Generalized Farey Tree Network with Small-World

Jin-Qing Fang and Yong Li

China Institute of Atomic Energy, Beijing 102413
fjq96@126.com

Abstract. Generalized Farey tree network (GFTN) model with small-world is proposed, and the topological characteristics are studied by both theoretical analysis and numerical simulations, which are in good accordance with each other. Analytical results show that the degree distribution of the GFTN is exponential. As the number of network nodes increasing with time interval (or level number), *t*, the clustering coefficient of the networks tends to a constant, ln2; the diameter of the network is increasing with *t*, the resulting networks are evolved from disassortative to assortative and show assortative coefficient tends to 0.25 for large *t*.

Keywords: Generalized Farey tree network, topological properties, theoretical analysis, small world, numerical simulation.

1 Introduction

Due to diversity and complexity of real-world networks in the nature and human society, real-world network structures often are of rich and varied architecture. How to find a more suitable and another kind of complex network in nature and human society has been a challenging and significant issue in network science. According to our knowledge, generalized Farey tree and its pyramid have been applied to describe complexity of nonlinear complex systems [1-3] since 1980s. In previous study Kim and Ostlund constructed a Farey triangle to obtain rational approximants of pair of irrational number motivated by the interest in nonlinear systems. They have shown that frequency locking in maps can be organized by the Farey arithmetic that provides rational approximants of irrational numbers, and showed that frequency locking on a three torus can be organized by the generalized Farey arithmetic that gives rational approximants for pairs of mutually irrational numbers. Then Maselko and Swinney [2] found that the Farey triangle can provide a natural compact description of sequences of periodic state observed in their experiments on the Belousov-Zhabotinskii reaction. Fang [3] have also found that in nonlinear dynamical system mode locking and complicated multi-peaked periodic oscillations can be described by the generalized Farey tree, triangle and pyramid, which demonstrated that this classification provides a natural, compact and elegant means of organizing complexity and reveals the self-similar structure of the nonlinear dynamics. Calvo and his cooperators [4] have also investigated the hierarchical structure of three-frequency resonances in nonlinear dynamical systems with three interacting frequencies, in which they hypothesize an ordering of these resonances based on a generalization of the Farey tree organization

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1678 – 1689, 2009.

from two frequencies to three, and the experiments and numerical simulations demonstrated that their organization may describe the hierarchies of three-frequency resonances in representative dynamical systems, and may be universal across a large class of three-frequency systems.

Could we use generalized Farey organization to describe complex networks from a view point of network science [5-6]? That is our motivation for this paper. The answer is positive. We propose and construct another family of complex network: generalized Farey tree network (GFTN) which can depict and reveal complexity and universality of complex network systems. Main theoretical results of topological properties in the GFTN, including degree distribution, clustering coefficient, diameter and degree-degree correlation coefficient (assortative coefficient), are deduced theoretically. This work can be extended to generalized Farey organized network pyramid, which will be appeared elsewhere.

2 Construction Method of the Generalized Farey Tree Network (GFTN)

In the ordinary Farey analysis an infinite tree of rational numbers can be constructed from a pair of rational numbers: the Farey sum of the pair *p/q* and *p'/q'* is (*p*+*p'*)/(*q*+*q'*), which is the rational mediant between *p/q* and *p'/q'* with largest denominator. This Farey addition can be continued infinitely, yielding the Farey tree.

Constructional rule of the GFTN is as follows. (1) It is similar to ordinary Farey tree above, but in the GFTN three nodes of network start from the adjacent three values of (0/1, 1/1, 1/0), which represent first generation, also so-called first level. (2) First level (generation) is (0/1, 1/1), (1/1, 1/0) and (1/0, 0/1), respectively. Thus they format three families (branches). (3)Second level of the GFTN has (0/1+1/1=1/2), (1/1+1/0=2/1) and (1/0+0/1=1/1), the GFTN can be constructed in a recursive way. Fig.1 illustrates the diagram of the first three steps of the GFTN and Fig.2 gives the GFTN with the 6 levels which consists of the three branches (families).

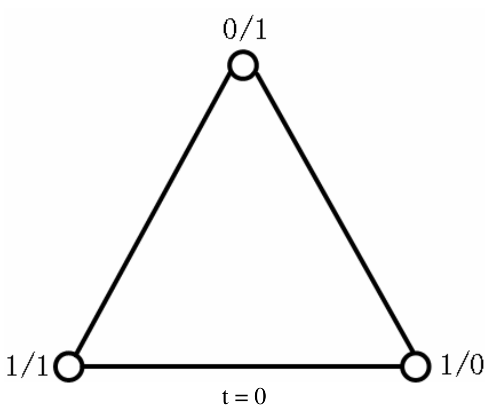


Fig. 1. The first three levels (steps) of the GFTN

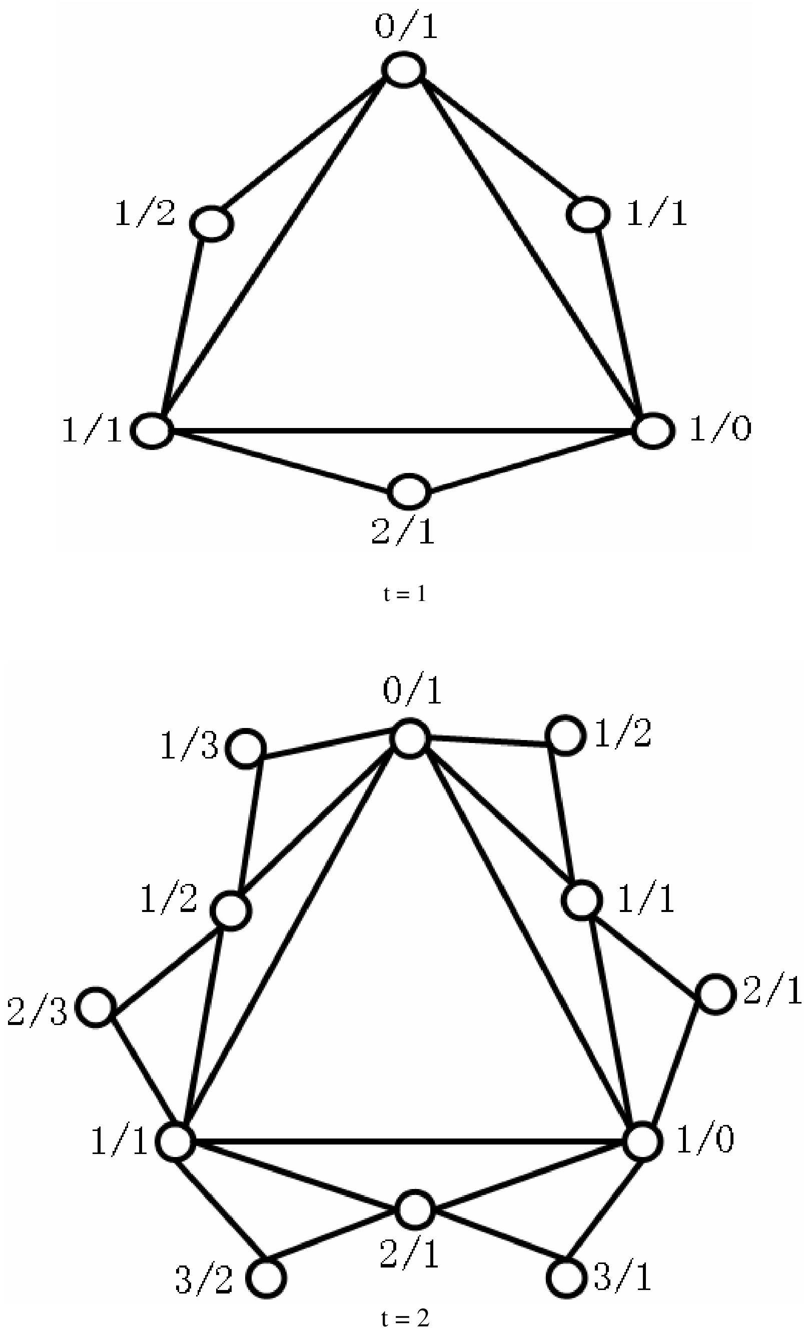
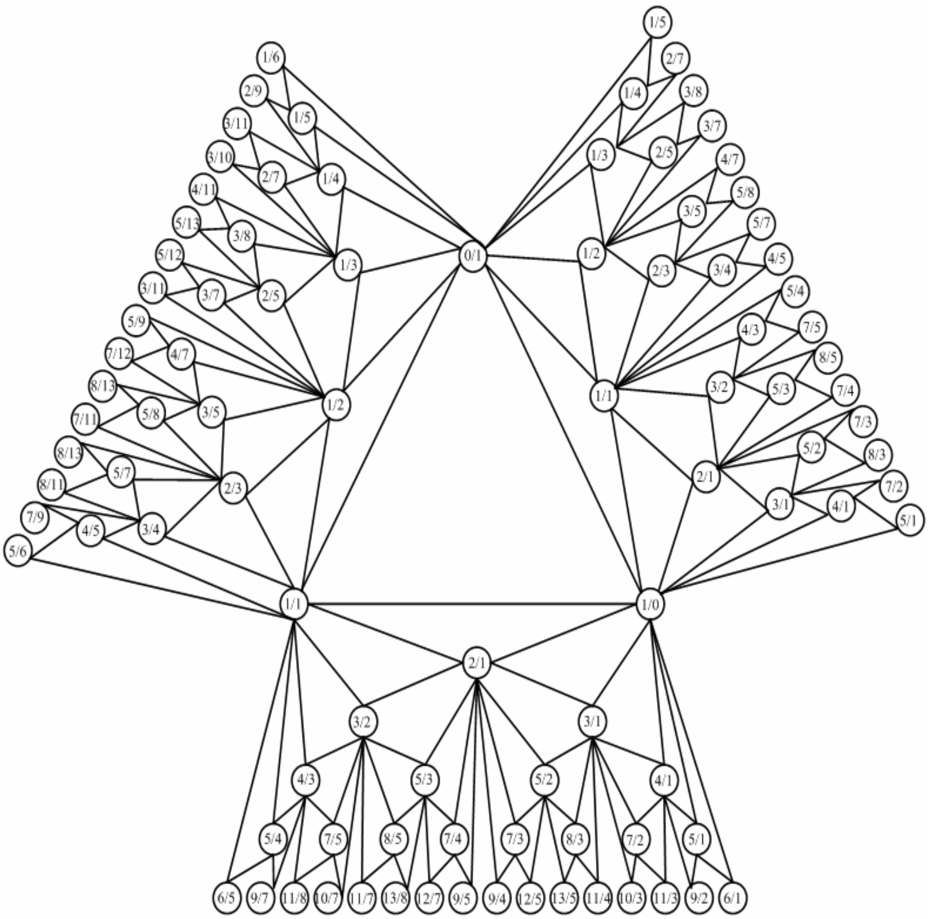


Fig. 1. (*continued*)



t = 5

Fig. 2. The GFTN with the 6 levels, which consists of huge three families

3 Topological Properties of the GFTN

In the following sections, we denote the network after t steps by $G_t, t \geq 0$, and the total numbers of nodes and edges in G_t are n_t and e_t respectively. For t=0, G_0 is a triangle composed by the first level's three nodes (0/1, 1/1, 1/0), which link to each other. When $t \geq 1$, the number of next level's nodes added to the network Δn_t is three times the number of new nodes appearing in a single Farey sequence. So we have

$$\Delta n_t = 3 \cdot 2^{t-1} . \quad (1)$$

Each new node has two edges that link the new node to two nodes. The p/q of each new node is taken in increasing level from the GFTN. Thus at time interval $t \geq 1$, according to the evolving rules, the numbers of total nodes and total edges in the GFTN are respectively

$$\begin{aligned} n_t &= n_{t-1} + \Delta n_t, n_0 = 3 \\ e_t &= e_{t-1} + 2(n_t - n_{t-1}), e_0 = 3 \end{aligned}. \tag{2}$$

Combining with Eq. (1), the solution of Eq. (2) is

$$\begin{aligned} n_t &= 3 \cdot 2^t \\ e_t &= -3 + 6 \cdot 2^t = 3 \cdot (2^{t+1} - 1) \end{aligned}. \tag{3}$$

Topology properties are of fundamental significance to understand the complex dynamics of real-world systems. Here we focus on four important characteristics: degree distribution, clustering coefficient, diameter, and degree-degree correlations.

3.1 Degree Distribution

Degree k_i of a node i is the number of edges connected to it. In general, the importance of a node is in proportion to its degree. From Eq. (3), we have the average number of degree of network at time interval t,

$$\bar{k} = \frac{2e_t}{n_t} = \frac{2(-3 + 6 \cdot 2^t)}{3 \cdot 2^t} \xrightarrow{t \to \infty} 4. \tag{4}$$

Degree distribution P_k, which is the probability that a node chosen uniformly at random has degree k, is one of the most important statistical characteristics of a network. Let $k_{i,t}$ is the degree of node i at step t, and t_j is the current step j at which node i is added to the network. Then by construction, initial degree of each new node at t_j, is 2, and the increment of old node at t_{j-1} is 2 for $j \geq 1$. It is not difficult to find following relation:

$$\begin{cases} k_{i,t} = k_{i,t-1} + 2, t \geq 1 \\ k_{i,0} = 2 \end{cases}. \tag{5}$$

The solution of Eq.(5) is

$$k_{i,j} = 2(j+1), j = 0, 1, \cdots, t. \tag{6}$$

Therefore, the degree in the GFTN at step t is discrete and the nodes have maximum degrees are the first three nodes in G_0. Combing with Eq.(6), we can give the relationship between k and N_k, which is the number of total nodes whose degree are all k in G_t of the GFTN, we have

$$N_k = N_{2j} = \begin{cases} 3 \cdot 2^{t-j}, j = 1, 2, \cdots, t \\ 3, j = t+1 \end{cases}. \tag{7}$$

Since the number of all nodes in network at step t is n_t, we can deduce the degree distribution P_{2j} according to the knowledge of classical probability.

$$P_{2j} = \frac{N_{2j}}{n_t} = \begin{cases} \dfrac{3\cdot 2^{t-j}}{n_t} = \dfrac{3\cdot 2^{t-j}}{3\cdot 2^t} = \dfrac{1}{2^j}, j = 1, 2, \cdots, t \\ \dfrac{3}{n_t} = \dfrac{3}{3\cdot 2^t} = \dfrac{1}{2^t}, j = t+1 \end{cases} . \tag{8}$$

Let k=2j, we have j=k/2, and $2^{-j} = 2^{-\frac{k}{2}}$. Through substitution of variable,

$$P_k = P_{2j} = \begin{Bmatrix} 2^{-\frac{k}{2}}, k = 2, 4, \ldots, 2t \\ 2^{-t}, k = 2(t+1) \end{Bmatrix} \xrightarrow{t->\infty} 2^{-\frac{k}{2}}, k = 2, 4, \ldots, 2t, 2(t+1). \tag{9}$$

From Eq. (9), degree distribution of the GFTN follows the form of exponential. Fig.3 shows the curve of degree distribution at step t=10 and t=100 and theoretical results fit well with numerical ones.

Theoretic curve is taken as

$$P_k = 2^{-\frac{k}{2}} . \tag{10}$$

It is seen from Fig.3 that the theoretical result of degree distribution is in agreement with numerical simulation.

3.2 Clustering Coefficient

While average clustering coefficient(C) is considered as another characteristic of a network, we also derive the analytical expression of the C. In a network the local clustering coefficient C_i of a node i is defined as Eq. (11) and C is the algebraic average of all C_i.

$$\begin{cases} C_i = \dfrac{2E_i}{k_i(k_i - 1)} \\ C = \dfrac{1}{n_t}\sum_{i\in G_t} C_i \end{cases} . \tag{11}$$

where E_i is the number of links presenting among its neighbors of the given node i in a network and k_i is the degree of node i, n_t is the total number of nodes in network.

By construction, at step 0, it is straightforward to calculate exactly the C_i and C=1. When a node i joins the network at step j, k_i and E_i are 2 and 1 respectively. After that, if the degree k_i of node i increases by one, its new neighbor must connect to one of its presenting neighbors. So E_i increases by one at the same time. All the edges among neighbors of node i can almost be a closed polygon except an edge is lacked. Anyway, for a node i which degree is k_i, there is only k_i-1 links among its neighbors.

$$E_i = k_i - 1. \tag{12}$$

So, combing with Eq. (3), (7), and (12), Eq. (11) can be rewritten as

$$C_i = \frac{2(k_i - 1)}{k_i(k_i - 1)} = \frac{2}{k_i}. \tag{13}$$

And for $t \geq 1$

$$\begin{aligned} C &= \frac{1}{n_t}\sum_{k_i} N_{k_i} C_i = \frac{1}{3\cdot 2^t}(3\cdot 2^{t-1}\frac{2}{2} + 3\cdot 2^{t-2}\frac{2}{4} + 3\cdot 2^{t-3}\frac{2}{6} + \cdots + 3\cdot 2^{t-t}\frac{2}{2t} + 3\cdot\frac{2}{2(t+1)}) \\ &= \frac{1}{2^t}(2^{t-1}\frac{1}{1} + 2^{t-2}\frac{1}{2} + 2^{t-3}\frac{1}{3} + \cdots + 2^{t-t}\frac{1}{t} + \frac{1}{t+1}) = \frac{1}{2^t}(\sum_{i=1}^{t}\frac{2^{t-i}}{i} + \frac{1}{t+1}) \end{aligned} \tag{14}$$

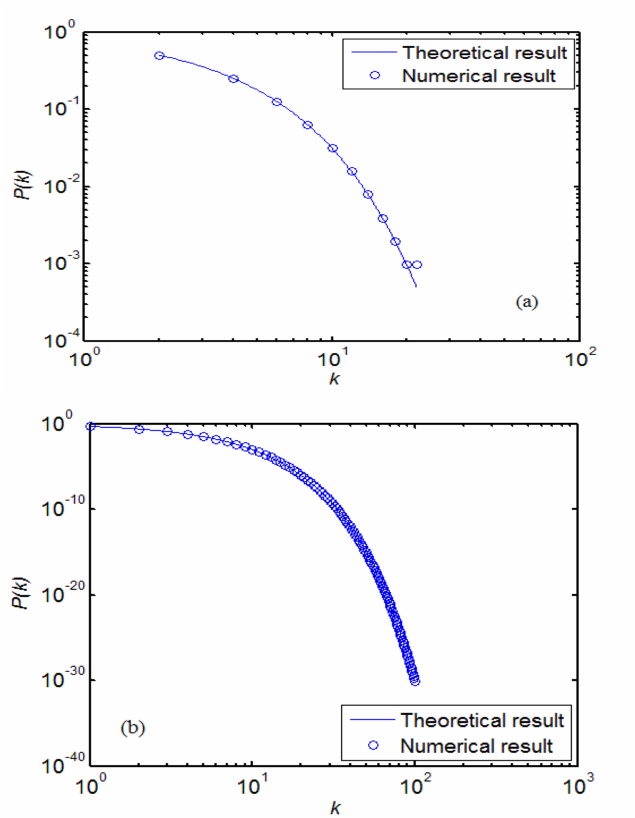


Fig. 3. Comparison of the degree distribution of network between theoretic result (circle) and numerical one (line). (a) t=10, (b) t=100.

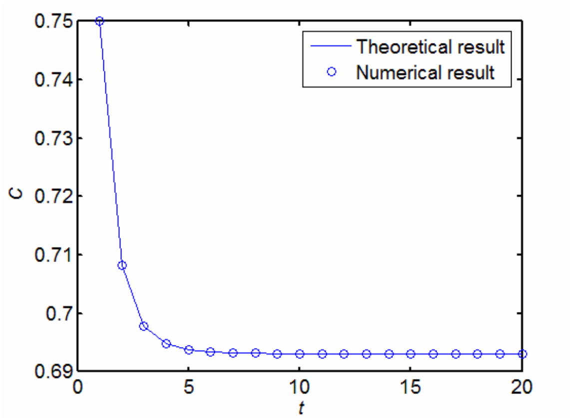


Fig. 4. Comparison of the cluster coefficient of the GFTN between theoretical result (circle) and numerical one (line) with the increasing of step t

In the limit of large step t,

$$\begin{aligned}\lim_{t\to\infty} C &= \lim_{t\to\infty}\frac{1}{2^t}\left(\sum\nolimits_{i=1}^{t}\frac{2^{t-i}}{i}+\frac{1}{t+1}\right)=\lim_{t\to\infty}\frac{1}{2^t}\sum\nolimits_{i=1}^{t}\frac{2^{t-i}}{i}+\lim_{t\to\infty}\frac{1}{2^t(t+1)} \\ &=\lim_{t\to\infty}\sum\nolimits_{i=1}^{t}\frac{1}{i2^i}+\lim_{t\to\infty}\frac{1}{2^t(t+1)}=\lim_{t\to\infty}\sum\nolimits_{i=1}^{t}\frac{1}{i2^i}\end{aligned} \tag{15}$$

According to the D'Alembert discriminance of convergence of positive term series, we have

$$\lim_{i\to\infty}\frac{\frac{1}{(i+1)2^{i+1}}}{\frac{1}{i2^i}}=\lim_{i\to\infty}\frac{i}{2(i+1)}=\frac{1}{2}<1\,. \tag{16}$$

So Eq. (12) is constringent. With the expanding series formula of logarithmic,

$$\begin{aligned}ln(1+x)&=x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4+....,(-1<x\le 1) \\ ln(1-x)&=-x-\frac{1}{2}x^2-\frac{1}{3}x^3-\frac{1}{4}x^4-....,(-1<x\le 1)\end{aligned}\,. \tag{17}$$

we have

$$\begin{aligned}&\frac{[ln(1+x)-ln(1-x)]-[ln(1+x)+ln(1-x)]}{2} \\ &=x+\frac{1}{2}x^2+\frac{1}{3}x^3+\frac{1}{4}x^4+....=\sum\nolimits_{i=1}^{\infty}\frac{x^i}{i}\end{aligned}\,. \tag{18}$$

Taking the value x=2^{-1}, Eq. (15) can be rewritten as

$$\lim_{t\to\infty} C = \sum_{i=1}^{t} \frac{1}{i2^i} = \frac{1}{2}[ln(1+\frac{1}{2}) - ln(1-\frac{1}{2})] - [ln(1+\frac{1}{2}) + ln(1-\frac{1}{2})] = ln2 \tag{19}$$

Fig.4 shows the relationship between C and t from theoretical and numerical results, which are consistent. Theoretic curve is calculated with Eq. (14).

3.3 Diameter and Small World

The length of the geodesic, l_{ij}, is defined as the minimum number of edges from node i to node j in a network. The maximum value of l_{ij} between any pair of its nodes is called the diameter of the network at step t, which is denoted by D_t. Diameter plays an important role in the transport and communication within a network and it can measure maximum delay of them. By construction, when t=0 there are the first three nodes in the network and D_0 =1. As increasing of step t, D_t is associated with the new nodes joined to the network at step t, and we can observe the relationship between D_t and t as following

$$D(t) = t+1. \tag{20}$$

Fig.5 shows the relationship between D_t and t from theoretical and numerical results, which are consistent. We observe Eq.(20) that D_t= t+1 for the GFTN. Since at time t the network diameter is also equal to t+1, so the average path length(APL) of the network is less than t+1. In the mean time, the logarithm value of total number of nodes n_t is $\ln(3\cdot 2^t) = \ln 3 + t\ln 2$, which equals approximately to $t$ln2 for large t. It shows that D grows logarithmically with the network size n_t, so its APL grows slower than ln(n_t). It means obviously that the GFTN has the small-world characteristic.

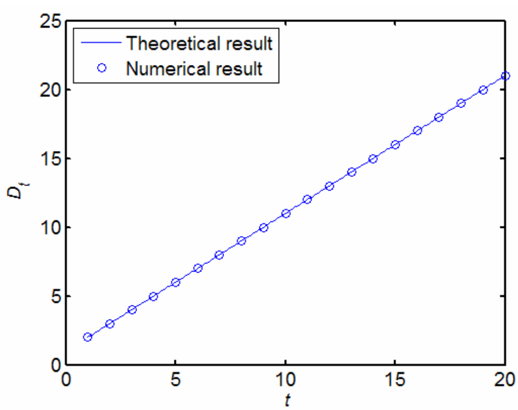


Fig. 5. Comparison of the diameter of networks between theoretic result (circle) and numerical one (line) with the increasing of step t

3.4 Degree-Degree Correlations

Besides above three network characteristics, degree-degree correlations (or so-called assortative coefficient) are usually used to measure the remaining degree—the number of edges leaving the node other than the one we arrived along. A widely accepted formula is the one of assortative coefficient r_c proposed by Newman[7] as following.

$$r_c = \frac{e_t^{-1}\sum_i p_i q_i - [e_t^{-1}\sum_i \frac{1}{2}(p_i + q_i)]^2}{e_t^{-1}\sum_i \frac{1}{2}(p_i^2 + q_i^2) - [e_t^{-1}\sum_i \frac{1}{2}(p_i + q_i)]^2} . \quad (21)$$

where p_i and q_i are the degrees of the vertices at the ends of the ith edge, with i=1, 2, …, e_t. And r_c lies in the range $-1 \le r_c \le 1$.

A network is said to show assortative mixing when $r_c>0$, and it indicates that the nodes in the network that have many connections tend to be connected to other nodes with many connections. A network is said to show disassortative mixing when $r_c<0$, and it indicates that the nodes in the network that have many connections tend to be connected to other nodes with few connections. The value of r_c is zero for no assortative mixing or disassortative mixing.

By construction above, the GFTN is composed of three Farey trees, and each of them has similar topology structure. An arbitrary edge e_{pq} in a single Farey tree can be analyzed through the two nodes of it connects, p and q. According to the current steps t_p and t_q, at which the two nodes joined to the network, we can easily obtain their characteristics as filled in the Tab.1, where $N_{kp,kq}$ is the number of edges e_{pq}.

Table 1. Characteristics of edge and two nodes that it connects in the network

t_p	t_q	$N_{kp,kq}$	k_p	k_q
	0	1·3	2(t+1)	2(t+1)
	1	2·3	2(t+1)	2t
0	2	2·3	2(t+1)	2(t-1)
	…	…	…	…
	t	2·3	2(t+1)	2[t-(t-1)]
	2	2·3	2t	2(t-1)
1	3	2·3	2t	2(t-1)
	…	…	…	…
	t	2·3	2t	2[t-(t-1)]
…	…	…	…	…
	i+1	$2·2^{i-1}·3$	2[t-(i-1)]	2(t-i)
	…	…	…	…
i (i≥1)	i+j	$2·2^{i-1}·3$	2[t-(i-1)]	2[t-(i+j-1)]
	…	…	…	…
	t	$2·2^{i-1}·3$	2[t-(i-1)]	2[t-(t-1)]
…	…	…	…	…
t-1	t	$2·2^{t-2}·3$	4	2

Combing with Tab.1, Eq.(21) can be rewritten as following

$$\begin{cases} M = \dfrac{1}{3\cdot(2^{t+1}-1)} \\ A = 3\cdot\sum_i p_i q_i \\ = \sum_{i=1}^{t}[2\cdot 2(t+1)\cdot 2(t+1-i)] + [2(t+1)]^2 + \sum_{j=1}^{t-1}\sum_{i=j+1}^{t}[2^j\cdot 2(t-j+1)\cdot 2(t+1-i)] \\ B = 3\cdot\sum_i \dfrac{1}{2}(p_i+q_i) \\ = 3\cdot\dfrac{2\sum_{i=1}^{t}[2(t+1)+2(t+1-i)]+2\cdot 2(t+1)}{2} + 3\cdot\dfrac{\sum_{j=1}^{t-1}\sum_{i=j+1}^{t}\{2^j[2(t-j+1)+2(t-i+1)]\}}{2} \\ C = 3\cdot\sum_i \dfrac{1}{2}(p_i^2+q_i^2) \\ = 3\cdot\dfrac{2\sum_{i=1}^{t}[2^2(t+1)^2+2^2(t+1-i)^2]+2\cdot 2^2(t+1)^2}{2} + 3\cdot\dfrac{\sum_{j=1}^{t-1}\sum_{i=j+1}^{t}\{2^j[2^2(t-j+1)^2+2^2(t-i+1)^2]\}}{2} \\ r_c = \dfrac{MA-(MB)^2}{MC-(MB)^2} \\ = \dfrac{6+2^{3t+3}+2t^2-2^{4t+4}+9t+28(4^t)-38t2^t-26(2^t)-2^{3t+3}t-8t^2 4^t+44t4^t+16t^2 8^t-4t^2 2^t}{[6t^2 2^t+t^2+6t2^t+5t-16(4^t)+16(2^t)](2^{t+1}-1)^2} \end{cases} \quad (22)$$

From the expression of r_c in Eq. (22), we can easily calculate its value with t trends to infinity.

$$\lim_{t\to+\infty} r_c = \frac{1}{4} = 0.25 \quad (23)$$

Fig.6 gives the relationship of $\boldsymbol{r_c}$ with t and compares of theoretical with numerical results, which are consistent each other. While $t<4$, $r_c<0$ and network show disassortative mixing. While $t \geq 4$, $r_c > 0$ and r_c is tend to a constant, the network show assortative mixing.

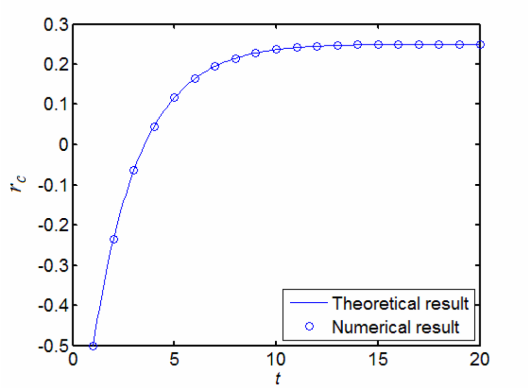


Fig. 6. The r_c vs t and comparison of theoretical with numerical results

4 Conclusions

In summary, we have introduced and studied another kind family of complex network: generalized Farey tree network of the time evolution with geometrically growth. We have derived the analytical expressions of characteristic quantities of the GFTN and computed corresponding numerical results, which are in good accordance with each other. These characteristic quantities include degree distribution, average clustering coefficient, diameter, as well as the degree-degree correlations, which are determined by the step (level number) t and in accordance with large amount of real observations. We have revealed the main features of the GFTN as follows. The degree distributions is exponential form with discrete. The average clustering coefficient of the GFTN is decreasing as step t (or level number) increasing and is tendency to a constant, ln2 for large time t. The GFTN has the small world since the *APL* is less than $ln(n_t)$. The degree-degree correlations are independent when t is large and the networks almost show assortative mixing.

The features above of the GFTN provide a new way to study the complexity-diversity and universality-simplicity of the real-world networks. Further work will study weighted GFTN, and it can be extended to investigate another kind of generalized Farey organized pyramid [8].

Acknowledgment. This work is supported by Nature Science Foundation of China (Grand Nos. 60874087, 70431002, 60773120 and 10647001). Nature Science Foundation of Beijing (Grand No. 4092040).

References

1. Kim, S. H., Ostlund, S.: Simultaneous rational approximations in the study of dynamical systems. Phys. Rev. A. 34, 3426--3434 (1986)
2. Maselko, J., Swinney, H. L.: A Farey triangle in the Belousov-Zhabotinskii reaction. Phys. Lett. A. 119, 403--406 (1987)
3. Fang, J. Q.: Generalized farey organization and generalized winding number in a 2-D DDDS. Phys. Lett. A. 146, 35--44 (1990)
4. Calvo, O., Cartwright, J. H. E., Gonzalez, D.L., et al.: Three-frequency resonances in coupled phase-locked loops. Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions. 47, 4, 491--497 (2000)
5. Fang, J. Q., Wang, X. F., Zheng, Z. G., et al.: New interdisciplinary science: network science(I) (In Chinese). Prog. In Phys. 27, 3, 239--343 (2007)
6. Fang, J. Q., Wang, X. F., Zheng, Z. G., et al.: New interdisciplinary science: network science(II) (In Chinese). Prog. In Phys. 27, 4, 361--548 (2007)
7. Newman, M. E. J.: Assortative Mixing in Networks. Phys. Rev. Lett. 89, 208701 (2002)
8. Fang, J. Q.: Briefly review on complex network pyramid and their universality-complexity. In: Fang, J. Q(eds.) Proceedings of CCAST(WL) Workshop Series:, Forth National Forum on Network Science and Graduate Student Summer School, vol.191, pp. 204-221. CCAST, Beijing (2008)