Community structure in audio clip sharing

Gerard Roma
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
gerard.roma@upf.edu

Perfecto Herrera
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
perfecto.herrera@upf.edu

Abstract—**This paper describes the application of several network analysis techniques to the networks of users in audio clip sharing sites. Two kinds of networks are described: one based on the file downloading activity, and the other based on semantic similarity in the use of free tags to annotate sounds. The properties of these networks as well as their community structure are analyzed using data from Freesound, a very popular audio clip sharing site. Both are characterized as "small worlds". The component structure of the file downloading network provides a general overview of the user community which resembles the structure of the world wide web. On the other hand the analysis of the community structure of the semantic network using modularity optimization appears to be a good way to identify the different semantic areas in the site.**

Keywords-**Multimedia databases; Personal communication networks; Music**

I. Introduction

Since the appearance of digital audio samplers in the 80s, audio clips have been collected by musicians and sound designers for using them in their creations. Many companies sell audio sample CDs that could be roughly divided in two large markets: musical loops and instrument sounds for electronic musicians, and sound effects for sound designers. This differentiation of communities was relatively clear in the pre-internet era from the content production industry point of view. However, while many of these companies now offer their catalog online, it has also become common for users themselves to share audio clips online. Many sites currently allow people to upload and share their recordings, usually with some small textual description and using different licenses, such as Creative Commons (CC) licenses. This of kind user-driven activity is usually related to the possibilities of socializing on the web. Thus, as audio clip sharing communities grow, it is increasingly important to detect the different cultures and sensibilities that exist in these communities, and to relate people with similar interests.

In this paper we describe the application of several concepts and techniques from network science to the characterization of audio clip sharing sites and the discovery of their different user communities. We demonstrate this approach by analyzing the user community at Freesound[1], one of the most popular sites for sharing sounds under CC license. This analysis allows us to understand more clearly the characteristics of the audio clip sharing activity and the different topics involved. Potential applications include search and recommendation of audio clips as well as identification of potential peers in collaborative applications based on shared audio repositories. Beyond audio clip sharing, the studied community is a good example of altruistic generation of resources that have some economic and cultural value, such as (at a different scale) Wikipedia.

II. Related work

Methods for community discovery have been extensively used in several domains involving networks of users. A general trend consists in analyzing if common properties of real world networks appear in these specific cases. In some cases, these studies are motivated by practical applications, while in others the interest of community detection is more related to cultural analysis.

Some classic work analyzed networks of co-authorship in research papers [1], which showed how network analysis techniques help understand collaboration in scientific research. Network theory has also been extensively applied to music recommendation [2], [3]. In [4], communities of music listeners with similar tastes are found through modularity optimization. Several studies have analyzed the network of musicians at Myspace [5], [6] and found that it correlates with the different musical genres. Also, community detection has ben shown to increase accuracy in the identification of cover songs [7]. Finally, some works have explored networks embedded in the creation process. For example [8] describes the network of video creations around the social phenomenon of virtual japanese singer Hatsune Miku in the video sharing site *Nico Nico Douga*. Similarly, the network of collaborative works at *CC-Mixter* (a community of musicians that share sounds under CC) is studied in [9].

Our study shares the focus on users of music recommendation, but here users are actually authors of their sounds. On the other hand, unlike networks of musicians mentioned

---

[1]http://freesound.org

above, we don't study explicit friendship or collaboration relations. Our work is also close to [8] and [9] but instead of networks of documents, we analyze networks of users. To our knowledge, no previous study has analyzed user networks in audio clip sharing.

III. AUDIO CLIP SHARING NETWORKS

The information stored in repositories of audio clips can be summarized in three groups: resources (audio files), users, and textual descriptions. Most sites use free tags as textual descriptions of the audio content. This structure is very similar to tagging systems, which are traditionally studied as tripartite hypergraphs [10]. In tagging systems, an annotation is seen as a link between a user, a resource and a label. From this representation, several networks can be extracted, for example between users on the basis of their shared resources or on the basis of shared tags.

Audio clip sharing sites are different in that users normally describe the files they create and not those of other users. For example in Freesound tags are mainly generated as part of the file uploading process, and while it is possible for any user to tag sounds, 99% of the tags are assigned by the author of a sound. A possible explanation may be that, while in tagging sites such as delicio.us or music radios like last.fm users act as active receivers of information and their main activity is consuming and describing that information, in audio clip sharing there is a more practical motivation in using these files (i.e, the creation of music or audiovisual content) and less in directly enjoying them online. Thus, in audio clip sharing sites it is not possible to link users on the basis of labeling the same resource as in other tagging systems. Instead, it is possible to derive a network from the activity of downloading each other's sounds, and also to create a network of authors based on their similarity in the use of tags.

A. Download network

We define the *download network* of an audio clip sharing site as a graph where each node is a user and each edge represents a file has been downloaded between two users. A user may download many files from the same other user, and thus the network is a multigraph. For some operations, we convert this graph to an undirected one. In this case, downloads between two users in both directions become a number of undirected edges that represent the strength of the similarity (or coincidence in interest) between both users.

B. Semantic network

The *semantic network* of users of an audio clip sharing site can be derived from the concepts they employ to describe their sounds. In this case, nodes are also users, but edges represent the similarity between two users based on their tags. Thus, if we represent a user with the vector $v_i = v_{i0}...v_{in}$ where v_{ij} is the number of files that the user

i has annotated with tag j, the adjacency matrix can be defined as $a_{ij} = sim(v_i, v_j)$. We use cosine similarity,

$$sim(v_i, v_j) = \frac{v_i \cdot v_j}{||v_i|| ||v_j||},$$

to assign values to the edges of the semantic network. This network can also be seen as a multigraph by considering a minimum weight unit [11] so that similarity is expressed in integer values.

C. Summary of analyzed concepts

In the next sections we analyze both the download and the semantic network using concepts from network theory. We now briefly introduce these concepts.

A component of a graph is a set of nodes such that a path exists between any pair of them. A strongly connected component in a directed graph is a set of nodes where any pair is linked in the two directions. It is common to find that real world networks are dominated by a *giant component* that connects a majority of nodes. In general, an analysis of the components of the network helps in describing its global structure. Other general properties are the degree (number of links of each node) and assortativity, which measures the tendency of nodes to link with nodes of the same degree.

Another common property in real world networks is that many can be characterized as *small worlds*. This concept was introduced by Watts and Strogatz [12] who characterized it as a crossover between a regular lattice and a random graph. In random graphs generated according to the Erdős-Rényi (ER) model (i.e. by connecting nodes according to a given probability), the average distance between any two nodes (the average shortest path length) is low, but the clustering coefficient is also low. The clustering coefficient measures how nodes tend to cluster together, by dividing the number of closed triples of nodes in the network (fully connected groups of three nodes) by the total number of (fully or not) connected triples. In small world networks, the average distance between nodes is also low, but the clustering coefficient is much higher. This means that often in real world networks nodes tend to group together forming clusters with high link density.

A group of such densely connected nodes is called a *community*, and a network is considered to have community structure if it can be divided into such groups. The most common measure used to evaluate community structure is *modularity*. Modularity evaluates a given partition of the network by counting the number of links between nodes in the same partition compared to the number of links that would naturally exist due to their degree. Here we consider modularity of an undirected multigraph, which can be used for the two described networks. Given the adjacency matrix A of a graph where A_{ij} is the number of links between nodes i and j, modularity is defined as

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i c_j)$$

where k_i is the degree of node i, m is the total number of edges in the network, and $\delta(c_i c_j)$ is a function that returns 1 if the group of node i and node j are the same and 0 otherwise.

Given this definition, many algorithms have been described for partitioning a network by directly optimizing modularity. One established method was proposed in [13] that works for undirected weighted networks (multigraphs) and is currently used in a number of network analysis packages. The so-called Louvain method works as follows: each node in the network is initially assigned its own community. Then each iteration is divided in two phases. In the first one, each node is moved to its neighboring community that produces the highest modularity. In the second phase, communities are aggregated into nodes. The iterations continue until no further gain in modularity is obtained. This process typically produces a dendrogram that allows the network to be analyzed as a hierarchy of communities. We focus on the top level, which represents a broad partition of the network and also the one with the highest modularity.

While modularity allows to determine the quality of the partition (with values above 0.3 empirically considered to indicate a community structure), adding some external information can help to characterize and understand the found communities. We use the tags employed by users to describe their sounds as indicators of what the main themes of the communities are. We expect that the semantic network, built using those tags, will have more clearly defined themes in this respect, but it is also interesting to see if the communities of the download network are related with different semantic topics. We use the probability of a tag in a community $P_{i|C_k} = \frac{n_{i|C_k}}{n_{C_k}}$ (the number of documents annotated by tag i in community k divided by the total number of documents in that community) and the analogue for the whole network $P_i = \frac{n_i}{n}$. The ratio P_C/P_G indicates if a tag is commonly used in the community with respect to its normal use in the whole site. Values above 1 indicate a more specific use in this community, and tags with large values are used to characterize the community. On the other hand, the entropy of a community with respect to tags can be described using the local probabilities:

$$S_{C_k} = - \sum_{i \in C_k} P_{i|C_k} log P_{i|C_k},$$

in the same sense as the genre entropy used in [5]. An average of this quantity over all communities (S_C) provides a hint about the quality of the partition with respect to tags. A low value of the entropy of tags in the community indicates a greater homogeneity with respect to tags. Note that, usually, tags are not necessarily as general and comprehensive as genres are in music. However, for this measures we limit

part	n of nodes	percent	average in-degree	average out-degree
SCC	1836	82%	32.33	31.68
IN	145	6.5%	0	21.04
OUT	187	8.4%	11.24	0
other	66	2.9%	10.47	14.32

Table I
GLOBAL STRUCTURE OF THE DOWNLOAD NETWORK: STRONGLY CONNECTED COMPONENT (SCC), NODES WITH ZERO IN-DEGREE (IN) AND NODES WITH ZERO OUT-DEGREE (OUT)

tags to the most popular ones, with a frequency of more than 1000 in the database. On the other hand, since the value of entropy depends on the size of the community, we compute the entropy of a proportional random partition of the network as a baseline.

IV. THE FREESOUND COMMUNITY

We extracted the download and semantic networks from Freesound. We now analyze both networks using the concepts explained in the previous section.

A. Download network

One thing that stands out when building the download network of Freesound, is that a vast majority of users are only interested in downloading files. The reason is that, in order to download files at their original quality, the site requires people to sign in. These users could be represented by nodes with zero in-degree. However, the purpose of this article is to characterize the network of users who upload sounds, and we discard users who did not upload sounds. This gives us a reduced community of 2234 nodes (out of the more than a million of registered users) and 60342 edges. In this reduced community, we find a strongly connected component of 1836 nodes, about 80% of people who have uploaded some sound. The remaining 20% is mainly split among users who have zero in-degree (this is, people whose sounds have not been downloaded by any other author) or zero out-degree (users who have uploaded sounds but that have not downloaded any sound), with a remainder of nodes that mostly have either one in-degree or one out-degree. This structure mimics the *bow tie* structure that has been found to characterize the World Wide Web [14] (Table I).

On the other hand, the network of authors can be characterized as a small world. Table II shows the average shortest path length and clustering coefficient of the network compared to the equivalent ER graph. While the average distance between nodes is roughly the same than for the random graph, the clustering coefficient is an order of magnitude larger. Interestingly, assortativity is negative, which means that nodes tend to associate with nodes of different degree.

The application of the Louvain method for community detection to this network reveals a structure dominated by a large community with a modularity of 0.29. Table III shows the different communities and their characteristic

Nodes	2234
Edges	61697
Mean degree	27.62
Diameter	5
Assortativity	-0.181
SPL	2.23 (2.22)
Clustering Coefficient	0.28 (0.024)

Table II

MAIN PROPERTIES OF THE DOWNLOAD NETWORK (VALUES FOR THE EQUIVALENT ER GRAPH)

tags. While this structure may seem similar to the global structure described by the component analysis, as a matter of fact the proportion of nodes of each community that belong to the strongly connected component is very similar (around 80% for all communities). This discards that the modularity partition merely replicates the division between the giant component and the rest. At the second level, this large component loses some nodes to small sibling communities, but the trend is preserved. An analysis of the most characteristic tags in the top level (Table IV) shows, however, that there are some differences in the topics preferred by each community. The main community, while more heterogeneous in the tags, is characterized by one of the most popular tags of the site: *field recording*, and generally concepts related to environmental sounds. Contrastingly, in the rest of communities, concepts related with musical uses of sounds are more salient. However, the general entropy of the partition is not much lower than the entropy of the equivalent random partition.

communities	modularity	S_C	S_C (random)
5	0.29	3.02	3.17

Table III

NUMBER OF COMMUNITIES, MODULARITY, ENTROPY AND AVERAGE ENTROPY OF EQUIVALENT RANDOM PARTITION IN DOWNLOAD NETWORK

B. Semantic network

The semantic network is itself an undirected, weakly connected component, although it contains less nodes (2161) than the download network. The missing 27 users are disconnected from this network because they use few and very specific tags that nobody else uses. On the other hand, the number of edges is higher (75607) which makes a very dense network.

An analysis of the average shortest path length and the clustering coefficient shows that this network can also be characterized as a small world (Table V). This feature is even more pronounced than in the download network. Again, assortativity is strongly negative, which is understandable from the construction of the network: users who use many words to describe their sounds will be linked to users who

Community 0 , Size: 33 , Entropy: 2.58	
5.49	human
8.67	acoustic
2.92	short
2.01	percussion
1.55	field-recording
1.28	noise
Community 1 , Size: 294 , Entropy: 3.26	
5.26	analog
2.52	drums
2.10	percussion
1.91	drum
1.80	beat
1.80	glitch
Community 2 , Size: 32 , Entropy: 2.38	
6.90	percussion
6.18	metal
5.45	drum
4.66	hit
2.45	guitar
2.30	drums
Community 3 , Size: 1644 , Entropy: 3.48	
1.59	birds
1.59	nature
1.53	male
1.50	ambience
1.42	field-recording
1.37	techno
Community 4 , Size: 231 , Entropy: 3.41	
2.84	reaktor
2.76	multisample
2.22	drone
2.08	space
1.61	fx
1.54	electronic

Table IV

CHARACTERISTIC TAGS IN COMMUNITIES OF DOWNLOAD NETWORK (PROBABILITY RATIO)

use less (but similar) words and therefore have a smaller degree.

Nodes	2161
Edges	75607
Mean degree	69.97
Diameter	5
Assortativity	-0.25
SPL	2.07 (2.07)
Clustering Coefficient	0.54 (0.032)

Table V

MAIN PROPERTIES OF THE SEMANTIC NETWORK (VALUES FOR THE EQUIVALENT ER GRAPH)

Modularity-based community detection reveals a more pronounced and also more balanced division of the network with respect to the download network, with a modularity of 0.35 (Table VI). Here, the distribution of topics is even clearer among the different communities (Table VII), as could be expected from the construction of the network (for example, *techno* appears in a community with related topics, while in the download network it appeared in the larger

community along with concepts related to environmental sounds). Also, the average entropy is significantly lower than the value for the equivalent random partition. Several groups are related to percussive sounds used for creating rhythms in electronic music, while others seem more related with voice (another popular tag of the site) or environmental sounds. The case of voice samples is characteristic: a tradition exists in Freesound to provide recordings of voices with specific utterances upon request. This community could be considered to reflect this activity. Some of the detected communities are very small and do not contain any of the more popular tags, so they are omitted from Table VII.

communities	modularity	S_C	S_C (random)
6	0.36	1.94	2.79

Table VI
NUMBER OF COMMUNITIES, MODULARITY, ENTROPY AND AVERAGE ENTROPY OF EQUIVALENT RANDOM PARTITION IN SEMANTIC NETWORK

Community 0 , Size: 757 , Entropy: 3.04	
8.00	metal
7.34	water
6.35	hit
3.58	birds
2.56	percussion
2.38	ambience
Community 1 , Size: 128 , Entropy: 2.62	
11.90	human
7.91	male
7.71	voice
3.18	short
2.29	acoustic
1.90	effect
Community 3 , Size: 332 , Entropy: 2.5	
5.53	nature
5.40	field-recording
4.62	birds
2.97	male
2.78	water
1.79	voice
Community 5 , Size: 990 , Entropy: 3.46	
1.32	techno
1.32	multisample
1.31	electro
1.31	reaktor
1.29	analog
1.28	glitch

Table VII
CHARACTERISTIC TAGS (PROBABILITY RATIO) IN COMMUNITIES OF SEMANTIC NETWORK (COMMUNITIES WITH LESS THAN 10 NODES ARE OMITTED)

C. Confusion of communities in both networks

Table VIII shows the amount of users in each community of both the semantic and download networks. Row and column indexes correspond to indexes of the communities in each network. Both partitions seem to be related, with the main group of the download network splitting equally into two of the semantic communities. For the rest, a majority of users of a community in one network belongs to a community of the other. A χ^2 test on this matrix, as a contingency table with 20 degrees of freedom returns a very small value ($p = 9.5 \cdot 10^{-24}$) which gives support to the rejection of the null hypothesis stating that both assignments are independent.

	0	1	2	3	4	5
0	5	3	0	7	0	17
1	61	12	1	36	2	171
2	5	2	0	1	0	24
3	640	98	6	268	3	601
4	34	11	0	19	0	162

Table VIII
CONFUSION MATRIX: COUNT OF NODES IN THE DETECTED COMMUNITIES OF THE DOWNLOAD (ROWS) AND SEMANTIC (COLUMNS) NETWORKS

V. CONCLUSIONS

Audio clip databases are usually analyzed as collections of resources (audio files), often taking into account semantic labels or tags. However, in audio clip sharing websites users are a key aspect as they generate the sounds and their descriptions. The motivations of users are often related to social aspects of sharing sounds online. Finding communities with related interests is an important step in order to add socialization features to this kind of sites, especially given the wide diversity of interests related to different contexts of sound design and the creation of different genres of music. We have studied two networks of users that can be extracted from such sites: the download network (based on the activity of downloading files among users) and the semantic network (based on similarity in the use of the tags). Both networks have a high number of edges and thus are very dense, more than many real world networks where often the amount of edges is in the same order of magnitude than the amount of nodes. However, both networks exhibit properties common in real-world networks: nodes tend to cluster more than in a random graph, and thus they can both be characterized as *small worlds*. The analysis of the component structure of the download network allows to obtain a general overview of the community, dominated by a large, strongly connected component. The division of this network with highest modularity produces a large community with about 73% of nodes. This seems to reflect the perception when visiting the web and its forum that there is a community of very active users, and that the generated wealth is related to the existence of this central and connected community. A study of the temporal evolution of this network could help in further understanding this process.

The semantic network makes it possible, using modularity optimization, to identify a balanced partition into communities which clearly reflect different themes or semantic areas

of the database. This partition makes it easier to understand different interests in the complex mix of tags that form the Freesound folksonomy. This partition seems to be very correlated with the partition obtained from the download network, which indicates a relation between the activity of downloading someone else's file and the tendency to use similar words to describe sounds. This can be interpreted in two directions: a user may feel inclined to download from another user because of the proximity in the tags utilized, or when tagging sounds a user may be influenced by the labels of sounds she has downloaded. At the same time, the observation that the download network is related but less partitioned seems to reflect a common interests among members of different semantic communities. In music and audiovisual compositions, sounds are used very often because of their acoustic properties, independently from the semantics. A sound that is acoustically similar to the desired effect can be used even if its actual source is completely unrelated to the creative context, which means that people do not necessarily take into account semantic tags as the only factor when downloading sounds. This could be taken as a strong advice to include query interfaces that can exploit the acoustic relatedness or similarity between sounds, and not only high-level or semantic concepts.

VI. Acknowledgements

References

[1] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," in *Proc. National Academy of Sciences*, 2004, pp. 5200–5205.

[2] M. Zanin, P. Cano, J. M. Buldú, and O. Celma, "Complex networks in recommendation systems," in *Proc. 2nd WSEAS Int. Conf. on Computer Engineering and Applications*, World Scientific Advanced Series In Electrical And Computer Engineering. Acapulco, Mexico: World Scientific Advanced Series In Electrical And Computer Engineering, 2008.

[3] J. Park, O. Celma, M. Koppenberger, P. Cano, and J. Martin-Buldú, "The social network of contemporary popular musicians," *Int. Journal of Bifurcation and Chaos (IJBC)*, vol. 17, pp. 2281 – 2288, 2007.

[4] A. Anglade, M. Tiemann, and F. Vignoli, "Virtual communities for creating shared music channels," in *Proc. 8th Int. Conf. on Music Information Retrieval*, 2007.

[5] K. Jacobson, B. Fields, and M. Sandler, "Using audio analysis and network structure to identify communities in online social networks of artists," *Proc. 9th Int. Conf. on Music Information Retrieval*, 2008.

[6] B. Fields, K. Jacobson, M. Casey, and M. Sandler, "Do you sound like your friends? exploring artist similarity via artist social network relationships and audio signal processing," in *Proc. of ICMC*, August 2008.

[7] J. Serrà, M. Zanin, and P. Herrera, "Cover song networks: analysis and accuracy increase," in *Net-Works 2010*, Zaragoza, Spain, 2010.

[8] M. Hamasaki, H. Takeda, and T. Nishimura, "Network analysis of massively collaborative creation of multimedia contents: case study of hatsune miku videos on nico nico douga," in *Proc. 1st Int. Conf. on Designing Interactive User Experiences for TV and Video*, 2008, pp. 165–168.

[9] G. Cheliotis and J. Yew, "An analysis of the social structure of remix culture," in *C and T 09: Proc. 4th international conference on Communities and technologies*. New York, NY, USA: ACM, 2009, pp. 165–174.

[10] P. Mika, "Ontologies are us: A unified model of social networks and semantics," in *Int. Semantic Web Conf.*, ser. Lecture Notes in Computer Science, vol. 3729, International Semantic Web Conference 2005. Springer, November 2005, pp. 522–536.

[11] M. E. J. Newman, "Analysis of weighted networks," *Phys. Rev. E*, vol. 70, no. 5, p. 056131, Nov. 2004.

[12] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, June 1998.

[13] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of community hierarchies in large networks," *CoRR*, vol. abs/0803.0476, 2008.

[14] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," *Comput. Netw.*, vol. 33, no. 1-6, pp. 309–320, 2000.