Supporting Information

Barthélemy et al. 10.1073/pnas.0910259107

SI Text

SI Model and Method

1 Determining the A_is by Using Eigenvectors of the Correlation Matrix. The data correlation matrix C_{ij} is known to provide useful information, in particular for the analysis of financial time series (1–3) or in other fields; e.g., in protein structure analysis (4). The first, largest, eigenvalue is related to a global trend, and usually one is interested in the small number of intermediate eigenvalues: The associated eigenvectors give the relevant correlations in the data—e.g., allows to extract the sectors in financial time series. Here, making explicit use of our hypotheses, we extract from the first eigenvector of the correlation matrix the A_i factors that show how the global trend is amplified or reduced at the local level.

We have

$$C_{ij} = A_i A_j + D_{ij} \qquad \textbf{[S1]}$$

where $D_{ij} = \langle G_i G_j \rangle$. If ψ is a normalized eigenvector ($\psi \cdot \psi = 1$) of C with eigenvalue λ: $C \cdot \psi = \lambda \psi$, we have

$$C \cdot \psi = (A \cdot \psi) A + D \cdot \psi. \qquad \textbf{[S2]}$$

We can have $A \cdot \psi$ equal to zero, which implies that ψ is also eigenvector for D, which in general is unlikely (there are no reasons that eigenvectors of D are orthogonal to A). If $A \cdot \psi \neq 0$ we then obtain

$$\lambda = A \cdot A + \frac{A \cdot D \cdot \psi}{A \cdot \psi} \qquad \textbf{[S3]}$$

and

$$\psi = \frac{A \cdot \psi}{\lambda} A + \frac{D \cdot \psi}{\lambda}. \qquad \textbf{[S4]}$$

For the largest eigenvalue, we will neglect at first order the second term of the right-hand side of this last equation, which leads to $\psi \propto A$. Because ψ is normalized, we obtain

$$\psi \approx \frac{A}{\sqrt{A \cdot A}}. \qquad \textbf{[S5]}$$

This approximation is justified if $A \cdot D \cdot \psi$ is small compared to $A \cdot A$ and thus

$$\frac{A \cdot D \cdot A}{(A^2)^2} \ll 1. \qquad \textbf{[S6]}$$

Because $A \cdot A = \mathcal{O}(N)$, this approximation is justified if $A \cdot D \cdot A$ is of order N and not of order N^2. This is correct if D is diagonal (which means that the external components are not correlated $\langle G_i G_j \rangle \propto \delta_{ij}$), but also if the number of nonzero terms of D_{ij} is finite compared to N, or in other words if D is a sparse matrix.

We compared the values of A_i computed with the method exposed in the text and with the eigenvector method. Results are reported in the Figs. S1, S2, and S3.

We see that indeed for the crime rates in the United States and in France, D_{ij} is indeed negligible, which demonstrate that the correlations of the internal contributions between different states in the United States are negligible. This is not the case for the stocks in the Standard and Poor's 500 Index (S&P 500) where we can observe (small) discrepancies between the two methods, a result which supports the idea of sectors in the S&P 500.

2 Scaling. We show that the scaling $\sigma_i^{\text{ext}} \sim \langle f_i \rangle$ observed by de Menezes and Barabasi in (5, 6) is actually built in the method proposed by these authors: it is a direct consequence of their definitions of the internal and external parts, and it does not depend on the data structure.

Indeed, let $f_i(t), t = 1, \ldots, T, i = 1, \ldots, N$ be an arbitrary dataset such that $\langle \bar{f} \rangle \neq 0$. For $i = 1, \ldots, N$, following (5) define A_i^{MB} by

$$A_i^{\text{MB}} \equiv \frac{\langle f_i \rangle}{\langle \bar{f} \rangle} \qquad \textbf{[S7]}$$

and $f_i^{\text{MB,ext}}(t)$ by

$$f_i^{\text{MB,ext}}(t) \equiv A_i^{\text{MB}} \bar{f}(t). \qquad \textbf{[S8]}$$

Then, from these definitions and without any hypothesis or constraint on the data other than $\langle \bar{f} \rangle \neq 0$, one has

$$\langle f_i^{\text{MB,ext}} \rangle = A_i^{\text{MB}} \langle \bar{f}(t) \rangle = \langle f_i \rangle \qquad \textbf{[S9]}$$

and

$$\langle (f_i^{\text{MB,ext}})^2 \rangle = (A_i^{\text{MB}})^2 \langle \bar{f}(t)^2 \rangle. \qquad \textbf{[S10]}$$

Hence

$$(\sigma_i^{\text{MB,ext}})^2 = (A_i^{\text{MB}})^2 \sigma_f^2 = \langle f_i \rangle^2 \frac{\sigma_f^2}{\langle \bar{f} \rangle^2} \qquad \textbf{[S11]}$$

with

$$\sigma_f^2 \equiv \langle \bar{f}(t)^2 \rangle - \langle \bar{f}(t) \rangle^2. \qquad \textbf{[S12]}$$

Hence, one has always

$$\sigma_i^{\text{MB,ext}} = \frac{\sigma_f}{|\langle \bar{f} \rangle|} |\langle f_i \rangle|. \qquad \textbf{[S13]}$$

The dispersion of the external component, if defined from **[S7]** and **[S8]**, is thus exactly proportional to the mean value of the local data.

3 Synthetic Series: Correlated Random Walkers. We considered the case where the external trend is

$$F(t) = \sin(\omega t). \qquad \textbf{[S14]}$$

The Gaussian noises are given by

$$\xi_i(t) = \alpha \sum_{j=1}^{M} u_j^{(0)}(t) + \sum_{j=M+1}^{N} u_j^{(i)}(t) \qquad \textbf{[S15]}$$

where the $u_j^{(0)}(t)$ and $u_j^{(i)}(t)$ are independent, uniform random variable of zero mean and variance equal to 1/12. In this case, the correlation between different noises are governed by the parameters α and M

$$\overline{\xi_i \xi_j} = \frac{\alpha^2 M}{12} + \frac{N - M}{12} \delta_{ij}. \qquad \textbf{[S16]}$$

When $M = 0$, the variables ξ_i and ξ_j are independent (for $i \neq j$) and we can monitor the correlations by increasing the value of M. We plot $N = 100$ random walkers in the usual uncorrelated case in Fig. S4 and in presence of correlations in Fig. S5.

In this simple case the exact result is given by $w(t) = F(t)$, $a_i = 1$, and $g_i(t) = \xi_i(t)$. The important condition for the validity of the method is given by $A_i A_j \gg \langle G_i G_j \rangle$ and is given here by

$$1 \gg \alpha^2 M. \qquad \textbf{[S17]}$$

For $M = 0$, the random noises are independent and our method is very accurate as shown in the main text.

More generally, to assess quantitatively the efficiency of the method, we compute the Pearson correlation coefficient between the exact $f_i^{\text{int}}(t)$ and the estimate g_i computed with the method. We plot, in Fig. S6, this coefficient versus $\alpha^2 M$. This figure confirms the fact that our method is valid and very precise provided that the correlations between local contributions are not too large (here $\alpha^2 M < 4$).

4 Dependence of the a_i on the Time Interval. We can compute the quantities a_i for the interval $[t_0, t]$ and by letting t vary. We then obtain for the crime in the United States (in the case of the crime rates in France, the dataset is not large enough) the Fig. S7. This figure shows that in the case of the crime rate in the United States, the a_i converge to a stationary value, independent of the time interval, provided it is large enough. Our method will then lead to reliable results constant in time.

We also tested our method on the financial time series given by the 500 most important stocks in the United States economy (Fig. S7), and which composition leads to the S&P 500 index. Here the "local" units are the individual stocks ($i = 1, \ldots, N = 500$), and the (naive) average—analogue to a national average—is precisely the S&P 500 index time series. We study the time series for these stocks on the 252 days of the period October 2007 through October 2008 and we compute the global pattern $w(t)$, the coefficients a_i, and the parameters η_i (defined in the text) computed for the time window $[10/2007, t]$ for t varying from April 2008 to September 2009. These quantities η_i measure quantitatively the importance of local versus external fluctuations for the stock i. The results for the η_is are shown in Fig. S8 and display large variations, particularly when we approach October 2008, a period of financial crisis. It is therefore not completely surprising that the η_i (and the a_is) in this case fluctuate a lot. In some sense, we can conclude that the a_is correspond to an average susceptibility to the global trend, are not invariable quantities and can vary for different periods. We thus see on this example, that it is important to check the stability of the coefficients a_i which is an crucial assumption in our method. The variations of these coefficients is however interesting and further studies are needed to understand these variations.

5 Obesity in the United States: Variances for the External and Internal Contribution. For the obesity rate series, we compare the variances of the internal (g_i) and the external ($a_i w$) contributions. We observe in Fig. S9 that the variance of the external contribution became dominant after the year ≈2000.

1. Mantegna R, Stanley HE (2000) *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge Univ Press, Cambridge, UK).
2. Bouchaud J-P, Potters M (2003) *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management* (Cambridge Univ Press, Cambridge, UK).
3. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Stanley HE (1999) Universal and nonuniversal properties of cross correlations in financial time series. *Phys Rev Lett* 83:1471–1474.
4. Halabi N, Rivoire O, Leibler S, Ranganathan R (2008) Protein sectors: Evolutionary units of three-dimensional structure. *Cell* 138(4):774–786.
5. de Menezes MA, Barabasi A-L (2004) Fluctuations in network dynamics. *Phys Rev Lett* 92:028701.
6. de Menezes MA, Barabasi A-L (2004) Separating internal and external dynamics of complex systems. *Phys Rev Lett* 93:068701.
7. Standard and Poor's. Historical Data for S&P 500 stocks. Available at http://biz.swcp.com/stocks/.

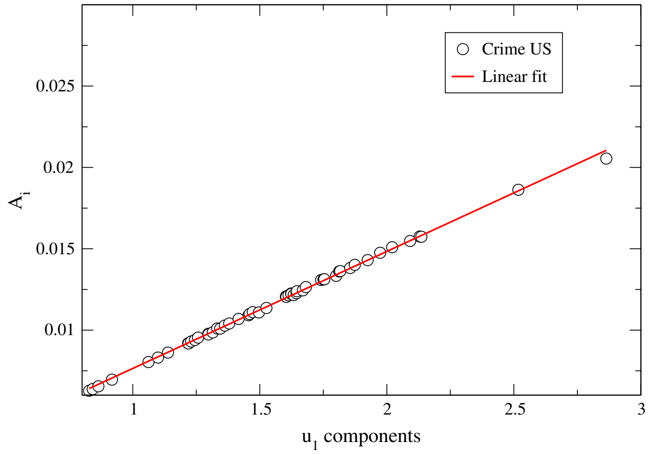

Fig. S1. Comparison of the A_i computed with expressions in the text (Eq. **16**) and with the components of the eigenvector corresponding to the large eigenvalue of C_{ij} in the case of crime rates in the United States.

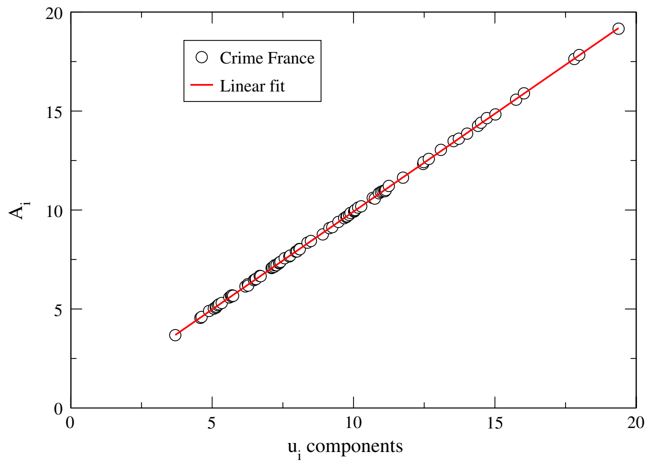

Fig. S2. Comparison of the A_i computed with expressions in the text (Eq. **16**) and with the components of the eigenvector corresponding to the large eigenvalue of C_{ij} in the case of crime rates in France.

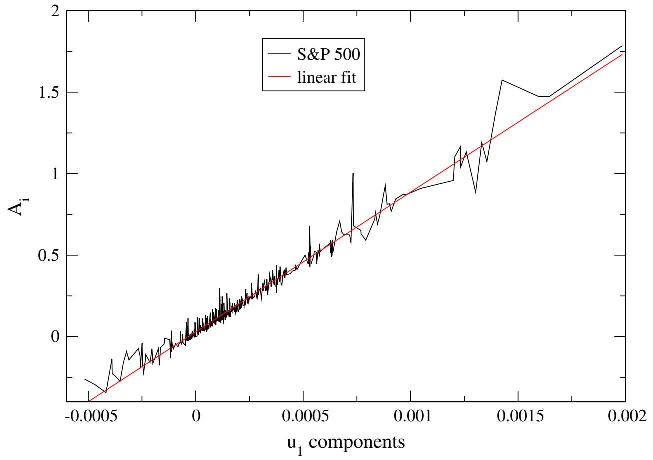

Fig. S3. Comparison of the A_i computed with expressions in the text (Eq. **S16**) and with the components of the eigenvector corresponding to the large eigenvalue of C_{ij} in the case of the S&P 500.

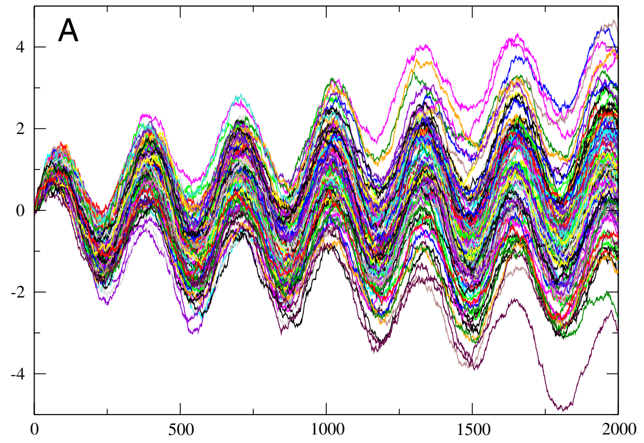

Fig. S4. *N* Uncorrelated random walkers ($N = 100$, $\alpha^2 M = 0$).

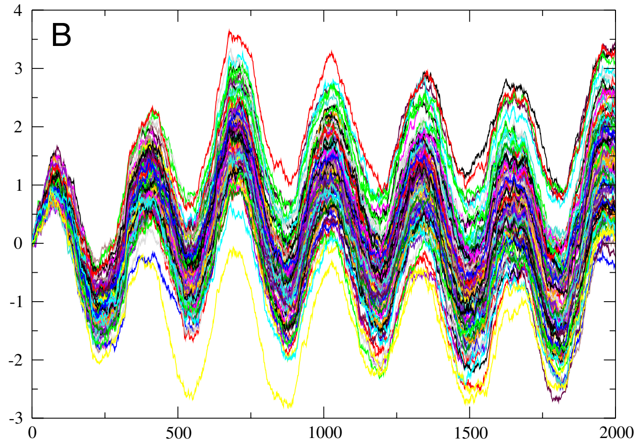

Fig. S5. Random walkers with correlations ($\alpha^2 M = 10$).

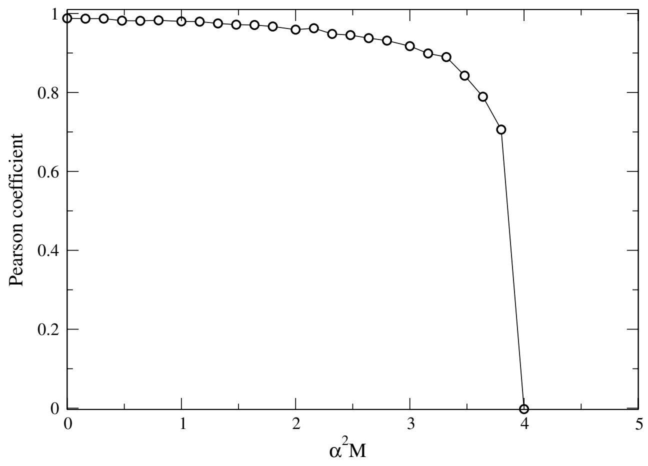

Fig. S6. Pearson correlation coefficient between the exact local contribution and the local contribution computed with our method computed for different values of the correlation ($N = 100$, results averaged over 100 realizations).

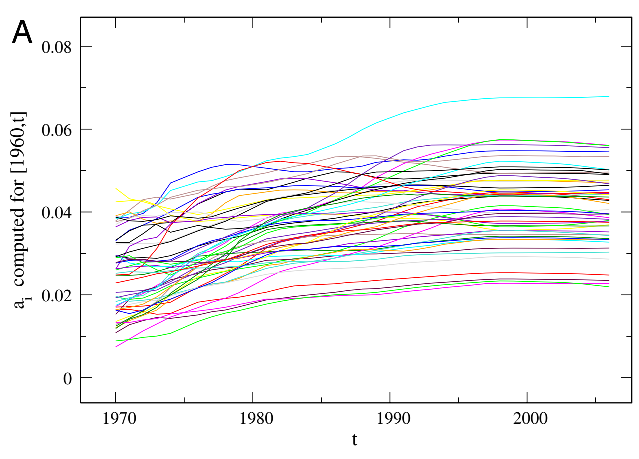

Fig. S7. Coefficients a_i computed in the case of US crime for the interval $[1960, t]$ with varying t (in years).

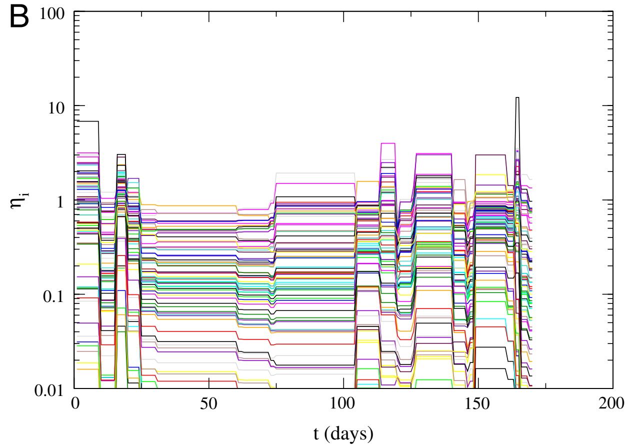

Fig. S8. Coefficients η_i computed for the S&P 500 in the interval $[0, 125 + t]$ (t is in days in this case).

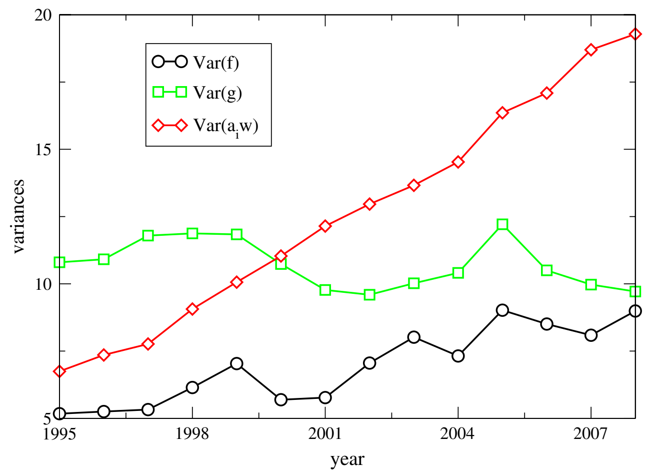

Fig. S9. Comparison of internal and external fluctuations for the obesity in the United States. We represent the total variance of the signal (f), the external ($a_i w$), and the internal contribution (g). We observe that for the external contribution is dominating since the year 2000.