Percolation Theory Identifies Critical Constraints on White Matter Tract Development

Rostam M. Razban[1], Ken A. Dill[1,4,6], and Lilianne R. Mujica-Parodi[1,2,3,4,5,*]

[1]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY, USA
[2]Dept. of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA
[3]Program in Neuroscience, Stony Brook University, Stony Brook, NY, USA
[4]Dept. of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA
[5]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[6]Dept. of Chemistry, Stony Brook University, Stony Brook, NY, USA
[*]correspondence mujica@lcneuro.org

June 20, 2022

Abstract

To better understand fundamental constraints on the global structural organization of the brain, we apply *percolation theory* as a quantitative measure of global communication across a network. The largest present sub-network across which *all* sets of regions are able to communicate defines a *giant cluster*. By novel analytic solution, we prove that if constructed tracts are constrained to emanate from those regions already in the giant cluster, a giant cluster grows smoothly without requiring a critical point. The predicted structure is highly consistent with brain data, among human diffusion MRI across adults (UK Biobank, N=19,380), adolescents (Adolescent Brain Cognitive Development Study, N=15,593) and neonates (Developing Human Connectome Project, N=758), as well as mouse viral tracing (Allen Institute). This suggests a fundamental mechanism for neurodevelopment in which the earliest tracts also become the longest and densest.

Summary

During neural growth and specialization there is a risk that regions may become siloed, yet the biological strategy by which neurodevelopmental processes avoid this trap is unknown. Here, we study brain structure 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 a

rule of brain structure, in which earlier tracts become longer and denser. This signature, invariant to age or mental health, may suggest a fundamental condition that must be met for the brain to function as an emergent whole.

Keywords: dMRI, percolation, neurodevelopment, white matter

1 Introduction

From liquid-liquid phase transitions of proteins [1] to the flocking behavior of birds [2, 3], biology abounds with examples in which the widespread application of simple rules gives rise to highly complex emergent structures. The human brain is one such complex structure, with ∼86 billion neurons [4], organized by connectivity, modularity, and nested control processes. Yet, as the brain develops, this functionally adaptive complexity must be balanced against the risk that regions become siloed from the brain as a whole.

This risk can be quantified by *percolation theory*. Percolation theory measures the probability (P) of a node to be in the largest group of nodes, such that: for any two nodes, there exists some path of edges connecting the first node to the second. This largest group of nodes is the *giant cluster*; for any arbitrary region to communicate with any other region, P must equal 1 [5, 6]. Previous studies have illustrated the ability to differentiate between graph classifications through P curves generated by *targeted attack*: sequential removal *in silico* of nodes or edges based on the rank order of some node or edge property [7, 8, 9]. Yet, in spite of widespread interest by the neuroscience field in applying graph theoretic features, such as *degree distribution*, *path length*, *clustering coefficient* and *rich-club organization*, to the brain [10, 11, 12, 13, 14, 15, 16, 17], the most basic property: *percolation probability*, has yet to be fully understood. This matters because, if true, percolation would need to be observed as a fundamental self-organizing principle of the brain, which must be considered in conjunction with the other identified constraints.

With respect to *functional* connectivity, it has been shown that when attacking edges based on correlations of functional MRI (fMRI) signals between regions, the brain's P curves uniquely differ from those of standard graphs, such as random and scale-free graphs [11]. Others have used percolation theory as a starting point to infer the brain's modularity and fractal nature [18, 19], as well as critical nodes responsible for keeping networks intact [20]. With respect to *structural* connectivity, targeted attack has also been applied to diffusion MRI (dMRI) [21], which measures the topology of large bundles of axons known as *white matter tracts* [22]. This study removed nodes based on their degree [21], motivated by the logic that their targeted attack procedure directly simulates neurodegeneration, which is more likely to be located in metabolically costly hubs [23]. Yet while the above studies generated P curves, they did not address their underlying mechanism.

Here, our primary aim is to identify the generative processes that give rise to percolation properties seen in the brain. First, we use statistical mechanics to derive an analytic equation for P under the constraint that at least one of the two regions that the tract connects is already in the giant cluster. The analytically-predicted P curve is then validated against P curves calculated through targeted attack of tract length and density from N=35,731human dMRI scans acquired from the UK Biobank [24], the Adolescent Brain Cognitive Development (ABCD) Study [25] and the Developing Human Connectome Project (dHCP) [26]. To test uniqueness, we consider alternative graph theoretical structures, such as random graphs, scale-free graphs and a graph constrained by brain volume. Finally, to confirm that our conclusions are not specific to human brains or the dMRI data modality, we replicate our human results on mouse viral tracing experiments [27].

2 Results

2.1 Targeted attack on increasing tract lengths and densities produce nonrandom percolation probability curves.

We run a targeted attack procedure on the two main outputs from a generic dMRI data analysis: *tract length* and *tract density* (Figure 4 in the Methods). Tract density, also known as streamline count, is the number of tracts connecting two grey 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. Figure 1 demonstrates that 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 (Figure S1, right). Decreasing tract length removal results in a P curve qualitatively similar to that of random graphs (Figure S1, left), consistent with previous results [21]. For the remainder of the article, we focus on characterization of the unique mechanism underlying the P curves from increasing tract densities and increasing tract lengths under targeted attack (below, abbreviated as *the P curves*).

The most striking observation is that immediately at an average degree ($\langle k \rangle$) of 0, P starts increasing above 0 (Figure 1). In the physical sense, the P curves do not exhibit a 'critical point' for positive $\langle k \rangle$, i.e. P does not equal zero at $\langle k \rangle > 0$ and then suddenly jump in value. This observation might seem surprising in light of statistical mechanical applications that have demonstrated the importance of critical points in biological systems [28]. Biological systems are thought to evolve to be situated near critical points because it poises them to respond most efficiently to small environmental changes [28, 29, 30]. However,

unlike with dynamical systems, in which there are mechanisms to amplify and attenuate signals; once a region is structurally siloed, there is no mechanism to signal the siloed region to catalyze tract formation. The fact that at no point along the attack can a critical point be reached is not a limitation of the P metric. P curves of random graphs exhibit a critical point at $\langle k \rangle = 1$ and classes of lattices exhibit critical points at positive $\langle k \rangle$ [9, 31].

One important consequence of lacking a positive valued critical point is that the brain is uniquely robust to targeted attacks by increasing tract length or density at small $\langle k \rangle$. Total network failure is only reached when all tracts are destroyed; thus, there will always be some communication between brain regions despite almost all tracts being removed. Compared to random graphs, however, this benefit comes at a cost for larger $\langle k \rangle$. While P for random graphs quickly reaches 1 by $\langle k \rangle \sim 4$, the rise in P to reach 1 for the brain is more gradual (Figure 1). Thus, perfect communication among regions represented by $P = 1$ for larger values of $\langle k \rangle$ seems to be sacrificed in order to maintain some possible communication as $\langle k \rangle$ approaches 0.

To identify the mechanism governing the brain's P curves, we look at other clusters besides the largest. We find that the distribution of cluster sizes in the brain differ from those corresponding to a random graph. At $\langle k \rangle = 1$, the giant component in the brain is essentially the only cluster with non-negligible size; nodes not in the giant cluster are siloed and do not form secondary clusters. This contrasts sharply with results in the random graph; at $\langle k \rangle = 1$, the giant cluster is just one of many non-negligible clusters present (Figure S2). We confirm this observation across different values of $\langle k \rangle$ and conclude that the brain's giant cluster continuously degrades one node at a time, rather than by sudden fragmentation into secondary clusters from one large cluster under the targeted attack procedure. This lack of fragmentation explains the lack of a critical point in the brain's P curve.

2.2 Tracts only add from regions already in the giant cluster.

We take the surprising result that the brain mostly lacks secondary clusters and encode it in an analytical model for P, where each node and edge are defined to be a brain region and white matter tract, respectively. Inspired by percolation theory, which has successfully derived, at a minimum, limits of P with respect to $\langle k \rangle$ for many graphs [9, 31], we provide a novel self-consistent equation for P and solve for the following expression over the entire range of $\langle k \rangle$. Percolation theory is the opposite of targeted attack; rather than constructing P curves from right to left, it predicts P curves based on encoded edge addition rules from left to right along the $\langle k \rangle$ axis.

$$P = 1 + W\left(-e^{-1-\langle k\rangle/\alpha}\right) \tag{1}$$

Equation 1 is a reproduction of Equation 10 and copied here for the reader's convenience. For derivation details, please refer to the equations leading up to Equation 10 in the Methods. W corresponds to the Lambert W function, also known as the product logarithm. Equation 1 contains only one parameter, α, which we fit to the experimental data. The parameter captures the modular nature of brain development, by presuming that nodes already in the giant cluster are more favored to form edges. As shown in Figure S3, α is sensitive to the resolution of the parcellation, with finer parcellations having larger α.

As demonstrated in Figure 1, Equation 1 captures experimental P curves with appropriate α fits. In Figure S7 and S10, we demonstrate that its predictive power extends to other individuals. The accuracy of our theory drops off at larger $\langle k\rangle$ for both tract length and tract density targeted attack P curves. This is expected because of increased inaccuracy in the approximation in Equation 7 of the Methods at larger P (Equation S1). To rule out other potential mechanisms that capture the P curves (unconstrained by analytical tractability), we performed simulations of networks formed by a null model based on the volume constraints of the brain, preferential attachment and the canonical small-world network. Although these simulations exhibit some qualitative agreement at coarser parcellations (Figure S4), their accuracies do not transfer to results shown in the main text, which are analyzed at a finer parcellation (Figure S5 and Figure S6). Thus, our model based on preservation of the giant cluster under attack, or inversely no secondary cluster formation, is more accurate and general in capturing trends across different parcellations.

2.3 The percolation probability curves reflect white matter development.

Although our theory accurately captures the P curves, it is unclear whether the inferred tract addition rule reflects developmental processes, rather than being related to brain maturation or even deterioration. From the 19,380 individuals with dMRI scans from the UK Biobank, we tabulate P curve variability through the fitted α parameter in Equation 1. Figure S7 shows how α fits vary for six other individuals. We test whether α systematically varies as functions of age or mental health status based on corresponding biometric properties also provided by the UK Biobank. We find no strong systematic changes in α as a function of age (Figure S8). Although the Spearman correlation coefficient between α and age is highly significant for those α values corresponding to tract density with a p-value $\sim 10^{-19}$, 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. As a control, we

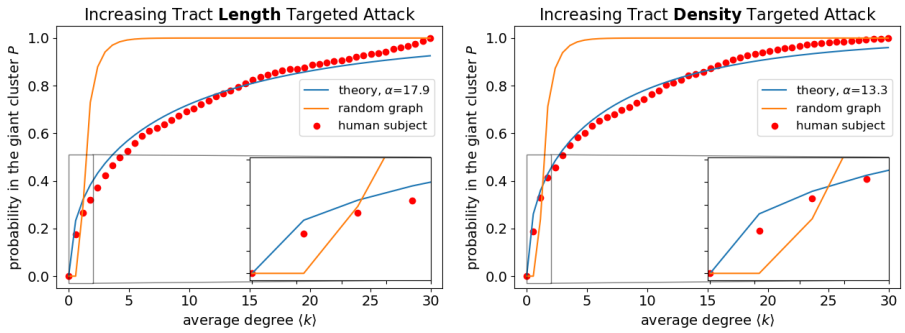

Figure 1: Our theory captures the probability for a brain region to be in the giant cluster across the average degree range with an appropriately fitted α parameter for both targeted attack of increasing tract length (left) and increasing tract density (right). Points are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360) [32]. The inset zooms in on the first four equidistant sampled experimental points to highlight the lack of a critical point.

show that both sexes have statistically indistinguishable α values (Figure S9) and that α fits are not systematically affected by age (Figure S10). In addition, no differences are found in α for individuals diagnosed with diabetes compared to healthy individuals (Figure S11) or for individuals diagnosed with bipolar disorder or depression compared to healthy individuals (Figure S12). While previous studies based on fMRI found modularity to be sensitive to age and disease, α is a coarse modularity proxy that, like the total number of modules across the brain, has been found to be fairly constant with respect to age and disease [33, 34]. This result is consistent with our hypothesis that percolation is a necessary condition for brain function, rather than a parameter that tunes performance.

To further probe and clarify α's relationship with development, we analyze dMRI data from the ABCD Study [25], which contains individuals aged 9 through 13 years old. We find α values for the corresponding tract density and length targeted attack strikingly similar to those of adults studied in the UK Biobank and relatively insensitive to age, although more sensitive than those of the UK Biobank (Figure S13). Furthermore, we analyze dMRI data from dHCP [26] for recently conceived newborns and reach the same conclusion, although the similarity with adult α values is not as high (Figure S14). Figure 2 illustrates the similarity in P curves across individuals of different ages. Taken together, these results imply that α does not reflect postnatal white matter development. Could the P curves then be capturing a necessary constraint that all brains must satisfy, thus reflecting a highly evolutionary conserved, early life stage developmental process? To this end, we access viral tracing data on mice

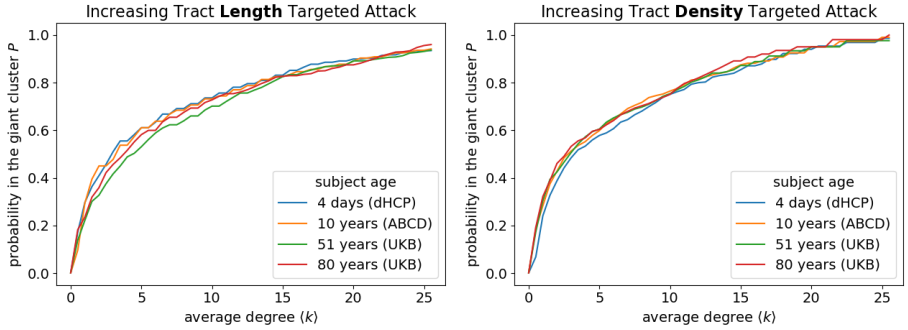

Figure 2: The probability for a brain region to be in the giant cluster as a function of average degree upon targeted attack of increasing tract length (left) and increasing tract density (right) are strikingly similar across human subjects of different ages. Plots are derived from dMRI data from the respective data set for subjects of age 4 days (Developing Human Connectome Project, subject ID: CC00063AN06), 10 years (Adolescent Brain Cognitive Development Study, subject ID: NDARINVNVF8N71U), 51 and 80 years old (UK Biobank, subject IDs: 6025360 and 4482035, respectively).

from the Allen Institute [27] and find that our theory is still valid; however, fitted α values are markedly different from those of humans (Figure S15). These differences in details could be attributed to a combination of factors, such as the particular parcellation of the mouse brain into regions and being limited to distances rather than tract lengths as in dMRI (further discussion in Figure S15's caption). Nonetheless, these results suggest that the P curves reflect a generic developmental signal that extends across different species.

The claim that our percolation analysis on white matter tracts captures early-stage brain developmental rules, perhaps during the *in utero* time frame, is not without precedent. P curves of active neuronal samples extracted from mice were shown to follow an integrate-and-fire model which the authors termed 'quorum percolation' [35, 36]. Notably, the 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 in our results. Nonetheless, this study showed that P equals to 1 near the expected birth of the fetal mouse [35]. In addition, their P curves of adult mice and fetal mice were similar, indicating that P curves reflect a developmental signature that persists throughout organisms' lifetimes [35, 37].

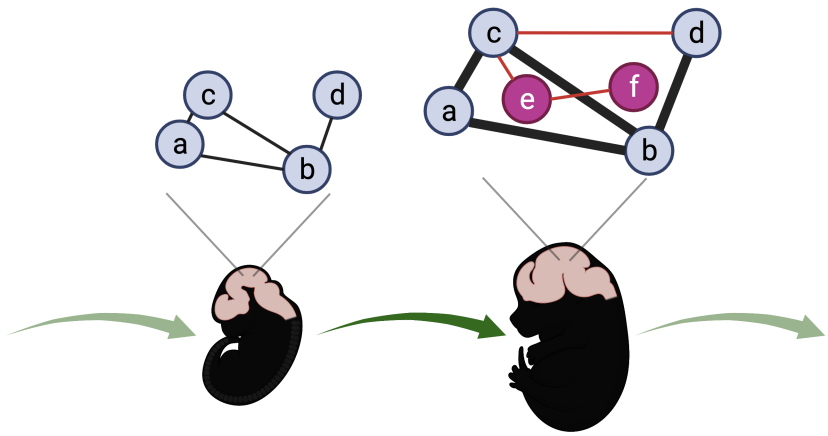

Figure 3: Schematic of our model for brain development. New edges, marked in red, start off short and thin (less dense) and become longer and wider (more dense) during the course of development. New nodes added between the newly expanded space are demarcated in pink. Figure created with biorender.com.

2.4 A model of brain development predicts that the earliest tracts become the densest and longest.

One outstanding question is the observation that targeted attack yields non-random P curves when attacking edges based on *increasing* tract lengths and densities. This is especially relevant because as just discussed in the previous paragraph, it has been shown to approach 1 *in utero* [35], a developmental stage in which the brain is rapidly growing [38, 39]. In Figure 3, we propose a model accounting for concurrent brain growth as the giant cluster increases in size. When the fetal brain is just 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 (lengthening and weighting of edges, known as *scaling* and *fasciculation*, respectively [40]). Although fasciculation is well-established [41, 38, 42], the idea of scaling is relatively new and shown to be present in the nervous system development of *C. elegans* [40, 43].

We perform a simulation of our proposed model incorporating scaling and fasciculation, alongside the critical constraint that edges only add between regions where at least one region is already in the giant cluster. In Figure S17, we find P curves consistent with our theory that recapitulates experiment. We also show that our model makes a prediction about the correlation between tract distances and densities. Numerous studies have shown that this correlation is negative and interpreted it to represent energetic constraints [44, 15, 45, 4]. Long-range connections are particularly metabolically expensive, and thus would be expected to drive evolutionary selection towards low densities [44, 15, 46, 47, 16]. Our simulation predicts no correlation between the

two properties (Figure S18), hinting that the negative result may originate from local constraints [48, 49, 50, 51] rather than from the fundamental constraint of giant cluster formation required for brain function. In turn, our simulation can potentially inform why this correlation is not very strong with a Spearman correlation of around -0.25 (Figure S19). If the giant cluster constraint results in no correlation, then other more local constraints are limited in swaying the density-distance relationship.

Discussion

Our results provide two fundamental insights into white matter tract development. First, tracts are constrained to form solely from regions already in the giant cluster. Second, the first tracts constructed become the longest and the densest on average. Results predicted by these principles are experimentally confirmed by three large-scale independent samples of human individuals' dMRI scans, as well as mouse viral tracing. We find that accurate prediction with the experimental data is obtained only when targeted attack sequentially removes from the shortest to longest or least to most dense, suggestive of a neurodevelopmental mechanism.

Our decision to experimentally generate P curves from right to left along the $\langle k \rangle$ axis reflects a theoretical, rather than purely experimental, limitation. Ideally, we could also experimentally generate P curves from left to right; i.e. *build up* the giant cluster rather than *break it down*. Recent advances in fetal dMRI would seem to already make this possible [52]. However, it is important to note that our model would predict a constant line at $P = 1$ (Figure 3). Therefore, all available regions based on the stages of brain growth would be expected to be in the giant cluster based on the functional constraint that no regions are siloed. In other words, it is only after the fact that we can accurately characterize the drop in P because all regions at the end of development have to be present for proper normalization. A simple division of giant cluster sizes across development by the number of regions at the end of development is unfeasible because of changing parcellations at different developmental stages [53].

Only P curves separately generated by targeted attack on increasing tract lengths or densities are markedly nonrandom (Figure 1). Unlike random graphs and lattices, which demonstrate a sharp phase transition at $\langle k \rangle > 0$ [9], the percolation probability for the brain's P curves only goes to zero when $\langle k \rangle = 0$. This behavior is reminiscent of how a scale-free network behaves during random attack [8, 9]. However, in deriving our analytic expression, we did not have to assume that the probability of edge addition depends on the degree of the corresponding node. Instead, the probability of forming an edge is uniform for all edges connecting giant cluster nodes in the same class. For edges connecting one giant cluster node and a node outside of the giant cluster, the probability is also uniform but smaller than that between two giant cluster nodes. The main

constraint enforced in our derivation is that two nodes not in the giant cluster have no chance of forming an edge.

Our analytic expression has only one parameter α, which we fit to experimental P curves. It captures the difference in probability values of edge formation (P) between the two possible classes of edges: giant cluster (gc) node to giant cluster node and giant cluster node to non-giant cluster (ngc) node. For example, $\alpha = 11$ corresponds to $p_{\text{gc-gc}} = 5 * p_{\text{gc-ngc}}$ (discussed in Figure S22). This parameter captures the modular nature of brain development at a global scale. A large α indicates an increased preference for edge formation among nodes already in the giant cluster, i.e. larger modularity. We find that α is sensitive to the resolution of the brain atlas used to derive the connectivity matrices from dMRI data (Figure S3). However, across individuals analyzed within the same atlas, the α value is roughly constant and insensitive to age or diseases (Figures S8, S11, S12 and S13).

The lack of variability in the P curves is consistent with a fundamental self-organizing constraint required for brain function. This fundamental aspect distinguishes percolation from well-established graph-theoretic properties of the brain, such as small-worldness, which likely reflect more nuanced tuning parameters for network efficiency and thus have useful diagnostic applications [54, 15, 47, 55]. Yet the distinction between function versus non-function matters from not only mechanistic but also theoretical perspectives, as neuroscience considers at what point proto-brains such as organoids "function" as brains [56, 57, 58].

Methods

United Kingdom Biobank

The United Kingdom (UK) Biobank is an ongoing epidemiological study aimed at discovering early markers for brain diseases [24]. UK Biobank is an excellent data source because of its large sample size (current release at our disposal: 19,380 people, ages 45 to 80 (mean age is 62.6±7.4 years old)) and data standardization procedures – all corresponding measurements are collected on the same machines across several sites operating the same software [32]. More details on the minimally preprocessed dMRI images publicly assessed from UK Biobank can be found in Alfaro-Almagro et al. [32].

Adolescent Brain Cognitive Development Study

The Adolescent Brain Cognitive Development (ABCD) Study is an ongoing longitudinal study of ten years focused on understanding brain development during the critical period of adolescence [25]. The ABCD Study complements

the UK Biobank by having individuals with an age range of 8.9 to 13.3 (mean age is 10.6±1.1 years old), for the current data release at our disposal of 15,593 scans (release 4.0, NIMH Data Archive Digital Object Identifier 10.15154/1526472). More details on the minimally preprocessed dMRI images of the ABCD Study publicly assessed from the National Institute of Mental Health Data Archive can be found in Hagler et al. [59].

Developing Human Connectome Project

The Developing Human Connectome Project (dHCP) acquires MRI from neonates within weeks of birth to probe early development [26]. The dHCP contains the youngest human subjects publicly accessible, gestational ages 26.7 to 45.1 weeks old (mean gestational age 39.7 ± 3.5 weeks old). We have 758 scans at our disposal from the current data release (third data release). More details on the minimally preprocessed dMRI images publicly assessed from dHCP can be found in Christiaens et al. [60].

Calculating the connectivity matrix

Connectivity matrices are calculated using the Diffusion Imaging in Python (Dipy) software [61]. Briefly, we input a brain atlas that distinguishes between white and grey matter, and parcellate the grey matter into an arbitrary number of regions. In the Results, we use the Talairach atlas [62], however, we also show results for the Harvard-Oxford [63] and the modified Desikan-Killiany atlases [61, 64] in the Supplement (Figure S20).

We only seeded voxels in the white matter and counted tracts that ended in the grey matter. The constant solid angle method of order 6 is employed to estimate fiber orientation. Our protocol to obtain connectivity matrices closely follows a Dipy tutorial found on their website under streamline tools. We also tried different parameters and methods to calculate the connectivity matrix and found similar P curves as those shown in the main text (Figure S21).

The analysis outputs the number of tracts that connects two grey 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 grey matter regions and call that the tract length.

Targeted attack procedure

The graph theoretical properties that we focus on are P as a function of average degree (denoted by $\langle k \rangle$). Degree is calculated by counting the number of edges of a certain node. For example, node c in Figure 4 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

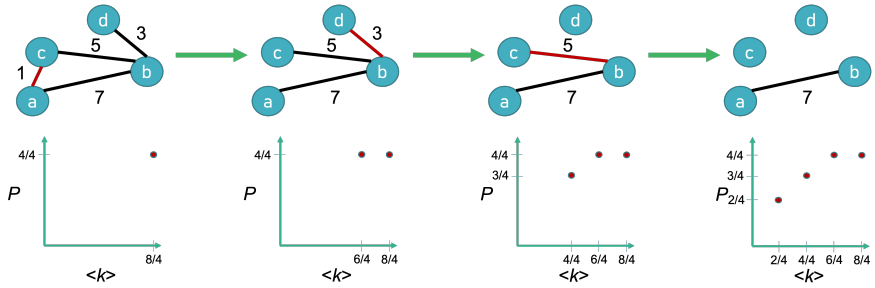

Figure 4: Schematic illustrating the targeted attack procedure on increasing tract lengths. By starting with all edges present on the left, we sequentially remove those edges whose length (listed above the edge) are currently the shortest (colored in red). P stands for the probability a node is in the giant cluster; $\langle k \rangle$ stands for the average degree.

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 $P = 2/4$ (Figure 4). In a static setup, $P = 1$ is a necessary condition for nodes from all across the graph to be able to communicate because it guarantees that all nodes can be reached either directly or indirectly by an edge [6]. Nodes found to form zero edges when calculating the connectivity matrix from dMRI data are removed from consideration such that the P value equals to 1 when all edges are considered.

Figure 4 is a schematic demonstrating how the P curves are generated. We threshold edges over the range of possible tract lengths or densities to generate P as a function of $\langle k \rangle$, which we call a P curve. If continuous thresholding is not performed, we would only have a single point. Thresholding provides more information on the organization of tract lengths and densities in the brain by demonstrating the sensitivity of P as more and more edges are removed. It allows us to peel back the layers of the onion so to speak, where we move shorter or less dense edges first, and then proceed sequentially with longer or denser edges to see how quickly P drops from a starting value of 1 to a final value of 0. This thresholding procedure is called targeted attack in the percolation theory literature [9]. However, instead of targeting based on graph theoretical properties as is common, such as targeting nodes based on their degree, we target edges based on one of the two experimentally derived measurements.

In addition to performing targeted attack from small to large tract lengths or densities as is the focus in this text, we could also do the opposite direction. Crossley et al. [21] in fact did targeted attack on tract lengths from largest to smallest. We are able to reproduce their results when performing the same targeted attack procedure. However, the mechanism underlying this P curve is essentially a random graph (Figure S1, left). In addition, we are not able to

get much meaningful information on targeted attack from large to small tract densities because when edges of tract density equal to 1 are removed, $\langle k \rangle$ drops from 6 to 0 (Figure S1, right). For these reasons, as well as the fact that the mechanism of targeted attack from small to large tract length and density is markedly nonrandom, we focus on targeted attack from small to large tract lengths and densities in the main text.

Derivation of the percolation probability curve

Inspired by percolation theory [65, 31], which is based in statistical mechanics, we derive an expression for the P curves by setting up a self-consistent equation for P. First, we derive the corresponding P curve for a random graph to contextualize our giant cluster preservation mechanism.

Random graph

The self-consistent equation for a random graph is the probability of a node not to be in the giant cluster equal to the probability of a node not forming an edge with a node in the giant cluster. Taking p to signify the probability to form an edge and N as the total number of nodes, we translate the previous sentence into mathematical symbols as

$$1 - P = \sum_{j=0}^{N-1} \binom{N-1}{j} (1-p)^{N-1-j} p^j (1-P)^j \tag{2}$$

The summation on the right-hand side of Equation 2 can be solved according to the binomial theorem.

$$1 - P = (1 - pP)^{N-1} \tag{3}$$

$$\approx e^{-\langle k \rangle P} \tag{4}$$

where we employed the definition of $\langle k \rangle = Np$ to reach Equation 4.

We can analytically solve for P in Equation 4 by using the Lambert W function [31], also known as the product logarithm.

$$P = 1 + \frac{W(-\langle k \rangle e^{-\langle k \rangle})}{\langle k \rangle} \tag{5}$$

Our theory of the brain

We set up a self-consistent equation for P as we did for the random graph case (Equation 2). The main difference is that the right-hand side of the equation is now equal to the probability of the node not to form an edge, multiplied by

the probability that the nodes in the giant cluster are arranged on other nodes. We constrain cluster formation in this manner because we find that the brain has negligible secondary cluster formation (further discussed in the Results and Figure S2). Compared to Equation 2, this corresponds to only keeping the $j = 0$ term in the sum times the probability of the giant cluster being arranged among the $N - 1$ other nodes.

$$1 - P = (1 - p)^{N-1} \prod_{i=1}^{NP} \left(1 - \frac{i}{N}\right) \tag{6}$$

$$1 - P \approx (1 - p)^{N-1} \left(1 - \frac{1}{N}\right)^{NP} \tag{7}$$

$$\approx e^{-\langle k \rangle} e^{-P} \tag{8}$$

We show in detail the approximation of the product sum from Equation 6 to 7 in the Supplement (Equation S1).

Just as in the random graph case, we can analytically solve Equation 8 for P by using the Lambert W function.

$$P = 1 + W(-e^{-1-\langle k \rangle}) \tag{9}$$

Equation 9 has no parameters. However, when comparing with experimental data at finer parcellations, we find that $\langle k \rangle$ needs to be rescaled. For a discussion of the form of P at coarser parcellations, see Equation S4 and Figure S20. The rescaling means that the probability of forming an edge that grows the giant cluster is less than the probability of forming an edge between two nodes in the giant cluster. This could be a consequence of the brain's modular nature [66, 18, 67, 68, 69, 70]. At any point in time, those nodes in the giant cluster can be thought of as representing a module that are more likely to accumulate more and more edges between themselves, rather than bring in a new node into the giant cluster that previously formed no edges.

Introducing the rescaling factor as the parameter α, Equation 9 becomes

$$P = 1 + W(-e^{-1-\langle k \rangle/\alpha}) \tag{10}$$

The α parameter is the only parameter in our expression and it is determined by fitting to experimental data. In Figure S22, we demonstrate that the analytic expression matches simulation results with an appropriate α based on the mechanism described.

Code and data availability

Scripts and processed data necessary to reproduce figures and conclusions reached in the text can be found at github.com/rrazban/percolating_brain.

Acknowledgments

The authors thank Elisha Moses and Botond Antal 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 LRMP and KD) and the White House Brain Research Through Advancing Innovative Technologies (BRAIN) Initiative (NSFNCS-FR 1926781 to LRMP and KD).

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 Development^{SM} (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 age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health 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, 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 Digital Object Identifier 10.15154/1526472].

Author Contributions

R.M.R., K.A.D., L.R.M.P. designed research, analyzed data, and wrote the manuscript. R.M.R. performed the research.

Declaration of Interests

The authors declare no competing interests.

References

[1] Gregory L Dignon, Robert B Best, and Jeetain Mittal. Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annual review of physical chemistry*, 71:53–75, 2020.

[2] William Bialek, Andrea Cavagna, Irene Giardina, Thierry Mora, Edmondo Silvestri, Massimiliano Viale, and Aleksandra M Walczak. Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences*, 109(13):4786–4791, 2012.

[3] Iain D Couzin. Collective cognition in animal groups. *Trends in cognitive sciences*, 13(1):36–43, 2009.

[4] Suzana Herculano-Houzel. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences*, 109(Supplement 1):10661–10668, 2012.

[5] Shi Gu, Fabio Pasqualetti, Matthew Cieslak, Qawi K Telesford, B Yu Alfred, Ari E Kahn, John D Medaglia, Jean M Vettel, Michael B Miller, Scott T Grafton, et al. Controllability of structural brain networks. *Nature communications*, 6(1):1–10, 2015.

[6] Aming Li, Sean P Cornelius, Y-Y Liu, Long Wang, and A-L Barabási. The fundamental advantages of temporal networks. *Science*, 358(6366):1042–1046, 2017.

[7] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378–382, 2000.

[8] Reuven Cohen, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. Resilience of the Internet to Random Breakdowns. *Phys. Rev. Lett.*, 85(21):4626–4628, 2000.

[9] Albert-László Barabási. *Network Science*. Cambridge University Press, 2016.

[10] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

[11] Sophie Achard, Raymond Salvador, Brandon Whitcher, John Suckling, and ED Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72, 2006.

[12] Martijn P Van Den Heuvel and Olaf Sporns. Rich-club organization of the human connectome. *Journal of Neuroscience*, 31(44):15775–15786, 2011.

[13] Danielle S Bassett and Edward T Bullmore. Small-world brain networks revisited. *The Neuroscientist*, 23(5):499–516, 2017.

[14] Danielle Smith Bassett and ED Bullmore. Small-world brain networks. *The neuroscientist*, 12(6):512–523, 2006.

[15] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.

[16] Richard F Betzel, Andrea Avena-Koenigsberger, Joaquín Goñi, Ye He, Marcel A De Reus, Alessandra Griffa, Petra E Vértes, Bratislav Mišic, Jean-Philippe Thiran, Patric Hagmann, et al. Generative models of the human connectome. *Neuroimage*, 124:1054–1064, 2016.

[17] Danyal Akarca, Petra E Vértes, Edward T Bullmore, and Duncan E Astle. A generative network model of neurodevelopmental diversity in structural brain organization. *Nature communications*, 12(1):1–18, 2021.

[18] Lazaros K Gallos, Hernán A Makse, and Mariano Sigman. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proceedings of the National Academy of Sciences*, 109(8):2825–2830, 2012.

[19] Paolo Moretti and Miguel A Muñoz. Griffiths phases and the stretching of criticality in brain networks. *Nature communications*, 4(1):1–10, 2013.

[20] Gino Del Ferraro, Andrea Moreno, Byungjoon Min, Flaviano Morone, Úrsula Pérez-Ramírez, Laura Pérez-Cervera, Lucas C Parra, Andrei Holodny, Santiago Canals, and Hernán A Makse. Finding influential nodes for integration in brain networks using optimal percolation theory. *Nature communications*, 9(1):1–12, 2018.

[21] Nicolas A Crossley, Andrea Mechelli, Jessica Scott, Francesco Carletti, Peter T Fox, Philip McGuire, and Edward T Bullmore. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain*, 137(8):2382–2395, 2014.

[22] Derek K Jones, Thomas R Knösche, and Robert Turner. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion mri. *Neuroimage*, 73:239–254, 2013.

[23] Dardo Tomasi, Gene-Jack Wang, and Nora D Volkow. Energetic cost of brain functional connectivity. *Proceedings of the National Academy of Sciences*, 110(33):13642–13647, 2013.

[24] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, 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 medicine*, 12(3):e1001779, 2015.

[25] Nora D Volkow, George F Koob, Robert T Croyle, Diana W Bianchi, Joshua A Gordon, Walter J Koroshetz, Eliseo J Pérez-Stable, William T Riley, Michele H Bloch, Kevin Conway, et al. The conception of the abcd study: From substance use to a broad nih collaboration. *Developmental cognitive neuroscience*, 32:4–7, 2018.

[26] Emer J Hughes, Tobias Winchman, Francesco Padormo, Rui Teixeira, Julia Wurie, Maryanne Sharma, Matthew Fox, Jana Hutter, Lucilio Cordero-Grande, Anthony N Price, et al. A dedicated neonatal brain imaging system. *Magnetic resonance in medicine*, 78(2):794–804, 2017.

[27] Seung Wook Oh, Julie A Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M Henry, et al. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, 2014.

[28] Thierry Mora and William Bialek. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302, 2011.

[29] Miguel A Munoz. Colloquium: Criticality and dynamical scaling in living systems. *Reviews of Modern Physics*, 90(3):031001, 2018.

[30] John M Beggs and Nicholas Timme. Being critical of criticality in the brain. *Frontiers in physiology*, 3:163, 2012.

[31] Abbas Ali Saberi. Recent advances in percolation theory and its applications. *Physics Reports*, 578:1–32, 2015.

[32] Fidel Alfaro-Almagro, Mark Jenkinson, Neal K Bangerter, Jesper LR Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Stamatios N Sotiropoulos, Saad Jbabdi, Moises Hernandez-Fernandez, Emmanuel Vallee, et al. Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank. *Neuroimage*, 166:400–424, 2018.

[33] Xue Chen, Joe Necus, Luis R Peraza, Ramtin Mehraram, Yanjiang Wang, John T O'Brien, Andrew Blamire, Marcus Kaiser, and John-Paul Taylor. The functional brain favours segregated modular connectivity at old age unless affected by neurodegeneration. *Communications biology*, 4(1):1–16, 2021.

[34] Yinghui Zhang, Yin Wang, Nan Chen, Man Guo, Xiuzhen Wang, Guangcai Chen, Yongchao Li, Lin Yang, Shan Li, Zhijun Yao, et al. Age-associated differences of modules and hubs in brain functional networks. *Frontiers in aging neuroscience*, 12:511, 2021.

[35] Jordi Soriano, María Rodríguez Martínez, Tsvi Tlusty, and Elisha Moses. Development of input connections in neural cultures. *Proceedings of the National Academy of Sciences*, 105(37):13758–13763, 2008.

[36] Or Cohen, Anna Keselman, Elisha Moses, M Rodríguez Martínez, Jordi Soriano, and Tsvi Tlusty. Quorum percolation in living neural networks. *EPL (Europhysics Letters)*, 89(1):18008, 2010.

[37] Jean-Pierre Eckmann, Ofer Feinerman, Leor Gruendlinger, Elisha Moses, Jordi Soriano, and Tsvi Tlusty. The physics of living neural networks. *Physics Reports*, 449(1-3):54–76, 2007.

[38] Hugo Lagercrantz. *Infant brain development*. Springer, 2016.

[39] Richard AI Bethlehem, Jakob Seidlitz, Simon R White, Jacob W Vogel, Kevin M Anderson, Chris Adamson, Sophie Adler, George S Alexopoulos, Evdokia Anagnostou, Ariosky Areces-Gonzalez, et al. Brain charts for the human lifespan. *Nature*, pages 1–11, 2022.

[40] Marcus Kaiser. Mechanisms of connectome development. *Trends in Cognitive Sciences*, 21(9):703–717, 2017.

[41] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven A Siegelbaum, and A James Hudspeth. *Principles of Neural Science*. New York: MacGrawHill Medical, 2013.

[42] Debasish Chaudhuri, Peter Borowski, and Martin Zapotocky. Model of fasciculation and sorting in mixed populations of axons. *Physical Review E*, 84(2):021908, 2011.

[43] Sreedevi Varier and Marcus Kaiser. Neural development features: spatio-temporal development of the caenorhabditis elegans neuronal network. *PLoS computational biology*, 7(1):e1001044, 2011.

[44] Sophie Achard and Ed Bullmore. Efficiency and cost of economical brain functional networks. *PLoS computational biology*, 3(2):e17, 2007.

[45] Patric Hagmann, Maciej Kurant, Xavier Gigandet, Patrick Thiran, Van J Wedeen, Reto Meuli, and Jean-Philippe Thiran. Mapping human whole-brain structural networks with diffusion mri. *PloS one*, 2(7):e597, 2007.

[46] Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford university press, 2006.

[47] Ed Bullmore and Olaf Sporns. The economy of brain network organization. *Nature reviews neuroscience*, 13(5):336–349, 2012.

[48] Mikail Rubinov, Rolf JF Ypma, Charles Watson, and Edward T Bullmore. Wiring cost and topological participation of the mouse brain connectome. *Proceedings of the National Academy of Sciences*, 112(32):10032–10037, 2015.

[49] Mikail Rubinov. Constraints and spandrels of interareal connectomes. *Nature communications*, 7(1):1–11, 2016.

[50] Richard F Betzel and Danielle S Bassett. Specificity and robustness of long-distance connections in weighted, interareal connectomes. *Proceedings of the National Academy of Sciences*, 115(21):E4880–E4889, 2018.

[51] Marcus Kaiser and Claus C Hilgetag. Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS computational biology*, 2(7):e95, 2006.

[52] Siân Wilson, Maximilian Pietsch, Lucilio Cordero-Grande, Anthony N Price, Jana Hutter, Jiaxin Xiao, Laura McCabe, Mary A Rutherford, Emer J Hughes, Serena J Counsell, et al. Development of human white matter pathways in utero over the second and third trimester. *Proceedings of the National Academy of Sciences*, 118(20), 2021.

[53] Shadab Khan, Lana Vasung, Bahram Marami, Caitlin K Rollins, Onur Afacan, Cynthia M Ortinau, Edward Yang, Simon K Warfield, and Ali Gholipour. Fetal brain growth portrayed by a spatiotemporal diffusion tensor mri atlas computed from in utero images. *NeuroImage*, 185:593–608, 2019.

[54] Danielle S Bassett and Edward T Bullmore. Human brain networks in health and disease. *Current opinion in neurology*, 22(4):340, 2009.

[55] Hannelore Aerts, Wim Fias, Karen Caeyenberghs, and Daniele Marinazzo. Brain networks under attack: robustness properties and the impact of lesions. *Brain*, 139(12):3063–3083, 2016.

[56] Arun S Mahadevan, Nicolas E Grandel, Jacob T Robinson, Kevin R Francis, and Amina A Qutub. Living neural networks: Dynamic network analysis of developing neural progenitor cells. *bioRxiv*, page 055533, 2018.

[57] Harpreet Setia and Alysson R Muotri. Brain organoids as a model system for human neurodevelopment and disease. In *Seminars in cell & developmental biology*, volume 95, pages 93–97. Elsevier, 2019.

[58] Abed AlFatah Mansour, J Tiago Gonçalves, Cooper W Bloyd, Hao Li, Sarah Fernandes, Daphne Quang, Stephen Johnston, Sarah L Parylak, Xin Jin, and Fred H Gage. An in vivo model of functional and vascularized human brain organoids. *Nature biotechnology*, 36(5):432–441, 2018.

[59] Donald J Hagler Jr, SeanN Hatton, M Daniela Cornejo, Carolina Makowski, Damien A Fair, Anthony Steven Dick, Matthew T Sutherland, BJ Casey, Deanna M Barch, Michael P Harms, et al. Image processing and analysis methods for the adolescent brain cognitive development study. *Neuroimage*, 202:116091, 2019.

[60] Daan Christiaens, Lucilio Cordero-Grande, Maximilian Pietsch, Jana Hutter, Anthony N Price, Emer J Hughes, Katy Vecchiato, Maria Deprez,

A David Edwards, Joseph V Hajnal, et al. Scattered slice shard reconstruction for motion correction in multi-shell diffusion mri. *Neuroimage*, 225:117437, 2021.

[61] Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, 8:8, 2014.

[62] Jack L Lancaster, Marty G Woldorff, Lawrence M Parsons, Mario Liotti, Catarina S Freitas, Lacy Rainey, Peter V Kochunov, Dan Nickerson, Shawn A Mikiten, and Peter T Fox. Automated talairach atlas labels for functional brain mapping. *Human brain mapping*, 10(3):120–131, 2000.

[63] Nikos Makris, Jill M Goldstein, David Kennedy, Steven M Hodge, Verne S Caviness, Stephen V Faraone, Ming T Tsuang, and Larry J Seidman. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia research*, 83(2-3):155–171, 2006.

[64] Fischl Bruce. Freesurfer. *NeuroImage*, 62(2):774–781, 2012.

[65] Béla Bollobás. *Random graphs*. Number 73. Cambridge university press, 2001.

[66] David Meunier, Renaud Lambiotte, and Edward T Bullmore. Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, 4:200, 2010.

[67] Shu Guo, Xiaoqi Chen, Yimeng Liu, Rui Kang, Tao Liu, and Daqing Li. Percolation analysis of brain structural network. *Frontiers in Physics*, 9:345, 2021.

[68] Cécile Bordier, Carlo Nicolini, and Angelo Bifone. Graph analysis and modularity of brain functional connectivity networks: searching for the optimal threshold. *Frontiers in neuroscience*, 11:441, 2017.

[69] Christopher W Lynn and Danielle S Bassett. The physics of brain network structure, function and control. *Nature Reviews Physics*, 1(5):318–332, 2019.

[70] Muhua Zheng, Antoine Allard, Patric Hagmann, Yasser Alemán-Gómez, and M Ángeles Serrano. Geometric renormalization unravels self-similarity of the multiscale human connectome. *Proceedings of the National Academy of Sciences*, 117(33):20244–20253, 2020.

[71] Francesca Melozzi, Eyal Bergmann, Julie A Harris, Itamar Kahn, Viktor Jirsa, and Christophe Bernard. Individual structural features constrain the mouse functional connectome. *Proceedings of the National Academy of Sciences*, 116(52):26961–26969, 2019.

[72] Giorgia Grisot, Suzanne N Haber, and Anastasia Yendiki. Diffusion mri and anatomic tracing in the same brain reveal common failure modes of tractography. *NeuroImage*, page 118300, 2021.

Supplementary Information

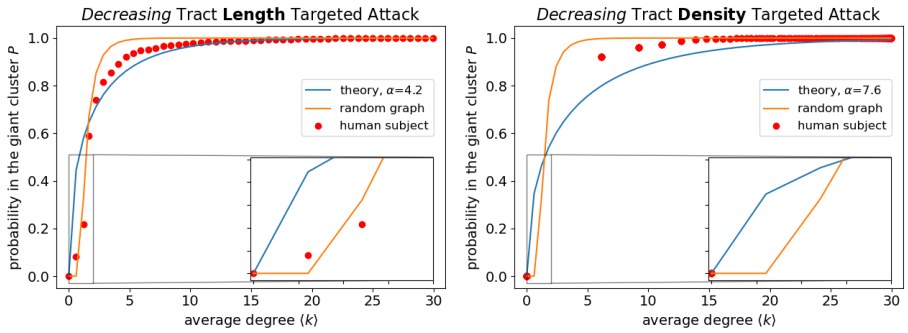

Figure S1: Targeted attack on *decreasing* tract lengths and tract densities. Note that throughout the main text, we focus on targeted attack of increasing tract lengths and densities and derive an analytic solution based on our theory (Equation 1), which is also shown in this figure for comparison. Targeted attack on decreasing tract lengths essentially follows a random graph. It is not possible to fully assess the corresponding P curve of tract density because of the large gap at low average degree. Curves are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1.

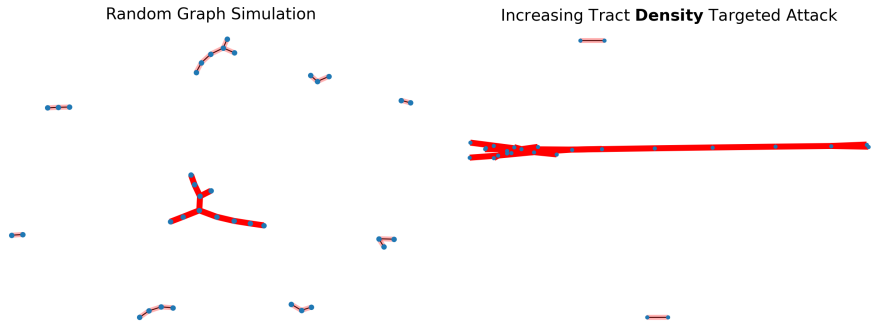

Figure S2: Visualizing cluster formation at average degree $\langle k \rangle \sim 1$ for both a random graph simulation (left) and the brain under increasing tract density targeted attack (right). The random graph simulation has several non-negligible secondary clusters, while the brain essentially has only one densely connected, large giant cluster. The random graph simulation has the same number of nodes as that of the brain shown analyzed under the Harvard-Oxford atlas, 64 total regions, however we exclude siloed nodes in both graphs for ease of visualization. The brain's cluster distribution is derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1, however, analyzed at a coarser parcellation to easily visualize clusters.

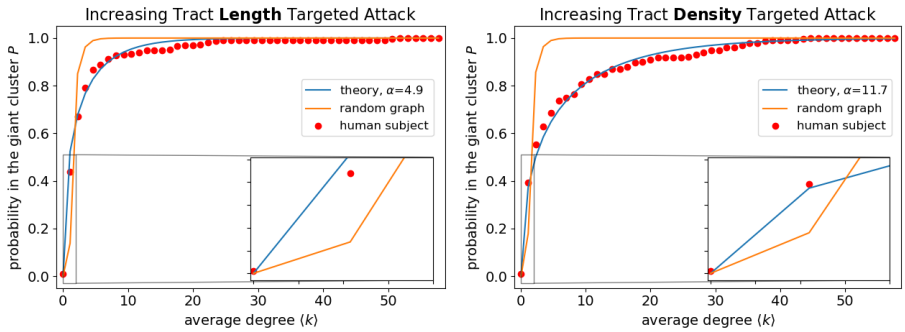

Figure S3: A coarser parcellation leads to a smaller fitted α parameter. The coarser parcellation has 99 regions and is created by randomly merging neighboring Talairach atlas regions. Curves are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1.

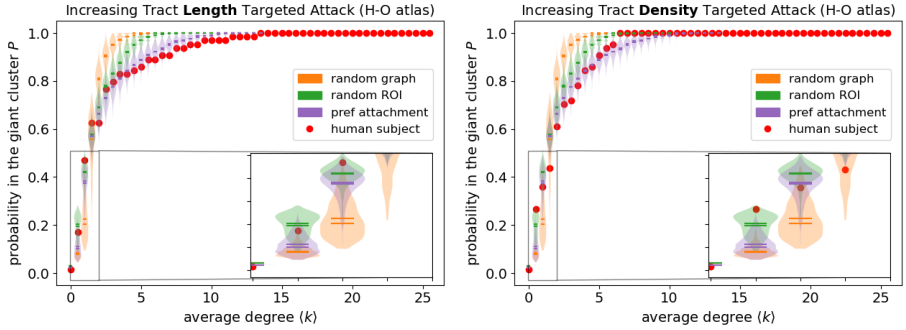

Figure S4: Competing models are able to roughly capture P curves at a coarse parcellation, the Harvard-Oxford (H-O) atlas. Competing models are presented as violin plots of simulations (1,000 independent runs) because some models such as the random ROI and preferential attachment models do not have known analytic expressions across the entirety of $\langle k \rangle$. The random region of interest (ROI) model is a random graph model constrained by center of mass distances between regions determined by the atlas. Experimental curves are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1.

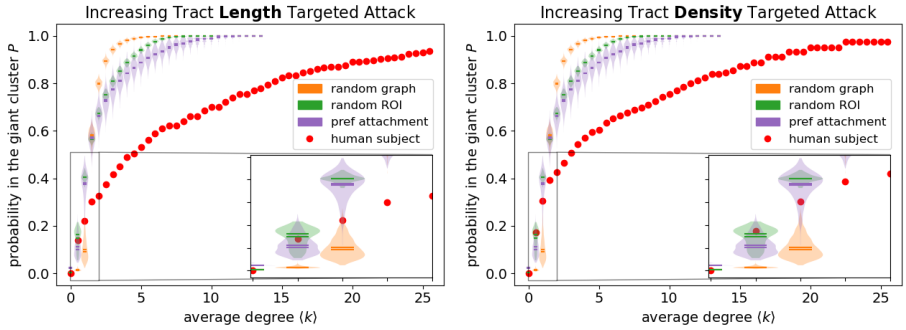

Figure S5: Competing models are not able to capture P curves tabulated under the Talairach atlas. Violin plots for the competing models are determined as in Figure S4 for the Talairach atlas. Curves are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1.

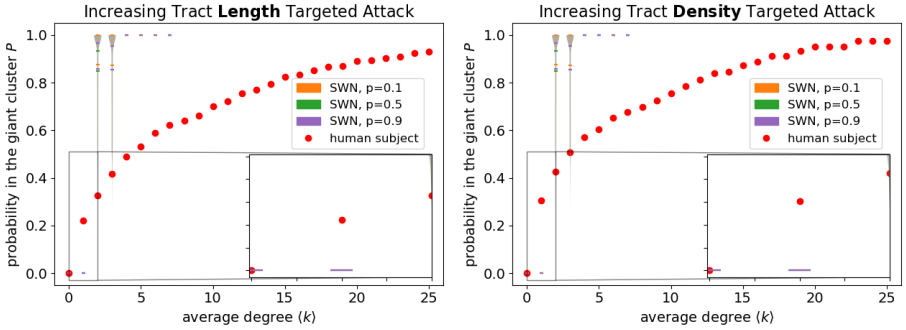

Figure S6: Competing small-world network (SWN) models are not able to capture P curves tabulated under the Talairach atlas. Small-world networks are defined as in [10] and the parameter p in the legend corresponds to the probability of rewiring each edge. A sharp transition in SWN P curves are seen because SWN is essentially a 2-dimensional ring lattice. Lattices are known to have critical points in their P curves [9]. Violin plots for the competing models are determined as in Figure S5. Curves are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1.

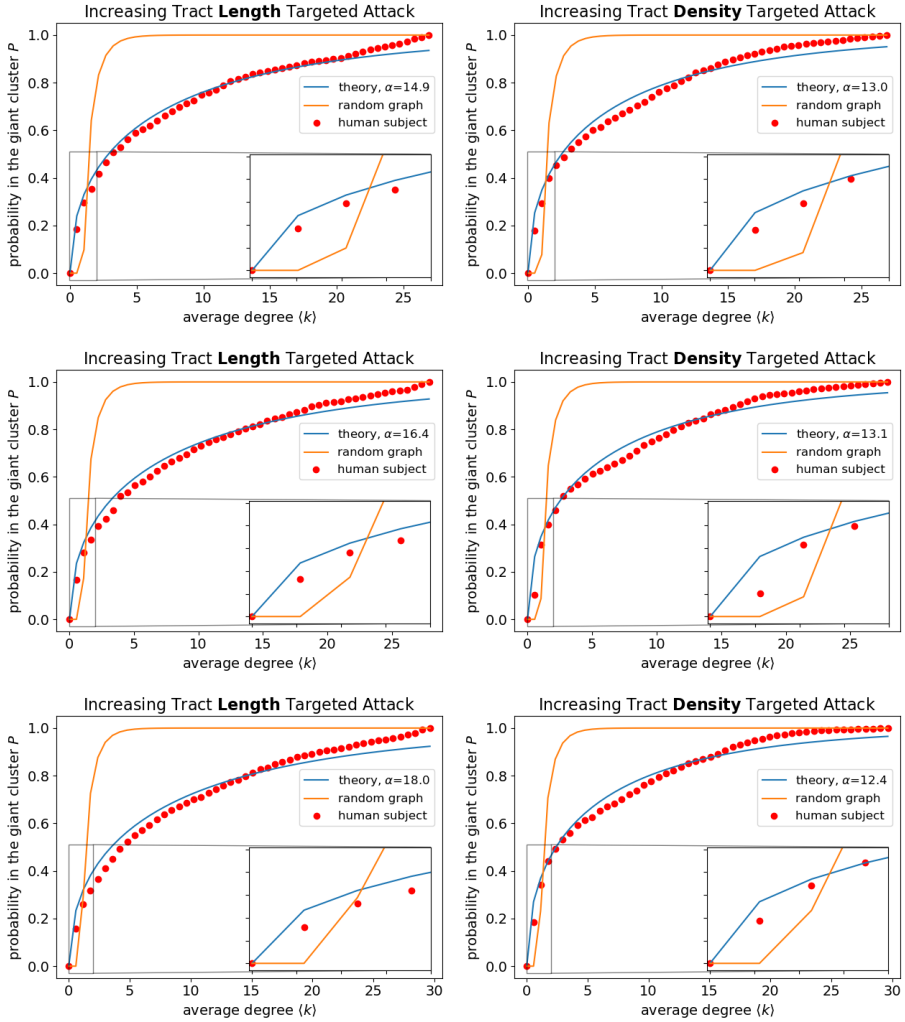

Figure S7: P curves for additional individuals besides that shown in Figure 1. From the top to bottom row, the UK Biobank subject IDs are 1701210, 1363111 and 2460658. The fitted α parameters are fairly similar among different subjects.

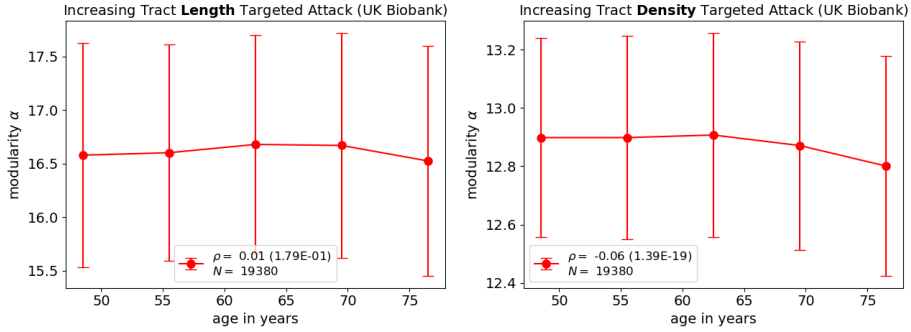

Figure S8: Our fitted α parameter does not systematically vary with age for both targeted attack by increasing tract length and tract density. Binned data is presented with a line connecting means and error bars correspond to standard deviations. The variable ρ corresponds to the Spearman correlation coefficient calculated over all individuals, in parenthesis is the p-value. The variable N corresponds to the total number of individuals from the UK Biobank included in the analysis.

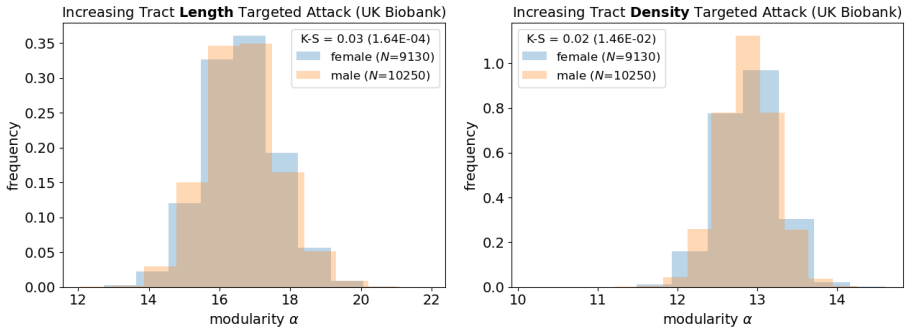

Figure S9: Distributions of our fitted α parameter does not systematically vary between individuals of different sexes. K-S corresponds to the Kolmogorov-Smirnov test statistic, in parenthesis is the p-value. The variable N corresponds to the total number of individuals with the corresponding sex from the UK Biobank.

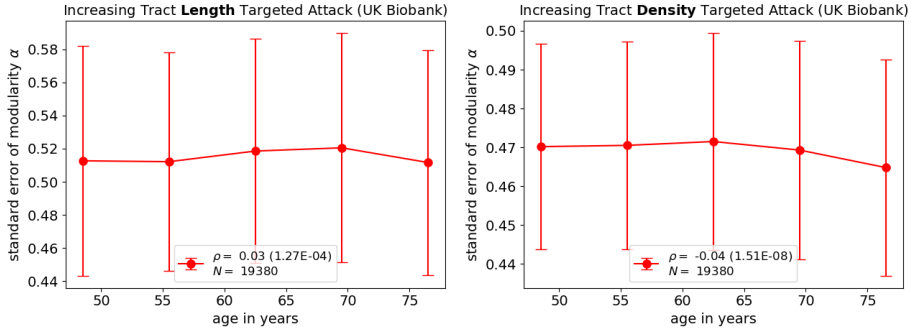

Figure S10: Quality of α parameter fits in Figure S8 does not greatly vary as a function of age. Binned data is presented with a line connecting means and error bars correspond to standard deviations. The variable ρ corresponds to the Spearman correlation coefficient calculated over all individuals, in parenthesis is the p-value. The variable N corresponds to the total number of individuals from the UK Biobank included in the analysis.

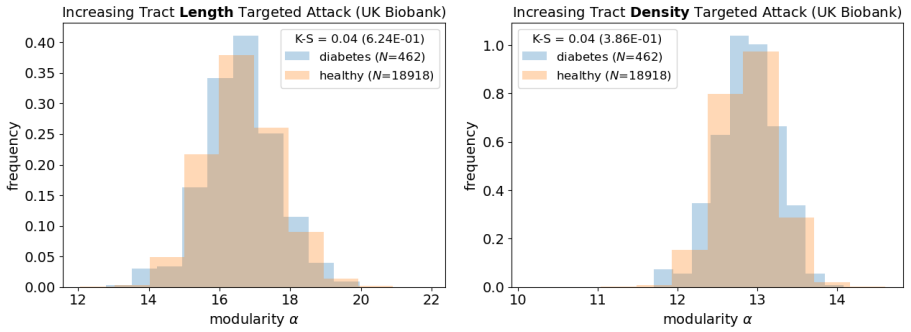

Figure S11: Distributions of our fitted α parameter does not systematically vary between individuals diagnosed with diabetes and those not diagnosed with diabetes (healthy). K-S corresponds to the Kolmogorov-Smirnov test statistic, in parenthesis is the p-value. The variable N corresponds to the total number of individuals with the corresponding diabetes diagnosis from the UK Biobank.

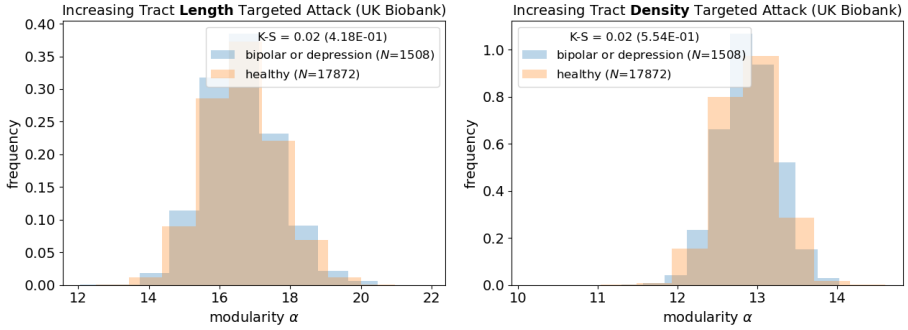

Figure S12: Distributions of our fitted α parameter does not systematically vary between individuals diagnosed with bipolar disorder or depression and those not diagnosed (healthy). K-S corresponds to the Kolmogorov-Smirnov test statistic, in parenthesis is the p-value. The variable N corresponds to the total number of individuals with the corresponding bipolar disorder and depression diagnosis from the UK Biobank.

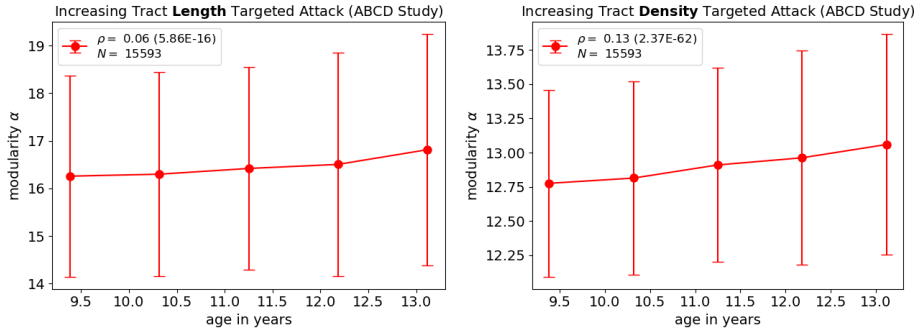

Figure S13: Our fitted α parameter does not vary much with age for both targeted attack by increasing tract length and density for the Adolescent Brain Cognitive Development (ABCD) Study. Binned data is presented with a line connecting means and error bars correspond to standard deviations. The variable ρ corresponds to the Spearman correlation coefficient calculated over all individuals, in parenthesis is the p-value. The variable N corresponds to the total number of brain scans from the ABCD Study. Remarkably, α values are similar to those seen in the UK Biobank despite ABCD being an independent data set containing much younger individuals (Figure S8).

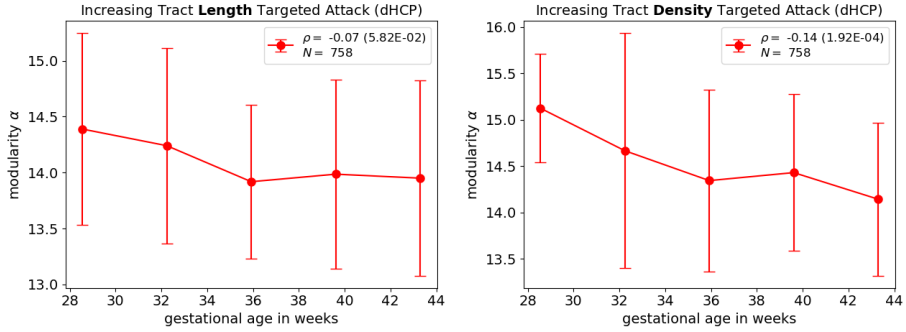

Figure S14: Our fitted α parameter does not vary much with age for both targeted attack by increasing tract length and density for the Developing Human Connectome Project (dHCP). Binned data is presented with a line connecting means and error bars correspond to standard deviations. The variable ρ corresponds to the Spearman correlation coefficient calculated over all individuals, in parenthesis is the p-value. The variable N corresponds to the total number of brain scans from the dHCP. The α values are not as similar as those between the ABCD Study and UK Biobank (Figure S13 and S8), perhaps reflecting larger preprocessing differences (details in the Methods). Nonetheless, α values are seen to be stable across ages just as in the other two data sets.

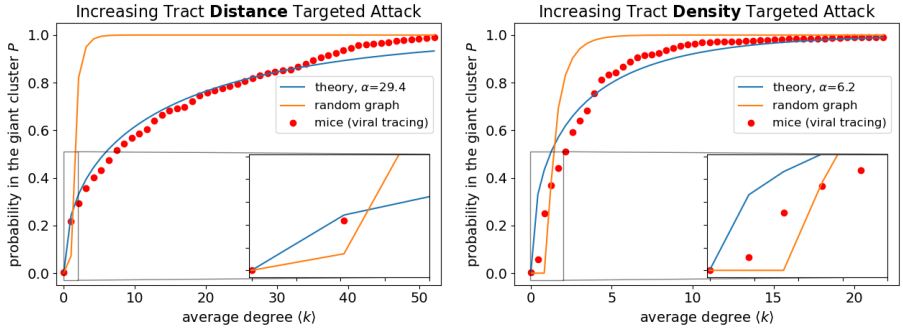

Figure S15: Our theory captures the probability for a brain region to be in the giant cluster across the average degree range with an appropriately fitted α parameter for both targeted attack by increasing distance between regions' center of masses and tract density. Curves are derived from viral tracing data by the Allen Institute averaged over ∼1000 mice [27]. While α values for the two human P curves are comparable at 17.9 and 13.3 (Figure 1), the α value needed to fit the distance P curve is around five times larger than that of the tract density for mouse (Figure 15). The relative differences do not seem to be a result of viral tracing being limited to distances; we find similar trends if we look at corresponding distances for human results (Figure S16). Also, the resolution of the atlas does not seem to be factor; the atlas for the presented mouse results contains 426 grey matter regions, while that for human contains 727 grey matter regions. Adjusting the atlas used in processing the dMRI to calculate the connectivity matrix to 99 regions, still only yields α(distance)/α(density) $\sim$ 2 (Figure S3). Further investigation is needed to elucidate whether relative α differences are a result of differences between the dMRI and viral tracing data modalities, as has been recently highlighted in other studies [71, 72].

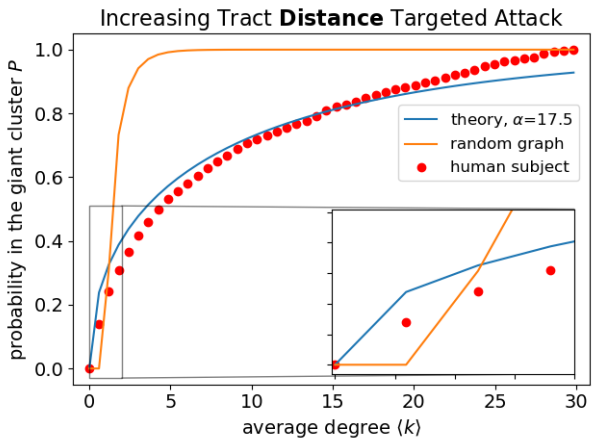

Figure S16: Targeted attack based on center of mass distances of Talairach atlas regions yields similar results as compared to tract lengths. Curves are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1.

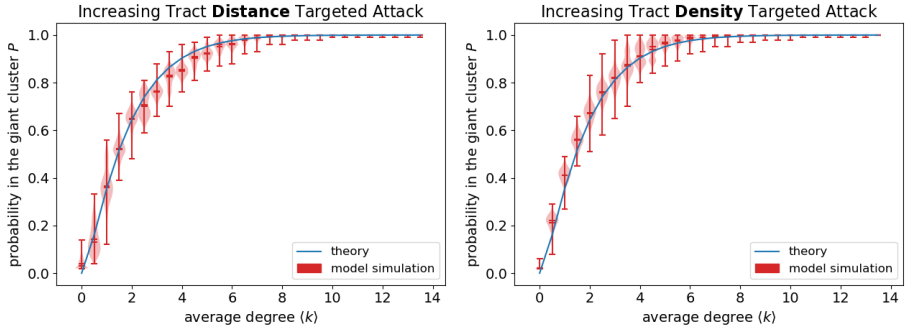

Figure S17: Corresponding P curves for simulations of our model (Figure 3) matches our theory and thus, are consistent with experimental P curves (Figure 1). Violin plots represent 1000 independent runs. There are many ways of encoding growth mechanism details; however, similar results are obtained as long as the main concepts of scaling fasciculation are present. For simplicity, we encode the growth mechanism in two dimensions as an expanding circle with a fixed growth rate. The first nodes are randomly added in the first rung. When that rung fills up, all the placed nodes are moved up one rung, and then new nodes are continually added to fill up the now empty lower rung and the partially filled upper rung. We arbitrarily assume that each rung can hold $\text{int}(2 * \pi r/3)$, where r is the rung radius from the origin and assumed to be a positive integer. We continue adding nodes to rungs in this manner until 100 nodes are placed. Edges are concurrently added as nodes are placed and are assumed to lengthen as nodes move apart during the course of growth. The final edge distance is calculated as the Euclidean distance between the two nodes. All present edges become stochastically denser by a number sampled uniformly from 1 to 10 at each growth step. Stochasticity is necessary to avoid many edges having the same density.

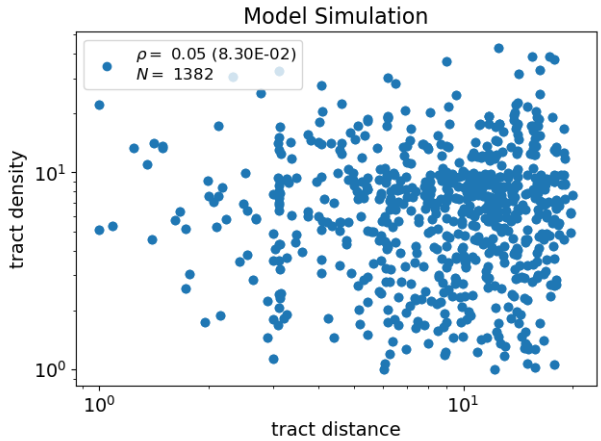

Figure S18: Tract distance and density relationship for one simulation of the model outlined in Figure 3 and described in detail in Figure S17. The Spearman correlation coefficient (ρ) is close to 0, where the p-value is in parenthesis. The variable N corresponds to the total number of edges. Although absolute values of the model's density and length do not match corresponding experimental values, we can trivially rescale them without changing the rank correlation.

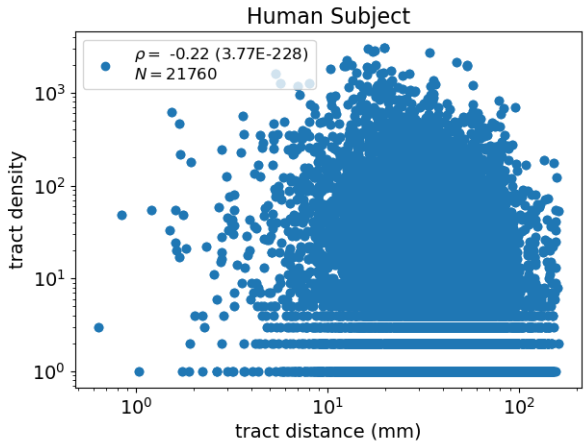

Figure S19: Tract distance and density are negatively correlated with a not very strong magnitude Spearman correlation coefficient (ρ), where the p-value is in parenthesis. The variable N corresponds to the total number of tracts for a given individual derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1.

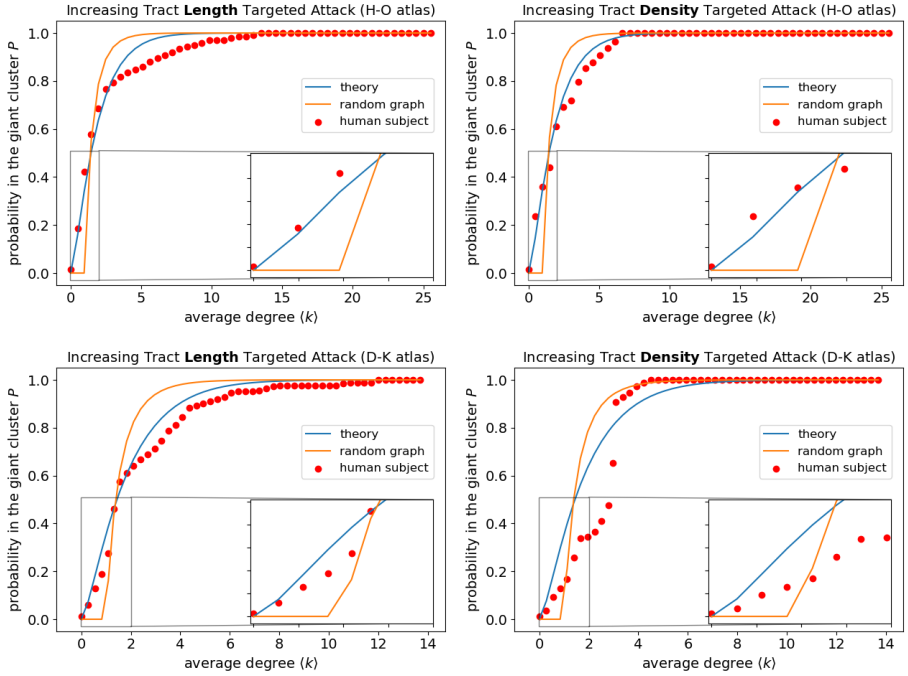

Figure S20: A simpler form of our theory without α mostly captures P curves generated from different atlases: Harvard-Oxford (H-O) and the modified Desikan-Killiany (D-K). Both atlases are much more coarse than the Talairach atlas used throughout the main text; Harvard-Oxford and the modified Desikan-Killiany atlases have 64 and 84 regions considered in our analysis, respectively. Curves are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1. The simpler form of our theory is discussed in detail in Equation S4 and requires no α parameter. It is not clear why the Desikan-Killiany tract density targeted attack deviates from our theory. The results look somewhat similar to a random graph and demonstrate a sharp rise in its P curve. It seems that coarser atlases can miss unique features identified in the main text when using finer atlases.

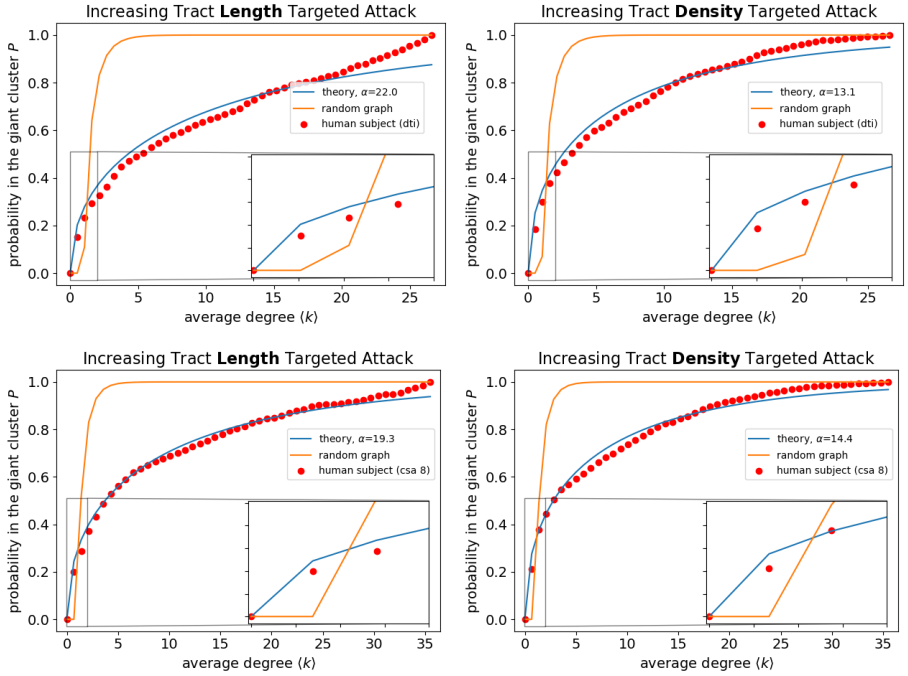

Figure S21: Changing the default procedure in the Dipy software to calculate the connectivity matrix from dMRI data has little effect on the P curves. In the top two figures, we use the diffusion tensor imaging (dti) model, rather than the constant solid angle method used in the main text. In the bottom two figures, we alter the constant solid angle (csa) method to order 8, compared to order 6 in the main text. Curves are derived from dMRI data collected by the UK Biobank for one human individual (subject ID: 6025360), the same subject as in Figure 1. Comparing to Figure 1, fitted α values capture the P curves.

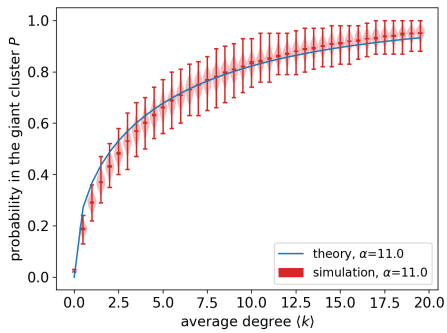

Figure S22: Simulation results verify Equation 10. Simulation violin plots represent 1000 independent runs for a graph with 100 nodes. The probability that an edge is added between nodes in the giant cluster is set to p, while the probability that an edge is added between a giant cluster node and a non-giant cluster node is $p/5$, resulting in $\alpha = 5 * 2 + 1 = 11$. Based on the giant cluster preservation mechanism, any two nodes not in the giant cluster have no chance of forming an edge.

Approximation from Equation 6 to 7

We first expand the product sum in Equation 6 of the Methods,

$$\prod_{i=1}^{NP}\left(1-\frac{i}{N}\right) = \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)...\left(1-\frac{NP}{N}\right) \tag{S1}$$

$$= \left(1-\frac{1}{N}\right)\left(1-\frac{1}{N}-\frac{1}{N}\right)...\left(1-\frac{1}{N}-\frac{NP+1}{N}\right) \tag{S2}$$

$$\approx \left(1-\frac{1}{N}\right)^{NP} \tag{S3}$$

The approximation in the final line is valid for small P. As P approaches 1, the approximation breaks down. This explains the relatively poor agreement with experiment at large P values (Figure 1). This effect is also seen in simulations of our theory (Figure S22).

Theory for giant cluster preservation with no modules

At coarser parcellations (Figure S20), we find that the α parameter is not necessary and the form of the P curve closely follows that of Equation 8, with an additional multiplicative factor on the right hand-side.

$$1 - P \approx e^{-\langle k\rangle} e^{-P}(1 + \langle k\rangle + \langle k\rangle^2/2) \tag{S4}$$

The $\langle k \rangle$ and $\langle k \rangle^2/2$ terms can be reasoned to be present based on the Molloy-Reed criterion, which states that each node must have more than two edges to propagate a giant cluster [9]. Thus, even if a node has one edge ($\langle k \rangle$ term) or two edges ($\langle k \rangle^2/2$ term) on average, the giant cluster cannot exist. It is worth noting that the same giant cluster preservation mechanism as described in the main text is still at play, however, without modularity. Equation S4 is verified by corresponding simulation results (Figure S23).

Interestingly, our theory in the main text seems to violate the Molloy-Reed criterion by lacking both $\langle k \rangle$ and $\langle k \rangle^2/2$ terms in Equation 8. We reason that the modular structure enables giant cluster formation with the presence of just a single edge on a single node because the giant cluster is so densely connected.

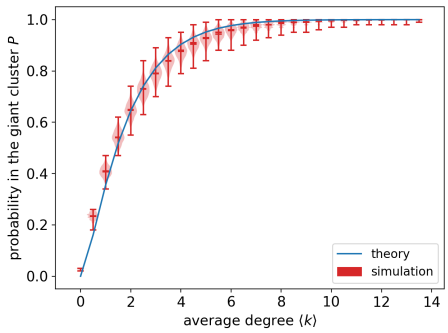

Figure S23: Simulation results verify Equation S4. Simulation violin plots represent 1000 independent runs for a graph with 100 nodes. The probability that any edge between at least one giant cluster node is set to $p = 14/100$. Based on the giant cluster preservation mechanism, any nodes not in the giant cluster have no chance of forming an edge.