

SCIENTIFIC REPORTS



OPEN

Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market

Higor Y. D. Sigaki¹, Matjaž Perc^{2,3} & Haroldo V. Ribeiro¹

The efficient market hypothesis has far-reaching implications for financial trading and market stability. Whether or not cryptocurrencies are informationally efficient has therefore been the subject of intense recent investigation. Here, we use permutation entropy and statistical complexity over sliding time-windows of price log returns to quantify the dynamic efficiency of more than four hundred cryptocurrencies. We consider that a cryptocurrency is efficient within a time-window when these two complexity measures are statistically indistinguishable from their values obtained on randomly shuffled data. We find that 37% of the cryptocurrencies in our study stay efficient over 80% of the time, whereas 20% are informationally efficient in less than 20% of the time. Our results also show that the efficiency is not correlated with the market capitalization of the cryptocurrencies. A dynamic analysis of informational efficiency over time reveals clustering patterns in which different cryptocurrencies with similar temporal patterns form four clusters, and moreover, younger currencies in each group appear poised to follow the trend of their 'elders'. The cryptocurrency market thus already shows notable adherence to the efficient market hypothesis, although data also reveals that the coming-of-age of digital currencies is in this regard still very much underway.

The efficient market hypothesis in financial economics asserts that asset prices fully reflect all available information¹. A market is thus said to be informationally efficient if asset prices are always fully up-to-date, such that any new information about a particular firm is certain and immediately priced into its stock. This of course has important implications for financial trading in that it is impossible to beat an informationally efficient market with expert stock selection or arbitrage. Moreover, the efficient market hypothesis effectively prevents economic bubbles, which are amongst the main culprits behind stock crashes and market instability^{2,3}. On the other hand, critics of the efficient market hypothesis claim that it is precisely the (false) belief in rational markets that is to blame for the 07'-08' financial crisis, as well as for many others unwanted developments in the World economy⁴.

Whether the efficient market hypothesis is the Holy Grail of financial economics, or whether it is just a simplification of reality that does not always hold but is still valid for most investment purposes – fact is that it is an immensely important concept that, even half a century since its inception, still critically shapes today's economic thought. Not surprisingly, recent research has been hard at work in probing whether the newly introduced digital currencies stand the test of the efficient market hypothesis⁵⁻¹² (as well as other questions¹³⁻¹⁷). One of the earliest publications in this direction by Urquhart⁵ reported that Bitcoin returns are significantly inefficient over the whole studied sample, but when split into two subsample periods, tests indicated that efficiency is there in the second period. Subsequently, Bariviera⁸ arrived at similar conclusions using a dynamic approach, which revealed that the Bitcoin market seems to be more informationally efficient from 2014 onwards. Informational efficiency of the Bitcoin market was also confirmed by Nadarajah and Chu⁹ and by Tiwari *et al.*¹⁰.

The growing popularity of cryptocurrencies, despite volatile prices, indicates that decentralized control through the blockchain technology, along with secure financial transactions due to strong cryptography, are highly valued among customers worldwide. A better overall understanding of the digital currency market is therefore much needed. To that effect, we here conduct a large-scale efficiency analysis of this market, analyzing the daily price time series of 437 cryptocurrencies. Methodologically, we rely on the permutation entropy¹⁸ and statistical complexity^{19,20} over sliding time-windows of price log returns. Past research has shown that such a

¹Departamento de Física, Universidade Estadual de Maringá, Maringá, PR 87020-900, Brazil. ²Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, SI-2000, Maribor, Slovenia. ³Complexity Science Hub Vienna, Josefstädterstraße 39, A-1080, Vienna, Austria. Correspondence and requests for materials should be addressed to M.P. (email: matjaz.perc@uni-mb.si)

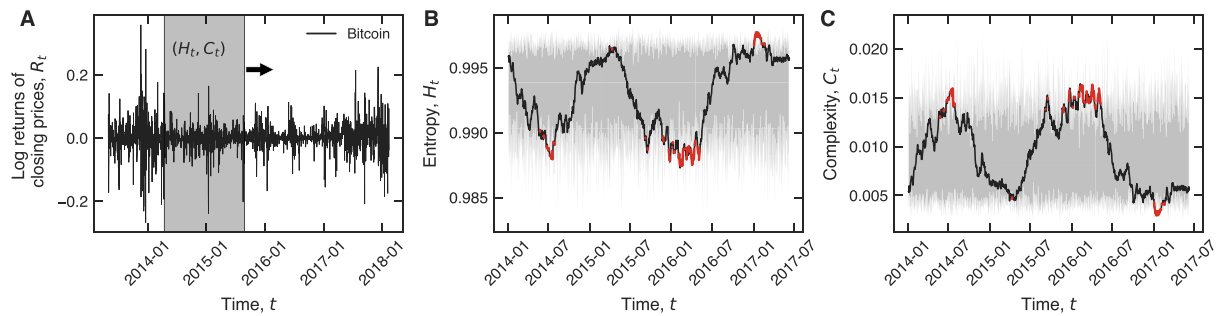


Figure 1. Quantifying the informational efficiency of crypto-assets with permutation entropy and statistical complexity. **(A)** Time series of log returns R_t of the closing prices for Bitcoin (BTC) since April 28, 2013. The shaded area shows a 500-day (about two years of economic activity) sliding window, which moves forward by 1 data point. The black curves in panels **(B,C)** show the time evolution of the permutation entropy H_t and statistical complexity C_t , estimated within the sliding window, respectively. The embedding dimension used was $d = 4$ (Methods Section for details). The shaded areas represent the 95% random confidence intervals obtained by shuffling the data in each window and calculating the values of the entropy and complexity over several independent realizations. The red segments indicate the values of H_t and C_t that are outside the random confidence band. We define the overall informational efficiency E as the fraction of time (days) that both complexity measures stay within the 95% random confidence band. The estimated efficiency for Bitcoin is $E \approx 0.85$ in the period under analysis.

physics-inspired approach works very well on economic data²¹, in particular for quantifying the stock market inefficiency^{22–26}. Once determining the permutation entropy and statistical complexity, we consider that a cryptocurrency is informationally efficient within a time-window when both measures are within the 95% random confidence interval. We obtain the latter by shuffling the time series in each window and calculating the permutation entropy and statistical complexity for several independent realizations.

As we shall show in what follows, our research reveals that 37% of the 437 cryptocurrencies are informationally efficient in over 80% of the time, while 20% stay efficient in less than 20% of the time. We also find that efficiency is not correlated with the market capitalization of the cryptocurrencies. Moreover, a dynamic analysis of informational efficiency over time reveals clustering patterns in which different cryptocurrencies are grouped together due to their similar temporal profiles. For the clustering analysis, we rely on dynamic time warping²⁷, which has the advantage that distances between time series of different lengths and scale of values can still be determined. Based on this analysis, we find four groups of cryptocurrencies: (i) those that begin at a higher efficiency level but evolve to a lower one (12% of market); (ii) those that improve the efficiency over time (19% of market); (iii) those displaying a nearly constant efficiency level (43% of market); and those that start at a higher efficiency level, decrease to a lower level, and then increase their efficiency level (26% of market). This clustering analysis also indicates that younger currencies in each group appear poised to follow the trends of their 'elders'. Overall, we thus find that large parts of the cryptocurrency market satisfy the efficient market hypothesis most of the time, but also that the coming-of-age of digital currencies is in this respect still in progress.

Results

Our results are based on data comprising the daily closing prices and market capitalization of 437 cryptocurrencies, including the most popular and important crypto-assets in circulation today (Methods Section for details). We begin by presenting our methodology on the Bitcoin (BTC) time series, arguably still the most well-known and popular cryptocurrency. As shown in Fig. 1A, the log return R_t of the closing price time series is sampled with a sliding window (shaded gray) comprising 500 data points, which corresponds to approximately two years of economic activity. The sliding window moves ahead one day at a time, and for each, we determine the permutation entropy¹⁸ H_t and the statistical complexity^{19,20} C_t , as shown in Fig. 1B,C, respectively (Methods Section for details). These two complexity measures are estimated from the local ordering patterns among consecutive values of R_t . The permutation entropy measures the degree of randomness in the occurrence of these patterns, ranging from $H \approx 0$ for a completely regular series to $H \approx 1$ for a completely random series. The statistical complexity, in turn, quantifies the structural complexity in this ordering dynamics: $C \approx 0$ in both extremes of order and disorder, whereas $C > 0$ when the ordinal patterns occur in a more complex fashion.

We consider that a cryptocurrency adheres to the efficient market hypothesis when the values of H_t and C_t within a time-window cannot be distinguished from those obtained by chance. To determine the previous condition, we calculate the 95% random confidence interval (shaded gray in Fig. 1B,C) by shuffling the data in each time-window and calculating H and C for several independent realizations (Methods Section for details). Finally, we define the overall informational efficiency E of a cryptocurrency as the fraction of time at which H_t and C_t are simultaneously within the 95% random confidence band. The values of E thus range from zero to one, with the lower bound indicating a very low-efficiency cryptocurrency, while the upper bound represents a high-efficiency cryptocurrency. It is worth noting that this definition maps well the ideas underlying the efficient market hypothesis, in the sense that the price of a crypto-asset having a high value of E is expected to be very robust against profitable trading strategies, while the price of a cryptocurrency having a low value of E is more likely to be predicted and vulnerable to betting strategies. For the Bitcoin market, we find $E \approx 0.85$, indicating that this crypto-asset is

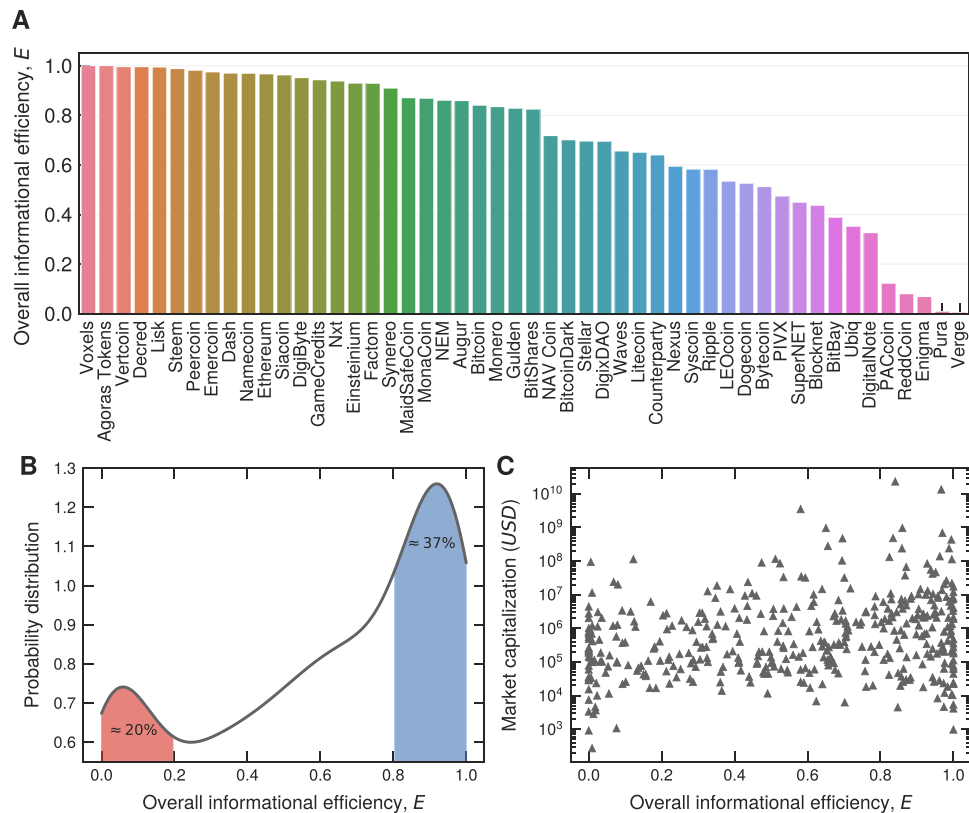


Figure 2. The overall efficiency of the cryptocurrency market and its detachment from the market capitalization mean. **(A)** Overall efficiency E for the top-50 largest cryptocurrencies according to the market capitalization mean. **(B)** Kernel density estimation of the probability distribution function of the informational efficiency E . We note that this distribution is bimodal. The first peak (for E smaller than 0.2) contains $\approx 20\%$ of the cryptocurrencies (red shaded area), indicating the most informationally inefficient cases, whereas in the second peak (for E greater than 0.8) we have $\approx 37\%$ of the cryptocurrencies (blue shaded area), the most efficient cases. **(C)** Scatter plot depicting the values of informational efficiency E versus the market capitalization mean in a linear-log representation. We observe no correlation between these variables, indicating a detachment of informational efficiency from the market capitalization mean. The Pearson linear correlation coefficient is ≈ 0.07 and the 2-tailed p -value is $\approx 0.16\%$, indicating that the null hypothesis of no linear correlation cannot be rejected. For this analysis, we have considered all the 437 cryptocurrencies with more than 600 observations of R_t .

remarkably adherent to the efficient market hypothesis, much in line with preceding research^{9,10}. This in turn also validates our approach and invites a large-scale analysis along the same lines.

To that effect, we calculate the overall efficiency E for all the 437 cryptocurrencies in our dataset. Figure 2A shows this quantity for the top-50 largest cryptocurrencies according to the market capitalization mean, while Fig. 2B shows the probability distribution of E , as obtained for all the 437 cryptocurrencies in our dataset. It can be observed that this distribution features a strongly pronounced bimodal shape such that the first peak contains $\approx 20\%$ of the cryptocurrencies, while the second peak contains $\approx 37\%$. The first peak corresponds to the most informationally inefficient cryptocurrencies, while the second peak corresponds to the most informationally efficient cryptocurrencies. Evidently, there are significantly more informationally efficient than inefficient digital currencies on the market. Nevertheless, a hard case for a pervading adherence to the efficient market hypothesis would still be quite far-fetched.

To further corroborate our conclusions thus far, we show in Fig. 2C how the mean of the market capitalization depends on the informational efficiency E . The first thing to notice is that there is hardly any correlation between market capitalization mean and informational efficiency. Indeed, a statistical analysis reveals that the Pearson linear correlation coefficient is ≈ 0.07 , and the 2-tailed p -value is ≈ 0.16 . Accordingly, it is impossible to reject the null hypothesis of no linear correlation. Since the market capitalization determines the market value of the cryptocurrency, i.e., the number of crypto-coins multiplied by their current price on the market, we thus find that the adherence of the cryptocurrency market to the efficient market hypothesis is hardly dependent on the volatile price variations in recent years.

We now focus on a fine-grained view of the dynamical behavior of the informational efficiency. To do so, we select the 167 cryptocurrencies having more than 460 observations for H_t and C_t . From these time series, we define the time-dependent informational efficiency E_t by using a 360 data points sliding window which moves

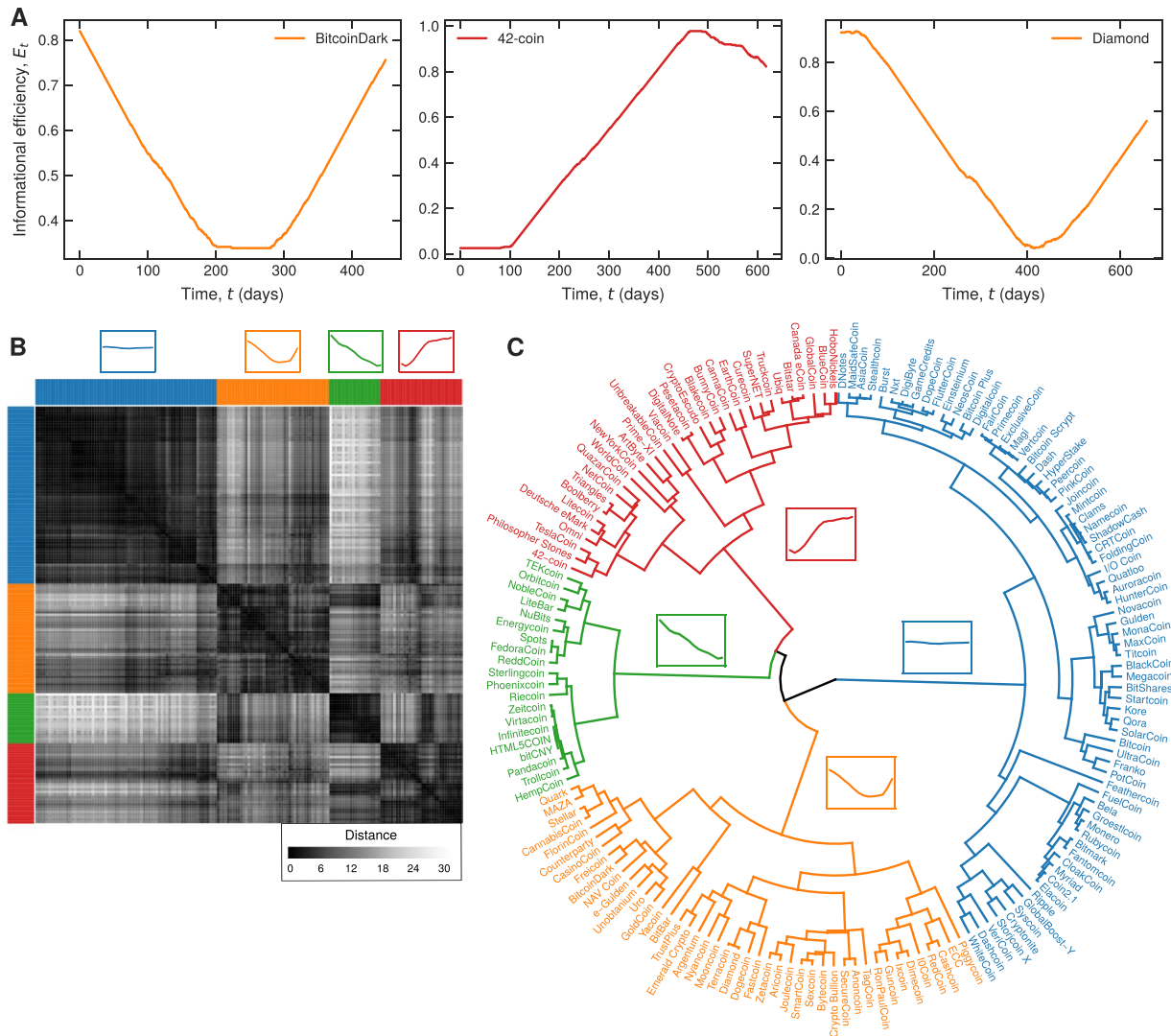


Figure 3. Clustering patterns in the dynamics of informational efficiency. **(A)** Three examples of the time evolution of informational efficiency E_t . The dynamics of E_t is obtained by using a 360-day sliding window over the time series H_t and C_t and by calculating the fraction of points (days) that both complexity measures are within the 95% random confidence interval. We note that the motifs of E_t look very similar for BitcoinDark (BTCD) and Diamond (DMD), while 42-coin (42) displays a different trajectory for E_t . **(B)** Matrix plot of the dynamic time warping (DTW) distance among all pairs of the 167 cryptocurrencies having more than 460 observations for H_t and C_t . **(C)** Dendrogram showing the result of the hierarchical clustering based on the DTW distances and using the average linkage criteria. The colored branches indicate the four groups of cryptocurrencies exhibiting similar motifs for the dynamics of E_t . These groups are obtained by cutting the dendrogram at the threshold distance that maximizes the silhouette coefficient (Methods Section for details). The order of rows and columns in the matrix plot **(B)** is the same used in the dendrogram, and the colored insets of that panel represent the average pattern of E_t for each group.

forward one day at a time and calculating the fraction of days at which H_t and C_t are simultaneously within the 95% random confidence interval. These new time series E_t thus provide a detailed view about how the informational efficiency of cryptocurrencies changes over time, which in turn gives us a chance to find those currencies with similar temporal characteristics. Figure 3A shows three examples of the time evolution of informational efficiency E_t for BitcoinDark (BTCD), 42-coin (42) and Diamond (DMD). This small sample already indicates that the shape of E_t can be very similar among some cryptocurrencies such as for BitcoinDark and Diamond. To extend this comparative analysis to all 167 cryptocurrencies, we use the dynamic time warping algorithm²⁷ (DTW, Methods Section for details) for measuring similarity among all possible pairs of E_t time series. To state briefly, DTW is a distance-like and shape-based similarity measure between two time series that can be used to compare time series of different lengths. For the examples of Fig. 3A, the DTW distance between BitcoinDark and Diamond is 3.64, while the distance between 42-coin and these two cryptocurrencies are 7.73 and 11.72 respectively. Thus, the closer the DTW distance between a pair of cryptocurrencies, the more similar is the profile

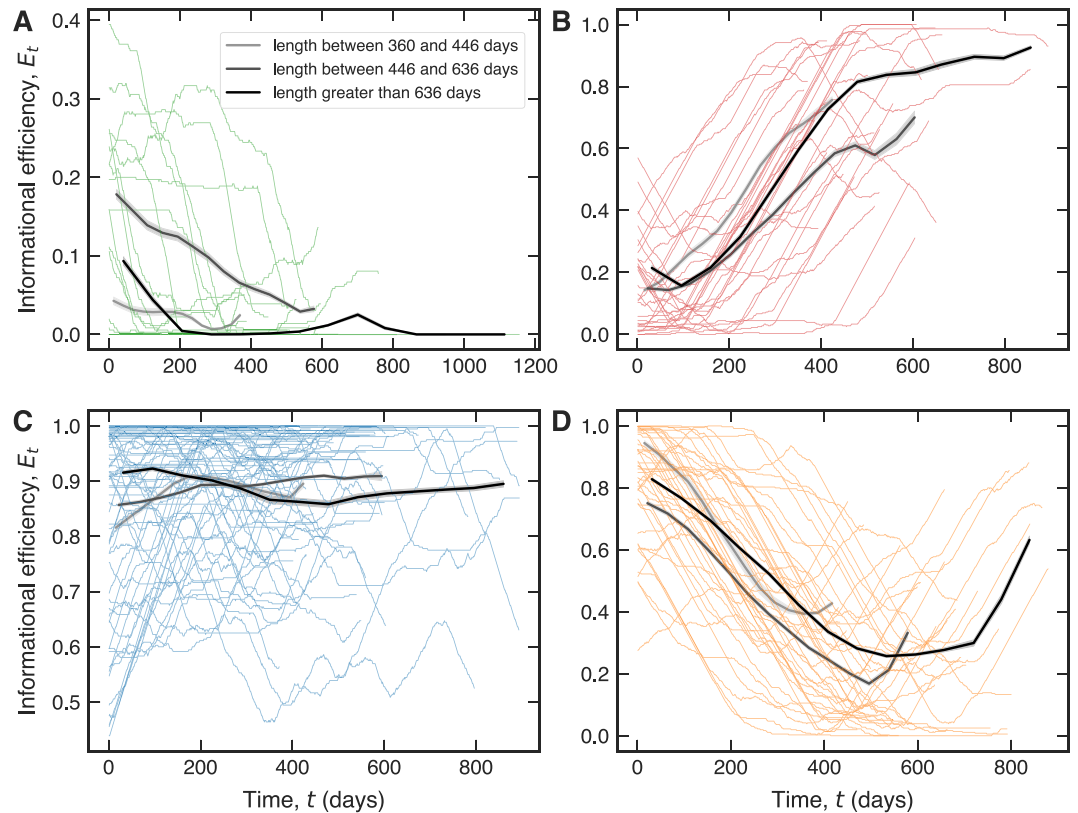


Figure 4. Temporal evolution of informational efficiency within the four different groups of cryptocurrencies. Panels (A–D) show the time series of the informational efficiency E_t for the four groups of cryptocurrencies obtained by the hierarchical clustering procedure. The colored curves represent each one of the 167 cryptocurrencies, while the gray curves show the average behavior of E_t in each group for three intervals of lengths of these series. These length intervals were chosen to contain approximately the same number of cryptocurrencies. In (A) we observe predominantly a decreasing pattern for E_t , while in (B) we have the prevalence of an increasing trend for the youngest cryptocurrencies in that group until it starts saturating, in particular for the older cryptocurrencies. In (C) there is a varying behavior of the informational efficiency E_t over time, that on average is nearly constant and highly efficient. Finally, in (D) we observe that in the beginning most cryptocurrencies are highly efficient, but then follow a decreasing trend before they recover towards the end of the observation time. These results suggest that the younger cryptocurrencies in each group are likely to follow the trend of their elders.

of the evolution of E_t between them. Figure 3B shows the matrix plot of the DTW distance between every pair of the selected 167 cryptocurrencies.

To probe for a possible hierarchical organization of cryptocurrencies regarding their informational efficiency evolution, we use the average linkage criteria to build up a dendrogram representation of the DTW distance matrix. This clustering procedure iteratively merges pairs of clusters having the smallest average distance, in which the latter is defined as the average value of the distance between all pairs of elements among the two clusters. Figure 3C depicts this dendrogram that supports the idea that cryptocurrencies are hierarchically organized regarding the profile of the evolution of their informational efficiency E_t . This hierarchical organization also gives us a chance to find groups of currencies with similar temporal characteristics. To do so, we need to find an optimal threshold distance for cutting the dendrogram and splitting the cryptocurrencies into groups. A ‘natural’ approach to specify this threshold distance is by maximizing the silhouette coefficient, a clustering evaluation metric that simultaneously grades the cohesion and the separation of the produced groups (Methods Section for details). By maximizing the silhouette coefficient, we find that 5.7 is the optimal DTW distance that splits the dendrogram into four groups of cryptocurrencies indicated by the colored branches in Fig. 3C. Thus, cryptocurrencies belonging to the same group have roughly the same temporal profile of the informational efficiency E_t , whose average behavior corresponds to the four insets at the top of Fig. 3B. These four average profiles can be qualitatively described by: (i) an almost constant and high-efficiency level (blue group, 43% of cryptocurrencies); (ii) an informational efficiency that starts at a higher level, decreases to a lower one, and increases again (orange group, 26% of cryptocurrencies); (iii) a decreasing informational efficiency (green group, 12% of cryptocurrencies); and (iv) an increasing informational efficiency (red group, 19% of cryptocurrencies).

Lastly, it is revealing to look at the temporal evolution of informational efficiency E_t for each of the four different clusters, such that currencies with different ages are distinguished, as shown in Fig. 4. We observe in

all four groups that the trends of older currencies appear to be followed up by mid-age and younger currencies. Accordingly, this dynamic analysis of informational efficiency over time reveals that younger currencies in each group are poised to follow the trend of their 'elders'. It is also worth noting that 81% of the cryptocurrencies belong to the groups characterized by a constant and high informational efficiency or by an increasing efficiency level (blue, red, and orange groups). In addition to that, among the top-10 largest currencies according to the market capitalization mean, we verify that six belong to the blue group [BitShares (BTS), Bitcoin (BTC), Dash (DASH), MaidSafeCoin (MAID), Monero (XMR), Ripple (XRP)], three to the orange group [Bytecoin (BTS), Dogecoin (DOGE), Stellar (XLM)] and one to the red group [Litecoin (LTC)]. This in turn suggests that the younger cryptocurrencies, which currently do not abide to the efficient market hypothesis, might very well do so in the foreseeable future, and at that time render the digital currency market equally compliant with the efficient market hypothesis as the traditional financial markets usually are. The relatively small deviations of younger currencies in comparison to the informational efficiency trends of the older currencies further suggest that this coming-of-age of the digital currency market might come rather soon.

Discussion

We have presented a large-scale efficiency analysis of the cryptocurrency market, focusing on temporal clustering patterns and the coming-of-age of the digital currency market in general. By using permutation entropy and statistical complexity over sliding time-windows of price log returns, our research reveals that the cryptocurrency market is to a quite significant degree compliant with the efficient market hypothesis, with only 20% of cryptocurrencies being less than 20% informationally efficient. On the other hand, over half of the cryptocurrency market is over 60% informationally efficient, and 37% of the 437 cryptocurrencies in our study is over 80% informationally efficient. This is in accord with preceding research on the same subject^{9,10}, where authors have reported similar conclusions for the Bitcoin currency.

Moreover, based on a dynamic analysis of informational efficiency over time, and by clustering together cryptocurrencies with similar temporal profiles, we have revealed which factions of the digital currency market follow the same ups and downs with respect to their informational efficiency. While the particularities of the makeup of different clusters might be of interest to investors seeking a diversified portfolio, we have shown that within the four identified clusters the younger currencies appear to follow the trend of their 'elders'. We have argued that this can be interpreted as the coming-of-age of digital currencies, such that younger cryptocurrencies, which as of yet do not abide to the efficient market hypothesis, might do so in the future. The similarities in temporal trends between younger and older cryptocurrencies led us to conclude that the coming-of-age in this regard is certainly more imminent than far-fetched.

Taken together, we hope that our physics-inspired approach aimed at determining the efficiency of the cryptocurrency market in its current state will continue to inspire further fruitful synergies between physics and economics²¹, and we hope that the presented results will contribute to the lively research environment of the digital currency market.

Methods

Data. We have obtained the daily closing prices and the market capitalization of 1509 cryptocurrencies by crawling the website coinmarketcap.com on 13 February 2018. These data span different periods ranging from a few days (e.g., for the cryptocurrencies Medicalchain and Farstcoin) to almost five years (for Bitcoin). We have selected the 437 cryptocurrencies having more than 600 observations for the analysis of the overall informational efficiency, while the results related to the dynamical behavior of the informational efficiency are based on 167 cryptocurrencies spanning more than 960 days. These filters are necessary to have a reliable estimate of the informational efficiency in each analysis, ensuring that the estimation of the overall efficiency is based on at least 100 observations of entropy and complexity, and that the time series of the informational efficiency are longer than 100 days.

Permutation entropy and statistical complexity. The permutation entropy¹⁸ and the permutation statistical complexity^{19,20} are complexity measures originally proposed for characterizing time series. Both quantities are based on a probability distribution related to the local ordering patterns among consecutive time series elements. To define these measures, let us consider a time series $\{x_t\}_{t=1,2,\dots,n}$ and overlapping partitions of length $d > 1$ (the embedding dimension) represented by

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

where $s = d, d+1, \dots, n$. For each one of these $(n-d+1)$ partitions, we investigate the $d!$ permutations $\pi = (r_0, r_1, \dots, r_{d-1})$ of $(0, 1, \dots, d-1)$ defined by the ordering $x_{s-r_{d-1}} \leq x_{s-r_{d-2}} \leq \dots \leq x_{s-r_0}$. These permutations represent all $d!$ possible ordering patterns among the d elements of the (\vec{s}) partitions. We thus calculate the relative frequency of each one of these $d!$ permutations

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

defining the probability distribution of the ordinal patterns $P = \{p(\pi_i)\}_{i=1,2,\dots,d!}$.

The permutation entropy is thus defined as a normalized Shannon entropy³⁸ of P , that is,

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

where the $\ln d!$ is a normalization constant (the maximum value of the Shannon entropy). The values of H quantify the degree of disorder in the ordering dynamics of the time series elements. Values of $H \approx 1$ indicate that elements of x_t are locally randomly ordered, while values of $H \approx 0$ show that these elements appear almost always in a particular order.

The permutation statistical complexity is defined as

$$C(P) = \frac{D(P, U)H(P)}{D^*}, \quad (4)$$

where $D(P, U)$ is a relative entropic measure (the Jensen-Shannon divergence) between $P = \{p(\pi_i)\}_{i=1, 2, \dots, d!}$ and the uniform distribution $U = \{u_i = 1/d!\}_{i=1, 2, \dots, d!}$, that is,

$$D(P, U) = S\left(\frac{P+U}{2}\right) - \frac{S(P)}{2} - \frac{S(U)}{2}, \quad (5)$$

and D^* is a normalization constant obtained by calculating $D(P, U)$ when $P = \{p_i = \delta_{1,i} \mid i = 1, \dots, d!\}$. The statistical complexity quantifies the degree of structural complexity present in the time series. A value of $C \approx 0$ occurs for both extremes of order and disorder, whereas $C > 0$ represents more complex patterns in the arrangement of the series elements.

We note that both measures depend only on the choice of the embedding dimension d . However, this choice is not entirely arbitrary because the condition $d! \ll n$ must be satisfied to obtain reliable statistics. Because of their simplicity, intuitive meaning, and scalability from the computational point of view, this framework has been successfully used in several applications with time series^{29–34} and image analysis^{35–38}.

Measuring information efficiency. The information efficiency of each cryptocurrency is estimated from the logarithmic daily price returns series defined by²¹

$$R_t = \log P_t - \log P_{t-1}, \quad (6)$$

where $\log P_t$ and $\log P_{t-1}$ are the natural logs of closing prices at time t and $t-1$.

We calculate the values of H and C within a sliding window of size w moving forward in daily steps over the return series. This procedure defines new time series representing the values of H_t and C_t in each window, where t stands for center of time-windows (Fig. 1B,C). We have used $w = 500$ days to satisfy the condition $d! \ll n$. We have also estimated the 95% random confidence intervals by shuffling the time series in each window and calculating the values of H and C for 30 independent realizations. We define the overall information efficiency E of a given cryptocurrency as the fraction of time that the values of entropy and complexity stay simultaneously within the random confidence band. The information efficiency time series E_t is estimated by calculating the same fraction within a sliding window of 360 days, as illustrated in Fig. 3A. We also observe that this procedure is robust against different sizes for the sliding window, for instance, we have verified that very similar results are obtained for windows with 30 and 180 days.

Hierarchical clustering procedure. We have used the dynamic time warping (DTW)³⁹ algorithm for estimating the similarities between the dynamical behavior of E_t among the cryptocurrencies. DTW is a distance-like and shape-based similarity measure that can be applied to time series of different lengths and range of values. This method calculates an optimal alignment between two time series by minimizing a cost function (or distance)²⁷ and it is also widely used for time-series clustering⁴⁰. We construct a matrix of the DTW distances between every pair of cryptocurrencies (Fig. 3B) and use the average linkage criteria⁴¹ to hierarchically cluster the cryptocurrencies, as shown in Fig. 3C. This clustering procedure iteratively merges clusters having the smallest average distance. The threshold distance used to determine the number of clusters was obtained by maximizing the silhouette coefficient⁴². This coefficient quantifies the consistency of the clustering procedure and is defined by the average value of

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (7)$$

where a_i is the cohesion (the average intracluster distance) and b_i is the separation (the average nearest-cluster distance) for the i -th cryptocurrency. The higher the average value of the silhouette for all cryptocurrencies, the better the cluster configuration. We find that the maximum value of the silhouette coefficient occurs for the threshold distance 5.7, which is the distance used to obtain the four clusters depicted in Fig. 3B,C. The algorithms used for evaluating the DTW and performing the hierarchical clustering procedure are implemented in the *Python* modules *dtaidistance*⁴³ and *SciPy* library⁴⁴ respectively.

References

- Malkiel, B. G. & 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).
- Preis, T. & Stanley, H. E. Bubble trouble. *Phys. World* **24**, 29, <https://doi.org/10.1088/2058-7058/24/05/34> (2011).
- Sornette, D. *Why stock markets crash: Critical events in complex financial systems* (Princeton University Press, Princeton, 2017).
- Fox, J. & Sklar, A. *The myth of the rational market: A history of risk, reward, and delusion on Wall Street* (Harper Business, New York, 2009).
- Urquhart, A. The inefficiency of bitcoin. *Econ. Lett.* **148**, 80–82, <https://doi.org/10.1016/j.econlet.2016.09.019> (2016).
- Bariviera, A. F., Basgall, M. J., Hasperu e, W. & Naiouf, M. Some stylized facts of the bitcoin market. *Phys. A* **484**, 82–90, <https://doi.org/10.1016/j.physa.2017.04.159> (2017).

7. Zhang, W., Wang, P., Li, X. & Shen, D. The inefficiency of cryptocurrency and its cross-correlation with dow jones industrial average. *Phys. A* **510**, 658–670, <https://doi.org/10.1016/j.physa.2018.07.032> (2018).
8. 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).
9. 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).
10. 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).
11. 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).
12. Alvarez-Ramirez, J., Rodriguez, E. & Ibarra-Valdez, C. Long-range correlations and asymmetry in the bitcoin market. *Phys. A* **492**, 948–955, <https://doi.org/10.1016/j.physa.2017.11.025> (2018).
13. ElBahrawy, A., Alessandretti, L., Kandler, A., Pastor-Satorras, R. & Baronchelli, A. Evolutionary dynamics of the cryptocurrency market. *Royal Soc. Open Sci.* **4**, 170623, <https://doi.org/10.1098/rsos.170623> (2017).
14. Javarone, M. A. & Wright, C. S. From bitcoin to bitcoin cash: A network analysis. In *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, CryBlock'18, 77–81, <https://doi.org/10.1145/3211933.3211947> (ACM, New York, NY, USA, 2018).
15. Ermann, L., Frahm, K. M. & Shepelyansky, D. L. Google matrix of bitcoin network. *The Eur. Phys. J. B* **91**, 127, <https://doi.org/10.1140/epjb/e2018-80674-y> (2018).
16. Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R. & Fu, F. Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *arXiv:1805.00558* (2018).
17. Alessandretti, L., ElBahrawy, A., Aiello, L. M. & Baronchelli, A. Anticipating cryptocurrency prices using machine learning. *Complexity* **2018**, 8983590 <https://doi.org/10.1155/2018/8983590> (2018).
18. 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).
19. Lopez-Ruiz, R., Mancini, H. L. & Calbet, X. A statistical measure of complexity. *Phys. Lett. A* **209**, 321, [https://doi.org/10.1016/0375-9601\(95\)00867-5](https://doi.org/10.1016/0375-9601(95)00867-5) (1995).
20. Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A. & Fuentes, M. A. Distinguishing noise from chaos. *Phys. Rev. Lett.* **99**, 154102 (2007).
21. Mantegna, R. N. & Stanley, H. E. *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 1999).
22. Zunino, L., Tabak, B. M., Pérez, D. G., Garavaglia, M. & Rosso, O. A. Inefficiency in latin-american market indices. *The Eur. Phys. J. B* **60**, 111–121, <https://doi.org/10.1140/epjb/e2007-00316-y> (2007).
23. Zunino, L. *et al.* A multifractal approach for stock market inefficiency. *Phys. A* **387**, 6558–6566, <https://doi.org/10.1016/j.physa.2008.08.028> (2008).
24. Zunino, L., Zanin, M., Tabak, B. M., Pérez, D. G. & Rosso, O. A. Forbidden patterns, permutation entropy and stock market inefficiency. *Phys. A* **388**, 2854–2864, <https://doi.org/10.1016/j.physa.2009.03.042> (2009).
25. 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* **389**, 1891–1901, <https://doi.org/10.1016/j.physa.2010.01.007> (2010).
26. Zunino, L., Bariviera, A. F., Guercio, M. B., Martinez, L. B. & Rosso, O. A. On the efficiency of sovereign bond markets. *Phys. A* **391**, 4342–4349, <https://doi.org/10.1016/j.physa.2012.04.009> (2012).
27. Sakoe, H. & Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on acoustics, speech, and signal processing* **26**, 43–49, <https://doi.org/10.1109/TASSP.1978.1163055> (1978).
28. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379, <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x> (1948).
29. Ribeiro, H. V., Zunino, L., Mendes, R. S. & Lenzi, E. K. Complexity–entropy causality plane: A useful approach for distinguishing songs. *Phys. A* **391**, 2421–2428, <https://doi.org/10.1016/j.physa.2011.12.009> (2012).
30. 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).
31. 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).
32. Jovanovic, T., Garc a, S., Gall, H. & Mej a, A. Complexity as a streamflow metric of hydrologic alteration. *Stoch. Environ. Res. and Risk Assess.* 1–13, <https://doi.org/10.1007/s00477-016-1315-6> (2016).
33. Stosic, T., Telesca, L., de Souza Ferreira, 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).
34. Ribeiro, H. V., Jauregui, M., Zunino, L. & Lenzi, E. K. Characterizing time series via complexity-entropy curves. *Phys. Rev. E* **95**, 062106, <https://doi.org/10.1103/PhysRevE.95.062106> (2017).
35. Schlemmer, A., Berg, S., Shajahan, T., Luther, S. & Parlitz, U. Quantifying spatiotemporal complexity of cardiac dynamics using ordinal patterns. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 4049–4052, <https://doi.org/10.1109/EMBC.2015.7319283> (IEEE, 2015).
36. 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).
37. Sigaki, H. Y. D., Perc, M. & Ribeiro, H. V. History of art paintings through the lens of entropy and complexity. *Proc. Natl. Acad. Sci. USA* **115**, E8585–E8594, <https://doi.org/10.1073/pnas.1800083115> (2018).
38. Antonelli, A., Meschino, G. & Ballarin, V. Mammographic density estimation through permutation entropy. In *World Congress on Medical Physics and Biomedical Engineering 2018*, 135–141, https://doi.org/10.1007/978-981-10-9035-6_24 (Springer, 2018).
39. Sakoe, H. & Chiba, S. Dynamic-programming approach to continuous speech recognition. In *1971 Proc. the International Congress of Acoustics, Budapest* (1971).
40. Aghabozorgi, S., Shirkhorshidi, A. S. & Wah, T. Y. Time-series clustering—a decade review. *Inf. Syst.* **53**, 16–38, <https://doi.org/10.1016/j.is.2015.04.007> (2015).
41. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: Data mining, inference, and prediction*. Springer Series in Statistics (Springer, New York, 2013).
42. 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](https://doi.org/10.1016/0377-0427(87)90125-7) (1987).
43. Meert, W. & Craenendonck, T. V. Time series distances: Dynamic time warping (DTW), <https://doi.org/10.5281/zenodo.1314205> (2018).
44. Jones, E. *et al.* SciPy: Open source scientific tools for Python (2001).

Acknowledgements

This research was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Grants 440650/2014-3 and 303642/2014-9, and by the Slovenian Research Agency Grants J1-7009, J4-9302, and J1-9112.

Author Contributions

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

Additional Information

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019