Modeling and Algorithms on Releasing Range of Traffic Guidance Information

Lijun Pei
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
Email: 08120300@bjtu.edu.cn

Dewang Chen
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
Email: dwchen@bjtu.edu.cn

***Abstract*—The traffic guidance system is an important subsystem in the intelligent transportation system which can timely release alerts regarding traffic congestion, incidents and other information to road-users. However, it is very hard to decide the releasing range exactly where and for whom to release such information. With the correlation analysis of road traffic flow in time and space domain, the releasing range model for traffic guidance information is formulated based on combinatorial mathematics in this paper. As it is difficult to get the optimal solution in a limited time through analysis for the model, a simple greedy algorithm, an algorithm based on simulated annealing and an algorithm based on M-Chameleon clustering are presented to divide the road links into several groups. Some traffic flow data detected by remote microwave sensors in some road links in Beijing urban expressway are employed to compare the algorithms. The results show that the algorithm based on M-Chameleon achieved the best performance, and it can be used to improve the pertinence, validity and automation of releasing traffic guide information.**

***Index Terms*—traffic guidance, information releasing range, clustering, simulated annealing, M-Chameleon**

I. INTRODUCTION

Nowadays, due to the continuous increase in vehicle numbers, traffic congestion has become one of the most serious and costly problems in the world, especially in many big cities and metropolitan areas. Take Beijing as an example, there are 60 to 110 typically congested roads in rush hours [1]. ITS (Intelligent Transport System) has been proposed to solve today's traffic problems and leads to more convenient and comfortable transportation by applying electronics, information, control and communication technologies [2]. The traffic guidance system is one of the core components of ITS, which helps drivers in choosing better routes by supplying convenient information. Therefore, the guidance system is worth studying since it is able to more effectively release traffic information to the travelers in big cities.

There are many ways to release traffic guidance information, including radio, VMS (variable message signs), navigation systems, television, internet, telephone, and etc [3]. For the big cities, a difficult point in the study of traffic guidance systems lies on deciding the range of releasing the guidance information. Because the impact of traffic congestion and incidents is restricted in a certain area, there is no need to release the local traffic congestion and incidents information to the whole city. Deep research on releasing range model and algorithms for traffic guidance information can make the release of traffic information more effective and purposeful. However, in the literature, there is almost no quantitative study about the release range of traffic guidance information at present. After the correlation analysis of road traffic flow in time and space domain, a mathematical model of this problem is built by grouping the road links.

As far as our goals in this paper are concerned, some hierarchical clustering techniques in data mining can be applied in grouping the road links automatically. These techniques are aimed at discovering natural divisions of networks into groups, based on various metrics of similarity or strength of connection between vertices. In an agglomerative method, similarities are calculated by one method or another between vertex pairs, and edges are then added to an initially empty network starting with the vertex pairs with highest similarity.

After considering this problem as a hierarchical clustering one, an algorithm based on data mining is presented to group the road links to quantitatively decide the range of releasing the information in this paper [4]. Some traffic flow data detected by remote microwave sensors in some road links in Beijing urban expressways are employed to validate the algorithm and compare it with other heuristic algorithms.

doi:10.4304/jcp.5.12.1856-1861

II. MODELING

Data source: Traffic speed flow data of 30 road links in second ring urban expressway (Fig. 3) in Beijing that were detected by remote microwave sensors are employed to prove the algorithm. Data acquisition time is from 0:00 to 12:00 of July 16, 2008 and the sampling interval is two minutes.

Similarity matrix: Suppose there are n road links and every two of them have a correlation coefficient r_{ij}, then we get a matrix $R_O[n,n]$.

$$R_o(i,j)=\frac{\mathrm{cov}(f_i[t_1,t_2],f_j[t_1,t_2])}{\sqrt{\mathrm{cov}(f_i[t_1,t_2],f_i[t_1,t_2])\times\mathrm{cov}(f_j[t_1,t_2],f_j[t_1,t_2])}}\ (i=1:n,j=1:n) \quad (1)$$

where $f_i[t_1,t_2]$ is the speed sequence of road link i between time t_1 to t_2. As the range of correlation coefficient is [-1,1], it is usual to normalize all correlation coefficient values to lie between 0 and 1, by calculating

$$R(i,j)=\frac{R_o(i,j)+1}{2} \quad (2)$$

Moreover, the distance between two road links has some influence on the correlation coefficient. Clearly, the impact is greater between two closer road links. We define spatial interaction coefficient between road link i and j as d_{ij}. It can be measured by using the space distance or number of road links k between them. In another way, by means of fuzzy sets，Gaussian function fuzzy sets to describe the correlation between the road links in this paper, then we can get a $n\times n$ matrix $D[n,n]$.

$$d_{i,j}=e^{-\frac{k^2}{2*5^2}} \quad (3)$$

Then we multiply matrix $R[n,n]$ and $D[n,n]$ to get the correlation matrix $R^{'}[n,n]$ based on time and space. If α>1, then greater importance is given to the space correlation, and when α<1, it gives a greater importance on time correlation. In this paper, we assume α is 1.

$$R^{'}=R\times D^{\alpha} \quad (\alpha>0) \quad (4)$$

Ⅲ. ALGORITHMS

Suppose there are n road links, and we group them into m groups with k road links in every group (Suppose m can be divisible by k). Performance function is shown as (5). Then the total calculation number is C.

$$S=\sum_{i=1}^{m}s_i=\sum_{i=1}^{m}\sum_{j=1}^{k}\sum_{r=j+1}^{k}(R^{'}[j,r]+R^{'}[r,j]) \quad (5)$$

$$C=\frac{mk(k-1)}{2}*Z=\frac{mk(k-1)n!}{2(k!)^m} \quad (6)$$

For example, if n=30, m=5, k=6, then C is 1.0282×10^{20}.C is too big that the method will consume a lot of time. In this paper, three algorithms are used to solve this problem.

Considering road links as a network, and every road link as a node (vertex), so there is a correlation coefficient between every two roads and what we need to do is to divide them into several groups (clusters).

A. Greedy algorithm

The greedy algorithm is a non-pursuit of the optimal solution, focusing only on a more satisfactory solution. Generally, it can find a better solution quickly. The greedy algorithm often starts from local optimization [5]. The steps are as follows:

Step 1: Randomly select the initial road i_1, and then find road i_2 which is most relevant to i_1.

Step 2: Find road i_3 which is most relevant to road i_1 and road i_2, and so on, until road i_k.

Step 3: Repeat step 1 and step 2 until finishing m groups.

B. Simulated Annealing algorithm

The simulated annealing algorithm is a random search method that simulates the annealing process of metal heuristic. It has been used in a variety of combinatorial optimization problems including the traveling salesman problem [6]. The steps are as follows:

Step 1: First randomly generate an initial solution and calculate the performance function value.

Step 2: Create a new solution based on the initial solution, the method that create the new solution is reversing the order of the data which are randomly selected. Calculate the performance function value and compare with the value obtained from step 1.

Step 3: If the performance function value is reduced, then receive the new solution; otherwise receive the new solution under a certain degree of probability based on the annealing temperature.

Step 4: Repeat steps 1 to 3 until the termination conditions are met.

C. M-Chameleon algorithm

We use the Chameleon algorithm from Long et al. [7]. It uses Modularity Q [8] and structural equivalence similarity degree S [7] to solve the problem.

Modularity: A measure of the quality of a particular division of a network. This measure is based on a previous measure of assortative mixing proposed by Newman [9]. Consider a particular division of a network into k clusters. Let us define a k×k symmetric matrix e whose element e_{ij} is the fraction of all edges in the network that link vertices in cluster i to vertices in cluster j. Here we consider all edges in the original network—even after edges have been removed by the community structure algorithm our modularity measure is calculated using the full network.

The trace of this matrix $Tr\,\mathrm{e} = \sum_{i} e_{ii}$ gives the fraction of edges in the network that connect vertices in the same cluster, and clearly a good division into clusters should have a high value of this trace. The trace on its own, however, is not a good indicator of the quality of the division since, for example, placing all vertices in a single cluster would give the maximal value of $Tr\,\mathrm{e} = 1$ while giving no information about cluster structure at all.

We further define the row (or column) sums $a_i = \sum_j e_{ij}$, which represent the fraction of edges that connect to vertices in cluster i. In a network in which edges fall between vertices without regarding for the clusters they belong to, we would have $e_{ij} = a_i a_j$. Thus we can define a modularity measure by

$$Q = \sum_i (e_{ii} - a_i^2) = Tr\,\mathrm{e} - \left\| \mathrm{e}^2 \right\| \quad (7)$$

where $\|X\|$ indicates the sum of the elements of the matrix X. This quantity measures the fraction of the edges in the network that connect vertices of the same type (i.e., within-cluster edges) minus the expected value of the same quantity in a network with the same cluster divisions but random connections between the vertices. If the number of within-cluster edges is no better than random, we will get Q = 0. Values approaching Q = 1, which is the maximum, indicate strong cluster structure.

Structural equivalence similarity degree: If two nodes have the same adjacent nodes, we call them structural equivalence. Therefore, we define structural equivalence similarity degree S measure by

$$S_{i,j} = \sqrt{\sum_{k \neq i,j} (W_{ik} - W_{jk})^2} \quad (8)$$

where W_{ij} is the edge weight of i and j, in this paper is also the correlation coefficient between node i and node j. For structural equivalence nodes, the structural equivalence similarity degree is 0; for two nodes that have no common adjacent nodes, the structural equivalence similarity degree is maximum.

For two clusters, we define structural equivalence similarity degree as follows:

$$S_{r,h} = \frac{1}{n_r n_h} \sum_{i=1}^{n_r} \sum_{j=1}^{n_h} S_{x_{ri}, x_{hj}} \quad (9)$$

where n_r is the number of nodes in cluster r, and n_h is the number of nodes in cluster h.

The steps of M-Chameleon algorithm are as follows:

Step 1: For network G, initialize the clusters n as beginning, that is every vertex is a sub-cluster.

Step 2: Calculate $S_{i,j} = \sqrt{\sum_{k \neq i,j} (W_{ik} - W_{jk})^2}$ or $S_{r,h} = \frac{1}{n_r n_h} \sum_{i=1}^{n_r} \sum_{j=1}^{n_h} S_{x_{ri}, x_{hj}}$, select two sub-clusters that have the minimum structural equivalence similarity degree and then merge them.

Step 3: Calculate modularity Q.

Step 4: Repeat steps 2 and 3 until modularity Q begin to decline, then stop and output the result. Or repeat steps 2 and 3 until all the vertices are in one cluster, then choose the best result which has the biggest modularity, stop and output the result.

Ⅳ. EXPERIMENT RESULTS

Assume that every road link as a vertex, then we can consider these road links compose a network which have n vertices and $R'[n,n]$ is the $n \times n$ similarity matrix that (i, j) element of the matrix represents the similarity between *ith* and *jth* road links. In M-Chameleon algorithm, $R'[n,n]$ can be seen as the edge weights. Then we use M-Chameleon to group these road links and the result is as follows, see Fig. 1.

From Fig. 1, we can see that the modularity is best when the number of cluster is 5, and the result is showed by Table Ⅰ and Fig. 2. In Fig. 2, same color and shape boxes represent the same group. The close road links have large correlation, comply with the common sense.

For comparison，we also use the greedy algorithm and simulated annealing algorithm to solve this problem. The number of groups wanted must be known before using these two algorithms.

Here we assume that the group number is 5 in order to compare these three algorithms. The results are shown in Table Ⅱ and Table Ⅲ. The comparison result is shown in Table Ⅳ.

From Table Ⅳ, we can see that greedy algorithm consume least time (1.748s), but also has the worst Modularity (0.6574). The simulated annealing algorithm has the best Modularity (0.8691), but consume a lot of time (25.019s). The M-Chameleon algorithm has the Modularity (0.8395) that smaller than simulated annealing algorithm, but it needs a short time (3.178s). In this paper, we only use thirty road links. We prefer to use the M-Chameleon algorithm when there are many road links needing to be grouped because it can save a lot of time and has higher degree of automation.

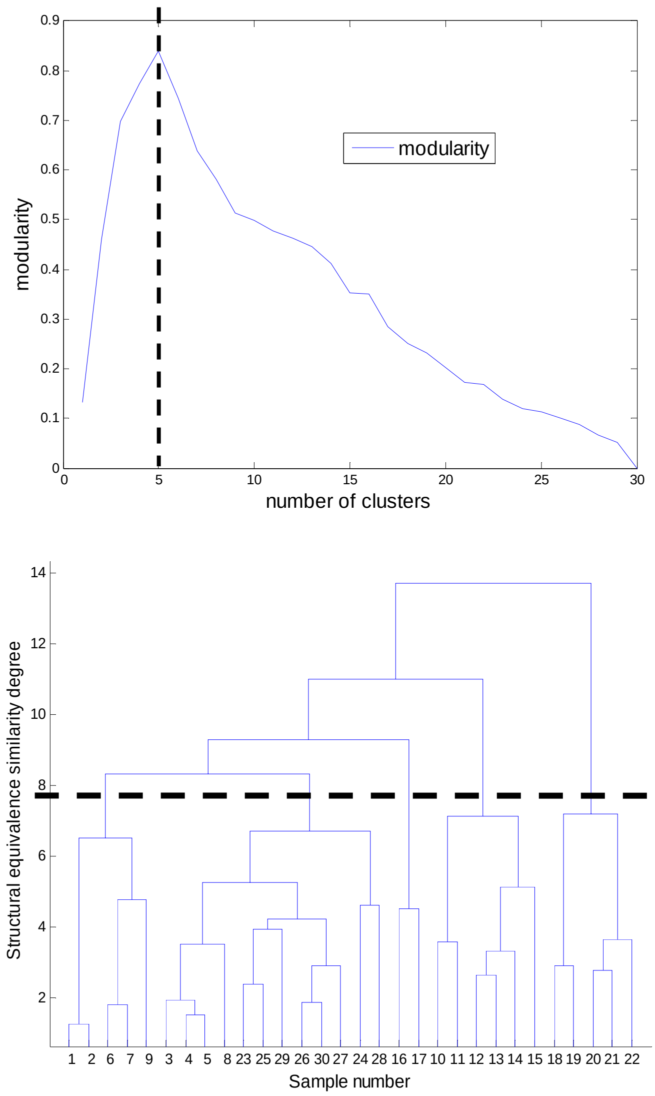

Figure 1. Plot of the modularity and dendrogram for the thirty road links.

TABLE Ⅰ.

RESULT FOR THE M-CHAMELEON ALGORITHM

1st group	02034	02035	02036	02040	02060	02061
	02062	02063	02064	02065	02066	02067
2nd group	02042	02043	02045	02046	02047	02049
3rd group	02052	02054	02055	02056	02057	
4th group	02032	02033	02037	02038	02041	
5th group	02050	02051				

TABLE Ⅱ.

RESULT FOR THE GREEDY ALGORITHM

1st group	02056	02057	02061	02063	02066	02067
2nd group	02032	02033	02037	02038	02042	02043
3rd group	02034	02035	02036	02040	02041	02046
4th group	02045	02047	02049	02052	02054	02055
5th group	02050	02051	02060	02062	02064	02065

TABLE Ⅲ.

RESULT FOR THE SIMULATED ANNEALING ALGORITHM

1st group	02042	02043	02045	02046	02047	02049
2nd group	02060	02062	02063	02065	02066	02067
3rd group	02052	02054	02055	02056	02057	05061
4th group	02032	02033	02037	02038	02041	02064
5th group	02034	02035	02036	02040	02050	02051

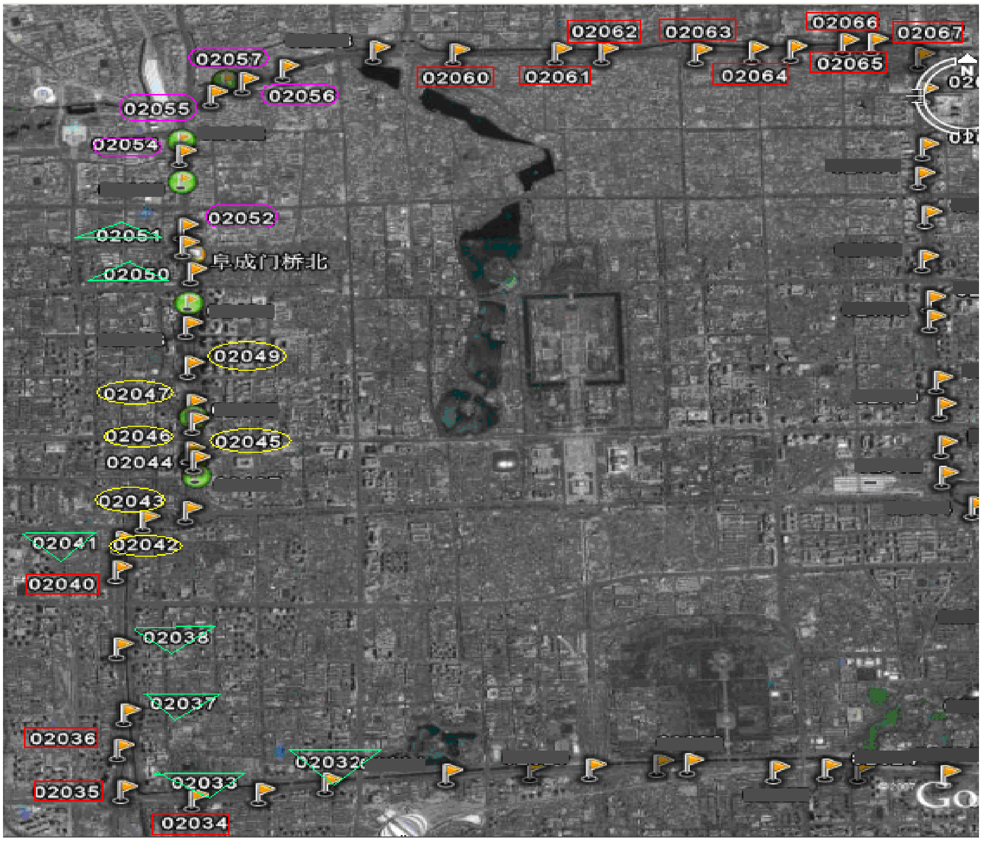

Figure 2. Schematic diagram of grouping result.

TABLE Ⅳ.

RESULTS OF THE COMPARISON OF THREE ALGORITHMS

Performance / Algorithm	Modularity	Time Consumption	Degree of Automation
Greedy	0.6574	1.748s	low
Simulated annealing	0.8691	25.019s	low
M-Chameleon	0.8395	3.178s	high

Algorithm validation: The following Fig. 3 is the speed flow chart of road links in the 2nd group and the 3rd group for M-Chameleon algorithm. From the figure, we can see that the result is good, the speed flow of each group has its own characteristics, and the difference between the groups is large, reflecting the relevance of the road links, which verifies the rationality of the algorithm.

Ⅴ. CONCLUSION AND FUTURE WORK

The paper develops a mathematical model for the releasing range of traffic guidance information and applies some quantitative calculation methods. It can provide a basis and reference for real-life traffic guidance information releasing. The result of validation using measured data shows that the model and the algorithm are useful. The M-Chameleon algorithm has two advantages: 1) it can obtain good results in short time. 2) it can find the best result automatically instead of setting up the initial conditions as in the greedy algorithm and simulated annealing algorithm.

Once traffic congestion or traffic incident has happened, the information should be released in the same group in order to improve the pertinence, validity and automation of releasing traffic information. Whether the influence between road links is symmetric, further study is needed. This paper only uses the data measured by microwave detector, Also worthy of study is how to use other type of traffic data or even data measured by other detectors. In addition, for algorithm research, the future task is mainly to find better algorithms to solve the problem.

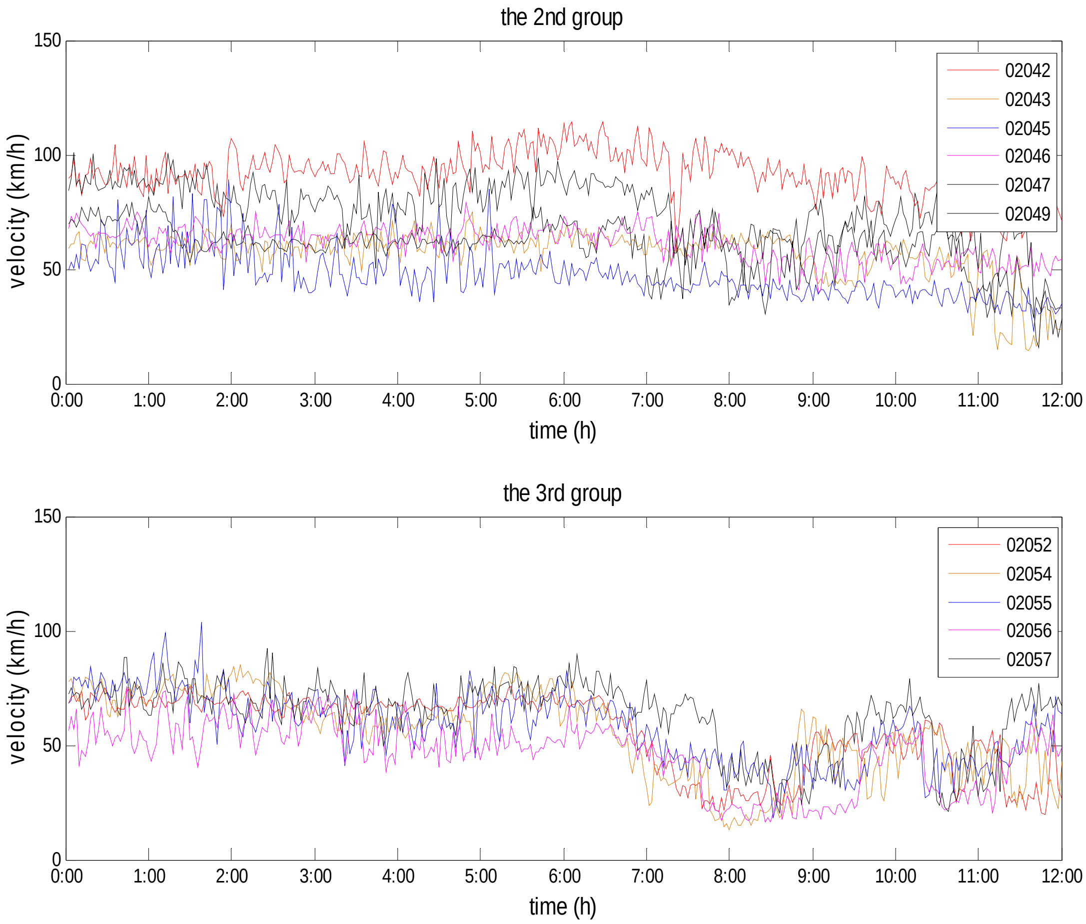

Figure 3. Speed flow chart of two groups for M-Chameleon algorithm.

ACKNOWLEGMENT

This work was partially supported by the National High Technology Research and Development Program (863 Program) (2007AA11Z212), the State Key Laboratory of Rail Traffic Control and Safety (Contract No.RCS2008ZZ001, No.RCS2009ZT004).

REFERENCES

[1] M. Zhao. "Urban Traffic Flow Guidance System based on data driven," *2009 International Conference on Measuring Technology and Mechatronics Automation*, Hunan, China, pp.653-657, 2009.

[2] T. Sugimoto. "Current Status of ITS and its International Cooperation," *Vehicle Electronics Conference,(IVEC '99) Proceedings of the IEEE International*, pp.462-467, 1999.

[3] X. Bao, X.L. Zhu, Y.Sh. Wang, J.Zh. Guan. "Research on Dynamic Real-time Traffic Routing Information Distribution based on VMS," Advanced Display, Vol.64, pp.34,63-65, 2006.

[4] G. Karypis, E. H. Han and V. Kumar. "Chameleon: Hierarchical Clustering Using Dynamic Modeling," Computer, Vol.32, No.8, pp. 68-75, 1999.

[5] Ch.Ch. Song, W. Wei. "Study of Multi-machine Scheduling Problem Based on Greedy Algorithm," Information Technology and Informatization, Vol.5, pp.45-46, 2009.

[6] Y. Liu, Sh.W. Xiong. "Hybrid tabu-simulated approach to solve traveling salesman problem," Computer Engineering and Applications,Vol.45, No.31,pp.43-45, 2009.

[7] Zh.Zh. Long, C. Zhang, F.Y. Liu and Zh.W. Zhang. "Improved Chameleon Algorithm," Computer Engineering, Vol.35, No.20, pp. 189-191. 2009.

[8] M. E. J. Newman and M. Girvan. "Finding and Evaluating Community Structure in Networks," Phys. Rev. E, Vol.69, No.2, pp. 113-127, 2004.

[9] M. E. J. Newman. "Mixing patterns in networks," Phys. Rev. E 67, 026126, 2003.