Eur. Phys. J. B **76**, 179–183 (2010)
DOI: 10.1140/epjb/e2010-00208-1

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Information explosion on complex networks and control

X.J. Ma[1,2], W.-X. Wang[1,a], Y.-C. Lai[1,3], and Z. Zheng[2]

[1] School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA
[2] Department of Physics, Beijing Normal University, Beijing 100875, P.R. China
[3] Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

Received 28 October 2009 / Received in final form 11 May 2010
Published online 2 July 2010 –

Abstract. When a piece of information spreads on a complex network, error or distortion can occur. For a high error probability, the phenomenon of information explosion can occur where the number of distinct pieces of information on the network increases continuously with time. We construct a physical model to address this phenomenon. The transition to information explosion as the error probability increases through a critical value is uncovered and elucidated, and a control strategy is articulated to maximize the robustness of the network against information explosion, which is validated by both numerical computations and a mean-field based analysis.

1 Introduction

Spreading and transportation processes are fundamental and ubiquitous in a variety of complex systems: the Internet, biological networks, and social networks. Most existing works have addressed how the underlying network structure affects the spreading and transportation dynamics, with efforts ranging from routing data traffic on the Internet [1–9] to the spreading of epidemic [10–15], opinions and rumors [16–20] on either social networks or communication networks. Often, one focus of analysis and computation is on the asymptotic extent of the spreading process, as characterized by the percentage of infected nodes after the termination of the process. In this regard, various processes such as those described by the two-state spreading model (SIS) [11,12,14], the voter model [16–18], and the rumor-spreading model [19,20] have been studied. In most previous works, the entity of spreading, such as a particular type of virus or a piece of information, is assumed to be invariant during the process. In realistic situations, distortion of the entity during the spreading process can be expected, such as mutations of viruses, errors in transported data packets and distorted opinion or rumors [21]. The problem of information distortion is particularly relevant when human behaviors are involved in the spreading and transportation process [22]. The distortions can lead to a significant increase, or even a divergence in the number of messages on the network over the time. The purpose of this paper is to address the problem of information explosion on complex networks. Information explosion has occurred in the modern time. There has been an unprecedented growth in the number and variety of data collections as technology and network connectivity become increasingly affordable [23,24]. Distortion in communication is inevitable and may contribute partially to the growth of data information. How to hold and release information becomes an issue of increasing importance, with implications ranging from personal privacy to national security [24,25].

The starting points of our consideration are the following: (i) a node (or an agent) accepts or discards a message based on the existent information content in its memory, and (ii) information distortion can occur during the spreading process, which can be quantified by the probability p that a message is distorted after passing through an agent. In principle, the values of p can be different for different agents, but for simplicity we shall assume that the spread in the probabilities is small and can be neglected. To gain insight, we examine the case where one message is set out to spread on the network initially. In the error-free case of $p = 0$, the number of messages is simply one. For $p \gtrsim 0$, we expect the number of messages to be greater than one. However, since p is small, a steady state may arise where the average number of messages on the network tends to a constant. For large values of p, due to the frequent mutations, the number of distinct messages can increase with time. This introduces a positive feedback mechanism that generates an increasing amount of difficulty for agents to distinguish between the true and modified messages. As a result, different versions of the true message can accumulate in the memories of agents, generating even more distorted messages. As a result, the number of messages can keep increasing with time, leading to information explosion. In general, we expect that as p is increased from zero and passes through a critical point p_c, a transition can occur from steady state to

[a] e-mail: wenxuw@gmail.com

information explosion. A main finding of this paper is that this scenario can indeed occur on complex networks.

Another result of this paper concerns a network's robustness to information explosion, which can be measured by the value of p_c, where a higher value indicates that the network is more robust. An issue then concerns about some control strategy to increase the network robustness. We shall demonstrate in this paper that a process that controls an agent's selection of a neighbor to spread the message to can be used to maximize the value of p_c. We shall derive a theoretical analysis of the controlled strategy of selection and provide numerical support.

Qualitatively, spreading dynamics leading to information explosion can be understood, as follows. Each agent is endowed an internal memory to store what they have received from their neighbors. The incorporation of the memory is motivated by previous works on task queueing model and naming game [22], in which memories for agents are a key characteristic. When an agent receives a message that is different from any of the messages in its memory, the new message will be stored in the memory. Messages that are different from the original ones are a result of distortion during the spreading process. In each communication, the message received by an agent differs from that in the memory of the agent that sends the message with probability p. Physically, the distortion may be due to noise, errors in communication, mutation, or cheating and rumoring. Every message in an agent's memory is weighted by the frequency that the message is received. More frequently communicated messages are more likely to stay in the memories of various agents for a longer time, and less frequently received messages are more likely to be disregarded. For a small value of p, mutated messages are few in number and their frequencies are low, so they are more likely to be removed. In this case, the number of messages in the network can reach a steady state. However, for large p values, the balance between message mutation and removal may be broken, leading to information explosion.

2 Model

Our model for information explosion in the absence of control can then be described, as follows. All agents possess unlimited memory lengths and start with empty memories, and an agent is randomly chosen to initiate an original message with weight one. At each time step, every agent i whose memory is not empty selects, completely randomly, a neighboring agent j, after which agent i picks the message, denoted by a, with the largest weight in its memory and sends it to j. With probability p, j receives a different message, say a', which is included in its memory with weight $w_{a'} = 1$. With probability $1 - p$, j receives the same message a. If a already exists in j's memory before the communication, a's weight w_a becomes $w_a + 1$ (weight strengthening). At the same time, the message with the lowest weight is removed from j's memory. If a is a new message to j, a will be added to j's memory with weight 1 and no deletion occurs. Note that mutated messages are always new to all agents and if there are

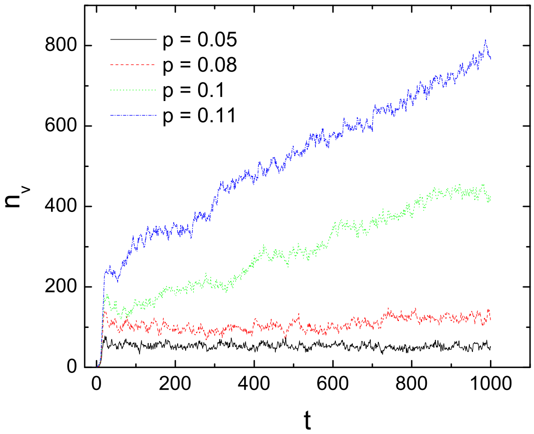

Fig. 1. (Color online) For a scale-free network of 500 nodes and average degree $\langle k \rangle = 6$, time evolutions of the total numbers of different messages for different values of the distortion probability p, where an agent selects randomly one of its neighbors to transmit the message.

more than one message with the same largest or the lowest weight, one of them is picked randomly. The weight-strengthening mechanism describes the fact that one may trust more a message if it is received more times than other messages. Since different messages tend to contradict each other in an agent's memory, if an agent trusts a message more, others will carry less weights. Because of this, weight strengthening and message deletion occur simultaneously. In a social network, people usually tell others what they believe. In our model, this fact is incorporated by allowing agents to transmit messages with the largest weight to their neighbors.

To provide a numerical example for the scenario to information explosion, we implement our model on a standard scale-free network [26]. Suppose a seed message is initiated at time $t = 0$. We monitor the evolution of the total number of distinct messages, n_v, for different values of p, as shown in Figure 1. We see that, after a short transient phase, n_v exhibits two characteristically different types of behavior, depending on the value of p. For small values of p, the numbers of distorted and removed messages are balanced, resulting in a steady state in which n_v is characterized by small random fluctuations about a constant mean value. When p exceeds the critical value $p_c \approx 0.087$, on average n_v keeps increasing with time, signifying information explosion, as the cases for $p = 0.10$ and $p = 0.11$ in Figure 1.

A key issue in spreading dynamics is the selection of a neighbor for an agent to transmit a message to. This process is usually not completely random. For example, an agent may choose a particularly "important" neighbor, i.e., a neighboring node with relatively large degree, to pass the message onto. Motivated by this consideration, we propose the following controlled selection strategy. At each time step, every agent i whose memory is not empty selects a neighboring agent j according to the probability

$$\Phi = k_j^{\beta} \Big/ \sum_{l \in \Gamma_i} k_l^{\beta}, \tag{1}$$

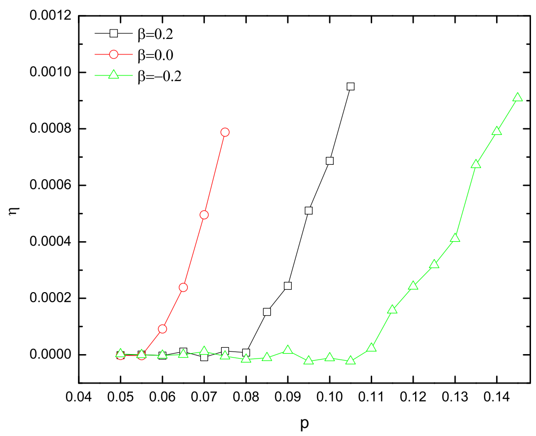

Fig. 2. (Color online) Order parameter η as a function of the distortion probability p for different values of the control parameter β. The network ensemble is the same as in Figure 1.

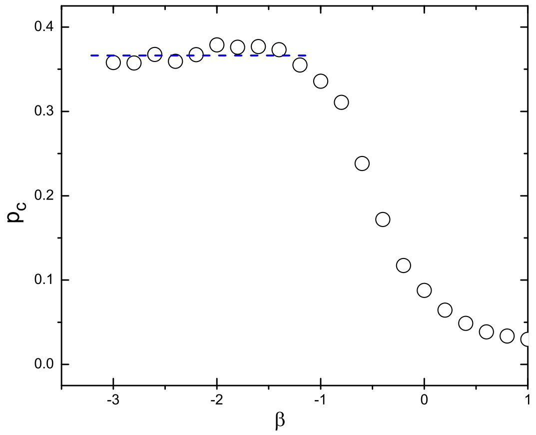

Fig. 3. (Color online) Critical distortion probability p_c as a function of the control parameter β. The network ensemble is the same as in Figure 1. The data points are obtained by numerical simulations and the dash line is the analytical estimation.

where Γ_i is the set of neighboring nodes of i, k is the node degree, and β is a tunable parameter. For $\beta > 0$ ($\beta < 0$), neighbors with larger (smaller) degrees are more likely to be chosen. Assuming that information explosion is undesirable and is to be prevented, we ask whether the network robustness to the explosion can be enhanced by adjusting β. That is, can the value of the critical point p_c be made larger by choosing value of β in a proper range?

To quantify the transition from steady state to information explosion so that the value of p_c can be obtained accurately , we introduce the following order parameter:

$$\eta = \lim_{t\to\infty} \frac{n_v(t+\Delta t) - n_v(t)}{N\Delta t}, \tag{2}$$

where N is the maximum possible number of distorted messages generated per unit time so that the value of η is confined in $[0, 1]$. In a steady state, $n_v(t + \Delta t) - n_v(t)$ does not increase as Δt increases so that η is zero for large enough t and Δt. When an information explosion occurs, $n_v(t+\Delta t)-n_v(t)$ increases with Δt so that η will be larger than zero. Thus, the transition can be characterized by a sudden increase in η from zero to a positive value. Simulation results of η as a function of p for different values of β are shown in Figure 2. It can be seen that, for any given value of β, when p exceeds a critical value p_c, η increases rather abruptly from approximately zero to some positive value, as anticipated. Figure 3 shows p_c versus β. We see that, as β is decreased from some positive value, the value of p_c increases. This trend continues until β reaches the value $\bar{\beta} \approx -1$. For $\beta < \bar{\beta}$, p_c remains approximately at a constant value. The network is thus most robust to resist information explosion for $\beta < \bar{\beta}$. This means that, when small-degree nodes are frequently chosen to receive messages, the likelihood for information explosion can be reduced significantly as compared with the case where high-degree nodes are selected. The value $\bar{\beta} \approx -1$ can thus be regarded as the *onset* of network robustness against information explosion. In what follows we present a mean-field approach to analytically predicting this onset value.

For an arbitrary agent i, the content of information stored in its memory changes with time during the spreading process, which can be characterized by L_i, the length of the occupied portion of its memory. In a steady state, we have $dL_i/dt = 0$ for $i = 1, \ldots, N$. To gain insights into the dynamical behaviors of the spreading process in terms of the value of p_c for positive and negative values of β, we consider two extreme cases: β is positively large and negatively large, respectively. For $\beta > 0$, nodes with large degrees are more likely to be chosen to receive an message. If β is positive and sufficiently large, the agent with the largest degree in Γ_i will be chosen by node i (if there is more than one neighbor of i with the same largest degree, they have the same probability to be chosen). The largest degree agent in the network will then be chosen by all its neighbors. For simplicity, we assume that there is only one largest degree agent in the network. Let R_i be the number of messages received by node i from its neighbors at each time step. For the largest-degree agent, we then have $R_{k_{\max}} \approx k_{\max}$. For $t > 0$, all new messages are due to distortions and an old message is removed for this agent at each time step. We thus have $dL_{k_{\max}}/dt \approx k_{\max}p - (1 - k_{\max}p)$, where the first term describes the increase in $L_{k_{\max}}$ and the second term represents the decrease of a message with probability $1 - k_{\max}p$ averagely. For a standard scale-free network, we have $k_{\max} \approx k_{\min}\sqrt{N}$. The condition $dL_i/dt = 0$ thus leads to

$$p_c(\beta \to \infty) \sim \frac{1}{2k_{\max}} \approx \frac{1}{2k_{\min}\sqrt{N}}. \tag{3}$$

This means that for sufficiently large values of β, p_c tends to a small constant, regardless of the values of β.

Similarly, for the smallest-degree agent, we have

$$p_c(\beta \to -\infty) \sim \frac{1}{2k_{\min}}. \tag{4}$$

Indication that $p_c(\beta < 0) > p_c(\beta > 0)$ can be seen by comparing equation (4) to (3), where we see that

$p_c(\beta \to -\infty)$ is much larger than $p_c(\beta \to \infty)$. This comparison is only valid in a qualitative sense because a scale-free network typically has many nodes that share the same smallest degree.

To derive the value of $\bar{\beta}$ for the onset of the plateau in p_c for $\beta < 0$, we note that, for an arbitrary node i, two factors contribute to changes in L_i: receipt of new messages and removal of the existing message with the lowest weight. We have

$$\frac{dL_i}{dt} = R_i p + (1-p)cR_i - (1-p)(1-c)1, \qquad (5)$$

where c denotes the probability that the received message is new to i but no distortion occurs in the transmission. Averagely, c can be estimated by

$$c \approx \frac{n_v}{n_t}, \qquad (6)$$

where n_v is the number of different messages in the network and n_t is the total number of messages. This estimation can be validated by considering the fact that if all the messages are the same, i.e. $n_v = 1$, c tends to be zero because $n_t \geq N$; while if all messages are different, i.e. $n_v = n_t$, all the messages transmitted from neighbors are new as confirmed by $c = 1$. In the absence of information explosion, both n_v and n_t are independent of time, as exemplified in Figure 1. We can then write

$$\begin{aligned}\frac{dL_i}{dt} &= \sum_j \frac{A_{i,j}k_i^\beta}{\sum_l A_{j,l}k_l^\beta} p + \sum_j \frac{A_{i,j}k_i^\beta}{\sum_l A_{j,l}k_l^\beta} \\ &\quad \times (1-p)c - (1-p)(1-c)1 \\ &= \sum_j \frac{A_{i,j}k_i^\beta}{\sum_l A_{j,l}k_l^\beta}(p + (1-p)c) - (1-p)(1-c)1\end{aligned} \qquad (7)$$

for $j, l = 1, \ldots, N$. The first term on the right represents the increase in L_i resulting from the distortion during the transmission from node i's neighbors to i, the second term is the increase in L_i induced by the non-distorted messages that are new to node i, and the last term denotes the decrease caused by removal of messages. To reach a steady state, the condition $dL_i/dt = 0$ should be satisfied, which leads to the critical value p_c for any value of β. We use the mean-field approximation to simplify equation (7). For a network with negligible degree-degree correlation among nodes, a common property observed in real-world networks, we have

$$\begin{aligned}\sum_l A_{jl}k_l^\beta &\approx k_j \sum_{k=k_{\min}}^{k_{\max}} p(k'|k_j)(k')^\beta \\ &\approx \frac{k_j}{\langle k \rangle} \sum_{k=k_{\min}}^{k_{\max}} (k')^{\beta+1} p(k') \approx \frac{k_j \langle k^{\beta+1} \rangle}{\langle k \rangle}.\end{aligned} \qquad (8)$$

Substituting equation (8) into equation (7), after some algebra we have

$$\frac{dL_i}{dt} = W\left[p + (1-p)\frac{n_v}{n_t}\right] - (1-p)\left(1 - \frac{n_v}{n_t}\right),$$

where $W = k_i^{\beta+1}/\langle k^{\beta+1} \rangle$. Setting $dL_i/dt = 0$ yields

$$p_c(i) = \frac{1 - \frac{n_v}{n_t} - W\frac{n_v}{n_t}}{(W+1)(1 - \frac{n_v}{n_t})}. \qquad (9)$$

The critical value of p_c should be the lowest values of $p_c(i)$ among all agents. To attain the highest value of p_c, $p_c(i)$ for all agents should assume values to keep the system as far away as possible from the state of information explosion. A heuristic choice $\beta = -1$, for which $p_c(i)$ is independent of degree of i. This analytical estimate is consistent with the numerical result, as shown in Figure 3.

The message removal strategy plays a key role in achieving possible information consensus in the presence of information distortion as, in the absence of message removal, convergence can never occur for any values of p. In general, there can be alternative ways to construct the message-removal scheme in the convergence process. For example, one can exploit "weakest message weakening" instead of the "weakest message removal" in the current model. In particular, in the weakest-message weakening scheme, it can be assumed that at each time, a randomly selected individual tends to decrease the weight of the message with the lowest weight in its memory. This alternative rule can avoid the situation where a message with relatively low weight in the memory is deleted right after it is strengthened. While this scenario is likely, the natural convergent process in the presence of information distortion is well captured by the current model with respect to decision making. The weakest-message weakening can result in similar consensus behavior and information explosion, which depend on the information distortion probability in the same way as in the current model. The advantage of the current scheme lies in the fast achievement of information consensus attributed to the removal of a large number of weakest message at each time. This can effectively reduce redundant information stored in the memories of individuals, which could otherwise generate interference to reduce the reliability of information.

It is noteworthy that the phenomenon of information explosion reported in this paper is quite different from the congestion phenomenon in data traffic on complex networks, e.g., in references [1–9], although both correspond to a continuous phase transition. In particular, our model for information explosion differs from the existing network traffic models in several aspects. Firstly, in traffic models, there is a pair of origin destination nodes for each data packet transmitted according to some routing protocol. After a packet arrives at its destination, it is removed from the network. This scenario is characteristic of, e.g., information transmission on the Internet. In our model, however, it is not necessary to distinguish between origin and destination for messages and routing strategies to guarantee successful transmission. Secondly, along a packet-delivering path, all nodes are passive in the sense that they act as routers that function to forward packets. In this case, the content of the packet cannot be modified. In our model, however, nodes are active because they represent agents that serve to evaluate the messages received from their neighbors and determine whether a message should be preserved or discarded. Thus, all messages are weighted and the weights of repeated messages are strengthened and messages with low weights are discarded with nonzero probability. Distortion, weight strengthening and abandonment can occur at any time on any node. Thirdly, in existing network traffic models, data packets are usually different, regardless of whether the system is in a free-flow or congested state. In contrast, in our model, in the absence of information explosion, most messages are identical in the network. In fact, distorted messages are minority and their weights are relatively small. Under these conditions, all agents can reach consensus about information or opinions in the system. This is a unique feature in the dynamics of opinion propagation and spreading on complex networks, which is not captured by any existing network-traffic models.

3 Conclusion

In conclusion, we have constructed a physical model to address the phenomenon of information explosion on complex networks due to inevitable distortion and errors in the messages occurring during the spreading process. The essence of our approach is to incorporate a memory-based decision making mechanism. Our computations and analysis reveal a transition from steady state to information explosion as the error probability passes through a critical value. We have also proposed a control strategy to direct messages to small-degree nodes so as to maximize the network robustness against information explosion, and we have obtained quantitative prediction based on the mean-field approximation for the optimal range of the control parameter, characterized by higher probabilities for selecting smaller-degree nodes to transmit messages. This is consistent with the intuition that hub nodes are generally capable of spreading errors on a large scale so that they should be avoided as much as possible. Information explosion can be particularly relevant to social networks where the chances for information distortion can be significant due to human behaviors.

WXW and YCL are supported by AFOSR under Grant No. FA9550-10-1-0083. XJM's visit to Arizona State University is sponsored by the State Schorlarship Fund of China. XJM and ZGZ are partially supported by NNSF of China Nos. 10575010 and 10875011, the 973 Program No. 2007CB814805, the Foundation of Doctoral Training No. 20060027009, and the Science Fund for Creative Research Groups of BNU.

References

1. A. Arenas, A. Díaz-Guilera, R. Guimerà, Phys. Rev. Lett. **86**, 3196 (2001)
2. P. Echenique, J. Gómez-Gardeñes, Y. Moreno, Phys. Rev. E **70**, 056105 (2004)
3. P. Echenique, J. Gómez-Gardeñes, Y. Moreno, Europhys. Lett. **71**, 325 (2005)
4. L. Zhao, Y.-C. Lai, K. Park, N. Ye, Phys. Rev. E **71**, 026125 (2005)
5. D. De Martino, L. Dall'Asta, G. Bianconi, M. Marsili, Phys. Rev. E **79**, 015101(R) (2009)
6. W.-X. Wang, B.-H. Wang, C.-Y. Yin, Y.-B. Xie, T. Zhou, Phys. Rev. E **73**, 026111 (2006)
7. W.-X. Wang, C.-Y. Yin, G. Yan, B.-H. Wang, Phys. Rev. E **74**, 016101 (2006)
8. M.-B. Hu, W.-X. Wang, R. Jiang, Q.-S. Wu, Y.-H. Wu, Phys. Rev. E **75**, 036102 (2007)
9. M.-B. Hu, W.-X. Wang, R. Jiang, Q.-S. Wu, Y.-H. Wu, Europhys. Lett. **79**, 14003 (2007)
10. R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001)
11. M. Boguñá, R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003)
12. M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. **92**, 178701 (2004)
13. V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, Proc. Natl. Acad. Sci. USA **103**, 2015 (2006)
14. L. Huang, K. Park, Y.-C. Lai, Phys. Rev. E **73**, 035103 (2006)
15. R. Yang, L. Huang, Y.-C. Lai, Phys. Rev. E **78**, 026111 (2008)
16. H.-U. Stark, C.J. Tessone, F. Schweitzer, Phys. Rev. Lett. **101**, 018701 (2008)
17. V. Sood, S. Redner, Phys. Rev. Lett. **94**, 178701 (2005)
18. C. Nardini, B. Kozma, A. Barrat, Phys. Rev. Lett. **100**, 158701 (2008)
19. Y. Moreno, M. Nekovee, A.F. Pacheco, Phys. Rev. E **69**, 066130 (2004)
20. M. McDonald, O. Suleman, S. Williams, S. Howison, N.F. Johnson, Phys. Rev. E **77**, 046110 (2008)
21. For convenience, we shall refer a data packet, a virus, an opinion, or a rumor, collectively to as a piece of information or a *message*)
22. The role of human decision making in opinion dominance has been studied, for example, through the task-queueing model [e.g., A.-L. Barabási, Nature **435**, 207 (2005) and A. Vázquez, Phys. Rev. Lett. **95**, 248701 (2005)] and the naming-game approach [e.g., A. Baronchelli, L. Dall'Asta, A. Barrat, V. Loreto, Phys. Rev. E **73**, 015102 (2006) and A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, L. Steels, J. Stat. Mech. P06014 (2006)], which take into account a memory effect. However, how to incorporate human behaviors that can often lead to information distortions into spreading dynamics on social networks remains to be an open issue.
23. J.C. Green, Science **144**, 646 (1964)
24. L. Sweeney, *Information Explosion. Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, edited by L. Zayatz, P. Doyle, J. Theeuwes, J. Lane (Urban Institute, Washington, DC, 2001)
25. M.J. Eppler, J. Mengis, Inf. Soc. **20**, 325 (2004)
26. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)