From hits to niches? or how popular artists can bias music recommendation and discovery

Òscar Celma
Music Technology Group
Universitat Pompeu Fabra
Barcelona, SPAIN
oscar.celma@iua.upf.edu

Pedro Cano
Barcelona Music and Audio Technologies
Llacuna 162, 08018
Barcelona, SPAIN
pedro.com

ABSTRACT

This paper presents some experiments to analyse the popularity effect in music recommendation. Popularity is measured in terms of total playcounts, and the Long Tail model is used in order to rank music artists. Furthermore, metrics derived from complex network analysis are used to detect the influence of the most popular artists in the network of similar artists.

The results from the experiments reveal that—as expected by its inherent social component—the collaborative filtering approach is prone to popularity bias. This has some consequences on the discovery ratio as well as in the navigation through the Long Tail. On the other hand, in both audio content–based and human expert–based approaches artists are linked independently of their popularity. This allows one to navigate from a mainstream artist to a Long Tail artist in just two or three clicks.

Categories and Subject Descriptors

H3.3 [**Information Search and Retrieval**]: Information filtering, Selection process; G.2.2 [**Graph Theory**]: Graph algorithms

Keywords

recommender systems, popularity, long tail, evaluation, complex network analysis

1. INTRODUCTION

The Long Tail is composed by a very few popular items, the well–known *hits*, and the rest, located in the heavy tail, that does not sell *that well* [1]. The Long Tail offers the possibility to explore and discover—using automatic tools; such as recommenders—from vast amounts of data. Until now, the world was ruled by the *Hit or Miss* classification, due in part to the shelf space limitation of the brick–and–mortar stores. A world where a music band could only succeed selling millions of albums, and touring worldwide. Nowadays, we are moving towards the *Hit vs. Niche* idea, where there is a large enough availability of choice to satisfy even the most 'Progressive–obscure–Swedish–metal" fan. The problem, though, is to filter and present the *right* artists to the user, according to her musical taste.

Indeed, in his book [1], Chris Anderson introduces a couple of (among others) very important conditions to exploit the niche markets. These are: *(i)* make everything available, and *(ii)* help me find it. It seems that the former condition is already fulfilled; the distribution and inventory costs are nearly negligible. Yet, to satisfy the latter we need recommender systems that exploit the *"from hits to niches"* paradigm. The main question, though, is whether current recommendation techniques are ready to assist us in this discovery task, providing recommendations of the *hidden jewels* in the Long Tail. In fact, recommenders that appropriately discount popularity may increase total sales [9], as well as potentially increase the margins by suggesting more novel or less known tunes.

Some answers are provided in this paper, in the context of the music domain. The analysis is not performed in terms of classic precision and accuracy of the recommendations, but focusing on how the algorithms behave regarding item popularity. Actually, popularity is the element that defines the characteristic shape of the Long Tail.

This paper is structured as follows: section 2 introduces the Long Tail model. This is the first step needed in order to use artist popularity information. Section 3 focuses on analysing the artists' similarity graph, created using any item–based recommendation algorithm. The metrics allows us to characterise the intrinsic topology of the artist network (e.g. are the hubs in the recommendation network the most popular artists?). Then, section 4 presents the experiments performed in the context of the music domain, comparing three algorithms (collaborative filtering, content–based, and human experts). Finally, in section 5 we discuss the main findings, and conclude with future work in section 6.

2. THE LONG TAIL MODEL

The Long Tail of a catalog is measured in terms of frequency distribution (e.g. purchases, downloads, etc.), ranked by item popularity. It has been largely acknowledged that item popularity can decrease user satisfaction and novelty detection in the recommendation workflow, by providing obvious recommendations [10, 15].

As an example, Figure 1 (left) depicts the Long Tail for 260,525 music artists[1]. The horizontal axis contains the list

[1]The data was gathered from *last.fm* website during July,

2nd Netflix-KDD Workshop, August 24, 2008, Las Vegas, NV, USA.

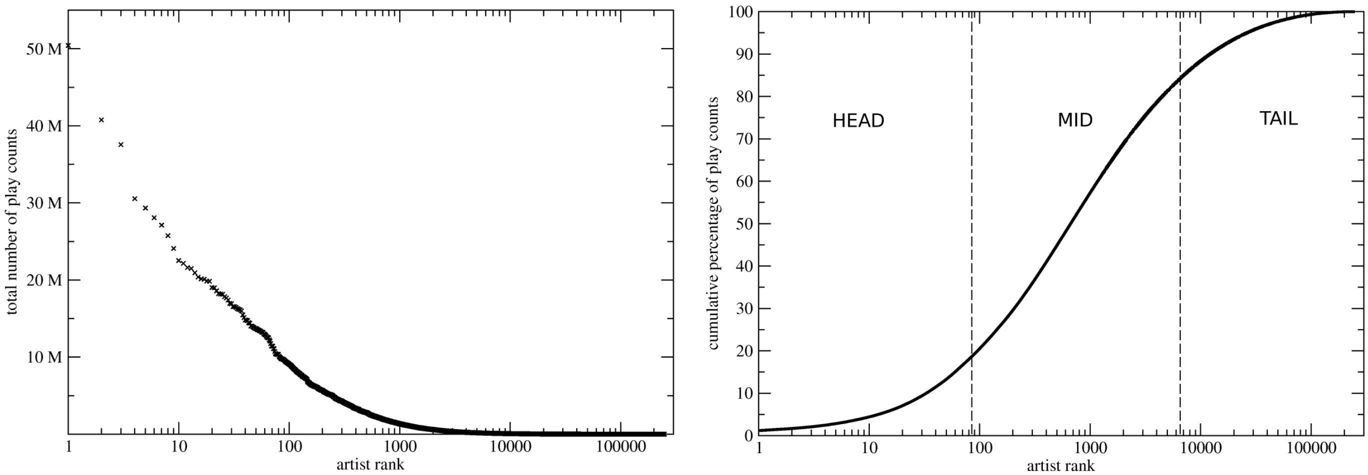

Figure 1: (left) The Long Tail of 260,525 music artists. A log–linear plot depicting artist rank in terms of total playcounts (e.g. at top–1 there is The Beatles with more than 50 million total playcounts). Data gathered from *last.fm*, during July 2007. (right) Cumulative percentage of playcounts from the left figure. Top–737 artists accumulates the 50% of total playcounts (N_{50}). The curve is divided in three parts: head, mid and tail ($X_{head \to mid} = 82$, and $X_{mid \to tail} = 6,655$). The fitted model, $F(x)$, has $\alpha = 0.73$ and $\beta = 1.02$.

of artists ranked by total playcounts. E.g. The Beatles, at position 1, has more than 50 million playcounts.

The Long Tail model, $F(x)$, simulates any Long Tail curve [12]. It models the cumulative distribution of the Long Tail data. $F(x)$ equals to the share of total volume covered by objects up to rank x:

$$F(x) = \frac{\beta}{(\frac{N_{50}}{x})^{\alpha} + 1} \tag{1}$$

where α is the factor that defines the S–shape of the function, β is the total volume share (and also describes the amount of latent demand), and N_{50} is the number of objects that cover half of the total volume, that is $F(N_{50}) = 50$.

Once the Long Tail is modelled using $F(x)$, we can divide the curve in three parts: head, mid, and the tail. The boundary between the head and the mid part of the curve is defined by:

$$X_{head \to mid} = N_{50}^{2/3} \tag{2}$$

Likewise, the boundary between the mid part and the end of the tail is:

$$X_{mid \to tail} = N_{50}^{4/3} \simeq X_{head \to mid}^{2} \tag{3}$$

Figure 1 (right) depicts the cumulative distribution of the Long Tail of 260,525 music artists. Interestingly enough, the top–737 artists account for 50% of the total playcounts, $F(737) = 50$, and only the top–30 artists hold around 10% of the plays. In this sense, the *Gini coefficient* measures the inequality of a given distribution, and it determines the degree of imbalance. In our Long Tail example, 14% of the artists hold 86% of total playcounts, yielding a Gini coefficient of 0.72. This value denotes an imbalanced distribution, higher than the 80/20 Pareto rule (0.6). Figure 1 (right) shows the head of the curve, $X_{head \to mid}$ which consists of only 82 artists, whereas the mid part has 6,573 ($X_{mid \to tail} = 6,655$). The rest of the artists are located in the tail part.

An interesting work is to analyse artist similarity according to the popularity. In our case, this is performed in the context of a network that links the artists (nodes) according to their resemblance. The following section is devoted to explain the metrics that we use.

3. COMPLEX NETWORK ANALYSIS

We propose several metrics to analyse an item–based recommendation graph; $G := (V, E)$, being V a set of nodes, and E a set of unordered pairs of nodes, named edges. In our case, the items (i.e. music artists) are nodes, and the edges denote the (weighted) similarity among the items, using any item–based recommendation algorithm. The metrics used are derived from Complex Network and Social Network analysis.

3.1 Metrics

3.1.1 Navigation

The **average shortest path** (or mean geodesic length) measures the distance between two vertices i and j. They are connected if one can go from i to j following the edges in the graph. The path from i to j may not be unique. The minimum path distance (or geodesic path) is the shortest path distance from i to j, d_{ij}. The average shortest path in the network is:

$$\langle d \rangle = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i,j \in V, i \neq j} d_{ij} \tag{4}$$

In a random graph, the average path approximates to:

$$\langle d_r \rangle \sim \frac{logN}{log \langle k \rangle}, \tag{5}$$

2007. *Last.fm* provides plugins for virtually any desktop music player to track users' listening behaviour.

where $N = |V|$, and $\langle k \rangle$ denotes the mean degree of all the nodes.

The longest path in the network is called its **diameter** (D). In a recommender system, average shortest path and diameter inform us about the global navigation through the network of items.

The **strong giant component**, SGC, of a network is the set of vertices that are connected via one or more geodesics, and are disconnected from all other vertices. Typically, networks possess one large component that contains a majority of the vertices. It is measured as the % of nodes that includes the giant component. In a recommender system, SGC informs us about the catalog coverage, that is the total percentage of available items the recommender recommends to users [10].

3.1.2 Connectivity

The **degree distribution**, p_k, is the number of vertices with degree k:

$$p_k = \sum_{v \in V \,|\, \deg(v)=k} 1, \tag{6}$$

where v is a vertex, and $\deg(v)$ its degree. More frequently, the *cumulative degree distribution* (the fraction of vertices having degree k or larger), is plotted:

$$P(k) = \sum_{k'=k}^{\infty} p_{k'} \tag{7}$$

A cumulative plot avoids fluctuations at the tail of the distribution and facilitates the evaluation of the power coefficient γ, in case the network follows a power law. In a directed graph, that is when a recommender algorithm only computes the top–n most similar items, $P(k_{in})$ and $P(k_{out})$, the cumulative incoming (outcoming) degree distribution, are more informative. Cumulative degree distribution detects whether a recommendation network has some nodes that act as hubs. That is, that they have a large amount of attached links. This clearly affects the recommendations and navigability of the network.

Another metric used is the **degree correlation**. It is equal to the average nearest–neighbour degree, k^{nn}, as a function of k:

$$k^{\text{nn}}(k) = \sum_{k'=0}^{\infty} k' p(k'|k), \tag{8}$$

where $p(k'|k)$ is the fraction of edges that are attached to a vertex of degree k whose other ends are attached to vertex of degree k'. Thus, $k^{\text{nn}}(k)$ is the mean degree of the vertices we find by following a link emanating from a vertex of degree k.

A closely related concept is the **degree–degree correlation coefficient**, also named *assortative mixing*, which is the Pearson r correlation coefficient for degrees of vertices at either end of a link. A monotonically increasing (decreasing) k^{nn} means that high–degree vertices are connected to other high–degree (low–degree) vertices, resulting in a positive (negative) value of r [17]. In recommender systems, it measures to which extent nodes are connected preferentially to other nodes with similar characteristics.

3.1.3 Clustering

The clustering coefficient, C, estimates the probability that two neighbouring vertices of a given vertex are neighbours themselves. C is defined as the average over the *local measure*, C_i [21]:

$$C_i = \frac{2|E_i|}{k_i(k_i-1)}, \tag{9}$$

where E_i is the set of existing edges that are direct neighbours of i, and k_i the degree of i. C_i denotes, then, the portion of actual edges of i from the potential number of total edges.

For random graphs, the clustering coefficient is defined as $C_r \sim \langle k \rangle / N$. Typically, real networks have a higher clustering coefficient than C_r.

3.2 Related work in music recommendation

During the last few years, complex network analysis has been applied to music information retrieval in general, and music recommendation in particular. In [6], we compared different music recommendation algorithms based on the network topology. The results are aligned with our main findings: social based recommenders present a scale–free network topology, whereas human expert–based controlled networks does not.

An empirical study of the evolution of a social network constructed under the influence of musical tastes, based on playlist co-occurrence, is presented in [14]. The analysis of collaboration among contemporary musicians, in which two musicians are connected if they have performed in or produced an album together, is presented in [18]. In [2], the authors present a user clustering algorithm that exploits the topology of a user–based similarity network.

A network of similar songs based on timbre similarity is presented in [3]. Interestingly enough, the network is scale–free, thus a few songs appear in virtually any list of similar tracks. This has some problems when generating automatic playlists. [11] presents an analysis of the Myspace social network, and conclude that artists tend to form on-line communities with artists of the same musical genre.

3.3 Network analysis and the Long Tail model

Once each item in the recommendation network is located in the head, mid, or tail part (see section 2), the next step is to combine the similarity network with the Long Tail information. Two main analysis are performed: first, we measure the similarity among the items in each part of the curve. That is, for each item that belongs to the head part, compute the percentage of similar items that are located in the head, mid and tail part (similarly, for the items in the mid and tail part). This measures whether the most popular items are connected with other popular items, and vice versa. Second, we measure the correlation between an item's rank in the Long Tail and its indegree. This allows us to detect whether the hubs in the network are also the most popular items. Section 4 presents the experiments regarding popularity analysis, comparing three different music artists recommendation algorithms: collaborative filtering (CF) from *last.fm*, content–based audio filtering (CB), and expert–based recommendations from *Allmusic.com* (AMG) musicologists.

Property	CF (*Last.fm*)	CB	Expert–based (*AMG*)
N	122,801	59,583	74,494
$\langle k \rangle$	14.13	19.80	5.47
$\langle d_d \rangle$ ($\langle d_r \rangle$)	5.64 (4.42)	4.48 (4.30)	5.92 (6.60)
D	10	7	9
SGC	99.53%	99.97%	95.80%
γ_{in}	2.31($\pm$0.22)	1.61($\pm$0.07)	NA (exp. decay)
r	0.92	0.14	0.17
C (C_r)	0.230 (0.0001)	0.025 (0.0002)	0.027 (0.00007)

Table 1: Artist recommendation network properties for *last.fm* collaborative filtering (CF), content–based audio filtering (CB), and *Allmusic.com* (AMG) expert–based. N is the number of nodes, and $\langle k \rangle$ the mean degree, $\langle d_d \rangle$ is the avg. shortest directed path, and $\langle d_r \rangle$ the equivalent for a random network of size N, and D is the diameter of the network. SGC is the size of the strong giant component, γ_{in} is the power–law exponent of the cumulative indegree distribution, r is the indegree–indegree Pearson correlation coefficient (assortative mixing). C is the clustering coefficient, and C_r for the equivalent random network.

4. EXPERIMENTS

In order to put into practice the Long Tail model, and the properties of item–based recommendation networks, we performed several experiments in the music recommendation field. It is worth noting that music is somewhat different from other entertainment domains, such as movies, or books. Tracking users' preferences are mostly done implicitly, via their listening habits. Moreover, a user can consume an item (i.e. a track, or a playlist) several times, even repeatedly and continuously. Regarding the evaluation process, music recommendation allows us instant feedback with a, say, 30 seconds excerpt.

The experiments aim at evaluating the popularity effect using three (music artists) recommendation approaches: collaborative filtering (CF), content–based audio similarity (CB), and human expert–based resemblance. We measure the popularity effect by contrasting the properties from the network with the Long Tail information of the catalog (e.g. are the hubs in the recommendation network the most popular items? Are the most popular items connected with other popular items, and vice versa?).

4.1 Datasets

CF artist similarity was gathered from *last.fm*, using Audioscrobbler web services[2], and selecting the top–20 similar artists. *Last.fm* has a strong social component, and their recommendations are based on the classic item–based algorithm[3] [20].

To compute artist similarity in the CB network, we apply content–based audio analysis in a music collection ($\mathcal{T}$) of 1.3 Million tracks of 30 seconds samples. Our audio analysis considers not only timbral features (e.g. Mel frequency cepstral coefficients), but some musical descriptors related to rhythm and tonality, among others [7]. Then, to compute artist similarity we used the most representative tracks, $\mathcal{T}_a$, of an artist a, with a maximum of 100 tracks per artist. For each track, $t_i \in \mathcal{T}_a$, we obtain the most similar tracks (excluding those from artist a):

$$sim(t_i) = \operatorname*{argmin}_{\forall t \in \mathcal{T}} (distance(t_i, t)), \quad (10)$$

and get the artists' names, $\mathcal{A}_{sim(t_i)}$, of the similar tracks. The list of (top–20) similar artists of a is composed by all $\mathcal{A}_{sim(t_i)}$, ranked by frequency and weighted by the audio similarity distance:

$$similar\_artists(a) = \bigcup \mathcal{A}_{sim(t_i)}, \forall t_i \in \mathcal{T}_a \quad (11)$$

Finally, we gather expert recommendations from *All Music Guide* (AMG)[4]. AMG makes use of professional editors to interconnect artists, according to several aspects, such as: *influenced by*, *followers of*, *similar artists*, *performed songs by*, etc. In order to create an homogeneous network, we only make use of the *similar artists* links. Artists from both CB and expert–based networks are a subset of the CF artists.

4.2 Network analysis

The network properties of the three datasets are shown in Table 1. All the networks present the *small–world* phenomena [21]. They have a small average directed shortest path, $\langle d_d \rangle$, similar than its equivalent random network, $\langle d_r \rangle$. Also the clustering coefficients, C, are significantly higher than the equivalent random networks C_r. This is an important property, because recommender systems can be structurally optimised so as to allow users surfing to any part of a music collection with a small number of mouse clicks, and so that they are easy to navigate using only local information [13].

AMG network has a giant component, SGC, smaller than CF and CB networks. Around 4% of their artists are isolated, and cannot be reached from rest (in the giant component). This has strong consequences with regard to the coverage of the recommendations, as well as the navigation for the artists located in the "small islands".

Regarding cumulative indegree distribution, AMG has an exponential decay, whereas CF and CB follow a power law. CF has a power–law exponent, $\gamma = 2.31$, similar to those detected in many scale–free networks, including the world wide web linking structure [5]. These networks are known to show a right–skewed power law distribution, $P(k) \propto k^{-\gamma}$ with $2 < \gamma < 3$, relying on a small subset of hubs that control the network [4].

[2]http://www.audioscrobbler.net/data/webservices/

[3]Although, is quite possible that they are using, as well, some information gathered from social tagging.

[4]http://www.allmusic.com

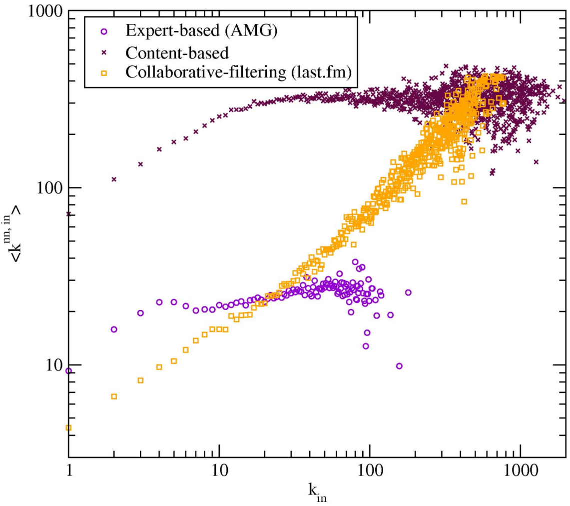

Figure 2: Indegree–indegree correlation (assortative mixing) for the three artist recommendation networks: *last.fm* collaborative filtering (CF), Content–based (CB), and *Allmusic.com* experts. CF clearly presents the assortative mixing phenomenon ($r_{CF} = 0.92$). Neither CB nor Expert–based present any correlation ($r_{CB} = 0.14$, $r_{Expert} = 0.17$).

Another difference is the assortative mixing, or indegree–indegree correlation, presented in Figure 2. CF presents a high assortative mixing ($r = 0.92$). That means that the most connected artists are prone to be similar to other top connected artists. Neither CB nor Expert–based present indegree–indegree correlation, thus artists are connected independently of their inherent properties.

4.3 Popularity analysis

We have outlined, in the previous section, the main topological differences among the three networks. Now, we add the popularity factor (measured in terms of total playcounts per artist), by combining artists' rank in the Long Tail with the results from the network analysis. Two experiments are performed. The former reports the relationships among popular (and unknown) artists. The latter experiment aims at analysing the correlation between artists' indegree and its popularity.

4.3.1 Artist similarity

Figure 3 depicts the correlation among artist's total playcounts and the total playcounts of its similar artists. That is, given the total playcounts of an artist (x axis) it shows, in the vertical axis, the average playcounts of its similar artists. CF network has a clear correlation ($r_{CF} = 0.46$); the higher the playcounts of a given artist, the higher the avg. playcounts of its similar artists. Neither CB nor AMG present any correlation ($r_{CB} = 0.08$, $r_{EX} = 0.09$). Thus, artists are linked independently of their popularity.

Table 2 presents artist similarity divided into the three sections of the Long Tail curve. Given an artist, a_i, it shows (in %) the Long Tail location of its similar artists (results are averaged over all artists). In the CF network, given a very popular artist, the probability of reaching (in one click) a similar artist in the tail is zero. Actually, half of the similar

Method	$a_i \rightarrow a_j$	Head	Mid	Tail
CF top–20	**Head**	45.32%	54.68%	0%
	Mid	5.43%	71.75%	22.82%
	Tail	0.24%	17.16%	82.60%
CB top–20	**Head**	6.46%	64.74%	28.80%
	Mid	4.16%	59.60%	36.24%
	Tail	2.83%	47.80%	49.37%
Expert	**Head**	5.82%	60.92%	33.26%
	Mid	3.45%	61.63%	34.92%
	Tail	1.62%	44.83%	53.55%

Table 2: Artist similarity and their location in the Long Tail. Given an artist, a_i, it shows (in %) the Long Tail location of its similar artists (results are averaged over all artists). Each row represents, also, the Markov chain transition matrix for CF, CB, and expert–based methods.

artists are located in the head part, that contains only 82 artists, and the rest in the mid area. Artists in the mid part are tightly related to each other, and only 1/5 of the similar artists are in the tail part. Finally, given an artist in the tail, its similar artists remain in the same area. Contrastingly, CB and expert–based promote much more the mid and tail parts in all the cases (specially in the head part).

Moreover, a Markovian stochastic process [16] is used to simulate someone surfing the recommendation network. Indeed, each row in Table 2 can be seen as a Markov chain transition matrix, M, being the head, mid and tail parts the different states. The values of M denote the transition probabilities, $p_{i,j}$, between two states i, and j (e.g. $p^{CF}_{head,mid} = 0.5468$). The Markovian transition matrix, M^k, denotes the probability of going from any state to another state in k steps (clicks). The initial distribution vector, $P^{(0)}$, sets the probabilities of being at a determined state at the beginning of the process. Then, $P^{(k)} = P^{(0)} \times M^k$, denotes the probability distribution after k clicks, starting in the state defined by $P^{(0)}$.

Using $P^{(k)}$ and defining $P^{(0)} = (1_H, 0_M, 0_T)$, we can get the probability of reaching the tail, starting in the head part. Table 3 shows the number of clicks needed to reach the tail from the head, with a probability $p_{head,tail} \geq 0.4$. In CF, one needs five clicks to reach the tail, whereas in CB and expert–based only two clicks are needed.

Finally, the stationary distribution π is a fixed point (row) vector whose entries sum to 1, and that satisfies $\pi = \pi M$. The last two columns in Table 3 present the stationary distribution vector for each algorithm, and the number of steps to converge to π, with an error $\leq 10^{-6}$. CF needs more than three times the number of steps of CB or expert–based in order to reach the steady state. Even though the probability to stay in the tail in CF is higher than CB and expert–based, this is due to the high probability to remain in the tail once is reached ($p^{CF}_{tail,tail} = 0.8260$).

4.3.2 Artist indegree

Up to now, we have analysed popularity in terms of relationships among the artists. Now we analyse the correlation between artists' indegree (potential hubs in the network) and its popularity. As a starting point, we present in Table 4 the top–10 indegree artists for each network. CF and

Method	k	$\mathbf{P^{(k)}}$, with $P^{(0)} = (1_H, 0_M, 0_T)$ and $p_{head,tail} \geq 0.4$	π	n
CF	5	$(0.075_H, 0.512_M, 0.413_T)$	$(0.044_H, 0.414_M, 0.542_T)$	26
CB	2	$(0.038_H, 0.562_M, 0.400_T)$	$(0.037_H, 0.550_M, 0.413_T)$	7
Expert	2	$(0.030_H, 0.560_M, 0.410_T)$	$(0.027_H, 0.544_M, 0.429_T)$	8

Table 3: Long Tail navigation in terms of a Markovian stochastic process. Second and third columns depict the number of clicks (k) to reach the tail from the head part, with a probability $p_{head,tail} \geq 0.4$. Fourth and fifth columns show the stationary distribution π, as well as the number of steps, n, to reach π.

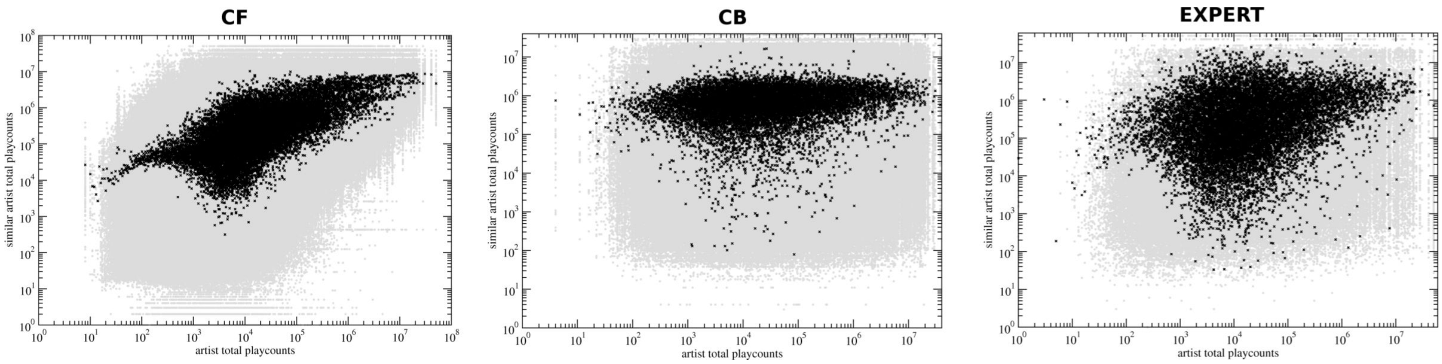

Figure 3: A log–log plot depicting the correlation between artist total playcounts and its similar artists (average values are depicted in black, whilst grey dots display all the values). Pearson r values are: $r_{CF} = 0.46$ $r_{CB} = 0.08$, and $r_{EX} = 0.09$.

CF			CB			Expert		
k_{in}	Artist	LT pos	k_{in}	Artist	LT pos	k_{in}	Artist	LT pos
976	Donald Byrd	6,362	1,955	George Strait	2,632	180	R.E.M.	88
791	Little Milton	19,190	1,820	Neil Diamond	1,974	157	Radiohead	2
772	Rufus Thomas	14,007	1,771	Chris Ledoux	13,803	137	The Beatles	1
755	Mccoy Tyner	7,700	1,646	The Carpenters	1,624	119	David Bowie	62
755	Joe Henderson	8,769	1,547	Cat Stevens	623	117	Nirvana	19
744	R.E.M.	88	1,514	Peter Frampton	4,411	111	Tool	17
738	Wayne Shorter	4,576	1,504	Steely Dan	1,073	111	Pavement	245
717	U2	35	1,495	Lynyrd Skynyrd	668	109	Foo Fighters	45
712	Horace Silver	5,751	1,461	Toby Keith	2,153	104	Soundgarden	385
709	Freddie Hubbard	7,579	1,451	The Charlie Daniels Band	22,201	103	Weezer	51

Table 4: Top–10 artists with higher indegree (k_{in}) for each recommendation network (spikes in Figure 4). The table shows too, the artist ranking in the Long Tail (LT pos).

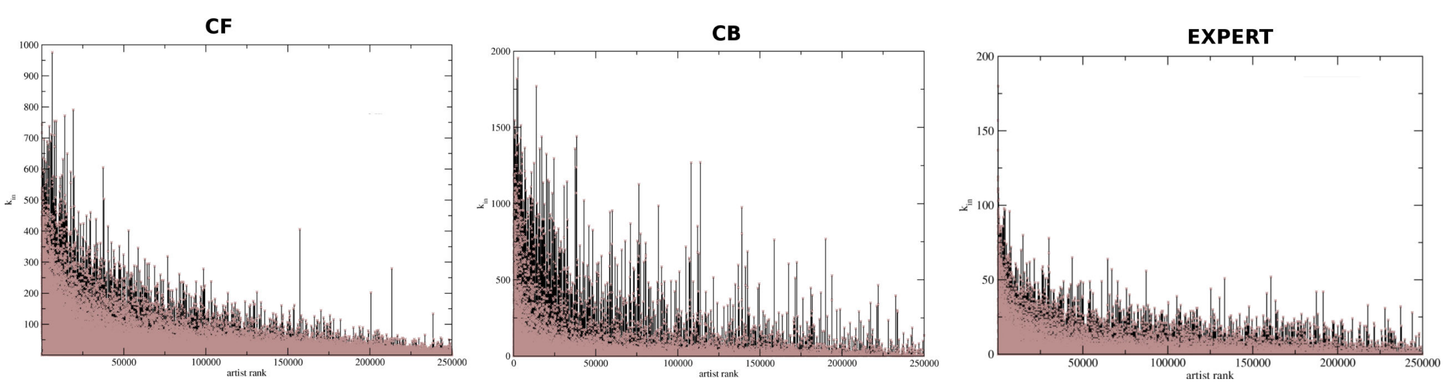

Figure 4: Artist rank in the Long Tail and its indegree, k_{in} (y axis). CF (left) concentrates most of the hubs in the most popular artists—head and mid parts—, whilst in CB (mid), and expert–based (right) hubs are spread out through the whole Long Tail.

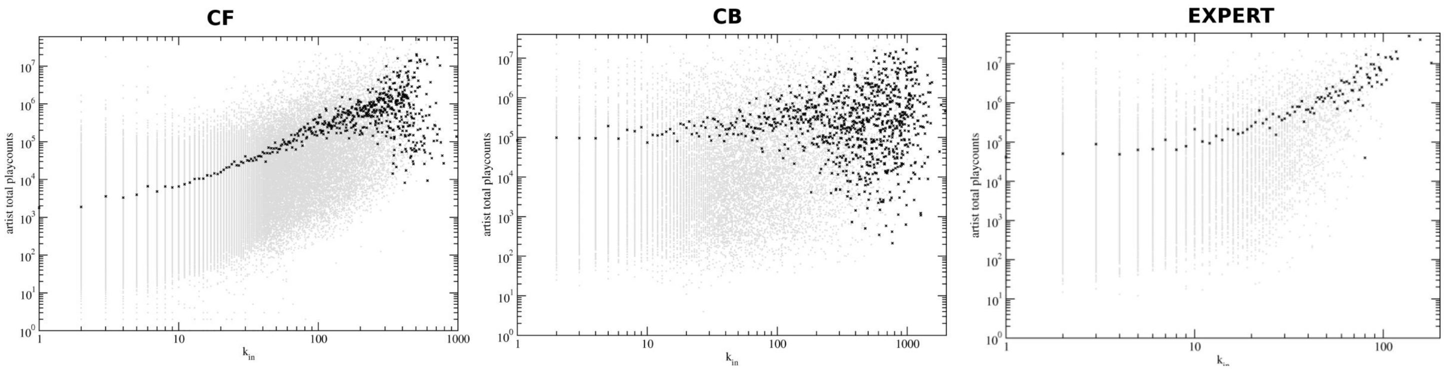

Figure 5: A log–log plot showing the correlation between artist indegree (k_{in}, in horizontal axis) and its total playcounts (avg. values in black), in vertical axis. Pearson r values are: $r_{CF} = 0.38$, $r_{CB} = 0.10$, and $r_{EX} = 0.69$.

expert–based contains two and eight mainstream artists, respectively. CF contains *U2* and *R.E.M.*, but the rest of the list if made of more or less well known Jazz musicians, including some in the—top of the—tail part. The whole list in expert–based AMG is made of very popular artists. Our guess is that the editors connect Long Tail artists with the most popular ones, either for being influential or because a lot of bands are *followers* of these mainstream artists. On the other hand, CB has a more eclectic top–10 list, as one could expect. Oddly enough, there is no new or actual artists, but some classic bands and artists ranging several musical genres. Some bands are, in fact, quite representative of a genre (e.g. *Lynyrd Skynyrd*, and *The Charlie Daniels Band* for Southern Rock, *The Carpenters* for Pop in the 70's, *George Strait* for Country, and *Cat Stevens* for Folk Rock). Probably, the indegree is due to being very influential in its respective musical styles. In some sense, there are some bands that "cite" their music (i.e. sound similar). Although, these results might be somewhat biased; CF and AMG networks are subsets of their whole similar artists' graph, thus our sampling could not be a good representation of the whole dataset. Furthermore, the differences in the maximum indegree value (k_{in} for top–1 artist) among the three networks are due to the different sizes (N) and average degree $\langle k \rangle$, but mostly to the topology of the networks; CF and CB follow a power–law cumulative indegree distribution, whereas AMG has an exponential decay. Therefore, AMG maximum indegree, k_{in}, is much smaller than the CF and CB ones.

Figure 4 depicts the artist rank in the Long Tail and its indegree in the network. The figure shows whether the artists with higher indegree in the network (hubs) are the most popular artists, in terms of total playcounts. In both cases, CF and expert-based networks, the artists with higher indegree (hubs) are mostly located in the head and mid part, whereas in CB they are more spread out through all the curve. In a similar way, Figure 5 presents the correlation between artist indegree (k_{in}), and total playcounts. Again, both CF and AMG expert–based confirm the expectations, as there is a clear correlation between the artist indegree and its total playcounts ($r_{CF} = 0.38$, $r_{EX} = 0.69$). Artists with high indegree are the most popular ones. In CB, given a high indegree value it contains—on average—artists ranging different levels of popularity ($r_{CB} = 0.10$).

5. DISCUSSION

The results show that *last.fm* CF tends to reinforce popular artists, at the expense of discarding less–known music. Thus, the popularity effect derived from the community of users has consequences in the recommendation network. This reveals a somewhat poor discovery ratio when just browsing through the network of similar music artists. It is not easy to reach relevant Long Tail artists, starting from the head or mid parts (see Table 3). Moreover, given a long tail artist, its similar artists are all located in the tail area, too. This do not always guarantee novel music; a user that knows quite well an artist in the Long Tail is likely to know most of the similar artists, too (e.g. the solo project of the band's singer, collaborations with other musicians, and so on). Thus, these might not be considered good novel recommendations to that user, but familiar ones. CF contains, then, all the elements to conclude that popularity has a strong effect in the recommendations because: *(i)* presents assortative mixing (indegree–indegree correlation) in Figure 2, *(ii)* there is a strong correlation between an artist total playcounts and the total playcounts of its similar artists (see Figure 3), *(iii)* most of the hubs in the network are popular artists (see Figure 5), and *(iv)* it is not easy to reach relevant Long Tail artists, starting from the head or mid parts (see Table 3).

Human expert–based recommendations are more expensive to create, and also have a smaller Long Tail coverage compared to automatically generated recommendations like CF and CB. In terms of popularity, the hubs in the expert network are comprised by mainstream music, thus potentially creating a network dominated by popularity (see Table 4 and Figure 5). However, the topology—specially the exponential decay in the indegree distribution—indicates that these artists do not act as hubs. Moreover, it does not present assortative mixing (see Figure 2), so artists are linked in an heterogeneous way; popular artists are connected with other less–known artists (see Table 2 and Figure 3). According to the stationary distribution π (see Table 3), the key Long Tail area in CB and expert–based AMG are the artists located in the mid part. These artists allow to navigate inside the Long Tail acting as entry points, as well as main destinations when leaving the Long Tail. Users that listen to mainly very unknown music are likely to discover artists that are in the mid part, and that are easily reachable from the artists in the tail. One should pay attention, too, to the quality data in the Long Tail. Assuming that there exists some extremely poor quality music, CB is not able to clearly discriminate against it. In some sense, the popularity effect drastically filters these low quality items. Although, it has been proved in [19] that increasing the strength of social influence increased both inequality and unpredictability of success and, as a consequence, popularity was only partly determined by quality.

Finally, we need to evaluate the quality of the relationships among artists, as well as the popularity effect when providing novel, unknown recommendations to the users. Without any user intervention, then, it is impossible to evaluate the quality and user satisfaction of the recommendations, which does not necessarily correlate with predicted accuracy [15]. In this sense, our incoming work [8] presents a user–centric experiment done with 288 subjects and 5,573 rated songs. The results indicate that even though CF recommends less novel items than CB and expert–based, the users' perceived quality is better than those recommended by CB and human expert methods.

6. CONCLUSIONS

Recommender systems should assist us in the process of filtering and discovering relevant information hidden in the Long Tail. In our experiments, popularity is the element that defines the characteristic shape of the Long Tail. In this sense, we have analysed the popularity effect in three different music recommendation approaches. We measure popularity in terms of total playcounts, and the Long Tail model is used in order to rank all music artists. As expected by its inherent social component, the collaborative filtering approach is prone to popularity bias. This has some consequences on the discovery ratio as well as in the navigation through the Long Tail.

Future work includes expanding the analysis of the recommendation network, taking into account its dynamics. This could be used, for instance, to detect "hype" items, that become popular in a very short period of time.

7. ACKNOWLEDGMENTS

This work was partially funded by the Pharos (IST-FP6-45035) and Variazoni (e-Content plus) projects, sponsored by the European Commission.

8. REFERENCES

[1] C. Anderson. *The Long Tail. Why the future of business is selling less of more*. Hyperion Verlag, 2006.

[2] A. Anglade, M. Tiemann, and F. Vignoli. Complex-network theoretic clustering for identifying groups of similar listeners in p2p systems. In *Proceedings of the ACM conference on Recommender systems*, pages 41–48, Minneapolis, USA, 2007. ACM.

[3] J.-J. Aucouturier and F. Pachet. A scale-free distribution of false positives for a large class of audio similarity measures. *Pattern Recognition*, 41(1):272–284, 2008.

[4] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[5] A.-L. Barabási, R. Albert, H. Jeong, and G. Bianconi. Power-law distribution of the world wide web. *Science*, 287:2115a, 2000.

[6] P. Cano, Ò. Celma, M. Koppenberger, and J. Martin-Buldú. Topology of music recommendation networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(013107), 2006.

[7] P. Cano, M. Koppenberger, and N. Wack. An industrial-strength content-based music recommendation system. In *Proceedings of 28th International ACM SIGIR Conference*, Salvador, Brazil, 2005.

[8] Ò. Celma and P. Herrera. Evaluating the quality of novel recommendations. In *Proceedings of the 2nd ACM International Conference on Recommender Systems*, Laussane, Switzerland, 2008.

[9] D. M. Fleder and K. Hosanagar. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *SSRN eLibrary*, 2007.

[10] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, 2004.

[11] K. Jacobson and M. Sandler. Musically meaningful or just noise? an analysis of on-line artist networks. In *Proceedings of the 6th International Symposium on Computer Music Modeling and Retrieval*, Copenhagen, Denmark, 2008.

[12] K. Kilkki. A practical model for analyzing long tails. *First Monday*, 12(5), May 2007.

[13] J. M. Kleinberg. Navigation in a small world. *Nature*, 406:845, 2000.

[14] J. Martin-Buldú, P. Cano, M. Koppenberger, J. Almendral, and S. Boccaletti. The complex network of musical tastes. *New Journal of Physics*, 9(172), 2007.

[15] S. M. Mcnee, J. Riedl, and J. A. Konstan. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *Computer Human Interaction*, pages 1097–1101, New York, USA, 2006. ACM.

[16] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer–Verlag, 1993.

[17] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20), 2002.

[18] J. Park, Ò. Celma, M. Koppenberger, P. Cano, and J. Martin-Buldú. The social network of contemporary popular musicians. *International Journal of Bifurcation and Chaos*, 17(7):2281–2288, 2007.

[19] M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, February 2006.

[20] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In A. Press, editor, *Proceedings of 10th International World Wide Web Conference*, pages 285–295, Hong Kong, 2001.

[21] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.