nature neuroscience

Article

https://doi.org/10.1038/s41593-023-01299-3

Functional brain networks reflect spatial and temporal autocorrelation

In the format provided by the authors and unedited

Supplementary Note 1: Mathematical derivations and proofs

1.1 Proof of correlated spectral sampling

Correlated spectral sampling generates a set of timeseries $x_i[t]$ with arbitrary frequency spectra $|X_i|^2$ and arbitrary spatial structure given by the covariance matrix, C. To show that $\{x_1[t], \ldots, x_N[t]\}$ have the desired spatial structure, we show that for any i, j, $\mathbb{E}[\text{corr}(x_i[t], x_j[t])] = C_{i,j}$. As an intermediate step, we show that $\mathbb{E}[\text{corr}(x_i[t], x_j[t])] = \text{corr}(\Re(X_i[k]), \Re(X_j[k])))$.

Let $x_i[t]$ and $x_j[t]$ be two of the timeseries resulting from this algorithm. Note that, for large N, $\text{corr}(x_i[t], x_j[t]) = \text{cov}(x_i[t], x_j[t])/(\text{var}(x_i[t]) \cdot \text{var}(x_j[t]))^{1/2}$, where var and cov are the empirical estimates of the variance and covariance, respectively. We first seek to expand $\text{cov}(x_i[t], x_j[t])$, $\text{var}(x_i[t])$, and $\text{var}(x_j[t])$. Without loss of generality, we may assume that $x_i[t]$ and $x_j[t]$ have a mean of 0, as their mean does not influence the spatial or temporal structure of the set of timeseries, so we can estimate $\text{cov}(x_i[t], x_j[t]) = 1/N \sum_t x_i[t] x_j[t]$.

Recall that $X_i[k]$ is the discrete Fourier transform $\mathcal{F}(x_i[t])$. We denote the even and odd components $x_i^e[t] = \mathcal{F}^{-1}(\Re(X[k])$ and $x_i^o[t] = \mathcal{F}^{-1}(i\Im(X[k])$. Thus, our estimator of the covariance can be written as

$$\text{cov}(x_i[t], x_j[t]) = 1/N \left(\sum_t x_i^e[t] x_j^e[t] + \sum_t x_i^e[t] x_j^o[t] + \sum_t x_i^o[t] x_j^e[t] + \sum_t x_i^o[t] x_j^o[t] \right).$$

For the even part of the function, $\sum_t x_i^e[t] x_j^e[t] = \sum_t x_i^e[t] \overline{x_j^e[t]}$, and by Parseval's theorem, this equals $1/N \sum_k \Re(X_i[k]) \overline{\Re(X_j[k])}$. This can be written as $1/N \sum_k \Re(X_i[k]) \Re(X_j[k]) = \text{cov}(X_i[k], X_j[k])$. By the same logic, $\sum_t x_i^o[t] x_j^o[t] = \text{cov}(\Re(X_i[k]), \Re(X_j[k]))$.

The real and imaginary components of $X_i[k]$ are independent by construction, which means $\mathbb{E}[\sum_t x_i^e[t] x_j^o[t]] = 0$ and $\mathbb{E}[\sum_t x_i^o[t] x_j^e[t]] = 0$. Thus, we can rewrite the covariance estimator as

$$\begin{aligned} \mathbb{E}[\text{cov}(x_i, x_j)] &= 1/N(\text{cov}(\Re(X_i[k]), \Re(X_j[k])) + 0 + 0 + \text{cov}(\Re(X_i[k]), \Re(X_j[k]))) \\ &= 2 \cdot \text{cov}(\Re(X_i[k]), \Re(X_j[k]))/N. \end{aligned}$$

This gives $\mathbb{E}[\text{var}(x_i)] = \mathbb{E}[\text{cov}(x_i, x_i)] = 2 \cdot \text{var}(\Re(X_i[k]))/N$, so

$$\begin{aligned} \mathbb{E}[\text{corr}(x_i[k], x_j[k])] &= \frac{2 \cdot \text{cov}(\Re(X_i[k]), \Re(X_j[k]))/N}{\sqrt{2 \cdot \text{var}(\Re(X_i[k]))/N} \cdot \sqrt{2 \cdot \text{var}(\Re(X_j[k]))/N}} \\ &= \frac{\text{cov}(\Re(X_i[k]), \Re(X_j[k]))}{\sqrt{\text{var}(\Re(X_i[k]))} \cdot \sqrt{\text{var}(\Re(X_j[k])}} \\ &= \text{corr}(\Re(X_i[k]), \Re(X_j[k])). \end{aligned}$$

A similar procedure applies to the imaginary component of $X_i[k]$. ■

We now show that $\mathbb{E}[\text{corr}(\Re(X_i[k]), \Re(X_j[k]))] = C_{i,j}$. We note that, by construction, $a_i^R[k]$ and $|X_i[k]|^2$ are independent, and that $\mathbb{E}(a_i^R[k]) = 0$, meaning $\mathbb{E}[\Re(X_i[k])] = \mathbb{E}[|X_i[k]|a_i^R[k])] = \mathbb{E}[|X_i[k]|] \cdot \mathbb{E}[a_i^R[k]] = 0$. This allows us to expand as before, giving

$$\mathbb{E}[\text{corr}(\Re(X_i[k]), \Re(X_j[k]))] = \frac{\mathbb{E}\left[\sum_k (|X_i[k]||X_j[k]|)a_i^R[k]a_j^R[k]\right]}{\mathbb{E}\left[\sum_k (|X_i[k]|a_i^R[k])^2\right]^{1/2} \cdot \mathbb{E}\left[\sum_k (|X_j[k]|a_j^R[k])^2\right]^{1/2}}$$

Because $a_i^R[k]$ and $a_j^R[k]$ have mean 0 and variance 1 and have correlation equal to approximately $\Sigma_{i,j}$, the product $a_i^R[k]a_j^R[k]$ has an expected value of $\Sigma_{i,j}$. By construction, $(a_i^R[k]a_j^R[k])$ and $(|X_i[k]||X_j[k]|)$ are independent, so

$$\mathbb{E}\left[\sum_k |X_i[k]||X_j[k]|a_i^R[k]a_j^R[k]\right] = \Sigma_{i,j} \cdot \mathbb{E}\left[\sum_k |X_i[k]||X_j[k]|\right] = \Sigma_{i,j} \cdot \sum_k |X_i[k]||X_j[k]|$$

When $i = j$, $C_{i,j} = 1$. By the same logic,

$$\mathbb{E}\left[\sum_k (|X_i[k]|a_i^R[k])^2\right]^{1/2} \cdot \mathbb{E}\left[\sum_k (|X_j[k]|a_j^R[k])^2\right]^{1/2} = \left(\sum_k |X_i[k]|^2\right)^{1/2} \cdot \left(\sum_k |X_j[k]|^2\right)^{1/2},$$

and so

$$\begin{aligned}
\mathbb{E}[\text{corr}(x_i, x_j)] &= \Sigma_{i,j} \cdot \frac{\sum_k |X_i[k]||X_j[k]|}{\left(\sum_k |X_i[k]|^2\right)^{1/2} \cdot \left(\sum_k |X_j[k]|^2\right)^{1/2}} \\
&= \Sigma_{i,j} \cdot \text{sim}(|X_i[k]|, |X_j[k]|) \\
&= C_{i,j}.
\end{aligned}$$

■

1.2 Relationship between TA-Δ_1 and power spectrum

Suppose we have a finite timeseries $x[t]$ with discrete Fourier transform $X[k]$. We claim that its TA-Δ_1 is

$$\text{corr}(x[t], x[t+1]) = \frac{1}{\sum_{k=1}^{N-1} |X[k]|^2} \sum_{k=1}^{N-1} |X[k]|^2 \cos(2\pi k/N).$$

Let $ACF(x)[\tau]$ denote the autocovariance of X at lag τ, i.e. $\text{cov}(x[t], x[t+\tau])$. By the Wiener-Khinchin theorem, $ACF(x)[\tau]$ is the inverse Fourier transform of an estimate of the

power spectrum, $|X[k]|^2$. Without loss of generality, let us suppose $x[t]$ has zero mean, and hence, $X[0] = 0$, so that

$$ACF(x)[\tau] = \sum_{k=1}^{N-1} |X[k]|^2 e^{i2\pi k\tau/N}.$$

Because $x[t]$ is real, $e^{i2\pi k\tau/N} = \cos(2\pi k\tau/N)$. Making the approximation that $\text{var}(x[t]) = \text{var}(x[t+1])$, we have

$$\begin{aligned}
\text{corr}(x[t], x[t+1]) &= \frac{\text{cov}(x[t], x[t+1])}{\sqrt{\text{var}(x[t]) \cdot \text{var}(x[t+1])}} \\
&= \frac{\text{ACF}(x)[1]}{\sum_{k=1}^{N-1} |X[t]|^2} \\
&= \frac{1}{\sum_{k=1}^{N-1} |X[t]|^2} \sum_{k=1}^{N-1} |X[k]|^2 \cos(2\pi k/N).
\end{aligned}$$

■

1.3 Relationship between TA-Δ_1 and noisy power spectrum

Additionally, we would like to know how the addition of white noise to a power spectrum will change the TA-Δ_1. We prove that, for an finite timeseries $y[t] = x[t] + w[t]$ where $w[t] \sim N(0, \sigma)$ of length N,

$$\mathbb{E}[\text{corr}(y[t], y[t+1])] = \frac{1}{N\sigma^2 + \sum_{k=1}^{N-1} |X[k]|^2} \sum_{k=1}^{N-1} |X[k]|^2 \cos(2\pi k/N).$$

The Fourier transform is linear, so we can write

$$\begin{aligned}
|Y[k]|^2 &= |X[k] + W[k]|^2 \\
&= |X[k]|^2 + |W[k]|^2 + 2|X[k]||W[k]| \cos(\angle X[k] - \angle W[k]).
\end{aligned}$$

Since $w[t]$ is white noise, $W[k]$ has the properties $\mathbb{E}[|W[k]|] = \sigma\sqrt{N}$ and $\angle W[k] \sim U(0, 2\pi)$. Thus, $\mathbb{E}[\cos(\angle X[t] - \angle W[t])] = 0$, so

$$\begin{aligned}
\mathbb{E}[|Y[k]|^2] &= |X[k]|^2 + \mathbb{E}[|W[k]|^2] + 2|X[k]||W[k]|\mathbb{E}[\cos(\angle X[k] - \angle W[k])] \\
&= |X[k]|^2 + \sigma^2 N + 0.
\end{aligned}$$

Applying our result from Supplement 1.2, we see that

$$\begin{aligned}\mathbb{E}[\text{corr}(y[t], y[t+1])] &= \frac{1}{N\sigma^2 + \sum_{k=1}^{N-1} |X[k]|^2} \sum_{k=1}^{N-1} (|X[k]|^2 + N\sigma^2)\cos(2\pi k/N) \\ &= \frac{1}{N\sigma^2 + \sum_{k=1}^{N-1} |X[k]|^2} \cdot \left[N\sigma^2 \sum_{n=1}^{N-1} \cos(2\pi k/N) + \sum_{n=1}^{N-1} |X[k]|^2 \cos(2\pi n/N) \right] \\ &= 0 + \frac{1}{N\sigma^2 + \sum_{k=1}^{N-1} |X[k]|^2} \sum_{k=1}^{N-1} |X[k]|^2 \cos(2\pi k/N). \quad \blacksquare\end{aligned}$$

It directly follows that, for a given $\mathbb{E}[\text{corr}(y[t], y[t+1])] = \phi$, we can solve this equation for σ^2 as

$$\sigma^2 = \frac{1}{N^2\phi} \sum_{k=1}^{N-1} |X[k]|^2 \left(\cos(2\pi k/N) - \phi\right)$$

which is equivalent to Equation 3.

1.4 Proof of infeasibility of parameter-free fitting of the spatiotemporal model

In theory, it is possible to fit the spatiotemporal model in a parameter-free manner. In other words, the SA-λ, SA-∞, and TA-Δ_1 statistics can all be measured from the data, and therefore, it should be possible to choose SA-λ^{gen} and SA-∞^{gen}, with arbitrary power spectra estimators $|X_i[k]|^2$ for region i, such that the spatiotemporal model produces timeseries which have a given SA-λ, SA-∞, and TA-Δ_1. We describe the mathematical derivation of a procedure for selecting SA-λ and SA-∞ using this logic, and show why this procedure is infeasible in practice.

As a reminder, the spatiotemporal model has four steps: (1) use SA-λ^{gen} and SA-∞^{gen} in conjunction with the nodal Euclidean distances $D_{i,j}$ to generate a spatial correlation matrix C; (2) construct a power spectrum corresponding to high-pass filtered Brownian ($1/f^2$) noise; (3) construct timeseries $x_1, \ldots, x_N$ for brain regions $1, \ldots, N$ using correlated spectral sampling such that $\text{corr}(x_i, x_j) = C_{i,j}$ and all timeseries have the power spectrum specified in step (2); (4) for each time point t in each timeseries $x_i[t]$, add white noise $w_i[t] \sim N(0, \sigma_i^2)$, with σ_i^2 chosen according to Equation 3 such that $x_i[t]$ has some desired TA-Δ_1.

To perform the derivation, suppose that we have two timeseries x_i and x_j, $i \neq j$, such that $x_i'[t] = x_i[t] + w_i[t]$, where $x_i[t]$ has some power spectrum $|X_i[t]|^2$ for all i and $w_i[t] \sim N(0, \sigma_i^2)$ for given σ_i^2. Let the correlation $\text{corr}(x_i, x_j) = C_{i,j}$ for some unknown $C_{i,j}$. We derive an equation to select $C_{i,j}$ such that $\mathbb{E}(\text{corr}(x_i', x_j')) = C_{i,j}'$ for some desired $C_{i,j}'$

To begin, we rewrite

$$\text{corr}(x_i, x_j) = \frac{\text{cov}(x_i, x_j)}{\sqrt{\text{var}(x_i)\text{var}(x_j)}}$$
$$\text{corr}(x'_i, x'_j) = \frac{\text{cov}(x'_i, x'_j)}{\sqrt{\text{var}(x'_i)\text{var}(x'_j)}}$$

Since $\mathbb{E}(\text{cov}(x_i + w_i, x_j + w_j)) = \text{cov}(x_i, x_j)$ for $i \neq j$, and $\mathbb{E}(\text{var}(x_i + w_i)) = \text{var}(x_i) + \sigma_i^2$, we can write

$$\begin{aligned}
\mathbb{E}(\text{corr}(x'_i, x'_j)) &= \frac{\text{cov}(x_i, x_j)}{\sqrt{(\text{var}(x_i) + \sigma_i^2)(\text{var}(x_j) + \sigma_j^2)}} \\
&= \frac{\text{corr}(x_i, x_j)\sqrt{\text{var}(x_i)\text{var}(x_j)}}{\sqrt{(\text{var}(x_i) + \sigma_i^2)(\text{var}(x_j) + \sigma_j^2)}} \\
&= \frac{\text{corr}(x_i, x_j)}{\sqrt{\left(1 + \frac{\sigma_i^2}{\text{var}(x_i)}\right)\left(1 + \frac{\sigma_j^2}{\text{var}(x_j)}\right)}}
\end{aligned}$$

The variance of a timeseries x_i can be estimated by its power spectrum using Parseval's theorem, such that $\text{var}(x_i) = \sum_{k=1}^{N/2} |X_i[k]|^2/(N/2)^2$. Thus, substituting $C'_{i,j} = \text{corr}(x'_i, x'_j)$ and $C_{i,j} = \text{corr}(x_i, x_j)$ and rearranging terms, we have

$$\begin{aligned}
C_{i,j} &= C'_{i,j}\sqrt{\left(1 + \frac{\sigma_i^2}{\text{var}(x_i)}\right)\left(1 + \frac{\sigma_j^2}{\text{var}(x_j)}\right)} \\
&= C'_{i,j}\sqrt{\left(1 + \frac{\sigma_i^2(N/2)^2}{\sum_{k=1}^{N/2} |X_i[k]|^2}\right)\left(1 + \frac{\sigma_j^2(N/2)^2}{\sum_{k=1}^{N/2} |X_j[k]|^2}\right)}
\end{aligned}$$

■

Thus, we have derived a $C_{i,j}$ which can be used in our procedure to produce timeseries with correlation $C'_{i,j}$.

While mathematically interesting, the derivation is not useful in practice. Because we are computing a correlation, $C_{i,j}$ has an upper bound of 1, thereby bounding the potential values of $C'_{i,j}$ which can be produced. In practice this constraint is difficult to satisfy. In our data, σ_i^2 can take on values up to an order of magnitude larger than $\text{var}(x_i)$, meaning that only very small values of $C'_{i,j}$ (approximately $C'_{i,j} < 0.1$) can satisfy the constraint. Additionally, in the

analysis of neural data, we consider more than two timeseries. This means that C must be a correlation matrix, and thus C must be positive semi-definite, or equivalently, have all non-negative eigenvalues. In our data, approximately 25% of eigenvalues are negative after applying this procedure, showing that it is unlikely to be useful in practice.

1.5 Relationship between temporal autocorrelation and variance in correlation

Previous work established that temporal autocorrelation (TA) increases the variance of the Pearson correlation between pairs of timeseries[1–4]. Here, we quantified TA using TA-Δ_1, the TA at a single lag. We did this with the understanding that TA-Δ_1 correlates with TA at higher lags, or with an underlying long memory process (Figure ED3). Since higher lag TA may have an impact on the variance, it is insufficient to solely measure the effect of TA-Δ_1 on the variance in Pearson correlation. A more general approach is to quantify the effect of the entire power spectrum on the variance in Pearson correlation (Supplement 1.2). The power spectrum is a full-rank linear transformation of the autocorrelation function, and therefore, is a non-parametric analysis of TA for both short and long memory dynamics. Thus, we seek to understand the impact of an arbitrary power spectrum on the variance in Pearson correlation of two timeseries.

We focus on a specific case of particular importance by analyzing two independent timeseries with identical power spectra and phases distributed randomly around the unit circle. We show that the variance of the Pearson correlation between these two timeseries increases as the uniformity of the power spectrum decreases. In other words, Pearson correlation is lowest for a timeseries of iid white noise, which by definition has no TA. It increases as the power spectrum deviates from uniformity, and is highest for spectra consisting of a single frequency, i.e. a sinusoid.

The variance of a timeseries is closely connected with the power spectrum. Consider the real, discrete, finite timeseries $x_1[t]$ and $x_2[t]$ of length N, with Fourier transform $X_1[k]$ and $X_2[k]$ and variance σ_1^2 and σ_2^2. Without loss of generality, assume $x_1[t]$ and $x_2[t]$ have mean 0. We can estimate the covariance for large N as

$$\mathrm{cov}(x_1, x_2) = \frac{1}{N}\sum_{t=0}^{N} x_1[t]x_2[t].$$

Then, by Parseval's theorem, we can write the covariance as

$$\mathrm{cov}(x_1[t], x_2[t]) = \frac{1}{N^2}\sum_{k=1}^{N} |X_1[k]\overline{X_2[k]}|.$$

Therefore, by noting that the variance of a timeseries is simply self-covariance, the empirical Pearson correlation corr can be written as

$$\begin{aligned}\mathrm{corr}(x_1[t], x_2[t]) &= \frac{\frac{1}{N^2}\sum_k |X_1[k]\overline{X_2[k]}|}{\sqrt{\frac{1}{N^2}\sum_k |X_1[k]|^2}\sqrt{\frac{1}{N^2}\sum_k |X_2[k]|^2}} \\ &= \frac{1}{\sigma_1\sigma_2 N^2}\sum_k |X_1[k]\overline{X_2[k]}| \\ &= \frac{1}{\sigma_1\sigma_2 N^2}\sum_{k=1}^{N} |X_1[k]||X_2[k]|\cos(\theta_k).\end{aligned}$$

Note that the quantities $|X_1[k]|^2$ and $|X_2[k]|^2$ are estimates of the power spectra of $x_1[t]$ and $x_2[t]$. Since we assume the angles θ_k are selected randomly from the uniform distribution from $[0, 2\pi]$, we can easily compute

$$\begin{aligned}\mathbb{E}\left(\mathrm{corr}(x_1[t], x_2[t])\right) &= 0 \\ \mathrm{var}\left(\mathrm{corr}(x_1[t], x_2[t])\right) &= \frac{1}{2\sigma_1\sigma_2}\sum_{k=1}^{N}\sqrt{|X_1[k]|^2|X_2[k]|^2} \qquad (1)\end{aligned}$$

The variance given in Equation 1 is maximized at $1/2\sigma_1\sigma_2$ when both power spectra have unit spikes at the same location, corresponding to sinusoids with identical frequency. Since we assume that $x_1[t]$ and $x_2[t]$ have identical power spectra, i.e. $|X_1[k]|^2 = |X_2[k]|^2$, Equation 1 is minimized at $1/2\sigma_1\sigma_2 N$ when $x_1[t]$ and $x_2[t]$ are white noise, i.e. $|X_1[k]| = 1/N$. (If we had not assumed the power spectra were identical, the variance would be minimized at zero for any non-overlapping power spectra, indicating that it is impossible for these timeseries to show a non-zero Pearson correlation.) Hence, any deviation from a uniformly-distributed power spectrum, as occurs in temporally-autocorrelated timeseries, will increase the variance in the Pearson correlation.

Supplementary Note 2: Economic clustering model

2.1 Overview

Our analysis focused on models at the level of time series. However, another popular type of generative model simulates network formation by directly adding edges to a set of nodes. We refer to these models as "generative graph models" to distinguish them from timeseries-based generative models. Generative graph models have been especially useful in modeling real-world networks[5,6]. For example, a generative graph model can capture the property that nodes which are already well-connected are likely to gain even more connections, i.e., that the "rich get richer"[7]. It effectively models the generative process of real-world networks such as social networks, where people with a large number of friends are more likely to gain new friends over time. By simulating how real networks are formed, generative graph models reveal the mechanisms that drive the high-level structure of those networks. For this reason, these models have quickly become ubiquitous in the natural and social sciences.

Brain networks share many properties with other complex biological and physical systems, so they can be analyzed using similar graph-theoretic methods. Recently, generative graph models have been used in neuroscience to model the human connectome[8–10], opening up new avenues of analysis of both functional[11,12] and structural[13–17] connectivity data. One reason why generative graph models are particularly attractive in the field is that they appear, at first glance, to model the trade-off between wiring cost and topological efficiency in brain networks[18]. In particular, generative graph models that construct networks based on the physical distance between brain regions and a few basic topological properties of the network have been especially effective[11,15]. In addition, generative graph models are relatively simple to construct and conceptualize, and they can be analyzed using computational tools from network science and graph theory.

Despite their popularity and conceptual simplicity, most generative graph models are removed from the underlying biological processes that occur in the brain. This is because the generative process used to construct those models is starkly different from the actual pipeline used to extract connectivity data from brain signals. Nevertheless, these models, such as the economical clustering (EC) model[11], are exceedingly effective at explaining the topology of brain connectivity.

In order to explain this effectiveness using a more biologically-motivated timeseries-based generative process, we compare the EC model from Ref.[11] directly to our spatiotemporal model.

The EC model is controlled by two parameters: one controls the connection probability with respect to distance, and the other with respect to homophily (i.e. shared neighbors, or clustering) between nodes[11]. There are a few apparent symmetries between the EC model and our spatiotemporal model. Both models produce realistic synthetic networks, and they are each driven by two features, one of which relates connectivity with physical distance. These initial similarities suggest the possibility that the second parameter in each model may also capture similar underlying properties. In fact, our findings show that both pairs of parameters in the two models are indeed tightly correlated, which opens up a new biologically-motivated interpretation of the generative graph model.

2.2 Economic clustering model

We compared the spatiotemporal model to the "economic clustering" (EC) model[11], which was chosen because of its good performance and its apparent parallels to our timeseries-based model. We briefly summarize the model here. We constructed a minimum spanning tree of the mean FC matrix using Kruskal's algorithm[19] to ensure that our procedure does not produce disconnected components. Then, edges were added one at a time according to a probabilistic wiring rule, until the number of edges in the model network and observed network were equal. The relative probability of adding a connection between node u and node v was determined by:

$$P(u,v) \propto D_{u,v}^{\eta} \times K(u,v)^{\gamma}$$

where $D_{u,v}$ is the Euclidean distance between nodes u and v, $K(u,v)$ is the number of shared neighbors between nodes u and v, described below, and η (the distance parameter) and γ (the clustering parameter) are free parameters. The exponent η controls the extent to which distance impacts connection probability. For highly negative η, short-range connections are much more likely; for values of η close to 0, all connections regardless of distance are almost equally likely to be added.

$K(u,v)$ represents the number of shared neighbors between nodes u and v, or equivalently, the number of nodes that are connected to both node u and node v. We can compute this as $K(u,v) = \sum_w A_{uw}A_{wv}$ where A is the adjacency matrix of the graph, i.e. $A_{ij} = 1$ if nodes i and j are connected, and $A_{ij} = 0$ otherwise. Note that $K(u,v)$ must be recomputed at each iteration of the algorithm, since the number of shared neighbors can change after an edge is added. The exponent γ controls the extent to which the number of shared neighbors impacts connection probability. For $\gamma < 0$, nodes with few shared neighbors are more likely to be connected; for

$\gamma > 0$, nodes with many shared neighbors are more likely to be connected.

Note that this procedure does not generate networks which vary smoothly with the parameters η and γ. In other words, for a given seed, a small change in either of these parameters could cause major changes in the topology of the generated graph. This occurs because each edge addition causes a cascading change in connection probabilities, and thus, even slight differences in the first edges added will have a large impact on the generated graph.

It has been shown that generative graph models which only specify connection probability as a function of distance do not capture features such as small-worldness, modularity, and degree distribution of real brain networks; rather, a second parameter is necessary to emulate these topological properties[11]. These findings are consistent with the results of our spatiotemporal model. In particular, our spatiotemporal model has two main distinguishing features: spatial autocorrelation and temporal autocorrelation. This apparent symmetry between the spatiotemporal model and the EC model motivates our comparison of the underlying parameters across the two models.

2.3 Fitting the spatiotemporal model to the EC model

To determine whether the parameters of the spatiotemporal model and the EC model are correlated, we developed a fitting procedure linking the two models. Because the model is determined by probabilistic wiring rules, a small perturbation in a model parameter may cause a large change in graph structure. Thus, the highly stochastic nature of the EC model produces a rugged optimization landscape, which prevents us from fitting the parameters η and γ using conventional optimization algorithms. In order to circumvent this problem, we instead fit the spatiotemporal model to instances of the EC model, since the spatiotemporal model has a smooth optimization landscape.

Overall, the procedure consists of the following steps. For a given pair of (η, γ), we generate ten instances of the EC model. Then, we fit the spatiotemporal model to the population of simulated EC models and record the optimal parameters SA-λ and TA-Δ_1. We repeat these steps over a predetermined range of η and γ values that generally produce reasonable networks. On a high level, this procedure returns the parameters SA-λ and TA-Δ_1 of the spatiotemporal model that fit best to the EC model determined by given values of η and γ. This allows us to conduct further analyses to reveal how changes in the EC model parameters affect changes in the corresponding spatiotemporal model parameters, thereby linking the two models for comparison.

In order to perform this fitting procedure, we need to establish a similarity metric between

the multiple instances of the EC model and the single instance of the spatiotemporal model. We also require an optimization method for fitting parameters that maximize this similarity metric.

To evaluate the similarity between two networks, we use the energy function presented in Ref.[15],

$$E = \max(KS_k, KS_c, KS_b, KS_e)$$

where each argument in the max function is the two-sample Kolmogorov-Smirnov test statistic between the networks' degree (k), clustering (c), betweenness centrality (b), and edge length (e) distributions. Note that for consistency with Ref.[15], this energy function measures fitness based on nodal graph metrics, in contrast to the energy function for fitting the spatiotemporal model to data, which is based on eigenvalues.

To reduce noise and to produce population-level fits, we collectively minimize the average energy between all instances of the EC model and a single instance of the spatiotemporal model,

$$E_{avg} = \frac{1}{n}\sum_{i=1}^{n} \max(KS_k^i, KS_c^i, KS_b^i, KS_e^i)$$

where n is the number of subjects in our dataset, and each argument to the max function is the Kolmogorov-Smirnov statistic between the distributions of ith instance of the generative graph model and the spatiotemporal model for the corresponding graph metric. Instances of the EC model differ only in the initial random seed. We used differential evolution to determine the optimal model parameters.

2.4 Results

One apparent similarity between the construction of the EC model and the homogeneous spatiotemporal model is that each model includes a parameter that controls connection probability with respect to distance between nodes: the distance parameter η in the EC model and the SA-λ parameter in the homogeneous spatiotemporal model. As $\eta \to 0^-$, distance plays a smaller role in determining connection probability, so long- and short-range connections are more equally likely to be chosen. As SA-$\lambda \to 0^+$, the spatial autocorrelation approaches zero for all non-zero distances, so distance has a diminishing effect on the correlation. Therefore, we expect the EC distance parameter η to be negatively correlated with SA-λ. To test this empirically, we fix the EC clustering parameter at several different values and vary the EC distance parameter, fitting the homogeneous spatiotemporal model to each of the EC model's simulated networks. We find a strong negative relationship between the EC distance parameter η and SA-λ (Figure ED9a). This confirms the similarities expected between the two spatial parameters.

The ability of both models to capture network topology, combined with the strong association of SA-λ and EC distance parameter, suggests that the second parameters of these two models may also be related. Specifically, we test whether the clustering parameter γ in the EC model and TA-Δ_1 in the homogeneous spatiotemporal model are associated as well. After applying the same procedure of simulating from the EC model and fitting with the homogeneous spatiotemporal model, we found a tight relationship between TA-Δ_1 and the EC clustering parameter (Figure ED9b). Indeed, varying these parameters in the two models results in similar changes in graph metrics (Figure ED9). By contrast, the TA-Δ_1 parameter is not strongly associated with the EC distance parameter (Figure ED9c), and the SA-λ parameter is not strongly associated with the EC clustering parameter (Figure ED9d). Thus, increases in SA breadth reduce the relative probability of nearby connections, and increases in TA increase the probability of connections occurring preferentially within clusters. This confirms our intuition of how SA and TA influence graph topology.

Supplementary References

1. Yule, G. U. Why do we Sometimes get Nonsense-Correlations between Time-Series?–A Study in Sampling and the Nature of Time-Series. *Journal of the Royal Statistical Society* **89,** 1 (Jan. 1926).
2. Arbabshirani, M. R. *et al.* Impact of autocorrelation on functional connectivity. en. *NeuroImage* **102,** 294–308. ISSN: 1053-8119 (Nov. 2014).
3. Schaworonkow, N., Blythe, D. A. J., Kegeles, J., Curio, G. & Nikulin, V. V. Power-law dynamics in neuronal and behavioral data introduce spurious correlations. en. *Human Brain Mapping* **36,** 2901–2914. ISSN: 1097-0193 (2015).
4. Afyouni, S., Smith, S. M. & Nichols, T. E. Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. en. *NeuroImage* **199,** 609–625. ISSN: 1053-8119 (Oct. 2019).
5. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393,** 440–442 (June 1998).
6. Barabási, A.-L. & Albert, R. Emergence of Scaling in Random Networks. *Science* **286,** 509–512 (Oct. 1999).

7. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74,** 47–97 (Jan. 2002).

8. Fornito, A., Zalesky, A. & Bullmore, E. *Fundamentals of Brain Network Analysis* en. ISBN: 9780124081185 (Academic Press, Mar. 2016).

9. Bassett, D. S. & Sporns, O. Network neuroscience. *Nature Neuroscience* **20,** 353–364 (Feb. 2017).

10. Betzel, R. F. *et al.* The modular organization of human anatomical brain networks: Accounting for the cost of wiring. *Network Neuroscience* **1,** 42–68 (Feb. 2017).

11. Vértes, P. E. *et al.* Simple models of human brain functional networks. en. *Proceedings of the National Academy of Sciences* **109,** 5868–5873. ISSN: 0027-8424, 1091-6490 (Apr. 2012).

12. Morgan, S. E., Achard, S., Termenon, M., Bullmore, E. T. & Vértes, P. E. Low-dimensional morphospace of topological motifs in human fMRI brain networks. *Network Neuroscience* **2,** 285–302 (Jan. 2018).

13. Hinne, M., Heskes, T., Beckmann, C. F. & van Gerven, M. A. Bayesian inference of structural brain networks. *NeuroImage* **66,** 543–552 (Feb. 2013).

14. Klimm, F., Bassett, D. S., Carlson, J. M. & Mucha, P. J. Resolving Structural Variability in Network Models and the Brain. *PLoS Computational Biology* **10** (ed Hilgetag, C. C.) e1003491 (Mar. 2014).

15. Betzel, R. F. *et al.* Generative models of the human connectome. *NeuroImage* **124,** 1054–1064 (Jan. 2016).

16. Tang, E. *et al.* Developmental increases in white matter network controllability support a growing diversity of brain dynamics. *Nature Communications* **8.** doi:10.1038/s41467-017-01254-4 (Nov. 2017).

17. Akarca, D., Vértes, P. E., Bullmore, E. T. & and, D. E. A. A generative network model of neurodevelopment. *Biorxiv.* doi:10.1101/2020.08.13.249391 (Aug. 2020).

18. Bullmore, E. & Sporns, O. The economy of brain network organization. *Nature Reviews Neuroscience* **13,** 336–349 (Apr. 2012).

19. Kruskal, J. B. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society* **7,** 48–48 (Jan. 1956).

Supplementary tables

Table S1

Graph metric fits for all models. For each model, in each dataset, the single subject fits are shown for assortativity, clustering, local efficiency, global efficiency, modularity, and transitivity (top) and SA-λ, SA-∞, global TA-Δ_1, mean-FC, var-FC, and kurt-FC (bottom). Fits are quantified using Lin's concordance, Spearman correlation r_s, and coefficient of determination R^2.

Dataset	Model	Assortativity			Clustering coef.			Local efficiency			Global efficiency			Modularity			Transitivity		
		Lin	r_s	R^2	Lin	r_s	R^2	Lin	r_s	R^2	Lin	r_s	R^2	Lin	r_s	R^2	Lin	r_s	R^2
HCP	Data (retest)	0.39	0.39	-0.22	0.37	0.39	-0.26	0.42	0.44	-0.17	0.4	0.38	-0.22	0.45	0.47	-0.12	0.53	0.52	0.04
	Spatiotemporal model	0.43	0.52	-0.4	0.63	0.63	0.33	0.74	0.79	0.54	0.2	0.32	-0.77	0.6	0.7	0.38	0.45	0.6	-0.66
	Intrinsic timescale + SA	0.34	0.41	0.13	-0.01	-0.17	-98.83	-0.07	-0.16	-4.5	0.01	0.46	-33.47	-0.1	-0.3	-3.39	0.02	0.37	-12.54
	SA only	0.02	0.17	-1.35	0.0	0.0	-5.24	0.0	-0.05	-9.36	0.01	0.2	-2.19	0.0	0.23	-5.7	0.0	0.51	-30.29
	TA only	-0.01	-0.13	-3.02	-0.0	0.01	-89.13	-0.01	-0.06	-3.06	0.0	0.23	-30.6	-0.01	-0.16	-3.78	0.0	0.17	-12.59
	Phase randomization	-0.02	-0.13	-1.99	-0.01	-0.39	-128.6	-0.06	-0.42	-7.96	0.01	0.31	-37.51	-0.06	-0.43	-5.11	-0.0	-0.06	-14.67
	Eigensurrogate	0.27	0.31	-0.03	-0.29	-0.49	-4.23	-0.08	-0.45	-2.67	0.17	0.29	-6.66	-0.05	-0.32	-1.51	-0.04	-0.15	-8.55
	Zalesky matching	0.16	0.37	-0.15	-0.02	-0.48	-75.8	-0.23	-0.56	-2.71	0.02	0.37	-28.21	-0.02	-0.27	-5.23	-0.0	-0.04	-10.98
	Edge reshuffle	0.0	-0.06	-6.79	-0.05	-0.64	-77.28	-0.1	-0.7	-10.35	0.03	0.44	-14.24	0.02	0.91	-10.81	0.01	0.26	-12.25
HCP-GSR	Data (retest)	0.43	0.43	-0.15	0.39	0.39	-0.23	0.44	0.45	-0.13	0.42	0.41	-0.16	0.42	0.43	-0.18	0.53	0.53	0.06
	Spatiotemporal model	0.34	0.46	-0.57	0.37	0.37	-0.27	0.52	0.68	-0.34	0.16	0.28	-1.2	0.36	0.49	-0.36	0.23	0.37	-1.95
	Intrinsic timescale + SA	0.28	0.43	-0.18	0.0	0.01	-181.85	0.01	0.06	-16.87	0.01	0.32	-50.21	-0.01	-0.06	-11.23	0.01	0.42	-20.14
	SA only	0.01	0.15	-0.08	0.0	-0.04	-5.33	-0.0	-0.09	-13.2	0.01	0.14	-3.09	0.0	0.1	-5.12	0.0	0.23	-25.09
	TA only	-0.02	-0.23	-4.48	0.0	0.14	-173.28	0.01	0.19	-14.99	0.0	0.32	-45.63	0.0	0.05	-14.31	0.0	0.32	-19.98
	Phase randomization	-0.03	-0.28	-2.85	-0.0	-0.08	-257.15	-0.01	-0.11	-33.41	0.0	0.23	-57.72	-0.01	-0.21	-17.76	0.0	0.13	-23.19
	Eigensurrogate	0.22	0.3	0.05	-0.05	-0.25	-17.47	-0.27	-0.49	-1.43	0.2	0.37	-3.74	-0.23	-0.45	-0.54	-0.02	-0.02	-2.67
	Zalesky matching	0.12	0.28	-0.45	-0.0	-0.25	-181.39	-0.03	-0.28	-14.85	0.0	0.23	-48.29	0.0	0.18	-16.84	0.0	0.06	-19.88
	Edge reshuffle	-0.02	-0.32	-6.93	-0.01	-0.4	-193.51	-0.03	-0.6	-41.27	0.01	0.34	-25.04	0.01	0.76	-28.89	-0.0	-0.05	-19.2
Yale-TRT	Data (retest)	0.13	0.15	-0.94	0.31	0.28	-0.42	0.41	0.4	-0.29	0.21	0.21	-0.68	0.36	0.36	-0.33	0.24	0.24	-0.6
	Spatiotemporal model	-0.12	-0.17	-1.65	0.16	0.55	-4.14	0.2	0.5	-4.13	0.22	0.55	-1.47	0.02	0.24	-32.27	0.1	0.31	-3.24
	Intrinsic timescale + SA	0.02	0.04	-1.42	0.01	0.34	-29.85	0.03	0.5	-16.04	0.01	0.23	-15.13	0.01	0.17	-28.83	0.01	0.32	-12.63
	SA only	0.01	0.43	-3.84	0.02	0.75	-5.49	0.0	0.47	-19.72	0.02	0.8	-6.5	-0.0	0.03	-0.63	0.01	0.82	-11.45
	TA only	0.0	0.05	-4.24	0.0	0.02	-29.49	-0.0	-0.01	-11.93	-0.0	-0.05	-14.58	0.0	0.11	-23.11	-0.0	-0.02	-12.26
	Phase randomization	-0.0	-0.02	-6.2	0.0	0.32	-75.2	0.0	0.27	-72.3	0.0	0.3	-35.41	0.0	0.22	-68.85	0.0	0.28	-24.73
	Eigensurrogate	0.26	0.3	-0.3	0.17	0.75	-6.59	0.22	0.41	-1.46	0.61	0.88	-0.24	0.19	0.33	-2.2	0.69	0.83	0.18
	Zalesky matching	0.06	0.1	-2.19	0.01	0.65	-57.6	0.02	0.58	-36.39	0.01	0.63	-28.49	0.01	0.4	-55.84	0.01	0.68	-21.15
	Edge reshuffle	-0.01	-0.27	-15.41	0.01	0.47	-103.56	-0.0	-0.11	-196.35	0.03	0.71	-27.9	-0.0	-0.13	-126.97	0.01	0.43	-24.62
Cam-CAN	Spatiotemporal model	0.16	0.39	-5.84	0.08	0.11	-0.5	0.11	0.13	-0.56	0.16	0.25	-2.08	0.1	0.16	-1.08	0.21	0.4	-3.24
	Intrinsic timescale + SA	-0.01	-0.04	-1.28	0.2	0.2	-0.42	0.14	0.25	-0.95	-0.02	-0.05	-1.71	0.16	0.28	-1.02	0.14	0.09	-0.72
	SA only	0.03	0.41	-1.92	0.0	0.13	-4.37	0.0	0.1	-8.3	0.03	0.24	-0.41	-0.0	0.01	-4.09	0.01	0.4	-8.48
	TA only	-0.0	-0.04	-6.79	0.01	0.06	-2.27	0.01	0.05	-2.36	-0.0	-0.04	-0.42	-0.0	-0.04	-3.31	-0.02	-0.07	-0.92
	Phase randomization	-0.02	-0.03	-5.63	-0.23	-0.32	-4.66	-0.28	-0.32	-2.03	0.04	0.14	-6.02	-0.2	-0.32	-1.95	-0.3	-0.35	-1.52
	Eigensurrogate	0.15	0.31	-2.01	0.04	0.11	-1.59	0.01	0.06	-2.9	0.04	0.23	-4.38	0.02	0.1	-2.23	0.14	0.49	-2.8
	Zalesky matching	0.03	0.04	-1.48	-0.24	-0.28	-1.86	-0.26	-0.31	-1.13	0.1	0.31	-4.12	-0.23	-0.28	-1.19	-0.15	-0.18	-2.69
	Edge reshuffle	-0.0	-0.02	-17.84	-0.08	-0.63	-22.34	-0.15	-0.7	-8.35	0.01	0.45	-17.11	0.03	0.77	-8.02	0.01	0.3	-12.15

Dataset	Model	SA-λ			SA-∞			Global TA-Δ_1			Mean-FC			Var-FC			Kurt-FC		
		Lin	r_s	R^2	Lin	r_s	R^2	Lin	r_s	R^2	Lin	r_s	R^2	Lin	r_s	R^2	Lin	r_s	R^2
HCP	Data (retest)	0.68	0.69	0.36	0.6	0.63	0.18	0.75	0.75	0.5	0.61	0.65	0.21	0.59	0.6	0.19	0.59	0.6	0.17
	Spatiotemporal model	0.39	0.82	-5.67	0.84	0.94	0.53	0.99	1.0	0.97	0.85	0.97	0.55	0.21	0.83	-1.9	0.81	0.93	0.48
	Intrinsic timescale + SA	1.0	1.0	1.0	-0.08	-0.47	-3.1	1.0	1.0	1.0	-0.06	-0.17	-3.24	0.03	0.84	-8.78	-0.01	-0.07	-43.16
	SA only	0.03	0.72	-171.82	0.69	0.81	0.17	-0.0	-0.27	-33.2	0.89	0.97	0.75	0.07	0.26	-2.48	0.12	0.57	-16.55
	TA only	0.0	0.57	-14.43	0.0	0.52	-3.24	1.0	1.0	1.0	0.0	0.39	-3.8	0.03	0.79	-7.64	0.41	0.78	0.35
	Phase randomization	0.0	-0.01	-14.54	-0.0	-0.02	-3.28	1.0	1.0	1.0	-0.0	-0.03	-3.8	0.02	0.78	-9.15	0.09	0.25	0.01
	Eigensurrogate	0.0	-0.0	-14.41	-0.0	-0.09	-3.3	0.0	0.02	-33.14	-0.0	-0.16	-3.8	0.15	0.8	-35.08	0.09	0.89	-4.54
	Zalesky matching	0.0	0.03	-14.42	0.98	0.99	0.95				1.0	1.0	1.0	0.99	0.99	0.98	-0.01	-0.16	-1.16
HCP-GSR	Data (retest)	0.72	0.73	0.44	0.54	0.59	0.06	0.78	0.78	0.56	0.56	0.63	0.11	0.67	0.67	0.34	0.6	0.6	0.19
	Spatiotemporal model	0.47	0.87	-4.91	0.59	0.64	-0.27	0.97	1.0	0.94	0.67	0.81	-0.11	0.17	0.84	-2.63	0.64	0.89	-0.33
	Intrinsic timescale + SA	1.0	1.0	1.0	-0.18	-0.25	-1.39	1.0	1.0	1.0	-0.13	-0.23	-1.64	0.02	0.84	-9.75	-0.02	-0.28	-67.02
	SA only	0.03	0.77	-166.95	0.44	0.32	-0.25	0.0	0.01	-34.62	0.59	0.81	-0.15	0.22	0.67	-1.42	0.13	0.75	-22.34
	TA only	0.0	0.14	-16.36	0.0	0.18	-2.22	1.0	1.0	1.0	-0.0	-0.18	-3.23	0.02	0.8	-8.68	0.28	0.74	-0.05
	Phase randomization	-0.0	-0.0	-16.4	-0.0	-0.03	-2.28	1.0	1.0	1.0	0.0	-0.0	-3.24	0.01	0.75	-10.32	0.04	0.23	-0.88
	Eigensurrogate	0.0	-0.0	-16.29	-0.0	0.0	-2.29	-0.0	-0.04	-34.8	-0.0	-0.09	-3.24	0.32	0.78	-5.76	0.04	0.77	-7.64
	Zalesky matching	0.0	0.04	-16.29	0.94	0.97	0.86				1.0	1.0	1.0	1.0	1.0	0.99	0.01	0.41	-3.61
Yale-TRT	Data (retest)	0.54	0.63	0.02	0.55	0.54	0.05	0.62	0.7	0.21	0.35	0.37	-0.4	0.33	0.38	-0.45	0.38	0.44	-0.4
	Spatiotemporal model	0.18	0.49	-11.93	0.02	0.13	-33.06	0.96	0.99	0.92	0.0	0.09	$<-10^5$	0.31	0.6	-0.64	-0.0	-0.14	-9.87
	Intrinsic timescale + SA	0.96	0.96	0.93	0.03	0.07	-1.68	0.83	0.98	0.63	0.01	0.17	-422.87	0.04	0.54	-5.92	-0.0	0.04	-7.74
	SA only	0.01	0.35	-120.14	0.01	0.06	-0.85	-0.0	-0.29	$<-10^4$	0.0	0.19	$<-10^5$	0.03	0.9	-13.43	0.06	0.35	-15.38
	TA only	0.0	0.12	-11.89	0.0	0.05	-1.0	0.86	1.0	0.7	0.0	0.05	-3.66	0.03	0.47	-3.21	0.0	0.53	-10.97
	Phase randomization	-0.0	-0.11	-14.4	-0.0	-0.01	-1.36	0.93	0.99	0.85	-0.03	-0.05	-2.18	0.01	0.51	-21.3	0.05	0.53	-4.52
	Eigensurrogate	0.0	-0.02	-14.28	0.0	0.03	-1.32	0	-0.03	$<-10^4$	-0.04	-0.07	-2.92	1.0	1.0	1.0	0.01	0.23	-20.46
	Zalesky matching	0.0	0.03	-14.34	0.0	0.01	-1.38				0.16	0.24	-2.86	0.99	0.96	0.97	0.01	0.46	-10.55
Cam-CAN	Spatiotemporal model	-0.02	0.14	-1.18	0.66	0.83	-0.08	0.37	0.94	-3.66	0.77	0.98	0.03	0.45	0.63	-0.23	0.21	0.65	-0.2
	Intrinsic timescale + SA	0.0	0.07	-1.08	-0.01	-0.19	-4.11	0.86	0.99	0.65	0.06	0.3	-4.41	-0.05	-0.2	-0.4	-0.01	-0.11	-0.66
	SA only	0.0	-0.02	-10.33	0.32	0.83	-2.51	0.0	0.56	-532.71	0.97	0.98	0.95	0.06	0.61	-1.92	-0.0	-0.71	-0.73
	TA only	-0.0	-0.1	-1.58	0.0	0.36	-9.96	0.7	1.0	-0.06	0.0	0.33	-22.98	-0.01	-0.21	-4.03	-0.01	-0.4	-1.16
	Phase randomization	0.0	-0.02	-1.57	-0.0	-0.01	-9.95	0.95	0.98	0.88	0.0	-0.02	-22.92	0.03	-0.0	-0.67	-0.07	-0.5	-1.0
	Eigensurrogate	-0.0	-0.02	-1.54	0.0	0.06	-9.99	-0.0	-0.04	-628.92	-0.0	-0.12	-22.95	-0.03	-0.58	-128.4	-0.02	-0.75	-2.28
	Zalesky matching	0.0	-0.02	-1.51	0.71	0.87	0.52				1.0	1.0	1.0	0.93	0.94	0.86	0.41	0.67	0.21