Lawrence Berkeley National Laboratory

Title
The role of geography in the complex diffusion of innovations.

Permalink
https://escholarship.org/uc/item/9s84485v

Journal
Scientific Reports, 10(1)

Authors
Lengyel, Balázs
Bokányi, Eszter
Di Clemente, Riccardo
et al.

Publication Date
2020-09-15

DOI
10.1038/s41598-020-72137-w

Copyright Information

Peer reviewed

SCIENTIFIC REPORTS

natureresearch

Check for updates

OPEN

The role of geography in the complex diffusion of innovations

Balázs Lengyel[1,2,3,4✉], Eszter Bokányi[3,4], Riccardo Di Clemente [1,5,6], János Kertész [7] & Marta C. González [1,8,9]

The urban–rural divide is increasing in modern societies calling for geographical extensions of social influence modelling. Improved understanding of innovation diffusion across locations and through social connections can provide us with new insights into the spread of information, technological progress and economic development. In this work, we analyze the spatial adoption dynamics of iWiW, an Online Social Network (OSN) in Hungary and uncover empirical features about the spatial adoption in social networks. During its entire life cycle from 2002 to 2012, iWiW reached up to 300 million friendship ties of 3 million users. We find that the number of adopters as a function of town population follows a scaling law that reveals a strongly concentrated early adoption in large towns and a less concentrated late adoption. We also discover a strengthening distance decay of spread over the life-cycle indicating high fraction of distant diffusion in early stages but the dominance of local diffusion in late stages. The spreading process is modelled within the Bass diffusion framework that enables us to compare the differential equation version with an agent-based version of the model run on the empirical network. Although both model versions can capture the macro trend of adoption, they have limited capacity to describe the observed trends of urban scaling and distance decay. We find, however that incorporating adoption thresholds, defined by the fraction of social connections that adopt a technology before the individual adopts, improves the network model fit to the urban scaling of early adopters. Controlling for the threshold distribution enables us to eliminate the bias induced by local network structure on predicting local adoption peaks. Finally, we show that geographical features such as distance from the innovation origin and town size influence prediction of adoption peak at local scales in all model specifications.

Collective behavior, such as massive adoption of new technologies is a complex social contagion phenomenon[1]. Individuals are influenced both by media and by their social ties in their decision-making. This feature was first modelled in the 1960s with the Bass model of innovation diffusion[2]. The model distinguishes between exogenous and peers' influence and reproduces the observation that few early adopters are followed by a much larger number of early and late majority adopters, and finally, by few laggards[3]. The differential equations of the Bass model have been extensively used to describe the diffusion process and forecast market size of new products and the time of their adoption peaks[4].

Only in the past two decades, the importance of the social network structure has become increasingly clear in the mechanism of peers' influence[5]. In spreading phenomena, individuals perform a certain action only when a sufficiently large fraction of their network contacts have performed it before[6–10]. Complex contagion models, in which adoption depends on the ratio of the adopting neighbors, often referred to as adoption threshold[1,11],

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]International Business School Budapest, Budapest 1037, Hungary. [3]Agglomeration and Social Networks Lendület Research Group, Centre for Economic- and Regional Studies, Institute of Economics, Budapest 1097, Hungary. [4]Institute of Advanced Studies, Corvinus University of Budapest, Budapest 1093, Hungary. [5]Computer Science Department, University of Exeter, Exeter EX4 4QF, UK. [6]The Bartlett Centre for Advanced Spatial Analysis, University College London, London WC1E 6BT, UK. [7]Department of Network and Data Science, Central European University, Budapest 1051, Hungary. [8]Department of City and Regional Planning, University of California at Berkeley, Berkeley, CA 94720, USA. [9]Energy Analysis and Environmental Impacts Division, Lawrence Berkeley National Laboratory, Berkeley, Ca 94720, USA. ✉email: lengyel.balazs@krtk.mta.hu

have been efficiently applied to characterize the diffusion of online behavior[12] and online innovations[13,14]. In order to incorporate the role of social networks in technology adoption, the Bass model has been implemented through an agent-based model (ABM) version[15]. This approach is similar to other network diffusion approaches regarding the increasing pressure on the individual to adopt as network neighbors adopt; however, spontaneous adoption is also possible in the Bass ABM[16]. The structure of social networks in diffusion, such as community or neighborhood structure of egos, are still topics of interest[17,18]. Nevertheless, understanding how physical geography affects social contagion dynamics is still lacking[1].

Early work on spatial diffusion has highlighted that adoption rate grows fast in large towns and in physical proximity to initial locations of adoption[19,20]. It is argued that spatial diffusion resembles geolocated routing through social networks[20]. Social contagion—similar to geolocated routing[21]—occurs initially between two large settlements located at long distances and then becomes more locally concentrated reaching smaller towns and short distance paths. Facilitated by the observation of a large scale Online Social Network (OSN) over a decade, we capture for the first time these dynamics and provide insights into social network diffusion in its geographical space.

In this paper, we analyze the adoption dynamics of iWiW, a social media platform that used to be popular in Hungary, over its full life cycle (2002–2012). This unique dataset allows us to investigate two major geographical features that characterize spatial contagion dynamics: town size described by the urban scaling law[22] and distance decay described by the gravity law[23]. We find empirical evidence that early adoption is concentrated in large towns and scales super-linearly with town population but late adoption is less concentrated. Diffusion starts across distant big cities such that distance decay of spread is slight and becomes more local over time as adoption reaches small towns in later stages when distance decay becomes strong.

To better understand the spatial characteristics of complex contagion in social networks, we develop a Bass ABM of new technology's adoption on a sample of the empirical network preserving the community structure and geographical features of connections within and across towns. The data allows us to measure individual adoption thresholds that we can use to parameterize the likelihood of adoption at given fractions of infected social connections. We compare how the ABM and the Bass differential equation (DE) model fit to the empirical urban scaling and distance decay characteristics. Finally, we evaluate model accuracy in predicting the time of local adoption peaks and assess the bias induced by local network structures, or geographical features of towns. These analyses enable us to evaluate the role of geography in complex contagion models at local scales.

We find that the scaling of the number of earliest adopters with town population is best reflected by the ABM when threshold parameters are incorporated. None of our models can reproduce the high probability of diffusion across distant peers in the early stages of the life-cycle. Certain features of the network within towns—e.g. high network density and transitivity—accelerate the ABM diffusion and make predictions of adoption peaks early, which can be overcome when controlling for threshold distributions. Meanwhile, other features of the network—e.g. modularity and average path length—delay the prediction of adoption peaks, and cannot be eliminated with the threshold control. Nonetheless, we assess that contagion models cannot cure the bias of physical geography, such as distance from the innovation origin and town size, on the predictions of adoption peaks.

The threshold mechanisms introduced to the Bass ABM allow us to reproduce aggregated effects in relation to the number of adopters per population size. However, as expected, it is hard to predict the location of the social ties when an adoption occurs. This is in turn, affects the prediction of when the different towns reach their tipping point. Unfolding these aforementioned empirical features, we were able to capture the limitations of the standard model of complex contagion in predicting adoption at local scales and to describe key elements of diffusion in geographical space through the contact of local and distant peers.

Data

The social platform analyzed in this work is iWiW, which was a Hungarian online social network (OSN) established in early 2002. The number of users was limited in the first three years, but started to grow rapidly after a system upgrade in 2005 in which new functions were introduced (e.g. picture uploads, public lists of friends, etc.). iWiW was purchased by Hungarian Telecom in 2006 and became the most visited website in the country by the mid-2000s. Facebook entered the country in 2008, and outnumbered iWiW daily visits in 2010, which was followed by an accelerated churn. Finally, the servers of iWiW were closed down in 2014. All in all, more than 3 million users (around 30% of the country population) created a profile on iWiW over its life-cycle and reported more than 300 million friendship ties on the website. Until 2012, to open a profile, new users needed an invitation from registered members. Our dataset covers the period starting from the very first adopters (June 2002) until the late days of the social network (December 2012). Additionally, it contains home locations of the individuals, their social media ties, invitation ties, and their dates of registration and last login for each of the 3,056,717 users. The last two variables can be used to identify the date of adoption and disadoption (also referred to as churn) on individual level. Spatial diffusion and churn of iWiW have been visualized in Movie S1.

In previous studies, the data has demonstrated that the gravity law applies to the spatial structure of social ties[24], that adoption rates correlate positively both with town size and with physical proximity to the original location[25], that users central in the network churn the service after the users who are on the periphery of the network[26], and that the cascade of churn follows a threshold rule[27]. Socio-economic outcomes such as local corruption risk[28] and income inequalities[29] have been also investigated with the use of iWiW data.

Results

In the first step of the analysis, we empirically investigated the spatial diffusion over the OSN life-cycle. We categorized the users based on their adoption time for which we applied the rule proposed by Rogers[3] that divides adopters as follows: (1) innovators: first 2.5%, (2) early adopters: next 13.5%, (3) early majority: following 34%,

(4) late majority: next 34%, and (5) laggards: last 16%. Figure 1A illustrates the number of new users and the cumulative adoption rate (top plot), the spatial distribution of registered users (maps over white background) and the spatial patterns of accepted invitations to register (maps over black background). In the Innovator phase that lasted for 3 years (in red), adoption occurred in the metropolitan area of Budapest from where the innovation spouted over long distances, reaching the most populated towns first. In the Early Adopters phase (in green) and later in the Majority and Laggards phases (in blue), adoption became spatially distributed and more towns started to spread invitations.

The data allow us to demonstrate two major empirical characteristics of spatial diffusion proposed by previous literature[20]. First, by regressing the number of adopters with town population (both on logarithmic scale)[22] we find in Fig. 1B that the number of Innovators and Early Adopters ($\beta_{Innovators} = 1.41$, CI [1.23; 1.59], $\beta_{EarlyAdopters} = 1.28$, CI [1.18; 1.37], and $\beta_{Majority\&Laggards} = 1.07$, CI [1.04; 1.10]) are strongly and significantly concentrated in large towns. Second, in Fig. 1C we illustrate the gravity law[23] by stages of the life-cycle by depicting the probability of invitations sent to a new user at distance d formulated by ($P_d^t = L_d^t / N_a^t \times N_b^t$), where L_d^t refers to the number of invitations sent at d over stage t while N_a^t and N_b^t denote the number of users who registered in stage t in towns a and b separated by d. The strengthening distance decay of invitation links demonstrates that diffusion first bridges distant locations but becomes more and more local over the life-cycle.

Adoption in the bass diffusion framework. The Bass diffusion model[2] enables us to investigate adoption dynamics at global and local scales. This can be done by fitting the cumulative distribution function (CDF) of adoption (shown in Fig. 1A) with model CDF. The Bass CDF is defined by $dy(t)/dt = (p + q \times y(t))(1 - y(t))$, with $y(t)$ the number of new adopters at time t (months), p innovation or advertisement parameter of adoption (independent from the number of previous adopters), and q imitation parameter (dependent on the number of previous adopters). This nonlinear differential equation can be solved by:

$$y(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}, \tag{1}$$

with m size of adopting population. Equation 1 described the CDF empirical values with residual standard error $RSE = 0.0001398$ on $df = 125$ and empirical values $q = 0.108$, CI [0.097; 0.12]; $p = 0.00016$, CI [10^{-4}; 2×10^{-4}]. We repeated these estimations of the diffusion parameters for every geographic settlement i (called towns henceforth) and consequently estimated p_i and q_i.

The time of adoption peak[30], defined by the maximum amount of adoption per month, is an important feature of adoption dynamics. To evaluate the Bass model accuracy on local scales, we investigate Prediction Error, the peak month predicted by the model minus the empirical peak month (smoothed by a 3-month moving average that helps to eliminate noise). Prediction Error is illustrated in Fig. 2A. Towns' differences in terms of the time of adoption peak indicate a wide distribution of local deviations from the global diffusion dynamics (Fig. 2B), which can be used in statistical analysis. The Bass model estimation of the adoption peak t_i^* for every town i is:

$$t_i^* = \frac{\ln p_i + \ln q_i}{p_i + q_i} \tag{2}$$

and is positively correlated with the empirical peaks in Fig. 2D ($\rho = 0.742$, CI [0.725; 0.759]).

In case we keep one of p_i and q_i parameters fixed, adoption becomes faster as the other increases (Fig. 2C). Furthermore, towns diverge from Eq. (2) for peak times in months 50–60 (Fig. 2D), corresponding to low p_i and large q_i (Fig. 2C). This suggests that the innovation term in the Bass model is lower and the process is driven by imitation in towns where diffusion happens at the primitive stage. On average, peaks in towns predicted by Eq. (2) are 1.76 months later, with a 95% confidence interval [1.54; 1.98], than empirical peaks (Fig. 2E). Prediction is late in large towns but is early in towns distant from Budapest that are also smaller than average (correlation between population and distance is $\rho = -0.32$) CI [-0.35; -0.28] (Fig. 2F). Population correlates with both Eq. (2) parameters (with p_i, $\rho = 0.11$ CI [0.07; 0.14] and with q_i, $\rho = -0.34$ CI [-0.38; -0.30]). The correlation between Bass parameters, peak prediction and town characteristics are reported in Supporting Information 2.

Although parameters are estimated for every town separately, physical geography still influences model prediction. An important limitation of modelling local adoption with Bass DE is that towns are handled as isolates. To disentangle the role of geography in diffusion, we need models that can consider connections between locations.

A complex diffusion model. We further investigated the spreading of adoption on a social network embedded in geographical space connecting towns and also individuals within these towns via the ABM version of the Bass model. We used the social network observed in the data by keeping the network topology fixed at the last timestamp without removing the churners, using this as a proxy for the underlying social network. This approximation is a common procedure to model diffusion in online social networks when the underlying social network cannot be detected[13]. The ABM is tested on a 10% random sample of the original data (300K users) by keeping spatial distribution and the network structure stratified by towns and network communities. The latter were detected from the global network using the Louvain method[31]. We show in Supporting Information 4 that samples of different sizes have almost identical network characteristics and these are very similar to the full network as well.

In the ABM, each agent j has a set of neighbors n_j taken from the network structure (Fig. 3A) and is characterized by a status $F_j(t)$ that can be susceptible for adoption S or infected I (already adopted). Once an agent reaches
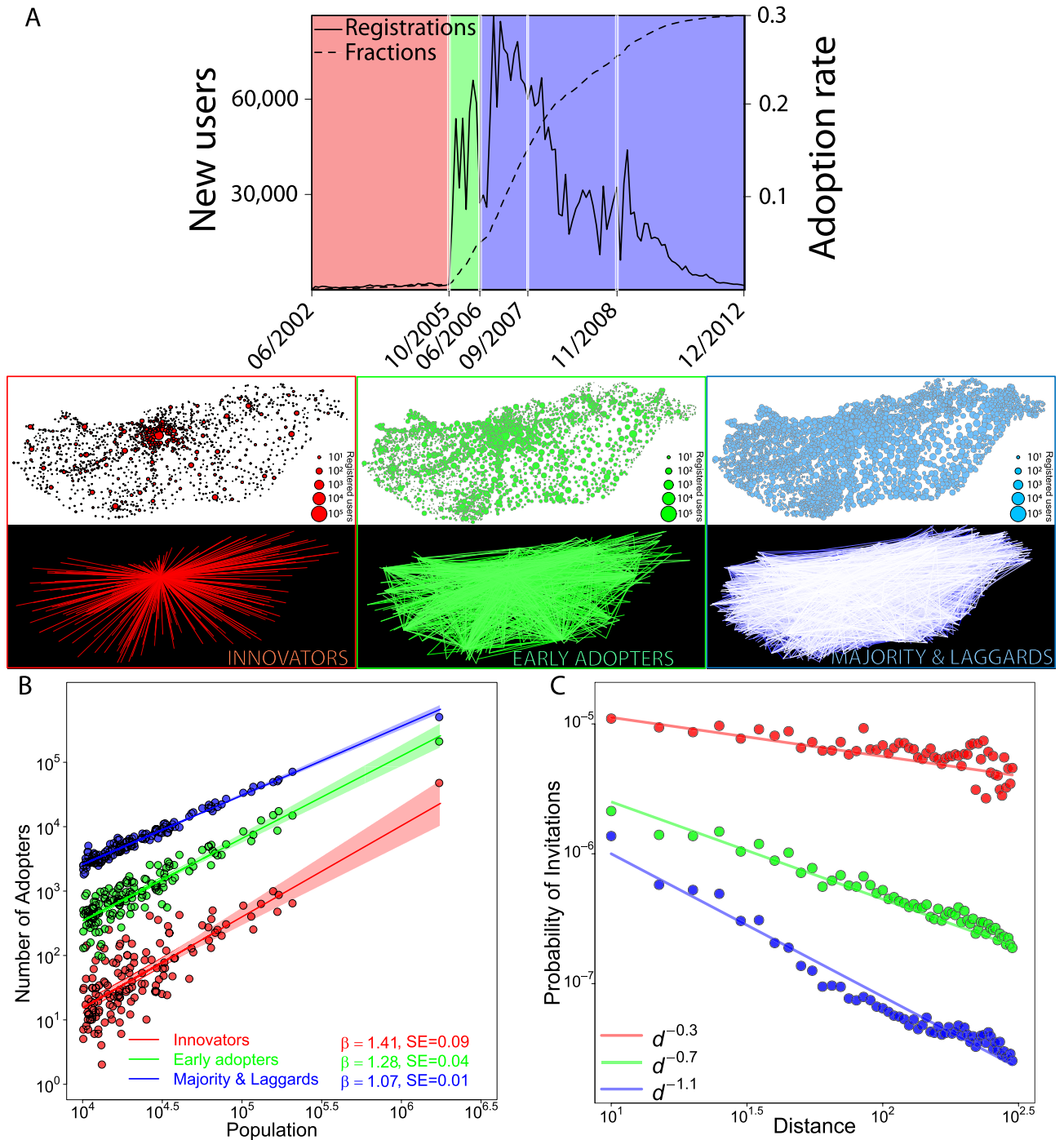
Figure 1. Spatial diffusion over the OSN life-cycle. (**A**) Top: Number of new users and the cumulative fraction of registered individuals among total population over the OSN life-cycle. Users are categorized by the time of their registration into Rogers's adopter types: (1) innovators: first 2.5%, (2) early adopters: next 13.5%, (3) early and late majority and laggards. White background maps: coloured dots depict towns; their size represent the number of adopters over the corresponding period. Black background maps: links depict the number of invitations sent between towns over the corresponding periods. (**B**) Adoption scales super-linearly with town population. The β coefficient of urban scaling denotes very strong concentration of adoption in the Innovator stage and decreases gradually in later stages. Fitted lines explain the variation in Number of Adopters (log) by $R^2 = 0.63$ (red), $R^2 = 0.83$ (green), $R^2 = 0.97$ (blue). (**C**) The Probability of Invitations to distant locations is relatively high in the Inventors stage but decreased over the product life-cycle while diffusion became more local. Exponent fits explain the variation of Probability of Invitations (log) with $R^2 = 0.24$ (red line), $R^2 = 0.85$ (green line), $R^2 = 0.92$ (blue line).
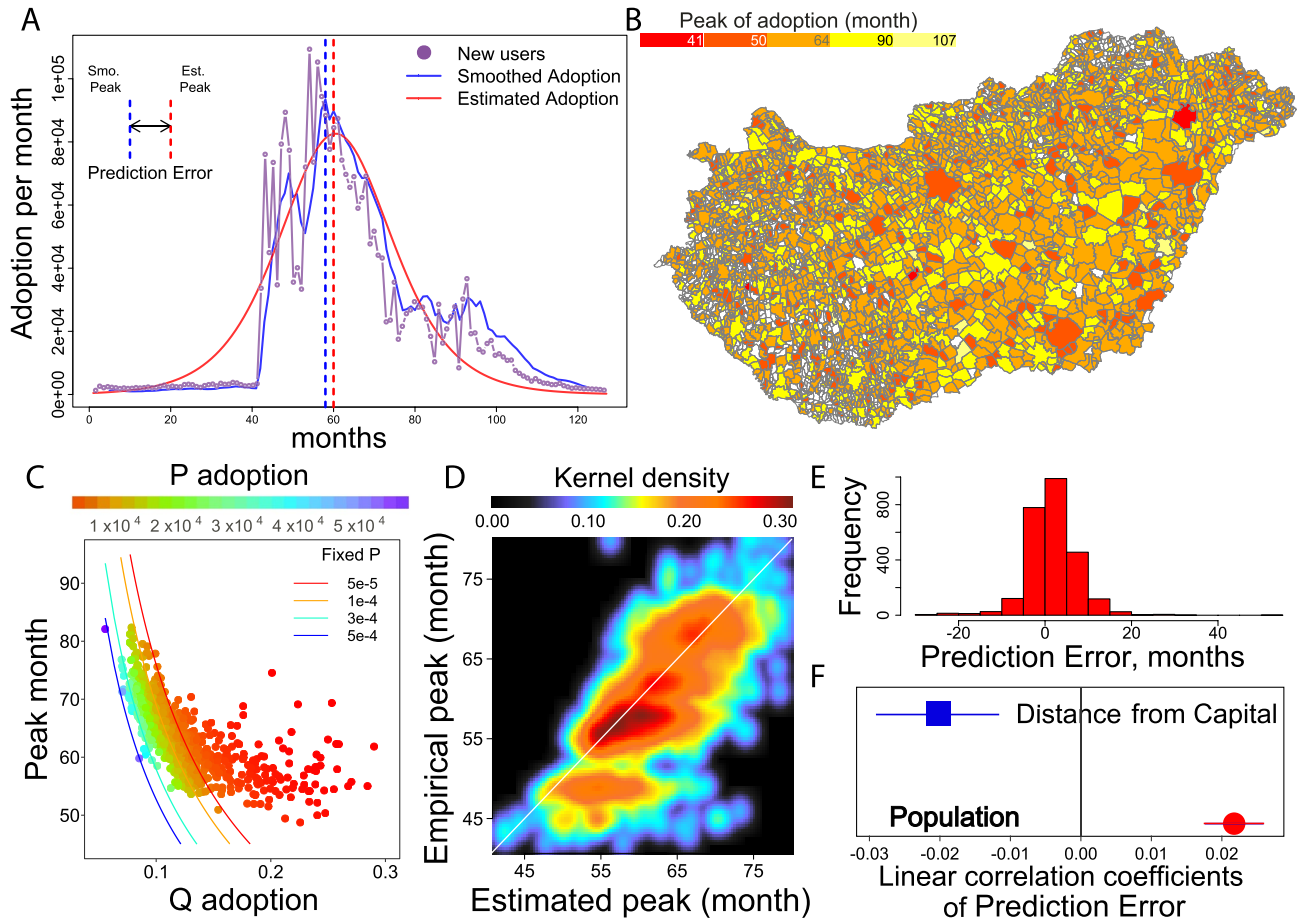
Figure 2. Adoption peak prediction on local scales with the Bass model. (**A**) The Bass DE model estimates on the monthly adoption trend and a smoothed empirical adoption trend (3-month moving average) are compared. We investigate the difference between estimated peak month and the peak of the smoothed trend. (**B**) Times of adoption peaks vary across towns. (**C**) Estimated p_i and q_i result in same adoption peak with fixed p_i, except in early adoption cases when q_i is high. (**D**) Estimated peaks of adoption correlate with empirical peaks of adoption ($p = 0.74$). (**E**) Prediction Error in town i is the predicted month of adoption peak by Equation 2 minus the empirical month of smoothed adoption peak. (**F**) Dots are point estimates of linear univariate regressions and bars depict standard errors. Dependent variable is scaled with its maximum value and independent variables are log-transformed with base 10.

the status I, it cannot switch back to S. To reflect reality, the users that adopted in the first month in the real data were set as infected I in $t = 1$. The process of adoption $F_j(t) = S \rightarrow F_j(t + 1) = I$ is defined as:

$$F_j(t + 1) = \begin{cases} I & \text{if } U(0,1)_{jt} < \hat{p}^{\text{ABM}} + T(\mathcal{N}_j(t), h, l) \times \mathcal{N}_j(t) \times \hat{q}^{\text{ABM}} \\ S & \text{otherwise} \end{cases} \tag{3}$$

where $U(0,1)_{jt}$ is a random number picked from a uniform distribution for every agent j in each t. $\hat{p}^{\text{ABM}}$ denotes adoption probability exogenous to the network and $\hat{q}^{\text{ABM}}$ is adoption probability endogenous to the network. In order to focus on the role of network structure in diffusion, $\hat{p}^{\text{ABM}}$ and $\hat{q}^{\text{ABM}}$ are kept homogenous for all j in the network. Consequently, the process is driven by the neighborhood effect $\mathcal{N}_j(t)$ defined as:

$$\mathcal{N}_j(t) = \frac{\# n_j^I(t)}{\# n_j^I(t) + \# n_j^S(t)} \tag{4}$$

where $\# n_j^I(t)$ is the number of infected neighbors and $\# n_j^S(t)$ is the number of susceptible neighbors at t.

The distribution of $\mathcal{N}_j(t)$ at the time of adoption carries information about adoption dynamics in the social network[13]. Figure 3B suggests that the probability of adoption in our case is the highest when $\mathcal{N}_j(t)$ is around 0.5 (in case $10 \leq k \leq \infty$) and decreases when $\mathcal{N}_j(t)$ is close to 0 or 1. To reflect on this empirical finding in the ABM, we introduce the transformation function $T(\mathcal{N}_j(t), h, l)$ on $\mathcal{N}_j(t)$ defined by

$$T(\mathcal{N}_j(t), h, l) = -4(h + l) \cdot \mathcal{N}_j(t)^2 + 4(h + l) \cdot \mathcal{N}_j(t) + 1 - l. \tag{5}$$
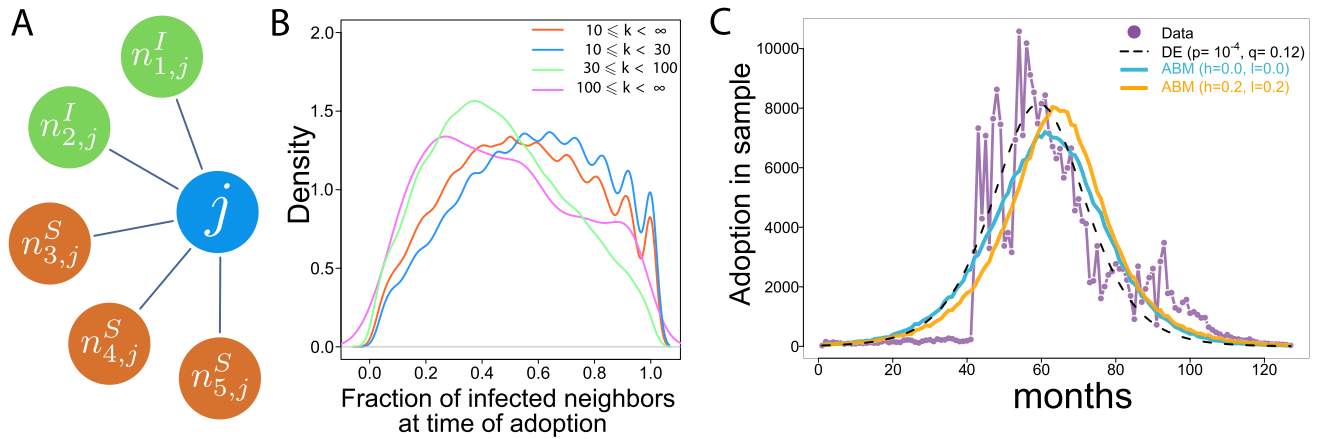
Figure 3. Model of complex contagion. (**A**) Network topology and peers influence in the Bass ABM. A sample individual j has two infected neighbors n_j^I who have already adopted the innovation and three susceptible neighbors who have not adopted yet n_j^S. (**B**) The distribution of adoption thresholds. Fraction of infected neighbors at time of adoption illustrate that most individuals adopt when half of their neighbors have already adopted. This fraction is smaller for high degree ($k > 30$) individuals. (**C**) ABM adoption curves assuming linear ($h = 0.0, l = 0.0$ in blue) and non-linear ($h = 0.2, l = 0.2$ in orange) functions of infected neighbor ratio predict slower adoption than Bass DE.

where h controls the relative importance of $\mathcal{N}_j(t) = 0.5$ and l controls the decrease of the adoption probability at $\mathcal{N}_j(t) = 0$ and $\mathcal{N}_j(t) = 1$. Both of parameters h and l are considered in order to find optimum model descriptions of spatial adoption.

This definition of the process implies that users are assumed to be identically influenced by advertisements and other external factors and are equally sensitive to the influence from their social ties that are captured by the fraction of infected neighbors $\mathcal{N}_j(t)$. The decision regarding adoption of innovation or postponing this action is an individual choice that is assumed to be random. This model belongs to the complex contagion class[1,12] because adoption over time is controlled by the fraction of infected neighbors[9,13]. As the fraction of infected neighbors increases, the agent becomes more likely to adopt the innovation. Supporting Information 4 describes the calibration of $\hat{p}^{ABM}$ and $\hat{q}^{ABM}$, and explain how h and l parameters were selected.

We set Bass parameters in the ABM to their calibrated values $\hat{p}^{ABM} = 0.0002$ and $\hat{q}^{ABM} = 0.12$ that are close to the estimated values using Eq. (1) on the ABM sample (reported in Fig. 3B) as suggested by[32]. Two ABMs are considered. ABM ($h = 0.0, l = 0.0$) assumes that adoption probability increases linearly with $\mathcal{N}_j(t)$. ABM ($h = 0.2, l = 0.2$) assumes a non-linear influence of $\mathcal{N}_j(t)$ on adoption probability. Supporting Information 4 illustrates $T(\mathcal{N}_j(t), h, l)$ with parameters h = 0.2 and l = 0.2, and it's relation with the empirical threshold distribution and explains how parameters ($h = 0.2, l = 0.2$) change adoption probability in the ABM compared to the case when $h = 0.0$ and $l = 0.0$.

In Fig. 3C, we report global adoption trends after running both ABM 10 times and calculating average values of these realizations over time-steps $t = (1, 120)$ that reflects the months taken from the real data. Both ABM ($h = 0.0, l = 0.0$) (solid blue line) and ABM ($h = 0.2, l = 0.2$) (solid orange line) are faster in the early phase (before month 40) than in reality, which is due to the extraordinary tipping point around month 40 that is difficult to fit. ABM ($h = 0.2, l = 0.2$) is closer to reality in this early phase while ABM ($h = 0.0, l = 0.0$) follows the DE trend until month 40. Comparing to ABM ($h = 0.0, l = 0.0$), ABM ($h = 0.2, l = 0.2$) is faster from month 40, has an adoption volume at its peak comparable to the DE estimate, and decline faster after it's peak. The peak predicted by DE is at month 59, by ABM ($h = 0.0, l = 0.0$) is at month 61, and by ABM ($h = 0.2, l = 0.2$) is at month 63; whereas the empirical peak smoothed with 3 months moving average is at month 58. Adoption in ABM ($h = 0.0, l = 0.0$) fit to adoption in DE with $\chi^2 = 15,621, p = 4^{-4}$; while ABM ($h = 0.2, l = 0.2$) fit to adoption in DE with $\chi^2 = 15,748, p = 4^{-4}$. These initial comparisons suggest that ABM ($h = 0.2, l = 0.2$) can capture early adoption dynamics better than ABM ($h = 0.0, l = 0.0$), while the peak of adoption might be better reproduced by ABM ($h = 0.0, l = 0.0$).

Local adoption in the ABM. To better understand the differences between DE and ABM versions, we move now from the global trend to local scales and compare DE that is informed by location-specific p_i and q_i but cannot incorporate networks with ABM that can control networks but has homogenous $\hat{p}^{ABM}$ and $\hat{q}^{ABM}$. The introduction of $T(\mathcal{N}_j(t), h, l)$ enables us to investigate how controlling for the threshold distribution improves ABM predictions at local scales compared to data and the DE estimations.

A major challenge in spatial diffusion modeling is the unknown spatial distribution of Innovators and Early Adopters that need to be predicted by the model; however, as a paradox, this spatial distribution is a prerequisite of accurate prediction of local adoption peaks in social networks[30]. To overcome this limitation, we empirically analyze how the ABM captures spatial distribution of adoption in three phases of product life-cycle. In Fig. 4A,C we compare how the number of adopters observed in the data and predicted by the model scale with the town population[22,33] by using the β coefficient of the linear regression in towns with more than 10^4 inhabitants. Because
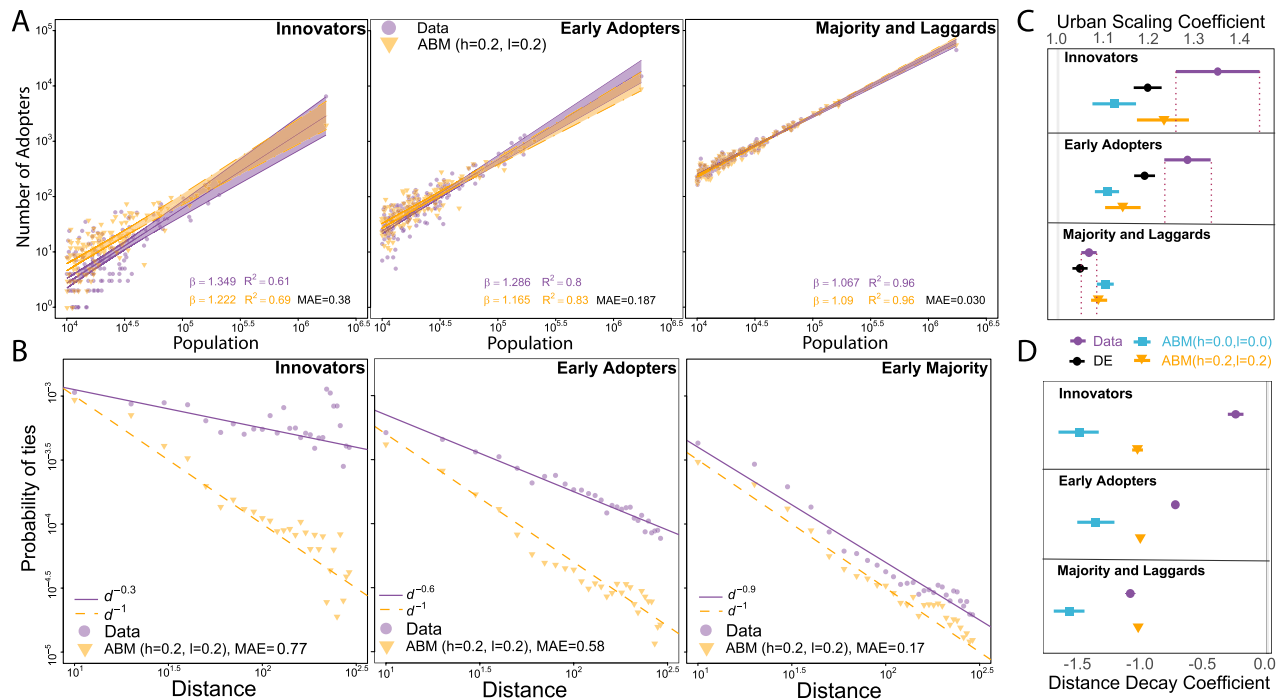
Figure 4. Urban scaling and distance decay in the ABM. (**A**) Urban scaling of adoption in the ABM (h = 0.2, l = 0.2) and in the empirical data across the product life-cycle. Solid lines denote linear regression estimation and shaded areas are 95% confidence intervals. The ABM (h = 0.2, l = 0.2) significantly over-predicts the number of Innovators in small towns. Urban scaling β in the Early Adopters phase is still smaller in the ABM (h = 0.2, l = 0.2) than in the empirical data. (**B**) The distance decay of social ties of Innovators and Early Adopters is larger in the ABM (h = 0.2, l = 0.2) than in reality and only becomes similar in the Early Majority phase. (**C**) Empirical urban scaling coefficients are declining over the life-cycle that is best captured by the ABM (h = 0.2, l = 0.2) prediction. Markers denote point estimates and horizontal lines denote standard errors. (**D**) Empirical distance decay coefficients are declining over the life-cycle that are not captured by the models. Markers denote point estimates and horizontal lines denote standard errors.

both ABM (h = 0.0, l = 0.0) and ABM (h = 0.2, l = 0.2) are faster than real adoption in the first 40 months but are slower than DE and following Rogers[3] we define Innovators and Early Adopters as the first 2.5% and the next 13.5% of adopters. This enables us to compare spatial distribution of Innovators and Early Adopters between the ABMs, Bass DE and reality regardless of temporal differences in the global trend.

An empirical superlinear scaling measured in the sampled Data in the Innovator and Early Adopter phases indicates strong urban concentration of diffusion during the early phases of adoption, already reported in Fig. 1 on the full network. Supporting Information 5 demonstrates that the urban scaling estimation is robust against introducing various indicators of town development or demographics. To compare Bass ABM and Bass DE approaches, we re-estimate Eq. (1) for every town in the sample and estimate monthly adoption that can enter the scaling regression. Figure 4C reveals that ABM (h = 0.2, l = 0.2) follows the changes in empirical urban scaling somewhat better both in terms of β and fit to empirical adoption than ABM (h = 0.0, l = 0.0) that has an urban scaling β of adoption around 1.1 in all phases of the life-cycle. The scaling coefficient of ABM (h = 0.2, l = 0.2) is within the margin of error in the Innovator and Majority and Laggards phases; in the former this is due to the large standard error of empirical scaling coefficient. ABM (h = 0.2, l = 0.2) partly outperforms the DE estimation that only captures scaling of Early Adopters better. However, we find in Fig. 4A that in the Innovator phase of the life-cycle, the ABM predicts more adoption in small towns and less in large towns compared to reality and predicts smaller adoption volumes in large towns in the Early Adopters stage. What happens is that the ABM interchanges individuals' early adoption in large towns with early adoption in small towns such that much more small town users get into the first 2.5% than in reality. This is a bit less striking when adoption probability is increased at most frequent individual thresholds in ABM (h = 0.2, l = 0.2), which probably slows ABM adoption down in small towns. Confidence intervals of urban scaling coefficients plotted in Fig. 4C can be found in Supporting Information 6.

Turning to the role of distance in diffusion over the life-cycle, Fig. 4B compares the distance of influential peers, measured as the probability that Innovators, Early Adopters, and Early Majority[3] have social connections at distance d[23,24,34–36] in the ABM (h = 0.2, l = 0.2) versus in the empirical data. Ties of Innovators have a very week distance decay, which intensifies for Early Adopters and even more for Early Majority. The intensifying role of distance measured here resembles distance decay measurement by invitation data (see Fig. 1) and confirms that innovation spreads with high propensity to distant locations during the early phases of the life-cycle[20]. However, neither ABM (h = 0.2, l = 0.2) nor ABM (h = 0.0, l = 0.0) are able to handle the changing role of distance. Instead, distance decay in both ABMs are rather stable across these three phases of the life-cycle (Fig. 4C). Unfortunately,
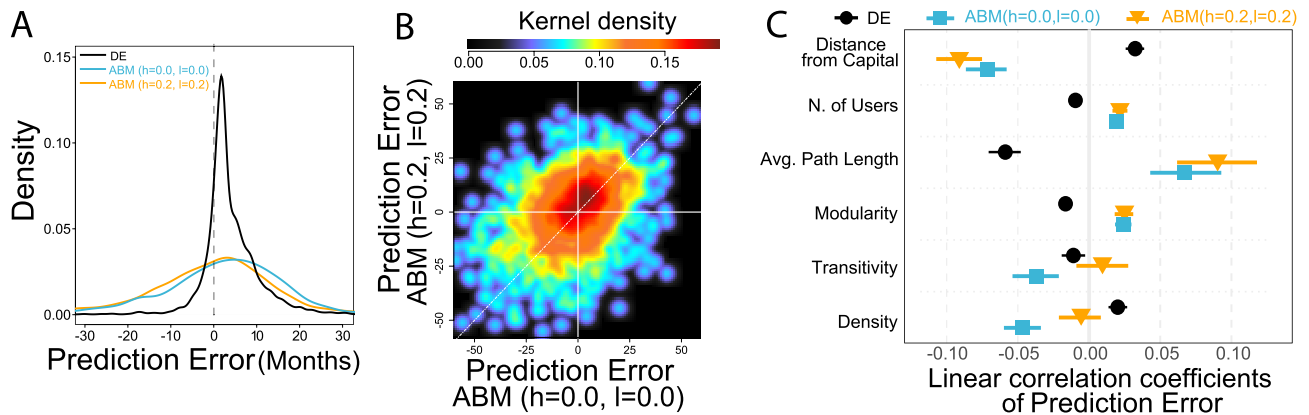
Figure 5. ABM in predicting adoption peaks in towns. (**A**) Prediction Error in town i is the peak month predicted by the ABM minus the month of empirical peak. Negative Prediction Error denotes early prediction and positive means late prediction. (**B**) Correlation of Prediction Error of ABM (h = 0.0, l = 0.0) and ABM (h = 0.2, l = 0.2) ($\rho = 0.39$, CI [0.36; 0.43]). (**C**) Estimation of town-level Prediction Error of ABM versions via a simple linear regression. Independent variables are characteristics from networks within towns and geographical characteristics of towns. Symbols represent point estimates and horizontal lines denote standard errors.

we are not able to compare these patterns to DE estimations, since the distance decay of social connections cannot be inferred on with the DE method due to the lack of individual predictions. Our findings imply that ABM replaces distant contagion with proximate contagion in the early phases of the life-cycle. Innovators are mostly found in distant large towns. Even though they are connected to each other, these connections might be bridges across communities that slows complex contagion in the ABM[1].

Adoption peaks typically happen in the Early- and Late Majority phases of the life-cycle, for which ABM (h = 0.2, l = 0.2) adoption predicts the aggregated number of adopters in towns well (Fig.4A). To understand how accurate the peak time predictions are, we analyze determinants of ABM Prediction Error as already done in Fig.2E for the Bass model on the full network. In case of ABM (h = 0.0, l = 0.0), the predicted month of adoption peak matches the observed month of adoption peak in the data with 95% confidence interval [−1.69; −0.46]; indicating that the ABM (h = 0.0, l = 0.0) predicts adoption peaks early in most towns (Fig. 5A). However, peaks predicted by ABM (h = 0.2, l = 0.2) are 1.74 months late on average with 95% confidence interval [1.16; 2.32]. Peaks predicted by the Bass DE are 3.89 months late on average with 95% confidence interval [3.62; 4.16]. Prediction error values of the ABMs are correlated (Fig. 5B). However, there are towns, where prediction is early in ABM (h = 0.0, l = 0.0) and is late in ABM (h = 0.2, l = 0.2) and vice versa.

In order to analyze the role of network structure in local adoption dynamics in the Majority phase, we correlated the town-level Prediction Errors with several town-level network properties (Fig. 5C). Density, the fraction of observed connections among all possible connections in the town's social network; and Transitivity, the fraction of observed triangles among all possible triangles in the town's social network, are claimed to facilitate diffusion[1]. On the other hand, complex contagion is more difficult in networks with modular structure, when social links between network communities are sparse, and in networks with long paths, when the distance of nodes within the town's social network is large. In fact, the ABM (h = 0.0, l = 0.0) predicted the peak of adoption early in the towns where Density and Transitivity are relatively high (Fig. 5C). Influencing the probability of adoption according to the adoption threshold distribution in ABM (h = 0.2, l = 0.2), however, cures this bias as the co-efficients of Density and Transitivity become non-significant. ABM modification does not cure the delaying influence of Modularity and Average Path Length. These latter co-efficients of ABM (h = 0.0, l = 0.0) and ABM (h = 0.2, l = 0.2) are within estimation error. We also find that Assortativity, the index of similarity of peers in terms of adoption time[37] delays adoption of large towns, which we discuss in detail in Supporting Information 7. DE Prediction Error estimations are illustrated for the reasons of comparison. We find that local network estimations on DE Prediction Error are not corresponding with ABM estimations and are even counterintuitive from a network diffusion perspective. These are in line with expectations because DE prediction is not allowed to use information on the local network structure. This finding support our claim that network-based models are needed to better understand diffusion on networks. Confidence intervals of coefficients plotted in Fig. 5C can be found in Supporting Information 8.

Finally, we observe that geographical characteristics, Population (measured here by number of users in the ABM sample) and Distance (measured by Euclidean distance from Budapest) influence the accuracy of ABM peak prediction. Like we found in the case of the Bass DE model on the full network in Fig. 2F, prediction is late in large towns but is early in towns distant from Budapest, that are significantly smaller in terms of population than average (see multiple regression results in Supporting Information 9). Point estimates of ABM (h = 0.0, l = 0.0) and ABM (h = 0.2, l = 0.2) are not significantly different from each other but are significantly different from DE estimates on the sample. These latter estimations are reported only for the sake of comparison. The DE coefficients seem to be biased by the sampling process, and thus the difference between coefficients in Figs. 2F and 5C, and are not robust against regressing them together in a multiple regression framework (see Supporting Information 9). The ABM coefficients confirm that geography has a role in the complex diffusion of innovations.

We suggest social contagion models to incorporating town size and geographical distance between peers in order to improve accuracy of local adoption prediction.

Discussion

Taken together, we studied spatial diffusion over the life-cycle of an online product on a country-wide scale. By combining complex diffusion with empirical threshold distribution, we proposed a stochastic modeling framework that allows for spontaneous adoption in the network and is able to explore how geography influences model accuracy in capturing local adoption trends. The model does not perfectly predict how adoption rates scale with a city's population, especially in the early stages of the life cycle. This is to some extent due to the fact that the standard model assumes a linear relation between adoption probability and the share of neighbors already active on the OSN. In reality, the relation between individual adoption probability and adoption rates of neighbors is nonlinear: we observe that adoption rates accelerate for intermediate, but decelerate at very high adoption levels by neighbors. Once the ABM takes this into consideration, it's fit to the observed urban scaling of adoption in the early life cycle periods improves. This step eliminates the influence of dense and transitive local networks as well that would otherwise accelerate adoption peaks in towns too early.

One of our most important empirical findings is the changing distance decay of diffusion. In fact, contagion in the early stages of the product life-cycle occurs mostly between distant locations with larger populations. This new aspect could not be captured by the model, indicating that it needs theoretical extension. The superlinear relation of Innovators and Early Adopters as a function of the town population highlights the importance of urban settlements in the adoption of innovations that corresponds with the early notion of Haegerstrand[20]. Adoption peaks initially in large towns and then diffuses to smaller settlements in geographical proximity. We find that town population and distance from the original location of innovation bias predictions of adoption peak in all models. These findings call for incorporating geography into future models of complex contagion.

Unlike many of the previous work on social networking cites that investigate a large selection of OSNs[38] or a dominant OSN entering many countries[39], our results are limited to a specific product in a single country. In this regard, future research shall investigate how various types of online products diffuse across space and social networks and in different countries. For example, complex products, which has been reported to scale super-linearly with city size[40,41] might diffuse across locations differently than non-complex products due to the difficulties to adopt complex technologies and knowledge. Technologies compete with each other, which is completely missing from our understanding on spatial diffusion in social networks. Some of the technologies dominate over long periods but when quitting becomes collective, their life-cycle ends[38,42,43]. Recent studies have shown that both adopting and quitting the technology follow similar diffusion mechanisms[27,44]. However, the geography of how churning is induced by social networks is still unknown.

Future work on spatial diffusion of innovation in social networks has to tackle the difficulty of modeling individual adoption behavior embedded in geographical space. One of the challenges is that individuals are heterogenous regarding adoption thresholds that is non-trivially related to the formation and spatial structure of social networks. Individuals who are neighbors in the social network are likely to be located in physical proximity as well, but this is not always the case[23]. Further, network neighbors typically are alike in terms of adoption thresholds[45]. Thus, it is not clear whether social influence has a geographical dimension or we can think of it using a space-less network approach. We propose that investigating and incorporating the distance decay in social influence modeling might help us understanding spatial diffusion of innovation better.

Methods

Nonlinear least-square regression with the Gauss–Newton algorithm was applied to estimate the parameters in Eq. (1). In order to identify the bounds of parameters search, this method needs starting points to be determined, which were $p_i = 0.007$ and $q_i = 0.09$ for Eq. (1).

Identical estimations were applied in a loop of towns, in which the Levenberg–Marquardt algorithm[46] was used with maximum 500 iterations. This estimation method was applied because the parameter values differ across towns, and therefore town-level solutions may be very far from the starting values set for the country-scale estimation. Initial values were set to $p_i = 7 \times 10^{-5}$ and $q_i = 0.1$ in Eq. (1).

To characterize urban scaling of adoption in Figs. 1 and 4, we applied the ordinary least squares method to estimate the formula $y(t) = \alpha + \beta x$, where $y(t)$ denotes the logarithm (base 10) of accumulated number of adopters over time period t, and x is the logarithm (base 10) of the population in the town. R-squared values have been applied to the variance of the log-transformed dependent variable.

Data availability

Data tenure was controlled by a non-disclosure agreement between the data owner and the research group. The access for the same can be requested by email to the corresponding author.

Code availability

ABM simulation and parameter calibration codes have been written in Python and have been reposited at https://github.com/bokae/spatial_diffusion. All other codes to produce the results have been written in R. These latter codes are available upon request at the corresponding author.

References

1. Centola, D. & Macy, M. Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702–734 (2007).
2. Bass, F. M. A new product growth for model consumer durables. *Manag. Sci.* **15**, 215–227 (1969).
3. Rogers, E. M. *Diffusion of Innovations* (Simon and Schuster, New York, 2010).
4. Mahajan, V., Muller, E. & Bass, F. M. New product diffusion models in marketing: A review and directions for research. *J. Market.* **54**(1), 1–26 (1990).
5. Centola, D. *How Behavior Spreads: The Science of Complex Contagions* Vol. 3 (Princeton University Press, Princeton, 2018).
6. Schelling, T. C. *Micromotives and Macrobehavior* (WW Norton, New York, 1978).
7. Granovetter, M. Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978).
8. Valente, T. W. Social network thresholds in the diffusion of innovations. *Social Netw.* **18**, 69–89 (1996).
9. Watts, D. J. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci.* **99**, 5766–5771 (2002).
10. Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The diffusion of microfinance. *Science* **341**, 1236498 (2013).
11. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Modern Phys.* **87**, 925 (2015).
12. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
13. Karsai, M., Iñiguez, G., Kikas, R., Kaski, K. & Kertész, J. Local cascades induced global contagion: How heterogeneous thresholds, exogenous effects, and unconcerned behaviour govern online adoption spreading. *Sci. Rep.* **6**, 27178 (2016).
14. Katona, Z., Zubcsek, P. P. & Sarvary, M. Network effects and personal influences: The diffusion of an online social network. *J. Market. Res.* **48**, 425–443 (2011).
15. Rand, W. & Rust, R. T. Agent-based modeling in marketing: Guidelines for rigor. *Int. J. Res. Market.* **28**, 181–193 (2011).
16. Watts, D. J. & Dodds, P. S. Influentials, networks, and public opinion formation. *J. Consumer Res.* **34**, 441–458 (2007).
17. Ugander, J., Backstrom, L., Marlow, C. & Kleinberg, J. Structural diversity in social contagion. *Proc. Natl. Acad. Sci.* **109**, 5962–5966 (2012).
18. Aral, S. & Nicolaides, C. Exercise contagion in a global social network. *Nat. Commun.* **8**, 14753 (2017).
19. Griliches, Z. Hybrid corn: An exploration in the economics of technological change. *Econometrica*, **25**(4), 501–522 (1957).
20. Hagerstrand, T. Innovation diffusion as a spatial process. *University of Chicago Press, 1967*.
21. Leskovec, J. & Horvitz, E. Geospatial structure of a planetary-scale social network. *IEEE Trans. Comput. Social Syst.* **1**, 156–163 (2014).
22. Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci.* **104**, 7301–7306 (2007).
23. Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P. & Tomkins, A. Geographic routing in social networks. *Proc. Natl. Acad. Sci. USA* **102**, 11623–11628 (2005).
24. Lengyel, B., Varga, A., Ságvári, B., Jakobi, Á. & Kertész, J. Geographies of an online social network. *PloS One* **10**, e0137248 (2015).
25. Lengyel, B. & Jakobi, Á. Online social networks, location, and the dual effect of distance from the centre. *Tijdschrift voor economische en sociale geografie* **107**, 298–315 (2016).
26. Lőrincz, L., Koltai, J., Győr, A. F. & Takács, K. Collapse of an online social network: Burning social capital to create it? *Social Netw.* **57**, 43–53 (2019).
27. Török, J. & Kertész, J. Cascading collapse of online social networks. *Sci. Rep.* **7**, 16743 (2017).
28. Wachs, J., Yasseri, T., Lengyel, B. & Kertész, J. Social capital predicts corruption risk in towns. *R. Soc. Open Sci.* **6**, 182103 (2019).
29. Tóth, G. *et al.* Inequality is rising where social network segregation interacts with urban topology. arXiv preprint arXiv:1909.11414 (2019).
30. Toole, J. L., Cha, M. & González, M. C. Modeling the adoption of innovations in the presence of geographic and media influences. *PloS One* **7**, e29528 (2012).
31. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
32. Xiao, Y., Han, J. T., Li, Z. & Wang, Z. *A fast method for agent-based model fitting of aggregate-level diffusion data* Tech. Rep, SSRN, (2017).
33. Deville, P. *et al.* Scaling identity connects human mobility and social interactions. *Proc. Natl. Acad. Sci.* **113**, 7047–7052 (2016).
34. Scellato, S., Mascolo, C. & Musolesi, M. & Latora, V. *Geo-social metrics for online social networks*. In WOSN, Distance matters, (2010).
35. Onnela, J.-P., Arbesman, S., González, M. C., Barabási, A.-L. & Christakis, N. A. Geographic constraints on social network groups. *PLoS One* **6**, e16939 (2011).
36. Wang, P., González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding the spreading patterns of mobile phone viruses. *Science* **324**, 1071–1076 (2009).
37. Newman, M. E. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).
38. Ribeiro, B. Modeling and predicting the growth and death of membership-based websites. in *Proceedings of the 23rd International Conference on World Wide Web*. 653–664 (ACM, 2014).
39. Kassa, Y. M., Cuevas, R. & Cuevas, Á. A large-scale analysis of facebookâ™s user-base and user engagement growth. *IEEE Access.* **6**, 78881–78891 (2018).
40. Gomez-Lievano, A., Patterson-Lomba, O. & Hausmann, R. Explaining the prevalence, scaling and variance of urban phenomena. *Nat. Hum. Behav.* **1**, 0012 (2016).
41. Balland, P.-A. *et al.* Complex economic activities concentrate in large cities. *Nat. Hum. Behav.* **4**, 248–254 (2020).
42. Kairam, S. R., Wang, D. J. & Leskovec, J. The life and death of online groups: Predicting group growth and longevity. in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, 673–682 (ACM, 2012).
43. Kloumann, I., Adamic, L., Kleinberg, J. & Wu, S. The lifecycles of apps in a social ecosystem. in *Proceedings of the 24th International Conference on World Wide Web*, 581–591 (International World Wide Web Conferences Steering Committee, 2015).
44. Garcia, D., Mavrodiev, P., Casati, D. & Schweitzer, F. Understanding popularity, reputation, and social influence in the twitter society. *Policy Internet* **9**, 343–364 (2017).
45. Aral, S., Muchnik, L. & Sundararajan, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci.* **106**, 21544–21549 (2009).
46. Moré, J. J. The Levenberg–Marquardt algorithm: Implementation and theory. In: Watson G.A. (eds) *Numerical Analysis*. Lecture Notes inMathematics, vol. 630. (Springer, Berlin, Heidelberg, 1978).

Acknowledgements

Author contributions

B.L. and M.G. designed the research, B.L., E.B. and R.D.C. conceived the experiments, B.L., E.B., R.D.C., J.K and M.G. analyzed the results. All authors wrote and reviewed 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/s41598-020-72137-w.

Correspondence and requests for materials should be addressed to B.L.

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

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