SCIENCE CHINA
Information Sciences

• LETTER •

May 2017, Vol. 60 059102:1–059102:3
doi: 10.1007/s11432-015-0567-9

Common patterns of online collective attention flow

Yong LI[1,2], Xiaofeng MENG[1*], Qiang ZHANG[2], Jiang ZHANG[3] & Changqing WANG[4]

[1] *School of Information, Renmin University of China, Beijing* 100872*, China;*
[2] *College of Computer Science and Engineering, Northwest Normal University, Lanzhou* 730070*, China;*
[3] *School of Systems Science, Beijing Normal University, Beijing* 100875*, China;*
[4] *DNSLAB, China Internet Network Information Center, Beijing* 100190*, China*

Received December 9, 2015; accepted August 30, 2016; published online December 15, 2016

Citation Li Y, Meng X F, Zhang Q, et al. Common patterns of online collective attention flow. Sci China Inf Sci, 2017, 60(5): 059102, doi: 10.1007/s11432-015-0567-9

Dear editor,

During the past two decades, we have witnessed the explosive growth of the World Wide Web. To understand the common patterns of how this complex artificial software system evolves has been a more and more important and active research area for scientists [1]. If we see the Web as a virtual tissue, according to the metabolic theory, the website must absorb "energy" to grow, reproduce, and develop. We are interested in the following questions: (1) Where does the "energy" come from? (2) What are the common patterns of this energy flow?

Inspired by metabolic theory, we view the Web as virtual living tissues that grow at the cost of online collective users' attention flow as the "energy". We quantify the metabolic rate of websites, and obtain the Web version of Kleiber's law [2] (see Figure S1), which describes how the collective attention as energy is fed into sites and dissipated out of sites.

We obtain data from China Internet Network Information Center (CNNIC) and establish a site-level attention flow network which is a weighted directed graph (see Appendix B in the supporting information and Figure S2 for detail). As shown in Table 1, we define six basic variables related to weight of each site to infer the common patterns of co-evolution of the Web and online collective attention flow. Based on the similarity of the discrete version of a generalized beta distribution (DGBD) [3] curves (see Figure S3), we make an educated guess that there may exist a connection between the basic variables. The analysis reveals several surprising findings, such as allometric scaling laws, dissipation laws, gravity laws, and Heaps' law.

Allometric scaling laws. A general expression describing the relation between attention flow intensity A_i and node degree D_i may be found if the logarithm of A_i is plotted against the logarithm of D_i. As shown in Figure S4, the fitting result is a straight line and the explained variance of the OLS (ordinary least square) regression is 0.90, indicating that the logarithm of A_i is proportional to the logarithm of D_i, which means that the relation between A_i and D_i satisfies a power law:

$$A_i \sim kD_i^{\alpha}, \tag{1}$$

where k is a normalization constant and the exponent α is a parameter to be estimated. We often ignore the normalization constant k and write $A_i \sim D_i^{\alpha}$. In our case, α=1.33.

As another Kleiber's law [2], Eq. (1) can be viewed as the allometric scaling law at the Web level. From the exponent α, we can determine the relationship between the attention flow intensity

* Corresponding author (email: xfmeng@ruc.edu.cn)
The authors declare that they have no conflict of interest.

© Science China Press and Springer-Verlag Berlin Heidelberg 2017

Table 1 Symbols and definitions of the basic variables

Symbol	Definition
A_i	Total inflow and outflow intensity of a given site i
T_i	Total dwelling time of collective users on a site i
P_i	Total page views of a site i on an observing duration
D_i	Degree of node i, total inward edges and outward edges
I_i	Total attention flow from "source" node (node 0), $(W_{0,i})$
H_i	Attention flow dissipated to "sink" node (node $n+1$), $(W_{i,n+1})$

and the node degree. Our finding of α greater than one shows that A_i scales supper-linearly against D_i; this phenomenon is called the accelerating growth in ecology [4]. This allometric scaling relationship indicates that the core nodes (sites) of attention flow network have the greatest nodes degree and attention flow intensity. For example, Baidu[1)], being not only a searching engine but also an information and contents provider site in China, has the greatest node degree and attention flow intensity. Figure S5 shows the allometric scaling relationship $T_i \sim P_i^{1.36}$, which also indicates accelerating growth pattern. We can infer from this pattern that the core sites have more page views which can attract more online user dwelling time.

In addition, we also find sub-linear scaling relationship in attention flow network. For example, we analyse the relationship between the attention flow intensity A_i and the collective page views P_i. As shown in Figure S6, $A_i \sim P_i^{0.73}$ and R^2 is greater than 0.8. Note that α being 0.73 smaller than one, which is called the "decelerating growth" in ecology. This result is not consistent with the common sense that "the more page views being taken up, the greater attention flow of a site is". From the supper-linearly scaling relationship of $T_i \sim P_i^{1.36}$ and sub-linear scaling relationship of $A_i \sim P_i^{0.73}$, we can conclude that the more page views can attract more online user dwelling time but cannot attract more online attention flow intensity. Therefore, the core factor of a site is its contents to attract more online collective attention flow rather than the click rate. These results are consistent with what we found in [5].

Dissipation laws. As pointed by the earlier ecological studies [4], in the whole ecosystem, a large fraction of energy flow dissipates to the environment. The dissipative processes have different forms, such as respiration and excretion. In complex systems, all the open systems need to dissipate energy flow to the environment to maintain the process of existence, reproduction, and evolution. For maintaining moving toward an equilibrium state, the systems must not only increase their negative entropy by consuming energy, but also eliminate the positive entropy that naturally accumulates over time as systems are trying to sustain their existence and evolution [6].

As noted by famous physicist Schrodinger: "what an organism feeds upon is negative entropy" [7]. We view the online collective attention flow as "energy" for the Web evolution. The development of the Web system, as well as the process by which they come into existence, growth, reproduction, and expire, is governed by the transfer of energy from online collective attention flow. The information (or contents) on the Web will gradually become obsolete and no longer concerned by online collective users, which can be said to be the positive entropy generated by the evolution of time. The Web system has a potential to offset the increasing entropic trend by consuming online collective attention flow and absorbing information from the outside world, thus creating negative entropy, which prevents the Web system from moving toward an equilibrium state.

In this perspective, as shown in Figure S2, the "energy" infuses into network from the "source" or other sites and dissipates to the "sink" node. The dissipated attention flow can be captured by the variable H_i, which indicates the attention flow intensity of a site i infusing into the "sink" node. Empirically, if a site has a stronger stickiness for collective attention flow dwelling time(T_i), it has a smaller dissipated attention flow. This phenomenon can be formulated mathematically as

$$H_i \sim kT_i^{\beta}, \tag{2}$$

where k and β are parameters to be estimated and Eq. (2) is called the dissipation law in this paper.

As shown in Figure S7, the estimated exponent β is 0.81 which is less than one and R^2 is 0.85; thus the empirical data supports our conjecture. Note that β less than one shows a sub-linear pattern: the dissipation rate (dissipation per stickiness) increases with a lower rate compared to the dwelling time T_i, which means that the stronger of the site stickiness, the smaller dissipated attention flow.

In Figure S3, we find that I_i satisfies a similar distribution to one of H_i. We again make a conjec-

1) http://www.baidu.com/.

ture that there is a connection between I_i and H_i. As shown in Figure S8, the empirical data confirms our conjecture. $H_i \sim kI_i^{\beta}$ and $\beta = 0.96 \simeq 1$, which means that if a site has more inward attention flow from "source", it has more dissipated attention flow. This pattern implies that the dissipation rate decreases as the inward attention flow from other sites (non-"source" vertex) increases.

Gravity laws. In classical physics, the Newton's Law of universal Gravitation states that any two bodies attract each other with a force g_{ij}, where $g_{ij} = Gm_i m_j/r^2$, G is the gravity constant, m_i and m_j are their masses, and r is the distance between them. In recent years, researchers found that the gravity law has long been one of the most successful empirical models in economics and other complex systems [8]. In attention flow network, we still do not know how the attention flow distributes among different pairs of sites. It turns out that the attention flow W_{ij} between sites i and j is as follows:

$$W_{ij} \sim \frac{k(A_i A_j)^{\gamma}}{D_i D_j}, \tag{3}$$

where k is a constant, A_i and A_j are total flow intensity of the sites, D_i and D_j are degree of the sites, and γ is fitting exponent.

Figure S10 shows the gravity law phenomenon and the parameter in attention flow network. The R^2 has value of 0.89 indicates that this pattern is significant. In addition, we fit other basic variables with the gravity law, where γ is 0.58, R^2 is 0.75 for dwelling time T_i; γ is 0.74 and R^2 is 0.83 for overall page views P_i. But these two patterns are not as significant as the total flow intensity A_i because the R^2s are smaller than 0.85. From gravity law of Eq. (3), we could predict the attention flow between two sites by flow intensity and degree of each site.

Heaps' law. Heaps' law is commonly observed in characterizing natural language processing [9, 10]. Similar results were found in online resources and programming languages, such as Java, C and C++. However, as far as we know, there is no report about the Heaps' law in online collective behaviors. We found that the growth of distinct sites in online collective page views satisfies the Heaps law in attention flow network. As shown in Figure S11, the Heaps' law is formulated as

$$N_s \sim pv^{\theta},$$

where N_s is the number of distinct sites when the collective page views length is pv, and $\theta = 0.78$ is the Heaps' exponent. The explained variance of the regression for fitted line is $R^2 = 1.00$, which means that this pattern is quite significant.

Conclusion. We found four interesting common patterns in online collective attention flow. These patterns have some potential applications, such as advertisement targeting, websites ranking, etc. However, we still do not know the underlying mechanisms of these patterns. The theoretical model and underlying mechanisms of the online collective attention flow are left for the future work.

Acknowledgements This work was supported by National Natural Science foundation of China (Grant Nos. 71263045, 61662070, 61379050, 91224008, 61532010, 61532016), Research Funds of Renmin University of China (Grant No. 11XNL010), and Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20130004130001).

Supporting information Figures S1–S8, S10, S11, and Appendix B. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Huberman B A, Pirolli P L T, Pitkow P J, et al. Strong regularities in world wide web surfing. Science, 1998, 280: 95–97

2 Mackenzie D. New clues to why size equals destiny. Science, 1999, 284: 1607–1609

3 Martnez-Mekler G, Martnez R A, del Rio M B, et al. Universality of rank-ordering distributions in the arts and sciences. PloS ONE, 2009, 4: e4791

4 Zhang J, Feng Y. Common patterns of energy flow and biomass distribution on weighted food webs. Phys A Stat Mech Appl, 2014, 405: 278–288

5 Li Y, Zhang J, Meng X, et al. Quantifying the influence of websites based on online collective attention flow. J Comput Sci Tech, 2015, 30: 1175–1187

6 Hubler A, Crutchfield J. Order and disorder in open systems. Complexity, 2010, 16: 6–9

7 Schrodinger E. What is Life? The Physical Aspect of the Living Cell and Mind. Cambridge: Cambridge University Press, 1967. 24–25

8 Anderson J E. The gravity model. Ann Rev Econ, 2011, 3: 133–160

9 Lv L, Zhang Z, Zhou T. Zipf's law leads to Heaps'law: analyzing their relation in finite-Size systems. PLoS ONE, 2010, 5: e14139

10 Liang X, Zhao J C, Xu K. A general law of human mobility. Sci China Inf Sci, 2015, 58: 100103