Warsaw University

Faculty of Mathematics, Informatics and Mechanics

Jarosław Piersa

Scale-freeness and small-world phenomenon in information-flow graphs of geometrical neural networks

PhD dissertation

Supervisor:
prof. dr hab. Piotr Bała
Faculty of Mathematics and Computer Science
Nicolaus Coppernicus University

Co-supervisor:
dr Filip Piękniewski
Faculty of Mathematics and Computer Science
Nicolaus Coppernicus University

June 25, 2012

Author's declaration:
aware of legal responsibility I hereby declare that I have written this dissertation myself and all the contents of the dissertation have been obtained by legal means.

.....	.....
date	Author's signature

Supervisor's declaration:
the dissertation is ready to be reviewed.

.....	.....
date	Supervisor's signature

Contents

Abstract

In this dissertation we set out to study a simplified model of activation flow in artificial neural networks with geometrical embedding. The model provides a mathematical description of abstract neural activation transfer in terms, which bear resemblances to multi-value Boltzmann-like evolution. The activation-preserving constraint mimics a critical regime of the dynamics and, along with accounting for geometrical location of the neurons, makes the system more feasible for modelling of real-world networks.

We focus on *scale invariance* or *scale-freeness* and *small-world phenomena* in the said networks. Our results clearly confirm presence of both features at the functional level of the activity-flow graph. We show that the degree distribution preserves a power-law shape with the exponent value approximately equal to $\gamma \simeq -2$. In addition, we present our results concerning *characteristic path length* in the said graphs, which grows roughly logarithmically with the size of the network, while the *clustering coefficient* turns out to be relatively high. Taken together, the clustering and path length ratios are surprisingly high, and thus confirm large both local and global efficiency of the network.

Finally, we compare the properties of activation-flow model to those reported in neurobiological analyses of brain networks recorded with functional magnetic resonance imagining (fMRI). There is a strong agreement between the shape and exponent value of degree distribution also the clustering and characteristic path lengths are comparable in both the model and medical data.

Keywords: neural networks, small-world phenomenon, scale-free network

AMS Mathematics Subject Classification 2010: 92B20 (Neural networks, artificial life and related topics), 82C32 (Neural nets), 05C80 (Random graphs), 05C82 (Small world graphs, complex networks).

Streszczenie

Celem niniejszej rozprawy jest analiza uproszczonego modelu przepływu aktywności w sztucznych sieciach neuronowych zanurzonych w przestrzeni geometrycznej. Przedstawiony model dostarcza matematycznego opisu transferu aktywności w terminach zbliżonych do wielowartościowych maszyn Boltzmanna. Wymóg zachowania stałej sumarycznej aktywności odzwierciedla krytyczność dynamiki i wraz z uwzględnieniem wpływu lokalizacji geometrycznej neuronów sprawia, że system jest bardziej adekwatny do modelowania rzeczywistych sieci.

Badania koncentrują się na *bezskalowości* oraz *fenomenie małego świata* w wyżej wymienionych sieciach. Uzyskane rezultaty potwierdzają obecność obu własności w omawianych grafach. Pokażemy, że rozkład stopni wejściowych wierzchołków zachowuje się jak funkcja potęgowa z wykładnikiem równym $\gamma = -2$. Ponadto prezentujemy wyniki dotyczące *charakterystycznej długości ścieżki*, który rośnie logarytmicznie wraz z wielkością systemu, podczas gdy *współczynnik klasteryzacji* okazuje się dość duży. W konsekwencji stosunek klasteryzacji do długości ścieżek jest zaskakująco wysoki, co jest dystynktywną własnością sieci małego świata.

Wreszcie, dokonujemy porównania cech omawianego modelu przepływu aktywności z neuro-biologicznymi rezultatami, przedstawionymi w badaniach grafów mózgowych z danych uzyskanych z funkcjonalnego obrazowania z wykorzystaniem rezonansu magnetycznego (fMRI). Wskazujemy silną odpowiedniość pomiędzy kształtem i wartością wykładnika rozkładu stopni, zaś klasteryzacja i charakterystyczna długość ścieżki są porównywalne w modelu i danych medycznych.

Słowa Kluczowe: sieci neuronowe, grafy bezskalowe, grafy małego świata

Klasyfikacja tematyczna AMS 2010: 92B20, 82C32, 05C80, 05C82.

Chapter 1

Introduction

1.1 Aim of the thesis

In this work we set out to study a simplified model of activation-flow in the context of question whether scale-freeness and small-world phenomena occur in sparse recurrent neural networks with geometrical embeddings.

We expect the answer to this question to be positive in both cases (i.e. both scale-freeness and small-world phenomenon). In such case, this would make the model comparative and adequate to mimic medical results of the brain imagining. So far, the medical data of the functional graphs of human brain confirm these features, but there is a limited explanation of them.

In order to address the raised question, we adopt numerical simulations and, whenever possible, mathematical analysis tools. The quantitative description of the model is highly reliant on random graph theory, which is a frequently chosen for large-scale graphs. The computational aspect of the simulation seems to match the algorithmic definition of the model, so the numerical analyses are also reasonable. Moreover, we take advantage of the mathematical tools, mainly from the area of statistical mechanics, to discuss some of the properties of the model.

Tu summarise

Aims of the work:

- to study a mathematical model of activation-flow in neural networks with geometrical embedding,
- to investigate the presence of the scale-free phenomenon in the said networks,
- to investigate the presence of small-world phenomenon in the networks.

Methodology

- description in terms of random graph theory,
- numerical simulations to empirically test the properties of the model,
- mathematical analysis of some of the predictions.

Results

- both phenomena have been confirmed.

The dissertation itself can be seen as a continuation of the work of Piękniewski [Piękniewski 2008], however the main results are obtained with different mathematical means and the model is more elaborate. In addition, the focus of this work is also extended into small-world analyses and, as mentioned above, accounts for geometrical embedding.

1.2 Motivation

Artificial neural networks are biologically inspired mathematical models for organic neural cells and brain systems. The area of the topic is also extended to their application in problem solving and machine learning. Depending on required accuracy and available computing power, the models vary from discreet algebraic operations, covered extensively in book [Rojas 1996], to set of coupled differential equations integrated over time [Izhikevich 2007]. Though the expressive power of the single neuron is quite limited, the real capabilities are at their peak when the neurons are connected into a complex network. The interconnection dynamics can again be expressed in terms ranging from weighted averages to differential diffusion equations.

Other aspect of artificial networks is their emergent tendency to self-organization. This led to arising a range of stochastic algorithms such as unsupervised learning [Hebb 1949], principal component analysis, k-means [MacQueen 1967], Kohonen mappings [Kohonen 1982] etc.

This self-organisation phenomenon is a subject of both theoretical and empirical research Not only neural networks, but also a collection of large-scale real-world graphs was studied in this context and resulted in a formulation of their properties in a mixed statistical and graph-theoretical terms. The possibility of obtaining the said properties from simple graph-construction rules led to Watts-Strogatz, Albert-Barabasi and other self-organisation-driven graph models [Albert & Barabasi 2002].

The above cases are vital extension of the random graph theory. While the graph theory dates back to XVII-th century, the random graph theory is

connected to work of Erdős and Rényi from the half of the twentieth century [Erdős & Rényi 1959]. The branch sets out to analyse graphs, which are too large to be precisely described by sole quantitative features. The list of examples includes, but is not limited to linguistic networks [Kello & Beltz 2009], metabolic networks [Csemely 2009], co-authoring, co-starring, acquaintanceship, power lines etc. Note that, all of these are not a result of strictly supervised process, quite opposite — they spontaneously evolved into their final shape (and in some cases they still evolve).

Analyses of their properties led to rapid development of the random graph theory in the second half of the twentieth century. Due to gigantic sizes and almost continuous growth, they required a description in qualitative rather than quantitative terms. For instance shape of the degree distribution of the vertices, instead of fixed numbers. The uniform or gaussian structure would seem natural for such graphs, but they turn out to be highly hierarchical. As a result their degree distributions are frequently characterised by heavy tails of high values. Such sort of behaviour is preserved throughout all ranges of scale and therefore is referred as scale-invariance or scale-freeness.

Another frequently examined feature is high global and local 'efficiency' of the network along. The global one — a characteristic path length is related to number of edges, needed to reach any pair of vertices, the local one — a clustering coefficient is derived from the presence of local clique structures in the graph. Clearly, the both concepts are trivial in full graphs, however most of the empirical networks are sparse. Despite their sparsity most of reports in large-scale graph confirm short average path length and high local clustering.

This phenomenon is best known in social networks as so called *six handshakes distance*. As the name suggests the maximum distance between any pair of people is six handshakes, which means one can pass a message for anyone just by at most five middle-men of which the sender knows the first one, the first middle-man knows the second one, etc. and the last one knows the recipient. This phenomenon was deeply studied in the twentieth century [Gurevitch 1961] and led to the development of the *small-world graph* concept [Watts & Strogatz 1998].

In 2004 the empirical data obtained from human brain activity was analysed with random graph theory tools. Eguiluz et al. confirmed the scale-free dependencies in activity graphs obtained by functional magnetic resonance imagining (fMRI) devices from human patients [Eguiluz et al. 2005]. While it was neither first nor the last power-law report, it brought the attention of neuroscientists to the notion of random graphs. In addition, it raised the problem of distinction between the functional and structural brain networks.

More precisely the availability of some component (a neural cell, a synapse, etc.) in brain structure does not indicate, that it is always used during the dynamics. This is due to a unsupervised dynamics process and brain plasticity, which is mandatory for learning. Hence this is the point, there the notions of self-organisation in neural networks and random graphs meet.

1.3 Thesis structure

The dissertation is organised as follows.

In Chapter 2 we briefly sketch the main concepts of the random graph theory, which will be used in following chapters. We provide a formal definition of characteristic path lengths, clustering coefficient and other mathematical tools, which are used to provide a qualitative description of large scale networks. Finally, we define the Erdős-Rényi and construction-based random graph models, and discuss their properties.

In Chapter 3 we provide a brief reiteration of artificial neural networks, in whose context an activity-flow model will be built and analysed. We list classical spiking (Hodgkin-Huxley, Quadratic Integrate-and-Fire, Izhikevich) and firing-rate (perceptron, feed-forward, Hopfield networks) models of neurons and networks. The chapter is concluded with a summary of the state of the arts in contemporary neuroscience.

In Chapter 4 the main results of the thesis is discussed. We define a *simplified activity-flow model* in the recurrent network, discuss the geometrical embedding and formulate the stochastic dynamics. Then we discuss the scale-freeness and small-world phenomenon in the activity-flow graph, and compare these results to those obtained in medical counterparts from fMRI data.

Finally, we summarise our results in Chapter 5 and point out potential aims of forthcoming researches.

Technical remarks, focused mainly of parallelisation, are outlined in Appendix section.

Chapter 2

Elements of random graph theory

In this Section we formulate the basic elements random graph theory, which branched out from the graph theory in the mid of twentieth century. We provide a formulation and properties of Erdős-Rényi model, discuss the characteristics of the random graphs and briefly list contemporary models, which will be used in analyses of the small world phenomenon and scale invariances analyses in further chapters.

For any additional information concerning random graphs the reader is referenced to the monograph [Chung & Lu 2006]. Note, that the 'classic' graph theory is deliberately omitted and should the reader require such knowledge, we reference him to handbooks for instance [Deistel 2006] or in the case of algorithmic applications to [Sysło et al. 2006].

2.1 Random graph theory

The 'classical' graph theory dates back to XVIII-th century and Leonhard Euler. As the technical development continued, the analysed graphs became larger and larger and reached the point, where they turned out too complex to be analysed quantitatively.

In second half of the twentieth century, a Hungarian mathematician Paul Erdős looked on large-scale graphs in a statistical way. Having assumed that the edges appear randomly, he analysed probability that a graph is connected [Erdős & Rényi 1959], its degree distribution etc. With his numerous works he laid foundations of the random graph theory. This branch of mathematics has been developing quite rapidly since then, finding its applications in real-world large scale networks modelling and being extended onto more elaborate models.

Among the most characteristic figures used to describe random graph one

can mention the degree distribution, characteristic and maximal path length, clustering coefficient etc.

2.1.1 Definition of the ER model

Definition Let $n \in \mathbb{N}_{\geq 1}$ and $p \in (0,1)$. A Erdős-Rényi random graph with n vertices and probability p is a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, ..., n\}$ and each edge is added independently with probability p — $\forall_{u,v \in \mathcal{V}} \quad \mathbb{P}(\{u, v\}) \in \mathcal{E}) = p$.

Erdős-Rényi graphs assume almost complete homogeneity in the network with little, to no variety of the vertices. On one hand this makes the model relatively easy to analyse and implement. On the other, mathematical models based on this graph might turn out to omit subtle dependencies such as hierarchy or outsiders. In addition ER model assumes a static graph behaviour. Nonetheless, the ER model is still a useful tool for analyses of the real-world processes.

2.1.2 Random graph characteristics

Degree distribution

Definition The degree distribution of the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a discreet probability distribution of the degrees of the graph vertices

$$\mathbb{P}(deg(v) = k). \tag{2.1}$$

When a graph is given (for instance a real network) the notion is a bit abused and applied to *empirical degree distribution* or *normalised histogram* of the vertices.

With size of the graphs approaching infinity, the degrees can also became large. Some features like average connectivity (assuming the distribution has finite mean value) can be inferred from the degree distribution. With some additional assumptions one can tell if the graph is connected.

Characteristic path length

A path in graph is defined as:

Definition A path in graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ between vertices u and v is a sequence of distinct edges $p_{uv} = (e_1 = \{u_1, v_1\}, e_2 = \{u_2, v_2\}, ..., e_k = \{u_k, v_k\})$, such that: $u_1 = u$, $v_k = v$, $\forall_{i=1..k-1} v_i = u_{i+1}$ and $\forall_{i \neq j} u_i \neq u_j, v_i \neq v_j$ and $\forall_{i=1..k} e_i \in \mathcal{E}$.

Definition A characteristic path length in graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ or an average path length is defined as an average of the shortest path lengths between any pair of vertices in $\mathcal{V}$. We denote it as $l(\mathcal{G})$ or simply l.

Definition A maximal path length in graph or graph diameter $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined as an maximum of the shortest path lengths between any pair of vertices in $\mathcal{V}$. We denote it as $d(\mathcal{G})$.

The characteristic path length can be used as rigorous formulation of so-called *six degrees of separation phenomenon*, which was noticed in early XX-th century. First, it was related to general social graphs, with people as vertices and an event of exchanging a handshake as an edge [Csemely 2009]. The famous 'six degrees of separation' term estimates the l of the handshake graph for the whole planet to be 6. These results have been partially confirmed with survey analysis [Gurevitch 1961] and Monte Carlo simulations, new researches carried on the Facebook social graphs also confirm this hypothesis [Ugander et al. 2011], we should however note, that the precise results for such wide population have not been calculated yet.

Humble attempts were made to calculate this value for smaller, specialised graphs, of which the best-known examples are scientific collaboration and actor co-starring graphs. In the former the edge is defined as a co-authorship of at least one scientific publication [Barabasi et al. 2002], in the latter — the event of casting in at least one cinema film [Herr et al. 2007]. Pretty much as in previous researches, the characteristic path lengths turned out to be short compared to the size of the sample. In 1998 Barabasi et al. obtained $l \simeq 9$ for mathematical collaboration graph, consisting of over 70 000 authors. While this is above 6 one should keep in mind that, this is still a subgraph of the social network and the lifetime of mathematical productivity is shorter than the lifetime of acquaintanceship.

Rapidly and quite spontaneously developing Internet soon became a centre of attention for graph scientists. The directed nature of the hyperlinks requires to separately consider input and output interlinks in documents, which in graph term needs input $deg_{in}(v)$ and output degrees $deg_{out}(v)$. Nonetheless, despite its directed nature, the characteristic path length in World Wide Web is still orders of magnitude lower than the network size. Albert et al. obtained $l \simeq 19$ for a sub-network of WWW counting over $3 \cdot 10^5$ documents, see [Albert et al. 1999]. As mentioned above, Ugander et al. analysed social graph of Facebook counting $721 \cdot 10^6$ active users and obtained $l \simeq 4.3$, see [Ugander et al. 2011].

Recent research with set out to analyse characteristic path length in functional graphs of brain activity. The obtained value l was roughly between

2.5 and 11, which is still a very low result, see [Basset & Bullmore 2006]. It is worth emphasising, that fMRI imagining offers low-to-medium spatial resolution and the amount of obtained functional areas was not particularly high.

The short characteristic path length implies a high average efficiency of the transport throughout the network (the diameter can be taken as a pessimistic efficiency). One should however tell the difference between existence of the short path, finding a proper next point on the route, while knowing the *local network only* and finding the shortest path with knowledge of the *whole network*. Clearly the latter is a polynomial time problem [Sysło et al. 2006]. The former is a subject of heuristic techniques [Russel & Norvig 1995] and whenever possible substituted with a full data collection and a classic search algorithm. For instance the BGP and OSPF routing protocols, except for transmitting packets, gather and exchange routing information in order to find the best path [RFC 4271, RFC 2328].

Clustering Coefficient

Definition A clustering coefficient of the node $v \in \mathcal{V}$ is defined as a fraction of all edges present in neighbourhood of v to all possible edges in the neighbourhood.

$$C(u) = \frac{|\{(w,v) \in \mathcal{E} : (u,w) \in \mathcal{E} \vee (u,v) \in \mathcal{E}\}|}{\frac{|\{w \in \mathcal{V}:(u,w)\in\mathcal{E}\}|(|\{w\in\mathcal{V}:(u,w)\in\mathcal{E}|-1)}{2}} \tag{2.2}$$

Definition A clustering coefficient of the graph $G = (V, E)$ is defined as as an average of all clustering coefficients of the graph vertices.

$$C = \frac{1}{|V|} \sum_{v \in V} C(v) \tag{2.3}$$

The clustering coefficient has a clear interpretation of local cliquishness or local transport efficiency of the graph. Effectively, it describes how well is the graph connected in the neighbourhood of the vertices. The full graph K_n has its maximum value $C = 1$. The empty graph, all trees and forests (unions of trees) have the lowest possible $C = 0$. Regular lattices (except for rectangular grid) usually have relatively high local clustering coefficient despise being arbitrary sparse.

Since the diagonal of the cube of the adjacency matrix $B = [b_{ij}]_{i,j=1}^{n} = A^3$ consists of numbers of closed paths of length 3 the clustering coefficient can be calculated as:

$$C(v_i) \simeq \frac{2b_{ii}}{deg_i \cdot (deg_i - 1)}. \tag{2.4}$$

Another scheme after [Newman et al. 2001]

$$C \simeq \frac{3(\text{number of closed triangles})}{\text{number of paths of length 2 edges}}. \tag{2.5}$$

In sociological terms clustering coefficient is an equivalent to a close ring of friends, of who everyone is familiar with each other. As a result, the clustering in social and collaboration network is known to be high [Barabasi et al. 2002]. Quite similar results were obtained in already mentioned analyses Facebook graph, see again [Ugander et al. 2011].

High clustering might seem to affect short path length in graph. It means that lots of edges are 'wasted' in the neighbourhood, which is already connected, 'effectively' reducing the path by just one edge. It is worth noting that both figures are independent, and one can generate graphs with various combinations of these values. For instance, regular triangular lattices tend to have high clustering and large average path length, while balanced binary trees have low clustering and short path lengths. The full or nearly full graphs have large C and very short diameter. More random and less 'degenerate' examples are provided further in this chapter.

In case of adapting the clustering onto the collaboration graph (computing nodes, human collaboration) the high C would suggest an existence of local cooperation or competition. Thus, when designing the network architecture, this figure can on one hand denote a 'global' local communication, i.e. every vertex communicates with large number of neighbours, but on the other suggests a natural aggregation of vertices into a larger unit (for instance scheduling computing tasks onto a single node, in order to reduce the network communication).

2.1.3 Properties of the ER graphs

Below we provide basic facts about Erdős-Rényi graph.

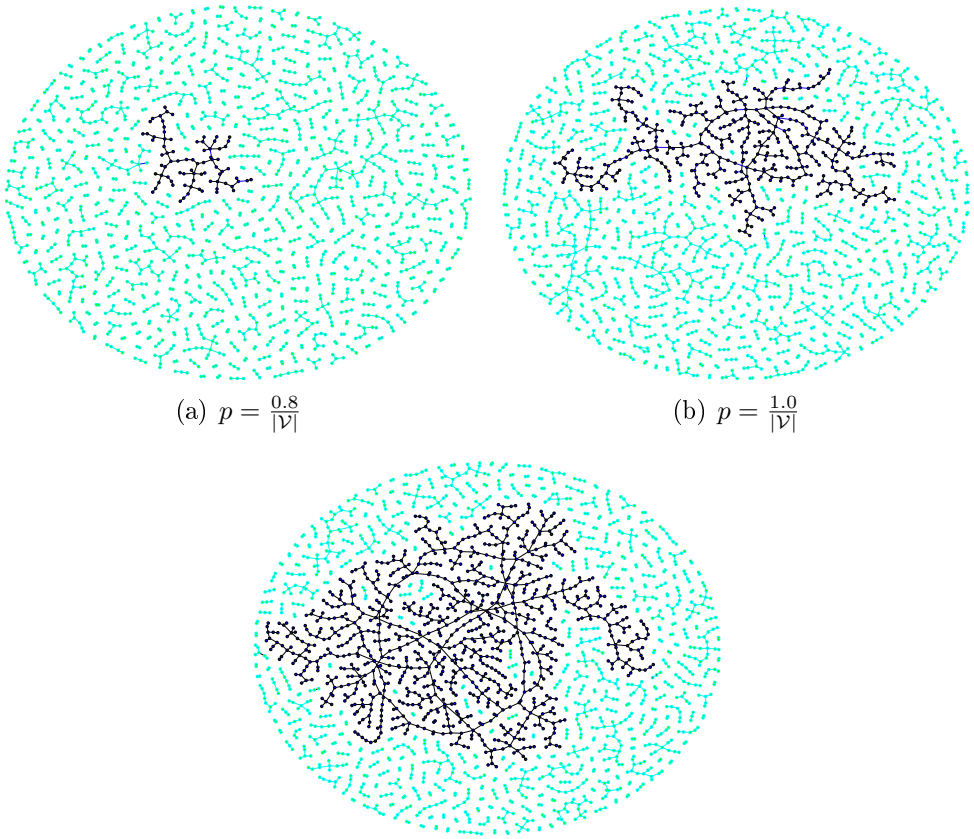
Average connectivity The average connectivity in ER model is $\langle k \rangle = p(n-1)$, where p is connectivity probability and $n = |\mathcal{V}|$ is a number of vertices.

Connectedness For $p < \frac{1}{n}$ the graph is almost surely disconnected, built from small trees and uni-cyclic structures of size $O(\log n)$. For $p > \frac{1}{n}$ the graph almost surely consist of one large connected component of size $O(n)$ and small components of size $O(\log n)$. The large one is unique and referred

to as a *giant component* . Additionally for $p > \frac{c+\log n}{n}$, for $c > 1$ the graph is almost surely connected.

For $p = \frac{1}{n}$ the structure is quite complex. The graph is almost surely disconnected, but as p passes $\frac{1}{n}$ the small components merge into a large one of size $O(n^{2/3})$.

The proof can be found in the work of Erdős and Rényi, see [Erdős & Rényi 1959]. The estimations concerning size of the components, also for wider ranges of p, are available in [Chung & Lu 2006] in Chapter 5.



(a) $p = \frac{0.8}{|\mathcal{V}|}$ (b) $p = \frac{1.0}{|\mathcal{V}|}$

(c) $p = \frac{1.2}{|\mathcal{V}|}$

Figure 2.1: A raise of the large component in Erdős-Rényi random graphs, while the connectivity probability p passes the value $\frac{1}{|\mathcal{V}|}$. All the graphs count approximately 3200 vertices and were remapped onto a circle. For the case of visibility all the separate vertices were removed from the plot, they constitute roughly a fraction of $\frac{1}{e}$ of the whole set of vertices $\mathcal{V}$. The largest component is marked with dark colour.

Degree distribution One can easily calculate the degree distribution in ER model. Since the probability of edge addition is identical and independent, the resulting σ-field for the vertex degree is a product of $n-1$ σ-fields of additions of incident edges. All of these have zero-one distributions with probability p, so their sum has a *binomial distribution* $B(n-1,p)$. Asymptotically, for large n and the probability p, such that $np \to \lambda$, the degree distribution becomes poissonian $\mathcal{P}(np)$.

Average path length Assuming the graph is has a giant component (for $p > \frac{1}{n}$), its characteristic path length grows logarithmically with the number of vertices

$$l \propto \frac{\ln n}{\ln\langle k\rangle} = \frac{\ln n}{\ln np}. \tag{2.6}$$

The proof was provided by Chung and Lu in [Chung & Lu 2001], some more results concerning concentration of the length distribution can be found in their monograph [Chung & Lu 2006].

Clustering coefficient Quite similarly one can estimate the clustering coefficient for ER graph. Due to independence of edge addition, in the neighbourhood of size m there can be at most $\frac{m(m-1)}{2}$ edges, of which an average fraction of p is added to $\mathcal{E}$. Hence, the expected clustering coefficient of any vertex is

$$C(v) = \frac{\frac{m(m-1)}{2}\cdot p}{\frac{m(m-1)}{2}} = p \tag{2.7}$$

and the expected clustering coefficient of the whole graph is still p.

2.2 Scale-free phenomenon in graphs

A power law probability distribution is a distribution for which $\mathbb{P}(X=k) \propto x^\gamma$, $\gamma < -1$, and in terms of continuous measures with its density function $d_X(k) \propto k^\gamma$.

The clear distinction from poissonian (discreet), exponential or gaussian (continuous) distributions is a lack of exponent term in the formula (poissonian distribution formula has a factorial in its denominator). Hence the decay rate of the power law distribution is much slower. As a result the mass of the distribution is skewed towards large values, while in the case of exponential distribution large values are so rare that they are taken as 'outsiders'. Figure 2.2 presents shapes of power-law and exponential distributions, note

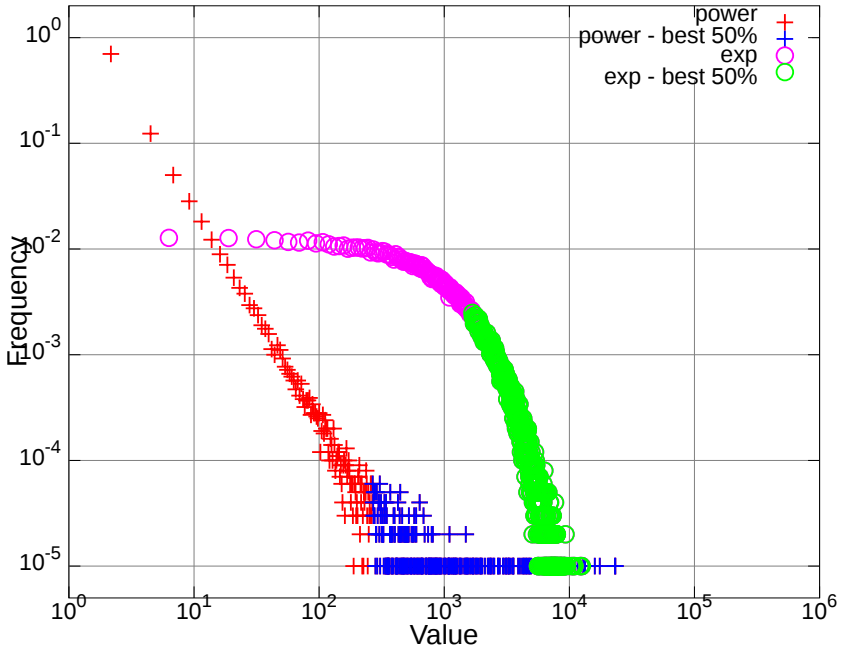


Figure 2.2: A comparison between power law k^{-2} and exponential $c \cdot \exp(-k)$ distributions. Each of highlighted parts contains 50% mass of the sample. For power law these are 380 best values (out of 10^5), for exponential distribution roughly $18.6 \cdot 10^3$ best values (again out of 10^5).

that the plot is renormalised for visibility purposes. It is not hard to see that a $f(x) = \frac{1}{x^k}$ function, when plotted in double logarithmic scale looks like a straight line, this fact is sometimes useful when estimating the distribution. Clearly the power-law distributions may fail to have variance or even the expected value (for $-2 \leq \gamma < -1$).

Upon translation of this considerations onto a graph theory (a *power-law* or *scale-free graphs*) most frequently the (input) degree of the node is taken as the random variable.

Definition A random graph is scale free if its degree distribution follows a power law

$$\mathbb{P}(deg(v) = k) \propto k^{\gamma}, \tag{2.8}$$

where $\gamma < -1$ and v is a randomly selected vertex of the graph.

The notion of scale-freeness in graphs was brought into attention in [Barabasi et al. 1999]. One should keep in mind, that the scale-freeness can be defined as an independence from spatial and / or temporal scale of the graph, i.e. regardless of size, all its sub-graphs should yield the same shape

of the degree distribution. This more generalised definition is provided by Chung and Lu, though it is not so useful to decide whether or not the graph is scale-free, see [Chung & Lu 2006].

A quick example for losing the properties on scale changing, let us consider an random graph G with n vertices and the independent probability of edge inclusion p (see also Section 2.1.3). For $p > \frac{1}{n}$ the graph is almost surely composed from a single giant component and small number of separate vertices [Erdős & Rényi 1959]. Let us consider G_1 — induced subgraph of G with $m \ll n$ vertices. If $\frac{1}{n} < p < \frac{1}{m}$ than G_1 becomes almost surely disconnected and composes from small separate components, see ibidem. Hence, the selection of the scale can dramatically affect the properties of the graph.

2.3 Contemporary random graph models

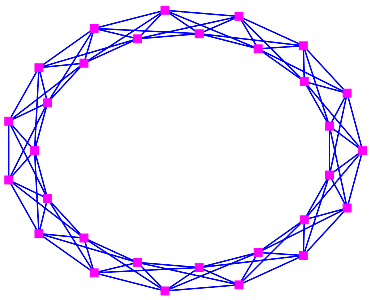
2.3.1 Watts-Strogatz random graph

This model was put forward in 1998 and was accompanied by first formal notions of clustering coefficient and small-world graphs [Watts & Strogatz 1998, Watts 2004]. The main motivation was the inability of ER model yo yield a high clustering without losing its sparsity. To address this problem Watts and Strogatz developed a graph model with fixed number of vertices and edges, but heavy reliant on random rewiring.

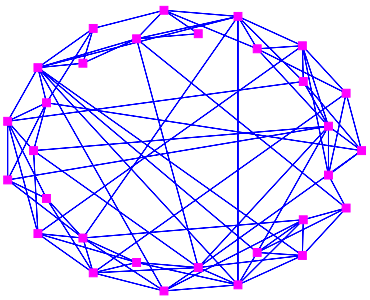
Definition Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a regular graph with n vertices organized into unit ring each connected to k nearest neighbours. A Watts-Strogatz random graph with rewiring probability $p \in [0, 1]$ is obtained from $\mathcal{G}$ by applying a rewiring procedure independently with a probability p to every edge present in $\mathcal{E}$. Rewiring removes the current edge from the graph and includes another one, picked randomly with uniform probability, but excluding multi-connections and self-loops.

Degree distribution If the target vertex of the rewiring is picked uniformly, the degree distribution is approximately binomial, with the variance falling with p. Clearly, for $p = 0$ the distribution degenerates at number of connected nearest neighbours in starting phase $\mathbb{P}(deg(v) = k) = 1$.

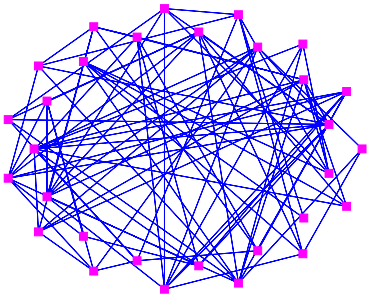
Path length and clustering The starting graph with the topology of periodic lattice has high clustering (for $k > 2$) but its diameter is large (approximately $d \simeq \frac{n}{k}$). The aim of the rewiring part of the construction



(a) $p = 0$, $k = 6$ — regular lattice.



(b) $p = 0.3$, $k = 6$ — small world.



(c) $p = 1$, $k = 6$ — a random Erdős-Rényi graph.

Figure 2.3: A plot of three Watts-Strogatz random graphs with various rewiring probabilities p.

is to create small number of 'short-cuts' in the graph without damaging its local clustering. Clearly for $p = 1$ the model is equivalent to ER model with connection probability equal to average connectivity (except for the fact that the number of edges is exactly equal to the theoretical value) and the clustering structure is destroyed in the process. On the other hand, for $p = 0$ the model leaves predefined regular graph unaltered with large path length. Both situations depend on the rewiring probability and seem to oppose each other, large p increases the chance of the short-cut, while small p preserves local structure. Contrary to these expectations, Watts ans Strogatz found quite a long range of p, for which the average path length l is short and the clustering coefficient C is high.

This sort of regime of efficient global transfer and local connectivity is referred to as *small-world phenomenon*.

2.3.2 Albert-Barabasi construction scheme

In 2002 Albert and Barabasi put forward a graph model which accounts for continuous evolution, clearly inspired social and World Wide Web networks, see [Albert & Barabasi 2002]. Their model assumes two vital features, which are present in real large scale network — a growth and preferential attachment.

Definition Albert-Barabasi random graph generation scheme starts from small regular graph (for instance a single vertex) and iteratively performs two actions according to following mechanisms:

- *growth* adds a single new vertex v to the graph,
- *preferential attachment* randomly wires the newly added node to the graph with the probability of choosing a pairing node equal to

$$\mathbb{P}(\{v,u\} \in E) = \frac{deg(u)}{\sum_{w \in V} deg(w)}. \tag{2.9}$$

An example of graph constructed by Albert-Barabasi scheme is depicted in Figure 2.4. The growth reflects an ubiquitous property of the networks, that they are not rigid structures, and new vertices are born with higher or lower frequency, for instance new Internet sites or young authors of scientific publications. The model gradually gains popularity and is applied for modelling connectivity patterns in researches concerning cellular automata or interaction in game theory, for instance [Matuszak & Miękisz 2011].

The preferential attachment is a step towards winner-take-most postulate. Indeed the existing vertex with high popularity is way more probable to gain new neighbour (depending on the network: citation, hyperlink, associate, etc.) rather than completely unknown one.

The model yields a degree distribution which is a power law with exponent value approximately equal to -3. By modifying the generation scheme one can (asymptotically) obtain any exponent below -2. Some alterations to the growth result in exponentially truncated power law or strictly exponential degree distribution. Below we provide a sketch calculation of the obtained probability distribution after [Albert & Barabasi 2002].

Denote time as t, time of addition of node v_i as t_i, a degree of vertex v_i at time t as $k_i(t)$, number of new connections introduced along with the new vertex as m, initial number of vertices as m_0. Due to the preferential attachment, by considering the time-scale to be continuous we can estimate

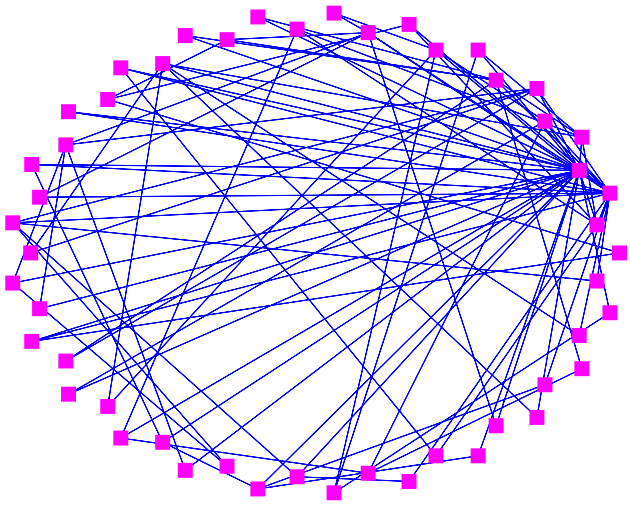


Figure 2.4: A plot of Albert-Barabasi random graphs with two edges added per vertex in preferential attachment. The vertices were added counter-clockwisely starting from right.

the rate of obtaining a new edges as

$$\begin{cases} \frac{dk_i}{dt} & = \frac{k_i}{2t} \\ k_i(t_i) & = m \end{cases}. \tag{2.10}$$

The scalar 2 in denominator roots from two 'degrees' being added per edge. The set of equations gives $k_i(t) = m\sqrt{t/t_i}$. Now we calculate the probability, that the degree of v_i is less than k

$$\mathbb{P}(k_i < k) = \mathbb{P}(t_i > m^2 \frac{t}{k^2}). \tag{2.11}$$

Here we assume that the vertices are added towards the system with uniform distribution. Now we obtain

$$\mathbb{P}(t_i > m^2 \frac{t}{k^2}) = 1 - \mathbb{P}(t_i \leq m^2 \frac{t}{k^2}) = 1 - \frac{tm^2}{k^2(t+m_0)}. \tag{2.12}$$

For large graphs $t \to \infty$ and $\frac{t}{t+m_0} \to 1$ and we obtain

$$\mathbb{P}(t_i > m^2 \frac{t}{k^2}) = 1 - \frac{m^2}{k^2}. \tag{2.13}$$

Finally, we differentiate over k ($\mathbb{P}(k) := d\mathbb{P}(t_i \geq k)/dk$) in order to obtain the distribution (instead of its cumulative form):

$$\mathbb{P}(k) \simeq \frac{-(-2)m^2}{k^3}. \tag{2.14}$$

Which results in power law degree distribution with the exponent value $\gamma = -3$.

□

The number of edges grows linearly, while the preferential attachment leads to appearance of *hubs* — vertices with large number of connections which greatly reduce the number of hops necessary to reach other vertices. Thus, pretty much like Watts-Strogatz, the Albert-Barabasi graph model can generate sparse graph with short-diameter. More formally, the characteristic path lengths of the AB model is proportional to logarithm of the number of vertices n:

$$l \propto \frac{\ln n}{\ln \ln n}. \tag{2.15}$$

Additionally, the obtained graphs tend to have relatively high clustering. To our knowledge, this value has not been calculated analytically (see again [Albert & Barabasi 2002]), however the numerical estimation suggests that the clustering decays as

$$C \sim n^{-\frac{3}{4}}. \tag{2.16}$$

2.4 Critical regime and power laws

Complex systems and large networks frequently undergo their unsupervised dynamics and reorganisation process. Such self-organisation requires an abundance of degrees of freedom available in the system [Chialvo 2004]. Their evolution dramatically differ from carefully designed devices, computers for instance, where there is almost no freedom left in command execution protocol. On the other hand, complex systems as landslides [Bak et al. 1987] are energy-driven and tend relieve the force in order to reach local energy minimum.

Even more striking, the newly reached minimum is barely stable. It is like staying in the local minimum at the border of two regimes. Hence, the complex system can operate in such *critical point* for quite long time.

For instance, human brain has a constant moderate activity even during its 'idle' phases. The activity may raise for an instant as a result of input stimulus, but it will not explode and neither vanish. Any of these cases would lead to a disaster. Vanishing of all the activity would lead to ceasing of cognitive abilities, as the brain would not be able to reignite the information flow. On the other hand, in extreme situation an over-excitation would cause fast chaos propagation and as a result malfunctioning of the whole brain and epilepsy.

As a counter-example, a computer in its 'critical state' (i.e. high load of tasks), either becomes blocked by vast amount of long computations or completes them (or aborts in order to save system stability) quickly to become idle again. Neither of these cases is critical in the sense discussed above.

Except for the different scales, this seems to be held in mathematical models of spiking neurons (see Section 3.2). The sophisticated set of equations, describing the dynamics admit the coexistence of two or more different states of the neuron: resting and spiking. The resting state is (barely) stable, so in the absence of external stimuli the neuron tends stabilise in it. However, after an application of relatively low input current the dynamics leaves the resting state and performs a cycle of rapid increase and drop of voltage on the membrane of the neurons (which corresponds to spike generation) and again returns to the resting state, for detail see [Izhikevich 2007]. In this case a single neural cell operates in critical or nearly critical dynamics (close to a bifurcation of the dynamics).

Another example is stochastic Boltzmann machine dynamics, see [Ackley et al. 1985] and Section 3.3.3 of this dissertation. In a fixed high temperature regime the network acts completely at random, in a low temperature the dynamics freezes in a local or (hopefully) global minimum. Although, there exists a small range of temperature in which the dynamics becomes complicated and the obtained temporal correlation activity graph becomes asymptotically scale free [Fraiman et al. 2009]. This, however, occurs for rather carefully tuned temperature variable, while in most of complex systems the critical regime is approached spontaneously, see again [Bak et al. 1987].

Quite frequently, the systems in their critical states are reported to exhibit in some of their activity statistics a lack of characteristic scale and / or a power low distribution in their activity. Recall, that a *scale-free* or *scale-invariant graph* is a graph, whose degree distribution is the same as degree distribution of some of its random subgraphs for various scales of its size [Chung & Lu 2006]. The best known examples are graphs obeying a power low degree distribution. Note, that one can also consider graph invariants on spatial or temporal scales, see ibidem.

Eguiluz findings about scale-free networks in fMRI data may relate to a criticality in human brain networks, see [Eguiluz et al. 2005]. The work of Chialvo [Chialvo 2004] also account for this hypothesis. This behaviour has not been fully understood yet, but it is believed that this phenomenon can be a key to the understanding how the brain works.

Chapter 3

Mathematical models of neural networks

In this section we briefly sketch the state of the arts in the field of contemporary artificial neural networks, which are the very foundation of this thesis. We reiterate various neural models heading for the simplified and mathematically tractable ones. Finally we recall the results of Piękniewski and Schreiber, which are the direct inspiration of our research aims.

Artificial neural networks are biologically-inspired mathematical models for organic brain cells and nerve systems. Depending on desired accuracy and available computing power they can vary in means of description from simple weighted averages to dynamical systems of differential equations. Though the expressive power of the single neuron is quite limited, the real capabilities are at their peak when the neurons are connected into a complex network. The interaction dynamics can again be expressed in terms ranging from vector and matrices algebra to differential diffusion equations.

The aim of modelling of a neural cell and a whole brain is two-fold. First one is development of said biologically inspired algorithms. Second is to shed some light on the principles of the work of human nerve system, perception and higher cognitive abilities.

3.1 Biological neural cell

Biological brain is the centre of the nerve system in animals and humans. It consists of huge number of *neural cells* or simply *neurons* connected with *synapses*. The number of neurons in human brain is roughly put at $10^{10}-10^{11}$, while the number of synapses 10^{14}, which yields an average connectivity of $10^3 - 10^4$ synapses per neuron, see [Russel & Norvig 1995]. This means, the

the brain is a very sparse network.

If we were to compare this vast power to the Internet (a network of computing units), we would conclude, that the whole WWW network, with its 40 billion of webpages, slowly approaches a size of single human brain[1]. Crude approximations of brain computing capacity vary between 10^{15} and 10^{18} FLOPS (floating-point operations per second), that is 1 petaFLOPS to 1 exaFLOPS, see [Hilbert & Lopez 2011, Russel & Norvig 1995]. These numbers might be even more affected depending on the accuracy of the dynamics and translating (or not) continuous neural activity into set of floating-point operations. As of the date of writing the top supercomputer achieved as much as 10.5 petaFLOPS[2]. On the other hand the human brain operates at the power far below 100W, while the every contemporary computing centre requires a small power plant for calculation and cooling.

With some simplifications, each cell can be viewed as a small electric circuit operating on ionic, rather than electric, current. The diffusion and electrostatic potentials cause the current (mainly sodium, potassium, calcium and chloride) to flow through the cellular membrane, propagate along the synapses and excite or inhibit neighbouring cells. These impulses or *spikes*, or at least their firing frequency, are believed to be the main mechanism of information passing in the nerve system. They are generated by perception cells as a result of input (for instance visual) stimuli, transmitted and 'processed' through the nerve system, finally they reach effectors tissues (for instance muscles) as a reaction to the stimuli.

An interesting and essential feature of the brain is its *plasticity*. It is capable of altering its properties in order to adjust to the environmental input. An explanation of learning in terms of altering the parameters of neurons and their interconnections was postulated by Donald Hebb in half of the twentieth century [Hebb 1949]. There is a common agreement, that the learning process modifies the strength of the synapses depending on the activation of both pre-synaptic and post-synaptic neurons.

The understanding of the processes of thinking and other higher cognitive processes is a long-time program, that may take decades, although some modest achieves have been accomplished. For instance, recordings of auditory cortex activity was applied to speech reconstruction by Pasley et al., see [Pasley et al. 2012]. Nishimoto et al. used fMRI recordings of visual areas activity along with bayesian inferring for an image reconstruction, see [Nishimoto et al. 2011]. Vast number of contemporary research in the field

[1] The data after http://www.worldwidewebsize.com. The methodology paper by Maurice de Kunder *Geschatte grootte van het geindexeerde World Wide Web* is available in Dutch at http://www.dekunder.nl.

[2] Information after http://top500.org/, valid as of date 2012.03.06.

of neuroscience sets out to shed some light at small pieces of the complex dynamics and step-by-step add some scarce information to the knowledge.

3.2 Hodgkin-Huxley model

Historically the earliest neural cell model was named after Hodgkin and Huxley. In the middle of the twentieth century they analysed the axon cell of the giant squid Loligo. As a result of a several publications in *Journal of Physiology* [Hodgkin & Huxley, 1952a, Hodgkin & Huxley, 1952b], the excitability of the neural membrane was described in terms of dynamical system consisting of a set of differential equations. The formulae of the functions and values of the parameters were adjusted to make the system approximately close to the biological data. Along with development of the neuroscience, the system was put into use in neural simulations at voltage level. Here we present the contemporary version of this system with adjustments suggested by Izhikevich in his monograph [Izhikevich 2007].

The dynamics of the neural cell is described as a set of differential equations evolving with time variable t:

$$\begin{cases} \frac{d}{dt}V(t) &= \frac{1}{C}\left(I - \bar{g}_{\mathrm{l}}(V - E_{\mathrm{l}}) - \bar{g}_{\mathrm{Na}}n^4(V - E_{\mathrm{Na}})\right. \\ & \quad \left. -\bar{g}_{\mathrm{K}}m^3h(V - E_{\mathrm{K}})\right) \\ \frac{d}{dt}n(t) &= \frac{n_\infty(V) - n(V)}{\tau_n(V)} \\ \frac{d}{dt}m(t) &= \frac{m_\infty(V) - m(V)}{\tau_m(V)} \\ \frac{d}{dt}h(t) &= \frac{h_\infty(V) - h(V)}{\tau_h(V)} \end{cases} \tag{3.1}$$

where $V(t)$ stands for *membrane voltage*, $n(t)$, $m(t)$ and $h(t)$ are *gating variables*. E_{Na}, E_{K} and E_{l} are rest equilibrium potentials for sodium, potassium and leak currents respectively. I is an input current and C is a membrane capacitance. $\bar{g}_{\mathrm{Na}}$, $\bar{g}_{\mathrm{K}}$ and $\bar{g}_{\mathrm{l}}$ are maximum conductivity of sodium, potassium and leak ionic gates.

The $n_\infty(V), m_\infty(V)$ and $h_\infty(V)$ are stationary gating functions of sigmoid shape, see Formula (3.8). The $\tau_n(V), \tau_m(V)$ and $\tau_h(V)$ are unimodal speed functions, which determine the convergence rates of the gating variables to their stationary values.

The values of parameters, adjusted to fit the model to the data, are equal

to (again after [Izhikevich 2007]):

$$\begin{aligned} E_{\mathrm{Na}} &= 55mV & \bar{g}_{\mathrm{Na}} &= 120mS/cm^2 \\ E_{\mathrm{K}} &= -77mV & \bar{g}_{\mathrm{K}} &= 36mS/cm^2 \\ E_{\mathrm{Leak}} &= -55.4mV & \bar{g}_{\mathrm{Leak}} &= 0.3mS/cm^2 \\ C &= 1\mu F/cm^2 \end{aligned} \tag{3.2}$$

The formulae of the rate-of-convergence functions:

$$\begin{aligned} \tau_n &= \frac{1}{\alpha_n + \beta_n} & n_\infty &= \frac{\alpha_n}{\alpha_n + \beta_n} \\ \tau_m &= \frac{1}{\alpha_m + \beta_m} & m_\infty &= \frac{\alpha_m}{\alpha_m + \beta_m} \\ \tau_h &= \frac{1}{\alpha_h + \beta_h} & h_\infty &= \frac{\alpha_h}{\alpha_h + \beta_h} \end{aligned} \tag{3.3}$$

The formulae of the steady-state value functions:

$$\begin{aligned} \alpha_n(V) &= 0.01\frac{10 - V}{\exp(\frac{10-V}{10}) - 1} & \beta_n &= 0.125\exp(\frac{-V}{80}) \\ \alpha_m(V) &= 0.1\frac{25 - V}{\exp(\frac{25-V}{10}) - 1} & \beta_m &= 4\exp(\frac{-V}{18}) \\ \alpha_h(V) &= 0.07\exp(\frac{-V}{20}) & \beta_m &= \frac{30 - V}{\exp(\frac{30-V}{10}) + 1} \end{aligned} \tag{3.4}$$

The main variable responsible for spiking is a *membrane potential* V, see Figure 3.1. Gating variables play a secondary, but still important part of auto-adjusting the in-flowing and out-flowing ionic currents.

Accurate though the system might be, a set of four differential equations turned out to be computationally demanding to simulate, in particular the exponent functions present in steady-state and convergence functions, which must be calculated every step. Therefore a number of simpler models were developed. Among the best known simplifications one can note a persistent sodium and potassium (Na-K) model, quadratic integrate-and-fire (QIF) model, FitzHugh-Nagumo, Izhikevich simple model etc. Some of them will be discussed in Section 3.4. In the second half of the twentieth century, in a pursuit for speed optimization the differential equations-driven dynamic was abandoned in favour of easily implemented weighted averages such as the perceptron, see Section 3.3.1. With the development of numerical methods and parallel computing, the accurate spiking models made their comeback into the neuroscience few decades later.

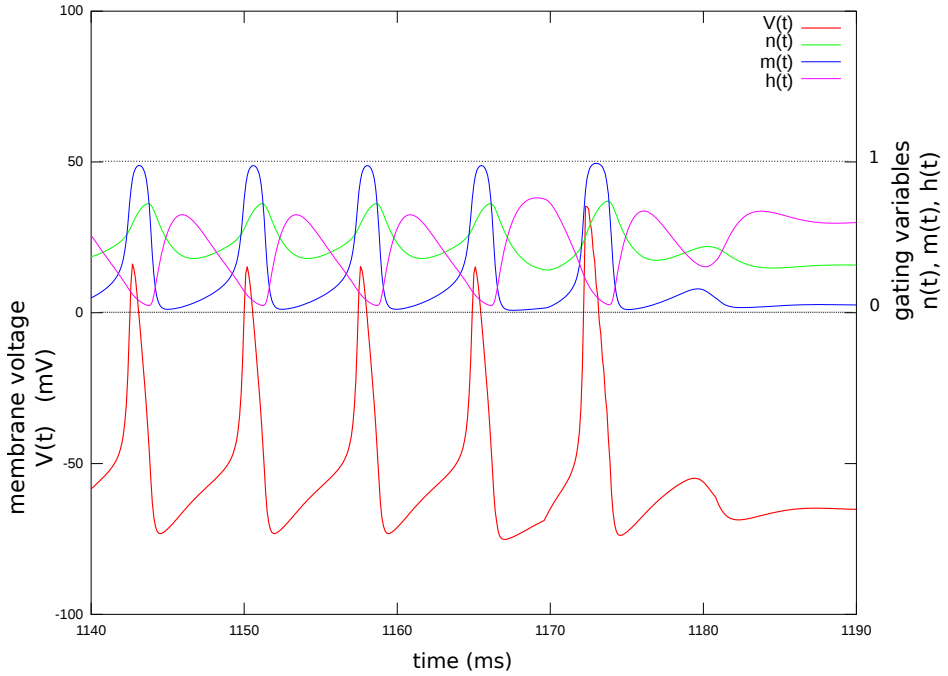


Figure 3.1: A plot of Hodgkin-Huxley neuron model in periodic spiking regime. Membrane voltage V has its scale on the left, while gating values are between 0 and 1 (right scale was renormalised for better visibility).

3.3 Models of neural firing rates

The spike sequences are sometimes a bit chaotic and hard to replicate experimentally. Hence instead of single spikes, a *spiking frequency* is more frequently considered as a main information transfer mechanism. This is a level of abstraction higher above the dynamics of a single synaptic impulse.

The firing rate models describe the average activity of the neuron or even the collection of neurons in response to external stimuli. They frequently trade the accuracy of the spike generation for the computational efficiency. The result can be interpreted as a binary, discreet or continuous response to the input signal. These models found their applications in various problems of computational intelligence, for instance classification and image recognition, see Chapters 3–5 in [Bishop 1995].

3.3.1 Simple perceptron model

One of the most basic models of a firing rate single neural cell is *simple perceptron* developed by Frank Rosenblatt in 1958, defined as follows

[Rosenblatt 1958]:

Definition A simple perceptron is a system consisting of n inputs $x_1, .., x_n$, $x_i \in \mathbb{R}$, $n+1$ weights $w_0, w_1, ..., w_n$, $w_i \in \mathbb{R}$ associated with the inputs $x_0 = +1, x_1, ..., x_n$ and an activation or response function $f : \mathbb{R} \to \mathbb{R}$. Given the input vector $\bar{x} = [x_1, .., x_n]$, the perceptron returns a response equal to

$$Out(\bar{x}) = f(\sum_{i=0}^{n} x_i w_i). \tag{3.5}$$

The complexity of the model was dramatically reduced to a weighted average over the input signal and computation of one response function f, which frequently is also not particularly demanding to compute. As a result, the dynamics can be very quickly calculated even for floating point arguments. It is worth emphasising, that perceptron works synchronously i.e. it returns its value immediately after all the input signals are available.

Assuming the step formula of the response function

$$f(s) = \begin{cases} +1 & s \geq 0 \\ 0 & s < 0 \end{cases}, \tag{3.6}$$

the perceptron acts an an integrator, namely it accumulates incoming impulses and produces zero-or-one spike response. From the view of data mining, the perceptron can be considered as a binary classifier capable of assigning the input data $\bar{x}$ one of two complementary classes — class A, when the input stimuli evoked a spike and class B otherwise.

Before the perceptron can be applied in computational intelligence problem, it must undergo a *learning process*. First class of learning algorithms form the *supervised* or *error correction*. Their biological motivation might be argued, as they require an external mechanism, which 'knows' the correct response and can compare it to the obtained result. The examples of supervised learning algorithms are pocket algorithm ([Peretto 1994, Rojas 1996]), back-error propagation [Werbos 1974] with their modifications.

The second class are unsupervised learning algorithms. Due to lack of external control mechanism, the perceptron is taught by competition between its weights. Hebb postulated that interactions between input signal and the response lead to strengthening or weakening the effect of the connecting synapse. Like a self-perpetuating loop, spikes along already strong synapse cause their effect to be even stronger. The idea is referred as a *winner-take-all* or *competitive learning* and widely used in learning algorithms as Hebbian rule

$$w_i := w_i + \eta \cdot x_i Out(\bar{x}), \tag{3.7}$$

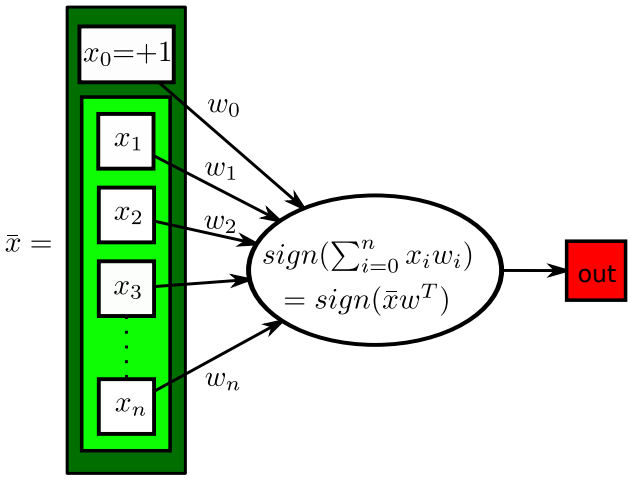


Figure 3.2: A schematic figure of the simple perceptron with a *sign* response function.

where x_i is the i-th input signal, w_i is a weight of the synapse between the cell and i-th input, $Out()$ is a post-synaptic response and $\eta > 0$ is a learning rate, [Hebb 1949]. Due to the numerical instability, the Hebbian rule was later improved by Oja [Oja 1982].

3.3.2 Feed forward perceptron network

A directed, acyclic and topology sorted graph of perceptrons is referred as a *feed forward perceptron network*. If two perceptrons v and u are connected with a directed edge $v \rightarrow u$, then the output value returned by v is a part of the input signal for u with its associated weight ascribed to the edge.

In such simple case passing the information between neurons is reduced to adding the output value of pre-synaptic unit multiplied by its weight to the activation signal of the post-synaptic neuron. Assuming the vector form of input $\bar{x}$ and weights $\bar{w}$, the response of the neuron is expressed as $f(\bar{x}^{tr}\bar{w})$, where f is the response function.

The linear algebra is exploited even deeper in multi-layer perceptron (MLP) networks. In this case the neurons are organised into pairwise disjoint *layers*. The first layer is referred as *input layer* and contains only the input data. The last one is an *output layer* and the remaining are said to be *hidden*.

The network connections are allowed only between neighbouring layers, that is form $l-th$ to $l+1$-th and are considered to constitute a full bipartite graph. Let us denote $X = [x_1, .., x_n] \in \mathbb{R}^n$ as the activations of neurons in

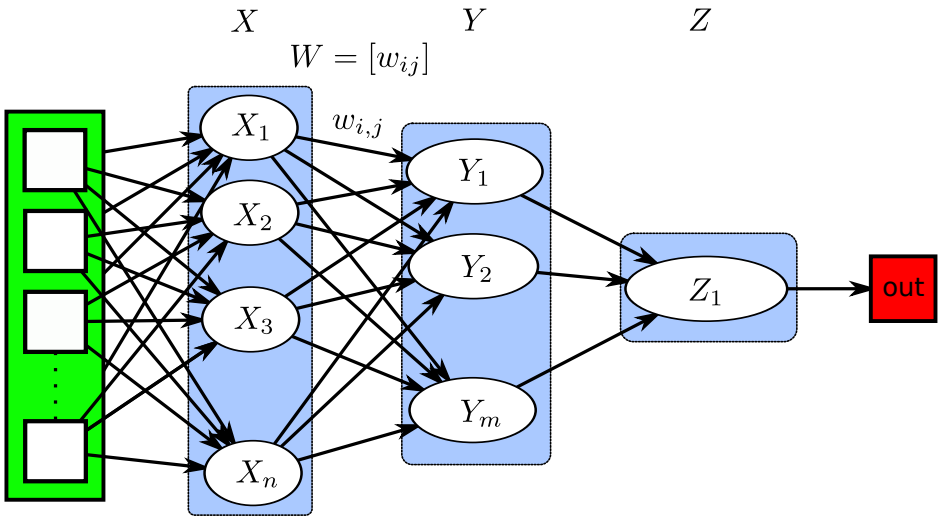


Figure 3.3: A schematic of multi-layer network with two hidden layers.

l-th layer, $Y = [y_1, .., y_m] \in \mathbb{R}^m$ — activations of neurons in $l+1$-st layer and $W = [w_{i,j}]_{i=1..n, j=1..m} \in R^{n \times m}$ — a weight matrix. The propagation of data in the MLP network can be expressed as vector-matrix multiplication $Y = \bar{f}(W^{tr}X)$, where $\bar{f}$ stands for coordinate-wise application of the response function.

By assigning the sigmoid formula of f

$$f(s) = \frac{1}{1 + exp(-\beta s)}, \tag{3.8}$$

the response can be adjusted to be continuous function of the stimuli $\bar{x}$. This should not be uniquely interpreted as the spike or its lack, but rather an average intensity of impulsing (hence the name — *firing rate*). On the other hand this sort of response function allows an application of gradient descent algorithm to error-correction driven learning. The best known learning algorithm for feed-forward networks is *back error propagation* developed in PhD dissertation of Paul Werbos, [Werbos 1974]. However, it was not until 1986 and the work of [Rumelhart et al. 1986], when this algorithm became popular.

The alternative paradigm is to build the network from scratch by subsequent perceptron addition. Each addition is usually supposed to increase the expressing power of the network up to the point, where all the training examples are properly categorized. The literature [Peretto 1994] list pyramid, tiling, up-start algorithms etc. They usually yield a more sophisticated graph, but with still regular structure.

3.3.3 Hopfield network and Boltzmann machine

The Hopfield network is a simple example of recurrent neural network for firing rate. Unlike feed-forward, recurrent networks admit cycles and self-loops. This makes them more feasible to model biological systems. As a matter of fact, restriction of acyclicness was needed to ease mathematical tractability of the inferring and learning algorithms. For instance the back-error propagation algorithm for recurrent network requires a few tricks like 'unfolding' it into few copies and proper selection of number of time steps [Werbos 1990]. In the end, the recurrent network loses its uniformity, i.e. the functionalities of some parts may be predefined during the network design rather than learning process. Recurrent networks by their very definition provide a support for a possible cyclic information propagation. Their main issue is the stability of learning and reasoning process.

As described in [Hopfield 1982], the Hopfield network found applications in solving constraint based problems. The network consists of n neurons, each with binary *spin* $\sigma_i \in \{-1, +1\}$, for $i = 1..n$. In addition each neuron has assigned *bias* $h_i \in \mathbb{R}$, which can be interpreted as innate preference of the neuron towards a positive or a negative spin value. The neurons are connected with all-to-all symmetric *synapses*. Each synapse has its *weight* $w_{ij} \in \mathbb{R}$, which describes its excitatory (when positive) or inhibitory nature. In case of not-fully connected neurons, the 'weights' of not included synapses are assumed to be zero. Though the self loops are not permitted, the symmetric weights and clear lack of order of neurons make Hopfield network recurrent one, i.e. the information sent by a neuron can after some time return to it (possibly modified by a number of weights and thresholds).

For a given configuration of the spins $\bar{\sigma} = (\sigma_1, ..., \sigma_n)$ we define an *energy* (or Hammilonian) of the network as follows:

$$E(\bar{\sigma}) = -\frac{1}{2}\sum_{i \neq j} w_{i,j}\sigma_i\sigma_j + \sum_{i=1}^{n} h_i\sigma_i. \tag{3.9}$$

The energy function has an interpretation of total disagreement in the network. In this case, a pair of neurons σ_i, σ_j, which are connected with excitatory synapse $w_{ij} > 0$, 'prefer' to be in the same state ($\sigma_i = \sigma_j$), if the weight is negative, they tend to be in opposite spins and the 'tendency' is proportional to the absolute value of the weight. Additionally the bias can be interpreted, as an intrinsic attitude of the unit towards positive (for $h_i > 0$) or negative spin.

The network dynamics iteratively swaps one or all, depending on the synchronisation policy, of the spins in $\bar{\sigma}$ obtaining new configuration $\bar{\sigma}'$ leading

towards lower energy states. In synchronous case this can be written as

$$\bar{\sigma'} := \text{sign}(W \cdot \bar{\sigma} + H \cdot_{coord} \bar{\sigma}), \tag{3.10}$$

where $W = [w_{ij}]_{i,j=1}^{n}$ is a square weight matrix, $H = [h_1, ..., h_n]^{tr}$ is a bias vector, sign() is a coordinate-wise sign function and $\cdot_{coord}$ is a coordinate-wise multiplication of vectors.

In practical application, the spin configuration $\bar{\sigma}$ can be an encoded solution of given discrete optimization problem. In such case the energy is expressed as a sum of *quality component* and *penalty component*. The quality part is the value to be optimized, for instance the total length of the Hammiltonian cycle, see Figure 3.4. The penalty component can be interpreted as an additional energy 'punishment' for violating the formal requirements, given to the solution. In the case of Hammiltonian cycle this may happen when the 'solution' encoded by the set of spins omits a town on its route, see again Figure 3.4.

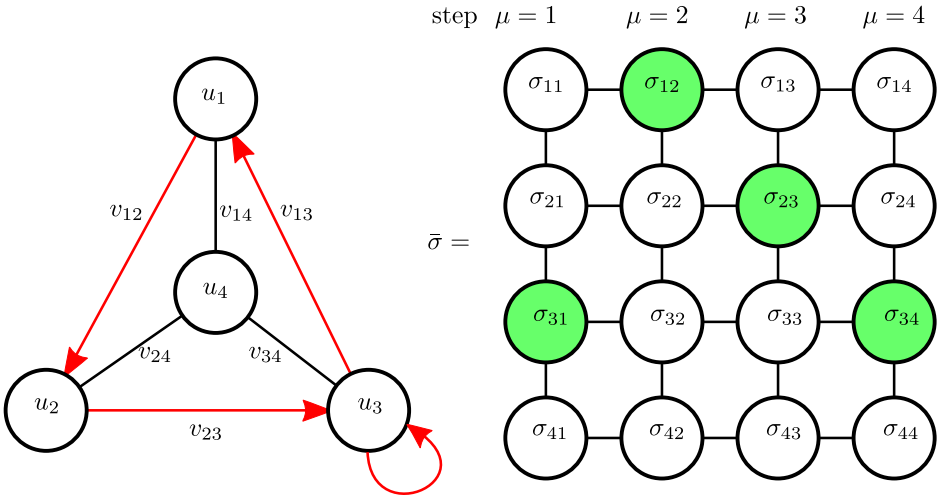


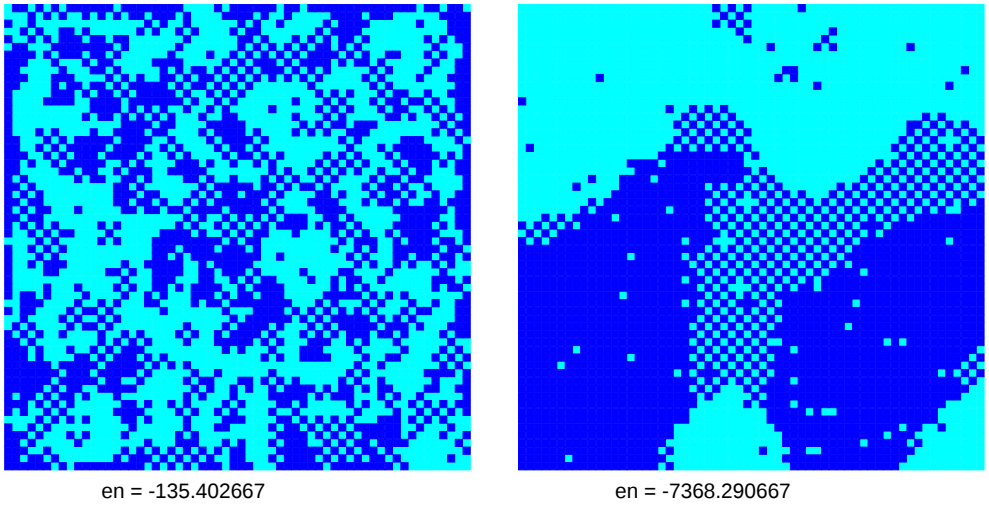
Figure 3.4: An application of Hopfield network to Hammiltonian cycle problem. In this case there are four cities $u_1, ..., u_4$ to be visited. The configuration $\bar{\sigma}$ has sixteen double-indexed spins. The spin $\sigma_{u,\mu} = +1$ denotes visiting the city u in step μ. Note, that the current solution $\bar{\sigma}$ is not a strict Hammiltonian cycle, as it omits the city u_4, while u_3 is visited twice.

Hinton, Ackley and Sejnowski developed a stochastic version of Hopfield network, see [Ackley et al. 1985]. It was called a *Boltzmann machine* after Boltzmann distribution, which is a vital part of its theoretical background. Despite similar structure of the network and, to lesser extent, its dynamics, the Boltzmann machine theory is based on Markov Chain [Markov 1971] and is inspired by thermodynamics [Newell & Montroll 1953].

While the dynamics of Hopfield networks is vulnerable to local energy minima, stochastic evolution of Boltzmann machines allows to jump to higher energy configuration and, by extension, to leave a suboptimal solution.

The dynamics accounts for a stochastic nature of the particles. The new configuration $\bar{\sigma}'$ is obtained from previous $\bar{\sigma}$ by swapping a random spin σ_i. Then $\bar{\sigma}'$ is accepted as new state of the Boltzmann Machine if it causes the energy to drop or with the probability $\mathbb{P} = \exp(-\beta(E(\bar{\sigma}') - E(\bar{\sigma})))$ otherwise (when $E(\bar{\sigma}') - E(\bar{\sigma}) \geq 0$) and is rejected with complementary probability. The parameter $\beta > 0$ stands for an *inverse temperature*, $\beta = \frac{c}{T}$, where $T \geq 0$ is a *temperature*.

For the zero temperature $T = 0$ (infinite β) the dynamics Boltzmann machine is the same as in asynchronous Hopfield network. On the other hand, for T asymptotically approaching infinity (β approaching 0) the dynamics becomes chaotic — namely every single configuration $\bar{\sigma}'$ is accepted and the system becomes a white noise. Clearly both these extreme cases do not occur in thermodynamics.



(a) High temperature. (b) Low temperature.

Figure 3.5: Boltzmann machines in low and high temperature regimes.

In addition, the Boltzmann machine can be adjusted to account for multi-valued spins, as presented by Lin and Lee [Lin & Lee 1995]. Another modification can be found in Chapter 4 in the definition of the activation-flow model.

3.4 Spiking models

In this section we briefly list models which generate spikes, rather than measure average spiking activity. They offer a bit more precise description of the cell, but frequently at the expense of the time cost. Some drawback of spiking models is lack of understanding and shortage of theoretical backgrounds for learning algorithms, which would put such systems into a use in computational intelligence.

3.4.1 Integrate-and-fire models

A *leaky integrate-and fire* (LIF) model is one of the simplest model of neural activity in terms of differential equation. It describes a dynamics of membrane voltage in terms of differential equation:

$$\begin{cases} \dfrac{d}{dt}V = I - g_{\mathrm{L}}(V - E_{\mathrm{L}}) \\ \text{if } \ V \geq v_{\mathrm{thresh}} \ \text{, then produce a spike; } V := v_{\mathrm{reset}} \end{cases} \tag{3.11}$$

A word of explanation is needed here, as LIF is not actually a spiking model. To be more precise — it can only model the voltage at sub-threshold regime. Upon hitting its threshold value v_{thres} it is said to 'produce a spike' and than simply resets the voltage V. The dynamics does not replicate neither a rapid raise (depolarisation) nor the downfall (repolarisation) of the voltage.

Its simple extension is *quadratic integrate-and-fire* (QIF) neuron model [Hansel & Mato 2001]. Similarly, the only equation describes membrane voltage, but a quadratic term does allow the system to produce a fast upstroke. However it is still unable to produce a re-polarisation (bringing the voltage down after the rise) of the membrane, hence the presence of conditional resetting.

In LIF hitting the threshold value indicates a production of the whole spike, while in QIF it is substitution only of the downstroke mechanism.

$$\begin{cases} \dfrac{d}{dt}V = V^2 + I \\ \text{if } \ V \geq v_{\mathrm{thresh}}, \text{ then } \quad V := v_{\mathrm{reset}} \end{cases} \tag{3.12}$$

The voltage V dynamics during the simulation of QIF model is presented on Figure 3.4.1. Note, that the instantaneous drops of the membrane potential are an effect of 'manual' resets. Otherwise the voltage would grow to infinity in finite time.

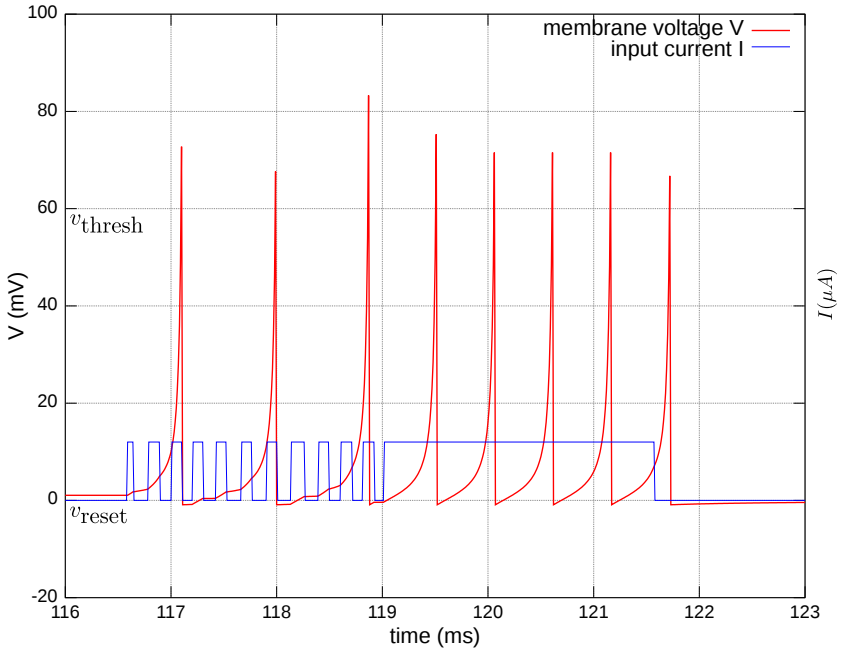


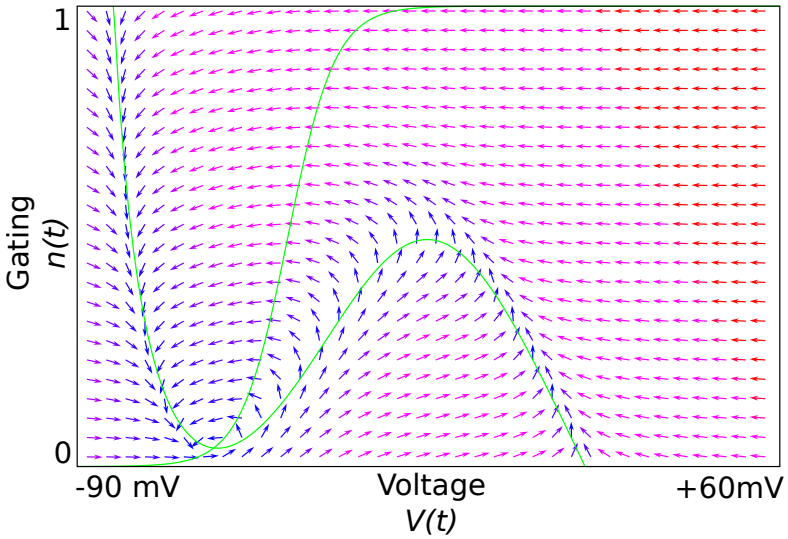
Figure 3.6: A plot of voltage variable in response to input stimuli in quadratic integrate-and-fire neuron model. Membrane voltage V, v_{thresh} and v_{reset} are marked on the left scale. Input current on the right scale.

The major flaw of both models is their inability to mimic any sort of spiking pattern, other than integrator and cyclic spiking, see discussion in [Izhikevich 2007].

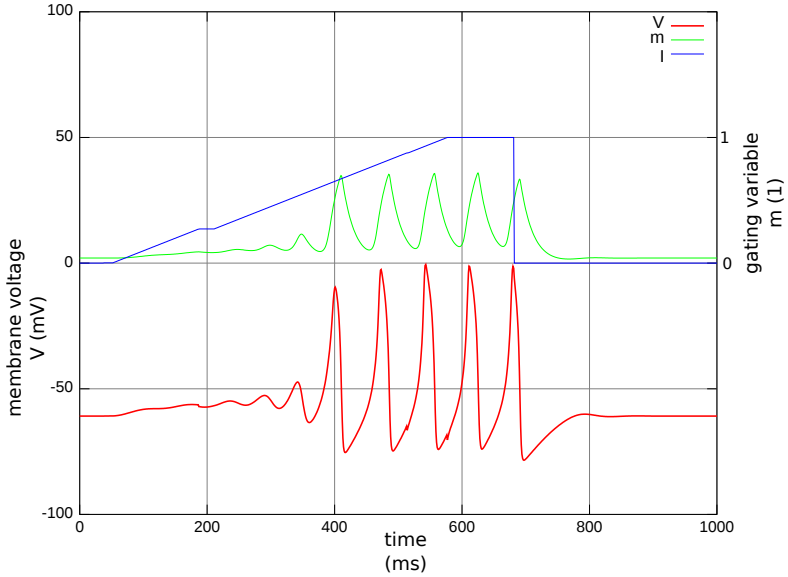
3.4.2 Two-dimensional models

By adding a second equation to the dynamical one can obtain a cyclic behaviour patterns, which can model both voltage upstroke and downfall without a need of manual voltage adjustments. Here we provide two brief examples of such models.

An *instantaneous sodium plus potassium neuron model* accounts for two simplifications from Hodgkin and Huxley set of rules, see Section 3.2. First one is making sodium current instantaneous, which removes the necessity of gating variable m. Additionally the inactivation potassium gate h is also removed.



(a) A plot of phase portrait of instantaneous sodium and potassium neuron model in rest stare (no input current applied). Nullclines are marked with green solid lines. The unique stable equilibrium is at the intersections of the nullclines.



(b) A plot of instantaneous sodium and potassium neuron model on gradual increase of input current I. Membrane voltage V has its scale on the left, while gating value n is between 0 and 1 on right scale

Figure 3.7: The dynamics of instantaneous sodium and potassium neuron model: (a) phase portrait, (b) spike origination.

The model is described by a following set of differential equations:

$$\begin{cases} \frac{d}{dt}V(t) &= \frac{1}{C}\left(I - \bar{g}_{\mathrm{l}}(V - E_{\mathrm{l}}) - \bar{g}_{\mathrm{Na}} m_{\infty}(V)(V - E_{\mathrm{Na}})\right. \\ & \quad \left. -\bar{g}_{\mathrm{K}} n(V)(V - E_{\mathrm{K}})\right) \\ \frac{d}{dt}n(t) &= \frac{n_{\infty}(V) - n(V)}{\tau_n(V)} \end{cases} \tag{3.13}$$

The functions $n_{\infty}(), m_{\infty}(), \tau_n()$ have the same meaning and formulae as those in Hodgkin-Huxley model, see Equations 3.1.

Compared to QIF model the Na-K is a *two-dimensional* (it is described by two variables: V and n). As a result it can produce a spike as continuous response of the system to external stimulation, without 'hand-made' resets.

By adjusting its parameters it admits an existence of dramatic qualitative phase space shifts, also referred as *bifurcations*. This allow to change the spiking pattern by gradually changing one or more of the parameter values. In addition the number of replicable spiking regimes is larger, compared to QIF.

The major drawback of this model is still its computational complexity. It requires a computation of two exponent functions at each step of numerical Euler simulation. This makes it hard to use in real-time neuro-dynamic models.

A *FitzHugh-Nagumo model* was developed in 1961 [FitzHugh 1961] as a simplified version of a neuron. The dynamics is described by a following set of equations:

$$\begin{cases} \frac{d}{dt}V(t) &= V(a - V)(V - 1) - w + I \\ \frac{d}{dt}w(t) &= bV - cw \end{cases} \tag{3.14}$$

V is, as before, a membrane voltage, w stands for recovery current, I is an input current, a and b are parameters. When compared to ionic current models, the right-hand side of the equation for V was approximated with a cubic polynomial and w with a linear function. Despite those simplifications, the model is still capable of generating a spiking responses to the input stimuli described by current I. This idea of getting rid of computationally expensive functions was further used in the development of simple neuron model.

3.4.3 Izhikevich simple model

In a pursuit for accurate and efficient description of spiking process, a mathematician Eugene Izhikevich developed a *simple neuron*. It was first published

in his work in 2003, see [Izhikevich 2003]. This is a sort of phenomenological neuron, in other words it has not been derived from a biological system, but simple dynamics and parameters were adjusted in order to match biological recordings. The dynamics is still expressed in terms of differential equations system, but number of equations is kept at two, in addition all divisions and exponential functions were done away with, in order to facilitate the computation on floating-point arithmetic. To compensate a non-linear if-clause was added to the dynamics.

The model is defined as follows:

$$\begin{cases} \dfrac{dV}{dt} & = & I + 0.04V^2 + 5V + 140 - u \\ \dfrac{du}{dt} & = & a(bu - V) \\ \text{if } V \geq 30 & & V := c; \quad u := u + d; \end{cases} \tag{3.15}$$

where V is a membrane voltage potential and u is a dimensionless membrane recovery variable. a, b, c and d are parameters. The values for the case of regular spiking cell modelling the dynamics of cortical pyramidal cells of a mammal (see ibidem):

$$\begin{array}{ll} a & = 0.02 \\ b & = 0.2 \\ c & = -65(mV) \\ d & = 2. \end{array} \tag{3.16}$$

The parameter a stands for time scale of recovery variable u, b is its sensitivity, c is after-spike hyper-polarisation reset value of the potential and d denotes after-spike recovery of u.

The model and its properties were deeply studied in the monograph [Izhikevich 2007]. The most significant is an ability to mimic many of spiking regimes, noted in the cortical cells including: regular spiking, integrate-and-fire, resonate-and-fire, burst spiking, etc. In addition it can demonstrate wide range of qualitative changes in the spiking patterns by manipulating the parameters a, b, c and d.

The model found its applications in computational neurobiology, in large scale networks, which require vast resources.

3.4.4 Cable equation

However complex the dynamics of the single neural cell might be, the capabilities of single neuron are still limited. In order to put the models into a good use, one must connect a number of neurons into a *network*. To achieve

that, the network topology and inter-neural interactions need to be specified. The network topology is frequently described as a directed (or not) graph, see Section 2. Pretty much as the variety of neuron models, the inter-neural dynamics can also be described in a number of means. In this section we present the best recognised.

The diffusion equation is a dynamical system model for heat diffusion in solid objects. In this case it was adapted to model the diffusion of voltage along the membrane surface. The equation formula is

$$C\frac{\partial V}{\partial t} = \frac{2a}{R}\frac{\partial^2 V}{\partial x^2} + I - I_{\mathrm{Na}} - I_{\mathrm{K}} - I_{\mathrm{L}}, \tag{3.17}$$

where C stands for capacitance, a is a radius of the cable, R is a resistivity of the material, x is a geometrical coordinate, t is a time and of course the heat T was substituted with voltage V. In addition we account for ionic currents I_{Na}, I_{K}, I_{L} and external input current I.

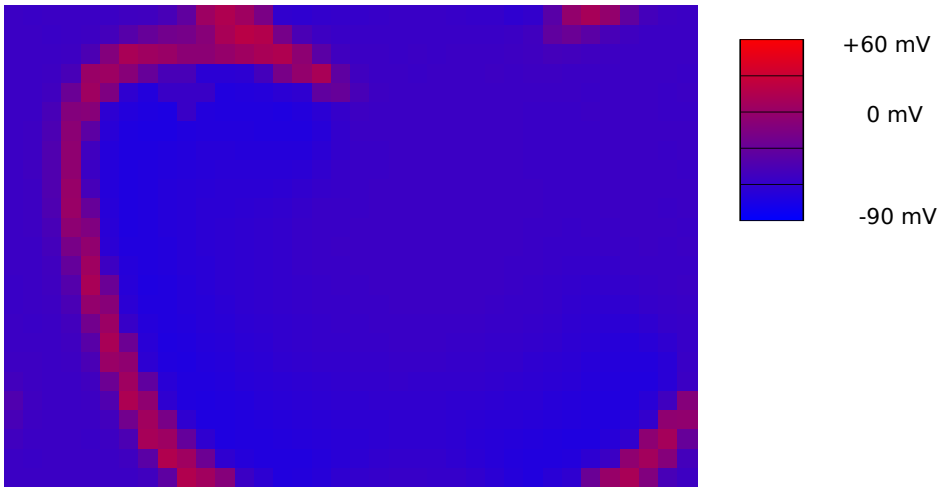


Figure 3.8: Propagation of the spike on the membrane surface with toroidal topology according to the cable equation dynamics.

The system is mainly used for impulse propagation along the membrane in axon or dendrite tree, though it can also be used for computing a current passed between the neurons, though there exist more elaborate means like neurotransmitter simulations. Note, that the equation takes into account spatial location of the cells. This makes the formula unsuitable for abstract artificial networks without geometrical embedding.

3.5 The state of the art in functional brain analysis

In 2005, Eguiluz in [Eguiluz et al. 2005] reported a power-law input degree distribution in correlation-based activity graph obtained from functional magnetic resonance imagining (fMRI) of natural brain activity during execution of simple tasks. This phenomenon continued the trend of finding scale free graphs (i.e. whose degree distribution obeys a power law [Albert & Barabasi 2002]), which had been reported in linguistic networks [Kello & Beltz 2009], www [Albert et al. 1999], social [Barabasi et al. 2002, Ugander et al. 2011], metabolic [Jeong et al. 2000] [Wagner & Fell 2011], computer programs [de Moura et al. 2003] etc. This somehow changed the perception of the brain network. Rather than a set of random or predefined all-to-all wires, a more hierarchical and functional-level based approach have been gaining popularity.

In 2009 the first analyses of scale-free phenomenon in artificial recurrent neural networks were carried by Piękniewski and Schreiber [Piękniewski 2008, Piękniewski & Schreiber 2008]. They showed both theoretically and experimentally, that in fully connected recurrent neural networks with stochastic driven dynamics the input (and output as well) degree distribution obeys the power law x^{γ}. In addition they estimated the exponent γ to the value -2.

In next chapter we extend (or maybe rather reforge) the approach to account for a not all-to-all connected graphs but, instead, make the neural network more dependent on its geometrical features as it is in the cortex [Eguiluz et al. 2005].

3.5.1 Emergence of scale-free phenomenon in recurrent neural networks

The *spike* or *activation flow model* for neural activity in recurrent neural networks was coined by Piękniewski and Schreiber in 2007 – 2008, see [Piękniewski & Schreiber 2007]. The assumption they took, was that it ought to be simple and mathematically tractable. The neurons are able to evoke their interactions depending on the previously accumulated activity. In fact, the restriction to the neural networks can be lifted and the model fits any abstract or real network, whose nodes are capable of accumulating some sort of quantity and exhibits this quantity-dependent interactions. As a result the model resembles the Boltzmann Machine with its range for the allowed spin values (a stored charge) extended to $\{0, .., M\}$, for some positive number M. The dynamics is based on charge-conserving and Hammilonian-driven spike

flow between a pair of neurons and does not have any obvious bias towards a scale-free or any hierarchical organisation.

By application of the mean field theory in asymptotic conditions they analysed the in-degree sequence of the neuron visits throughout the travel of the charge and showed that it obeys a power law with the exponent value equal to -2. Below we provide a sketch of the proof for recurrent setup after [Piękniewski & Schreiber 2008].

The said travel takes place according to the direction indicated by a growing *support* of the neurons (see definition in Section 4.1.4). Suppose that a unit of charge travels through neurons k_l, $l \geq 0$, starting from k_0. Hitting $k_l = 1$ indicates, that the unit of charge reached the neuron with the best support (first in a sorted decreasingly list). Consider a sequence of random variables $X_1, X_2, ...$, where X_1 is chosen form a uniform at $(0, 1)$ distribution: $X_1 \sim U_{(0,1)}$ and $X_{i+1} \sim U_{(0,X_i)}$ for $i > 1$. Now we can approximate $k_l = \lceil X_l \cdot |\mathcal{V}| \rceil$, where $|\mathcal{V}|$ is a number of neurons.

Define π_i to be a probability that a unit of charge visits a neuron u_i, that is $\pi_i := \mathbb{P}(\exists_l \quad k_l = i)$. For large $|\mathcal{V}|$ we have $\pi_i \simeq \mathbb{E}(l : X_l \in [\frac{l-1}{|\mathcal{V}|}, \frac{l}{|\mathcal{V}|}])$, and $\pi_1 = 1$. The input degrees are distributed according to binomial distribution with parameters π_i and n equal to the number of charge in the system.

Since X_i are uniformly distributed on $(0, X_{i-1})$, than $T_i := -\log X_i$ is exponentially distributed. Also $-\log X_{i+1} - (-\log X_i) = -\log \frac{X_{i+1}}{X_i}$, hence $T_{i+1} - T_i$ is also exponentially distributed. As a result T_i is a Poisson process and

$$\pi_i \simeq \mathbb{E}|\{l : T_l \in [-\log(\frac{i}{|\mathcal{V}|}, -\log(\frac{i-1}{|\mathcal{V}|})]\}| \simeq \frac{1}{i}.$$

We end with $D_i \simeq \frac{1}{i}$ and $|i : D_i > k| \simeq \frac{n}{k}$. And in non-cumulative form: $|i : D_i = k| \simeq \frac{n}{k^2}$.

Which concluded the proof of the main theorem of the dissertation of Piękniewski, see [Piękniewski 2008]:

Theorem In recurrent neural networks the input degree distribution obeys a power law

$$\mathbb{P}(deg_v = k) \propto \frac{1}{k^2}.$$

In following chapter we show, that similar property holds for sparsely connected networks with geometrical embedding.

Chapter 4

Thesis results

In this chapter we present main results of our work, namely the emergence of scale invariance and the small world phenomenon in neural networks. The contents concerning the former are based on our (co-authored with Tomasz Schreiber and Filip Piękniewski) publication [Piersa et al. 2010] and for the latter — on publication [Piersa 2011] and new preliminary results to be published.

In following sections we define a mathematically tractable model for functional activation networks in geometrical space. Then we analyse its structure in the context of said terms of random graph properties. Finally, we compare the results with their medically obtained counterparts.

As pointed out in the introduction the selected methodology mainly focuses on numerical simulations and, whenever possible, analytical predictions, both of them in the terms of random graph theory as sketched in Section 2.1. The theory seems to be a reasonable choice for quantitative description a medium-to-large scale graphs, while the simulation seems to be in tact with the model described in algorithmic terms. Of course, the analytical results are important part, which completes the simulation, though it is not always available. We note, that the direct analyses of the fMRI data, in the context described above, is beyond the scope of this work, due to lack of necessary devices, not to mention a permission to carry a research on human-acquired data. Hence, all the fMRI results, used for comparisons, origin form external researches and are referenced.

4.1 Basic model for neural activity

We start with definition of the formal model for neural activity in geometrical networks as introduced by Piersa, Piękniewski and Schreiber in

[Piersa et al. 2010].

4.1.1 Structural network

The underlying *structural network* is defined as a undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ constructed as follows.

Suppose we are given a two-dimensional sphere S^2 with radius R and expected density of units per square unit of the surface of the sphere $\rho \gg 1$. The number of *abstract neurons* n is picked from a Poisson distribution $\mathcal{P}(|S^2|\rho)$. Each neuron v is generated independently from a uniform distribution on the surface of S^2 and added to the vertices set $\mathcal{V}$, so it receives its euclidean coordinates $(x_i, y_i, z_i) \in \mathbb{R}^3$. In addition, an initial *abstract activity level* or an *accumulated charge* σ_i is independently assigned to the neuron. The activity is restricted only to non-negative integer values and the initial charge is assumed to be strictly positive $\sigma_i > 0$. The set of vertices $\mathcal{V}$ is defined as all generated neurons.

Now we generate edges or *synapses* of the network $\mathcal{E}$. For every pair of neurons $\{v, u\} \quad v, u \in \mathcal{V}$, a symmetric *edge* $e = \{v, u\}$ is included into $\mathcal{E}$ independently with probability

$$\mathbb{P}(e \in \mathcal{E}) = \begin{cases} 1 & d(e) \leq 1 \\ d(e)^{-\alpha} & \text{otherwise,} \end{cases} \tag{4.1}$$

where $d()$ is a euclidean distance and $-\alpha$ is a decay exponent. The above formula of the connectivity function was put forward in [Eguiluz et al. 2005]. In most of our considerations the value of $-\alpha$ is fixed about the value of the dimension of the sphere. Note, that assuming large density $\rho \gg 1$, the Formula (4.1) with overwhelming probability yields a connected graph. Indeed, suppose otherwise, that is the graph is disconnected. In the simplest case, this would require sampling a graph with a neuron v, such that no other was picked within a radius at least 1. For large R we can approximate the empty spherical cap with a circle. Hence, we have the probability of disconnected graph $\mathbb{P} \leq (1 - \frac{2\pi 1^2}{4\pi R^2})^{(\rho \cdot 4\pi R^2)}$, which for large R becomes $e^{-\rho/2\pi}$ and this approaches zero as the density ρ increases. The probability is even less, as we omitted the second case of (4.1), which can connect v with other neurons lying further than 1 unit of distance from it.

Each synapse $\{u, v\}$ receives its symmetric *weight* w_{uv} independently from the gaussian distribution $N(0, 1)$. The weight indicates the excitatory (when positive) or inhibitory (when negative) nature of the synapse. For simplicity the weight of edge, which is not included in the network, is assumed to be zero (neutral) $\forall_{e \notin \mathcal{E}} \quad w_e = 0$. In particular self-loop weights are zero $w_{uu} = 0$.

Having defined the structural graph we define an *energy* (Hammiltonian) of the system:

$$E(\bar{\sigma}) = \sum_{\{u,v\}\in\mathcal{E}} w_{u,v}|\sigma_v - \sigma_u|. \tag{4.2}$$

Te formula can be rewritten as a sum over pairs of neurons since weights of synapses absent in $\mathcal{E}$ are zero.

Clearly, it bears resemblance to the energy of the Hopfield network, see [Hopfield 1982] and Equation (3.9) in Section 3.3.3 in previous chapter, and has straightforwardly the same interpretation. Namely, it prefers neurons, which have similar amount of activity σ, for positive weight synapse $w_{uv} > 0$ and large differences in inhibitory case.

4.1.2 Evolution of the model

The network undergoes a following, charge-conserving, Kawasaki-style dynamics (put forward in [Kawasaki 1966]):

1. Iterate multiple times. Terminate the simulation once a stable state is reached.

   a. Randomly pick a pair of neurons (u, v) connected with a synapse, such that $\sigma_u \geq 1$.

   b. Attempt to transfer a single unit of charge from u to v along the synapse (i.e. $\sigma_u -= 1$ and $\sigma_v += 1$). Suppose $\bar{\sigma}$ is an initial activity vector and $\bar{\sigma}'$ is obtained from $\bar{\sigma}$ by performing the transfer.

   c. If the transfer reduces the energy of the network $E(\bar{\sigma}') < E(\bar{\sigma})$, than accept the transfer. In such case the unit of activity is said to *flow through a synapse* ($\bar{\sigma} := \bar{\sigma}'$). Otherwise accept the transfer with probability

   $$\mathbb{P}(\bar{\sigma} := \bar{\sigma}') = \exp(-\beta\Delta E), \tag{4.3}$$

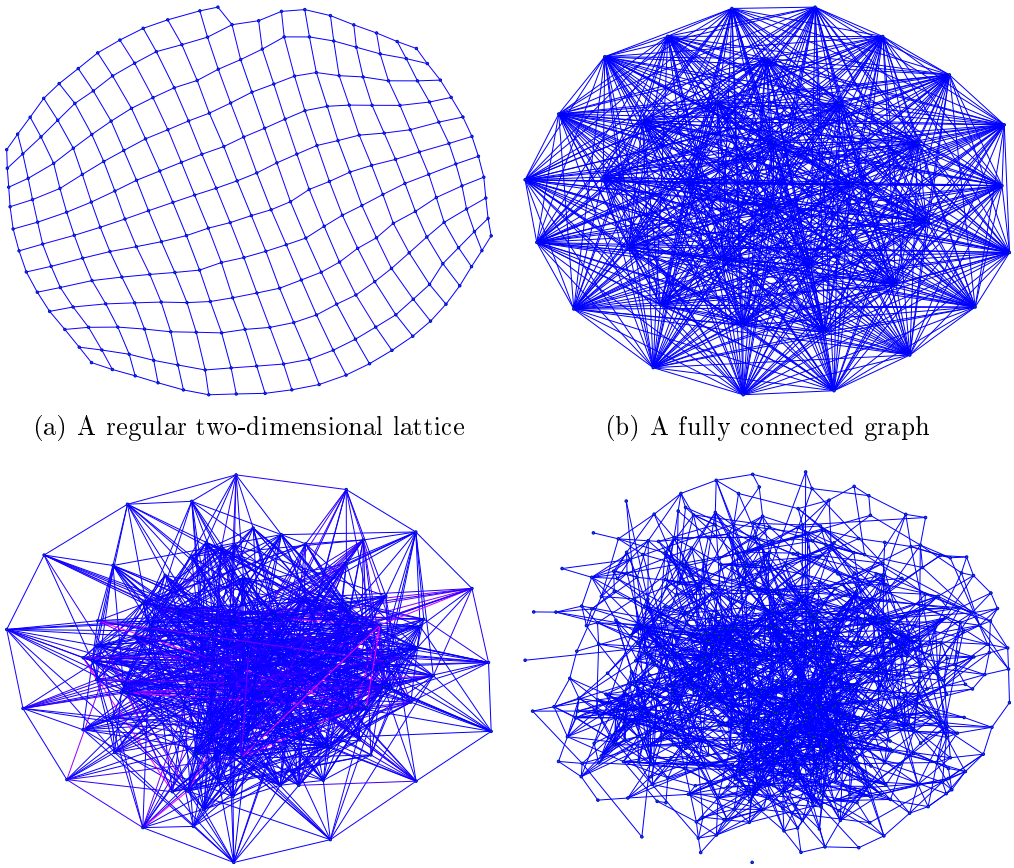
   where ΔE is an increase of energy caused by the transfer ($\Delta E = E(\bar{\sigma}') - E(\bar{\sigma})$).

The parameter $\beta \gg 0$ is an inverse temperature and is assumed to be large. Its interpretation is pretty much the same as an interpretation of the temperature in Boltzmann machines, see Section 3.3.3.

4.1.3 Functional network

The flow of activity through synapses is recorded throughout the dynamics. Let f_{uv} denote the total charge transferred from u to v, clearly $f_{uv} = 0$ if $\{u, v\} \notin \mathcal{E}$.

Definition The spike-flow activity graph or functional graph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ is defined as a subgraph of $\mathcal{G}$, consisting of all neurons $\mathcal{V}' = \mathcal{V}$ and the edges of $\mathcal{E}$, whose transfer was positive $\mathcal{E}' = \{e \in \mathcal{E} : \ f_e \geq 1\}$.



(a) A regular two-dimensional lattice

(b) A fully connected graph

(c) Erdős-Rényi graph

(d) Obtained functional graph

Figure 4.1: A comparison of regular lattice, a clique, ER graph and obtained activation-flow network. All the graphs were remapped onto a unit circle for better visibility.

These spike-flow graphs are the main focus of this work. Bassett and Bullmore [Basset & Bullmore 2006, Bullmore & Sporns 2009] discussed the

differences between structural and functional networks in fMRI scannings of the brain activity. Piękniewski [Piękniewski 2008] analysed a degree distribution in functional artificial neural networks assuming that the structural was a full graph. As it turns out, the difference can be substantial.

4.1.4 Evolution of the system

Since the dynamics is a charge conserving i.e. $\sum_{v\in\mathcal{V}} = const$, the possible charge configurations $\bar{\sigma} = (\sigma_1, ..., \sigma_n)$ form a space $\mathcal{F} = \{(s_1, ..., s_n) \subset \mathbb{Z}^n : \forall_i\ s_i \geq 0 \wedge \sum_i s_i = const\}$. In addition a jump between two states $\bar{\sigma}$ and $\bar{\sigma}'$ is possible if and only if they differ on exactly two positions and only by 1, that is $\exists_{i\neq j}\ \sigma_i = \sigma'_i + 1 \wedge \sigma_j = \sigma'_j - 1$ and $\sigma_k = \sigma'_k$ for all other $k \neq i, k \neq j$. In this case the dynamics, can be interpreted as a Markov Chain walk on the state space $\mathcal{F}$.

If the inverse temperature β is constant, the chain is uniform. In addition, in this case the chain is connected since the probability of passing between neighbouring states is always positive $\mathbb{P}(\bar{\sigma}'|\bar{\sigma}) \geq \frac{1}{|\mathcal{V}|}\exp(-\beta\Delta E) > 0$ and reaching every state of $\mathcal{F}$ is possible via a sequence of neighbouring swaps.

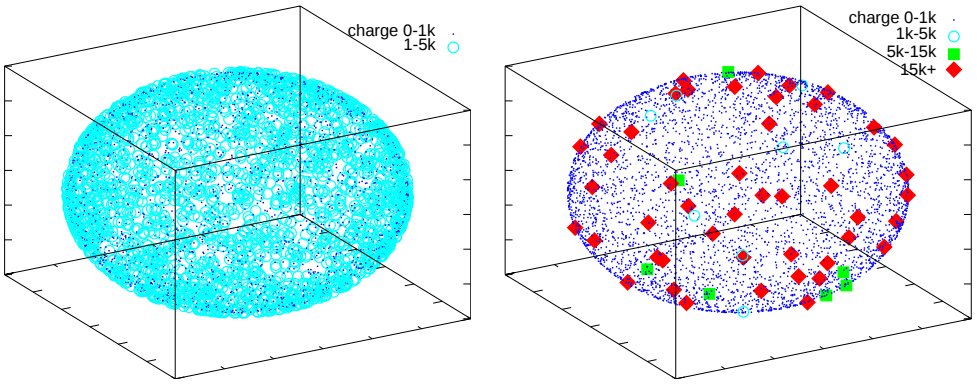
The chain is also irreducible, to see it we just need to notice that, the probability of staying in the same state is positive $\mathbb{P}(\neg(\bar{\sigma}' := \bar{\sigma})) = 1 - \exp(\Delta E) > 0$, at least for a minimal energy state $\bar{\sigma}$. Now the period of $\bar{\sigma}$ is equal to a one, $k = \gcd\{1, 2, ...\} = 1$. And from connectedness, 1 is a period of every state of $\mathcal{F}$.

Thus, there exists a unique stationary distribution of the states. We can approximate it as $\mathbb{P}(\bar{\sigma}) = \dfrac{\exp(-\beta E(\bar{\sigma}))}{\sum_{\hat{\sigma}}\exp(-\beta E(\hat{\sigma}))}$. For large inverse temperature β this distribution degenerates to one-point distribution centred on $\bar{\sigma}_{min}$ with minimal energy. However, due to the finite simulation time, it is more reasonable to consider a collection of minimum-energy states, which do not necessary are focused on single neuron.

For further considerations let us introduce a following definition:

Definition A support of the neuron u is a minus sum of the weights of the incident synapses $S_u := -\sum_i w_{ui}$.

We can sketch the dynamics with a modified *elite* and *bulk* picture, as described by Piękniewski, see [Piękniewski 2008]. The neurons, with especially high support values S_v are referred to as an *elite* and the remaining with low-to-average support as a *bulk*. With large inverse temperature β and in late, stable phases the dynamics asymptotically becomes a winner-take-all with the flow towards the neurons with higher support, see definition above.



(a) Charge distribution before the simulation.

(b) Charge distribution after the simulation.

Figure 4.2: Activity distribution in the network counting approximately 4k neurons before and after the evolution. The charge was concentrated in small number of neurons leaving the rest drained.

The number of opposing transfers becomes negligible. The elite accumulates all of the activity by draining it from the bulk through the available synapses. From the activity point of view, this can be seen as a sequence of leaps of the charge to neurons which are higher in the support hierarchy. Eventually, the unit of charge reaches the top of the growing-support chain and than is stuck in there. In fully connected networks there is a single maximum-support unit and the process can be seen as ascending a hill, though gathering all the charge in a single neuron can be long. In the case of sparse nets, we are more likely to end up with a collection of elite neurons, which are not connected to each other, and the whole process is rather ascending a poset, which has a number of distinct maximum elements.

The starting (roughly equal) setup of the network can be seen in Figure 4.2 left, while the ending phase in the right plot.

4.2 Geometrical embedding

The human cortex is organised in a highly folded surface. In addition the fraction of synapses of length d in biological brain decays as $\frac{1}{d^{\alpha}}$, see [Eguiluz et al. 2005]. Similar results for graph of functional brain areas were obtained in [Salvador et al. 2005]. Therefore we decided to account an impact of the geometrical location of the neurons on the results from the very beginning.

In addition to lengths of the synapses in the cortex, there are differences between purposes of the left and right hemispheres. The biological organisation of the brain reveals specialised areas, in which most of specific functional processes are carried for instance visual perception in occipital lobe. Going down the hierarchy we reach the *cortical column*, a group of few hundred highly coupled neurons. As a result the effective organisation of the brain, it seems to be highly dependent on its the local structure.

As defined in Section 4.1.1, the embedding should affect only the structural level of the network. We picked a unit sphere, mainly in order to ease some theoretical predictions. As the S^2 has full rotational symmetry, no neuron, which was picked from a uniform distribution, is biased towards any higher or lower number of neighbours. The decay exponent value of the connectivity function in Equation 4.1 was picked to be roughly equal to the dimension of the embedding space. Extremely high exponents would degenerate connectivity function to the case where any synapse longer than one unit would be rejected. This would eventually lead to locally-connected regular grid, which might be an abnormal structure.

For comparison purposes we provide the results from other embeddings: ball, cube and two-dimensional square as well as spheres with varying decay exponents (see Section 4.4.1). However, unless stated otherwise, the discussed results were obtained from two-dimensional spherical topology with decay exponent $\alpha = -2.5$.

4.3 Scale-freeness in activation flow model

In this section we discuss the origination of scale-free phenomenon in the activation-flow model.

4.3.1 Theoretical analysis

Consider a structural network in the model as described in Section 4.1.1. Below we provide a formal calculus by Schreiber (see [Piersa et al. 2010]) for input degree distribution.

Let us denote a cumulative distribution of (input) degrees as:

$$G(k) = \mathbb{P}(deg_{in}(v) \geq k), \tag{4.4}$$

where v is randomly selected neuron. Each neuron is ascribed its visiting intensity $\Phi(v)$ defined as number of units of charge, which visited v throughout

their transfer in the network. We can write Φ as

$$\Phi(v) = a + \sum_{u \to v} \frac{\Phi(u)}{|N^{\uparrow}(u)|}, \tag{4.5}$$

where $u \to v$ stands for $\{u, v\} \in \mathcal{E}$ and $S_u < S_v$ (for a definition of the support see Section 4.1.4), and $N^{\uparrow}(u) = \{v \in \mathcal{V} : \ u \to v\}$. By using a random variable $\mathbb{1}_{u \to v}$, which is equal to $+1$ if holds $u \to v$ and 0 otherwise, we can rewrite (4.5) as

$$\Phi(v) = a + \sum_{u \in \mathcal{V}} \frac{\mathbb{1}_{u \to v} \Phi(u)}{\sum_{y \in \mathcal{V}} \mathbb{1}_{v \to y}}. \tag{4.6}$$

Note, that the construction of the network ensures us, that $\mathbb{1}_{u \to v}$ are independent. In addition the symmetric setup has no bias towards any of the neurons. In such case we argue that Φ is in fact a function of μ_v rather than v where μ_v is a chosen uniformly between $[0, 1]$ *support mark* indicating what fraction of neurons has lower support than v $\mu_v > \mu_u \iff S_v > S_u$.

$$\Phi(v) = \varphi(\mu_v) \tag{4.7}$$

Now we can rewrite

$$\begin{aligned} \mathbb{P}(u \to v) &= (1 - \mu(u)) g(|u - v|) \\ \mathbb{P}(v \to u) &= \mu(u) g(|u - v|), \end{aligned} \tag{4.8}$$

where g is a connectivity probability function (recall Equation 4.1): $g(x) = \begin{cases} 1 & x \leq 1 \\ x^{\alpha} & x > 1 \end{cases}$.

$$N^{\uparrow}(v) \approx \rho(1 - \mu_v) \int_{S_{d-1}} g(|u - v|) \xi(dv), \tag{4.9}$$

where ξ is a Lebesgue sphere surface measure. We rewrite (4.9) as:

$$N^{\uparrow}(v) \approx \rho(1 - \mu_v)\gamma, \tag{4.10}$$

with γ equal to the spherical integral $\gamma = \int_{S_{d-1}} g(|u - v|)\xi(dv)$. This requires the density ρ to be large $\rho(1 - \mu_v) \gg 1$. Now we put together (4.7) to (4.9):

$$\begin{aligned} \varphi(\mu_v) &= a + \rho \int_{S_{d-1}} \int_0^{\mu_v} g(|v - u|) \frac{\varphi(\mu_u)}{\rho(1 - \mu_u)\gamma} d\mu_u \xi(dv) \\ &= a + \int_0^{\mu_v} \frac{\varphi(\mu_u)}{(1 - \mu_u)} d\mu_u \ \frac{\int_{S_{d-1}} g(|v - u|)\xi(dv)}{\gamma} \end{aligned} \tag{4.11}$$

which reduces to

$$\varphi(x) = a + \int_0^x \frac{\varphi(y)}{(1-y)} dy. \tag{4.12}$$

After solving

$$\varphi(x) = \frac{a}{1-x}. \tag{4.13}$$

From the construction of the network we have, that the average number of neurons $v \in \mathcal{V}$ is $\rho \cdot \xi(S^2)$ and the number of neurons v such that $\mu_v \in (s, s+t)$ is $\rho \cdot t \cdot \xi(S^2)$. Assuming large density of the neurons $\rho \to +\infty$ we can approximate $deg_{in}(v) \simeq \Phi(v) \simeq \varphi(\mu_v)$. As a result (from (4.4) (4.7) and (4.12))

$$G(k) = \mathbb{P}(deg_{in}(v) \geq k) \simeq \int\limits_{s:\varphi(s)\geq k} ds \simeq \int\limits_{1-\frac{1}{k}}^{1} 1 ds = \frac{a}{k} \tag{4.14}$$

and since G is a complementary cumulative function, we conclude that

$$\mathbb{P}(deg_{in}(v) = k) \propto \frac{1}{k^2}. \tag{4.15}$$

Which ends the proof of following theorem.

□

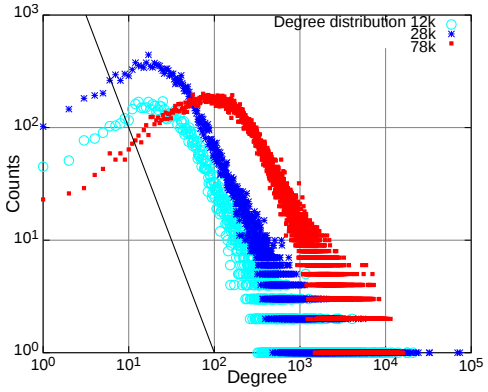
Theorem In geometrically embedded spike-flow network the input degree distribution function of the vertices obeys power law with the exponent value equal to -2

$$\mathbb{P}(deg_{in}(v) = k) \propto \frac{1}{k^2}. \tag{4.16}$$

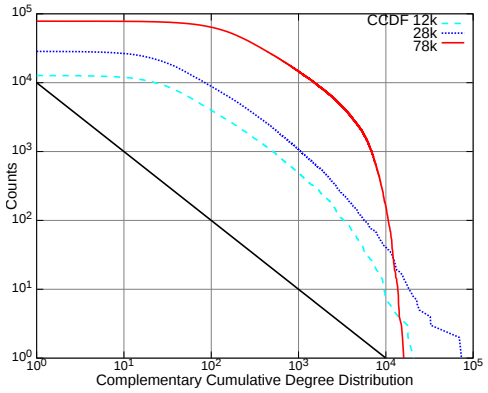
Since the difference between visiting and leaving activation is equal to the initial σ, (except for small number of neurons which gathered more charge, but their number is negligible), the same property is held for the out-degree distribution.

4.3.2 Numerical results

The model was implemented in computer simulation. The obtained empirical input degree and complementary cumulative input degree distribution is presented in Figure 4.3. In both cases the (exponentially truncated) power-law dependency is clearly observable. The truncation in the tail of the distribution and seems to originate from a finite sample size. The peak of the distribution centred in $deg \simeq 100$ is approximately equal to initial mean



(a) A degree distribution function $d(k) = \#\{v : \deg(v) = k\}$



(b) A complementary cumulative degree distribution function $d_c(k) = \#\{v : \deg(v) \geq k\}$

Figure 4.3: A logarithmic plot of input degree distribution and complementary cumulative distribution function in obtained functional network counting 38k neurons. For a reference a solid straight line denotes exact power law functions $c \cdot \frac{1}{x^2}$ (left) and $\frac{c}{x}$ (right), which in log-log plot are straight lines with their slope equal to the exponent value in original formula.

value of the activity σ. For the cases, in which the initial charge was equal to one unit, the maximum moved to $deg = 1$ and the peak disappeared, see Figure 4.4.

The exact values are presented in Table 4.1. Note, that the slope was calculated for complementary cumulative degree distribution function (CCDF), so the exponent value is lower by one. The slope was estimated with least square linear regression and the mean value is $\bar{\gamma} = -2.0912$, which seems to be a good approximation for a theoretical value. In most of the cases the dynamics led to draining of over 98 per cent of the neurons from their charge, which seems to be a reasonable convergence of the system dynamics.

Interestingly, quite similar results were obtained in [Eguiluz et al. 2005] form functional Magnetic Resonance Imagining (fMRI) data. Recall, that their functional graph was obtained by thresholding a correlation matrix of small voxels. The obtained (asymmetric) degree distribution decayed as a power low x^γ, with γ estimated between -2.0 and -2.2 depending on the threshold. In addition, this is in agreement with Ising-like dynamics reported in [Fraiman et al. 2009], though the scaling exponent was not provided. Similar result, but in dense networks without accounting for embedding, was obtained by Piękniewski, see [Piękniewski 2008]. It turned out that geometrical embedding, yields strikingly comparable results to those obtained in the

Table 4.1: Approximated exponent value obtained in simulations. Table columns include number of neurons, number of connections, number of iterations, approximated slope value, network geometry (S^2 is a sphere, $(0..d)^3$ is three-dimensional cube) and fraction of nodes storing all network charge. Abbreviations: k = $\times 10^3$, M = $\times 10^6$, G = $\times 10^9$. The table is reprinted after [Piersa et al. 2010].

Neurons	Connections	Iterations	Slope	Geometry	Units with charge
9k	580k	70M	-1.062	S^2	0.0065
9k	560k	100M	-1.060	S^2	0.01
11k	700k	100M	-1.100	S^2	0.009
12k	800k	100M	-1.118	S^2	0.008
12k	800k	150M	-1.100	S^2	0.009
12k	800k	150M	-1.053	S^2	0.0087
13k	960k	150M	-1.066	$(0..d)^3$	0.008
19k	1.3M	150M	-1.082	S^2	0.011
21k	1.7M	200M	-1.170	$(0..d)^3$	0.008
21k	1.7M	200M	-1.129	$(0..d)^3$	0.01
32k	2.1M	300M	-1.102	S^2	0.0076
33k	2.7M	200M	-1.096	$(0..d)^3$	0.019
40k	3.4M	500M	-1.066	$(0..d)^3$	0.0065
45k	3.0M	500M	-1.103	S^2	0.0082
48k	4.2M	800M	-1.063	$(0..d)^3$	0.0056
48k	4.2M	600M	-1.143	$(0..d)^3$	0.019
50k	3.4M	800M	-1.080	S^2	0.016
50k	3.5M	800M	-1.058	S^2	0.0078
58k	4M	900M	-1.091	S^2	0.0082
58k	5M	1G	-1.081	$(0..d)^3$	0.006
78k	5.4M	1.2G	-1.063	S^2	0.0084
81k	5.6M	1.4G	-1.040	S^2	0.0082
88k	6.2M	1.6G	-1.058	S^2	0.0078
95k	6.7M	1.8G	-1.081	S^2	0.0081

mean field model.

In Table 4.2 we provide a comparison of exponent values for power law in degree distributions reported in medical researches and computer simulations, including listed above. Despite the varying size of the networks, most of them confirm a power law, the only exception is [He et al. 2007], in whose case the dependency was exponentially truncated. Moreover all the reported exponents are approximately equal to each other.

Table 4.2: Comparison of exponent values in power law formula of the degree distribution in fMRI data and mathematical models of neural activity. Columns from the left: type of the network / model, size of the network, estimated value of the exponent, notes (for instance tuned parameters), source of the data.

Network	Size	γ	Notes	Source
Brain fMRI *	$31 \cdot 10^3$	-2.0	$r_c = 0.6$	[Eguiluz et al. 2005]
Brain fMRI *	$17 \cdot 10^3$	-2.1	$r_c = 0.7$	[Eguiluz et al. 2005]
Brain fMRI *	$4.8 \cdot 10^3$	-2.2	$r_c = 0.8$	[Eguiluz et al. 2005]
Cortex fMRI **	54	-1.34	truncated	[He et al. 2007]
Ising model ***	$40 \cdot 10^3$	$\simeq -2$	$T = 2.3$ (cr.)	[Fraiman et al. 2009]
Mean field	$4 \cdot 10^3$	-2 ± 0.03	Full graph	[Piękniewski 2008]
Activity flow$^+$	$58 \cdot 10^3$	-2.098	S_2	[Piersa et al. 2010]
Activity flow^{++}	$10 \cdot 10^3$	-1.8004	low σ_u	

* Depending on the threshold parameter r_c.
** Obtained exponentially truncated power law i.e. $\mathbb{P}(deg = k) \propto k^{\gamma} e^{k/k_c}$.
*** The exact value of γ was not calculated. Yields a scale-free distribution only for a critical value of temperature T.
$^+$ Spherical embedding, moderate initial activity.
$^{++}$ Averaged over various embeddings, small initial activity σ_u.

4.3.3 Scale-freeness in geometrical embeddings

To ensure the validity of the predictions we performed additional simulations in more varying embeddings. First, we lifted the restriction concerning strictly spherical topology and extended the possibilities to cubes, balls (two- and three-dimensional) and spheres (one- and two- dimensional). Secondly, we checked wider range of decay exponents α, see Equation (4.1) between -2 and -8. Clearly the lower exponent, the less edges in underlying structural network.

For the sake of simplicity we reduced amount of activation per neuron in starting set-up down to one unit and restricted the size of the network to small samples counting roughly $1.0 - 1.2 \cdot 10^4$ units. This allowed us to apply a linear regression directly to the logarithm of the distribution (unlike complementary cumulative distribution as in previous section), although at the expense of accuracy of the simulation.

Nonetheless the power-law shape of the degree distribution was preserved, see Figure 4.4. This turned out to be true even for the most extreme values of the decay exponent α. As noted before, the characteristic peak of the

distribution at the mean starting activation has gone, or rather 'shifted' to the value 1.

Table 4.3: A comparison of various embeddings and obtained exponent values in degree distribution. Columns form the leftmost: the embedding topology, a decay exponent value α (see Equation (4.1)), size of the network, average connectivity, approximated value of the degree distribution exponent.

Top.	α	n	$\langle k \rangle$	γ	Top.	α	n	$\langle k \rangle$	γ
	-2.5	12782	18	-1.901		-2.5	12683	66	-1.675
	-3.5	12561	15	-1.944		-3.5	12233	36	-1.788
Sphere	-4.5	12599	14	-2.005	Sphere	-4.5	12301	28	-1.758
1d	-5.5	12498	13	-2.069	2d	-5.5	12663	26	-1.810
	-6.5	12480	13	-2.032		-6.5	12530	24	-1.849
	-7.5	12637	13	-2.064		-7.5	12589	22	-1.809
	-2.5	12726	60	-1.746		-2.5	11414	156	-1.700
	-3.5	12714	36	-1.823		-3.5	11516	77	-1.702
Ball	-4.5	12402	28	-1.782	Ball	-4.5	11516	51	-1.742
2d	-5.5	12519	25	-1.767	3d	-5.5	11518	41	-1.774
	-6.5	12515	23	-1.818		-6.5	11478	35	-1.708
	-7.5	12495	22	-1.814		-7.5	11239	31	-1.803
	-2.5	10889	58	-1.729		-2.5	10033	146	-1.777
	-3.5	10870	35	-1.755		-3.5	10116	73	-1.733
Cube	-4.5	10904	28	-1.735	Cube	-4.5	10062	49	-1.678
2d	-5.5	10915	25	-1.842	3d	-5.5	9815	39	-1.706
	-6.5	10810	23	-1.836		-6.5	10039	34	-1.674
	-7.5	10979	22	-1.825		-7.5	9990	31	-1.643

The results are summarised in Table 4.3 and in Figure 4.5. The small size of the sample and limited amount of charge have their impact on precise value of the exponent. The average value is $\bar{\gamma} = -1.8004$ and with the standard deviation $s_\gamma = 0.10832$. These values vary from theoretical value -2, but we conclude that, they still provide a reasonable approximation, especially when take into account limited sample. Interestingly, the results obtained form a a circle (one-dimensional sphere) resulted in much better approximations of the theoretical value.

It might seem surprising that the geometrical embedding preserves scale-freeness of the degree distribution, as it dramatically reduces the number of available edges in the system. On the other hand the considerations in

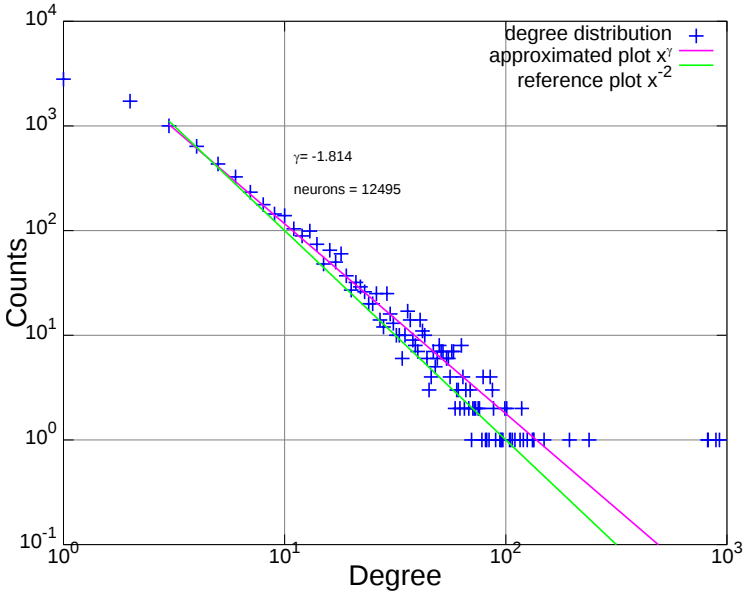


Figure 4.4: A plot degree distribution for two-dimensional ball, with decay exponent $\alpha = -7.5$. The scale-free shape is clearly preserved, although the estimated slope deviates from theoretical value -2.

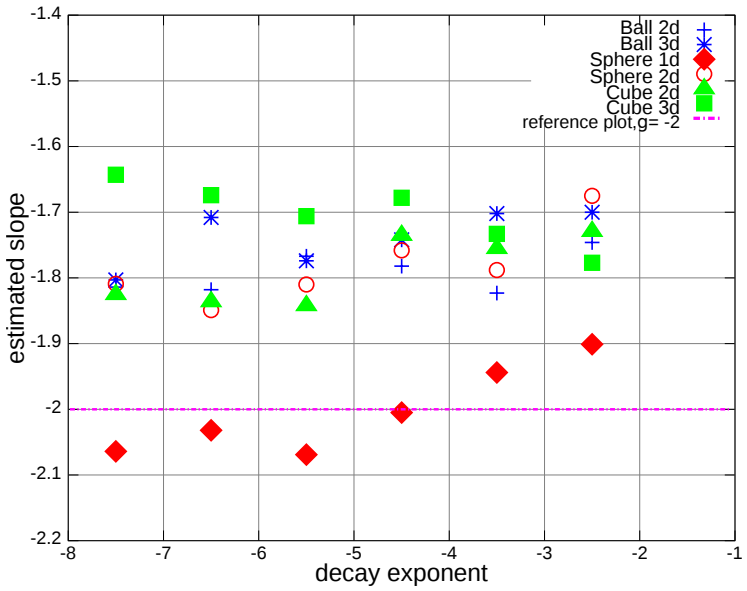


Figure 4.5: A plot of estimated exponent value in power law degree distribution vs a decay exponent α for embeddings: two-dimensional spherical, three-dimensional ball, three cube. See also Table 4.3.

Section 4.3.1 were completely independent from the formula of g, we only assumed it produces a connected network. Hence we confirmed that the model yields a power-law degree distribution in functional network even for sparse graphs with limited long-range connections for various embeddings.

4.4 Small-world phenomenon in artificial neural networks

In this section we discuss the small-world phenomenon in spike-flow graphs of geometrical neural networks, in the model described in Sections 4.1.1-4.1.3. The results are a compilation of our works concerning the diameter of the graph [Piersa & Schreiber 2010] and [Piersa 2011], as well as some new results focused on the clustering, yet to be published. Compared to previous sections, the results of this one are slightly more numerically-oriented.

The emergence of small-world structures had been reported earlier in a wide range of networks starting from acquaintanceship [Gurevitch 1961], collaboration and co-authorship [Barabasi et al. 2002], film co-starring [Herr et al. 2007], social network portals like Facebook [Ugander et al. 2011]. Bullmore and Basset in their work [Basset & Bullmore 2006] confirm the existence of small-world organisation on the structural level of brain activity obtained from the fMRI imagining, obtained in a similar way as in the work of Eguiluz, see [Eguiluz et al. 2005].

The term *small-world* was coined by Watts and Strogatz in their publication in 1998 [Watts & Strogatz 1998]. Unlike previous social-based analyses, they turned attention not only to the availability of the short path, but also a local structure of the network. The characteristic or average path length of the Erdős-Rényi random graph scales as $\log(|\mathcal{V}|)$, see [Chung & Lu 2006], Chapter 5. This however does not make it small world, especially in a sparse cases. This is where the graph clustering coefficient (or simply clustering) steps in. A small-world graph should be characterised not only by a short average path length, but also it ought to have a high clustering.

4.4.1 Numerical results

Humphries et al. in [Humphries et al. 2006] analysed small-world phenomenon in a brain stem formation. They defined a *path length ratio* is defined as $\lambda = \dfrac{L_{\text{real}}}{L_{\text{teo}}}$, where L_{real} is an average path length of the analysed graph and L_{teo} is the mean path length in Erdős-Rényi graph, counting the same number of vertices and edges. This ratio should approximately be equal to a

unity.

Quite similarly they defined a *clustering ratio*[1] as $\gamma = \frac{C_{\text{real}}}{C_{\text{teo}}}$, where C_{real} is a clustering of the analysed graph and C_{teo} is a equivalent Erdős-Rényi graph. Since the clustering coefficient in ER graph is approximately equal to the connectivity probability p_c, this ratio should be large, especially for sparse small-world graphs.

Finally they defined a *'small-world-ness' indicator* as $\sigma = \frac{\gamma}{\lambda}$, namely small-world graph should have this indicator value large $\sigma \gg 1$.

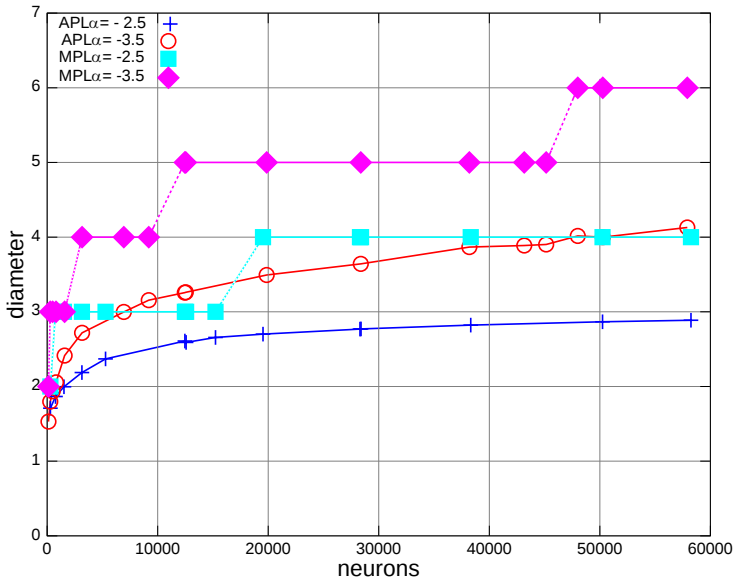
Characteristic path length

In our considerations we apply the model as coined in previous sections. The obtained spike flow graph was analysed for its characteristic path length. In addition we provide results concerning a maximum path length in the network (diameter).

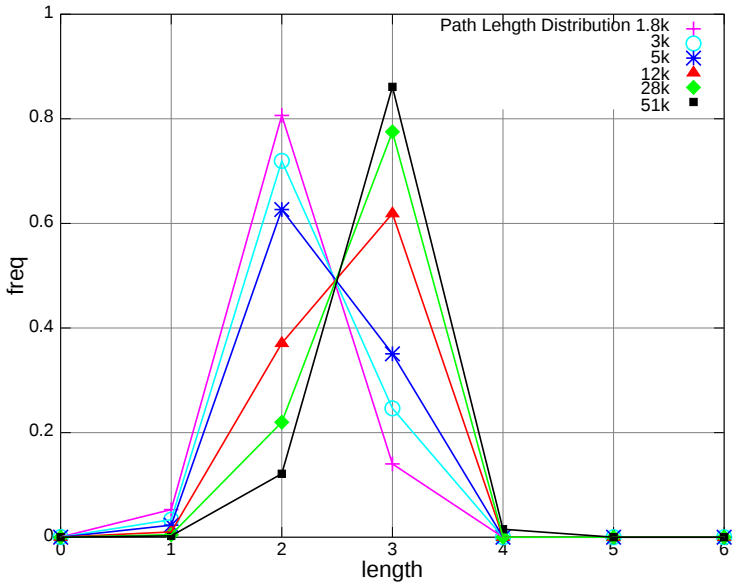
The results are presented in Figure 4.6(a). As it is visible, both the characteristic and the maximum path length grow quite slowly compared to the networks size. The diameter does not exceed 6 hops for $\alpha = -3.5$ and 4 hops for $\alpha = -2.5$ (decay exponent in underlying structural graph, see Equation (4.1) and the discussion in [Eguiluz et al. 2005]) for networks counting up to $6 \cdot 10^4$ neurons. The characteristic path length barely reached 4 and roughly 3 for cases $\alpha = -3.5$ and $\alpha = -2.5$ respectively. The size of obtained networks have is still orders of magnitude smaller than social graph, nonetheless the diameter is still short.

When focus on exact path length distribution (Figure 4.6(b)), it is not surprising to find out that most of the distribution mass is concentrated on 3 points. The remaining values up to the diameter are negligible, though positive. The shape of the distribution slowly moves right (grows) as the number of neurons increases.

[1] From now onwards, the symbol γ is used as a clustering ratio and should not be mistaken for the exponent value in degree distribution. Both these notations are adopted from existing works [Humphries et al. 2006] (clustering) and [Piękniewski 2008] (exponent), and we concluded that altering one of them would bring even more confusion.



(a) A plot of characteristic and maximum path length in obtained spike-flow graph after the evolution of the network (after [Piersa 2011]



(b) A Distribution of the path lengths in the spike-flow network. Note that the number of paths equal to 1 is non-zero (though still negligible)

Figure 4.6: Up: A plot of characteristic and maximum path length. Bottom: A plot of distribution of the path lengths in the spike-flow network.

Clustering coefficient

The clustering coefficient as defined in Section 2.1.2 can be considered as a cliquishness indicator. In social terms this is a equivalent to groups in which everyone knows each other. In neural networks this might be taken as a brain area, in which all neurons cooperate (or compete) with each other.

As presented in Figure 4.7(a), the clustering sharply drops for small network size (which is quite reasonable as with radius R approaching 1, the graph becomes almost fully connected) and stabilises around 10^{-1} for larger samples, though small decay is still noticeable. This is high, when compared to the network size and average connectivity.

The plot of the distribution of the clustering coefficient is depicted in Figure 4.7(b). It has a shape of unimodal function with its peak around its expected value.

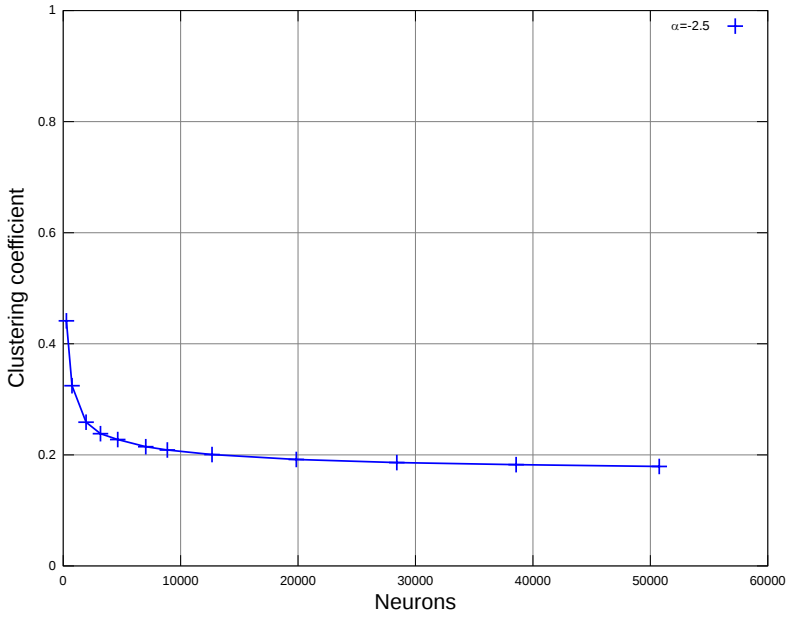
The most interesting feature is, that the clustering is one to two orders of magnitude higher than in equivalent ER model (see next section for more details) and this difference seem to be even larger as the size of the network increases. Thus, the local structure of the functional graph is highly non-trivial. The flow of the activity brings about the synchronisation of the whole groups of nodes, which in the end pick one of them as a representative, who sends the information further, up the support-induced hierarchy ladder.

Path-length and clustering ratios

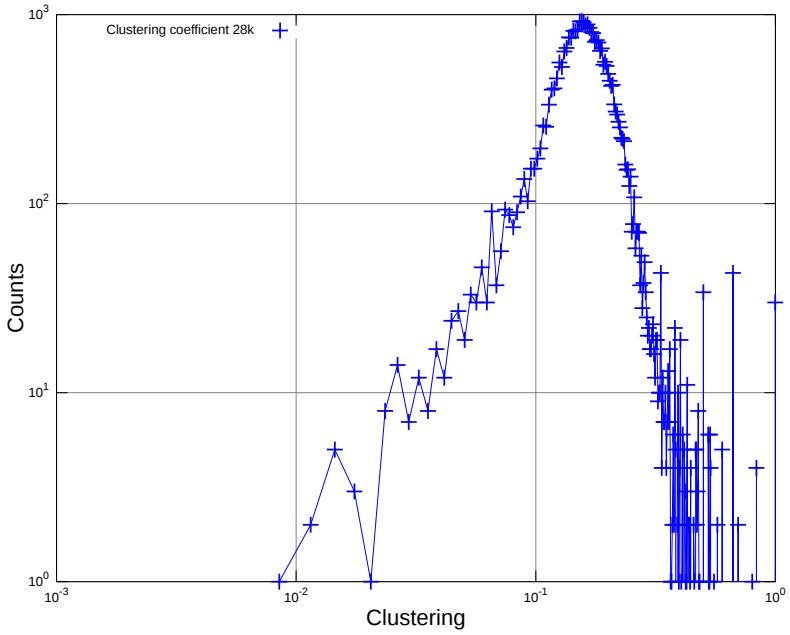
In order to get the larger picture we provide the real and theoretical connectivity and path length values $L_{\text{real}}, L_{\text{teo}}, C_{\text{real}}, C_{\text{teo}}$. The C_{teo} can be easily estimated as an average connectivity in the graph. The remaining were empirically computed from obtained graphs for sizes varying up to $5 \cdot 10^4$ and the decay exponent (see Equation (4.1)) equal to -2.5, which is somewhere 'in between' the dimension of the sphere and those of the embedding space $\mathbb{R}^3$.

According to the definitions in Section 4.4.1 this requires to compute clustering and path length values for ER graph with the same number of vertices and edges. The plots for various network sizes are presented in Figure 4.8 and the numerical values are provided in Table 4.4.

Not only are the path lengths for both network types of similar magnitude, but also they are roughly equal in values. As a result path length ratio $\lambda = \dfrac{L_{\text{real}}}{L_{\text{teo}}}$ is almost constant and equal to the unity. This means that the flow through the obtained spike-flow network is very efficient and any information can reach its destination as quickly as in random graph model.

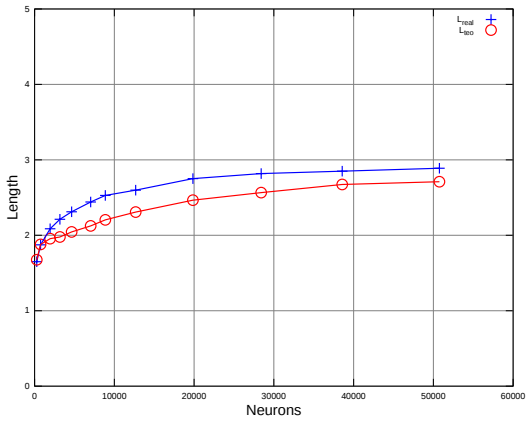


(a) A plot of clustering coefficient in obtained spike-flow graph vs size of the network

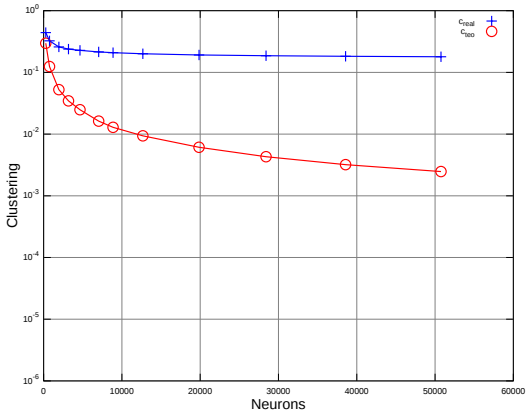


(b) A plot of empirical distribution of clustering coefficient.

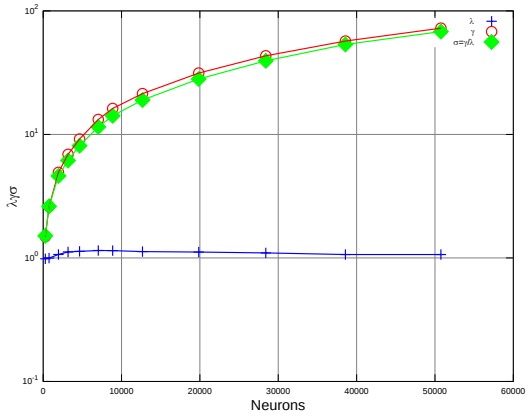
Figure 4.7: A plot of clustering coefficient of the spike-flow activity graph for various network sizes and a typical distribution of clustering in the network



Real (i.e. from spike-flow graph) and theoretical (i.e. equivalent to ER model) characteristic path length $L_{\text{real}}, L_{\text{teo}}$



Real and theoretical clustering coefficient $C_{\text{real}}, C_{\text{teo}}$



Path length ratio λ, clustering ratio γ and 'small-world-ness' indicator $\sigma = \dfrac{\gamma}{\lambda}$

Figure 4.8: A comparison of plots of characteristic path length and clustering coefficient for obtained spike-flow graph and equivalent Erdős-Rényi graph model — top and middle. A path length, clustering ratios and 'small-world-ness' indicator — bottom plot.

Table 4.4: Numerical values of the properties of obtained functional graphs. The columns from the leftmost include: sample size, characteristic path length, theoretical path length (in ER graph), clustering coefficient, theoretical clustering coefficient (in ER graph), path length ratio, clustering ratio, 'small-world-ness' indicator (see description in text).

Neurons	L_{real}	L_{rand}	C_{real}	C_{rand}	λ	γ	σ
274	1.64	1.67	0.44	0.29	0.98475	1.48	1.51
751	1.87	1.87	0.32	0.12	0.99716	2.60	2.61
1948	2.08	1.95	0.25	0.052	1.06722	4.92	4.61
3176	2.21	1.97	0.23	0.034	1.11862	6.89	6.16
4644	2.31	2.04	0.22	0.024	1.13089	9.16	8.10
7035	2.44	2.12	0.21	0.016	1.14908	13.21	11.50
8865	2.52	2.20	0.20	0.012	1.14681	16.22	14.14
12677	2.59	2.30	0.20	0.0093	1.12570	21.38	18.99
19854	2.75	2.46	0.19	0.0061	1.11605	31.39	28.13
28408	2.81	2.56	0.18	0.0042	1.09826	43.31	39.44
38567	2.84	2.67	0.18	0.0031	1.06611	57.00	53.46
50757	2.88	2.70	0.17	0.0024	1.06608	72.50	68.01

On the other hand, the clustering coefficients of the activation-flow equivalent ER networks and are approximately equal only for small number of neurons. The clustering of the functional graph drops slowly, but it did not fell under 10^{-1}), while this value for ER graph decreases by 2 to 3 orders of magnitude.

This dramatically alters the properties of the resulting graph. The theoretical ER it is a sort of random casual-effect scenario, in which there is a hardly any structure in the flow graph. However, from the high clustering we conclude, that there is a local competition in the neighbourhood of the unit. Not only the best-supported one (see definition in Section 4.1.4) drains the activity from the set of the neighbours, but also the incident neurons compete for the activity.

This can be read otherwise. The well-supported units cooperate with each other in gathering the charge scattered along the entire neighbourhood. In the end they transfer the collected activity towards a single designated 'leader'.

As a result we conclude, that the local cooperation in activity-flow graphs exhibit high level of 'small-world-ness', which can be key for obtaining high efficiency of the system, which emerged throughout the stochastic dynamics.

4.4.2 Comparison with numerical data

The results, obtained in presented activity-flow model, are consistent with those reported by neuroscientists. We used results obtained and discussed in works [Achard et al. 2006], [Eguiluz et al. 2005], [He et al. 2007] and [Salvador et al. 2005]. In addition we provide the results obtained from Ising-based correlation model in critical temperature, which was put forward in [Fraiman et al. 2009].

The network sizes vary from about 50 in graphs of brain regions, up to 10^4 in the case of fMRI voxels and computer simulation models. However we point out lack of 'medium sized' samples (of orders 10^2 and 10^3), which makes the comparison a bit tricky. Both the path length and clustering coefficient gathered from mentioned above medical analyses of brain networks as well as theoretical models are summarised in Table 4.5 and Figure 4.9

In all of the cases the measured path length scales logarithmically with the network size and is bounded between 2 and 14 edges. Interestingly, in most of the cases L_{rand} for equivalent random model is slightly smaller than L_{real}, but the difference seems to be negligible and, by extension, the path length ratio λ for most of the data is varies between the values one and three.

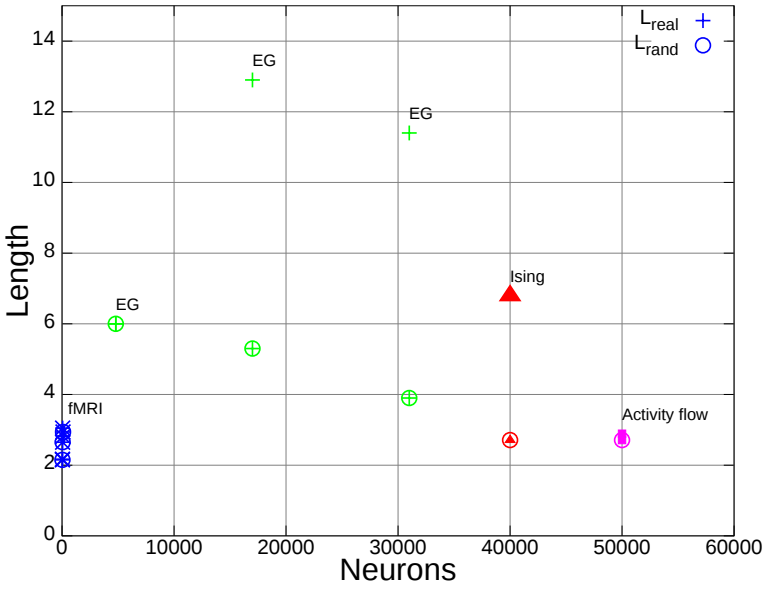
The clustering ratio γ has wider range, between 2 and 350 depending on the source and acquisition method, but this value is much more affected by size of the sample. Assuming approximately constant value (this is very crude approximation) for C_{real} and $\frac{\langle k \rangle}{|\mathcal{V}|}$ as an estimation for C_{rand} we obtain a sub-linear growth. Though, that the above reasoning is highly heuristic.

A very similar property holds for a 'small-world-ness' indicator σ, which is between 1.9 and abut 150. Note, that by definition the theoretical 'small-world-ness' indicator for Erdős-Rényi graphs is equal to 1 (which means no small-world at all).

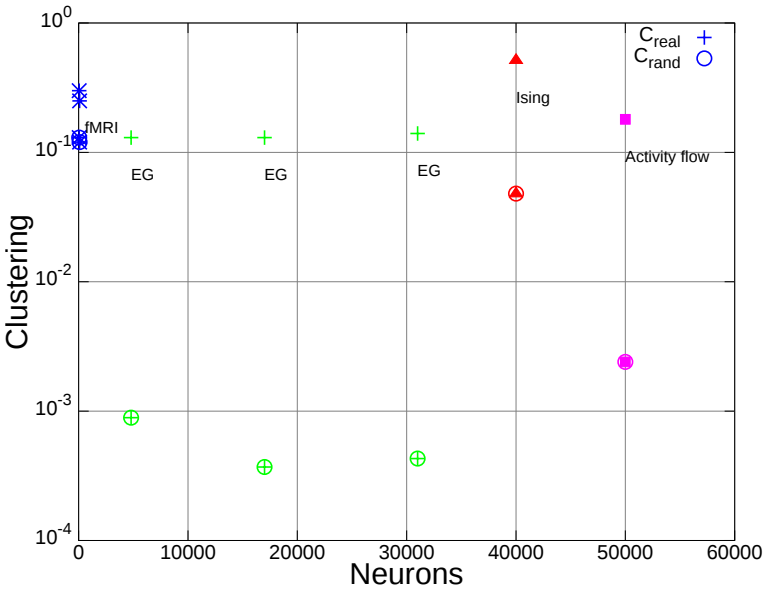
Taken together, the activity-flow model gives strikingly similar effect as the medical imagining data, and the emergence of small world phenomenon is present functional networks obtained from both of in these sources.

The direct comparison with small samples can be hard due to a dramatical difference of the number of units. On one hand, the activation-flow model turns out to be a surprisingly good approximation of clustering results of Eguiluz. On the other hand the path length ratio resembles more the values obtained for small sizes graphs as those in work of Salvator et al.

Interestingly, the Ising model of Fraiman et al. seems to have an oppose properties, in other words path length ratio λ is larger than one, as in Eguiluz et al., while the clustering ratio is somewhere in the middle between two extremes of medical values.



(a) A plot of characteristic path length vs size of the network.



(b) A plot of clustering coefficient vs size of the network.

Figure 4.9: Comparison of the analysed characteristics in medical data and simulations. Notations: Eg — [Eguiluz et al. 2005], fMRI — brain regions, Ising — spin-glass model [Fraiman et al. 2009], AF — presented activity-flow model.

Table 4.5: Comparison of characteristic path lengths and clustering coefficients of functional graphs from fMRI data and simulations. Columns from left include: type of the network, size of the network, obtained characteristic path length, path length for equivalent ER graph with identical number of vertices, path length ratio λ, obtained value of clustering coefficient, value of clustering coefficient for equivalent ER graph, clustering ratio γ, 'small-world-ness' indicator σ, notes, source.

Network	Size	L_{real}	L_{rand}	λ	C_{real}	C_{rand}	γ	σ	Notes	Source
Macaque cortex	39	2.17	2.15	1.01	0.38	0.15	2.53	2.50		[Stephan et al. 2000]
Brain fMRI	$31 \cdot 10^3$	11.4	3.9	2.92	0.14	$4.3 \cdot 10^{-4}$	325.5	111.3	$r_c = 0.6$	[Eguiluz et al. 2005]
Brain fMRI	$17 \cdot 10^3$	12.9	5.3	2.43	0.13	$3.7 \cdot 10^{-4}$	351.3	144.6	$r_c = 0.7$	[Eguiluz et al. 2005]
Brain fMRI	$4.8 \cdot 10^3$	6.0	6.0	1.0	0.13	$8.9 \cdot 10^{-4}$	144.1	144.1	$r_c = 0.8$	[Eguiluz et al. 2005]
Human fMRI	90	2.82	2.58	1.09	0.25	0.12	2.08	1.9		[Salvador et al. 2005]
Human fMRI	90	2.49	2.31	1.08	0.53	0.22	2.38	2.20	$r_c = 0.44$	[Achard et al. 2006]
Cortex	54	3.05	2.65	1.15	0.3	0.13	2.31	2.0		[He et al. 2007]
Ising model	$40 \cdot 10^3$	6.8	2.71	2.51	0.516	0.048	10.75	4.3	$T = 2.3$	[Fraiman et al. 2009]
Activity flow	$50 \cdot 10^3$	2.89	2.71	1.09	0.18	$2.4 \cdot 10^{-3}$	75	68.8		to be published

We are aware, that the major flaw of this section is a lack of a strictly analytical methods for the 'small-world-ness' indicator, thus the numerical simulations will have to make up for them. Nonetheless, as a result we can provide a positive answers to both raised questions, namely in complex artificial neural systems the functional network exhibits both scale-free and small-world phenomena.

Chapter 5

Discussion

5.1 Conclusion

In Chapter 4 we analysed the activity-flow model and showed, that is able to mimic functional patterns typical to biological systems. The model built a layer of abstraction between functional and neural dynamics, where the underlying system consists of fairly simple units which operate in discreet state space. Moreover, we got rid of vast number of connections, putting their number down to less than 2% of the all available in full graph. To compensate a geometrical-driven synapse generation was applied and we only required that the local connectivity is efficient enough. Finally a pretty straightforward asynchronous dynamics with energy-driven set of rules was defined.

The results of this thesis are nowhere near to the precise explanation of the large-scale brain dynamics and origination of the behaviour, they give however a not overcomplicated nor computationally demanding tool to model such complex systems. To summarise, we formulate a short list of conclusions concerning mesoscopic-level neural dynamics.

- The energy driven dynamics leads to emergence of self-organisation. This appearance of the hierarchy occurred without any bias towards any unit and was a result of the system evolution.

- The number of connections required to admit an occurrence of complex self-organization does not have to be high and not necessarily require a nearly full global density. This is in agreement with some researches concerning optimal neighbourhood function in self-organising mappings with Kohonen algorithm, according to which the region size does not necessarily have to cover the whole network, especially at the ending

phases of the learning, the local and sparse neighbourhood function yields a good quality mappings, see [Keith-Magee et al. 1999].

- The short diameter of the resulting functional network provides a good transfer of information. This is vital feature of large-scale graphs such as social or the Internet, as it makes the flow of data efficient.

- Heavy tail of the degree distribution can be an issue, as it suggests that there exist units, which have an extraordinary connectivity and require efficient communication protocols. Moreover due to the spontaneous appearance of these hubs, it might be hard to predict which nodes would become this *elite* and require special maintenance. This would be useful information when implementing such system, as the hubs could be supplied with special efficient hardware to facilitate information processing. On the other hand, elites and hubs do exist in real-world graphs and seem to be a vital parts of them, and their limited capacity seems to play a secondary part.

5.2 Spectral properties

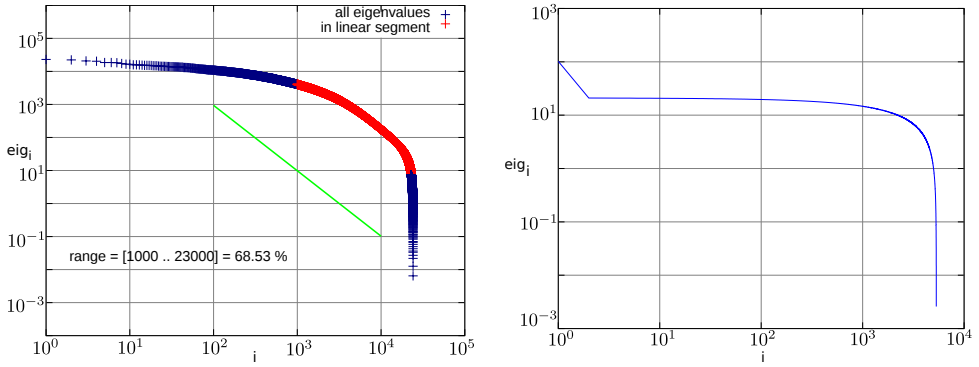
It would seem interesting to take a closer look at some statistics which are bound to classical graph theory such as *graph spectrum*. The spectrum is defined as a set of eigenvalues of the adjacency matrix[1] of the graph: $\{\lambda \in \mathbb{C} : Ad \cdot x = \lambda \cdot x, \text{for some } x \in \mathbb{R}^n\}$, [Chung 2007, Cvetković et al. 1997].

Clearly, the spectrum does not describe a graph uniquely up to the isomorphism, but a set of properties such as connectedness or bi-partitioning can be obtained from it. However, the spectrum can be extended into a multivariable polynomial obtained as a determinant of the adjacency matrix with non-zero entries substituted by distinct variables, see [Krishnamurthy 1981]. This generalisation does define a isomorphism-unique description.

Neural networks are infrequently analysed in terms of spectral graph theory. Schreiber in [Schreiber 2008] analysed spectra of winner-take-all dynamics in neural networks and concluded that the i-th eigenvalue, when sorted decreasingly, vanquish like $c\frac{1}{i^2}$. This prediction was numerically confirmed by Piękniewski in [Piękniewski 2009], who analysed spectra of stochastic dynamics in recurrent neural networks.

The spectral analysis of the activity-flow model is one of the aims of the forthcoming studies, though the first preliminary results are quite encouraging. In [Piersa & Schreiber 2012] we have computed a spectral density of

[1] Other definitions take eigenvalues of the laplacian matrix of the graph $L := I_n - Ad$.



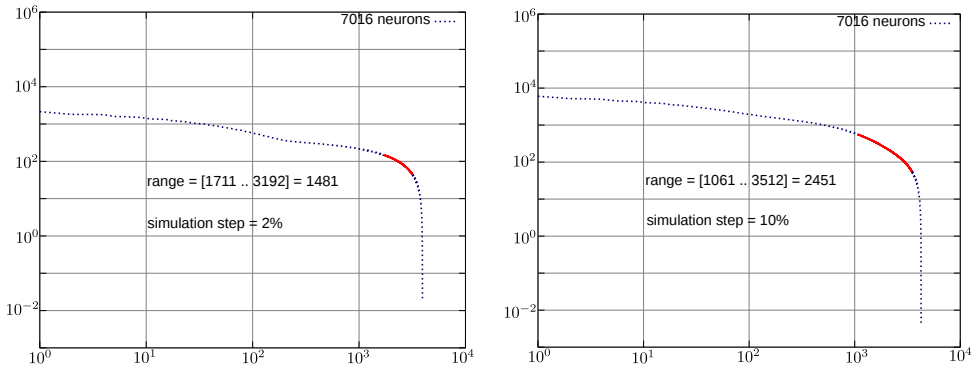
(a) Spectrum of the obtained functional graph. (b) A spectrum of the Erdős-Rényi graph

Figure 5.1: Left: a plot of i vs i-th eigenvalue (sorted decreasingly) of the obtained functional graph. The middle highlighted part of the plot indicates a segment, for whom the scaling is valid. The solid reference line is $\frac{c}{i^2}$. Figure reprinted after [Piersa & Schreiber 2012]. Right: a comparison plot of the spectrum in the ER random graph model.

medium-size sample graph obtained from the model, as described in Section 4.1. The results confirm a scale free dependency with an exponential truncation of the tail. Interesting feature is that, while in [Piękniewski 2009] the dependency arouse for first 10 principal eigenvalues, we have obtained this in the middle part of the log-log plot and it was valid for approximate 40 to 60% of the spectrum, clearly excluding the principal ones, see Figure 5.1. It is hard to tell if the truncation is just an artefact caused by a limited sample size or it is a innate to the geometrical embedding. To some extent the prediction of [Schreiber 2008] still holds although the WTA dynamics can be only approximated with high inverse temperature β (see Section 4.1.4 and Equation 4.3) and the presence of synapses is determined by geometrical embedding rather than taken for granted.

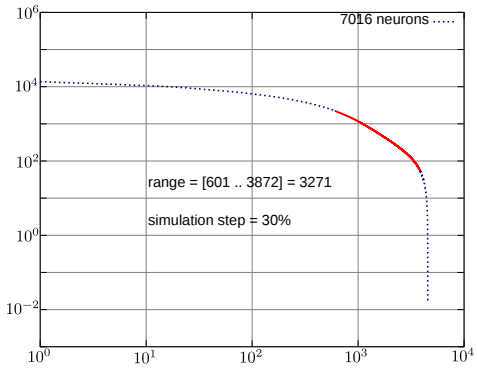
The estimation of the segment, for whom the power law approximation is valid, was computed by obtaining beginning and ending indices of the segment i_1 and i_2, such that the linear regression of the log() of the data between the indices yields a linear dependency and the sum of squared residuals lays beneath a fixed threshold value.

When observing the spectrum throughout the dynamics in Figure 5.2, the scale-free dependency quite clearly appears and strengthens as the evolution continues. Although in early stages of the dynamics the spectrum bears resemblances to those of ER graph, the linear segment (in double-logarithmic

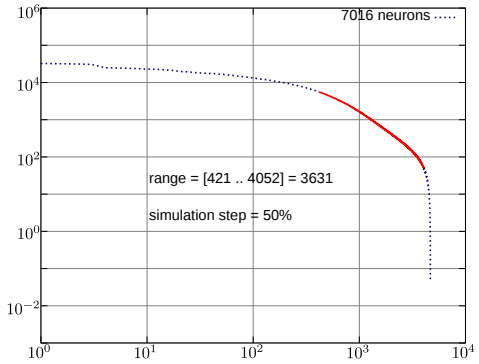


(a) After 2% of the iterations.

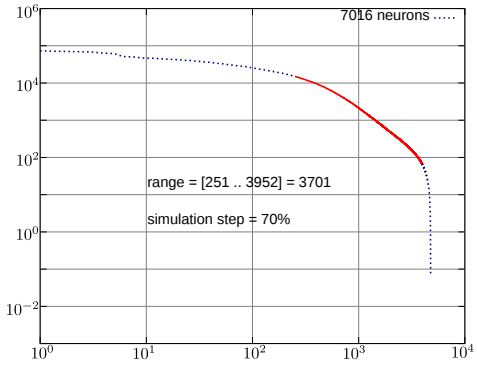
(b) After 10 % of the iterations.



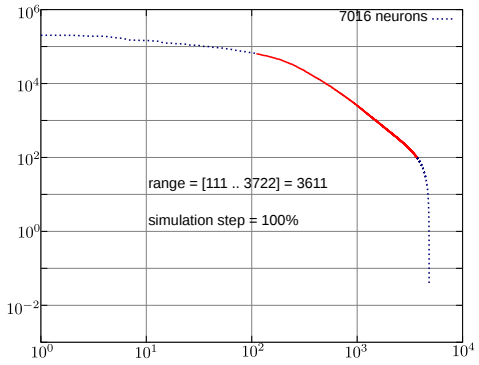
(c) After 30 % of the iterations.



(d) After 50 % of the iterations.



(e) After 70% of the iterations.



(f) After 100 % of the iterations.

Figure 5.2: Plots of the spectrum of the functional graphs obtained along the progress of the dynamics. Dotted lines denote whole spectrum, solid lines — the estimated linear segments ($\frac{1}{i^2}$ in regular plot). Figure reprinted after [Piersa & Schreiber 2012].

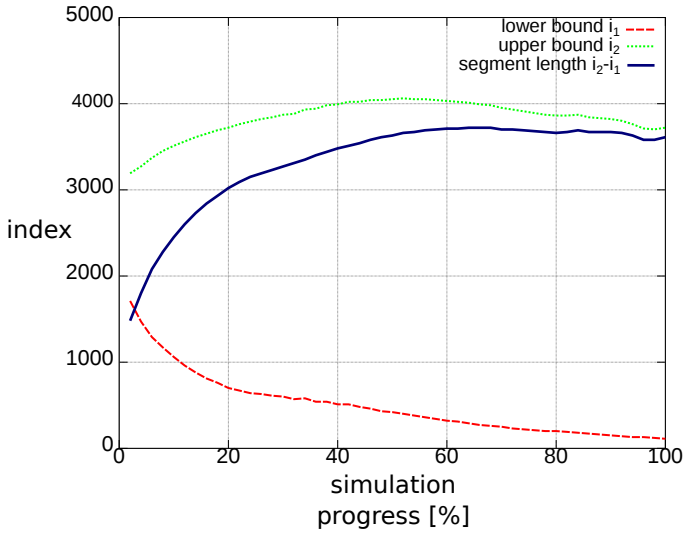


Figure 5.3: A plot of estimated indices i_1, i_2 and length of the segment, for whom the scale-free dependency of the spectrum is valid vs the progress of the network dynamics. The data is measured every 2%.

plot, thus $\frac{c}{i^2}$ in the original plot) arises and, as the dynamics continue, grows in length.

This is even more visible in Figure 5.3, which presents a plot of both indices and the length of the segment during the simulation. The estimation scheme, while a bit straightforward, confirms an emergence of scaling in the set of eigenvalues. The functional graph only at the beginning is a random ER model, than as the dynamics continue, it gradually shifts towards more sophisticated structure. The saturation of the plot at the 60% of the iterations might indicate that our terminating condition is a bit too strong. The fluctuations at the ending phase seem to stem from a competition between the 'elite' units, which have gathered the activity and now 'turn against' each other

An interesting extension of this research would be a comparison to those obtained, from fMRI data. However, to our knowledge correlation-based functional graphs from medical imagining have not been analysed with spectral methods.

5.3 Further research

The discussed properties of the activity flow model are crucial, but by no means complete. In our research we assumed a full invulnerability of the

units to noise and random damages. While this is beyond the scope of the dissertation is would be meaningful to take a closer look at this problem. The human brain is known for its plasticity and ability to reorganize itself after minor-to-moderate damages. One might expect, that a good model should also be resilient to some damage.

Interestingly, large scale real networks such as the Internet or power lines have their 'fail-safe' protocols, which within reasonable limits keep them alive despite some breakdowns. These properties are however dependent on the structure of the network rather than self-organisation. Should too many nodes in the Internet become damaged, the whole system will eventually malfunction, see discussion in [Albert & Barabasi 2002]. Analyses of the resiliency of the obtained network seem to yield a promising results. Especially, that, unlike spectral analyses, some results concerning noise resistance in brain network have been reported, for instance [Achard et al. 2006]. Other results concerning a reorganisation after the damage were discussed in [Grefkes et al. 2011].

Our results concerning spectral properties cannot be considered as complete without comparison to medical counterparts. This concern can be, to some extent, addressed with modern trends in the science, namely open-access availability of research data. Not only are the reports published in freely-accessed repositories more often, but also they are accompanied by the raw data.

With an access to fMRI records, we would be able to provide an answer to the raised problem about spectral properties of brain networks without a need of having our own devices. Such data is, however, striped from any potentially relevant information about the subject and his/her health state and, therefore, should be taken with caution when inferring about functional properties of the *healthy* brain.

Appendix A

Technical challenges

In this Appendix we briefly list technical and numerical aspects of the carried simulation part of this dissertation.

A.1 Technical details

The simulation of the model and its dynamics was implemented in C++ programming language. The data post-processing was carried in GNU Octave[1] numerical environment. The plots were prepared in Gnuplot[2] package with Octave front-end. In addition the software used Linear Algebra Package (LaPack)[3] binding libraries.

The thread-wise parallelisation allowed us to take advantage of high performance computing systems. The largest samples were run on the infrastructure of the PL-Grid Project[4], whom the author is grateful for sharing the computational resources. Some of those challenges were discussed in [Piersa & Schreiber 2010] and [Piersa 2011].

In Figure A.1 we outline modular shape and dependencies of the simulations. Clearly some of them should be easy to 'switch off' whenever they are out of focus of ongoing tests. Due to the modularity of the simulation, we were able to focus on the most time-consuming parts of the computation that is the dynamics itself, generation of the structural network and calculation of the characteristic and maximal path length.

[1] http://www.gnu.org/software/octave/
[2] http://www.gnuplot.info/
[3] http://www.netlib.org/lapack/
[4] http://www.plgrid.pl/

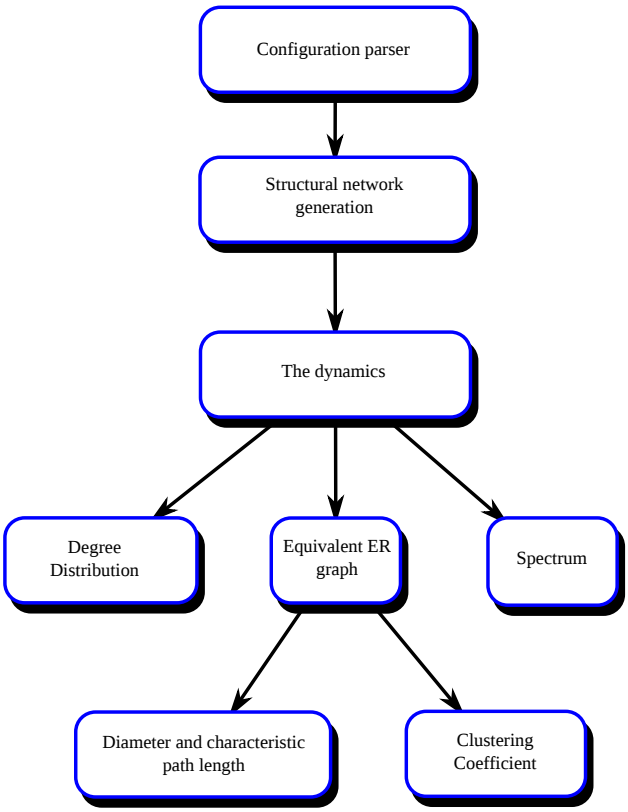


Figure A.1: A schematic diagram of the simulation program.

A.2 Parallelisation

Structural graph generation requires $O(n^2)$ symmetric tests (for each pair $(v_i, v_j), v_i, v_j \in \mathcal{V}, \ |\mathcal{V}| = n$) and can be easily domain-partitioned. Quite similarly calculating a shortest paths in functional graph between every pair of vertices is a handbook Multiple Data Single Instruction stream (MDSI) problem.

The biggest challenge was parallelisation of the dynamics itself, as the activity alterations require synchronised access to the network structure. We decided to split the dynamics among t threads. Since each thread manipulates registers in two nodes, we end up with extended version of the *five dining philosophers problem*[5]. Clearly in this case each thread (a philosopher) can reach towards any pair of connected vertices (forks), rather than just two assigned in advance. Similarly, each vertex (a fork) can be grabbed

[5]The formulation of the problem can be found in work *Two starvation-free solutions of a general exclusion problem* of E. Dijkstra, available at http://www.cs.utexas.edu/users/EWD/ewd06xx/EWD625.PDF

by any thread (philosopher). The semaphore-controlled access turned out to quite efficient for low t, although the theoretical analogy to birthday paradox (suppose t threads randomly lock one of n vertices, what is the probability that a single vertex will be locked by two threads?) clearly indicates, that the number of threads t of order $\sqrt{n}$ can result in high fraction of collisions and waits on mutex.

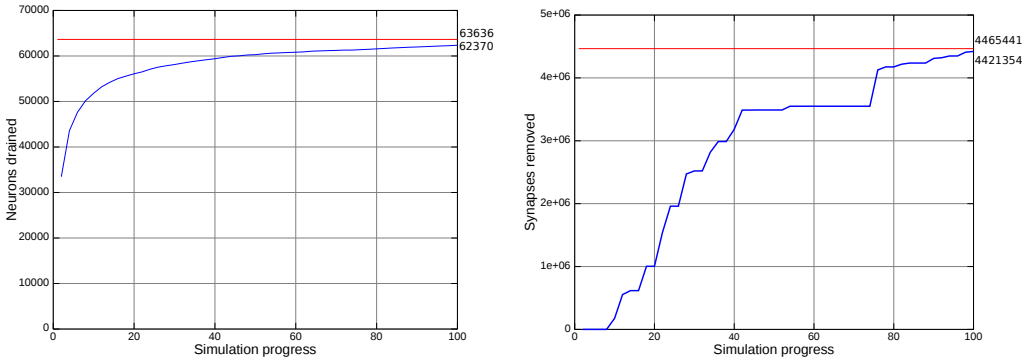
Even more important form of optimization is picking a pair of neurons which are connected with a synapse, such that one of them has positive activity level, see Section 4.1.2. Due to sparsity of the graph the naive *pick a random pair until it satisfies requirements* is highly inefficient. Moreover, as the dynamics continues and the charge becomes concentrated in a small number of nodes (see discussion in Section 4.1.4) even *picking a random edge* frequently becomes in vain, since vast majority of synapses connects a pair of drained units. Indeed, if the number of remaining useful (i.e. at least one of incident nodes has positive charge) edges k is small compared to all available ones (denote their number as l), then l times random sampling over the set gives $(1 - \frac{k}{l})^l \to e^{-k}$ probability of failing to find a useful synapse (in l consecutive samplings). This can become an issue in the ending phases of the simulations, when k is just a few.

Therefore, if the first selected edge is not useful, then following edge is picked as the next one present on pre-defined list, instead of random. This ensures us that we can find a valid synapse in at most l steps, instead of having just chances $1 - e^{-k}$ to find such edge within l steps, where l is a number of edges on the list.

In addition, we decided to facilitate this step by periodically purging the list of void synapses. As it can be seen in Figure A.2, this reduces the pessimistic number of synapses to search l by 50% after a quarter of the iterations. This saving is even larger in later parts. This procedure requires an exclusive access to the list and, thus, it is performed infrequently, every 1-5% of the iterations.

As a last improvement, we require only one of the incident neurons has a positive activation, to make the synapse useful. The direction of the transfer is automatically defined towards drained node. Opposite choice would have been discarded anyway, so this only saves iterations needed to find a valid synapse. If both have some activity left, than the direction is picked at random. It might seem, that it artificially increases probability of using a synapse with one neuron drained. However, as we discussed in Section 4.1.4, as the dynamics tend to drain the activity from neurons and the number of affected synapses (i.e. incident to two positive activation neurons) quite rapidly drops, also see Figure A.2. Taken together, these improvements al-

lowed us to dramatically enhance speed of the simulation and, by extension, increase the sample size.



(a) Number of drained neurons

(b) Number of edges, removed from the sampling set

Figure A.2: Number of drained neurons and edges removed from the sampling list throughout the simulation, measured every 2%. Horizontal reference line is the total number of neurons (left) and synapses (right).

A trade-off of this is some periodical maintenance of the sampling list of edges. On the other hand, any sort of statistic tracking also requires suspension of the computation due to exclusive access to the data structures. Nonetheless, the results turned out to be satisfying, as the parallelisation allowed to achieve a speed-up of the simulation with an efficiency between 0.66 to 0.8, see [Piersa & Schreiber 2010].

Acknowledges

The author would like to mention an unparalleled help, inspiration and motivation from prof. Tomasz Schreiber (1975–2010), a brilliant mathematician, physicist and computer scientist, professor at the Faculty of Mathematics and Computer Science, Nicolaus Coppernicus University. It was His passion which had the greatest influence on direction of scientific interests of the author and, by extension, the subject of this work.

The author acknowledges the help of supervisors — prof. Piotr Bała and dr Filip Piękniewski, who managed to point the way to the lucky end.

The author would like to thank to Michał Meina, Krzysztof Rykaczewski, Piotr Przymus, Michał Gawarkiewicz and Anna Gogolińska for brief hints and hours of discussions concerning the work; Marcin Gąsiarek and Katarzyna Zając (and their carpet) for providing an outstanding motivation and reason to start writing of this thesis.

Special thanks and warming words for Maja Czoków and Michał Matuszak, who carry their own parts of prof. Schreiber's scientific legacy.

The author acknowledges the support of PL-Grid project, for an access to the computing resources for simulations.

The work has been partially supported by the European Social Fund and Government of Poland as a part of Integrated Operational Program for Regional Development, ZPORR, Activity 2.6 „Regional Innovation Strategies and Knowledge Transfer" project „Scholarships for PhD Candidates" of Kuyavian-Pomeranian Voivodeship in Poland (2008 –2009) and by the Polish Ministry of Science and Higher Education and National Science Centre, research grant DEC-2011/01/N/ST6/01931 (2011 – 2013).

List of Figures

List of Tables

Index

Bibliography

[Achard et al. 2006] S. Achard, R. Salvador, B. Whitcher, J. Suckling, E. Bullmore, *A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs*, The Journal of Neuroscience, no 26(1) 2006, pp.63-72, January 2006, doi:10.1523/JNEUROSCI.3874-05.2006 68, 70, 78

[Ackley et al. 1985] D. H. Ackley, G. E. Hinton, T. J. Sejnowski, *A Learning Algorithm for Boltzmann Machines*, Cognitive Science, 9(1):147-169, 1985, doi:10.1016/S0364-0213(85)80012-4. 26, 36

[Albert et al. 1999] R. Albert, H. Jeong, A. L. Barabasi, *Diameter of the World-Wide Web*, Nature, Vol 401, 9 September 1999, doi:10.1038/43601. 15, 44

[Albert & Barabasi 2002] R. Albert, A. L. Barabasi, *Statistical mechanics of complex networks*, Reviews of modern physics, Vol 74, January 2002, doi:10.1103/RevModPhys.74.47. 10, 23, 25, 44, 78

[Bak et al. 1987] P. Bak, C. Tang, K. Wissenfeld, *Self-Organised Criticality: An Explanation of* $1/f$ *Noise*, Physical Review Letters, vol 59, no 4, July 1987, doi:10.1103/PhysRevLett.59.381. 25, 26

[Barabasi et al. 1999] A. L. Barabasi, R. Albert, *Emergence of Scaling in Random Networks*, Science, vol 286, 15 October 1999, doi:10.1126/science.286.5439.509. 20

[Barabasi et al. 2002] A. L. Barabaci, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, T. Vicsek, *Evolution of the social network of scientific collaborations*, Physica A 311, (3-4) (2002), pp. 590-614, doi:10.1016/S0378-4371(02)00736-7. 15, 17, 44, 61

[Basset & Bullmore 2006] D. S. Bassett, E. Bullmore, *Small-World Brain Networks*, The Neuroscientist, Volume 12, Number 6, 2006, doi:10.1177/1073858406293182. 16, 50, 61

[Bishop 1995] Ch. Bishop, *Neural networks in pattern recognition*, Clarendon Press, 1995. 31

[Bullmore & Sporns 2009] E. Bullmore, O. Sporns, *Complex brain networks: graph theoretical analysis of structural and functional systems*, Nature Reviews, Neuroscience, vol 10, March 2009, doi:10.1038/nrn2575. 50

[Chialvo 2004] D. R. Chialvo, *Critical brain networks*, Physica A: Statistical Mechanics and its Applications, Vol. 340, Issue 4, September 2004, doi:10.1016/j.physa.2004.05.064. 25, 26

[Chung & Lu 2001] F. Chung, L. Lu, *The Diameter of Sparse Random Graphs*, Advances in Applied Mathematics 26, 257–279 (2001), doi:10.1006/aama.2001.0720 19

[Chung & Lu 2006] F. Chung, L. Lu, *Complex graphs and networks*, American Mathematical Society, 2006. 13, 18, 19, 21, 26, 61

[Chung 2007] F. Chung, *Spectral graph theory*, American Mathematical Society, 2007. 74

[Csemely 2009] P. Csermely, *Weak links: the universal key to the stability of networks and complex systems*, Springer Verlag (Heidelberg, Germany) 2009. 11, 15

[Cvetković et al. 1997] D. Cvetković, P. Rowlingson, S. Simić, *Eigenspaces of graphs*, Cambridge University Press, 1997. 74

[Deistel 2006] R. Diestel, *Graph theory*, Springer, Berlin, 2006. 13

[Eguiluz et al. 2005] V. Eguiluz, D. Chialvo, G. Cecchi, M. Baliki, V. Apkarian, *Scale-free brain functional networks*, Physical Review Letters, PRL 94 018102, January 2005, doi:10.1103/PhysRevLett.94.018102. 11, 26, 44, 48, 52, 56, 58, 61, 62, 68, 69, 70

[Erdős & Rényi 1959] P. Erdős, A. Rényi, *On random graphs I*, Publ. Math. Debrecen 6, 290–297, 1959. 11, 13, 18, 21

[FitzHugh 1961] R. FitzHugh, *Impulses and Physiological States in Theoretical Models of Nerve Membrane*, Biophysical Journal, vol 1, no 6, pp. 445–466, July 1961. 41

[Fraiman et al. 2009] D. Fraiman, P. Balenzuela, J. Foss, D. R. Chialvo, *Ising-like dynamics in large-scale functional brain networks*, Physical Review E, Volume 79, Issue 6, June 2009. doi:10.1103/PhysRevE.79.061922 26, 56, 58, 68, 69, 70

[Grefkes et al. 2011] Ch. Grefkes, G. R. Fink, *Reorganization of cerebral networks after stroke: new insights from neuroimaging with connectivity approaches*, BRAIN — A Journal of neurology, March 16, 2011, doi:10.1093/brain/awr033. 78

[Gurevitch 1961] M. Gurevitch, *The social structure of acquaintanceship networks*, phd dissertation, MIT, 1961. 11, 15, 61

[Hansel & Mato 2001] D. Hansel, G. Mato, *Existence and Stability of Persistent States in Large Neuronal Networks*, Physical Review Letters, vol 86, no 18, pp. 4175-4178, April 2001, doi:10.1103/PhysRevLett.86.4175 38

[He et al. 2007] Y. He, Z. J. Chen, A. C. Evans, *Small-World Anatomical Networks in the Human Brain Revealed by Cortical Thickness from MRI*, Cerebral Cortex; 17:2407-2419 October 2007, doi:10.1093/cercor/bhl149. 57, 58, 68, 70

[Hebb 1949] D. Hebb, *The Organisation of Behaviour*, New York: Wiley, 1949. 10, 28, 33

[Herr et al. 2007] B. W. Herr, W. Ke, E. Hardy, K. Börner, *Movies and Actors: Mapping the Internet Movie Database*, Proceedings of 11th Annual Information Visualization International Conference, pp. 465-469, IEEE Computer Society Conference Publishing Services, doi:10.1109/IV.2007.78, April 2007. 15, 61

[Hilbert & Lopez 2011] M. Hilbert, P. Lopez, *The World's Technological Capacity to Store, Communicate, and Compute Information*, Science, Vol. 332 no. 6025 pp. 60-65, April 2011, doi:10.1126/science.1200970 28

[Hodgkin & Huxley, 1952a] A. L. Hodgkin, A. F. Huxley, *Currents Carried By Sodium and Potassium Ions Through the Membrane of the Giant Axon of Loligo*, J. Physiol. 116, pp. 449-472, 1952. 29

[Hodgkin & Huxley, 1952b] A. L. Hodgkin, A. F. Huxley, *A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve*, J. Physiol. 117, pp. 500-544, 1952. 29

[Hopfield 1982] J. J. Hopfield, *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, Proceedings of the National Academy of Sciences of the USA, vol. 79 no. 8 pp. 2554-2558, January 1982. doi:10.1073/pnas.79.8.2554 35, 49

[Humphries et al. 2006] M. D. Humphries, K. Gurney, T. J. Prescott, *The brainstem reticular formationis a small-world, not scale-free, network*, Proceedings of the Royal Society, Biological Sciences, no. 273 (2006), pp. 503–511 doi:10.1098/rspb.2005.3354, November 2005. 61, 62

[Izhikevich 2007] E. Izhikevich, *Dynamical Systems in Neuroscience*, The MIT Press, Cambridge, Massachusetts, London, England, 2007. 10, 26, 29, 30, 39, 42

[Izhikevich 2003] E. Izhikevich, *Simple model of spiking neurons*, IEEE Transactions on Neural Networks, vol. 14, no 6., November 2003. doi:10.1109/TNN.2003.820440 42

[Jeong et al. 2000] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A. L. Barabasi, *The large-scale organization of metabolic networks*, Nature, vol 407, pp. 651-654, October 2000, doi:10.1038/35036627. 44

[Kawasaki 1966] K. Kawasaki, *Diffusion Constants near the Critical Point for Time-Dependent Ising Models I*, Physical Review. 145, May 1966. doi:10.1103/PhysRev.145.224 49

[Keith-Magee et al. 1999] R. Keith-Magee, S. Venkatesh, M. Takatsuka, *An empirical study of neighbourhood decay in Kohonen's self organising map*, International Joint Conference on Neural Networks, Proceedings, Vol 3, Publ. Ieee, pp: 1953-1958, July 1999, doi:10.1109/IJCNN.1999.832682 74

[Kello & Beltz 2009] C. T. Kello, B. C. Beltz, *Scale-Free Networks in Phonological and Orthographic Wordform Lexicons*, I. Chitoran, C. Coupe, E. Marsico, F. Pellegrino (Eds.), *Approaches to Phonological Complexity*, Mouton de Gruyter, Berlin 2009. 11, 44

[Kohonen 1982] T. Kohonen, *Self-Organized Formation of Topologically Correct Feature Maps*, Biological Cybernetics, Vol 47, No 1, pp. 59-69, 1982, doi:10.1007/BF00337288 10

[Krishnamurthy 1981] E. V. Krishnamurthy, *A form invariant multivariable polynomial representation of graphs*, Lecture Notes in Mathematics, Volume 885/1981, pp. 18-32, 1981, doi:10.1007/BFb0092252. 74

[RFC 4271] T. Li, S. Hares, *RFC 4271 — A Border Gateway Protocol 4 (BGP-4)*, available on-line `http://www.ietf.org/rfc/rfc4271.txt` 16

[Lin & Lee 1995] C. T. Lin, C. S. Lee, *A multi-valued Boltzmann machine*, IEEE Transactions on Systems, Man and Cybernetics, Volume 25, Issue 4, pp. 660 - 669, April 1995, doi:10.1109/21.370198. 37

[MacQueen 1967] J. MacQueen, *Some methods for classification and analysis of multivariate observations*, Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 281-297, University of California Press, 1967. 10

[Markov 1971] A. A. Markov, *Extension of the limit theorems of probability theory to a sum of variables connected in a chain*, (reprint from Russian) R. Howard, *Dynamic Probabilistic Systems*, vol. 1, John Wiley and Sons, 1971. 36

[Matuszak & Miękisz 2011] M. Matuszak, J. Miękisz, *Coevolution of Graphs and Strategies*, preprint, January 2011, [available on-line] http://ssdnm.mimuw.edu.pl/student/16 23

[de Moura et al. 2003] A. de Moura, Y.-C. Lai, A. Motter; *Signatures of small-world and scale-free properties in large computer programs*, Physical Review E, vol. 68, 017102, July 2003, doi:10.1103/PhysRevE.68.017102 44

[RFC 2328] J. Moy, *RFC 2328 — OSPF Version 2*, April 1998, available on-line `http://www.ietf.org/rfc/rfc1247.txt`. 16

[Newell & Montroll 1953] G. F. Newell, E. W. Montroll, *On the Theory of the Ising Model of Ferromagnetism*, Reviews of Modern Physics vol. 25, issue 2, pp. 353–389, April 1953. doi:10.1103/RevModPhys.25.353 36

[Newman et al. 2001] M. Newman, S. Strogatz, D. Watts; *Random graphs with arbitrary degree distributions and their applications*, Physical Review E, vol 64, 026118, July 2001, doi:10.1103/PhysRevE.64.026118 17

[Nishimoto et al. 2011] S. Nishimoto, A. Vu, T. Naselaris, Y. Benjamini, B. Yu, J. Gallant, *Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies*, Current Biology, Vol 21, No 19, pp. 1641–1646, October 2011, doi:10.1016/j.cub.2011.08.031 28

[Oja 1982] E. Oja, *A Simplified Neuron Model as a Principal Component Analyser*, Journal of Mathematical Biology, Vol 15, pp. 267-273, November 1982, doi:10.1007/BF00275687 33

[Pasley et al. 2012] B. Pasley, S. David, N. Mesgarani, A. Flinker, S. Shamma, N. Crone, R. Knight, E. Chang, *Reconstructing Speech from Human Auditory Cortex*, Public Library of Science Biology, Vol. 10, No. 1, e1001251, January 2012, doi:10.1371/journal.pbio.1001251 28

[Peretto 1994] P. Peretto, *Introduction to Modeling Neural Networks*, Cambridge University Press 1994. 32, 34

[Piękniewski & Schreiber 2007] F. Piękniewski, T. Schreiber, *Emergence of scale-free spike flow graphs in recurrent neural networks*, Proc. IEEE symposium series in computational intelligence — foundations of computational intelligence. pp. 357–362, April 2007, doi:10.1109/FOCI.2007.371496 44

[Piękniewski 2008] F. Piękniewski, *Spontaneous scale-free structures in spike flow graphs for recurrent neural networks*, Ph.D. dissertation, Warsaw University, Warsaw, Poland, 2008. 10, 44, 45, 51, 56, 58, 62

[Piękniewski & Schreiber 2008] F. Piekniewski, T. Schreiber, *Spontaneous scale-free structure of spike flow graphs in recurrent neural networks*, Neural Networks vol. 21, no. 10, pp. 1530 – 1536, December 2008, doi:10.1016/j.neunet.2008.06.011. 44, 45

[Piękniewski 2009] F. Piękniewski, *Spectra of the spike flow graphs of recurrent neural networks*, Artificial Neural Networks — ICANN 2009, 19th International Conference, 2009, Proceedings, Part II, doi:10.1007/978-3-642-04277-5_61. 74, 75

[Piersa et al. 2010] J. Piersa, F. Piękniewski, T. Schreiber, *Theoretical model for mesoscopic-level scale-free self-organization of functional brain networks*, IEEE Transactions on Neural Networks, vol. 21, no. 11, November 2010. doi:10.1109/TNN.2010.2066989 47, 48, 53, 57, 58

[Piersa & Schreiber 2010] J. Piersa, T. Schreiber, *Scale-free degree distribution in information-flow graphs of geometrical neural networks. Simulations in concurrent environment* Postproceedings of Mathematical Methods in Modelling and Analysis of Concurrent Systems (MASYW 2010), pp. 143–151, IPI PAN, Warsaw, July 2011. 61, 79, 82

[Piersa 2011] J. Piersa, *Diameter of the spike-flow graphs of geometrical neural networks*, accepted for postproceedings 9th International Conference on Parallel Processing and Applied Mathematics, Lecture Notes on Computer Science, September 2011. 47, 61, 63, 79

[Piersa & Schreiber 2012] J. Piersa, T. Schreiber, *Spectra of the Spike-Flow Graphs in Geometrically Embedded Neural Networks*, accepted for Proceedings of 11th International Conference on Artificial Intelligence and Soft Computing, ICAISC 2012, Lecture Notes in Computer Science, vol. 7267/2012, pp. 143–151, May 2012, doi:10.1007/978-3-642-29347-4_17 74, 75, 76

[Rojas 1996] R. Rojas, *Neural Networks, A Systematic Introduction*, Springer 1996. 10, 32

[Rumelhart et al. 1986] D. E. Rumelhart, G. E. Hinton, R. J. Williams; *Learning internal representations by error propagation*, in *Parallel distributed processing: explorations in the microstructure of cognition*, vol. 1, MIT Press Cambridge, MA, USA, 1986. 34

[Rosenblatt 1958] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, Psychological Review, Vol 65(6), pp 386-408, November 1958, doi:10.1037/h0042519 32

[Russel & Norvig 1995] S. J. Russell, P. Norvig, *Artificial Intelligence, A modern Approach*, Prentice Hall, Englewood Cliffs, New Jersey, 1995. 16, 27, 28

[Salvador et al. 2005] R. Salvador, J. Suckling, M. Coleman, J. Pickard, D. Menon, E. Bullmore, *Neurophysiological Architecture of Functional Magnetic Resonance Images of Human Brain*, Cerebral Cortex, September 2005;15, pp. 1332–1342, September 2005, doi:10.1093/cercor/bhi016 52, 68, 70

[Schreiber 2008] T. Schreiber, *Spectra of winner-take-all stochastic neural networks*, 3193(0810), pp. 1–21. arXiv [available on-line], October 2008, http://arxiv.org/abs/0810.3193 74, 75

[Stephan et al. 2000] K. E. Stephan, C. Hilgetag, G. Burns, M. O'Neill, M. Young, R. Kötter, *Computational analysis of functional connectivity between areas of primate cerebral cortex*, Philosofical Transactions of Royal Society B, London, vol. 355 no. 1393, pp. 111–126, January 2000, doi:10.1098/rstb.2000.0552 70

[Sysło et al. 2006] M. M. Sysło, N. Deo, J. Kowalik, *Discrete Optimization Algorithms With Pascal Programs*, Dover Publications, Inc. Mineola, New York, 2006. 13, 16

[Ugander et al. 2011] J. Ugander, B. Karrer, L. Backstrom, C. Marlow, *The Anatomy of the Facebook Social Graph*, November 2011, available online: http://arxiv.org/abs/1111.4503v1. 15, 17, 44, 61

[Wagner & Fell 2011] A. Wagner, D. A. Fell, *The small world inside large metabolic networks*, The Royal Society: Biological Sciences, vol. 268 no. 1478, pp. 1803–1810, September 2001, doi:10.1098/rspb.2001.1711. 44

[Watts & Strogatz 1998] D. Watts, S. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, vol 393, pp. 440–442, 4 June 1998, doi:10.1038/30918. 11, 21, 61

[Watts 2004] D. J. Watts, *Small worlds: the dynamics of networks between order and randomness*, Princeton University Press, New Jersey, 2004. 21

[Werbos 1974] P. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciences*, PhD dissertation, Harvard University, November 1974. 32, 34

[Werbos 1990] P. Werbos, *Back-propagation Through Time: What It Does and How to Do It*, Proceedings of The IEEE, vol. 78, no 10, pp. 1550–1560, October 1990, doi:10.1109/5.58337. 35