

SCIENTIFIC REPORTS



OPEN

The Robustness and the Doubly-Preferential Attachment Simulation of the Consensus Connectome Dynamics of the Human Brain

Balázs Szalkai¹, Bálint Varga¹ & Vince Grolmusz^{1,2} 

Consensus Connectome Dynamics (CCD) is a remarkable phenomenon of the human connectomes (braingraphs) that was discovered by continuously decreasing the minimum confidence-parameter at the graphical interface of the Budapest Reference Connectome Server, which depicts the cerebral connections of $n = 418$ subjects with a frequency-parameter k : For any $k = 1, 2, \dots, n$ one can view the graph of the edges that are present in at least k connectomes. If parameter k is decreased one-by-one from $k = n$ through $k = 1$ then more and more edges appear in the graph, since the inclusion condition is relaxed. The surprising observation is that the appearance of the edges is far from random: it resembles a growing, complex structure. We hypothesize that this growing structure copies the axonal development of the human brain. Here we show the robustness of the CCD phenomenon: it is almost independent of the particular choice of the set of underlying connectomes. This result shows that the CCD phenomenon is most likely a biological property of the human brain and not just a property of the data sets examined. We also present a simulation that well-describes the growth of the CCD structure: in our random graph model a doubly-preferential attachment distribution is found to mimic the CCD.

The plenty of high-quality structural MRI data from the human brain makes possible of studying the cerebral anatomy in an unprecedented way today. Among other large projects, the NIH-funded Human Connectome Project¹ records and publishes multimodal MRI data from hundreds of healthy individuals. High-angular resolution diffusion imaging (HARDI) data can be processed to discover connections, consisting of axonal fibers, between anatomically identified² gray matter areas. Consequently, a braingraph or connectome can be constructed that contains the connections as follows: the nodes or vertices of the graph correspond to the anatomically identified gray matter areas, and two nodes are connected if fibers of axons are discovered between them by processing the diffusion weighted data^{3–5}.

If we have braingraphs from several hundred subjects, then, since the vertices of the different braingraphs are corresponded to the very same brain map², we can describe the diversity of the edges between different subjects and in different lobes or smaller brain areas as in⁶, or we can just describe the common edges through numerous subjects, as in the Budapest Reference Connectome Server <http://connectome.pitgroup.org>^{7,8}. The data source of these studies was the Human Connectome Project¹.

Distinguishing the frequently and rarely appearing connections within the human brain may help the neuroscientist in identifying the normally appearing, usual, standard and non-standard connections. These non-standard connections can cause or can be caused by some disease, or just can be the result of the personal variability with or without psychological consequences. Therefore, the mapping of the frequent and the infrequent connections by the Budapest Reference Connectome Server can have straightforward clinical significance.

Very surprisingly, we have discovered a phenomenon, called Consensus Connectome Dynamics (CCD), on the Budapest Reference Connectome Server, which may open up new horizons in the study of the development of the human brain⁹. This discovery was very surprising since the server was not built to study brain development: the imaging data is originated from adults between 22 and 35 years of age, so the age-span is — seemingly — inadequate for studying the early development of brain connections, occurring in several months just before and

¹PIT Bioinformatics Group, Eötvös University, H-1117, Budapest, Hungary. ²Uratim Ltd., H-1118, Budapest, Hungary. Correspondence and requests for materials should be addressed to V.G. (email: grolmusz@pitgroup.org)

after the birth¹⁰. To clarify this discovery, we need to cover some details of the Budapest Reference Connectome Server <http://connectome.pitgroup.org>.

The server is capable of computing and visualizing the consensus connectomes with setting several parameters. The braingraphs of $n = 418$ subjects are processed on the server. Let k be an integer such that $1 \leq k \leq n$. Let us call an edge e k -frequent if the edge e is present in at least k braingraphs out of the maximum n graphs. Let us call a connectome k -consensus connectome if it contains all the k -frequent edges. The k -consensus connectomes, consequently, contains all edges that are present in at least k connectomes. For $k = 1$, the 1-consensus connectome contains all the edges that are present in at least one subject's braingraph out of the n graphs. The n -consensus connectome contains the edges that are present in all subject's connectomes. Clearly, the n -consensus connectome contains much fewer edges than the 1-consensus connectome. Let $1 \leq i \leq j \leq n$ then it is also obvious that the edges of the j -consensus connectome are also present in the i -consensus connectome. This means that the i -consensus connectome contains more and more edges by the decrease of i .

The fascinating observation⁹ is that the new edges that appear in the i -consensus connectome, relative to the $i+1$ -consensus connectome, are not placed randomly, they seem to connect to the edges of the $i+1$ -consensus connectome. Consequently, if we consider the k -consensus connectomes, for decreasing k values from $k = n, n-1, \dots, 3, 2, 1$ then we get more and more edges, and the edges form a growing, complex, tree-like structure.

The observation is visualized on a video at <https://youtu.be/yxlyudPaVUE> for the whole brain and at https://youtu.be/wBciB2eW6_8 restricted for the frontal lobe only^{9,11}. The observation is statically visualized on a very large component-tree at http://pitgroup.org/static/graphmlviewer/index.html?src=connectome_dynamics_component_tree.graphml, which is described in detail in⁹. The observation is analyzed quantitatively on Fig. 2 in⁹ for the whole brain and on Fig. 1 of¹¹ for the frontal lobe only.

The interested reader can also experience the Consensus Connectome Dynamics phenomenon on the Budapest Reference Connectome Server <http://connectome.pitgroup.org> by (i) choosing the "Show options" button and (ii) moving the "Minimum edge confidence" slider to the rightmost position, and (iii) slowly moving the "Minimum edge confidence" slider from right to left.

In^{9,11} we hypothesized that the Consensus Connectome Dynamics (CCD) phenomenon describes the order of the development of the connections of the brain: the deviation of the oldest, first developed connections are the smallest, and, gradually, the newer and newer developed connections cumulate the deviations of the connections that they connect to, and, because of this, their deviation will be higher and higher; that is, they will appear only in k -consensus connectomes for smaller values of k .

Results

In the present contribution we examine two relevant questions concerning the CCD phenomenon:

- Robustness:** The CCD phenomenon can be characterized by the order of appearance of the edges in the growing graphs of k -consensus connectomes, for the decreasing k parameters. For showing that this order has any biological meaning and it is not just the product of the specific choice of the dataset processed, we need to demonstrate the independence of the appearance of the edges from the particular choice of the underlying dataset.
- Random graph model for CCD:** The main reason for preparing a random network model for a known, interesting graph is uncovering the possible mechanism involved in the development (or evolution) of the graph. As the most famous example, the Barabási-Albert model for the description of the degree distribution of the webgraph^{12,13} uses the "preferential attachment" principle in the description of the development of the webgraph. In the Barabási-Albert model, roughly, in every step one new vertex appears, and it connects to the older ones with probabilities proportional to the degree of the older vertices. It was shown first by computer simulation¹² and, later by an exact mathematical proof¹³ that this process led to the power law degree distribution with exponent -3 . Most importantly, since the random simulation process well-described the degree-distribution of the webgraph, the model uncovers also the mechanism that guides the users of the web in hyperlinking the new vertices (web pages) in the webgraph.

Discussion

In what follows we show that the CCD phenomenon is robust in the sense described above, and we also define a probabilistic graph model with a "doubly preferential attachment" that well-describes the CCD phenomenon.

Robustness. Here we examine the independence of the CCD phenomenon of the particular choice of the underlying datasets of braingraphs. For this goal, we partitioned the 418 braingraphs into 4 disjoint sets of almost equal size with a ± 1 margin (let us call these sets "quarter sets"), and we computed the order of appearance of the edges in the CCD, according to each quarter set. Next, we have compared the order of appearance of the edge-pairs as follows:

We have chosen those edges that are present in (at least one graph) in all the four quarter sets; there are 31,873 such edges. Then we - randomly - chose two of the quarter sets out of the four, say X and Y , and also randomly two edges, say e and f out of the 31,873 ones that are present in all four quarter sets. Next, we compute the experimental probability that in the X dataset-based edge e appears strictly before f and in the Y based experiment f appears strictly before e (that is, their order of appearances differs).

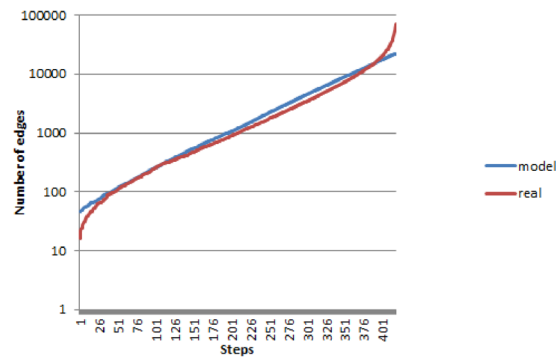


Figure 1. The comparison of the edge numbers in the CCD phenomenon with the edge numbers in the doubly-preferential attachment random model we suggest. The horizontal axis contains the numbering of the steps: step i corresponds to the $(n + 1 - i)$ -consensus connectome.

If in the CCD the edges just appeared randomly this probability would be equal to $1/2$. The smaller is the probability, the more robust is the CCD phenomenon. We have got this probability to be 0.104 .

Similarly, we have also computed the order of the connections of the vertices in the consensus connectomes, and for a randomly chosen u, v vertex-pair we computed the experimental probability that in the X dataset-based CCD vertex u appears strictly before v and in the Y based experiment v appears strictly before u . We have got that this probability is 0.053 .

Dealing with possible artifacts: It is known that the algorithmic details in the workflow of processing the MR images may influence the connectomes constructed from these images (e.g.¹⁴, compares the effects of the choice of different tractography algorithms to connectomes). The 418 graphs in our computations above have been constructed using a deterministic fiber-tracking algorithm (SD_STREAM option in MRtrix 0.2). For excluding processing artifacts, we have re-computed all the 418 graphs with probabilistic tractography (MRtrix 0.2, with 1 million streamlines, white matter seeding/masking and probabilistic fiber-tracking [the SD_PROB option]). Next, we compared the new graphs (constructed with the application of probabilistic tractography) with the old ones (constructed with deterministic tractography) in several ways:

The probability that two random edges appeared in a different order in the CCD in two random quarter sets in the new graphs: 0.076 (it is better than in the old graphs, there this value was 0.104).

The probability that two random vertices are connected in different order to the consensus connectomes in the growing CCD structure in the new graphs: 0.067 (it is a little bit worse than the value for the old graphs: 0.053).

We have also compared the old and the new graphs as follows: we have taken the edges that were present in the new and the old graphs, in all 4 quarter sets of the old and new graphs, then we took two random edges and two random quarter sets, one from the old, the other from the new quarter sets, and have found that 0.101 is the probability that the order of the appearance of these two edges differ in the old and the new graphs in the CCD phenomenon.

Similarly, we have found that the probability of the connection of two random vertices in the CCD phenomenon by two random quarter sets (one from the old, one from the new quarter sets) is 0.085 (for the old graphs this value was 0.053).

Therefore, we can conclude that for the edges and the vertices, the order of their appearance in the CCD phenomenon is almost independent of the underlying dataset, so, in our opinion, this order describes a property of the brain, and not of the datasets.

Random graph model for the CCD. There are three significant differences — relative to the webgraph-simulation — that need to be addressed in developing the random graph model for the CCD phenomenon:

- (i) In CCD, we have the data of the buildup of the graph; in other words, in CCD we have a dynamic process of the appearances of the new edges, while in the case of the webgraph only a static image: the degree distribution of the vertices in the graph;
- (ii) In CCD we observe new edges between those “old” vertices that were connected to some edges in the previous steps (i.e., with larger k s in k -consensus connectomes). In the Barabási-Albert model, new edges are always connected to the new vertex, and they never appear between two old nodes.
- (iii) We do not intend to model an unboundedly growing graph as in^{12,13}; our goal is to model the 1015-vertex CCD phenomenon.

Here we suggest a doubly-preferential attachment probability distribution for the new edges: the probability of the appearance of a new uv edge between vertices u and v is proportional to the sum $\deg u + \deg v$, i.e., the sum of their degrees. We call this rule “doubly-preferential attachment”, because in the Barabási-Albert model¹² the new vertex u was connected to old vertex v with a probability, proportional to $\deg v$ (the “preferential attachment model”). The mathematical details and the parameter choices are detailed in the “Methods” section.

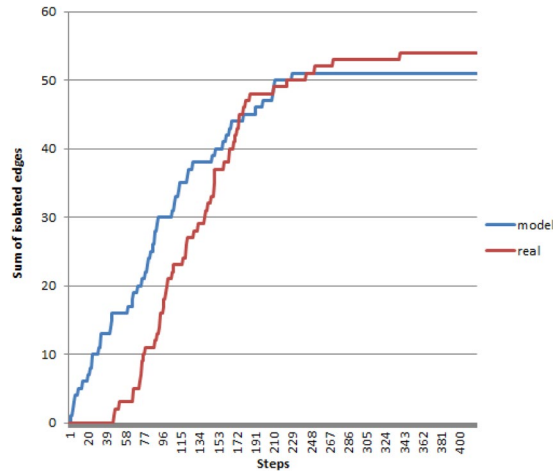


Figure 2. The comparison of the sum of the isolated edges in the CCD phenomenon with the corresponding value in the doubly-preferential attachment random model we suggest. The horizontal axis contains the numbering of the steps: step i correspond to the $n - i$ -consensus connectome.

Figure 1 compares the increase of the edge numbers in the real CCD phenomenon and in the doubly-preferential attachment model we suggest. Step i on the horizontal axis correspond to the $(n + 1 - i)$ -consensus connectome.

Figure 2 compares the sum of the isolated edges in the CCD and the random, doubly-preferentially attached model. An edge is called “isolated” in the k -consensus connectome, if it does not connect to any other edges, and it was not present in the $k + 1$ -consensus connectome. One quantitative characterization of the CCD phenomenon is the very small number of isolated edges (c.f. Figure 2 in⁹ and Fig. 1 of¹¹). Therefore the sum of the isolated edges is a appropriate measure of the good characterization of the CCD phenomenon.

Conclusions

We have shown that the CCD phenomenon is robust, in other words, most probably it describes a biological phenomenon, and it is not just the property of the particularly chosen datasets.

We have also shown that the doubly-preferentially attached model well-describes the CCD phenomenon. This fact also strengthens our hypothesis described in^{9,11} that the CCD phenomenon copies the axonal development of the brain on a macroscopic level: there we hypothesized that the new axonal connections prefer connecting to neurons with numerous existing connections; the success of the doubly-preferentially attached model is in line with this assumption, since here new edges appear more probably between – already – high-degree nodes.

Methods and Materials

We used a Barabási-Albert-like model for approximating the connectome distribution. First, we observed that the number of edges is approximately an exponential function of k , with sharp increases at the beginning and at the end. An exponential regression yielded the equation $46.37e^{0.014k}$ ($R^2 = 0.99$). Let $A := 46.37$ and $B := 0.014$. From this equation we have derived the following simple model: we start from a $[A]$ -edge graph (i.e. a 46-edge graph), generated randomly over D selected nodes (where $D \leq N$ is a parameter of the model) then, in each step, we add each uv edge with the probability

$$p_{uv} := \frac{B}{2(N - 1)}(\text{deg } u + \text{deg } v), \tag{1}$$

where $\text{deg } u$ denotes the degree of node u in the previous step.

This indeed yields an exponentially growing number of edges. Observe that the expected number of new edges in the next step is (if we do not account for multiple edges)

$$\sum_{u \in V} \sum_{\substack{v \in V \\ v > u}} \frac{B}{2(N - 1)}(\text{deg } u + \text{deg } v) = \frac{B}{2(N - 1)} 2 \frac{N - 1}{2} \sum_{u \in V} \text{deg } u = B|E|, \tag{2}$$

where $|E|$ is the number of edges in the previous step. Thus our model indeed generates an exponential expected number of edges, namely approximately $[A]e^{Bk} = 46e^{0.014k}$ edges in the k th step, which is consistent with the exponential regression.

Unfortunately, this model does not allow adding new edges between zero-degree (isolated) nodes, because p_{uv} becomes 0 for those nodes. To circumvent this problem, we have modified the equation so that we allow a certain probability for the inclusion of these “isolated” edges. We added a constant to p_{uv} , that is, in our new model, the probability of inclusion for the edge uv has now become

$$p_{uv} := \frac{B}{2(N-1)}(\deg u + \deg v) + C, \quad (3)$$

where C is the inclusion probability for isolated edges.

This causes the number of edges to be not $[A]e^{Bk}$, but approximately $\left([A] + \left(\frac{N}{2}\right)C/B\right)e^{Bk}$. This is because the expected number of new edges of C is $\left(\frac{N}{2}\right)C$ in each step, and the number of edges is multiplied by about $1 + B$ in each step. Therefore, by using the $1 + z + z^2 + \dots = \frac{1}{1-z}$ formula for $z := \frac{1}{1+B}$, we can derive the number of edges relative to e^{Bk} . Based on this formula, we need to decrease the initial number of edges from $[A]$ to $\left\lceil A - \left(\frac{N}{2}\right)C/B + 0.5 \right\rceil$ (0.5 is added for rounding to the nearest integer).

Since it is unlikely that two isolated edges appear from the same node at once, the value of C influences the total number of new isolated edges in an almost linear fashion. So we can start from a value of C , count the total number of new isolated edges, then compare it with the desired total number of isolated edges, and divide C with this ratio. This way, we determined the optimal value for C as 7.6×10^{-7} . We found that, in reality, the cumulative number of new isolated edges is 0 up until step 47, then increases in an approximately linear fashion up until step 186, and after that it levels off at 50–55 edges in total. In comparison, the average curve of the cumulative number of new isolated edges in our simulation increased linearly until about step 130, had a concave section until about step 230, where it leveled off at 56 edges.

We can thus conclude that our model for CCD not only approximates the number of edges in each step well, but also the cumulative number of isolated edges. Furthermore, to avoid over-fitting, the model is simple and only has 4 parameters.

Data availability. The Human Connectome Project's MRI data is accessible at: <http://www.humanconnectome.org/documentation/S500>¹.

The graphs (both undirected and directed) that were prepared by us from the HCP data can be downloaded at the site <http://braingraph.org/download-pit-group-connectomes/>.

References

- McNab, J. A. *et al.* The Human Connectome Project and beyond: initial applications of 300 mT/m gradients. *Neuroimage* **80**, 234–245 (2013).
- Fischl, B. Freesurfer. *Neuroimage* **62**, 774–781 (2012).
- Daducci, A. *et al.* The connectome mapper: an open-source processing pipeline to map connectomes with MRI. *PLoS One* **7**, e48121 (2012).
- Gerhard, S. *et al.* The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Front Neuroinform* **5**, 3 (2011).
- Tournier, J., Calamante, F., Connelly, A. *et al.* Mrtrix: diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* **22**, 53–66 (2012).
- Kerepesi, C., Szalkai, B., Varga, B. & Grolmusz, V. Comparative connectomics: Mapping the inter-individual variability of connections within the regions of the human brain. *Neuroscience Letters* **662**, 17–21 (2018).
- Szalkai, B., Kerepesi, C., Varga, B. & Grolmusz, V. Parameterizable consensus connectomes from the human connectome project: The Budapest Reference Connectome Server v3.0. *Cognitive Neurodynamics* **11**, 113–116 (2016).
- Szalkai, B., Kerepesi, C., Varga, B. & Grolmusz, V. The Budapest Reference Connectome Server v2.0. *Neuroscience Letters* **595**, 60–62 (2015).
- Kerepesi, C., Szalkai, B., Varga, B. & Grolmusz, V. How to direct the edges of the connectomes: Dynamics of the consensus connectomes and the development of the connections in the human brain. *PLoS One* **11**, e0158680 (2016).
- Lewis, T. L. Jr., Courchet, J. & Polleux, F. Cell biology in neuroscience: Cellular and molecular mechanisms underlying axon formation, growth, and branching. *J Cell Biol* **202**, 837–848 (2013).
- Kerepesi, C., Varga, B., Szalkai, B. & Grolmusz, V. The dorsal striatum and the dynamics of the consensus connectomes in the frontal lobe of the human brain. *arXiv* **1605.01441** (2016).
- Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Bollobas, B., Riordan, O., Spencer, J. & Tusnady, G. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms* **18**, 279–290 (2001).
- Zhan, L. *et al.* Comparison of nine tractography algorithms for detecting abnormal structural brain networks in alzheimer's disease. *Frontiers in Aging Neuroscience* **7**, 48 (2015).

Acknowledgements

Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. VG acknowledges the funding of OTKA KH-126472 and VEKOP-2.3.2-16-2017-00014 of Ministry for National Economy, Hungary.

Author Contributions

S.B. and V.G. wrote the main manuscript text; S.B. discovered that the doubly-preferential attachment model well-describes the CCD phenomenon, analyzed data and prepared Figures 1 and 2. V.G. initiated the robustness studies, B.V. performed tractography computations and constructed the braingraphs. All authors have reviewed the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017