scientific reports

Check for updates

OPEN ## Collective dynamics of stock market efficiency

Luiz G. A. Alves [1], Higor Y. D. Sigaki [2], Matjaž Perc [3,4,5✉] & Haroldo V. Ribeiro [2]

Summarized by the efficient market hypothesis, the idea that stock prices fully reflect all available information is always confronted with the behavior of real-world markets. While there is plenty of evidence indicating and quantifying the efficiency of stock markets, most studies assume this efficiency to be constant over time so that its dynamical and collective aspects remain poorly understood. Here we define the time-varying efficiency of stock markets by calculating the permutation entropy within sliding time-windows of log-returns of stock market indices. We show that major world stock markets can be hierarchically classified into several groups that display similar long-term efficiency profiles. However, we also show that efficiency ranks and clusters of markets with similar trends are only stable for a few months at a time. We thus propose a network representation of stock markets that aggregates their short-term efficiency patterns into a global and coherent picture. We find this financial network to be strongly entangled while also having a modular structure that consists of two distinct groups of stock markets. Our results suggest that stock market efficiency is a collective phenomenon that can drive its operation at a high level of informational efficiency, but also places the entire system under risk of failure.

The efficient market hypothesis is a paradigm in financial economics and a widespread belief among stock market agents[1,2]. The hypothesis states that asset prices fully reflect all available information in an ideally efficient stock market[3]. As new information about individual stocks or the stock market becomes available, it is immediately priced by the stock market agents producing new values for the assets[4]. This idea is frequently associated with the concept of a random walk, in which future price changes represent random departures from previous prices. Accordingly, attempts to predict future prices in informationally efficient markets are likely no better than a random guess. The efficient market hypothesis thus imposes substantial limitations to financial trading and makes it impossible to surpass efficient markets via arbitrage or trading strategy[4].

While the efficient market hypothesis is a crucial concept in financial economics, real stock markets are not ideally efficient[5]. Stock prices in real-world markets can become auto-correlated during short-term periods[6], corroborating the more holistic idea that making short-term predictions and arbitrage are possible. Another evidence that stock markets are not ideally efficient are the non-Gaussian fluctuations (fat-tailed distributions) of the log-returns of asset prices[7,8], the difficulty of simple random walk models in predicting stock market crashes[9], and the existence of successful trading strategies[10]. Furthermore, while it is believed that efficient markets prevent economic bubbles and crashes[11], it is precisely the long-range correlations present in such events that make them more predictable[12].

Whether the lack of efficiency in stock markets is an opportunity for profit, or whether it represents a systemic risk for financial systems that must be fixed—fact is that the efficient market hypothesis is still an ubiquitous concept among economic agents and academics working with financial data and models. This interest translates into a wealth of works that try to quantify the degree of efficiency in different stock markets[13–21] and new forms of investment such as cryptocurrencies[22–31]. Despite the increasing engagement in better understating different facets of the efficient market hypothesis, a large part of these studies assume that market efficiency remains unchanged throughout the investigated period. Thus, the more realistic possibility of having stock markets with time-varying degree of efficiency remains considerably less explored. This is indeed regrettable since time-dependent efficiency measures allow us to probe for collective motions in the evolution of market efficiency, as well as to quantify the stability of different efficiency ranks in stock markets.

Here we investigate the dynamical behavior of the efficiency of 43 major world stock markets during the past 20 years. We use a physics-inspired approach for defining a time-varying efficiency from log-returns of

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA. [2]Departamento de Física, Universidade Estadual de Maringá, Maringá, PR 87020-900, Brazil. [3]Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia. [4]Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan. [5]Complexity Science Hub Vienna, Josefstädterstraße 39, 1080 Vienna, Austria. ✉email: matjaz.perc@gmail.com
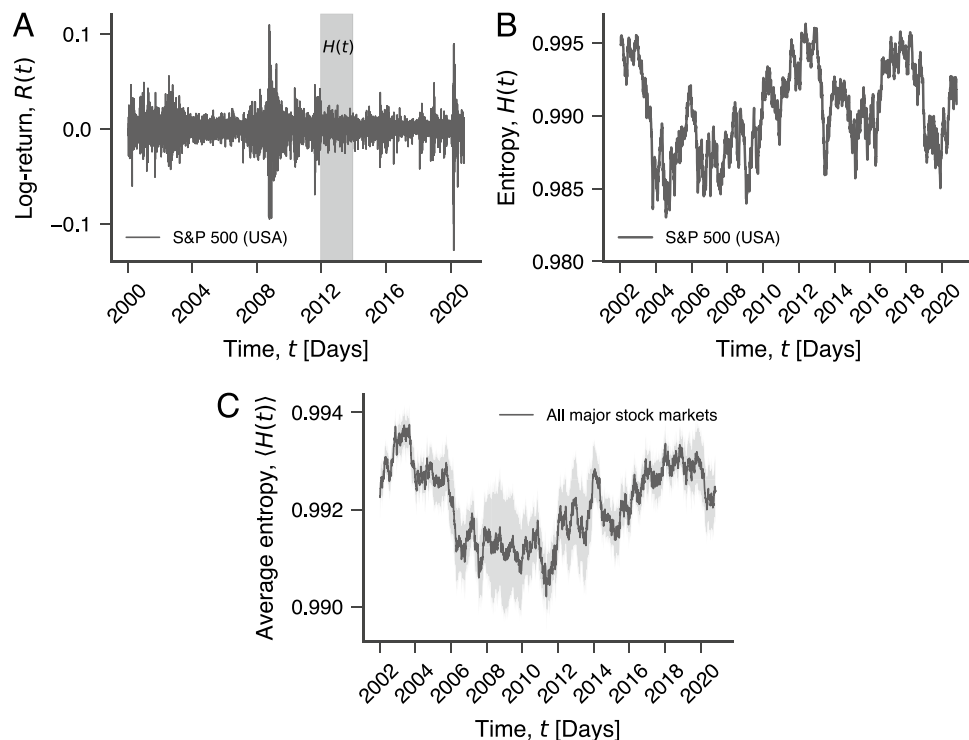
Figure 1. Defining the informational efficiency of stock markets with permutation entropy. (**A**) Log-return time series $R(t)$ of the S&P 500 daily closing prices from January 1, 2000 to October 31, 2020. The shaded area illustrates a 500-day sliding window (2 years of stock market operation) used for calculating the permutation entropy $H(t)$. (**B**) Time evolution of the permutation entropy $H(t)$ with embedding dimension $d = 4$ ("Methods" section for details) of the S&P 500 index. (**C**) Time evolution of the average value of the efficiency $\langle H(t) \rangle$ of all 43 stock markets in our study (shaded band represents the standard error of the mean).

stock market indices. Specifically, we define the time-dependent efficiency of a stock market as the permutation entropy[32] calculated within sliding time-windows of log-returns. Because the permutation entropy estimates the degree of randomness in a time series, we expect this measure to be close to one when a market is in an informationally efficient state. Conversely, values smaller than one indicate a departure from the random behavior such that the smaller the entropy value, the less efficient the stock market.

Our research shows that major world stock markets can be hierarchically classified into several clusters according to the similarity in the long-term evolution of their efficiency degrees. However, we find that the efficiency ranks of stock markets and clusters of markets with similar efficiency trends are only stable during short periods of time, typically lasting no longer than a few months. This result indicates that grouping markets by their long-term behavior may hide important aspects of their interactions with other markets. To unveil such features, we use a dynamical clustering approach that is able to find groups of stock markets with similar efficiency patterns within a time window. By using these time-varying clusters, we construct a weighted network where nodes represent stock markets, edges indicate markets that appear together in the same group at least once, and edge weights are proportional to the number of times a pair of markets appears in the same cluster. This network allows us to identify the most influential markets, as well as their modular structure, which comprises two distinct market groups that have similar efficiency trends. We further observe that the obtained network of stock markets is very dense and entangled. Thus, we conclude that the efficiency dynamics of stock markets is a collective phenomenon that can drive the entire financial system to operate at a very high level of informational efficiency—a level that might be unattainable by any other means—but that also places the entire system under continuous and systemic risk of failure.

Results

Our results are based on the daily closing indices of 43 major world stock markets from January 2000 to October 2020 ("Methods" section for details and Table S1 for the list of market indices). These historical time series represent the stock market indices after adjustments for all applicable splits and dividend distributions (daily adjusted closing prices). From these time series, we estimate the log-returns $R(t)$ ("Methods" section for details) of each market index where t stands for the closing date. Figure 1A illustrates the time evolution of $R(t)$ for the S&P 500 index (an influential indicator of the USA equity market), while Fig. S1 shows the behavior of $R(t)$ for all stock markets in our study. Next, we sample the log-return series with a 500-day sliding window (shaded gray in Fig. 1A), roughly corresponding to 2 years of economic activity. The sliding window moves with daily steps, and
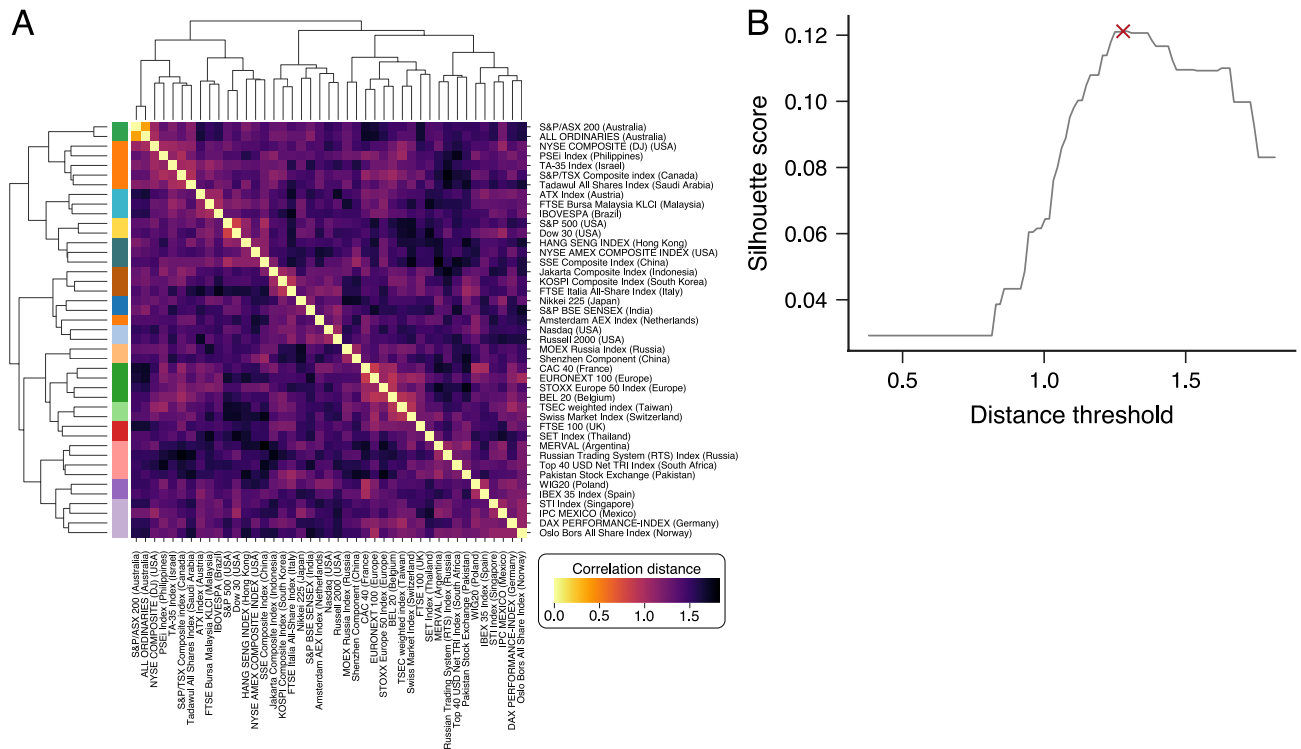
Figure 2. Long-term hierarchical organization of efficiency patterns of major world stock markets. (**A**) Matrix plot of the correlation distance among all pairs of entropy time series of stock markets. The dendrograms that go along this matrix show the hierarchical clustering result based on Ward's minimum variance method. (**B**) Silhouette score calculated from clusters obtained by cutting the dendrogram at different threshold distances. The colored squares located below the dendrogram branches in panel (**A**) indicate the 16 clusters obtained by cutting the dendrogram at the threshold distance that maximizes the silhouette score (indicated by the red cross).

for every step, we estimate the normalized permutation entropy $H(t)$[32]. This procedure creates a time series of the permutation entropy $H(t)$ for each stock market, as shown in Fig. 1B for the S&P 500 (see Fig. S2 for all markets).

As detailed in the "Methods" section, the permutation entropy is estimated from ordering patterns among consecutive values of $R(t)$ and quantifies the degree of randomness in the occurrence of these patterns. We expect an entirely regular series to have $H \approx 0$, while a completely random series displays $H \approx 1$. Thus, the higher the value of $H(t)$, the more random the log-return series is around time t, and so the more informationally efficient is the stock market at that particular time. Conversely, a decrease in $H(t)$ indicates the emergence of a more regular (and possibly more predictable) behavior of the log-return series, and thus a less efficient period of the stock market. We have also estimated the average behavior of the efficiency $\langle H(t) \rangle$ over all stock markets. Figure 1C shows that the aggregate behavior is smoother than the behavior observed for individual stock markets and appears to reflect major financial events such as the "global financial crisis" (2007–2008), as around this period $\langle H(t) \rangle$ displays lower values of entropy.

Investors' behavior tends to synchronize during stock market crashes[5], and investment strategies can propagate shocks through financial networks and lead to the emergence of strong correlations among financial markets[33]. Similarly, we expect these collective behaviors to strongly affect the efficiency dynamics of stock markets and produce joint movements in $H(t)$ that may organize markets into hierarchical structures with similar efficiency trends. To investigate this possibility, we first estimate the correlation distance of the efficiency time series among all pairs of markets ("Methods" section for details), creating the correlation distance matrix shown in Fig. 2A. Next, we use Ward's minimum variance method ("Methods" section for details) to build up a dendrogram representation of the distance matrix, which is also depicted in Fig. 2A. Our results indicate that stock markets form a hierarchical structure regarding their long-term efficiency evolution. However, we do not observe large clustering structures in the distance matrix. Indeed, by determining the number of clusters by maximizing the silhouette score (as shown in Fig. 2B, see "Methods" section for details), we end up with 16 clusters in which 15 consist of only a few markets (the largest cluster consists of 5 markets) and 1 with only a single market. These groups of markets exhibit similar long-term temporal profiles of $H(t)$ (see Fig. S3), but this global analysis does not capture short-term movements of $H(t)$ among stock markets.

To start investigating short-term collective behavior in the efficiency of stock markets, we sample the time series of $H(t)$ with a 1-year sliding window and create a ranking of markets based on the average efficiency within each time-window. Next, we investigate the stability of these efficiency rankings by estimating the Kendall rank correlation coefficient (Kendall-τ) among all possible pairs of time-windows. This analysis results in the correlation matrix shown in Fig. 3A, where rows and columns represent the last date of each time-window. By
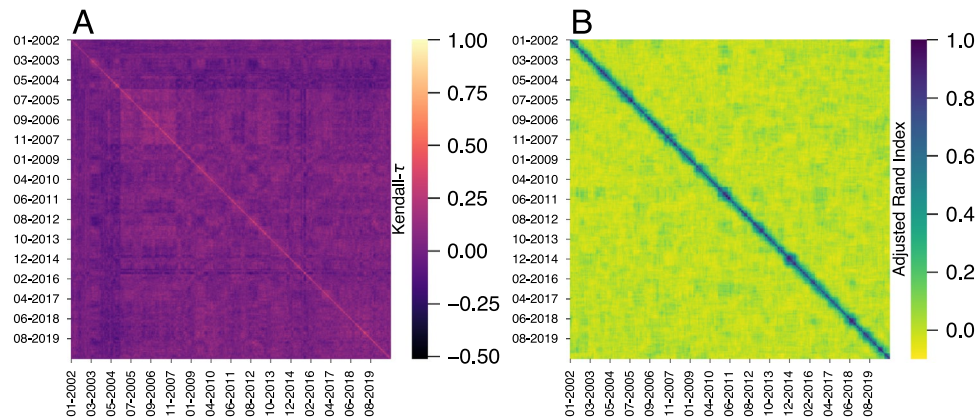
Figure 3. Short-term stability of efficiency rankings and clusterings of stock markets. (**A**) Matrix plot of the Kendall rank correlation coefficient (Kendal-τ) among all pairs of efficiency ranks of stock markets calculated within a 1-year sliding window. We note the formation of small diagonal blocks with widths of 1 or 2 months, indicating that the efficiency ranks change over time. (**B**) Matrix plot of the adjusted rand index estimating the agreement among clustering of markets with similar temporal efficiency profiles at different time-windows. We also observe the existence of small diagonal blocks with widths of approximately 4 months, showing that groups with similar efficiency profiles do not remain stable during long periods.

definition, the diagonal elements of this matrix are unitary (that is, the efficiency rank in a given time window is perfectly correlated to itself). Values along a given row or column indicate how similar is the efficiency rank of that particular date with past and future rankings. Thus, we would expect large diagonal blocks with high Kendal-τ values if these efficiency ranks were stable over long periods. However, we observe small diagonal blocks with about 1-month width, indicating that these efficiency ranks are stable only under very short periods.

We have also calculated the correlation distance of the efficiency $H(t)$ among all pairs of markets for each time-window and applied the same clustering approach used for the entire time series of $H(t)$ (that is, Ward's minimum variance dendrogram segmented at the maximum silhouette score). This approach produces groups of markets with similar efficiency evolution in each time-window. We compare the temporal stability of these groups by estimating the adjusted rand index among all clustering results in each time-window. This coefficient measures the agreement between two clusterings by counting pairs of elements assigned in the same group while controlling for the overlap expected by chance. Values of adjusted rand index around zero indicate that the two clusterings are no more similar than random partitions, while 1 indicates perfect agreement between them. Figure 3B shows the matrix plot of the adjusted rand index for all pairs of clustering results. The small block-diagonal structures of the adjusted rand index matrix indicate that groups of markets with similar $H(t)$ profiles remain stable only for approximately 4 months.

The results of Fig. 3 demonstrate that short-term collective patterns in the efficiency evolution of stock markets change with time and are different from those obtained at long time scales. The dynamical patterns we have found indicate that simple partitions are not enough to capture the complex interactions among financial markets and motivate the use of a different approach that considers these entangled interactions. To do so, we propose to build a complex network where nodes represent the stock markets, and links indicate two markets that clustered together at least once over time. We also assume that the connection between two stock markets is weighted by the number of times those particular markets are grouped in the same cluster. This representation allows us to aggregate the short-term information into a global and coherent picture where stock markets whose efficiency dynamics are correlated during some period are connected; furthermore, the strengths of these connections indicate the intensity of the interactions among the markets.

Figure 4A shows this network representation for the 43 stock markets in our study. We observe that this complex network forms a complete graph as it presents all possible connections among all stock markets. Thus, world stock markets are strongly globalized regarding their efficiency such that simultaneous periods of high or low efficiency may involve a large number of markets. This result suggests the existence of systemic risk for the "spreading" of low-efficient states but, at the same time, it indicates that high-efficient states can also globally emerge. Although the density of this financial network is maximum, the interactions among the stock markets are not uniformly distributed. The Gini coefficient of the edge weights is 0.18 (on a scale where zero means perfect equality and one maximal inequality) and strengthens the idea that some markets may have a higher impact on the efficiency dynamics of the entire system. Figure 4B shows a centrality ranking based on PageRank[34] where the Amsterdam AEX Index (Netherlands) and KOSPI Composite Index (South Korea) emerge as the most influential markets, while the two Russian index (MOEX Russia Index, Russian Trading System (RTS) Index) are the less influential for the efficiency dynamics of the world stock markets.

The inequality in the edge weights distribution also suggests that the financial network of Fig. 4A may have a modular structure in which groups of markets are more similar among themselves than with other groups. To probe for this possible modular structure, we use the stochastic block model approach[35–37]. As detailed in "Methods" section, we have fitted different stochastic block models to our network data, finding that the nested
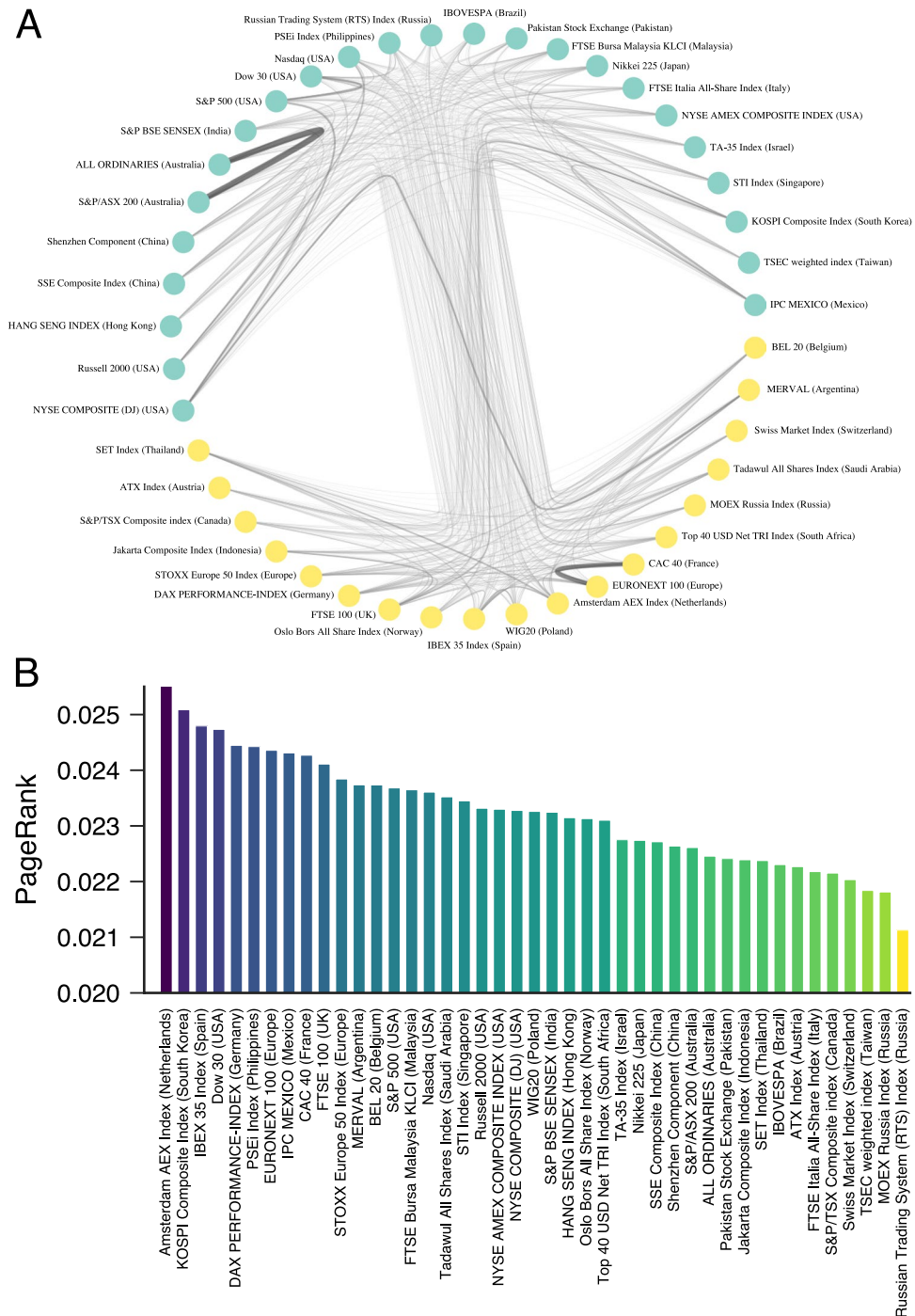
Figure 4. Financial network of stock markets exhibiting similar short-term trends of efficiency. (**A**) Nodes represent stock markets, and links are drawn among markets appearing at least once in the same cluster regarding the short-term similarity in their evolution of efficiency. These links are further weighted by the number of times pairs of markets are grouped in the same cluster. By using the nested stochastic block model ("Methods" section for details), we identify two network modules indicated by the different node colors. In this visualization edge widths are proportional to the weights. (**B**) Centrality ranking based on PageRank indicating the most influential stock markets for the efficiency dynamics of the stock market indices.

stochastic block model without degree correction is the best model description for our network. This model yields two network modules represented by the different node colors in Fig. 4A.

The largest network module (with 24 indices) comprises markets from USA (6 indices), Asia-Pacific countries (14 indices), and 4 other markets from Brazil, Italy, Israel, and Mexico. The second major module includes the

remaining 19 indices: 12 from Europe, and other markets from Argentina, Canada, Indonesia, Russia, Saudi Arabia, South Africa, and Thailand. Despite the existence of many exceptions, the geographical distance appears to play a role in these partitions. However, more important than understanding the particularities of each module is the emergence of this modular structure. Although the associations among stock markets are quite entangled, this modular structure suggests that some groups of markets are more similar to each other such that low- or high-efficiency states are more likely to encompass these modules.

Discussion

We have presented an investigation of dynamical efficiency patterns of 43 major world stock markets during the past 20 years. To do so, we have relied on a physics-inspired approach in which the efficiency of a stock market at a particular time is defined by estimating the permutation entropy within sliding time-windows of log-returns of stock market indices. Our results indicate that stock markets can be hierarchically organized into groups of markets having similar long-term efficiency trends. However, we have also found that these long-term clusters are not enough to fully understand the collective dynamics of market efficiency. Indeed, our research has revealed that short-term collective patterns in the evolution of efficiency vary with time, and are different from those at longer time scales. We have observed that efficiency ranks of stock markets are stable only during relatively short periods of time, no longer than a month or two. Similarly, the clustering of markets with similar efficiency profiles is stable only for approximately 4 months.

Because of these facts, we have proposed a complex network representation where nodes are stock markets, and connections among them indicate markets that are clustered together at least once during the 20 years of our data. We have further considered that links between a pair of markets are weighted by the number of times they appear in the same cluster regarding their short-term efficiency dynamics. Our results show that this financial network is fully-connected, indicating that the efficiency of stock markets is strongly entangled and globalized. Previous works have already demonstrated that systemic failures in stock markets emerge from a synchronization process[9] that takes place in social networks[38,39] and may cause bubbles that eventually burst[12]. Studies also suggest the existence of strong correlations between fund investment strategies as a cause of systemic risks and shock propagation[33]. In this context, the intricate financial network uncovered by our work suggests a systemic risk for the "spreading" of low-efficiency states but, at the same time, indicates that high-efficiency states can also emerge at a global level. Although our financial network forms a complete graph, the link weights among stock markets are not uniformly distributed, allowing us to identify the most influential markets. Furthermore, we have found this financial network to have a modular structure that comprises two market groups whose efficiency dynamics is more similar within the groups as it is with markets outside of the groups. Therefore, despite stock markets being quite entangled in terms of their efficiency profiles, the modular structure indicates that low-efficient and high-efficient states are more likely to emerge within these groups.

Because efficiency in stock markets can be an opportunity for profit as well as an early signal of an impending financial crisis, it would be interesting to incorporate our approach into models that try to predict stock market prices to measure financial transaction risks. For instance, researches used *Google Trends* to inform investors when to buy or sell stocks and found that their strategy was 326% better than the traditional buy-and-hold strategy[10]. Despite the impressive improvement of this strategy, it does not inform the buying or selling volumes, or the risks that are involved in the proposed transactions. Thus, quantifying the time dependence of the stock efficiency may help economic agents to quantify their transaction risks.

Naturally, our study has its limitations. The correlations among stock markets expressed in our network representation do not carry information about the direction of the influence. It would therefore be interesting to consider other measures that are capable of determining the causality of these associations. While we have used the longest available period in our data set (20 years), further investigations may include longer historical data and probe the effects of financial market age and other features that the financial network carries. Similarly, it would be enlightening to create other financial networks with high-frequency intra-day data, or to use a multi-scale approach. It is also worth noticing that stock market indices comprise the aggregated prices of several stocks, and it would be fascinating to investigate a similar financial network composed of a large number of individual stocks. We believe these limitations open several possibilities for future research, and we hope they will inspire other studies with a goal to better understand financial data with physics-inspired approaches.

Methods

Data.   The data set used in our study was obtained from the *Yahoo!* finance historical data API (via the Python module *yfinance*[40]), the Wall Street Journal market data[41], and *investing.com* (via the Python module *investpy*[42]). We have first collected the ticker symbol of all 43 major stock market indices (Table S1), and next retrieved the adjusted daily closing prices of each one from *Yahoo!* finance in the period from January 1, 2000 to October 31, 2020 (each time series has 5204 data points). The tickers missing or with incomplete data in the *Yahoo!* finance database were then retrieved from the Wall Street Journal markets. For tickers that were not available or remained incomplete with the previous data source, we retrieved data from *investing.com*. Thus, all data necessary to reproduce our findings are freely available.

Permutation entropy.   The permutation entropy[32] is a complexity measure originally proposed for characterizing time series. This measure is calculated from a probability distribution related to local ordering patterns among consecutive time series elements. To define the permutation entropy, we consider a generic time series $\{x_t\}_{t=1,2,\ldots,n}$ and overlapping partitions of length $d > 1$ (the embedding dimension) represented by

$$(\vec{s}\,) \mapsto (x_{s-(d-1)}, x_{s-(d-2)}, \ldots, x_{s-1}, x_s), \tag{1}$$

where $s = d, d+1, \ldots, n$. First, for each one of these $(n-d+1)$ partitions, we investigate the $d!$ permutations $\pi = (r_0, r_1, \ldots, r_{d-1})$ of the symbols $(0, 1, \ldots, d-1)$ defined by the ordering $x_{s-r_{d-1}} \leq x_{s-r_{d-2}} \leq \ldots \leq x_{s-r_0}$. These permutations symbols represent all $d!$ possible ordering patterns among the d elements of the $(\vec{s}\,)$ partitions. Next, we estimate the relative frequency of each one of these $d!$ permutation symbols

$$p(\pi_i) = \frac{\text{the number of } s \text{ that has type } \pi_i}{(n-d+1)}, \tag{2}$$

to define the probability distribution of the ordinal patterns $P = \{p(\pi_i)\}_{i=1,2,\ldots,d!}$. Finally, the permutation entropy[32] is simply the normalized Shannon entropy[43] of P, that is,

$$H(P) = -\frac{1}{\ln(d!)} \sum_{i=1}^{d!} p(\pi_i) \ln p(\pi_i), \tag{3}$$

where $\ln(d!)$ is a normalization constant representing the maximum value of the Shannon entropy.

The embedding dimension d defining the length of the overlapping partitions is the only parameter of the method. In practice, the choice of d is limited by the time series length N such that the condition $d! \ll n$ must hold to have a reliable estimate of $P = \{p(\pi_i)\}$. We have used $d = 4$ in all our results as the length of the sliding window is 500 trading days. By definition, the normalized permutation entropy is bounded to the interval $0 \leq H \leq 1$, where values close to the upper bound ($H \approx 1$) occur for random time series and lower values of H indicate that the time series exhibits a more complex ordering dynamics. Mainly because of its simplicity, discrimination capabilities, and fast computational evaluation, the permutation entropy framework has successfully been used in many applications[44–52].

Efficiency degree of stock markets. To define the informational efficiency $H(t)$ used in our study, we first estimate the logarithmic daily price returns (or log-return) series defined by[53]

$$R(t) = \log P(t) - \log P(t-1), \tag{4}$$

where $\log P(t)$ and $\log P(t-1)$ are the natural logarithm of the closing indices at time t and $t-1$. Next, we sample the log-return series with a 500-day sliding window that moves ahead one trading day at a time. For each of these time-windows, we calculate the normalized permutation entropy $H(t)$; here, t stands for the last date in the time-windows. This procedure defines a new time series representing the permutation entropy (our measure of informational efficiency) in each window $H(t)$ (Fig. 1B). The permutation entropy estimates the degree of randomness in a time series, therefore, entropy values close to one indicate markets at a high informational efficiency state. Conversely, values smaller than one indicate less efficient states of stock markets.

Hierarchical clustering procedure. To compute the similarities in the time evolution of the efficiency degree among stock markets, we use the correlation distance matrix

$$d(H_i, H_j) = \sqrt{2(1 - \rho(H_i, H_j))}, \tag{5}$$

where $\rho(H_i, H_j)$ is the Pearson correlation coefficient between the entropy time series H_i of i-th stock market and H_j the same quantity for j-th stock market. For the dynamical clustering analysis, H_i and H_j represent the time series segments of stock markets i and j obtained by sampling the efficiency time series with a 1-year sliding window. From this correlation distance matrix (Fig. 2A), we use the Ward linkage criteria[54,55] to hierarchically cluster the stock markets (dendrogram in Fig. 2A). This clustering procedure recursively merges pair of clusters that minimally increase within-cluster variance.

The threshold distance used to cut the dendrogram and determine the number of clusters is obtained by maximizing the silhouette score[56]. This coefficient quantifies the consistency of the clustering procedure and is defined by the average value of

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \tag{6}$$

where a_i is the cohesion (the average intra-cluster distance) and b_i is the separation (the average nearest-cluster distance) for the i-th index. The higher the average value of the silhouette for all indices, the better the cluster configuration. We use the Python module *scikit-learn*[57] to compute the silhouette scores and the *SciPy*[58] package to compute the correlation distance matrix.

Stochastic block models. We estimate the modular structure of Fig. 4B by using the stochastic block modeling approach[35–37]. This method has the advantage of directly estimating the marginal probabilities that the network is partitioned by a certain number of groups and the probability that a node belongs to a particular group during the inference process. We have tested different stochastic block models (SBM) to fit our network data: usual SBM, degree-corrected SBM (DCSBM), nested SBM, and nested DCSBM. We have considered the best model as the one with the smallest minimal description length[37], based on the statistical evidence that the model is not mistaking stochastic fluctuations for actual modular structure. This means that groups in the

network could not have arisen from stochastic fluctuations, as they do in fully random graphs[59]. We have found that the best model describing our network is the nested stochastic block model without degree correction (nested SBM, see Table S2). We have further collected the partitions for 10,000 sweeps of a Metropolis-Hastings acceptance-rejection Markov Chain Monte Carlo[60] with multiple moves to sample hierarchical network partitions, at intervals of 10 sweeps. Using these partitions, we estimate the marginal probability of the number of network modules (Fig. S4), reinforcing that a modular structure with two modules is the most likely structure of our financial network. We have also estimated the marginal probabilities of node membership in our financial network (Fig. S5). The results indicate that the vast majority of markets are almost always assigned to the same partition. All SBM models were implemented with the Python library *graph-tool*[61].

References

1. Cootner, P. *The Random Character of Stock Market Prices* (MIT Press, Cambridge, 1964).
2. Fama, E. F. Efficient capital markets: a review of theory and empirical work. *J. Finance* **25**, 383–417. https://doi.org/10.2307/2325486 (1970).
3. Fama, E. F. The behavior of stock-market prices. *J. Bus.* **38**, 34–105 (1965).
4. Malkiel, B. G. The efficient market hypothesis and its critics. *J. Econ. Perspect.* **17**, 59–82. https://doi.org/10.1257/089533003321164958 (2003).
5. Sornette, D. *Why Stock Markets Crash: Critical Events in Complex Financial Systems* (Princeton University Press, Princeton, 2017).
6. Stanley, H. E., Amaral, L. A., Gabaix, X., Gopikrishnan, P. & Plerou, V. Similarities and differences between physics and economics. *Phys. A Stat. Mech. Appl.* **299**, 1–15. https://doi.org/10.1016/S0378-4371(01)00351-X (2001).
7. Gopikrishnan, P., Plerou, V., Amaral, L. A. N., Meyer, M. & Stanley, H. E. Scaling of the distribution of fluctuations of financial market indices. *Phys. Rev. E* **60**, 5305. https://doi.org/10.1103/physreve.60.5305 (1999).
8. Mantegna, R. N. & Stanley, H. E. Scaling behaviour in the dynamics of an economic index. *Nature* **376**, 46–49. https://doi.org/10.1038/376046a0 (1995).
9. Johansen, A., Ledoit, O. & Sornette, D. Crashes as critical points. *Int. J. Theor. Appl. Finance* **3**, 219–255. https://doi.org/10.1142/S0219024900000115 (2000).
10. Preis, T., Moat, H. S. & Stanley, H. E. Quantifying trading behavior in financial markets using Google Trends. *Sci. Rep.* **3**, 1684. https://doi.org/10.1038/srep01684 (2013).
11. Preis, T. & Stanley, H. E. Bubble trouble. *Phys. World* **24**, 29. https://doi.org/10.1088/2058-7058/24/05/34 (2011).
12. Sornette, D. Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes, and human birth. *Proc. Natl. Acad. Sci.* **99**, 2522–2529. https://doi.org/10.1073/pnas.022581999 (2002).
13. Zunino, L., Tabak, B. M., Pérez, D. G., Garavaglia, M. & Rosso, O. A. Inefficiency in Latin-American market indices. *Eur. Phys. J. B* **60**, 111–121. https://doi.org/10.1140/epjb/e2007-00316-y (2007).
14. Zunino, L. *et al.* A multifractal approach for stock market inefficiency. *Phys. A Stat. Mech. Appl.* **387**, 6558–6566. https://doi.org/10.1016/j.physa.2008.08.028 (2008).
15. Zunino, L., Zanin, M., Tabak, B. M., Pérez, D. G. & Rosso, O. A. Forbidden patterns, permutation entropy and stock market inefficiency. *Phys. A Stat. Mech. Appl.* **388**, 2854–2864. https://doi.org/10.1016/j.physa.2009.03.042 (2009).
16. Zunino, L., Zanin, M., Tabak, B. M., Pérez, D. G. & Rosso, O. A. Complexity-entropy causality plane: a useful approach to quantify the stock market inefficiency. *Phys. A Stat. Mech. Appl.* **389**, 1891–1901. https://doi.org/10.1016/j.physa.2010.01.007 (2010).
17. Zunino, L., Bariviera, A. F., Guercio, M. B., Martinez, L. B. & Rosso, O. A. On the efficiency of sovereign bond markets. *Phys. A Stat. Mech. Appl.* **391**, 4342–4349. https://doi.org/10.1016/j.physa.2012.04.009 (2012).
18. Szarek, D., Bielak, L. & Wyłomańska, A. Long-term prediction of the metals' prices using non-Gaussian time-inhomogeneous stochastic process. *Phys. A Stat. Mech. Appl.* **555**, 124659. https://doi.org/10.1016/j.physa.2020.124659 (2020).
19. Wang, L. & Li, L. Long-range correlation and predictability of Chinese stock prices. *Phys. A Stat. Mech. Appl.* **549**, 124384. https://doi.org/10.1016/j.physa.2020.124384 (2020).
20. Filho, T. M. R. & Rocha, P. M. Evidence of inefficiency of the Brazilian stock market: the IBOVESPA future contracts. *Phys. A Stat. Mech. Appl.* **543**, 123200. https://doi.org/10.1016/j.physa.2019.123200 (2020).
21. Sánchez-Granero, M., Balladares, K., Ramos-Requena, J. & Trinidad-Segovia, J. Testing the efficient market hypothesis in Latin American stock markets. *Phys. A Stat. Mech. Appl.* **540**, 123082. https://doi.org/10.1016/j.physa.2019.123082 (2020).
22. Urquhart, A. The inefficiency of Bitcoin. *Econ. Lett.* **148**, 80–82. https://doi.org/10.1016/j.econlet.2016.09.019 (2016).
23. Bariviera, A. F., Basgall, M. J., Hasperué, W. & Naiouf, M. Some stylized facts of the Bitcoin market. *Phys. A Stat. Mech. Appl.* **484**, 82–90. https://doi.org/10.1016/j.physa.2017.04.159 (2017).
24. Zhang, W., Wang, P., Li, X. & Shen, D. The inefficiency of cryptocurrency and its cross-correlation with Dow Jones industrial average. *Phys. A Stat. Mech. Appl.* **510**, 658–670. https://doi.org/10.1016/j.physa.2018.07.032 (2018).
25. Bariviera, A. F. The inefficiency of Bitcoin revisited: a dynamic approach. *Econ. Lett.* **161**, 1–4. https://doi.org/10.1016/j.econlet.2017.09.013 (2017).
26. Nadarajah, S. & Chu, J. On the inefficiency of Bitcoin. *Econ. Lett.* **150**, 6–9. https://doi.org/10.1016/j.econlet.2016.10.033 (2017).
27. Tiwari, A. K., Jana, R., Das, D. & Roubaud, D. Informational efficiency of Bitcoin—an extension. *Econ. Lett.* **163**, 106–109. https://doi.org/10.1016/j.econlet.2017.12.006 (2018).
28. Bariviera, A. F., Zunino, L. & Rosso, O. A. An analysis of high-frequency cryptocurrencies prices dynamics using permutation-information-theory quantifiers. *Chaos* **28**, 075511. https://doi.org/10.1063/1.5027153 (2018).
29. Alvarez-Ramirez, J., Rodriguez, E. & Ibarra-Valdez, C. Long-range correlations and asymmetry in the Bitcoin market. *Phys. A Stat. Mech. Appl.* **492**, 948–955. https://doi.org/10.1016/j.physa.2017.11.025 (2018).
30. Sigaki, H. Y., Perc, M. & Ribeiro, H. V. Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market. *Sci. Rep.* **9**, 1–9. https://doi.org/10.2307/2325486 (2019).
31. Dimitrova, V., Fernández-Martínez, M., Sánchez-Granero, M. & Trinidad Segovia, J. Some comments on Bitcoin market (in)efficiency. *PLoS One* **14**, e0219243. https://doi.org/10.1371/journal.pone.0219243 (2019).
32. Bandt, C. & Pompe, B. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* **88**, 174102. https://doi.org/10.1103/PhysRevLett.88.174102 (2002).
33. Delpini, D., Battiston, S., Caldarelli, G. & Riccaboni, M. Systemic risk from investment similarities. *PLoS One* **14**, e0217141. https://doi.org/10.1371/journal.pone.0217141 (2019).

34. Brin, S. & Page, L. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Comput. Netw.* **56**, 3825–3833. https://doi.org/10.1016/j.comnet.2012.10.007 (2012).
35. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: first steps. *Soc. Netw.* **5**, 109–137. https://doi.org/10.1016/0378-8733(83)90021-7 (1983).
36. Funke, T. & Becker, T. Stochastic block models: a comparison of variants and inference methods. *PLoS One* **14**, e0215296. https://doi.org/10.1371/journal.pone.0215296 (2019).
37. Peixoto, T. P. Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X* **4**, 011047. https://doi.org/10.1103/PhysRevX.4.011047 (2014).
38. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591. https://doi.org/10.1103/RevModPhys.81.591 (2009).
39. Perc, M. *et al.* Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51. https://doi.org/10.1016/j.physrep.2017.05.004 (2017).
40. Aroussi, R. *et al.* Yahoo! Finance market data downloader (2020). https://github.com/ranaroussi/yfinance.
41. Journal, W. S. Wall Street Journal market data (2020). https://www.wsj.com/market-data.
42. del Canto, A. B. investpy—financial data extraction from Investing.com with Python (2020). https://github.com/alvarobartt/investpy.
43. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x (1948).
44. Ribeiro, H. V., Zunino, L., Mendes, R. S. & Lenzi, E. K. Complexity-entropy causality plane: a useful approach for distinguishing songs. *Phys. A Stat. Mech. Appl.* **391**, 2421–2428. https://doi.org/10.1016/j.physa.2011.12.009 (2012).
45. Zanin, M., Zunino, L., Rosso, O. A. & Papo, D. Permutation entropy and its main biomedical and econophysics applications: a review. *Entropy* **14**, 1553–1577. https://doi.org/10.3390/e14081553 (2012).
46. Li, Q. & Zuntao, F. Permutation entropy and statistical complexity quantifier of nonstationarity effect in the vertical velocity records. *Phys. Rev. E* **89**, 012905. https://doi.org/10.1103/PhysRevE.89.012905 (2014).
47. Jovanovic, T., García, S., Gall, H. & Mejía, A. Complexity as a streamflow metric of hydrologic alteration. *Stoch. Environ. Res. Risk Assess.* **31**, 2107–2119. https://doi.org/10.1007/s00477-016-1315-6 (2017).
48. Stosic, T., Telesca, L., de SouzaFerreira, D. V. & Stosic, B. Investigating anthropically induced effects in streamflow dynamics by using permutation entropy and statistical complexity analysis: A case study. *J. Hydrol.* **540**, 1136–1145. https://doi.org/10.1016/j.jhydrol.2016.07.034 (2016).
49. Antonelli, A. P., Meschino, G. J. & Ballarin, V. L. Permutation entropy: texture characterization in images. In *Information Processing and Control (RPIC), 2017 XVII Workshop on*, 1–7. https://doi.org/10.23919/RPIC.2017.8211650 (IEEE, 2017).
50. Antonelli, A., Meschino, G. & Ballarin, V. Mammographic density estimation through permutation entropy. In *Congress on Medical Physics and Biomedical Engineering 2018*, **135–141** (Springer, 2018). https://doi.org/10.1007/978-981-10-9035-6_24.
51. Sigaki, H. Y. D., Perc, M. & Ribeiro, H. V. History of art paintings through the lens of entropy and complexity. *Proc. Natl. Acad. Sci.* **115**, E8585–E8594. https://doi.org/10.1073/pnas.1800083115 (2018).
52. Sigaki, H. Y. D., de Souza, R. F., de Souza, R. T., Zola, R. S. & Ribeiro, H. V. Estimating physical properties from liquid crystal textures via machine learning and complexity-entropy methods. *Phys. Rev. E* **99**, 013311. https://doi.org/10.1103/PhysRevE.99.013311 (2019).
53. Mantegna, R. N. & Stanley, H. E. *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 1999).
54. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2013).
55. Ward, J. H. Jr. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244. https://doi.org/10.1080/01621459.1963.10500845 (1963).
56. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7 (1987).
57. Pedregosa, F. *et al.* Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
58. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272. https://doi.org/10.1038/s41592-019-0686-2 (2020).
59. Guimerà, R. & Sales-Pardo, M. Missing and spurious interactions and the reconstruction of complex networks. *Proc. Natl. Acad. Sci.* **106**, 22073–22078. https://doi.org/10.1073/pnas.0908366106 (2009).
60. Peixoto, T. P. Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models. *Phys. Rev. E* **89**, 012804. https://doi.org/10.1103/PhysRevE.89.012804 (2014).
61. Peixoto, T. P. The graph-tool Python library. figshare. https://doi.org/10.6084/m9.figshare.1164194 (2014).

Acknowledgements

Author contributions

L.G.A.A., H.Y.D.S., M.P, and H.V.R. designed research, performed research, analyzed data, and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-78707-2.

Correspondence and requests for materials should be addressed to M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.