The Topology of Music Artists' Graphs

Pedro Cano[1], Oscar Celma[1], Markus Koppenberger[1] and Javier M. Buldú[2]
[1] *Music Technology Group, Universitat Pompeu Fabra, Ocata 1, 08003 Barcelona, Spain.*
[2] *Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Colom 11, 08222 Terrassa, Spain*

We study the topology of several music networks, namely citation in AllMusicGuide (AMG), Amazon, Launch-Yahoo! and MSN-Entertainment. The analysis uncovers the emergence of complex network phenomena in music artist similarity networks built considering artists as nodes and its relations as edges. The properties provide some hints on searchability and possible optimizations in the design of music recommendation systems. It may also provide a deeper understanding on the similarity measures that can be derived from existing music knowledge sources.

Nowadays access to music is possible by querying artists or song names —editorial data— or browsing recommendations generated by collaborative filtering —i.e: recommender systems that exploit information such as *"users that bought this album also bought this album"*. An obvious drawback is that consumers need to know the name of the song or the artist, or a number of consumers must have heard and rated the music. This situation makes it difficult for users to access and navigate through the vast amount of music composed and performed by unknown new artists, which is available on-line in an increasing number of sites. As an alternative, content-based music similarity systems are being pursued. Content-based similarity systems perform an analysis of the musical meaningful attributes of audio music in order to recommend music. A difficulty arises when developing such systems: how to evaluate the goodness of a similarity measure. Similarity metrics are subjective and it is not easy to devise an psychoacoustic experiment. Several evaluation metrics have been proposed to validate content-based similarity using as a ground truth data from AllMusicGuide[1], where experts recommend similar artists, or analysis in P2P networks, where there is the assumption that the artists that co-occur in somebody's machine are somehow similar. In this work, complex network measurements are used to analyze the topology of networks underlying these information systems.

The properties of the different networks under study raise a discussion on the underlying forces driving collaborative systems and music expert networks. For example, how much of the network structure is due to content similarity and how much to the self-organization of the network. Therefore, it can shed new light on the design and validation of music similarity measures on perceptual similarity measures and its evaluation[2]. Secondly, it also hints possible optimizations when designing music information systems, as a number of artists might have very few or no links at all with other artists[3]. Content-based techniques can add links optimized in a way to allow navigating from any artist on the database to any other in a short number of *clicks* (the number of content-based derived edges is a design decision). Finally, it may help to understand the dynamics of certain aspects of music evolution, e.g: how did an artist get popular?.

We have gathered information from four different sources, namely artist relationships, in AllMusicGuide, Amazon[4], Launch-Yahoo![5] and MSN-Entertainment[6], and we have created a graph for each source. A graph is constructed as follows: each node represents a music artist whereas an edge denotes a *"relationship"* among them.

All networks under study are Small Worlds. Regarding cumulative degree distributions, $P_c(k_{in})$ and $P_c(k_{out})$, we have noticed that only MSN-Entertainment and Launch-Yahoo! networks are "scale-free". The mechanism for collaborative filtering-based networks is known given the distribution of user ratings[7]. The deviations from power-law behavior, in the AllMusicGuide network, which is the only one purely created by human editors may have its origin on information filtering[8]. Two of the networks, Launch-Yahoo! and MSN-Entertainment are clearly dissortative which is not at all common in social networks[9]. The reasons why the structure of these network is different among them is unclear. There are other open more practical issues, such as: how can the large scale architecture of musical artist information networks be exploited to improve user interaction with a vast music collection?[10]

---

[1] , AllMusicGuide, (2004), http://www.allmusic.com.
[2] D. P. Ellis, B. W. A. Berenzweig, and S. Lawrence, *Proc. of the Intl. Symp. on Music Information Retrieval*, pp. 170–177 (Paris,2002).
[3] J. M. Kleinberg, Nature, **406**, 845, (2000).
[4] Amazon, (2004), http://www.amazon.com.
[5] Launch-Yahoo!, (2004), http://launch.yahoo.com.
[6] MSN, (2004), http://music.msn.com.
[7] , G. Linden, B. Smith, and J. York, IEEE Internet Computing, **4**, (2003), http://dsonline.computer.org/0301/d/wp1lind.htm.
[8] S. Mossa, M. Barthélemy, H. E. Stanley, and L. A.Ñ. Amaral, Phys. Rev. Letters, **88**, (2002).
[9] M. E. J. Newman, Phys. Rev. Letters, **89** (2002).
[10] M. E. J. Newman, SIAM Review, **45**, 167 (2003).