Maximizing Network Capacity with Optimal Source Selection: A Network Science Perspective

Chunxiao Jiang, *Member, IEEE*, Yan Chen, *Senior Member, IEEE*, Yong Ren, *Member, IEEE*, and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—How to enhance the network capacity is one of the most important issues. To achieve this, the existing works have focused on improving either the network structure or routing strategies with a common assumption of uniformly distributing the replicas of information among the nodes in the network. The nodes associated with information replicas are considered as source nodes (or server). However, for many networks such as the Internet, some nodes have much more traffics than the others, exhibiting an asymmetric phenomenon. In this letter, we study the optimal source selection strategy to enhance the network capacity, where an optimization model is proposed to find the optimal source selection probability distribution. Simulation results show that in homogeneous networks, most of the nodes can be the sources. While in heterogeneous networks such as the scale-free networks, only a small number of the nodes can be the sources. Moreover, an interesting phenomenon is observed that the optimal proportion of source nodes in Erdős-Rényi random network and Barabási-Albert scale-free network exhibits a power law relationship with the network size.**

***Index Terms*—Client-server, network capacity, network science, source selection.**

I. Introduction

TRAFFIC dynamics on complex networks have attracted a lot of attentions [1]. Most of the interests come from some real-world complex networks including the Internet, the World Wide Web (WWW) and the high-way networks [2]. To deal with the ever-increasing amount of traffic in these networks, one needs to have a deep understanding upon the traffic behaviors in order to avoid congestion. Recent studies have presented some models to describe the traffic routing on complex networks by introducing the concepts of packet generating rate and randomly selected sources and destinations of each packet [3], [4]. These models use the critical packet generating rate, R_c, at which the phase transition from the free flow phase to the congestion phase occurs, to define the network capacity. These studies have proposed several ways to improve network capacities either by making changes to network structure [5]–[7] or by using different routing strategies [8]–[10].

Manuscript received October 27, 2014; revised December 01, 2014; accepted December 02, 2014. Date of publication December 08, 2014; date of current version December 15, 2014. This work was supported by NSFC China under Project 61371079, 61273214, 61271267, and 91338203. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael Rabbat

C. Jiang and Y. Ren are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: jchx@tsinghua.edu.cn; reny@tsinghua.edu.cn).

Y. Chen and K. J. R. Liu are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: yan@umd.edu; kjrliu@umd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2378264

All the aforementioned studies have made a common assumption: the source nodes in the networks are selected uniformly, i.e., the probability of one node being a source or not is equivalent. However, traffic in most of the practical networks is heterogeneous. For example, big web sites have much more visitors than the small ones. Although the original goal of the Internet was to interconnect existing networks [11], the Internet users are more interested in accessing resources rather than connecting to machines. The resource allocation within the complex network has expedited the network caching technology, i.e., some nodes serve as servers to cache and distribute popular resources [12]. In other words, these caching technologies use distributed replicas to change the distribution of the resources. By doing this, the network traffic can be changed to improve network capacity and enhance user experience. Therefore, uniform source selection strategy, i.e., uniformly distributing the resource replicas, may not be optimal for network capacity improvement. A nature question is what is the most efficient distribution of source nodes to maximize the capacity.

To answer this, we study the optimal sources selection strategy in this letter by formulating a network capacity optimization model. Through solving the optimization problem, the optimal source selection strategy and network capacity are found and analyzed. Simulation results show that in homogenous networks, the sources should be distributed uniformly; while in heterogeneous networks, only a few nodes can be the sources. Moreover, we also observe that the optimal proportions of source nodes in Erdős-Rényi random network and Barabási-Albert scale-free network exhibits a power law relationship with the network size. The problem solutions in this letter can be widely applied in network caching (how to control the proportion of nodes with caches), content-centric-networks (how to control the proportion of servers with popular resources), datacenter deployment (how to control the proportion of datacenter), etc. In the following, we will first illustrate the optimization model formulation, and then give the simulation results and the insights.

II. Optimal Source Selection Model

In the traffic routing model, all nodes in the network can be both hosts and routers for generating and delivering packets. The shortest path routing strategy is commonly used to forward packets. Each node has a packet queue that works on a "first-in-first-out" basis. When a packet is generated at a node or arrives at a node along its path, it is appended at the end of the queue. Once a packet reaches its destination, it would be removed from the network. Let us denote C as the delivery capacity of each node, i.e., the maximal number of packets a node can deliver at one time step. Suppose there are totally R packets generated in a N-node network at each time step, where the packets are generated at a node s with probability $p(s)$ and the destinations are chosen uniformly from the other $N-1$ nodes.

The network capacity can be characterized by the order parameter presented in [4] as follows

$$\eta(R) = \lim_{t\to\infty} \frac{C}{R}\frac{\langle \Delta W\rangle}{\Delta t}, \tag{1}$$

where $\langle \Delta W\rangle = W(t+1) - W(t)$ indicates the average number of packets during time windows Δt, and $W(t)$ is the total number of generated packets within the network at time t. When $R < R_c$, i.e., the critical point, we have $\langle \Delta W\rangle = 0$ and $\eta = 0$, which indicates that the network system is under the free-flow state. When $R > R_c$, η is above zero, which indicates that packets are accumulating in the network and the network will become congested. Therefore, R_c represents the maximal packets generated per time step for the network to maintain in the free-flow state, and is used as a measure of the overall capacity of the network system.

The critical point R_c is related to the betweenness centrality of the nodes. Betweenness centrality (BC) [13], is a measure of a node's centrality in a network, which equals to the number of shortest paths from all nodes to all others that pass through that node. The BC of a node v is given by

$$g(v) = \sum_{s\neq t, v\neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \tag{2}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through node v. If the sources and destinations are chosen uniformly, the probability of any packet to pass node v is

$$q(v) = \sum_{s\neq t, v\neq t} p(s,t)\frac{\sigma_{st}(v)}{\sigma_{st}} = \frac{g(v)}{N(N-1)}, \tag{3}$$

where $p(s,t) = \frac{1}{N(N-1)}$ is the probability of a packet to choose node s as its source and node t as its destination. Under the shortest path routing strategy, since the network congestion occurs at the node with the largest betweenness, R_c can be estimated by [4]

$$R_c = \frac{CN(N-1)}{g_{max}} = \frac{C}{\max_v q(v)}, \tag{4}$$

where g_{max} is the largest betweenness among all nodes.

In our proposed model, we meticulously select the source nodes to maximize the network capacity instead of uniform strategy. Since the destination nodes are mostly common network users, as in previous models [5]–[10], we also assume that the destinations of packets (or clients) are uniformly distributed. For example, in the network caching technology, the general network users who require the popular resources are destination nodes, and the network users in a complex network are statistically uniformly distributed. On the other hand, different from the previous models where the sources are also chosen uniformly, we select a node s as the source according to its probability $p(s)$. The probability of a packet starting from node s and ending in node t is

$$p(s,t) = p(s)p(t) = \frac{p(s)}{N-1}. \tag{5}$$

In such a case, the probability of any packet to pass node v can be calculated as follows

$$q(v) = \sum_{s\neq t, v\neq t} p(s,t)\frac{\sigma_{st}(v)}{\sigma_{st}} = \frac{1}{N-1}\sum_{s\neq t, v\neq t} p(s)\frac{\sigma_{st}(v)}{\sigma_{st}}. \tag{6}$$

We can calculate the source dependent betweenness matrix $\mathbf{A}$ by

$$A(s,v) = \frac{1}{N-1}\sum_{t(t\neq s, t\neq v)} \frac{\sigma_{st}(v)}{\sigma_{st}}, \tag{7}$$

where $A(s,v) = p(v|s)$ measures the conditional probability of any node which starts from node s to pass node v. Here, the $\frac{1}{N-1}$ is a normalized parameter and $g(v) = (N-1)\sum_s A(s,v)$. We can also calculate the BC of a node v with non-uniform source selection strategy as

$$h(v) = \sum_{s\neq t, v\neq t} \frac{p(s)p(t)}{\frac{1}{N(N-1)}}\cdot\frac{\sigma_{st}(v)}{\sigma_{st}} = \sum_{s\neq t, v\neq t} \frac{p(s)N\sigma_{st}(v)}{\sigma_{st}}, \tag{8}$$

which we call generalized BC with respect to the BC of nodes with uniform source selection in (2). In such a case, R_c can be estimated by

$$R_c = \frac{CN(N-1)}{h_{max}} = \frac{C}{\max\limits_v\{\sum\limits_s p(s)A(s,v)\}}. \tag{9}$$

To maximize the capacity R_c of a network with N nodes is to minimize the maximum probability of any packet to pass node v: $q(v)$, which is a min-max problem as follows:

$$\begin{aligned} \min \quad & \max_v q(v) = \textstyle\sum_s p(s)A(s,v), \\ \text{s.t.} \quad & 0 \le p(s) \le 1, \quad \textstyle\sum_s p(s) = 1. \end{aligned} \tag{10}$$

By introducing an auxiliary variable $z = \max_v q(v)$ ($v = 1,\cdots,N$) representing the maximum probability that a packet will pass a node v, the optimization problem (10) can be casted as a linear programming problem as follows

$$\begin{aligned} \min \quad & z, \\ \text{s.t.} \quad & \mathbf{Ap} - z\mathbf{1} \le \mathbf{0}, \mathbf{p}^T\mathbf{1} = 1, \mathbf{p} \ge \mathbf{0}. \end{aligned} \tag{11}$$

where $\mathbf{A} = [A(s,v)]$ as shown in (7), $\mathbf{p} = [p(s), s = 1,\cdots,N]^T$ and $\mathbf{1} = [1,\cdots,1]^T$. Thus, we can easily find the minimal $z^\star$ by linear programming algorithms.

A. Remarks

Note that when the following condition holds

$$\mathbf{A^{-1}1} \ge \mathbf{0}, \tag{12}$$

the optimal source selection strategy $\mathbf{p}^\star$ can be found by

$$\mathbf{p}^\star = z^\star\mathbf{A^{-1}1}. \tag{13}$$

Homogeneous networks, e.g., the lattice and random regular networks, has the property (12). The optimal source selection strategy suggests that the generalized betweenness centrality of every node can be uniformed. However, the property (12) usually does not hold for heterogenous networks, e.g., random Erdős-Rényi network (ER) [14] and Barabási-Albert scale free network(BA) [15]. Nevertheless, we can add a slack variable $\mathbf{y}$ to this linear programming problem and re-write it as follows

$$\begin{aligned} \min \quad & z, \\ \text{s.t.} \quad & \mathbf{A}\mathbf{p} + \mathbf{y} - z\mathbf{1} = \mathbf{0}, \\ & \mathbf{p}^T\mathbf{1} = 1, \mathbf{p} \geq \mathbf{0}, \mathbf{y} \geq \mathbf{0}. \end{aligned} \tag{14}$$

This new linear programming problem has $2N + 1$ variables: $\mathbf{p}, \mathbf{y}, z$ ($\mathbf{p}, \mathbf{y}$ are both vectors of length N) and $N + 1$ equality constraints: $\mathbf{A}\mathbf{p} + \mathbf{y} - z\mathbf{1} = \mathbf{0}, \mathbf{p}^T\mathbf{1} = 1$. The simplex method indicates that in the optimal result, at least N of these $2N + 1$ variables are 0. Since it is obvious that $z > 0$, we have

$$\mathcal{L}(\mathbf{p} = \mathbf{0}) + \mathcal{L}(\mathbf{y} = \mathbf{0}) \geq N, \tag{15}$$

where $\mathcal{L}(\cdot = 0)$ represents the number of 0 in one vector. Thus, the optimal number of nodes with non-zero probabilities to be selected as sources $N_{p(s)>0}$ satisfies that

$$N_{p(s)>0} = N - \mathcal{L}(\mathbf{p} = \mathbf{0}) \leq \mathcal{L}(\mathbf{y} = \mathbf{0}), \tag{16}$$

where $y(v) = 0$ means the generalized betweenness centrality of node v equals to the upper bound, i.e., $h(v) = \max_i h(i)$. The (16) indicates that the optimal number of source nodes in the network is bounded by the number of nodes that can reach the upper bound of betweenness. **In homogeneous networks, all of the nodes can have the same generalized betweenness after the optimization process, and thus most of the nodes can be source nodes. In heterogeneous networks, the variance of betweenness can be so large that only a small number of nodes' betweenness centralities can reach the upper bound, and thus only a small number of nodes can be sources.**

III. Simulation Results and Discussions

Due to the discovery of small-world [16] and scale-free [15] phenomena, there are extensive studies about network structures. We conduct simulation on four typical types of networks: square lattices with periodic boundary condition, random regular graph (RG), random Erdős-Rényi network (ER) [14] and Barabási-Albert scale free network(BA) [15] and present a comparative analysis of results from these networks. Note that the first two networks are homogeneous networks since every node has the same degree. The square lattice is in a two-dimensional Euclidean space with each vertex connected to its four neighbors while in RG each node chooses its neighbors randomly. The latter two networks are heterogeneous networks, where the ER has a Poisson degree distribution and the BA has a power-law (or scale-free) degree distribution. All these networks are configured with the same size $N = 1225$. The average degree $\overline{k} = 4$ for the lattice and $\overline{k} = 6$ for the other three networks. We set the node deliver capacity $C = 1$ for all the nodes for simplicity and 100 independent realizations are averaged for all simulations. By solving the optimization problem (10), the optimal source selection probability (i.e. $p(s)$) distribution can be obtained. The results are compared with uniform source selection strategy, i.e. $p(s) = \frac{1}{N}$.

Fig. 1 shows the order parameter η versus the packet-generating rate R in four different networks, respectively. We can see that the optimal strategy and uniform strategy behave nearly the same regarding network capacities R_c in the lattice and RG networks. When it comes to the random ER network and BA scale-free network, the network capacity R_c of optimal strategy outperforms that of the uniform strategy significantly. Therefore,

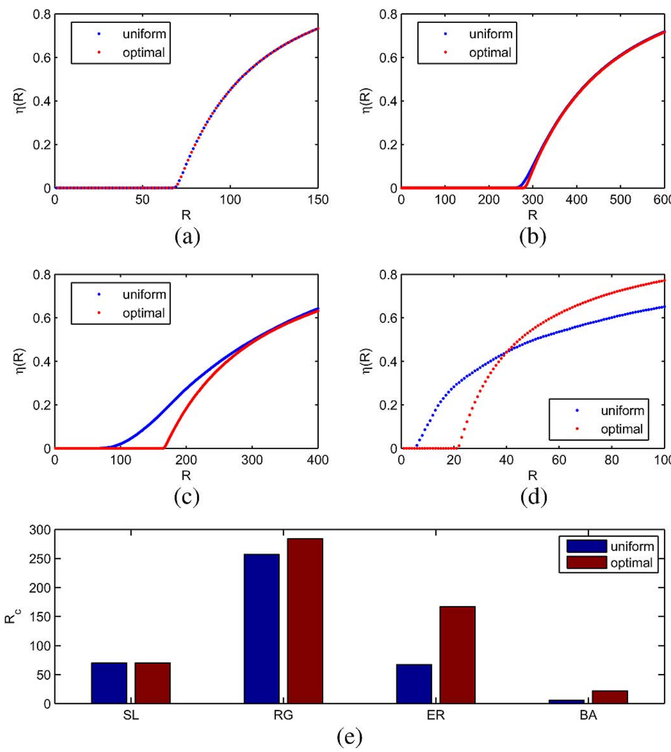

Fig. 1. The order parameter η versus the packet-generating rate R and the R_c for the following: (a) lattice, (b) random regular network (RG), (c) random Erdős-Rényi network (ER) and (d) Barabási-Albert scale free network (BA), where the blue and red dots correspond to uniform and optimal source selection strategies, respectively.

it can be concluded that the uniform source selection strategy is optimal or near optimal for the homogenous networks like square lattice and RG networks, while the ER and BA networks can benefit significantly from the optimal source selection strategy.

The generalized betweenness centralities (BC), i.e, $g(v)$ defined in equation (2) and $h(v)$ in equation (8), under these two strategies of all the four networks are compared in Fig. 2. Fig. 2-(a,b) show that all the nodes have the same generalized BC in the lattice and the RG has a narrow BC distribution range. Meanwhile, in those two homogenous networks, the generalized BC of all the nodes become exactly the same by the optimization process. In the ER and BA networks as shown in Fig. 2-(c,d), the ranges of BC are so large that they cannot be unified by optimization. Nevertheless, the optimization process can significantly narrow down the ranges and the upper bounds can be clearly observed. Recall that in Section II, we have proved that the optimal number of source nodes in heterogenous network is bounded by the number of nodes that can reach this upper bound.

Fig. 3 shows the optimal source selection probability distribution $p(s)$, i.e., the probability of any packet to choose a node s as its source, where the rank in x-axis means the number of nodes that are selected as source nodes. The $p(s)$ is plotted in the descend order. We can see that the uniform strategy is optimal for the lattice. In the RG networks, every node has a positive $p(s)$, but the values are different. In the ER and BA networks, only a small portion of the nodes have non-zero probabilities. **These results indicate that in homogeneous network such as lattice and RG, the sources (or servers) should distribute uniformly like the P2P network system. However, in heteroge-**

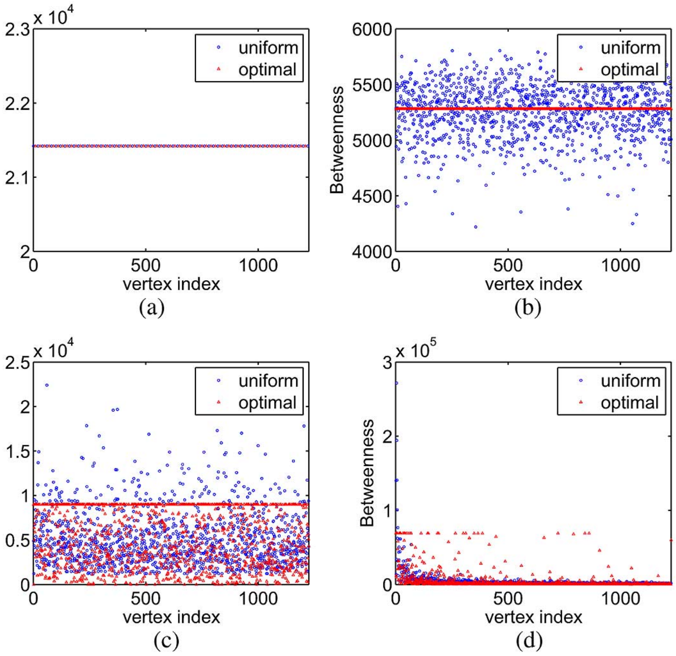

Fig. 2. Distribution of node betweennesses before (blue dots) and after (red dots) optimization for (a) the lattice and (b) random regular (RG) (c) the random Erdős-Rényi (ER) and (d) Barabási-Albert (BA) networks.

neous network such as ER and BA, only a few nodes should act as sources, which means in these networks, the centralized client-server architecture can perform better than the decentralized P2P-like architecture if the servers are distributed properly. In practical scenario, e.g., for content-centric-networks, the content cache should not be distributed uniformly. Instead, only a few nodes should act as content servers and the optimal proportion of the source nodes can be found in this letter. Moreover, the optimal probability distribution is not directly related with the degree of each node, but determined by the source dependent betweenness matrix $\mathbf{A}$ defined in (7) and the optimization process. Nodes with high degrees do not necessarily have the high probabilities to be as sources, as shown in Fig. 3.

As shown above, only a few nodes should act as sources in heterogeneous networks like the ER and BA networks. What is the optimal proportion of source nodes in the network? Fig. 4 shows the optimal proportion of nodes with non-zero source probabilities, i.e., $n_{p(s)>0} = N_{p(s)>0}/N$, versus network size N. We can see that the optimal proportion decreases with the increase of the network size N and their relation is approximately power law. Moreover, Fig. 5 shows the average path length of packets using the two strategies. The average path length $\overline{L}$ is the average shortest path length of all the source-destination pairs, which can be calculated by

$$\overline{L} = \sum\nolimits_{s,t} p(s)p(t)L_{s,t} = \frac{1}{N-1}\sum\nolimits_{s,t} p(s)L_{s,t}. \qquad (17)$$

The average path length $\overline{L}$ follows the small world property, i.e., $\overline{L} \sim lnN$. From Fig. 5, we can see that $\overline{L}$ in the optimal strategy are slightly larger than that in the uniform strategy for both ER and BA networks. This is because through optimal source selection, the packets bypass the nodes with high betweenness.

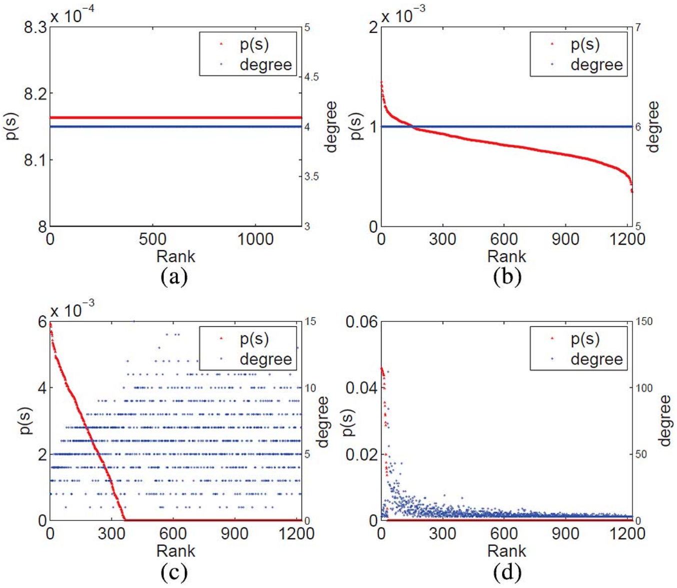

Fig. 3. The optimal source selection probability $p(s)$ distribution in the four networks, where $p(s)$ means the probability of any packet to choose s as the source node. We plot $p(s)$ in the descend order, and also plot the corresponding degree of each node (a) Lattice (b) RG (c) ER (d) BA.

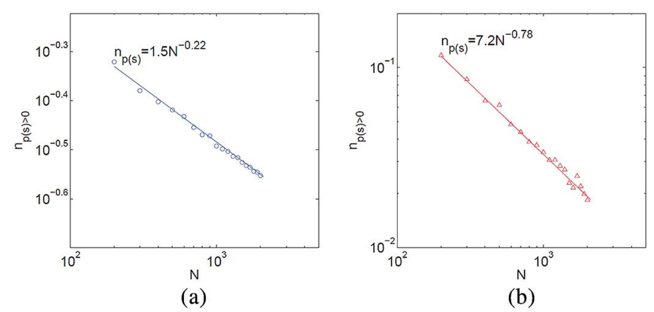

Fig. 4. The optimal proportion of nodes with nonzero source probabilities to all the nodes versus network size N for the random Erdős-Rényi (ER) and Barabási-Albert (BA) networks. $N = 200 - 2000$.

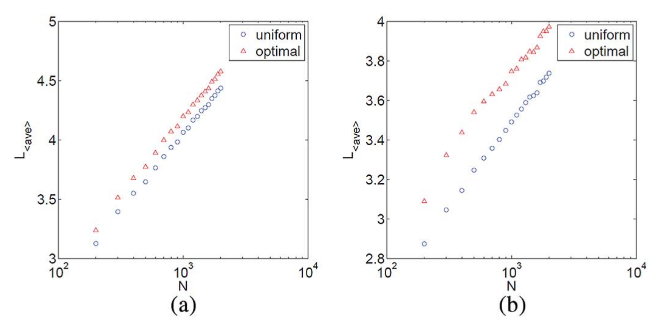

Fig. 5. The average path length $\overline{L}$ versus network size N using uniform and optimal strategies in the ER and BA networks. $N = 200 - 2000$.

IV. Conclusion

In this letter, we studied the optimal source selection strategy to improve the network capacity. The optimal selection probability distribution was found through solving a min-max optimization problem. Simulation results on four different types of network showed that in heterogeneous networks, the source nodes should be distributed within a small number of the nodes. When applied to the Internet information sharing service, as the Internet is a gigantic heterogeneous network, our work suggests that the client-sever structure with properly selected servers can improve the network capacity.

REFERENCES

[1] S. Chen, W. Huang, C. Cattani, and G. Altieri, "Traffic dynamics on complex networks: A survey," *Math. Probl. Eng.*, vol. 2012, 2011.

[2] L. Kocarev and V. In, "Network science: A new paradigm shift," *IEEE Network*, vol. 24, no. 6, pp. 6–9, 2010.

[3] T. Ohira and R. Sawatari, "Phase transition in a computer network traffic model," *Phy. Rev. E*, vol. 58, no. 1, p. 193, 1998.

[4] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, "Onset of traffic congestion in complex networks," *Phys. Rev. E*, vol. 71, no. 2, p. 026125, 2005.

[5] G.-Q. Zhang, S. Zhou, D. Wang, G. Yan, and G.-Q. Zhang, "Enhancing network transmission capacity by efficiently allocating node capability," *Phys. A: Statist. Mech. Applicat.*, vol. 390, no. 2, pp. 387–391, 2011.

[6] W.-B. Du, Z.-X. Wu, and K.-Q. Cai, "Effective usage of shortest paths promotes transportation efficiency on scale-free networks," *Phys. A: Statist. Mech. Applicat.*, vol. 392, no. 17, pp. 3505–3512, 2013.

[7] Z.-Y. Jiang, M.-G. Liang, and W.-J. An, "Effects of efficient edge rewiring strategies on network transport efficiency," *Phys. A: Statist. Mech. Applicat.*, vol. 394, pp. 379–385, 2014.

[8] X. Ling, M.-B. Hu, R. Jiang, and Q.-S. Wu, "Global dynamic routing for scale-free networks," *Phys. Rev. E*, vol. 81, no. 1, p. 016113, 2010.

[9] Z.-Y. Jiang and M.-G. Liang, "Incremental routing strategy on scale-free networks," *Phys. A: Statist. Mech. Applicat.*, vol. 392, no. 8, pp. 1894–1901, 2013.

[10] H.-X. Yang and M. Tang, "Adaptive routing strategy on networks of mobile nodes," *Phys. A: Statist. Mech. Applicat.*, vol. 402, pp. 1–7, 2014.

[11] D. Clark, "The design philosophy of the DARPA internet protocols," *Comput. Commun. Rev.*, vol. 18, pp. 106–114, 1988.

[12] A. Passarella, "A survey on content-centric technologies for the current internet: CDN and P2P solutions," *Comput. Commun.*, vol. 35, no. 1, pp. 1–32, 2012.

[13] A. Tizghadam and A. Leon-Garcia, "Betweenness centrality and resistance distance in communication networks," *IEEE Network*, vol. 24, no. 6, pp. 10–16, 2010.

[14] P. Erdős and A. Rényi, "On random graphs," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.

[15] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[16] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.