Google matrix analysis of C.elegans neural network

V. Kandiah, D.L. Shepelyansky *

Laboratoire de Physique Théorique du CNRS, IRSAMC, Université de Toulouse, UPS, F-31062 Toulouse, France

1. Introduction

The human brain neural network has an enormous complexity containing about $10^{11}$ neurons and $10^{14}$ synapses linking various neurons [1]. Such a complex network can only be compared with the World Wide Web (WWW) which indexed size is estimated to be of about $10^{10}$ pages [2]. This comparison gives an idea that the methods of computer science, developed for WWW analysis, can be suitable for the investigations of neural networks. Among these methods the PageRank algorithm of the Google matrix of WWW [3] clearly demonstrated its efficiency being at the heart of Google search engine [4]. Thus we can expect that the Google matrix analysis can find useful applications for the neural networks. This approach has been tested in [5] on a reduced brain model of mammalian thalamocortical systems studied in [6]. However, it is more interesting to perform the Google matrix analysis for real neural networks. In this Letter we apply this analysis to characterize the properties of neural network of C.elegans (worm). The full connectivity of this directed network is known and documented at [7]. The number of linked neurons (nodes) is $N = 279$ with the number of synaptic connections and gap junctions (links) between them being $N_C = 2990$. This network is significantly smaller compared to the one studied in [5] but now we are working not with a model network but with the real worm network. Also, we use several new rank-based methods of network analysis comparing to those used in [5].

Recently, there is a growing interest to the complex network approach for investigation of brain neural networks [8–12]. Generally these networks are directional but it is difficult to determine directionality of links by physical and physiological measurements. Thus, at present, the worm network is practically the only example of neural network where the directionality of all links is established [7]. The analysis of certain properties this directed network has been reported recently in [11,12], however, the approach based on the Google matrix has not been used yet.

In the last years there is a clear trend to apply various advanced methods of network science to understand in a deeper way the connectivity properties of brain. Thus the properties of network centrality were used to characterize the human brain functional graphs [13]. A study of the whole connectivity matrix of the mouse brain has been reported recently [14]. Thus we think that our study will allow to highlight the features of worm network using recent advancements of computer science and push forward such methods for investigation of more complex brain networks.

2. Google matrix construction

The Google matrix $G$ of C.elegans is constructed using the connectivity matrix elements $S_{ij} = S_{syn,ij} + S_{gap,ij}$, where $S_{syn}$ is an asymmetric matrix of synaptic links whose elements are 1 if neuron $j$ connects to neuron $i$ through a chemical synaptic connection and 0 otherwise. The matrix part $S_{gap}$ is a symmetric matrix describing gap junctions between pairs of cells, $S_{gap,ij} = S_{gap,ji} = 1$ if neurons $i$ and $j$ are connected through a gap junction and 0 otherwise. Following the standard rule [3,4], the matrix elements $S_{ij}$ are renormalized ($S_{ij} = S_{ij} / \sum S_{ij}$) for each column with non-zero elements; the columns with all zero elements are replaced by...
The eigenvector at $\lambda = 1$ is known as the PageRank vector. According to the Perron–Frobenius theorem it is positive and its sum is normalized to unity. Thus $P(j)$ gives a probability to find a random surfer on a node $j$. All nodes can be ordered in a decreasing order of probability $P(K_j)$ with highest probability at top values of PageRank index $K_j = 1, 2, \ldots$. For large matrices $P(j)$ can be found numerically by the iteration method [4] but for C.elegans case it can be obtained by a direct matrix diagonalization. We note that it is well established and verified for various complex networks that the PageRank distribution is stable in respect to variation of damping factor $\alpha$ in a range $0.5 \leq \alpha < 0.95$ [4]. We also checked that it is the case for our network and thus we used the usual value $\alpha = 0.85$.

It is also useful to consider the Google obtained from the network with inverted directions of links (see e.g. [16–18]). The matrix $G^*$ for this network with inverted direction of links is constructed following the same definition (1). The PageRank vector of this matrix $G^*$ is called the CheiRank vector with probability $P^*(K_j)$ and CheiRank index $K^*$. According to the known results [3,4] the top nodes of PageRank are the most popular pages, while the top nodes of CheiRank are the most communicative nodes [17,18].

The structure of the matrix elements of $G$, presented in the PageRank ordering of nodes, and $G^*$, presented in the CheiRank ordering of nodes, is shown in Fig. 1. The number of nonzero elements $N_C$ with indexes less than $K$ is determined for various values of $K = 10, 100$. We find the values of ratio $N_C/K$ is approximately $1.2$ at $K = 10, 100$. These values correspond approximately to those of WWW networks of Universities of Cambridge and Oxford being

\begin{equation}
\sum_j G_{ij} \psi_j(j) = \lambda \psi_i(j).
\end{equation}

The eigenspectrum $\lambda_i$ and right eigenvectors $\psi_i(j)$ of $G$ satisfy the equation

\begin{equation}
\sum_j G_{ij} \psi_j(j) = \lambda \psi_i(j).
\end{equation}

The eigenspectrum of eigenvectors $\psi_i(j)$ of $G$ and $G^*$ as shown in top panel of Fig. 2. The imaginary part of eigenvalues is distributed in a range $-0.2 < \text{Im} \lambda < 0.2$ which is more narrow than for the networks of Wikipedia and UK universities [19]. This is related to a significant number of symmetric links. On the other side the networks of Le Monde or Python have comparable width for Im $\lambda$. We find that the second by modulus eigenvalues are $\lambda_2 = 0.8214$ for $G$ and $\lambda_2 = 0.8608$ for $G^*$. Thus the network relaxation time $\tau = 1/|\ln \lambda_2|$ is approximately $5, 6.7$ iterations of $G$, $G^*$.

The properties of eigenstates $\psi_i$ can be characterized by the Inverse Participation Ratio (IPR) $\xi_i = (\sum_j |\psi_i(j)|^2)^2/\sum_j |\psi_i(j)|^4$, which is broadly used in analysis of electron conductivity in disordered systems (see e.g. [19, 20]). This quantity effectively determines the number of nodes on which is located an eigenstate $\psi_i$. We see that some eigenstates have rather large $\xi$ on about ten nodes. We will return to the discussion of properties of eigenstates later.

Fig. 1. (Color on-line.) Google matrix $G$ (left) and $G^*$ (right) for the neural network of C.elegans for $N = 279$ connected neurons. Matrix elements $G_{kk'}$ are shown in the basis of PageRank index $K$ (and $K'$) and elements $G_{s_{syn}, ij}$ are shown in the basis of CheiRank index $K^*$ (and $K^{*'}$) at $\alpha = 0.85$. Here, $x$ and $y$ axes show $1 \leq K, K' \leq N$ and $1 \leq K^*, K^{*'} \leq N$; the elements $G_{11}, G^*_{11}$ are placed at the top left corner; color is proportional to the square root of matrix elements which are changing from black at minimum value $(1-\alpha)/N$ to light yellow at maximum.

Fig. 2. (Color on-line.) Top panel: spectrum of eigenvalues $\lambda$ for the Google matrices $G$ and $G^*$ at $\alpha = 0.85$ (black and red symbols). Bottom panel: IPR $\xi$ of eigenvectors as a function of corresponding Re $\lambda$ (same colors).
The dependence of probabilities of PageRank and CheiRank vectors on their indexes $K$ and $K^*$ is shown in Fig. 3. A formal fit for power law dependence $P \propto 1/K^\nu$, $P^* \propto 1/K^{\nu^*}$ in the range $1 \leq K, K^* \leq 200$ gives $\nu = 0.33 \pm 0.03$ for PageRank and $\nu^* = 0.50 \pm 0.03$ for CheiRank. Of course, the number of nodes is small compared to the WWW or Wikipedia networks but on average we can say that a power law provides a satisfactory description of data. We note that the values of $\nu$ are notably smaller than the usual exponent value $\nu \approx 0.9$ (in $K$), $0.6$ (in $K^*$) found for the WWW or Wikipedia networks (see e.g. [14,17]). Also, in our neural network we find that the exponent in $K$ is smaller then in $K^*$ while usually one finds the opposite situation. At the same time due to a small size of the network we do not claim that the exact value of $\nu$ is so important. It is better to say that its values give an indication of tendency. We think that for large size brain network this exponent can be determined with a better precision.

We also find that $P^* \approx 85$ for $P$ and $\xi \approx 23$ for $P^*$ so that PageRank is distributed over a larger number of neurons. It is possible that such an inversion is related to a significant importance of outgoing links in neural systems: in a sense such links transfer orders, while ingoing links bring instructions to a given neuron from other neurons. We note that somewhat similar situation appears for networks of Business Process Management (BMP) where Principals of a company are located at the top CheiRank position while the top PageRank positions belong to company Contacts [21].

We note that our network is a directional network and as a result we have a significant asymmetry between ingoing and outgoing links. As a result the ranking nodes of PageRank and CheiRank have different probabilities and thus the top nodes have different functions. This fact is well known for directed networks (see e.g. [14,18,20,21]).

The correlations between PageRank and CheiRank vectors is convenient to characterize by the correlator $\kappa = N \sum_i P(i)P^*(i) - 1 = 0.125$. For C. elegans network the value of correlator is relatively small compared to those found for Wikipedia ($\kappa \approx 4$) and WWW of UK universities ($\kappa \sim 3$) [18]. In a sense for C. elegans neural network the situation if more similar to the networks of Linux Kernel ($\kappa \approx 0$) [16] and BMP ($\kappa = 0.164$) [21]. Thus, the C. elegans network has practically no correlations between ingoing and outgoing links. It is argued in [16,18] that such a network structure allows to perform a control of information flow in a more efficient way. Namely, it allows to reduce the propagation of errors in software codes. It seems that the neural networks also adopt such a structure.

Each neuron $i$ belongs to two ranks $K_i$ and $K^*_i$ and it is convenient to represent the distribution of neurons on the two-dimensional plane (2D) of PageRank–CheiRank indexes $(K, K^*)$ shown in Fig. 4. The plot confirms that there are little correlations between both ranks since the points are scattered over the whole plane. Neurons ranked at top $K$ positions of PageRank have their soma located mainly in both extremities of the worm (head and tail) showing that neurons in those regions have important connections coming from many other neurons which control head and tail movements. This tendency is even more visible for neurons at top $K^*$ positions of CheiRank but with a preference for head and middle regions. In general, neurons, that have their soma in the middle region of the worm, are quite highly ranked in CheiRank but not in PageRank. The neurons located at the head region have top positions in CheiRank and also PageRank, while the middle region has some top CheiRank indexes but rather large indexes of PageRank (Fig. 4 left panel). The neuron type coloration (Fig. 4 right panel) also reveals that sensory neurons are at top PageRank positions but at rather large CheiRank indexes, whereas in general motor neurons are in the opposite situation.

The top 20 neurons of PageRank and CheiRank vectors are given in the first two columns of Table 1. We note that both rankings favor important signal relaying neurons such as AVA and AVB that integrate signals from crucial nodes and in turn pilot other crucial nodes. Neurons AVA, AVB, AVBR, AVBL and AVEL, AVER are considered to belong to the rich club analyzed in [12]. The right panel...
in Fig. 3 and second two columns of Table 1 represent ImpactRank which is discussed below.

We can also use 2DRank index $K_2$ discussed in [17], which counts nodes in order of their appearance on ribs of squares in $(K, K^*)$ plane with the square size growing from $K = 1$ to $K = N$. The top neurons in $K_2$ are AVAR, AVBR, AVBL, PVB. Thus at the top $K_2$ values we find dominance of interneurons. More detailed listings are available at [15].

It may be also useful to consider renormalized equal opportunity rank recently discussed in [22]. In this approach PageRank probability of node $i$ is replaced by $P(i)/d(i)$ where $d(i)$ is indegree of node $i$. For the Google matrix this recipe should be replaced by $P(i) \rightarrow P(i)/\sum_j G_{ij}$ and respectively for CheiRank by $P^*(i) \rightarrow P^*(i)/\sum_j G^*_{ij}$. The corresponding rank indexes $K, K^*$ rank the neurons in the decreasing order of these renormalized probabilities. The distribution of nodes in the plane $(K, K^*)$ is shown in Fig. 5. In this ranking the top $K$ nodes correspond to important sensory neurons rather than information relaying centers, whereas the top nodes of $K^*$ are composed mainly by motor neurons. Thus such an approach allows to highlight additional features of Celegans network being complementary to PageRank and CheiRank properties discussed above. Tables for neuron renormalized ranking are available at [15].

### Table 2
Top ten neurons of the eigenvectors of $G$ (left panel) and $G^*$ (right panel) corresponding to the 10th largest eigenvalues $|\lambda|$; IPR are respectively $\xi$ and $\xi^* = 4$.

| $\lambda_{10} = -0.49446$ | $|\psi_i|$ |
|-----------------|-------|
| 1 | AIAR 0.11986 |
| 2 | AIAL 0.11159 |
| 3 | ASIL 0.096475 |
| 4 | ASIR 0.096236 |
| 5 | AWAR 0.024228 |
| 6 | ASHR 0.022241 |
| 7 | RMGR 0.018502 |
| 8 | AIMR 0.018387 |
| 9 | ADLL 0.018377 |
| 10 | PVQL 0.017547 |

| $\lambda_{10} = -0.45784$ | $|\psi_i^*|$ |
|-----------------|-------|
| 1 | AVAR 0.10651 |
| 2 | AVAL 0.079403 |
| 3 | AVBR 0.036779 |
| 4 | AVIL 0.025866 |
| 5 | VA09 0.02438 |
| 6 | VA06 0.020977 |
| 7 | VA08 0.020242 |
| 8 | AVBL 0.019225 |
| 9 | DO02 0.018684 |
| 10 | PD2 0.016485 |

### 4. ImpactRank

In certain cases it is useful to determine an influence or impact of a given neuron on other neurons. A recent proposal of ImpactRank, described in [20], is based on the probability distribution of a vector $v_f = (1 - \gamma)(1 - \gamma G)^{-1}v_0$, $v_f = (1 - \gamma)(1 - \gamma G^*)^{-1}v_0$, where $v_0$ is initially populated neuron. The vector $v_f$ can be viewed as a Green function propagator. The computation of $v_f$ is obtained numerically by a summation of geometrical expansion series which are convergent within approximately first 200 terms at $\gamma \sim 0.7$ (see also [20]). The distributions of probabilities of ImpactRank $P(v(i)) = P(i)$, $P^*(i) = v_f^*(i)$ versus the corresponding ImpactRank indexes $K, K^*$ are shown in Fig. 3 (right panel) for the initial state neuron $RMGL$. The corresponding top 20 ImpactRank neurons influenced by $RMGL$ are given in columns $IMPR, IMCR$ of Table 1. The analysis of neurons linked to $RMGL$ shows that indeed, ImpactRank correctly selects neurons influenced by $RMGL$. The neurons in the top list of $P(i)$ are those pointed by outgoing links of $RMGL$ while those in the top list of $P^*(i)$ are those that have ingoing links to $RMGL$. Such a method can be easily applied to other initial neuron states of interest showing a contamination propagation over the neural network starting from initial neuron $RMGL$.

### 5. Properties of eigenstates

The Google matrix analysis of the Wikipedia hyperlink network [19] demonstrated that the eigenstates with large values of $|\lambda|$ select well defined communities. Thus we can expect that other eigenstates of matrices $G$ and $G^*$ with $|\lambda| < 1$ correspond to certain physiological functions of worm neural network. It is convenient to order index of eigenstates in a decreasing order of $|\lambda|$ with $\lambda_1 = 1$.

The top ten neurons in eigenfunction amplitude for four specific eigenstates of $G$ and $G^*$ are given in Table 2, Table 3. In Table 2 we have eigenstates with low value of IPR so that the corresponding wavefunctions are localized essentially on only about 4 neurons being AIAR, AIAL, ASIR and AVAL, AVAR, AVBR for $\lambda_{10}$ of $G$.
and \(G^*\) respectively. In Table 3 the values of IPR are rather large and these eigenstates are spread over many neurons.

To determine if some eigenvectors are localized on a certain group of neurons, we plot in Fig. 6 the amplitude of each eigenstate horizontally in the basis of neurons ordered by indexes of \(K\) and \(K^*\) of PageRank and CheiRank vectors. The eigenstates of \(G\) matrix show four distinct vertical stripes at \(K = 149\), \(K = 165\), \(K = 185\), \(K = 261\) which correspond respectively to neurons \(PVDR\), \(IL2DR\), \(IL2DL\), \(PLNR\). The same plot for \(G^*\) matrix shows a larger number of stripes which have less pronounced amplitudes. These stripes of \(G^*\) are located on the following neurons \(K^* = 116\) (\(RIPR\)), \(K^* = 123\) (\(RIPR\)), \(K^* = 120\) (\(AS07\)), \(K^* = 122\) (\(AS10\)), \(K^* = 135\) (\(DB06\)), \(K^* = 137\) (\(DB05\)), \(K^* = 215\) (\(DA07\)), \(K^* = 162\) (\(VA1O\)), \(K^* = 172\) (\(SIADL\)), \(K^* = 181\) (\(SIAVL\)), \(K^* = 199\) (\(SIAVR\)), \(K^* = 221\) (\(SIADR\)).

We think that an identification of eigenstates with certain physiological functions of worm can be an interesting task which however requires further more detailed studies in collaboration with physiologists. The tables of top 20 nodes of eigenstates with 50 largest \(|\hat{\lambda}_i|\) values are available at [15].

6. Discussion

In this Letter, we analyzed the neural network of C. elegans using Google matrix approach to directed networks which proved its efficiency for the WWW. We classify worm neurons using PageRank and CheiRank probabilities corresponding to the principal vectors of the Google matrix with direct and inverted links. Thus neurons in the head region take top positions in PageRank, CheiRank and combined 2DRank. Also, interneurons occupy top ranking positions. We show that influences and interdependency between neurons can be studied using the ImpactRank propagator. We argue that the eigenvectors with large modulus of eigenvalues of the Google matrix may select specific physiological functions. This conjecture still need to be investigated in more detailed studies. Our research shows that the Google matrix analysis represents a useful and powerful method of neural network analysis.

Acknowledgements

We thank Emma K. Towlson and Petra E. Vértes for useful discussions and for providing us the links between neurons available from C. elegans neural network data set at [7]. This work is supported in part by EC FET Open project “New tools and algorithms for directed network analysis” (NADINE No. 288956).

References