Networks on Musical String

Debashis Mondal (05CS6005)
Himika Biswas (05CS6027)
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Abstract

One of the fundamental elements of Hindustani classical music is *raaga*. *Raaga* is the intricate system of scales and associated melodic patterns. *Raagas* express melodic structure. In their numerical ratios, the scales and melodic patterns correspond with moods, colors, seasons, and hours of day and night. There are about 200 main *raagas*, each of which is defined by its unique combination of scale-pattern, dominant notes, specific rules to be followed in ascending or descending and certain melodic phrases associated with it.

In this work we have tried to find out some interesting characteristics about Hindustani Classical Music by creating Networks on Musical Strings. For creating these networks here we have considered a small set of *raagas*. Then by analyzing the networks we have founded few interesting features of these *raagas* and some correlations among them.

1 Introduction

Here we are considering a set of *Raagas*[1][4] from Hindustani Classical Music. Then we are building Networks from them to analyze and find out certain interesting characteristics.

Note	*Position*	*Note*	*Position*	*Note*	*Position*
$r <$	1	r	13	$r >$	25
$R <$	2	R	14	$R >$	26
$g <$	3	g	15	$g >$	27
$G <$	4	G	16	$G >$	28
$m <$	5	m	17	$m >$	29
$M <$	6	M	18	$M >$	30
$P <$	7	P	19	$P >$	31
$d <$	8	d	20	$d >$	32
$D <$	9	D	21	$D >$	33
$n <$	10	n	22	$n >$	34
$N <$	11	N	23	$N >$	35
S	12	$S >$	24		

Table 1: Note Positional Table

Each Raaga, which we have considered, constitutes of several compositions without any distinct composition boundary. But for our convenience we are assuming that consecutive 4 lines correspond to a composition. The basic notes are represented as **S r R g G m M P d D n N**, where the capital letters denote sudhdha swaras, whereas small letters represent komal swaras. The only exception is madhyam (ma), where 'm' stands for sudhdha madhyam, and M stands for tivra madhyam (kadi ma).

[1]The Raaga is the basic unit of Indian Classical Music. It is defined to be a set of notes, and the adjacency relationship of the notes is governed by the Raaga's grammar.

A note in lower octave is followed by '<', e.g. "d<" means komal dhaivat in the mandra saptak. Similarly, "d>" means komal dhaivat of the tad saptak (upper octave).

For building the network, we are considering a sliding window of size k. k consecutive notes are taken from a line of a composition and considered as a node of the network. Then we are sliding the window to one position to the right and taking the next node. In this way we are creating the nodes. Then with unique nodes we are building the network.

For example, we are considering a file, say Bhairavi.txt, where the compositions are written. We are taking a window of size *four* and considering them as nodes in the network. Let, "g g r S n< S d d n d d g m d m P" be a line of a composition. So, the nodes are:

- g g r S
- g r S n<
- r S n< S
- S n< S d
- n< S d d
- ....

Raaga	*Thaat*
Alhaiya	*Bilaval*
Asabari	*Asabari*
Basant	*Purvi*
Bhairavi	*Bhairavi*
Bhairav	*Bhairav*
Kafi	*Kafi*
Khamaj	*Khamaj*
Marwa	*Marwa*
Paraj	*Purvi*
Purvi	*Purvi*
Todi	*Todi*
Yaman	*Kalyan*

Table 2: Raagas and their *Thaats*

Here we are looking at overlapping windows, because that way we do not have to justify where we start counting the window. In other words this is just the note 4-gram. In this way we are finding the unique nodes from these compositions. Here we have created three types of networks.

1. **Nodes as note sequences (Network1):** Here each *node* is a k consecutive unique note sequence of compositions. Between any two nodes there exists an *edge* if they belong to the same composition. *Weight* of an edge is defined as the number of compositions where the end nodes of the edge co-occur. Here for each composition the nodes present in that composition will be completely connected.

2. **Nodes as compositions (Network2):** Here each *node* is a composition. Between any two nodes there exists an *edge* if they have any k length unique note sequence in common. *Weight* of an edge is defined as the number of such k sequences common in between two nodes.

3. **Nodes as Jumping distance sequences (Network3):** *Jumping distance* between two consecutive notes of a composition is defined as the distance between the notes. For example, a sequence of notes is $S\ n{<}\ r\ G\ m\ N\ S{>}$, here the jumping distance between $S\ n{<}$ is 2, between $n{<}\ r$ is 2 and so on. We have defined the positions of each note of each octave as shown in

Table 1. Jumping distance is nothing but the positional distance between two notes. We have created *nodes* as the unique consecutive k jumping note sequence of the compositions. *Edges* and *weights* are defined as it was in network type 1. Here also for each composition the nodes present in that composition will be completely connected.

Section 2 discusses about the objective of this work, section 3 elaborates the experimental design. In sections $4-6$ we have discussed about the obtained results from the analysis of 3 different networks. Finally conclusion has been drawn in section 7.

$Features$	Threshold = 5			Threshold = 10			Threshold = 15			Threshold = 20		
	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$
$\#Nodes$	1278	2918	5021	1278	2918	5021	1278	2918	5021	1278	2918	5021
$\#DN$	841	2284	4409	1002	2618	4820	1076	2747	4950	1121	2815	4999
$\#Clusters$	122	104	63	72	55	25	55	32	12	44	20	6
$AverageCC$	0.26	0.15	0.07	0.16	0.07	0.02	0.12	0.03	0.003	0.08	0.02	0.001
$Diameter$	4	7	9	6	9	17	5	7	9	6	6	3
$ASPL$	2.25	2.97	3.91	2.36	3.4	6.56	2.49	2.9	3.59	2.69	2.74	1.39

Table 3: Experimental Results for Network1. $\#DN = $ #Disconnected Nodes, $AverageCC = $ Average Clustering Coefficient, $ASPL = $ Average Shortest Path Length.

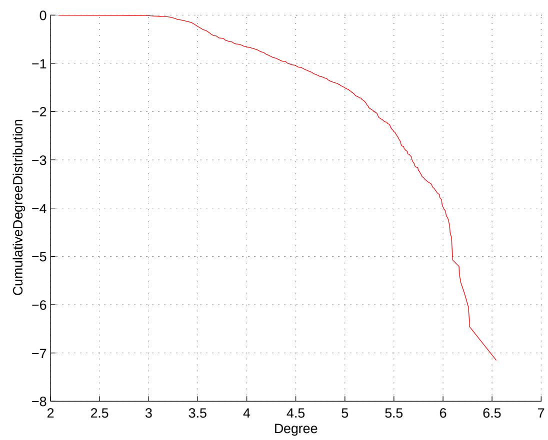

Figure 1: Loglog plot of Cumulative Degree Distribution for $k=3$ for Network1.

2 Objective of this Work

Our main objective of this work is to analyze the created networks in search of certain interesting characteristics about Hindustani Classical Music. For this purpose we have followed the following steps:

- Calculate the Diameter and average shortest path length of the networks with varying window size and different threashold[2] values.
- Calculate the average clustering coefficients of the networks with varying window size and threashold values.
- Analyze the cumulative degree distribution curves for different window size; i.e., why the nature of the curve is so.

[2]we have considered a threashold value on the edge weight; if the weight of an edge is more than the threshold value then we have considered that edge for analysis of the networks.

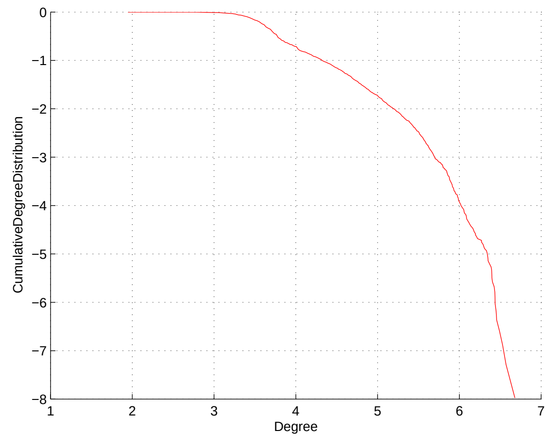

Figure 2: Loglog plot of Cumulative Degree Distribution for $k = 4$ for Network1.

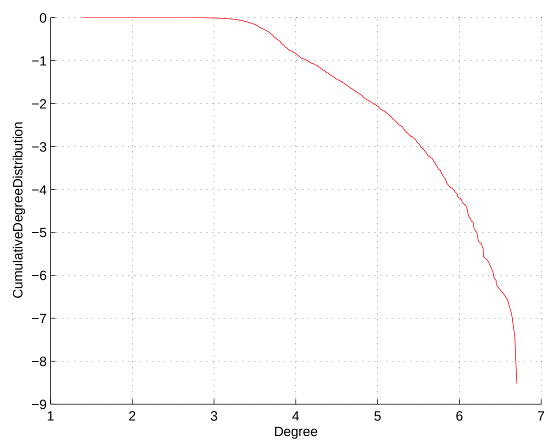

Figure 3: Loglog plot of Cumulative Degree Distribution for $k = 5$ for Network1.

- Apply clustering on the networks so that we can analyze what these clusters signify? Whether they are the note sequences corresponding to the same *Raaga* (i.e. one cluster corresponds to one *raaga*) or are they of similar *Raagas* (i.e. one cluster corresponds to a set of similar *Raagas* belonging to same *Thaat*)? The answer will depend on:
  - The set of compositions that we choose.
  - The clustering algorithm, as well as the clustering parameters.
  - The inherent nature of the networks and Hindustani Classical Music.
- Whether there exist some nodes having very high degree. If exists then whether they belong to most of the chosen Raagas and whether these note sequences are soothing or not. If they are soothing then what are the specialities about these sequences and whether there are any similarities among these nodes.

3 Experimental Design

For building the networks we have first reformed the compositions such that successive same notes are replaced by one occurance of the note. For example if the note sequence is $SSnnPd < d <$, then the reformed note sequence will be $SnPd <$. We have taken this approach as

Note Sequence	Present in Raagas
$NS > N$	Basant, Bhairav, Kafi, Khamaj, Marwa, Paraj, Purvi, Todi, Yaman
$S > NS >$	Basant, Bhairav, Kafi, Khamaj, Marwa, Paraj, Purvi, Todi, Yaman
$S > r > S >$	Alhaiya, Asabari, Basant, Bhairav, Bhairavi, Marwa, Paraj, Purvi, Todi
dPd	Alhaiya, Asabari, Bhairav, Bhairavi, Paraj, Todi
PmP	Alhaiya, Asabari, Bhairav, Bhairavi, Kafi, Khamaj, Paraj, Todi
dPm	Alhaiya, Asabari, Bhairav, Bhairavi, Paraj, Purvi, Todi
$SN < S$	Bhairav, Kafi, Marwa, Paraj, Purvi, Todi, Yaman
PdP	Alhaiya, Asabari, Basant, Bhairav, Bhairavi, Paraj, Purvi, Todi
GmP	Bhairav, Kafi, Khamaj, Paraj
PMG	Basant, Purvi, Yaman
$NS > NS >$	Basant, Bhairav, Kafi, Khamaj, Marwa, Paraj, Todi
$S > NS > N$	Basant, Bhairav, Kafi, Khamaj, Marwa, Paraj, Todi, Yaman
$S > r > S > N$	Basant, Bhairav, Paraj, Purvi, Todi
$rSN < S$	Basant, Bhairav, Marwa, Paraj, Purvi, Todi
$MGrS$	Basant, Marwa, Paraj, Purvi
$dPmP$	Alhaiya, Asabari, Bhairav, Bhairavi, Paraj, Todi
$GMGr$	Basant, Marwa, Paraj, Purvi
$dNS > N$	Basant, Bhairav, Paraj, Purvi, Todi
$MGMG$	Basant, Marwa, Paraj, Purvi, Yaman
$NdPM$	Basant, Paraj, Purvi, Todi
$GMGrS$	Basant, Marwa, Paraj, Purvi
$dNS > r > S >$	Basant, Bhairav, Paraj, Todi
$MGMGr$	Basant, Marwa, Paraj, Purvi
$NS > r > S > N$	Basant, Bhairav, Paraj, Todi
$S > r > S > Nd$	Basant, Bhairav, Paraj, Purvi, Todi
$dNS > NS >$	Basant, Bhairav, Paraj, Todi
$NS > NS > N$	Bhairav, Kafi, Khamaj, Paraj, Todi
$MdNS > N$	Basant, Paraj, Purvi, Todi
$NdNdP$	Basant, Bhairav, Paraj, Purvi, Todi
$S > NS > NS >$	Bhairav, Kafi, Khamaj, Marwa, Paraj, Todi

Table 4: Top 10 note sequences for k=3, 4 and 5 for Network1.

otherwise we are getting the note sequences like SSS, PPP, $\ldots$ which are not much interesting from the point of view of analysis. We have calculated average clustering coefficients also for each network with varying size of window and with varying threshold.

Next for finding the *diameter*, *average shortest path length* we have used the *Pajek*[1] software. For using the *Pajek* software the generated network structure output is made compatible with the *Pajek* input *.net* file. Using the *Pajek* we have found *hierarchical clustering*, *partitioning* of all the three types of the networks. From the *partitioning* and the *heirarchical clustering* of the networks we have tried to find whether there is any significant cluster(s) in the networks or not. Pajek uses clustering algorithm where more weightage is given to the edges with less weights. But in our network, clusters should be formed between the nodes whose edge weights are high. For resolving this, we have just inverted the edge weights in the input to the Pajek. So now edges with higher weights will get higher weightage.

Features	Threshold = 5			Threshold = 10			Threshold = 15			Threshold = 20		
	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$
$\#Nodes$	381	381	381	381	381	381	381	381	381	381	381	381
$\#DN$	0	1	9	6	21	75	42	113	193	177	236	281
$\#Clusters$	133	106	63	94	55	27	52	29	15	23	14	9
$AverageCC$	0.75	0.68	0.56	0.72	0.53	0.30	0.51	0.25	0.10	0.17	0.09	0.03
$Diameter$	4	7	13	10	7	9	9	11	7	7	6	4
$ASPL$	2.00	2.81	4.83	3.81	2.60	3.42	2.89	3.77	2.59	2.89	2.30	1.66

Table 5: Experimental Results for Network2. $\#DN$ = #Disconnected Nodes, $AverageCC$ = Average Clustering Coefficient, $ASPL$ = Average Shortest Path Length.

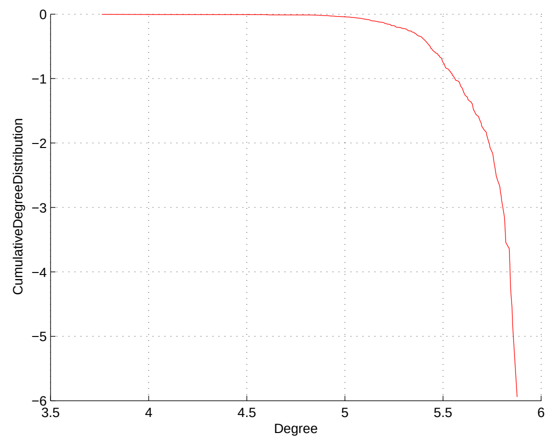

Figure 4: Loglog plot of Cumulative Degree Distribution for $k = 3$ for Network2.

3.1 Data Set

We have taken 12 popular *Raagas* for our experiments. These *raagas* are of different *Thaats*[3] [3]. The considered *raagas* and their corresponding *thaats* are given in Table2.

We can see that each of these *raagas* belong to different *Thaats* (except *raagas* Purvi, Basant and Paraj belong to the same *Thaat* Purvi. In each raaga we have taken on an average 25 compositions from [2].

4 Results obtained for Network1

In Table 3 we can see the obtained data for network1. Here we have observed that with the increasing size of window as well as the threshold the graph is becoming sparse and there we are finding several disconnected components. We have not found any cluster in the thresholded graphs. The clustering coefficient values are also supporting the absense of clusters in the network. The cumulative degree distribution plots can be seen for $k = 3, 4\&5$ in the figures 1, 2 and 3 respectively. All these figures follows power law and along with that the diameter and average shortest path length of the network shows the small world property of the graph as infered in [5]. From the *Arohan* [3] and *Avrohan* [3] of the selected twelve *raagas* we have found that some of the note sequences are common in different *raagas*, e.g. note sequence $S >$

[3]Thaat is a way of classification of *Raagas*. Each Thaat is a set of seven notes in an octave. Classification of *Raaga* into a Thaat is empirical rather than strictly formal, because a *raaga* might use more than 7 notes. For example, thaat Khamaj contains the notes S, R, G, m, P, D, n but the *raaga* desh belonging to this thaat uses both n(komal ni) and N(suddha ni). Therefore, thaat is not strictly computable - though it may be computable for some *raagas*.

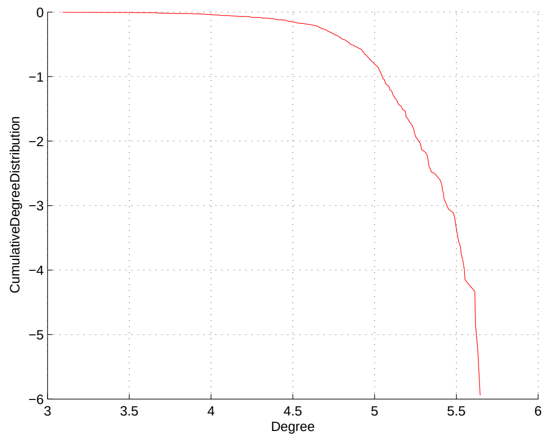

Figure 5: Loglog plot of Cumulative Degree Distribution for $k = 4$ for Network2.

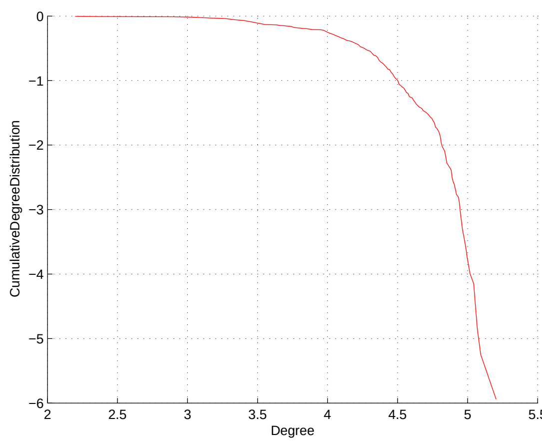

Figure 6: Loglog plot of Cumulative Degree Distribution for $k = 5$ for Network2.

$NDPMG$ is common in the *Avrohan* of both *raagas* Todi, Yaman; note sequence $S > NdPM$ is common in *Avrohan* of the *raagas* Basant, Purvi, Paraj, etc. As in this network we have considered an edge between two nodes if they cooccure in same composition. On an average there are more than 20 compositions in a *raaga* in our dataset. So, most of the edges are having weights greater than the threashold we have used and the thresholded network is showing the small world property.

Table 4 shows the top 10 note sequences in the network for varying window size of 3, 4 and 5. We are finding interestingly that all these note sequences are basically consecutive notes like $SNSN$, $NS > R > S$ etc. This may indicate that the core of the music pattern is just like this, i.e. soothing note sequences are the consecutive note sequences.

We have also observed from Table 4 that, for all the highest frequency note sequences *Raaga* Asabari and Alhaiya are co-occuring. Similarly, we found that *Raaga* Kafi, Khamaj and *Raaga* Basant, Paraj, Purvi are also co-occuring in almost all the highest frequency note sequences. From this we can say that *raaga* Asabari, Alhaiya and *raaga* Kafi, Khamaj are of similar nature. As *raaga* Basant, Paraj and Purvi belong to the same *Thaat* Purvi so trivially they also follow the similar characteristics.

We have chosen the 12 *raagas* which are mostly from different *thaats* and that is why we have not found any significant cluster in the network as the characteristics of all these *raagas* are

Features	Threshold = 5			Threshold = 10			Threshold = 15			Threshold = 20		
	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$
$\#Nodes$	423	1000	2103	423	1000	2103	423	1000	2103	423	1000	2103
$\#DN$	293	701	1624	327	805	1865	342	852	1993	352	881	2036
$\#Clusters$	82	125	110	66	83	52	52	63	33	49	52	23
$AverageCC$	0.27	0.23	0.14	0.20	0.15	0.06	0.16	0.11	0.03	0.14	0.09	0.02
$Diameter$	3	4	6	3	5	8	3	5	6	3	4	6
$ASPL$	1.63	1.94	2.54	1.57	1.96	2.94	1.56	2.04	2.58	1.54	2.00	2.62

Table 6: Experimental Results for Network3. $\#DN$ = #Disconnected Nodes, $AverageCC$ = Average Clustering Coefficient, $ASPL$ = Average Shortest Path Length.

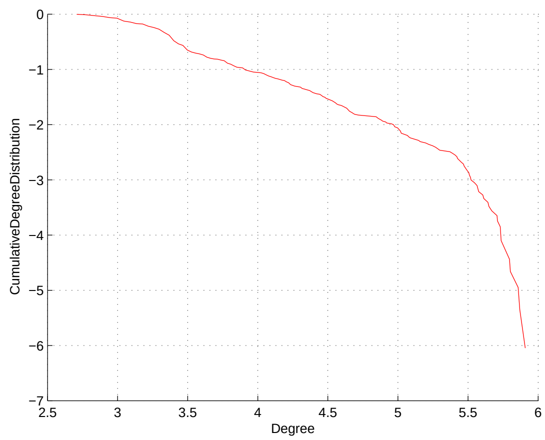

Figure 7: Loglog plot of Cumulative Degree Distribution for $k = 3$ for Network3.

completely different.

5 Results obtained for Network2

From Table 5 and figures 4 - 6, we can conclude that this network is not a small world as the cumulative degree distribution plots are not following power law. Here in this network we define an edge between two compositons if they contain same note sequence in common. In this case we are not considering the edges within intra-composition note sequences. As a result the edge weights are not so high as in Network1. That is why the diameter of the network is high and it is not following the small world property.

Here from Table 5 we can see that for threashold 5 and 10 the average clustering coefficients of the network is quite high, whereas for threshold 15 and 20 it is low. From these values we can conclude that, there are very few note sequences which have cooccured in more than 20 compositions. On an average we have taken about 20 compositions in a *Raaga*. So we can conclude that these 20 compositions are from the same *Raaga* and/or are from the same *Thaat*. If we increase the threshold more than 20 then the graph is fragmented into small components. Here also from the values of #Nodes and #clusters in the network, we can see that there is no significant cluster in the graph.

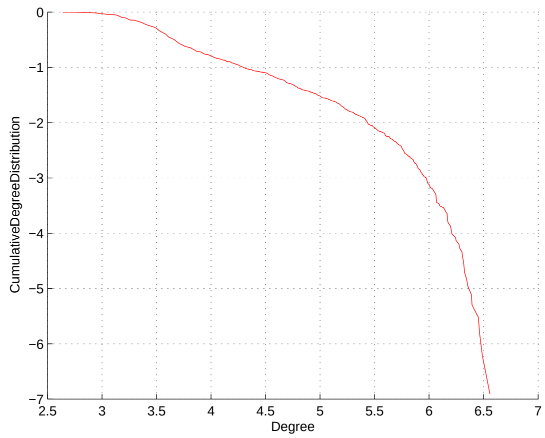

Figure 8: Loglog plot of Cumulative Degree Distribution for $k = 4$ for Network3.

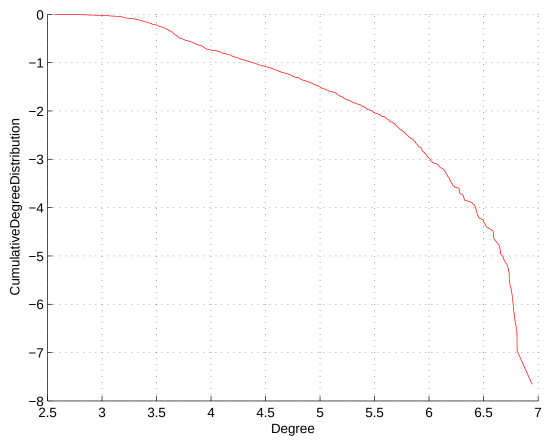

Figure 9: Loglog plot of Cumulative Degree Distribution for $k = 5$ for Network3.

6 Results obtained for Network3

In Table 6 we can see the obtained data for network3. Here also we have observed that with the increasing size of windows and increasing value of the threshold the graph is becoming sparse and there we are finding several disconnected components. We have find no cluster in the thresholded graphs. The clustering coefficient values are also supporting the absense of clusters in the network. The cumulative degree distribution plots can be seen for $k = 3, 4\&5$ in the figures 7, 8 and 9 respectively. All these figures along with the diameter and average shortest path length of the network shows the small world property of the graph here also. We can justify the small world property of this network as we have justified in Network1.

Table 7 shows the top 10 note sequences in the network for varying window size of 3, 4 and 5. Here also we are finding that all these note sequences are basically consecutive notes, i.e. jumping distance is too small for all the sequences. This also indicates that the soothing note sequences are the consecutive note sequences and this may be the core of the music.

As before we can conclude that as we have chosen the 12 *raagas* which are mostly from different *thaats* and that is why we have not found any significant cluster in the network as the characteristics of all these *raagas* are completely different.

Note Sequence	Note Sequence	Note Sequence
1 2 2	1 2 2 2	1 2 2 2 2
2 2 2	1 1 2 2	1 1 2 2 2
2 3 1	2 2 2 2	2 2 2 2 2
1 1 2	1 1 1 2	1 1 1 1 2
2 2 3	2 2 2 1	2 2 2 1 2
1 1 1	2 3 1 1	3 1 1 1 1
2 2 1	2 2 2 3	2 3 1 1 1
2 1 1	1 1 1 1	2 2 2 2 3
1 2 3	2 2 1 1	2 1 1 2 2
3 1 2	2 2 3 1	2 2 1 2 2

Table 7: Top 10 jumping distance note sequences for $k = 3, 4, 5$ for Network3.

7 Conclusion

We have found that all the highest frequency note sequences are basically consecutive notes. This shows a very important characteristic about Hindustani Classical Music. We can say from this observation that, the soothing note sequences are the consecutive note sequences and this is the core of the Hindustani Classical Music.

We have also observed that, *Raagas* Asabari and Alhaiya are very close in nature. Similarly we found that *Raagas* Kafi and Khamaj are also closed in their nature. *Raagas* Basant, Paraj and Purvi belong to the same *Thaat* Purvi and that is why trivially they also follow the similar characteristics.

We have chosen 12 *raagas* of different *thaats*. May be for this reason we have found that there is no cluster in the network. If we carry on our experiments with different sets of *raagas* then we may get some more interesting characteristics about the *raagas*.

Acknowledgement

We take the opprtunity to express our deep sense of gratitude to Prof. Niloy Ganguly and Monojit Choudhury for their encouraging guidance, support and inspiration. We are also thankful to Animesh Mukherjee for his kind help and cooperation. We are grateful to all the faculty members and staffs of the department of Computer Science & Engineering, IIT Kharagpur for their support whenever we required.

References

[1] Vladimir Batagelj. **Pajek: Package for Large Network Analysis**, http://vlado.fmf.uni-lj.si /pub/networks/pajek/.

[2] Pt. V. A. Bhatkhande. **Hindustani Sangeet Paddhati KRAMIK PUSTAK MALIKA**, Sangeet Karyalaya, India, 1923.

[3] Monojit Choudhury and Pradipta Ranjan Ray. **Measuring Similarities Accross Musical Compositions: An Approach Based on the Raga Paradigm**, Journal of ITC Sangeet Research Academy, ITC-SRA, 17:1-10, 2003.

[4] Dipanjan Das and Monojit Choudhury. **Finite State Models For Generation of Hindustani Classical Music**, Proceedings of the International Symposium on Frontiers of Research in Speech and music (FRSM - 2005), pp. 59-64, Bhubaneswar, India, 2005.

[5] Pablo M. Gleiser and Leon Danon. **Community Structure in Jazz**, Advances in Complex Systems , Vol. 6, No. 4, pp. 565-573, 2003.