



City Research Online

City, University of London Institutional Repository

Citation: Bracci, A., Boehnke, J., ElBahrawy, A., Perra, N., Teytelboym, A. & Baronchelli, A. (2022). Macroscopic properties of buyer–seller networks in online marketplaces. *PNAS Nexus*, 1(4), pgac201. doi: 10.1093/pnasnexus/pgac201

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/29300/>

Link to published version: <https://doi.org/10.1093/pnasnexus/pgac201>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

Macroscopic properties of buyer–seller networks in online marketplaces

Alberto Bracci ^a, Jörn Boehnke ^b, Abeer ElBahrawy^c, Nicola Perra ^d, Alexander Teytelboym ^e and Andrea Baronchelli ^{a,f,g,*}

^aDepartment of Mathematics, City, University of London, London EC1V 0HB, UK

^bGraduate School of Management, University of California Davis, 1 Shields Ave Davis, CA 95616, USA

^cChainalysis Inc, New York, NY 10011, USA

^dSchool of Mathematical Sciences, Queen Mary University of London, London E1 4NS, UK

^eDepartment of Economics, University of Oxford, Oxford OX1 3UQ, UK

^fUCL Centre for Blockchain Technologies, University College London, London WC1E 6BT, UK

^gThe Alan Turing Institute, London NW1 2DB, UK

*To whom correspondence should be addressed: Email: Andrea.Baronchelli.1@city.ac.uk

Edited By: Karen E. Nelson

Abstract

Online marketplaces are the main engines of legal and illegal e-commerce, yet their empirical properties are poorly understood due to the absence of large-scale data. We analyze two comprehensive datasets containing 245M transactions (16B USD) that took place on online marketplaces between 2010 and 2021, covering 28 dark web marketplaces, i.e. unregulated markets whose main currency is Bitcoin, and 144 product markets of one popular regulated e-commerce platform. We show that transactions in online marketplaces exhibit strikingly similar patterns despite significant differences in language, lifetimes, products, regulation, and technology. Specifically, we find remarkable regularities in the distributions of transaction amounts, number of transactions, interevent times, and time between first and last transactions. We show that buyer behavior is affected by the memory of past interactions and use this insight to propose a model of network formation reproducing our main empirical observations. Our findings have implications for understanding market power on online marketplaces as well as intermarketplace competition, and provide empirical foundation for theoretical economic models of online marketplaces.

Keywords: online marketplaces, complex networks, buyer–seller networks, dark web marketplaces

Significance Statement:

Online marketplaces have become very successful in the past 20 y, with dark web marketplaces (DWMs) also proliferating on the dark web to satisfy the demand for illicit goods. However, the networks that result from interactions between buyers and sellers on these platforms are poorly understood. Here, we investigate these networks by analyzing 245M transactions on one e-commerce platform and 28 DWMs. Despite many differences between the marketplaces, we find remarkable regularities in user behavior and propose a simple model reproducing the main empirical observations. The results shed light on buyer–seller networks and their formation mechanisms, highlighting the central role of buyer memory and preferential attachment mechanism, and have important implications for the understanding, design, and regulation of these platforms.

Introduction

Much of online trade happens on regulated and unregulated online marketplaces. Regulated online marketplaces include Amazon, Craigslist, eBay, Walmart, Alibaba (China), Rakuten (Japan), Gumtree (UK), and Mercado Libre (South America). Unregulated online marketplaces, such as Silk Road, AlphaBay, and Hydra, which specialize in the sale of illicit goods, have proliferated on (and disappeared from) the dark web since the introduction of Bitcoin (1–4). The amount of transactions in online marketplaces is vast and growing. For example, in 2020 Amazon reported a net revenue of 386B USD (5), while in 2019 the ecosystem of

dark web marketplaces (DWMs) had reached a total volume of 4B USD (2).

Online marketplaces are commercial websites that allow participating buyers and sellers to exchange information about prices and products and to execute transactions (6–8). Sellers can usually post an ad for their product that includes a product description, a price and a shipping method. Buyers instead can see all relevant product ads matching search keywords, and have access to reviews and seller ratings. When a purchase is made, the payment is processed through the platform, while shipping is usually taken care of by the seller.

Competing Interest: The authors declare no competing interest.

Received: April 14, 2022. **Accepted:** September 29, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Despite the importance of online marketplaces for e-commerce and global trade (9, 10), little is known about their empirical properties, transaction patterns and the resulting buyer–seller networks. The properties of the transaction network could, however, provide important insights into the presence of market power (11, 12), the nature of interplatform competition (13, 14), product design (15), the effects of reputation on sellers' revenue growth (16), and the long-run sustainability of the platforms (17). Moreover, measuring properties of the buyer–seller networks could help provide empirical foundations for theoretical models of online marketplaces, from the estimation of model parameters to suggesting specific model mechanisms. However, buyer–seller networks in online marketplaces have specific features that make them different from other networks (e.g. social networks): they exhibit a naturally bipartite structure; most transactions (links) occur between anonymous agents; transaction activity might be infrequent and sporadic. Moreover, the structure of buyer–seller networks could depend on the nature of the traded products, on the types of buyers and sellers, on the user experience on the marketplace, or even on the legal, institutional and geographic constraints.

One strand of prior work relevant to our paper has touched on various aspects of regulated online marketplaces. For example, the role of reputation and feedback (18–20) has been identified as one of the main drivers of the worldwide success of regulated online platforms (21). Other work has looked at consumer search and the effect of rankings on product choice (22–25), online auction markets (26–29), market microstructure (30, 31), and price formation in online markets (32–36). [For a more complete but less up-to-date review, see ref. (6).] Another strand of research has studied unregulated marketplaces. This work has focused on country-specific studies (37–39), the effects of closures and law enforcement raids (2, 3, 40–42), the characterization of the trade of specific goods (43–45), the importance of geography (46, 47), or sociological interview-based studies (39, 48). However, most work on unregulated online marketplaces was limited to specific markets, and focused on information available from public listings (e.g. using crawled data) (1, 40, 43–46).

In this paper, we focus on patterns in transactions, which typically cannot be publicly observed either on regulated or unregulated online marketplaces. We analyze two datasets. The first dataset contains 220M transactions between 99M buyers and 7.4M sellers, which occurred in 144 randomly sampled product markets of one regulated e-commerce platform between 2010 and 2020, for a total volume of over 10B USD. The second dataset contains 25M transactions involving 17M entities with a total volume of 4.2B USD, which occurred in 28 major DWMs between 2011 and 2021, for a total volume of 4.2B USD (for more details on the datasets see the “Materials and methods” Section). In both cases, the datasets cover all transactions, which occurred in each corresponding market.

We observe striking similarities in user behavior across online marketplaces, despite their significant differences. First, we find that the number of transactions, amount, interevent time and time between first and last transaction are highly heterogeneous across users but follow consistent fat-tailed distributions across all marketplaces. Then, we show that individual behavior is influenced by past purchases similarly (albeit less strongly) to what is observed in the renewal of past ties in social networks (49, 50). Finally, we propose a simple model of buyer–seller interactions that reproduces the main stylized facts of the data and emphasizes the

critical role of preferential attachment (51, 52) and memory in the market dynamics.

Results

Empirical properties of buyer–seller networks

In order to characterize the buyer–seller networks, we start by analyzing different aggregate user-level quantities. First, we study the distributions (for each market) for the number and amount of user transactions. Results for all DWMs and for each e-commerce market are shown in Figs. 1(a) to (d), where black and yellow lines are obtained by aggregating all users in the respective datasets. Single distributions display common behavior, spanning several orders of magnitudes. It is important to note that distributions are computed without any rescaling or manipulation of the data, and that higher values generally reached by the regulated platform in all distributions are exclusively due to the different platform sizes. The slight discrepancy between the distributions in the total number of received transactions can be ascribed to the different nature of the two datasets: While in the DWM dataset sellers can withdraw the earnings from several market trades at once, in the e-commerce data each transaction corresponds to a single purchase.

We then analyze the temporal dimension of the data. We focus on the distribution of user lifetimes, defined as the time between the first and last user transaction in the market, and the interevent time between two successive transactions of the same user. Again, we find remarkable regularities across different DWMs and different regulated product markets, as shown in Figs. 1(e) to (h). In these distributions, as before, we also observe the effects of different sizes of marketplaces. The similarity between different distributions is particularly pronounced in the meaningful timescales between an hour and a month/year. Discrepancies for longer periods are due to the different lifetimes of the markets, whereas discrepancies for shorter timescales can be explained by the different nature of the two datasets: precise timestamps on transaction data for the regulated marketplaces vs. times at which the transaction is actually added to the Bitcoin blockchain (which depends on its algorithm) for the DWM dataset.

Having considered buyers and sellers separately, we now investigate the dynamics of buyer–seller relationships and the evolution of the buyer–seller network. We limit this analysis to e-commerce markets, since DWMs data do not contain buyer–seller links (see the “Materials and methods” Section for more details). We first consider how single users distribute their purchases across sellers: for example, buyers could purchase equally from multiple sellers or, alternatively, buyers could show loyalty to one seller from which they do most of their purchases. A standard way to quantify how distributed or concentrated this pattern is to compute the normalized entropy for the purchases of each buyer i as in Eq. 1, and then compute its distribution for all markets. The normalized entropy is defined as

$$e_i = - \sum_{j=1}^J n_i^j \log_2(n_i^j) / \log_2(J), \quad (1)$$

where n_i^j is the share of buyer i 's purchases from seller j and we sum over the J sellers the buyer purchased from. Fig. 2(a) shows that the distributions, computed for each market, are fat-tailed, with buyers populating the full $[0,1]$ support but with most of the mass toward the top, meaning that most buyers buy a similar number of times from the different sellers they purchase

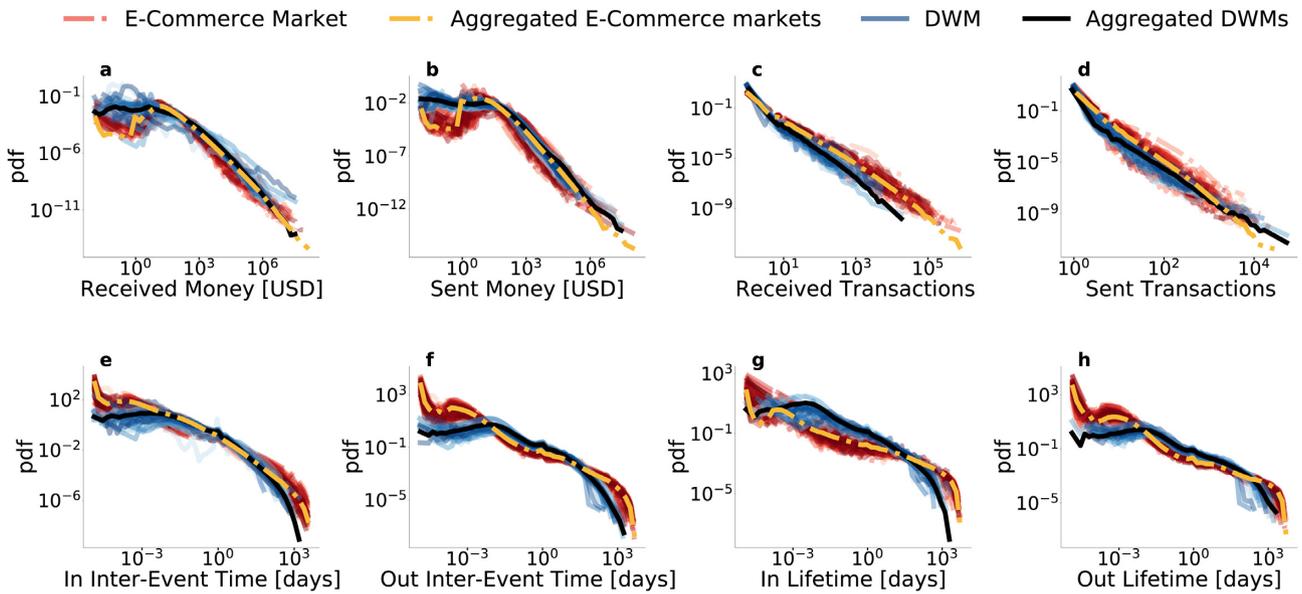


Fig. 1. Online marketplaces show strikingly similar patterns according to different aggregated quantities. Top (a to d): distributions of 4 analyzed users aggregate quantities: money and number of transactions both sent and received. Bottom (e to h): distributions of four analyzed users temporal quantities: the interevent time (time between successive transactions) and the lifetime (time between first and last transaction) both measured in days. Each blue line represents one DWM, the black line is the distribution built aggregating all DWMs together, each dashed red line represents one e-commerce platform market, while the dashed yellow line is the distribution aggregating all e-commerce markets. Similar patterns are observed between different markets in the same platform, but also across regulated and unregulated online marketplaces.

from. Buyers with zero entropy, who buy from just one seller, were excluded from the figure for visual clarity, but these were almost exclusively buyers who only made a single purchase (see Fig. S1 in the Supplementary Material). In Fig. S2 in the Supplementary Material, we further compare the distributions against a null model obtained reshuffling the transactions in the dataset (preserving buyers activity and sellers attractiveness), where we find lower heterogeneity and higher tendency toward high values of entropy. This implies that the empirical entropy distributions show a broader non trivial range of behaviors interpolating between perfect exploration and exploitation.

The observed normalized entropy distributions are compatible with different kinds of temporal patterns produced by two possible choices: either buyers choose to engage with new sellers they have never purchased from (i.e. exploration) or they return to a known seller (i.e. exploitation). We investigate these dynamics by leveraging insight from the social networks literature, where several studies have investigated how users explore and exploit social connections by renewing previously activated ties or by establishing new ones (49, 50). Indeed, across different types of social networks, the temporal evolution of links that a person forms with their contacts can be captured by the following expression:

$$P(n) = (1 + n/c(k_{min}))^{-\beta(k_{min})}, \quad (2)$$

where— now using the language of online marketplaces — $P(n)$ is the probability that a buyer (node) of degree n (who has already bought from n different sellers) chooses to buy from a new seller, while c and β are positive constants, depending on the final degree of the buyer, which measure their propensity to explore new sellers and thus the effect of memory. Following the procedure proposed in (49) (see Supplementary Material for more details), we group nodes in different classes according to the final degree: a buyer is in class k_{min} if the final degree k satisfies $k_{min} \leq k \leq 2k_{min} - 1$, starting from $k_{min} = 2$. For this computation, we aggregate

all markets together in order to have a representative sample in classes with higher k_{min} . If a user is present in multiple markets, we keep its activity in different markets separated (i.e. effectively considering her as different users). We then fit Eq. 2 to each node class obtaining a value of $c(k_{min})$ and $\beta(k_{min})$ (see Table S1 in the Supplementary Material).

Results are shown in Fig. 2(b). Since different classes feature different values of β and c , we plot a rescaled $P(x)^{1/\beta}$ as a function of n/c . Indeed, Eq. 2 becomes $1/(1+x)$ (dashed line in Fig. 2b) for every degree class k_{min} , where $x = n/c$. In other words, we re-scale both axes assuming the empirical behavior is captured by Eq. 2. As shown in Table S1 in the Supplementary Material, the parameter values are independent of the degree class and suggest a weaker ($\beta \sim 10^{-1}$) effect than previously observed in social networks ($0.48 \leq \beta \leq 2$) (50). The close fit of the data to the predicted memory for different k_{min} indicates the applicability of Eq. 2 in the dynamics of buyer–seller relationships. While users have different propensities to explore new sellers, they follow the same mechanism: the more sellers a user has bought from, the less likely is their next purchase from a new seller.

Modeling buyer–seller networks

In order to understand possible mechanisms that drive the properties of buyer–seller networks, we propose an agent-based model aimed at capturing the patterns observed in the previous section. The main features of the model are:

- (1) *Activity.* The rate at which buyers make transactions. As shown in Fig. 1, in both e-commerce and DWMs buyers feature heterogeneous propensities to make purchases.
- (2) *Memory.* When making new transactions buyers can either choose a seller they already bought from or pick a new one. As shown in Fig. 2(b) and discussed above, buyers have a memory of the sellers they had interacted with, and this memory affects their future purchases.

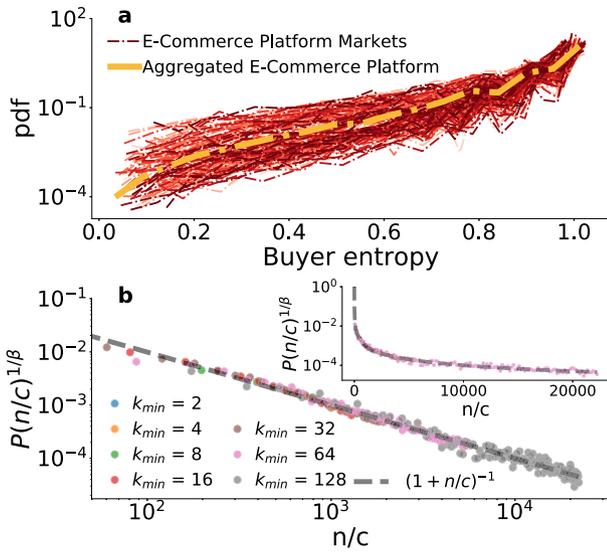


Fig. 2. Buyer memory affects their purchase decisions. (a) Normalized buyer entropy distribution for each e-commerce market (red), and the whole e-commerce platform (yellow), excluding users with zero entropy (mostly users with one transaction, see Fig. S1 in the Supplementary Material) for visual clarity. The fat-tailed distributions span the full [0,1] range, with most buyers almost equally buying from multiple sellers. (b) $P(n)$ is the probability to buy from a new seller after a buyer has already bought from n different ones. Each degree class $k_{min} \leq k \leq 2k_{min} - 1$ is rescaled according to the fitted value of c and β (see Table S1 in the Supplementary Material for the values), with Eq. 2 (dashed line) well reproducing the memory effect on the buyers' behavior: the more sellers they try, the less likely they are to buy from a new one.

(3) *Preferential attachment.* The attractiveness (i.e. popularity) of a seller is proportional to the number of their sales. This attractiveness captures the fact that, in online marketplaces, sellers are rated based on the feedback they receive from the buyers (i.e. customer reviews), and buyers prefer sellers with higher ratings, other things equal (18, 19, 53–55). Here, we focus on the number of sales rather than sale volume to capture the fact that it is mainly frequency of transactions that matters for seller reputation.

Given these three ingredients the model dynamics is as follows. The system consists of N buyers and M sellers. At $t = 0$ we assign the activity a_i to each buyer i . Each seller j starts with attractiveness $A_j = 1$. At each time step t , each buyer makes a purchase with probability $a_i \cdot \Delta_t$, where Δ_t is the simulation time step (fixed to 1 from now on). A buyer who interacted with n sellers in the past has probability $P(n) = (1 + n/c)^{-\beta}$ of choosing a new seller and $1 - P(n)$ of returning to a known one. In the first case, the buyer selects a new seller j proportionally to their attractiveness (56) A_j , in the latter, the buyer selects it proportionally to the number of previous interactions. In other words, buyers select sellers either according to past purchases or to their popularity. In both cases the attractiveness of the seller is increased by μ . This model produces a bipartite temporal network: at each time step t we build a network in which two types of nodes—buyers and sellers—are linked if the buyer has purchased from that seller at time t . These networks are then combined together in an aggregated network, where each buyer–seller link is weighed according to the number of purchases between that buyer and that seller across time.

Compared to other activity-driven models developed to capture the temporal evolution of different social networks (49, 50, 57), our model extends the framework to bipartite networks of buyers

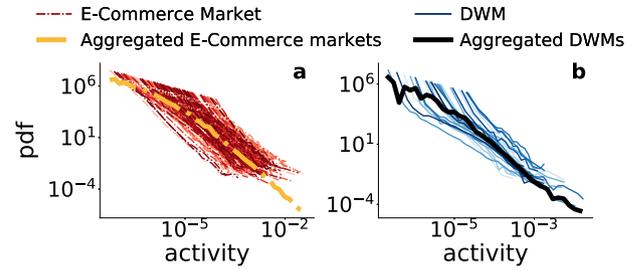


Fig. 3. Empirical activity distributions. (a) Activity distribution for all e-commerce markets in red, and activity distribution of all aggregated markets in yellow. (b) Activity distribution for all DWMs in blue, and activity distribution of all aggregated DWMs in black.

and sellers and introduces the preferential attachment guiding the buyer selection process. Henceforth, we will refer to the model lacking preferential attachment, proposed in (49), as *Model NoPA*. We will also consider a version of the model that does not include the memory element (*Model NoMem*). Comparing these versions of the model will allow us to identify the role played by the different mechanisms.

A standard way to define and measure user activity in a (social) network is $a_i = n_i / \sum_l n_l$, where n_i is the number of purchases made by buyer i , where the sum is over all buyers in their market. In Fig. 3, we show the activity distributions of all e-commerce markets (a) and all DWMs (b). While curves exhibit fat-tailed behavior, they no longer overlap due to different activity ranges and shapes in different product markets. As a result, we need to use market-specific empirical activity distributions as inputs for our model.

We now fit the model to the e-commerce data. As mentioned above, since the DWM dataset does not contain the full bipartite buyer–seller network, we cannot test all the model predictions on the DWM data. We employ a data-driven approach, fine-tuning the model to each single market so we can more faithfully compare the simulation results with the empirical buyer–seller networks. In the main text, we show results for two different product markets, 26 more are shown in Figs S3 to S6 in the Supplementary Material, for a total of 28 markets (see Supplementary Material for the sampling procedure). We fix parameters $\beta = 0.1$ and $c = 0.001$, which we fitted previously (see Table S1 in the Supplementary Material), and use the empirical activity distributions as measured in the data (see Fig. 3a) to reflect the observed heterogeneity across different markets. The value of μ is determined with Maximum Likelihood Estimation for each market (see Supplementary Material for more details, and Table S3 for the fitted values).

Results are in Fig. 4. We first compare the model's output with the empirical distributions of the final seller attractiveness and degree. The attractiveness of a seller j is their market share $A_j = s_j / \sum_l s_l$, where s_j is the total number of sales of seller j and the sum is over all the sellers. Fig. 4 shows that the model reproduces both distributions well, while the NoPA variation of our main model (without preferential attachment) fails to capture the heterogeneity (up to six orders of magnitude) of these curves, emphasizing how preferential attachment is key to reproducing the presence of very active sellers. We then consider the buyer side of the network. We first study the degree distribution. Fig. 4 shows that the model captures the empirical distributions, while the absence of buyer memory generally leads to a small overestimation of the tails, since it does not induce the repetition of past interactions with a subset of buyers.

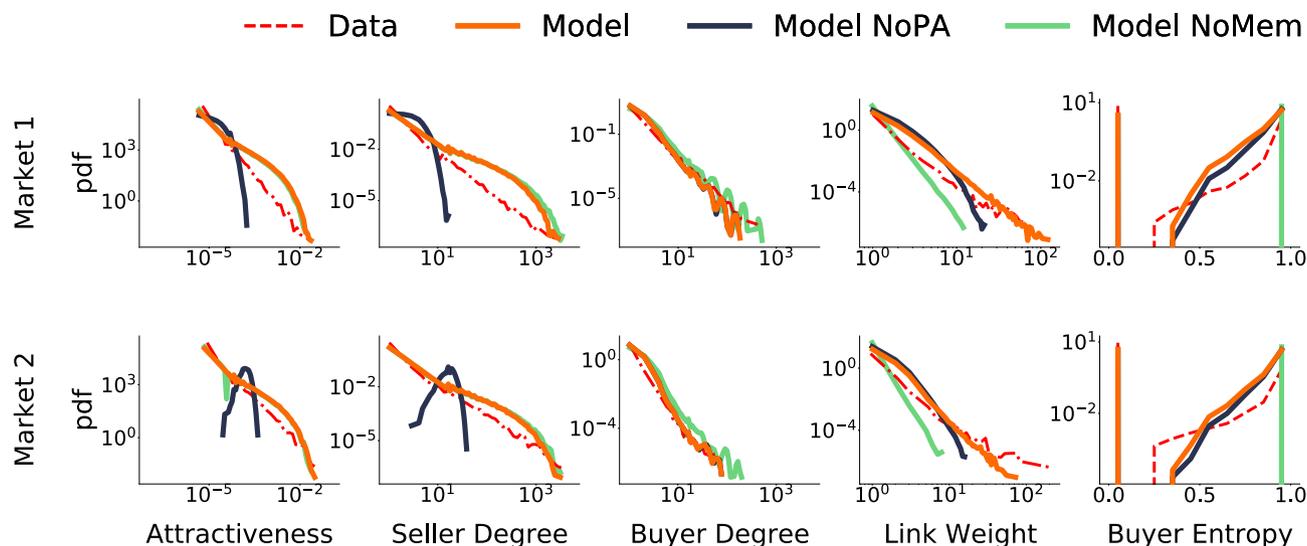


Fig. 4. The model reproduces different properties of buyer–seller networks. Each row corresponds to a different market (see Figs. S3 to S6 in the Supplementary Material for other markets), whose simulation parameters are individually calibrated as detailed in the main text. From left to right, we show distributions for different quantities: attractiveness, seller degree, buyer degree, link weight and seller entropy. The comparison with the two model variations, without preferential attachment or without memory, shows the key role of both parameters in shaping the network: preferential attachment is crucial in reproducing highly active sellers, whereas buyer memory is fundamental to capture the heterogeneity of buyer–seller relationships.

Thus far, we have considered node-level properties aggregating detailed information on the links. For example, the attractiveness only accounted for the total number of links, whereas the degree only captures the total number of different buyers or sellers that the user has interacted with. To better understand how the model performs in reproducing finer details of the buyer–seller network structure, we test our model against two other properties of the aggregated network: link weight—the number of transactions between a buyer and a seller—and the buyer entropy, as defined in Eq. 1. Our main model outperforms its two variations in reproducing the shape and tails of the link weight distribution. In particular, the memory mechanism appears to be fundamental in reproducing repeated transactions between a buyer and a seller. The buyer entropy distribution is again well-captured by the model and shows how the memory mechanism is key to capturing the diversity of relationships buyers establish with different sellers. Indeed, the *NoMem* model produces only entropy values close to 0 and 1; this happens because without memory, a buyer almost never finds any previous seller, hence buyers making more than one purchase almost always buy from new sellers.

We have seen that our model is able to capture various aspects of the final aggregated buyer–seller network. The next step is to see whether our model can also reproduce the temporal evolution of the buyer–seller network. To investigate this, we focus on the degree of top sellers since we previously showed these sellers generate the largest activity and volume on these markets. We measure time by the total number of purchases made. Results are shown in Figs. 5(a) to (c), where we plot the temporal evolution of the top 50 (a), 100 (b), and 200 (c) seller degree distribution for one illustrative product market. In doing so, we compare the model to its two variations and the data. Results for more product markets are shown in Figs. S7 to S10 in the Supplementary Material. The main model is able to reproduce the temporal evolution of the distributions, as clearly shown by the cores (i.e. interquartile ranges) overlapping at different times. We further compute the absolute value of the difference between the mean of the models' distribution and the mean of the data, for each of the nine equally spaced

time steps and for all 28 simulated product markets. As shown in Figs. 5(d) to (f), the model is better able to reproduce the temporal dynamics for all simulated markets. Indeed, the median of the distance distributions is always smaller in the main model than the two other model variations.

Discussion

In this paper, we have analyzed 244M (25B USD) transactions occurring on regulated and unregulated online marketplaces. First, we have revealed remarkable regularities in the aggregate static and temporal properties of the buyer–seller networks, both for buyers and sellers. Then, we have revealed how buyers are affected by the memory of past interactions. Finally, we have proposed a model, which captures the main stylized facts of the data, based only on three well-known network formation mechanisms of online marketplaces: buyers have different propensity to make purchases, they remember the sellers they purchased from, and they are more likely to buy from successful sellers.

It is important to highlight the limitations of our study, which also represent directions for further work. First, while our study is based on (pre-processed) blockchain data, access to DWM server logs could provide more detailed information on some specific markets, for instance, the directed buyer–seller links, which are not observable in our data. Second, the model could be further enriched with other known mechanisms: pricing dynamics (58, 59), product search ranking (22–25), variable (e.g. also negative) customer reviews (20), sellers entering or leaving the platform (60), and recommendation algorithms (61). Finally, including richer economic incentives (e.g. strategic behavior) to model buyers' and sellers' decisions could shed light on how agents could exploit their market power. In particular, the inclusion of strategic behavior would also drop phenomenological rules such as preferential attachment, which would naturally result from the agents' behavior (62, 63). A deeper understanding of economic incentives and equilibrium behavior in buyer–seller networks could ul-

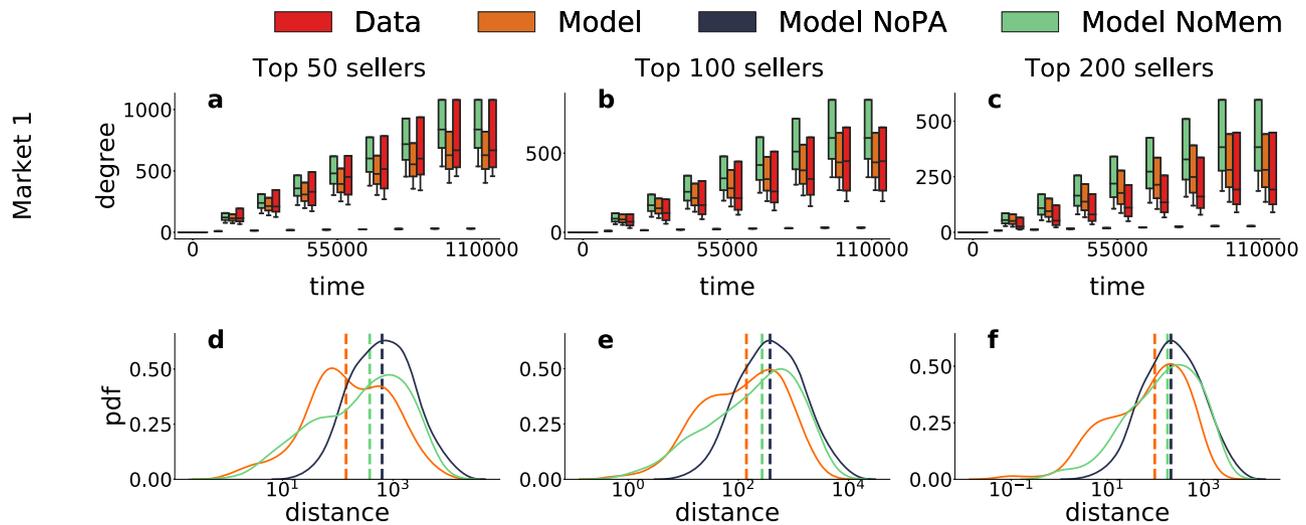


Fig. 5. Model reproduces the temporal evolution of the top sellers degree distribution. Top (a to c): Temporal evolution of the degree distribution of the top 50 (a), 100 (b), and 200(c) sellers, representing the distribution at nine equally spaced time steps with boxplots ranging from the first to the third quartiles, whiskers extending from 2.5th to 97.5th percentiles. Results are shown for one product market, all other markets are shown in Figs. S7 to S10 in the Supplementary Material. Bottom (d to f): Distribution of the distance between the empirical and model(s) median degree of the top 50(left), 100(center), and 200(right) sellers, for all product markets and time steps, and the three considered models. Vertical lines represent the distributions median, showing that the model median is always smaller than the alternatives. The model better captures the temporal evolution of the top sellers degree for all product markets than the alternatives neglecting either the preferential attachment or the memory mechanism.

timely inform market design and regulation of online marketplaces.

Nevertheless, our work supports and extends previous findings. The fat-tailed heterogeneous curves in Figs. 1(a) to (d) substantiate previous observations of high concentration in DWMs: wholesale (47), few sellers (40), or few buyers (46) were found responsible for the largest part of volumes in smaller samples of data. The fat-tailed interevent time distributions, spanning times between a second and a year, are compatible with the bursty nature of several social activities (64, 65), and the finding about a shared memory kernel further points to a similarity between social and economic activities (49, 50). Taken together, our results could inform and enrich economic models where heterogeneity assumptions are now commonplace (14) but empirical evidence on the structure of buyer–seller networks has not yet been introduced.

The regularities observed in Fig. 1 are surprising given the differences in the marketplaces covered by our data: transactions on the clear web with state enforcement of contracts (66) vs. transactions on the dark web that rely mainly on reputation and self-governance (54); the sale of only regulated products vs. mainly unregulated products; and the use of fiat vs. the use of cryptocurrencies. And, indeed, there is both substantial heterogeneity in product markets in the e-commerce dataset and several differences across marketplaces in the DWM dataset (e.g. existence time period, geography, product focus, etc.). Our model suggests specific mechanisms that drive the regularities across the two datasets. Sellers build a reputation that makes them more attractive to buyers who, in turn, are affected by their memory of the sellers they already purchased from. In particular, the presence of both memory and preferential attachment is fundamental in reproducing both local and global properties of the buyer–seller network, as already shown for the intrinsically different social networks (49, 51, 52, 64). However, commercial interactions exhibit important differences compared to social interactions, with preferential attachment playing a dominant role in the market dynamics.

Our results point toward alternative strategies to attempt to reduce trading of illicit goods on DWMs. Historically, DWMs have been closed after long and expensive operations targeting the market admins in order to arrest them and shut down the servers (67). However, the high degree of concentration, the importance of preferential attachment, and the memory kernel in the buyer dynamics, all suggest that limited observations of the market dynamics could give a clear enough picture of who the key actors of these networks are. For instance, key sellers will most likely attract most of the observed purchases from the more active buyers, and stopping them would effectively stop a large part of the market trade. In this regard, our model could also be used to produce candidate synthetic DWM buyer–seller networks to quantitatively study and simulate the effects of targeting “key players” on marketplaces (68).

Finally, a better understanding of buyer–seller network formation could have consequences for market design and regulation. For example, fat-tailed distributions show a high degree of concentration on both buyer and seller sides of the marketplaces: just a few agents (both on the buyers and seller sides) are responsible for a vast majority of the transaction volume. While buyer market power appeared in analyses of labor monopsony and retailers (69, 70), our empirical finding of buyer concentration calls for a deeper understanding of buyer power in online marketplaces. Moreover, these observations can also inform theoretical economic models of online marketplaces, providing empirical backing to heterogeneity assumptions and suggesting specific values for parameters or shapes for distributions. Also, we find signs of both local (memory) and global (reputation) mechanisms in the structure and evolution of buyer–seller relationships. Thus, the inclusion of memory and reputation in previously developed models can improve our understanding of the pricing of network effects (12), interplatform competition (14) and long-run sustainability of the platforms (17).

Materials and methods

DWMs

DWMs are illegal unregulated commercial websites. They operate similarly to other online marketplaces, such as Gumtree or Craigslist. To improve anonymity DWMs are reached through browsers supporting the onion protocol, and use cryptocurrencies, mainly Bitcoin, as the main currency. While all Bitcoin transactions are publicly available, they record money exchanges between addresses, and a user can generate a new address (identifier) at each transaction to favor anonymity. As a result, the data need to be preprocessed to cluster addresses into individual entities in order to perform any economic analysis. Our dataset has been preprocessed by Chainalysis Inc. (71) to map addresses into entities (see Supplementary Material for more details).

Our dataset contains the entire transaction data of 28 entities corresponding to DWMs between 2011 June, and 2021 February (see Fig. S11 and Table S2 in the Supplementary Material). These markets all have an average daily volume of more than 15,000 USD, in order to be able to reliably measure different observables, and include all relevant DWMs as identified by law enforcement agencies (67). The data contain all transactions received or sent by DWMs, excluding services such as exchanges (Bitcoin trading exchanges allow users to trade Bitcoin). Note that the data hide the direct buyer–seller link, because the money pass through the platform during the transaction.

E-commerce platform

E-commerce platforms are regulated online marketplaces where sellers can post ads for products. Buyers and sellers can generally be both individuals or businesses. The payment is usually processed by the platform, but the shipping is handled by the seller. Sellers receive feedback from buyers, which together with product categorization helps people navigate the platform and choose what to buy.

The data used in this study consist in all the purchases made on 144 product markets from a popular e-commerce platform since 2010. The 144 product markets have been randomly selected from the markets that were active during the entire time period. The data cover only one geographical region. Similarly to the DWM data, the transaction data include: timestamp of the transaction, pseudonyms for buyer and seller, and the amount spent in the transaction. One key difference is that the data show the direct link between buyer and seller, forming a bipartite buyer–seller network and allowing for a more fine-grained analysis. For more details on the data see Fig. S12 in the Supplementary Material.

Model simulation

Each simulation is tuned to simulate one specific product market. We fix the agents population according to the data: number of sellers N , number of buyers M , and simulation total number of time steps T , to fix the average total number of transactions in our simulations as in the data: $(a_i \cdot \Delta t \cdot T \cdot N) = t$, where a_i is the buyer activity as defined in the main text, Δt is the simulation time step (fixed to 1) and t is the total number of transactions present in the data. We realize 30 different realizations for each parameter set, and aggregate the final results.

Preprints

A preprint of this article is published at <https://arxiv.org/abs/2112.09065>.

Supplementary Material

Supplementary material is available at [PNAS Nexus](#) online.

Funding

The research was partially supported by the Alan Turing Institute and the UK National Institute for Data Science and AI, and also thanks to a generous gift from PayPal. J.B. gratefully acknowledges the support of the Center of Mathematical Sciences and Applications at Harvard University.

Authors' Contributions

All authors designed the research; A.Br., J.B., and A.E. acquired the data. A.Br. and J.B. performed the measurements. All authors analyzed the data. All authors wrote the manuscript, discussed the results and commented on the manuscript.

Data Availability

Synthetic data reproducing the main properties of the original dataset are available in the public OSF repository dataset for the paper "Macroscopic properties of buyer–seller networks in online marketplaces" at <https://osf.io/rw2ce/>

References

- Christin N. 2013. Traveling the silk road: a measurement analysis of a large anonymous online marketplace. Proceedings of the 22nd International Conference on World Wide Web—WWW '13; Rio de Janeiro, Brazil.
- ElBahrawy A, et al. 2020. Collective dynamics of dark Web Marketplaces. *Sci Rep.* 10(1):18827.
- Décary-Héту D, Giommoni L. 2017. Do police crackdowns disrupt drug cryptomarkets? A longitudinal analysis of the effects of operation onymous. *Crime Law Soc Change.* 67(1):55–75.
- Bracci A, et al. 2021. Dark Web Marketplaces and COVID-19: before the vaccine. *EPJ Data Sci.* 10(1):6.
- Chevalier S. 2021. Annual net revenue of Amazon from 2004 to 2020. <https://www.statista.com/statistics/266282/annual-net-revenue-of-amazon/> [Accessed Sept 29, 2021]
- Standing S, Standing C, Love PED. 2010. A review of research on E-marketplaces 1997–2008. *Decis Support Syst.* 49(1):41–51.
- Bakos JY. 1997. Reducing buyer search costs: implications for electronic marketplaces. *Manag Sci.* 43(12):1676–1692.
- Choudhury V, Hartzel KS, Konsynski BR. 1998. Uses and consequences of electronic markets: an empirical investigation in the aircraft parts industry. *MIS Quarterly.* 22(4):471–507.
- Terzi N. 2016. The impact of E-commerce on international trade and employment. In: Özşahin Mehtap , editors. Encyclopedia of E-commerce development, implementation, and management, IGI Global. Amsterdam, the Netherlands: Elsevier. p. 2271–2287.
- Martin J. 2014. Drugs on the dark net: how cryptomarkets are transforming the global trade in illicit drugs. New York City, USA: Springer.
- Elliott M, Galeotti A. 2019. The role of networks in antitrust investigations. *Oxford Rev Econ Pol.* 35(4):614–637.
- Fainmesser IP, Galeotti A. 2016. Pricing network effects. *Rev Econ Stud.* 83(1):165–198.
- Armstrong M. 2006. Competition in two-sided markets. *RAND J Econ.* 37(3):668–691.

14. Ambrus A, Argenziano R. 2009. Asymmetric networks in two-sided markets. *Am Econ J Microecon*. 1(1):17–52.
15. Parker GG, Van Alstyne MW. 2005. Two-sided network effects: a theory of information product design. *Manage Sci*. 51(10):1494–1504.
16. Dellarocas C. 2003. The digitization of word of mouth: promise and challenges of online feedback mechanisms. *Manage Sci*. 49(10):1407–1424.
17. Eisenmann T, Parker G, Van Alstyne MW. 2006. Strategies for two-sided markets. *Harvard Bus Rev*. 84(10):92.
18. Luca M. 2017. Designing online marketplaces: trust and reputation mechanisms. *Innov Policy Econ*. 17:77–93.
19. Strader TJ, Ramaswami SN. 2002. The value of seller trustworthiness in C2C online markets. *Commun ACM*. 45(12):45–49.
20. Cabral L, Hortaçsu A. 2010. The dynamics of seller reputation: evidence from ebay. *J Indus Econ*. 58(1):54–78.
21. Tadelis S. 2016. Reputation and feedback systems in online platform markets. *Ann Rev Econ*. 8(1):321–340.
22. Chen Y, Yao S. 2017. Sequential search with refinement: model and application with click-stream data. *Manage Sci*. 63(12):4345–4365.
23. Ursu RM. 2018. The power of rankings: quantifying the effect of rankings on online consumer search and purchase decisions. *Mkt Sci*. 37(4):530–552.
24. De los Santos B, Koulayev S. 2017. Optimizing click-through in online rankings with endogenous search refinement. *Mkt Sci*. 36(4):542–564.
25. Koulayev S. 2014. Search for differentiated products: identification and estimation. *RAND J Econ*. 45(3):553–575.
26. Xu B, Lin Z, Shao B. 2010. Factors affecting consumer behaviors in online buy-it-now auctions. *Internet Res*. 20(5): 509–526.
27. Houser D, Wooders J. 2006. Reputation in auctions: theory, and evidence from eBay. *J Econ Manage Strat*. 15(2):353–369.
28. Bodoh-Creed AL, Boehnke J, Hickman B. 2021. How efficient are decentralized auction platforms? *Rev Econ Stud*. 88(1):91–125.
29. Bajari P, Hortacsu A. 2003. The winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions. *RAND J Econ*. 34(2):329.
30. Dall'Amico L, Fosset A, Bouchaud JP, Benzaquen M. How does latent liquidity get revealed in the limit order book? *J Stat Mech Theory Exp*. 2019(1):013404.
31. Bouchaud JP, Farmer JD, Lillo F. 2009. CHAPTER 2 - How markets slowly digest changes in supply and demand. In: Hens T, Schenk-Hoppé KR, editors. *Handbook of financial markets: Dynamics and evolution handbooks in finance*. San Diego (CA): North-Holland, p. 57–160.
32. Gorodnichenko Y, Talavera O. 2017. Price setting in online markets: basic facts, international comparisons, and cross-border integration. *Am Econ Rev*. 107(1):249–282.
33. Bodoh-Creed A, Boehnke J, Hickman B. 2020. Using machine learning to predict price dispersion. Working Paper.
34. DiRusso DJ, Mudambi SM, Schuff D. 2011. Determinants of prices in an online marketplace. *J Prod Brand Manage*. 20(5): 420–428.
35. Kim HW, Xu Y. 2007. Drivers of price premium in E-markets. *Commun ACM*. 50(11):91–95.
36. Lucking-Reiley D, Bryan D, Prasad N, Reeves D. 2007. Pennies from Ebay: the determinants of price in online Auctions. *J Indus Econ*. 55(2):223–233.
37. Broséus J, et al., 2016. Studying Illicit drug trafficking on darknet markets: structure and organisation from a Canadian perspective. *Forensic Sci Int*. 264:7–14.
38. Cunliffe J, Martin J, Décary-Héту D, Aldridge J. 2017. An Island Apart? Risks and prices in the Australian cryptomarket drug trade. *Int J Drug Pol*. 50:64–73.
39. Van Buskirk J, et al., 2016. Characterising dark net marketplace purchasers in a sample of regular psychostimulant users. *Int J Drug Pol*. 35:32–37.
40. Soska K, Christin N. 2015. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In: Jung Jaeyeon, editor. *Proceedings of the 24th USENIX conference on security symposium SEC'15*, Washington (DC); USENIX Association, p. 33–48.
41. Buxton J, Bingham T. 2015. The rise and challenge of dark net drug markets. *Policy Brief*. 7:1–24.
42. Nadini M, et al., 2022. Emergence and structure of decentralised trade networks around dark Web Marketplaces. *Sci Rep*. 12(1):1–9.
43. Martin J, Cunliffe J, Décary-Héту D, Aldridge J. 2018. Effect of restricting the legal supply of prescription opioids on buying through online illicit marketplaces: interrupted time series analysis. *BMJ (Clin Res Ed)*. 361 <https://www.bmj.com/content/361/8157>
44. Van Wegberg R, et al. 2018. Plug and prey? Measuring the commoditization of cybercrime via online anonymous markets. In: Enck William, Porter Felt Adrienne, editor. *Proceedings of the 27th USENIX conference on security symposium SEC'18*, Washington (DC); USENIX Association, p. 1009–1026.
45. Miller S, El-Bahrawy A, Dittus M, Graham M, Wright J. 2020. Predicting drug demand with wikipedia views: evidence from dark-net markets. In: *Proceedings of The Web Conference 2020*. Taipei, Taiwan: ACM, p. 2669–2675.
46. Norbutas L. 2018. Offline constraints in online drug marketplaces: an exploratory analysis of a cryptomarket trade network. *Int J Drug Pol*. 56:92–100.
47. Aldridge J, Décary-Héту D. 2016. Hidden Wholesale: the drug diffusing capacity of online drug cryptomarkets. *Int J Drug Pol*. 35:7–15.
48. Van Hout MC, Bingham T. 2014. Responsible vendors, intelligent consumers: Silk road, the online revolution in drug trading. *Int J Drug Pol*. 25(2):183–189.
49. Karsai M, Perra N, Vespignani A. 2015. Time varying networks and the weakness of strong ties. *Sci Rep*. 4(1):4001.
50. Ubaldi E, et al. 2016. Asymptotic theory of time-varying social networks with heterogeneous activity and tie allocation. *Sci Rep*. 6(1):35724.
51. Barabási AL, et al. 2002. Evolution of the social network of scientific collaborations. *Phys A Stat Mech Appl*. 311(3-4):590–614.
52. Jeong H, Néda Z, Barabási AL. 2003. Measuring preferential attachment in evolving networks. *Europhys Lett (EPL)*. 61(4):567–572.
53. European Monitoring Centre for Drugs and Drug Addiction. 2016. Reputation is everything: the role of ratings, feedback and reviews in cryptomarkets. In: Mounteney Jane, Bo Alessandra, Oteo Alberto, editor. *The internet and drug markets*. Publications Office, P. 49–54
54. Hardy RA, Norgaard JR. 2016. Reputation in the internet black market: an empirical and theoretical analysis of the deep web. *J Inst Econ*. 12(3):515–539.
55. Duxbury SW, Haynie DL. 2018. Building them up, breaking them down: topology, vendor selection patterns, and a digital drug market's robustness to disruption. *Soc Netw*. 52:238–250.
56. Alessandretti L, Sun K, Baronchelli A, Perra N. 2017. Random walks on activity-driven networks with attractiveness. *Phys Rev E*. 95(5):052318.

57. Perra N, Gonçalves B, Pastor-Satorras R, Vespignani A. 2012. Activity driven modeling of time varying networks. *Sci Rep.* 2(1):469.
58. Goolsbee AD, Klenow PJ. 2018. Internet rising, prices falling: measuring inflation in a world of e-commerce. *AEA Pap Proc.* 108:488–92.
59. Gorodnichenko Y, Sheremirov V, Talavera O. 2018. Price setting in online markets: Does it click? *J Eur Econ Assoc.* 16(6):1764–1811.
60. Nosko C, Tadelis S. 2015. The limits of reputation in platform markets: an empirical analysis and field experiment. National Bureau of Economic Research (US). Working Paper 20830.
61. Li SS, Karahanna E. 2015. Online recommendation systems in a B2C E-commerce context: a review and future directions. *J Assoc Inf Syst.* 16(2):2.
62. Mandel A, Venel X. 2021. Sequential competition and the strategic origins of preferential attachment. *Int J Game Theo.* 1–26. <https://link.springer.com/article/10.1007/s00182-021-00801-0>
63. Hsieh CS, König M, Liu X. 2017. Network formation with local complements and global substitutes: the case of R&D networks. Department of Economics, University of Zurich, Working Paper. 217.
64. Barabási AL. 2005. The origin of bursts and heavy tails in human dynamics. *Nature.* 435(7039):207–211.
65. Moinet A, Starnini M, Pastor-Satorras R. 2015. Burstiness and aging in social temporal networks. *Phys Rev Lett.* 114(10): 108701.
66. Wehinger F. 2011. The dark net: self-regulation dynamics of illegal online markets for identities and related services. In: 2011 European Intelligence and Security Informatics Conference. Athens: IEEE, p. 209–213.
67. European Monitoring Centre for Drugs and Drug Addiction. 2017. Drugs and the darknet: perspectives for enforcement, research and policy. Luxembourg: Publications Office, P. 1–90
68. Ballester C, Calvó-Armengol A, Zenou Y. 2006. Who's who in networks. Wanted: the key player. *Econometrica.* 74(5): 1403–1417.
69. Boal WM, Ransom MR. 1997. Monopsony in the labor market. *J Econ Lit.* 35(1):86–112.
70. Dube A, Jacobs J, Naidu S, Suri S. 2020. Monopsony in online labor markets. *Am Econ Rev Insights.* 2(1): 33–46.
71. Chainalysis Inc., 2020. <https://www.chainalysis.com/>