

The Distribution of Online Reputation: Audience and Influence of Musicians on MySpace

Beuscart Jean-Samuel, Couronné Thomas

Orange Labs, 38-40 rue Général Leclerc, F-92794 Issy Les Moulineaux Cedex 9, France

{JeanSamuel.Beuscart, Thomas.Couronne}@Orange-ftgroup.com

Abstract

The user-generated content platforms like MySpace, YouTube or Flickr produce emergent sociability patterns and behaviors, which are well observed in various studies. Few of them deal with the nature of the reputation in these large social networks. In this work, we extract a sample of MySpace Music by parsing the artists' pages and following the "best friend" links. The study of the network shows that the online fame is composed by two independent features: the audience, dominated by the major companies' artists, and the influence, more associated with independent labels. We observe that most of the audience is focused on a few stars: the tail of MySpace Music is long, but flat. The network structure shows that artists tend to link to artists of similar reputation.

Introduction

In recent years, many online publication tools have been developed, providing an easy way for both amateurs and professionals to broadcast their written productions, photography, music, or video. The success of such platform is undeniable, and the biggest platforms (MySpace, YouTube, Flickr, Wordpress, etc.) are now among the most visited and traffic-generating websites. This success raises questions about the nature of the audience on User-Generated Content portals, and about the way the whole audience is distributed among the proposed contents:

- Is there a focus on a few particular stars, as observed in the offline cultural markets? Or, consistently with the intuition of the "long tail", do we observe that attention to online contents is more evenly spread all over the available content?
- Which contents receive the largest attention? Do the amateur productions receive the same audience as those of established professionals?
- Does the success on platforms generate a coherent and interconnected online elite?

These issues are important to understand better the ways the new media are transforming the structure of the cultural industries.

1 – The distribution of online reputation: related works

Several studies were devoted to the distribution of the online attention, in a context of maximal offer. Researches in the field of computer sciences have shown that, for many Internet products (web pages, catalogs of music and video online content, UGC), in an abundant supply context, the demand distribution is highly focused (Cho and Roy, 2004; Anderson, 2006; Cha *et al*, 2007). This distribution can often be modeled by a power law. This demand curve is similar to that of the traditional demand for cultural goods, usually summarized by the cultural industry actors as the "80-20 rule", which states that 80% of the demand covers 20% of products. It remains uncertain, however, whether the Internet enhances or reduces this unequal distribution.

In his book about "the long tail", Anderson (2006) argues that the search and recommendation tools (rather than the traditional marketing tools) will naturally bring some of the demand from the top of the curve (star products) to the bottom (the long tail) and will cause the curve thickening. At this point, the reasoning is purely speculative. Nevertheless, empirical studies (Elberse, Oberholzer, 2007) suggest that the thickening of the curve is minor. None of these studies considered the UGC portals such as MySpace or YouTube, which are specific in the sense of compiling the works of professional artists, amateurs, and a myriad of fans in the same catalog.

One major explanation of this power law trend is based on the differential exposure of products and its consequences. The intuition suggests that "the most demanded products are the most viewed and, therefore, the most exposed". There are several classical formulations of this idea, ranging from the mathematical one, based on Yule process of "preferential attachment", to "the Matthew effect", proposed by Merton in the field of sociology of scientific reputation (Merton, 1968). Intuitively, it may be suggested that certain automated scoring tools (number of views, votes, page rank) will enhance the Matthew effect, while

others (personal recommendations, social networks) will decrease this effect.

Thus, evaluating the recommendation tools appears to be crucial for understanding reputation distribution. On the one hand, some studies on YouTube or Flickr reveal clear correlation between the assessment and the content popularity (Cha *et al.*, 2007; Prieur *et al.*, 2008). On the other hand, the study of Herring *et al.* (2005) of the visibility in the blogosphere shows that the most popular blogs (A-list) tend to be highly cited by each other, while the less-known blogs cite the popular ones without being cited by them. In other words, the recommendation mechanisms can increase rather than decrease the exposure differences between the elites and the mass.

As an extension of the studies discussed above, the present work helps to understand a) what kind of attention distribution is produced by the recommendation and network tools provided by MySpace and b) the links between the online and the offline attention distributions.

2 – Data and Methodology

In this work, we examine the questions formulated above on a subgraph of 13,859 MySpace users, including 8517 musicians, interconnected by recommendation links.

Our previous qualitative studies (Beuscart, 2008) have shown that the main horizontal recommendation tool on MySpace is to mention an artist as "best friend": when a MySpace member (artist or not) declares an artist as "best friend", the declaration appears visibly on its own page, and creates de facto a link to the artist's page, exactly like a blogroll. This link is interpreted by MySpace users as a recommendation or a musical affinity. In line with our objective, we sought to gather information on how the recommendation works on the site, constructing a network sample by following the links between individuals, *i.e.*, using the «snowball » method (Herring *et al.*, 2005; Carverlee and Webb, 2008).

In order to verify that the sample is not unusual, we have made several networks aspirations, varying the entry points numbers (from 3 to 10) and the parsing depth (from 2 to 4). We obtain between 9000 and 27000 nodes, with an average music profile equal to 0.53. A correlation test is then applied on the four variables: comments, friends, in-degree and hits, for each sample. Then a Mantel test is performed on the four correlation table, showing that the coefficients are significantly similar, *i.e.* the variables of each sample are correlated in the same proportions ($r > 0.99$; $p < 0.001$). The network is finally formed by 13,859 nodes (accounts), in which 8517 musicians, linked together by 76,017 edges (best friendship declaration). The network's density is 0.0004, and 27% of links are reciprocal.

The data also contains information printed on the MySpace pages about the artists. We record the musician's position in the industry: 1191 are labeled as "major", 4604 as "independent label", while 2664 are declared "unsigned". For each artist, three indicators are recorded: artist page hits (total number of visits), total number of friends, and

total number of comments left on the page. A fourth indicator is extracted from the network's sample, the In-degree of the node, which is in fact the number of users declaring the artist as "best friend" (recommendation activity). The table below provides a brief description of these indicators.

		Friends	Hit	In-Degree	Comments
Mean		22656,8	754424,8	6,6	7676,7
Median		3885,5	66600,0	4,0	888,5
Standard-dev.		91358,9	4411430,4	9,0	78204,8
Minimum		1,0	0,0	1,0	0,0
Maximum		3485616,0	165677295,0	183,0	3791342,0
Centiles	25	1442,5	24296,5	2,0	321,0
	50	3885,5	66600,0	4,0	888,5
	75	11637,8	215495,8	8,0	2470,0

Table. 1: Statistical description of the reputation variables

These four reputation indicators are unsurprisingly characterized by a strong dispersion. The standard deviation for each variable is above the mean of the variable. For example, artists in our sample have an average of 750 000 pages hits, but one fourth have less than 25 000.

Beyond further analysis of the determinants of this dispersion, it is interesting to establish the correlations between the variables. The correlations analysis of the four variables shows if there is one or more dimensions describing the online reputation. Three indicators are highly correlated with each other: the total number of page hits, the total number of friends and the received comments. In contrast, the In-degree (best friend declaration) is weakly correlated with these three indicators: the average correlation is around 0.17. One may object that the two kinds of indicators have not been constructed in the same way: the first three measure the amount of actions (visits, friendship, comment) of all the MySpace members on the artist profile, while the fourth reflects the actions of the MySpace music sample ("best friendship" declaration). The effect is nevertheless strong enough to suggest that the online reputation has two features: the first one is the audience, characterized by the hits, comments, and the overall number of friends; the second one is the influence, characterized by the in-degree, indexing the position in the recommendation network.

3 – The distribution of online reputation

Online Reputation depends on the position in the music industry

A variance analysis of the label factor (three modalities: Major, Independent, None) is performed on the variables comments, friends, hits and in-degree, and demonstrates a significant effect of the label factor on the variables amount of comments, amount of friends, pages views (hit), in-degree ($p < 0.05$). Thus, an artist reputation on MySpace depends on its position in the music industry. Artists

signed on a “major label” are more visited, commented and influential than independent artists, which are themselves more visited and influential than “unsigned” artists (fig 2).

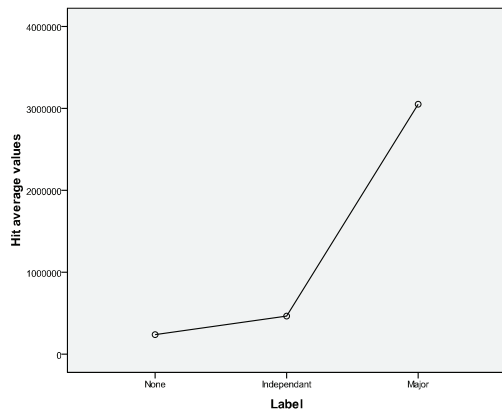


Fig. 1: Hits average per artist's label

MySpace does not overturn the established offline hierarchies. This result, if it contrasts with the classical reputation of MySpace as a place promoting new emerging artists, is not surprising. Indeed, it makes sense that people on MySpace primarily look and listen to artists they know, even if they secondarily follow the recommendation system suggestions. Moreover, there are many ways to access to the MySpace page of an artist, which do not necessarily pass via the homepage of the site, its search engine, links or rankings. It is therefore understandable that the site statistics reflect the reputation of artists established in the classical (offline and non-social) media.

MySpace's Tail is long, but flat

The musical content of MySpace follows the same rule as most of the online content: a minority of artists receives the main part of the audience. The distribution trend of our variables demonstrates a power law (fig.3), so 10% of the artists represent over 90% of hits (page views) of our sample. A similar distribution was found for the other indicators, the declination being relatively lower for the In-degree, because of a smallest standard deviation.

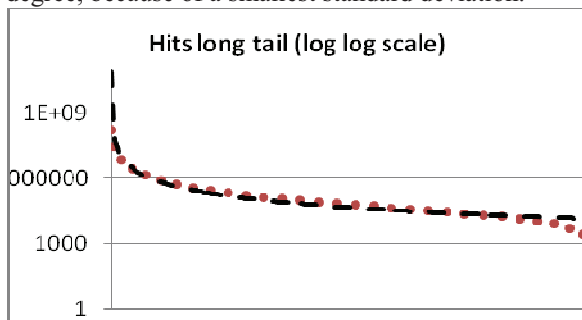


Fig. 2: Log-log plot of Hits distribution; red dot : Hits distribution; black dash: power law trend

Who is the elite on MySpace Music?

In the aim to know who belongs to the MySpace's elite, according to our two dimensions, we focus on the upper

percentile of two main indicators: the page hits (audience) and the In-degree (influence).

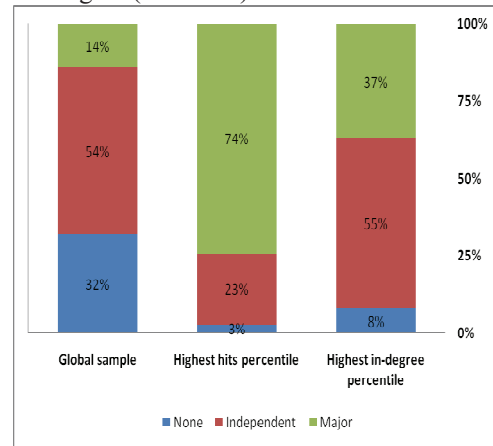


Fig.3 Label distribution per sub-sample

If the artists labeled as “majors” are the undisputed stars of the audience on MySpace (they constitute 73% of the upper hit percentile, against 14% in the sample), the artists labeled as independent constitute an important part of the most influential artists (55% of the upper influential percentile, against 54% of the sample – fig.9).

Among the ten most visited artists of the sample, 8 artists belong to a major, essentially American R&B superstars in European and US charts at the extraction period: Lil'Wayne, Chris Brown, TI, Akon, etc. The two non-major artists of this top ten are, surprisingly, artists without label: Tila Tequila and Jeffree Star.

Among the ten most influential artists of the sample we find eight independent artists and only two majors. These artists belong to independent labels with innovative artistic and promotion strategy, based on an intensive and smart use of social tools such as MySpace. They form a coherent music scene (electro-rock: EdBanger, Justice, MIA...), highly inter-linked.

Artists from majors are more visited, independent more recommended

We have seen that the most visited artists belong to major labels, whereas the most recommended ones are associated with independent labels. This trend is also verified for the mass. To avoid distortion engendered by the high dispersion in the values distribution, we exclude artists belonging to the hit and In-degree upper deciles; we obtain a new sample of 6594 "ordinary" artists.

A hierarchical ascendant classification (Ward criteria – Euclidian distance) is then applied on the variables, identifying the structure of the sample. The optimal typology provides 5 classes of artists.

The first class, the largest (41% of the sample), gathers people whose notoriety is low for all dimensions (friends, hits, comments, recommendations) and is significantly associated with being a not-signed artist. The second class (23% of the sample) consists of individuals who are highly

influential, with a relatively limited audience, strongly associated with belonging to an independent label. Conversely, the fifth group (4% of the sample) consists of artists that have a high audience (in terms of hits, comments and friends), and are labeled as major. Classes 3 and 4, 17% and 9%, are less significant: they include respectively influential and highly visited artists, but are not explained by the "label" modality.

4 – Understanding the network

We finally examine if the elite is interconnected. In their papers about the blogs, Herring *et al.* show that the popular blogs tend to be interconnected and to ignore the less popular blogs. In the case of MySpace, do the well-known artists promote lesser-known artists, or are the popular artists connected reciprocally but exclusively together? Conversely, do the confidential artists recommend each other depending on their musical style, or do they prefer to display prestigious friendship?

We divide our population in ten deciles based on the number of hits; we compare, for each edge of the network, the decile rank of the origin node and the decile rank of the receiver node. 24% of all the recommendations of the sample are directed to the elite (highest decile), which is consistent with the previous results, but 53% of the recommendation issued by these elites are directed to another elite member. A clear trend appears, the "stars" recommending each other inside an elitist cast. We can generalize this result with a density graph, ie. the matrix $M(i,j)=n$ where n is the number of links from the artists belonging to the i -th decile to artists belonging to the j -th decile. Finally the matrix is normalized by the total number of links. It provides a density estimation of the inter/intra-deciles edge distribution.

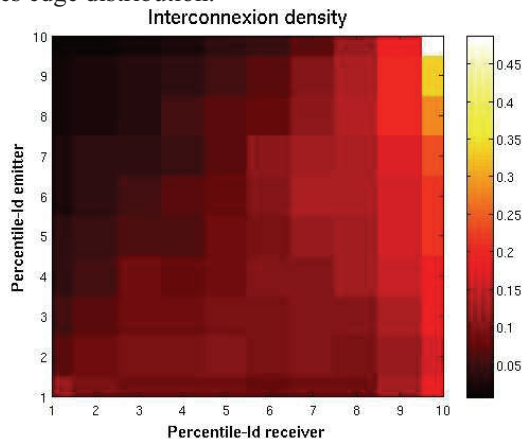


Fig.4: Density distribution of the links between artists labeled per Hit's percentile

In abscise is the decile rank of the link receiver; in ordinate is the decile rank of the link emitter. The lighter the color is, the more the density is strong, the more the occurrence number is high. It means that the top deciles mention only very few the lower deciles (black color), and much more the highest deciles (red, yellow and white).

Conclusion

Our exploration of a snowball sample of MySpace Music provides several results. First, we show that the distribution of reputation among artists fits a power law: most of the audience is concentrated on a few stars. Second, we observe that these stars are mostly established music stars: artists signed on major labels receive the most attention. Third, we identify two de-correlated dimensions of reputation: audience and influence. High audience values are strongly associated with artists belonging to major labels, whereas influence (understood as "being the object of recommendation") is more associated with independent artists. Fourth, we show that artists with a strong audience tend to link one to each other. But the network is also structured by musical affinities. MySpace appears as the echo chamber of artistic reputations built offline by the cultural industries; nevertheless, the platform provides social recommendation tools that tend to favor independent labels, and that draw relevant musical maps.

References

- Anderson, C. 2006. *The Long Tail: How the Future of Business is Selling Less of More*, New York, Hyperion Books.
- Beuscart, J.S. 2008. "Online interactions, virtual reputation and artistic career. The uses of MySpace by self-produced musicians", *Réseaux*, n°152.
- Caverlee, J. & Webb, S. 2008. "A Large-Scale Study of MySpace: Observations and Implications for Online Social Networks", presented at the ICWSM 2008 Conference.
- Cha, M., Harwoon, K., Rodriguez, P., Ahn, Y.Y., Moon, S. 2007. "I Tube, You Tube, Everybody Tubes: Analysing the World's Largest User Generated Content Video System", presented at the IMC'07, San Diego, California.
- Cho, J & Roy, S. 2004. "Impact of Search Engines on Page Popularity", presented at the WWW2004 Conference, New York.
- Elberse, A. & Oberholzer, F. 2008. "Superstars and Underdogs: an Examination of the Long Tail Phenomenon in Video Sales", Harvard Business School Working Papers Series.
- Herring, S.C et al.. 2005. "Conversations in the Blogosphere: An Analysis "From the Bottom Up", *Proceeding of the Thirty-Eighth Hawaii International Conference on System Sciences*, Los Alamitos, IEEE Press.
- Merton, R.K. 1968. "The Matthew Effect in Science", *Science*, n°159, pp.56-63.
- Prieur, C., Cardon, D., Beuscart, J.S., Pissard, N., Pons, P. 2008. "The Strenght of Weak Cooperation: A Case Study of Flickr", arXiv:0802.2317.