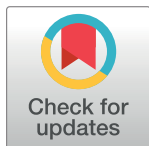


RESEARCH ARTICLE

Effects of parcellation and threshold on brainconnectivity measures

T. C. Lacy^{1,2}, P. A. Robinson^{1,2*}**1** School of Physics, University of Sydney, Sydney, NSW, Australia, **2** Center for Integrative Brain Function, University of Sydney, Sydney, NSW, Australia* robinson@physics.usyd.edu.au

Abstract

It is shown that the statistical properties of connections between regions of the brain and their dependence on coarse-graining and thresholding in published data can be reproduced by a simple distance-based physical connectivity model. This allows studies with differing parcellation and thresholding to be interrelated objectively, and for the results of future studies on more finely grained or differently thresholded networks to be predicted. As examples of the implications, it is shown that the dependences of network measures on thresholding and parcellation imply that chosen brain regions can appear to form a small world network, even though the network at finer scales, or ultimately of individual neurons, may not be small world networks themselves.

OPEN ACCESS

Citation: Lacy TC, Robinson PA (2020) Effects of parcellation and threshold on brainconnectivity measures. PLoS ONE 15(10): e0239717. <https://doi.org/10.1371/journal.pone.0239717>

Editor: Niels Bergsland, University at Buffalo, UNITED STATES

Received: June 1, 2020

Accepted: September 14, 2020

Published: October 1, 2020

Copyright: © 2020 Lacy, Robinson. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: There are no data. This is a theoretical paper. Results are provided in the paper.

Funding: PAR received grants CE140100007, DP130100437, and FL140100025 to support this work, from the Australian Research Council. The funders had no role in study design, data collection, nor preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

It is important to understand the structure and function of the brain and its large network of 10^{11} neurons each with 10^4 synapses to other neurons. In order to simplify this network to a manageable level, one of the main approaches used is to coarse-grain the network by grouping large numbers of neurons into regions (typically 30 – 1000) based on spatial proximity and then study the connections between these regions. This grouping is partly due to the limited ~ 1 mm spatial resolution of MRI measurements, but MRI voxels are often further parcellated into larger, cm-scale regions of interest. The hope is that the properties of the coarse-grained network will be useful in themselves and will also be similar to the original network of neurons, so that studying the former may provide useful insight into the latter; however, there is no proof that this similarity of properties will be the case. The connections between regions are usually determined by estimating the numbers of axonal fibers between them, but can also be based on functional correlations in their activity. From these measurements, either a connection strength is assigned between each pair of regions, or the network is thresholded to simplify the analysis, sometimes binarizing by keeping connections between regions only if they are above a certain strength and discarding them if they are weaker.

These inter-regional networks are then typically characterized [1–4] by using graph theory and calculating graph measures of the network, such as the clustering coefficient and the average path length. From this, conclusions about the structure of the overall network of neurons

are often drawn, usually including a conclusion that the regions of the brain form a small-world network. This is done by comparing the network measures to those of the model in [5], which is widely used as a model for small-world networks. However, this analysis is done without directly acknowledging that the coarse-grained network of regions is an approximation to the physical network of the brain, and that information is also lost by binarizing connections. In addition, there is a large degree of variation in the properties reported between studies [1–3, 6, 7]. There has been preliminary work on how the coarse-graining and thresholding affect the network properties [8–11], suggestions that the spatial embedding of regions will influence the path length, clustering, and node degree [11, 12–15], and work on spatial networks in general [13, 16, 17], but no systematic analysis has yet been done for the brain. Without ensuring that the analysis is robust with respect to changes in the parcellation and threshold, it is entirely possible that as the number of regions increases, the graph properties may change in ways that invalidate the analysis and its conclusions. For example, that a network of 100 brain regions forms a small-world network does not necessarily mean that finer networks, nor ultimately the 10^{11} underlying neurons, form small-world networks too [18–20]. For this conclusion to be drawn, a causal link would need to be found explicitly, instead of just being assumed to exist.

In addition, most network studies do not have any model for the process of obtaining connections between regions from the network of synapses when it is parcellated and thresholded. Many models have been put forward for the structure of the connections between regions [21], usually favoring preferential attachment [discussed in detail in [13]], but sometimes following other rules. However, they are usually proposed ad hoc to match the data on coarse-grained connections between the regions, instead of modeling the underlying neuronal connections (which are ignored). To overcome this deficiency, we must start with a plausible physical model and coarse-grain it instead of jumping directly to modeling the regions. In order to focus on the main effects of parcellation and thresholding, our model is chosen to reproduce the approximately exponential fall-off of connectivity with distance and the overall spherical topology of a cortical hemisphere. Refinements due to cortical folding and bihemispheric structure are not included.

In this paper we show that by simulating and examining the underlying network of individual neurons within the brain, we can predict how the coarse-grained network changes properties with parcellation and threshold. In particular, we predict how the clustering coefficient and average path length change with parcellation and threshold, and reproduce the data of [8]. This analysis is then extended, as a further example, to determine how the small-world properties of the coarse-grained networks vary with parcellation and threshold.

Theory and methods

In this section, we outline the procedure used for generating networks of regions from a simulated synaptic network with a simple distance-based connection rule in which the probability of connections decreases exponentially with distance, in accord with anatomical data [9].

In order to show all of the steps of the modeling procedure before applying them to real brain geometries, we illustrate them explicitly for a small network (36 regions of 5 neurons each), with the regions spread evenly on a 2D square (instead of approximately evenly throughout a 3D sphere, like an actual brain). This makes it easier to show how the synaptic network is thresholded to yield a regional network. Fig 1 shows the first step of placing the neurons throughout a simulated brain. Each of the 36 regions contains exactly 5 neurons, and there is some clumping of neurons occurring due to the random placement of them within the regions.

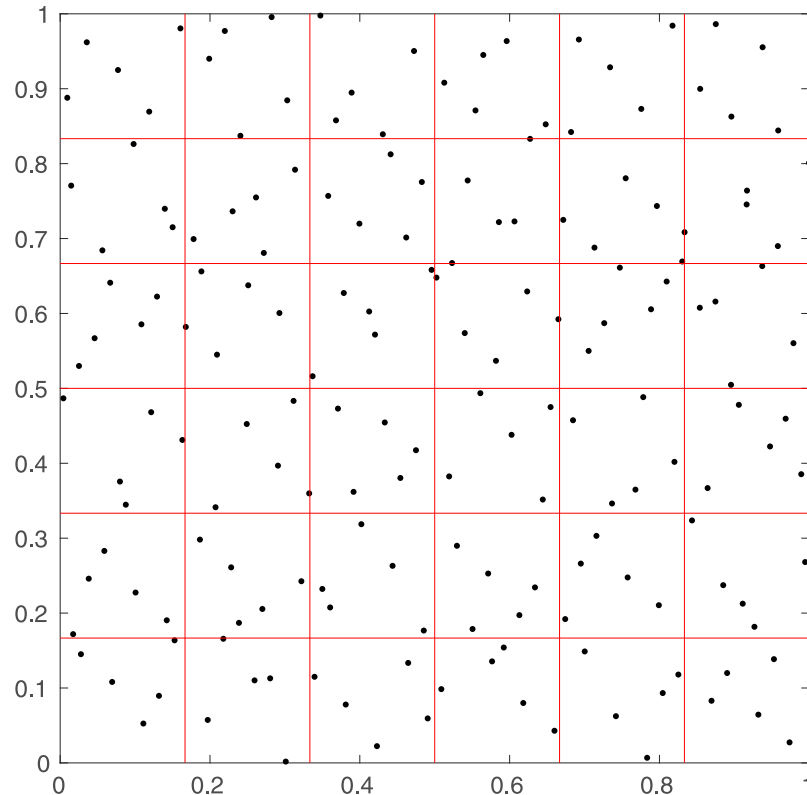


Fig 1. Simulated neural network. Locations of the neurons (black crosses) in the simulated network on a square cortex, in 36 regions of 5 neurons each (red lines show the borders of each region, axes show spatial coordinates on the square cortex with side length normalized to 1).

<https://doi.org/10.1371/journal.pone.0239717.g001>

Once the neuron locations are determined, the next step is to determine connections between the neurons. For pairs of neurons, the probability of there being a synapse connecting them decreases the larger the further separated they are. Because longer range connections are rarer, possibly due to the cost of growing long axons, we assume the probability of connection between neurons follows a simple exponential distribution, although other distributions can be used. If the neurons are at positions \mathbf{a} and \mathbf{b} , the probability they are connected is approximated as

$$P(\mathbf{a}, \mathbf{b}) = C e^{-|\mathbf{a}-\mathbf{b}|/r_0}, \quad (1)$$

where C is a base connection probability between neurons located in very close proximity, and r_0 is a typical connection range. The connections from neurons in one of the regions to all the other neurons can be seen in Fig 2. This shows that the neurons are often connected to neurons within the same or adjacent regions, and rarely connected to neurons in distant regions.

We now have a full simulated network of neurons and regions with a network of connections present between neurons. However, for a typical brain with 10^{11} neurons, this network is too hard to measure directly, so papers such as those by [1, 2], and [3] measured the connections between regions instead, by coarse-graining the network. The most common procedure is to examine the degree to which each pair of regions is connected (either through correlations of activity or the number of connections between neurons). Then, all connections between regions that are above a certain threshold are kept, and the rest are discarded. This

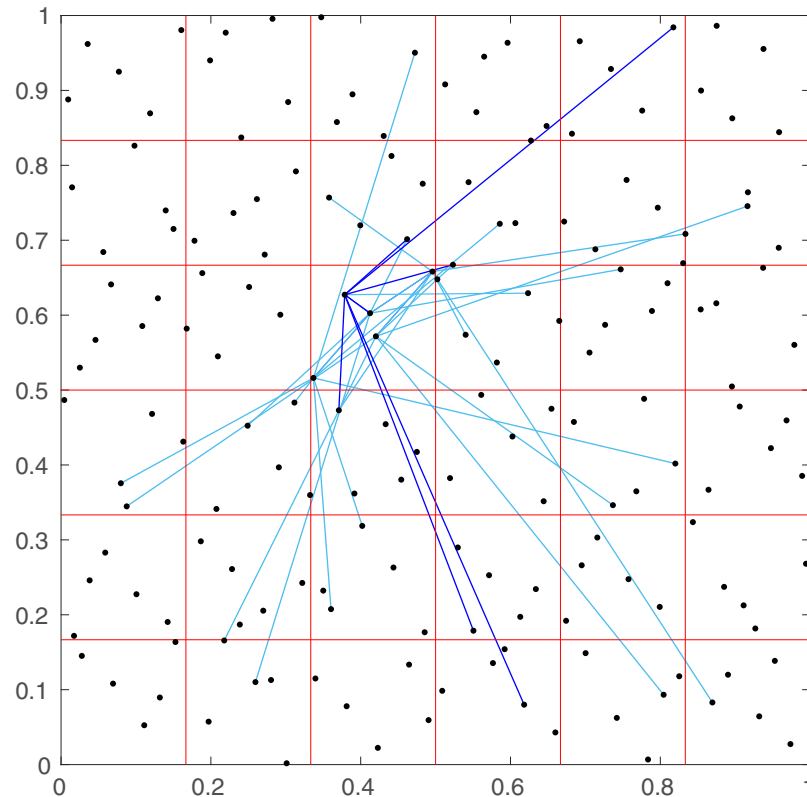


Fig 2. Simulated neural connections. Connections (blue lines, darker for one particular neuron) from neurons in one region to ones in the rest of the simulated network, with the same neurons and regions as Fig 1. Axes show spatial coordinates, as in Fig 1.

<https://doi.org/10.1371/journal.pone.0239717.g002>

gives a much simpler network that involves only connections between approximately 1000 regions instead of connections between all of the 10^{11} neurons.

The effects of thresholding can be seen in Fig 3, where the number of interregional synaptic connections required for pairs of the 36 regions to be considered to be connected is set to either 1, 2, or 3, respectively. Each point in this figure corresponds to one of the red boxes in Fig 1. The number of connections between the regions decreases with increasing threshold, and the majority of the connections are between regions in close proximity.

To test the behavior for larger networks, we increase the number of neurons from 5 per region to 100, and compare to the previous results. This gives Fig 4, where the threshold for the connections required is 200, 600, or 1000, respectively. This shows networks that are much more grid-like, because the correlation between proximity and connection strength between regions is much stronger; decreasing the threshold gradually allows for longer connections between regions to be present. This behavior also tends to hold for other probability distributions for connections that decrease with distance, because the large number of neurons per region makes it very unlikely that distant regions will be strongly connected.

The above behavior can be used to enable a very simple but effective model-based analysis of the coarse-graining. Because there will be approximately at least 10^7 neurons present in regions at the parcellation scales used in experiments, thresholding will tend to just leave the closest pairs of regions connected. This means that the coarse-graining of the neurons can then be modeled without having to simulate every neuron present.

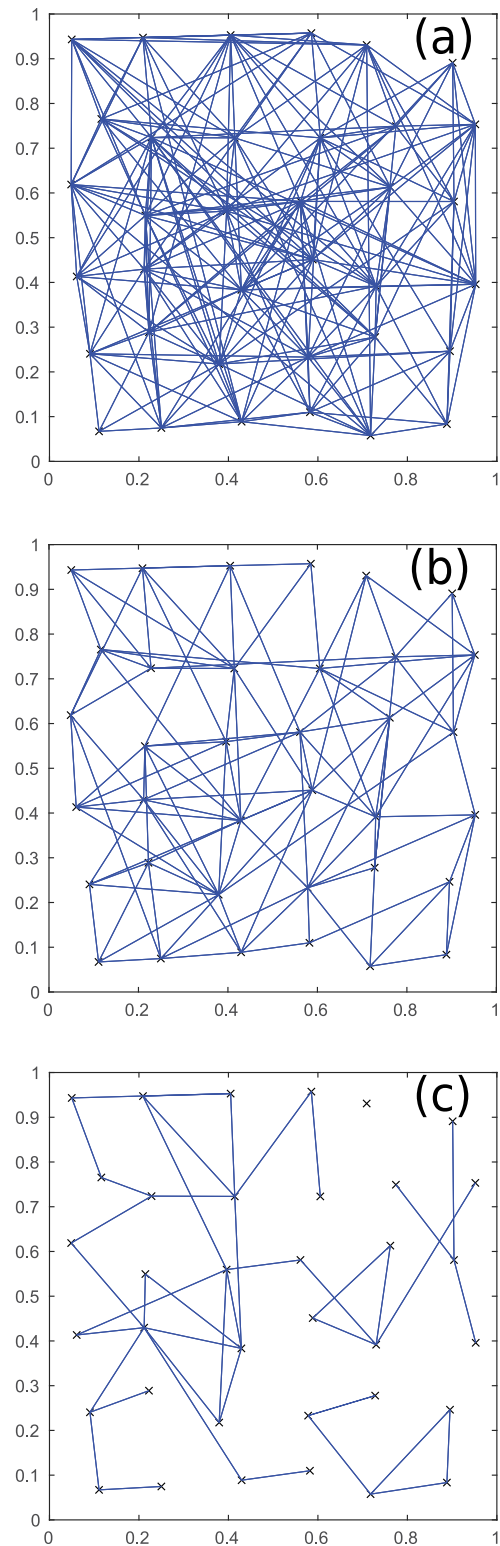


Fig 3. Thresholded neural connections. Connections remaining after thresholding between the 36 regions in Fig 1 for three different thresholds. (a) Threshold of 1 connection. (b) 2 connections. (c) 3 connections. Axes show spatial coordinates, as in Fig 1.

<https://doi.org/10.1371/journal.pone.0239717.g003>

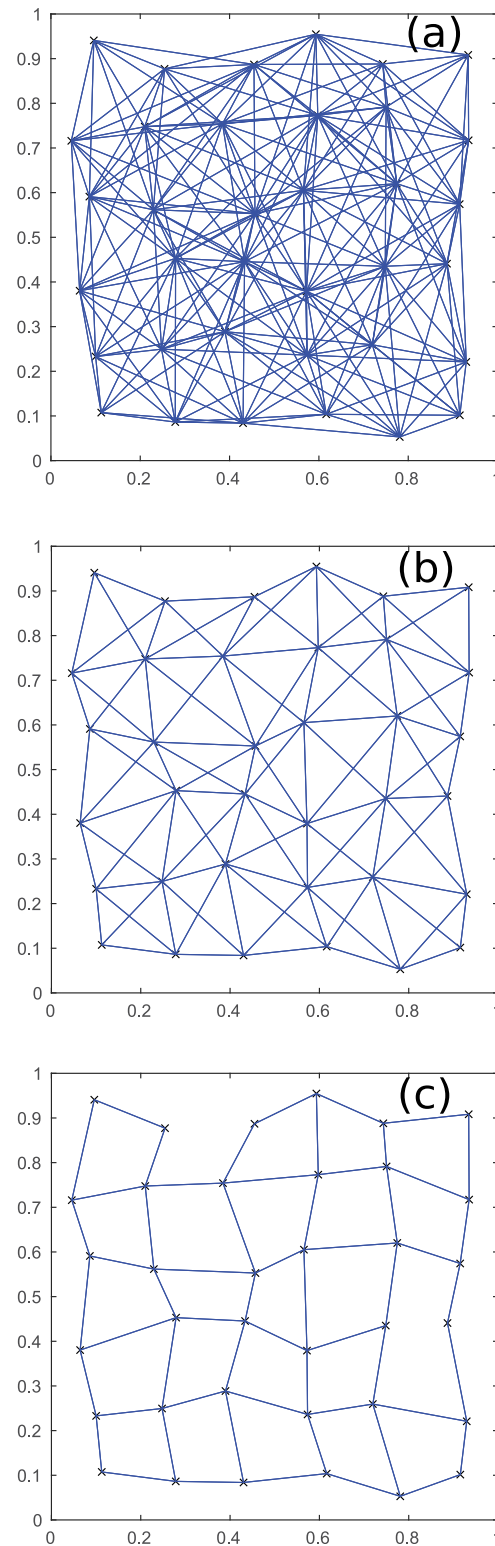


Fig 4. Thresholded inter-regional connections. Connections between regions for 36 regions of 100 neurons, for three thresholds. (a) Threshold of 200 connections. (b) 600 connections. (c) 1000 connections. Axes show spatial coordinates, as in Fig 1.

<https://doi.org/10.1371/journal.pone.0239717.g004>

Unlike the above example, which used square regions to cover a 2D plane, the model analyzed here distributed the centers of regions using a quasirandom sequence so they are approximately uniformly separated without much random clumping [22] to fill a 3D sphere, before each of them is moved in a random direction up to 10% of the region's radius, to avoid forming an almost perfectly regular grid. This gives them an approximately uniform spread throughout the sphere (which approximates the geometry of the brain or a hemisphere), while also allowing for some variation in how they are spread, as seen in experiments. Other geometries can easily be used, such as allocating the regions to cover the surface of a 3D sphere (which would match the surface connections more closely) or on a 2D disk, and these are used in the Results section to determine which geometry best reproduces the data of [8].

For the calculation of the clustering coefficient and average path length, we use the same procedures as in [8]. In particular, if a region is disconnected from the largest connected component of the network, then a finite value for the path length has to be artificially assigned to it for the average path length even to be defined. Here and in [8], such nodes are defined to have an average path length from that node equal to the longest path length observed between a pair of nodes in the largest connected component of the rest of the network. However, because the real brain is a fully connected network, any combination of R (number of regions) and θ (fraction of connections retained after thresholding) values that lead to a disconnected network indicate that the resulting graph is too sparse to provide a valid representation, so any conclusions derived when this is the case are highly questionable.

Results and discussion

In this section, we explain the main predictions made by the model that relate to coarse-graining neuronal networks, specifically relating to how network measures change with the number of regions and the threshold and what quantities are invariant, and compare these predictions with published experimental data.

Clustering coefficient

One of the most common measures used for determining how a network is structured is its clustering coefficient C , which is the probability of two neurons or regions being connected provided they are both connected to a particular third neuron or region. This has been measured for regional networks many times previously [1–3] and is usually used, along with the average path length, to argue that the regions of the brain form a small world network (a network with high clustering but low average path length). For example, [8] used resting-state functional magnetic resonance imaging (fMRI) to measure the interregional covariances of activity at 7 different parcellation scales (corresponding to $R = 84, 91, 230, 438, 890, 1314,$ and 4320). By thresholding their connections at various covariance levels, networks of the interregional connections were obtained.

Motivated by this experiment, we examine how the clustering coefficient in the model networks changes with both threshold θ (the proportion of connections remaining) and number of regions R , as shown in Fig 5a. Each C - θ curve in this figure corresponds to a particular R , averaged over 100 simulations because network details change with the locations of the regions, which are pseudorandomly distributed. The curves approach a limiting form as R increases for the range of thresholds used. However, if R is decreased, the clustering of higher density networks ($\theta > 0.2$) stays approximately the same, while the clustering at lower densities drops significantly. This means that for studies with $\theta < 0.1$, the clustering coefficient obtained from network analysis will depend strongly on R , and this needs to be taken into account when comparing results of different studies.

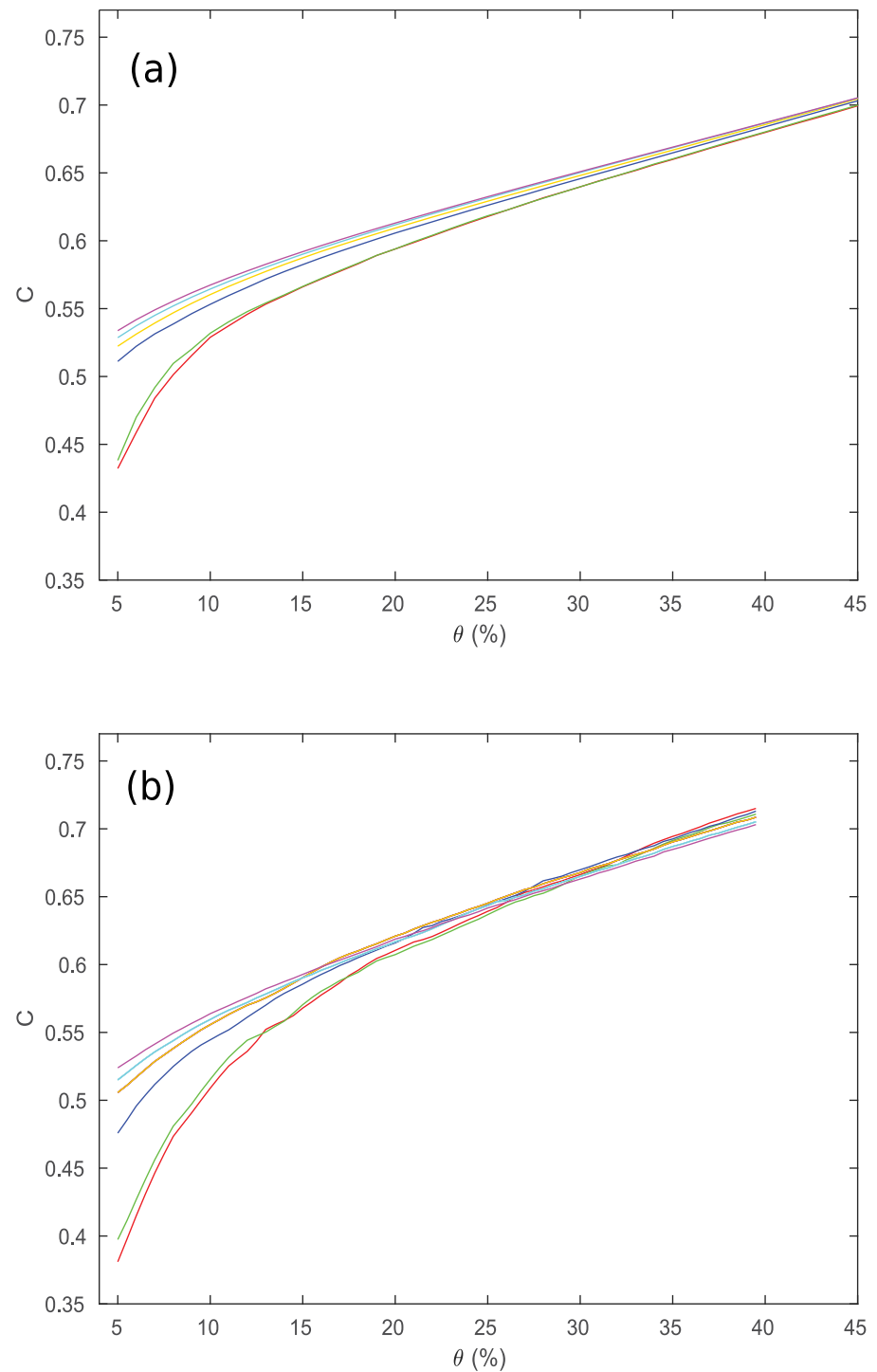


Fig 5. Clustering coefficient vs. θ . C vs. θ for different values of R , where $R = 84$ is red, 91 is green, 230 is blue, 438 is orange, 890 is cyan, 1314 is magenta. (a) Results from the simulated networks. (b) Results from the observed networks, adapted from [8].

<https://doi.org/10.1371/journal.pone.0239717.g005>

We next compare the curves produced by the thresholding procedure with the data from [8], which showed the dependence of the clustering coefficient C on the threshold θ (called ‘Cost’ in their paper); their curves are shown in Fig 5b. The main trends of the curves in both figures are the same, with the clustering being highest at high θ , and gradually decreasing as θ is decreased. Furthermore, the curves in Fig 5b from the thresholded functional networks show the decrease in C for low θ as R decreases, as predicted by our model. The main discrepancy between the two figures is a slight disagreement in exactly how fast C decreases for very low θ values (the difference in values for $\theta > 0.15$ is less than 0.02 for all parcellations) where the network is becoming disconnected and so the network analysis is not valid anyway, but the overall features of the plots are very similar in both figures.

For the previous simulation, we assumed that the regions were spread quasirandomly throughout a 3D sphere. However, it is possible that the regions are distributed differently, such as if they are concentrated on the surface, making the topology more like a 2D spherical shell. We test this hypothesis by adjusting the model to change how the regions are spread throughout the brain volume. Instead of obtaining the C curves seen in Fig 5a, we then obtain Fig 6b if the regions are spread on a 2D disk, and Fig 6c if the regions are spread on a spherical shell. As can be seen by comparing these plots to Fig 5a and 5b, both the surface and disk topologies are significantly worse at predicting the experimental C , with both predicting much higher C at low θ for all R than Fig 5b. This gives strong evidence that the changes in C seen in Fig 5b are partially determined by the geometry of the brain regions as well as by R and θ . We note that the work of [15] obtained a good match to C and L for paths that followed the 2D convoluted cortex through its 3D folding, which matches the present work most closely.

It is also important to note that any single simulated network may have C significantly higher or lower than that predicted by the above plots, which were averaged over many simulations. To get an idea of how much they vary, we simulate 400 networks and then instead of showing only the mean of C for each value of θ , plot the median, values separated by one standard deviation from the median, and values separated by two standard deviations from the median, which are shown in Fig 7. This also shows C over the entire range of θ . This figure demonstrates that although the average of the clustering of the networks is highly constrained, any particular network can vary significantly. For example, only 90% of the networks simulated had C between 0.44 and 0.58 at $\theta = 10\%$. This further helps explain the variance of C observed in various studies, because not only will different values for R and θ significantly change C , but there will be variation beyond this, due to other random factors (such as the distribution of regions chosen). As an example, [2] obtained $C = 0.30$ for their regional network, whereas [3] reported $C = 0.53$ on a different network, which is much higher. However, their values for R were 54 and 90, respectively, and their corresponding θ were 0.073 and 0.10. This explains the discrepancy, as C decreases in Fig 5 as both R and θ decrease. This means that the discrepancies are consistent with being due to the different choices of R and θ , as opposed to actually having detected significantly different networks. Thus this goes beyond the work of [8] by uncovering why C changes with R and θ .

We can explore the changes in C with R and θ further by simulating networks with intermediate values of R as well, and then examining the full 2D surface of C values, which can be seen in Fig 8. With log scales on both axes, it shows that for $\theta > 0.1$, the clustering does not vary much with R , which is supported by Fig 5. However, as θ continues to decrease, the network starts becoming disconnected, making the clustering decrease rapidly to 0. The points at which this begins to happen fall on a straight line in Fig 8, implying that it occurs for a constant value of $R\theta$. This can be checked by examining Fig 9, which shows the clustering vs. R and $R\theta$; we see that C falls rapidly toward zero as $R\theta$ decreases below ≈ 5 , consistent with theory [23]. By checking the size of the largest connected component against $R\theta$ as in Fig 10, it can be

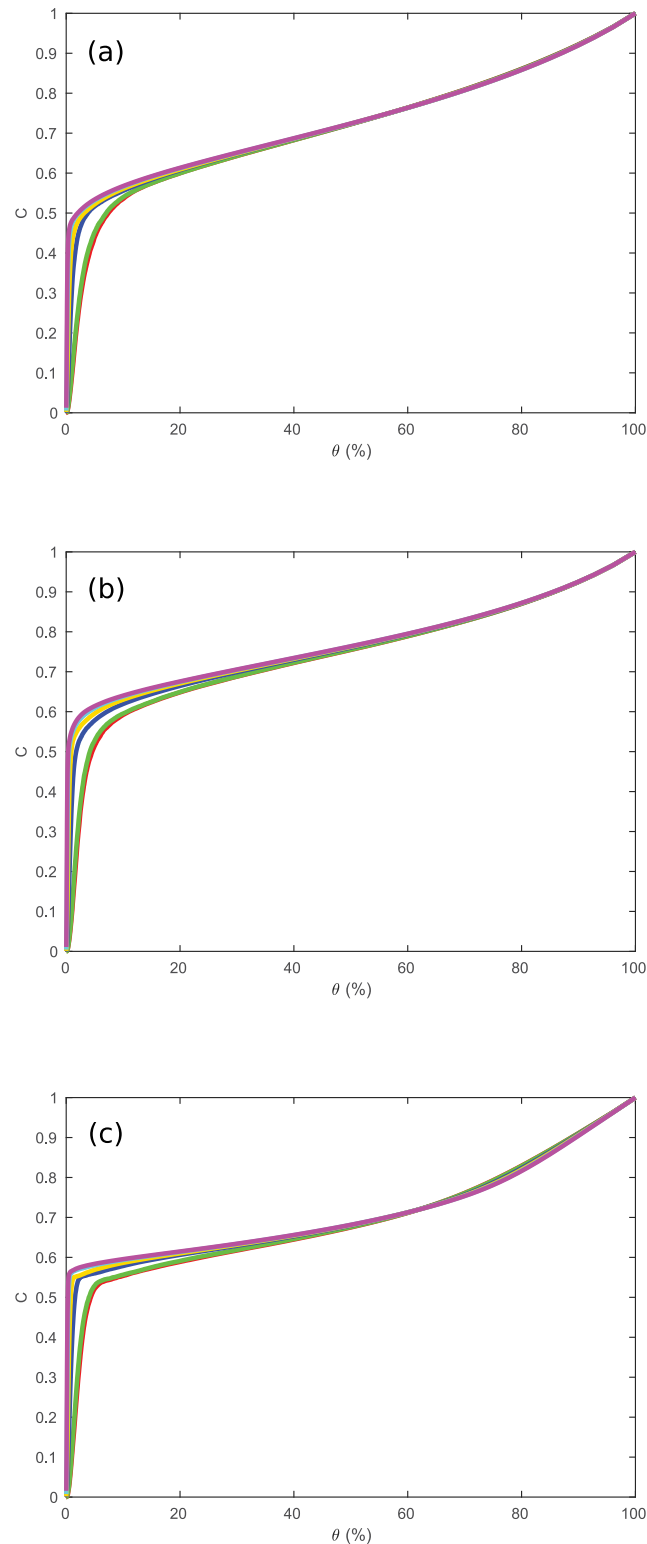


Fig 6. Clustering coefficient vs. θ . Same as Fig 5 for the full range of thresholds, where lines correspond to the number of regions, $R = 84, 91, 230, 438, 890,$ and 1314 from bottom to top. (a) Regions in a sphere. (b) Regions in a disk. (c) Regions on a spherical shell.

<https://doi.org/10.1371/journal.pone.0239717.g006>

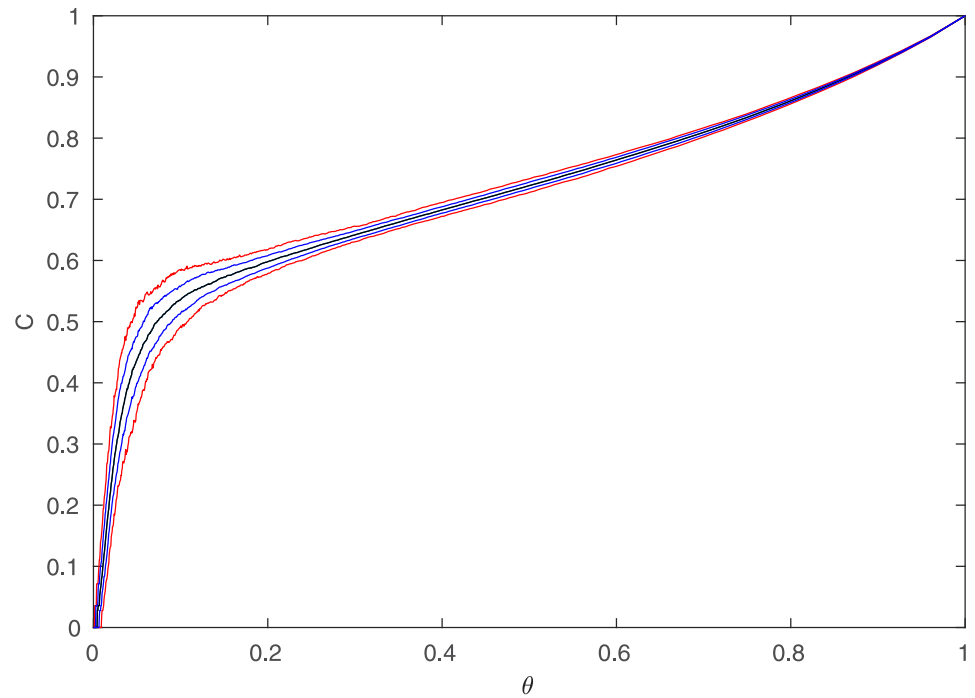


Fig 7. Simulated clustering coefficient vs. θ . The median (black), one standard deviation from median (blue) and two standard deviations from median (red) clustering coefficients vs. threshold from 400 simulated regional networks, with $R = 84$.

<https://doi.org/10.1371/journal.pone.0239717.g007>

confirmed that this is the point at which this component starts rapidly decreasing in size, as large numbers of regions become disconnected from the network and the analysis is no longer relevant to the (fully connected) brain. This all suggests that the different rates of decrease of C seen as θ decreases in Fig 5 are primarily due to the cases with lower R values reaching the $R\theta \approx 5$ threshold faster as θ decreases.

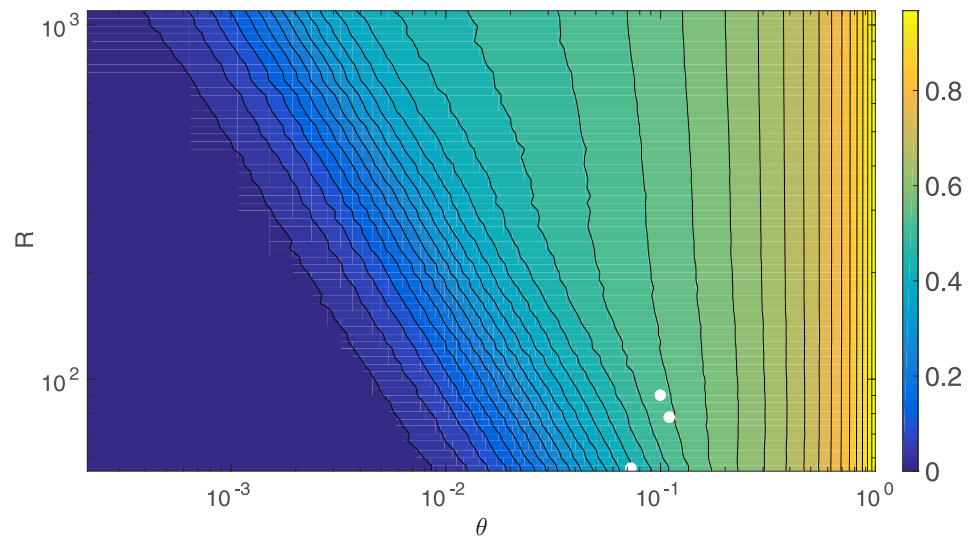


Fig 8. Clustering coefficient vs. R and θ . C vs. R and θ , with log scales on both axes and the color bar indicates the value of C . The white dots show the R and θ values for three selected studies. In order of decreasing R (top to bottom), they are from [3], [1], and [2].

<https://doi.org/10.1371/journal.pone.0239717.g008>

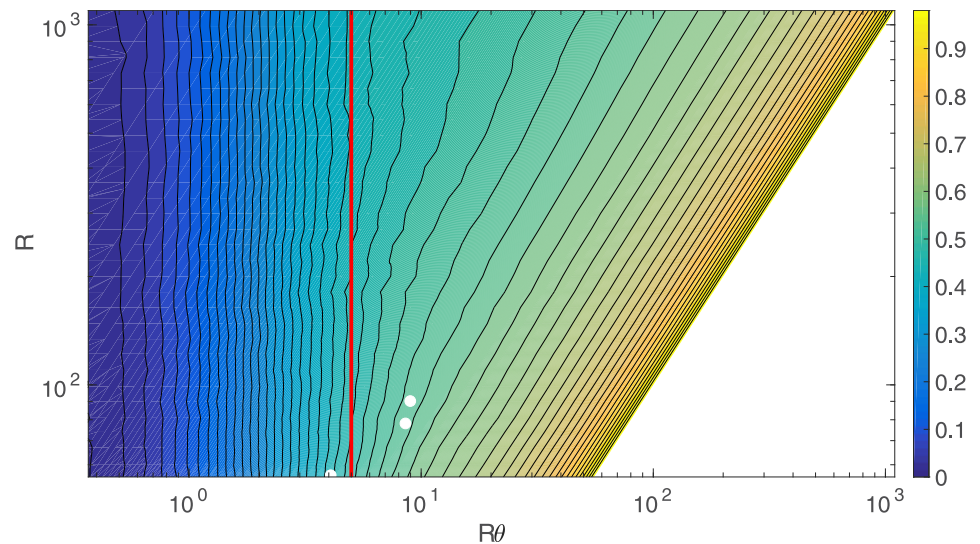


Fig 9. Clustering coefficient vs. R and $R\theta$. C vs. R and $R\theta$, with log scales on both axes and the color bar indicates the value of C . The white dots show the R and $R\theta$ values for three selected studies, and the vertical red line is $R\theta = 5$. In order of decreasing R (top to bottom), they are from [3], [1], and [2].

<https://doi.org/10.1371/journal.pone.0239717.g009>

To compare these predictions of the model with experimental data other than the data from [8], we can overlay some of the R , θ and C values obtained in other experiments over Fig 8, which are the white dots. In order of decreasing R , these are $R = 90$, $\theta = 0.1$, $R\theta = 9$, $C = 0.53$ from [3], $R = 78$, $\theta = 0.11$, $R\theta = 8.58$, $C = 0.49$ from [1], and $R = 54$, $\theta = 0.073$, $R\theta = 3.94$, $C = 0.30$ from [2]. The model's values for C at these points are 0.54, 0.54, and 0.44 respectively. This is a good agreement overall, the higher two values are reasonably close, while the discordant, lower value from [2] is, from the data obtained, below the $R\theta = 5$ boundary, as shown in Fig 9, indicating that the resulting network is significantly disconnected anyway, making any conclusions drawn unreliable. Even for the other two studies, the $R\theta$ value is not that much

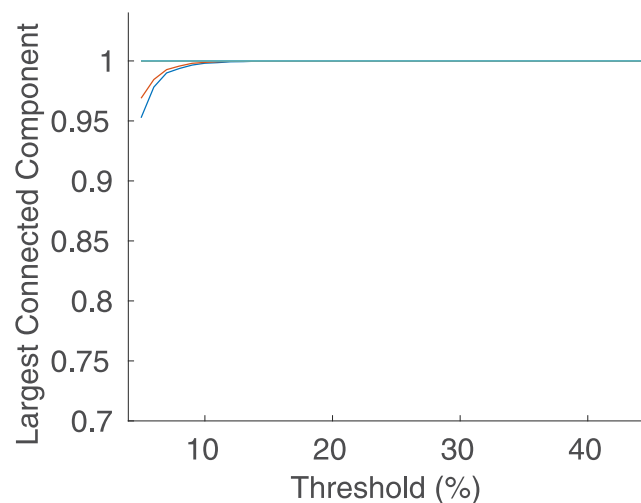


Fig 10. Largest connected component. Size of the largest connected component of the network as a percentage of the entire network vs. θ for the same networks as in Fig 5(a). The bottom-most curve is for $R = 84$, the middle is for $R = 91$, and the upper combines all the cases from with $R \geq 230$, which are indistinguishable on this scale.

<https://doi.org/10.1371/journal.pone.0239717.g010>

higher than the boundary, due to the relatively small R used, demonstrating the general need to use larger networks for these kinds of studies.

In addition, Figs 8 and 9 allow us to model much higher R networks accurately for $\theta > 0.1$, because the C - θ curves approach a limiting curve quickly as R increases for that range, and for $R\theta < 5$, because this is consistently the point at which the network becomes disconnected and the analysis is no longer applicable to real brains. This can be applied to compare the model with [24], which has $R = 10000$, significantly higher than the largest network simulated here. Their Fig 2(c) shows $C \approx 0.50 \pm 0.03$ when $\theta = 0.15$, decreasing to $C \approx 0.44 \pm 0.02$ when $\theta = 0.02$. Estimating C from Fig 8 by extrapolating the trends to higher R gives $C \approx 0.59$ for $\theta = 0.15$ and $C \approx 0.49$ for $\theta = 0.02$, overestimating C by a small amount in each case. Improving the model to account for the actual structure of the brain as opposed to the very simplified one used here would allow for very accurate extrapolation to large R networks.

Path length

Another widely used measure for analyzing brain networks is the average path length between regions, which is the average number of connections required to join a random pair of regions in the network, and this is the other quantity usually used for determining small-worldness of a network. We can modify the above analysis to examine how the model predicts the dependence of average path length L on θ and R . Fig 11a shows L vs. θ for the same values of R as in Fig 5a. This shows that L increases as θ decreases, with it increasing faster with θ for smaller R . This is as expected, because as θ decreases, the number of connections available to link a random pair of regions decreases, and so it will take on average more steps to link them to each other.

We next compare the simulated L with the values found experimentally by [8], which are shown in Fig 11b. The results are seen to agree very well with regards to the rapid increase of L as θ decreases (this comparison is done in more detail below). Furthermore, in both cases, decreasing R gives a much higher L , due to regions becoming disconnected and the parcellation failing. The main difference is that the highest L seen in the model for $R = 84$ is only about 5.3 while in Fig 11b it reaches 6.6 at $\theta = 0.05$. The discrepancy is due to the model disagreeing with the experimental network over the size of the largest connected component and artificial contributions from disconnected components. For such θ values, any conclusions drawn are unreliable, because trying to determine the properties of the completely connected brain by examining a coarse-grained network which is not completely connected is invalid from the outset.

Extending the range of θ , similarly to Fig 6, gives Fig 12. This clearly shows that $L \rightarrow 1$ as $\theta \rightarrow 1$, and also demonstrates that L dramatically increases as $\theta \rightarrow 0$. Low θ values also demonstrate the behavior of L for networks that become disconnected (i.e. not real brain networks), because as the size of the largest connected component approaches 0, $L \rightarrow 0$ from the artificial definition in this case.

We now use the model to explain the behavior of L . To a first approximation, if R is large and θ is small, each region will be connected to all regions within a small sphere around it, with a volume approximately equal to θ times the volume of the brain, because θ is the proportion of connections present; again, to a first approximation, connections beyond this distance will be removed, even though some are functionally important [10]. Hence, because the average distance between randomly selected points within the brain is a constant, L should be proportional to the average distance between two randomly chosen regions divided by the average distance that can be covered each step (D_{step}). Because we have

$$D_{step} = K\theta^{1/3}, \quad (2)$$

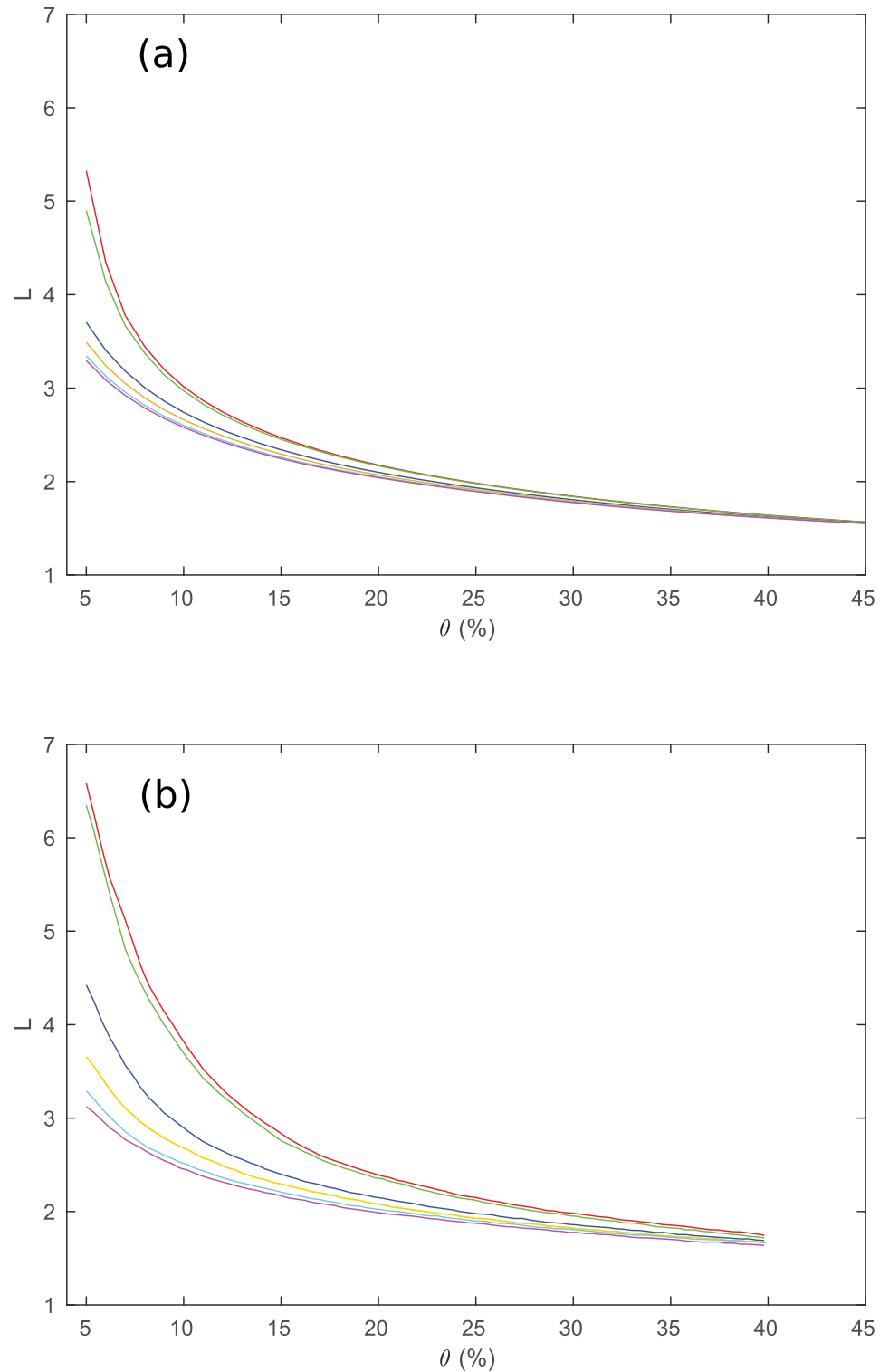


Fig 11. Path length vs. threshold. L vs. θ for different values of R ($R = 84$ is red, 91 is green, 230 is blue, 438 is orange, 890 is cyan, 1314 is magenta). (a) Results from the simulated networks. (b) Results from the networks adapted from [8].

<https://doi.org/10.1371/journal.pone.0239717.g011>

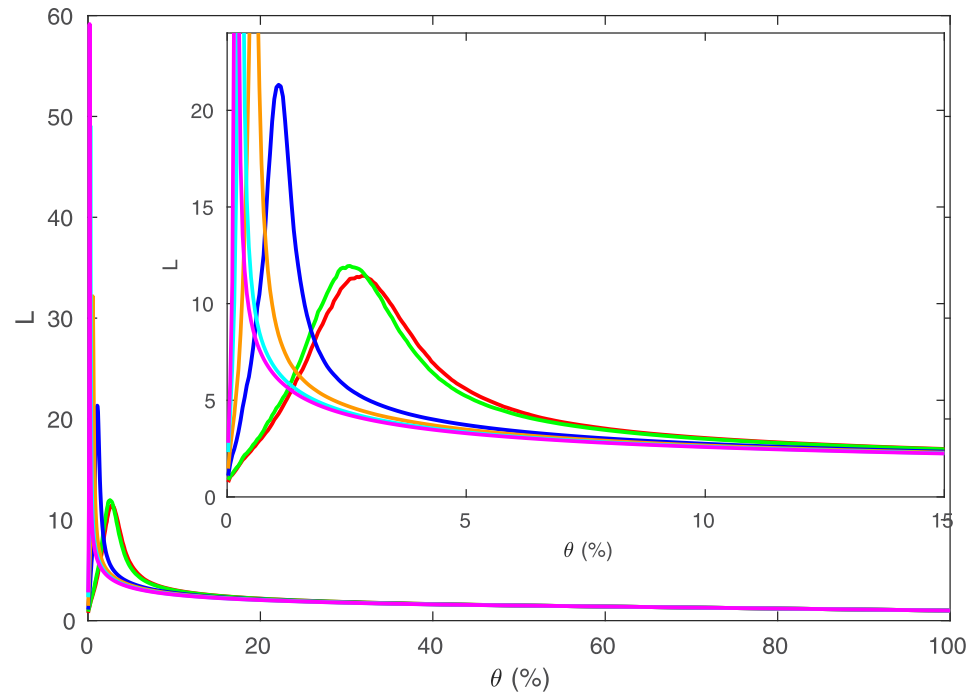


Fig 12. Path length vs. threshold. L vs. θ as in Fig 11a, but extended to cover all θ . The inset shows a closeup for low θ and L .

<https://doi.org/10.1371/journal.pone.0239717.g012>

for a constant K (because the volume is proportional to the cube of the radius), L should obey

$$L^{-3} = A\theta, \tag{3}$$

for a constant A . From the average separation of a random pair of points within a sphere, which is $36R_s/35$ where R_s is the radius of the sphere that represents one brain hemisphere [25] and Eq (3), we have

$$A = \left(\frac{35D_{step}}{36R_s} \right)^3, \tag{4}$$

where D_{step} is the actual average step size between regions. In Fig 13 we show the prediction (3) for the networks used in Fig 11 by plotting L^{-3} against θ , while also testing other geometries as in Fig 6. This shows a clear linear relationship between these two quantities for a spherical geometry (for $\theta > 0.1$, above the level below which the network becomes disconnected and the thresholding becomes invalid), with the line shifting upwards (i.e., A increases) and approaching exact proportionality with increasing R ; curves at small R are downshifted due to the larger fraction of points near the surface where possible paths are more restricted and shorter on average. When the network starts becoming disconnected at low θ , the path length drops suddenly, so L^{-3} diverges. The relationships for the disk and spherical shell geometries differ significantly from the sphere, and are clearly not linear over the range of thresholds shown. However, as mentioned above, [15] reproduced a number of graph measures for a convoluted 2D cortical surface embedded in 3D.

We compare the simulated curves with data from [8] in Fig 14. The curves shift up less with R than in Fig 13a, but the experimental data clearly exhibit the same general features as the model in spherical geometry, and significantly differ from the behavior for disk and spherical

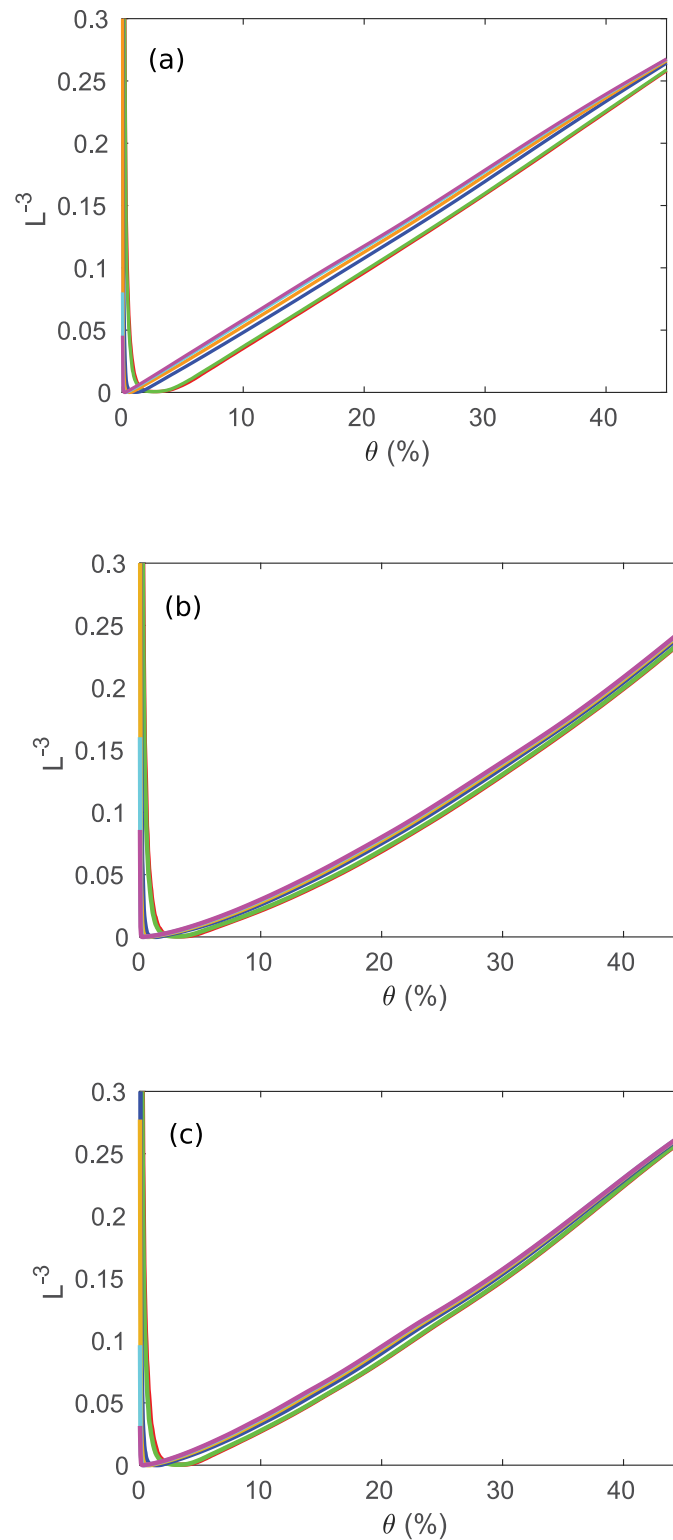


Fig 13. Simulated path length scaling. L^{-3} in the simulated regional network (averaged over 100 networks for each curve) vs. θ for changing R where $R = 84$ is red, 91 is green, 230 is blue, 438 is orange, 890 is cyan, 1314 is magenta. (a) Regions in a sphere. (b) Regions on a disk. (c) Regions on a spherical shell.

<https://doi.org/10.1371/journal.pone.0239717.g013>

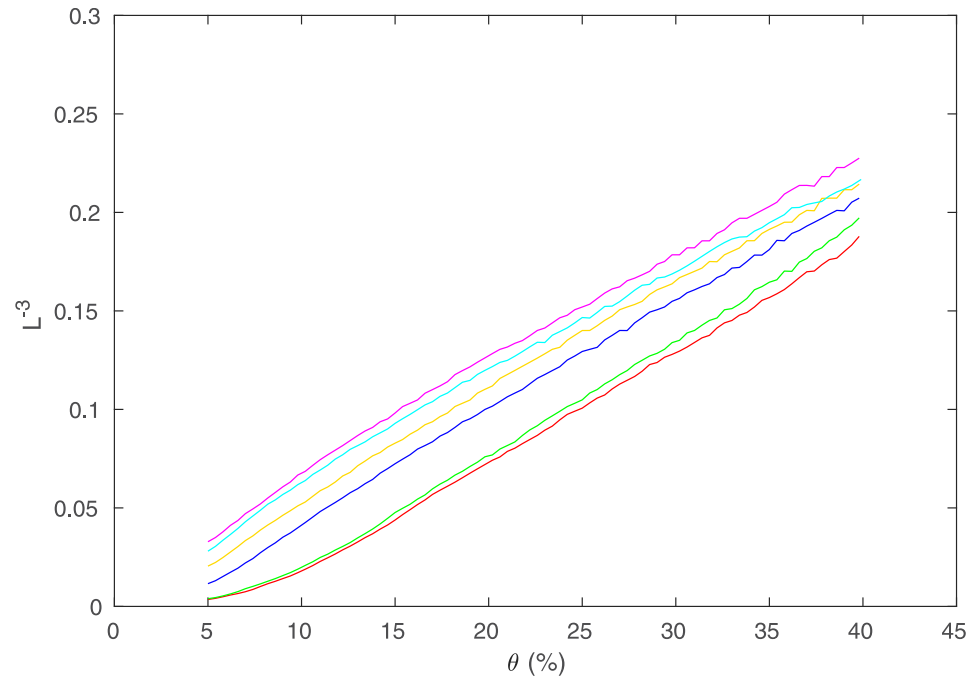


Fig 14. Experimental path length scaling. L^{-3} vs. θ for same data as in Fig 11 for different values of R , $R = 84$ is red, 91 is green, 230 is blue, 438 is orange, 890 is cyan, 1314 is magenta.

<https://doi.org/10.1371/journal.pone.0239717.g014>

shell geometries. This strongly indicates that the behavior of L vs. R and θ in [8] is mainly due to how the regions are arranged in space, reinforcing the conclusion from the behavior of C as well.

Finally, we estimate how closely the measured and predicted values of A from Eq (3) match. Taking the slope of the $R = 1314$ lines for both the model and the data give $A = 0.57 \pm 0.03$. This means that for the model and data that $D_{step} \approx 0.86R_s$, where the fact that D_{step} is smaller than the mean $36R_s/35$ in the continuum limit is because connections between distant regions cannot always follow straight lines, due to other discrete regions not being present exactly on the straight line path in some cases. This forces paths to detour via more, smaller steps. This also suggests that D_{step} will increase toward $36R_s/35$ with R , because higher R results in higher numbers of discretized points lying close to this line. The scaling seen in Fig 14 can be used to interrelate results obtained for different θ , especially at large R where D_{step} is approximately constant and the curves approach a limiting form.

Small-worldness

The original measure of whether a graph is small-world was

$$\sigma = \frac{\gamma}{\lambda} = \frac{C/C_{rand}}{L/L_{rand}}, \tag{5}$$

where C_{rand} and L_{rand} are respectively the clustering and average path length for a random graph of the same size and connection density [5]; $\sigma > 1$ implies small-world characteristics. However, [26] argued that σ was overly sensitive to changes in C_{rand} , and that larger networks tend to have larger σ , thereby making comparisons more difficult. Hence, they introduced an

alternative measure ω ,

$$\omega = \frac{L_{rand}}{L} - \frac{C}{C_{lat}}, \tag{6}$$

where C_{lat} is the clustering coefficient of an equivalent lattice graph, which is obtained by repeatedly swapping edges between nodes while preserving the node degrees until C reaches a maximum. Here, we will assume that the equivalent (same degree and number of nodes) lattice graph is topologically a 1D ring lattice, as discussed in [5], where the nodes are arranged in a ring and connected to their closest neighbors, which will be the case as long as all the regions have approximately the same degree. However, this is not spatially embedded, unlike the simulated brain networks.

We can examine how σ and ω change for the network of regions as θ varies, as shown in Figs 15 and 16. Fig 15 shows that when less than 10% of connections are present, the value of σ is significantly above 1, indicating that in this region, the network would be considered to be small-world according to this measure. As $\theta \rightarrow 0$, σ dramatically increases to at least 20 in all parcellations, with $\sigma > 600$ for $R = 1314$. When the regions start disconnecting as $\theta \rightarrow 0$, C decreases faster than L , thereby decreasing σ in the process, but disconnection makes these thresholded parcellations invalid.

In [26] it was argued that $|\omega| \leq 0.5$ should be taken to indicate a small-world graph, with $\omega > 0.5$ indicating a random graph, and $\omega < -0.5$ indicating a lattice graph. Fig 16 shows that $|\omega| \leq 0.1$ for $0.05 < \theta < 0.6$, so that networks at these values for R and θ will always appear to be small-world graphs, and θ smaller than this leads to a disconnected network for most R and the results are thus not relevant to real brains.

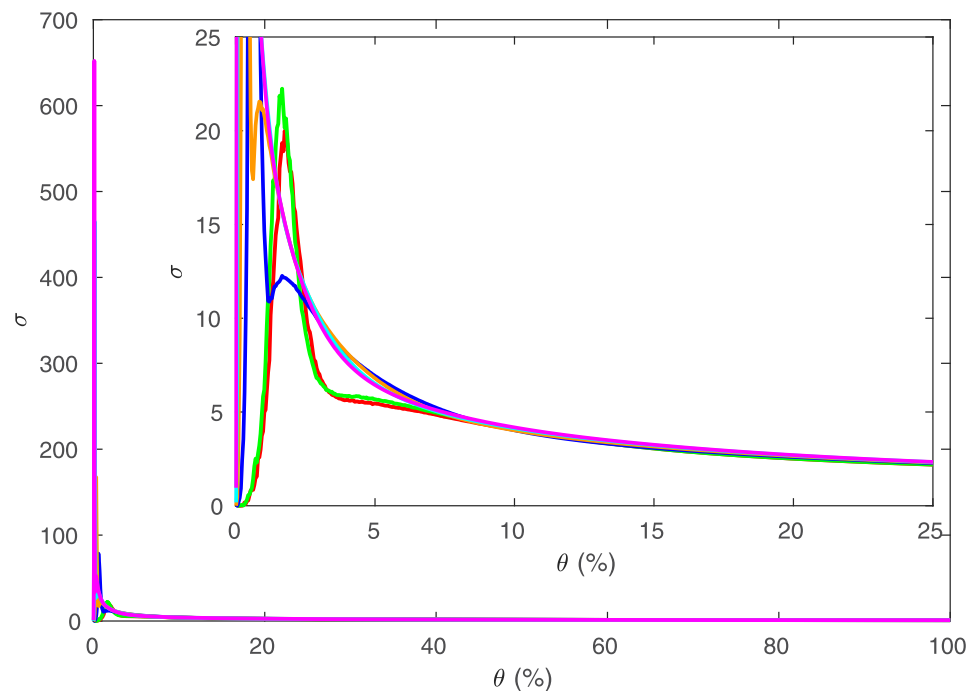


Fig 15. Small-worldness vs. threshold. σ vs. θ for the same data as Fig 6, where $R = 84$ is red, 91 is green, 230 is blue, 438 is orange, 890 is cyan, 1314 is magenta. Inset is zoomed to show low θ .

<https://doi.org/10.1371/journal.pone.0239717.g015>

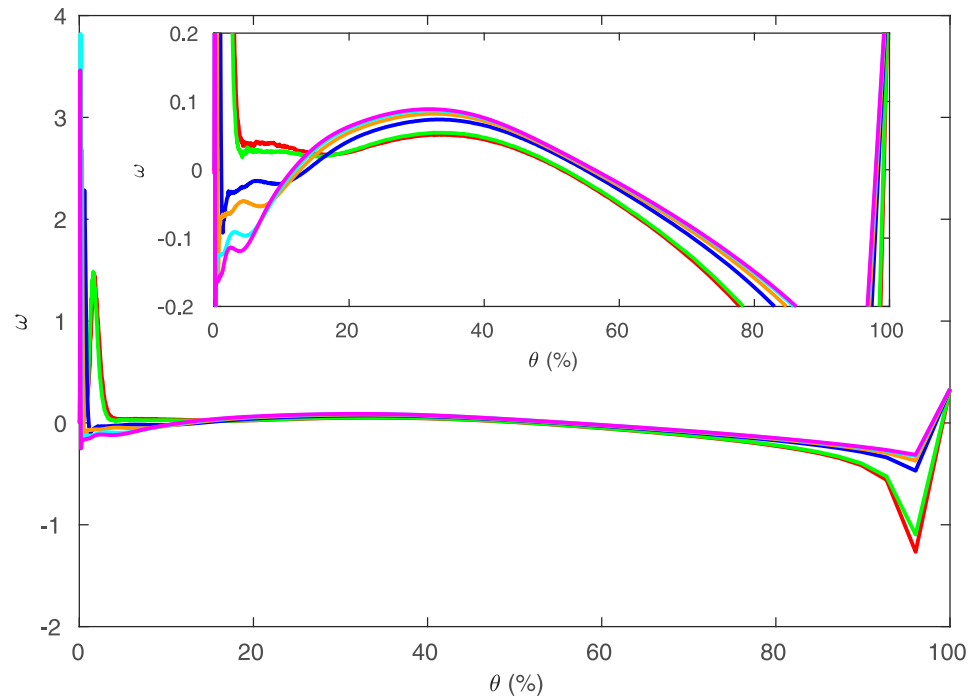


Fig 16. Small-worldness vs. threshold. ω vs. θ for the same data as Fig 6, where $R = 84$ is red, 91 is green, 230 is blue, 438 is orange, 890 is cyan, 1314 is magenta. Inset is zoomed to show low $|\omega|$.

<https://doi.org/10.1371/journal.pone.0239717.g016>

The networks here are not small-world in the usual sense, because the simulated connections between regions are based on proximity and thresholding means that long-range connections are removed and only nearby regions are connected; long-range connections are usually thought to be required in small-world networks, to reduce L . The simulated networks' large σ and small $|\omega|$ values (which we have also checked for fine-scale networks) are at least partially due to L being much lower here than typical comparison networks, because the real network is embedded in 3D space. Hence L only increases as $R^{1/3}$, in contrast to a typical lattice comparison network such as a ring lattice in which $L \propto R$, making the brain network appear small-world due to the dimensionality of the network. This highlights the misleading inferences that can be obtained by comparing the brain to graphs that are not spatially embedded [10, 11, 14, 15, 17, 27].

In addition, the small-worldness of the coarse-grained network of the brain doesn't provide much information about the network structure at finer resolutions. When the thresholding is done, the majority of long range neuronal connections end up being ignored because the much more numerous short range connections dominate the connectivity between regions, leading to important connections between distant regions being ignored [10]. Hence, it is not possible to determine whether the neuronal network is small-world or not from studying the usual thresholded regional networks, because it is not possible to determine how many long range connections between neurons are present in the real brain by examining the thresholded network, from which they have been almost entirely deleted. However, [10] and [28] showed that an alternate method is to keep the strongest edges at each distance, not just the strongest overall. This preserves important long distance connections between regions, and helps mitigate the effects of the spatial embedding on the network measures.

Summary and conclusion

We have studied the dependence of widely used graph properties of regional-level brain networks on parcellation scale R , threshold θ , and geometry by simulating the underlying neuronal network from a physical starting point and comparing with data. The main results are:

- (i) The simple physically based model produces networks of regions based solely on the distances between pairs of regions by assuming that neurons have a connection probability that exponentially decays with distance. When thresholded, this results in regions always connecting to their closest neighbors before any other regions. Despite the simplicity of these assumptions, the properties of these networks are very similar to the properties observed experimentally after thresholding experimental networks obtained via structural or functional measurements.
- (ii) The dependence of network properties such as the clustering coefficient C and average path length L on R and θ are accurately reproduced by the model, with only small discrepancies. This allows the results of different studies to be interrelated and discrepancies resolved, because many use only a single set of R and θ values and end up with different network properties. The ability to accurately simulate the network properties means that results for R and θ values other than those that have already been tested experimentally can be predicted, which is especially important for large R values, which are becoming increasingly relevant as experimental methods improve.
- (iii) The density of the final network can be freely set anywhere between completely connected and completely disconnected by varying θ . This means that information is lost by thresholding at any level and the information on variable connection strengths between regions should be retained. However, this also means that new measures must be used in order to quantify connectivity, because some existing ones such as L do not generalize directly to variable connection strengths.
- (iv) If $R\theta < 5$, the coarse-grained network becomes disconnected. Thus because the real brain is completely connected, all conclusions drawn in this region are unreliable and cannot be trusted because the network has been made too sparse to represent this fundamental aspect of the brain. Thus, care should be taken to not threshold any regional networks too strongly by choosing a very small θ , and the criterion $R\theta > 5$ for connectedness must be tested and satisfied for any study to be valid.
- (v) The typical measures used to examine regional networks of the brain, including C and L , are not robust to changes in either θ or R , varying rapidly when either R or θ is small. This means that care needs to be taken when comparing the measures between different studies, to make sure that the variations due to R and θ are not changing the conclusions. Alternatively, they can use other measures that are more robust to changes than C or L , such as $L^3\theta$, so that these measures can be used for comparison between experiments instead.
- (vi) The measures σ and ω both indicate that the simulated networks are heavily small-world, despite not having any of the typical properties of a small-world network. The apparent small-worldness of these regional networks is mainly due to the spatial embedding of the regions in the brain into 3D space, because this embedding means that $L \propto R^{1/3}$, so the regions appear to form a small-world network even without any long range connections. Thus, when determining whether a network is small world or not, the reference networks it is compared to should be spatially embedded in the same way, so as to not let the spatial effects invalidate the comparison.

(vii) The appropriateness of using thresholded networks to study the brain's neuronal network is highly questionable. Conclusions about the network of regions, such as whether it is a small-world network, do not necessarily apply to the network of neurons, so conclusions about the architecture of the brain cannot be drawn from these measurements directly. Hence, when attempting to determine how the underlying neuronal network is structured, it is very important to understand how the structure of the neuronal network influences the thresholded regional network obtained. The network of regions should thus be seen as a simplified and transformed version of the network to be examined, not as the most fundamental object of interest, although its properties may still be relevant to some applications. However, methods such as those in [10] allow for a more robust method of thresholding which is not dominated by geometric effects, decreasing their influence on the results.

Overall the above results imply that standard thresholded network studies chiefly detect the properties of a near-neighbor network in 3D space.

Author Contributions

Conceptualization: P. A. Robinson.

Formal analysis: T. C. Lacy.

Funding acquisition: P. A. Robinson.

Investigation: T. C. Lacy, P. A. Robinson.

Methodology: P. A. Robinson.

Project administration: P. A. Robinson.

Resources: P. A. Robinson.

Software: T. C. Lacy.

Supervision: P. A. Robinson.

Writing – original draft: T. C. Lacy.

Writing – review & editing: P. A. Robinson.

References

1. Gong G, He Y, Concha L, Lebel C, Gross DW, Evans AC, et al. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb Cortex*. 2009; 19(3):524–536. <https://doi.org/10.1093/cercor/bhn102> PMID: 18567609
2. He Y, Chen ZJ, Evans AC. Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb Cortex*. 2007; 17(10):2407–2419. <https://doi.org/10.1093/cercor/bhl149> PMID: 17204824
3. Achard S, Salvador R, Whitcher B, Suckling J, Bullmore ET. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J Neurosci*. 2006; 26(1):63–72. <https://doi.org/10.1523/JNEUROSCI.3874-05.2006> PMID: 16399673
4. Bullmore ET, Sporns O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Rev Neurosci*. 2009; 10(3):186–198. <https://doi.org/10.1038/nrn2575> PMID: 19190637
5. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998; 393(6684):440–442. <https://doi.org/10.1038/30918> PMID: 9623998
6. Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, Bullmore ET. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb Cortex*. 2005; 15(9):1332–1342. <https://doi.org/10.1093/cercor/bhi016> PMID: 15635061

7. Iturria-Medina Y, Sotero RC, Canales-Rodriguez EJ, Alemán-Gómez Y, Melie-García L. Studying the human brain anatomical network via diffusion-weighted MRI and graph theory. *NeuroImage*. 2008; 40(3):1064–1076. <https://doi.org/10.1016/j.neuroimage.2007.10.060> PMID: 18272400
8. Fornito A, Zalesky A, Bullmore ET. Network scaling effects in graph analytic studies of human resting-state fMRI data. *Front Syst Neurosci*. 2010; 4:1197–1207. <https://doi.org/10.3389/fnsys.2010.00022> PMID: 20592949
9. Roberts JA, Perry A, Lord AR, Roberts G, Mitchell PB, Smith RE, et al. The contribution of geometry to the human connectome. *NeuroImage*. 2016; 124:379–393. <https://doi.org/10.1016/j.neuroimage.2015.09.009> PMID: 26364864
10. Roberts JA, Perry A, Roberts G, Mitchell PB, Breakspear M. Consistency-based thresholding of the human connectome. *NeuroImage*. 2017; 145:118–129. <https://doi.org/10.1016/j.neuroimage.2016.09.053> PMID: 27666386
11. Robinson PA. Physical brain connectomics. *Phys Rev E*. 2019; 99:012421. <https://doi.org/10.1103/PhysRevE.99.012421> PMID: 30780309
12. Nunez PL. *Neocortical dynamics and human EEG rhythms*. New York: Oxford University Press. 1995.
13. Barthélemy M. Spatial networks. *Phys Rep*. 2011; 499(1):1–101. <https://doi.org/10.1016/j.physrep.2010.11.002>
14. Henderson JA, Robinson PA. Using geometry to uncover the relationships between isotropy, homogeneity, and modularity in cortical connectivity. *Brain Conn*. 2013; 3:423–437. <https://doi.org/10.1089/brain.2013.0151> PMID: 23802922
15. Henderson JA, Robinson PA. Relations between the geometry of cortical gyrification and white-matter network architecture. *Brain Conn*. 2014; 4:112–130. <https://doi.org/10.1089/brain.2013.0183> PMID: 24437717
16. Barthélemy M. Crossover from scale-free to spatial networks. *Europhys Lett*. 2003; 63(6):915. <https://doi.org/10.1209/epl/i2003-00600-6>
17. Samu D, Seth AK, Nowotny T. Influence of wiring cost on the large-scale architecture of human cortical connectivity. *PLoS Comput Biol*. 2014; 10:e1003557. <https://doi.org/10.1371/journal.pcbi.1003557> PMID: 24699277
18. Hilgetag CC, Goulas A. Is the brain really a small-world network? *Brain Struct Funct*. 2016; 221:2361–2366. <https://doi.org/10.1007/s00429-015-1035-6> PMID: 25894630
19. Bassett DS, Bullmore ET. Small-world brain networks revisited. *The Neuroscientist*. 2017; 23:499–516. <https://doi.org/10.1177/1073858416667720> PMID: 27655008
20. Muldoon SF, Bridgeford EW, Bassett DS. Small-world propensity and weighted brain networks. *Sci Rep*. 2016; 6:22057. <https://doi.org/10.1038/srep22057> PMID: 26912196
21. Vértes PE, Alexander-Bloch AF, Gogtay N, Giedd JN, Rapoport JL, Bullmore ET. Simple models of human brain functional networks. *PNAS* 2012; 109(15):5868–5873. <https://doi.org/10.1073/pnas.1111738109> PMID: 22467830
22. Train K. Halton sequences for mixed logit. Department of Economics, University of California at Berkeley. 2000.
23. Bollobás B. *Random Graphs*. London: Academic Press; 1995.
24. van den Heuvel MP, Stam CJ, Boersma M, Pol HH. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *NeuroImage*. 2008; 43(3):528–539. <https://doi.org/10.1016/j.neuroimage.2008.08.010> PMID: 18786642
25. Burgstaller B, Pillichshammer F. The average distance between two points. *Bull Aust Math Soc*. 2009; 80(3):353. <https://doi.org/10.1017/S0004972709000707>
26. Telesford QK, Joyce KE, Hayasaka S, Burdette JH, Laurienti PJ. The ubiquity of small-world networks. *Brain Conn*. 2011; 1(5):367–375. <https://doi.org/10.1089/brain.2011.0038> PMID: 22432451
27. Henderson JA, Robinson PA. Geometric effects on complex network structure in the cortex. *Phys Rev Lett*. 2011; 107(1):018102. <https://doi.org/10.1103/PhysRevLett.107.018102> PMID: 21797575
28. Betzel RF, Medaglia JD, Papadopoulos L, Baum GL, Gur R, Gur R, et al. The modular organization of human anatomical brain networks: Accounting for the cost of wiring. *Network Neurosci*. 2017; 1:42–68. https://doi.org/10.1162/NETN_a_00002 PMID: 30793069