7-1-2013

Using The Wisdom Of The Crowd To Predict Popular Music Chart Success

Dennis M Steininger
University of Mannheim, Mannheim, Germany, dennis.steininger@wiwi.uni-augsburg.de

Simon Gatzemeier
University of Mannheim, Mannheim, Germany, simon.gatzemeier@gmail.com

USING THE WISDOM OF THE CROWD TO PREDICT POPULAR MUSIC CHART SUCCESS

Steininger, Dennis M., University of Augsburg, Universitaetsstr. 16, 86135 Augsburg, Germany, dennis.steininger@wiwi.uni-augsburg.de

Gatzemeier, Simon, University of Mannheim, Schloss, 68131 Mannheim, Germany, simon.gatzemeier@bwl.uni-mannheim.de

Abstract

The peculiarities of the recording industry system, such as fashion cycles, the hedonic nature of music, socio-network effects, informal heuristics in the decision-making process of recording companies, and the opaque selection process of media gatekeepers have created uncertainty about the chart potential of musical products. With respect to the on-going digital transformation and shift in power from organizations to consumers, we leverage the principles of crowdsourcing to build a prediction model for understanding chart success. Therefore, we investigate the causal relationship between crowd evaluations based on listening experience and popular music chart success. We use 150 music songs and track their live cycles with entry and peak positions in reported music charts. Additionally, we carry out about 20 evaluations by the crowd for each song resulting in a total of 2.852 observations. Our findings indicate that the crowd has predictive relevance concerning popular music chart success. However, this predictive relevance is bound to certain conditions, namely the composition of the crowd, the underlying chart and market mechanisms and the novelty of the musical material. In sum we find that crowd-based mechanisms are especially suited for predicting the performance of novel songs from unknown artists, which makes them a powerful decision support instrument in very uncertain contexts with limited historical data availability.

Keywords: Crowdsourcing, Wisdom of the Crowd, Recording Industry, Popular Music, Chart Success, Hedonic Consumption Model, Music Consumption.

1 Introduction

"I'm not interested in forcing my music on people, and that's what the whole music industry nowadays is based on is forcing stations to play it, forcing people to listen to it." (John Frusciante, Red Hot Chili Peppers)

Even though it is overly critical, this quote basically summarizes the traditional model about the diffusion of innovations in cultural industries, such as the recording industry (Hirsch, 1972). A selection from the vast supply of creative content is developed, bundled into marketable entertainment products and distributed to the consumers under high promotional efforts with the hope of passing the media gatekeepers, such as Radio or Television (Young and Collins, 2010). Once a product has passed these gatekeepers it will be repeatedly exposed to the largest possible audience, whose attention is then focused on this elevated set and ideally transformed into commercial transactions. As a result, the commercial success of recorded music depends on several influential factors, which are not in the artists' or recording companies' locus of control.

First, the consequences of the digital transformation, such as music piracy behaviour or lower per unit revenues of digital content, have diminished the total profitability of the industry (cf. Lam and Tan, 2001). Second, the selection criteria of the recording companies are based on informal heuristics, since the hedonic nature of music prevents an objective measurement of its quality. Third, direct and large-scale consumer engagement has been expensive. In order to increase the chances of reaching the consumer, the products are therefore developed and bundled according to the expected consumer taste formulated by the media gatekeepers (cf. Hirsch, 1972). Fourth, consumer taste is susceptible to fashion cycles, socio-network effects and decision difficulties due to increasing choice of consumption, which result in diversity and uncertainty about expected demand. Fifth, the selection criteria of media gatekeepers are arbitrary and opaque from the perspective of artists and recording companies (Ordanini, 2006). It is "working knowledge" in the recording industry, that revenues are distributed according to the power law, meaning that the revenues of few commercial successes have to compensate for the losses of the large amount of failures (Howe, 2009). Hence, there is high uncertainty about the outcome of the production process in the recording industry. Consequently, cultural producers, i.e. all actors involved in creating, producing and marketing an experiential good, such as musicians, writers, directors, record labels (Seiter, 1986), need more accurate and complete information about the potential of their creative outputs.

Several researchers have tried to accommodate this information need by developing econometric prediction models in order to find patterns and regularities of successful experiential goods based on their original value, stemming from the intrinsic (e.g., song features such as tempo, loudness) or extrinsic features (e.g., chart history of an artist) (Bhattacharjee et al., 2007). Nevertheless, objective approaches fail to take into account the derivative value arising from the interaction between consumers and music (Lorenzen and Frederiksen, 2005; Prince, 1972). We tie in on this notion and intend to close this research gap by building a prediction model based on large-scale subjective crowd evaluations of musical products (Celma and Serra, 2008; Holbrook, 2006). Crowdsourcing, i.e. giving a complex problem to an anonymous mass of people (the crowd), is seen as a tool to support management decisions due to its ability to "balance [a crowd] between diversity and expertise" (Bonabeau, 2009, p. 47). As commercial success is determined through chart rankings, we formulate the following research question using these mechanisms: *Can the wisdom of the crowd forecast chart success and thereby support artist investing decisions of the music industry?* We aim to contribute valuable insights about the predictive ability of the crowd to the body of research concerned with Crowdsourcing, Human Computation and Decision Support Systems. We are also convinced about the practical relevance of this approach to strengthen the position of recording companies. The remainder of the paper is structured as follows. We explain the foundations of crowdsourcing and chart success and our hypotheses development in the second section and give details on our research design in section three. Thereafter, we report and discuss our results in the sections four and five and conclude with section six.

2 Background and Hypotheses Development

We begin our endeavour through qualitative interviews with experts from the music industry. These first insights inspire our hypotheses development and therefore guide our literature reviews on popular music chart success, the decision-making process on new songs in the music industry, and crowdsourcing.

2.1 Success and Decision-Making in the Music Industry

Commercial success in the recording industry is tracked in charts, reflecting the relative performance of creative input in comparison to other material. Being successful in the charts is the most important key performance indicator in the industry. Recording companies track peak positions and duration on the charts in order to get insights about the quality and lifecycle of their musical material (Bhattacharjee et al., 2007). Since musical material rather has a short lifecycle, promotional resources are allocated to the initial release in order to elevate the entry position, which is also often the peak position, because the material then falls to lower ranks and eventually drops out of the charts (Anand and Peterson, 2000). According to (Hirsch, 1972), music is an intangible experiential good "directed at a public of consumers, for whom [it] generally serve[s] an [a]esthetic or expressive, rather than a clearly utilitarian function" (p. 642). In addition, music is an integral part of human communication and fulfils emotional, social or cognitive functions (Lacher and Mizerski, 1994). Research has criticized the traditional view on consumption and proposed the alternative hedonic consumption paradigm. "Consumption has begun to be seen as involving a steady flow of fantasies, feelings, and fun encompassed by what we call the 'experiential view'. This experiential perspective is phenomenological in spirit and regards consumption as a primarily subjective state of consciousness with a variety of symbolic meanings, hedonic responses, and [a]esthetic criteria" (Holbrook and Hirschman, 1982, p. 132). Consequently, since the evaluation of music is bound to the experience, there is no standard measure to compare music quality, because the derived value is due to the subject (Styvén, 2007).

The managerial subsystem is characterized by a de facto oligopolistic structure with few multinational organizations, the major labels, and a long tail of small and medium independent labels (Rayna and Striukova, 2009). These organizations bundle and transform creative raw material into entertainment products. Thereby, the core asset is the input interface, which contains the Artist & Repertoire (A&R) department. Similar to the function of R&D departments in organizations of other industries, the A&R department is responsible for the discovery, selection and management of talents (Negus, 1992). Depending on the corporate strategy, special composed A&R teams are screening the vast amount of creative input for promising and trending talents to invest in, which will complement the existing portfolio. Nowadays, A&R managers mostly rely on their professional network in order to find new talents, since recommendations from peers and acquainted opinion leaders communicate credibility (Neelamegham and Jain, 1999). The initial step is then the evaluation of the talent's marketability. Another important criterion is the quality of music. But there is no standard indicator for the quality of hedonic experiential goods. Therefore, A&R managers use surrogate measures to describe the quality level, such as the innovativeness, uniqueness and authenticity of a song (Zwaan and Ter Bogt, 2009). In addition, to better circumscribe the quality, A&R managers often anchor the talents to examples of similar and already established artists (Negus, 1992). Furthermore, intuition is commonly applied in ascertaining the potential of talents. A&R managers trust their gut feeling, which they have acquired from professional experience, knowledge and personal interest in music. Also, the career maturity of the talent is incorporated into the selection decision. An already established or mature artist may bear less risk than does a newcomer. Finally, criteria with respect to the personal characteristics of the artist are taken into account, such as the artistic skill level, the working attitude, charisma and the visual appearance (Negus, 1992). Based on these criteria selection decisions are made. However, there are some inherent biases and drawbacks in this process.

The decision to select a potential talent is not based on the actual ascertainment of consumer taste but on the expectations of consumers' taste by media gatekeepers. Even though the creative output passes the media gatekeeper, it might therefore fail to be appreciated by the audience (Hirsch, 1972). The full responsibility lies at the A&R manager then (Young and Collins, 2010). In addition, the process of sorting out the vast quantity of input is very time-consuming. Also, past economic failures are exerting pressure on future selections. The likelihood increases that A&R managers are focussing on skimming short-term potential by relying on proven properties and imitating successful competitors instead of developing a long-term career for the artist and creating a unique style (Steinkrauß et al., 2008).Even though it is backed up by experience in the recording industry, decision-making based on intuition is always accompanied by informal heuristics, such as wishful thinking or pure randomness (Seifert and Hadida, 2006). Hence, large sums are invested in artists without the certainty of a return. Therefore scoring a commercial success is very speculative. Increasing the amount of accurate and complete information during this process would strengthen the position of the managerial subsystem. Hence, there is a need for new business models and innovations to overcome this information bottleneck (Steinkrauß et al., 2008).

2.2 Leveraging Crowdsourcing for Decision-Making

Since music is an experiential good and therefore no objective measures on good or bad can be applied, we suggest that crowdsourcing, i.e. using the wisdom of the crowd to evaluate current and planned song releases, might be a way for making better decisions by predicting chart success. The essence of Crowdsourcing is a generic innovation process of distributing the creation of value to a crowd via open online marketplaces and aggregating the results in order to accommodate an information need (Amit and Zott, 2001). Since Crowdsourcing is not restricted to a certain industry, we suggest that it is well suited to solve the pending need for accurate and complete information in the recording industry, namely the extension of the managerial subsystem by an external support subsystem based on the principles of Crowdsourcing (Hirsch, 1972; Mason and Suri, 2011). Filtering and ranking the vast amount of incoming creative material can support A&R managers in their decision-making by reflecting the taste of the crowd and resembling a de facto objective standard for quality evaluation. Consequently, the application of a crowd support subsystem is an alternative (bottom-up) approach to dis-intermediate traditional taste making and gatekeeping mechanisms. Instead of releasing songs and hoping that song releases will break even, A&R managers could rely on the crowd to evaluate the potential of this material prior to conducting large investments in terms of Marketing and Artist Development.

2.3 Hypotheses Development

Based on these ideas, our suggested model contains four building blocks, namely Chart Success, the aggregated Crowd Evaluation, Extrinsic Song Features such as the artist's chart history, and Intrinsic Song Features such as tone or tempo. While the focus lies on the relationship between the Crowd Evaluation and Chart Success, we have incorporated the other blocks as well in order to control for their influence. We are now able to formulate the null hypothesis:

H0: Crowd song evaluations do not have any predictive relevance

As there are no official global popular music charts, the definition of Chart Success is bound to a certain music market. Thus, based on the summary of the top 20 music markets in the 2011 IFPI annual report of the global music industry, we have selected a relevant set of markets with a trade value greater than one billion US dollars (IFPI, 2011). The resulting markets are the United States ($4.1677 bn), Japan ($3.9586 bn), Germany ($1.4122 bn) and the United Kingdom ($1.3785 bn).Therefore, we decided to investigate the conservative German market, where the tipping point of the digital transformation has not yet been reached and the majority of value is still generated by physical sales (Lam and Tan, 2001). Furthermore, the music charts are solely compiled on the basis of physical and digital sales. Hence, the uncertainty of outcome is higher, since traditional gatekeeping mechanisms are mostly intact, meaning consumers have less choice to influence the selection of musical material. Finding a causal relationship between German Chart Success and crowd evaluations would yield valuable information especially for the recording companies in the managerial subsystem. Given the experiential nature of music, we need a framework, which reflects the Hedonic Consumption Paradigm and can be connected to Chart Success.

In addition, the construct for the crowd evaluation has to allow crowd workers to discriminate the quality of the respective songs based on the post-consumption evaluation of the listening experience. Thereby, we would like to focus our attention solely on the evaluation of the musical stimuli. Even though it has been acknowledged to influence Chart Success, we are exempting any visual appearance of the cultural producer from our study (Hargreaves and North, 1997). Since the aggregated crowd evaluation is calculated on the basis of individual consumer evaluations we have screened the consumer research literature for adequate constructs. Predicting success in consumer research is concerned with the performance or non-performance of admired behaviour. As data of observed and actual behaviour is absent in most research studies, it seems reasonable to suggest, that the widely used Intention to Perform certain behaviour should serve as a surrogate for and predictor of actual behaviour (Morwitz and Schmittlein, 1992). Behavioural intention is defined by Ajzen (1991, p. 181) as an "[indication] of how hard people are willing to try, of how much an effort they are planning to exert in order to perform the behavio[u]r". Thus, the stronger an individual's motivation to perform a specific behaviour, the more likely should the individual convert the intention into actual behaviour. There is an on-going debate about behavioural intention constructs due to empirical concerns about the predictive validity (Kalwani and Silk, 1982). Several researchers argue extensively about the discrepancy in the intention-behaviour relationship, meaning that consumers intending to perform certain behaviour essentially do not act afterwards (Sheeran, 2002). We therefore decided to include an additional evaluative post-consumption measure into the model in order to see which of these measures predicts Chart Success best.

Among the few relevant frameworks, we suggest the Hedonic Consumption Model by Lacher and Mizerski (1994) because it is the first exploratory attempt to explain why consumers purchase music and therefore connects a detailed model of the hedonic consumption paradigm to behavioural intention measures. Even though this model has been established prior to the digital transformation of the recording industry, recently, Dilmperi et al. (2011) have successfully applied this model to explain individual differences in the profiles of music pirates and genuine music consumers. Further, the Hedonic Consumption Model focuses on the listening experience solely evoked by the musical stimuli, which is in accordance with our intention. The model contains a global evaluative construct as well. Consequently we adopt the two relevant constructs from the model, namely the Need-to-Re-experience and the Overall Affective Response that were found to influence consumers' music purchase intentions.

According to Zajonc (1980), there are three aspects to consider when it comes to the judgemental character of the Overall Affective Response. First, the value of hedonic experiential goods is holistic. This 'Gestalt approach' refers to the notion that the value is derived from the overall impression about the object and its global features rather than from the individual ones (Agarwal and Malhotra, 2005). Second, affective judgements about an object are reflecting the inner self of the individual, meaning the respective values and personality (Zajonc, 1980). Third, individuals are not able to verbalize the reasons behind their affective judgments (Mittal, 1988). Hence, we define the Overall Affective Response as the global valence judgement of a crowd worker about liking or disliking a piece of music based on the listening experience (Lacher and Mizerski, 1994). Incorporating the above-mentioned selection criteria of the A&R managers, the subjective evaluation of a crowd worker about the hedonic value of a song constitutes a performance benchmark in comparison to other songs and therefore shows the unique selling point of a song. In addition to the assumed relationship in the Hedonic Consumption Model, there have been other researchers confirming the significant relationship to purchase intention (Mizerski et al., 1988). We generalize these findings to behavioural intentions related to music consumption and formulate the following hypothesis:

H1: The Overall Affective Response is positively related to Chart Success

The Need-to-Re-experience is an individual's desire to consume a piece of music again (Lacher and Mizerski, 1994). However, the consumer does not necessarily have to purchase the music in order to control the re-experience nowadays, since there are several choices for consuming music. Therefore, we do not emphasize the temporal control as being the major aspect of this construct, but the multitude of behavioural intentions related to music consumption. Consumers might directly influence Chart Success by re-experiencing the music or indirectly influence it by recommending music to friends. This construct is a self-reflection about the personal fit of the song with the lifestyle and status expressions of the consumer, since only credible and meaningful information is recommended to others (Lacher and Mizerski, 1994). Electronic Word-of-Mouth is the most important promotion strategy for artists, since on the one hand, music is evolving more and more into a service and on the other hand, it is an indicator for growth in artist equity (Howe, 2009; Styvén, 2007). Positive Word-of-Mouth leads to a potential increase in the fan base of the artist, which in turn increases the attention of and exposure to the mass audience as well as increases the likelihood of reaching the critical mass for entering the charts (Hayes, 2008). We assume that converting crowd workers into consumers and advocates of a particular artist is an indicator for the Chart Success potential of a song. Therefore, we define the Need-to-Re-experience as a consumer's desire to re-experience and recommend a song based on the listening experience and formulate the following hypothesis:

H2: The Need-to-Re-experience is positively related to Chart Success

As mentioned above, there are several factors in the production process of a piece of music, which may not be controlled by the cultural producers. Hence, we do not only control for traditional items such as age, gender, country but also for mood, music experience, musical affinity, formal music training, music employment, music preferences, purchase preferences, and consumption preferences of crowd workers. Existing research also suggests criteria for prediction models (e.g., Asai, 2008) based on different facets. Hence, we also control for the awareness a song has due to marketing efforts or 'buzz' (Goel et al., 2010), the label association (major vs. independent) (Steinkrauß et al., 2008), the chart history of the song's artist (DiMaggio, 1977), repertoire (Legrand, 2012), and the positive musical features (intrinsic features) that are concerned with the elementary properties of a song, such as tone, tempo or rhythm (Ni et al., 2011).

3 Methodology

We use structural equation modelling (SEM) since this enables us to transform our set of hypotheses into a path diagram to test and estimate the causal relationship between the crowd workers' song evaluations through a survey on our constructs and controls and Chart Success at the same time. We are more interested in the predictive relevance of these evaluations than in the accurate estimation of the respective model parameters. Therefore, we apply the PLS-SEM approach.

We operationalize our dependent variable Chart Success with the peak position as the single-item manifest indicator for each song. For the period of three months, we are provided with weekly rankings consisting of the top 100,000 sold digital and physical single tracks by the world's leading entertainment data provider. These rankings have to be weighted in order to reflect a composite approximation of the actual chart position. The weighting scheme is defined according to the digital and physical share of the German market's trade value (IFPI, 2011). Nevertheless, we only have to calculate the composite chart position for songs from the sample where physical as well as digital sales ranks are available (see equation 1). Hence, the endogenous variable is measured on an ordinal level and is operationalized as follows:

$$\mathrm{Comp}_{ij} = \begin{cases} x_{ijd}, & x_{ijd} > 0, x_{ijp} = 0, \\ (w_p * x_{ijp}) + (w_d * x_{ijd}), & x_{ijd}, x_{ijp} > 0 \\ x_{ijp}, & x_{ijp} > 0, x_{ijd} = 0 \\ 0, & x_{ijd}, x_{ijp} = 0 \end{cases} \tag{1}$$

with:
$i = 1 \dots N$ (number of song); $j = 1 \dots M$ (number of week); x_{ijd} = Digital sales rank of song i in week j;
x_{ijp} = Physical sales rank of song i in week j; $w_d = 0.13$ (Digital share of Trade value); $w_p = 0.81$ (Physical share of Trade value); Comp_{ij} = Composite approximated chart position of song i in week j

Since the Overall Affective Response is concerned with a global feeling and subjective evaluation of a sample song after consumption, we intended to find a construct best suited for this kind of judgement. Nevertheless, the construct should also reflect all the different facets of the selection criteria of the A&R managers and should cover the hedonic evaluation of the listening experience. Consequently, as in the case of the Hedonic Consumption Model, we have adopted the global index of evaluation (Lacher and Mizerski, 1994). This construct is a summary of bipolar adjective pairs, which are correlating strongly with the "good-bad" pair and thus have an affective overtone. There are nine bipolar adjective pairs, which are reflectively measured on a 7-point Semantic Differential scale, namely *bad-good*, *distasteful-tasty*, *dull-exciting*, *tasteless-tasteful*, *unimaginative-creative*, *untalented-talented*, *unpleasant-pleasant*, *forgettable-memorable* and *boring-interesting* (Osgood et al., 1957). Thus, the level of measurement is interpreted as an interval scale.

The operationalization of the Need-to-Re-experience shall display the crowd workers self-reflection about his support for the artist based on the listening experience. Thus, according to Hayes (2008) post-consumption behaviours, such as the recommendation intention, will ideally reflect a crowd worker's advocacy for the artist. Therefore, we decide to use the existing measurement model from the Hedonic Consumption Model as well, because it also has high reliability (Coefficient Alpha of .90) and covers the re-experience as well as recommendation aspect (Lacher and Mizerski, 1994). This construct measures three reflective items on a 7-point Likert scale (strongly disagree/strongly agree), such as "*I would enjoy listening to this song again*", "*I would like to play this song for my friends*" and "*I want to be able to listen to this song whenever I feel like it*" (Likert, 1932). In the same way, the scale is considered to be of interval level of measurement.

The operationalization of the control variables can be distinguished in two groups. On the one hand, we have dichotomized (binary) variables indicating the presence or absence of a certain attribute. This applies for the Label Affiliation (LAB), which attests, whether a major label releases a sample song or not. We acquired the label information from the iTunes Music store and compared them with the information provided in the sales rankings. Further, the Chart History (CHH) reflects the career status of an artist of being either a newcomer or mainstream artist, based on the previous chart listings and achievements. Hence, we queried several online chart databases for any previous chart listings of the Artists. Besides, we included the origin of the Repertoire (REP). Therefore, we assigned samples to the international repertoire category based on the appearance on the iTunes Music stores of multiple countries. Further, a sample is regarded as domestic repertoire, if it is only listed in the local iTunes Music store. On the other hand, there are variables, which are measured on ordinal scales. For the Awareness (AWA) variable, we have chosen the Google Trends search volume index, which is used to indicate the relative importance of a certain search term in comparison to other search terms in the same category and period of time. Finally, we measure the potential of the intrinsic features (FAN) of the respective sample song with the hit score computed by Ni et al. (2011).

We collect data on 150 songs and let each of them be evaluated by 20 crowd workers on Amazon Mechanical Turk resulting in a total of 2.852 observations. We use the averaged values of the crowd evaluations for each song. Since there is strong cultural proximity of the German music market to the Anglo-American markets in terms of popular music culture, we focus on crowd workers originating in Anglo-American countries with similar popular music culture (i.e. US, UK, Canada). We use 90-second samples of the songs. Furthermore, we are facing the same uncertainty as the A&R managers in determining which new releases will make it into the charts. Thus, in order to determine sample eligibility, we have to follow a lenient strategy. First, we restrict the language parameter to “English” due to its characteristic as global language and wide application in international repertoire. Second, we include only songs from most popular genre dimensions (Rentfrow and Gosling, 2003). As we have mentioned in the introduction, previous research focused their attention mainly on historic data in order to build their prediction models (e.g., Asai, 2008; Gazley et al., 2011; Ni et al., 2011). However, our intention is to predict Chart Success based on evaluations of currently released songs. Even more vital is to assure that the crowd workers have not yet been exposed or stumbled upon the respective song since this familiarity bias may flaw the true evaluation of the song (Zajonc, 1980). Thus, we have to test the samples on the day of the release and actively control for familiarity with the particular song and/or artist.

We follow recommended practices of previous research studies in order to establish appropriate quality mechanisms prior, during and after the task execution (see Table 1). In order to indicate that the crowd workers actually completed the survey, in the end, they are presented a unique completion code or a similar generated token which they have to paste into the respective textbox (Mason and Suri, 2011). Apart from the qualification requirements, crowd workers are randomly selected. Novel crowd workers have to pass the additional qualification test, since we have to ensure that they are physically able to hear. Therefore we asked them to listen to several animal sounds and to tick the corresponding answer (Sprouse, 2010). In order to further protect our survey from satisficing, we also implemented an attention check in the form of a trick question, which is randomly positioned among the scale items. Further, after evaluating the sample song, we also checked the answers according to their patterns (Mason and Suri, 2011).

Phase	**Ex ante HIT**	**During HIT**	**Ex post HIT**
Mechanisms	# of HITs approved >= 50	Geo-location	Pattern Flag
	Approval rate >= 95%	Qualification Test	Familiarity Flag
	Adult Content qualification = 1	Age verification	Completion Time
	Location is NOT India	Previous Participation	Completion Code
		Listening time	
		Attention check	

Table 1. Applied Quality Mechanisms

4 Results

After HIT publication we recorded 3,890 attempts to work on our task. The incorporation of the chosen quality management mechanisms proved valuable, because it separated the useful submissions from the useless ones. Consequently, the final 2,852 approved submissions resemble 73.32% of the initial attempts. By adding up the weekly completion times, the entire song evaluation task was accomplished within 3.32 days. 594 unique crowd workers generated the approved evaluations, whereby, on average, one crowd worker rated 4.80 samples. As a consequence of the quality check, the average ratings per song reduced from 20 to 19.01. The average completion time for a HIT corresponds to 7.22 minutes and is about half of the initially allotted time. Based on the average time, the effective hourly wage increased to $3.33, which is a fair reward and complies with related work on AMT (Sorokin and Forsyth, 2008; Ipeirotis, 2010b; Grady and Lease, 2010). In terms of the composition of the crowd, 74% of the participating crowd workers are in the age groups between 18 and 34, which are highly relevant for the recording industry (Steinkrauß et al., 2008). Further, 60% of the crowd workers are female and 96% of them earn less than or equal to $60,000 per year. With the focus laid on crowd workers from Western countries with similar popular music culture, we already expected that a large share (97%) of them originates from the US. The crowd is also very active on AMT. 40% of crowd workers spend between 5 and 15 hours per week on AMT and another 22% works between 15 and 30 hours per week on AMT. Also, we have found that the heavy contributors, meaning crowd workers submitting more than 10 song evaluations, amounted to 15% whereas 50% only worked on a single HIT.

The absence of crowd workers with profound musical background let us assume that the appraisal of music seems to be a universal phenomenon (Kazai, 2010). Further, the crowd workers mainly consume music by digitally accessing (37%) and digitally owning (51%) the music. On the aggregated song-level, we have observed that the values for the dependent variable are primarily determined by the digital peak position, because the majority of songs (78%) did not enter physical sales ranks. We have also found that songs can be classified into three chart ranges, namely the Top 500 (43%), the Long Tail (47%) and the Non-Entries (10%) (Anderson, 2006). In general, major label material is dominating the dataset (59%), especially the Top 500 (86%). This is in accordance with the division of power in the global recording industry, where the market shares of the big four labels outperform the remaining labels (Billboard, 2010). Also, mainstream artists (73%) outweigh the newcomer artists (27%) in the dataset, whereby the mainstream artists are leading the Top 500 (83%) and the Long Tail (64%). Approximately nine out of ten samples (86%) are associated as being of international repertoire. This ratio is also consistent throughout all chart ranges. The distribution of Awareness is skewed towards lower values, whereby the average Awareness of the Top 500 (15.56) is higher than the Long Tail (4.81), which in turn is higher than the average Awareness for the Non-Entries (0.00). In terms of the Feature Analysis score, our dataset shows roughly a normal distribution around the mean score of 5.81. Hence, the audio features of the majority of songs are very similar but do not necessarily comprise hit potential. The average score for the Non-Entries (6.69) was greater than for the Top 500 (6.03), which was greater than the score from the Long Tail (5.60).

For the evaluation of our model, we excluded the Non-Entries from the estimation, because they did neither enter the digital nor the physical sales ranks and therefore do not provide information for the dependent variable. With the remaining sample songs (n = 134), we focused on evaluating the reliability and validity of the measurement model of our two crowd constructs, because the single-item control variables are not part of the core model (Ringle et al., 2012). A preliminary factor analysis resulted in two factors with Eigenvalues greater than 1, whereas the first factor accounts for 59% of the variance and resembles the Overall Affective Response. The second factor explains another 21% of the variance and is associated with the Need-to-Re-experience. All indicator loadings were high on their intended factor and exceeded the suggested threshold (0.700). Further, the values of Cronbach's Alpha (OAR: 0.962; NTR: 0.956) and the Composite Reliability (OAR: 0.967; NTR: 0.972) surpass the required threshold values for acceptable results (0.800) (Cronbach, 1951). The values for the Average Extracted Variance of the Overall Affective Response is considerably lower than the one of the Need-to-Re-experience. Nevertheless, it is still above the critical value (0.500) (Vinzi et al., 2010). Internal consistency is also reflected with indicator reliabilities ranging from 0.527 to 0.971 and exceeding the suggested decision criterion (0.400) (Bagozzi and Yi, 1988). In addition, all indicators have very high significance. Consequently, we assume that both latent variables have very high levels of internal consistency and convergent validity. Moreover, we tested the Fornell-Larcker Criterion against our data in order to ascertain discriminant validity. Thereby, the squared correlation between Overall Affective Response and the Need-to-Re-experience (0.045) is considerably lower than the AVE of both variables (Fornell and Larcker, 1981). Thus, we also assume discriminant validity of our measurement model.

In contrast to the exemption from the analysis of the measurement model, we have incorporated all paths and variables in the analysis of the structural model (cf. Figure 1). The execution of the PLS algorithm with the parameters set to path weighting scheme and standardized values showed that our proposed model has an explanatory power (R^2) of 43.5% of the variance in Chart Success. Corrected for possible biases stemming from the number of predictors, the adjusted explanatory power ($R^2_{adj.}$) reduces to 40.4%. Further, the blindfolding procedure with an omission distance of 7 resulted in a predictive relevance (Q^2) of 44.8% (Fornell and Bookstein, 1982). Compared to an estimation of a model consisting only of the control variables, the inclusion of the crowd evaluation constructs increases the explanatory power by approximately 10%. It seems that the core model is determined by label affiliation, prior chart history and our crowd evaluation constructs.

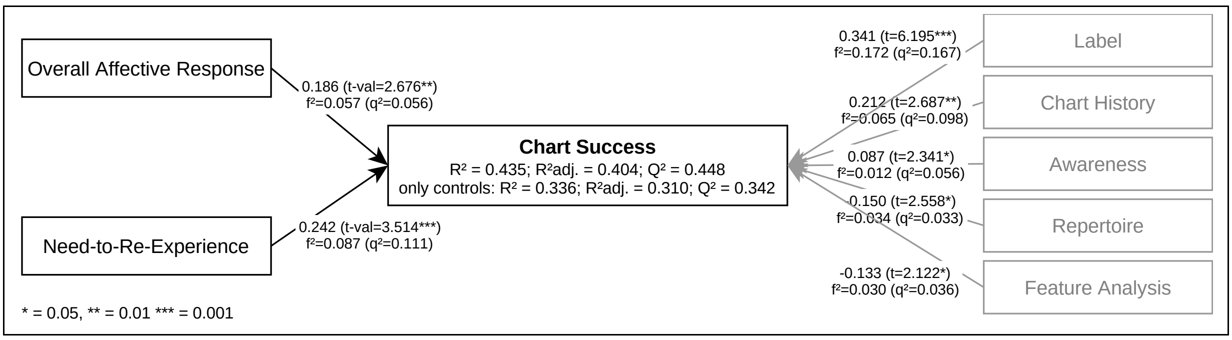

Figure 1. Model Estimation

5 Discussion

In the beginning, we posed the research question, whether the crowd is able to forecast popular music Chart Success. Based on our results, we have found significant and positive relationships between the crowd evaluation constructs and popular music chart success. The results from the estimation have shown, that the path between the Overall Affective Response and Chart Success has predictive relevance ($q_{OAR}^2 = 0.056$). In addition, the path between the Need-to-Re-experience and Chart Success has predictive relevance as well ($q_{NTR}^2 = 0.111$). By summing up both values, the predictive relevance of the crowd is greater than zero ($Q_{crowd}^2 = 0.056 + 0.111$) with the Need-to-Re-experience having twice as much predictive power than the Overall Affective Response. Hence, we have acquired empirical evidence for the rejection of the null hypothesis ($0.166 > 0$) and are able to answer our research question: Yes, the crowd can forecast Chart Success. However, even though we almost replicated the reliability of the measurement models of the latent constructs from the hedonic consumption model, this statement has to be qualified with respect to the overall impact of our model. It seems that there are several aspects influencing the magnitude of this forecasting ability. First, the degree of industry manipulation is confounding the impact of the crowd, meaning that even though the songs are rated on an equal and unbiased basis, traditional gatekeeping mechanisms and mass promotion dynamics are still intact (Hirsch, 1972). There is an imbalance in the allocation of resources and exposure, which is not in the crowd's locus of control but influences its predictive relevance. With increasing chart position the impact of the crowd evaluation becomes less important and industry manipulation is dominant in these chart ranges. Hence, we have had a closer look at the degree of industry manipulation per chart range, whereby we distinguish between songs that have received any manipulation, songs with prior Chart History, songs with Label Affiliation and songs with prior Chart History and Label Affiliation.

We found that the crowd shows sufficient discriminatory power in evaluating songs without any industry manipulation. In this case, we observed that the better the evaluation of the crowd the higher is the chart position. In fact, there were two hidden gems in the dataset, which were neither affiliated to a major label nor previously exposed to the masses but highly appreciated by the crowd and ranking high in the Top 500 chart range (25th and 368th position). Targeting investments and marketing budgets towards these songs might have been far more profitable than diversifying the investment on a myriad of failures while blowing up the marketing budgets in order to distinguish the songs from other releases. In our study, the share of songs from unsigned newcomers only amounted to 16%, which may explain the small predictive relevance. This result has been confirmed by the feature analysis algorithm, which has also assigned a high score to the hidden gems. In fact, the hit score of the two hidden gems outperformed the other songs from the Top 500. Thus, we also assume that the hit score works best for material without any industry manipulation. Second, a gap between the chart mechanisms in the respective market and the consumption patterns of the crowd leads to small predictive relevance.

The German market may still be regarded as conservative, because the compilation of the charts relies solely on physical and digital sales (Lam and Tan, 2001). However, due to the on-going digital transformation, individuals nowadays have an increased choice of consuming music without the necessity of purchasing physical records or digital downloads (Fox and Wrenn, 2001). In fact, there are several other revenue streams, which may have the potential to supplant these sales in the future, such as revenues generated from subscription services or live performances. In our study, we found that the crowd already consists in large part of digital natives, since digital consumption patterns (88%) are dominating their physical counterparts (12%). In addition, the focus of consumption also switches from owning (62%) to accessing music (38%). Hence, even though a song is positively evaluated and there is a strong Need-to-Re-experience, the predictive impact of the crowd is rather low, because the current mechanisms of the German music charts do not take into account the consumption preferences from a large portion of the crowd. Nonetheless, as being a market information regime for the recording industry, the charts should reflect all possible consumption choices in order to determine the true overall popularity of a piece of music (Anand and Peterson, 2000). The large share of digital-only releases in our dataset supports this view. Third, apart from the underlying mechanism, the lack of data availability hinders the profound application of accurate indicators for Chart performance. The predictive relevance may be small due to the application of our composite score for measuring Chart Success. It still only relies on the best possible approximation of the true chart score, because the exact process of how to calculate the charts is not publicly available and the true chart positions are only compiled for a limited range (usually between 20 and 200). Thus, even though the true score would have fit for the first 100 chart positions, we are not able ascertain the fit for the remaining chart positions. In addition, we are facing the same uncertainty as the music managers in choosing relevant songs for the study. Hence, by taking only the Top 100 positions, only 20% of our selected songs would have made it into the Top 100, wherefore we encounter problems concerning the minimum samples size and the estimation of the model.

With regards to the discussed results, we suggest adding an external support subsystem in the form of an open marketplace to the recording industry system. The integration of this subsystem is re-intermediating the recording industry bottom-up relying on the principles of Crowdsourcing (Steinkrauß et al., 2008). Based on this notion, we have identified three possible areas of application. First, the crowd may fulfil an on demand music discovery function by filtering out the hidden gems from the vast amount of creative input. Thereby, the crowd either supports the talent selection process of the A&R departments of recording companies from the managerial subsystem or resembles a quality control mechanism for the musical material destined for the direct distribution over the Internet. Resources of the cultural producers can be allocated to more value-adding and attention-building processes, such as Marketing and Promotion. Second, the cost-efficient and direct interaction with the crowd can be used to include large-scale feedback about the quality and marketability of musical material into an iterative bundling and development process. Therefore, enhanced market intelligence increases the chances that the musical material is tailored to the needs of the consumers and meets current stylistic expectations, which in turn increases the likelihood of passing the media gatekeepers. In addition, the crowd may also support the development process by deciding which musical material shall be nurtured and receive funding from the managerial subsystem. Third, the crowd may constitute a semantic recommendation system by leveraging on the weak ties between musically interested influencers and mainstream followers. This crowd-induced pull distribution strategy may weaken the power of traditional gatekeepers by offering consumers an extended variety of musical material without increasing their search effort. Apart from the implications, there are also some limitations to our study.

First, the necessity to restrict the analysis to a certain music market requires the adaptation of the research design to the underlying peculiarities and mechanisms in this respective market. Second, the crowd evaluation should become more accurate with increasing sample size, which is only accomplished through a substantial investment of cost and time. Third, even though the sample is more diverse than traditional convenience samples, the focus on Internet users and digital natives still limits the general representativeness. Fourth, since we have faced the same uncertainty about the future chart performance of the sample songs as A&R managers in the managerial subsystem, we had to follow a lenient sampling strategy for the song selection, which resulted in an imbalanced song distribution of the dataset. Fifth, for reasons of data availability, the point in time for the data collection was chosen to be the day of the release instead of a point in time prior to the release.

6 Conclusion

We have conducted this study in order to find out, whether the disruptive potential of Crowdsourcing can be used for the success prediction of popular music in the German recording industry. Therefore, we aimed at finding empirical evidence for the causal relationship between aggregated crowd song evaluations and the peak chart position. We have used two constructs from the hedonic consumption model to reflect the crowd evaluation. Together with control variables from the literature related to the recording industry, such as Label Affiliation or prior Chart History, we have applied PLS to build, test and estimate our model, whereas we have used Amazon Mechanical Turk to build the song evaluation tasks. Our results show that we are able to explain large parts of chart success through our model (R^2=0.435). Even though we find evidence for reliable and valid measurement model, the structural model has average explanatory power and moderate predictive relevance. We proposed two explanations for these results. First, there is a mismatch between the consumption patterns of the crowd and the underlying chart mechanisms of the music market. Second, the effects of the industry manipulation outweigh the influence of the crowd on chart success. Consequently, we analysed the predictive relevance of novel songs from unknown artists and found that the crowd is well suited to discriminate the quality and predict the performance of this kind of musical material.

Overall, we have found that the inclusion of our crowd evaluation constructs increased the explained variance by approximately 10%. Thus, opening up the value chain of the recording industry may result in more complete and accurate market information about the potential of musical material. We recommend integrating a support subsystem in the value chain of the recording industry in order to reduce the uncertainty by filtering the vast amount of creative input. Since this has been an initial attempt to investigate the applicability of Crowdsourcing in the recording industry, there are several directions for future research. First, the validity of our findings has to be replicated with further datasets. Second, the applicability of the findings could be tested in other cultural industries as well. Third, the findings from the study may be compared with other music markets in order to find general patterns of crowdsourced success prediction or to prove our assumption that the predictive relevance of the crowd will be larger in already digitally transformed music markets.

References

Agarwal, J., Malhotra, N. K. (2005). An Integrated Model of Attitude and Affect. Journal of Business Research, 58 (4), 483–493.

Ajzen, I. (1991). The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes, 50 (2), 179–211.

Amit, R., Zott, C. (2001). Value Creation in E-Business. Strategic Management Journal, 22 (6-7), 493–520.

Anand, N., Peterson, R. A. (2000). When Market Information Constitutes Fields: Sensemaking of Markets in the Commercial Music Industry. Organization Science, 11 (3), 270–284.

Anderson, C. (2006). The Long Tail: Why the Future of Business is Selling Less of More. Hyperion, New York, NY, USA.

Asai, S. (2008). Factors Affecting Hits in Japanese Popular Music. Journal of Media Economics, 21 (2), 97–113.

Bagozzi, R. P., Yi, Y. (1988). On the evaluation of Structural Equation Models. Journal of the Academy of Marketing Science, 16 (1), 74–94.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., Telang, R. (2007). The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts. Management Science, 53 (9), 1359–1374.

Bonabeau, E. (2009). Decisions 2.0: the Power of Collective Intelligence. MIT Sloan Management Review, 50 (2), 45–52.

Celma, Ò., Serra, X. (2008). FOAFing the Music: Bridging the Semantic Gap in Music Recommendation. Web Semantics: Science, Services and Agents on the World Wide Web, 6 (4), 250–256.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. Psychometrika, 16 (3), 297–334.

Dilmperi, A., King, T., Dennis, C. (2011). Pirates of the Web: The Curse of Illegal Downloading. Journal of Retailing and Consumer Services, 18 (2), 132–140.

DiMaggio, P. (1977). Market Structure, the Creative Process, and Popular Culture: Toward an Organizational Reinterpretation of Mass-Culture Theory. The Journal of Popular Culture, 11 (2), 436–452.

Fornell, C., Bookstein, F. L. (1982). Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. Journal of Marketing Research, 19 (4), 440–452.

Fornell, C., Larcker, D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research, 18 (1), 39–50.

Fox, M., Wrenn, B. (2001). A Broadcasting Model for the Music Industry. International Journal on Media Management, 3 (2), 112–119.

Gazley, A., Clark, G., Sinha, A. (2011). Understanding Preferences for Motion Pictures. Journal of Business Research, 64 (8), 854–861.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., Watts, D. J. (2010). Predicting Consumer Behavior with Web Search. Proceedings of the National Academy of Sciences, 107 (41), 17486–17490.

Hargreaves, D. J., North, A. C. (1997). The Social Psychology of Music. New York, NY, USA.

Hayes, B. E. (2008). Measuring Customer Satisfaction and Loyalty : Survey Design, Use, and Statistical Analysis Methods. ASQ Quality Press, Milwaukee, WI, USA.

Hirsch, P. M. (1972). Processing Fads and Fashions: An Organization-Set Analysis of Cultural Industry Systems. American Journal of Sociology, 77 (4), 639–659.

Holbrook, M. B. (2006). Audience Judgments as the Potential Missing Link Between Expert Judgments and Audience Appeal: An Illustration Based on Musical Recordings of “My Funny Valentine”. Journal of the Academy of Marketing Science, 34 (1), 8–18.

Holbrook, M. B., Hirschman, E. C. (1982). The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun. Journal of Consumer Research, 9 (2), 132–140.

Howe, J. (2009). Crowdsourcing : Why the Power of the Crowd is driving the Future of Business. Crown Business, New York, NY, USA.

IFPI (2011). The Recording Industry in Numbers. IFPI.

Kalwani, M. U., Silk, A. J. (1982). On the Reliability and Predictive Validity of Purchase Intention Measures. Marketing Science, 1 (3), 243–286.

Kazai, G. (2010). An Exploration of the Influence that Task Parameters have on the Performance of Crowds. In 1–9, San Francisco, CA, USA.

Lacher, K. T., Mizerski, R. (1994). An Exploratory Study of the Responses and Relationships Involved in the Evaluation of, and in the Intention to Purchase New Rock Music. Journal of Consumer Research, 21 (2), 366–380.

Lam, C. K. M., Tan, B. C. Y. (2001). The Internet Is Changing the Music Industry. Communications of the ACM, 44 (8), 62–68.

Legrand, E. (2012). Monitoring the cross-border circulation of European music repertoire within the European Union. European Music Office.

Likert, R. (1932). A Technique for the Measurement of Attitudes. Archives of Psychology, 22 (140), 1–55.

Lorenzen, M., Frederiksen, L. (2005). The Management of Projects and Product Experimentation: Examples from the Music Industry. European Management Review, 2 (3), 198–211.

Mason, W., Suri, S. (2011). Conducting Behavioral Research on Amazon's Mechanical Turk. Behavior Research Methods, 44 (1), 1–23.

Mittal, B. (1988). The Role of Affective Choice Mode in the Consumer Purchase of Expressive Products. Journal of Economic Psychology, 9 (4), 499–524.

Mizerski, R., Pucely, M. J., Perrewe, P., Baldwin, L. (1988). An Experimental Evaluation of Music Involvement Measures and their Relationship with Consumer Purchasing Behavior. Popular Music and Society, 12 (3), 79–96.

Morwitz, V. G., Schmittlein, D. (1992). Using Segmentation to Improve Sales Forecasts Based on Purchase Intent: Which "Intenders" Actually Buy? Journal of Marketing Research, 29 (4), pp. 391–405.

Neelamegham, R., Jain, D. (1999). Consumer Choice Process for Experience Goods: An Econometric Model and Analysis. Journal of Marketing Research, 36 (3), 373–386.

Negus, K. (1992). Producing Pop : Culture and Conflict in the Popular Music Industry. E. Arnold ; Distributed in the USA by Routledge, Chapman, and Hall, New York, NY, USA.

Ni, Y., Santos-Rodríguez, R., McVicar, M., De Bie, T. (2011). Hit song science once again a science? In 1–2, Melia Sierra Nevada & Melia Sol y Nieve, Sierra Nevada, ES.

Ordanini, A. (2006). Selection Models in the Music Industry: How a Prior Independent Experience May Affect Chart Success. Journal of Cultural Economics, 30 (3), 183–200.

Prince, W. F. (1972). A Paradigm for Research on Music Listening. Journal of Research in Music Education, 20 (4), 445–455.

Rayna, T., Striukova, L. (2009). Monometapoly or the Economics of the Music Industry. Prometheus, 27 (3), 211–222.

Rentfrow, P. J., Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. Journal of Personality and Social Psychology, 84 (6), 1236–1256.

Ringle, C. M., Sarstedt, M., Straub, D. W. (2012). A Critical Look at the Use of PLS-SEM in MIS Quarterly. MIS Quarterly, 36 (1), iii–xiv.

Seifert, M., Hadida, A. L. (2006). Facilitating talent selection decisions in the music industry. Management Decision, 44 (6), 790–808.

Seiter, E. (1986). Stereotypes and the Media: A Re-evaluation. Journal of Communication, 36 (2), 14–26.

Sheeran, P. (2002). Intention—Behavior Relations: A Conceptual and Empirical Review. European Review of Social Psychology, 12 (1), 1–36.

Sprouse, J. (2010). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. Behavior Research Methods, 43 (1), 155–167.

Steinkrauß, N., Gmelin, H., Günnel, S. (2008). Wettbewerbsanalyse. In Clement, M., Schusser, O., Papies, D. (Eds), Ökonomie der Musikindustrie, 27–44, Gabler Verlag, Wiesbaden, GER.

Styvén, M. (2007). The Intangibility of Music in the Internet Age. Popular Music and Society, 30 (1), 53–74.

Vinzi, V. E., Chin, W. W., Henseler, J., Wang, H. (2010). Handbook of partial least squares concepts, methods and applications. Springer, Berlin; New York.

Young, S., Collins, S. (2010). A View from the Trenches of Music 2.0. Popular Music and Society, 33 (3), 339–355.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35 (2), 151–175.

Zwaan, K., Ter Bogt, T. F. M. (2009). Research Note: Breaking into the Popular Record Industry: An Insider's View on the Career Entry of Pop Musicians. European Journal of Communication, 24 (1), 89–101.