Supporting Information

Vértes et al. 10.1073/pnas.1111738109

SI Text 1. Effect of Model Parameters η and γ

1.1. Distance Penalty (η) in the Exponential Decay Model. Distance penalty (η) in the exponential decay model is addressed in Fig. S1.

1.2. Effect of Parameters η and γ in the Economical Clustering Model. Effect of parameters η and γ in the economical clustering model is illustrated in Figs. S2 and S3.

SI Text 2. Simulated Annealing

Simulated annealing (SA) is a widely used and generally robust method of optimization (1). It is most often used for optimizing multiparameter functions, which may be difficult to optimize by more computationally expedient methods. Its main disadvantages are that it can be computationally costly and (like most optimization methods) it can get stuck in local minima. Here we used SA to estimate the generative model parameters that minimized an energy function or, equivalently, maximized similarity between simulated and experimental [functional MRI (fMRI)] networks (*SI Text 4*).

To mitigate the risk of local minimization, each optimization was estimated by 10 different annealing processes, starting twice from five different initial parameter values: $(\eta,\gamma)=(0,0)$; $(\eta,\gamma)=(0,5)$; $(\eta,\gamma)=(5,5)$; $(\eta,\gamma)=(5,0)$; and $(\eta,\gamma)=(2.5,2.5)$. Fig. S4 shows that the jitter in parameter values obtained by these separate optimization runs is small compared with the differences in parameters between the healthy volunteer and childhood onset schizophrenia (COS) datasets.

We note that the residual jitter in parameter values estimated by different annealing runs likely results from two independent sources of variability. First, SA may not find exactly the same parameters for two runs on a given energy landscape. The second factor is the stochastic nature of our generative model, which leads to variations in the network structures—and therefore the value of the energy function—obtained in repeated runs (even if the model parameters are identically estimated by different annealing runs). In effect, this means that the energy landscape is fluctuating over time during the optimization process, and we are looking for a point that is optimal on average. SA is therefore a convenient optimization method because it avoids calculating local gradients and instead visits various points in parameter space.

SI Text 3. Evaluation of Generative Models

In the main text, we have focused on three possible generative models: (*i*) a simple, one-parameter model of connection probability as a function of distance (Eq. **1**); (*ii*) an economical preferential attachment, two-parameter model of connection probability as a function of distance and the degrees of the connected nodes (Eq. **2**); and (*iii*) an economical clustering, two-parameter model of connection probability as a function of distance and the clustering of connectivity between nodes (Eq. **3**). In addition, we evaluated nine other possible generative models that were variants on these thematic exemplars.

For each of the 12 models considered in total, we summarize in Table S1 its name, abbreviation, specification, optimal parameter values, and goodness-of-fit statistics. As detailed in *SI Text 4*, below, the goodness-of-fit measures include individual *P* values for the difference in each of four key network properties between simulated and experimental networks, as well as the minimum of the energy function (the inverse of the product of these four *P* values).

Fig. S5 shows that among the four best models from Table S1 ($E_2<10^6$), the economical clustering model provides the best fit to the distance distribution, which was left unconstrained during annealing. Note that all variants of the economical preferential attachment model (two of which are shown in Fig. S5) yield similarly bad fits to the distance distribution.

SI Text 4. Choice and Validation of an Energy Function

To measure how well a set of model networks fits the data, we take the (rescaled) difference between the mean values of each measure in the two samples (multiple probabilistic realizations of the generative model vs. multiple experimental (fMRI) datasets):

$$T_C=\frac{\left|\langle C_{data}\rangle-\langle C_{model}\rangle\right|}{SE_C}$$

$$T_E=\frac{\left|\langle E_{data}\rangle-\langle E_{model}\rangle\right|}{SE_E}$$

$$T_M=\frac{\left|\langle M_{data}\rangle-\langle M_{model}\rangle\right|}{SE_M},$$

where C, E, and M denote clustering, efficiency, and modularity, respectively, and SE in the denominator is the standard error on the mean of the numerator (this is required to take model and data variability into account in the comparison). These distance measures, T, are therefore simply the t statistics for the t tests between model and data for each measure. The degree distribution is an exception to this because it is far from normally distributed, and its mean is not an appropriate characterization of the distribution. For this case, we use a Kolmogorov-Smirnov (KS) test to calculate the distance KS_{deg} between the observed and simulated degree distributions.

For an energy function, we now need to combine four distance terms (T_C, T_E, T_M, and KS_{deg}) into just one expression. An obvious choice is to use a product (or sum) of the inverse of the four terms: $E_1=1/(T_C.T_E.T_M.KS_{deg})$. Note that the inverse operation is needed only if we prefer to minimize the energy function rather than maximize it.

However, because KS_{deg} is not scaled to the other measures it may, in this case, be more suitable to use directly the P value of the t test (or KS test) instead of the t statistic (or KS distance) for each measure. We then have: $E_2=1/(p_C.p_E.p_M.p_{deg})$ or $E_3=1/p_C+1/p_E+1/p_M+1/p_{deg}$.

In this article, we have principally used E_2 as the energy function minimized by SA. We have found the parameter values that minimize E_2 for any given generative model; and we have used the minimum value of E_2 (associated with optimal parameters) as a goodness-of-fit metric to benchmark different generative models in terms of the closeness of their approximations to experimental brain networks derived from the same set of fMRI data. Generative models that more closely approximate experimental networks will have relatively low energy functions at optimization. As shown in Table S1, four models (including economical clustering) set themselves apart by yielding very low energy values. Among the four best-fit models, we have focused on economical clustering because this model was the only one (of all 12 models considered) also capable of approximately capturing the distance distribution of connections and the probability distribution of nodal topological properties—two requirements not built into the annealing procedure (Fig. S5).

However, we could have used either E_1 or E_3, among other possible choices of energy function. To demonstrate that the results reported on the basis of minimizing E_2 were not unduly biased by the precise form of the energy function, we also used E_1 and E_3 as a basis for SA of the parameters for all 12 probabilistic models considered.

As shown in Fig. S6, the relative rankings of the models in terms of energy minimization is fairly consistent across different choices of energy function, especially among the better-fitting models.

We found that, as shown in Fig. S7, the variation between parameters of the same model estimated by annealing over different energy functions (E_1, E_2, and E_3) was small compared with the variation between parameters of different models.

Both these results suggest that the specific details of the energy function do not materially affect the key results reported on the basis of optimizing E_2.

SI Text 5. Effects of Network Connection Density

In the main text, we have focused on comparing modeled networks with experimental networks at a connection density of 4%. This connection density was chosen because it maximized the topological differences in network organization between healthy volunteer and schizophrenic groups. However, to demonstrate that other aspects of the analysis, such as the relative goodness of fit of various generative models, do not depend sensitively on the connection density at which modeled and experimental networks are compared, we additionally report these results for networks at connection densities of 8% and 16%.

As shown in Table S2, the economical clustering model (Eq. **3**) provides a better approximation of the experimental data than either the exponential decay model (Eq. **1**) or the economical preferential attachment model (Eq. **2**) at all connection densities considered. Note that denser networks become more random and therefore less particular to the data being analyzed. We also avoid overly sparse networks, which are largely disconnected, and therefore focus on the range 4–16%.

We also note that we focused on modeling the probability of functional connection in a single (right) hemisphere of the cerebral cortex. As briefly mentioned in *Materials and Methods*, we did so because we assumed that the distance between functionally connected regions is more likely to faithfully represent the length of an anatomical connection between them when both regions are located in the same hemisphere.

SI Text 6. Validation of Generative Models

6.1. Unconstrained Degree Distribution. Unconstrained degree distribution is illustrated in Fig. S8.

6.2. Validation of Generative Models on an Independent Dataset of Normal fMRI Networks. In the main text, the performance of the economical clustering model was compared with other possible generative models by analysis of a set of fMRI data collected on 20 healthy volunteers. This approach is potentially open to criticism in that the model parameters were estimated on the same experimental dataset that was also used to quantify model goodness of fit. To address this potential circularity, we acquired fMRI data (using identical experimental procedures) on a second independent sample of normal volunteers. Using the generative model parameters estimated from the first fMRI dataset, we then evaluated the goodness of fit of the modeled networks in comparison with the brain networks derived from the second experimental dataset.

As shown in Fig. S9, the economical clustering model with parameter values estimated from the first experimental dataset was able to approximate closely the clustering, efficiency, modularity, degree distribution, and distance distribution estimated in the second experimental dataset. Thus, the model is able to provide a good fit to experimental brain networks even when it is evaluated in the context of experimental data that were not used to optimize its parameters.

SI Text 7. Effects of Schizophrenia on Generative Model Parameters

Effects of schizophrenia on generative model parameters are illustrated in Fig. S10.

SI Text 8. Optimization and Tradeoffs

If we knew the function optimized by brain networks, we could use this optimality function to inform our analysis. However, the optimality function for brain network selection is not certainly known. Some degree of cost minimization is considered likely to be an important consideration for brain network formation. Brain networks are also topologically efficient. However, here we show that simple optimization of either connection distance (a measure of cost) or efficiency does not accurately reproduce the properties of brain networks.

To minimize cost (for a given number M of edges), one simply includes the M shortest edges available. If enforcing full connectivity, one begins by finding the minimum spanning tree over the edge lengths, then one adds edges in ascending order of connection distance until a total of M edges is reached. As shown in Table S4, this results in a highly clustered network that is very far from the data.

Brain networks are also significantly less efficient than random networks with the same number of nodes and edges ($E_{brain} = 0.29 \pm 0.02$, whereas $E_{rand} = 0.36 \pm 0.002$), ruling out the idea that the brain simply optimizes for efficiency.

In short, simple optimization of one or other of a number of possible selection criteria (cost, efficiency) does not generate graphs that are even statistically similar to brain graphs. Somewhat more surprisingly, the brain is also not wired so as to achieve a specific desired efficiency at minimum cost. Indeed, it is easy to generate networks with efficiencies comparable to that of the brain at much lower cost. This can be seen, for example, by examining the distance distribution for the exponential decay model (shown in green) in Fig. 1.

It is possible that there exists a simple network measure that embodies the subtle tradeoff between cost and function sought by brain networks. However, it is currently not known what properties (if any) the brain is strictly optimized for, so simple optimization is not currently an option for generating brain-like graphs. Instead, we can (as is done in this article) posit simple growth rules that embody some plausible tradeoffs between cost and topological properties, then test how brain-like the resulting networks are. We find that some such simple, two-parameter rules generate networks that share many topological properties observed in the data.

The idea that two simple constraints can generate networks with a given clustering, modularity, efficiency, and degree distributions also highlights the fact that the large (and growing) numbers of network measures are not independent of one another. Ultimately, we hope that the type of approach adopted in our article will provide insights into the question of what independent topological features (if any) the brain has evolved to optimize and under what constraints. Once such hypotheses exist, and provided they are simple enough, they could of course be tested by a direct optimization method. We therefore view these two approaches as complementary.

Note also that, even if the brain were found to have evolved to optimize a small set of network properties, it would still be interesting to discover what local growth rules may be used by an evolving network to ensure the emergence of this set of desirable features.

Importantly, however, the growth rule approach does not rule out the possibility that the brain is in fact not optimal in any simple sense. For example, the quantity to optimize may be a complicated global property, such that evolution has instead optimized simpler,

more local properties only achieving an approximation of the optimal topology. This situation is quite different from the engineering context, where products are designed to directly optimize a set of desired attributes and are then accurately wired to achieve this state.

1. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ Press, New York), 2nd Ed.

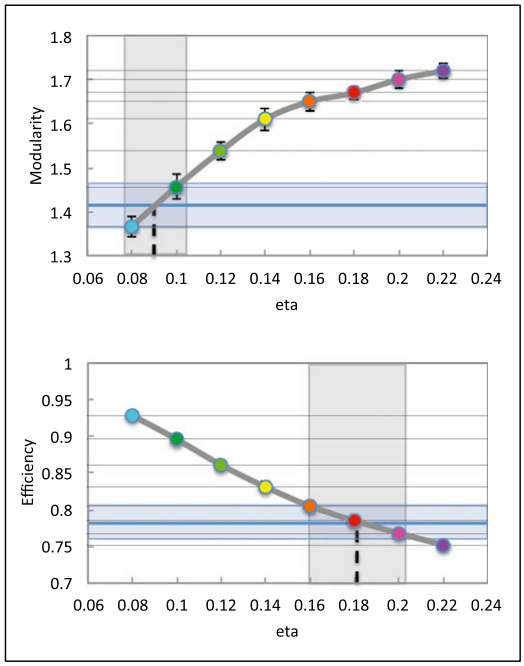

Fig. S1. To match the observed value of modularity within a 95% confidence interval (marked by the blue horizontal line and blue box in the top plot), the parameter η needs to be set at $\eta = 0.09 \pm 0.01$ (marked by the dashed line and gray box). However, matching the observed value of efficiency within a 95% confidence interval (marked by a blue line and blue box in the bottom plot) requires $\eta = 0.18 \pm 0.02$ (marked by the dashed line and gray box), which corresponds to a lower value of modularity. The exponential decay model is therefore unable to match the observed values of modularity and efficiency simultaneously. Note that in practice we used SA for optimizing the parameter η in this simple model, purely for consistency with the procedure adopted for the more complex, two-parameter models used later on.

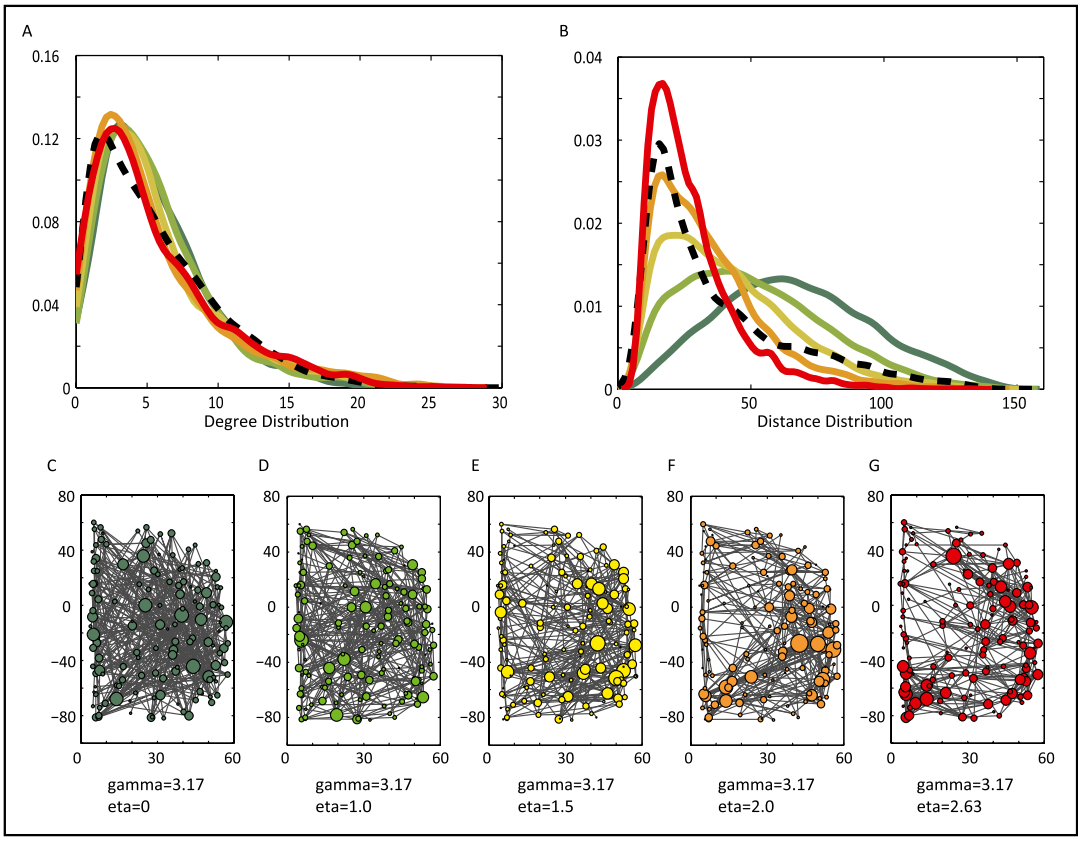

Fig. S2. Varying η with γ set at the value used to model normal human fMRI data ($\gamma = 3.17$) has little effect on the degree distribution (*A*) but greatly influences the distance distribution (*B*). Larger values of η correspond to sharper distance penalties and therefore result in fewer long-distance links. This can also be seen in the schematic representation of the brain network for each parameter setting (*C*–*G*). Throughout the figure, values of η are color-coded so that networks with $\eta = 0, 1, 1.5, 2, 2.63$ are represented in dark green, light green, yellow, orange, and red respectively. This final network (*G*) with $\eta = 2.63$ and $\gamma = 3.17$ (red) corresponds to the best-fitting model of these data.

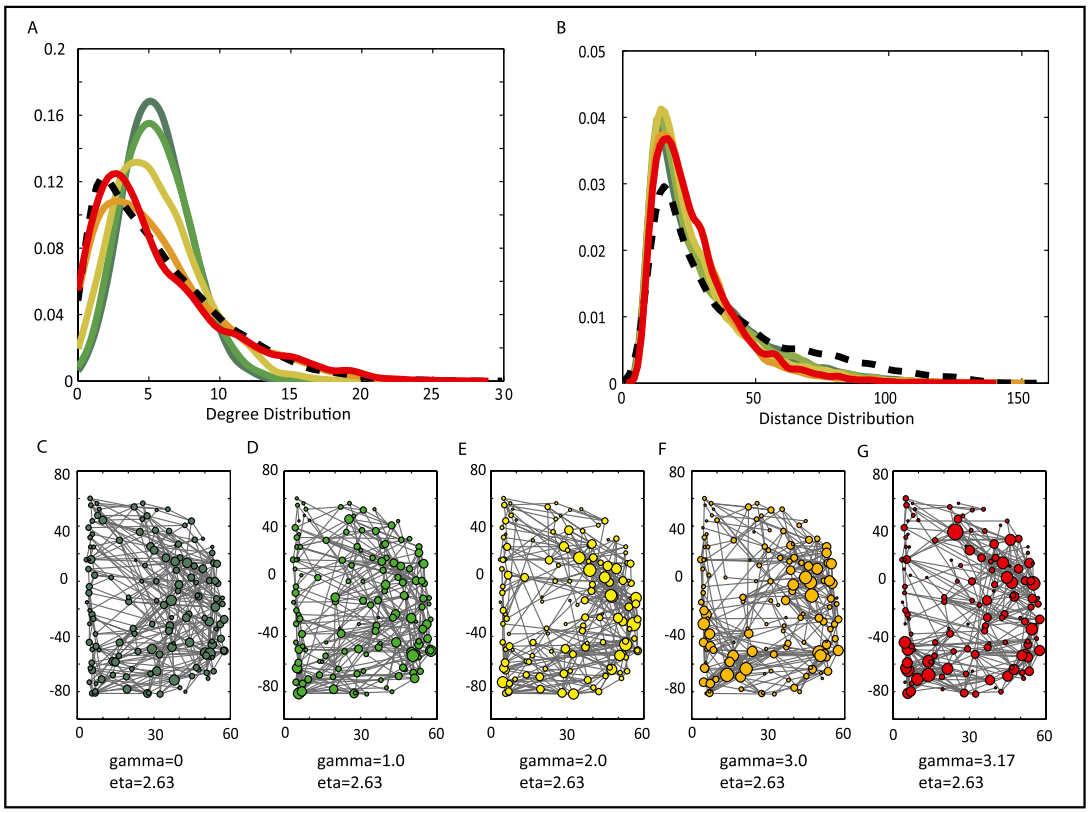

Fig. S3. Varying γ with η set at the value used to model normal Human fMRI data ($\eta = 2.63$) has little effect on the distance distribution (*B*) but greatly influences the degree distribution (*A*). Larger values of γ correspond to more skewed degree distributions and therefore the occurrence of more and larger hubs. This can also be seen in the schematic representation of the brain network for each parameter setting (*C* and *D*). Throughout the figure, values of γ are color-coded so that networks with $\gamma = 0, 1, 2, 3, 3.17$ are represented in dark green, light green, yellow, orange, and red, respectively (*C*–*G*). This final network (*G*) with $\eta = 2.63$ and $\gamma = 3.17$ (red) corresponds to the best-fitting model of these data.

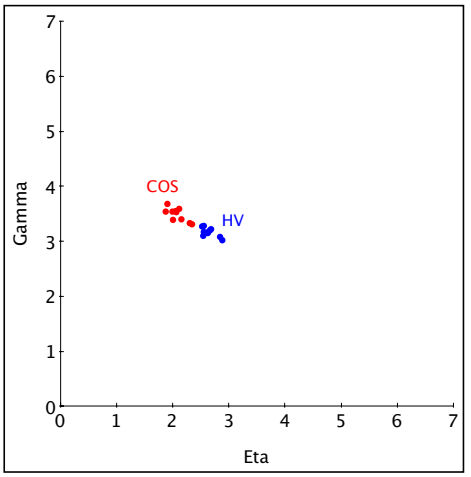

Fig. S4. This plot shows the jitter in parameter values obtained when we perform SA 10 times for both the healthy volunteer group (HV) and the COS group. We start the annealing process for a given dataset twice from each of the following five initial positions: $(\eta, \gamma) = (0, 0)$; $(\eta, \gamma) = (0, 5)$; $(\eta, \gamma) = (5, 5)$; $(\eta, \gamma) = (5, 0)$; and $(\eta, \gamma) = (2.5, 2.5)$. Note that the resulting jitter in parameter values is small compared with the differences in parameter setting between the healthy volunteer and COS datasets.

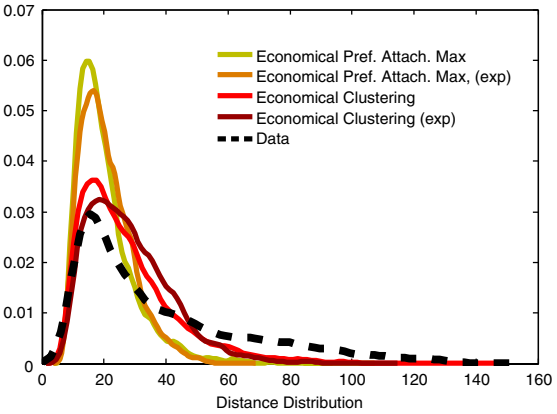

Fig. S5. Among the four best models from Table S1 ($E_2 < 10^6$), the economical clustering model with power-law distance penalty provides the best fit to the distance distribution, which was left unconstrained during annealing.

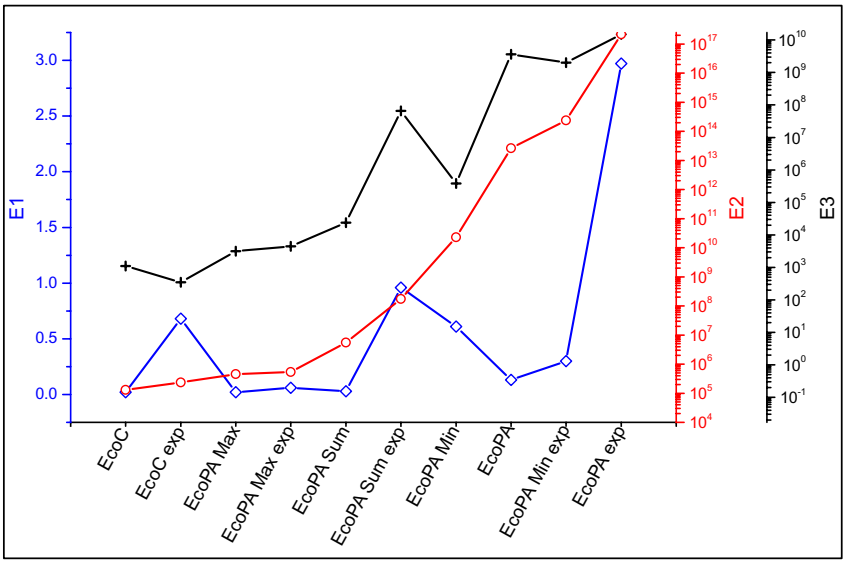

Fig. S6. This plot shows energy values minimized by SA for all 10 two-parameter generative models (we have excluded the simple exponential and power-law decay models because they consistently yielded much higher energies and would have affected the axes considerably). The models are ordered along the *x* axis according to their ranking in terms of *E2* energy as shown in Table S1. The blue diamonds, red circles, and black crosses correspond to $E_1 = 1/(T_C.T_E.T_M.KS_{deg})$, $E_2 = 1/(p_C.p_E.p_M.p_{deg})$, and $E_3 = 1/p_C + 1/p_E + 1/p_M + 1/p_{deg}$, respectively. Generative models are abbreviated as in Table S1. EcoC, economical clustering (Eq. **3**); EcoPA, economical preferential attachment (Eq. **2**).

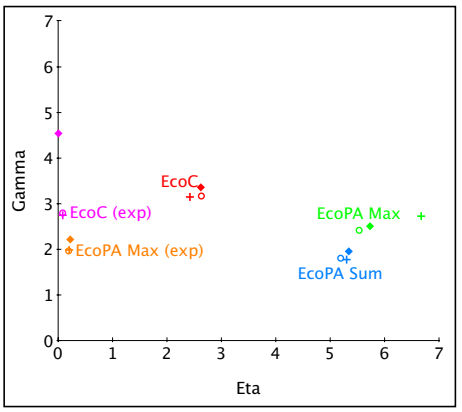

Fig. S7. This plot shows the optimal parameter values (η and γ) for the five best generative models, estimated by SA on three alternative energy functions. The five models are the economical clustering (EcoC) with power-law distance penalty (red), the EcoC model with exponential distance penalty (magenta), the economical preferential attachment (EcoPA) model using the sum of degrees (blue), and the EcoPA Max models using the maximum of degrees with both power-law and exponential distance penalties (green and orange, respectively). The shape of the markers represents the energy functions used for annealing. The diamond, the open circle, and the cross correspond to $E_1 = 1/(T_C.T_E.T_M.KS_{deg})$, $E_2 = 1/(p_C.p_E.p_M.p_{deg})$, and $E_3 = 1/p_C + 1/p_E + 1/p_M + 1/p_{deg}$, respectively.

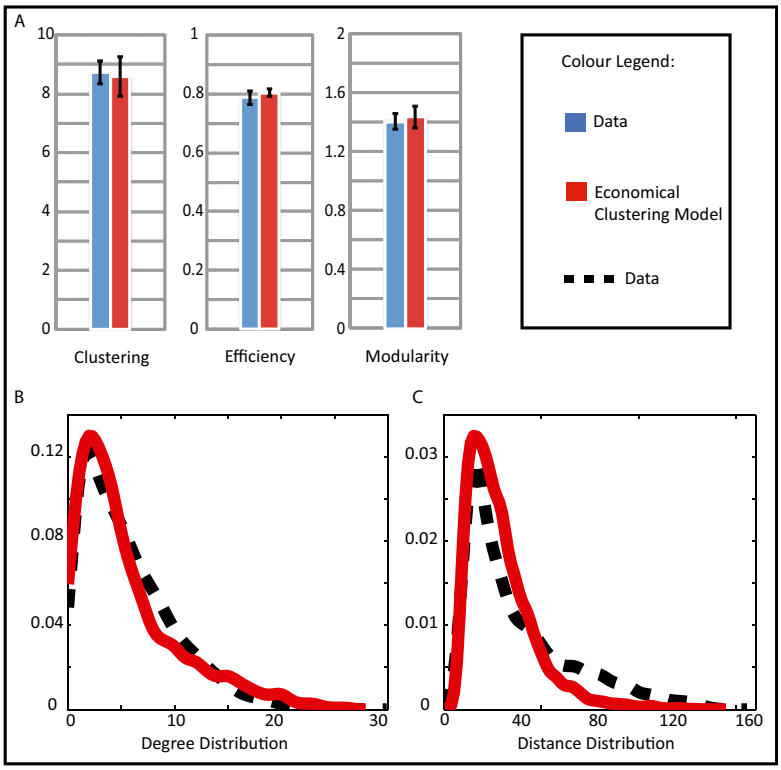

Fig. S8. Annealing over only clustering, efficiency, and modularity (without using any information about the degree distribution) still yields best-fit parameters that enable relatively accurate modeling of human fMRI data, both in terms of matching the key network measures (*A*), as well as the degree (*B*) and distance (*C*) distributions.

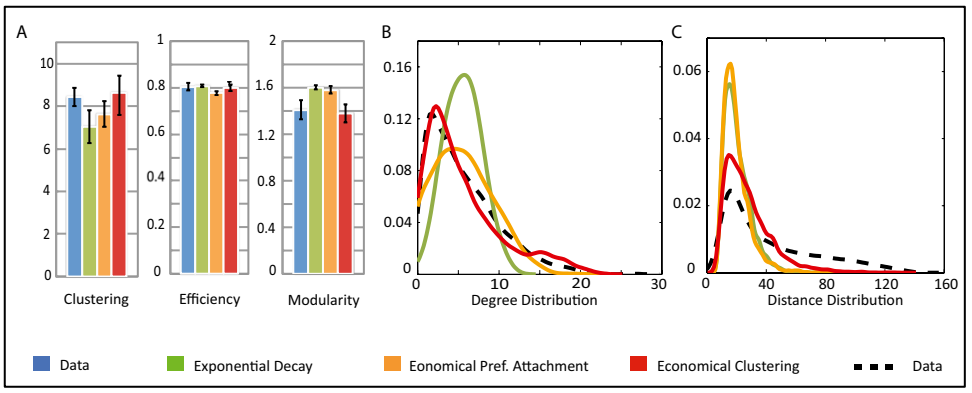

Fig. S9. Results from Fig. 1 are replicated in a second, independent group of 12 healthy volunteers. We show the comparison of networks simulated by three generative models with brain functional networks derived from experimental fMRI data (blue). Once again, both the model based on an exponential distance penalty (green) and the economical preferential attachment model (orange) fail to simultaneously capture several topological characteristics of functional brain networks. In contrast, the economical clustering model (red) yields significantly more realistic networks by all of the following measures. (*A*) Normalized clustering coefficient, global efficiency, and modularity of brain functional networks are all well matched by the economical clustering model. All values are averaged over 12 instantiations of each network, and error bars represent the 95% confidence interval for the mean. Degree (*B*) and distance (*C*) distributions are shown in solid colored and dashed black lines for the models and data, respectively. Both distributions are more closely approximated by the economical clustering model (red) than by the exponential decay (green) or economical preferential attachment (orange) models. All networks have an overall connection density of 4%.

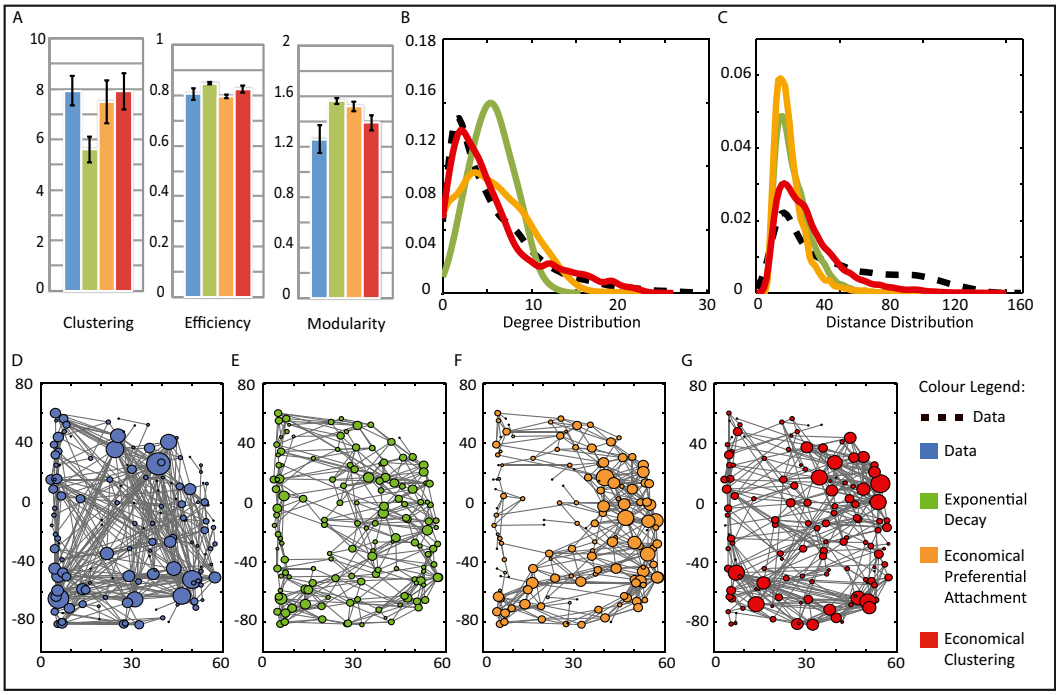

Fig. S10. Comparison of networks simulated by three generative models with brain functional networks derived from experimental fMRI data on a group of participants with COS (blue). Both the simple one-parameter model based on an exponential distance penalty (green) and the more sophisticated economical preferential attachment model (orange) fail to simultaneously capture several topological characteristics of functional brain networks. In contrast, the economical clustering model (red) yields significantly more realistic networks by all of the following measures. (*A*) Normalized clustering coefficient, global efficiency, and modularity of brain functional networks are all well matched by the economical clustering model. All values are averaged over 19 instantiations of each network, and error bars represent the 95% confidence interval for the mean. Degree (*B*) and distance (*C*) distributions are shown in solid colored and dashed black for the models and data, respectively. Both distributions are significantly better captured by the economical clustering model (red) than by the exponential decay (green) or economical preferential attachment (orange) models. (*D*–*G*) Schematic representation of the right hemisphere of the fMRI brain network for one participant (blue) and of a representative network generated by a single instantiation of each model. To ensure that these networks are representative, the single participant and the specific model instantiations displayed were each chosen to have the median value of skew in their degree distributions. The size of each node represents the degree of the corresponding brain region within the network. All networks have an overall connection density of 4%.

Table S1. Value of the minimum energy attained via annealing [using the energy function $E = 1/(p_C \cdot p_E \cdot p_M \cdot p_{deg})$] can be used to quantify the suitability of various one- and two-parameter models for simulating our data

Model name	Abbreviation	Model	η	γ	P_C	P_E	P_M	$K S_{deg}$	Energy
Exponential decay	ExpD	$P_{ij} \propto exp(-\eta d_{i,j})$	0.16	NA	5.10^{-9}	0.04	4.10^{-8}	8.10^{-68}	10^{84}
Power-law decay	PLD	$P_{ij} \propto exp(d_{i,j})^{-\eta}$	3.38	NA	3.10^{-11}	7.10^{-8}	2.10^{-4}	2×10^{70}	5.10^{-91}
Economical clustering	EcoC	$P_{ij} \propto exp(k_{i,j})^{\gamma}(d_{i,j})^{-\eta}$	2.63	3.17	**0.79**	**0.74**	**0.10**	2.10^{-4}	10^{5}
Economical clustering with exponential distance penalty	EcoC (exp)	$P_{ij} \propto (k_{i,j})^{\gamma}exp(-\eta d_{i,j})$	0.08	2.8	**0.85**	0.02	**0.24**	10^{-3}	2×10^{5}
Economical preferential attachment	EcoPA	$P_{ij} \propto (k_i k_j)^{\gamma}(d_{i,j})^{-\eta}$	5.37	1.81	0.01	**0.80**	0.001	3×10^{-9}	3×10^{13}
Economical preferential attachment with exponential distance penalty	EcoPA (exp)	$P_{ij} \propto (k_i k_j)^{\gamma}exp(-\eta d_{i,j})$	0.19	1.42	0.005	**0.51**	0.0001	1.5×10^{-11}	2×10^{17}
Economical summed preferential attachment	EcoPA Sum	$P_{ij} \propto (k_i + k_j)^{\gamma}(d_{i,j})^{-\eta}$	5.19	1.81	**0.27**	**0.78**	**0.12**	7×10^{-6}	6×10^{6}
Economical summed preferential attachment with exponential distance penalty	EcoPA Sum (exp)	$P_{ij} \propto (k_i + k_j)^{\gamma}exp(-\eta d_{i,j})$	0.18	1.42	0.007	**0.95**	0.008	10^{-4}	2×10^{8}
Economical minimum preferential attachment	EcoPA Min	$P_{ij} \propto (min[k_i, k_j])^{\gamma}exp(d_{i,j})^{-\eta}$	5.55	2.03	**0.05**	**0.34**	0.0003	9×10^{-6}	2×10^{10}
Economical minimum preferential attachment with exponential distance penalty	EcoPA Min (exp)	$P_{ij} \propto (min[k_i, k_j])^{\gamma}exp(-\eta d_{i,j})$	0.18	1.49	0.001	**0.65**	0.004	10^{-9}	2×10^{14}
Economical maximum preferential attachment	EcoPA Max	$P_{ij} \propto (max[k_i, k_j])^{\gamma}(d_{i,j})^{-\eta}$	5.53	2.42	**0.95**	**0.08**	**0.05**	5.5×10^{-4}	5×10^{5}
Economical maximum preferential attachment with exponential distance penalty	EcoPA Max (exp)	$P_{ij} \propto (max[k_i, k_j])^{\gamma}exp(-\eta d_{i,j})$	0.19	1.97	**0.41**	**0.88**	**0.33**	1.5×10^{-5}	5×10^{5}

The lower the energy value the closer an approximation the model can yield, with optimal parameters, to real data. Here we report energy values for a set of alternative models that were found to be less suitable than the economic clustering rule presented in this article. In each case we display the values of the parameters η and γ minimizing the energy function, the energy value obtained (median value over 15 independent runs for a given set of parameters), and the *P* values for the comparison of each network metric between groups of experimental data and multiple realizations of the simulated networks. The *P* values that show no significant difference between model and data ($P > 0.05$) are highlighted in bold.

Table S2. Key results presented in this article also hold at 8% and 16% connection cost, in healthy volunteers

Cost	η	γ	P_C	P_E	P_M	P_k	Energy
Cost = 8%							
Exponential decay	0.05	NA	5.10^{-25}	5.10^{-15}	10^{-7}	10^{-54}	3.10^{-99}
Economical preferential attachment	4.47	1.78	10^{-4}	4.10^{-3}	**0.05**	6.10^{-5}	5.10^{11}
Economical clustering	2.27	2.42	**0.21**	**0.75**	**0.43**	5.10^{-3}	3.10^{3}
Cost = 16%							
Exponential decay	0.08	NA	5.10^{-7}	0.02	9.10^{-9}	2.10^{-44}	8.10^{-59}
Economical preferential attachment	3.39	1.54	5.10^{-4}	2.10^{-4}	0.04	10^{-3}	2.10^{11}
Economical clustering	1.89	1.76	**0.20**	**0.35**	**0.07**	0.05	4.10^{3}

In each case we display the values of the parameters η and γ minimizing the energy function, the energy value obtained, and the *P* values for the comparison of each network metric between groups of experimental data and multiple realizations of the simulated networks (with the three models discussed in the main text). The *P* values that show no significant difference between model and data ($P > 0.05$) are highlighted in bold. These results are fully consistent with the results presented in Table S1 and throughout the article at cost = 4%.

Table S3. Simulated annealing results for several possible generative models of brain networks in COS

Childhood onset schizophrenia	η	γ	P_C	P_E	P_M	P_k	Energy
Exponential decay	0.13	NA	2.10^{-6}	0.002	10^{-5}	8.10^{-87}	3.10^{99}
Economical preferential attachment	5.23	1.93	0.006	**0.60**	0.009	10^{-12}	3.10^{16}
Economical clustering	2.30	3.33	**0.95**	**0.21**	**0.05**	8.10^{-5}	10^{6}

P values that show no significant difference between model and data ($P > 0.05$) are highlighted in bold. We also report the values of the parameters η and γ minimizing the energy function for each model.

Table S4. Values of clustering, efficiency, and modularity observed in brain functional networks compared with those obtained in a model enforcing minimum wiring cost

Parameter	Brain data	Minimum cost model
Clustering	0.35 ± 0.03	0.47
Efficiency	0.29 ± 0.02	0.25
Modularity	0.53 ± 0.05	0.66