



Extracting the multi-timescale activity patterns of online financial markets

Kobayashi, Teruyoshi

Sapienza, Anna

Ferrara, Emilio

(Citation)

神戸大学経済学研究科 Discussion Paper, 1809

(Issue Date)

2018

(Resource Type)

technical report

(Version)

Version of Record

(URL)

<https://hdl.handle.net/20.500.14094/81010099>



Extracting the multi-timescale activity patterns of online financial markets

Teruyoshi Kobayashi

Anna Sapienza

Emilio Ferrara

February 2018

Discussion Paper No.1809

GRADUATE SCHOOL OF ECONOMICS

KOBE UNIVERSITY

ROKKO, KOBE, JAPAN

Extracting the multi-timescale activity patterns of online financial markets

Teruyoshi Kobayashi,¹ Anna Sapienza,² Emilio Ferrara^{2,*}

¹*Department of Economics, Kobe University, Kobe, Japan*

²*University of Southern California, Information Sciences Institute, Los Angeles (CA), USA*

*Corresponding author: emiliofe@usc.edu

February 21, 2018

Abstract

Online financial markets can be represented as complex systems where trading dynamics can be captured and characterized at different resolutions and time scales. In this work, we develop a methodology based on non-negative tensor factorization (NTF) aimed at extracting and revealing the multi-timescale trading dynamics governing online financial systems. We demonstrate the advantage of our strategy first using synthetic data, and then on real-world data capturing all interbank transactions (over a million) occurred in an Italian online financial market (e-MID) between 2001 and 2015. Our results demonstrate how NTF can uncover hidden activity patterns that characterize groups of banks exhibiting different trading strategies (normal vs. early vs. flash trading, etc.). We further illustrate how our methodology can reveal “crisis modalities” in trading triggered by endogenous and exogenous system shocks: as an example, we reveal and characterize trading anomalies in the midst of the 2008 financial crisis.

Introduction

Social activity of individuals follows certain rhythms at different time scales. Many of the individual activities, such as sending e-mails and making phone calls, are likely to be done in particular time intervals within a day (i.e., diurnal or circadian cycles), and the total daily activity could heavily depend on the day of the week (i.e., weekly cycles) [1–6]. In general, these social activity rhythms emerging at different time scales may be correlated with each other; for instance, it has been shown that face-to-face contacts between classmates in a primary school follow a common diurnal cycle driven by the daily class schedule [7, 8], yet at the same time they would also share activity rhythms at longer scales such as weekly and monthly, reflecting the annual school schedule.

As social communications between humans form temporal social networks, financial transactions between banks also shape time-varying networks [9–13]. In the interbank market, for instance, overnight bilateral lending and borrowing between banks organize temporal networks whose structure changes on a daily basis, because the overnight financial contracts last for only one day [11, 12]. Thus, financial markets, similarly to social networks, can be interpreted as complex systems where each agent’s activity should be captured and characterized at appropriate time scales [11, 12]. In fact, many commonalities have recently been found between social communication patterns of humans and financial interaction patterns of banks at particular temporal resolutions [12, 14, 15].

In recent years, *non-negative tensor factorization* (NTF) has been frequently used to extract temporal activity patterns in various social contexts, such as face-to-face contacts [8, 16], Twitter posts [17] and players’ matches in online games [18]. Among these studies, Gauvin et al. [8] showed that NTF is highly effective in detecting diurnal rhythms of students’ activity in a primary school [7, 19]. Characterizing diurnal rhythms lead the authors to uncover the multi-timescale community structure formed by classmates: students’ temporal activity cycles, rather than the aggregated history of contacts, allows to reveal meaningful patterns that explain the complexity of children contact networks and communities.

Given the similarity between social and financial temporal dynamics, justified by underlying human factors, and the fact that multiple activity cycles are present at different frequencies in human social activities, our question is whether similar rhythms exist also in financial systems. In this work, we uncover hidden multi-timescale patterns of banks’ activities in the Italian online interbank market, e-MID [20]. In previous studies, it has been recognized that there exist activity patterns in banks’ financial transactions at a particular time scale, such as inter-day [12] and intra-day scales [21–23]. However, it is still unknown whether these patterns coexisting at different time scales are dependent on each other, in which case banks exhibiting a given interday activity pattern are likely to follow a particular intraday pattern.

We employ NTF as a tool to detect multi-timescale patterns, in which banks’ activities are captured by a tensor having three dimensions: the list of banks, time of the day, and date. We will show that the PARAFAC decomposition of the 3-way tensor [24, 25] indicates that banks’ trading activities can be classified into several patterns over the data period of 2001–2015. In particular, the NTF allows us to identify an anomalous pattern around the Lehman Brothers’ collapse in September 2008. Our multi-timescale NTF approach reveals not only how banks facing the financial crisis changed their diurnal trading rhythms, but also on what dates such anomalies emerged. Our framework to extract multi-timescale patterns advances the understanding of when and how banks react to systemic shocks, and in the future it could contribute to controlling the spread of financial systemic risk [9, 10, 26–28].

Data

We use the time-stamped data on bilateral interbank transactions occurred between January 2001 and December 2015 in the Italian online interbank market (e-MID). e-MID is an online platform for financial institutions provided by the Italian company e-MID SIM S.p.A. based in Milan, Italy (henceforth, we call financial institutions as “banks” for brevity). Banks in need of lending or borrowing funds may find a counterpart by posting an order on the online platform, which makes e-MID a marketplace in which lenders and borrowers are matched. Each transaction record in this data represents when a loan contract agreement has been reached between two banks and how much funds have been lent (in million Euros). In addition, the data contains information about the type of each loan agreement, i.e., a lender-proposed transaction or a borrower-proposed transaction. We define a lender-proposed (resp., borrower-proposed) transaction to be a transaction proposed by a bank that lends to (resp., borrows from) its counterpart.

As in the other interbank markets, the vast majority of transactions ($> 86\%$) in the e-MID market are overnight lending and borrowing (i.e., a loan contract lasts just for one day), while there are other types of transactions having longer maturity lengths such as two weeks, three month, etc. To eliminate the influence of differences in maturity length, we focus on overnight transactions of unsecured Euro deposits labeled as “ON” (i.e., overnight) or “ONL” (i.e., overnight large, namely overnight transactions greater than 100 million Euros). Table 1 summarizes the basic statistics of this dataset.

There are 3,839 business days over the data period between Jan 2, 2001 and Dec 31, 2015, on each of which transactions are made between 8:00 and 18:00. There are 289 banks that conducted at least one transaction over this period: 194 of them are Italian banks while the rest are foreign banks. The total number of transactions (ON and ONL) that we use in the analysis amounts to 1,148,699. It

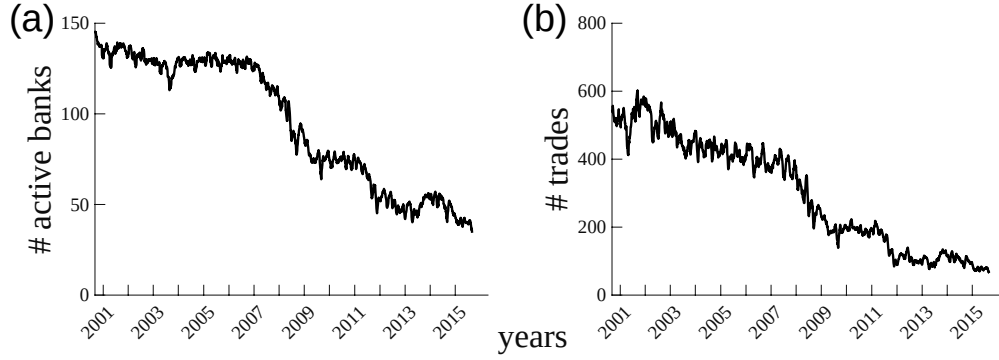


Figure 1: Daily activities in the e-MID market. 20-day moving average is shown. (a) Number of active banks having at least one transaction on a day. (b) Number of trades.

Table 1: Summary statistics of the Italian interbank market, e-MID.

Data period	Jan 2, 2001 – Dec 31, 2015
No. days	3,839
Trading time	8:00 – 18:00
Maturity length	overnight (ON and ONL)
No. banks (Italian banks)	289 (194)
No. total transactions	1,148,699
Average no. daily participants	95.10
Average no. daily transactions	299.22

should be noted that the 289 banks are not always active over the data period (*cf.*, Fig. 1a). In fact, the daily average of the number of participating banks is around 95, and the average number of daily transactions is approximately 300 (*cf.*, Fig. 1b). The dataset is commercially available from e-MID SIM S.p.A (<http://www.e-mid.it/>).

Method

As noted in the previous section, banks are not always active in the interbank market. At the daily scale, some banks have transactions only on a certain fraction of business days while having no transactions on the other days [12, 21]. If we look at intraday time scales, on the other hand, the frequency of a bank having at least one transaction is not uniformly distributed over the time intervals [21, 22].

Since the maturity of loan contracts is overnight, it is important to identify what are the days in which banks are more likely to be active. Conversely, since the participation of banks in the trading activity can change over the course of a day, it is also crucial to understand intraday activity changes. Due to the multi-scale nature of bank trading activity, directly studying the number of transactions of each bank over time would not allow us to detect inter- and intra-day dynamics. Thus, we need to disentangle each bank’s activity across the two dimensions (i.e., time scales): interday and intraday. To this aim, we represent our data in a three-dimensional array (i.e., tensor), whose entries represent the amount of trades each bank performs at a given time of a given day. Given this multi-dimensional representation, our main goal is to extract some meaningful correlated multi-timescale activity patterns related to groups of banks sharing a similar number of transactions at the same time (respectively, intraday and interday). This can be achieved by means of *non-negative tensor factorization* (NTF), as described in the following.

Non-negative tensor factorization (NTF) at different time scales

Let us consider a three-dimensional tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, where $I = N$ is the number of banks, $J = T$ is the number of time intervals of bank activity in a day (here between 8:00 and 18:00), and $K = D$ is the number of days. The entry $x_{ijk} \in \mathbb{Z}_+$ of the tensor $\mathcal{X}$ denotes the total amounts of trades (in million Euros) conducted by the i -th bank during j -th time interval of the k -th day. To extract the intra- and inter-day trading patterns characterizing banks with similar number of transactions over time, we rely on the NTF, which decomposes the tensor $\mathcal{X}$ into the sum of several rank-one tensors, namely components.

Here, we implement the NTF by applying the well-established method called PARAFAC (parallel factor analysis) or CANDECOMP (canonical decomposition) with non-negativity constraints [24, 25, 29]. The decomposition can be written as

$$\mathcal{X} \approx \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (1)$$

$$x_{ijk} \approx \sum_{r=1}^R a_{ir} b_{jr} c_{kr}, \quad (2)$$

where $R \in \mathbb{N}$ denotes the number of components (i.e., the rank of the tensor), and the operator $\circ$ represents outer product. Eq. (1) is the so-called canonical polyadic decomposition (CP) of a tensor $\mathcal{X}$, where $\mathbf{a}_r \in \mathbb{R}_+^N$, $\mathbf{b}_r \in \mathbb{R}_+^T$ and $\mathbf{c}_r \in \mathbb{R}_+^D$ represent r -th component factors that respectively encode the membership of a bank to the component, the intervals in a day and the days in which the component is active (i.e. $\mathbf{b}_r$ and $\mathbf{c}_r$ are the intra- and inter-day activity patterns of the groups of banks in $\mathbf{a}_r$).

Let $\mathbf{A} \in \mathbb{R}_+^{N \times R}$, $\mathbf{B} \in \mathbb{R}_+^{T \times R}$ and $\mathbf{C} \in \mathbb{R}_+^{D \times R}$ denote the factor matrices, whose r -th columns are the vectors $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$, respectively. The factor matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are calculated by solving the following minimization problem with non-negativity constraints:

$$\min_{\mathbf{A} \geq 0, \mathbf{B} \geq 0, \mathbf{C} \geq 0} \|\mathcal{X} - [\mathbf{A}, \mathbf{B}, \mathbf{C}]\|_{\text{F}}^2, \quad (3)$$

where $[\mathbf{A}, \mathbf{B}, \mathbf{C}]$ represents the Kruscal form of the tensor decomposition (i.e., the RHS of Eq. (1)), and $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm.

This minimization problem can be rewritten to perform the minimization with respect to one factor matrix at a time. Let $\mathbf{X}_{(n)}$ denote a matrix created by the mode- n matricization (or flattening) of tensor $\mathcal{X}$, where each row of $\mathbf{X}_{(n)}$ consists of a vector corresponding to a given index of mode- n . That is, $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}$ and $\mathbf{X}_{(3)}$ are $N \times TD$, $T \times ND$ and $D \times NT$ matrices, respectively. They can be expressed in terms of factor matrices as follows:

$$\mathbf{X}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^\top, \quad (4)$$

$$\mathbf{X}_{(2)} = \mathbf{B}(\mathbf{C} \odot \mathbf{A})^\top, \quad (5)$$

$$\mathbf{X}_{(3)} = \mathbf{C}(\mathbf{B} \odot \mathbf{A})^\top, \quad (6)$$

where $\odot$ denotes the Khatri–Rao product, which is the “matching columnwise” Kronecker product defined as $\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \dots, \mathbf{a}_R \otimes \mathbf{b}_R]$. The minimization problem for PARAFAC decomposition Eq. (3) can be now reformulated as

$$\left\{ \min_{\mathbf{A} \geq 0} \|\mathbf{X}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^\top\|_{\text{F}}^2, \quad \min_{\mathbf{B} \geq 0} \|\mathbf{X}_{(2)} - \mathbf{B}(\mathbf{C} \odot \mathbf{A})^\top\|_{\text{F}}^2, \quad \min_{\mathbf{C} \geq 0} \|\mathbf{X}_{(3)} - \mathbf{C}(\mathbf{B} \odot \mathbf{A})^\top\|_{\text{F}}^2 \right\}. \quad (7)$$

The minimizers of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ of this problem are computed using the non-negative alternate least squares method (ANLS) combined with the Block Principal Pivoting method (BPP) developed by [30]. Our implementation is based on the Tensor Toolbox [31, 32] and the MATLAB codes available from [33].

Rank size

To determine the number of components R (i.e., rank size) used in the NTF model, we rely on the Core Consistency Diagnostic [24]. Given the tensor element x_{ijk} , we can rewrite Eq. (2) as

$$x_{ijk} = \sum_{n=1}^R \sum_{m=1}^R \sum_{p=1}^R \lambda_{nmp} a_{in} b_{jm} c_{kp}, \quad (8)$$

where λ_{nmp} denotes the (n, m, p) element of the superdiagonal binary tensor $\mathcal{L}$ (i.e., $\lambda_{nmp} = \delta_{nm} \delta_{mp} \delta_{np}$). Now let g_{ijk} denote the (i, j, k) element of the core tensor $\mathcal{G}$, which is obtained by fitting the data to the Tucker3 model [24]. In the Tucker3 model, the minimization problem reads

$$\min_{\mathbf{G}} \|\mathbf{X} - \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^\top\|_F^2, \quad (9)$$

where $\mathbf{X}$ is a $N \times TD$ matrix converted from tensor $\mathcal{X}$, and $\mathbf{G}$ is also a matricized version ($R_n \times R_m R_p$) of the core tensor $\mathcal{G}$. With the Tucker3 decomposition, the tensor is written as

$$x_{ijk} = \sum_n^{R_n} \sum_m^{R_m} \sum_p^{R_p} g_{nmp} a_{in} b_{jm} c_{kp}. \quad (10)$$

From the lemma of Bro and Kiers [24], the core tensor $\mathcal{G}$ will be identical to the superdiagonal tensor $\mathcal{L}$ if the Tucker3 model is perfectly fitted and the factors have full column rank. If $\mathcal{G}$ is significantly different from the superdiagonal tensor $\mathcal{L}$, by contrast, it means that there are non-negligible interactions between factors and the PARAFAC model is not appropriate. Thus, we could assess whether the data should be fitted to the PARAFAC model or the Tucker3 model by measuring the distance between $\mathcal{G}$ and $\mathcal{L}$.

We employ the Core Consistency (CC) value proposed by Bro and Kier [24] as a measure of the distance between $\mathcal{G}$ and $\mathcal{L}$:

$$\text{CC} = 100 \times \left(1 - \frac{\sum_{n=1}^R \sum_{m=1}^R \sum_{p=1}^R (g_{nmp} - \lambda_{nmp})^2}{R} \right), \quad (11)$$

where we imposed the constraint $R_n = R_m = R_p = R$ in implementing the Tucker3 decomposition. CC takes 100 if the PARAFAC model perfectly fits the data and less than 100 (possibly a negative value) if the model does not fit perfectly. Note that a rise in R will reduce fitting errors while increasing the possibility of over-fitting. In general, as R increases, there arises more interactions between components, which makes the core tensor $\mathcal{G}$ far from superdiagonal, resulting in a low value of CC. Therefore, it is reasonable to stop increasing R before the symptom of over-fitting emerges. Our determination of R is given as

$$R = \max R' \in \{R' : \text{CC}(R') > L_{cc}, R' \in \mathbb{N}, \}, \quad (12)$$

where $\text{CC}(\cdot)$ denotes the core consistency as a function of the number of components, and L_{cc} is a threshold parameter. In line with [16], we set $L_{cc} = 85$, and to minimize randomness created by a PARAFAC decomposition, we implement PARAFAC decomposition 20 times for a given rank size and use their mean as the value for $\text{CC}(R')$. Our implementation uses the MATLAB code available from [34] which is based on the alternating nonnegativity-constrained least squares with block principal pivoting [35].

Results

Synthetic temporal financial markets

To validate the accuracy of our multi-timescale decomposition, we first implement NTF on a synthetic financial market whose properties are fully known *ex ante*. Suppose that there are three groups of banks,

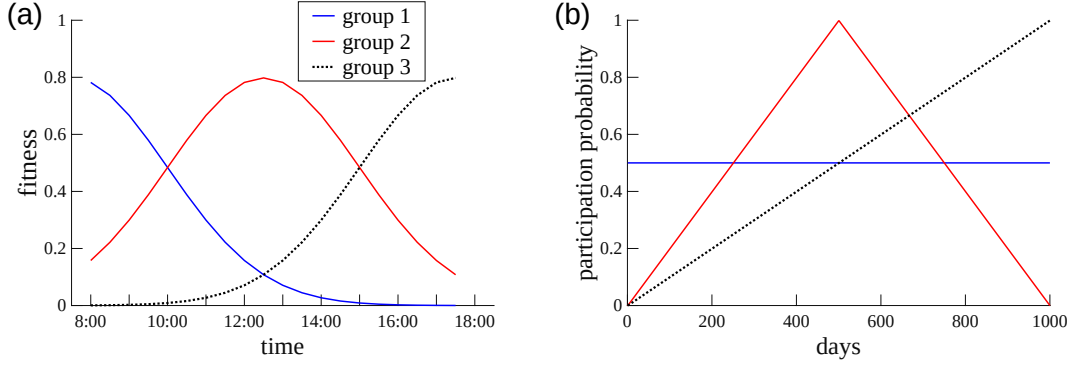


Figure 2: Predefined intra- and interday trading patterns for the synthetic financial market. (a) Predefined intraday activity patterns, represented by the fluctuation of fitness. (b) Predefined interday activity patterns, given by the participation probability.

each of which having particular intra- and inter-day trading patterns. First, we introduce *intraday* trading patterns by considering the “fitness” of banks. In the fitness model, the probability that banks i and j trade at time t is given by

$$p_{ij,t} = a_{i,t}a_{j,t}, \quad (13)$$

where $a_{i,t} \in [0, 1]$ denotes the fitness of bank i within the time interval $[t - \Delta, t]$ [12, 36]. Δ is the length of a time interval (in a day) expressed in minutes. Differences in the intraday trading patterns across groups are then represented by different values of fitness. Let us assume that the fitness evolves as illustrated in Fig. 2a. Banks belonging to group 1 are most active in the early morning (blue solid), and banks in group 2 exhibit the highest trading activity around noon (red solid), and banks in group 3 are most active at the end of the day (black dotted). More specifically, the fitness value of a bank belonging to group s at time interval t is given as

$$a_{i,t}^s = f(t; \mu_s, \sigma), \quad t = 1, \dots, T, \quad (14)$$

where $f(t; \mu_s, \sigma)$ is the p.d.f. of normal distribution with mean μ_s and standard deviation σ . We set $(\mu_1, \mu_2, \mu_3) = (0, T/2, T)$ and $\sigma = T/4$.

Second, the *interday* trading patterns are captured by variations in a bank’s participation probability, which is the probability that a bank participates in the market on a given day, apart from whether or not the bank is able to find a trading partner *ex post*. For day $d = 1, \dots, D$, the participation probability of bank i , denoted by $q_{i,d}$, is given as (Fig. 2b)

$$q_{i,d} = \begin{cases} 0.5 & \text{if } i \in \text{group 1,} \\ \frac{2}{D}(d-1) & \text{if } i \in \text{group 2 and } d \leq D/2, \\ \frac{-2}{D}(d-D) & \text{if } i \in \text{group 2 and } d > D/2, \\ \frac{1}{D}(d-1) & \text{if } i \in \text{group 3.} \end{cases} \quad (15)$$

For simplicity, we assume that trade volumes are identical for all trades so that the (i, j, k) element of the 3-dimensional tensor is equal to the total number of trades that bank i has during time interval j on day k . We set $N = 120$, $T = 20$ (i.e., resolution $\Delta = 30$ for hours between 8:00 and 18:00 in each day) and $D = 1000$.

Using the above synthetic tensor, we can now implement our multi-timescale tensor factorization. The factorization of tensor $\mathcal{X}$ allows us to measure the activity of each individual element: a bank, an intra-day time interval, and a day. The activity of intra-day time interval j of component r is given by

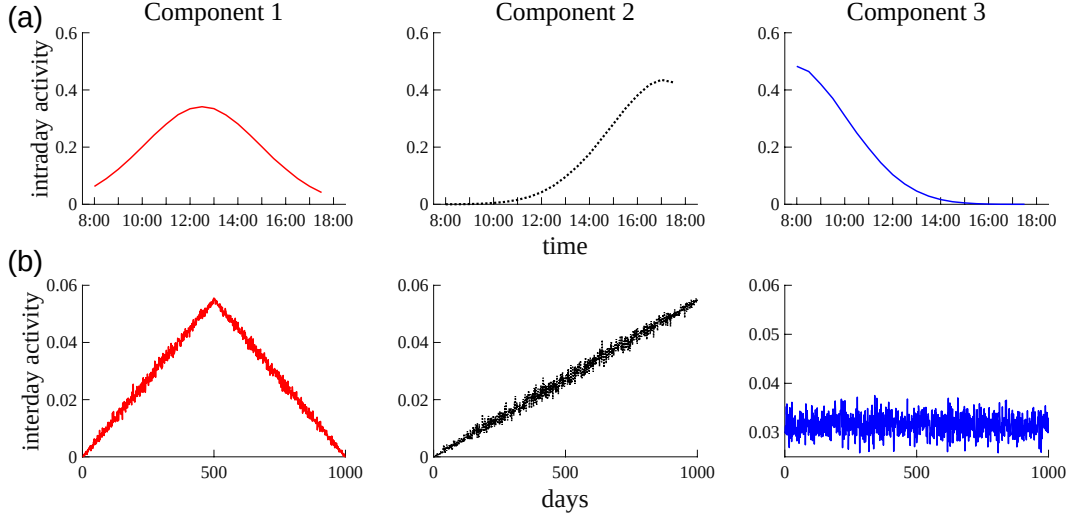


Figure 3: PARAFAC decomposition for the synthetic financial markets. (a) Intra-day activity for each component. Activity in time interval j of component r is given by b_{jr} . (b) Inter-day activity for each component. Activity of day k of component r is given by c_{kr} .

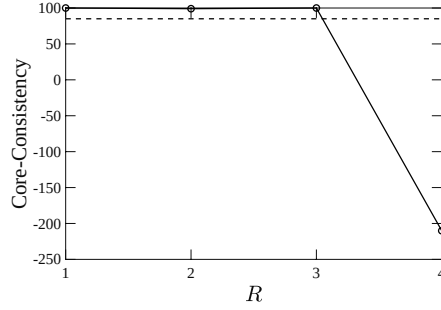


Figure 4: Core-Consistency for the synthetic markets. Dashed line denotes $L_{cc} = 85$.

b_{jr} , and analogously the activity of day k of component r is represented by c_{kr} . Fig. 3 illustrates that PARAFAC decomposition well extracts the true multi-timescale patterns. The core-consistency value strongly suggests $R = 3$ since $CC(4)$ takes a large negative value.

Empirical data

Now we analyze empirical data. Given the accuracy of our method in the synthetic model, we expect the NTF method to extract latent multi-timescale patterns in the real-world financial market, e-MID. Here, we implement the $N \times T \times D$ tensor factorization described in Method, where we set intraday time resolution $\Delta = \{3, 5, 10, 15, 30, 45, 50\}$, and D is set at the total number of business days between January 2001 and December 2015 (i.e., $D = 3,839$). The number of active banks N , which traded at least once during the period, is 289. The Core-Consistency will reveal the number of components R , if any, which is equivalent to the the variety of existent activity patterns.

Core-Consistency

The CC value for a given rank size is illustrated in Fig. 5a for $\Delta = 15$. The CC value becomes lower than 85 for $R \geq 4$, suggesting that $R = 3$ should be selected. In general, imposing different intraday

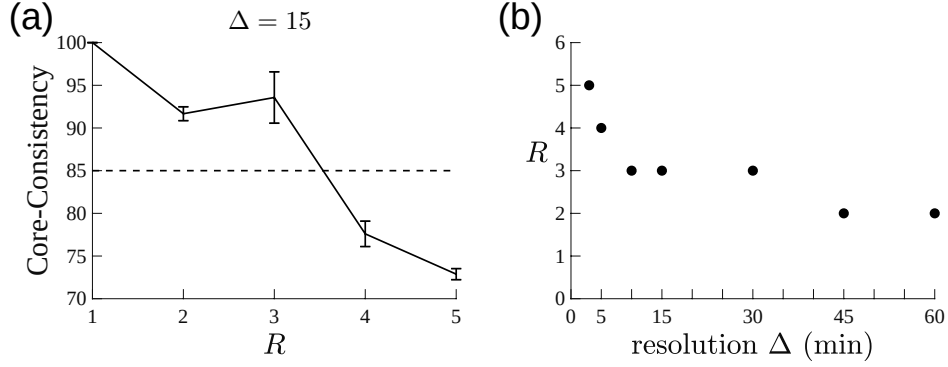


Figure 5: Determination of rank size. (a) Core-Consistency values for $\Delta = 15$. Error bar denotes the 95% confidence interval of the mean Core-Consistency value, calculated by 20 runs of PARAFAC decomposition. (b) Selected value of R for a given time resolution $\Delta \in \{3, 5, 10, 15, 30, 45, 60\}$.

resolutions may lead us to choose different values of R (Fig. 5b). There is a tendency that a higher resolution requires a larger number of components for the PARAFAC decomposition to be appropriate. In the following, we employ $\Delta = 15$ and thereby $R = 3$ as the benchmark case, the rationale of which will be discussed in the next subsection.

Intra-day and inter-day activity

Fig. 6a shows *intra-* and *inter-day* activities for $\Delta = 15$, in which case there are three different patterns. The intra-day activity of Component 1 is characterized by its bimodal pattern.¹ It has two peaks around 10:00 and 17:00. On the other hand, both Components 2 and 3 have a single distinct peak in the early morning, after which their activities are very low. However, the distribution of intraday activity of Component 3 is more skewed than that of Component 2, which makes Components 1 and 3 two polar cases and Component 2 in between them. We would call the trading patterns of Components 1, 2 and 3 as *normal trading*, *early trading*, and *flash trading*, respectively.

The activity patterns for different temporal resolutions are also presented in Fig. S1. For the case of a finer intra-day resolution (e.g., $\Delta = 5$, Fig. S1a), there are more than three components, and the activity of an additional component (i.e., Component 4) turns out to be very similar to the flash-trading pattern we saw in Fig. 6a. On the other hand, when the temporal resolution is low (e.g., $\Delta = 45$, Fig. S1c), there are only two components and the flash-trading pattern is no longer detected. Given these observations, it is natural to set $\Delta = 15$ as a benchmark case, where apparently independent intra-day patterns can be captured. It should be noted that the following results are not sensitive to the temporal resolution level as long as $R = 3$ is to be selected (e.g., $\Delta = 30$, Fig. S1b).

Fig. 6b illustrates the *inter-day* activity of each component. For Components 1 and 2, the activity decreased radically during the global financial crisis of 2007–2009, which was initiated by a significant decline of the US house prices (the peak date is indicated by blue dotted). In contrast, the activity of Component 3 spiked around the collapse of Lehman Brothers in September 2008. Fig. 6c shows how the share of a given component evolved over time.

These observations suggests that Component 3 captures the “crisis mode” of bank trading while Components 1 and 2 represent more normal trading patterns. The crisis-mode interpretation is reinforced by the intra-day trading pattern of Component 3, in which banks trade only in the very early morning. In fact, we will show that, although we did not use information about the directionality of trades, this

¹Since the assignment of component index is originally arbitrary in the NTF implementation, we re-assign index in the ascending order of cumulative intra-day activity between 8:00 and 10:00.

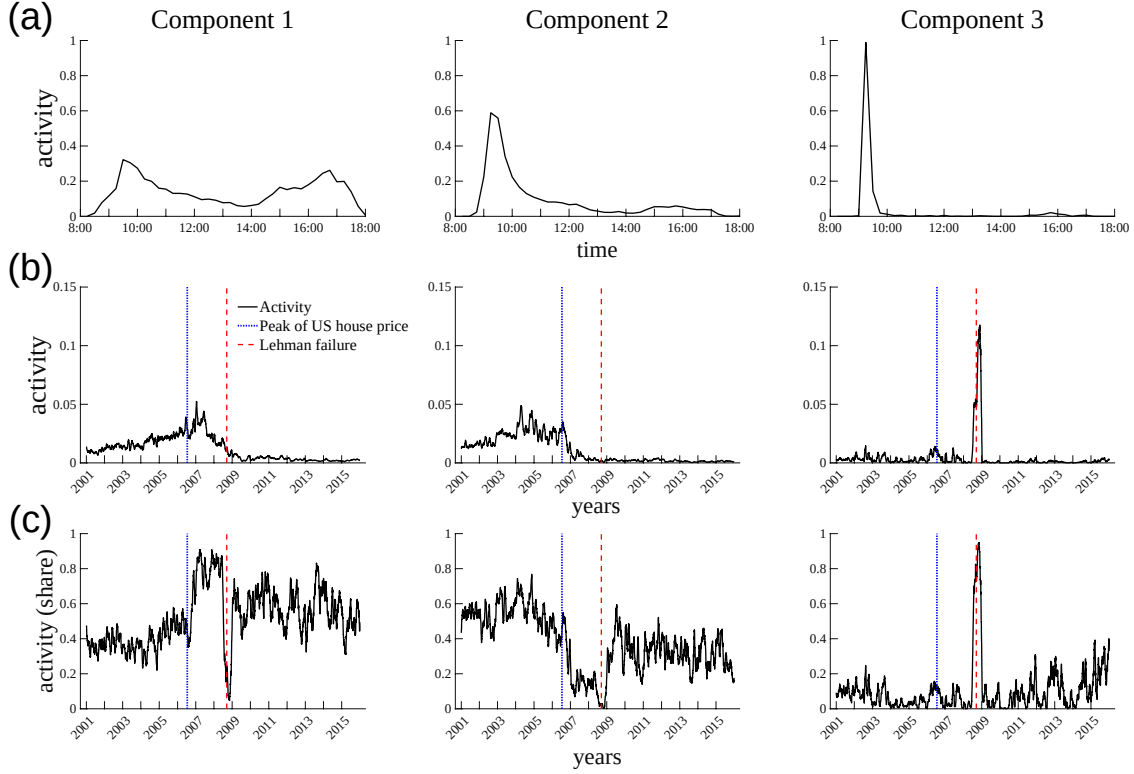


Figure 6: Intra- and inter-day activities. Peak of the US house price (July 2006) and Lehman collapse (September 2008) are indicated by blue dotted and red dashed, respectively. The data for the US house price is taken from FRED (S&P/Case-Shiller U.S. National Home Price Index) [37] (a) Intra-day activity. (b) Inter-day activity. (c) Share of a component in the daily activity. In panels (b) and (c), 20-day moving average is shown.

anomaly comes from the fact that banks in need of liquidity tried to obtain loans, rather than to extend credit, in the midst of the financial crisis.

Banks' affiliation to components

Each element of vector $\mathbf{a}_r$ represents the extent to which a bank's trading pattern is captured by Component r . Thus, we can extract a subgroup of banks whose trading patterns are characterized by Component r at least to some extent. Here, the set of banks belonging to component r is given by $I_r = \{i'\}$ for which $a_{i'r}$ is within the 90th percentile of $\{a_{ir}\}_{i=1}^N$ (Fig. 7a). It should be noted that a bank may belong to multiple components if its trading pattern has multiple features (Fig. 7b).

The bank activity patterns emerging at the daily scale, captured by the row average of $\mathbf{a}_r \mathbf{c}_r^\top$, exhibit explicit differences across components (Fig. 7c). While all types of banks have not been active since 2009, the normal-trading feature of banks belonging to Component 1 represents their behavior until 2008, and the trading pattern of banks belonging to Component 2 are dominated by the early-trading pattern up to 2006. On the other hand, the banks belonging to Component 3 exhibit daily patterns similar to the other types of banks prior to 2008, but suddenly their activity level spiked around the collapse of Lehman Brothers.

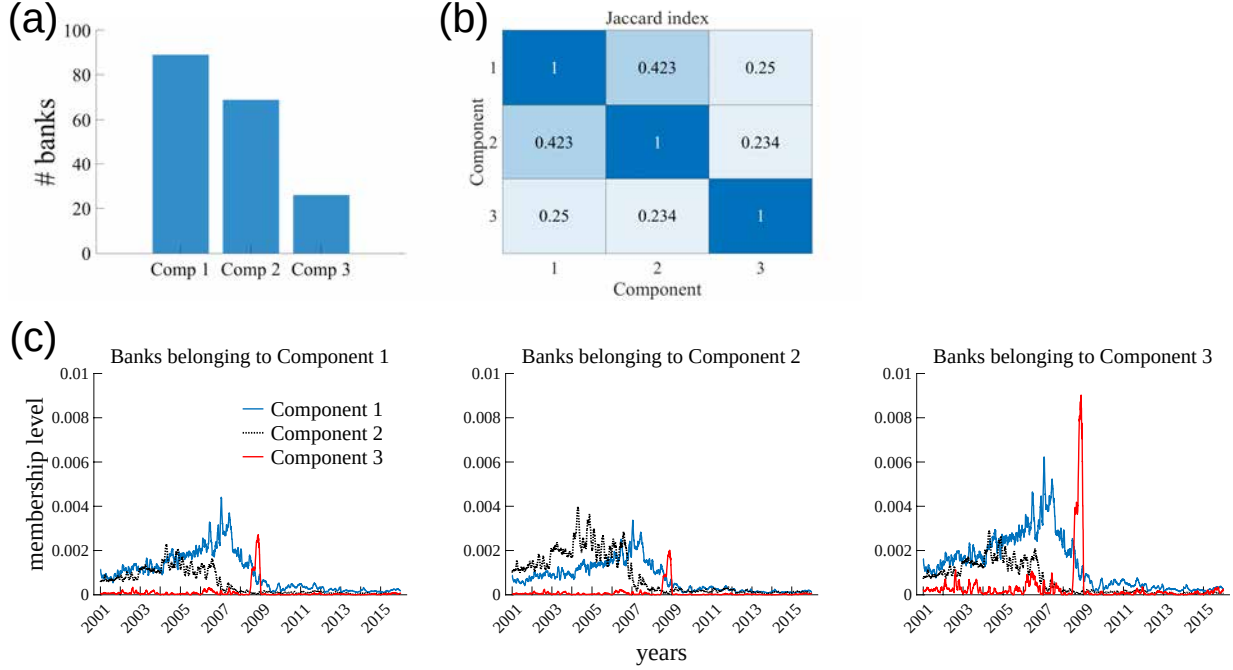


Figure 7: Affiliation of banks to components. The set of banks belonging to component r is given by $I_r = \{i'\}$ for which $a_{i'r}$ is within the 90th percentile of $\{a_{ir}\}_{i=1}^N$. (a) Number of banks that belong to a component. (b) Jaccard index for the overlap of banks belonging to multiple components. (c) Average membership level on a given day. Membership level for a bank is given by a row of $\mathbf{a}_r \mathbf{c}_r^\top$.

Characterization of components

Aside from the intra- and inter-day trading patterns, it would be of interest to see whether we may characterize each component by inherent attributes of the belonging banks. In the e-MID data set, a trading bank is classified into either of the following four types: aggressor lender, quoter lender, aggressor borrower and quoter borrower. In the e-MID online platform, a bank posts a request for loans or a proposal for lending, and a loan contract is made if another bank accepts the posted request. A bank is called “aggressor lender” when the bank accepts a request for loan posted on the e-MID platform, and the bank posted the request and borrowed fund is called “quoter borrower”. “quoter lender” and “aggressor borrower” can also be understood analogously. In addition, we can also ask to what extent the nationality of banks can explain the differences in trading patterns.

In fact, the majority of transactions in the e-MID market were done between aggressor lenders and quoter borrowers, which is a common property to all the component types (Fig. 8a). However, more than half of the banks belonging to Component 3 are quoter borrowers and the share of aggressor lenders is significantly lower than that. As for the nationality, the share of Italian banks among all the banks belonging to Component 1 is significantly low while the corresponding fraction is significantly large for the banks in Component 2 (Fig. 8b) at 90% statistical significance, respectively. This suggests that foreign banks (i.e., banks from countries outside Italy) are likely to exhibit the normal-trading pattern, and Italian banks are more likely to employ the early-trading pattern than foreign banks do.

These observations give us important implications about the multi-timescale trading patterns in the financial market. First, flash trading, which represents the intra-day trading pattern of Component 3, can be attributed mostly to quoter borrowers. Recall that flash trading was observed mostly in the period around the Lehman collapse in September 2008, but we do not know *ex ante* whether such a trading pattern was driven either by demand or supply. The fact that Component 3 is largely attributed

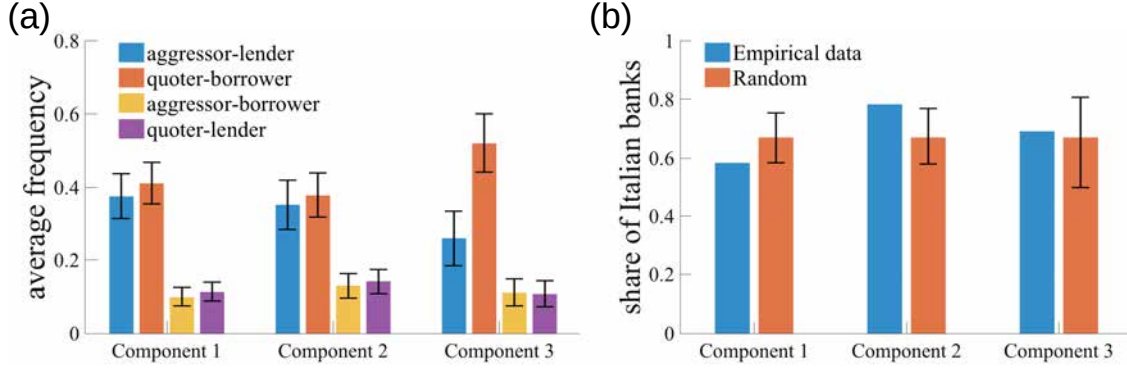


Figure 8: Role of bank attributes in the characterization of components. (a) Average frequency of a bank being an aggressor lender, a quoter borrower, an aggressor borrower and a quoter lender. Error bar denotes 95% confidence interval. (b) Share of Italian banks in each group. “Random” denotes the average share of Italian banks under random chance, and error bar denotes the 90% interval calculated from a binomial distribution with parameter (N_r, p) , where N_r is the number of banks belonging to component r and $p \equiv \# \text{Italian banks} / N$.

to quoter borrowers suggests that the flash-trading pattern during the financial crisis was conducted by banks that attempted to obtain liquidity as early as possible by posting quotes for loans. This may be regarded as evidence that some banks in fact faced a serious liquidity shortage at the time of the Lehman collapse [38, 39]. Our multi-timescale NTF approach reveals not only how banks reacted to the fear of liquidity shortage (i.e., intra-day pattern), but also specific dates on which the fear was most evident (i.e., inter-day pattern).

Second, differences in the nationality of banks can lead to the variety of trading patterns. It turns out that foreign banks (Italian banks) are more likely to employ normal trading (early trading) than early trading (normal trading). On the other hand, the source of the crisis modality cannot be explained by the nationality of banks (Fig. 8b), suggesting that banks had similar chance of being in a crisis mode regardless of their nationalities.

Discussion

In this work, we presented an analytic framework based on *non-negative tensor factorization* (NTF) to extract temporal activity patterns of financial systems. Despite previous studies on online financial markets recognized the existence of trading activity patterns at specific time scales (e.g., inter-day [12], or intra-day [21–23]), we demonstrated—to the best of our knowledge for the first time—that activity patterns coexist at different time scales and depend upon each other. Our methodology allowed us to uncover the hidden multi-timescale patterns of trading activities in an online interbank market (e-MID [20]).

Within our framework, banks’ activities were represented by means of a tensor of three dimensions, namely the list of banks, time of the day, and date. Leveraging the power of *non-negative tensor factorization* (based on the PARAFAC decomposition of the 3-way tensor [24, 25]), we uncovered the multi-timescale nature of financial trading dynamics in e-MID, suggesting that banks’ trading patterns could be classified into subgroups exhibiting significantly different multi-timescale patterns. Our framework also allowed us to attribute roles to banking institutions (aggressor lender/borrower, or quoter lender/borrower) based on their trading patterns, yielding interpretable analytic insights.

By modeling the multi-timescale dynamics occurred over the period covered by our data (2001–2015), the proposed NTF-based framework allowed us to identify trading anomalies in the midst of the financial

crisis. For example, around the time of Lehman Brothers’ collapse in September 2008, our approach showed how banks changed their trading rhythms, and on what dates such anomalies emerged.

Understanding trading patterns in financial systems is of fundamental importance for many reasons: first, a rigorous characterization of such systems’ trading dynamics could help central banks to effectively intervene in the interbank market. In turn, this would contribute to reducing systemic risk in the financial system as a whole. Financial linkages created by bilateral transactions between banks can lead to a global network of interconnected risk, in which a failure of one bank can immediately spread over the entire system [9, 26, 27]. Extracting multi-timescale patterns of financial markets allows for a better understanding of when and how banks react to systemic shocks.

Over the last couple of years, many proposals for measuring and controlling systemic risk have been presented, yet most of them focus on the static nature of financial networks [40–42]. However, as we showed in this study of the Italian e-MID interbank market, various patterns and anomalies systematically emerge at different time scales, which should be taken into account in the design of financial regulations. This work lays out the foundations for other researchers to study how patterns and anomalies emerge in financial systems at different temporal resolutions: in the future, such a framework could contribute to controlling the spread of financial systemic risk [9, 10, 26–28].

In this spirit, we will investigate whether our method could be used both in predictive and prescriptive ways to help determine what could happen under certain risk conditions, as well as to provide possible intervention strategies prior to shocks or cascading collapses in financial systems.

References

- [1] Malmgren, R. D., Stouffer, D. B., Motter, A. E. & Amaral, L. A. N. A Poissonian explanation for heavy tails in e-mail communication. *Proc. Natl. Acad. Sci. USA* **105**, 18153–18158 (2008).
- [2] Malmgren, R. D., Stouffer, D. B., Campanharo, A. S. L. O. & Amaral, L. A. N. On universality in human correspondence activity. *Science* **325**, 1696–1700 (2009).
- [3] Jo, H. H., Pan, R. K. & Kaski, K. Emergence of bursts and communities in evolving weighted networks. *PLOS ONE* **6**, e22687 (2011).
- [4] Jo, H. H., Karsai, M., Kertész, J. & Kaski, K. Circadian pattern and burstiness in mobile phone communication. *New J. Phys.* **14**, 013055 (2012).
- [5] Aledavood, T., Lehmann, S. & Saramäki, J. Digital daily cycles of individuals. *Front. Phys.* **3**, 73 (2015).
- [6] Aledavood, T. *et al.* Daily rhythms in mobile telephone communication. *PLOS ONE* **10**, e0138098 (2015).
- [7] Stehlé, J. *et al.* High-resolution measurements of face-to-face contact patterns in a primary school. *PLOS ONE* **6**, e23176 (2011).
- [8] Gauvin, L., Panisson, A. & Cattuto, C. Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach. *PLOS ONE* **9**, e86028 (2014).
- [9] May, R. M., Levin, S. A. & Sugihara, G. Complex systems: Ecology for bankers. *Nature* **451**, 893–895 (2008).
- [10] Haldane, A. G. & May, R. M. Systemic risk in banking ecosystems. *Nature* **469**, 351–355 (2011).

- [11] Barucca, P. & Lillo, F. Disentangling bipartite and core-periphery structure in financial networks. *Chaos Solitons Fractals* **88**, 244–253 (2016).
- [12] Kobayashi, T. & Takaguchi, T. Social dynamics of financial networks. arXiv:1703.10832 (2017).
- [13] Caccioli, F., Barucca, P. & Kobayashi, T. Network models of financial systemic risk: A review. *J. Comput. Soc. Sci.* **1**, 81–114 (2018).
- [14] Starnini, M., Baronchelli, A. & Pastor-Satorras, R. Modeling human dynamics of face-to-face interaction networks. *Phys. Rev. Lett.* **110**, 168701 (2013).
- [15] Starnini, M., Baronchelli, A. & Pastor-Satorras, R. Model reproduces individual, group and collective dynamics of human contact networks. *Soc. Netw.* **47**, 130 – 137 (2016).
- [16] Sapienza, A., Barrat, A., Cattuto, C. & Gauvin, L. Estimating the outcome of spreading processes on networks with incomplete information: a mesoscale approach. arXiv:1709.01806 (2017).
- [17] Panisson, A., Gauvin, L., Quaggiotto, M. & Cattuto, C. Mining concurrent topical activity in microblog streams. In *Proceedings of the 4th workshop on ‘Making Sense of Microposts’*, World Wide Web Conference 2014 (2014).
- [18] Sapienza, A., Bessi, A. & Ferrara, E. Non-negative tensor factorization for human behavioral pattern mining in online games. arXiv:1702.05695 (2017).
- [19] Cattuto, C. *et al.* Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLOS ONE* **5**, 1–9 (2010).
- [20] <http://www.e-mid.it/>.
- [21] Iori, G., De Masi, G., Precup, O. V., Gabbi, G. & Caldarelli, G. A network analysis of the Italian overnight money market. *J. Econ. Dyn. Control* **32**, 259–278 (2008).
- [22] Beaupain, R. & Durré, A. The interday and intraday patterns of the overnight market: Evidence from an electronic platform. *ECB Working Paper* **988** (2008).
- [23] Kobayashi, T. & Takaguchi, T. Significant ties: Identifying relationship lending in temporal inter-bank networks. arXiv:1708.08594 (2017).
- [24] Bro, R. & Kiers, H. A. A new efficient method for determining the number of components in PARAFAC models. *J. Chemom.* **17**, 274–286 (2003).
- [25] Kolda, T. G. & Bader, B. W. Tensor decompositions and applications. *SIAM Rev.* **51**, 455–500 (2009).
- [26] Gai, P. & Kapadia, S. Contagion in financial networks. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **466**, 2401–2423 (2010).
- [27] Brummitt, C. D. & Kobayashi, T. Cascades in multiplex financial networks with debts of different seniority. *Phys. Rev. E* **91**, 062813 (2015).
- [28] Battiston, S., Caldarelli, G., May, R. M., Roukny, T. & Stiglitz, J. E. The price of complexity in financial networks. *Proc. Natl. Acad. Sci. USA* **113**, 10031–10036 (2016).
- [29] Lim, L.-H. & Comon, P. Nonnegative approximations of nonnegative tensors. *J. Chemom.* **23**, 432–441 (2009).

- [30] Kim, J. & Park, H. Fast nonnegative tensor factorization with an active-set-like method. In *High-Performance Scientific Computing*, 311–326 (Springer, 2012).
- [31] Bader, B. W., Kolda, T. G. *et al.* Matlab Tensor Toolbox version 2.6. (2015). URL <http://www.sandia.gov/~tgkolda/TensorToolbox/>.
- [32] Bader, B. W. & Kolda, T. G. Efficient MATLAB computations with sparse and factored tensors. *SIAM J. Sci. Comput.* **30**, 205–231 (2007).
- [33] <https://sites.google.com/site/jingukim/>.
- [34] <http://www.cs.ucr.edu/~epapalex/code.html>.
- [35] Papalexakis, E. E. & Faloutsos, C. Fast efficient and scalable Core Consistency diagnostic for the PARAFAC decomposition for big sparse tensors. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 5441–5445 (IEEE, 2015).
- [36] De Masi, G., Iori, G. & Caldarelli, G. Fitness model for the Italian interbank money market. *Phys. Rev. E* **74**, 066112 (2006).
- [37] <https://fred.stlouisfed.org/series/CSUSHPINSA>, accessed Feb 19, 2018.
- [38] Brunnermeier, M. K. Deciphering the liquidity and credit crunch 2007–2008. *J. Econ. Perspect.* **23**, 77–100 (2009).
- [39] Allen, F. & Carletti, E. An overview of the crisis: Causes, consequences, and solutions. *Intl. Rev. Finance* **10**, 1–26 (2010).
- [40] Beale, N. *et al.* Individual versus systemic risk and the regulator’s dilemma. *Proc. Natl. Acad. Sci. USA* **108**, 12647–12652 (2011).
- [41] Gai, P., Haldane, A. & Kapadia, S. Complexity, concentration and contagion. *J. Monet. Econ.* **58**, 453–470 (2011).
- [42] Kobayashi, T. & Hasui, K. Efficient immunization strategies to prevent financial contagion. *Sci. Rep.* **4**, 3834 (2014).

Acknowledgements

TK acknowledges financial support from the Japan Society for the Promotion of Science Grants no. 15H05729 and 16K03551. EF is grateful to DARPA for support (grant #D16AP00115). This project does not necessarily reflect the position/policy of the Government; no official endorsement should be inferred. Approved for public release; unlimited distribution.

Author contributions

TK and EF conceived the research and wrote the manuscript. TK performed the analysis. TK, AS and EF discussed the results and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Supplementary Information

“Extracting the multi-timescale activity patterns of online financial markets”

Teruyoshi Kobayashi, Anna Sapienza and Emilio Ferrara

S1 Intra- and inter-day activity of banks

The r -th rows of factor matrices $\mathbf{B}$ and $\mathbf{C}$ (i.e., $\mathbf{b}_r$ and $\mathbf{c}_r$) respectively represent the intra- and inter-day activities of component r . Fig. S1 illustrates intra- and inter-day activities for a given temporal resolution $\Delta \in \{5, 30, 45\}$. The selected rank size, based on the Core-Consistency measure, is $R = 4$ for $\Delta = 5$, $R = 3$ for $\Delta = 30$, and $R = 2$ for $\Delta = 45$. We note that the choice of Δ does not affect the activity distributions for a fixed value of R .

(a) $\Delta = 5$

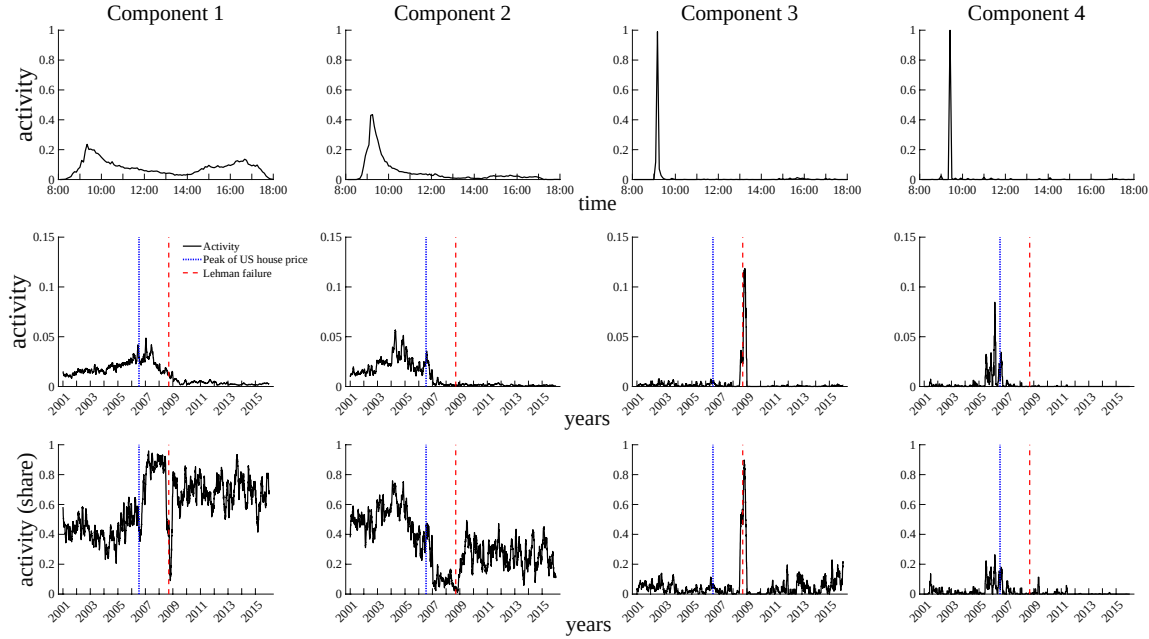
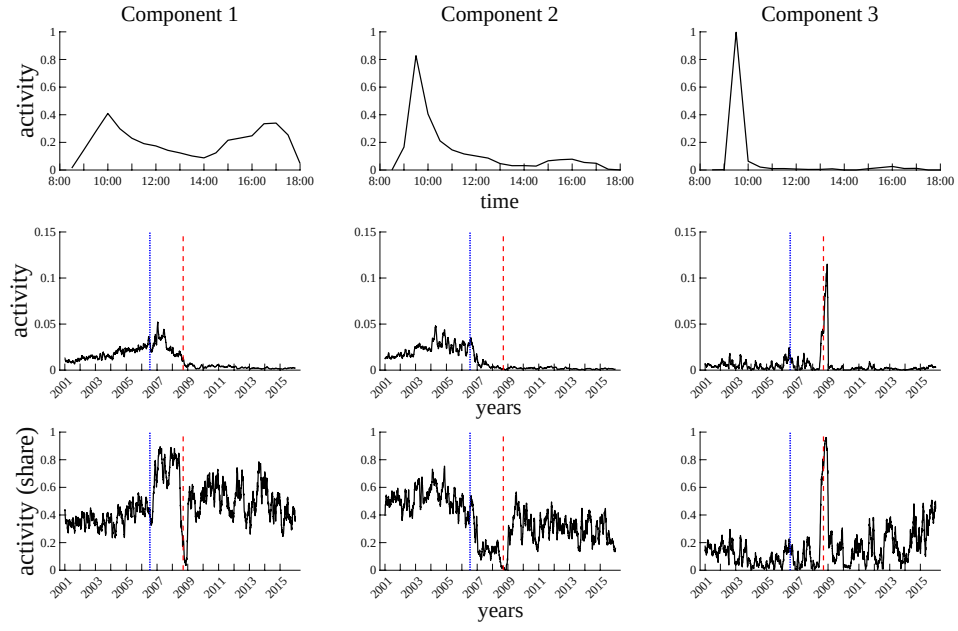


Figure S1: Intra- and inter-day activities. Upper: Intra-day activity. Middle: Inter-day activity. Lower: Share of a component in the daily activity. (a) $\Delta = 5$, (b) $\Delta = 30$, and (c) $\Delta = 45$. See the caption of Fig. 6.

(b) $\Delta = 30$



(c) $\Delta = 45$

