The Hidden Geometry of Attention Diffusion
Lingfei Wu\textsuperscript{1,2}, Cheng-Jun Wang\textsuperscript{4}, Marco Janssen\textsuperscript{1,2}, Jiang Zhang\textsuperscript{4}, Min Zhao\textsuperscript{5}
\textsuperscript{1} School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, U.S.
\textsuperscript{2} Center for the Study of Institutional Diversity, Arizona State University, Tempe, AZ 85281, U.S.
\textsuperscript{3} School of Journalism and Communication, Nanjing University, Nanjing 210093, P.R. China
\textsuperscript{4} School of Systems Science, Beijing Normal University, Beijing 100875, P. R. China
\textsuperscript{5} Baidu Inc., Baidu Campus, Haidian District, Beijing 100085, P. R. China

Abstract
It is difficult to develop quantitative hypotheses to describe the diffusion of information among users, because, in theory, information resources can have an infinite number of copies. And this theoretical possibility has become a reality in an age of information explosion, when a Twitter meme may spread rapidly to millions of users in a few hours. To address this issue, we propose to study the transportation of users’ attention between information resources. We use clicks as a proxy of attention and construct attention networks using the browsing log files from 1,000 Web forums. Our previous research discovered the scaling relationship between clicks and users with an exponent $\theta$ that characterizes the efficiency of forums in spreading information. In this paper we propose a geometric model to explain this scaling property. We view attention networks as $d$-dimensional flow-balls that satisfy $\theta = (d + 1)/d$ and we find that the time-invariant parameter $d$ can be estimated from the spatial distribution of flow dissipation in the balls. In other words, we are able to predict the temporal dynamics of Web forums ($\theta$) by analyzing the structure of attention networks.

Introduction
A majority of studies on information diffusion focus on the transmission of information among users. A strong limitation of this approach is that, in theory, information can be copied for infinite times, therefore, it is difficult to place constraints on these diffusion models in order to test empirical data against theoretical expectations. For example, the popular “S-curve” model based on a logistic function containing two or more parameters can always be used to fit the growth in user adoption at the aggregated level, no matter what structure the information diffusion network is, which also means that the model lacks falsifiability. As a consequence, scholars have to seek for exogenous factors such as information content or user properties to develop quantitative hypotheses.

As an alternative approach we suggest studying the flow of collective attention between information resources. The information resources can be news, threads, videos, etc., depending on the type of online system to be studied. Within a given time period, a user may interact (e.g., browse, vote, or comment) with many resources sequentially, generating a stream of attention connecting all visited resources. Putting together a large number of individual streams, we will have an “attention network” that shows how collective attention is transported between resources. Using clicks as an approximation of attention helps us focus on traceable activities in the cyber-world. But this also requires us to collect data sets that contain anonymized user information such that we can recognize the same identity of people on their activities across time.

Compared with the models of information flow, modeling attention flow will bring at least two bonuses. Firstly, we now have a conserved quantity in the system that allows us to write down equations as quantitative assumptions. This is because the total amount of collective attention (clicks) in an online
community during a short time period is limited. Secondly, the interpretation of the latent space of attention networks is easier and more straightforward than the interpretation of the latent space of user networks. In this paper, we show that the attention networks constructed from the browsing log files of 1,000 Web forums can be as viewed $d$-dimensional “flow balls” such that 1) the number of users and clicks corresponds to the dissipation $I$ and the flow storage $B$ of the balls, which naturally gives the scaling relationship between users and clicks with an exponent $\theta = (d + 1)/d$; 2) attention always flows along the radius $L_i$ from the center to the edge of the sphere, with the direction of attention flow revealing the collective preferences of users.

We are not the first researchers to explore the diffusion of attention. Cattuto et al. studied the flow of attention in semantic space using the data from social tagging systems [5]. Bollen et al. collected user’s clickstreams between journals to visualize a map of human knowledge [6]. There is also an increasing interest in the transport of attention between news [7], tweets [8], and scientific papers [9]. However, we extend these studies by suggesting a unified, geometric model of attention flow. This model not only has applied consequences, such as predicting long-term temporal dynamics of online systems from attention network snapshots, but also contributes to a unified theory of flow systems. For example, we suggest that attention networks, or virtual flow systems, are different from real-world flow systems such as food webs, although they exhibit very similar allometric scaling patterns [10,11]. In particular, the former are multi-source systems and the latter are single-source systems. Our simulation on 2-D flow balls shows that the value of $\theta$ decreases with the increase of the number of sources in the system. Our exploration for a unified theory of flow systems is greatly inspired by Banavar and Dreyer’s works on geometric models of flow systems [12,13].

Materials and Methods

Data Sources

Three data sets collected from Alexa ([www.alexa.com](http://www.alexa.com)), Digg (urlwww.digg.com), and Tieba ([http://tieba.baidu.com/](http://tieba.baidu.com/)) are used in the current study. The Alexa dataset contains the global user traffic between top websites. The constructed network contains 516 nodes and 2,490 links. The Digg dataset was collected by [14]. It includes 3,018,197 votes on 3,553 popular news stories made by 139,409 distinct users over a period of 36 days. The constructed network contains 3,552 nodes and 211,045 edges. We use these two datasets to demonstrate the direction of attention flow in online systems. Tieba is an online system containing a large number of Chinese forums of specific topics. We get access to the log files of the top 1,000 forums, whose size (the averaged number of daily clicks) varies from thousands to millions. For each forum, we select the data in two months from February 27 to April 27 in 2013 and construct 1,440 successive hourly-based networks. We use the Tieba networks to investigate the scaling relationship between the numbers of clicks and users.

Attention Networks and Key Variables

Figure 1 presents an example attention network, in which nodes are information resources and edges represent the switch of users between resources. More precisely, the nodes are websites in the Alexa network, news stories in the Digg network, and threads in the Tieba network. The Digg and Tieba data sets only contain individual records thus we have to aggregate all individual switches between a pair of nodes to derive the weight of the edge between them. After the attention networks have been constructed, we balance the flow on networks by adding two nodes “source” and “sink”, which represent the environment of the networks. For each node, we added a link from “source” if its weighted in-degree is smaller than its weighted out-degree and we add a link from this node to “sink” if otherwise. By doing this we retrieve the missing information of the exchange of flow between the networks and the
Figure 1. An example of an attention network in which nodes are information resources and edges represent users switching between resources. The supplied/dissipated flow $I$ of the network equals to 80, and the total amount of flow in the network equals 235. The node-level dissipation $D_i$ from the first to the fifth node is $\{0, 10, 25, 10, 35\}$ and the passing-through flow $T_i$ on these nodes is $\{80, 60, 35, 20, 40\}$. Note that $I$ equals the sum of $D_i$ across these five nodes and $B$ equals the sum of $T_i$, as suggested by Eq. 1 and Eq. 2. The flow distance $L_i$ of nodes from “sources” can be calculated by using Eq. 6 as $L_i = \{1, 365/164, 193/82, 529/164, 285/82, 63/16\}$, in which the distance from “source” to “sink” $L_6 = 63/16 \approx 4$. This means that an average random walker will go through four steps before leaving the system.

environment. The resulting attention networks satisfy the principle of “flow conservation” [15], i.e., input equal to output on every node and for the entire network.

We define two network-level variables $I$ and $B$ and three node-level variables $D_i$, $T_i$, and $L_i$. $I$ is the flow supplied to (or dissipated from) the entire network and $B$ is the total amount of flow stored in the network. They equal 80 and 235 in the example network presented by Figure 1 respectively. $D_i$ is the dissipation from node $i$ to “sink” and $T_i$ is the passing-through flow on node $i$. $D_i$ and $T_i$ equals $\{0, 10, 25, 10, 35\}$ and $\{80, 60, 35, 20, 40\}$ from the first to the last nodes in the example network, respectively. As the example network is balanced, we can derive that

$$I = \sum_{i=1}^{N} D_i, \quad (1)$$

and

$$B = \sum_{i=1}^{N} T_i, \quad (2)$$

in which $N$ is the number of nodes (not including “source” and “sink”) in the network.

We propose a novel indicator “flow distance” $L_i$ as a measure of the number of steps from “source” to node $i$ averaged over all possible paths. We convert the example network into a weighted matrix and
then normalize all entries by the sum of the corresponding column. In the normalized matrix $M$, an entry $m_{ij}$ indicates the probability that a random walker at $i$ comes from the connected node $j$. We can calculate $L_i$ as

$$L_i = 1 + \sum_{j=1}^{N+1} m_{ij} L_j.$$  (3)

This is because a random walker at $i$ either comes from “source” or comes from another node $j$. The first case occurs at a probability of $m_{i,0}$ and generates a “flow distance” that equals 1. For each candidate node $j$ in the second case, the corresponding probability is $m_{i,j}$ and existing flow length is $L_j$. By considering all possible paths in the second case, we obtain $\sum_{j=1}^{N+1} m_{ij}(1 + L_j)$, in which 1 denotes the extra step from $j$ to $i$. Putting together these two cases, we have

$$l_i = m_{i,0} \cdot 1 + \sum_{j=1}^{N+1} m_{ij}(1 + L_j) = m_{i,0} + \sum_{j=1}^{N+1} m_{ij} + \sum_{j=1}^{N+1} m_{ij}L_j = 1 + \sum_{j=0}^{N} m_{ij}L_j.$$  (4)

Eq.3 can be expressed and solved using matrix notation. Defining vector $L = (l_0, l_1, \cdots, l_N)^T$ gives us the matrix form of Eq.3 as

$$L = A + ML,$$  (5)

in which $A = (1, 1, \cdots, 1)^T$ is an all-ones vector of length $N + 1$. From Eq. 5 we derive that

$$L = (A - M)^{-1} \cdot A.$$  (6)

Using $U$ to denote the inverse matrix we write Eq.6 as

$$L = U \cdot A = \sum_{j=0}^{N} U_{ij}.$$  (7)

The result of Eq.7 is a vector $L$ of length $N + 1$, in which the $i$th element indicates the average steps from “source” to the $i$th node, i.e., the flow distance.

For empirical analysis when the network is extremely large, e.g., containing millions of nodes, we use an iterative calculation method based on Eq.3 instead of doing matrix calculation (Eq.7) directly. In particular, we set an initialized constant $L_i = 1$ for all nodes and iterate over their in-links repetitively to update the entries in $L_i$ until they converge. A similar technique has been used to calculate the Page Ranks of web sites in very large hyperlink networks [16].

### Results

#### The Hidden Order of Attention Flow

The Web is perhaps the most ideal free market ever in human history. It allows users to follow their own preferences in trading attention for information. As a consequence, the individual browsing activities are random-like and it is difficult to see patterns. However, when we put together a large number of individual records to explore how users act collectively, order begins to emerge from chaos.

We assume that like the flow of water in river networks [17], the flow of collective attention also has a direction. The only difference is that the former is shaped by the physical landscape in the real world, whereas the latter is constrained by the invisible landscape of user preference in the virtual world. Following this idea, we calculate $L_i$, the average length of paths from the environment to node $i$ in attention networks, and call it “flow distance.” We also examine whether the calculated flow distance has a theoretical interpretation.
For the empirical analysis we investigate two attention networks. The Alexa network contains top websites as nodes and the transportation of user traffic between these sites as edges. The Digg network contains news stories as nodes and the navigation of users between the stories as links. We find that in the Alexa network the attention of global Web users flows from English websites to Chinese websites (Figure 2A). This means that it is easier for information to diffuse from the former to the latter than the opposite. Meanwhile, in the Digg network attention flow is transported from latest news to the older news, which is consistent with our daily life experience (Figure 2B). We visualize these two networks using ball-like structures, in which attention always flows along the radius $L_i$ from the center to the edge of the sphere.

**The Allometric Growth of Attention Balls**

In our previous studies, we discovered a universal scaling relationship between the number of clicks and users across various online systems \[10\,18\,19\]. We also suggested that in attention networks these two variables correspond to the flow storage $B$ and dissipated flow $I$, respectively \[10\,11\]. In reviewing the literature of hydrology and biology, we find that the scaling between $B$ and $I$, also called “allometric growth,” has been widely observed in real-world flow systems such as river networks \[17\] and vascular systems \[20\].
Figure 3. The growth of an attention network as a $d$-dimensional flow-ball. This plot is based on Figure 2 in [13].

To explain this pattern, Dreyer proposed a simple model in which flow networks are viewed as “flow-balls” supplied by a central source [13]. Streams are imported into the system from the source and are transported and dissipated along the radius of the ball before vanishing at the sphere. Dreyer suggested that, in order to maintain a flow ball, the flow density $f(r)$ at distance $r$ from the center should be $R^d - r^d$, in which $R = r_{max}$ is the radius of the sphere and $d$ is the dimension of the embedding space (Figure 3). Therefore, the total amount of flow in the ball is equal to the integration of $f(r)$ over the entire volume of the ball, which is proportional to $R^{D+1}$. Meanwhile, as sinks are assumed to have a constant dissipating rate and distribute uniformly along the radius, the dissipated flow of the entire system equals the size of the ball, which is proportional to $R^d$. Putting these conditions together, we can derive the general relationship between dissipation $I$ and flow storage $B$ as

$$B \sim I^{\theta = (d+1)/d}. \quad (8)$$

$d = 2$ in river systems corresponds to Hack’s law [17] and $d = 3$ in vascular systems corresponds to Kleiber’s law [20]. In attention networks, $d$ could be much greater than 3 since the value of $\theta$ is found to be very close to 1 in some cases [10, 21] suggests that $\theta$ characterizes the storage capacity of flow systems. This conclusion can also be used to explain attention networks constructed from collective browsing activities, in which larger $\theta$ means a longer average surfing length. As a consequence, users are exposed to more information resources. From the perspective of information diffusion, we can also use $\theta$ as an indicator for the efficiency of systems in delivering information to users.

Our previous findings extended scaling laws from the real world to the virtual world [10, 11], but left two unexplored problems. Unlike real-world flow systems, attention networks have multiple sources, i.e., users may enter into or leave from the system at any node. Meanwhile, how to define the “radius” of dissipative nodes that corresponds to $r$ in Dreyer’s model was unknown. However, these two problems are addressed in the current study: the attention networks can be viewed as single-source systems after we balance the networks by adding weighted, directed edges from the “source” (which represents the environment) to the other nodes; and the distance $L$ defined in Eq. 6 can be used as a proxy for $r$. These two conditions allow us to use Dreyer’s model to describe the attention networks, which are viewed as “attention balls” embedded in $d$-dimensional spaces.

A significant benefit of using Dreyer’s model is that it allows us to predict the long-term value of $\theta$ by analyzing a single network snapshot. In previous studies, to estimate $\theta$ scholars usually need to obtain many attention networks (balls), as in Figure 4 A, where 24 hourly-based networks are constructed.
using the data from the “EXO” (a pop band in Korea) forum. However, if Dreyer’s model is true, then the attention balls formed in different hours should have the same embedding dimension. This is because this model requires attention networks to preserve self-similar structures during their growth. As a consequence, we should be able to obtain the value of $\theta$ and $d$ by analyzing the self-similarity of dissipations within a single network.

To test this assumption, in each of the 24 EXO networks, we integrate the passing-through flow $T_i$ and the dissipated flow $D_i$ along the radius $r = L_i$ to derive a scaling function that “simulates” the scaling relationship between storage $B$ and dissipation $I$ during the growth of networks (Figure 3):

$$B_{L_i} = \int_{0}^{L_i} T_i \sim (\int_{0}^{L_i} D_i)^\mu = I_{L_i}^\mu.$$  \hspace{1cm} (9)

Eq. 9 quantifies the self-similar geometry of attention balls. When $L_i = \text{max}(L_i) = R$, Eq. 9 degenerates to Eq. 8. But before $L_i$ approaches its maximum value, we can calculate from data many pairs of $B_{L_i}$ and $I_{L_i}$ and use them to fit the value of $\mu$ in a log-log scale plot.

In the empirical analysis we find that the size of forums has an effect on the correlation between $\mu$ and $\theta$. To remove this effect, we use the following function to obtain a size-adjusted value of $\mu$

$$B_{L_i} = A \ast D_{L_{max}}^L I_{L_i}^\mu.$$  \hspace{1cm} (10)

which gives the size-adjusted value of $\mu$ as $\mu + a$. For the EXO forum we observe that the “simulated” scaling properties of hourly networks (Figure 3 C) are highly consistent with the “real” scaling dynamics across 24 hours (Figure 4 A). This observation is further confirmed by the systematic investigation of the relationship between the size-adjusted $\mu$ and $\theta$ across all 1,000 forums in the Tieba dataset (Figure 4 D).

**Attention Network as Multi-source systems**

In the methodology section we introduced how to turn a multi-source flow network into a single-source one by adding two artificial nodes “source” and “sink” to balance the network. This technique can also be used to compare these two types of systems. As Dreyer’s model does not specify the topological structure of networks, we have to use another model to explore the consequences of multiple sources. We choose Banavar’s model, which has the same geometric properties as Dreyer’s model but provides more specific assumptions on linking topology [12].

Banavar’s model is a tree-like flow network supplied by a central source and is also embedded in a $d$-dimensional space. The nodes (sinks), which have a constant rate of dissipation that equals 1, are distributed uniformly within the space. And the average distance from the sinks to the central source is proportional to the linear size of the space, $L$. Therefore the dissipation of the entire network ($I$) is proportional to $L^d$, and the total flow on edges ($B$) is proportional to $L^d \ast L = L^{d+1}$. This gives the scaling relationship between $B$ and $I$ as in Dreyer’s model (Eq. 8). The first row of Figure 5 gives a 2-d Banavar network and also its dissipative and allometric patterns.

In the original version of Banavar’s model, flow can only be transported locally in the embedding space. More precisely, nodes on level $L$ only receive flow from nodes on level $L - 1$. This can be traced back to the source after $L$ steps, forming a hierarchical flow structure. Now we revise the model and introduce multiple sources into the system. In particular, we use the central source to represent the environment of the system, which supplies flow directly to $k$ percent of nodes randomly selected as new sources. By doing this, we break the flow hierarchy and obtain a new flow structure of multi-source systems. Figure 5 C gives a flow network in which 10% nodes become sources. Figure 5 E shows the dissipative patterns of the revised model. Note that with the increase of $k$, the decaying curve of the passing-through flow $T_i$ in the original model (Figure 5 B) turns into a unimodal curve that replicates the empirical pattern presented in Figure 4 B. This process leads to a decrease in the average surfing length of users and thus lowers the value of the scaling exponent $\theta$. Eventually, when all nodes import
flow directly from the environment, the function of $T_i$ on $L$ and the function of $D_i$ on $L$ are overlapped, and the scaling exponent $\theta$ collapses to 1. Remember that $\theta$ characterizes the efficiency of online systems in spreading information: our simulation on 2-d Banavar networks suggests that a system with more entrances could be less efficient in holding users within the system and delivering information to them.
Conclusion and Discussion

We explore the structure of attention networks using empirical data and find that they can be viewed as $d$-dimensional flow-balls. The center of the ball is the most likely focal attention of a random user, and the further away from the center the less attention a random user in the community spends on this information. This method could be used to analyze the performance of various online social systems. Do people spend attention on the topics the organizers want them to? Is there a difference between small and bigger communities in terms of the flow structure of attention? All these questions are relevant to
the method presented in the current study.

We also find that the widely observed allometric growth pattern of online communities [18,19] can be explained by the ball-like topology of attention diffusion. Using the \(d\)-dimensional flow-ball model we can predict the exponent of allometric growth by analyzing the dissipation patterns on network snapshots. This technique relates the long-term growth dynamics of online communities to their underlying geometric structures and thus paves the way for quantifying the evolution of online communities.

Finally, we conduct simulations to compare the dynamics between single-source and multi-source systems [12]. Our simulation on 2-D flow networks shows that both types of systems demonstrate the allometric scaling pattern but the scaling exponent decreases when the system has more sources. This analysis may contribute to the exploration for a unified theory of flow systems [12,13].

Acknowledgment

We acknowledge financial support for this work from the National Science Foundation, grant number 1210856.

References

1. Rogers EM (2010) Diffusion of innovations. Simon and Schuster.
2. Goel S, Watts DJ, Goldstein DG (2012) The structure of online diffusion networks. In: Proceedings of the 13th ACM conference on electronic commerce. ACM, pp. 623–638.
3. Hong L, Dan O, Davison BD (2011) Predicting popular messages in twitter. In: Proceedings of the 20th international conference companion on World wide web. ACM, pp. 57–58.
4. Pennacchiotti M, Popescu AM (2011) Democrats, republicans and starbucks afficionados: user classification in twitter. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 430–438.
5. Cattuto C, Loreto V, Pietronero L (2007) Semiotic dynamics and collaborative tagging. Proceedings of the National Academy of Sciences 104: 1461–1464.
6. Bollen J, Van de Sompel H, Hagberg A, Bettencourt L, Chute R, et al. (2009) Clickstream data yields high-resolution maps of science. PLoS One 4: e4803.
7. Wu F, Huberman BA (2007) Novelty and collective attention. Proceedings of the National Academy of Sciences 104: 17599–17601.
8. Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. Scientific Reports 2.
9. Kuhn T, Perc M, Helbing D (2014) Inheritance patterns in citation networks reveal scientific memes. Available at SSRN 2426903 .
10. Wu L, Zhang J, Zhao M (2014) The metabolism and growth of web forums. PloS one 9: e102646.
11. Zhang J, Wu L (2013) Allometry and dissipation of ecological flow networks. PloS one 8: e72525.
12. Banavar JR, Maritan A, Rinaldo A (1999) Size and form in efficient transportation networks. Nature 399: 130–132.
13. Dreyer O (2001) Allometric scaling and central source systems. Physical review letters 87: 038101.
14. Lerman K, Ghosh R (2010) Information contagion: An empirical study of the spread of news on digg and twitter social networks. ICWSM 10: 90–97.

15. Higashi M (1986) Extended input-output flow analysis of ecosystems. Ecological Modelling 32: 137–147.

16. Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems 30: 107–117.

17. Rinaldo A, Rodriguez-Iturbe I, Rigon R, Bras RL, Ijjasz-Vasquez E, et al. (1992) Minimum energy and fractal structures of drainage networks. Water Resources Research 28: 2183–2195.

18. Wu L, Zhang J (2011) Accelerating growth and size-dependent distribution of human online activities. Physical Review E 84: 026113.

19. Wu L, Zhang J (2013) The decentralized flow structure of clickstreams on the web. The European Physical Journal B 86: 1–6.

20. West GB, Brown JH, Enquist BJ (1997) A general model for the origin of allometric scaling laws in biology. Science 276: 122–126.

21. Zhang J, Guo L (2010) Scaling behaviors of weighted food webs as energy transportation networks. Journal of Theoretical Biology 264: 760–770.