PHRASE-BASED RĀGA RECOGNITION USING VECTOR SPACE MODELING

Sankalp Gulati[†], Joan Serrà*, Vignesh Ishwar[†], Sertan Şentürk[†], Xavier Serra[†]

[†]Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
*Telefonica Research, Barcelona, Spain

ABSTRACT

Automatic rāga recognition is one of the fundamental computational tasks in Indian art music. Motivated by the way seasoned listeners identify rāgas, we propose a rāga recognition approach based on melodic phrases. Firstly, we extract melodic patterns from a collection of audio recordings in an unsupervised way. Next, we group similar patterns by exploiting complex networks concepts and techniques. Drawing an analogy to topic modeling in text classification, we then represent audio recordings using a vector space model. Finally, we employ a number of classification strategies to build a predictive model for rāga recognition. To evaluate our approach, we compile a music collection of over 124 hours, comprising 480 recordings and 40 rāgas. We obtain 70% accuracy with the full 40-rāga collection, and up to 92% accuracy with its 10-rāga subset. We show that phrase-based rāga recognition is a successful strategy, on par with the state of the art, and sometimes outperforms it. A by-product of our approach, which arguably is as important as the task of rāga recognition, is the identification of rāga-phrases. These phrases can be used as a dictionary of semantically-meaningful melodic units for several computational tasks in Indian art music.

Index Terms— Rāga recognition, rāga motifs, melodic phrases, Indian art music, Carnatic music

1. INTRODUCTION

Rāga is the melodic framework of Hindustani and Carnatic music [1, 2], the two art music traditions of the Indian subcontinent. In addition, numerous compositions in folk and film music in India are based on rāgas. It is a core component for navigation, organization, and pedagogy in Indian art music (IAM). A computational approach to automatic rāga recognition can enable rāga-based music retrieval from large audio archives, semantically-meaningful music discovery and navigation, and several applications around music pedagogy. In this paper, we propose an approach to automatic rāga recognition in audio collections of IAM.

A rāga is characterized by a set of svaras (roughly speaking, notes), ārōhana-avrōhana (the ascending and descending melodic progression), and, most importantly, by a set of characteristic melodic phrases or motifs (also referred to as catch phrases). These melodic phrases capture the essence of a rāga, and are the building blocks of melodies in IAM [1], both composed and improvised. In addition, they serve as the main cues for rāga identification by experienced listeners [3].

Despite the importance of melodic phrases in rāga characterization, they are barely utilized by the computational approaches for

rāga recognition. This can be attributed to the challenges involved in a reliable extraction of melodic patterns in audio recordings of IAM. These challenges include the extraction of a reliable melodic representation, the lack of musically-meaningful melody segmentation models, the difficulty of deriving perceptually-relevant melodic similarity measures, and the ability to deal with large collections of long audio recordings. In recent years, research in computational modeling of IAM has been focused on addressing these challenges [4, 5].

Rāga recognition is one of the most researched topics within music information retrieval of IAM. A large number of approaches to this task use pitch distribution (PD)-based tonal features [6–12]. However, such approaches do not consider the temporal aspects of melody such as melodic progressions and phrases, which are some of the fundamental aspects that characterize rāgas. These aspects are even more relevant in distinguishing phrased-based rāgas [3]. Several studies address the limitations of PD-based methods by modeling temporal dynamics of the melody [13–17]. These methods generally use melodic progression templates [15], n-gram distributions [13], or hidden Markov models [16] to capture the sequential information. To the best of our knowledge, there are only two approaches [18] and [14] that explicitly use rāga phrases. However, these approaches rely on manually annotated phrases, and thus, they are susceptible to subjectivity, as musicians may not agree to the same set of characteristic phrases. This also limits the applicability of these approaches to even moderate-sized real-world collections.

In this paper, we propose a novel approach for rāga recognition that uses automatically discovered melodic phrases. For that, we use our earlier work on unsupervised discovery of melodic patterns in IAM [19]. We select relevant melodic phrases from the discovered patterns employing topic modeling techniques used in text classification. We draw an analogy between the unfolding of a rāga using melodic phrases in a recording, and the description of a topic using words in a text document. In this process, we also come up with a technique that uses network analysis to estimate a melodic similarity threshold. We evaluate our approach on a diverse and representative dataset of Carnatic music, which we make publicly available. It is the largest such dataset in terms of the number of recordings, the number of rāgas and total duration. Since our approach is based on melodic phrases, the extracted features and results can be easily interpreted. Furthermore, the selected set of melodic phrases can directly be used as a dictionary of rāga phrases in several applications.

2. METHOD

2.1. Melodic Pattern Discovery

The block diagram of the proposed approach is shown in Figure 1. In order to get a reliable input for our approach, we employ the IAM pattern discovery system we presented in [19]. This is one of the few unsupervised systems we are aware of that can discover

Fig. 1. Block diagram of the proposed approach.

Fig. 2. Evolution of clustering coefficients of $\mathcal{G}$ and $\mathcal{G}_r$ and their difference for different similarity thresholds.
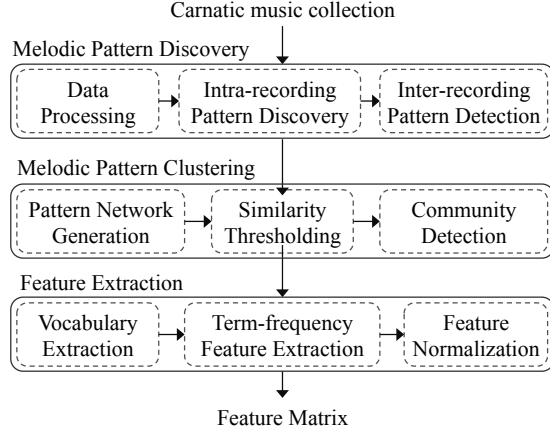
meaningful melodic patterns in large-scale collections of IAM. This pattern discovery method consists of three main processing blocks (Figure 1). In the data processing block, predominant melody is extracted and tonic normalized pitch is used as the melody representation. Here, all possible 2 second melodic segments are considered as pattern candidates. This duration is chosen based on the average human annotated phrase duration reported in a recent study [20]. In the intra-recording pattern discovery block, melodically similar seed patterns are discovered within each audio recording. Finally, in the inter-recording pattern detection block, these seed patterns are used to perform a search across all the recordings in the dataset. All the seed patterns, and their nearest neighbors across all the recordings is the output of this method. This method uses dynamic time warping based melodic similarity, and employs several lower-bounding techniques to reduce the computational complexity of the task. We use the same parameter settings and implementation as reported in the paper. For a detailed explanation of this approach we refer to [19]. All these discovered phrases are made available online[1] for listening.

2.2. Melodic Pattern Clustering

We proceed to cluster the melodic patterns obtained in the previous step. The objective is to group together all the patterns that are different occurrences of the same melodic phrase. For this, we propose to perform a network analysis in which the clustering is performed using a non-overlapping community detection method.

We start by building an undirected network $\mathcal{G}$ using the discovered patterns as the nodes of the network. We connect any two nodes only if the distance between them is below a similarity threshold T_s. Noticeably, this distance is computed using the same measure as used in the pattern discovery block (Section 2.1). The weight of the edge, when it exists, is set to 1. Non-connected nodes are removed.

Determining a meaningful similarity threshold is a complex task. Here, we define the optimal similarity threshold $T_s^\star$ as the one that maximizes the number of connections between the similar patterns and, at the same time, minimizes the number of connections between the dissimilar ones. With this in mind, we propose to estimate $T_s^\star$ by exploiting the topological properties of the network $\mathcal{G}$. We compare the evolution of the clustering coefficient C of the obtained network $\mathcal{G}$ with a randomized network $\mathcal{G}_r$ over different distance thresholds T_s. Clustering coefficient measures the degree to which nodes in a network cluster together [21]. The randomized
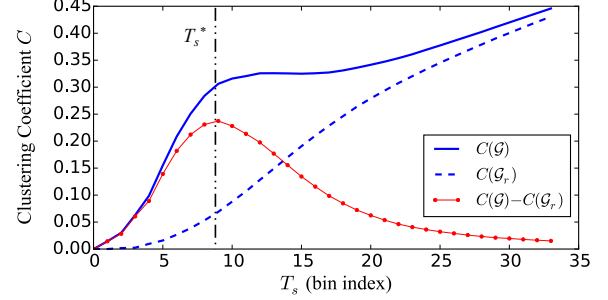
network $\mathcal{G}_r$ is obtained by swapping the edges between randomly selected pairs of nodes such that the degree of each node is preserved [22]. The optimal threshold $T_s^\star$ is taken as the distance that maximizes the difference between the two clustering coefficients. In Figure 2, we show $C(\mathcal{G})$, $C(\mathcal{G}_r)$ and $C(\mathcal{G}) - C(\mathcal{G}_r)$ for different values of T_s, and mark the optimal threshold $T_s^\star$.

The next step of our approach (Figure 1) groups similar melodic patterns. To do so, we detect non-overlapping communities in the network of melodic patterns using the method proposed in [23]. This method is based on optimizing the modularity of the network and is parameter-free from the user's point of view. This method is capable of handling very large networks and has been extensively used in various applications [24]. We use its implementation available in networkX [25], a Python language package for exploration and analysis of networks and network algorithms. Note that, from now on, the melodic patterns grouped within a community are regarded as the occurrences of a single melodic phrase. Thus, a community essentially represents a melodic phrase or motif.

2.3. Feature Extraction

As mentioned, we draw an analogy between rāga rendition and textual description of a topic. Some melodic patterns characterize a rāga (musically recognized as catch phrases), similar to the words that are specific to a topic. Some melodic patterns are used across rāgas (for example, Kampita[2]) similar to stop words in text documents, while some patterns are specific to a recording (composition-specific melodic phrases). Using this analogy we represent each audio recording using a vector space model. This process is divided into three blocks (Figure 1).

We start by building our vocabulary P, which translates to selecting relevant communities (melodic phrases) for characterizing rāgas. For this, we include all the detected communities except the ones that comprise patterns extracted from only a single audio recording. Such communities are analogous to the words that only occur within a document and, hence, are irrelevant for modeling a topic. The size of the obtained vocabulary P corresponding to the optimal threshold mentioned above ($T_s^\star = 9$) is 2594.

We experiment with three different sets of features F_1, F_2 and F_3, which are similar to the term frequency-inverse document frequency features typically used in text information retrieval. We denote our corpus by R comprising $N = |R|$ number of recordings. A melodic phrase and a recording is denoted by p and r, respectively

$$F_1(p, r) = \begin{cases} 1, & \text{if } f(p, r) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

---

[1] http://compmusic.upf.edu/node/278

[2] a specific type of an oscillatory melodic movement on a svara [3].

where, $f(p, r)$ denotes the raw frequency of occurrence of phrase p in recording r. F_1 only considers the presence or absence of a phrase in a recording. In order to investigate if the frequency of occurrence of melodic phrases is relevant for characterizing rāgas, we take $F_2(p, r) = f(p, r)$. As mentioned, the melodic phrases that occur across different rāgas and in several recordings are futile for rāga recognition. Therefore, to reduce their effect in the feature vector we employ a weighting scheme, similar to the inverse document frequency (idf) weighting in text retrieval.

$$F_3(p, r) = f(p, r) \times \text{irf}(p, R) \qquad (2)$$

$$\text{irf}(p, R) = \log\left(\frac{N}{|\{r \in R : p \in r\}|}\right) \qquad (3)$$

where, $|\{r \in R : p \in r\}|$ is the number of recordings where the melodic phrase p is present, that is $f(p, r) \neq 0$ for these recordings. We denote our proposed method by M

3. EVALUATION

3.1. Music Collection

The music collection used in this study is compiled as a part of the CompMusic project [26–28]. The collection comprises 124 hours of commercially available audio recordings of Carnatic music belonging to 40 rāgas. For each rāga, there are 12 music pieces, which amounts to a total of 480 recordings. All the editorial metadata for each audio recording is publicly available in Musicbrainz[3], an open-source metadata repository. The music collection primarily consists of vocal performances of 62 different artists. There are a total of 310 different compositions belonging to diverse forms in Carnatic music (for example kīrtana, varnam, viruttam). The chosen rāgas contain diverse set of svaras (note), both in terms of the number of svaras and their pitch-classes (svarasthānās). To facilitate comparative studies and promote reproducible research we make this music collection publicly available online[4].

From this music collection we build two datasets, which we denote by DB40rāga and DB10rāga. DB40rāga comprises the entire music collection and DB10rāga comprises a subset of 10 rāgas. We use DB10rāga to make our results more comparable to studies where the evaluations are performed on similar number of rāgas.

3.2. Classification and Evaluation Methodology

The features obtained above are used to train a classifier. In order to assess the relevance of these features for rāga recognition, we experiment with different algorithms exploiting diverse classification strategies [29]: Multinomial, Gaussian and Bernoulli naive Bayes (NBM, NBG and NBB, respectively), support vector machines with a linear and a radial basis function kernel, and with a stochastic gradient descent learning (SVML, SVMR and SGD, respectively), logistic regression (LR) and random forest (RF). We use the implementation of these classifiers available in scikit-learn toolkit [30], version 0.15.1. Since in this study, our focus is to extract a musically relevant set of features based on melodic phrases, we use the default parameter settings for the classifiers available in scikit-learn.

We use a stratified 12-fold cross-validation methodology for evaluations. The folds are generated such that every fold comprises equal number of feature instances per rāga. We repeat the entire experiment 20 times, and report the mean classification accuracy as

[3]https://musicbrainz.org
[4]http://compmusic.upf.edu/node/278

db	Mtd	Ftr	NBM	NBB	LR	SVML	1NN
DB10rāga	M	F_1	90.6	74	84.1	81.2	-
		F_2	**91.7**	73.8	84.8	81.2	-
		F_3	90.5	74.5	84.3	80.7	-
	S_1	PCD_{120}	-	-	-	-	82.2
		PCD_{full}	-	-	-	-	**89.5**
	S_2	PD_{param}	37.9	11.2	**70.1**	65.7	-
DB40rāga	M	F_1	69.6	61.3	55.9	54.6	-
		F_2	**69.6**	61.7	55.7	54.3	-
		F_3	69.5	61.5	55.9	54.5	-
	S_1	PCD_{120}	-	-	-	-	66.4
		PCD_{full}	-	-	-	-	**74.1**
	S_2	PD_{param}	20.8	2.6	**51.4**	44.2	-

Table 1. Accuracy (in percentage) of different methods (Mtd) for two datasets (db) using different classifiers and features (Ftr).

the evaluation measure. In order to assess if the difference in the performance of any two methods is statistically significant, we use the Mann-Whitney U test [31] with $p = 0.01$. In addition, to compensate for multiple comparisons, we apply the Holm-Bonferroni method [32].

3.3. Comparison with the state of the art

We compare our results with two state of the art methods proposed in [7] and [12]. As an input to these methods, we use the same predominant melody and tonic as used in our method. The method in [7] uses smoothened pitch-class distribution (PCD) as the tonal feature and employs 1-nearest neighbor classifier (1NN) using Bhattacharyya distance for predicting rāga label. We denote this method by S_1. The authors in [7] report a window size of 120 s as an optimal duration for computing PCDs (denoted here by PCD_{120}). However, we also experiment with PCDs computed over the entire audio recording (denoted here by PCD_{full}). Note that in [7] the authors do not experiment with a window size larger than 120 s.

The method proposed in [12] also uses features based on pitch distribution. However, unlike in [7], the authors use parameterized pitch distribution of individual svaras as features (denoted here by PD_{param}). We denote this method by S_2. The authors of both these papers courteously ran the experiments on our dataset using the original implementations of the methods.

4. RESULTS AND DISCUSSION

In Table 1, we present the results of our proposed method M and the two state of the art methods S_1 and S_2 for the two datasets DB10rāga and DB40rāga. The highest accuracy for every method is highlighted in bold for both the datasets. Due to lack of space we present results only for the best performing classifiers.

We start by analyzing the results of the variants of M. From Table 1, we see that the highest accuracy obtained by M for DB10rāga is 91.7%. Compared to DB10rāga, there is a significant drop in the performance of every variant of M for DB40rāga. The best performing variant in the latter achieves 69.6% accuracy. We also see that for both the datasets, the accuracy obtained by M across the feature sets is nearly the same for each classifier, with no statistically significant difference. This suggests that, considering just the presence or the absence of a melodic phrase, irrespective of its frequency of occurrence, is sufficient for rāga recognition. Interestingly, this finding is consistent with the fact that characteristic melodic phrases are
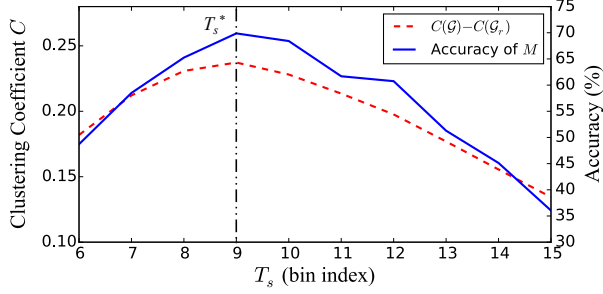
Fig. 3. Accuracy of M and $C(\mathcal{G}) - C(\mathcal{G}_r)$ for different similarity thresholds.

unique to a rāga and a single occurrence of such phrases is sufficient to identify the rāga [3]. As seen in Table 1, the performance of the proposed method is very sensitive to the choice of the classifier. We notice that for both the datasets, the best accuracy is obtained using the NBM classifier, and the difference in its performance compared to any other classifier is statistically significant. Note that, the NBM classifier outperforming other classifiers is also well recognized in the text classification community [33]. We, therefore, only consider the NBM classifier for comparing M with the other methods. It is worth noting that the feature weights assigned by a classifier can be used to identify the relevant melodic phrases for rāga recognition. These phrases can serve as a dictionary of semantically-meaningful melodic units for many computational tasks in IAM.

Before further analyses of our results, we verify here our approach to obtain the optimal similarity threshold $T_s^\star$ (Section 2.2). In Figure 3, we show the accuracy obtained by M, and $C(\mathcal{G}) - C(\mathcal{G}_r)$ as a function of similarity threshold T_s. We see that these curves are highly correlated. Thus, the optimal threshold $T_s^\star$, which we defined in Section 2.2 as the distance that maximizes the difference $C(\mathcal{G}) - C(\mathcal{G}_r)$, also results in the best accuracy for rāga recognition.

We now analyze the confusion matrix, and understand the type of classification errors made by M, (Figure 4). We observe that our method achieves near-perfect accuracy for several rāgas including Kalyāṇi, Pūrvikalyāṇi, Tōḍi and Varāḷi. This is consistent with the fact that these are considered to be phrase-based rāgas, that is, their identity is predominantly derived from melodic phraseology [3]. At the same time, we observe low accuracy for some other phrase-based rāgas such as Madhyamāvati, Kānaḍa and Śrī. On investigating further we find that such rāgas are confused often with their allied rāgas[5] [3]. Distinguishing between allied rāgas is a challenging task, since it is based on subtle melodic nuances. We also note that, among the other rāgas for which the obtained accuracy is low, several are considered as scale-based rāgas. This is in line with [3], where the authors remark that the identification of such rāgas is not based on melodic phraseology. Overall, this analysis of the classification errors indicates that our proposed method is more suitable for recognizing phrase-based rāgas compared to scale-based rāgas.

Finally, we compare M with the state of the art methods S_1 and S_2. From Table 1, we see that M outperforms S_2 for both the datasets, and the difference is found to be statistically significant. When compared with S_1, we see that M performs significantly better than the PCD_{120} variant of S_1 for both the datasets. However, the performance of the PCD_{full} variant of S_1 is comparable to M for DB10rāga, and, significantly better for DB40rāga. A comparison of the results of M and S_1 for each rāga reveals that their performance is complementary. M successfully recognizes several rāgas

---

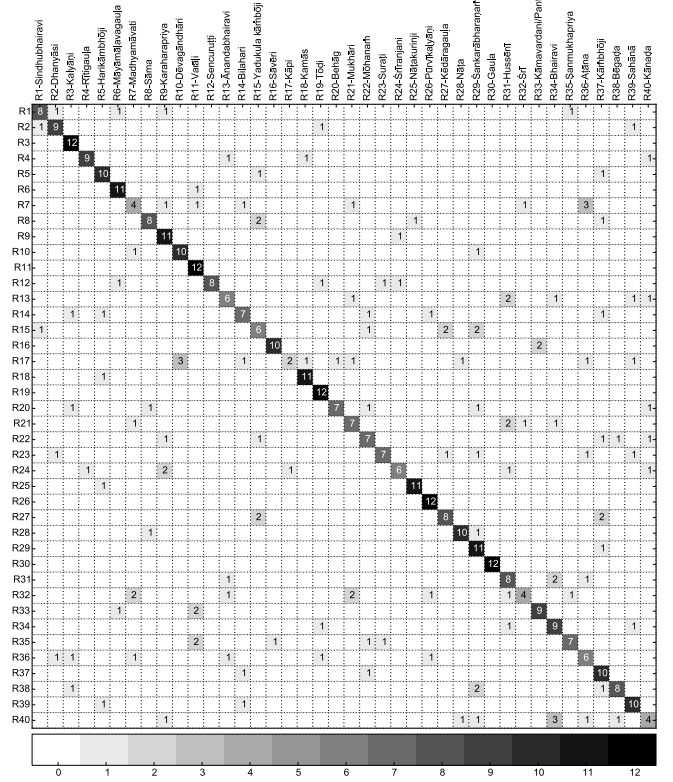[5]Allied rāgas have a common set of svaras and similar melodic movement

Fig. 4. Confusion matrix for the proposed method. The different shades of grey are mapped to different number of audio recordings.

with high accuracy for which S_1 performs poorly, and vice-versa. This suggests that the proposed phrase-based method can be combined with the pitch distribution-based methods to achieve a higher rāga recognition accuracy.

5. CONCLUSIONS

In this paper, we proposed a novel phrase-based approach to rāga recognition. To the best of our knowledge, no other method employs a fully automated methodology for discovery and selection of melodic phrases for rāga recognition. Melodic patterns are discovered in a collection of audio recordings using an unsupervised approach, and are clustered using the complex network concepts. A vector space model is employed to represent audio recordings using these melodic patterns. The features thus obtained are used to train a classifier. For evaluations we compiled a sizable representative Carnatic music collection, which we make publicly available. We experimented with a number of classification algorithms and found that the multinomial naive Bayes classifier outperforms the rest. Our results indicate that considering the mere presence or the absence of melodic phrases in audio recordings is sufficient for rāga recognition. Overall, we showed that phrase-based rāga recognition is a successful strategy, on par with the state of the art, and at times outperforming it. An analysis of the classification errors revealed that the type of errors made by the phrase-based and the pitch distribution-based methods are complementary. In the future, we plan to investigate if both these methodologies can be successfully combined to improve rāga recognition.

6. REFERENCES

[1] A. Danielou, *The ragas of Northern Indian music*, Munshiram Manoharlal Publishers, New Delhi, 2010.

[2] T. Viswanathan and M. H. Allen, *Music in South India*, Oxford University Press, 2004.

[3] T. M. Krishna and V. Ishwar, "Karṇāṭic music: Svara, gamaka, motif and rāga identity," in *Proc. of the 2nd CompMusic Workshop*, 2012, pp. 12–18.

[4] S. Gulati, J. Serrà, and X. Serra, "An evaluation of methodologies for melodic similarity in audio recordings of indian art music," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, 2015, pp. 678–682.

[5] P. Rao, J. C. Ross, K. K. Ganguli, V. Pandit, V. Ishwar, A. Bellur, and H. A. Murthy, "Classification of melodic motifs in raga music with time-series matching," *Journal of New Music Research*, vol. 43, no. 1, pp. 115–131, Jan. 2014.

[6] P. Chordia and A. Rae, "Raag recognition using pitch-class and pitch-class dyad distributions.," in *In Proc. of Int. Society for Music Information Retrieval Conf.*, 2007, pp. 431–436.

[7] P. Chordia and S. Şentürk, "Joint recognition of raag and tonic in north indian music," *Computer Music Journal*, vol. 37, no. 3, pp. 82–98, 2013.

[8] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Rāga recognition based on pitch distribution methods," *Journal of New Music Research*, vol. 41, no. 4, pp. 337–350, 2012.

[9] H. G. Ranjani, S. Arthi, and T. V. Sreenivas, "Carnatic music analysis: Shadja, swara identification and raga verification in alapana using stochastic models," in *IEEE Workshop on applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 29–32.

[10] P. Dighe, P. Agrawal, H. Karnick, S. Thota, and B. Raj, "Scale independent raga identification using chromagram patterns and swara based features," in *IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1–4.

[11] P. Dighe, H. Karnick, and B. Raj, "Swara histogram based structural analysis and identification of indian classical ragas.," in *In Proc. of Int. Society for Music Information Retrieval Conf. (ISMIR)*, 2013, pp. 35–40.

[12] G. K. Koduri, V. Ishwar, J. Serrà, and X. Serra, "Intonation analysis of rāgas in carnatic music," *Journal of New Music Research*, vol. 43, no. 1, pp. 72–93, 2014.

[13] V. Kumar, H Pandya, and C. V. Jawahar, "Identifying ragas in indian music," in *22nd Int. Conf. on Pattern Recognition (ICPR)*, 2014, pp. 767–772.

[14] R. Sridhar and T. V. Geetha, "Raga identification of carnatic music for music information retrieval," *International Journal of Recent Trends in Engineering*, vol. 1, no. 1, pp. 571–574, 2009.

[15] S. Shetty and K. K. Achary, "Raga mining of indian music by extracting arohana-avarohana pattern," *Int. Journal of Recent Trends in Engineering*, vol. 1, no. 1, pp. 362–366, 2009.

[16] P. V. Rajkumar, K. P. Saishankar, and M. John, "Identification of carnatic raagas using hidden markov models," in *IEEE 9th Int. Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2011, pp. 107–110.

[17] G. Pandey, C. Mishra, and P. Ipe, "Tansen: A system for automatic raga identification," in *In Proc. of the 1st Indian Int. Conf. on Artificial Intelligence*, 2003, pp. 1350–1363.

[18] S. Dutta, S. PV Krishnaraj, and H. A. Murthy, "Raga verification in carnatic music using longest common segment set," in *Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, Málaga, Spain, In Press.

[19] S. Gulati, J. Serrà, V. Ishwar, and X. Serra, "Mining melodic patterns in large audio collections of indian art music," in *Int. Conf. on Signal Image Technology & Internet Based Systems - MIRA*, Marrakesh, Morocco, 2014, pp. 264–271.

[20] S. Gulati, J. Serrà, and X. Serra, "Improving melodic similarity in indian art music using culture-specific melodic characteristics," in *Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, Málaga, Spain, 2015, pp. 680–686.

[21] M. EJ Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

[22] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, 2002.

[23] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, pp. P10008, 2008.

[24] S Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.

[25] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *In Proc. of the 7th Python in Science Conf.*, Pasadena, CA USA, Aug. 2008, pp. 11–15.

[26] X. Serra, "A multicultural approach to music information research," in *Proc. of Int. Conf. on Music Information Retrieval (ISMIR)*, 2011, pp. 151–156.

[27] Xavier Serra, "Creating research corpora for the computational study of music: the case of the Compmusic project," in *Proc. of the 53rd AES Int. Conf. on Semantic Audio*, London, 2014.

[28] A. Srinivasamurthy, G. K. Koduri, S. Gulati, V. Ishwar, and X. Serra, "Corpora for music information research in indian art music," in *Int. Computer Music Conf./Sound and Music Computing Conf.*, Athens, Greece, 2014, pp. 1029–1036.

[29] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, Springer, Berlin, Germany, 2nd edition, 2009.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[31] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[32] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[33] A. McCallum and K. Nigam, "A comparison of event models for naive bayes text classification," in *AAAI Workshop on learning for text categorization*, 1998, vol. 752, pp. 41–48.