Using online networks to analyse the value of electronic music

Anna Jordanous
Centre for e-Research
Department of Digital Humanities
King's College London
London WC2B 5RL, UK
anna.jordanous@kcl.ac.uk

Daniel Allington
Centre for Language and Communication
Faculty of Education and Language Studies
Open University
Milton Keynes, MK7 6BJ, UK
daniel.allington@open.ac.uk

Byron Dueck
Department of Music
Faculty of Arts
Open University
Milton Keynes, MK7 6BJ, UK
byron.dueck@open.ac.uk

Abstract

In evaluating how creative a program or an artefact is, a key factor to consider is the value inherent in that program or artefact. Our research investigates the process by which cultural products may be accorded a form of specifically cultural value independent of market value, focusing in particular on how that process has been transformed through mediation by online networks. To do this, we are studying a specific artform, i.e. music, and evidence from a specific website, i.e. SoundCloud, making a case study of a specific genre with a special association with that website: electronic music. Quantitative analysis ranges across all genres of music represented on the website, with social network graphs being constructed from relational data and corpus analysis being carried out on textual data. Interviews and observational research are being carried out with electronic music performers in order both to explore what interaction on the site means to them on a qualitative level and to study how the production and circulation of value on the site relates to the production and circulation of value in offline environments. This project will make available a methodology and supporting software for measuring creative value through relevant network analysis.

Introduction

Value is a key factor to consider when evaluating how creative an artefact or creative program is (Ritchie 2007; Jordanous 2012b). While evaluation of computational creativity systems usually includes some evaluation of value, this evaluation is often based on subjective data, which may need to be collected during a lengthy, time consuming testing process with a sample of users that may or may not be representative of the full target audience.

It is difficult to integrate the results of this kind of testing within an automated evaluation process, and feed the results back into the design and development of the creative system, particularly if (as often happens) system testing is carried out towards the end of research projects, rather than at earlier stages of system development. This makes it more difficult to efficiently implement autonomous self-evaluation by a system, despite the ongoing strand within computational creativity research to incorporate ongoing evaluation as part of the creative computational process (through employing *engagement-reflection*, *generate-and-explore* or evolutionary computational approaches to computational creativity, where the system cyclically engaging with creative production then critically reflecting on what it has produced, to inform the next stage of creative production (McGraw and Hofstadter 1993; Pérez y Pérez, Aguilar, and Negrete 2010; Saunders 2012), for example). Clearly there is a place for research exploring how to make value judgements in a more comprehensive and automated manner. As Boden (2004) says, creativity is not just about new ideas but also incorporates the development and refinement of these ideas.

Cultural value is one of those areas in which (as the saying goes) perceptions are also realities. Thus, sociologists have argued that the production of cultural value is actually the production of a form of belief. Although popular accounts of how art gets made tend to focus on brilliant individual creators, research has highlighted over and over again that their work typically emerges from a creative milieu, in which value (or belief in value) comes into existence. This highlights the complex relationship between professional, semi-professional, and amateur cultural production, and may explain why so many cultural producers create work primarily for appreciation by their peers.

The project reported in this paper focuses on a specific creative domain: music. This work is currently in-progress, with an exploratory focus. We are gathering evidence from the SoundCloud website, which many musicians use for commenting on one another's work. We focus on a specific genre that has a special relationship with that website, i.e. electronic music. We are combining social network analysis of evaluations implied by 'likes', 'follows', and 'comments' on the website with linguistic analysis of the kinds of language used in comments. We are also observing and interviewing musicians at gigs, to understand how they locate value in their relationships with one another, both online and off. Findings will be disseminated through academic and non-academic venues, including public engagement events.

Value measurement and evaluation in Computational Creativity

The term *value* encompasses many different aspects such as correctness, worthiness, quality and usefulness. Creativity is often treated as $novelty + value$ (Jordanous 2012a).

Although his empirical criteria for creativity evaluation depend heavily on having available some ratings of the value of a system's output, Ritchie (2007) is 'deliberately general and unconstrained about how value ratings might be arrived at'. Such an approach is understandable; it is difficult to find domain-independent heuristics to follow when ascertaining the value of products. Usefulness is relative; what is considered useful in products of one domain is not necessarily reproduced in the other and may not apply equally across that individual domain. To recognise the usefulness of a creative product, one must either know the product's domain well enough to appreciate value, or have access to the opinions of people who are experts in that domain.

In a review of evaluation of computational creativity systems, Jordanous (2011) found many examples of empirical measurements of value, as described below. She also found that for several systems, value was assessed through user evaluations. Evaluation data was either directly provided by the user, obtained through studies of aspects such as audience reactions and feedback at exhibitions, or obtained through qualitative tests with target users for usability and effectiveness of the system, or user evaluations and discussions. Many systems were evaluated by the correctness and validity of their products, such as calculating the percentage of material produced during runtime that can actually be used, or statistical tests for validity. Some systems were measured in terms of how interesting their products were, for example seeing if the products performed at a level above a given threshold for originality in the Wundt curve function (Saunders 2012) or using variables representing domain-specific interest or complexity measurements.

Aims of this work

This project directly investigates how cultural value is attributed, developing an evidence-based methodology for identifying and evaluating cultural value, that could in the future be incorporated into an autonomous (or more independent) creative system. The research focuses on interactions between creative producers and consumers, aggregating their evaluations and tracing the flow of value between them. Our focus on data from an online community shows how the internet mediates creative interaction and the production of cultural value. Online sites like SoundCloud open up ways of interacting with and evaluating cultural artefacts (and their producers); we study how cultural value is generated and perceived in these influential online communities.

Musical Cultural value and social networks

This current project echoes Csikszentmihalyi's sentiments of interactions between domain, individual and field (Csikszentmihalyi 1988) and is situated within the broader area of *field theory* (Bourdieu 1993). According to Bourdieu, the value of cultural goods is a form of belief produced through 'a vast operation of social alchemy... jointly conducted, with equal conviction and very unequal profits, by all the agents involved in the field of production, i.e. obscure artists and writers as well as "consecrated" masters; critics and publishers as well as authors; enthusiastic clients as well as convinced vendors.' (Bourdieu 1993, 81; emphasis in original) In field theory, this belief is often referred to as 'symbolic capital', and comes into being through cultural producers' esteem for one another's work.

Sites such as SoundCloud enable cultural producers to evaluate one another's work in public, both categorically, e.g. by clicking a 'like' button, and qualitatively, by leaving comments. These public evaluations provide direct evidence of how cultural value is co-produced. We will also incorporate 'offline' research into our investigations, interviewing musicians and contrasting our digital findings with other conceptions of musical value.

Methodology

The project combines digital research on the SoundCloud website with ethnographic research on electronic music producers in London who use the SoundCloud website. SoundCloud provides a good data source for technical reasons (there is a well-developed API available which provides access to all the necessary data) and for social reasons (it is widely used by amateur, semi-professional, and professional musicians for networking and for publishing music.

- Stage 1: Collect public data automatically from SoundCloud, using the SoundCloud API. Rather than study the entire network of users (which comprises over 20 million accounts, many of them inactive or controlled by bots), a snowball sampling method is used.
- Stage 2: Build networks of accounts and tracks, based on the 'follow', 'like', 'comment', 'share', and 'group' relationships accessible through the API.
- Stage 3: Identify and divide the corpus of English-language comments on tracks into subcorpora according to track genres and to commenters' locations, both in the network and in the real world. Use corpus analysis to identify evaluative vocabularies associated with particular genres, groups, and locations.
- Stage 4: Construct multiple networks of electronic music producers for closer analysis. Identify clusters within the networks for closer attention, with hand-coding and qualitative discourse analysis of commenting behaviour.

This work will be complemented by ethnographic research. Centred on electronic music performers identified as users of SoundCloud based in London, this research involves both observation and interviewing, focusing (a) on how value is produced in live performance of electronic music (for example, through allocation of better and worse slots, introductions for performers, audience behaviour), and (b) on how performers understand the role of SoundCloud and other digital sites in relation to physical sites such as live music venues in the production and circulation of value.

Results to date

Quantitative analysis

We have been collecting data from SoundClouds API (Application Programming Interface) which is the gateway to access SoundClouds data. We have also been updating code written by Daniel Allington for analysing network actions in Interactive Fiction communities,[1] to adapt it for analysing what happens between users on SoundCloud, using SoundCloud's SDK for Python. Upon exploring the data, we have been able to collect various data from SoundCloud including public data on their users, users tracks, groups they have joined, comments they have made on tracks, tracks they have favourited and tracks they have added to personal playlists, as well as who users follow and who follows them. We have collected two data samples of 500 users and 1500 users, plus their related data (tracks, groups, follows, etc), starting from a randomly chosen user and using snowball sampling to explore each users connections to other users. Currently a longer term larger-scale data collection process is underway.

Concentrating first on the information on how users follow each other, we have been able to use our updated version of the IF analysis code to analyse SoundCloud data, though these methods are not scaling up well to larger data samples. In the smaller data sample of 500 users, we were able to generate graph visualisations of follow relationships, producing the graphs in Figure 1, though even these diagrams are difficult to read. Learning from previous experience, it is useful to reduce our sample of users to include only those users who are a. followed by at least one other user and b. following at least one user. This leaves us with users who are both recognised by other users and who are themselves participating in the network. For example, in our sample of 1500, an account under the name of 'Justin Timberlake' has over 400000 followers, but as this user is not following any other users, they are not interacting with other users and do not therefore dynamically affect how value is attributed *within the network*. Similarly, several inactive accounts and 'bot' accounts can be filtered out this way.

Graphs for samples larger than 500 users would be unreadable. Instead, for our larger sample of 1500 users, we applied ranking methods to identify the top users in terms of influence, again focusing solely on the networks built by follow links between users, for ease in experimentation. Measuring recommendation and influence through indegree rankings (a measure based around how many users follow another user), we identify key players in our samples (Table 1). This ranking does find some key players in electronic music whose data have been captured in our sample, such as Tiésto. While indegree does not directly match the ordering of accounts with the most followers, there is some similarity. The afore-mentioned 'Justin Timberlake' account, for example, comes in at position 20, despite not interacting at all on SoundCloud. Hence we are exploring more sophisticated methods such as PageRank and eigenvector rankings to help identify key players in SoundCloud's networks.

The next steps are to construct networks with larger data samples, using more of the information within the SoundCloud data such as favourited tracks and comments, and to make sure our analysis is able to deal with these much larger collections of data. Our work-in-progress code is available at https://github.com/ValuingElectronicMusic/network-analysis and we will post regular updates on our progress on http://www.open.ac.uk/vem.

[1] https://github.com/ValuingElectronicMusic/ifdb-analysis

Table 1: Top 10 users (by indegree) in the 1500 user sample.

Position	Username	Indegree score
1	diplo	357
2	HARDWELL	320
3	steveaoki	277
4	Tiésto	277
5	A-Trak	263
6	Porter Robinson	259
7	Flosstradamus	246
8	DILLONFRANCIS	242
9	Martin Garrix	233
10	Zedd	230

A final area to report within the data analysis relates to the third stage of the methodology as described above, where English-language comments are to be divided into subcorpora according to track genres and to commenters' locations, both in the network and in the real world. Currently our code determines English language comments to reasonable accuracy by using the Open Office dictionaries for English, French, Spanish and Italian, though it still does not pick up comments such as 'wooooot!!!' or 'loveeeeeeeeee', the type of which occur surprisingly frequently in our sample.

Creating sub corpora based on genre and location has proven to be more troublesome than expected, partly because many users do not provide location information, and partly because of difficulties in identifying appropriate genres, as we will discuss during the next section.

Quantitative analysis and Qualitative research

We have found that questions arising during quantitative analysis can often be addressed during qualitative analysis, and that questions arise during qualitative analysis which we can use quantitative analysis to investigate. To date, four loosely-structured interviews have been conducted with London-based producers of electronic music. These interviews have generated thought-provoking discussion, suggesting many avenues for the research ahead: on the one hand for the ethnographic components of our work and on the other for the kinds of data we might examine from SoundCloud. For example, when we asked questions about valuing and appreciation, people often answered about relationships. This helps justify our approach of investigating relationships in quantitative analysis to analyse value.

One of the more interesting areas where we need to combine both qualitative and quantitative findings is in determining what musical genres qualify for our case study on 'electronic music'. A recurring question has been: what is electronic music? In quantitative analysis, the question has been how how to identify electronic music on SoundCloud.

> 'I find myself wondering where the 'edges' (no doubt feathered) are to electronic music as a category. Does it help to discuss these? Or do we just 'know it when we hear it?' Byron Dueck (during project discussions)

Genres are often not marked as 'electronic' but as a specific type of electronic music. In fact, one initial outcome

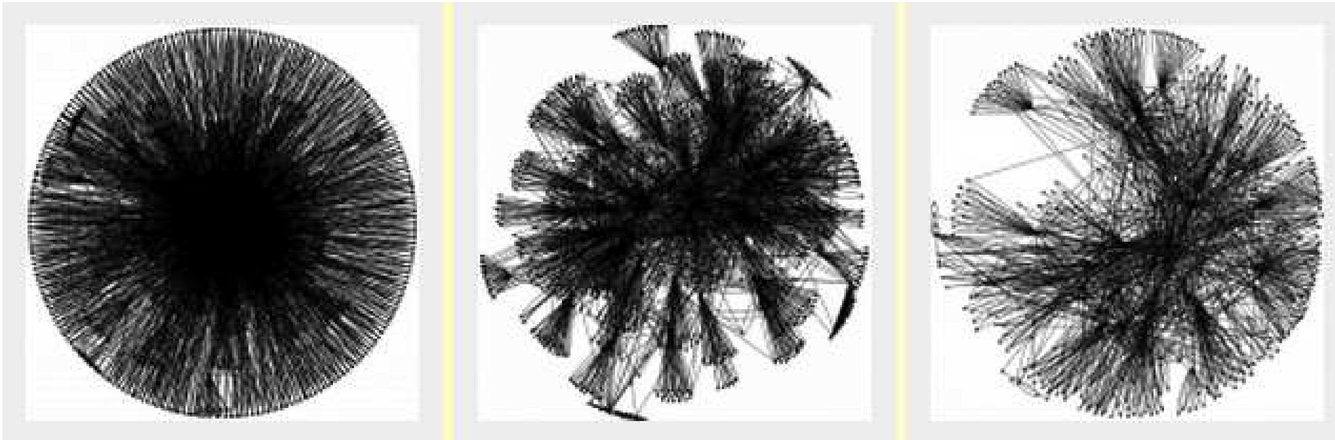

Figure 1: Graphs of all users (the dots) and who they follow (the lines between the dots). The leftmost graph shows all SoundCloud users in the sample of 500 users. The middle graph is the set of users in the sample who are all followed by others in the sample. The rightmost graph is the set of users who are all followed by and following others in the sample.

of our preliminary data explorations was that - unless the sample was very atypical - most of what's on SoundCloud *is* electronic music. Qualitative feedback from SoundCloud users has corroborated this conclusion. The variety of terms used as genres of electronic music is impressively vast, including: house, trance, dance, techno, electro, step, trap, dj, edm, idm, ambient, grime. One step we are now investigating in qualitative analysis is whether we can narrowing our data down to include only certain subgenres of electronic music,or perhaps include all genres but differentiate them in the network analysis. For example we have found 'house', 'trance', and 'edm' are likely to be fairly commercial, with 'step', 'grime', and 'trap' fairly underground. Do these distinctions constitute two different networks to be explored and analysed separately? One way to define our category of interest would be to base it empirically on a combination of interviews and factors arising from the analysis on SoundCloud. Another would be to come up with some (inevitably arbitrary) set of guidelines. This is still under investigation.

Conclusions and future goals

The purpose of this project is to investigate the role of inter-artist networking and peer evaluation in producing the value of cultural works - specifically music, and more specifically electronic music - looking in particular at how this happens via a specific digital platform, and qualitatively situating interactions on that platform with regard to interactions offline. Value judgement is a vital part of creativity, and in computational research on creativity, autonomous or integrated judgements of value are often desired but only occasionally realised. We intend that the methodology developed during the project will be adaptable to carrying out assessments of factors relating to value in a range of cultural contexts, providing an instrument for evidence-based investigation of cultural value.

Acknowledgments

This project is funded under the AHRC Cultural Value call. Thanks to our advisory board members and participants, and to an anonymous reviewer of this paper.

References

Boden, M. A. 2004. *The creative mind: Myths and mechanisms*. London, UK: Routledge, 2nd edition.

Bourdieu, P. 1993. *The Field of Cultural Production: Essays on Art and Literature*. Cambridge, UK: Polity Press.

Csikszentmihalyi, M. 1988. Society, culture, and person: a systems view of creativity. In Sternberg, R. J., ed., *The Nature of Creativity*. Cambridge, UK: Cambridge University Press. chapter 13, 325–339.

Jordanous, A. 2011. Evaluating evaluation: Assessing progress in computational creativity research. In *Proceedings of the Second International Conference on Computational Creativity (ICCC-11)*.

Jordanous, A. 2012a. *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application*. Ph.D. Dissertation, University of Sussex, Brighton, UK.

Jordanous, A. 2012b. A Standardised Procedure for Evaluating Creative Systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation* 4(3):246–279.

McGraw, G., and Hofstadter, D. 1993. Perception and creation of diverse alphabetic styles. *AISB Quarterly* 85:42–49.

Pérez y Pérez, R.; Aguilar, A.; and Negrete, S. 2010. The ERI-Designer: A computer model for the arrangement of furniture. *Minds and Machines* 20(4):533–564.

Ritchie, G. 2007. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines* 17:67–99.

Saunders, R. 2012. Towards autonomous creative systems: A computational approach. *Cognitive Computation* 4(3):216–225.