UNCOVERING AFFINITY OF ARTISTS TO MULTIPLE GENRES FROM SOCIAL BEHAVIOUR DATA

Claudio Baccigalupo **Enric Plaza**
IIIA, Artificial Intelligence Research Institute
CSIC, Spanish Council for Scientific Research
{claudio,enric}@iiia.csic.es

Justin Donaldson
Indiana University
School of Informatics
jjdonald@indiana.edu

ABSTRACT

In organisation schemes, musical artists are commonly identified with a unique 'genre' label attached, even when they have affinity to multiple genres. To uncover this hidden cultural awareness about multi-genre affinity, we present a new model based on the analysis of the way in which a community of users organise artists and genres in playlists.

Our work is based on a novel dataset that we have elaborated identifying the co-occurrences of artists in the playlists shared by the members of a popular Web-based community, and that is made publicly available. The analysis defines an automatic social-based method to uncover relationships between artists and genres, and introduces a series of novel concepts that characterises artists and genres in a richer way than a unique 'genre' label would do.

1 INTRODUCTION

Musical genres have typically been used as a 'high level' classification of songs and artists. However, many musical artists cannot be described with a single genre label. Artists can change style throughout their career, and perform songs that do not entirely fall into single categories such as 'Rock', 'R&B' or 'Jazz'. In general terms, artists can be said to have a certain degree of *affinity* to each specific genre.

In this paper we propose to model relationships from artists to genres as *fuzzy sets*, where the *membership degree* indicates the affinity of an artist to a genre. This is in contrast to conventional labelling approaches, in which artists either belong or do not belong to a genre, and allows for a sentence like "Madonna is Pop and R&B, not Jazz" to be rephrased as "Madonna belongs with a membership degree of 0.8 to Pop genre, with 0.6 to R&B and with 0.1 to Jazz".

To uncover the affinity of different artists to different genres, we follow a *social* approach: we exploit information about how people aggregate and organise artists and genres in *playlists*. Playlists are the predominant form of casual music organisation. When playlists are collected in aggregate over a large sampling of the music listening populace, they reflect current cultural sensibilities for the association of popular music. Our assumption about *large repositories* of playlists is that when two artists or genres co-occur together and closely in many playlists, they have some form of shared cultural *affinity*. For example, if we observe that a large number of songs by an artist x co-occur often and closely with songs whose artists are classified in the 'Jazz' genre, then we argue that x has a certain affinity with Jazz, independently from the original 'genre' label attached to x.

On the basis of this assumption, we build a model of *multi-genre affinity* for artists, grounded on data about the *usage* of music objects by real people. We suggest this kind of analysis can provide results which are complementary to those provided by other content-based or social techniques, which we discuss in the following subsection.

In this paper, we first discuss different approaches towards musical genre analysis, and provide arguments for the consideration of a community-based approach. Next, we report on the analysis we performed on a novel dataset, which lists the co-occurrences of 4,000 artists in a large collection of playlists, to uncover associations between artists and genres. Finally, we depict possible applications, such as affinity-based generation of 'smooth sequences' of songs.

1.1 Related work about genre analysis

Genre analysis has been a constant theme in the MIR community, and researchers have had some success in classifying or predicting genre using either *content-analytic* approaches or *social behaviour* approaches.

Content-analytic methods are broken into two distinct approaches. The 'acoustic-analytic' approach was introduced by Tzanetakis and Cook [9], and investigates the acoustic signature of music according to a battery of relevant cognitive and musicological descriptors, such as beats per minute, timbre, and frequency spectrum. The 'score-analytic' approach concerns itself with symbolic representations of music as sequences of notes and events making up a musical score, generally as MIDI files, as in McKay and Fujinaga [6].

Social behaviour approaches classify music objects

This research is partially supported by the MID-CBR (TIN2006-15140-C03-01) project, by the Generalitat de Catalunya under the grant 2005-SGR-00093, and by a MyStrands scholarship.

based upon the actions that a community of people perform with them. A large amount of effort has been directed towards harvesting textual correlations (of names of artists, songs, etc.) that co-occur on public Web sites. This has been the focus of Schedl et al. [8], who analysed artist-based term co-occurrences, Knees et al. [4], and Whitman and Smaragdis [10]. Text-based Web mining, however, can suffer from a lack of precision, mistaking band name terms with non-associated content as mentioned in [8].

Other attempts at providing richer representations for musical artists include describing an artist with a set of tags [2], with a set of moods [3] or identifying term profiles that characterise a specific cluster of similar artists [4]. All of these techniques maintain a 'Boolean' approach, identifying whether each artist matches specific tags, moods or clusters, while we seek to identify 'fuzzy' membership values relating each artist to each genre.

The social-based approach that we follow is the analysis of co-occurrences in playlists, which is currently not a common method of analysis. This is because playlists are not often publicly indexed or made available. Previously, a smaller public collection of about 29,000 playlists was used by Logan et. al [5] to analyse a set of 400 popular artists, and by Cano and Koppenberger [1] to provide a study of artists networks, while we work on a novel and much larger dataset of 1,030,068 human-compiled playlists.

Playlists have a special utility for social analysis of genres in that they are *specific* to music, with no extraneous "noise" information such as on arbitrary Web pages. They are also *intentional* in their creation, as opposed to a simple record of an individual's play history which may have been generated randomly.

2 A DATASET OF ARTISTS CO-OCCURRENCES

MusicStrands (http://music.strands.com) is a Web-based music recommendation community which allows members to publish playlists from their personal media players, such as Apple iTunes and Windows Media Player. Every published playlist is made of a sequence of IDs that univocally identify the songs and artists that the playlist contains.

On July 31st, 2007, we gathered the 1,030,068 user playlists published so far. We discarded playlists made by a single song or by a single artist. We also removed any information related to the creators (user name, creation date, playlist title, rating) to focus on the sequential nature of playlists in terms of co-occurrences of artists.

For each pair of artists (x, y), we extracted the number of times that a song by x occurs together with a song by y, and at what *distance* (i.e., how many other songs are in between). To reduce the dimension of the dataset, we considered only the co-occurrences of songs from the 4,000 most popular artists, where the popularity of an artist equals the number of playlists in which occurs. We also excluded non-specific popular artist labels such as "Various Artists" or "Original Soundtrack", and any co-occurrence with a distance larger than 2, to obtain the dataset sketched in Table 1, where each record contains: the ID of the first artist x, the ID of the second artist y, and the numbers $d_0(x, y)$, $d_1(x, y)$, and $d_2(x, y)$ of playlists where y co-occurs after x at distance 0, 1, and 2 respectively.

x	y	$d_0(x, y)$	$d_1(x, y)$	$d_2(x, y)$
11	11	137	96	77
11	27	0	0	1
11	91	0	1	2
...	...	...	...	...

Table 1. A portion of the provided dataset

An auxiliary table in the dataset lists the 4,000 artists IDs, together with their genre IDs and their Musicbrainz artist IDs. Both artist IDs and genre IDs can be easily translated to actual names using the OpenStrands developers API. [1]

Thanks to the support of MusicStrands, we now make public the complete dataset that we have elaborated, which is freely available to researchers wishing to perform social based analysis on music behaviour data from real users. [2]

3 GENRES AND AFFINITY

The artists in the previously described dataset already have a genre label attached. However, our motivation is to provide a richer representation, which overcomes the limitation of having only one genre per artist.

We denote with $\mathcal{G}$ the set of m genres, and with $\mathcal{A}$ the set of n artists in the dataset, where $m = 26$ and n = 4,000. Our goal is to describe each artist x as a vector $[M_x(g_1), M_x(g_2), \ldots, M_x(g_m)]$, where each value $M_x(g_i) \in [0, 1]$ indicates how much x has affinity to the genre g_i. We call $M_x(g_i)$ the **genre affinity degree** of artist x to genre g_i.

The process to calculate the genre affinity degrees is the following. First we measure the association from each artist $x \in \mathcal{A}$ to each other artist, based on their co-occurrences in the authored playlists. Then we aggregate these associations *by genre* to obtain the degree in which each artist x is associated with each genre $g \in \mathcal{G}$. Finally we combine artist-to-artist and artist-to-genre associations to measure the degree $M_x(g)$ of affinity between artist x and genre g.

The first step is to consider how much an artist x is associated with *any other* artist y. Let $\mathcal{X}$ be set of $n - 1$ artists other than x: $\mathcal{X} = \mathcal{A} - \{x\}$. We define the **artist-to-artist** association $A_x : \mathcal{X} \to \mathbb{R}$ as an aggregation of the number

[1] API available at http://my.strands.com/openstrands

[2] Dataset available at http://labs.strands.com/music/affinity

of co-occurrences of x with any artist $y \in \mathcal{X}$:

$$\begin{aligned} A_x(y) = \quad & \alpha \cdot [d_0(x,y) + d_0(y,x)] \\ + \quad & \beta \cdot [d_1(x,y) + d_1(y,x)] \\ + \quad & \delta \cdot [d_2(x,y) + d_2(y,x)] \,. \end{aligned}$$

This metric aggregates the number of playlists in which artists x and y co-occur, with different weights according to the *distance* at which x and y occur. Co-occurrences at distance 0 (d_0), distance 1 (d_1), and distance 2 (d_2) are weighted with α, β and δ respectively. These parameters, with values in $[0,1]$, can be tuned to reflect the importance that the authors of the playlists assign to the distance between associated artists in their playlists. In the case of MusicStrands, we have observed that some authors compile playlists grouping together associated artists *one after the other* (for example, DJs for dance playlists), while most users create playlists as *unordered sets* of associated songs, where the distance between two artists is not related to their association. Based on this mixed behaviour, in our analysis, we have assigned values of $\alpha = 1$, $\beta = 0.8$, and $\delta = 0.64$.

According to the metric defined, the values of A_x are heavily influenced by the popularity of x. In fact, an artist x who is present in many playlists co-occurs with many artists, and has a generally higher artist-to-artist association than an uncommon artist. Therefore, we perform a normalisation and obtain an artist-to-artist association $\widehat{A}_x : \mathcal{A} \rightarrow [-1,1]$ independent from the popularity of x, such that:

$$\widehat{A_x}(y) = \frac{A_x(y) - \overline{A_x}}{|max(A_x(y) - \overline{A_x})|}$$

where $\overline{A_x} = \frac{1}{n-1} \sum_{y \in \mathcal{X}} A_x(y)$, and $\widehat{A}_x(x) = 0$ by convention. A positive (resp. negative) value $\widehat{A_x}(y)$ indicates that the artist-to-artist association of x with y is above (resp. below) the average artist-to-artist association of x.

Now we turn to estimate the affinity of artists with respect to genres. First, let us consider how much an artist x is associated with any genre g. We define the association $P_x : \mathcal{G} \rightarrow \mathbb{R}$ by cumulating the artist-to-artist association A_x from x to *any artist of genre g*:

$$P_x(g) = \sum_{y \in \mathcal{X} : \gamma(y) = g} A_x(y)$$

where $\gamma(y)$ denotes the genre of y. Again, we normalise this association to counter-effect the uneven distribution of genres in the playlists (e.g., Rock/Pop artists occur in almost any playlist, while Folk artists are quite rare). The result is a normalised association $\widehat{P_x}(g) : \mathcal{G} \rightarrow [0,1]$, independent from the popularity of genre g, defined as:

$$\widehat{P_x}(g) = \frac{\sum_{y \in \mathcal{X} : \gamma(y) = g} A_x(y)}{\sum_{y \in \mathcal{X}} A_x(y)} \,.$$

This function $\widehat{P}$ measures the degree in which artists associated with artist x belong to the genre g. Finally we obtain the **genre affinity degree** $M_x : \mathcal{G} \rightarrow [0,1]$ by weighting the association $\widehat{P}$ with the artist-to-artist association $\widehat{A}$, and normalising the result to the range $[0,1]$:

$$M_x(g) = \frac{1}{2} \left(\frac{\sum_{y \in \mathcal{A}} \widehat{A_x}(y) \widehat{P_y}(g)}{n} + 1 \right) .$$

Notice that $M_x(g)$ is high when artists that often co-occur with x belong to genre g and when artists that rarely co-occur with x do not belong to genre g. Each artist can be characterised by a **genre affinity vector** $[M_x(g_1), M_x(g_2), \ldots, M_x(g_m)] \; \forall g_i \in \mathcal{G}$, which is a multidimensional representation of an artist in terms of genres.

For example, Table 2 shows , for three famous artists of the dataset labelled as 'Rock/Pop' (Madonna, Metallica and Bruce Springsteen), their artist-to-artist association $\widehat{A_x}(y)$, their association $\widehat{P_x}(g)$ and their genre affinity degrees $M_x(g)$ with respect to three popular genres (Rock/Pop, Country and R&B).

The 'Rock/Pop' genre is very common in our data (labelling 2,286 of the 4,000 artists in the dataset), so having these three artists generically labelled as 'Rock/Pop' is not very informative of their style. However, by observing the differences in their genre affinity degrees $M_x(g)$, we can spot their differences. For instance, Bruce Springsteen has a higher genre affinity to Country, while Madonna has a higher affinity to R&B. By plotting their $M_x(g)$ values with respect to these three genres on a graph, we can visually observe such difference in Fig. 1.

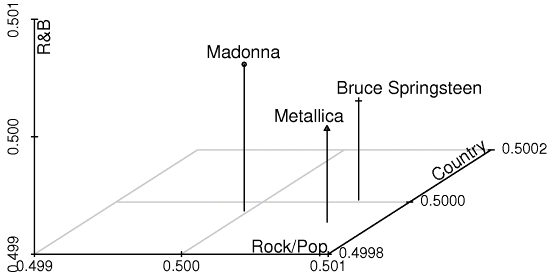

Figure 1. Genre affinity of 3 artists with 3 genres

4 ANALYSIS AND DISCUSSION

4.1 Genres and core artists

One advantage of describing artists in terms of affinity vectors, rather than with unique genre labels, is that we can identify how *central* an artist is to a genre. For each artist x, we define its **genre-centrality** $C_g : \mathcal{A} \rightarrow [0,100]$ to a genre $g \in \mathcal{G}$ as the percentage of artists whose genre affinity to g

$\widehat{A}_x(y)$	Madonna	Metallica	Bruce Spr.
Mad.	0	0.0026	0.0516
Met.	0.0322	0	0.0106
B. Spr.	0.0602	0.0063	0

$\widehat{P}_x(g)$	Rock	Country	R&B
Mad.	0.628	0.028	0.065
Met.	0.892	0.013	0.034
B. Spr.	0.790	0.041	0.026

$M_x(g)$	Rock	Country	R&B
Mad.	0.49997	0.49998	**0.50007**
Met.	**0.50040**	0.49995	0.49997
B. Spr.	0.50038	**0.50001**	0.49991

Table 2. Associations and affinities between artists and genres, shown only for a subset of 3 artists and 3 genres

is lower or equal than the genre affinity of x to genre g:

$$C_g(x) = \frac{100}{n} \cdot card\,\{y \in \mathcal{A} : M_y(g) \leq M_x(g)\}\ .$$

For instance, the *Country*-centrality of Bruce Springsteen is 79.3%, because as many as 3,172 out of the 4,000 artists in the dataset have an affinity degree to Country lower than his, while his *R&B*-centrality is 39.5%, since only 1,580 artists have a genre affinity degree to R&B lower than his.

We will call those artists which are the most central to a genre g the **core artists** of g. Core artists are good representatives of a genre, since they occur very often in the set of playlists with artists from that genre and seldom with artists from other genres. For instance, we interestingly observe that the top core artists for the genre labelled 'Soundtrack' are James Horner, Alan Silvestri and Michael Giacchino, who are famous composers of original movie scores (e.g., James Horner's *"Titanic Original Soundtrack"*), and not Pop artists who have only sporadically performed famous songs which appeared in movies (e.g., Celine Dion's *"My Heart Will Go On"*).

The top core artists for other three genres are: (Folk) Luka Bloom, Greg Brown, The Chieftains; (Blues) Muddy Waters, Taj Mahal, Dr. John; (Jazz) Bob James, Donald Byrd, Peter White; (Electronic) Kaskade, Sasha, Junkie XL.

4.2 Centrality and cross-genre artists

Table 3 shows four artists originally labelled as Electronic and their genre-centrality with respect to four genres. From this table, we learn that St. Germain has a high *Jazz*-centrality (94%), Soul II Soul have a high *R&B*-centrality (94%) and M83 have a high *Rock/Pop*-centrality (95%). This is not surprising, since St. Germain is "a French electronica and *nu jazz* musician", [3] Soul II Soul is a "U.K. *R&B* collective", [4] and M83 offer "a luscious blend of shoegaze aesthetics, ambient *pop*, and progressive textures". [5] This is an example of how, using genre-centrality, we can obtain richer descriptions of the affinity of artists to multiple genres, without employing expert knowledge, but analysing the way in which each artists' works were *employed* by real users in playlists, in combination with other artists.

Table 3 also shows that Cameron Mizell, originally labelled as Electronic, has a higher genre-centrality to Jazz (99%) than to Electronic (90%). The reason is that Cameron Mizell is a cross-genre artist, whose creations mostly fall in the Electronic category, and mostly occur together with Jazz songs. Indeed, Cameron Mizell is labelled as 'Electronic' in MusicStrands, but as 'Jazz' according to other external musical sources, such as AMG (http://allmusic.com), iTunes Music Store (http://apple.com/itunes) and Amazon (http://amazon.com). This is not such a strange case, for several artists exist which behave as a sort of *musical bridge* from a genre to another. Formally, we call **bridge artist** any artist x whose original genre label $\gamma(x)$ differs from the genre $\phi(x)$ for which x has the highest genre-centrality, where $\phi(x) = g \in \mathcal{G} : C_g(x) \geq C_h(x)\ \forall h \in \mathcal{G}$.

$C_g(x)$	St. Germain	Soul II Soul	M83	C.Mizell
Electronic	98%	96%	98%	90%
Jazz	**94%**	10%	6%	**99%**
R&B	41%	**94%**	8%	62%
Rock/Pop	37%	1%	**95%**	17%

Table 3. Genre-centrality of 4 artists to 3 genres

Table 4 reports the number of bridge artists, aggregated by genre, for six of the 26 genres (with the main diagonal containing artists which are *not* bridge artists). According to our data, the distribution of bridge artists is uneven and depends on genres; for instance there are no bridge artists from Reggae to Country and vice versa, while there are 90 bridge artists from Rap to R&B (37% of all the Rap artists). This suggests that genres like Country and Reggae are "distinct", in the sense that artists from the two genres tend not to "play well together", while genres like R&B and Rap are "adjoining", meaning that artists from the two genres tend to "play well together". These abstract concepts are defined more precisely hereafter, with different typologies of associations between genres characterised.

$g \setminus h$	Country	Blues	Jazz	Reggae	R&B	Rap
Country	162	-	1	-	-	-
Blues	-	3	-	-	-	-
Jazz	-	2	101	-	-	-
Reggae	-	-	-	24	-	-
R&B	1	1	9	6	334	11
Rap	-	-	-	9	90	226

Table 4. Number of artists labelled as genre g, whose genre-centrality is maximum for genre h, for 6 different genres

[3] Excerpt from http://last.fm/music/St.+Germain

[4] Excerpt from http://music.strands.com/artist/15246

[5] Excerpt from http://music.aol.com/artist/m83/1655884

4.3 Correlation between two genres

After having analysed associations from artists to genres, we draw some observations about *how genres relate* with each other. The goal is to discern for each pair of genres (g, h) whether they are or not correlated. The idea is that two genres are correlated when artists belonging to one genre naturally tend to be played together with artists of the other genre. Using the genre affinity degree $M_x(g)$, we can more precisely state that two genres g and h are **correlated** (resp. **exclusive**) when artists with a high $M_x(g)$ (genre affinity degree to g) tend to have a high (resp. low) $M_x(h)$ (genre affinity degree to h); and two genres are **independent** when their affinity degrees do not show any sign of correlation.

As an example, consider Figure 2 (above) showing, for each of the 4,000 artists in the dataset, its affinity degree to Country (x axis), to Reggae (y axis), and its original genre label (point type). Artists with high affinity to Country (in the lower right section) tend to have a 'neutral' affinity to Reggae (with most values close to 0.5). Similarly, artists with high affinity to Reggae (in the upper left section) tend to have a 'neutral' affinity to Country. This is an example of independent genres, since the genre affinity degree to Country and to Reggae of an artist look unrelated. Figure 2 (below) shows instead that artists with high affinity to R&B often have a high affinity to Rap as well (in the upper right section), suggesting that the two genres are correlated.

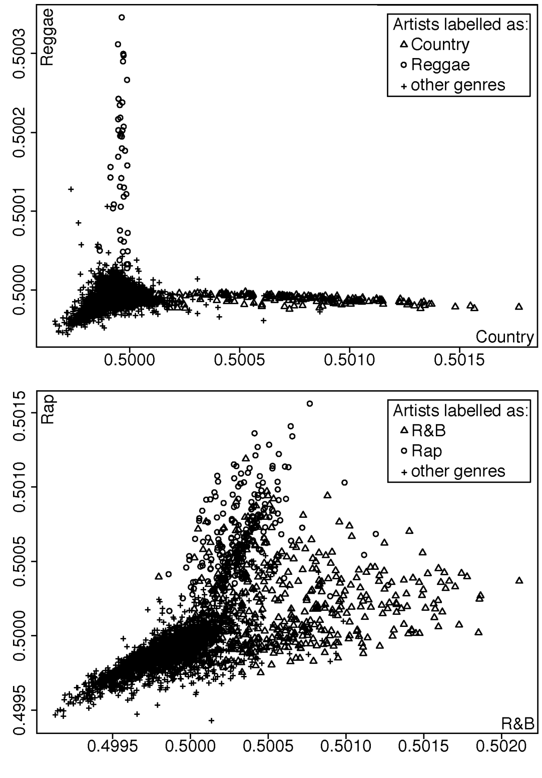

Figure 2. Comparing *Country* vs. *Reggae* (above), and *R&B* vs. *Rap* (below) genre-centralities

To measure the correlation between any two genres g and h, we employ Pearson's coefficient $\rho_{g,h} \in [-1, 1]$, which assesses the linear correlation between the affinity degree of artists from two genres, and has values close to 1 in the case of a positive relationship (i.e., artists with high affinity to g also have high affinity to genre h), and close to -1 in the case of a negative relationship (i.e., artists with high affinity to g have low affinity to genre h). Table 5 reports these coefficients for six genres, showing for instance that R&B and Rap (0.6) are positively correlated (as suggested by Fig. 2), as well as Jazz and Blues (0.4), while Blues and Rap (-0.2) are negatively correlated: artists with a high affinity to Blues tend to have a low affinity to Rap.

This social analysis of genres exposes cultural correlations between music that is often not present in an acoustic or score analysis. Most importantly, this analysis does not require expert human knowledge of genre relationships. Instead, relationships emerge from a larger cultural consensus.

$\rho_{g,h}$	Country	Blues	Jazz	Reggae	R&B	Rap
Country	1	0.2	0.1	0	0.1	-0.1
Blues	0.2	1	**0.4**	0.1	0	**-0.2**
Jazz	0.1	0.4	1	0.1	0.2	-0.1
Reggae	0	0.1	0.1	1	**0.4**	**0.4**
R&B	0.1	0	0.2	0.4	1	**0.6**
Rap	-0.1	-0.2	-0.1	0.4	0.6	1

Table 5. Pearson coefficient ρ for 6 genre-centrality vectors

5 APPLICATIONS

The analysis we have performed allows to describe artists in terms of affinity vectors and to uncover relations among genres. This knowledge is useful for many applications, some of which are presented hereafter.

First, consider the case of a Web radio, with a large repository of available music, and with the need to schedule different channels with different sequences of songs. The most common approach is to programme each channel with a randomly ordered selection of songs from the repository, matching the definition of each channel (e.g., a random sequence of jazz songs is played on a 'Jazz channel'). This method may work well for channels limited to a single genre, but may sound inadequate when multiple genres are permitted (e.g., in a channel generically defined as "Music from the '80s"), since unpleasant musical transitions could occur from one genre to another (e.g., a song by Abba, followed by a song by Metallica, and by a song by B.B. King). To procure *smooth* transitions from a song to the next one, the knowledge we have uncovered about affinity of each artist to each genre can be exploited, to generate better musical sequences that move, for example, from the core artists of 'Rock', to less central and cross-genre artists

('Rock'/'R&B'), to core artists of 'R&B', and so on, avoiding disturbing disruptions in the path. The same approach can work to generate playlists.

Another possible application of our analysis is to perform similarity assessments among artists. Having artists described as affinity vectors, we can obtain a measure of how close two artists are, independently from their 'genre' label. For instance, we can spot for each artist its nearest neighbour using Euclidean distance among these vectors. The results are interesting: in most cases we find that the nearest neighbour artists have the same genre label attached (e.g., Donna Summer goes to Diana Ross, both R&B), while sometimes they are differently labelled but still culturally associated, for example Nelly Furtado (Rock/Pop) goes to Missy Elliott (Rap), and Bob Sinclair (R&B) goes to Sonique (Rock/Pop). In this way, we exploit affinity fuzzy values to uncover associated artists, independently from their 'genre' label.

A third possible application is the creation of user profiles in music recommender systems. Rather than asking a user which musical genres she prefers, the system could ask her to name a few of her favourite artists, and then locate them on a multi-dimensional map of affinity degrees, to identify which styles of music the user seems to favour.

6 CONCLUSIONS

Music is a complex form of information that can be addressed both as an acoustic and a cultural phenomenon. Genre labels serve as simple and easily communicated classifications of musical style, but it can be shown that popular music often does not avail itself to such simple classifications. Musicians can share affinity with different genres, although in most organisational schemes (such as record stores) they have a unique genre label attached.

Our first contribution is providing a way to automatically uncover a richer relationship between artists and genres, based on social data in the form of playlists. Our proposal is moving from a Boolean concept of genres as exclusive categories to a view where artists have a degree of affinity with respect to each genre, and thus genres are conceptualised as fuzzy categories where affinity is represented as membership degrees in fuzzy set theory.

Our second contribution is developing a richer ontology for describing artist-to-genre and genre-to-genre associations, defining new concepts useful to better organise and understand large collections of music metadata. To enrich the characterisation of artists based on our first contribution (moving from Boolean categories to the notion of genre affinity), the ontology introduces the concepts of genre affinity degree, genre affinity vector, and genre centrality. To better characterise genres, the ontology introduces the concepts of core artists of a genre, bridge artists, and correlation or independence among genres.

To promote further social-based music analysis, we make publicly available the dataset of artists co-occurrences that we have produced from a very large repository of playlists. Since we identify each artist with its unique Musicbrainz ID, researchers have the chance to mash-up this dataset with a large number of other inter-linked music-related semantic Web datasets, following the methodology described in [7].

Future work includes expanding this approach to elucidate relationship between *songs*, *tags* and *genres*, viewing tags as user-provided fuzzy categories to determine the affinity of artists to tags, the core tags of genres and artists, the affinity and centrality of songs to genres, the relationship of songs with their artists, and the relationships among tags.

7 REFERENCES

[1] P. Cano and M. Koppenberger. The emergence of complex network patterns in music artist networks. In *Proc. of ISMIR*, pages 466–469, 2004.

[2] D. Eck, T. Bertin-Mahieux, and P. Lamere. Autotagging music using supervised machine learning. In *Proc. of ISMIR*, pages 367–368, 2007.

[3] X. Hu and J. Downie. Exploring mood metadata: Relationships with genre, artist and usage metadata. In *Proc. of ISMIR*, pages 67–72, 2007.

[4] P. Knees, E. Pampalk, and G. Widmer. Artist classification with web-based data. In *Proc. of ISMIR*, pages 517–524, 2004.

[5] B. Logan, D. Ellis, and A. Berenzweig. Toward evaluation techniques for music similarity. *The MIR/MDL Evaluation Project White Paper Collection*, 3:81–85, 2003.

[6] C. McKay and I. Fujinaga. Automatic genre classification using large high-level musical feature sets. In *Proc. of ISMIR*, pages 525–530, 2004.

[7] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. The Music Ontology. In *Proc. of ISMIR*, pages 417–422, 2007.

[8] M. Schedl, P. Knees, and G. Widmer. Discovering and visualizing prototypical artists by web-based co-occurrence analysis. In *Proc. of ISMIR*, pages 21–28, 2005.

[9] G. Tzanetakis, G. Essl, and P. Cook. Automatic musical genre classification of audio signals. In *Proc. of ISMIR*, pages 205–210, 2001.

[10] B. Whitman and S. Lawrence. Inferring descriptions and similarity for music from community metadata. In *Proc. of ISMIR*, pages 591–598, 2002.