



Early path dominance as a principle for neurodevelopment

Rostam M. Razban^a, Jonathan Asher Pachter^{a,b}, Ken A. Dill^{a,b,c,1}, and Lilianne R. Mujica-Parodi^{a,b,d,e,f,1}

Contributed by Ken A. Dill; received October 26, 2022; accepted March 15, 2023; reviewed by Michael Breakspear, Marcus Kaiser, and Amina Qutub

We perform targeted attack, a systematic computational unlinking of the network, to analyze its effects on global communication across the brain network through its giant cluster. Across diffusion magnetic resonance images from individuals in the UK Biobank, Adolescent Brain Cognitive Development Study and Developing Human Connectome Project, we find that targeted attack procedures on increasing white matter tract lengths and densities are remarkably invariant to aging and disease. Time-reversing the attack computation suggests a mechanism for how brains develop, for which we derive an analytical equation using percolation theory. Based on a close match between theory and experiment, our results demonstrate that tracts are limited to emanate from regions already in the giant cluster and tracts that appear earliest in neurodevelopment are those that become the longest and densest.

percolation theory | dMRI | connectomics | statistical mechanics | network neuroscience

Brains are networks of neuronal regions (nodes) that are linked together by bundles of axons (edges). Much more is known about static topologies of brain networks than of the dynamics of how topologies emerge through neurodevelopment. Known topological properties include their degree distributions, path lengths, clustering coefficients, and rich-club coefficients (1–9). These graph-theoretic features can be captured by modeling and compared to other known types of networks. For example, the high average clustering coefficients and low average path lengths in brains are similar to what is found in a lattice of sites with introduced random connections, called small-world networks (1, 5).

However, these topological properties of brains are often considered as snapshots at a given time (10). Our interest here is in the developmental trajectories through which brains come to have its unique topology. There are no publicly available datasets yet rich enough to give the full trajectories of topologies of developing brains*. Thus, as an alternative, we use data to construct a generative model of a quasidynamical sequence of events of developing brain networks. We do this using statistical mechanical percolation theory with a procedure called targeted attack (13–15).

Targeted Attack Analysis of Brain Data

Targeted attack is a computational procedure that can be performed on any network having a known topology. Targeted attack entails sequential removal in silico of nodes or edges based on the rank order of some node or edge property (13, 14, 16), followed by analysis of how the network changes step-by-step throughout the decimation of the net by the attack. Consider the metaphor of a city's roads. One metric is road lengths: You might remove them in order of shortest to longest, for example. By a different metric, you might remove links in order of how many cities they connect. The nature of how a network topology diminishes throughout the attack can give insights about the network beyond what single network measures can give. In particular, we focus on the giant cluster (the largest group of connected nodes) because it is a proxy for global communication. We compute the probability P that an arbitrary node in the network is in the giant cluster. If $P = 1$, then any arbitrary region can communicate with any other region (17, 18).

Various targeted attack analyses have previously been performed on brain network data (2, 19–24). It has been found that when attack sequences are based on correlations of functional connectivities, e.g., functional MRI (fMRI) signals between regions, the brain does not decimate in the same way as random or scale-free graphs (2). The attack procedure has also been applied to structural connectivities, e.g., diffusion MRI (dMRI), which measures the topology of large bundles of axons known as white matter tracts (25). This study removed nodes based on their degrees (26), motivated by the logic that their targeted attack procedure directly simulates

Significance

As brains develop through neural growth and specialization, what mechanism ensures that new neurons are integrated into a fully connected brain, avoiding “bridges to nowhere”? Here, we study brain structure development from the perspective of percolation, a global measure of communication. Analyzing over 35,000 diffusion MRI scans on human individuals, from newborns to adults, we identify the following rules of brain neurogenesis through percolation theory: Earlier tracts become longer and denser while maintaining a giant cluster. This signature, invariant to age or mental health, suggests a fundamental condition for the brain to function as an emergent whole.

Author contributions: R.M.R., J.A.P., K.A.D., and L.R.M.-P. designed research; R.M.R. and J.A.P. performed research; J.A.P. contributed new reagents/analytic tools; R.M.R., J.A.P., K.A.D., and L.R.M.-P. analyzed data; and R.M.R., J.A.P., K.A.D., and L.R.M.-P. wrote the paper.

Reviewers: M.B., The University of Newcastle - Newcastle City Campus; M.K., University of Nottingham; and A.Q., University of Texas at San Antonio.

The authors declare no competing interest.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹To whom correspondence may be addressed. Email: dill@laufcenter.org or mujica@lcnuro.org.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218007120/-/DCSupplemental>.

Published April 13, 2023.

*Recent brain scans of fetuses in the womb are promising but not yet publicly accessible (11, 12).

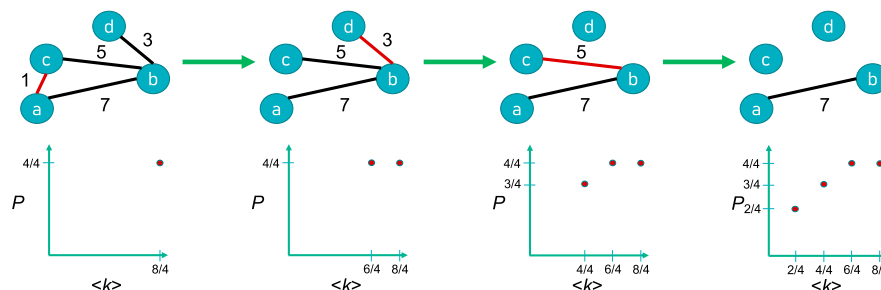


Fig. 1. The targeted attack procedure on increasing tract lengths. By starting with all edges present on the *Left*, we sequentially remove those edges whose length (listed above the edge) are currently shortest (colored in red). P is the probability a node is in the giant cluster; $\langle k \rangle$ is the average degree.

neurodegeneration, which is more likely to be located in metabolically costly hubs (27).

Here, we seek insights through the following aspects: 1) two new decimation strategies: tract lengths and densities; and 2) an analytical theory for time-reversing targeted attack, which can shed light on how the brain network builds up in this process. The logic is that time-reversal of the decimation process gives a plausible hypothesis for how brain structure emerges through early development, which we validate on $N=35,731$ human dMRI scans of increasing targeted attack of tract length and density acquired from the UK Biobank (28), the Adolescent Brain Cognitive Development (ABCD) Study (29) and the Developing Human Connectome Project (dHCP) (30). We compare brain results to alternative graph theoretical structures, such as random, scale-free, and small-world graphs, as well as a graph constrained by brain volume. And, to confirm that our conclusions are not specific to the dMRI data or human brains, we replicate our results on mouse viral tracing experiments (31).

The Computational Procedure of Targeted Attack. At a given step of decimation, the various nodes of the remaining graph will have different degrees of connectivity (numbers of other nodes to which it connects). At that stage of decimation, the graph's average connectivity will be $\langle k \rangle$. The calculation we make here is of $P = P(\langle k \rangle)$. For example, node c in Fig. 1 initially has a degree (k) of 2 in the left-most realization of the graph "abcd." $\langle k \rangle$ is the average over all nodes' degrees. P is calculated by the number of nodes in the giant cluster divided by the total number of nodes. The left-most realization of the graph "abcd" first has $P = 4/4$, then sequential attack from small to large tract lengths results in $P = 4/4$, $P = 3/4$, and $P = 2/4$ (Fig. 1).

We run a targeted attack procedure on two main outputs from a generic dMRI data analysis: tract length and tract density (Fig. 1). Tract density, also known as streamline count, is the number of tracts connecting two gray matter regions; tract length is the average length of those connections (*Methods*). Since we can remove edges in increasing (smallest to largest) or decreasing (largest to smallest) order, each physical property has two possible P curves. Fig. 2 demonstrates that the removal of increasing tract lengths and tract densities yield P curves qualitatively different from those corresponding to random graphs. In contrast, decreasing tract density removal results in a poorly resolved curve because a high proportion of edges have a tract density of 1 (*SI Appendix, Fig. S1, Right*). Decreasing tract length removal results in a P curve qualitatively similar to that of random graphs (*SI Appendix, Fig. S1, Left*), consistent with previous results (26). Here, we analyze the shapes of the curves of $P(\langle k \rangle)$ from increasing tract length and tract density-targeted attack and compare to other known types of networks.

Analytical Theory for Brain Network Formation. The attack procedure describes the P curve from right to left, by a breakdown process. Instead, looking at the same P curve now from left to right describes a hypothetical build-up process that provides a growth trajectory on network development from early to late. One well-known trajectory in percolation theory is a random network. Given in terms of the Lambert W function (15), an analytical solution is known for $P(\langle k \rangle)$ in the limit of infinite numbers of nodes (15, 32).

$$P(\langle k \rangle) = 1 + \frac{W(-\langle k \rangle e^{-\langle k \rangle})}{\langle k \rangle}. \quad [1]$$

We are not aware of any other known networks for which $P(\langle k \rangle)$ is fully derived[†]. Here, we analytically characterize a mechanism for growth of the giant cluster by restricting secondary cluster formation, based on experimental evidence discussed in the next section. Secondary clusters are defined as other clusters besides the giant cluster of size greater than one. We compute the trajectory $P(\langle k \rangle)$ of what we call the Giant Cluster Self Preference network over its full range of $\langle k \rangle$, as (derived in *Methods*):

$$P(\langle k \rangle) = 1 + \left(\frac{1}{1 - 2/\alpha} \right) W \left[\left(\frac{2}{\alpha} - 1 \right) e^{-(1-2/\alpha)} e^{-\langle k \rangle/\alpha} \right]. \quad [2]$$

Eq. 2 contains only one parameter, α , which we fit to the experimental data. The parameter reflects how many times more likely a new link lands in the giant cluster, relative to landing in a newly formed isolated node during the process of brain growth in early development.

Results

Attack Trajectories Depend on Tract Lengths and Densities.

Fig. 2 demonstrates that for both tract length and tract density, random graph theory is not consistent with the experimental data. Experiments show that $P(\langle k \rangle)$ changes much more gradually all the way down to $\langle k \rangle = 0$ at the left end of the figure. In one perspective, this might be surprising. Percolation behaviors often have steep cliffs: no global communication below a certain critical threshold, then a jump to a finite value of communication above it (14, 15). These two curves are for one adult individual; *SI Appendix, Fig. S2* shows that other adult brains have similar features.

We then probe more deeply into the difference of the brain network vs. random graphs. At $\langle k \rangle = 1$, the giant component

[†]Limits are known for many graphs such as scale-free and lattices; however, the full trajectory from small to large $\langle k \rangle$ remains analytically unsolved (14, 15).

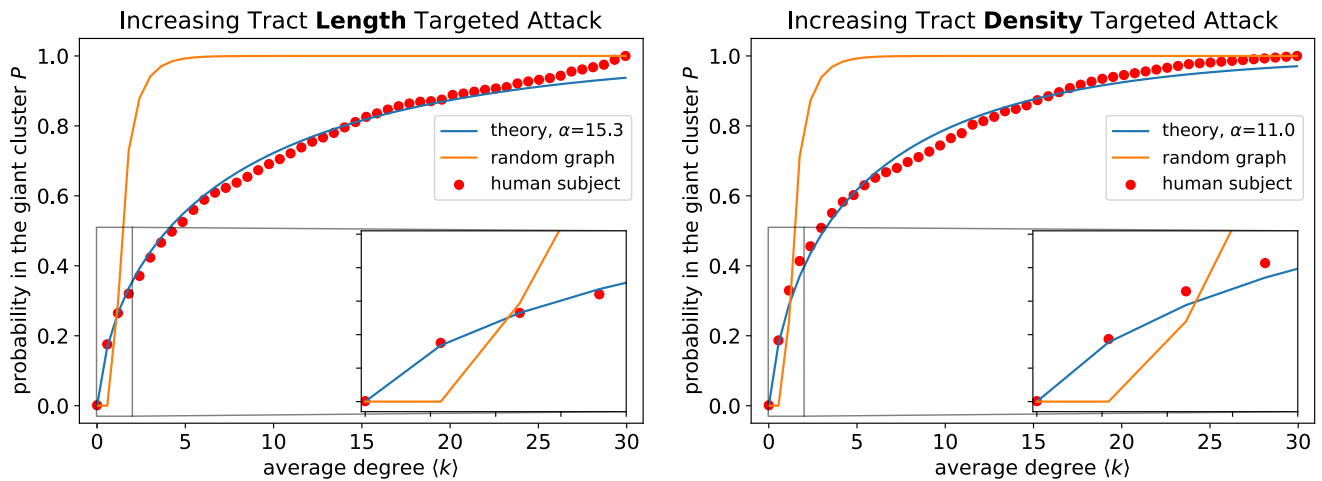


Fig. 2. Attack curves, $P(\langle k \rangle)$, theory vs. experiments. (Red points) dMRI data collected by the UK Biobank for one arbitrarily chosen human individual (subject ID: 6025360) (33) under the Talairach atlas (34). (Blue line) The presented Giant Cluster Self Preference theory. (Orange line) Prediction from random graph theory. The *Inset* zooms in on the first four equidistant sampled experimental points to highlight the lack of a sharp transition.

in the brain is essentially the only cluster; nodes not in the giant cluster are isolated and do not form secondary clusters. In contrast for the random graph at $\langle k \rangle = 1$, the giant cluster is just one of many nonnegligible clusters present (*SI Appendix, Fig. S3*). The lack of secondary clusters in the brain network holds across different values of $\langle k \rangle$ (*SI Appendix, Fig. S4*). We conclude that in this particular computational attack, the brain's giant cluster continuously degrades one node at a time, rather than by sharp fragmentation into secondary clusters from one large cluster. This lack of fragmentation explains the lack of a critical point in the P curves. We encode this mechanism into our Giant Cluster Self Preference theory (Eq. 2) and find good agreement of the full attack curves $P(\langle k \rangle)$, for both tract length and tract density attack variables (Fig. 2).

To rule out other potential mechanisms, we perform simulations of a network build-up by preferential attachment (35). We find a match with experimental P curves at a coarser parcellation (*SI Appendix, Fig. S5*). Scale-free networks built by preferential attachment are known to lack critical points during random attack (14, 16, 19). However, the match fails to extend to finer parcellations (*SI Appendix, Fig. S6*) even when considering general preferential attachment models where edge addition is not simply linear in nodes' degrees (*SI Appendix, Fig. S7*). The Giant Cluster Self Preference theory better captures trends across different parcellations (Fig. 2 and *SI Appendix, Fig. S8*).

Rather than building up graphs, we also create final networks and perform targeted attack to test whether corresponding P curves match those of real brains. Two models we consider are small-world networks (1) and random edges with distances based on the center of mass coordinates of regions in the parcellation. Both are limited to tract distance; they provide no information on tract lengths nor on tract densities. Neither model exhibits a match with the corresponding experimental data for increasing targeted attack on tract distance at either a coarse or fine parcellation (*SI Appendix, Figs. S9 and S10*). In the case of small-world networks, this is not unexpected because it is essentially a 2-dimensional ring lattice and lattices are known to have critical points (14, 15).

α as a Universal Property of Brains. The Giant Cluster Self Preference theory uses a fit parameter α , which measures the relative rates of a new network edge landing inside the giant cluster versus landing outside the giant cluster (*Methods*). Does

α reflect some universal feature of brains or depend on different brain states? We test whether α depends on the age or mental health status of adults based on the biometric properties in the UK Biobank dataset. We find no strong systematic changes in α as a function of age (Fig. 3)[‡]. We also find that α values are indistinguishable by gender (*SI Appendix, Fig. S11*). In addition, no differences are found in α for individuals diagnosed with diabetes compared to healthy individuals (*SI Appendix, Fig. S12*) or for individuals diagnosed with bipolar disorder or depression compared to healthy individuals (*SI Appendix, Fig. S13*). The shape of the attack/trajectory curves and the α values for each of them appear to be a relatively universal feature of human adult brains.

To extend our results to younger subjects, we analyzed dMRI data from the ABCD Study (29), which contains individuals aged 9 through 13 y. We find α values for the corresponding tract density and length-targeted attack remarkably similar to those of adults studied in the UK Biobank and relatively insensitive to age, although more sensitive than those of the UK Biobank (Fig. 3). Given the widespread presence of synaptic pruning during adolescence (36–38), our results indicate that synaptic pruning plays a minor role in affecting percolation because there are many edges responsible for maintaining $P = 1$. Furthermore, we analyze dMRI data from dHCP (30) for recently conceived newborns and reach the same conclusion, although the similarity with adult α values is not as high. Fig. 3 illustrates the similarity in P curves across individuals of different ages. Taken together, these results imply that α does not reflect postnatal human white matter development and is more consistent with prenatal development.

To check that our results are not an artifact of the dMRI data modality, we studied viral tracing data of mice from the Allen Institute (31); *SI Appendix, Fig. S14*. Again, the present Giant Cluster Self Preference theory gives excellent fits to the attack curves, but now with larger-than-human values of α for distance-based attacks and smaller-than-human for density-based attacks[§].

[‡]Although the Spearman correlation coefficient between α and age is highly significant for those α values corresponding to tract density with a P value $\sim 10^{-15}$, the effect size is low with a Spearman correlation coefficient of -0.06 . The Spearman correlation coefficient between α and age is not significant for those α values corresponding to tract length.

[§]These differences in details could be attributed to a combination of factors, such as the particular parcellation of the mouse brain into regions and viral tracing being limited to distances (see *SI Appendix, Fig. S14's* caption). Nonetheless, these results suggest that the P curves reflect a generic developmental signal that extends across different species.

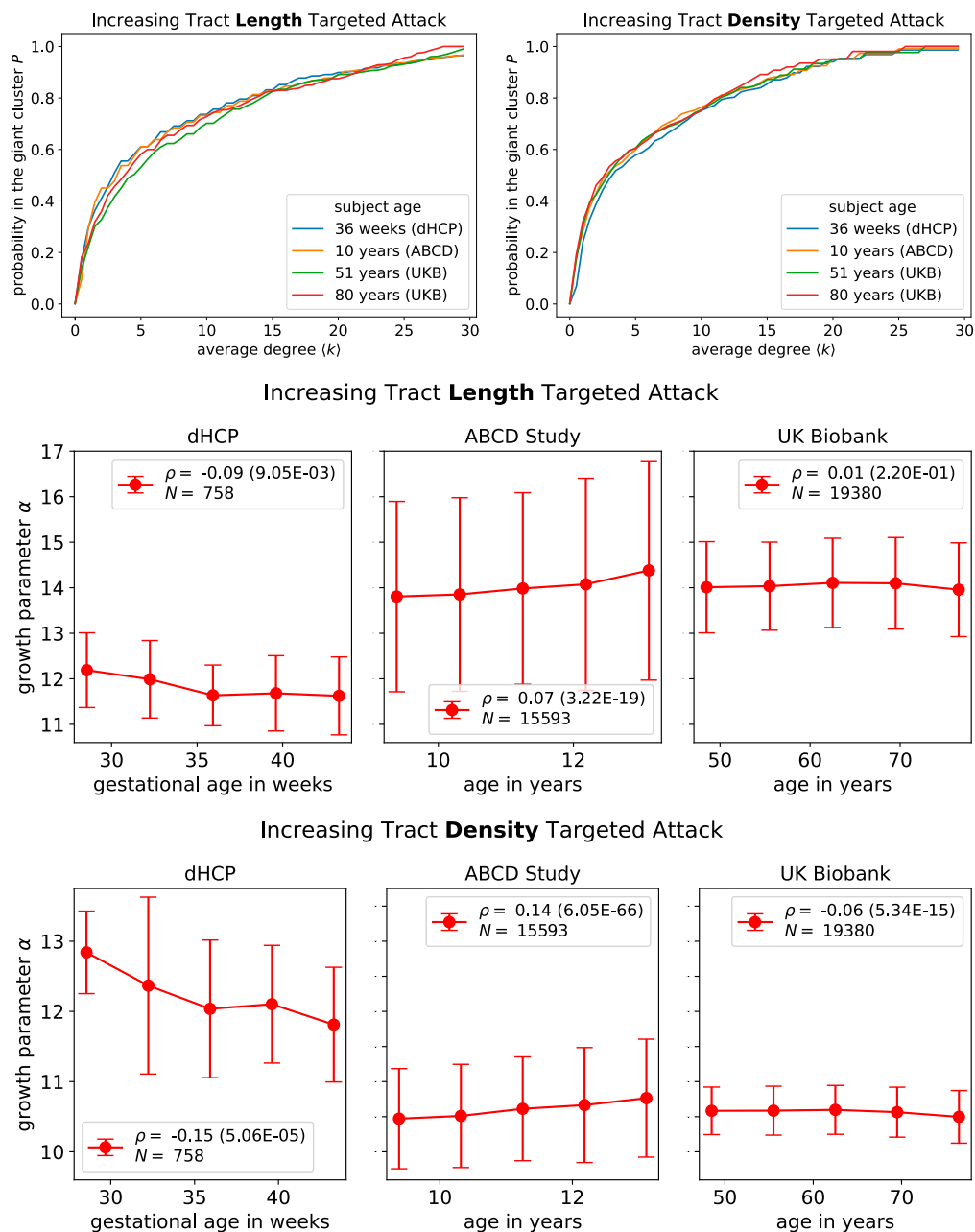


Fig. 3. Attack curves for different human subjects of different ages. $P(k)$ is remarkably similar across different ages within and across different datasets. To concisely capture the behavior of all P curves, we show fitted α parameters as a function of age in the *Bottom* two rows of plots. Binned data are presented with a line connecting means, and error bars correspond to SD. The variable ρ corresponds to the Spearman correlation coefficient calculated over all individuals between age and α , with P value in parentheses. The variable N corresponds to the total number of individuals from the respective dataset included in the analysis. An arbitrarily chosen sample of subjects are shown in the top row for subjects of age 36 wk (Developing Human Connectome Project, subject ID: CC00063AN06, $\alpha_{\text{length}}=12.7$, $\alpha_{\text{density}}=11.9$), 10 y (Adolescent Brain Cognitive Development Study, subject ID: NDARINNVF8N71U, $\alpha_{\text{length}}=13.1$, $\alpha_{\text{density}}=10.8$), 51 and 80 y old (UK Biobank, subject IDs: 6025360 and 4482035, $\alpha_{\text{length}}=15.3$, $\alpha_{\text{density}}=11.0$ and $\alpha_{\text{length}}=14.2$, $\alpha_{\text{density}}=10.4$, respectively).

Our claim that such a targeted attack approach can provide insights into prenatal development is not without precedent. It has direct support at the neuronal scale of mice (39–41)[‡]. Taken together, our results across humans of different ages and mental health diagnoses, as well as on mice using a different

data modality, support the claim that reversing the respective targeted attack procedure provides a hypothesis for how brains develop.

During Neurodevelopment, the Earliest Tracts Become Densest and Longest. One observation that the Giant Cluster Self Preference theory does not address is that targeted attack yields nonrandom P curves when attacking edges based on increasing tract lengths and densities. Results from the previous subsection indicate that the P curves reflect a special tract formation order during fetal development in which the brain

[‡]Notably, results qualitatively differ from ours, in that a critical point in their P curves is observed. It is not clear whether the neuronal scale, the in vitro sample, and/or the presence of neuronal activity are responsible for differences found compared to our results. Nonetheless, this study showed that $P = 1$ near the expected birth of the fetal mouse (40). In addition, P curves of neurons initially derived from adult mice and fetal mice were similar, indicating that P curves reflect a developmental signature that persists throughout organisms' lifetimes (39, 40).

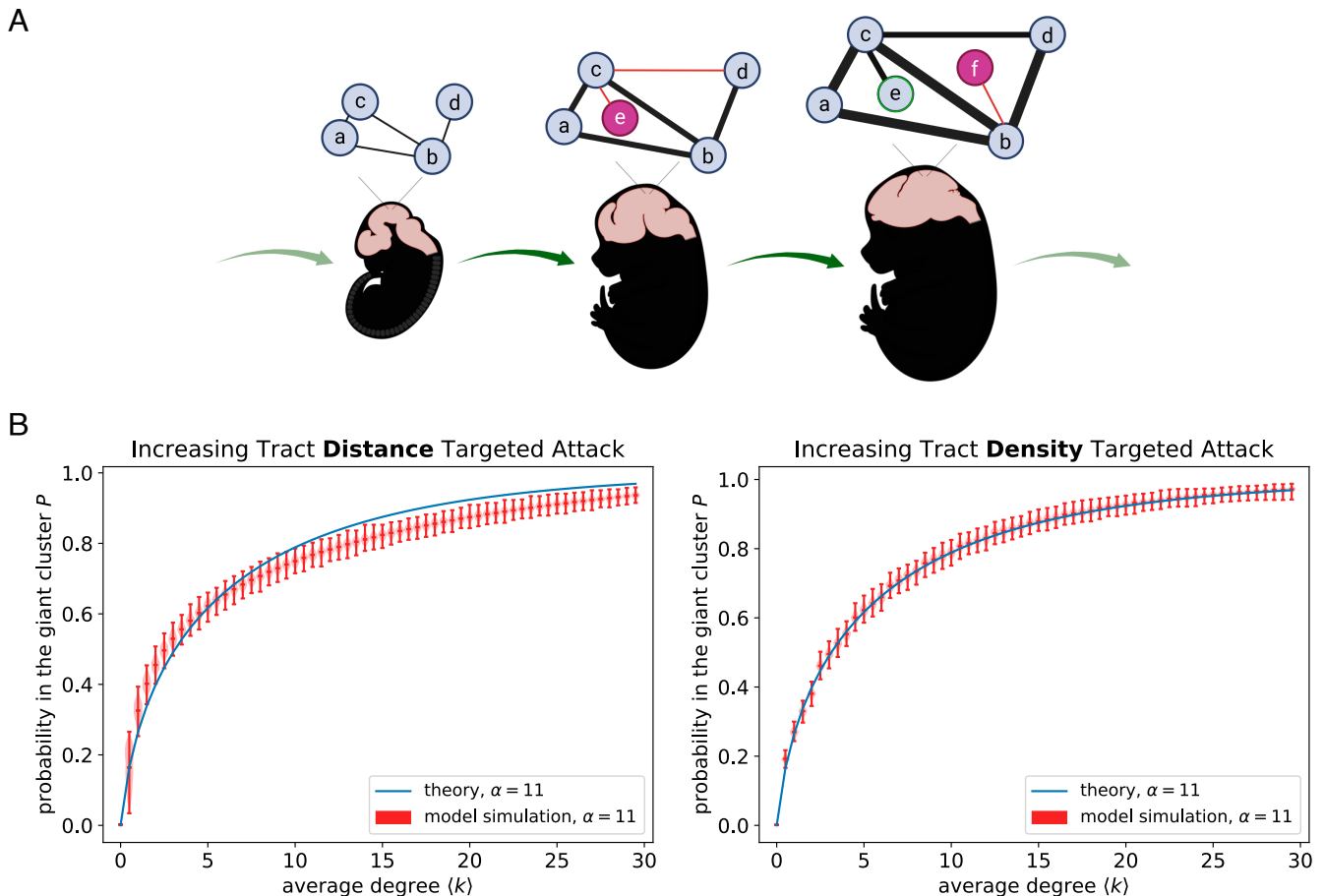


Fig. 4. Early Path Dominance Model: Proposed sequence of topological growth in brain development. (A) New edges, marked in red, start off short and thin (less dense) and become longer and wider (more dense) with each growth step. Figure created with biorender.com. (B) Simulations of the model are consistent with theory for the same α parameter. Violin plots represent 1,000 independent runs of a graph with 727 nodes, the same number as in the Talairach atlas.

is rapidly growing (36, 42). In Fig. 4A, we propose a model, Early Path Dominance, that supplements the Giant Cluster Self Preference theory by accounting for concurrent brain growth as the giant cluster increases in size. As the fetal brain is emerging, initial edges are short with small tract densities; however, as the brain grows in size and new regions appear, those same edges lengthen and become more dense. Tract lengthening has been called scaling and found to be present in the nervous system development of *C. elegans* (9, 43–45). In the context of dMRI, there are multiple reasons that tracts can appear more dense. These include both fasciculation (increasing fiber count) (42, 46, 47) as well as increasing myelination (25).

We perform a simulation of the Early Path Dominance model, incorporating scaling and fasciculation/myelination alongside the critical constraint that edges only add between regions where at least one region is already in the giant cluster (Giant Cluster Self Preference theory). Physical growth is encoded by randomly assigning coordinates to nodes within a unit sphere and assuming existing coordinates uniformly increase by a fixed constant every time an edge is added (*Methods*). Fig. 4B demonstrates that the P curves from the Early Path Dominance model agree well with the Giant Cluster Self Preference theory, which itself showed good agreement with experiments (Fig. 2). The Giant Cluster Self Preference theory (Eq. 2) makes no specific reference to tract length and density in so much that edge addition (increasing $\langle k \rangle$) reflects the corresponding tract order formation (further discussed in *Methods*).

An important feature seen in brain data is higher average clustering coefficients but similar average path lengths to random graph networks (6, 48–50). The Early Path Dominance model captures this feature to a certain extent (*SI Appendix, Fig. S16*), despite not being specifically designed to do so[#]. Small-world networks are capable of achieving the higher clustering coefficients seen in brain data at finer resolutions (1). For the same adult individual as in Fig. 2 under the Talairach atlas, we find that the average clustering coefficient is 0.60, while a random graph with the same average degree of 29.9 has an average clustering coefficient of 0.04. However, it is unclear whether highly elevated average clustering coefficients at fine parcellations are artifacts of the edge sparsity (51, 52).

Wider Perspectives

The relative invariance of the P curves implicates that this percolation mechanism may matter for brain function and that it may be a relatively universal mechanism of brain growth. Moreover, fitting brain data require values of $\alpha > 1$, meaning new neurons have greater propensity to connect inside the giant cluster than outside it. Taken together, these points lead us to the following speculations. First, perhaps the preference of neurons to

[#]Note that we only need the Giant Cluster Self Preference theory aspect of the Early Path Dominance model to achieve these results since clustering coefficient is a graph theoretical property and edge addition does not depend on growth mechanism in our model.

link to the giant cluster is because of a “Hebbian principle” (53), where neurons have greater recruitment and growth into regions of higher neuronal activity, which, in this case, may be associated with larger clusters. Future work will test this speculation with activity data of neuronal regions from functional MRI. Second, this mechanism has the efficiency advantage to the organism of “no neuron left behind”. That is, neurons mostly link to clusters that are active and connected. And third, this mechanism is opportunistic, generating new connections stochastically that lead to different brain wirings in detail, even while constrained by a common principle.

Conclusions and Discussion

We have analyzed human brain network topologies by performing an attack analysis, systematically removing links in the computer, and watching how it changes the probability of being in the giant cluster, $P(\langle k \rangle)$ as a function of the average degree $\langle k \rangle$ of the network at that stage. We use two different attack variables: tract length and density. These curve shapes under increasing targeted attack are universal across postnatal age and disease, and in mice, but are very different than for random networks. We hypothesize that time-reversing the attack procedure may mimic the physical neuronal development of fetal brains. On that basis, we derive an analytical equation that grows the brain network and provides two fundamental insights into white matter tract development. First, tracts form primarily from regions already in the giant cluster. Second, on average, the first tracts constructed become longest and densest.

Materials and Methods

Calculating the Connectivity Matrix. Connectivity matrices are calculated using the Diffusion Imaging in Python (DIPY) software (54). Alongside the already preprocessed dMRI images from the respective dataset (discussed in *SI Appendix*), we input a brain atlas that distinguishes between white and gray matter, as well as parcellates the gray matter into an arbitrary number of regions. In the Results, we use the Talairach atlas (34), however, we also show results for the Harvard-Oxford (55) and the modified Desikan-Killiany atlases (54, 56) in *SI Appendix, Fig. S8*. Nodes found to form zero edges when calculating the connectivity matrix are removed from consideration such that $P = 1$ when all edges are considered.

We perform a deterministic tracking method in DIPY to generate the connectivity matrix (54). Reconstruction of the orientation distribution function is done using Constant Solid Angle (Q-Ball) with a spherical harmonic order of 6 (57). The relative peak threshold is set to 0.8 with a minimum separation angle of 45°. We only seed voxels in the white matter and count tracts that ended in the gray matter. Minimum step size of tracts is 0.5 mm. Our protocol to obtain connectivity matrices closely follows a DIPY tutorial found on their website under streamline tools. We tried a different parameter (spherical harmonic order = 8) and reconstruction method (diffusion tensor imaging) to calculate connectivity matrices and found similar P curves as those shown in the main text (*SI Appendix, Fig. S17*). Results are also robust to variations in number of nodes or average degrees seen across individuals' brain networks (*SI Appendix, Fig. S18*).

Tractography outputs the number of tracts that connects two gray matter regions, which we call the tract density, and the individual lengths of each tract. Because each tract can have its own unique length, we take the average over all lengths to represent a given connection between two gray matter regions and call that the tract length.

The Giant Cluster Self Preference Theory. We seek a network with specific characteristics which quantitatively displays agreement with the wealth of percolation curves generated from targeted attack on dMRI data. In this section, we derive an analytical expression for $P(\langle k \rangle)$ —the fraction of nodes residing in the giant cluster as a function of the average degree $\langle k \rangle$ —for this network.

In the first subsection of the Results, we observed that brain networks under increasing tract length and density-targeted attack display the peculiar feature of essentially never having secondary clusters; there is just the one giant cluster and then either isolated nodes (nodes that form no other edges) or very small clusters (*SI Appendix, Figs. S3 and S4*). To develop our analytical theory, we take this to the extreme and approximate brain networks as having only one cluster. For a growing brain network, we even expect no isolates. When performing the targeted attack in order to generate percolation curves, sometimes removing a link causes a node to become isolated, like an island that has been set free from a mainland, and those are the isolates we see in the deconstructed network. In our theory, we consider the reverse of targeted attack, i.e., the growth and development of the network over time. From this perspective, there is always just one cluster, and the appearance of a new node coincides with a new link that emerges from a random node in the cluster (nodes e and f in Fig. 4A).

Now consider a growing brain network currently containing n nodes, all in one cluster, with E edges. Each growth step corresponds to the addition of exactly one new edge—the only question is, will this new edge reside in the existing cluster, providing more direct connections for the nodes already there, or will it branch outward into a new node? Here, we introduce the factors defining our analytical theory of brain networks.

There are $n(n-1)/2 - E$ available edge spots within the cluster. If a new node forms with some probability $1/\alpha$, then there are $N - n$ possible new nodes that can connect to any of the n nodes in the cluster. N corresponds to the final number of nodes in the full-sized network. The α parameter captures brain growth with respect to parcellation. Although growth rates change throughout early development (58), α is simply taken to be a constant and fitted to the data[†]. An alternative yet equivalent interpretation is that an added edge is α times more likely to show up in one of the available spots within the giant cluster versus those connecting to a new region.

With the complete number of possible edges tabulated, we can calculate the probability $p(n \rightarrow n+1 | E \rightarrow E+1)$ that we add a new node to the cluster during this growth step:

$$p(n \rightarrow n+1 | E \rightarrow E+1) = \frac{\frac{1}{\alpha} n(N-n)}{\frac{1}{2} n(n-1) - E + \frac{1}{\alpha} n(N-n)}. \quad [3]$$

Since there are only two options at each growth step, the probability to place the new edge within the existing cluster rather than creating a new node is simply $p(n \rightarrow n | E \rightarrow E+1) = 1 - p(n \rightarrow n+1 | E \rightarrow E+1)$.

We can now use these expressions for the transition probabilities to study how the probability $p(n|E)$ of having n nodes evolves as more edges are added to grow the network:

$$\begin{aligned} p(n|E+1) &= p(n|E) p(n \rightarrow n | E \rightarrow E+1) \\ &\quad + p(n-1|E) p(n-1 \rightarrow n | E \rightarrow E+1) \\ &= p(n|E) [1 - p(n \rightarrow n+1 | E \rightarrow E+1)] \\ &\quad + p(n-1|E) p(n-1 \rightarrow n | E \rightarrow E+1). \end{aligned} \quad [4]$$

Rearranging slightly and plugging in the expression from Eq. 3 yields a discrete master equation:

$$\begin{aligned} p(n|E+1) - p(n|E) &= p(n-1|E) p(n-1 \rightarrow n | E \rightarrow E+1) \\ &\quad - p(n|E) p(n \rightarrow n+1 | E \rightarrow E+1) \\ &= p(n-1|E) \frac{\frac{1}{\alpha} (n-1)(N-n+1)}{\frac{1}{2} (n-1)(n-2) - E + \frac{1}{\alpha} (n-1)(N-n+1)} \\ &\quad - p(n|E) \frac{\frac{1}{\alpha} n(N-n)}{\frac{1}{2} n(n-1) - E + \frac{1}{\alpha} n(N-n)}. \end{aligned} \quad [5]$$

[†]Coarser parcellations yield smaller values of α (*SI Appendix, Fig. S8*); this makes sense since all that is needed to connect a new large region into a network is for one of the many subregions of that region to get connected, whereas with finer parcellations, it is harder to bring a new region into the network because it is like throwing a dart at a smaller target.

This difference equation can be solved numerically for any value of N to give the full evolution of $p(n|E)$, using the initial condition $p(n|0) = \delta_{n=1}$. The function δ corresponds to the Kronecker delta; $\delta_{n=1}$ takes a value of 1 when $n = 1$ and is 0 elsewhere.

We now consider the limit as $N \rightarrow \infty$, where both n and E also go to infinity but the ratios $\rho = n/N$ and $\kappa = 2E/N$ remain finite^{**}. In this limit, defining $\rho(\rho|\kappa) = \lim_{N \rightarrow \infty} Np(n = \rho N|E = \kappa N/2)$ we can rewrite Eq. 5:

$$\frac{\partial \rho(\rho|\kappa)}{\partial \kappa} = -\frac{\partial}{\partial \rho} \left\{ \frac{1 - \rho}{2 - (2 - \alpha)\rho} \rho(\rho|\kappa) \right\}. \quad [6]$$

With initial condition $\rho(\rho|0) = \delta(\rho)$, the full solution of this partial differential equations is (using, for example, the method of characteristics) $\rho(\rho|\kappa) = \delta[\rho - f(\kappa)]$, where the function $f(\kappa)$ satisfies:

$$f'(\kappa) = \frac{1 - f(\kappa)}{2 - (2 - \alpha)f(\kappa)}, \quad [7]$$

and $f(0) = 0$. The solution of this now ordinary differential equation is

$$f(\kappa) = 1 + \left(\frac{1}{1 - 2/\alpha} \right) W \left[\left(\frac{2}{\alpha} - 1 \right) e^{-(1-2/\alpha)} e^{-\kappa/\alpha} \right]. \quad [8]$$

In this limit, the stochasticity in $\rho(\rho|\kappa)$ is lost and the network size actually grows deterministically with growing edge density ratio κ , satisfying $\rho = f(\kappa)$. Recognizing that κ —the edge density ratio in the growth perspective—corresponds to the average degree $\langle k \rangle$ in the targeted attack perspective^{††} and that $\rho(\kappa)$ —the fractional size of the network compared to its full size as a function of κ in the growth perspective—corresponds to the probability $P(\langle k \rangle)$ to be in the giant cluster as a function of $\langle k \rangle$ in the targeted attack perspective, we arrive at Eq. 2, which we reproduce here for ease of reading:

$$P(\langle k \rangle) = 1 + \left(\frac{1}{1 - 2/\alpha} \right) W \left[\left(\frac{2}{\alpha} - 1 \right) e^{-(1-2/\alpha)} e^{-\langle k \rangle/\alpha} \right]. \quad [9]$$

Although this form bears some resemblance to the analogous expression for a random graph (Eq. 1), including the appearance of the Lambert W function, it produces significantly different features, particularly $P(\langle k \rangle)$ approaching 1 much more gradually with increasing $\langle k \rangle$. Eq. 9 matches the numerical solution of Eq. 5, validating our large N approximations (SI Appendix, Fig. S19).

This form matches targeted attack data on increasing tract density and length remarkably well, when fitted with an optimal α value (Fig. 2). This further validates the main assumptions: brain networks have a single connected cluster throughout development, and internal connections are added continuously during growth, whereas the network only branches out and creates new nodes every so often. These aspects constitute our theory of developing brain networks, which we call Giant Cluster Self Preference. The name encapsulates the central idea that internal connections are shored up before the cluster branches out and connects with new nodes.

One limitation is that the Giant Cluster Self Preference theory does not include any reference to tract length or tract density. Future work will extend the current theory by assigning weights to each edge corresponding to these quantities and explain differences in fitted α values between tract density and length-derived P curves. Tract length and density are known to anticorrelate (6, 59, 60) (SI Appendix, Fig. S20). However, the anticorrelation does not underlie results because increasing tract length- and tract density-targeted attack produce qualitatively similar results rather than length/density increasing and density/length decreasing (Fig. 2 and SI Appendix, Fig. S1). Furthermore, future work could build on the foundations here to explore other graph-theoretic aspects of networks governed by the Giant Cluster Self Preference

^{**} Requiring κ to remain finite corresponds to an early stage from the growth and development perspective and to sparsely populated edge spots from the targeted attack perspective. In either case, in the opposite limit—at late stages of growth or when most of the edges are still there in targeted attack, we get the expected result that the connected cluster spans the whole possible network. This approximation is critical to obtaining an analytical expression because it allows us to neglect the κ term in the denominator of the transition probability.

^{††} The average degree $\langle k \rangle_g$ when normalized by the number of nodes present in the growth perspective is equal to κ/ρ .

theory, such as degree distributions, and how these quantities evolve over time.

Simulating the Early Path Dominance Model. In Fig. 4B, we demonstrate that corresponding P curves for simulations of the Early Path Dominance model match the Giant Cluster Self Preference theory and, thus, are consistent with experimental P curves (Fig. 2). In principle, there are numerous ways of implementing the Early Path Dominance model according to the rules outlined in Fig. 4A. Here, we constrain nodes to originate randomly within a sphere centered at the origin of radius 1. Nodes' Cartesian coordinates are rescaled by 1.0001 each time an edge is added to the network, causing existing edges to lengthen by a factor of 1.0001. New edges are initially given a density of 1 that increases by a factor of 1.001 each time an edge is added to the network, causing existing edges to become more dense. Edges are added according to the Giant Cluster Self Preference theory with $\alpha = 11$. Simulations are run with 727 nodes, the same number as in the Talairach atlas, and a final average degree of 100 to ensure that at the conclusion of the simulation, all graphs have $P = 1$. Note that the additional randomness of coordinate placement seems to result in the distance-targeted attack P curve requiring a higher α value ($\alpha \approx 13$) to capture the simulation (Fig. 4B).

Data, Materials, and Software Availability. Scripts necessary to reproduce figures and conclusions reached in the text can be found at http://github.com/rzabzan/percolating_brain. Please refer to the respective publicly available diffusion MRI dataset to access previously published data (UK Biobank, ABCD Study and dHCP) (28–31).

ACKNOWLEDGMENTS. We thank Richard Granger, Elisha Moses, Botond Antal, Charles Kocher, and Mobolaji Williams for helpful discussions of the results, as well as Botond Antal and Syed Fahad Sultan for assistance with the dMRI data. The research described in this paper is funded by the W.M. Keck Foundation (to L.R.M.-P. and K.A.D.) and the White House Brain Research Through Advancing Innovative Technologies (BRAIN) Initiative (NSFNCs-FR 1926781 to L.R.M.-P. and K.A.D.) and the Stony Brook University Laufer Center for Physical and Quantitative Biology (K.A.D.). This research has been conducted using the UK Biobank Resource under Application Number 37462. Data were provided by the developing Human Connectome Project, KCL-Imperial-Oxford Consortium funded by the European Research Council under the European Union Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 319456. We are grateful to the families who generously supported this trial. Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9 to 10 y and follow them over 10 years into early adulthood. The ABCD Study[®] is supported by the NIH and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, and U24DA041147. A full list of supporters is available at <https://abcdstudy.org/federal-partners.html>. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from NIMH Data Archive DOI <https://doi.org/10.15154/1526472>.

Author affiliations: ^aLaufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794; ^bDepartment of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794; ^cDepartment of Chemistry, Stony Brook University, Stony Brook, NY 11794; ^dDepartment of Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794; ^eProgram in Neuroscience, Stony Brook University, Stony Brook, NY 11794; and ^fAthinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02129

1. D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
2. S. Achard, R. Salvador, B. Whitcher, J. Suckling, E. Bullmore, A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* **26**, 63–72 (2006).
3. M. P. Van Den Heuvel, O. Sporns, Rich-club organization of the human connectome. *J. Neurosci.* **31**, 15775–15786 (2011).
4. D. S. Bassett, E. T. Bullmore, Small-world brain networks revisited. *Neurosci.* **23**, 499–516 (2017).
5. D. S. Bassett, E. Bullmore, Small-world brain networks. *Neuroscientist* **12**, 512–523 (2006).
6. E. Bullmore, O. Sporns, Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
7. R. F. Betzel *et al.*, Generative models of the human connectome. *Neuroimage* **124**, 1054–1064 (2016).
8. D. Akarca, P. E. Vértés, E. T. Bullmore, D. E. Astle, A generative network model of neurodevelopmental diversity in structural brain organization. *Nat. Commun.* **12**, 1–18 (2021).
9. M. Kaiser, *Changing Connectomes: Evolution, Development, and Dynamics in Network Neuroscience* (MIT Press, 2020).
10. T. Zhao, Y. Xu, Y. He, Graph theoretical modeling of baby brain networks. *NeuroImage* **185**, 711–727 (2019).
11. S. Wilson *et al.*, Development of human white matter pathways in utero over the second and third trimester. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023598118 (2021).
12. M. Ouyang *et al.*, Differential cortical microstructural maturation in the preterm human brain with diffusion kurtosis and tensor imaging. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4681–4688 (2019).
13. R. Albert, H. Jeong, A. L. Barabási, Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
14. A. L. Barabási, *Network Science* (Cambridge University Press, 2016).
15. A. A. Saberi, Recent advances in percolation theory and its applications. *Phys. Rep.* **578**, 1–32 (2015).
16. R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626–4628 (2000).
17. S. Gu *et al.*, Controllability of structural brain networks. *Nat. Commun.* **6**, 1–10 (2015).
18. A. Li, S. P. Cornelius, Y. Y. Liu, L. Wang, A. L. Barabási, The fundamental advantages of temporal networks. *Science* **358**, 1042–1046 (2017).
19. M. Kaiser, R. Martin, P. Andras, M. P. Young, Simulation of robustness against lesions of cortical networks. *Euro. J. Neurosci.* **25**, 3185–3192 (2007).
20. M. E. Lynall *et al.*, Functional connectivity and brain networks in schizophrenia. *J. Neurosci.* **30**, 9477–9487 (2010).
21. J. Cabral, E. Hugues, M. L. Kringelbach, G. Deco, Modeling the outcome of structural disconnection on resting-state functional connectivity. *Neuroimage* **62**, 1342–1353 (2012).
22. K. E. Joyce, S. Hayasaka, P. J. Laurienti, The human functional brain network demonstrates structural and dynamical resilience to targeted attack. *PLoS Comput. Biol.* **9**, e1002885 (2013).
23. S. Guo *et al.*, Percolation analysis of brain structural network. *Front. Phys.* **9**, 345 (2021).
24. P. Kotlarz, J. C. Nino, M. Febo, Connectomic analysis of Alzheimer's disease using percolation theory. *Net. Neurosci.* **6**, 213–233 (2022).
25. D. K. Jones, T. R. Knösche, R. Turner, White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *Neuroimage* **73**, 239–254 (2013).
26. N. A. Crossley *et al.*, The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382–2395 (2014).
27. D. Tomasi, G. J. Wang, N. D. Volkow, Energetic cost of brain functional connectivity. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13642–13647 (2013).
28. C. Sudlow *et al.*, UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
29. N. D. Volkow *et al.*, The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev. Cognit. Neurosci.* **32**, 4–7 (2018).
30. E. J. Hughes *et al.*, A dedicated neonatal brain imaging system. *Magnet. Res. Med.* **78**, 794–804 (2017).
31. S. W. Oh *et al.*, A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
32. B. Bollobás, *Random Graphs* (Cambridge University Press, 2001), No. 73.
33. F. Alfaro-Almagro *et al.*, Image processing and quality control for the first 10,000 brain imaging datasets from UK biobank. *Neuroimage* **166**, 400–424 (2018).
34. J. L. Lancaster *et al.*, Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* **10**, 120–131 (2000).
35. A. L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
36. R. A. Bethlehem *et al.*, Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
37. V. J. Sydnor *et al.*, Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. *Neuron* **109**, 2820–2846 (2021).
38. G. Collin, M. P. Van Den Heuvel, The ontogeny of the human connectome: Development and dynamic changes of brain connectivity across the life span. *Neuroscientist* **19**, 616–628 (2013).
39. J. P. Eckmann *et al.*, The physics of living neural networks. *Phys. Rep.* **449**, 54–76 (2007).
40. J. Soriano, M. R. Martínez, T. Tlustý, E. Moses, Development of input connections in neural cultures. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13758–13763 (2008).
41. O. Cohen *et al.*, Quorum percolation in living neural networks. *EPL (Europhys. Lett.)* **89**, 18008 (2010).
42. H. Lagercrantz, *Infant Brain Development* (Springer, 2016).
43. M. Kaiser, Mechanisms of connectome development. *Trends Cognit. Sci.* **21**, 703–717 (2017).
44. S. Varier, M. Kaiser, Neural development features: Spatio-temporal development of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001044 (2011).
45. S. Lim, M. Kaiser, Developmental time windows for axon growth influence neuronal network topology. *Biol. Cybernet.* **109**, 275–286 (2015).
46. E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, A. J. Hudspeth, *Principles of Neural Science* (MacGrawHill Medical, New York, 2013).
47. D. Chaudhuri, P. Borowski, M. Zapotocky, Model of fasciculation and sorting in mixed populations of axons. *Phys. Rev. E* **84**, 021908 (2011).
48. D. S. Bassett, E. T. Bullmore, Human brain networks in health and disease. *Curr. Opin. Neurol.* **22**, 340 (2009).
49. E. Bullmore, O. Sporns, The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349 (2012).
50. H. Aerts, W. Fias, K. Caeyenberghs, D. Marinazzo, Brain networks under attack: Robustness properties and the impact of lesions. *Brain* **139**, 3063–3083 (2016).
51. A. Zalesky *et al.*, Whole-brain anatomical networks: Does the choice of nodes matter? *Neuroimage* **50**, 970–983 (2010).
52. T. C. Lacy, P. A. Robinson, Effects of parcellation and threshold on brain connectivity measures. *PLoS One* **15**, e0239717 (2020).
53. D. O. Hebb, Organization of behavior. New York: Wiley. *J. Clin. Psychol.* **6**, 307–335 (1949).
54. E. Garyfallidis *et al.*, Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* **8**, 8 (2014).
55. N. Makris *et al.*, Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr. Res.* **83**, 155–171 (2006).
56. F. Bruce, Freesurfer. *Neuroimage* **62**, 774–781 (2012).
57. I. Anjan *et al.*, Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle. *Magnet. Reson. Med.* **64**, 554–566 (2010).
58. S. Khan *et al.*, Fetal brain growth portrayed by a spatiotemporal diffusion tensor MRI atlas computed from in utero images. *Neuroimage* **185**, 593–608 (2019).
59. S. Achard, E. Bullmore, Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* **3**, e17 (2007).
60. P. Hagmann *et al.*, Mapping human whole-brain structural networks with diffusion MRI. *PLoS One* **2**, e597 (2007).