Human mobility in interest space and interactive random walk

Fanqi Zeng, Li Gong, Jing Liu, Jiang Zhang, Qinghua Chen, and Ruyue Xin

1 Beijing Normal University, School of Systems Science, Beijing 100875, People’s Republic of China
2 University of Bristol, Department of Engineering Mathematics, Bristol BS8 1UB, United Kingdom
3 Author to whom any correspondence should be addressed.

E-mail: zhangjiang@bnu.edu.cn and qinghuachen@bnu.edu.cn

Keywords: human mobility, Lévy Flight, collective attention, interest space, attention flow networks, interactive random walk

Abstract

Compared with the well-studied topic of human mobility in real geographic space, only a few studies focus on human mobility in virtual space, such as interests, knowledge, ideas, and so on. However, it relates to the issues like public opinion management, knowledge diffusion, and innovation. In this paper, we assume that the interests of a group of online users can span an Euclidean space which is called interest space, and the transfers of user interests can be modelled as Lévy Flight in the interest space. Considering the interaction between users, we assume that the random walkers are not independent but interacting with each other indirectly via the digital resources in the interest space. The proposed model in this paper successfully reproduced a set of scaling laws for describing the growth of attention flow networks of online communities, and obtaining similar ranges of users’ scaling exponents with empirical data. Further, we inferred parameters for describing the individual behaviours of the users according to the scaling laws of empirical attention flow network. Our model can not only provide theoretical understanding of human online behaviours but also has broad potential applications such as dissemination and public opinion management, online recommendation, etc.

1. Introduction

Everything is moving. Understanding the mobility patterns of humankind is of great importance since it relates to epidemics, online gaming strategies, mental and artificial searches, and other issues in modern city. Lots of studies on human mobility in real space have been conducted in past decades. For instance, it is found that Lévy Flight, one of the most famous random walk model, which is significantly distinguished from Brownian motion, can be used to characterize human movements. However, human mobility does not only take place in real space exclusively but also in virtual space. For example, our mind always jumps between different ideas, which can be understood as a virtual movement in interest space. A large number of users that surfing on an online community, and jumping between different posts, can also be understood as collective movements in interest space. Surprising connections have also been made between online gaming strategies and search behaviours in real world. Javarone introduces an analytical model to study the dynamics of evolutionary games with agents moving in a bi-dimensional Euclidean space. Notably, they find some kinds of random walk lead to cooperation, while others to defection. Although virtual space seems to be less stable than physical space, the study of it is of considerable significance because it helps us to understand the issues of psychological therapy, dissemination, and management of public opinion, online recommendation, etc.
a large amount of attention is paid on how people interact, correlate, and connect each other in the studies of social networks [30, 31]. There is an increasing realization that universal statistical laws could characterize human’s daily activities [8–10, 32, 33]. These statistical laws proved that human behaviours are powerfully cast by individual’s propensity in the complex interactive space.

Although human dynamics and complex networks have drawn wide attention, little concern is made with the sequential movements of users in interest space because the concept of the interest space is not apparent. Recently, some attempts have been made to visualize the virtual space by attention flow network model [34, 35], on which nodes represent digital resources (posts, tags, and articles, etc). Notably, the attention flow network is built according to the data of users [18, 36], i.e., the representation of the interest space rather than the space itself.

In this paper, we proposed a concise physical model to reproduce the scaling laws of human’s collective activities in virtual world. Wu [18] has observed some scaling laws in the online forums, but did not propose any suitable model to simulate the real-world activities. Here, we assume that all the possible interests of users span an Euclidean space, in which adjacent points stand for similar interests, and users perform random walks of Lévy Flight. Lévy Flight is a random walk, which has a heavy-tailed probability distribution for the step-lengths. However, naïve Lévy Flight model [12] can not reproduce the required scaling laws because of the absence of interactions. Thus, we propose an interactive Lévy Flight model, in which, interactions occur between users indirectly bridged by digital resources. For the computation and visualization, we first use an infinite lattice plane to simulate the interaction process. The model successfully reproduces all the concerned scaling laws, and the range of scaling exponents can also be calibrated by adjusting the parameters in the model. Our model can not only deepen the understanding of virtual mobility, but also could improve the accuracy of predictions on user behaviours [16, 37], leading to wide applications on recommendation [27, 38], searching [39], user profiling [40], etc.

2. Model

To understand the scaling phenomena of production and diversity, and the relationship between users and digital resources, we construct an interactive Lévy Flight model to simulate user behaviours. Let us consider an online community (such as Baidu Tieba, stack exchange or Flickr, etc), which containing a large number of registered users. All their interests can span an interest space which is modelled by a two-dimensional Euclidean plane (see figure 1(a)). In which, each cell represents one possible interest, such as a kind of music style, or a type of article, etc. Two adjacent cells standing for similar contents, and similar types of interests would be distributed in concentrated areas at last. Meanwhile, articles, tags, Q & As, and other digital resources generated by users can be projected onto this interest space. We use \( C(X, t) \) to denote the number of units projected at \( X \) and time \( t \), where \( X \) is the coordinate of the cell. The cell occupied by at least one unit of digital resource, i.e., \( C(X, t) > 0 \), is called an active site meaning that the resource has the corresponding theme as the interest.

In figure 1(a), \( S_0, S_1, \ldots, S_4 \) represent active sites and occupied by digital resources. When a user generates a digital resource, it will be projected on the space and be able to be visited or read by other users. Thus, all users interact with each other indirectly through the active sites.

User’s sequential behaviours such as browsing, posting, Q & A, and others in a session can be viewed as a walk in the interest space. We assume that the user’s walk satisfies the Lévy Flight law in \( d \) dimensional Euclidean space, meaning that the movement pattern of a user is random and the probability density distribution function of the movement distance \( l \) in one jump satisfies power law:

\[
P(l) \sim l^{-\lambda}. \tag{1}
\]

This movement rule describes how a user’s interest transfers: user’s interest will stay in a narrow area for a long time, but occasionally implement a long-range jump with a small probability. \( \lambda \) (values are in [1, 3]) is the exponent of the Lévy Flight, and it characterizes how broad the interests of users are. In equation (1), the range of \( \lambda \) means there are a broad choices of different random walks. Specially, when \( \lambda = 2 \) this Lévy Flight is actually Brownian motion that is a well-known classical random walk model. If \( \lambda \) is small, then the user may perform long-range random jump frequently, representing they have broad interests, and vice versa. Thus, \( \lambda \) is a parameter of a trade-off between familiarity and novelty. Users usually consume some familiar topics, but they may also occasionally require some new information to visitors. In figure 1(a), the arcs labelled with the same numbers represent the flights of one user. For example, the user with label 1 travels \( S_0, S_2, \) and \( S_4 \) sequentially. In the lattice space, we choose an integer approximation of an individual’s step length to simplify the modelling.

Next, we consider the interaction between users. We know that if a community already has abundant digital resources (such as many posts in a forum), users will continue to visit them. Otherwise, they will lose their
Figure 1. The interactive Lévy Flight model in a two-dimensional space. (a) The active sites and the trajectories. (b) Category of trajectories generated by the random walkers in (a). (c) Constructed attention flow network according to (b). In (a), $S_0, S_1, \ldots, S_4$ are active sites with digital resources; Circled numbers represent users who implementing Lévy Flights. Site $D$ represents the places where the users jumping to but no digital resources existing on, and the users quitting consequently.

interests and quit quickly. In order to characterize this feature, without loss of generality, we assume that the user can jump continually and randomly from any cell $X$ as long as $X$ is active (there is at least one unit of digital resource). Otherwise, he(she) will go out of the space from $X$. We denote the position of user $i$ at time $t$ is $X_t$, and it is $\emptyset$ when the user quit the community, so we have:

$$X_t^i = \begin{cases} X_{t-1}^i + \xi & \text{if } C(X_{t-1}^i, t-1) > 0 \\ \emptyset & \text{if } C(X_{t-1}^i, t-1) = 0, \end{cases}$$  

(2)

where, $\xi$ is a random number following equation (1). On the other hand, each user will generate new digital resource with a certain probability in the process of a random walk. That is, if the user $i$ jumps to the cell $X$, he(she) will add a new digital resource at $X$ with a certain probability $p$. Thus, we have:

$$C(X_t^i, t) = C(X_{t-1}^i, t-1) + \eta,$$  

(3)

where, $\eta$ is a random number following 04 distribution with a probability $p$ to be 1. Therefore, the resource is generated by the users. This emulates the process of posting comments and tags.

Next, we consider the situation with $N$ users. Suppose in one simulation (a session), at the beginning, $N$ users are set on the origin of the interest space, and they begin to implement Lévy Flights from the origin simultaneously. Although they do not interact directly, they can interfere with each other via the active sites. The simulation ends at time $T(N)$ when all users exit the space. Apparently, $T(N)$ will increase with $N$ implying that the indirect interactions can keep users living in the community for a longer time since the probability that they encounter each other increases. One trajectory can be generated for one user in his (her) lifespan in the community. We also set up other models (see the appendix) with different conditions such as different dimensions or step-length distribution, but the individual’s behaviour rules did not change.

3. Results

3.1. Attention flow network

The so-called attention flow network is an open flow network [41, 42] in which nodes represent digital resources (active sites) and weighted directed links represent transition (such as making a comment in a post and then making a comment in another post) flows between two nodes formed by the collective behaviours of the random walkers. Figure 2 shows a brief construction process of attention flow network.
Figure 2. The construction of the attention flow network. In the left column, we list 3 users’ clickstream data which records the sites (URLs) and the visiting time ticks as examples. Accordingly, we can construct an open flow network model to depict the collective behaviours of these users. In the network, nodes are sites (URLs), and links are jumps between sites. The weight on each edge between $i$ and $j$ is the number of users who jump to $j$ after his (her) visiting $i$. Notice that there are two special nodes: the source and the sink which represent the space (offline world). If the time gap between any two records is longer than 30 min, we assume that this user jumps offline, which leading to a flow from the last visited site to the sink, and a flow from the source to the first site after 30 min (from Lou et al [42]).

To observe the collective behaviours of these Lévy Flighters, we construct an attention flow network, as shown in figure 1(b). In the model, the weight of the edge connecting active sites $X$ and $Y$ can be defined as:

$$W(X, Y) = \sum_{i=1}^{T-1} \sum_{i=1}^{N} \delta(X_i^t - X) \cdot \delta(X_{i+1}^t - Y),$$

(4)

where, $\delta$ is the Dirac delta function, it equals to 1 only if its component is 0. Therefore, the weight linking $X$ and $Y$ equals to the number of jumps between them. Two particular nodes which are the source and the sink are added in to represent the space. When a random walker starts to jump to an active site $X$ in the interest space, a unit of flux from the source to the node $X$ is added to the attention flow network. On the other hand, a unit of flux from node $X$ to the sink is added if the last site of a random walker’s visit is $X$. The attention flow network can characterize the collective properties of a large number of users for both the simulated model and the empirical data.

To validate our model, we study how the network properties will change as the size of the system increases, and to see if our model can reproduce the same scaling laws. We use the total number of users $N$ as the measure of the size. In fact, this quantity is also the total influx of the attention network for a given simulation. Following that, we will focus on how the macroscopic properties change with $N$.

3.2. Macroscopic behaviour

Here, we focus on three basic macroscopic variables. First,

$$A = \sum_{X} \sum_{Y} W(X, Y)$$

(5)

measures the activity of the community, which is defined as the total number of transitions of interests (jumps). Second,

$$D = \sum_{X} \sum_{i=1}^{T} (1 - \delta(C(X, t)))$$

(6)

is the total number of active sites, or the total number of nodes in the network. It measures the diversity of interests for all members in the community. Third,

$$E = \sum_{X} \sum_{Y} (1 - \delta(W(X, Y)))$$

(7)

is the total number of edges of the network, and it measures the diversification of interests transitions. According to our simulated data, all three variables scale to $N$ with different exponents, i.e., $A \sim N^\alpha$, $D \sim N^\beta$, $E \sim N^\gamma$. 
Figure 3. The scaling of simulated attention flow networks constructed by interactive Lévy Flight model. Three quantities $A, D, E$ scale with $N$ with different exponents when $\lambda = 2$ and $p = 0.4$. Inset shows the densification law, i.e., the scaling of $E$ (the number of edges) with $D$ (the number of nodes) of the attention flow network. All the experiments are conducted on 2D dimensional interest space, and more experiments with other parameters are shown in the appendix.

Where $\alpha$, $\beta$, and $\gamma$ are exponents characterizing the relative growth speed of the quantities to the size of the system as shown in figure 3.

3.3. Scaling

The empirical anonymous data sets present the individuals’ daily traces, which contain anonymous user number, visiting time and visiting site. To compare with the simulated data, we plot the empirical scaling laws for the same quantities on three representative online communities, Baidu Tieba (each jump represents a click behaviour, see figures 4(a)–(c)) and stack exchange (each jump represents an answering behaviour, see figures 4(d)–(f)).

We observe that all the communities follow the same scaling laws as the simulated results, and the values of exponents show similar trends. We also notice that the exponents $\alpha$ are always larger than 1.0 for different $p$ values in simulations. This observation also holds for empirical data. As shown in figures 5(a) and (b), we systematically calculate the exponents of 1000 Baidu Tiebas and 136 communities in stack exchange and plot the distribution of exponents.

It is clear that the distribution of the four exponents are nearly normal distribution, in which $\alpha$ is right skew and the average value of $\theta$ is approximately $1.25$ which is larger than 1.0 significantly. Some small Tiebas’ exponents are less than 1.0 since their scaling properties are not statistically significant.

We further confirm the super-linear relationship between $A$ and $N$ for more online communities, as shown in table 1. All the exponents are larger than 1.0. Among which, the communities with intensive interactions between users always have larger exponents like Baidu Tieba, stack exchange, and Digg.
Figure 4. The scaling of empirical attention flow networks constructed by the behaviours of users in representative online communities. We plot $A(a, d), D(b, e),$ and $E(c, f)$ versus $N$ for both Baidu Tieba (a, b, c) and stack exchange (d, e, f). Three communities are selected from Tieba and stack exchange, respectively. For each community, we construct an attention flow network in each hour according to the approach mentioned in figure 3 from the empirical trajectories of all users recorded in the data. After that, the aggregated quantities $A,$ $D,$ and $E$ are calculated based on the attention flow network in each hour. For each community, we construct an attention flow network in each hour according to the approach mentioned in figure 3 from the empirical trajectories of all users recorded in the data. After that, the aggregated quantities $A,$ $D,$ and $E$ are calculated based on the attention flow network in each hour.

Actually, we can regard the exponent $\alpha$ as an indicator to measure the intensity of the social interactions between users for one community. According to equation (1), we derive

$$\frac{A}{N} \sim N^{\alpha - 1}.$$  \hspace{1cm} (8)

That indicates the average number of jumps increases with the size of the system $N$ if $\alpha > 1.$ And the relative speed increases with $\alpha.$ Therefore, if $\alpha$ is large, the average activities generated by the users will be sensitive to the total number of users. This characterizes the non-linearity of the interactions of users.

To compare, we investigate the possible intervals of $\alpha$ for our simulations. As shown in figure 6, when the exponent $\alpha$ increases with the probability $p$, it means that when the propensity that user generating activities increases, the average intensity of interaction also increases.

If we understand the activity as a kind of production of users, then the exponent $\alpha$ characterizes the productivity of the bunch of users. If it is easy for users to express their interests ($p$ increases), the online community is more productive.
Next, we analyse the exponent of $\beta$, and the scaling behaviour between the number of nodes of the attention flow network and the size of the system. This scaling law indicates how the diversification of the digital resources generated by the users' changes with the size. We find that both for empirical data (see figure 4) and simulation (see figure 5), the exponents are significantly less than one which indicates a sub-linear scaling between diversity and the size of the system. This sub-linear scaling is often observed in other complex systems.

The total number of edges $E$ on the attention flow network measures the diversification of distinct transitions between pairs of nodes. However, there is a large deviation for the exponent $\gamma$. Super-linear and
Figure 6. The dependency of all exponents on $p$ and how the relations change with $\lambda$. There is an increasing tendency for exponents $\gamma, \alpha, \beta$ when $p$ and $\lambda$ increase, which means the diversity of network’s edges grows synchronously. Especially, this tendency becomes sharply when $p$ is over 0.8. The tendency for $\theta$ is tiny different from the other three exponents. When $p > 0.8$, the increment of $\theta$ is slow or even negative, which is probably because of the network becomes too dense that the power law could not stand in this situation [43].

The dependency of all exponents on $p$ and how the relations change with $\lambda$. There is an increasing tendency for exponents $\gamma, \alpha, \beta$ when $p$ and $\lambda$ increase, which means the diversity of network’s edges grows synchronously. Especially, this tendency becomes sharply when $p$ is over 0.8. The tendency for $\theta$ is tiny different from the other three exponents. When $p > 0.8$, the increment of $\theta$ is slow or even negative, which is probably because of the network becomes too dense that the power law could not stand in this situation [43].

sub-linear are both possible for different communities. There is a transition from sub-linear to super-linear in simulations.

When $p$ increases, both exponents ($\beta$ and $\gamma$) for diversification increase. That means the propensity that user generating contents can accelerate the relative speeds that diverse contents are produced compared to the size of the system. Thus, the average distinct contents generated increase with the size of the system. It is interesting to observe another scaling behaviour between $E$ and $D$ as

$$E \sim D^\theta$$

a super-linear can be observed. This phenomenon is observed for a large number of networks which is named as densification phenomenon. Our model can successfully reproduce this phenomenon and the exponent $\theta$ fluctuates around 1.5. All the ranges of exponents for models are consistent with the ones in empirical data, which implies that our model can capture the scaling behaviours in data.

We further test how the exponent of Lévy Flight influences the other exponents. The results are shown in figure 6. The qualitative characteristics of the dependence between the exponents and $p$ do not change dramatically; however, the range of exponents change is different.

We also note that the range of the fluctuation of $\gamma$ is relatively small for different $p$, but it changes dramatically with different exponent $\lambda$. Thus, we guess that the exponent of $\gamma$ exclusively depends on $\lambda$, and we suppose that this dependence can be used to infer the value of $\lambda$ for a real community.

In the appendix, we proposed another four simulations for one-dimensional and three-dimensional Euclidean space, and wider range of $\lambda$ over 3.0, and change the step-length distribution into the exponential distribution. The results show up that the properties of scaling laws remain striking under different simulated conditions, although the exponents’ values differ.

3.4. Inferences for parameters $\lambda$ and $p$

Next, we will infer the parameters $\lambda$ and $p$ from empirical exponents by using the maximum likelihood principle for each community. These two parameters describe the micro behaviours of users in a community, wherein $p$ describes how active the users are and $\lambda$ describes how broad the users’ interests are. For individual community, these two parameters can be inferred from the community’s empirical data.

We suppose the real exponents $\alpha, \beta, \theta$ are randomly sampled from the model. And the exponents follow normal distributions with centres determined by the model and standard deviation $\sigma$ for given $p$ and $\lambda$, that is:

$$P(\alpha, \beta, \theta|p, \lambda) \propto \exp \left( \frac{-[(\alpha - \alpha(p, \lambda))^2 + (\beta - \beta(p, \lambda))^2 + (\theta - \theta(p, \lambda))^2]}{\sigma^2} \right),$$  

where, $\alpha(p, \lambda)$, $\beta(p, \lambda)$, and $\theta(p, \lambda)$ represent the exponents generated by the model for given parameter $p$ and $\lambda$ which can be read from the dependency of figure 6. To infer $p$ and $\lambda$ from given empirical measure of
Figure 7. Inferred individual parameters \( p \) and \( \lambda \) for all Baidu Tieba and stack exchange communities. All the possible values of \( \lambda \) and \( p \) are separated into 20 intervals, therefore, there are 200 combinations of \( p \) and \( \lambda \) in total. We find the closest combination for each community. The colour of the circle stands for the number of communities, and the size stands for the difference \( D \) between empirical data and inferred data.

\[ \alpha_i, \beta_i, \text{and } \theta_i, \] we attempt to maximize the likelihood probability (equation (11)), that is:

\[ p, \lambda = \arg \max_{p, \lambda} P(\alpha, \beta, \theta | p, \lambda). \] (11)

So, we need to minimize the distance:

\[ D = \sqrt{(\alpha - \alpha(p, \lambda))^2 + (\beta - \beta(p, \lambda))^2 + (\theta - \theta(p, \lambda))^2}. \] (12)

Then we should find the most probable parameters \( p \) and \( \lambda \) so that the simulated exponents are closest to the empirical ones.

Figure 7 shows all the inferred parameters for Baidu Tieba and communities of stack exchange. We notice that all the Tieba’s can be roughly separated into two groups according to their parameters, and they have similar \( p \) value (0.1) but different \( \lambda \) values. We know that \( \lambda \) indicates how dissimilar of the users’ interests for one transition. Thus, the users in Tieba with small \( \lambda \) always have relatively broad interests. All the Tiebas’ have very small \( p \) values means that the tendency for posting a new thread is much less than clicking. While the communities in stack exchange almost concentrate in the area of \( 1.0 < \lambda < 2.0 \) or \( 0 < p < 0.4 \). That means the users in stack exchange have broad interest and do not like to post questions in general. However, compared to Tieba, stack exchange communities always have larger \( p \) values meaning that it is easier for asking a question compared to answering it than for posting a thread compared to clicking threads.
4. Discussion

In this paper, we proposed an interactive Lévy Flight model to simulate the random walk behaviours of users in virtual interest space. We ran this model in different situations such as different dimensions, step-length distributions and ranges of $\lambda$ values, and found that the properties of power laws remain striking. Two important parameters controlled the Lévy Flight’s behaviour, i.e., the wideness that users’ interests distributes and the propensity that a user delivers a post determine the structures of the attention flow network. We compared the statistical properties of the attention flow network with empirical online communities from the perspective of scaling laws. Four different scaling laws characterize how the macroscopic quantities of activity, diversification of resources generated by users, and the diversification of interests transfer scale to the size of users. And the exponents characterize the relative growth speed. All the scaling behaviours and the range of exponents in simulation are in accordance with the empirical data. We then can infer the two critical parameters $p$ and $\lambda$ from the exponents.

Therefore, the interest transition of users may be characterized by a simple random walk model on a two-dimensional space spanned by the interests of users. The indirect interaction between users is the key to explain the origins of scaling laws which observed in empirical communities. In our model, we assume that the users may stay in the system only if they can find the published digital resources which feed their interests. This is essential to the indirect interaction and the super-linear scaling law of activity because when the number of users increases, the interactions between users also increase but at a faster rate.

This work not only provides theoretical understanding of online communities, but also implies potential applications. First, the scaling exponents can be treated as novel indicators to characterize the growth of communities. For example, the exponent $\alpha$ indicates the level of interactive stickness of a community since it increases with the intensity of the interactions between users. The merits of adopting the exponents to quantify the communities include the stability of the exponent and the independency on the size of the community. We can make a reasonable evaluation of a forum or a community when it is small.

Moreover, we can infer the parameters from the measured exponents. All these parameters describe the behaviours of users. And our work makes it possible to infer the individual behaviour only from their macroscopic performance of collective. It is also feasible that we can imply the macroscopic behaviour if we know the individual parameters.

Here we investigate the virtual mobility of human beings. Our model shows that the human mobility behaviours in the virtual world are similar to the real world. Various interactions among users play important roles in the virtual mobility, and further patterns of these phenomena could be studied in the future.

5. Data availability statement

Part of the data that support the findings of this study are openly available. The data of stack exchange can be obtained at https://meta.stackexchange.com/questions/198915/is-there-a-direct-download-link-with-a-raw-data-dump-of-stack-overflow-not-a-t/199303. The data of Digg can be obtained at https://www.isi.edu/integration/people/lorman/downloads.html. The data of Yelp can be obtained at https://www.yelp.com/dataset/challenge.

Acknowledgments

The authors gratefully acknowledge funding supports from the National Natural Science Foundation of China (Grant: 61673070), the Fundamental Research Fund for the Central Universities (Grant: 310421103), and the Beijing Normal University Interdisciplinary Project. We also gratefully thank the anonymous referees for providing earnest and constructive comments on the manuscript.

Appendix

We could see that the interactive model carried in the two-dimensional simulation reproduced four different statistical scaling laws as in the real world. Although the modelling patterns strongly characterized human activities in the online world, only one case might be not sufficient to state the robustness. The primary purpose of this analysis is trying to give a much clearer understanding of the original human behaviours in the interest space and the emerging universal scaling laws. We will discuss this problem in four extreme cases according to the dimensions, step-length distribution, and power law exponent.
First, we will review the simulation results of the two-dimensional interactive Lévy Flight model. Figures A1 and A2 show the sub-linear and super-linear scaling under different $\lambda$ and $p$ values. Figure A3 shows an example of four macroscopic quantities where $\lambda = 2.0$ and $p = 0.4$. Now, we will change the simulation conditions to compare with the above results.

(i). When $\lambda > 3$

According to our dynamical rules, the movement pattern of a user is basically random, and the probability density distribution function of the movement distance $l$ in one jump satisfies power law (equation (13)). $\lambda$ is the exponent of the Lévy Flight, and it characterizes how broad the interests of users are. Generally, the value of $\lambda$ is in the range of 1.0 and 3.0 [4, 12]. We will analyse the situation when extending the $\lambda$ value.

\[ P(l) \sim l^{-\lambda} \quad \text{(A.1)} \]

Figures A4 and A5 give the cases of $\lambda = 3.1$ and $\lambda = 4.0$ respectively. Meanwhile, figure A6 shows the example of four macroscopic quantities where $\lambda = 4.0$ and $p = 0.4$.

From figures A4 and A5 we can see that the macroscopic trends of four curves have not changed, but the end point of each curve is higher than figures A1 and A2 respectively.

(ii). When the step length of random walker satisfy exponential distribution($l \sim E(\lambda)$)

We changed the individual’s step length distribution form power law to the exponential distribution, and this can check the validity of our interactive rules in the model. Figures A7 and A8 give the cases of $\lambda = 1.0$ and $\lambda = 2.0$ respectively.

From figures A7 and A8, we can see that the macroscopic trends of four curves have not changed. The endpoint of each curve is higher than figures A1 and A2, and at the same time, they’re lower than the endpoints in case i. From this simulation, we can see that our interactive rules guaranteed the scaling laws even when the random walk pattern of individual changed. Besides, the random walk pattern could also change exponents values.

Figure A9 shows the example of four macroscopic quantities where $\lambda = 2.0$ and $p = 0.4$.

(iii). When the interactive Lévy Flight model placed onto one-dimensional Euclid space
Figure A2. The simulation of four exponents in two-dimensional space with $\lambda = 2.0$. Two-dimensional interactive Lévy Flight model.

In the above simulation, the active sites with digital resources are placed onto the two-dimensional Euclid space as users’ interest space. Here, we will change the dimension to analysis if the scaling laws are independent of the dimension. The interactive rules of users in the virtual space are the same as the above cases.

The results are presented in figures A10 and A11. One obvious fact we can seize from these two graphics is that the curves are a touch of flat in the range of $0 < p < 0.9$, which is very different from the curves in other cases. Another fact is that the endpoints of curves in figures A10 and A11 are higher than the two-dimensional cases. However, the macroscopic trends of four curves keep unchanged as before. Figure A12 shows the example of four macroscopic quantities where $\lambda = 2.0$ and $p = 0.4$.

(iv). When the interactive Lévy Flight model placed onto three-dimensional Euclid space

Similarly, we changed the interest space into 3-dimension. Figures A13 and A14 give the cases of $\lambda = 1.0$ and $\lambda = 2.0$ respectively. From figures A7 and A8 we can get that the macroscopic trends of four curves have not changed too. But the curves are steeper than the 2-dimension or 1-dimension simulations. Meanwhile, the endpoint of each curve is lower than the 2-dimension or 1-dimension simulations. However, the macroscopic trends of four exponents still keep steady. Figure A15 shows the example of four macroscopic quantities where $\lambda = 2.0$ and $p = 0.4$.

To sum up, in this research, we mainly focused on mining the universal statistical laws of human’s daily activities in the virtual space. We proposed an interactive Lévy Flight model to simulate user’s online behaviour, which reproduced four remarkable scaling laws as same as the calculation of real data sets. We also considered another four modelling cases with the same interactive rules but different conditions to check the robustness of simulation. Totally, the five simulation cases proofed the effectiveness of our interactive rules, though the exponents’ values have a slight difference. In the virtual space, two important parameters controlled the individual’s behaviour, i.e., how wide are users interests and the propensity that a user delivers a post.

Besides, we have got some new findings through the five simulation cases. We found that the shapes of exponents curves become steep as the dimension of model increase. In the 1-dimension case, the four exponents curves are flatter than 2-dimension. While in the 3-dimension case, the four exponents curves climb up fast. Another obvious fact is that as the dimension grows, the endpoints of four exponent curves will move down. For example, when $\lambda = 2.0$ and $p = 1.0$ the values of four exponents in figure A11 are larger than values in figure A2, and the values in figure A2 are larger than figure A14.
Figure A3. The scaling quantities of $N$, $A$, $D$, $E$ in two-dimensional space with $\lambda = 2.0$, $p = 0.4$. Two-dimensional interactive Lévy Flight model.
Figure A4. The simulation of four exponents in two-dimensional space with $\lambda = 3.1$. Two-dimensional interactive Lévy Flight model.
Figure A5. The simulation of four exponents in two-dimensional space with $\lambda = 4.0$. Two-dimensional interactive Lévy Flight model.
Figure A6. The scaling quantities of $N$, $A$, $D$, $E$ in two-dimensional space with $\lambda = 4.0$, $p = 0.4$. Two-dimensional interactive Lévy Flight model.
Figure A7. The simulation of four exponents in two-dimensional space with $\lambda = 1.0$. Two-dimensional interactive random walk model.
Figure A8. The simulation of four exponents in two-dimensional space with $\lambda = 2.0$. Two-dimensional interactive random walk model.
Figure A9. The scaling quantities of $N, A, D, E$ in two-dimensional space with $\lambda = 2.0, p = 0.4$. Two-dimensional interactive random walk model.
Figure A10. The simulation of four exponents in two-dimensional space with $\lambda = 1.0$. One-dimensional interactive Lévy Flight model.
Figure A11. The simulation of four exponents in two-dimensional space with $\lambda = 2.0$. One-dimensional interactive Lévy Flight model.
Figure A12. The scaling quantities of \(N, A, D, E\) in two-dimensional space with \(\lambda = 2.0, \rho = 0.4\). One-dimensional interactive Lévy Flight model.
Figure A13. The simulation of four exponents in two-dimensional space with $\lambda = 1.0$. One-dimensional interactive Lévy Flight model.
Figure A14. The simulation of four exponents in two-dimensional space with $\lambda = 2.0$. One-dimensional interactive Lévy Flight model.
Figure A15. The scaling quantities of $N$, $A$, $D$, $E$ in two-dimensional space with $\lambda = 2.0$, $p = 0.4$. One-dimensional interactive Lévy Flight model.
ORCID iDs

Fanqi Zeng  https://orcid.org/0000-0003-2233-5947
Qinghua Chen  https://orcid.org/0000-0001-8226-0006

References

[1] Grenfell B T, Bjørnstad O N and Kappey J 2001 Nature 414 716
[2] Belik V, Geisel T and Brockmann D 2011 Phys. Rev. X 1 3103–106
[3] Filippo R, Andrea B and Amaral L A N 2012 PLoS One 7 e29910
[4] Baronchelli A and Radicchi F 2013 Chaos Solitons Fractals 56 101–05
[5] Namboodiri V M, Levy J M, Mihalas S, Sims D W and Hussain Shuler M G 2016 Proc. Natl Acad. Sci. 113 8747–52
[6] Ratti C, Sobolevsky S, Calabrese F, Andris C, Beades J, Martino M, Claxton R and Strogatz S H 2010 PLoS One 5 e14248
[7] Santi P, Resta G, Szel G, Sobolevsky S, Strogatz S and Ratti C 2013 Proc. Natl Acad. Sci. 111 13290–4
[8] Brockmann D, Hufnagel L and Geisel T 2006 Nature 439 462–65
[9] González M C, Hidalgo C A and Barabási A L 2008 Nature 453 779–82
[10] Song C, Koren T, Wang P and Barabási A 2010 Nat. Phys. 6 818–23
[11] Berkolaiko G, Havlin S, Larralde H and Weiss G H 1998 Phys. Rev. E 53 5774–778
[12] Raposo E P, Buldyrev S V, Luz M G E D, Viswanathan G M and Stanley H E 2009 J. Phys. A: Math. Theor. 42 434003
[13] Rhee I, Shin M, Hong S, Lee K, Kim S J and Chong S 2011 IEEE/ACM Trans. Netw. 19 630–43
[14] Bartumeus F, Catalan J, Fulco U L, Lyra M L and Viswanathan G M 2002 Phys. Rev. Lett. 88 097901
[15] Lambiotte R, Blondel V D, Kerchove C D, Huens E, Prieur C, Smoreda Z and Dooren P V 2008 Physica A 387 5317–325
[16] Liang H, Silva R N, Ooi W T and Motani M 2009 Multimedia Tools Appl. 45 163–90
[17] Wu F and Huberman B A 2007 Proc. Natl Acad. Sci. 104 17599
[18] Wu L, Jiang Z and Min Z 2014 PLoS One 9 e102646
[19] Sasahara K, Hirata Y, Toyoda M, Kitsuregawa M and Aihara K 2013 PLoS One 8 e61823
[20] Kenett D Y, Morstatter F, Stanley H E and Liu H 2014 PLoS One 9 e102001
[21] Havelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P and Ratti C 2014 Cartogr. Geogr. Inf. Sci. 41 260
[22] Reynolds A M 2013 Am. Nat. 181 555–61
[23] Javarone M A 2016 Eur. Phys. J. B 89 42
[24] Watts D J and Dodds P S 2007 J. Consum. Res. 34 441–58
[25] Smith D, Menon S and Sivakumar K 2005 J. Interact. Market. 19 15–7
[26] Armano G and Javarone M A 2017 Sci. Rep. 7 1–8
[27] Liben-Nowell D and Kleinberg J 2003 The link prediction problem for social networks 12th Int. Conf. on Information and Knowledge Management (New Orleans, LA, November 2003) 556–59
[28] Cho E, Myers S A and Leskovec J 2011 Friendship and mobility-user movement in location-based social networks ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (San Diego, CA, USA, 21–24 August 2011) pp 1082–090
[29] Barabási A L 2005 Nature 439 207
[30] Nakamura T, Kyiono K, Yoshiuchi K, Nakahara R, Struzik Z R and Yamamoto Y 2007 Phys. Rev. Lett. 99 138103
[31] Shi P, Huang X, Wang J, Zhang J, Deng S and Wu Y 2015 PLoS One 10 e0136243
[32] Zeng W, Fu C W, Arisona S M, Schubiger S, Burkhard R and Ma K L 2017 IEEE Trans. Intell. Transport. Syst. 18 2271–84
[33] Frank M R, Mitchell L, Dodds P S and Danforth C M 2013 Sci. Rep. 3 2625
[34] Yuan T, Cheng J, Zhang X, Liu Q and Lu H 2015 Knowl.-Based Syst. 88 70–4
[35] Pennacchiotti M and Gurrumurthy S 2011 Investigating topic models for social media user recommendation Int. Conf. on World Wide Web, WWW 2011 (Hyderabad, India, March 28–April 1, 2011) pp 101–02
[36] Cole M J, Hendahewa C, Belkin N J and Shah C 2015 ACM Trans. Inf. Syst. 33 1
[37] Middleton S E, Shadbolt N R and Roure D C D 2004 ACM Trans. Inf. Syst. 22 54–8
[38] Guo L, Lou X, Shi P, Wang J, Huang X and Zhang J 2015 Physica A 437 235–48
[39] Lou X, Li Y, Gu W and Zhang J 2016 PLoS One 11 e0165240
[40] Blagus N, Subelj L and Bajec M 2012 Physica A 391 2794–802