nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0638-y

Emergence of scaling in complex substitutive systems

Ching Jin[1,2,3,7], Chaoming Song[4,7], Johannes Bjelland[5], Geoffrey Canright[5] and Dashun Wang[1,2,6*]

Diffusion processes are central to human interactions. One common prediction of the current modelling frameworks is that initial spreading dynamics follow exponential growth. Here we find that, for subjects ranging from mobile handsets to automobiles and from smartphone apps to scientific fields, early growth patterns follow a power law with non-integer exponents. We test the hypothesis that mechanisms specific to substitution dynamics may play a role, by analysing unique data tracing 3.6 million individuals substituting different mobile handsets. We uncover three generic ingredients governing substitutions, allowing us to develop a minimal substitution model, which not only explains the power-law growth, but also collapses diverse growth trajectories of individual constituents into a single curve. These results offer a mechanistic understanding of power-law early growth patterns emerging from various domains and demonstrate that substitution dynamics are governed by robust self-organizing principles that go beyond the particulars of individual systems.

Diffusion processes impact broad aspects of human society[1–5], ranging from the spread of biological viruses[3,6–8] to the adoption of innovations[4,9–14] and knowledge[15,16] and to the spread of information[17–19], cultural norms and social behaviour[20–23]. Despite numerous studies that span multiple disciplines, our knowledge is mainly limited to spreading processes in non-substitutive systems. Yet, a considerable number of ideas, products and behaviours spread by substitution—to adopt a new one, agents often need to give up an existing one. For example, the development of science hinges on scientists' relentlessness in abandoning a scientific framework once one that offers a better description of reality emerges[24]. The same is true for adopting a new healthy habit or other durable items, such as mobile phones, cars or homes.

While substitutions play a key role from science to economy, our limited understanding of such processes stems from the lack of empirical data tracing their characteristics. To study the dynamics of substitutions, we explore growth patterns in four different substitutive systems where detailed dynamical patterns are captured with fine temporal resolution (see Supplementary Note 1 for detailed data descriptions). Our first dataset captures, with daily resolution, 3.6 million individuals choosing among different types of mobile handsets, recorded by a Northern European telecommunication company from January 2006 to November 2014. As an individual is unlikely to keep more than one mobile phone at a time, their adoption of a new handset is typically associated with discontinuance of the old one. Here, we focus on handsets that have been released for at least 6 months and used by at least 50 users in total (885 different handset models). Our second dataset captures monthly transaction records of 126 automobiles sold in the North America between 2010 and 2016. These automobiles have been released for at least 4 months before the data were collected. Automobiles represent a similar example as mobile handsets, in which adoptions are largely driven by substitutions, given the limited number of automobiles a typical household may have.

While handset and automobile adoptions are relatively exclusive, in reality, there are also 'hybrid' substitutive systems, in which the definition of substitutions is less strict. To test whether results presented in this paper may apply to such systems, we collected two additional datasets: one dataset traces the number of daily downloads for new smartphone apps published in the App store (2,672 most popular apps in the iOS systems from November to December 2016), and the other dataset is a scientific publication dataset, recording 246,630 scientists substituting for 6,399 scientific fields from 1980 to 2018. Indeed, usages of smartphone apps are subject to constraints of time and device space; hence, a new app downloaded reduces the usage of other similar apps, if not replacing them all together. Yet, at the same time, apps may also be downloaded without involving substitutions. Similarly, while many scientists may focus on one research area at a time[25], where research direction shifts may be characterized by substitutions, there are also people who explore several directions simultaneously and, therefore, an increased focus on one direction does not necessarily imply a decreased attention to others.

Results

A common prediction by current modelling frameworks, from epidemiological models[3,6] to disordered systems[2] to diffusion of innovations[4,12], is that early growth patterns follow an exponential function. To test this prediction, we measure the impact of each mobile handset, automobile model, smartphone app and scientific field in our four datasets. More specifically, we calculated $I(t)$, which measures the number of individuals who bought the handset up to time t since its availability (Fig. 1a), cumulative sales of an automobile (Fig. 1b), daily downloads of an app (Fig. 1c) and the number of publishing scientists in a field (Fig. 1d), respectively. To compare across different constituents, we normalized $I(t)$ by its initial value $I(1)$ (that is, the first day or year when the constituent was introduced), and first focused on their early growth periods only (Supplementary Note 1).

[1]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA. [2]Kellogg School of Management, Northwestern University, Evanston, IL, USA. [3]Center for Complex Network Research, Northeastern University, Boston, MA, USA. [4]Department of Physics, University of Miami, Coral Gables, FL, USA. [5]Telenor Research and Development, Fornebu, Norway. [6]McCormick School of Engineering, Northwestern University, Evanston, IL, USA. [7]These authors contributed equally: Ching Jin, Chaoming Song. *e-mail: dashun.wang@kellogg.northwestern.edu

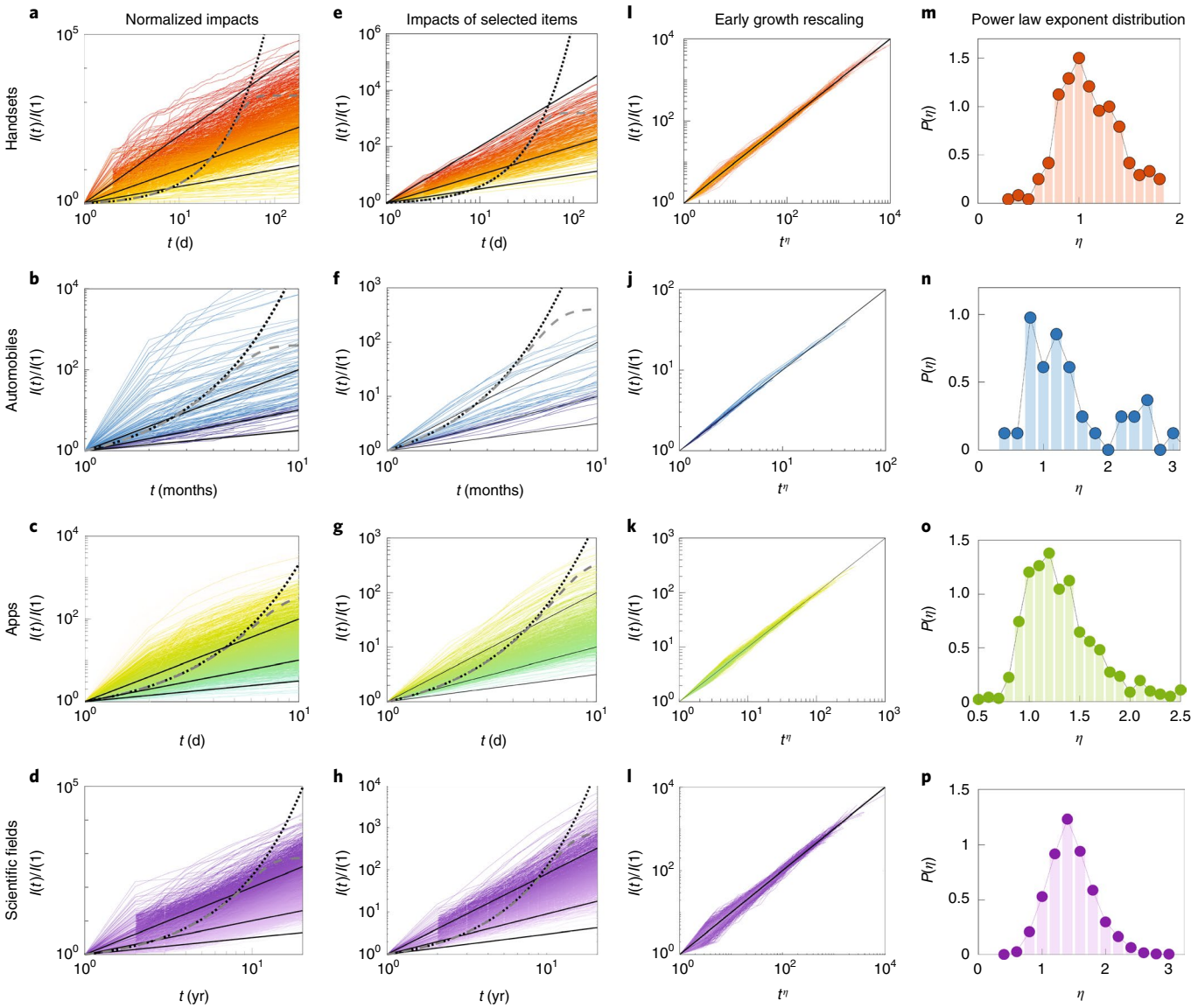

Fig. 1 | Power-law growth patterns in substitutive systems. a, Normalized impacts of all 885 handsets, which have been released for at least 6 months and used by 50 users in total (Supplementary Note 1). To compare different curves, we normalized $I(t)$ by $I(1)$, the number of users on the first day of release. We use the first 6 months to measure the early growth phase for each handset, finding that a considerable number of products do not follow exponential ($y \sim e^x$, black dotted line) or logistic growth ($y \sim L/(1+e^{-k(x-x_0)})$, grey dashed line). Instead, they prefer power-law growth patterns (for the statistical tests used for growth comparison, see Supplementary Note 1). Here we fitted each curve with a power law function $I(t) \sim t^{\eta}$. The colour of the line corresponds to the fitted power-law exponent for each handset. The solid lines are $y = x^{1/2}$, $y = x$ and $y = x^2$, respectively, as guides to the eye. **b**–**d**, Normalized impacts similar to those in **a** of 126 automobiles (**b**), 2,672 smartphone apps (**c**) and 6,399 scientific fields (**d**). Here we show the early growth pattern of all products whose records are longer than their early growth period (4 months for automobiles, 7 d for smartphones apps and 18 years for scientific fields), finding again that a large number of products prefer power-law growth patterns than exponential functions. Note that the exponential and logistic curves are shown as a guide to the eye, meant to highlight a conceptual difference between exponential and power-law functions. Interested readers should refer to Supplementary Figs. 4, 12 and 14 for more quantitative evaluations. **e**, Normalized impacts of 240 different handsets as a function of time. We find that, for a substantial fraction of handsets (240 of 885 handsets (27.12%)), their early growth patterns can be well approximated by power laws ($R^2 \geq 0.99$): $I(t) \sim t^{\eta}$. The colour of the line corresponds to the associated power-law exponent for each handset, η. The exponential and logistic curves are shown as guides to the eye as those in **a**, highlighting their fundamentally different nature compared with power-law growth patterns (see Supplementary Note 1 for the statistical tests used for fitting). **f**–**h**, Power-law growth patterns similar to those in **e** are observed in three other datasets, where we find that growth patterns for 37 of 126 cars (29.37%) (**f**), 1,022 of 2,672 apps (38.25%) (**g**) and 1,743 of 6,399 scientific fields (27.24%) (**h**) can be well approximated by power laws. **i**–**l**, We rescale the impact dynamics plotted in **e**–**h**, respectively, by t^{η}, finding that all curves collapse into $y = x$. **m**–**p**, Distribution of power-law exponents $P(\eta)$ for curves shown in **e**–**h**, respectively.

We find that, in contrast to the exponential curves predicted by canonical models, for many of the constituents across the four systems, their growth trajectories appear to follow straight lines on a log–log plot (Fig. 1a–d), suggesting that they may be described by power-law functions. This observation prompts us to systematically test whether power-law or exponential-class functions

(exponential or logistic) are preferred to describe the early growth curves observed in our four systems. Using the Akaike information criterion, we find that 98.6% handsets, 83.5% automobiles, 79.6% apps and 74.1% scientific fields favour power-law early growth patterns (Supplementary Note 1 and Supplementary Fig. 12). We further tested the robustness of this result by applying different statistical tests (Supplementary Note 1) and by varying the definition of early growth periods in each dataset (Supplementary Fig. 14), and for both cases, we arrived at the same conclusion.

Note that, although for a vast majority of the curves (80.18–99.21%) power law offers a better fit than exponential-class models (see Supplementary Note 1 and Supplementary Figs. 4, 12 and 14), there is variability in how well a power-law function fits different curves. Moreover, there is a small fraction (0.79–19.82%) of constituents whose early growth patterns can be described by exponential functions, suggesting that, for these constituents, their growth patterns are consistent with the predictions of existing models. To ensure that our fitting procedure is not biased against exponential functions, we analysed spreading patterns of 168 cases of flu pandemics in the United States, where early growth patterns are expected to follow exponential function. We find that the fitting results indeed systematically prefer exponential function to power law (Supplementary Note 1 and Supplementary Fig. 15). Together, Fig. 1a–d suggest the existence of a non-trivial fraction (74.1–98.6%) of constituents, whose early growth patterns follow a power-law function rather than an exponential function.

To examine whether there is indeed a fraction of growth trajectories that can be well described by power-law growth patterns, we further restrict the criteria for classifying power laws by selecting those with a high R^2 in fitting (for example, $R^2 > 0.99$). We find that, under the stricter criteria, a substantial fraction of constituents remained in each of the four systems (27.12% handsets, 29.37% automobiles, 38.25% apps and 27.24% scientific fields) (Fig. 1e–h). The results indicate that, for a substantial fraction of constituents across the four substitutive systems that we studied, their impacts grow following

$$I(t)/I(1) = t^{\eta_i} \quad (1)$$

We also noticed that within each system, the slopes of power-law curves shown in Fig. 1e–h differ across different constituents, suggesting that each of them is characterized by constituent-specific exponents (η_i). To test this hypothesis, we plotted each curve in Fig. 1e–h in terms of t^{η_i}. As constituents differ from each other, the rescaled curves show variations around the function $y = x$. Yet, we find that most curves are reasonably collapsed onto the same function (Fig. 1i–l). The rescaled growth patterns for all products across our four datasets are also shown in Supplementary Fig. 6. We find that, although as expected, their growth patterns show more variations around $y = x$, they are clearly different from exponential growth patterns.

The observations documented in Fig. 1a–l are somewhat unexpected for two main reasons. First, the four systems that we studied differ widely in their scope, scale, temporal resolution and user demographics. Yet, we find, independent of the nature of the system and the identity of the constituents, their early growth follows similar patterns, showing that a power-law scaling emerges across all four systems. Second, exponents η_i are mostly non-integers (Fig. 1m–p). Power-law growth with such non-integer exponents is rare because it corresponds to non-analytic behaviour. Indeed, due to the inability to express them in terms of Taylor series around $t = 0$, power laws with non-integer exponents indicate singular behaviour around the release time (the $\lceil \eta \rceil$-th order derivative diverges at $t = 0$). Current modelling frameworks[2–4,6,12] rely on functions without singularities and, hence, are unable to anticipate non-analytic solutions (detailed descriptions and comparisons to existing models are described in Supplementary Notes 1 and 3). Indeed, compared with exponential growth, power law encodes an early divergence, corresponding to an explosive growth at the moment when new constituents are introduced. Yet, following this brief singularity, the number of users grows much more slowly than what exponential functions predict, suggesting that substitutive innovations spread more slowly beyond the initial excitement.

Keep in mind, however, that not all curves follow power-law growth patterns, and a few of them can indeed be described by exponential functions, suggesting that substitutions and traditional adoptions may coexist in our systems. Nevertheless, these results document the existence of power-law early growth curves in the substitutive systems that we studied, a pattern that is not anticipated by traditional modelling frameworks, and suggests that substitutive systems may be governed by different dynamics.

To be sure, power laws can be generated in real networks due to the growth of the systems[26,27]. To check whether Fig. 1 may be explained by gradual addition of new users to the underlying network, we removed new mobile subscribers in the mobile phone dataset and again measured $I(t)$ for different handsets. We found that the power-law scaling holds the same (Supplementary Note 3 and Supplementary Fig. 19), indicating that the scaling observed in Fig. 1 is governed by mechanisms that operate within the system and is not driven by growth of the system. Another possible origin of power-law growth is rooted in the bursty nature of human behaviour[28,29], in which the inter-event time between adoptions follows a power-law distribution. We measured this quantity directly in the mobile phone dataset, finding that the data systematically reject power law as a viable function to describe the inter-event time distribution ($P < 10^{-3}$; Supplementary Fig. 20). It is also worth noting that sub-exponential growth patterns have recently been found in the spread of epidemics such as Ebola and human immunodeficiency virus (HIV) infection[30–32]. There are also phenomenological models of spreading dynamics that take power-law early growth as their assumptions[31,33,34], in addition to a large body of literature on modelling popularity dynamics[35–41]. While a mechanistic explanation is still lacking, these examples demonstrate that the power-law early growth patterns uncovered here may hold relevance to a broad array of areas. Together, these results raise a fundamental question: what is the origin of the power-law growth pattern.

Quantifying substitution patterns. A common characteristic of the four studied systems is that they evolve by substitutions. In this respect, mobile phones represent an ideal setting for the empirical investigation of substitutive processes. Indeed, each time a user purchases a handset, the transaction history is recorded by telecommunication companies. Anonymized phone numbers together with their portability across devices provide individual traces for substitutions. We examined detailed user histories in the mobile phone dataset, finding that the adoption and discontinuance histories are indeed predominantly represented by substitutions (Supplementary Note 2). Each type of handset is substituted by a large number of other handsets, hence substitution patterns are characterized by a dense, heterogeneous network that evolves rapidly over time ($\langle k \rangle = 73.6$; Supplementary Figs. 17b,c and 18). To visualize substitution patterns, we applied a backbone extraction method[42] to identify statistically significant substitution flows for each handset given its total substitution volumes (Fig. 2). While mobile handsets have changed substantially over the years, undergoing a ubiquitous shift from feature phones to smartphones, the rate at which new handsets enter the market remained remarkably stable (Fig. 3a), highlighting the highly competitive nature of the system: ensuing generations of new handsets enter the market in a somewhat regular manner, substituting for the incumbent, thereby affecting the rise and fall of their popularities (Supplementary Fig. 18a).

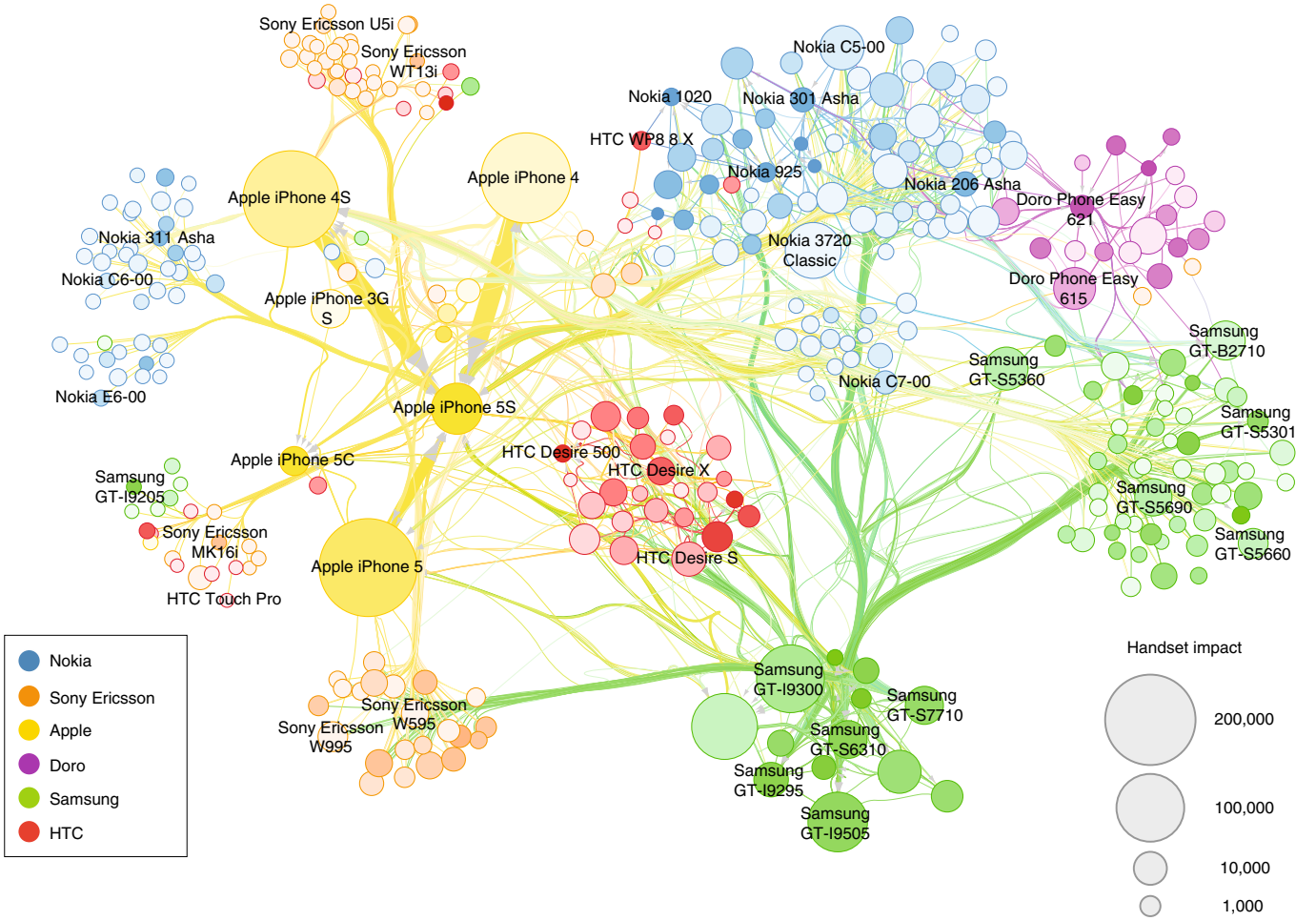

Fig. 2 | Empirical substitution network. We used the backbone extraction method[42] to construct a substitution network, capturing substitution patterns among handsets aggregated within a 6-month period (January–June 2014). Each node corresponds to one type of handset released before 2014 by one of the six major manufacturers. The node size captures its popularity, measured by the number of users of the particular handset at the time. Handsets are coloured based on their manufacturers (node colouring), which fade with the age of the handsets. If users substituted handset *i* with *j*, we add a weighted arrow pointing from *i* to *j*. The link weight captures the total substitution volumes between two handsets within the 6-month period. As the full network is too dense to visualize, here we only show the statistically significant links as identified by the method proposed in ref. [42] for a *P* value of 0.05. We colour the links based on the colour of the substituting handset. The network vividly captures the widespread transitions from feature handsets to smartphones. Indeed, most cross-manufacturer substitution links are either yellow or green, indicating their substitutions by iPhones or Android handsets. Substitution patterns are also highly heterogeneous. A few pairs of handsets have high substitution volumes, for example, between the successive generations of iPhones, but most substitutions are characterized by rather limited volumes. The structural complexity shown is further coupled with a high degree of temporal variability. Indeed, the system turns into a widely different configuration every year, even for the most dominant handsets (Supplementary Fig. 18b–e).

To uncover the mechanisms governing substitution dynamics, we note that the rate of change in $N_i(t)$, the number of users for handset i at time t, can be expressed in terms of the probability for individuals to transition from all other handsets (k) to i, $\Pi_{k\to i}$, subtracted by those leaving i for other handsets (j), $\Pi_{i\to j}$:

$$\frac{\mathrm{d}N_i(t)}{\mathrm{d}t} = \sum_k \Pi_{k\to i}(t)N_k(t) - \sum_j \Pi_{i\to j}(t)N_i(t) \qquad (2)$$

The key to solving the master equation (2) is to determine $\Pi_{i\to j}$, the substitution probability for a user to substitute handset i for j at time t. As we show next, $\Pi_{i\to j}$ is driven by three mechanisms: preferential attachment, recency and propensity.

Figure 3b shows that $\Pi_{i\to j}$ is independent of the number of individuals using i (N_i), but proportional to N_j: $\Pi_{i\to j} \sim N_j$. This result captures the well-known preferential attachment effect[15,26]: more popular handsets are more likely to attract new users than their less popular counterparts, consistent with existing models that can be used to characterize substitutions[43,44]. Yet N_j by itself is insufficient to explain $\Pi_{i\to j}$. Indeed, we further normalized $\Pi_{i\to j}$ by N_j, by defining $S_{i\to j} \equiv \Pi_{i\to j}/N_j$, the substitution rate at which handset j substitutes for i. We find that $P(S_{i\to j})$ follows a fat-tailed distribution spanning several orders of magnitude (Fig. 3c), indicating that substitution rates are characterized by a high degree of heterogeneity, in which $S_{i\to j}$ between some handset pairs are orders of magnitude higher than others.

To identify mechanisms responsible for the observed heterogeneity in $S_{i\to j}$, we grouped $S_{i\to j}$ based on the age of the substitutes t_j, the number of days elapsed since its release date, and measured the conditional probability $P(S_{i\to j}|t_j)$ for each group. We found that as substitutes grow older (increasing t_j), $P(S_{i\to j}|t_j)$ shifts systematically to the left (Fig. 3d), indicating that substitution rates decrease with the age of substitutes—newer handsets substitute for the incumbents

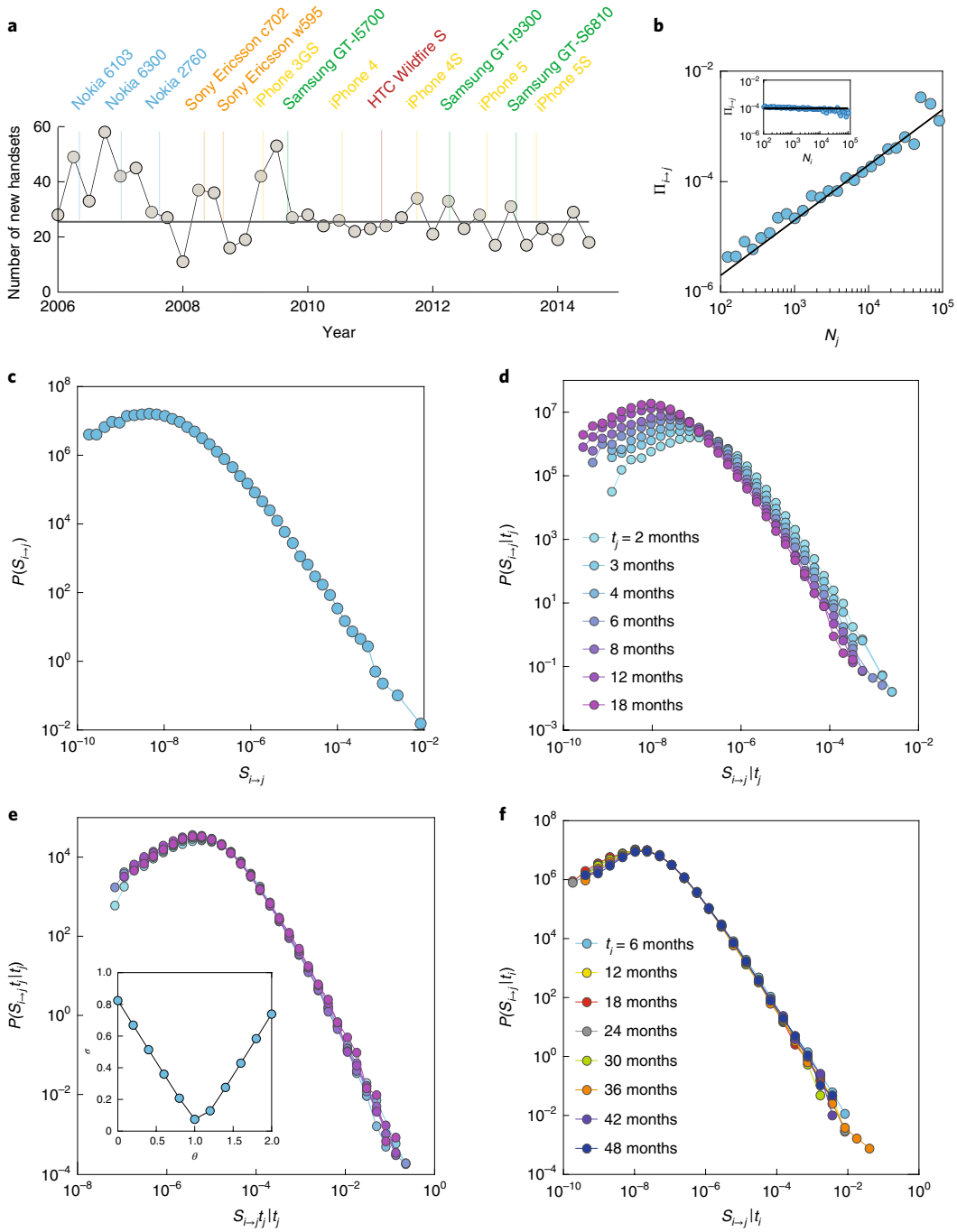

Fig. 3 | Empirical substitution patterns. a, The number of new handsets launched per quarter as a function of time. We find that new handsets are introduced at a constant rate. The most popular handset within each 8-month time window is highlighted and the colours correspond to those used in Fig. 2. **b**, Substitution probability $\Pi_{i\to j}$ is proportional to N_j, consistent with the preferential attachment effect. The inset shows that $\Pi_{i\to j}$ is largely independent of N_i. The measurement is based on eight snapshots of observations sampled uniformly in time. Specifically, we choose the first month of each year from 2007 to 2014 to measure substitution flows to test the preferential attachment hypothesis. **c**, Distribution of substitution rates $S_{i\to j}$. Here we show the distribution as a probability density function and we measured the substitution rates among handsets in January 2014. **d**, Distribution of substitution rates $S_{i\to j}$ conditional on the age of the substitute t_j. The distributions shift systematically to the left as t_j increases. **e**, After rescaling substitution rates by t_j^{-1}, we measure $P(S_{i\to j}t_j|t_j)$, finding that all seven curves in **d** collapse into one single curve. In the inset, we tested the relationship between $S_{i\to j}$ and t_j by rescaling $S_{i\to j}$ by $t_j^{-\theta}$, finding that the curves collapse onto each other when $\theta=1$. **f**, Distribution of substitution rates conditional on the age of the incumbent (t_i). All curves collapse automatically onto one single distribution, indicating an independence between substitution rates and the age of incumbent handsets.

at a higher rate. Yet, within each group, the heterogeneity of $S_{i\to j}$ persisted, as $P(S_{i\to j}|t_j)$ again follows a fat-tailed distribution. However, once we rescaled the distributions $P(S_{i\to j}|t_j)$ with t_j, we found that all seven distributions in Fig. 3d collapse into one single curve (Fig. 3e). To quantify the relationship between $S_{i\to j}$ and t_j, we took an ansatz $S_{i\to j}\sim t_j^{-\theta}$ and rescaled $S_{i\to j}$ by $t_j^{-\theta}$. As we varied θ, we monitored the diversity of the curves, finding that it reaches its minimum around $\theta=1$ (Fig. 3e, inset), indicating that $S_{i\to j}$ is inversely proportional

to t_j. The data collapse in Fig. 3e demonstrates that a single distribution characterizes substitution rates, independent of the age of substitutes:

$$P(S_{i\to j}|t_j) \sim t_j \mathcal{F}(S_{i\to j} t_j) \quad (3)$$

In other words, substitution rates $S_{i\to j}$ can be decomposed into two independent factors: one is the universal function $\mathcal{F}(x)$, which is independent of the substitute's age, capturing an inherent propensity-based heterogeneity among handsets. Denoting the propensity by $\lambda_{ij} \equiv S_{i\to j} t_j$, equation (3) indicates $S_{i\to j} \sim \lambda_{ij} \frac{1}{t_j}$. We repeated our analysis for t_i, that is, the age of incumbent handset i when substituted, finding that all curves of $P(S_{i\to j}|t_i)$ automatically collapsed onto each other (Fig. 3f). Hence, when incumbents are substituted, whether they were released merely a few months ago (small t_i) or have existed in the market for years (large t_i), their substitution rates follow the same distribution, documenting an independence between substitution rates and the age of the incumbents. Mathematically, Fig. 3f indicates $P(S_{i\to j}|t_i) = P(S_{i\to j})$.

Together, Fig. 3d–f helped us to uncover two more mechanisms governing substitutions, recency and propensity: substitution rates depend on the recency of substitutes, following a power law $1/t_j$. The uncovered power-law decay has a simple origin, documenting the role of competitions in driving the obsolescence of handsets. Indeed, when j first entered the system, being the latest handset (small t_j), it substitutes for the incumbent at its highest rate. Yet with time, more and more newer handsets are introduced. The constant rate of new arrivals (Fig. 3a) implies that the number of alternatives to j grows linearly with t_j. Hence, if we pick one handset randomly, the probability for handset j to stand out among its competitors decays as $1/t_j$. The temporal decay is further modulated by the inherent propensity λ_{ij} between two handsets, capturing the extent to which a certain handset is more likely to substitute for some handsets than others. Taken together, Fig. 3b–f predict

$$\Pi_{i\to j} = \lambda_{ij} N_j \frac{1}{t_j} \quad (4)$$

Minimal substitution model. Most importantly, equation (4) defines a minimal substitution model, which, as we show next, naturally leads to the observed power-law early growth patterns. In this model, the system consists of a fixed number of individuals, with new handsets being introduced constantly (Fig. 3a). In each time step, an individual substitutes their current handset i for new handset j with probability $\Pi_{i\to j}$, according to equation (4). The propensity λ_{ij} between handset i and j is drawn randomly from a fixed distribution. Our results are independent of specific distributions of λ_{ij}. We can solve our model analytically in its stationary state (Fig. 3a) by plugging equation (4) into equation (2), yielding (Supplementary Note 4):

$$N_i(t_i) = h_i t_i^{\eta_i} e^{-t_i/\tau_i} \quad (5)$$

indicating that the number of individuals using handset i is governed by three parameters: η_i, h_i and τ_i. $\eta_i \equiv \sum_k \lambda_{k\to i} N_k$ captures the fitness of a handset, measuring the total propensity for users to switch from all other handsets to i. The anticipation parameter h arises from the boundary condition at $t_i = 0$ when solving the differential equation (2), approximating the number of individuals using handset i when $t_i = 1$, which captures users' initial excitement for a particular handset. τ_i is the longevity parameter, as it captures the characteristic time scale for i to become obsolete. Indeed, defining t_i^* as the time when a handset reaches its maximum number of users, equation (5) predicts that the peak time t_i^* is proportional to its longevity parameter and fitness: $t_i^* = \eta_i \tau_i$.

The impact of handset i, that is, its cumulative sales, can be calculated by integrating all transition flows from other handsets to i before t_i: $I_i(t_i) = \int_0^{t_i} \sum_k \Pi_{k\to i} N_k \mathrm{d}t$, yielding:

$$I_i(t_i) = h_i \eta_i \tau_i^{\eta_i} \gamma_{\eta_i}(t_i/\tau_i) \quad (6)$$

where $\gamma_\eta(t) \equiv \int_0^t x^{\eta-1} e^{-x} \mathrm{d}x$ is the lower incomplete gamma function. Hence, in the early stage of a life cycle (small t_i), equation (6) predicts that the impact of handset i grows following a power law:

$$I_i(t_i) = h_i t_i^{\eta_i} \quad (7)$$

where the growth exponent is uniquely determined by the fitness parameter η_i, equivalent to the power-law exponent discovered in equation (1). Equation (7) indicates that the specific power-law exponent for each constituent is governed by its propensity to substitute for the incumbents in the system. The higher the fitness, the steeper the power-law slope, hence, the faster the take-off in the number of users. The power-law growth is further modulated by the anticipation parameter h, capturing the impact difference during the initial release. Note that it may take some time for model parameters to reach their stationary state, which may affect the validity of equation (5) and equation (7). To this end, we performed agent-based simulations of the model, finding that the parameters reach stationary states faster than the empirical time scale that we measured (Supplementary Notes 4 and Supplementary Fig. 21).

Universal impact dynamics. The minimal substitution model not only explains the early growth phase but it also predicts the entire life cycle of impacts (Supplementary Note 4). By using the rescaled variables: $\tilde{t}_i = t_i/\tau_i$ and $\tilde{I}_i = I_i/(h_i \eta_i \tau_i^{\eta_i})$, we obtain:

$$\tilde{I}_i = \gamma_{\eta_i}(\tilde{t}_i) \quad (8)$$

Therefore, for handsets with the same fitness, their impact dynamics can be collapsed into a single function after being rescaled by the three independent parameters (η, τ and h). Most interestingly, as the rescaling formula (equation (8)) is independent of the particulars of a system, it predicts that, constituents from different systems should all follow the same curve as long as they have the same fitness.

To test these predictions, we fit our model (equation (6)) to all four systems using maximum-likelihood estimation (Supplementary Note 4) to obtain the best-fitted three parameters (η_i, h_i and τ_i) for each handset, automobile, smartphone app and scientific field. We first selected from the four systems, those with similar fitness ($\eta = 1.5$). Although their impact dynamics appear different from each other (Fig. 4a–d), we found that all curves simultaneously collapsed into one single curve after rescaling (Fig. 4e–h). To test for variable fitness, we selected two additional groups of handsets ($\eta = 1.8$ and $\eta = 2.0$), finding that the rescaled impact dynamic in both groups can be well approximated by their respective universality classes predicted by equation (8) (Fig. 4i,j). The universal curves correspond to the associated classes of the incomplete gamma functions $\gamma_{\eta_i}(\tilde{t}_i)$, which only depend on the fitness parameter η (Fig. 4k). The model also predicts that, if we properly normalize out the effect by $\gamma_{\eta_i}(\tilde{t}_i)$, we can rescale the entire life cycle to a power law solely governed by η. Indeed, equation (6) indicates that, by defining $Q(t) \equiv [I(t)/h - \tau^\eta \gamma_{\eta+1}(t/\tau)] e^{t/\tau}$, Q should grow following a power law, $Q(t) = t^\eta$ (Supplementary Note 4). We find agreement across the four systems that we studied (Fig. 4l–o). Together, Fig. 4a–o documents regularities governing impact dynamics, which appear to hold both within a system and across different complex substitutive systems. Given the diversity of the studied systems and the numerous factors

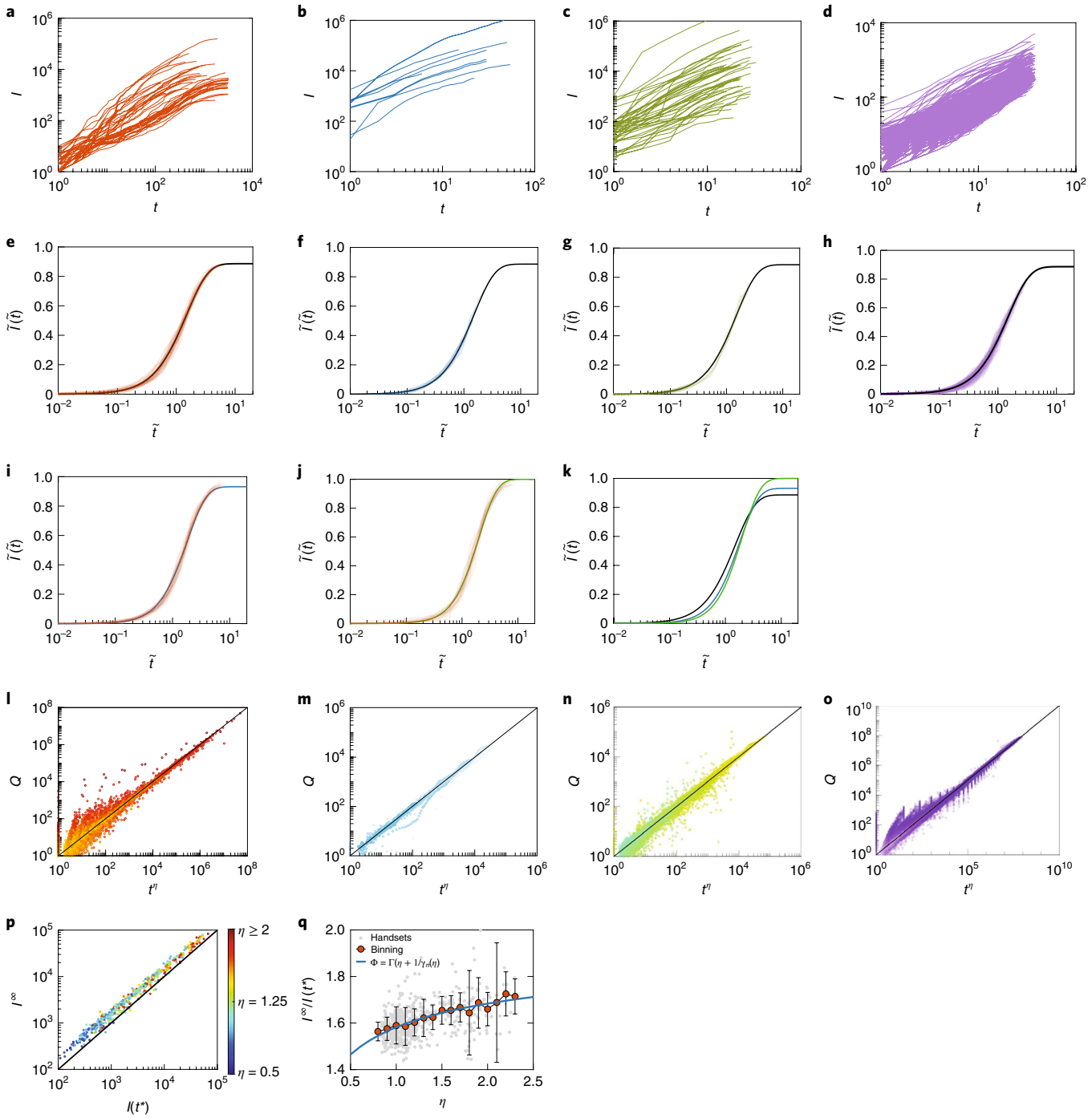

Fig. 4 | Universal impact dynamics. a–d, Impact dynamics for products with similar fitness ($\eta = 1.5 \pm 0.1$), including 40 handsets (**a**), 9 automobiles (**b**), 43 apps (**c**) and 505 scientific fields (**d**). **e–h**, Data collapse for products shown in **a–d**, respectively. After rescaling time and impact independently by $\tilde{t}_i = t_i/\tau_i$ and $\tilde{I}_i = I_i/(h_i \eta_i \tau_i^{\eta_i})$, we found that all curves from four systems collapse into the same universal curve, as predicted by equation (8). **i**, Data collapse for handsets with similar fitness $\eta = 1.8 \pm 0.1$ (30 handsets). **j**, Data collapse for handsets with similar fitness $\eta = 2.0 \pm 0.1$ (22 handsets). **k**, The universal functions shown in **e–j** are each associated with their respective universality classes that are solely determined by η. Here we visualize the analytical function $\tilde{I} = \gamma_{\eta}(\tilde{t})$, with $\eta = 1.5$ (black line), 1.8 (blue line) and 2.0 (green line). **l–o**, The entire life cycle can be rescaled as power laws if we properly normalize the effect from the incomplete gamma functions. Indeed, because the function $\gamma_{\eta}(x)$ has recurrence property $\gamma_{\eta+1}(x) = \eta\gamma_{\eta}(x) - x^{\eta}e^{-x}$, equation (6) predicts that, by defining $Q \equiv (I(t)/h - \tau^{\eta}\gamma_{\eta+1}(t/\tau))e^{t/\tau}$, we should expect $Q = t^{\eta}$. Here we plot Q as a function of t^{η} for all fitted products in the four systems (handset (**l**), automobile (**m**), app (**n**) and scientific field (**o**)), where the colour of each line corresponds to the learned fitness parameter (see also Supplementary Note 4 and Supplementary Figs. 25–28 for discussions about curve collapse and comparison with other models.) **p**, I^{∞} as a function of $I(t^{*})$ for the handsets with different fitness η shown in **l**. I^{∞} and t^{*} are calculated through the system parameters: h, η and τ. $I(t^{*})$ is the handset's impact at time t^{*} obtained from the empirical data (Supplementary Note 4). **q**, Scatter plot for the ratio $I^{\infty}/I(t^{*})$ as a function of η for the same handsets shown in **p**. The error bars indicate 1 s.d. The solid line corresponds to the analytical prediction by equation (10).

that determine the dynamics of spreading processes, ranging from initial seeds and timing[45,46] to social influence[13,22] to a large set of often unobservable factors[47], this level of agreement is somewhat unexpected.

Linking short-term and long-term impacts. The minimal substitution model predicts an underlying connection between short-term and long-term impact. Indeed, we can calculate the ultimate impact—the total number of a particular handset, automobile, smartphone app or scientific field, ever sold, downloaded or studied in its lifetime—by taking the $t \to \infty$ limit in equation (6), obtaining:

$$I_i^{\infty} = h_i \Gamma(\eta_i + 1) \tau_i^{\eta_i} \tag{9}$$

where $\Gamma(z) \equiv \int_0^{\infty} x^{z-1} e^{-x} \mathrm{d}x$ corresponds to the gamma function. A comparison of equation (6) and equation (9) reveals that ultimate impact and the impact at the peak number of users follow a simple scaling relationship

$$\frac{I_i^{\infty}}{I_i(t_i^{\star})} = \Phi(\eta_i) \tag{10}$$

where $\Phi(\eta) \equiv \frac{\Gamma(\eta)}{\gamma_{\eta}(\eta)}$. That is, I_i^{∞} scales linearly with peak impact $I_i(t_i^{\star})$, and their ratio is determined only by the initial power-law exponent η_i. To validate equation (10), we found that I_i^{∞} and $I_i(t_i^{\star})$ follow a clear linear relationship in our dataset for different values of η (Fig. 4p). In addition, Fig. 4p suggests that the relationship posts a slight shift as η increases. The rather subtle shift is also consistent with equation (10), as $\Phi(\eta)$ increases slowly with η (Fig. 4q). Therefore, the uncovered power-law growth patterns potentially offer a link between short-term and long-term impact in substitutive systems.

Discussion

In summary, here, we analysed a diverse set of large-scale data pertaining to substitutive processes, finding that early growth patterns in substitutive systems do not follow the exponential growth customary in spreading phenomena. Instead, they tend to follow power laws with non-integer exponents, indicating that they start with an initial explosive adoption process, followed by a much slower growth than expected in normal diffusion. Analysing patterns of 3.6 million individuals substituting for different mobile handsets, we uncovered three elements governing substitutions. Incorporating these elements allowed us to develop a minimal model for substitutions, which predicts analytically the power-law growth patterns observed in real systems, and collapses growth trajectories of constituents from rather diverse systems into universal curves.

Together, the results reported in this paper unpack the origin of robust self-organization principles emerging in complex substitutive systems, and demonstrate a high degree of convergence across the systems that we examined. Given the ubiquitous role that substitutions play in a wide range of important settings, our results may generalize beyond the instances that we studied. Potentially, these results could be relevant to our understanding and predictions of all spreading phenomena driven by substitutions, from electric cars to scientific paradigms, and from renewable energy to new healthy habits.

This work also opens up numerous directions for future investigations. For example, what is the role social network plays in substitutive dynamics? Unfortunately, regulations in the country from which the mobile phone dataset was collected prohibited us from obtaining any social network information. Nevertheless, the mobile phone setting may offer a distinctive opportunity to address this question, if mobile communication records could be collected in future studies to construct social connections among users[18,19,48–51]. Advances along this direction will further our understanding of substitutive dynamics and could also contribute meaningfully to the literature on social dynamics[23,28,29,39,52].

Furthermore, within each system, the obtained parameters for different constituents show interesting correlations (for example, we find negative correlations between the anticipation parameter h and fitness η, where the Pearson coefficient is −0.1642 for handsets, −0.5125 for automobiles, −0.13 for mobile applications and −0.416 for scientific fields, respectively). While such correlations do not affect the conclusion of the present paper, as our model estimates its parameters jointly and is compatible with any correlations that real systems might possess (Supplementary Note 4), the uncovered correlations suggest interesting directions for future studies. For example, one could better understand the different forces that may affect growth patterns by collecting auxiliary information on various constituents and inspecting their correlations with the model parameters. Such auxiliary information could also help us to better understand why diverse constituents differ from each other both within and across different systems.

On a theoretical level, it would also be interesting to explore further connections between our model with powerful theoretical tools offered by the epidemiology literature[3], such as recent findings on clustered epidemics[53,54] and multi-season models of outbreaks involving multiple pathogens with different levels of immunity[55].

It is important to note that, because our model is minimal, it ignores various contextual mechanisms, such as marketing campaigns, promotional activities or other platform-specific mechanisms, all of which could affect the studied phenomena. Although we analysed large-scale datasets from four different domains, to what degree our results can be extended beyond studied systems is a question that we cannot yet answer conclusively. However, the empirical and theoretical evidence presented in this paper provides a path towards the investigation of similar patterns in different domains, including re-examinations of familiar examples of spreading dynamics, as high-resolution data capturing early growth patterns become available. For example, there is growing evidence in the epidemiology community showing that the early spreading of certain diseases, such as Ebola and HIV infection, exhibits deviations from exponential growth, featuring sub-exponential growth patterns[31,32,56]. Although power-law early growth has not received as much attention, our results suggest that it may be more common than we realize, and that the power-law growth explained in our work may exist in even broader domains.

Methods

Details of the studied datasets are described in the main text and Supplementary Note 1. Empirical analyses of substitution patterns are detailed in Supplementary Note 2. Mathematical derivations of the minimal substitution model (equations (4)–(8)) are summarized in Supplementary Note 4. The handset-specific parameters are obtained through maximum-likelihood estimation, as described in Supplementary Note 4. The use of mobile phone datasets for research purposes was approved by the Northeastern University Institutional Review Board. Informed consent was not necessary because research was based on previously collected anonymous datasets.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data necessary to reproduce the results in the manuscript are available. The datasets for automobiles, smartphone apps and scientific fields are publicly available at https://chingjin.github.io/substitution/. The mobile phone dataset is not publicly available due to commercially sensitive information contained, but is available from the corresponding author on reasonable request.

Code availability

The custom code used is available at https://chingjin.github.io/substitution/.

Received: 4 March 2018; Accepted: 20 May 2019;
Published online: 8 July 2019

References

1. Barrat, A., Barthelemy, M. & Vespignani, A. *Dynamical Processes on Complex Networks* (Cambridge Univ. Press, 2008).
2. Ben-Avraham, D. & Havlin, S. *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge Univ. Press, 2000).
3. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
4. Rogers, E. M. *Diffusion of Innovations* (Simon and Schuster, 1962).
5. Gladwell, M. *The Tipping Point: How Little Things Can Make a Big Difference* (Little Brown, 2006).
6. Anderson, R. M., May, R. M. & Anderson, B. *Infectious Diseases of Humans: Dynamics and Control* Vol. 28 (Wiley, 1992).
7. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl Acad. Sci. USA* **103**, 2015–2020 (2006).
8. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
9. Bass, F. M. A new product growth for model consumer durables. *Manag. Sci.* **15**, 215–227 (1969).
10. Fisher, J. C. & Pry, R. H. A simple substitution model of technological change. *Technol. Forecast. Soc. Change* **3**, 75–88 (1972).
11. Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The diffusion of microfinance. *Science* **341**, 1236498 (2013).
12. Karsai, M., Iñiguez, G., Kaski, K. & Kertész, J. Complex contagion process in spreading of online innovation. *J. R. Soc. Interface* **11**, 20140694 (2014).
13. Aral, S., Muchnik, L. & Sundararajan, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl Acad. Sci. USA* **106**, 21544–21549 (2009).
14. Weiss, C. H. et al. Adoption of a high-impact innovation in a homogeneous population. *Phys. Rev. X* **4**, 041008 (2014).
15. Merton, R. K. *The Sociology of Science: Theoretical and Empirical Investigations* (Univ. Chicago Press, 1973).
16. Evans, J. & Foster, J. Metaknowledge. *Science* **331**, 721–725 (2011).
17. Granovetter, M. S. The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
18. Onnela, J.-P. et al. Structure and tie strengths in mobile communication networks. *Proc. Natl Acad. Sci. USA* **104**, 7332–7336 (2007).
19. Pentland, A. *Social Physics: How Social Networks Can Make Us Smarter* (Penguin, 2015).
20. Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379 (2007).
21. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
22. Morone, F. & Makse, H. A. Influence maximization in complex networks through optimal percolation. *Nature* **524**, 65–68 (2015).
23. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591 (2009).
24. Kuhn, T. *The Structure of Scientific Revolutions* (Univ. Chicago Press, 1996).
25. Jia, T., Wang, D. & Szymanski, B. K. Quantifying patterns of research-interest evolution. *Nat. Hum. Behav.* **1**, 0078 (2017).
26. Barabási, A.-L. *Network Science* (Cambridge Univ. Press, 2016).
27. Zang, C., Cui, P. & Faloutsos, C. Beyond sigmoids: the nettide model for social network growth, and its applications. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2015–2024 (ACM, 2016).
28. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
29. Kooti, F. et al. Portrait of an online shopper: understanding and predicting consumer behavior. In *Proc. Ninth ACM International Conference on Web Search and Data Mining* 205–214 (ACM, 2016).
30. Chowell, G., Viboud, C., Hyman, J. M. & Simonsen, L. The western Africa Ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates. *PLoS Curr.* https://doi.org/10.1371/currents.outbreaks.8b55f4bad99ac5c5db3663e916803261 (2015).
31. Chowell, G., Sattenspiel, L., Bansal, S. & Viboud, C. Mathematical models to characterize early epidemic growth: a review. *Phys. Life Rev.* **18**, 66–97 (2016).
32. Chowell, G., Viboud, C., Simonsen, L., Merler, S. & Vespignani, A. Perspectives on model forecasts of the 2014–2015 Ebola epidemic in West Africa: lessons and the way forward. *BMC Med.* **15**, 42 (2017).
33. Danon, L. & Brooks-Pollock, E. The need for data science in epidemic modelling: comment on: "Mathematical models to characterize early epidemic growth: a review" by Gerardo Chowell et al. *Phys. Life Rev.* **18**, 102–104 (2016).
34. Chowell, G., Sattenspiel, L., Bansal, S. & Viboud, C. Early sub-exponential epidemic growth: simple models, nonlinear incidence rates, and additional mechanisms: reply to comments on "Mathematical models to characterize early epidemic growth: a review." *Phys. Life Rev.* **18**, 114–117 (2016).
35. Wu, F. & Huberman, B. A. Novelty and collective attention. *Proc. Natl Acad. Sci. USA* **104**, 17599–17601 (2007).
36. Crane, R. & Sornette, D. Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Natl Acad. Sci. USA* **105**, 15649–15653 (2008).
37. Iribarren, J. L. & Moro, E. Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Lett.* **103**, 038702 (2009).
38. Gleeson, J. P., O'Sullivan, K. P., Baños, R. A. & Moreno, Y. Effects of network structure, competition and memory time on social spreading phenomena. *Phys. Rev. X* **6**, 021019 (2016).
39. Gleeson, J. P., Cellai, D., Onnela, J.-P., Porter, M. A. & Reed-Tsochas, F. A simple generative model of collective online behavior. *Proc. Natl Acad. Sci. USA* **111**, 10411–10415 (2014).
40. Shen, H. -W., Wang, D., Song, C. & Barabási, A. -L. Modeling and predicting popularity dynamics via reinforced Poisson processes. *In Proc. 28th AAAI Conference on Artificial Intelligence* **14**, 291–297 (2014).
41. Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013).
42. Serrano, M. Á., Boguná, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl Acad. Sci. USA* **106**, 6483–6488 (2009).
43. Lotka, A. J. Contribution to the theory of periodic reactions. *J. Phys. Chem.* **14**, 271–274 (1910).
44. Volterra, V. Variations and fluctuations of the number of individuals in animal species living together. *J. Cons. Int. Explor. Mer.* **3**, 3–51 (1928).
45. Salganik, M. J., Dodds, P. S. & Watts, D. J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856 (2006).
46. van de Rijt, A., Kang, S. M., Restivo, M. & Patil, A. Field experiments of success-breeds-success dynamics. *Proc. Natl Acad. Sci. USA* **111**, 6934–6939 (2014).
47. Watts, D. J. *Everything is Obvious:* Once You Know the Answer* (Crown Business, 2011).
48. Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proc. Natl Acad. Sci. USA* **106**, 15274–15278 (2009).
49. Valera, I. & Gomez-Rodriguez, M. Modeling adoption and usage of competing products. In *2015 IEEE International Conference on Data Mining (ICDM)* 409–418 (IEEE, 2015).
50. Dasgupta, K. et al. Social ties and their relevance to churn in mobile telecom networks. In *Proc. 11th International Conference on Extending Database Technology: Advances in Database Technology* 668–677 (ACM, 2008).
51. Sundsøy, P. R., Bjelland, J., Canright, G., Engø-Monsen, K. & Ling, R. Product adoption networks and their growth in a large mobile phone network. In *2010 International Conference on Advances in Social Networks Analysis and Mining* 208–216 (IEEE, 2010).
52. Deville, P. et al. Scaling identity connects human mobility and social interactions. *Proc. Natl Acad. Sci. USA* **113**, 7047–7052 (2016).
53. Hèbert-Dufresne, L. & Althouse, B. M. Complex dynamics of synergistic coinfections on realistically clustered networks. *Proc. Natl Acad. Sci. USA* **112**, 10551–10556 (2015).
54. Scarpino, S. V. et al. Epidemiological and viral genomic sequence analysis of the 2014 Ebola outbreak reveals clustered transmission. *Clin. Infect. Dis.* **60**, 1079–1082 (2014).
55. Scarpino, S. V., Allard, A. & Hèbert-Dufresne, L. The effect of a prudent adaptive behaviour on disease transmission. *Nat. Phys.* **12**, 1042–1046 (2016).
56. Viboud, C., Simonsen, L. & Chowell, G. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics* **15**, 27–37 (2016).

Acknowledgements

We thank B. Uzzi, J. Colyvas, J. Chu, M. Kouchaki, Q. Zhang, Z. Ma and all members of the Northwestern Institute on Complex Systems (NICO) for helpful comments. We are indebted to A.-L. Barabási for initial collaboration on this project and invaluable feedback on the manuscript. This work was supported by the Air Force Office of Scientific Research under award number FA9550-15-1-0162 and FA9550-17-1-0089, Northwestern University's Data Science Initiative, and National Science Foundation grant SBE 1829344. C.S. was supported by the National Science Foundation (IBSS-L-1620294) and by a Convergence Grant from the College of Arts & Sciences, University of Miami.

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

All authors designed the research. C.J., C.S. and D.W. conducted the analytical and numerical calculations. C.J., C.S., J.B. and D.W. analysed the empirical data. D.W. was the lead writer of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0638-y.

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

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

Peer review information: Primary Handling Editor: Stavroula Kousta.

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

nature research

Corresponding author(s): Dashun Wang

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	The mobile-phone dataset was obtained from the source. The Automobile dataset was collected from https://www.goodcarbadcar.net/ by using python 2.7. The smartphone app dataset was collected from https://apptopia.com by using python 2.7. The Scientific field dataset was obtained from Microsoft Academic Graph.
Data analysis	We use matlab for data analysis. The custom codes are available at https://github.com/chingjin/substitution.github.io.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data necessary to reproduce the results are available. The automobile, smart-phone apps and scientific fields datasets are publicly available at smart-phone apps

and scientific fields datasets are publicly available at https://github.com/chingjin/substitution.github.io. The mobile phone dataset is not publicly available due to commercially sensitive information contained, but are available from the corresponding author (dashun.wang@kellogg.northwestern.edu) on reasonable requests.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Behavioural & social sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is a quantitative study of diffusion and substitution patterns based on pre-existing datasets
Research sample	We assembled our datasets from a range of different sources: (1) Our first dataset captures, with daily resolution, 3.6 Million individuals choosing among different types of mobile handsets, recorded by a Northern European telecommunication company from January 2006 to November 2014. We focus on handsets that have been released for at least 6 months and used by at least 50 users in total (885 different handset models). (2) Our second dataset captures monthly transaction records of 126 automobiles sold in the North America between 2010 and 2016. These automobiles have been released for at least four months before the data was collected. (3) Our third dataset traces the number of daily downloads for new smart-phone Apps published in the App store (2,672 most popular Apps in the iOS systems from November to December 2016). (4) our fourth one is a scientific publication dataset, recording 246,630 scientists substituting for 6,399 scientific fields from 1980 to 2018.
Sampling strategy	No statistical methods were used to predetermine sample size.
Data collection	This study is based on pre-existing datasets.
Timing	The mobile phone dataset was collected in the end of 2014. The automobile and smart-phone all datasets were collected in the end of 2016. The scientific field dataset was collected in mid 2018.
Data exclusions	The analysis has no data exclusions. Selection criteria within a dataset are described in the supplementary information.
Non-participation	There are no participants in this study.
Randomization	This is a data driven study, not a randomized experiment.