IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **1** (2011), pp. 55–59

Cover song networks: analysis and accuracy increase

Joan Serrà[1,†], Massimiliano Zanin[2,†] and Perfecto Herrera[1]

[1] *Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*

[2] *INNAXIS Foundation & Research Institute, Madrid, Spain*

Abstract. The application of community detection in complex networks is explored within the framework of cover song identification, i.e. the automatic detection of different audio renditions of the same underlying musical piece. In the last years this particular task has been widely studied within the music information retrieval field as a query problem, where one song was submitted and a list of possible matches was created by the system. In this contribution we propose a new point of view: songs are embedded in a complex weighted network, whose links represent similarity (common musical content between songs). We analyze this network and find a strong modular structure, with well-defined communities and a clustering coefficient higher than expected. We then perform clustering and community detection to identify groups of songs that are versions the same musical piece. Importantly, the information gained through this process can be used to increase the overall accuracy of the system. Results show that accuracy increments of 5 percent points can be easily achieved. A further out-of-sample test provides evidence that this increase can be potentially higher.

Keywords: networks, community detection, information retrieval, cover songs
MSC 2000: 00A65, 05C90, 68P20, 68R10, 68U35, 90C35, 97M80, 97R50

1. Introduction

Audio cover song identification is the task of automatically detecting which songs are versions of the same underlying musical piece using only information extracted from the raw audio signal of each of them [1]. Cover song identification has been a very active area of study within the music information retrieval (MIR) community over the last years, and this is clearly due to the revolution which has lashed this field, intrinsically related to the introduction of digital ways to share and distribute information [2, 3, 4].

Traditionally, cover song identification has been set up as a typical information retrieval (IR) task where queries are processed in a batch mode [5]: the user submits a query (a song) and receives an answer back (a list of songs ranked by their relevance to the query; in this case a list of potential covers).

Therefore, the majority of efforts have been put in achieving a metric that faithfully captures cover song relationships [1]. However, a new approach can be considered: instead of isolated songs, systems may focus on groups of songs, with the new aim of identifying coherent sets of covers within a given music collection.

The reader may easily see the resemblance between the detection of cover sets and a more classical community detection task inside a complex network [6, 7]. This way, a set of nodes $\mathscr{N} \equiv \{n_1, n_2, \ldots, n_N\}$ represents the N recordings being analyzed, and the elements of the $N \times N$ weight matrix $\mathscr{W}$ represent the distance (dissimilarity) between any couple of nodes. Provided that the weights of this matrix are assigned with the help of a suitable cover song dissimilarity metric (e.g. the same one used to originally retrieve answers to a query), communities inside this complex network will represent sets of recordings with equivalent musical content.

2. The cover song network

The first step required by our proposal is to create a network and to embed nodes (songs) into it. We use an in-house music collection of 2125 songs comprising a variety of genres and styles. This collection is an extension of the one used in [8] and consists on 525 non-overlapping groups of cover songs.

Links between network's nodes should represent the cover song relationship between corresponding musical pieces (the dissimilarity between their musical content). Therefore, an algorithm to compute this dissimilarity is needed in order to calculate the elements $w_{i,j}$ of the matrix $\mathscr{W}$ for each couple of nodes n_i and n_j. Several alternatives for such dissimilarity measures have been proposed [1]. One of the most promising strategies is the $Q_{\max}$ measure presented in [8]. This measure allows to track curved and potentially disrupted traces in cross recurrence plots constructed from the state space representation of two time series. These time series might represent a certain musical aspect and are automatically extracted from the recordings.

The symmetric measure $Q_{\max}$ represents similarity. As in [9], this metric is converted to a dissimilarity measure to fill $\mathscr{W}$ following

$$w_{i,j} = \frac{\sqrt{|s_j|}}{Q_{\max}(s_i, s_j)}, \tag{1}$$

where $|s_j|$ is proportional to the duration of song s_j and $Q_{\max}(s_i, s_j) \in [1, \max(|s_i|, |s_j|)]$.

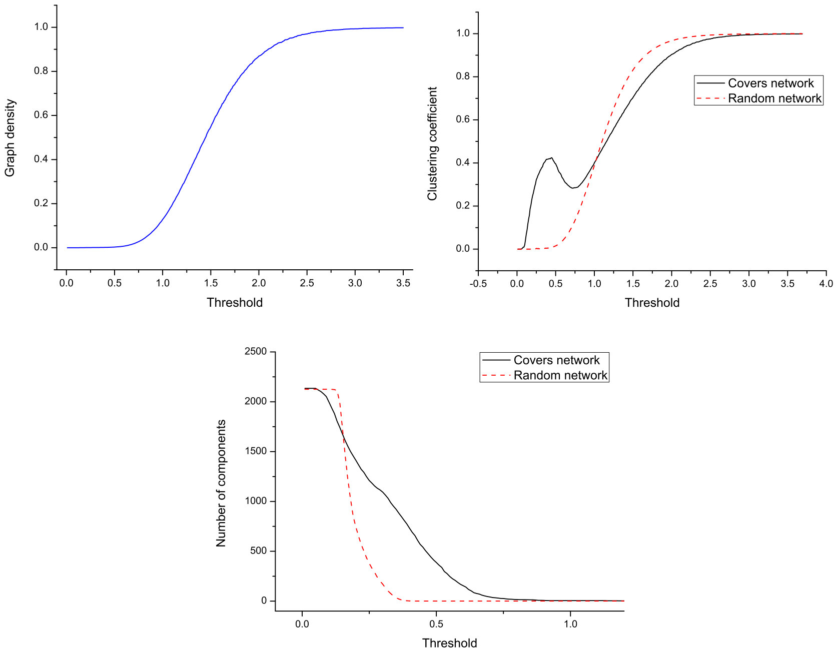

Figure 1: Evolution of 3 metrics of the network as a function of the applied threshold. These metrics are, from left to right: graph density, clustering coefficient, and number of independent components. Red dashed lines show the expected value in a random network with the same number of nodes and links.

3. Analysis of the network

The result of the previous procedure over all available data is a weighted directed graph expressing cover song relationships . In order to understand how the network evolves when the threshold is modified, we study a number of different classical metrics as a function of the threshold. These are [6]: graph density, number of independent components, size of the strong giant component, number of isolated nodes, efficiency [10], and clustering coefficient. Three of them are depicted in Fig. 1.

Importantly, we see that when reducing the threshold, the network splits into an higher number of clusters than expected, which represents the formation of cover song communities. When these communities are formed, they maintain a high clustering coefficient: that is, sub-networks of covers tend to be fully connected and maintain a triangular coherence. This process begins around a threshold of 0.5 (we could confirm it by the evolution of the size of

	SL	CL	AL	WL	MO	PM1	PM3
rppi	2.05	1.28	4.37	3.60	5.54	5.49	4.73

Table 1: Relative accuracy increase (rppi) for different clustering and community detection algorithms. First row corresponds to the acronyms of the different algorithms we use [11].

the strong giant component). It is also interesting to note that the number of isolated nodes remained lower than expected, except for high thresholds. This suggests that most of the songs are connected to some cluster, while a small group of them are different, with unique musical features.

4. Accuracy increase

Adding the information obtained through the detection of cover song communities to the final cover song identification system can lead to a certain accuracy increase. This way, $\mathscr{W}$ is updated with the community assignments so that elements detected to belong to the same community are close to each other. To show this we apply a number of clustering and community detection techniques [11] to our data and refine $\mathscr{W}$ accordingly. We compute the mean of average precisions measure [11], a standard IR evaluation metric, before and after updating $\mathscr{W}$. We subsequently calculate a relative percentual point increment (rppi). Results are shown in Table 1. Best algorithms were found to be the ones designed to exploit a complex network collaborative approach, either from a point of view of modularity optimization (MO) or connected components criterion (PM1) [11].

An out-of-sample test was done within the MIREX[1] audio cover song identification contest [3]. In the editions of 2008 and 2009 we submitted the same two versions (with and without a refinement of $\mathscr{W}$) of our system [9], which yielded 13.64 rppi. This increment is substantially higher than the ones we achieve with our data, most probably because the setup for the MIREX task includes more members per community than ours. In particular, when big communities are considered, one can think of the idea presented here to achieve more dramatic impacts in final system's accuracies.

[1]MIR evaluation exchange: `http://www.music-ir.org/mirexwiki/index.php/Main_Page`

5. Conclusion

We build and analyze a musical network that reflects cover song communities, where nodes correspond to different audio recordings and links between them represent a measure of resemblance between their musical content. We find that this network has a strong modular structure. We furthermore provide evidence that the knowledge acquired through community detection is valuable in improving the raw results of a query-based cover song identification system.

Acknowledgments

Joan Serrà has been partially funded by the Music 3.0 project (TSI-070100-2008-318) of the Spanish Ministry of Industry, Tourism, and Trade.

References

[1] J. Serrà, E. Gómez, & P. Herrera, *In Advances in Music Information Retrieval*. Springer-Verlag, vol. **274**, ch. 14, pp. 307–332, (2010).

[2] N. Orio, *Found. and Trends in Information Retrieval* **1** (1), pp. 1–90, (2006).

[3] J. S. Downie, *Acoustical Science and Technology* **29** (4), pp. 247–255, (2008).

[4] M. Casey, R. C. Veltkamp, M. Goto, M. Leman, C. Rhodes, & M. Slaney, *Proc. of the IEEE* **4** (96), pp. 668–696, (2008).

[5] R. Baeza-Yates & B. Ribeiro-Neto, *Modern information retrieval*, ACM Press Books, (1999).

[6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, & D. U. Hwang, *Phys. Rep.* **424** (4-5), pp. 175–308, (2006).

[7] L. da F. Costa, O. N. Oliveira Jr., G. Travieso, F. A. Rodrigues, P. R. V. Boas, L. Antiqueira, M. P. Viana, & L. E. Correa da Ocha, *Working manuscript*, arXiv:0711.3199v2, (2008).

[8] J. Serrà, X. Serra, & R. G. Andrzejak, *New Journal of Physics* **11**, 093017, (2009).

[9] J. Serrà, M. Zanin, & R. G. Andrzejak, *Music Information Retrieval Evaluation eXchange (MIREX) extended abstract*, (2009).

[10] V. Latora & M. Marchiori, *Phys. Rev. Lett.* **87**, 198701, (2001).

[11] J. Serrà, M. Zanin, C. Laurier, & M. Sordo, *Conf. of the Int. Soc. for Music Information Retrieval (ISMIR)*, pp. 225-230, (2009).