Large Scale Cross-Correlations in Internet Traffic

Marc Barthélemy[1], Bernard Gondran[2], and Eric Guichard[3]

[1] *CEA, Service de Physique de la Matière Condensée, BP12 Bruyères-Le-Châtel, France*

[2] *Réseau National de Télécommunications pour*
la Technologie, l'Enseignement et la Recherche
151, Bld de L'Hôpital, 75013 Paris, France

[3] *Equipe Réseaux, Savoirs & Territoires*
Ecole normale supérieure, 75005 Paris, France

(December 4, 2002)

The Internet is a complex network of interconnected routers and the existence of collective behavior such as congestion suggests that the correlations between different connections play a crucial role. It is thus critical to measure and quantify these correlations. We use methods of random matrix theory (RMT) to analyze the cross-correlation matrix C of information flow changes of 650 connections between 26 routers of the French scientific network 'Renater'. We find that C has the universal properties of the Gaussian orthogonal ensemble of random matrices: The distribution of eigenvalues—up to a rescaling which exhibits a typical correlation time of the order 10 minutes—and the spacing distribution follows the predictions of RMT. There are some deviations for large eigenvalues which contain network-specific information and which identify genuine correlations between connections. The study of the most correlated connections reveal the existence of 'active centers' which are exchanging information with a large number of routers thereby inducing correlations between the corresponding connections. These strong correlations could be a reason for the observed self-similarity in the WWW traffic.

PACS numbers: 02.50 -r, 05.45.Tp, 84.40.Ua, 87.23.Ge

I. INTRODUCTION

Internet connects different routers and servers using different operating systems and transport protocols. This intrinsic heterogeneity of the network added to the unpredictability of human practices [1] make the Internet inherently unreliable and its traffic complex [2,3,4,5,6]. Recently, there has been major advances in our understanding of the generic aspects of the Internet [7,8,9,10] and web [11,12,13,14,15,16] structure and development, revealing that these networks can exhibit emergent collective behavior characterized by scaling. Concerning data transport, most of the studies focus on properties at short time scales (usually < 1 min) or at the level of individual connections [2,17,18]. In particular, it has been shown that for wide- and local-area networks the self-similarity (for time correlations) applies. Possible reasons for this behavior were shown to be [17] the underlying distribution of WWW documents, the effects of user 'think time', and the addition of many such transfers.

Studies on statistical flow properties at a large scale [3,4,6,19] concentrate essentially on the phase transition from a 'fluid' regime to a 'congested' one for which the average packet travel time is very large [20]. The existence of such a collective behavior indicates the importance of spatial correlations between connections at a large scale in the system. In order to be able to understand and to model the traffic in the network, it is thus important to measure and to quantify the correlations between the flows in different connections.

In this paper, we analyze the correlations between different connections of a wide area network which is the French scientific network 'Renater'. We use random matrix theory (RMT) to study the corresponding empirical correlation matrix. RMT has been developed in the fifties for studying complex energy levels of heavy nuclei [21] and more recently it has also been used in the study of correlations of stocks [22,23] or statistics of atmospheric correlations [25].

We first demonstrate the validity of the universal predictions of RMT for the eigenvalue statistics of the cross-correlation matrix. However, we observe some deviations compared to the minimal hypothesis of random independent time-series. These deviations from the universal predictions of RMT identify system-specific, non-random properties of the network providing clues about the nature of the underlying interactions. This result allows one to distinguish genuine correlations in the network which are not just due to noise.

II. EMPIRICAL RESULTS

A. Data studied

We use data from the French network 'Renater' [26] which has about 2 million users and which consists of about 30 interconnected routers (Fig. 1). Most Research

institutes, technological, or educational institutions are connected to Renater.

The data consist of the real exchange flow (sum of Ftp, Telnet, Mail, Web browsing, etc.) between all routers even if there is not a direct (physical) link between all of them. For a connection (i, j) between routers i and j $(i \neq j)$, $F_{ij}(t)$ (in bytes per 5 minutes) is the effective information flow at time t going out from i to j (the flow going from i to k via j is excluded from F_{ij}). For technical reasons, data for a few routers were not reliable and we analyzed data for 26 routers which amounts in 26×26 matrices $F_{ij}(t)$ given for every sampling time scale $\tau = 5$ minutes during a two weeks period. We also exclude from the present study the internal flow F_{ii}, and the nights for which the flow is essentially due to machine activity. We thus studied data for days (8am-6pm), which amounts to a total of $N = 26 \times 25 = 650$ different connections given for $L = 12 \times 10 \times 14 \text{days} = 1680$ time counts. We choose as a measure of the magnitude of the time-series fluctuations the growth rate defined as the logarithm of the ratio of successive counts

$$g_{ij}(t) = \log \left[\frac{F_{ij}(t+\tau)}{F_{ij}(t)} \right] \qquad (1)$$

for $t = 0, \cdots, (L-1)\tau$. This measure has several nice properties. First, any multiplicative, time-independent sample bias cancels in the ratio. Second, this measure has a natural interpretation in terms of relative growth since for a small increase $g_{ij}(t) \simeq [F_{ij}(t+\Delta t) - F_{ij}(t)]/F_{ij}(t)$ is simply the relative increment. A large value of this quantity reflects a large activity (i.e. a large flow variation), while a small value corresponds to an almost constant flow. This measure is thus independent from the volume of information exchanged and thus does not eliminate the 'small' routers. The study of volume flow exchange will be published elsewhere [32] and in the present paper the quantity g allows us to study more subtle effects such as the activity of a regional router, independently of its 'size' measured in terms of exchanged information volume.

B. Correlation matrix

The simplest measure of correlations between different connections (i, j) and (k, l) is the equal-time cross-correlation matrix C which has elements

$$C_{(ij)(kl)} = \frac{\langle g_{ij} g_{kl} \rangle - \langle g_{ij} \rangle \langle g_{kl} \rangle}{\sigma_{ij} \sigma_{kl}} \qquad (2)$$

where $\sigma_{ij} = \sqrt{\langle g_{ij}^2 \rangle - \langle g_{ij} \rangle^2}$ is the standard deviation of the flow growth rate of the connection (i, j) and $\langle \cdots \rangle$ denotes a time average over the period studied. The correlation matrix is real symmetric and its elements are

comprised between -1 (anti-correlated connections) and 1 (correlated connections), while a null value denotes statistical independence.

The quantities g_{ij}/σ_{ij} have (by construction) a variance equal to one and a zero mean (for a sufficiently long time). It is thus natural to compare our empirical results with a mutual independent time-series—the 'null' hypothesis—described by the correlation matrix

$$\mathsf{R} = \frac{1}{L} \mathsf{A} \mathsf{A}^t \qquad (3)$$

where A (the so-called random Wishart matrix) is an $N \times L$ matrix containing N times series of L random independent elements with zero mean and unit variance (A^t denotes the transpose of A). Each element of R can be written as $R_{(ij)(kl)} = \langle a_{ij} a_{kl} \rangle$ where $a_{ij}(t)$ is a time series of independent elements with zero mean ($\langle a_{ij} \rangle = 0$) and unit variance ($\sigma_{ij} = 1$).

1. Eigenvalues

The probability distribution of the elements of C shows that most on the elements are positive (Fig. 2) which indicates a strong correlation among the whole network. For comparison, the elements of R are distributed according to a centered distribution with zero mean. We now study the statistical properties of C by applying RMT techniques. We first diagonalize C and obtain its eigenvalues λ_k $(k = 1, \cdots, N)$ which we sort from the largest to the smallest. We then calculate the eigenvalue distribution and compare it with the analytical result for a cross-correlation matrix generated from finite uncorrelated time series [28] in the limit $N \to \infty$, $L \to \infty$ where $Q = L/N \geq 1$ is fixed

$$P_{rm}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \qquad (4)$$

with $\lambda \in [\lambda_-, \lambda_+]$ and where

$$\lambda_\pm(Q) = 1 + 1/Q \pm 2/\sqrt{Q} \qquad (5)$$

The eigenvalue distribution of C is very different from Equ. (4) which predicts a finite range of eigenvalues depending on the ratio Q. The theoretical value is $Q = 2.58$ and we can reasonably fit the empirical curve with an effective value $Q^* = 1.1$ (Fig. 3a). This effective value can be explained as resulting from time correlations in the traffic of the order of $\frac{Q}{Q^*} \times \tau \simeq 11$ minutes. However, even this fit cannot reproduce the large eigenvalues observed: For $Q^* = 1.1$ the theoretical eigenvalues are distributed in the interval $2.17 \times 10^{-3} \leq \lambda_k \leq 3.82$ while few—a total of order 20—measured eigenvalues (not all shown on the graph) are found above $\lambda_+(Q^*) = 3.82$. The largest eigenvalue is of order $\lambda_1 \simeq 200$ namely approximately

hundred times larger than the maximum eigenvalue predicted for uncorrelated time series. As we will see, the empirical distribution of eigenvector components for the large eigenvalues is 'flat', all components being of the same order. This suggests that the largest eigenvalues are associated with strong correlations among the network.

We also calculate the distribution of the nearest-neighbor spacings $s = \lambda_{k+1} - \lambda_k$. We compare the empirical distribution of nearest-neighbor spacings with the RMT predictions for real symmetric random matrices. This class of matrices shares universal properties with the ensemble of matrices whose elements are distributed according to a Gaussian probability measure—the Gaussian orthogonal ensemble (GOE). We find good agreement (Fig. 3b) between the empirical data and Wigner's surmise

$$P_{\mathrm{GOE}}(s) = \frac{\pi s}{2} \, \exp\left(-\frac{\pi}{4} s^2\right) . \tag{6}$$

which indicates a 'level repulsion' existing in our system and means that the eigenvalues are correlated.

2. Eigenvectors and Inverse Participation Ratio

We now analyze the eigenvectors of C. We denote by u_k the eigenvector associated to the eigenvalue λ_k and if we normalize the eigenvectors such that $u_k^2 = N$, it can be shown that in the Wishart case the components u of the eigenvectors are distributed according to the so-called Porter-Thomas distribution

$$P(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \tag{7}$$

In agreement with this result we find that eigenvectors corresponding to most eigenvalues in the 'bulk' of the spectrum (λ_k not too large) follow this prediction (Fig. 4a).

On the other hand, eigenvectors with eigenvalues outside the bulk ($\lambda_k \geq \lambda_+(Q^*)$) show marked deviations from the Gaussian distribution (Fig. 4b,c). In particular, the vector corresponding to the largest eigenvalue λ_1 deviates significantly from the Gaussian distribution predicted by RMT (Fig. 4b). This eigenvector is the signature of a collective behavior—the network itself—for which all connections are correlated. This effect was already observed in the framework of stock correlations, the largest eigenvalue being in this case the entire market [22,23,24].

The distribution of the components of an eigenvector contains information about the number of connections contributing to it. In order to distinguish between one eigenvector with approximately equal components and another with a small number of large components we use the inverse participation ratio (IPR) introduced in the context of localization theory [29,30]

$$I_k = \frac{1}{N^2} \sum_{i=1}^{N} [u_{ki}]^4 \, , \tag{8}$$

where $u_{ki}, \; i = 1, \ldots, N = 650$ are the components of eigenvector u_k. When the components of a vector are of the same order and distributed according to Equ. (7), the average IPR is small and equal to $3/N$ whereas a vector with only few non zero components leads to a IPR of order unity. The quantity $\Upsilon_k = 3/I_k$ is thus a measure of the number of vector components significantly different from zero. We compared Υ_k for our empirical results and for uncorrelated time series with the same values of (N, L) (Fig. 5). For the latter case, Υ_k has small fluctuations around $N = 650$ indicating that all the vectors are extended [30] which means that almost all connections contribute to them. On the other hand, the empirical data show deviations of Υ_k from N for the smallest and largest eigenvalues (except for the largest eigenvalue). In these cases, the number of contributing connections is much smaller than N ranging from a few connections to a few hundreds. These deviations of few orders of magnitude of I_k from its average suggests that the vectors are localized [30] and that only a few connections contribute to them. As it will be illustrated on a simple example in the next section, these results have a clear meaning in the case of large eigenvalues for which the connections are correlated. In addition, it was also shown ([24] and see below) that strongly correlated pairs of routers (which correspond to large components in the eigenvectors) also appear with a relative negative sign in the eigenvector for *small* eigenvalues. This explains why the lower band edge also displays localized vectors but there is no clear connection with the spectrum observed in localization in electronic systems [30].

In addition, our empirical results exhibit 'quasi-extended' states in the center of the band. These states consist essentially of a group of $\simeq 300 - 400$ connections corresponding to eigenvalues of order $0.2 - 0.4$.

The physical picture which emerges is thus the following. The largest eigenvalue has an eigenvector which $\Upsilon_{k=1}$ is of order N and thus represents the whole network. The eigenvectors which correspond to eigenvalues which deviate from pure random matrix theory correspond to genuine correlations in the network. We have shown that these 'deviating' eigenvectors (of the order of 20) have a small value of Υ_k which means that these important correlations are localized and that a relatively small number of connections concentrate most of the activity [31].

3. Non-Universal Properties: Active Centers

The detail of the components of the 'deviating' eigenvectors give us information about the important correlations in the network. In particular, the largest components of the eigenvectors correspond to the most correlated connections. This can be seen on the simple following example of a 3×3 correlation matrix

$$\begin{pmatrix} 1 & c & 0 \\ c & 1 & c' \\ 0 & c' & 1 \end{pmatrix} \qquad (9)$$

where c (resp. c') denotes the strength of the $(1,2)$ (resp. $(2,3)$) correlation. If we denote the ratio of the correlation strengths $\eta = c'/c$, the eigenvectors u_1, u_2, and u_3 are respectively

$$\begin{pmatrix} 1 \\ \sqrt{1+\eta^2} \\ \eta \end{pmatrix}, \begin{pmatrix} -\eta \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -\sqrt{1+\eta^2} \\ \eta \end{pmatrix}, \qquad (10)$$

and correspond respectively to the eigenvalues (sorted in decreasing order)

$$1 + c\sqrt{1+\eta^2}, 1, 1 - c\sqrt{1+\eta^2} \qquad (11)$$

We thus see on this simple example that the components of the eigenvector u_1 (corresponding to the largest eigenvalue) identify the most correlated indices: For $\eta \ll 1$, $u_1 \simeq (1,1,0)$ and for $\eta \gg 1$ one obtains $u_1 \propto (0,1,1)$.

This remark shows that the eigenvectors are indeed important for identifying the most correlated connections in the network. We note that the large correlations are also reflected in the components—but with a relative minus sign—of the eigenvectors for small eigenvalues.

In the case of Renater, we have seen in the previous section that all the components of u_1 are positive which indicates a correlation among the whole network. Even if all the components of u_1 indicate correlations existing in the network, the simple example above shows that its largest components correspond to the most correlated connections. We thus looked at the largest components of u_1. A first fact is that a connection (i,j) is always (strongly) correlated with the connection (j,i). This result is not surprising since for most operations (Web browsing, Telnet, etc), there is always a 'outgoing' flow which is a significant part of the 'incoming' flow.

In order to look for other causes of correlations we plot on Fig. (6) the histogram of occurrences $h(i)$ of the router i in the set of the n most correlated connections (i,j) which are given by the first n components of the eigenvector u_1 corresponding to the largest eigenvalue. We compared the empirical results with the control case for increasing values of n (for n approaching the total number of components $N = 650$ all the connections appear and the histogram of occurrences is flat). We observe

marked differences between these two cases. In particular, in the control case the histogram tends to be uniform while for Renater we observe persistent peaks. On the last plot (Fig. 6c), it is apparent that there are still some fluctuations in the control case but much less than in the empirical one. The persistency of peaks and the fact that they appear to be much larger than the average value suggest that it is very unlikely that they are just fluctuations due to noise. Therefore, not all routers appear in the most correlated connections and the peaks can thus be identified as important 'active centers'. These centers are exchanging information with many other routers thereby inducing correlations between these connections.

It is interesting to note that occurrence peaks also appear in the components of the other deviating eigenvectors and would thus also correspond to active centers but at a lower level of correlation.

At this stage, we would like to emphasize that this analysis highlights active center independently of the volume of information exchanged. Indeed, in a volume flow analysis the 'small' routers even very active are completely hidden by the 'big' routers which are receiving and emitting huge amounts of bytes.

III. CORRELATIONS AND SELF-SIMILARITY IN THE WWW

The Internet is an example of a complex network which shows existence of a collective behavior such as a phase transition to a congested regime [3]. An important discovery was also the power-law decay of time correlations [2]. This self-similarity is usually explained on the basis of underlying distributions of WWW document sizes, effect of user 'think time' and the addition of many such effects in a network [17].

The present study shows that strong correlations between different connections exist in the traffic network. This result together with the existence of a phase transition, the existence of a power law decay of time correlation suggests that the large-scale data traffic dynamics could be described by a set of simple coupled stochastic differential equation, such as the Langevin equations with random interactions [33]. The equation for the Internet activity on a given connection (i,j) would thus be

$$\frac{\partial g_{ij}}{\partial t} = F(g_{ij}(t)) + \varepsilon_{ij}(t) + \sum_{kl} J_{(ij)(kl)} g_{kl}(t) \qquad (12)$$

where the function F is usually expanded for small g as [34]

$$F(g) \simeq -rg - ug^3 \qquad (13)$$

and describes the relaxation of a single isolated connection. The random noise ε is associated to the effect of

users and the quantity $J_{(ij)(kl)}$ is the coupling between connections (ij) and (kl). In the absence of interaction, the correlation function $<g(t)g(t+\tau)>$ decreases exponentially with a typical correlation time of order $1/r$ (for $u = 0$). When the coupling is strong enough, the system described by Equ. (12) undergoes a transition to an ordered state where all g's are centered around a non-zero value. At the transition point the correlation function is decaying as a power law [34].

In this model [Equ. (12)], the observed self-similarity in time is a consequence of the strong correlation existing in the network. This is in contrast with previous studies which explained the self-similarity as an effect of existing local power law distribution (such as the file size distribution). However, more data are needed for testing this hypothesis and the validity of Equ. (12) for the Internet traffic.

IV. CONCLUSIONS

In summary, the largest part of the correlation matrix of connections is random but also contains statistical information distinct from pure noise. The eigenvectors which correspond to eigenvalues outside of the RMT predictions contain information about genuine traffic correlation. In particular, the largest components of eigenvector u_1 (which corresponds to the largest eigenvalue) indicate the most correlated connections. We found different origins for the observed correlations. First, a connection (i, j) is always strongly correlated with (j, i) which is expected since for each process—such as web browsing for example—information is exchanged in both directions. Second, it appears that in the set of the strongly correlated connections there is only a small number of different routers which participate in different connections thereby inducing correlations. This support the idea of the existence of active centers which are either very active or very visited. More work and data—on larger space and time scales—are needed in order to understand more thoroughly the existence of such centers which seem to play an important role in the network traffic.

The approach presented in this study thus seems to allow one to extract relevant correlations between different connections and might have potential applications to traffic management and optimization. In particular, this analysis focus on activity independently of the volume of information exchanged and can thus reveal some very active routers which are usually hidden by 'big' routers exchanging very large flows.

Finally, the existence of strong correlations together with the existence of a phase transition and power-law decaying autocorrelation function suggest that the Internet traffic is similar to a spin glass close to the critical point. In this hypothesis, the self-similarity appears naturally as the result of a collective behavior without resorting to pre-existing power laws.

[1] B. A. Huberman and R. M. Lukose, Science **277**, 535 (1997).
[2] W. E. Leland *et al*, IEEE/ACM Transactions on Networking **2**, 1 (1994).
[3] I. Csabai, Journal of Physics A: Math. Gen. **27**, 417 (1994).
[4] K. Thompson, G. J. Miller, and R. Wilder, IEEE Network **11**, (Nov-Dec 1997).
[5] A. Feldmann, A. C. Gilbert, W. Willinger, and T. G. Kurtz, Computer Com. Rev. **28**, 5 (1998).
[6] M. Takayasu, H. Takayasu, and K. Fukuda, Physica A **277**, 248 (2000).
[7] For a study on the Internet structure and map, see the Internet Mapping Project: http://www.cs.bell-labs.com/who/ches/map/index.html.
[8] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *ACM SIG-COMM'99*, Comput. Com. Rev. **29**, 251 (1999)
[9] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52** 386 (2000).
[10] R. Pastor-Satorras, A. Vasquez, and A. Vespignagni, Phys. Rev. Lett. **87**, 258701 (2001). cond-mat/0105161.
[11] R. Kumar, P. Raghavan, S. Rajalopagan, and A. Tomkins, Proceedings of the 25th VLDB Conf., Edinburgh, 1999.
[12] A. Broder *et al. Proceedings of the 9th International WWW conference*, Amsterdam, The Netherlands, 309-320, Elsevier Science, May 2000.
[13] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[14] L. A. Adamic, Lect. Notes Comput. Sci **1696**, 443 (1999).
[15] B. .A. Huberman *et al*, Science **280**, 95 (1998).
[16] S. Mossa *et al*, Physical Review Letters **88**, 138701-1 (2002).
[17] M. Crovella and A. Bestavros, IEEE/ACM Transactions on Networking **5**, 835 (1997).
[18] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, IEEE/ACM Transactions on Networking **5**, 71 (1997).
[19] K.Fukuda, PhD Thesis: *A study on Phase Transition Phenomena in Internet Traffic*, Keio University (1999). Available at http://www.t.onlab.ntt.co.jp~fukuda/research/.
[20] T. Ohira and R. Sawatari, Physical Review E **58** 193 (1998).
[21] M. L. Mehta, *Random Matrices* (Academic Press, Boston, 1991).
[22] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
[23] V. Plerou *et al*, Phys. Rev. Lett. **83**, 1471 (1999).
[24] V. Plerou *et al*, Phys. Rev. E **65**, 066126 (2002).
[25] M. S. Santhanam and P. K. Patra, Phys. Rev. E **64**, 016102 (2001).

[26] For a map and more informations on this network, see the web page: http://www.renater.fr. For an animated version of flows, see: http://barthes.ens.fr/metrologie/Renater01.

[27] H. Bruus and J.-C Anglés d'Auriac, Europhys. Lett. **35**, 321 (1996).

[28] A. Edelman, SIAM J. Matrix Anal. Appl. **9**, 543 (1988) and references therein.

[29] F. J. Wegner, Z. Phys. B **36**, 209 (1980).

[30] B. Kramer and A. MacKinnon, Rep. Prog. Phys. **56**, 1469 (1993).

[31] This localization of *activity* has not to be confused with traffic localization (namely that a subgroup of the network concentrates most of the traffic volume) which was also found for this network [32].

[32] M. Barthélemy, B. Gondran, and E. Guichard, to appear in Physica A (2002).

[33] C. de Dominicis, Phys. Rev. B **18**, 4913 (1978). H. Sompolinsky and A. Zippelius, Phys. Rev. B **25**, 6860 (1982). K. H. Fisher and J. A. Hertz, *Spin Glasses*, Cambridge (1991).

[34] N. Goldenfeld *Lectures on Phase Transitions and the Renormalization Group*, Frontiers in Physics (1992).

FIG. 2. Probability distribution for the correlation coefficient calculated from 5-minutes flows in the Renater network for a 14 days period. The average value is positive indicating strong correlations among the whole network.
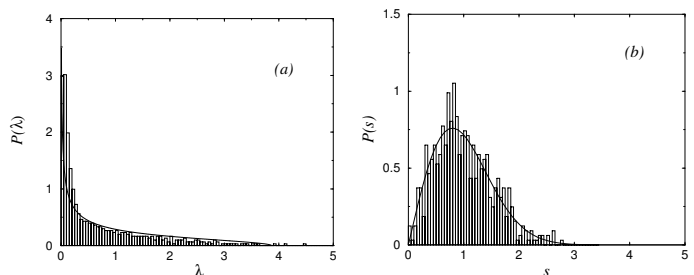
FIG. 3. **(a)** The probability density of the eigenvalues of the normalized cross-correlation matrix C for the 650 connections for a 2-weeks period. The results are reasonably fitted by the analytical result obtained for cross-correlation matrices generated from uncorrelated time series (solid line, obtained from Equ. 4 with $Q^* = 1.1$). There are however very large eigenvalues (not shown), the largest one being of order 200. **(b)** Nearest-neighbor spacing distribution of the eigenvalues of C after unfolding using the Gaussian broadening procedure [27]. The solid line is the RMT prediction for the spacing distribution for the Gaussian orthogonal ensemble (GOE).
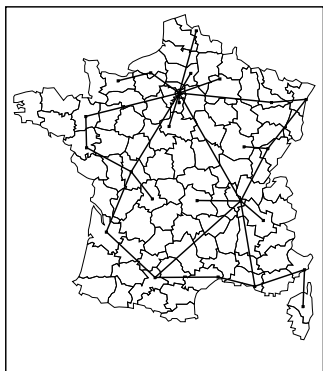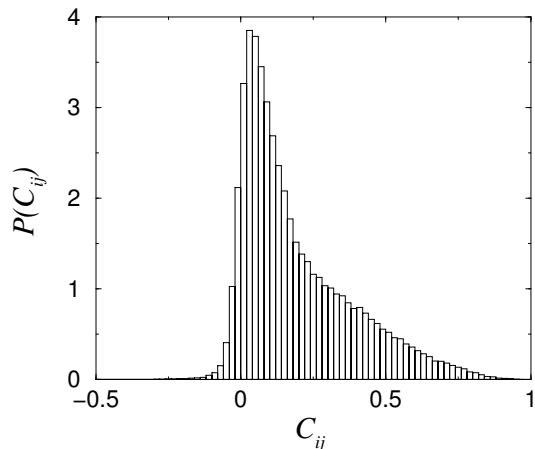

FIG. 1. Map of the Renater network. There is a total of about 30 interconnected routers (of which 26 are effectively studied). We show on this map the physical connections. The measured data consist in a flow matrix $F_{ij}(t)$ (with $t = \tau m$, $m = 0, \cdots, L - 1$ and $i, j = 1, \cdots, 26$) which gives the effective flow exchange between routers i and j. For more details on this network, see the web page http://www.renater.fr and for an animated version of flows, see http://barthes.ens.fr/metrologie/Renater01.
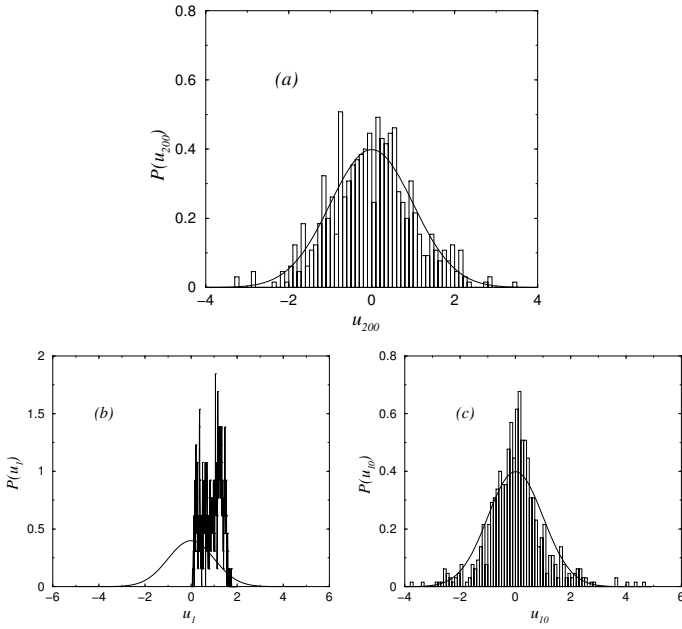
FIG. 4. Eigenvector component distribution (a) For eigenvalues in the center of the spectrum. In this case, the empirical results are in agreement with the results of RMT which is the Porter-Thomas distribution represented by a solid line. (b,c) For large eigenvalues there is a clear deviation compared to RMT predictions represented by the solid line (Porter-Thomas distribution). For the largest eigenvalue, most of the components is non-zero and positive which indicates correlations among the whole network.

FIG. 6. Number of occurrences of routers in the n most correlated connections (There is a total of 26 routers $i = 1, \cdots, 27$, the router 24 is excluded of the present study for technical reasons). In each plot, we compared the empirical results with the control case (histogram in red). The arrows indicate the two most frequent routers for Renater. In cases (a) $n = 30$ and (b) $n = 50$, it is clear that not all routers are participating equally. (c) Case $n = 100$. The control case still fluctuates around its average (which is $200/26 \simeq 7.70$) but much less than the empirical case. This fact and the observed persistency for increasing n suggest that it is very unlikely that the empirical peaks are just fluctuations due to noise. These peaks corresponds probably to routers which are very active and which are exchanging information with many other routers, thereby inducing correlations in the network.
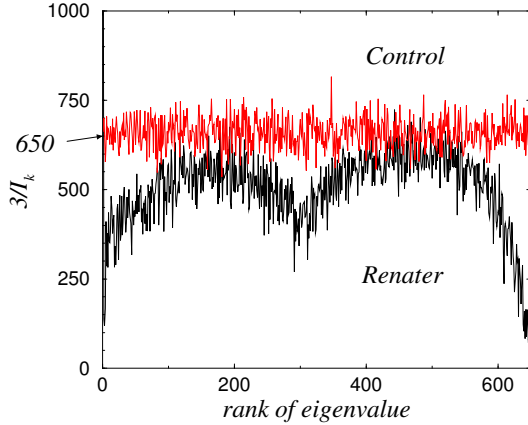

FIG. 5. Reciprocal inverse participation ratio for each of the 650 eigenvectors (sorted for decreasing eigenvalues). As a control case, we show the corresponding result for uncorrelated independent time series of the same length as the data. Empirical data show small values at both edges of the spectrum, whereas the control shows only small fluctuations around the average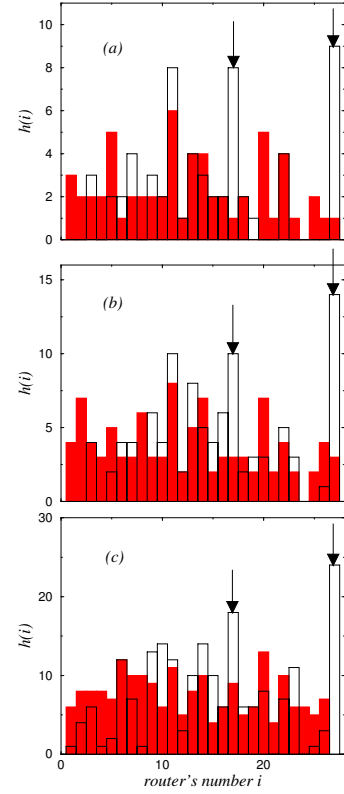 value $\langle 3/I \rangle = N = 650$.