Supporting Information

List of SI figures

List of SI tables

S1. Proof of Theorem 1.1

Preliminaries. To prove Theorem 1.1, we will require the notion of a contraction mapping:

Definition S1.1 (Contraction Mapping)**.** *Let $\mathcal{X}$ be a metric space with associated distance metric d. Then a mapping $F : \mathcal{X} \mapsto \mathcal{X}$ is called a contraction mapping if it satisfies for some $0 \leq \gamma < 1$:*

$$d(F(x), F(y)) \leq \gamma d(x, y).$$

Contraction mappings are known to have the following properties:

Theorem S1.2 (Contraction Mapping Theorem (34))**.** *Let $F : \mathcal{X} \mapsto \mathcal{X}$ be a contraction mapping. The F has a unique fixed point x^* satisfying $F(x^*) = x^*$. Furthermore, the sequence $x_{n+1} = F(x_n)$ (with x_0 arbitrarily initialized) satisfies $x_n \to x^*$.*

We will write the Eigenvector Smoothing iterations Eq. (3)-Eq. (5) in terms of an operator $G = (G_1, \ldots, G_T)$, given by:

$$G_1(\bar{U}_1, \ldots, \bar{U}_T) = \Pi_K(U_1 + \alpha \bar{U}_2) \quad \text{[S1]}$$

$$G_t(\bar{U}_1, \ldots, \bar{U}_T) = \Pi_K(U_t + \alpha \bar{U}_{t-1} + \alpha \bar{U}_{t+1}), \qquad t = 2, \ldots, T \quad \text{[S2]}$$

$$G_T(\bar{U}_1, \ldots, \bar{U}_T) = \Pi_K(U_T + \alpha \bar{U}_{T-1}), \quad \text{[S3]}$$

so that Eq. (3)-Eq. (5) can be written as $\bar{U}_{1:T}^{(\ell+1)} = G(\bar{U}_{1:T}^{(\ell)})$, where we have abbreviated $\bar{U}_{1:T} \equiv (\bar{U}_1, \ldots, \bar{U}_T)$. Our proof will use the following properties of G:

Lemma S1.3. *A necessary condition for $\bar{U}_1^*, \ldots, \bar{U}_T^*$ to be a global minimum of the optimization problem* Eq. (2) *is that $\bar{U}_{1:T}^* = G(\bar{U}_{1:T}^*)$.*

Lemma S1.4. *For $\alpha < \frac{1}{4\sqrt{2}+2}$, the mapping G is a contraction mapping.*

The proof of Lemma S1.4 will use the Davis-Kahan Theorem:

Theorem S1.5 (Davis and Kahan, (6, 21))**.** *Let Σ, Σ' be symmetric, suppose $S \subset \mathbb{R}$ is an interval, and suppose for some positive integer K that V, $V' \in \mathbb{R}^{n \times K}$, and the columns of V (V') form an orthonormal basis for the sum of eigenspaces of Σ (Σ') associated with the eigenvalues of Σ (Σ') in S. Let δ be the minimum of distance between any eigenvalues of Σ in S between any eigenvalues of Σ not in S. Then there exists an orthogonal matrix $R \in \mathbb{R}^{K \times K}$ such that*

$$||VR - V'||_F \leq \frac{\sqrt{2}}{\delta} ||\Sigma - \Sigma'||_F. \quad \text{[S4]}$$

Proof of Theorem. The proof of Theorem 1.1 is as follows:

Proof of Theorem 1.1. Eq. (2) must have a global minimum, since its feasible region is bounded and its objective function is bounded from below by 0. By Lemma S1.3, the global minimum of Eq. (2) must be a fixed point of G. By Lemma S1.4, G is a contraction mapping, and hence by Theorem S1.2 it has a unique fixed point. It follows that the unique fixed point of G is the global minimum of Eq. (2).

Since the Eigenvector Smoothing iterations are $\bar{U}_{1:T}^{(\ell+1)} = G(\bar{U}_{1:T}^{(\ell)})$, Lemma S1.4 and Theorem S1.2 imply that $\bar{U}_{1:T}^{(\ell)}$ converges to the fixed point of G as $\ell \to \infty$. Thus, $\bar{U}_{1:T}^{(\ell)}$ converges to the global minimum of Eq. (2), proving the theorem. □

Proof of Lemmas S1.3 and S1.4. The proof of Lemmas S1.3 and S1.4 are as follows:

Proof of Lemma S1.3. Let $\mathcal{V} = \{V \in \mathbb{R}^{n\times K} : V^TV = I\}$, so that $\mathcal{U} = \{VV^T : V \in \mathcal{V}\}$ denotes the feasible region of each $\bar{U}_t$ in Eq. (2). A necessary condition for $\bar{U}^*_{1:T}$ to be the global minimum of Eq. (2) is that each $\bar{U}^*_t$ for must minimize the objective when the other variables are held constant:

$$\bar{U}^*_1 = \arg\min_{\bar{U}_1\in\mathcal{U}} \|U_1 - \bar{U}_1\|_F^2 + \alpha\|\bar{U}_1 - \bar{U}^*_2\|_F^2 \tag{S5}$$

$$\bar{U}^*_t = \arg\min_{\bar{U}_t\in\mathcal{U}} \|U_t - \bar{U}_t\|_F^2 + \alpha\|\bar{U}_t - \bar{U}^*_{t+1}\|_F^2 + \alpha\|\bar{U}^*_{t-1} - \bar{U}_t\|_F^2, \qquad t = 2,\dots,T-1 \tag{S6}$$

$$\bar{U}_T = \arg\min_{\bar{U}_T\in\mathcal{U}} \|U_t - \bar{U}_T\|_F^2 + \alpha\|\bar{U}^*_{T-1} - \bar{U}_T\|_F^2. \tag{S7}$$

To prove the lemma, it thus suffices to show this condition is equivalent to $\bar{U}^*_t = G_t(\bar{U}^*_{1:T})$ for $t = 1,\dots,T$.

Using the identity $\|M\|_F^2 = \mathrm{Tr}(M^TM)$ and the fact that $\|U\|_F^2 = K$ for all $U \in \mathcal{U}$, algebraic manipulation of Eq. (S6) yields

$$\begin{aligned}\bar{U}^*_t &= \arg\max_{\bar{U}_t\in\mathcal{U}} \mathrm{Tr}(\bar{U}_t^TU_t) + \alpha\,\mathrm{Tr}(\bar{U}_t^T\bar{U}^*_{t+1}) + \alpha\,\mathrm{Tr}(\bar{U}_t^T\bar{U}^*_{t-1})\\ &= \arg\max_{\bar{U}_t\in\mathcal{U}} \mathrm{Tr}(\bar{U}_t^T(U_t + \alpha\bar{U}^*_{t+1} + \alpha\bar{U}^*_{t-1})),\end{aligned} \tag{S8}$$

Since $\bar{U}^*_t \in \mathcal{U}$, we can let $\bar{U}^*_t = \bar{V}^*\bar{V}^{T^*}$ for some $\bar{V} \in \mathcal{V}$, so that Eq. (S8) implies

$$\begin{aligned}\bar{V}^* &= \arg\max_{\bar{V}\in\mathcal{V}} \mathrm{Tr}(\bar{V}^T(U_t + \alpha\bar{U}^*_{t+1} + \alpha\bar{U}^*_{t-1})\bar{V})\\ &= \mathrm{Eigvec}_K(U_t + \alpha\bar{U}^*_{t+1} + \alpha\bar{U}^*_{t-1})\end{aligned} \tag{S9}$$

where Eigvec_K denotes the $n \times K$ matrix in $\mathcal{V}$ whose columns are the first K eigenvectors. Here we have used the fact that $U_t, \bar{U}^*_{t+1}$, and $\bar{U}^*_{t-1}$ are positive semidefinite matrices. Eq. (S9) implies that

$$\begin{aligned}\bar{U}^*_t &= \Pi_K(U_t + \alpha\bar{U}^*_{t+1} + \alpha\bar{U}^*_{t-1})\\ &= G_t\left(\bar{U}^*_{1:T}\right)\end{aligned}$$

for $t = 2,\dots,T-1$. Analogous arguments, using Eq. (S5) and Eq. (S7) instead of Eq. (S6), show the same for $t = 1$ and $t = T$. □

Proof of Lemma S1.4. Given $\bar{U}_{1:T}$ and $\bar{U}'_{1:T}$, let Σ_t and Σ'_t denote the matrices

$$\begin{aligned}\Sigma_1 &= U_1 + \alpha\bar{U}_2 \quad &&\text{and} \quad &&\Sigma'_1 = U_1 + \alpha\bar{U}'_2\\ \Sigma_t &= U_t + \alpha\bar{U}_{t-1} + \alpha\bar{U}_{t+1} \quad &&\text{and} \quad &&\Sigma'_t = U_t + \alpha\bar{U}'_{t-1} + \alpha\bar{U}'_{t+1}, \qquad t = 2,\dots,T-1\\ \Sigma_T &= U_T + \alpha\bar{U}_{T-1} \quad &&\text{and} \quad &&\Sigma'_T = U_T + \alpha\bar{U}'_{T-1},\end{aligned}$$

so that $G_t(\bar{U}_{1:T}) = \Pi_K(\Sigma_t)$.

Let V_t and V'_t denote orthonormal bases for the first K eigenvectors of Σ_t and Σ'_t, respectively. The following chain of equations can be seen to hold for $t = 2,\dots,T-1$:

$$\begin{aligned}\|G_t(\bar{U}_{1:T}) - G_t(\bar{U}'_{1:T})\|_F &= \|\Pi_K(\Sigma_t) - \Pi_K(\Sigma'_t)\|_F\\ &\overset{①}{=} \|V_tRR^TV_t^T - V'_tV'^T_t\|_F\\ &= \|(V_tR - V'_t)R^TV_t^T + V'_t(V_tR - V'_t)^T\|_F\\ &\le \|(V_tR - V'_t)R^TV_t^T\| + \|V'_t(V_tR - V'_t)^T\|_F\\ &\overset{②}{=} \|(V_tR - V'_t)R^T\|_F + \|V_tR - V'_t\|_F\\ &= 2\|V_tR - V'_t\|_F\\ &\overset{③}{\le} \frac{2\sqrt{2}}{\delta}\|\Sigma_t - \Sigma'_t\|_F\\ &= \frac{2\sqrt{2}}{\delta}\alpha\|\bar{U}_{t-1} + \bar{U}_{t+1} - \bar{U}'_{t-1} - \bar{U}'_{t+1}\|_F\\ &\le \frac{2\sqrt{2}}{\delta}\alpha(\|\bar{U}_{t-1} - \bar{U}'_{t-1}\|_F + \|\bar{U}_{t+1} - \bar{U}'_{t+1}\|_F)\\ &\overset{④}{\le} \frac{2\sqrt{2}}{1-2\alpha}\alpha(\|\bar{U}_{t-1} - \bar{U}'_{t-1}\|_F + \|\bar{U}_{t+1} - \bar{U}'_{t+1}\|_F)\end{aligned} \tag{S10}$$

where we have used the following steps:

①: Let $R \in \mathbb{R}^{K\times K}$ satisfy $RR^T = I$. (R will satisfy additional conditions in ③ below).
②: This step holds because $\|M^T V^T\|_F^2 = \mathrm{Tr}(M^T V^T V M) = \|M\|_F^2$ for any $V \in \mathcal{V}$.
③: This is the Davis-Kahan theorem. δ is the difference between the Kth and $(K+1)$th eigenvalues for Σ_t.
④: Given $U_1, U_2, U_3 \in \mathcal{U}$ (where $\mathcal{U}$ is defined in the proof of Lemma S1.3), it holds (by Weyl's inequality) that the Kth eigenvalue of $U_1 + \alpha(U_2 + U_3)$ is at least 1, and the $(K+1)$th eigenvalue is at most 2α. As a result, $\delta \geq 1 - 2\alpha$.

By analogous arguments, a similar condition holds for $t = 1$ and $t = T$:

$$\|G_1(\bar{U}_{1:T}) - G_1(\bar{U}'_{1:T})\|_F \leq \frac{2\sqrt{2}}{1-2\alpha}\alpha\|\bar{U}_2 - \bar{U}'_2\|_F \quad [S11]$$

$$\|G_T(\bar{U}_{1:T}) - G_T(\bar{U}'_{1:T})\|_F \leq \frac{2\sqrt{2}}{1-2\alpha}\alpha\|\bar{U}_{T-1} - \bar{U}'_{T-1}\|_F. \quad [S12]$$

Summing Eq. (S10)-Eq. (S12) over $t = 1, \ldots, T$ yields

$$\sum_{t=1}^{T}\|F_t(\bar{U}_{1:T}) - F_t(\bar{U}'_{1:T})\|_F \leq \frac{4\sqrt{2}\alpha}{1-2\alpha}\left(\sum_{t=1}^{T}\|\bar{U}_t - \bar{U}'_t\|_F\right),$$

establishing that when $\frac{4\sqrt{2}\alpha}{1-2\alpha} < 1$, then F is a contraction with metric $d(\bar{U}_{1:T}, \bar{U}'_{1:T}) = \sum_{t=1}^{T}\|\bar{U}_t - \bar{U}'_t\|_F$. □

S2. Derivation of Eq. (2) as Relaxation of Smoothed K-means Clustering

Preliminaries. Let $z \in [K]^n$ and $z' \in [K']^n$ denote cluster assignment vectors with K and K' clusters respectively. Let n_{ab}, n_a and n'_b for $a \in [K]$ and $b \in [K']$ denote class counts in z and z' as given by

$$n_{ab} = \sum_{i=1}^{n} 1\{z_i = a, z'_i = b\}, \qquad n_a = \sum_{i=1}^{n} 1\{z_i = a\}, \qquad n'_b = \sum_{i=1}^{n} 1\{z'_i = b\}.$$

The chi-squared distance between two partitions z and z' is defined in (27, 28) to be

$$\Delta(z, z') = \frac{K + K'}{2} - \chi^2(z, z'),$$

where $\chi^2(z, z')$ is defined to be

$$\chi^2(z, z') = \sum_{a\in[K]}\sum_{b\in[K']}\frac{n_{ab}^2}{n_a n'_b}.$$

Given data points $x_i \in \mathbb{R}^d$ for $i = 1, \ldots, n$, and cluster labels $z \in [K]^n$, let $\mu_1, \ldots, \mu_K \in \mathbb{R}^d$ denote the cluster means given by

$$\mu_k = \frac{1}{n_k}\sum_{i:z_i=k} x_i, \qquad k = 1, \ldots, K,$$

The K-means objective function is given by

$$\text{K-means}(z) = \sum_{i=1}^{n}\|x_i - \mu_{z_i}\|^2.$$

Given $z \in [K]$ and $z' \in [K']$, let $Z \in \{0,1\}^{n\times K}$ and $Z' \in \{0,1\}^{n\times K'}$ denote normalized indicator matrices given by

$$Z_{ia} = \frac{1}{\sqrt{n_a}}1\{z_i = a\} \qquad \text{and} \qquad Z'_{ib} = \frac{1}{\sqrt{n'_b}}1\{z'_i = b\},$$

observing that $Z^T Z = Z'^T Z' = I$.

Lemmas S2.1 and S2.2 are well-known (17, 28), and give alternative expressions for the K-means objective and chi-squared distance, as functions of Z and Z' instead of z and z':

Lemma S2.1. *Let $X \in \mathbb{R}^{n\times n}$ be given by $X_{ij} = x_i^T x_j$. It holds that*

$$\textit{K-means}(z) = \sum_{i=1}^{n}\|x_i\|^2 - \mathrm{Tr}(X, ZZ^T).$$

Lemma S2.2. *It holds that*

$$\|ZZ^T - Z'Z'^T\|_F^2 = 2\Delta(z, z').$$

Plugging Lemmas S2.1 and S2.2 into Eq. (6) gives the following identity:

Lemma S2.3. *The smoothed K-means objective function given by* Eq. (6)*,*

$$\min_{\{\mu_{tk}\},\{z_t\}} \sum_{t=1}^{T}\sum_{i=1}^{n} \|x_{ti} - \mu_{t,z_t(i)}\|^2 + \frac{\alpha}{2}\sum_{i=1}^{T-1}\Delta(z_t, z_{t+1}),$$

is equivalent to

$$\min_{Z_1,\dots,Z_T} \sum_{t=1}^{T}(-\operatorname{Tr}(X_t, Z_t Z_t^T)) + \frac{\alpha}{2}\sum_{t=1}^{T-1}\|Z_t Z_t^T - Z_{t+1}Z_{t+1}^T\|_F^2,$$

where $X_t \in \mathbb{R}^{n\times n}$ denotes the gram matrix $[X_t]_{ij} = x_{ti}^T x_{tj}$, and $Z_t \in \mathbb{R}^{n\times n}$ denotes the normalized indicator matrix of z_t, so that $[Z_t]_{ik} = n_{tk}^{-1/2} 1\{z_{ti} = k\}$, where $n_{tk} = \sum_{i=1}^n 1\{z_{ti} = k\}$.

Eq. (2) is Relaxation of Eq. (6). Similar to results found in (17, 28), here we show that Eq. (2) is a relaxation of Eq. (6). By Lemma S2.3, the smoothed K-means objective given by Eq. (6) can be written as

$$\min_{Z_1,\dots,Z_T\in\mathcal{Z}} \sum_{t=1}^{T}(-\operatorname{Tr}(X_t, Z_t Z_t^T)) + \frac{\alpha}{2}\sum_{t=1}^{T-1}\|Z_t Z_t^T - Z_{t+1}Z_{t+1}^T\|_F^2,$$

where $\mathcal{Z}$ denotes the set of normalized indicator matrices inducible by some $z \in [K]^n$. This is a difficult combinatorial problem. Since $Z^T Z = I$ for all $Z \in \mathcal{Z}$, a relaxation is to remove the constraint $Z_1,\dots,Z_T \in \mathcal{Z}$ and instead require $Z_1,\dots,Z_T \in \mathcal{V}$, where $\mathcal{V} = \{V \in \mathbb{R}^{n\times K} : V^T V = I\}$. This results in the relaxed optimization problem

$$\min_{V_1,\dots,V_T\in\mathcal{V}} \sum_{t=1}^{T}(-\operatorname{Tr}(X_t, V_t V_t^T)) + \frac{\alpha}{2}\sum_{t=1}^{T-1}\|V_t V_t^T - V_{t+1}V_{t+1}^T\|_F^2,$$

where $V_1,\dots,V_T$ denotes the relaxed optimization variable. Letting $\bar{U}_t = V_t V_t^T$, this is equivalent to optimizing Eq. (2),

$$\min_{\bar{U}_1,\dots,\bar{U}_T\in\mathcal{U}} \frac{1}{2}\sum_{t=1}^{T}\|X_t - \bar{U}_t\|_F^2 + \frac{\alpha}{2}\sum_{t=1}^{T-1}\|\bar{U}_t - \bar{U}_{t+1}\|_F^2,$$

where $\mathcal{U} = \{VV^T : V \in \mathcal{V}\}$, and we have used the identity $2\operatorname{Tr}(X_t, \bar{U}_t) = \|X_t - \bar{U}_t\|_F^2 - \|X_t\|_F^2 - \|\bar{U}_T\|_F^2$ and the fact that $\|\bar{U}_t\|_F^2 = K$ for all $\bar{U}_t \in \mathcal{U}$.

Proof of Lemmas S2.1, S2.2, and S2.3. These proofs follow closely (17) and (28), and are included for self-completeness.

Proof of Lemma S2.1. It holds that

$$\begin{aligned}
\text{K-means}(z) &= \sum_{i=1}^{n}\|x_i - \mu_{z_i}\|^2 \\
&= \sum_{i=1}^{n}\left(\|x_i\|^2 - 2x_i^T\mu_{z_i} + \|\mu_i\|^2\right) \\
&= \sum_{i=1}^{n}\|x_i\|^2 - 2\sum_{k=1}^{K} n_k \mu_k^T\mu_k + \sum_{k=1}^{K} n_k\|\mu_k\|^2 \\
&= \sum_{i=1}^{n}\|x_i\|^2 - \sum_{k=1}^{K} n_k\|\mu_k\|^2 \\
&= \sum_{i=1}^{n}\|x_i\|^2 - \sum_{k=1}^{K}\left(\frac{1}{n_k}\sum_{i:z_i=k}\sum_{j:z_j=k} x_i^T x_j\right) \\
&= \sum_{i=1}^{n}\|x_i\|^2 - \sum_{k=1}^{K}\left(\frac{1}{n_k}\sum_{i:z_i=k}\sum_{j:z_j=k} X_{ij}\right) \\
&= \sum_{i=1}^{n}\|x_i\|^2 - \operatorname{Tr}(X, ZZ^T)
\end{aligned}$$

where the third and fifth equalities use the identity that $\mu_k = \frac{1}{n_k}\sum_{i:z_i=k} x_i$, and the final equality uses the fact that $[ZZ^T]_{ij}$ is given by

$$[ZZ^T]_{ij} = \frac{1}{n_{z_i}} 1\{z_i = z_j\}.$$

□

Proof of Lemma S2.2. It can be seen that $\|ZZ^T - Z'Z'^T\|_F^2$ equals

$$\begin{aligned}\|ZZ^T - Z'Z'^T\|_F^2 &= \|ZZ^T\|_F^2 - 2\operatorname{Tr}\left(ZZ^TZ'Z'^T\right) + \|Z'Z'^T\|_F^2 \\ &= K - 2\sum_{a\in[K]}\sum_{b\in[K']}\frac{n_{ab}^2}{n_a n_b'} + K' \\ &= 2\Delta(z,z'),\end{aligned}$$

where we have used the identities

$$\begin{aligned}\|ZZ^T\|_F^2 &= \sum_{i=1}^n\sum_{j=1}^n [ZZ^T]_{ij}^2 \\ &= \sum_{i=1}^n\sum_{j=1}^n \frac{1}{n_{z_i}^2}1\{z_i = z_j\} \\ &= K,\end{aligned}$$

and

$$\begin{aligned}\operatorname{Tr}(ZZ^TZ'Z'^T) &= \operatorname{Tr}(Z^TZ'Z'^TZ) \\ &= \|Z^TZ'\|_F^2 \\ &= \sum_{a\in[K]}\sum_{b\in[K']}\frac{n_{ab}^2}{n_a n_b'}.\end{aligned}$$

□

Proof of Lemma S2.3. Use Lemmas S2.1 and S2.2 to rewrite the terms appearing in Eq. (6). □

S3. Convergence and Relaxations Results for Laplacian Smoothing

Convergence Result. Let $\lambda_{t,1} \geq \cdots \geq \lambda_{t,n} \geq 0$ denote the eigenvalues of the matrix $|L_t|$, and let $\delta_{|L_t|} = \lambda_{t,K} - \lambda_{t,K+1}$ denote the gap between the Kth and $(K+1)$th eigenvalues of $|L_t|$. Let $\delta_{\min} = \min_t \delta_{|L_t|}$.

Theorem S3.1 states that the Laplacian smoothing iterations given by Eq. (7)-Eq. (9) converge to the global minimum of Eq. (10), provided that α is not too large compared to $\delta_{\min}$.

Theorem S3.1. *For $\alpha < \frac{\delta_{\min}}{4\sqrt{2}+2}$, the iterations* Eq. (7)-Eq. (9) *converge to the global minimizer of* Eq. (10)*, under any feasible initialization*

A proof sketch is included at the end of this section.

Eq. (10) as Relaxation of Smoothed K-means Clustering. Let x_{ti} denote the ith row of the square root of $|L_t|$ (i.e., the eigencoordinates scaled by the square root of their eigenvalues), so that the gram matrix X_t is equal to $|L_t|$. The arguments of Section S2 make no assumptions on $\{x_{ti}\}$, and imply that

1. By Lemma S2.3 the smoothed K-means problem

$$\min_{\{\mu_{tk}\},\{z_t\}}\sum_{t=1}^T\sum_{i=1}^n \|x_{ti} - \mu_{t,z_t(i)}\|^2 + \frac{\alpha}{2}\sum_{t=1}^{T-1}\Delta(z_t, z_{t+1}),$$

   can be written as

$$\min_{Z_1,\ldots,Z_T\in\mathcal{Z}}\sum_{t=1}^T(-\operatorname{Tr}(|L_t|, Z_tZ_t^T)) + \frac{\alpha}{2}\sum_{t=1}^{T-1}\|Z_tZ_t^T - Z_{t+1}Z_{t+1}^T\|_F^2,$$

   where $\mathcal{Z}$ denotes the set of normalized indicator matrices inducible by some $z \in [K]^n$.

2. This problem can be relaxed in the same manner as Section S2, by replacing the constraint $Z_1,\ldots,Z_T \in \mathcal{Z}$ with $Z_1,\ldots,Z_T \in \mathcal{V}$, where $\mathcal{V} = \{V \in \mathbb{R}^{n\times K} : V^TV = I\}$. This yields the optimization problem given by Eq. (10),

$$\min_{\bar{U}_1,\ldots,\bar{U}_T\in\mathcal{U}}\sum_{t=1}^T\||L_t| - \bar{U}_t\|_F^2 + \alpha\sum_{t=1}^{T-1}\|\bar{U}_t - \bar{U}_{t+1}\|_F^2,$$

   where $\mathcal{U} = \{VV^T : V \in \mathcal{V}\}$.

Proof Sketch of Theorem S3.1. The Laplacian Smoothing iterations Eq. (7)-Eq. (9) can be written as an operator $H = (H_1, \ldots, H_T)$ very similar to G:

$$\begin{aligned} H_1(\bar{U}_1, \ldots, \bar{U}_T) &= \Pi_K(|L_1| + \alpha \bar{U}_2) \\ H_t(\bar{U}_1, \ldots, \bar{U}_T) &= \Pi_K(|L_t| + \alpha \bar{U}_{t-1} + \alpha \bar{U}_{t+1}), \qquad t = 2, \ldots, T \\ H_T(\bar{U}_1, \ldots, \bar{U}_T) &= \Pi_K(|L_T| + \alpha \bar{U}_{T-1}), \end{aligned}$$

for which analogous versions of Lemma S1.3 and S1.4 can be shown to hold:

Lemma S3.2. *A necessary condition for $\bar{U}^*_{1:T}$ to be a local minimum of* Eq. (10) *is that $\bar{U}^*_{1:T} = H(\bar{U}^*_{1:T})$.*

Lemma S3.3. *For $\alpha < \frac{\delta_{\min}}{4\sqrt{2}+2}$, the mapping H is a contraction mapping.*

Proof Sketch of Theorem S3.1. Theorem S3.1 follows from identical arguments as that of Theorem 1.1, except that we use the results of Lemmas S3.2 and S3.3 in place of Lemmas S1.3 and S1.4. □

Proof Sketch of Lemma S3.2. The proof of Lemma S3.2 is identical to Lemma S1.3, except that we write $|L_t|$ wherever the term U_t appears. □

Proof Sketch of Lemma S3.3. The proof of Lemma S3.3 is nearly identical to that of Lemma S1.4. Aside from writing $|L_t|$ in place of U_t, step ④ must be changed to:

> ④: To bound δ, the Kth eigengap of $\Sigma_t = |L_t| + \alpha \bar{U}_{t-1} + \alpha \bar{U}_{t+1}$, we observe that the Kth eigenvalue of Σ_t is at least $\lambda_{t,K}$, and the $(K+1)$th eigenvalue is at most $\lambda_{t,K+1} + 2\alpha$. It thus holds that $\delta \geq \delta_{|L_t|} - 2\alpha$.

Analogous to Eq. (S10), this shows for $t = 2, \ldots, T-1$ that

$$\|H_t(\bar{U}_{1:T}) - H_t(\bar{U}'_{1:T})\|_F \leq \frac{2\sqrt{2}}{\delta_{\min} - 2\alpha} \alpha (\|\bar{U}_{t-1} - \bar{U}'_{t-1}\|_F + \|\bar{U}_{t+1} - \bar{U}'_{t+1}\|_F),$$

with similar identities for $t = 1$ and $t = T$. Summing over all t then yields

$$\sum_{t=1}^{T} \|H_t(\bar{U}_{1:T}) - H_t(\bar{U}'_{1:T})\|_F \leq \frac{4\sqrt{2}\alpha}{\delta_{\min} - 2\alpha} \left(\sum_{t=1}^{T} \|\bar{U}_t - \bar{U}'_t\|_F \right),$$

establishing the contraction identity when $\alpha < \frac{\delta_{\min}}{4\sqrt{2}+2}$. □

S4. Cross-Validation to Choose α

We apply the cross validation method of (24) (Algorithm 1) to choose α, which is a general purpose cross validation procedure that requires a user-selected likelihood or loss function, and a model fitting procedure to evaluate the likelihood. We use this procedure with the DCBM likelihood function, and for the fitting procedure we choose the classes z_t by applying PisCES to the Laplacianized adjacency matrices and then fit the DCBM parameters given z_t and the unlaplacianized A_t in the obvious way.

Given our choice of likelihood and fitting method, Algorithm 1 of (24) divides the dyads into folds, and evaluates on the test fold a degree corrected blockmodel (DCBM) that is fit to the training folds. When fitting the DCBM to the training folds, matrix completion is used to impute the withheld dyads in the test fold.

The specific steps of the method are:

1. Randomly divide the $\binom{n}{2}T$ dyads of $A_1, \ldots, A_T$ into S folds. Let Q denote the set of candidate values of α under consideration.

2. For each fold $s = 1, \ldots, S$ and time $t = 1, \ldots, T$ let $A_t^{(s)} \in \mathbb{R}^{n \times n}$ be defined as

$$[A_t^{(s)}]_{ij} = \begin{cases} [A_t]_{ij} & (i,j,t) \notin s \\ 0 & (i,j,t) \in s, \end{cases} \qquad \text{[S13]}$$

and let $\hat{A}_t^{(s)}$ denote the results of applying a matrix completion method (As Eq (1) in (24)) to each $A_t^{(s)}$ to impute the zeroed entries. Let $\bar{U}_1^{(\alpha,s)}, \ldots, \bar{U}_T^{(\alpha,s)}$ denote the output of the PisCES algorithm using smoothing parameter α and imputed $\hat{A}_1^{(s)}, \ldots, \hat{A}_T^{(s)}$. (The first step of PisCES is to Laplacianize the imputed adjacency matrices.)

3. Evaluate $\mathcal{L}^{(\alpha,s)}$, the log likelihood of a fitted DCBM on the sth fold:

$$\mathcal{L}^{(\alpha,s)} = \sum_{(i,j,t)\in s} [A_t]_{ij} \log(P_{tij}) + (1 - [A_t]_{ij}) \log(1 - P_{tij}) \quad [S14]$$

where $P_{tij} \equiv P_{tij}^{(s)}$ is the estimated probability of connection between node i and node j at time t under the fitted DCBM, and is given by

$$P_{tij} = d_{ti} d_{tj} \cdot B_t(z_{ti}, z_{tj}),$$

where $z_{ti} \equiv z_{ti}^{(s)}$ is the estimated community of node i at time t, as given by K-means clustering on the eigenvectors of $\bar{U}_t^{(\alpha,s)}$ with $\kappa(\bar{U}_t^{(\alpha,s)})$ classes; where $d_{ti} \equiv d_{ti}^{(s)}$ is the estimated DCBM degree parameter for node i, and is given by

$$d_{ti} = \sum_j [\hat{A}_t^{(s)}]_{ij},$$

and where $B_t \equiv B_t^{(s)} \in \mathbb{R}^{K\times K}$ is the estimated DCBM density parameter matrix, and is given by

$$[B_t]_{kl} = \left(\sum_{(i,j):z_{ti}=k, z_{tj}=l} [\hat{A}_t^{(s)}]_{ij} \right) / \left(\sum_{(i,j):z_{ti}=k, z_{tj}=l} d_{ti} d_{tj} \right).$$

4. Return the smoothing parameter $\alpha \in Q$ giving the best value for $\sum_s \mathcal{L}^{(\alpha,s)}$.

Remark When performing matrix completion for unknown K, we estimate the rank of matrix by the 'nature approach' showed in section 2.3 of (24).

Remark If A_t is real-valued, instead of Eq. (S14) we can evaluate the likelihood of the thresholded matrix:

$$\widehat{\mathcal{L}}^{(\alpha,s)} = \sum_{(i,j,t)\in s} 1\{[A_t]_{ij} > \tau\} \cdot \log(P_{tij}) + 1\{[A_t]_{ij} \le \tau\} \cdot \log(1 - P_{tij}),$$

where τ is a threshold parameter. Note that P_{tij} remains unchanged from the discussion of Eq. (S14). In our data analysis, we choose $\tau = 0.7^6$, where the choice of 0.7 follows previous works such as (29) and (32), and the 6th power follows from Eq. (17).

S5. Further discussion of δ

Cartoon Illustration. Fig. S1 shows the eigenvalues $\{\lambda_i\}$ of the Laplacianized adjacency matrix L, in a static example ($T = 1$) where A is generated by a DCBM with $K = 4$ communities. The eigengaps $\lambda_{i+1} - \lambda_i$ become very small for $i > 4$, so our rule correctly identifies $K = 4$ using Eq. (12).

Usage of Eq. (12) assumes the Erdos-Renyi null model, which we choose because the DCBM has observation noise that is independent across dyads. We remark that a "DCBM with $K = 1$" null could also have been used, with similar results. The largest eigengap occurs at $i = 1$, so choosing the largest eigengap would incorrectly estimate K as 1.

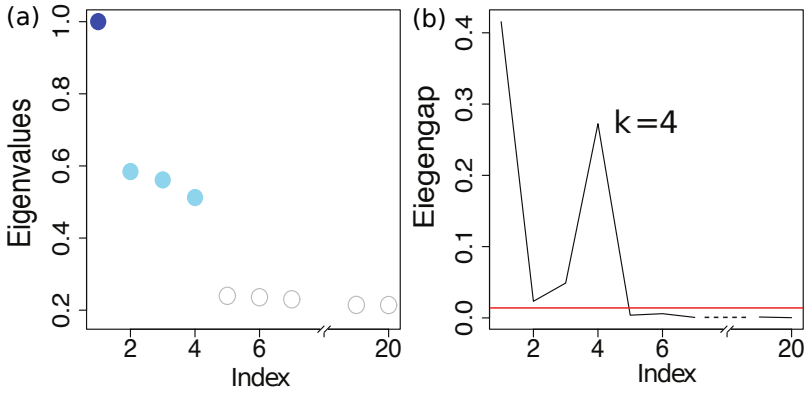

Fig. S1. (a) Largest 20 eigenvalues $\{\lambda_i\}$ of Laplacianized adjacency matrix L under DCBM with $K = 4$. (b) First 19 eigengaps $\lambda_{i+1} - \lambda_i$ corresponding to those 20 eigenvalues. The red line shows the threshold δ under Eq. (12).

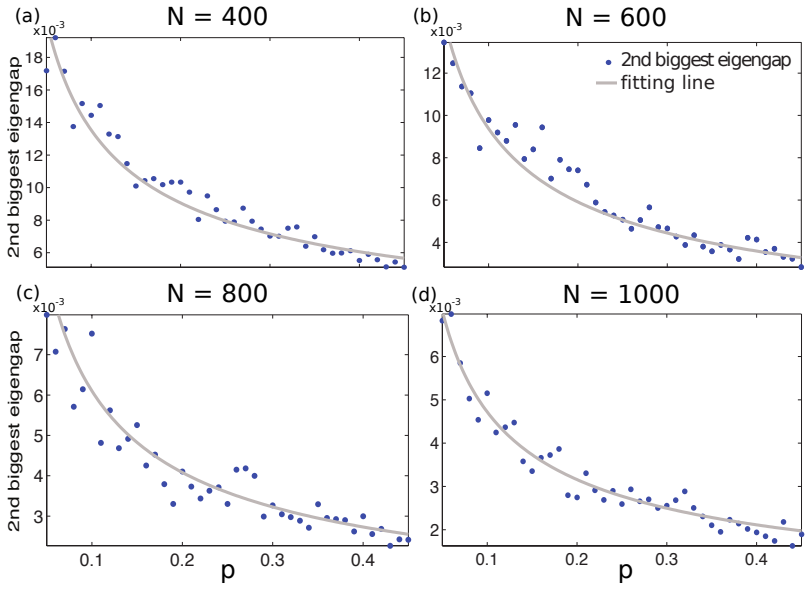

Fig. S2. Fitting lines for δ in different ER models with number of nodes n = 400 (a), 600 (b), 800 (c), and 1000 (d). and the density p ranges from 0.05 to 0.45.

Fast computation of δ under ER null. To speed up computation, we precomputed the threshold δ as given by Eq. (12), for degree-normalized inputs with n ranging between $0 < n \leq 1000$ and density $p = \frac{\#edges}{\binom{n}{2}}$ ranging between $0 < p \leq 0.5$, and found that the threshold is approximately given by the following curve (Fig. S2):

$$\delta \approx \frac{3.5}{p^{0.58} n^{1.15}}, \qquad \text{[S15]}$$

Null Model for Correlation Matrices. In Section 3, the adjacency matrices $A_1, \ldots, A_T$ are formed by transforming the empirical correlations between the n genes, by

$$[A_t]_{ij} = \begin{cases} |\operatorname{corr}(y_i, y_j)|^6 & i \neq j \\ 0 & i = j, \end{cases}$$

where $y_i \in \mathbb{R}^d$ denotes the vector of expression levels for gene i at time t.

We wish to compute a cutoff δ for the Laplacianized adjacency matrices L_t, below which the eigenvectors are discarded as being comparable to the "noise" eigenvectors of a null model. Our choice of null model for this setting is not an Erdos-Renyi (ER) random graph, but rather the following: let p denote the average of the correlations between the samples $\{y_i\}$:

$$p = \frac{1}{\binom{n}{2}} \sum_{i<j} \operatorname{corr}(y_i, y_j). \qquad \text{[S16]}$$

To generate a null model for A_t, we generate a random matrix $Y^{\text{null}} \in \mathbb{R}^{n \times d}$, whose columns are i.i.d. normal with covariance $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ given by

$$\hat{\Sigma} = (1-p)I + p11^T,$$

let $y_i^{\text{null}} \in \mathbb{R}^d$ denote the ith row of Y^{null}, and let $A_t^{\text{null}} \in \mathbb{R}^{n \times n}$ be given by

$$[A_t^{\text{null}}]_{ij} = \begin{cases} |\operatorname{corr}(y_i^{\text{null}}, y_j^{\text{null}})|^6 & i \neq j \\ 0 & i = j. \end{cases}$$

Computation of δ then follows as before,

$$\delta = \text{quantile}_{0.95} \left[\max\{|\lambda_i^{\text{null}}| - |\lambda_{i+1}^{\text{null}}|, i \geq 2\} \right], \qquad \text{[S17]}$$

where $|\lambda_1^{\text{null}}| \geq \cdots \geq |\lambda_n^{\text{null}}|$ are the eigenvalues of $|L_t^{\text{null}}|$, where L_t^{null} is the Laplacianized version of A_t^{null}, and the quantile is computed by simulation.

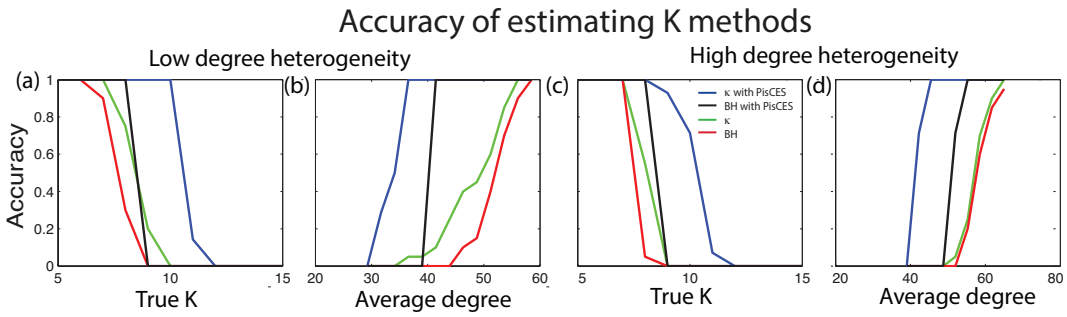

Fig. S3. Accuracy (measured by rate of correct estimation of K) of κ and Bethe Hessian method (BH) of (26). Methods are tested on static DCBM networks, with $n = 500$ and (a) $p_{\text{in}} = (0.3, 0.3)$, $p_{\text{out}} = 0.1$, $\psi \in (.75, 1.25)$ with K ranging from 5 to 15 and (b) $K = 10$, $p_{\text{in}} = (4p_{\text{out}}, 4p_{\text{out}})$, $\psi \in (.75, 1.25)$ and average degree ranging from 20 to 60. κ and BH are also tested within PisCES on dynamic DCBMs with $T = 6$. (c) and (d) are under identical settings as (a) and (b), but use $\psi \in (0.5, 1.5)$ instead of $(0.75, 1.25)$.

Remark: A Null Model for Dynamic Networks. The PisCES iterations are given by

$$\begin{aligned}
\bar{U}_1^{\ell+1} &= \Pi(|L_1| + \alpha\bar{U}_2^{\ell}) \\
\bar{U}_t^{\ell+1} &= \Pi(\alpha\bar{U}_{t-1}^{\ell} + |L_t| + \alpha\bar{U}_{t+1}^{\ell}) \qquad\qquad t = 2, \ldots, T-1 \\
\bar{U}_T^{\ell+1} &= \Pi(\alpha\bar{U}_{T-1}^{\ell} + |L_T|),
\end{aligned}$$

where the operator Π requires a model selection method κ that is applied not to $|L_t|$ alone, but rather to $|L_t| + \alpha(\bar{U}_{t-1}^{\ell} + \bar{U}_{t+1}^{\ell})$. For this reason, it may be of interest to explore null models for the matrix $|L_t| + \alpha(\bar{U}_{t-1}^{\ell} + \bar{U}_{t+1}^{\ell})$. One possible choice of null model is to generate $L_t^{\text{null}}, \bar{U}_{t-1}^{\text{null}}$, and $\bar{U}_{t+1}^{\text{null}}$ as follows:

1. Let L_t^{null} be generated either from an ER random graph or correlation matrix, whichever is more appropriate
2. Let $\bar{U}_{t-1}^{\text{null}}$ and $\bar{U}_{t+1}^{\text{null}}$ be random matrices in $\mathcal{U} = \{VV^T : V \in \mathbb{R}^{n\times K}, V^TV = I\}$ for K matching the rank of $\bar{U}_{t-1}^{\ell}$ and $\bar{U}_{t+1}^{\ell}$, respectively. $\bar{U}_{t-1}^{\text{null}}$ and $\bar{U}_{t+1}^{\text{null}}$ are generated independently of L_t^{null}.

Under this process, the cutoff δ equals the 0.95-level quantile of the eigengap (excluding λ_1) of $|L_t^{\text{null}}| + \alpha(\bar{U}_{t-1}^{\text{null}} + \bar{U}_{t+1}^{\text{null}})$ under this null distribution.

This generative process implies that $A_1, \ldots, A_T$ are assumed to be independent under the null. We investigate its performance under PisCES. In settings where the density of A_t is high (> 0.3) and the ER null is used for $|L_t|$ in step 1 above, we find that the additional randomness generated by adding random $\bar{U}_{t-1}$ and $\bar{U}_{t+1}$ results in more conservative (and improved) estimates; that is, δ is made smaller so that κ returns lower values for the estimated K. In other settings where A_t is noisier, the additional noise due to $\bar{U}_{t-1}$ and $\bar{U}_{t+1}$ is inconsequential in comparison, and does not change the estimated K; this is also the case when we try this null on the data described in Section 3.

S6. Simulations

Simulation Model (dynamic DCBM). In static settings (i.e., $T = 1$), simulated networks with n nodes are generated under a degree-corrected blockmodel (DCBM) with K classes, which is parameterized by a triplet (z, B, ψ), where $z \in [K]^n$ is the membership labels, $B \in [0, 1]^{K\times K}$ is a symmetric connectivity matrix, and $\psi \in [0.5, 1.5]^n$ models degree heterogeneity. Given (z, B, ψ), the entries of the adjacency matrix $A \in \{0, 1\}^{n\times n}$ are independently generated from the Bernoulli distribution with parameter as follows:

$$P(A_{ij} = 1) = \begin{cases} \psi_i\psi_j B_{z_i,z_j} & \text{if } i < j \\ 0 & \text{if } i = j, \end{cases} \qquad \text{with} \qquad A_{ij} = A_{ji} \text{ for } i > j$$

In dynamic settings, simulated networks are generated under a dynamic DCBM model of Eq. (13)-Eq. (15), with initial class labels $z^{(1)}$ chosen by

$$z_i^{(1)} \overset{\text{iid}}{\sim} \text{Multinomial}\left(\frac{1}{K}, \ldots, \frac{1}{K}\right).$$

Performance of κ. Fig S3 shows a comparison of κ (using ER null model as given by Eq. (12)) with the Bethe Hessian method (BH) of Eq. (2) of (26). In the static settings shown (which are dense DCBM networks), κ outperforms or performs comparably to BH. Fig S3 also shows that both κ and BH improve when they are used within PisCES (as part of the Π operator), suggesting that PisCES is flexible to the choice of model selection algorithm. We note that within PisCES we use $A_t + \alpha D_t^{1/2}\bar{U}_{t-1}D_t^{1/2} + \alpha D_t^{1/2}\bar{U}_{t+1}D_t^{1/2}$ as the input to BH, rather than $|L_t| + \alpha\bar{U}_{t-1} + \alpha\bar{U}_{t+1}$; this is done because BH is designed for the adjacency matrix as its input, rather than the Laplacian.

	n = 100			n = 500			n = 500, dense		
	hard	medium	easy	hard	medium	easy	hard	medium	easy
K	2	2	2	10	10	10	10	10	10
p_{in}	(0.2, 0.2)	(0.2,0.25)	(0.25, 0.25)	(0.2,0.35)	(0.2, 0.4)	(0.2, 0.45)	(0.55, 0.55)	(0.6,0.6)	(0.4, 0.8)
p_{out}	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.3	0.3

Table S1. Parameters for different cases in Fig. 1 and Fig. S4.

Overall Performance of PisCES. Table S1 gives DCBM parameter settings for 9 scenarios, where $n \in \{100, 500\}$ and the connectivity parameters p_{in} and p_{out} are chosen to create a range of signal-to-noise levels ("easy", "medium", and "hard"). Figures S4 and S5 show performance of PisCES and the comparison methods for these scenarios, under a range of values for r and T. For small networks where $n = 100$, Figure S4 shows that the methods behave similarly. For larger networks where $n = 500$, Figure S5 shows that PisCES often outperforms the others by a substantial margin provided that T is large – except when p_{in} corresponds to "hard" scenarios or $r = 0.5$, at which point all methods perform poorly. In nearly all cases, PisCES does comparable or better than the comparison methods.

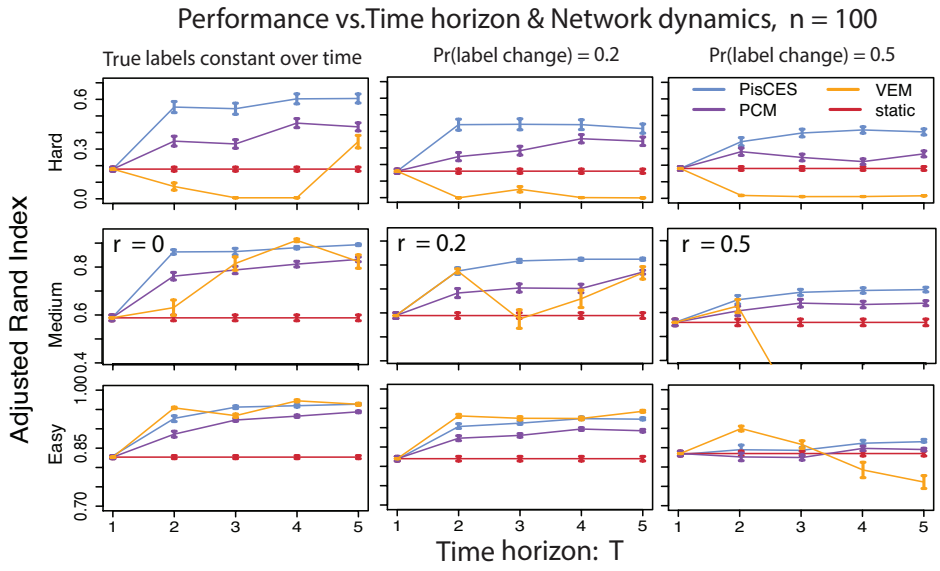

Fig. S4. Performance comparison on synthetic networks as a function of the time horizon and r, as measured by Adjusted Rand Index between true and estimated clusters. Networks are generated by dynamic DCBM with parameters corresponding to $n = 100$ in Table S1. 100 simulations per data point. Note that each row of subfigures uses different scales for the y-axis, as performance varies highly across the scenarios.

Sensitivity to α. Figure S6 shows performance of PisCES as a function of the smoothness parameter α. For each scenario, the value of α that gives the best average performance over all simulations is marked. For $r > 0$, this value is always ≤ 0.1, and is chosen by the cross-validation procedure of (24) in 100% of the simulated trials. For denser networks, performance peaked at smaller values of α; intuitively, each network has more signal so less smoothing is required. For large values of α, performance is poor and the method often does not converge, as suggested by Theorems 1.1 and S3.1; this can be seen in Figure S7.

S7. Identifying changing gene communities in developing primate brains

Data description. Bakken et al. (18) provide transcription for a dense temporal sampling of prenatal and postnatal periods measured on fine anatomical division of the rhesus monkey brain. The data contains 12,441 genes expressed in 1,936 distinct samples [‖].

To gain a high-level understanding of the difference in gene expression across ages and regions, we disply the similarity between samples using PCA. Region appears to be the dominant variable in the 724 postnatal samples projected onto the first two principle components (Fig. S8), which explain 41% of the variance. Focusing on the 214 postnatal mPFC samples, layer and age show high correlation with the first two PCs, which explain 40% of the variance (Fig. 4a). Likewise for 209 prenatal mPFC samples, layer and age are highly correlated with the first two PCs, which explain 46% of the variance (Fig. 4b).

Results. Table S3 summarizes several important statistics. Nodal degree is calculated for every adjacency matrix as degree of gene i, $d_i = \sum_j A_{ij}$. E40 and L4 have the highest median nodal values which might be due to the smaller number of samples available for these two periods. The number of communities in each age/layer is fairly similar over periods with the exception

‖ Available at https://github.com/AllenBrainAtlas/DevRhesusLMD/blob/master/cache/nhp_PrePost_StartingData.RData.

of the first developmental period in each series (E40/L6). Communities are defined as big when their size > 100 and dense when density > 0.1 for prenatal and > 0.06 for postnatal periods. At this initial phase the gene communities are larger but less dense, suggesting they are less developed as a network. For the prenatal phase, the number of big and dense communities generally increases over time. This pattern is not apparent in layers (L6 to L2). Finally, the high degree nodes are more likely to be included in paths.

Figure S11 displays the uncorrected enrichment p-values for each prenatal community with respect to SFARI ASD genes. Most communities are not nominally enriched, but a pattern of stronger nominal significance suggests ASD genes are most clustered in the mid to late fetal period.

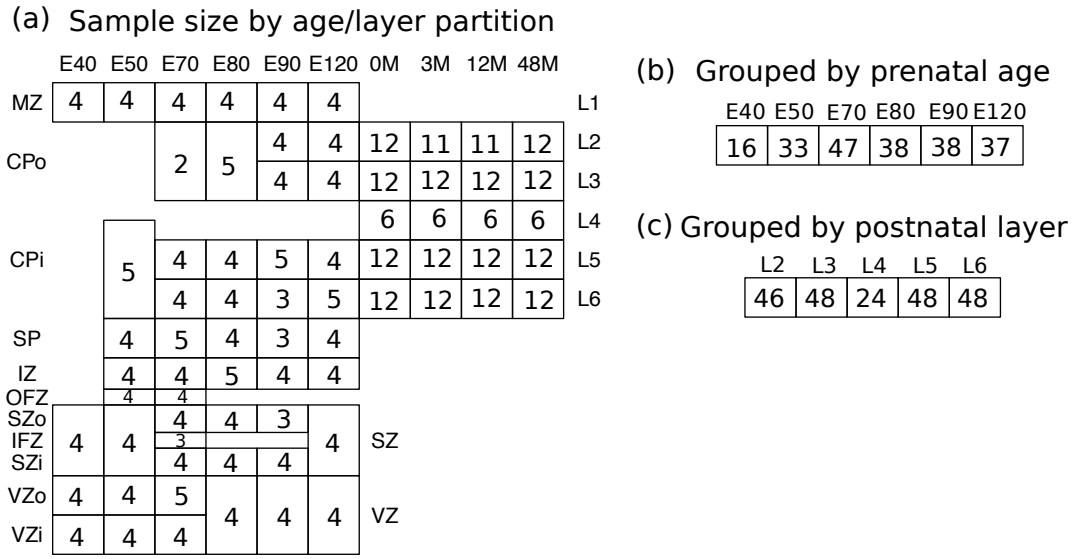

Table S2. Sample sizes under fine (a) and relatively coarse (b and c) partitioning schemes. In (a) the vertical axes define layers.

	nodal degree	number of communities	size of communities	density of communities	number of big and dense communities	number of big and dense communities in paths	number of high degree nodes in paths	number of non-high degree nodes in paths
E120	1st: 19 median: 60 3rd: 234	33	min: 49 median: 202 max: 1621	1st: 0.03 median: 0.05 3rd: 0.12	13	5(42%)	266 (29%)	1016 (12%)
E90	1st: 19 median: 50 3rd: 126	38	min: 9 median: 179 max: 1843	1st: 0.04 median: 0.06 3rd: 0.11	16	6 (38%)	702 (76%)	1064 (13%)
E80	1st: 20 median: 61 3rd: 143	31	min: 21 median: 257 max: 2080	1st: 0.03 median: 0.06 3rd: 0.09	10	5 (50%)	811 (88%)	1014 (12%)
E70	1st: 14 median: 40 3rd: 142	28	min: 27 median: 223 max: 2333	1st: 0.02 median: 0.04 3rd: 0.08	7	4 (57%)	845 (92%)	1200 (15%)
E50	1st: 19 median: 57 3rd: 225	26	min: 88 median: 229 max: 1944	1st: 0.03 median: 0.05 3rd: 0.10	9	3 (33%)	635 (69%)	975 (12%)
E40	1st: median: 156 3rd: 295	15	min: 144 median: 530 max: 1724	1st: 0.04 median: 0.07 3rd: 0.14	9	1 (11%)	98 (11%)	323 (0.04%)
total	1st: 22 median: 66 3rd: 195	-	-	-	78	35 (45%)	3183 (58%)	4619 (9%)
L2	1st: 20 median: 50 3rd: 122	35	min: 22 median: 188 max: 1683	1st: 0.03 median: 0.05 3rd: 0.07	10	4 (40%)	537 (58%)	1131 (14%)
L3	1st: 15 median: 39 3rd: 106	36	min: 34 median: 150 max: 1778	1st: 0.02 median: 0.03 3rd: 0.06	9	5 (56%)	734 (80%)	1240 (15%)
L4	1st: 36 median: 89 3rd: 222	22	min: 34 median: 227 max: 1535	1st: 0.03 median: 0.06 3rd: 0.11	14	8 (57%)	774 (84%)	2001 (24%)
L5	1st: 17 median: 40 3rd: 97	34	min: 13 median: 190 max: 1719	1st: 0.02 median: 0.04 3rd: 0.07	9	7 (78%)	686 (75%)	1643 (20%)
L6	1st: 19 median: 44 3rd: 96	25	min: 50 median: 261 max: 2091	1st: 0.02 median: 0.04 3rd: 0.07	7	4 (57%)	260 (28%)	969 (12%)
total	1st: 20 median: 50 3rd: 123	-	-	-	82	46 (56%)	3030 (66%)	5737 (14%)

Table S3. Summary of PisCES communities obtained from analysis of prenatal and postnatal mPFC samples. The median value and 1st and 3rd quantiles of nodal degree, size of communities and density of communities, as well as the number of communities for each network are shown in the first 4 columns. Big and dense communities are those with $\geq$ 100 nodes and within-community density exceeding 0.1 (prenatal) or 0.06 (postnatal). High degree nodes are those whose degree exceeds 90'th quantile. A path is a sequence of at least 3 big and dense communities that are connected by large flows (see Fig. 5, Fig S10), where a large flow comprises at least 50% of its smaller community.

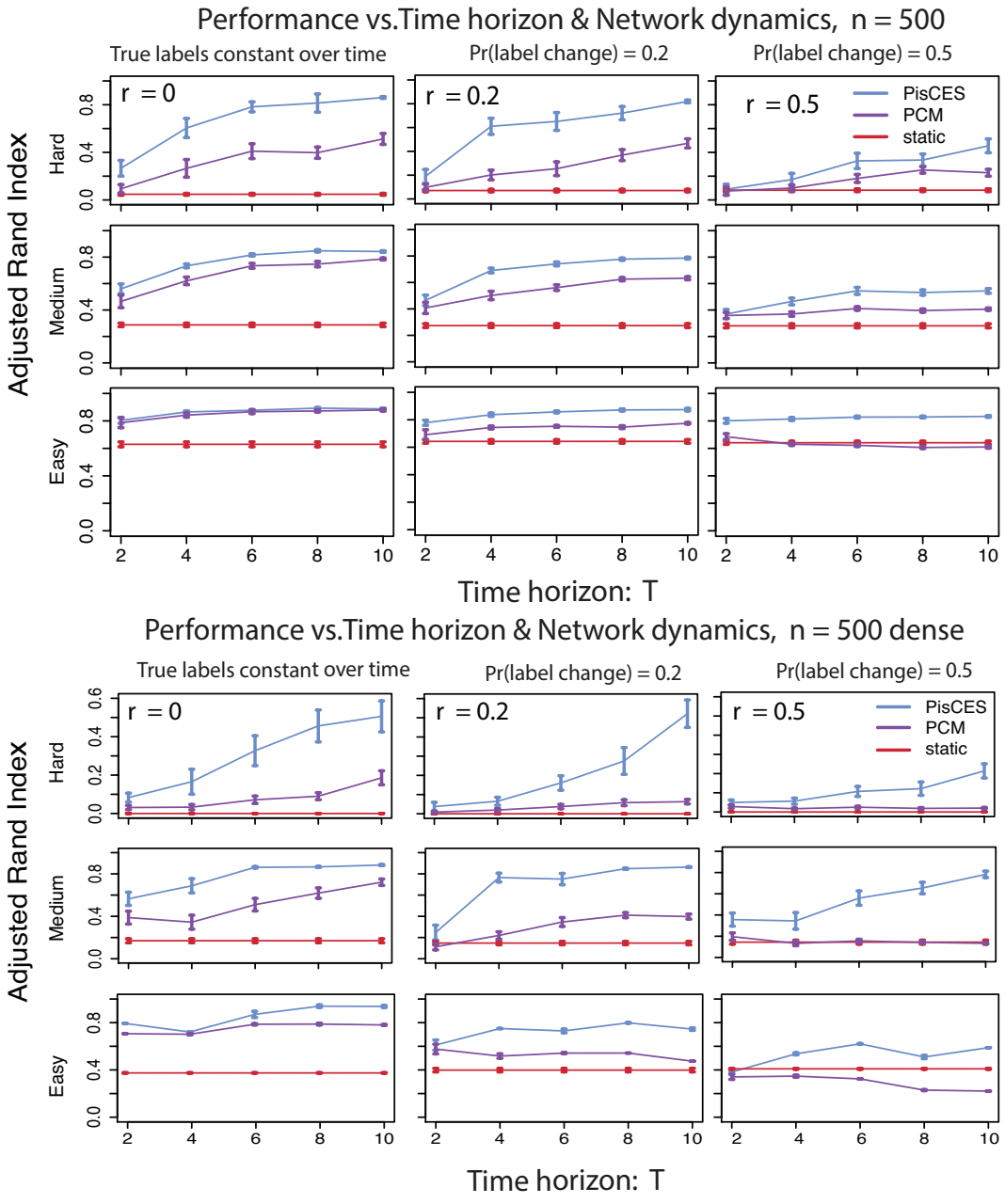

Fig. S5. Performance comparison on synthetic networks as function of time horizon and r, as measured by Adjusted Rand Index between true and estimated clusters. Networks are generated by dynamic DCBM with parameters corresponding to $n = 500$ and $n = 500$(dense) in Table S1. 100 simulations per data point. Note that each row of subfigures uses different scales for the y-axis, as performance varies highly across the scenarios.

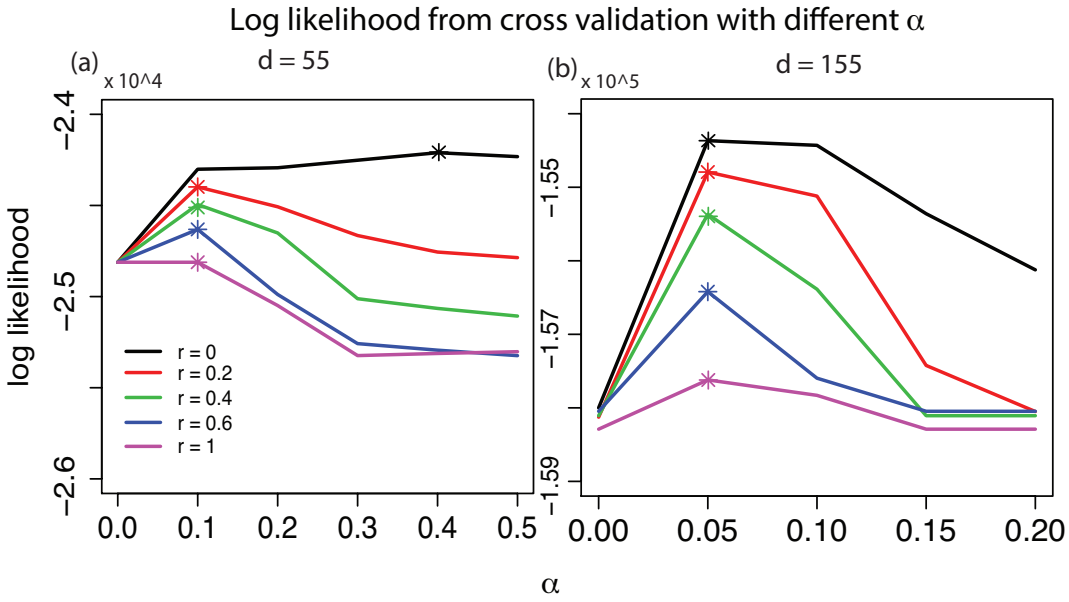

Fig. S6. Performance of PisCES as a function of α. Performance is measured by cross-validated log likelihood. Networks are generated by dynamic DCBM with $T = 10$, $K = 10$, $n = 500$, (a) $p_{\text{out}} = 0.1$, $p_{\text{in}} = (0.2, 0.2)$ and (b) $p_{\text{out}} = 0.3$, $p_{\text{in}} = (0.6, 0.6)$. The highest point of each curve is marked by a "*". 100 simulations per data point.

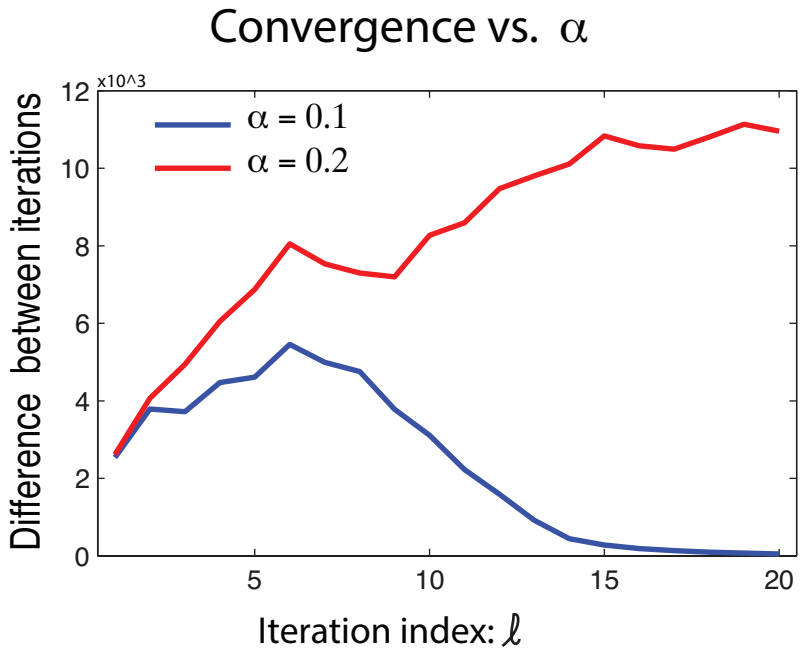

Fig. S7. Convergence status of PisCES with different α. Difference of eigenvectors between consecutive iterations is calculated as $||U^{\ell+1} - U^{\ell}||_1$, where U^{ℓ} is the output of PisCES in the ℓth iteration. The results shows one representative case under dynamic DCBM with $n = 500$, $K = 10$, $r = 0.1$, $p_{\text{in}} = (0.3, 0.3)$, $p_{\text{out}} = 0.1$ and $\alpha = 0.1$ or 0.2.

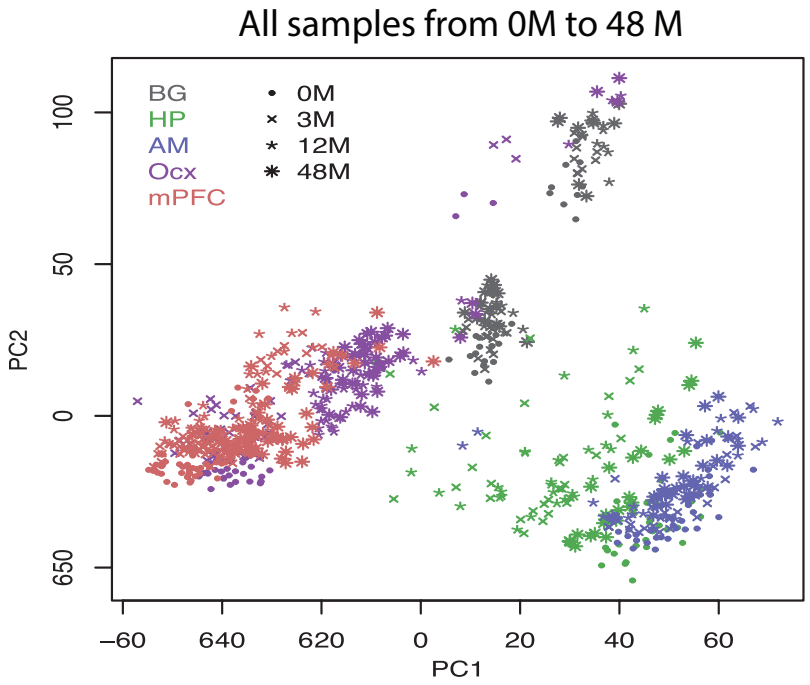

Fig. S8. Top two principal components for all samples from Age 0M to 48M. Age and region of each sample are depicted by shape and color, respectively (BG: basalganglia; HP: hippocampus; AM: amygdala; Ocx: occipital cortex; mPFC: medialprefrontal cortex).

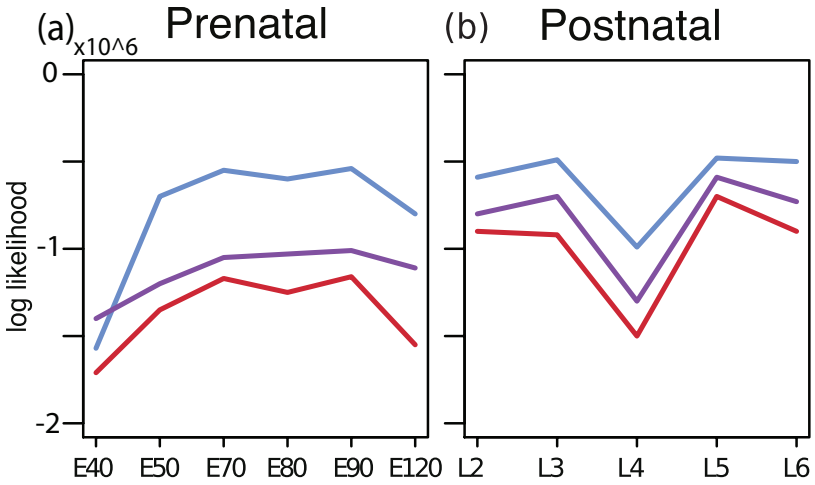

Fig. S9. Performance comparison for PisCES, PCM and static analysis on rhesus monkey brain data measured by log likelihood. The log likelihood values are calculated with the test data in 5-fold cross validation based on DCBM (Eq. Eq. (S14)) separately for each network in prenatal (a) and postnatal (b). The dynamic progression is informative, as illustrated by the inferior performance of the static analysis, compared to the two dynamic analyses. Performance at E40 is illuminating. The fetal brain is developing very rapidly at this stage and PisCES gains little from smoothing over age, but it still performs better than the unsmoothed, static method.

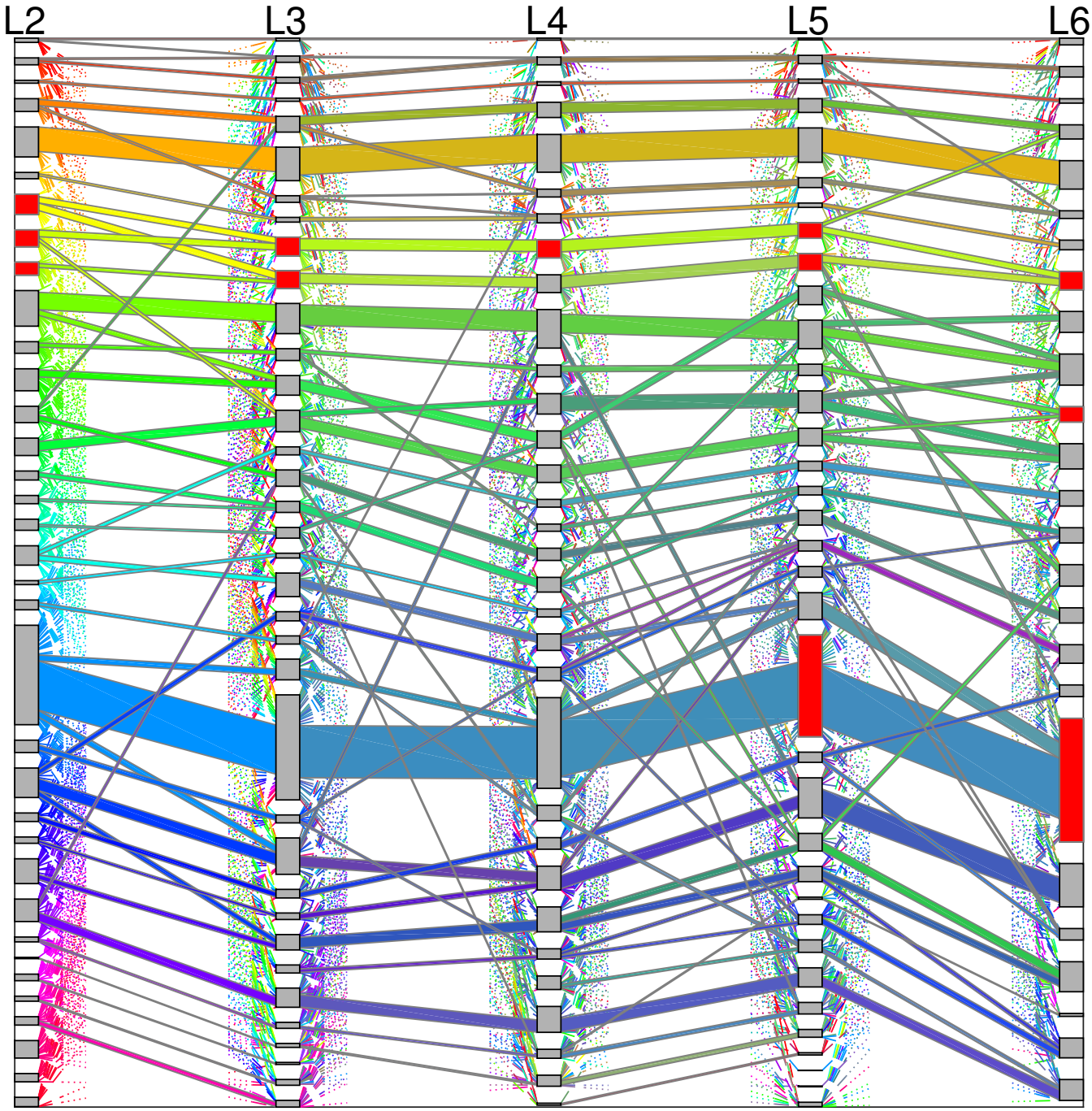

Fig. S10. Sankey plot for postnatal cases. Grey and red boxes denote communities, with height indicating community size. Colored "flows" denote groups of genes moving between communities, with height indicating flow size. To reduce clutter, only large flows (> 100 genes or $> 15\%$ of its source and destination community) are shown; smaller flows are partially drawn using dotted lines. Each flow's color is determined by its gene membership, and equals the mixture of the colors of its input flows. NPG-enriched communities are denoted by red boxes.

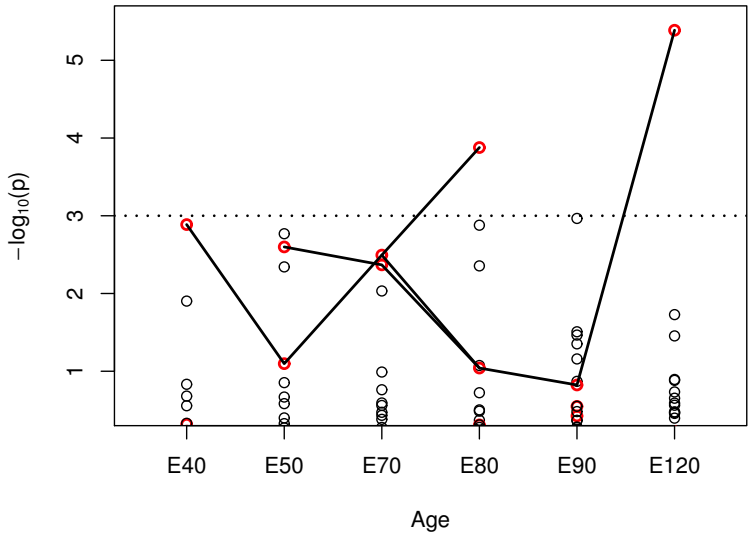

Fig. S11. Enrichment of SFARI genes in communities detected by PisCES. Each circle represents the $-\log_{10}(p)$ of SFARI enrichment of a community. The p-values are provided by single-sided Fisher's exact test and are not corrected for multiple testing. The most highly enriched cluster, at E120, includes 26 ASD genes out of 444 genes, including 4 NPG genes (CNTN6, CNTN4, CACNB2, KCNQ3), 9 synaptic transmission genes (GRIN2B, GABRB3, RIMS1, PRKCB, SLC12A5, CACNA1C, PLCB1, CACNB2, KCNQ3) and 2 synapse assembly genes (MYT1L, SHANK2). Nodes colored red are enriched for NPG genes. To help identify communities, connecting lines are drawn representing "flows" between communities as drawn in Fig. 5.

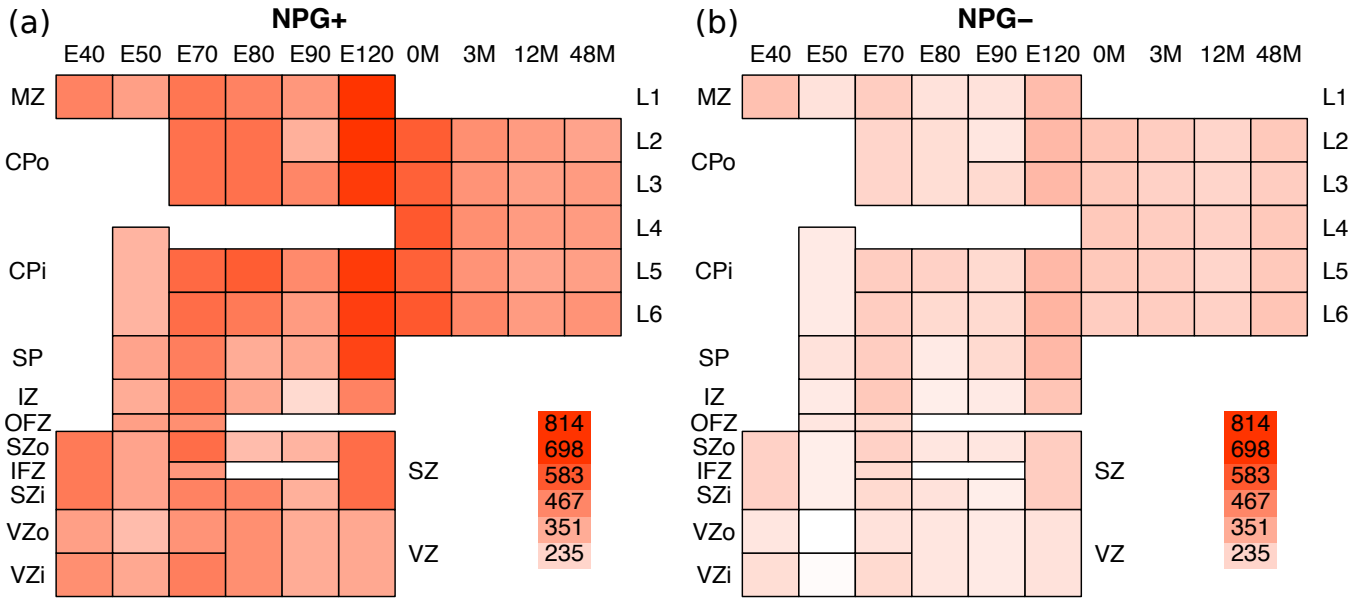

Fig. S12. Heat maps of mean expression levels across spatio-temporal partitions (18) of selective neuron projection guidance (NPG) genes in medial pre-frontal cortex (mPFC) samples partitioned by age and layer. The vertical axes define layers. The intensity of each grid represents the mean (to the power of 2) of either NPG+ (a) or NPG- (b) genes in its corresponding partition. The NPG+ genes are defined as those that appear in NPG-enriched communities for at least three consecutive periods. The remaining NPG genes define the NPG- group.

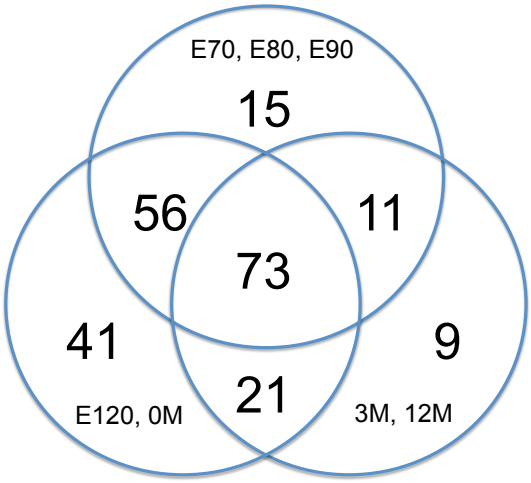

Fig. S13. Number of NPG genes that have at least one neighbor in the correlation networks (Fig. 7).

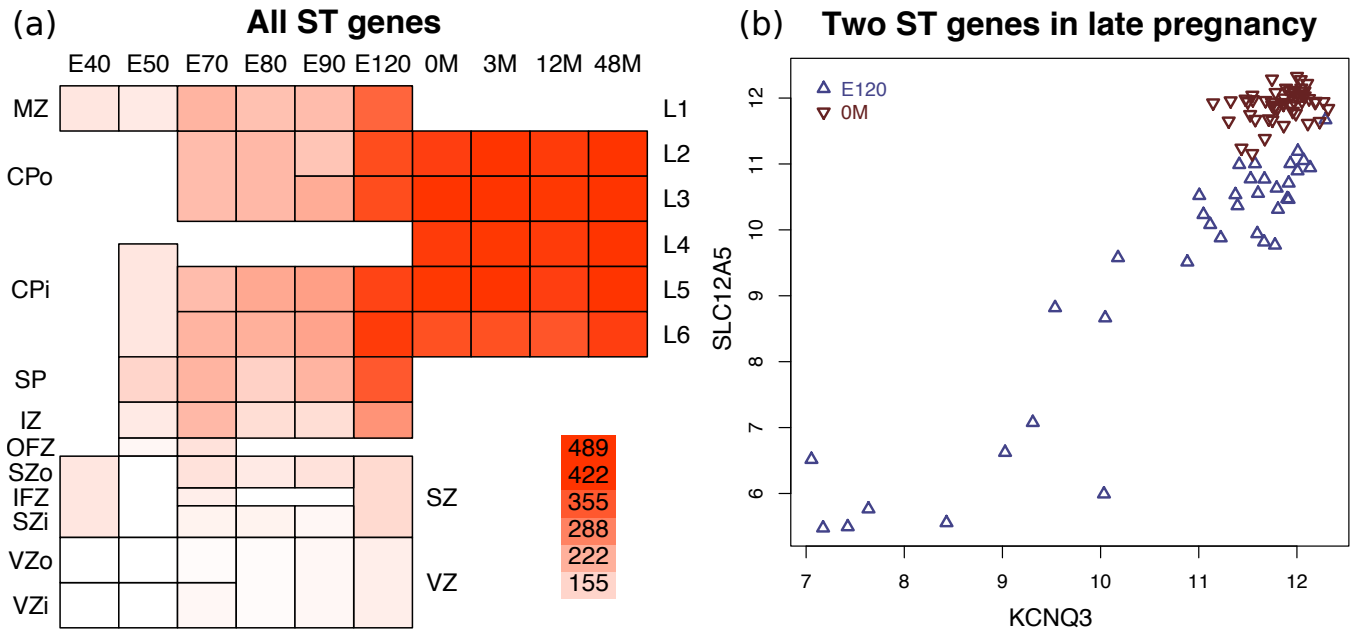

Fig. S14. (a), Heat map of mean expression levels of synaptic transmission (ST) genes in medial pre-frontal cortex (mPFC) samples partitioned by age and layer. In (a) the vertical axes define layers. The intensity of each grid represents the mean (to the power of 2) of ST genes in its corresponding parition. (b), Correlation between the expression levels of two synaptic transmission genes, KCNQ3 and SLC12A5, in mPFC samples at the ages of E120 and 0M. The axes are numbered by the expression levels of the two genes, respectively. Each blue triangle represents a sample at the age of E120 and each red triangle represents a sample at the age of 0M. The Pearson correlation of the blue triangles is 0.93 and that of the red triangles is 0.34.

		L6	L5	L4	L3	L2	
E40	1	1	0.25	0.19	0.21	0.15	L6
E50	0.11	1	1	0.32	0.28	0.17	L5
E70	0.07	0.21	1	1	0.27	0.16	L4
E80	0.07	0.13	0.27	1	1	0.21	L3
E90	0.05	0.10	0.18	0.27	1	1	L2
E120	0.05	0.09	0.14	0.19	0.23	1	
	E40	E50	E70	E80	E90	E120	

Table S4. Similarity of community assignments between age slices of prenatal samples (blue grids) and layer slices of postnatal samples (pink grids) measured by adjusted Rand index.