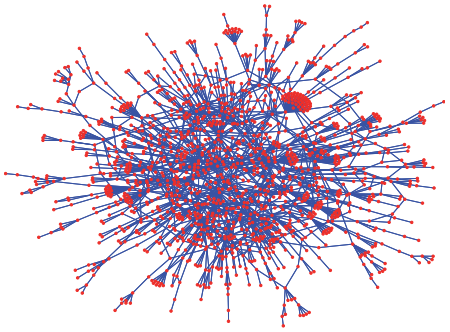


Fig. 4: (Colour on-line) Topological visualisation of the protein interaction network for the yeast cell *S. cerevisiae*.

Table 1: The cluster coefficients for the protein interaction network for the yeast cell *S. cerevisiae*, with $r=1$.

m	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
1	–	0.07083	0.01504	0.01697	0.00779
2	0.07083	0.01200	0.01504	0.01279	0.00701
3	0.01504	0.01504	0.00783	0.01250	0.00697
4	0.01697	0.01279	0.01250	0.01375	0.00779
5	0.00779	0.00701	0.00697	0.00779	0.00745

$r=1$ case the sub-networks will only contain the nearest neighbours of the node, and the size of the sub-networks will be in the order of the average connectivity $\bar{k}_{\rm out}$, here in the order of 3. For the $r=3$ case the number will be in the order of $\bar{k}^3_{\rm out}$, here in the order of 25, if we assume a uniform degree distribution. However, the degree distribution is not uniform, but scale free, and this produces hubs that significantly increases the size of the sub-networks for $r>1$. The average sub-network size for the giant component of the protein network is 2.7 for $r=1$ and 53.6 for $r=3$. Note that the result of $C^1(1,2)$ listed in table 1 for the giant component of the network strongly resembles the classical cluster coefficient. This value is found to be $7\cdot10^{-2}$, and the corresponding value is $5\cdot10^{-3}$ for a randomised version of the same network. These randomised networks have the same number of nodes and links and the same in- and out-degree distribution as the original network, but the links have been randomised. For a truly random undirected network (the Erdös-Rényi model [34]) the cluster coefficient is shown to be in the order of $\bar{k}/N$ [35], where $\bar{k}$ is the averages degree and N is the number of nodes. For the studied yeast network $\bar{k}/N$ is in the order of $2\cdot10^{-3}$. We can therefore conclude that the protein interaction network is rather well connected, having a cluster coefficient that is more than ten times that of a comparable random version.

In order to visualise any trends in the clustering, we compare the results from the real-world network with results from the randomised versions of the same network. The average clustering of the randomised networks is denoted $C^r_{\rm R}(m,n)$ and is calculated as an average over 100 samples. Any trends will then be apparent in the measure $W^r(m,n)=C^r(m,n)/C^r_{\rm R}(m,n)$. In fig. 5 $W^r(m,n)$ is plotted for $r=1$ and $r=3$. We see that for small r the number of paths in the original network with high m and n far exceeds the case for the randomised networks. For larger r more nodes are included in the calculation, and there is an abundance of small m large n paths (and vice versa), indicating that there are many paths to the near neighbours via further away neighbours. This seems to indicate that if protein a is directly linked to protein b, the indirect influence from a to b via some intermediate proteins is larger than in a random network. In terms of robustness, this could be seen as a sign of stability since the influence on b form a is upheld even though the direct link or some of the indirect links are broken, because there is an anomalous abundance of indirect links.

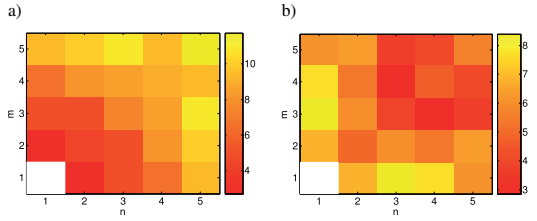


Fig. 5: (Colour on-line) Results from the calculation of the ratio $W^r(m,n)=C^r(m,n)/C^r_{\rm R}(m,n)$ for two different r and for m and n in the range 1 to 5. In plot a) for $r=1$ and in plot b) for $r=3$. The values plotted are the natural logarithm of $W^r(m,n)$.

Next we apply the clustering concept to the well-studied growing Barabási-Albert model based on preferential attachment [30,31]. Our Barabási-Albert networks were initialised with a cluster of $m_0=5$ fully connected nodes, and thereafter nodes were included node by node, each given two new links to the existing network using the preferential attachment rule. Results for the coefficients $C^1(1,2)$ and $C^2(2,2)$, as a function of network size, can be seen in fig. 6, each result is averaged over 200 samples. As can be seen in fig. 6, the coefficient $C^1(1,2)$ follows a power law $C^1(1,2)\sim N^{-\beta}$, with $\beta=0.73$. This is in good agreement with what has been seen previously with $\beta=0.75$ [35], and should come as no surprise since $C^1(1,2)$ is the classical cluster coefficient in a network with undirected links.

The coefficient $C^2(2,2)$, which also can be seen plotted in fig. 6, does not seem to decrease with increasing network size, but saturates at a finite value. This shows that the importance of "friend of a friend" type connections does not diminish as the system size increases, in contrast to the classical "my friend, your friend" connections. This may indicate that the Barabási-Albert model incorporates more clustering mechanisms that previously thought.

Community structures in networks is a feature describing the fact that there within groups of nodes is a higher link density than what one would expect in a randomised

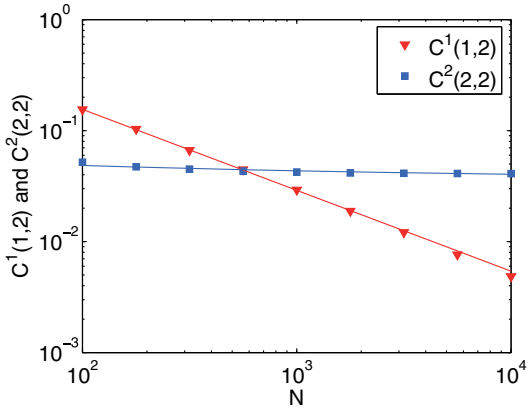


Fig. 6: (Colour on-line) The coefficients $C^1(1,2)$ and $C^2(2,2)$ as function of the system size for the BA model. Note that $C^1(1,2)$ follows a power law as is predicted for the classical cluster coefficient with a exponent of 0.73 and that $C^2(2,2)$ seem to asymptotically reach a finite value.

version of the same network. These nodes then form different communities which are interconnected by fewer links than one would expect if the links were placed on the basis of random chance [17].

The quantitative measure for community structures is known as modularity. Only differed by a multiplicative constant, the modularity measures the number of links within groups minus the expected number of links within the same groups if the links had been placed at random. This is related to randomised networks and not pure random networks. Randomised networks have the same number of nodes and links as the original network. In addition the degrees of the nodes are conserved, but the links have been randomised. Precise definitions of modularity can be found in [15] and [17].

It seems that our measure $W^r(m,n)$ might touch upon the concept of modularity since $W^r(m,n)$ measures the ratio between the generalised cluster coefficients $C^r(m,n)$ and cluster coefficient $C^r_R(m,n)$ in the corresponding randomised network. For a certain r_0, the measure $W^r(m,n)$ will be of the order of 1. At this point the average sub-network with radius r behaves as a sub-network of a randomised version of the original network. This indicates that we are including nodes outside the local module when increasing r beyond r_0, and hence r_0 could be interpreted as an indication of an average module radius of the network.

Summary and conclusion. – In this paper we have introduced a generalised version of the classical cluster coefficient. The classical concept can be retrieved as a special case of the generalised version. It should be stressed that our generalisation also can be applied to networks with directed links. The classic definition in eq. (1) does not apply to this very important class of networks, and it has therefore been difficult to quantify the clustering of such networks. Our definition can therefore be used to quantitatively describe the clustering of both directed and undirected networks. The new concept also gives additional information about the underlying structure within the networks, involving more than just the immediate neighbourhood of the nodes in the clustering idea. This gives information about more far-reaching structures, and looks at how the network is connected at a larger scale. To illustrate the usefulness and application of our generalisation, we have used it to investigate the clustering in the protein interaction network of the yeast cell *S. cerevisiae*. The network is shown to display a high order of clustering and non-trivial long-range path structures which depend on the size of the sub-networks used to calculate the local cluster coefficients. We also applied the measure to look at the clustering in the well-studied Barabási-Albert model of growing networks with preferential attachment. Our results may indicate that the Barabási-Albert model is better at modelling clustering mechanisms than previously thought. We have also shown that our measure can be related to modularity and community structures, with an indication of an average module radius of the network.

* * *

We thank the Center for Complex Network Research at University of Notre Dame for making the protein network data publicly available.

REFERENCES

[1] Boccaletti S. *et al.*, *Phys. Rep.*, **424** (2006) 175.
[2] Albert R., Jeong H. and Barabási A.-L., *Nature*, **401** (1999) 130.
[3] Newman M. E. J., *Phys. Rev. E*, **64** (2001) 016131; 016132.
[4] Liljeros F., Edling C. R., Amaral L. A. N., Stanley H. E. and Aberg Y., *Nature*, **411** (2001) 907.
[5] Albert R. and Barabási A.-L., *Phys. Rev. Lett.*, **85** (2000) 5234.
[6] Amaral L. A. N., Scala A., Barthélémy M. and Stanley H. E., *Proc. Natl. Acad. Sci. U.S.A.*, **97** (2000) 11149.
[7] Redner S., *Eur. Phys. J. B*, **4** (1998) 13.
[8] Maslov S. and Sneppen K., *Science*, **296** (2002) 910.
[9] Watts D. J. and Strogatz S. H., *Nature*, **393** (1998) 440.
[10] Milo R. *et al.*, *Science*, **298** (2002) 824.
[11] Shen-Orr S. S. *et al.*, *Nature Genetics*, **31** (2002) 64.
[12] Milo R. *et al.*, *Science*, **303** (2004) 1538.
[13] Albert I. and Albert R., *Bioinformatics*, **20** (2004) 3346.
[14] Girvan M. and Newman M. E. J., *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 7821.
[15] Newman M. E. J. and Girvan M., *Phys. Rev. E*, **69** (2004) 026113.
[16] Bagrow J. P. and Bollt E. M., *Phys. Rev. E*, **72** (2005) 046108.

[17] Newman M. E. J., *Proc. Natl. Acad. Sci. U.S.A.*, **23** (2006) 8577.
[18] Newman M. E. J., *Phys. Rev. E*, **70** (2004) 056131.
[19] Csermely P., *Trends Biochem. Sci.*, **29** (2004) 7.
[20] Serrano M. A., Boguna M. and Pastor-Satorras R., *Phys. Rev. E*, **74** (2006) 055101.
[21] Li M. *et al.*, *Physica A*, **375** (2007) 355.
[22] Wu A.-C. *et al.*, *Chin. Phys. Lett.*, **24** (2007) 577.
[23] Newman M. E. J., *SIAM Rev.*, **45** (2003) 167.
[24] Soffer S. N. and Vazquez A., *Phys. Rev. E*, **71** (2005) 057101.
[25] Bianconi G. and Capocci A., *Phys. Rev. Lett.*, **90** (2003) 078701.
[26] Caldarelli G., Pastor-Satorras R. and Vespignani A., *Eur. Phys. J. B*, **38** (2004) 183.
[27] Fronczak A. *et al.*, *Physica A*, **316** (2002) 688.
[28] Jiang B. and Claramunt C., *Environ. Plann. B*, **31** (2004) 151.
[29] Newman M. E. J., *Soc. Network*, **25** (2003) 83.
[30] Barabási A.-L. and Albert R., *Science*, **286** (1999) 509.
[31] Barabási A.-L. and Albert R., *Physica A*, **272** (1999) 173.
[32] Jeong H., Mason S. P., Barabási A.-L. and Oltvai Z. N., *Nature*, **411** (2001) 41.
[33] Network data was downloaded from `http://www.nd.edu/~networks/resources.htm`, on January 8th 2007.
[34] Erdös P. and Rényi A. R., *Publ. Math. Inst. Hung. Acad. Sci.*, **6** (1960) 17.
[35] Albert R. and Barabási A.-L., *Rev. Mod. Phys.*, **74** (2002) 47.

C. A. Andresen, H. F. Hansen, A. Hansen,
G. L. Vasconcelos, and J. S. Andrade Jr,
Correlations between political party size and voter memory:
A statistical study,
Int. Jour. Mod. Phys. C **19** (11) 1-11 (2008).

Is not included due to copyright

C. A. Andresen, R. Le Goc, P. Davy, and A. Hansen,
Network analysis of outcrop data,
Submitted to Journal of Geophysical Research - Solid Earth

Is not included due to copyright

C. A. Andresen, and A. Hansen,
Network properties of Discrete Fracture Network Model,

Is not included due to copyright