



OPEN

Spatial analysis of global Bitcoin mining

Wei Sun^{1,2}, Haitao Jin³, Fengjun Jin^{1,2}, Lingming Kong⁴, Yihao Peng⁴ & Zhengjun Dai³

Bitcoin mining is not only the fundamental process to maintain Bitcoin network, but also the key linkage between the virtual cryptocurrency and the physical world. A variety of issues associated with it have been raised, such as network security, cryptoasset management and sustainability impacts. Investigating Bitcoin mining from a spatial perspective will provide new angles and empirical evidence with respect to extant literature. Here we explore the spatial distribution of Bitcoin mining through bottom-up tracking and geospatial statistics. We find that mining activity has been detected at more than 6000 geographical units across 139 countries and regions, which is in line with the distributed design of Bitcoin network. However, in terms of computing power, it has demonstrated a strong tendency of spatial concentration and association with energy production locations. We also discover that the spatial distribution of Bitcoin mining is dynamic, which fluctuates with diverse patterns, according to economic and regulatory changes.

The validation of Bitcoin transactions is enabled by its proof-of-work (PoW) consensus mechanism¹. Bitcoin miners perform scanning for hash value to compete for obtaining the right of recording the block of transactions, and the successful creator of each block is rewarded by a certain amount of bitcoins. This process is called 'Bitcoin mining'^{2,3}. At the very beginning, mining activity was only supported by a few participants equipped with regular computers⁴. The surge of Bitcoin price and mining profitability incentivized increasing computing power to participate in the game. Moreover, specific mining rigs were quickly designed, manufactured and upgraded⁵. Mining sites were purposefully selected and developed. Huge amounts of energy and resources were put into mining industry^{6–8}.

Bitcoin and its mining activity have aroused attention in a variety of fields, including but not limited to blockchain technology^{2,3}, financial econometrics^{9,10}, and sustainability issues^{7,8,11–14}. Exploring the spatial distribution of Bitcoin mining will provide new angles and evidence with respect to a large portion of extant literature. In particular, the investigation from a spatial perspective will help to verify the decentralized design of blockchain technology, to identify certain kinds of price effects on cryptocurrencies and to make accurate estimations on energy consumption and carbon emissions from mining activity.

Some sustainability studies have brought valuable tracking ideas and provided interesting mapping outputs into spatial aspect of mining activity^{15–18}. Nevertheless, the spatial analyses as by-products from these studies are still limited in terms of data granularity and analytic methods. On the other hand, geographers and economists have a long tradition to describe geographical locations, patterns and dynamics of human production and trading activities^{19–22}. Bitcoin mining behaves quite differently in space when compared to conventional industrial activities. However, there is barely any novel idea published with regard to this nascent activity. Therefore, in this paper we aim to fill this gap by investigating the spatial patterns, characteristics and shaping forces of mining activity, as well as to understand, from a spatial perspective, the implications to the aforementioned topics from adjacent fields.

We carried out the research by extracting the hash rate data from million-level mining records and then desensitizing, geocoding and aggregating the data by hash rate, month and location (with unique longitude and latitude coordinates). To facilitate the spatial analysis, we divided the surface of the earth into hexagonal grids ($n = 7205$) and accommodated the hash rate data and the global power plant data²³ within the same grid system through multilayer spatial join. We then explored the statistical analysis of spatial measures over the processed data sets. We disclosed four kinds of spatial phenomena of mining activity: diffusion, concentration, association and fluctuation. Furthermore, we put the results in the context of the drivers and stages of Bitcoin mining to

¹Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. ²College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China. ³School of Computer, Beijing Information Science and Technology University, Beijing, China. ⁴BTC Team, BIT Mining Limited, Hong Kong, China. ✉email: sunw.17b@igsrr.ac.cn

better understand the causes for such spatial formations. The data sources and the step-by-step approaches are also detailed in the “Methods”.

Basics of mining activity. Prior to diving into spatial analysis, we explain some basics of mining activity up front. Three key factors that influence Bitcoin miners’ behaviour are economic incentives, technological progress and regulatory schemes. Although there are a number of studies on the economics of Bitcoin mining^{24–26}, we simplify the economic concepts of mining to better understand its relation with spatial choices as follows. In Eq. (1), P_{ij} is the mining profit for period i at location j , which is an important indicator for potential participants to determine whether they should enter the industry at the specific period and location. In Eq. (2), GM_{ij} is the gross margin for period i at location j , which is another indicator for miners to determine whether the mining rigs should be on or off.

$$P_{ij} = TR_{ij} - FC_{ij} - VCA_{ij} - VCB_{ij} \quad (1)$$

$$GM_{ij} = TR_{ij} - VCA_{ij} - VCB_{ij} \quad (2)$$

where TR_{ij} is the total mining revenue for period i at location j , which is determined by miner’s hash rate contribution, Bitcoins gained in the total network and exchange rate. FC_{ij} is the fixed cost for period i at location j , which consists of the amortization cost of hardware and initial settlement. VCA_{ij} is the variable cost (*Type A*) for period i at location j , which changes along with hash rate, mainly including the electricity cost. VCB_{ij} is the variable cost (*Type B*) for period i at location j , which also varies, but not strictly with hash rate, e.g., labour, bandwidth, cooling and other maintenance costs.

Three key takeaways are worth noting here: (i) any economic decision made by miners is based on the dynamics at a specific period and location but not on the static assumptions regardless of spatiotemporal factors; (ii) revenue factors are almost the same worldwide, while cost factors are highly localized. This means that miners obtain the same economic incentive regardless of where they are located. However, the cost breakdown of mining activity differs from location to location; (iii) it is difficult to achieve a real break-even point because of the high volatility of the Bitcoin price and the constant change in mining competition.

Technological progress intensifies the arm race of mining activity and makes it ‘portable’. Mining hardware has quickly upgraded from central processing units (CPUs), graphic processing units (GPUs) and field programmable gate arrays (FPGAs) to application-specific integrated circuits (ASICs), with an exponential increase in computational performance and energy efficiency⁵. This has apparently influenced the aforementioned economic equations on both the revenue and cost sides. Meanwhile, a set of modern technologies (including communication, engineering, logistics, etc.) make mining activity able to move and relocate easily in space, as a ‘portable industry’.

Regulatory attitudes towards Bitcoin mining vary significantly jurisdiction by jurisdiction²⁷. Some regulators take it favourable as data centre, cloud computing or fintech, while others treat it as a traditional energy-intensive industry or speculative bubble. Even within the same country, different sub-regions may hold totally different views. For example, mining activity was temporally banned in Plattsburgh, New York²⁸, while it became more favourable in Austin, Texas, due to cheap electricity and a relaxed regulatory environment²⁹. The lack of a clear global-level regulatory framework on how to define and regulate mining activity leaves room for Bitcoin miners to maneuver around the world.

Theoretically, mining activity is therefore free to move wherever it wants to exist. This is different from most industrial activities today, which are tightly constrained in space by two or more factors (e.g., resources, raw materials, talent and labour, market, transportation, regulatory permission). In addition, Bitcoin mining, to some extent, can be viewed as a prototype of the autonomous economy³⁰ (Supplementary Note 2). That is to say, the algorithm, the economic formula and the built-in technology determine the suitable locations for mining and drive human activity to move accordingly.

Spatial diffusion and concentration. It is natural to think that mining activity should be diffused all over the world due to its technical enablers and economic incentives. However, it is still astonishing to see how widely mining activity is distributed. By tracking the nodes connecting to one of the leading mining pools (“Methods”), we detected that mining activity existed in over 6000 geographical units from 139 countries and regions (Fig. 1). Except for well-known locations (e.g., China, Iceland, the US), mining activity was also detected at unexpected locations, such as Tahiti (the island in French Polynesia, the South Pacific archipelago) or Malawi (the landlocked country in Southeast Africa). If we divide the surface of the Earth into hexagonal grids ($n=7205$), we notice that 933 grids, namely, 44.3% of Earth’s land surface (Supplementary Note 3), have been found to have Bitcoin mining footprint (Fig. 2). Owing to the arm race of computing efficiency, nonspecific machines were squeezed out, such as desktops, laptops, consoles and smartphones. Otherwise, it will be overwhelming in terms of spatial presence if all the spare capacities of those devices are put into mining activity.

Although a small portion of miners are hobbyists or believers, the majority of miners nowadays are mining for economic purposes. Undoubtedly, they should tend to concentrate in locations with a competitive advantage for mining. Our results demonstrate this tendency by aggregating and counting all hash rates of individual locations within each grid (Fig. 2). Eighteen top-tier grids (share of hash rate $\geq 1\%$) accounted for 61.8% of the total computing power during our study period. In fact, miners not only concentrate in a few grids but also cluster with each other in adjacent grids. Moran’s I statistic is used to measure spatial concentration of mining activity (“Methods”). We find that the result suggests a strong rejection of the null hypothesis of spatial randomness ($I=0.65$, *pseudo p* = 0.001 for 999 permutations, $z=97.8$). In other words, mining activity demonstrated a strong tendency of concentration, in terms of computing power. We dig it further with Getis and Ord’s G_i^* statistic



Figure 1. Global presence of Bitcoin mining activity. All mining locations detected ($n=6062$) are mapped by their unique longitude and latitude coordinates. Details of each location are provided in Supplementary Table S2. The results are based on the monthly data from June 2018 to May 2019. The map is created by Geoda 1.18 (<http://geodacenter.github.io/download.html>).

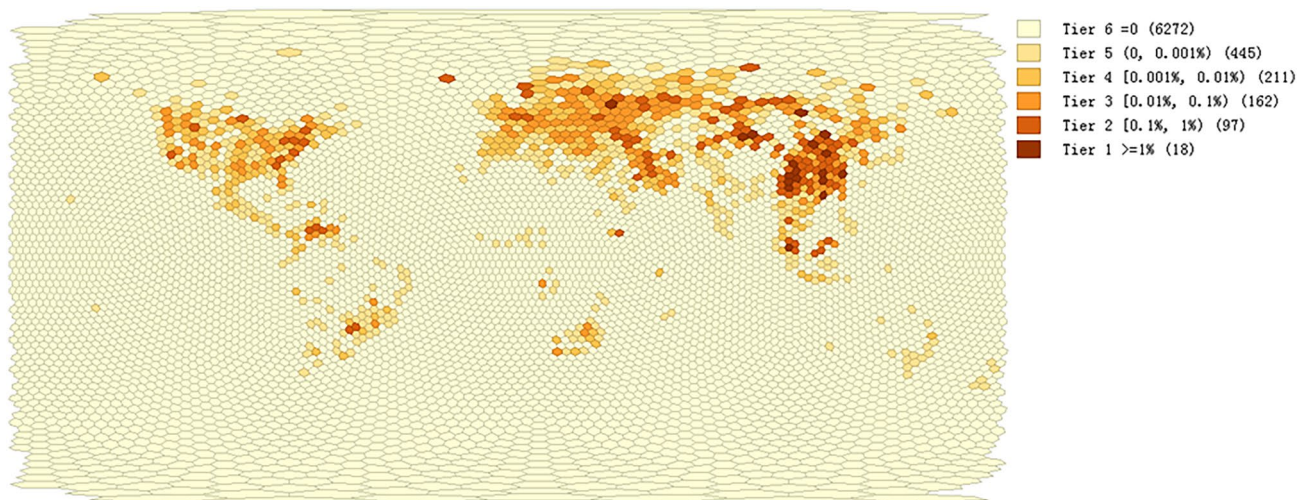


Figure 2. Share of computing power in terms of hash rate by grid. The share of computing power in each grid is represented as a percentage of total hash rates. All grids ($n=7205$) are divided into six tiers with Tier 1 grids ($n=18$, share of hash rate $\geq 1\%$), Tier 2 grids ($n=97$, $1\% >$ share of hash rate $\geq 0.1\%$), Tier 3 grids ($n=162$, $0.1\% >$ share of hash rate $\geq 0.01\%$), Tier 4 grids ($n=211$, $0.01\% >$ share of hash rate $\geq 0.001\%$), Tier 5 grids ($n=445$, $0.001\% >$ share of hash rate >0) and Tier 6 grids ($n=6272$, share of hash rate = 0). The results are based on the monthly data from June 2018 to May 2019. Details of the statistics are supplied in “Methods” and the repository as noted. The map is created by Geoda 1.18 (<http://geodacenter.github.io/download.html>).

(“Methods”) to identify the hot spots (High-High cluster cores) of mining activity under different significance (Fig. 3). Our data extended from June 2018 to May 2019. The maps for spatial concentration and hot spots may change afterwards, which will be addressed in section “Spatial fluctuation”. In addition, mining activity is virtually concentrated in the format of mining pools. An increasing number of miners are now joining pools to optimize the scanning of hash values and share returns based on their computing power contribution^{3,16}. In this analysis, we focus on the spatial phenomena in the physical world, so we will not pursue that in detail here.

Moran’s I statistic

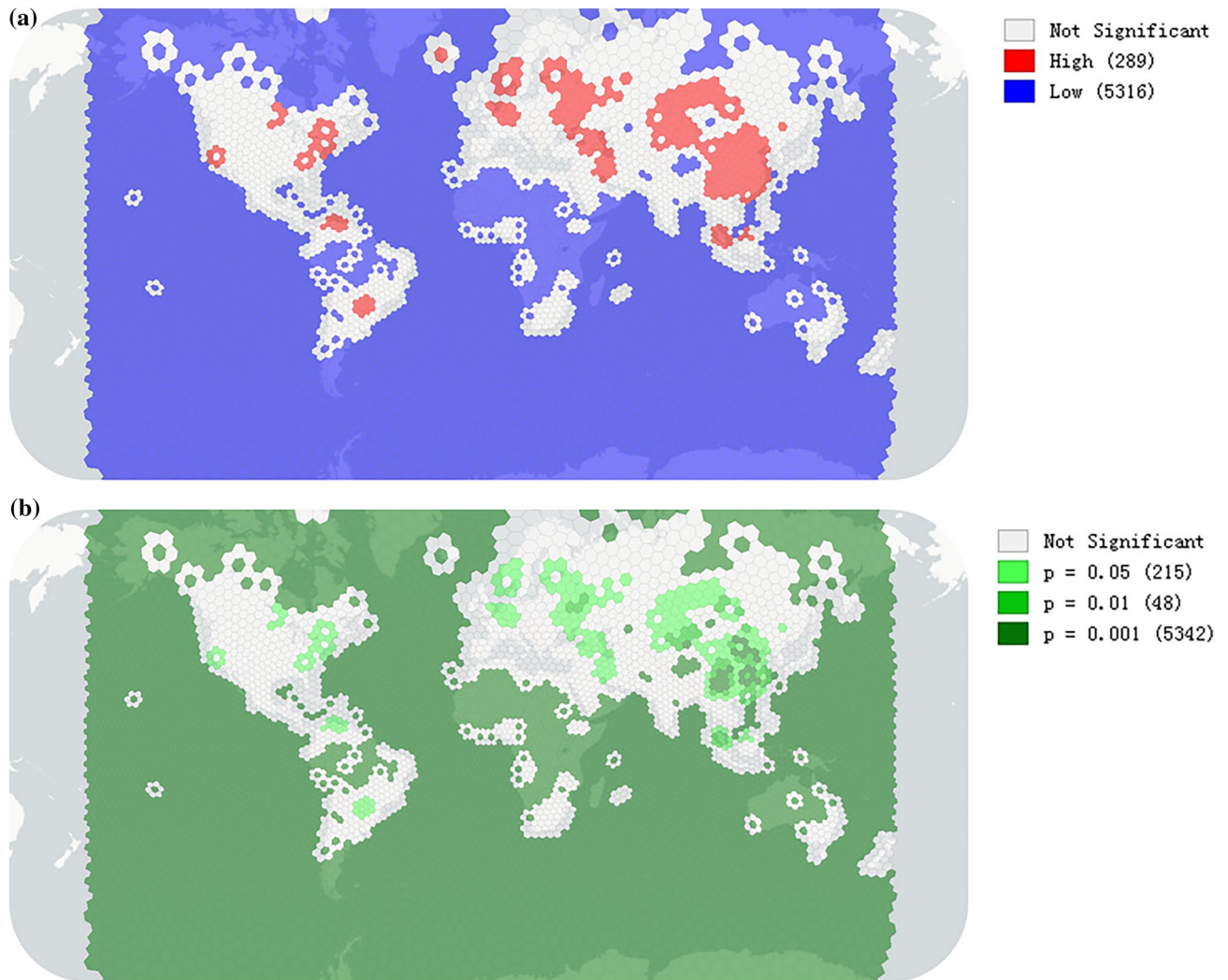


Figure 3. Hot and cold spots of Bitcoin mining activity with the corresponding significance map. **(a)** The hot spots (High-High clusters) and cold spots (Low-Low clusters) under the default setting of 999 permutations and a p-value ≤ 0.05 are marked in red and blue, respectively. **(b)** The corresponding significance map shows the clusters with the degree of significance reflected in increasingly darker shades of green, starting with $0.01 < p \leq 0.05$ ($n = 215$), then $0.001 < p \leq 0.01$ ($n = 48$) and $p \leq 0.001$ ($n = 5342$). The ‘Not Significant’ category with $p > 0.05$ remains the same in Maps **(a)** and **(b)**. Details of the statistics are supplied in “Methods” and the repository as noted. The results are based on the monthly data from June 2018 to May 2019. The maps are created by Geoda 1.18 (<http://geodacenter.github.io/download.html>).

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \tag{3}$$

where X_i and X_j are the hash rates for grids i and j , \bar{X} is the arithmetic mean of the hash rate for all grids, w_{ij} is the spatial weight between grids i and j , and n is equal to the total number of grids.

Getis and Ord’s G_i statistic

$$G_i = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} X_i X_j}{\sum_{i=1}^n \sum_{j=1}^n X_i X_j}, \quad \forall j \neq i \tag{4}$$

where X_i and X_j are the hash rates for grids i and j , w_{ij} is the spatial weight between grids i and j , and n is equal to the total number of grids.

Spatial association. As illustrated in Eqs. (1) and (2) and corroborated by our interviews and other studies^{7,11,15,16}, the most significant variable cost for mining activity is the electricity cost, which is used to power mining facilities. In this way, most miners should be inclined to locations that can provide cheap and constant sources of power. We put the global power plant data²³ into the aforementioned hexagonal grid system and explored the bivariate Moran’s I_{xy} statistics (“Methods”) between hash rate and all energy types, fossil, renewable

respectively. The results indicate a high significance of the spatial association between hash rate and all three energy variables (Fig. 4), though Moran's I between hash rate and fossil energy ($I_{hf}=0.57$) is slightly higher than that between hash rate and renewable energy ($I_{hr}=0.51$). Furthermore, we designed a 'Spatial-hit' index ("Methods") to identify areas suitable for renewable mining (Fig. 5), such as the Nordic (Hydro/Geothermal), US-Canada border areas (Hydro), US central (Wind), the Mekong River area (Hydro), and the Caucasus (Hydro).

Bivariate Moran's I_{xy} statistic

$$I_{xy} = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (X_i - \bar{X})(Y_j - \bar{Y})}{\sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X})(Y_j - \bar{Y})} \quad (5)$$

where X_i and Y_j are the hash rate for grid i and the power capacity for grid j , \bar{X} and \bar{Y} are the arithmetic mean of the hash rate and the power capacity for all grids, respectively, w_{ij} is the spatial weight between grids i and j , and n is equal to the total number of grids.

It is worth noting that it is an adaptive process that mining activity demonstrates a strong spatial association with renewable energy. Renewable energy is not always the cheapest power source and sometimes might be expensive when transmission costs are also included. However, most types of renewable energy (e.g., hydro) bear some kind of 'perishable' characteristics, similar to those of fruits (cheap in original place and value down to zero if rotted). Renewable energy providers are willing to offer miners with heavy discounts during peak seasons¹⁸. Therefore, it becomes a perfect match between the surplus of renewable energy and the 'portable' mining activity. Miners did not realize this at the early stage, while they learned and reacted through continuous testing and iteration. This will be further addressed in the next section.

Spatial fluctuation. When we drilled down to monthly data, we found that mining activity fluctuated in space based on the rolling twelve-month hash rate from June 2018 to May 2019. Here we use 1500 TH/s as the threshold to select grids with at least 100 mining rigs for our analysis (Supplementary Note 4). In terms of the characteristics of monthly fluctuation, grids with hash rate over 1500 TH/s ($n=229$) were observed and put into twelve clusters through cluster analysis with K-medoids ("Methods"). We further categorized twelve clusters into four groups with reference to the real operational environment: ascending, descending, relatively stable and seasonal fluctuation (Fig. 6).

Every fluctuating grid fluctuated in its own way, which might follow a combination of multiple patterns and can only be explicitly explained case by case. However, four primary patterns are studied and summarized here. (i) Price effect: the drop in the Bitcoin price drives mining profitability down, as illustrated in Eqs. (1) and (2). Large mining farms choose to migrate to locations with more cost advantages or update their mining machines, while most individual or small miners are reluctant to take immediate actions and wait for the suitable time to reopen their mining rigs. All these factors lead to a change in computing power in grids but to different degrees. (ii) Seasonal effect: some miners are accustomed to transfer periodically to leverage the discounts offered by suppliers within certain grids where there is surplus energy during the peak season (e.g., rainy season for hydro-power grids). It also happens when these miners move back to their original locations during the off-season. (iii) Regulatory effect: attitudes from regulators dramatically influence the behaviours of miners in related grids. Favourable measures (e.g., subsidies, tax benefits) encourage miners to move in, while adverse measures (e.g., bans, carbon taxation) drive miners out. (iv) Iterative effect: initial mining activity may start randomly from the grids where early believers, tech geeks or speculators inhabit. Miners (in particular large ones) continue to learn and search for better mining locations. The process is iterative for optimal solutions, and the radius of search is expanded to adjacent grids and then gradually to the global scale. Thus, a considerable portion of computing power at the original grids is relocated to the well optimized grids. Unfortunately, only part of this pattern can be observed within our study since the anonymity of the Bitcoin network makes it nearly impossible to recognize early mining locations.

Spatial fluctuation is never ending. We notice that the recent change in regulatory policy towards Bitcoin mining in some jurisdictions (e.g., China's crackdown in 2021) has intrigued a new round of spatial fluctuation and migration. Bitcoin mining activity is in the process of moving to achieve new spatial equilibrium^{31,32}. We believe that the spatial analysis here will still be applicable in new circumstances.

Discussion

We have presented an in-depth analysis of the spatial distribution and characteristics of Bitcoin mining through bottom-up tracking and geospatial statistics. Our research reveals that mining activity has been widely distributed all over the world. Meanwhile, the Moran's I statistics indicate a strong tendency of its spatial concentration and association with energy production locations. We also consider that the spatial distribution of Bitcoin mining fluctuates with diverse patterns, according to economic and regulatory changes. Moreover, as elaborated below, we believe that our findings from the spatial perspective will be beneficial to the related studies of blockchain technology, financial econometrics, sustainability and other fields.

First, our results show that in terms of presence, mining activity is widely distributed, which is in accord with the decentralized nature of blockchain technology. However, in terms of computing power, it has demonstrated a strong tendency of spatial concentration, particularly towards the locations with abundant and cheap energy supply. This phenomenon will increase the potential risk of a 51% attack¹¹ and make the whole network more vulnerable due to regulatory changes, disasters or other constraints at specific locations. By pointing out this pitfall, we hope that our analysis could help to inspire the improvement of blockchain design which shall further assess the concentration risk not only from ownership structure but also from geographical distribution.

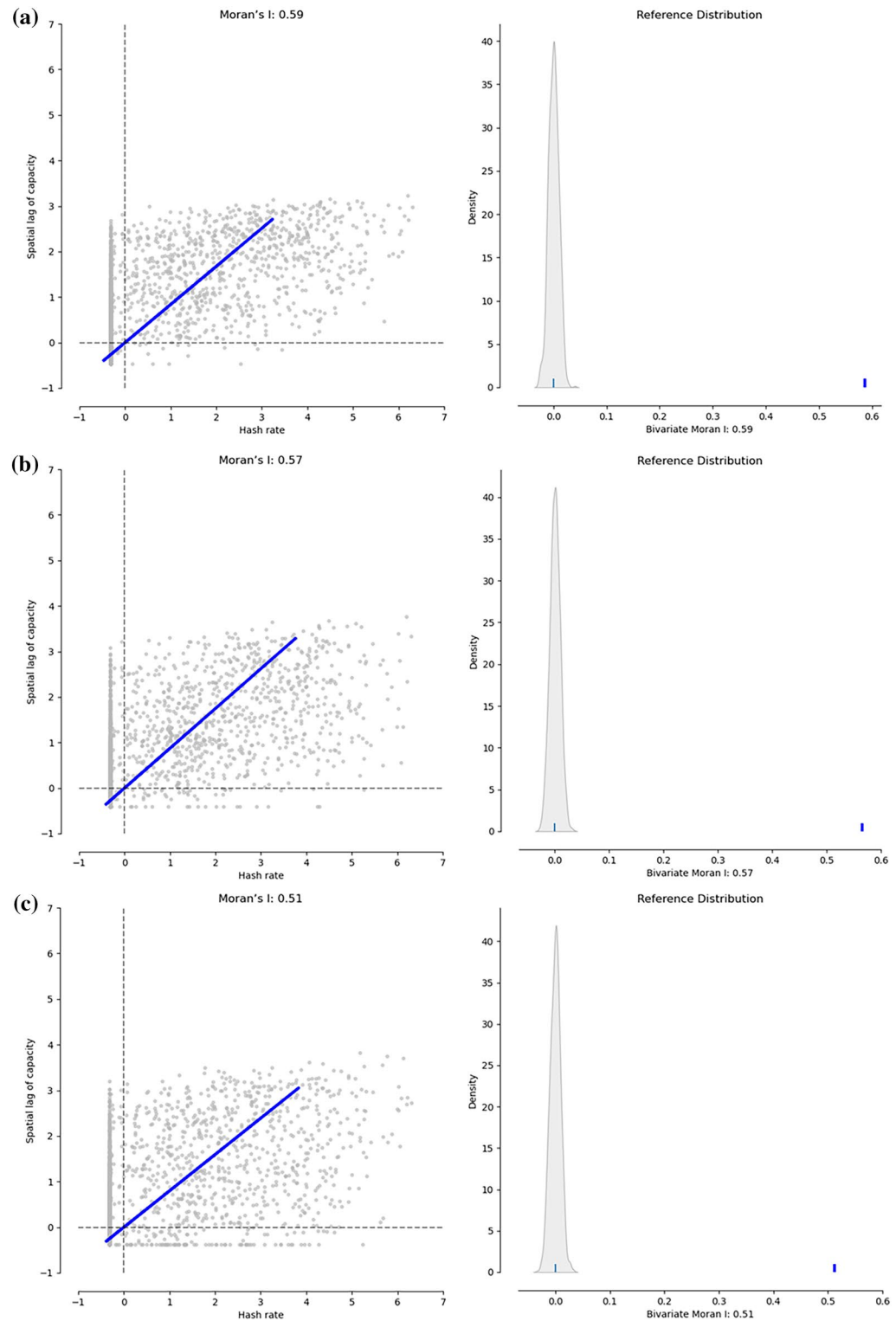


Figure 4. Bivariate Moran's scatter plots and reference distributions between hash rate and different energy variables. (a–c) Bivariate Moran's statistical results between the hash rate and capacity of all types of energy (a), fossil energy (b), and renewable energy (c) demonstrate the degree of spatial association between them. The scatter plot is depicted with the spatially lagged energy capacity on the y-axis and the original hash rate on the x-axis. The slope of the linear fit to the scatter plot equals Moran's I. The reference distribution demonstrates the result by randomly permuting the observed values over the locations, which is depicted as a distribution curve in the left. The short line shows the value of Moran's I, well to the right of the reference distribution. Details of the statistics are supplied in “Methods” and the repository as noted.

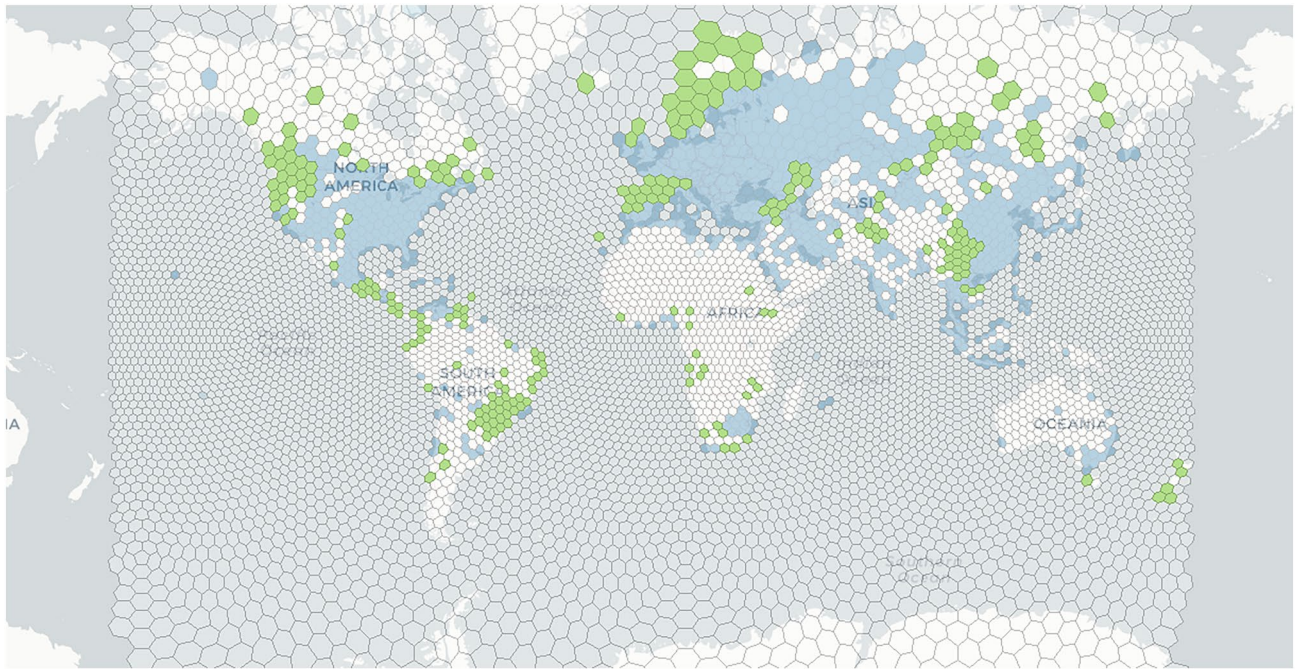


Figure 5. ‘Spatial hit’ index indicates the potential locations suitable for renewable mining. Grids with ‘spatial hit’ index = 2 (i.e. suitable for renewable mining) are highlighted in green (n = 247). Details of the definition and calculation of the index are provided in “Methods”. The results associated with this map are shown in Supplementary Table S4. The results are based on the monthly data from June 2018 to May 2019. The map is created by Geoda 1.18 (<http://geodacenter.github.io/download.html>).

Second, we find that our analysis may help to gauge market valuation on cryptocurrency assets and contribute as an input layer for neural network pricing models. The number of Bitcoin participants, the amount of computing power and the diversity of mining locations shall be viewed as the indicators of how widely the consensus is accepted. These factors can be used to measure the positive feedback loop between network effects and security: broader consensus and more participants lead to higher value, which in turn attracts more miners and increases the security of transaction validation²⁶. Moreover, the shift of mining capacity between Bitcoin and Altcoins shall indicate the price spillover effect among cryptocurrency assets and markets¹⁰. In fact, all of these can be quantified and structured as a separate input layer for neural network pricing models³³ and be further assessed the viability by backtesting with historical data. If the effectiveness is proved, we believe that our methodologies and empirical findings may contribute to fixing price misalignment and producing novel insights on cryptoasset management.

Third, understanding spatial distribution of Bitcoin mining is crucial for making accurate estimations on sustainability issues and for providing practical suggestions on more eco-friendly mining activity. Spatial footprint of Bitcoin mining is one of the fundamental factors influencing the accuracy of estimations on energy consumption and carbon emissions^{7,15}. Our results provide empirical evidence on the distribution and dynamics of mining activity, which help to estimate energy consumption and carbon emissions based on real situation instead of theoretical assumptions. Meanwhile, a series of studies have tried to give suggestions to reduce environmental impacts from cryptocurrency mining, such as less energy-intensive alternatives for the proof-of-work mechanism or guiding the computing power and energy to more meaningful usages^{3,11,34}. These are all good attempts, though they might face technical challenges or still lack enough consensuses. Based on the spatial analysis, we suggest that a feasible way in the short term is to build a global framework to regulate mining activity and motivate miners to the locations of abundant renewable energy, particularly where with surplus capacity.

Last but not least, Bitcoin mining, which was originated from the mail exchange of cryptographic community, has brought substantial impacts on our real world in terms of market capitalization, spatial presence, energy consumption and carbon emissions. It also provides researchers with a remarkable experiment³⁵ of the transparent rules, the general availability of data and the lively interaction. We hope that our analysis tackled from a spatial perspective will continue to inspire further cross-field communications among geography, sustainability, economics, finance and technology.

Methods

Data wrangling. The desensitized, geocoded and aggregated hash rate data (42,820 records in total) are given in Supplementary Table S1. The data were extracted from the million-level mining records of partnered pools. All sensitive or confidential data were desensitized to fully protect the privacy of all mining participants. Please note that the research team has no direct access to the original Internet Protocol (IP) addresses or any other privacy data. The IP addresses of mining activity were converted and geocoded into physical locations under the instructions given by the research team. IP addresses can be localized with ip2location.com or similar tools. The naming and classification of countries and regions are under the World Bank guidelines³⁶. The hash

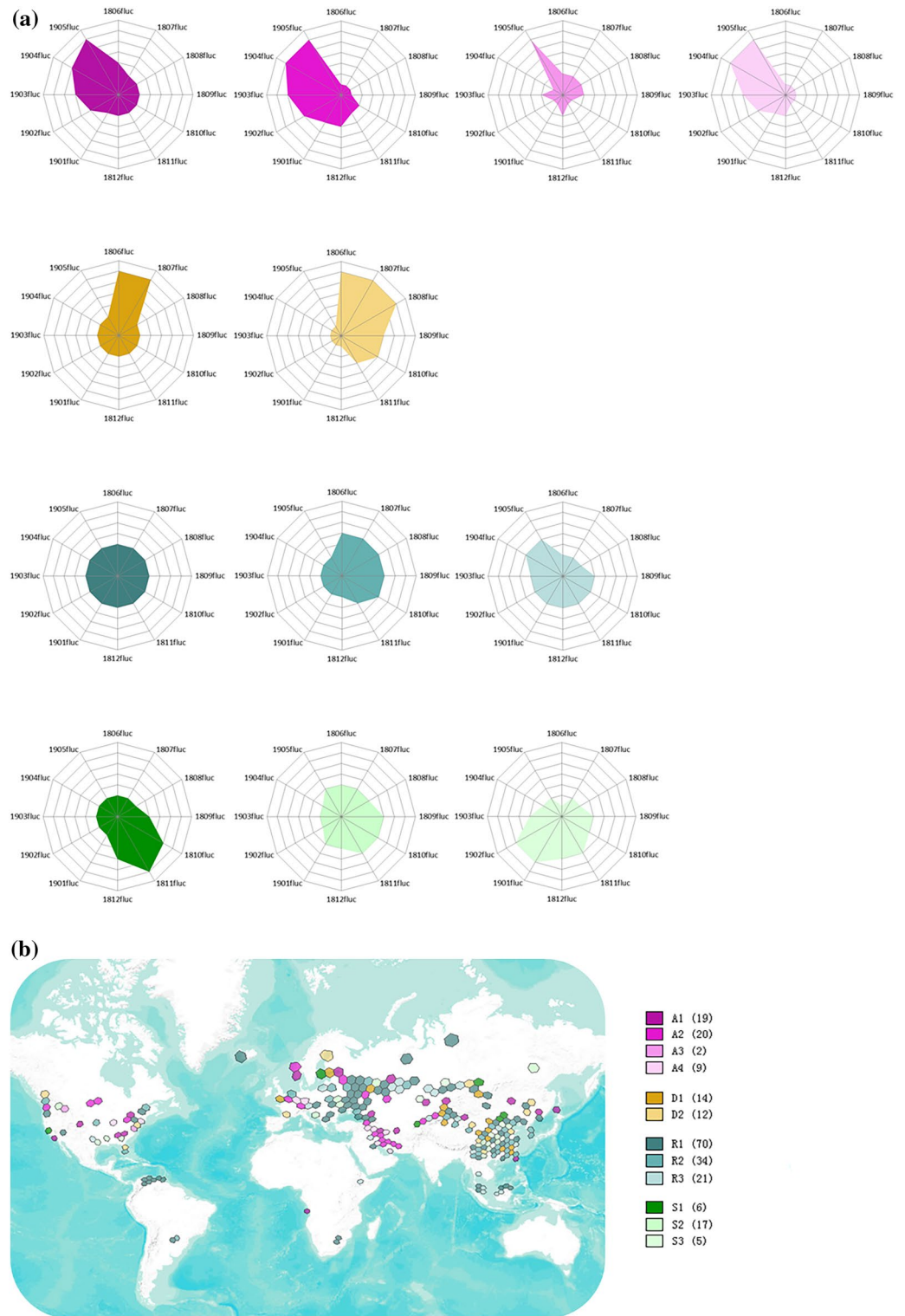


Figure 6. Classification of the grids with differentiated fluctuation patterns. **(a)** Grids with hash rate over 1500 TH/s ($n = 229$) are divided into twelve clusters in four groups. The twelve-month fluctuation indices of medoids are plotted in the radar chart as representatives of each cluster. **(b)** All the observed grids are plotted in Map **(b)** with their respective categories, sharing the sample colour scheme for each category in panel **(a)**. Details of the results are provided in Supplementary Tables S5, S6 and the repository. The results are based on the monthly data from June 2018 to May 2019. The map is created by Geoda 1.18 (<http://geodacenter.github.io/download.html>).

rate data in Supplementary Table S1 are finally organized by month, geographical name, location identifier (with unique longitude and latitude coordinates) and hash rate (max, min and average value of each month). By ticking the location identifier, we detect Bitcoin mining activity in 6062 unique locations from 139 countries and regions (Supplementary Tables S2, S3) and visualize the results in Fig. 1.

We use the global power plant data²³ (World Resources Institute, 2019) to indicate the abundance of different types of energy in specific locations. We selected the following fields for analysis: geographical name, location identifier (longitude and latitude coordinates, gppd_idnr), primary_fuel and capacity_mw. Based on the commonly used energy types for mining activity, we categorize oil, gas, petcoke and coal as fossil energy, hydro, geothermal, solar and wind as renewable energy and the remaining energy types as others.

Gridding method and preprocessing. In this subsection, we explain what kind of geogrid system is used for analysis and how we connect the hash rate and power data to the grid system. A discrete global grid system (DGGS), looking like mosaics that cover the entire Earth's surface, is often used as the geometric basis for the building of geospatial data structures^{37,38}. Our analysis builds on two publicly available hexagonal grid tools and resources^{39,40}. The blank grid datasets (with no attribute data) are provided at the GitHub repository. The original hash rate, power data and grid information are stored in different layers. We then explore spatial join by counting and aggregating the hash rate and power data of each location into the associated grid (i.e. the point location of longitude and latitude is within the given grid). Thus, the point-based hash rate and power data are merged into the blank grid datasets. The results are provided at the repository and the script for the cross-layer operation is shared on GitHub. This step can also be achieved in a common geographic information system (GIS) environment.

In addition, we preprocessed some of the merged grid data as follows. (i) 'HrateAvg' is calculated as the twelve-month average of hash rate for each grid; (ii) 'HrateShare' is calculated as the percentage of each grid's hash rate to the total; and (iii) both the hash rate and power capacity data of each grid are standardized as below.

$$X_ZS = ZSCORE(\log_{10}(X + 1)) \quad (6)$$

where X_ZS is the standardized value by the Z-score method ($Z = \frac{x-\mu}{\sigma}$, μ is the mean of the sample, σ is the standard deviation of the sample), and X is the raw value of the hash rate or power capacity.

Spatial statistics. Two types of spatial weight matrices assisting further analysis are given in the repository. We calculate the spatial weights based on the Queen contiguity (sharing a common edge or vertex), which express the neighbour structure between the grids. The difference between two spatial weight matrices is whether the in-situ relation (i.e. the diagonal of the weight matrix) is considered.

Moran's I statistic⁴¹ is the commonly used indicator of spatial concentration (autocorrelation) and illustrated in Eq. (3). In essence, Moran's I statistic demonstrates the relation between a variable (in deviation from the mean) and its spatial lag (weighted average of adjacent grids). We use Python's PySAL⁴² package to calculate Moran's I statistic for the hash rate and the related reference distribution under 999 permutations. The results ($I = 0.65$, $pseudo\ p = 0.001$, $z = 97.8$) suggest a strong rejection of the null hypothesis of spatial randomness. Similarly, the bivariate Moran's I_{xy} statistics between the hash rate and the capacity of all energy types, fossil and renewable are explored. The difference is that the bivariate statistic demonstrates the relation between one variable (hash rate) and the spatial lag of another variable (power capacity). In our case, the spatial lag of power capacity takes the in-situ value into consideration by applying the spatial weight matrix including diagonal. The results ($I_{hr} = 0.59$, $pseudo\ p = 0.001$, $z = 60.7$; $I_{hf} = 0.57$, $pseudo\ p = 0.001$, $z = 59.6$; $I_{hr} = 0.51$, $pseudo\ p = 0.001$, $z = 53.6$) indicate a high significance of the spatial association between hash rate and all three energy variables. The code, the settings for reproducibility and the outputs are archived on GitHub.

Getis and Ord's G_i^* statistic⁴³ is further used to indicate the location of the clusters of mining activity and illustrated in Eq. (4). The G_i^* statistic is fine-tuned by including the in-situ value at the given grid (i.e., the spatial weight matrix including diagonal is applied). The G_i^* statistic for each grid with an assessment of significance is calculated and compared with the global mean. A value larger than the mean suggests a hot spot (High-High cluster), while a value smaller than the mean indicates a cold spot (Low-Low cluster)⁴⁴. The calculations are implemented and the results are visualized in Geoda⁴⁴ (permutations = 999 and random seed = 123,456,789 as default for reproducibility).

We then introduce the concept of the 'spatial hit' index to determine the grids suitable for renewable mining, which is defined as

$$Spatial\ hit\ index = \begin{cases} 2, & \text{hash rate} > 0 \cup \text{capacity_sl of renewable} > \text{fossil} \\ 1, & \text{hash rate} > 0 \cup \text{capacity_sl of renewable} \leq \text{fossil} \\ 0, & \text{hash rate in the given grid} = 0 \end{cases} \quad (7)$$

The 'spatial hit' index is calculated to identify the grids where the value of the hash rate is greater than 0 and the spatial lagged capacity of renewable energy is greater than that of fossil energy. We can further use the original power data to trace back to the specific type of renewable energy (Hydro, Geothermal, Solar or Wind) for the grids with the 'Spatial hit' index equal to 2. The results (with the grids suitable for renewable mining highlighted in green) are also visualized in Geoda (Fig. 5).

Cluster analysis. We use 1500 TH/s as the threshold to select the grids for cluster analysis of spatial fluctuation since it is difficult to predict the mining behaviour in grids with only a few rigs. During the period of our study, the hash rate for a typical mining rig was approximately 15 TH/s. A grid with hash rate over 1500 TH/s

means that there are at least 100 mining rigs (equivalent to a small-sized mining farm) in the given grid. In this way, 229 grids out of all 933 grids with mining footprints are selected.

The fluctuation index measures the amplitude of the monthly fluctuation of hash rate in the given grid, defined as

$$Hrate_fluc = \frac{M_i - \bar{M}}{\bar{M}} \quad (8)$$

where M_i is the hash rate of each month (from June 2018 to May 2019) for the given grid and \bar{M} is the twelve-month average of hash rate in that grid.

The results are stored in the grid datasets and shared on GitHub. The fluctuation index is further categorized as

$$Hrate_fluc_code = \begin{cases} 6 \text{ (skyrocketing increase)}, & Hrate_fluc > 1 \\ 5 \text{ (significant increase)}, & 1 \geq Hrate_fluc > 0.75 \\ 4 \text{ (moderate increase)}, & 0.75 \geq Hrate_fluc > 0.25 \\ 3 \text{ (relatively stable)}, & 0.25 \geq Hrate_fluc \geq -0.25 \\ 2 \text{ (moderate decrease)}, & -0.75 \geq Hrate_fluc > -0.25 \\ 1 \text{ (significant decrease)}, & -1 \geq Hrate_fluc > -0.75 \end{cases} \quad (9)$$

We use a classical partitioning technique of clustering (k-medoids with PAM) and perform the analysis through Python's scikit-learn and scikit-learn-extra packages⁴⁵. The code and the settings for clustering are provided in the repository. The clustering results for each grid (Supplementary Table S5) are appended into the grid datasets and visualized in Geoda. Twelve medoids (Supplementary Table S6), as representatives for each cluster, are extracted and further categorized into four groups based on the understanding of real mining operations: ascending, descending, relatively stable and seasonal fluctuations (Fig. 6).

Robustness analysis. For spatial statistics, we use a computational approach based on permutation, which calculates a reference distribution for the statistic by randomly permuting the observed values over the locations. The statistic is computed for each of these randomly reshuffled data sets, which yields a reference distribution. This approach is not sensitive to potential violations of underlying assumptions, and is thus more robust⁴⁴. It can be easily verified by replicating the process based on the code we provided on Github (https://github.com/jricb/bitcoin_mining_spatial_analysis).

For cluster analysis, the k-medoids method is used to minimize the sum of the distances from the observations in each cluster to a representative center (medoid) for that cluster⁴⁴. In contrast to k-means and k-medians, k-medoids uses actual observations and will not move the center of the cluster towards the outlier⁴⁵. Thus it can be more robust to noise and outliers. For instance, we test different methods and find that Grid 444 and 673 are appropriately divided into category 10 under k-medoids.

There are several algorithms to compute k-medoids. The approach we used is the partitioning around medoids (PAM) algorithm, which consists of two steps: Build (greedy initialization of the medoids from the dataset) and Swap (compute the cost of swapping one medoid with any non-medoid point and stop when there is no change in the objective function)⁴⁵. We have observed that very similar results are yielded under different iterations and initialization settings. This can be also repeated by following the code and instructions we provided on Github.

Study limitations. We acknowledge that there are some limitations in our analysis. First, the sample of mining records may not be fully representative. The pools we partnered with are BTC.com (primary) and Ant-Pool (cross-check). Both were the subsidiaries of Bitmain (the former sold to BIT in 2021) and accounted for approximately 30% of the total hash rate during the period of our study. Large and medium-sized miners could allocate their computing power proportionally to the major pools, while small or individual miners might choose the pools nearby or perform solo mining. Thus, the sample may better represent large and medium miners instead of small miners. This implies that there might be more mining locations than detected and mining activity might distribute more evenly than demonstrated. Nevertheless, to the best of our knowledge, the dataset in this study should be the most comprehensive sample for mining activity at this level of detail.

Second, fake IP addresses may hinder the tracking of real locations. Some miners use virtual private networks (VPNs) to intentionally obfuscate their locations^{15,17} or manage their mining rigs through cloud services. However, it is reasonable to assume that the vast majority of miners have little incentive to use VPN given the resulting latency¹⁵. The connections from cloud services have been redirected to their original addresses as much as possible by the engineers of the partner pools. However, we cannot guarantee that the fake IP influence associated with the bottom-up approach has been completely eliminated.

Third, the approximation of renewable energy in each grid might be underestimated due to the data availability of global power plants. As noted by Byers et al.²⁶, it is difficult to collect information on relatively new and small renewable power plants since they are not always recorded in public documentation. When we aggregate the power plant data from each location to approximate the abundance of energy in the associated grid, this will lead to the fact that renewable energy is less extensively covered as fossil energy.

Finally, the spatial distribution and characteristics of mining activity are not stationary and continue evolving under different regulatory and economic scenarios. For instance, mining activity has shifted dramatically in space since a series of measures against Bitcoin mining was taken by the Chinese authorities in 2021. The latest data on mining actions have not yet been available. However, we are confident that our analytical methods and most of the results will still be applicable in new circumstances. Further studies are needed to observe the spatial dynamics of mining activity and draw conclusions about the ultimate form of similar self-driven production activity.

Data availability

All data used in the study are provided on GitHub (https://github.com/jricb/bitcoin_mining_spatial_analysis), Supplementary Information or publicly available online as noted.

Code availability

Python Spatial Analysis Library (PySAL) can be downloaded from <https://pysal.org/>. Geoda can be downloaded from <http://geodacenter.github.io/>. Python's scikit-learn and scikit-learn-extra packages can be downloaded from <https://scikit-learn.org/> and <https://scikit-learn-extra.readthedocs.io/>, respectively. The code for customized spatial analysis is available on GitHub at https://github.com/jricb/bitcoin_mining_spatial_analysis.

Received: 25 January 2022; Accepted: 16 June 2022

Published online: 23 June 2022

References

- Nakamoto, S. *Bitcoin: A Peer-to-Peer Electronic Cash System* (Bitcoin.org, 2008).
- Narayanan, A. et al. *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction* (Princeton University Press, 2016).
- Tschorsch, F. & Scheuermann, B. Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Commun. Surv. Tutor.* **18**(3), 2084–2123 (2016).
- Taylor, M. B. Bitcoin and the age of bespoke silicon. *2013 International Conference on Compilers, Architecture and Synthesis for Embedded Systems (CASES)*, 1–10 (2013).
- Taylor, M. B. The evolution of bitcoin hardware. *Computer* **50**(9), 58–66 (2017).
- O'Dwyer, K. J. & Malone, D. Bitcoin mining and its energy footprint. *25th IET Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies* 280–285 (2014).
- Krause, M. J. & Tolaymat, T. Quantification of energy and carbon costs for mining cryptocurrencies. *Nat. Sustain.* **1**, 711–718 (2018).
- De Vries, A. Bitcoin's growing energy problem. *Joule* **2**(5), 801–805 (2018).
- Sigaki, H. Y. D., Perc, M. & Ribeiro, H. V. Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market. *Sci. Rep.* **9**, 1440 (2019).
- Giudici, P. & Pagnottoni, P. High frequency price change spillovers in bitcoin markets. *Risks* **7**(4), 111 (2019).
- Vranken, H. Sustainability of bitcoin and blockchains. *Curr. Opin. Environ. Sustain.* **28**, 1–9 (2017).
- Giungato, P., Rana, R., Tarabella, A. & Tricase, C. Current trends in sustainability of bitcoins and related blockchain technology. *Sustainability* **9**(12), 2214 (2017).
- Mora, C. et al. Bitcoin emissions alone could push global warming above 2°C. *Nat. Clim. Change* **8**, 931–933 (2018).
- Foteinis, S. Bitcoin's alarming carbon footprint. *Nature* **554**, 169 (2018).
- Stoll, C., Klaaßen, L. & Gellersdorfer, U. The carbon footprint of bitcoin. *Joule* **3**(7), 1647–1661 (2019).
- Blandin, A. et al. *3rd Global Cryptoasset Benchmarking Study* (Cambridge Centre for Alternative Finance, 2020).
- Bitcoin Mining Map*. Cambridge Centre for Alternative Finance. https://cbeci.org/mining_map. Accessed 23 Nov 2021.
- Bendiksen, C., Gibbons, S. & Lim, E. *The Bitcoin Mining Network—Trends, Average Creation Cost, Electricity Consumption and Sources* (CoinShares Research, 2019).
- von Thünen, J. H. *The Isolated State English edition of Der Isolierte Staat* (Pergamon, 1966).
- Marshall, A. *Principles of Economics (1890)* 8th edn. (Macmillan, 1920).
- Krugman, P. Increasing returns and economic geography. *J. Polit. Econ.* **99**(3), 483–499 (1991).
- Dicken, P. *Global Shift: Reshaping the Global Economic Map in the 21st Century* 4th edn. (Sage, 2003).
- Byers, L. et al. *The Global Power Plant Database* (World Resources Institute, 2019). <https://datasets.wri.org/>
- Kroll, J. A., Davey, I. C. & Felten, E. W. The economics of Bitcoin mining, or Bitcoin in the presence of adversaries. *12th Workshop on the Economics of Information Security (WEIS)* 1–21 (2013).
- Houy, N. The Bitcoin mining game. *Ledger* **1**, 53–68 (2016).
- Catalini, C. & Gans, J. S. Some simple economics of the blockchain. *Commun. ACM* **63**(7), 80–90 (2020).
- Blandin, A. et al. *Global Cryptoasset Regulatory Landscape Study* (Cambridge Centre for Alternative Finance, 2019).
- Plattsburgh Lifts Ban on Cryptocurrency Mining Operations* (The Associated Press, 2019). <https://www.northcountrypublicradio.org/news/story/38175/20190304/plattsburgh-lifts-ban-on-cryptocurrency-mining-operations>
- Feng, Z. *Why China's Bitcoin Miners are Moving to Texas* (BBC, 2021). <https://www.bbc.com/news/world-us-canada-58414555>
- Arthur, W. B. Foundations of complexity economics. *Nat. Rev. Phys.* **3**, 136–145 (2021).
- Gent, E. Crypto's slow motion gold rush. *IEEE Spectrum* (2021).
- Cambridge data shows Bitcoin mining on the move. *BBC* <https://www.bbc.com/news/technology-57811959> (2021).
- Pagnottoni, P. Neural network models for Bitcoin option pricing. *Front. Artif. Intell.* **2**, 5 (2019).
- Fairley, P. Ethereum plans to cut its absurd energy consumption by 99 percent. *IEEE Spectrum* (2019).
- Böhme, R., Christin, N., Edelman, B. & Moore, T. Bitcoin: Economics, technology, and governance. *J. Econ. Perspect.* **29**(2), 213–238 (2015).
- Country Classification* (World Bank Data Help Desk). <https://datahelpdesk.worldbank.org/knowledgebase/topics/19280-county-classification>. Accessed 4 Aug 2021.
- Discrete global grid* (Wikipedia). https://en.wikipedia.org/wiki/Discrete_global_grid. Accessed 6 July 2021.
- Sahr, K., White, D. & Kimerling, A. J. Geodesic discrete global grid systems. *Cartogr. Geogr. Inf. Sci.* **30**(2), 121–134 (2003).
- Hexagon Discrete Global Grids* (CartoGeek, ArcGIS Online). <https://www.arcgis.com/home/item.html?id=38e3e0453e4040c69ff5e9fde306c12c>. Accessed 6 July 2021.
- Create Hexagon Tessellation* (Whiteaker, T., ArcGIS Online). <https://www.arcgis.com/home/item.html?id=03388990d3274160afe240ac54763e57>. Accessed 6 July 2021
- Moran, P. A. P. The interpretation of statistical maps. *J. R. Stat. Soc. Ser. B Methodol.* **10**(2), 243–251 (1948).
- Rey, S. J. & Anselin, L. PySAL: A Python library of spatial analytical methods. *Rev. Reg. Stud.* **37**(1), 5–27 (2007).
- Ord, J. K. & Getis, A. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* **27**(4), 286–306 (1995).
- Anselin, L., Syabri, I. & Kho, Y. GeoDa: An introduction to spatial data analysis. *Geogr. Anal.* **38**(1), 5–22 (2006).
- Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Acknowledgements

We would like to thank Chee-We Ng, Shun Zhang, Micree Zhan, Zhong Zhuang, Alice Wang, WWG, J.Y. Lu, Zilong Wang, Zhaoye Xie and Mola Gao for their valuable comments.

Author contributions

W.S. and H.J. conceived the study and designed the analysis approach. L.K. and Y.P. acquired and processed the original mining data. Z.D. developed part of the code. F.J. supervised and commented the study. W.S. analysed the data and drafted the manuscript.

Competing interests

The authors declare the following competing interests: H.J. declares financial holdings of less than US\$500 of Bitcoin. L.K. and Y.P. are employed by BTC.com. The other authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-14987-0>.

Correspondence and requests for materials should be addressed to W.S.

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.



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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022