Google matrix analysis of C.elegans neural network

Vivek Kandiah\textsuperscript{1}, Dima L. Shepelyansky\textsuperscript{1}

\textsuperscript{1} Laboratoire de Physique Théorique du CNRS (IRSAMC), Université de Toulouse, UPS, F-31062 Toulouse, France

\textbf{PACS} 84.35.+i – Neural networks
\textbf{PACS} 89.75.Hc – Networks and genealogical trees
\textbf{PACS} 89.20.Hc – World Wide Web, Internet

\textbf{Abstract} – We study the structural properties of the neural network of the C.elegans (worm) from a directed graph point of view. The Google matrix analysis is used to characterize the neuron connectivity structure and node classifications are discussed and compared with physiological properties of the cells. Our results are obtained by a proper definition of neural directed network and subsequent eigenvector analysis which recovers some results of previous studies. Our analysis highlights particular sets of important neurons constituting the core of the neural system. The applications of PageRank, CheiRank and ImpactRank to characterization of interdependency of neurons are discussed.

\section*{Introduction.} – The human brain neural network has an enormous complexity containing about $10^{11}$ neurons and $10^{14}$ synapses linking various neurons \cite{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 \cite{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 \cite{3} clearly demonstrated its efficiency being at the heart of Google search engine \cite{4}. Thus we can expect that the Google matrix analysis can find useful applications for the neural networks. This approach has been tested in \cite{5} on a reduced brain model of mammalian thalamocortical systems studied in \cite{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 \textit{C.elegans} (worm). The full connectivity of this directed network is known and documented at \cite{7}. The number of linked neurons (nodes) is $N = 279$ with the number of synaptic connections and gap junctions (links) between them being $N_L = 2990$.

Recently, there is a growing interest to the complex network approach for investigation of brain neural networks \cite{8,9,10,11,12}. 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 \cite{7}. The analysis of certain properties this directed network has been reported recently in \cite{11,12}, however, the approach based on the Google matrix has not been used yet. Thus we think that this study will allow to highlight the features of worm network using recent advancements of computer science.

\textbf{Google matrix construction.} – The Google matrix $G$ of \textit{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 \cite{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 columns with all elements $1/N$. Thus the sum of elements in each column is equal to unity and the Google matrix takes the form

$$ G_{ij} = \alpha S_{ij} + (1 - \alpha)/N. $$

\begin{equation} \tag{1} \end{equation}

Here $\alpha$ is the damping factor introduced in \cite{3}. In the context of the WWW, the last term of the equation describes a probability for a random surfer to jump on any node of the network \cite{4}. We use the usual value $\alpha = 0.85$ \cite{4}. All matrix elements $S_{syn,ij}, S_{gap,ij}, S_{ij}$ are given at \cite{7,8}.

The eigenspectrum $\lambda_i$ and right eigenvectors $\psi_i(j)$ of $G$
satisfy the equation
\[ \sum_{j'} G_{jj'} \psi_i(j') = \lambda \psi_i(j). \] (2)

The eigenvector at \( \lambda = 1 \) is known as the PageRank vector. According to the Perron-Frobenius theorem \([4]\) its elements \( P(j) \sim \psi_1(j) \) are positive and their 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.

It is also useful to consider the Google matrix obtained from the network with inverted directions of links (see e.g. \([14, 15]\)). The matrix \( G^* \) for this network with inverted direction of links is constructed following the same definition \([4]\). 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 \([14, 15]\) the top nodes of PageRank are the most popular pages, while the top nodes of CheiRank are the most communicative nodes \([14, 16]\).

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 \( N_G \) with indexes less than \( K \) is respectively \( N_G/K \approx 1.2, 10 \) at \( K = 10, 100 \). These values correspond approximately to those of WWW networks of Universities of Cambridge and Oxford being significantly smaller than the values of Twitter network characterized by a strong connectivity between top PageRank nodes with \( N_\theta/K \approx 100 \) for \( K = 100 \).

We note that the average number of links per neuron is \( \eta = N_\theta/N = 10.71 \) being approximately the same as for WWW of Universities of Cambridge and Oxford in 2006 \([16]\).

The global matrix structure is asymmetric. This leads to a complex spectrum of eigenvalues 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 \([17]\). 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 \( \text{Im}\lambda \) \([17]\). 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 \( \tau = 1/|\text{Im}\lambda| \) is approximately 5, 6.7 iterations of \( G, G^* \).

The properties of eigenstates \( \psi_i \) can be characterized by the Inverse Participation Ratio (IPR) \( \xi_i = \left( \sum_j |\psi_i(j)|^2 / \sum_j |\psi_i(j)|^4 \right)^2 \), which is broadly used in analysis of electron conductivity in disordered systems (see e.g. \([17, 18]\)). This quantity effectively determines the number of nodes on which is located an eigenstate \( \psi_i \). We see that some eigenstates have rather large \( \xi \approx N/3 \) while others have \( \xi \) located only on about ten nodes. We will return to the discussion of properties of eigenstates later.

**CheiRank versus PageRank.** The dependence of probabilities of PageRank and CheiRank vectors on their indexes \( K \) and \( K^* \) is shown in Fig. 3. A formal fit for a 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. \([4, 15]\)). Also in our neural network we find that the exponent in \( K \) is smaller then in
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 (\( \kappa \approx 3 \)) [10]. In a sense for C.elegans neural network the situation more similar to the networks of Linux Kernel (\( \kappa \approx 0 \)) [13] and BMP (\( \kappa = 0.164 \)) [19]. Thus, the C.elegans network has practically no correlations between ingoing and outgoing links. It is argued in [13,16] 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. 3. 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 neuronal contacts 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 [19].

### Table 1: Top twenty neurons of PageRank (PR), CheiRank (CR); ImpactRank of G (IMPR) and G* (IMCR) at initial state RMGL at \( \gamma = 0.7 \); following [7], the colors mark: interneurons (blue \( bu \)), motor neurons (green \( gn \)), sensory neurons (red \( rd \)), polyomodal neurons (purple \( pu \)).

| PR     | CR     | IMPR   | IMCR   |
|--------|--------|--------|--------|
| AVAR (bu) | AVAR (bu) | RMGL (bu) | RMGL (bu) |
| AVAR (bu) | AVAR (bu) | URXL (bu) | AVAR (bu) |
| PVCR (bu) | AVBR (bu) | ADEL (rd) | ASHL (rd) |
| RII (bu)  | AVBL (bu) | AIAL (local) | AVBL (bu) |
| ALAL (bu) | DD02 (gn) | ILZL (rd) | URXL (bu) |
| PHAL (rd) | DD02 (gn) | ADLL (rd) | AVEL (bu) |
| PHAR (rd) | DD01 (gn) | PVQL (bu) | RII (bu) |
| ADEL (rd) | RIBL (bu) | AMLL (rd) | RMDL (pu) |
| HSNR (gn) | RIVR (bu) | ASKL (rd) | RMDL (pu) |
| RMGR (bu) | DD04 (gn) | CEPL (rd) | RMDVL (pu) |
| RMDVL (pu) | DD03 (gn) | ASRL (rd) | AVAR (bu) |
| AIAR (bu) | DD01 (gn) | AWRL (bu) | SIVBR (bu) |
| AVBL (bu) | AVER (bu) | SADAR (bu) | AIBR (bu) |
| PIVPL (bu) | RMIE (gn) | RMDR (gu) | AADAL (bu) |
| AVM (rd)  | RMDVR (pu) | RMDL (gu) | RMDL (gu) |
| AVK (bu)  | AVEL (gu) | RII (bu) | AVBL (bu) |
| HSNL (gn) | DD05 (gn) | OQOVL (pu) | SIVBL (bu) |
| RMDL (gu) | SMDDR (pu) | AIML (gu) | ASKL (rd) |
| AVIR (bu) | DD03 (gn) | HSNL (gu) | RIV (bu) |
| AVF (bu)  | DD02 (gu) | SDQR (gu) | SMBSVL (gu) |

Fig. 3: (Colour on-line) Left panel: dependence of PageRank (CheiRank) probability \( P(K) (P^*(K^*)) \) on its index \( K (K^*) \) shown by black (red) curve. Right panel: dependence of ImpactRank probability \( P (P^*) \) on its index \( K (K^*) \), obtained via propagator of \( G (G^*) \) at \( \alpha = 0.85 \) and \( \gamma = 0.7 \) for the initial probability located on neuron RMGL (see text).

Fig. 4: (Colour on-line) PageRank - CheiRank plane \((K,K^*)\) showing distribution of neurons according to their ranking. Left panel: soma region coloration - head (red), middle (green), tail (blue). Right panel: neuron type coloration - sensory (red), motor (green), interneuron (blue), polyomodal (purple) and unknown (black). The classifications and colors are given according to WormAtlas [7].
rons such as $AVA$ and $AVB$ that integrate signals from crucial nodes and in turn pilot other crucial nodes. Neurons $AVAL$, $AVAR$, $AVBL$, $AVBR$ and $AVEL$, $AVER$ are considered to belong to the rich club analyzed in [12]. The right panel in Fig. 5 and second two columns of Table 1 represent ImpactRank which is discussed below.

![Fig. 5: (Colour on-line) Distribution of neurons in the plane $(K,K^*)$ of equal opportunity ranks (see text); colors are the same as in Fig. 3](image)

We can also use 2DRank index $K_2$, discussed in [15], 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 $AVAL$, $AVAR$, $AVBL$, $AVBR$, $PVCB$. Thus at the top $K_2$ values we find dominance of interneurons. More detailed listings are available at [13].

It may be also useful to consider renormalized equal opportunity rank recently discussed in [20]. In this approach PageRank probability of node $i$ is replaced by $P(i)/d(i)$ where $d(i)$ is in-degree of node $i$. For the Google matrix this recipe should be replaced by $P(i) \to P(i)/\sum_j G_{ij}$ and respectively for CheiRank by $P^\ast(i) \to P^\ast(i)/\sum_j G_{ij}$. The corresponding rank indexes $K, K^\ast$ rank the neurons in the decreasing order of these renormalized probabilities. The distribution of nodes in the plane $(K,K^\ast)$ 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 $K^\ast$ are composed mainly by motor neurons. Thus such an approach allows to highlight additional features of C.elegans network being complementary to PageRank and CheiRank properties discussed above. Tables for neuron renormalized ranking are available at [13].

**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 [18], is based on the probability distribution of a vector $v_f = (1 - \gamma)(1 - \gamma G)^{-1}v_0$, $v^\ast_f = (1 - \gamma)(1 - \gamma G^\ast)^{-1}v_0$, 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 [18]). The distributions of probabilities of ImpactRank $P(i) = v_f(i)$, $P^\ast(i) = v^\ast_f(i)$ versus the corresponding ImpactRank indexes $K, K^\ast$ are shown in Fig. 5 (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 $P^\ast(i)$ 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 $RMGL$.

**Properties of Eigenstates.** The Google matrix analysis of the Wikipedia hyperlink network [17] demonstrated that the eigenstates with large values of $|\lambda|$ select well defined communities. Thus we can expect that other eigenstates of matrices $G$ and $G^\ast$ 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 $|\lambda_i|$ with $\lambda_1 = 1$.

![Table 2: Top ten neurons of the eigenvectors of $G$ (left panel) and $G^\ast$ (right panel) corresponding to the 10th largest eigenvalues $|\lambda|$; IMPR are respectively $\xi \approx 5 \ell$ and $\xi \approx 4$.](image)

| $|\lambda| = -0.03646$ | $|\lambda| = -0.03784$ |
|-----------------|------------------|
| $AVAL$ | 0.11159 |
| $AVAR$ | 0.10651 |
| $AVBL$ | 0.096153 |
| $AVBR$ | 0.096153 |
| $RMGR$ | 0.050163 |
| $AVEL$ | 0.047016 |
| $AVER$ | 0.047016 |
| $PVQL$ | 0.047016 |
| $RMGL$ | 0.047016 |

![Table 3: Same as in Table 2 for 48th largest eigenvalue modulus $|\lambda|$; IMPR are respectively $\xi \approx 54$ and $\xi \approx 47$.](image)

| $|\lambda| = -0.03533 - 0.03533$ | $|\lambda| = -0.03533 - 0.03533$ |
|-----------------|------------------|
| $AVAR$ | 0.011072 |
| $AVBL$ | 0.011072 |
| $AVBR$ | 0.011072 |
| $PVQL$ | 0.011072 |
| $RMGL$ | 0.011072 |
| $AVEL$ | 0.011072 |
| $AVER$ | 0.011072 |

The top ten neurons in eigenfunction amplitude for four specific eigenstates of $G$ and $G^\ast$ are given in Table 2 and Table 3. In Table 2 we have eigenstates with low value of IMPR so that the corresponding wavefunctions are localized essentially on only about 4 neurons being $AIAR$, $AIAL$, $ASISL$, $ASIR$ and $AVAL$, $AVAR$, $AVBR$ for $\lambda_0$ of $G$ and $G^\ast$ respectively. In Table 3 the values of IMPR are rather large and these eigenstates are spread over many neurons.

---

p-4
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 \( K^* = 116(RIPR), K^* = 123(RIPL), K^* = 120(AS07), K^* = 122(AS10), K^* = 135(DB06), K^* = 137(DB05), K^* = 215(DA07), K^* = 162(VA10), K^* = 172(SIADL), K^* = 181(SIAVL), K^* = 199(SIAR), K^* = 221(SIADR) \).

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 \( |\lambda_i| \) values are available at [13].

**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.

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**

[1] Williams R.W. and Herrup K., *Ann. Rev. Neuroscience*, 11 (1988) 423

[2] http://www.worldwidewebsize.com/

[3] Brin S. and Page L., *Comput. Networks ISDN Syst.*, 30 (1998) 107

[4] Langville A. M. and Meyer C. D., *Google’s PageRank and Beyond* (Princeton University Press) 2006

[5] Shepelyansky D. L. and Zhirov O. V., *Phys. Lett. A*, 374 (2010) 3206

[6] Izhikevich E.M. and Edelman G.M., *Proc. Nat. Acad. Sci.*, 105 (2008) 3593

[7] Altun Z.F., Herndon L.A., Crocker C., Lints R. and Hall D.H. (Eds.), *WormAtlas*, http://www.wormatlas.org (2012)

[8] Arenas A., Fernández A. and Gómez S., *Bio-Inspired Comput. Commun. Lect. Notes in Computer Science*, 5151 (2008) 9

[9] Bullmore E., Sporns O., *Nature Rev. Neurosci.*, 10 (2009) 312

[10] Varier S. and Kaiser M., *PLoS Comput. Biol.*, 7(1) (2011) e1001044

[11] Varshney L.R., Chen B.L., Paniagua E., Hall D.H. and Chklovskii D.B., *PLoS Comput. Biol.*, 7 (2011) e1001066

[12] Towlson E.K., Vértes P.E., Ahnert S.E., Schafer W.R. and Bullmore E.T., *J. Neurosci.*, 33(15) (2013) 6380

[13] Kandiah V. and Shepelyansky D.L., http://www.quantware.ups-tlse.fr/QWLIB/wormgooglematrix/ (2013)

[14] Chepelianskii A. D., arXiv:1003.5455[cs.SE] (2010)

[15] Zhirov A. O., Zhirov O. V. and Shepelyansky D. L., *Eur. Phys. J. B*, 77 (2010) 523

[16] Ermann L., Chepelianskii A.D. and Shepelyansky D.L., *J. Phys. A: Math. Theor.*, 45 (2012) 275101

[17] Ermann L., Frahm K.M. and Shepelyansky D.L., *Eur. Phys. J. B*, 86 (2013) 193

[18] Frahm K.M., Eom Y.-H. and Shepelyansky D.L., arXiv:1310.5624 [physics.soc-ph] (2013)

[19] Abel M. and Shepelyansky D.L., *Eur. Phys. J. B*, 84 (2011) 493

[20] Banky D., Ivan G. and Grolmusz V., *PloS ONE*, 8(1) (2013) e54204

fig:6: (Colour on-line) Dependence of amplitude of eigenstates \( \psi_i(K) \) of \( G \) and \( \psi_i^*(K^*) \) of \( G^* \) on PageRank index \( K \) (left panel) and CheiRank index \( K^* \) (right panel); here \( x \)-axis shows values of \( K \) and \( K^* \), while \( y \)-axis shows index \( i \) of eigenstates being ordered in a decreasing order of \( |\lambda_i| \) (see text). The whole index range \( 1 \leq K, K^* \leq 279 \) is shown with PageRank (CheiRank) vector being at the bottom line of each panel. The color is proportional to \( |\psi_i(j)| \) changing from minimum blue value to maximum value in red.