■ 6224

An Optimal Routing Strategy Based on Specifying Shortest Path

Yonghua Xu[*1], Fei Shao[1,2]
[1]School of Information Technology, Jinling Institute of Technology,Nanjing, China
[2] Jiangsu Information Analysis Engineering Laboratory, Jinling Institute of Technology, Nanjing, China
*Corresponding author, e-mail: xyh@jit.edu.cn

Abstract

Unlike the shortest path is randomly chosen in the traditional shortest path routing strategy, a novel routing strategy to improve the network transportation capacity is proposed in this paper. According to the different characteristics of the nodes along actual paths, we specify the shortest paths of all pairs of nodes aiming at reducing the betweenness of those high-betweenness nodes. Simulations on both computer-generated and real-world networks show that the new routing strategy can enhance the network transportation capacity greatly. And it works better in those networks with the fuzzy community structure.

Keywords: *optimal routing strategy, network transportation capacity, community structure, complex networks*

1. Introduction

Due to the seminal work on the small-world phenomenon by Watts and Strogatz [1] and the scale-free property by Barabási and Albert [2], many people take tremendous interests in the structure and dynamics of complex networks. It has been widely proved that the topological structure have profound effects on the processes taking place on the networks. Specifically in recent years, the study of information traffic on complex networks is becoming more and more important, as a result of the constantly increasing importance of large communication networks such as the Internet and the World Wide Web. A particular focus of these studies is to make the network transportation capacity maximal so as to control the increasing traffic congestion.

In previous works, there are two ways to improve the networks transportation capacity: making appropriate adjustments to the network topology structure [3-4] or designing efficient routing strategies [5-9]. Since it is too expensive and too difficult to change the structure of some large-scale networks, most of previous works have been committed to the development of effective routing strategies. Some of them are based on the global information: the shortest path routing strategy that pass through the minimum number of nodes [5], the efficient path routing strategy whose degree sum of nodes is a minimum [6], and the global dynamic routing strategy based on the queue length of nodes [7]. Some others focus on local topological information since global information is usually unavailable in large-scale networks: neighbor information [8], next-nearest-neighbor information [9].

The shortest path routing strategy, as its name implies, has the shortest average path length which means a packet may reach its destination quicker than taking other paths. However, it is proved that shortest path routing strategy is not optimal for scale-free networks because congestion will occur in some nodes with higher degree [10]. As there may be more than one shortest path between some pairs of nodes and one of them will be selected randomly in the traditional routing strategy, we can specify the shortest path to enhance the network transportation capacity.

With the deepening of research of complex networks, the property of community structure appears to be common in lots of real networks [11-12], which means the tendency for nodes to divide into subsets within which node-node connections are dense but between which connections are sparser. Several real networks have more or less community structure which will impact the network transportation capacity [13]. The internal structure of networks will have influence on our new routing strategy.

2. Models

The traffic model can be described as follows:

1) All nodes can create packets with addresses of destination, receive packets from other nodes, and forward the packets to their destinations.

2) At each time step, there are R packets generated in the network, with randomly chosen sources and destinations. Once a packet is created, it is placed at the end of the queue if the node already has several packets waiting to be forwarded to their destinations.

3) At each time step, the first C_i packets at the head of the queue of each node are forwarded one step toward their destinations and then placed at the end of the queues of the selected nodes. We assume every node has the same packet delivery capability C_i and set C_i=1 for simplicity.

4) A packet, upon reaching its destination, is removed from the system.

As R is increased from zero, the numbers of created and forwarded packets are balanced, resulting in a steady free flow of traffic. Since packet delivery capacity of node is limited, traffic congestion will occur when R is large enough. The phase transition from the former to the latter is defined as the critical value R_c which is the criterion of the network transportation capacity. We are interested in resolving critical value R_c in order to address which kind of routing strategy is more susceptible to traffic congestion.

We introduce the algorithmic betweenness b_i to estimate the possible traffic passing through a node i under a given routing strategy which is defined as below:

$$b_i = \sum_{s,t} \frac{\sigma(s,i,t)}{\sigma(s,t)} \qquad (1)$$

where $\sigma(s,i,t)$ is the number of paths under the given routing strategy between nodes s and t that pass through node i and $\sigma(s,t)$ is the total number of paths under the given routing strategy between s and t and the sum is over all pairs s, t of all distinct nodes.

The probability a packet will pass through the node i is $b_i / \sum_{j=1}^{n} b_j$, and therefore the average number of packets that the node i will receive at each time step is $Rb_i / (n(n-1))$ where n is the nodes number of the whole network. When the number of incoming packets is equal to or larger than the outgoing packets at the node i, $Rb_i / (n(n-1)) \geq C_i = 1$, traffic congestion will occur. So the critical packet-generating rate R_c is

$$R_c = \min(n(n-1)/b_i) \qquad (2)$$

The most widely used algorithm betweenness is the shortest path betweenness[14], sb_i, base on the traditional shortest path routing strategy. Newman suggested another betweenness measure by the name random walk betweenness[15], rb_i, which is equal to the number of times that a random walk starting at s and ending at t passes through i along the way, averaged over all s and t. The random walk betweenness rb_i of node i is calculated as follows:

1) Construct the matrix $\mathbf{D-A}$, where $\mathbf{D}$ is the diagonal matrix with elements $\mathbf{D}_{ii}=k_i$ and $\mathbf{A}$ is the adjacency matrix (if there is an edge between i and j, $\mathbf{A}_{ij}$=1; otherwise, $\mathbf{A}_{ij}$=0).

2) Remove any single row and the corresponding column.

3) Invert the resulting matrix and then add back in a new row and column consisting of all zeros in the position where the row and column were previously removed. Call the resulting matrix $\mathbf{T}$, with elements $\mathbf{T}_{ij}$.

4) The random walk betweenness of node i is defined as

$$rb_i = \frac{\sum_{s<t} I_i(st)}{n(n-1)/2} \qquad (3)$$

$$I_i(st) = \begin{cases} \frac{1}{2} \sum_j A_{ij} \left| T_{is} - T_{it} - T_{js} + T_{jt} \right| & i \neq s,t \\ 1 & otherwise \end{cases} \qquad (4)$$

Our new routing strategy is described as follow:

1) Calculate the degree, the shortest path betweenness, and the random walk betweenness of every node as k_i, sb_i, rb_i accordingly.

2) Obtain all shortest paths between nodes s and t using breadth first search algorithm.

3) Calculate the sum of degrees k_i of all nodes on each shortest path, and select the one with minimum sum as our route path. We call this SK routing strategy, and SS routing strategy for the minimum sum of sb_i, SR routing strategy for the minimum sum of rb_i

4) Obtain route paths for all node pairs.

To observe the influence of network internal community structure on our strategies, we employ the modularity measure, Q, which is defined as follows[11]: consider a division of a network into *mods* communities, define an *mods×mods* symmetric matrix **E** whose element e_{ij} is the fraction of all edges that link nodes in community i to nodes in community j. So we get

$$Q = \sum_i (e_{ii} - a_i^2) \tag{5}$$

where we have $a_i = \sum_j e_{ij}$. Different division will result in different Q, the maximum of them, Q_{max}, can describe the internal community structure the network has.

3. Simulations and Analysis

To obtain the critical packet generating rate R_c in simulations, we use the order parameter [3]:

$$\eta = \lim_{t \to \infty} \frac{\langle \Delta\Theta \rangle}{R\Delta t} \tag{6}$$

where $\Delta\Theta = \Theta(t+\Delta t) - \Theta(t)$, with $\langle \cdots \rangle$ indicating average over time windows of width Δt, and $\Theta(t)$ is the total number of packets in the network at time t. At the early stage, when R is very small, the generated packets can be delivered, $<\Delta\Theta>$ is less than zero and so is η. Where η is greater than zero, we can obtain the critical packet generating rate R_c.

Firstly, we investigate the efficiency of our routing strategies in homogeneous networks. We generate a series of pseudo-random networks with n=128 nodes which are divided into *mods*=4 communities with *n/mods*=32 nodes in each community. Each node has on average Z_{in} edges connecting it to nodes of the same community and Z_{out} edges to nodes of other communities. While Z_{in} is varied, the value of Z_{out} is chosen to keep the total average degree $<k>$=16. We increase Z_{in} from 8 to 15 edges to generate the network, and use DA community detection algorithm [16] to gain the maximum modularity Q_{max} as shown in Figure 1.

As Figure 1 shown, when Z_{in} is 8, Z_{out}=8=Z_{in}, the network is random network with Q_{max}=0.2162. While Z_{in} is increasing, there are more edges connecting nodes of the same community which result in the stronger community structure and the higher Q_{max}. When Z_{in} is 15, Q_{max} reaches 0.6771.

The relationship between critical packet generating rates of three routing strategies and Z_{in} is shown in Figure 2. From Figure 2 we can see that all the three routing strategies can enhance the network transportation capacity and the routing strategy by specifying the shortest path with the minimum sum of the node shortest path betweenness is the most remarkable one. Figure 1 and Figure 2 also show that the stronger community structure the network has the more ineffective our strategies are.

As we all known, many real networks include the Internet display a heterogeneous structure with a scale-free degree distribution. We use BA model to validate our strategies in heterogeneous networks. We generate a series of scale-free networks using BA model with BA parameter m sets to 2 and also use DA algorithm to get the highest modularity Q_{max} as shown in Figure 3. And the relationship R_c and Z_{in} is shown in Figure 4.
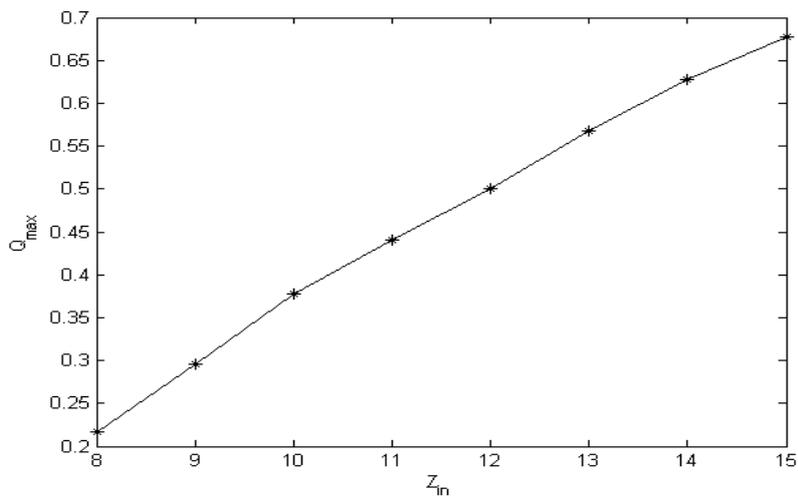
Figure 1. Maximum modularity Q_{max} versus Z_{in}
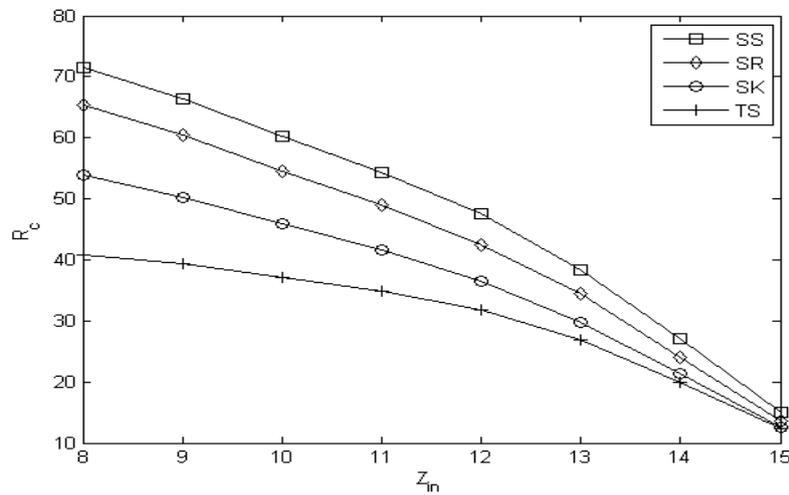
Figure 2. R_c versus Z_{in} for different routing strategies
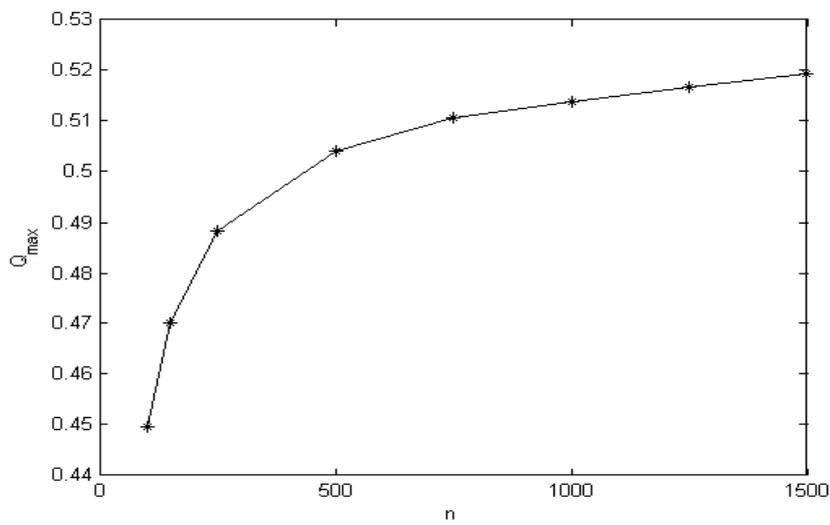
Figure 3. Maximum modularity Q_{max} versus nodes number n
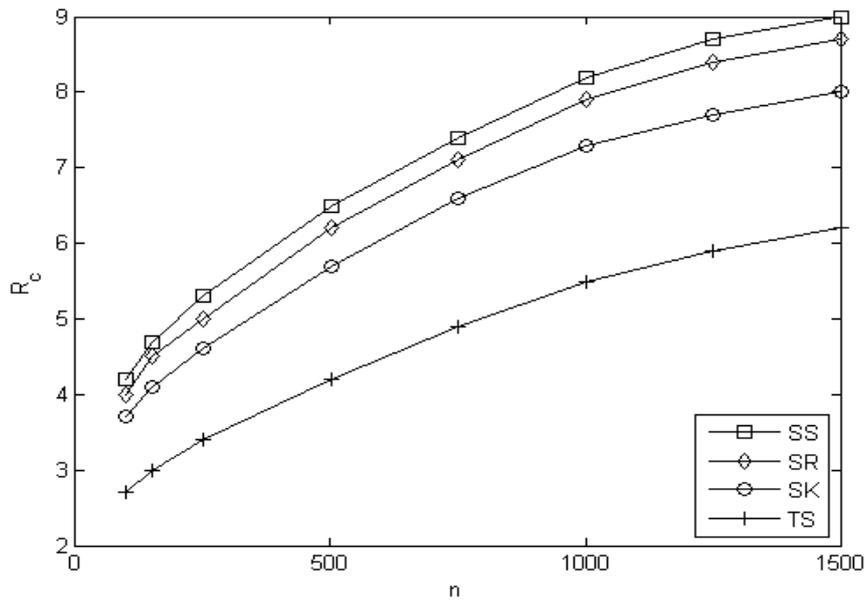
Figure 4. R_c versus nodes number n for different routing strategies

From Figure 3 and Figure 4, we can see that the SS strategy also gives the best results in heterogeneous networks. And three strategies work better in networks with fuzzy community structure which is consistent with analysis above.

Then we check the impact of BA parameter m on our routing strategies. We generate four BA networks with n=100 nodes and BA parameter m is 2, 3, 4, and 5. The maximum modularity Q_{max} is 0.4495, 0.3515, 0.2858, and 0.2592 correspondingly. The relationship between R_c and m is shown in Figure 5.
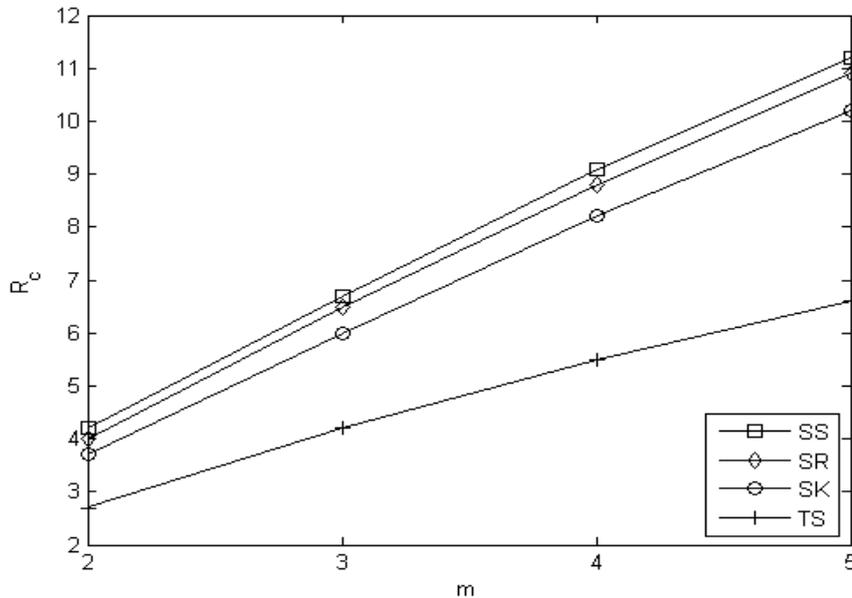
Fig. 5. R_c versus BA parameter m for different routing strategies

The results in Figure 5 also agree with our conclusion. The efficiency of our routing strategies decrease with the increasing of BA parameter m. When BA parameter m is 2, R_c is increase by 55.5% with SS routing strategy while nearly 69.7% when m is 5.

Figure 6. provides insight into how the routing strategy works by comparing the betweenness distributions with different routing strategies in the case of a BA network with 100 nodes and average node degree $<k>=4$. Figure 6.(a) shows the betweenness plotted against the node index while Figure 6.(b) shows histograms of the betweenness distribution. In TS routing strategy, the majority of the nodes have very low betweenness, but a small number of them have very higher betweenness. In SS routing strategy, node betweenness are confined to a narrow band.
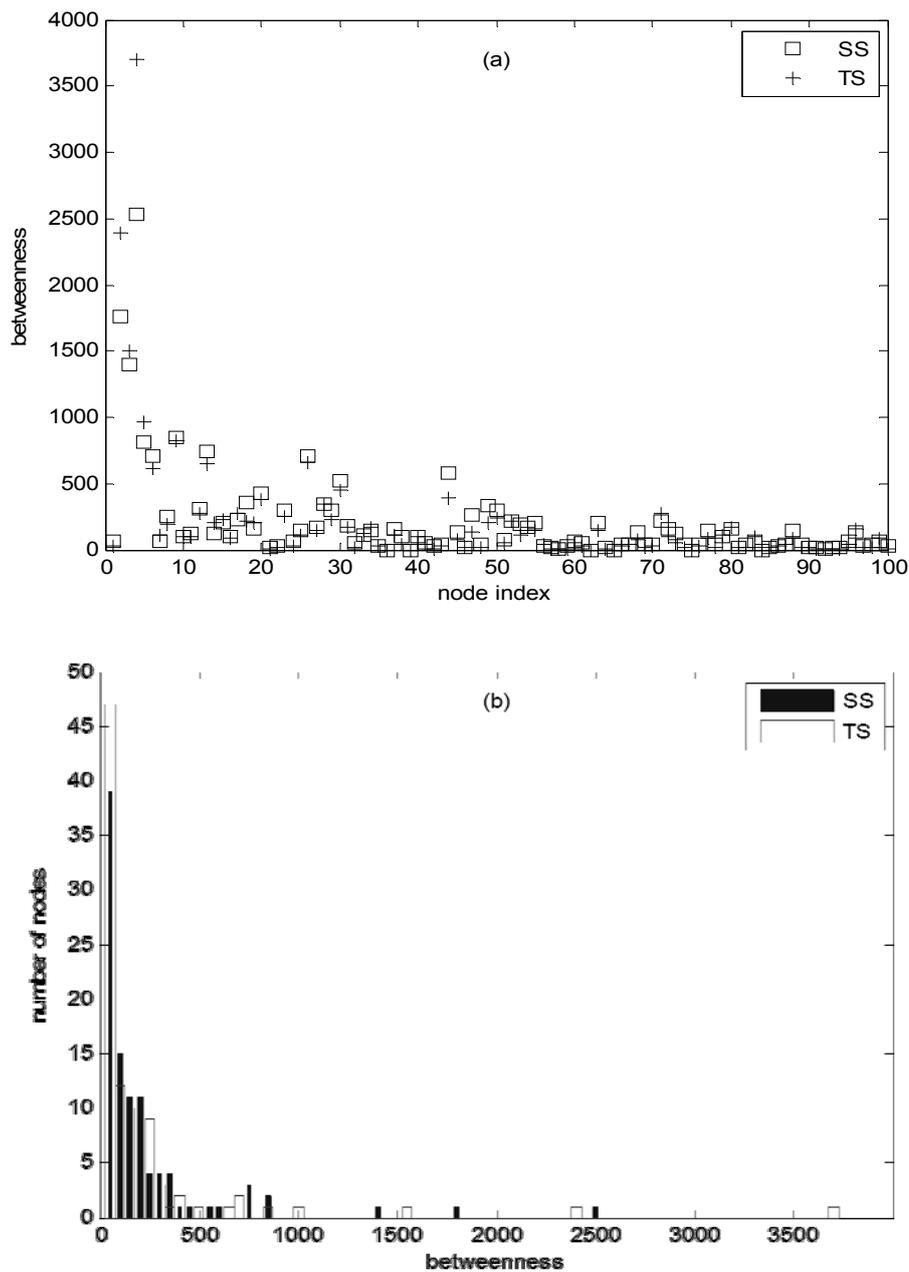
Fig. 6. Distribution of node betweenness with two different routing stragtegies
(a) betweenness of every node (b) distribution of node betweenness

Finally, we validate our routing strategies on the real world network. We employ the jazz network [17] with 198 nodes whose Q_{max} is 0.4452 and the email network [18] with 1133 nodes whose Q_{max} is 0.5738. The critical packet generating rates of three routing strategies are shown in Table 1.

Table 1. R_c of real world networks for different routing strategies

	SS	SR	SK	TS
Jazz	18.9	17.5	14.7	6.8
Email	37.8	36.1	31.2	26.2

As Table 1 shown, in real networks, the SS strategy can also enhance network transportation capacity more greatly than others, especially in networks with fuzzy community structure.

4. Conclusion

According to the different characteristics of the nodes along actual shortest path, we propose three routing strategies to enhance the network transportation capacity. The one by specifying the shortest path with the minimum sum of the node shortest path betweenness is proved to be the most remarkable one. All of our strategies are less effective in networks with stronger community structure.

Acknowledgments

References
[1]   D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*. 1998; 393(6684): 440-442.
[2]   A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*. 1999; 286(5439): 509-512.
[3]   A. Arenas, A. Díaz-Guilera, and R. Guimerà. Communication in networks with hierarchical branching. *Physical review letters*. 2001; 86(14): 3196-3199.
[4]   R. Guimerà, et al. Optimal network topologies for local search with congestion. *Physical review letters*. 2002; 89(24): 248701.
[5]   T. Zhou. Mixing navigation on networks. *Physica A: Statistical Mechanics and its Applications*. 2008; 387(12): 3025-3032.
[6]   G. Yan, et al. Efficient routing on complex networks. *Physical Review E*. 2006; 73(4): 046108.
[7]   X. Ling, et al. Global dynamic routing for scale-free networks. *Physical Review E*. 2010; 81(1): 016113.
[8]   W.X. Wang, et al. Traffic dynamics based on local routing protocol on a scale-free network. *Physical Review E*. 2006; 73(2): 026111.
[9]   C.Y. Yin, et al. Traffic dynamics based on an efficient routing strategy on scale free networks. *The European Physical Journal B*. 2006; 49(2): 205-211.
[10] S. Sameet, et al. Structural bottlenecks for communication in networks. *Physical Review E*. 2007; 75(3): 036105.
[11] M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. *PHYSICAL REVIEW E*. 2004; 69(2): 026113.
[12] M.E.J. Newman. Communities, modules and large-scale structure in networks. *Nature physics*. 2012; 8(1): 25-31.
[13] L. Danon, A. Arenas, and A. Díaz-Guilera. Impact of community structure on information transfer. *Physical Review E*. 2008; 77(3): 036103.

[14] L.C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*. 1977; 40(1): 35-41.
[15] M.E.J. Newman. A measure of betweenness centrality based on random walks. *Social networks*. 2005; 7(1): 39–54.
[16] J. Duch and A. Arenas. Community detection in complex networks using extremal optimization. *Physical Review E*. 2005; 72(2): 027104.
[17] P.M. Gleiser and L. Danon. Community structure in jazz. *Advances in Complex Systems*. 2003; 6(4): 565-573.
[18] R. Guimerà, et al. Self-similar community structure in a network of human interactions. *Physical Review E*. 2003; 68(6): 065103.