M2 project report, 2015
University of Warwick

The evolution of a new music genre as an expanding network; a quantitative systems approach to study innovation in the creative economy

Anna Keuchenius[1]*

Abstract
The creative industries are important for economic growth, especially due to their innovative spillovers to other sectors of the economy and society. This research provides a novel conceptual framework of innovation that explains this role of the creative industries, explicating the complex and interconnected nature of innovation. Moreover, this framework is applied in the study of the evolution of a new music genre. Where the focus of former quantitative research has been on best-sellers and sales data, this research aims to give a more complete view by including also the less popular items and taking a distinct systems approach. With real data from an online music platform a bipartite network is created consisting of tracks and users associated to a specific new music genre. Each track and user are connected when the user has bookmarked the track as their favourite. Using bipartite network statistics, the evolution of this network is tracked over a timespan of six years, revealing this genre's structure and trajectory in gaining popularity on the platform. Comparing the results with random networks with the same size and degree distribution, it is found that user-track relations converge to an equilibrium state. Albeit only applied to one small music genre this novel approach gives quantitatively substantiated new insight in the innovation diffusion process of a music genre, and more generally in the creative economy. Namely; innovation should be seen as a sequence of related events contributing to each others success and popularity should not be confused with importance.

Keywords
Innovation — Diffusion — Bipartite network — Music — Genre – Creative Economy

[1] *Centre for Complexity Science, University of Warwick, United Kingdom*

Contents

Introduction

The creative economy[1] (or creative industries) has gained striking attention, both from policymakers and in academia. This is partly due to the fact that the creative industries have continued to grow over the last years, even in times of economic crisis [20, pp 49]. However, even though the creative industries contribute to economic growth as industries as such, their positive influence on other sectors of the economy might be even more important. Authors like John Hartley[12] , Jason Potts[22][21], Stuart Cunningham[24] and Hasan Bakshi [4] have made the novel claim that the fundamental importance of the creative industry lies not just in its contribution to direct economic growth, but in its innovative spillovers to other industries. Hartley[12, pp 8] illustratively writes:

'it can be argued that the 'creative industries' are the empiri-

[1] The term creative economy refers to the socio-economic potential of activities that trade with creativity, knowledge and information. The term is an extension of the term creative industries, which are those parts of the economy that deal with goods and services that use creativity and intellectual capital as primary inputs. There are no formal definitions of creative economy and creative industries, but both terms are widely used and repeatedly defined by the parties using the term. See for example[30][5][20][31]

cal form taken by innovation in advanced knowledge-based economies, in which case their importance – like that of the media – exceeds their scale as a sector of the economy.'

Correspondingly, influential actors such as OECD[20], UNESCO[31] and NESTA[5] now use the term 'creative economy', instead of 'creative industries' to capture this shift in thinking from isolated creative sectors to an innovation fostering creative part of the economy that is intertwined with other parts of the economy.

Within this new way of thinking about the role of the creative economy in innovation, the importance of social networks, human capital and collective learning is particularly emphasized. Potts [24] speak of social network markets, in which consumers, commodities and producers are all connected in a network and value arises from interactions within this network. Hereby, these scholars reject the idea that the creative industries can be apprehended with the simple neo-classical supply demand paradigm[2]. Moreover, Potts et al. [25] claim that creativity is situated within these evolving relationships between consumers and producers, and is subsequently a dynamic entity rather than a static configuration. The view on the role of consumers in the creative innovation process is thus urged to change from passive to active.[3]

I claim that these arguments are based on a conceptual take on innovation that is not new in innovation theory, but has not been explicated yet within the debate. For this reason, I will here provide a novel conceptual framework of innovation that synthesis elements from multiple academic disciplines and explicates the complex and interconnected nature of innovation.

This conceptual take on innovation has two important implications for our understanding of the innovation process. Firstly, the object of innovation is rarely one well-defined item but should rather be seen as a sequence of events. Secondly, the consumers is given a more central role in the innovation process. As a result, answers to questions on how the innovation diffusion process can be studied and how policymakers can stimulate this process are in line with the suggestions made by scholars such as Potts and Hartley.

After the discussion of these theoretical ideas, I use this conceptual viewpoint to study the evolution of a music genre. Instead of focusing only on the most popular tracks of the genre, as is done in previous quantitative studies in music[4], the less popular tracks are also taken into consideration. Also, the consumer is explicitly given a central role. The main question of this case study is whether this method brings more insight into the evolution of a music genre, which serves as an illustrative example for innovation in the creative industries. Of particular interest is the role of hits; how important are they for the genre?

This paper proceeds as follows. Section 1 introduces a theoretical framework of innovation and diffusion. Subsequently, section 2 presents the evolution of a specific music genre in which the framework of section 1 serves as a cornerstone of the investigations. Section 2 is divided into four parts; the first part(2.1) introduces the data, the second part (2.2) discusses the research questions and methods, the third part (2.3)covers the methodology for an appropriate interpretation of the data and the fourth part (2.4) discusses the results. Finally, section 3 concludes what this small case study can tell us about the bigger picture of innovation and diffusion in the creative economy and ultimately across other sectors in the economy.

1. Innovation & diffusion ; a conceptual framework

The study of innovation can be roughly separated into two parts. The first parts has its focus on the producers' side of innovation, centred on creative innovator, the entrepreneur or enterprise. More recently, this centre has broadened to networks of firms, acknowledging the fact that 'an entrepreneur cannot innovate alone'[5] [11]. The second body of literature focusses on the consumers' side of innovation, studying the diffusion of innovations and the adoption behaviour of individuals, initiated by Everett Rogers in 1962.

I argue that these two sides of innovation, creation and diffusion, cannot be separated and studied as individual process, especially not in the case of innovation in creative industries or other meaning- or knowledge- based parts of the economy and society. The main reason for this is that innovation concerns the creation of a novel concept and requires a change in ones framework of thought[6]. Since concepts are mental constructs whose existence relies on a shared meaning and thus implicit agreement between individuals, creation and diffusion of innovation are tangled up by nature.

Consider the internet, one of the most important innovations till date. The internet was not once invented and immediately recognised as promising new technology. Instead, first ARPANET and ETHERNET were developed by a select group of computer scientists, and thereafter only used by government and armies. Only later, when WWW was developed and commercialized, its significance to the average citizen became clear and the term 'internet' has become a widespread concept.

In the following I will show that this take on innovation as the creation of a novel concept requiring as a shift in peoples framework of thought, is present in literature focussing on the creation side of innovation as well as in literature focussing

[2] Some of the basic assumptions of this paradigm are that people make independent choices and that they have rational preferences that can be identified and associated with their -stable- values. These assumptions are clearly violated in the view of Potts et al.

[3] See also [12].

[4] See for example [29] and [1].

[5] This is movement away from individual entrepreneurs or firms towards networks of firms and individuals is also stated by Potts et al., see [25, pp 1]

[6] Here I do not define innovation as the creation of a novel concept, but merely claim that innovations always concern the creation of a novel concept. In other words, this is a necessary condition for the definition of innovation, perhaps not a sufficient one. For example, many definitions of innovation also include an aspect of practical impact of the concept or some value judgement of positive development.

on the diffusion side of innovation.

Innovation from the angle of the supply side

In the vast amount of research on technological innovation, scholars pointed out already long ago that innovation and diffusion cannot be treated as two different processes. J. Stanley Metcalfe writes in 1994 [18, 936]: 'What is clear, is that the separate analysis of innovation and diffusion is no longer tenable, the two are inseparable and mutually reinforcing'. He moves on to argue that this is due to the fact that an innovation is rarely created and defined at once, where after it is diffused. Rather, innovation concerns a sequence of small creations that together lead to a shift in our thinking that creates an opening for all kinds of new things. Citing Metcalfe "what is being diffused is not a single innovation but rather a sequence of linked innovations which reflect the unfolding of a technological opportunity." [18, pp 939]. Rosenberg and Kline [27] [15] illustrate this same point by means of numerous examples of innovations, such as the steam engine, that cannot be connected to one event or one creation.

This technological opportunity then, that Metcalfe speaks of, is not so much an opportunity that arises from the invention of new physical products or techniques preventing the innovation from happening at any earlier point in time. As W. Brian Arthur describes in 'The nature of technology: What it is and how it evolves' [3], a single technology always consists of already existing parts of technologies that have been around for a much longer time, but the novelty lays in the way these parts are put together. The restriction, hence, on creating new technologies, is at the human ability to see the potential of possible combinations. Paul Krugman [16] summarises this, almost poetically, as 'The frameworks of thought which define technical opportunities act to offset the tyrannies of combinational explosion in their development.' Similarly, Metcalfe writes 'The innovation possibility frontiers are cognitive concepts, the anticipated opportunities and the cost of accessing them as expressed in the mind of technologists and managers'.

Both Metcalfe and Arthur can be said to be evolutionary- or complexity economists, and also Krugman finds himself at the edge of the neo classical paradigm, but we find the same sound from a different scientific angle of design studies and innovation: Norman and Verganti distinguish between incremental and radical innovation, where the first are described as improvements within a given frame of solutions and the latter are described as a change of frame, representing a discontinuity with the past. After years of studies in product design and innovation, Norman and Verganti find that radical innovation never results from careful research into persons' or society's needs, as both are trapped in the current context and cultural paradigm. Instead, they claim, "radical innovations have come about simply because their inventors thought they were interesting things to try."[7] [19, pp 3]

These scholars thus acknowledge the importance of a cultural paradigm and human sense making, but their focus is clearly in the supply side of innovation, where the entrepreneurs creativity is a driving force. In another body of literature we find the same conceptual arguments but now extended to the consumers, the demand side of innovation.

Innovation from the angle of the demand side

Everett M. Rogers (1962) [26, pp 16] defines a technology as "a design for instrumental action that reduces the uncertainty in the cause-effect relationships involved in achieving a desired outcome", and distinguishes between a hardware and software component of technology. For some technologies the hardware component is much more visible, but in other cases the technology consists almost entirely of information, such as with a political idea, a news event or a rumour. It might be a little unusual to consider a news event or rumour as a technology, but Rogers points out that they are subject to the same forces governing their diffusion as hardware dominant technologies. He noticed that when the growth of an innovation was set against time, a similar curve was found for all innovations, independently of its discipline. This curve has an S-shape, like a logistic function, see figure 1 .

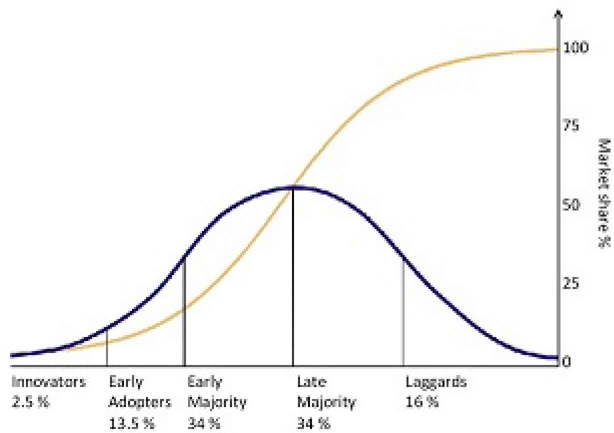


Figure 1. *A schematic example of the growth curve of an innovation per time unit (blue line) and cumulative (orange line). Rogers classified the people adapting at different times as shown on the x-axis.*

Rogers classified the adopters on the innovation according to their timing as 'innovators', 'early adaptors', 'late adaptors, 'early majority' and 'laggards'. The reason people belong to different adoption classes, according to Rogers, is not simply information asymmetry. Even though one is well aware of the existence of an innovation, he might not be willing to adopt, because the benefits of the innovation are not yet clear for him. Before adaption thus come "information-seeking and information-processing activities in which the individual is

[7] Norman and Verganti, however, do move on to say that this is the case for most technological innovations. They argue that there is another branch of radical innovation that results entirely from a change of meaning for the individuals.

motivated to reduce uncertainty about the advantages and disadvantages of the innovation" [26, pp 32], shortly described as the adaptive behaviour of individuals. This adaptation behaviour, according to Rogers, depends on individuals' social network and should be seen in the context of a particular society, since both determine the context in which information processing takes place.

Norman and Verganti similarly emphasize the importance of human sense making in the diffusion of innovation. They give the example of Apple's development of multi-touch interfaces and their associated gestures to control handheld and desktop systems. This development has been crucial to the success of the computer. However, Norman and Verganti stress that Apple's creations were not new, but that they were important for the understanding and meaning for the consumers. 'Although Apples ideas were not radical to the scientific community, they did come as radical, major shift in the world of products and how people interact with them and give meaning to them.' [19, pp 6] Norman and Verganti conclude that "radical innovation brings new domains, new paradigms and creates a potential for major changes." [19, pp 6]

To summarize, we find that innovation has been studied as having two sides to it; a supply side of innovation on the hand, where a combination of actions by creative entrepreneurs make room for a new technological opportunity, and one the other hand the demand side, where the individual's information processing and learning path enables him to see the advantage of the innovation. Both are dependent upon a framework of thought, and on the shared cultural paradigm in which this framework exists. Innovation, then, concerns a shift in this framework, which must take place for both producers and consumers.

With this in mind, the arguments of scholars and institutions stressing the importance of the creative economy can better be understood. The distinguishing feature of the creative economy is the socio-economic exploitation of the production of meaning, symbolism and experience that represent value for the consumers, hence this is how the creative economy can also foster innovation in other sectors. This is precisely what Potts argues, when he writes that the creative industries "drive, facilitate and engender the origination, adoption and retention of new ideas (the innovation process) into the sociocultural and economic system" [23, pp 2].

Moreover, this view on innovation also changes what methods are appropriate to study innovation. In the creative industries, quantitative studies have focused on the most popular items, using for example sales data [29]. In particular, in studies of music scholars focused on the hits using top-lists of the radio channels or Billboard chats [1]. This section, however, implies that studying innovation should include multiple events, not just the most successful ones, and an active role of the consumer. Ideally, product consumer interaction shall be taken into account as well. In the next section I aim to do just that, tracking the evolution of a music genre on a particular online music platform.

2. Case study ; Evolution of a new music genre

In this section the evolution of a new music genre is investigated. I do not consider this new music genre explicitly as an innovation, but I do argue that its evolution is similar to that of an innovation. As explicated in the previous section, innovation concerns a sequence of linked events that require a change of framework of thought for producers and consumers. The linked events are in this case the music tracks that are released in the new genre. The change in framework of thought for consumers are the willingness to listen to the new music and subsequently recognise and appreciate the novelty. Ultimately, a new (abstract) concept is created, which is in this case the music genre as distinct concept, that can be spoken about.

Motivated by the conceptual arguments of the previous section, this study does not limit itself to the most popular items. Instead, all tracks that were considered as part of the genre by the audience of a specific online music platform are taking into consideration. Furthermore, the consumers are given an explicit central role.

Since this is a new method, the main research aim is to see whether this method can give valuable new insight into the development of the genre. Of particular interest is the role of hits; How important are the hits for the growth and structure of the genre, and firstly how does one define growth and the structure of the genre?

Section 2.1 first introduces the data for this case study. Thereafter, section 2.2 describes the research questions and research method. Section 2.3 subsequently spends a few more words on the methodology to compare the results of the data to an appropriate null-model. Section 2.4 discusses the results.

2.1 Data

The music genre that will be taken as a case study in this research is Witch House. This is a dark electronic house music that emerged in the late 2000's. Witch House gained attention from music critics and became popular in underground scenes, but remained a small genre. Some say it has already 'died' by now.[8] Witch House is taken as a topic of this study for two important reasons. Firstly, since the genre emerged after 2007, there is data available from it's early days. Secondly, because the genre remained small, the amount of data is still managable. Moreover, there is little influence from main stream channels,such as radio and television, because the genre remained underground.

The music platform from which data is gathered is the HypeMachine. This website was launched in 2007 and became immediately popular. The platform has now grown to over one million users. On the HypeMachine users can listen to tracks and read what music bloggers have been writing about the tracks. In addition, users can follow blogs, follow other

[8] See for example `http://noisey.vice.com/blog/creep-were-glad-witch-house-is-dead`

users, become 'friends' of other users, and bookmark tracks as one of their favourites, called 'liking' or 'loving' in HypeMachine's terminology. Data is gather from the Hypemachine by means of a scraping program I wrote in Java.

On the HypeMachine, the tracks are listed with a number of tags -maximum ten- below it. These tags are user generated tags from another online website, Lastfm. Lastfm is also a music platform, with over 30 million users[9]. To identify the Witch House tracks, all the tracks of the Hype machine were selected that contained the tag "Witch House" as one of their five most ascribed tags. Thereafter, all the users that liked one of the Witch House tracks before 01-01-2015 were selected and the date of their like stored. All the data was stored into PostgreSQL database and later put into different formats to calculate statistics using Matlab, C++, Excel and Python.

The data covers the period from 01-01-2007 until 31-01-2014. At the end of 2014, the data consists of 1.248 tracks, 33.074 unique users and 64.495 likes within the genre.

2.2 Research Questions and Methods

An obvious first step to explore the evolution of the genre, is to check its growth trajectory. What this means, however, is less straightforward. Is the number of tracks a sufficient indicator for the size of the genre or should we rather look at their success -the total likes these tracks obtained- ? To investigate the growth I therefore look at three different measures: the number of tracks in the genre, the number of total likes these tracks obtained and the number of users responsible for those likes. Since innovations commonly show an S-shaped growth curve, as in figure 1, I expect to see something similar.

As a second step to explore this genre, is to looks at its structure. Firstly, I will ask how the number of likes is distributed over the different tracks. Do we see an extremely uneven distribution, such as the power law, as more frequently observed in popularity distributions? [9] [13] [10] [28] [8] Secondly, I am interested how the different tracks are related to another, from the perspective of the users. Does the genre, for example, consist of several subgenres that are not much related to another? How 'far away' are tracks from each other, in the perception of the users?

These questions can be answered by taking a network perspective. In fact, the data essentially is a network. This networks has two kinds of nodes; tracks (songs) and users. Each track and user are connected if the user has bookmarked the track as one of his favourites, see figure 2.

Since each connection is linked to a specific date, this network is expanding in time. The structure of this network, which is also changing over time, tells us something about the structure and formation of the genre. However, capturing the structure of a bipartite network –a network with two sets of nodes- is not a trivial task. This research therefore makes use of a set of statistics for the analysis of bipartite networks suggested by Latapy et al [17]. These statistics include the number of nodes in the giant component, the average distance, various statistics to capture clustering and several degree correlations. Before discussing the results, the next section spends a few more words on an appropriate methodology for interpretation of the data.

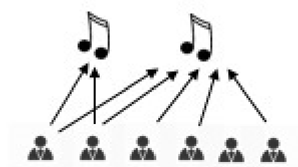


Figure 2. *A schematic example of a network with tracks and users. There are only edges between tracks and users, not between users and users or tracks and tracks.*

2.3 Methodology

The majority of existing network analysis is done on one mode networks, in which there is only one kind of node. This means that there is a more developed toolkit for analysis of one mode networks and better material for comparisons. Often, therefore, bipartite networks are 'projected' onto one of the node-types by linking two nodes when they have a common neighbour, and weighting the link relative to the number of common neighbours, see Figure 3.

Figure 3. *A schematic example of a bipartite network (left) projected on one of the nodes (right)*

This projection then forms a one mode network and the data is further analysed as such. However, a number of studies has stressed the information loss and biases that are introduced by this method [17]. Latapy et al. therefore developed a simple set of statistics to analyse bipartite networks while keeping their bipartite nature. This research makes use of this set of statistics.

In order to infer anything from the findings in the data, the statistics need to be compared to an appropriate null model. The null model should reflect a random realisation that lacks the effect for which the data is tested upon. However, what 'random' means is still a topic of discussion in network studies. One of earlier propositions was the Erdős Rényi Graph, that takes the number of nodes and the number of links (or the probability of a link) in the graph as fixed, and generates a network by randomly creating edges between each pair of nodes. However, the resulting degree sequence[10] is Poisson distributed and many empirical studies showed that real world networks rarely have Poisson distributed degree sequences.

[9]This is based on an estimate of Lastfm in 2009 (`http://blog.last.fm/2009/03/24/lastfm-radio-announcement`)

[10]The degree sequence of the network is a sequence of numbers, each number representing the degree of a specific node. The degree of the node is the number of connections/edges the node has.

Instead, real world networks, especially in social sciences, often have a degree sequence that follows a power law. In fact, as will be described in the next section, my data has a power law distributed degree sequences too -see appendix 2-. I have therefore constructed random networks by keeping the size and the degree sequences fixed. I used the algorithm by Hyunju Kim, Charo del Genio et al. [14] that creates random configurations of networks with fixed size and degree sequences, and average the obtained statistics over 100.000 samples. See appendix 1 for a more detailed explanation.

2.4 Results

2.4.1 Growth trajectory

The growth of Witch House, on music platform the Hypemachine, seems to be consistent with the diffusion of an innovation. Figure 4 shows the growth of the genre per year (upper chart) and the cumulative growth per month (lower chart). The green line shows the number of Witch House tracks, the blue line shows the corresponding likes these tracks obtained and the orange line shows the number of users that liked at least one Witch House track. All numbers are divided by their totals in January 2015 to allow for comparison. This figure shows that all growth curves have the S-shape which is well known for the diffusion of innovations, see figure 1.

One might suspect that the curve of the likes and users simply follows the curve of the tracks, and hence its shape is not an independent finding but rather an unavoidable consequence. However, the relationship between tracks and likes is not that simple, as the upcoming section 2.4.2 will show. Some tracks gain immense popularity whereas other are barely liked.

Judged by these growth curves, the first Witch House tracks became available on the Hypemachine in 2007 and their number grew at increasing rate until 2012. These Witch House tracks did not receive immediate popularity. Over time, more Witch House tracks circulated on the Hypemachine, retrieving more likes from existing but also from users new to the genre. The year 2012 could be identified as the peak of the genre, having the highest percentage of new tracks, likes and attracting the highest percentage of new users. After 2012, users, tracks and likes still increased but at a decreasing rate.

The number of artists new to the Witch House genre shows similar behaviour as the three growth curves of figure 4. Linking each artist to the date of her/his first track on the Hypemachine, the growth of Witch House artists is visualised in figure 16. Amongst these artists, the number of tracks available on the Hypemachine that are within the Witch House genre varies from 1 to 56. Also, the proportion of Witch House tracks amongst the total tracks on the Hypemachine of a particular artist varies widely (see appendix 3). Some artists have more than 80 per cent of their tracks within the Witch House genre. However, this doesn't mean artists with 80 per cent Witch House tracks are big Witch House artists; they might simply have only a few tracks on the Hypemachine. Clams Casino, for example, has 56 Witch House tracks out of his 61 Witch House tracks available at the Hypemachine. But many

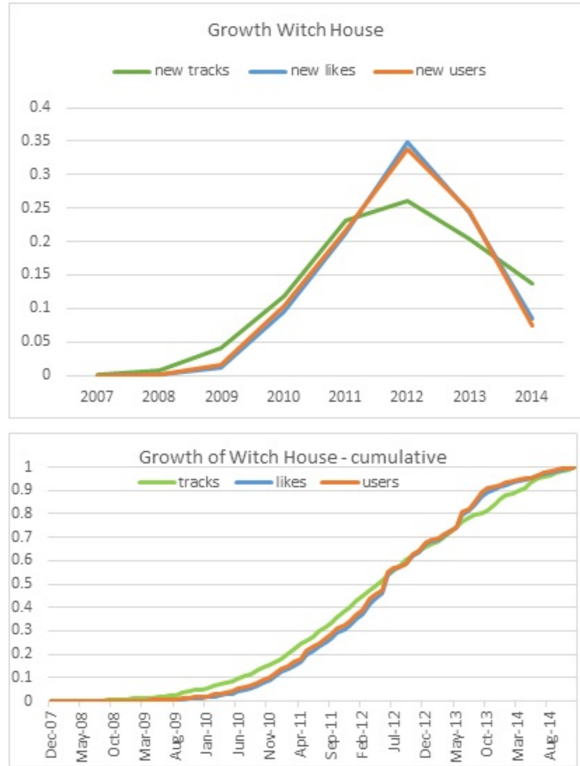


Figure 4. *This figure shows the growth of Witch House per year (above) and cumulative per month (below).The green line indicates the total number of tracks tagged as Witch House as a percentage of this total in January 2015. The blue line indicates the total of likes these tracks obtained, as a percentage of the total in January 2015. The orange line indicates the number of fans that liked at least one track in the genre, as a percentage of the total in January 2015*

other artists, 77 in total, have less than five tracks available on the Hypemachine, all of which are considered Witch House tracks.

I would like to consider these latter artists as a special category. Almost none of their tracks were available on the Hypemachine before 2011, and the year 2012 is responsible for most of them, see figure 16. In other words, these are not innovative artists producing Witch House tracks in the early days of the genre. Neither are these artists that produced some Witch House tracks besides their other tracks available on the Hypemachine. Rather, these seem to be artist 'free riding' on the popularity of Witch House in 2011 and 2012.

2.4.2 Track popularity

As mentioned earlier, some tracks within the genre are very popular, whereas the majority received few likes. To be more precise, the twenty percent most popular tracks are responsible for 84 per cent of the likes in the genre[11]. Figure 14 shows the cumulative distribution of track likes in January 2015. The data is plotted on a log-log scale for better visualisation

[11] Here the genre is defined as the 1 248 Which House tracks of the data, and the total likes are all the likes these 1 248 tracks obtained on the Hype machine until 01-01-2015.

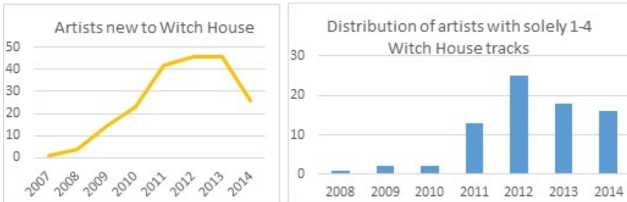


Figure 5. *This figure shows the growth of Witch House artists per year (left) and the number of 'free riding' artists per year (right). These latter artists have solely Witch House tracks available on the Hypemachine and no more than four tracks.*

purposes. Due to the low number of observations for statistical testing (the number of tracks is 1.248) the distribution cannot be identified with high accuracy, but the data seems to be consistent with a power-law of exponent 1.9[12].

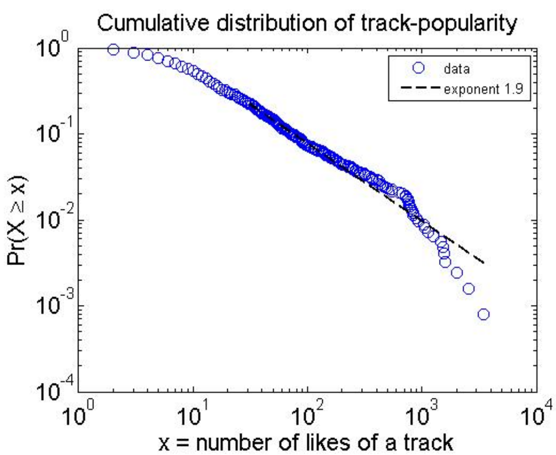


Figure 6. *This figures shows the cumulative distribution for 'likes' that tracks received (blue). It is plotted on a log-log scale. The data seems to be consistent with a power law with cut-off and exponent 1.9*

Such an uneven distribution in popularity is not surprising, especially not in the creative industries. Studies commonly find power laws in empirical data of the creative industries [9] [13] [10] [28] [8], and some scholars even speak of society being hit-driven (Chris Anderson [2] for example). What is new, however, is the finding that this uneven distribution also holds within a certain genre. In terms of direct market success, the hits thus seems to be important for the genre.

Yet, in other terms the hits might not be so important for the genre. For example, even though hits themselves received a lot of attention, this attention might have been of short notice. Consider for example the proportion of likes that were generated by users that only liked that one single Witch House track and besides that no other tracks in the genre. This proportion turns out to be increasing with popularity (see appendix 3). For the most popular track (*Alarm* by the artist Wise Blood) fifty percent of the likes were generated by users that solely liked that track, whereas this percentage is on average 0.21 for Witch House tracks.

This relationship between users and track popularity can be further investigated by counting the number of likes of each user per track and taking its average. This average of user likes indicates what kind of users are into the track; fans that like many Witch House tracks, or users that only like a few tracks in the genre. Figure 7 shows how this relates to track-popularity. The x-axis corresponds to the number of likes of a track, and the y-axis shows –on average- how much 'fan' the users are. A horizontal line would mean that there is no relation between track popularity and the kind of users liking the track. The Figure shows that unpopular tracks are much more often liked by fans that like many Witch House tracks. The more popular the track, the less 'fan' the users are.

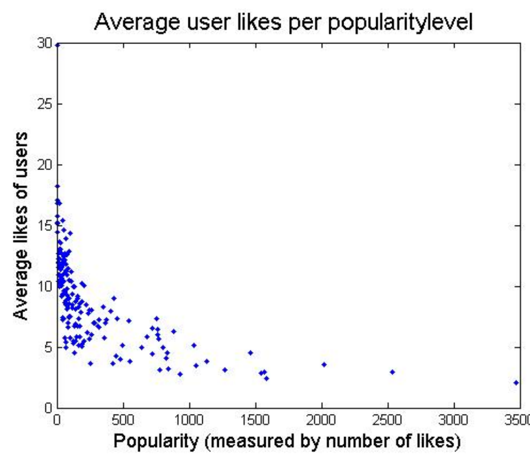


Figure 7. *This figure shows the number of Witch House tracks the average user likes that also likes a track corresponding to the popularity level on the x-axis.*

Perhaps it is not correct to simply call the most popular tracks 'hits', without further discrimination. Figure 8 shows how the most popular 13 tracks (1%) accumulated their likes over time. Each line represents a different track, and the height in each month indicates the corresponding number likes the track obtained as a proportion from its total likes. Figure 8 also shows the growth curve of total Witch House tracks –the thicker green line-, earlier shown in figure 4, so that the accumulation times can be seen in comparison to the overall growth rate of the genre in terms of tracks. It is striking that the three earliest popular tracks did not receive immediate attention, but rather accumulated their likes over a long period of time. The tracks in 2012, on the other hand, received immediate attention accumulating almost all their likes within the same month.

There are many possible explanations for differences in like- accumulation times. However, it is remarkable that the most popular tracks during the peak of the genre, which is roughly 2012, gained immediate attention, whereas the earlier popular tracks did not. This suggests that the timing of a track, seen from the perspective of the evolution of the genre,

[12] The power-law is tested with the method developed by Clauset & Newmann [7]. One can never proof that the data comes from a power-law distribution but with a significance level of 1% it cannot be rejected

influences the accumulation time.

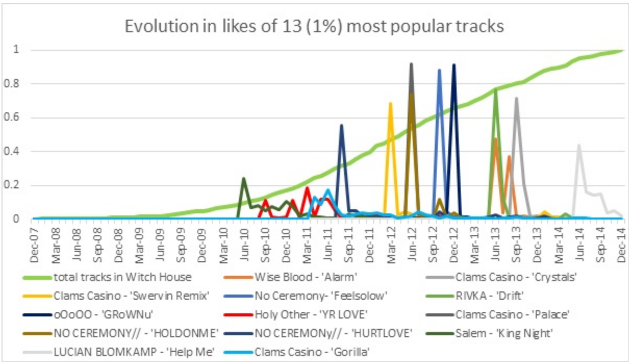


Figure 8. *Each line in this figure represents one of the 13 most popular tracks and the height corresponds to the number of likes this track obtained as a proportion from its total likes in January 2015. The thicker green line is the growth curve of tracks, earlier displayed in figure 4*

2.4.3 Relations between tracks&tracks and tracks&users

Static: By the end of 2014

Seen from the perspective of the users of the Hypemachine, Witch House does not consist of different subgenres that are little related to one another. In the network of users and tracks, almost all tracks (99%) are part of the giant component in January 2015. This means that almost all tracks are connected through some user preferences. For example, track A and track B are connected, because user 1 liked track A and track C (but not B), and user 2 liked track C and track B (but not A). The path from track A to B is hence of length 4, but this could be any other even number. In fact, this example not far from reality, since the average path length in January 2015 is 4.05. This suggest that there is some structure to the network causing this small-world property. However, if we compare this average distance to the random network with same degree sequences, we find a similarly low average distance, namely 3.95. This property is therefore a necessary result of the size of the network and the specific power law degree distributions and cannot be said to be the result of other structures in the network.

When the network is projected onto the tracks, creating a network with only track nodes and connections between the tracks in case one or more users liked both tracks, there is also no indication for different subgenres. The modularity of this network measures how well the network can be divided into smaller communities, representing subgenres. The modularity value is 0.23[13] which is low for networks of this size[14]. Changing the weights of the edges between tracks does not change this property; if the edge weight is proportional to the total number of users liking both tracks the modularity is 0.22 and when these edges are subsequently weighted by the sum the total likes both tracks received the modularity remains low.

As mentioned earlier, information is lost when these kind of projections are made. For example, a tracks degree in the projected network might not be proportional to its degree in the bipartite track-user network –which is the total likes a track obtained- due to overlap in user preferences. Multiplying the bipartite degree of the track with its average user likes gives the expected degree in the track projection network, in case there are no overlaps in user preferences. Figure 9 shows this expected degree, blue dots, and simultaneously the observed degrees, green dots, in this network. The figure shows that the overlaps in user preferences are significantly higher for more popular tracks.

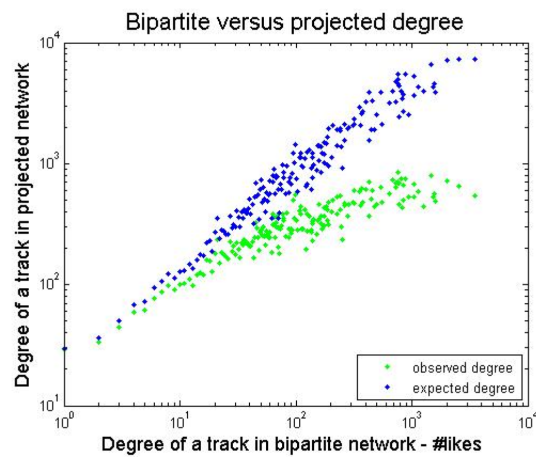


Figure 9. *This figure shows the relation between the degree of track in the bipartite network of tracks and users (which is equal to the number of likes the track received) and the degree of the track in the projected network. The blue dots are the expected degrees in the projection, based on the average user-likes of figure 7 and the green dots are the observed degrees.*

In order to quantify these overlaps, I use several clustering statistics proposed by Latapy et al **??** The first statistic measures the number of common neighbours between two nodes (intersection), as a proportion of their total neighbours (union), i.e. for nodes u and v:

$$cc_{\bullet}(u,v) = \frac{\mid N(U) \cap N(v) \mid}{\mid N(u) \cup N(v) \mid} \tag{1}$$

Averaging this over all neighbours of a node, gives the clustering coefficient per node, i.e:

$$cc_{\bullet}(u) = \frac{\sum\limits_{v \in N(N(u))} cc_{\bullet}(u,v)}{\mid N(N(u)) \mid} \tag{2}$$

The second statistic is similar, but considers the number of common neighbours between two nodes as a proportion of neighbours of the node with the smallest number in neighbours:

$$cc_{\underline{\bullet}}(u,v) = \frac{\mid N(u) \cap N(v) \mid}{min\left(\mid N(u) \mid, \mid N(v) \mid\right)} \tag{3}$$

[13] This is calculated by means of Louvains method, described in [6]

[14] It is difficult to determine what is high and what is low for this kind of statistic,but comparing the value to the modularity values of networks discussed in [6], 0.23 is a very low value.

This measure is useful in case the degree of nodes varies widely, which is the case for the data of this research. Then, it can often happen that the neighbourhood of one low-degree node is completely included in the neighbourhood of another high-degree node, see figure 10 for an illustration. However, due to the high degree of the latter node the clustering according to $cc_{\bullet}(u,v)$ remains low. Clustering according to $cc_{\underline{\bullet}}(u,v)$, on the other hand, would be high.

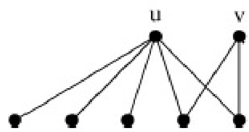


Figure 10. *An example of a low degree node (v) whose neighbourhood is completely included in the neighbourhood of a high degree node (u)*

A third way to measure clustering, is to measure the probability that, given four nodes with at least three edges, they are connected with four edges. This is measured by the fraction:

$$cc_{\rtimes}(G) = 2 * \frac{N_{\bowtie}}{N_{\rtimes}} \tag{4}$$

where G stand for the graph, $\rtimes$ stands for three links between four nodes and $\bowtie$ stand for four links between four nodes.

Notice that this last statistic is not node-centred, such as the other two statistics, and hence every network has one such corresponding clustering coefficient. To obtain the clustering of a network by means of the first two statistics, the node coefficients need to be averaged over all nodes. In this study I will focus on the clustering of the tracks, hence the presented values for $cc_{\bullet}(T)$ and $cc_{\underline{\bullet}}(T)$ are averages over all tracks.

The value of clustering found in the network in January 2015, according to the three measures, are given in the first column of table 1

Table 1. *Clustering coefficients*

Statistic	Observed	Random
$cc_{\bullet}(T)$	0.0380	0.0869
$cc_{\underline{\bullet}}(T)$	0.212	0.230
$cc_{\rtimes}(G)$	0.0176	0.00426

The fact that the value of $cc_{\underline{\bullet}}(T)$ (0.212) is much higher than the value of $cc_{\bullet}(T)$ (0.038) indicates that the neighbourhoods of many unpopular tracks are indeed included in the neighbourhoods of more popular tracks. In fact, it looks as if the neighbourhood of less popular tracks are mostly included into then neighbourhoods of the most popular tracks. Figure 11 shows the clustering coefficient, measured by $cc_{\bullet}(T)$ and $cc_{\underline{\bullet}}(T)$, per popularity level. As expected, $cc_{\bullet}(T)$ is decreasing as popularity increases, but $cc_{\underline{\bullet}}(T)$ instead increases at some point. This suggests that many users liking the least popular tracks also like the most popular tracks.

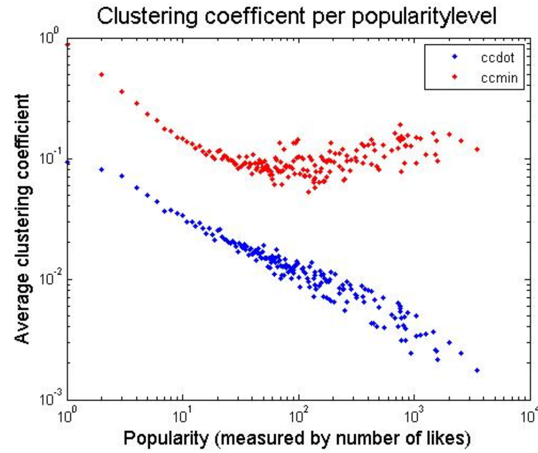


Figure 11. *This figure shows the average clustering (according to ccdot (blue) and ccmin (red) per popularity level, plotted on a loglog scale. By definition of ccdot, the average clustering is expected to decrease as popularity increases. However, this need not hold for ccmin.*

Comparing the found clustering values for $cc_{\bullet}(T)$ and $cc_{\underline{\bullet}}(T)$ to the clustering in a random network, displayed in the second column of table 1, gives the unexpected result that there is less clustering in this network than in a random network. The observed and random values of $cc_{\rtimes}(G)$, however, suggests the opposite. Although there are more explanations possible, I here provide explanation that only uses the earlier finding (of Figure 11) that users liking the unpopular tracks often like the most popular tracks too. Consider the situation in figure 12a) and b). Both network configurations have the same size and degree sequences, but in a) the clustering, according to $cc_{\bullet}(T)$ is higher (0.097) than in b) (0.03) simply because in b) most users liking unpopular tracks 1 or 2 also like the most popular track 3. The clustering according to $cc_{\rtimes}(T)$, on the other hand, is lower in a) (0.015) than in b) (0.051). I therefore suspect that the network of tracks and users consists of substructures that rather look like b) than –the more random- a).

Dynamic: Evolving over time

The evaluation of the network structure of time reveals that the relations between tracks and users converge to some kind of equilibrium. Figure 13 shows the proportion of tracks in the giant component, the number of tracks outside the giant component, the average distance between tracks, the cluster coefficients of $cc_{\bullet}(T)$, $cc_{\underline{\bullet}}(T)$ and $cc_{\rtimes}(T)$ over time. All these values, except the number of tracks outside the giant component, fluctuate until roughly 2011/2012 and thereafter converge to a stable value. The number of track outside the giant component, instead, increases until 2011 and thereafter almost steadily decreases.

This can be interpreted as though roughly by the end of 2012, the genre was established. From then on existing and new tracks were recognised and liked by Witch House likers, connecting them to the overall Witch House genre. Additional

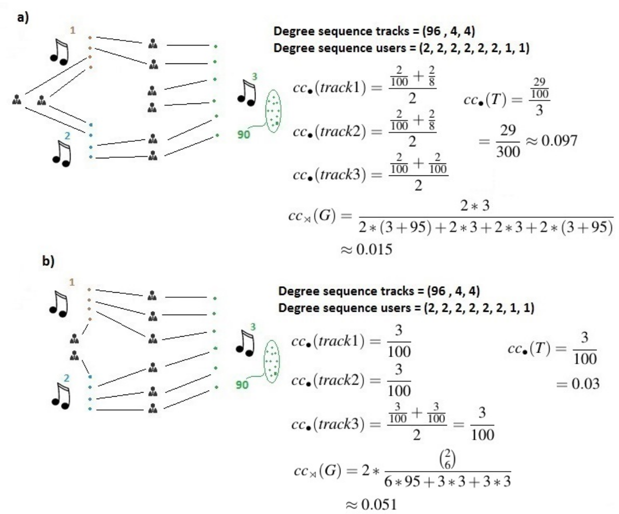


Figure 12. *This figure shows two configuration of a network with the same size and same degree sequences. This network has two unpopular tracks, with both four likes, and one popular track with 96 likes. The clustering coefficient $cc_\bullet(T)$ is higher in a) than in b). The clustering coefficient $cc_\rtimes(T)$, on the other hand, is higher in b) than in a).*

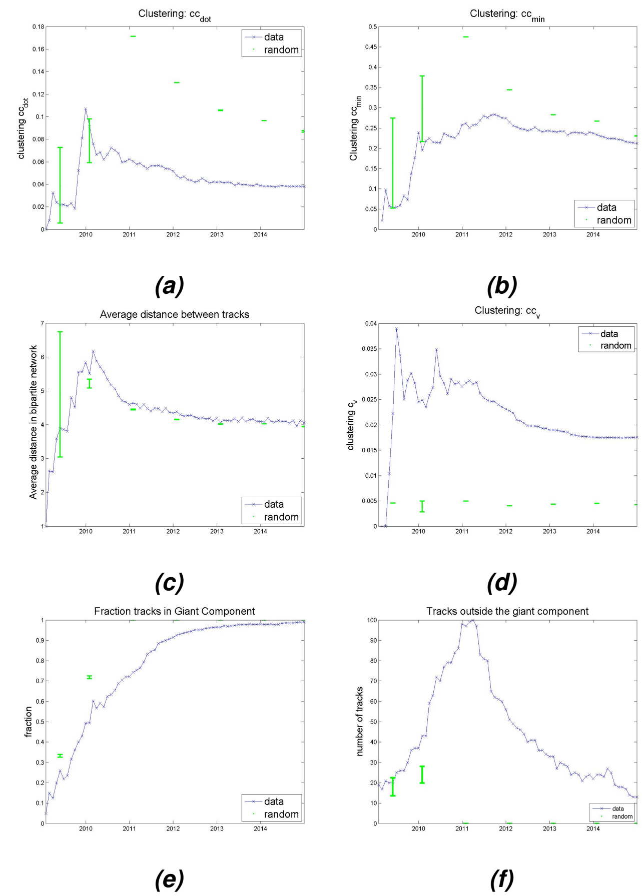


Figure 13. *This figure shows six different statistics of the bipartite network of tracks and users in time. The blue line represents the actual data, for which the statistics are calculated each month. The green points are the statistics of a random network, with same size and degree distribution. The random networks are only sampled for 6 different times throughout 2009-2014. The Figure also indicates errorbars on the random network statistics, which are very small due to high number of samples.*

tracks, likes and users might have shifted individual popularity or other node-characteristics (such as centrality or degree) but the macroscopic properties of the relations between tracks and users –the bipartite network- remained more or less equal.

This is an interesting finding, especially compared to the growth curves of the genre, displayed in Figure 4. Based on the growth in terms of the number of tracks that became available on the Hypemachine, the number of users that liked their first Witch House track and in terms of total likes the Witch House tracks obtained, the year 2012 -from 01/2012 till 01/2013- seems to be the most important year for Witch House. According to the networks statistics, however, this year is not very influential to the macroscopic relations between users and tracks.

3. Conclusions and Further Research

The main inspirational foundation of this research is the conceptual notion that innovations are sequences of related events that illustrate a change in peoples framework of thought. This change in framework must take place among both producers and consumers in order to be called an innovation. These innovations are often denoted by a single term, such as 'the internet' or 'the computer', even though none of such entities has ever been created at once and these innovations have consequences that can extend far beyond that single term. It follows, that innovation should be studied as a process in which producer, product and consumer interaction play a key role. The second part of this paper is a first attempt to do this in a quantitative fashion, by investigating the evolution of a music genre. Defining a genre as a collective of tracks that are all considered to be part of a genre through the eyes of consumers, the data under investigation reveals that the approach of comparing the evolution of a genre to the process of innovation is promising. The growth of the genre, in terms of tracks (songs) and popularity, follows the S-curve that is characteristic for the diffusion of innovations. The genre started small, then grew exponentially and after its highest growth, the peak, decreasingly grew until today. Along the same growth curve, also new artists started producing tracks in the genre.

The popularity within the genre is extremely unequally distributed; whereas most of the tracks were barely liked, a small number of tracks gained immense popularity. The unpopular track are mostly liked by consumers that like many tracks in the genre, including some of the highly popular ones. This results in a network structure between tracks and users that rather resembles a core-periphery structure, with popular tracks in the core surrounded by unpopular tracks, than a clustered network that can be roughly separated into communities of tracks and users, i.e. sub-genres. The evolution of this network structure in time shows that the consumer-product relations converge to an equilibrium, already before the peak of the genre. Although track- and consumer specific characteristics, such as popularity and centrality, may still change over time, the structure of the consumer product network does not endure any more large changes.

On the basis of this analysis, the early tracks, artists and consumers seem to determine the structure of the genre and the later tracks, artists and consumers only confirm the already existing structures. Although the later tracks and users might from a majority in numbers and their influence on popularity is significant, their influence in the establishment of the genre as a concept seems limited. The most important implication of this research is therefore that popularity should not be confused with importance. However, the conclusions of this research are meant to open a debate about the evolution of music genres and the study of the evolution of music genres as illustrating processes of innovations, not to close one. Further detailed investigation of the specific network structures of the analysed genre is needed to validate these conclusions more robustly. Also, a proposal for future research is to investigate other music genres with this approach and ultimately other 'products'.

References

[1] Peter J Alexander. Entropy and popular culture: product diversity in the popular music recording industry. *American Sociological Review*, pages 171–174, 1996.

[2] Chris Anderson. *The long tail*. Wereldbibliotheek, 2013.

[3] W Brian Arthur. *The nature of technology: What it is and how it evolves*. Simon and Schuster, 2009.

[4] Hasan Bakhshi, Eric McVittie, James Simmie, et al. *Creating Innovation: Do the creative industries support innovation in the wider economy?* Nesta London, 2008.

[5] Hasan Bakshi, Ian Hargreaves, and NESTA Mateos-Garcia, Juan. A manifesto for the creative economy. Technical report, april 2013.

[6] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[7] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[8] Raymond AK Cox, James M Felton, and Kee H Chung. The concentration of commercial success in popular music: an analysis of the distribution of gold records. *Journal of cultural economics*, 19(4):333–340, 1995.

[9] JA Davies. The individual success of musicians, like that of physicists, follows a stretched exponential distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 27(4):445–447, 2002.

[10] Arthur De Vany. *Hollywood economics: How extreme uncertainty shapes the film industry*. Routledge, 2004.

[11] Christian DeBresson. An entrepreneur cannot innovate alone: networks of enterprises are required. In *DRUID Summer Conference*, pages 9–12. Citeseer, 1999.

[12] John Hartley. From the consciousness industry to the creative industries: consumer-created content, social network markets, & the growth of knowledge. *Media Industries: History, Theory & Method*, pages 231–244, 2009.

[13] John Hartley, Jason Potts, Terry Flew, Stuart Cunningham, Michael Keane, and John Banks. *Key concepts in creative industries*. Sage, 2012.

[14] Hyunju Kim, Charo I Del Genio, Kevin E Bassler, and Zoltán Toroczkai. Constructing and sampling directed graphs with given degree sequences. *New Journal of Physics*, 14(2):023012, 2012.

[15] S.J. Kline and Nathan Rosenberg. An overview of innovation. *In R. Landau & N. Rosenberg (eds.), The Positive Sum Strategy: Harnessing Technology for Economic Growth.*, 10(1):275–305, 1986.

[16] Paul Krugman. Toward a counter-counterrevolution in development theory. *The World Bank Economic Review*, 6(suppl 1):15–38, 1992.

[17] Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48, 2008.

[18] J Stanley Metcalfe. Evolutionary economics and technology policy. *The economic journal*, pages 931–944, 1994.

[19] Donald A Norman and Roberto Verganti. Incremental and radical innovation: Design research vs. technology and meaning change. *Design Issues*, 30(1):78–96, 2014.

[20] OECD. Tourism and the creative economy. *OECD studies on tourism*, 2014.

[21] J Potts. Art & innovation: an evolutionary economic view of the creative, 2007.

[22] Jason Potts. *Creative industries and economic evolution*. Edward Elgar Publishing, 2011.

[23] Jason Potts. *Creative industries and economic evolution*. Edward Elgar Publishing, 2011.

[24] Jason Potts, Stuart Cunningham, John Hartley, and Paul Ormerod. Social network markets: a new definition of the creative industries. *Journal of cultural economics*, 32(3):167–185, 2008.

[25] Jason Potts, John Hartley, John Banks, Jean Burgess, Rachel Cobcroft, Stuart Cunningham, and Lucy Montgomery. Consumer co-creation and situated creativity. *Industry and Innovation*, 15(5):459–474, 2008.

[26] Everett M Rogers. *Diffusion of innovations*. Free Press, 2003.

[27] Nathan Rosenberg. Factors affecting the diffusion of technology. *Explorations in economic history*, 10(1):3–33, 1972.

[28] Sitabhra Sinha and Raj Kumar Pan. How a hit is born: The emergence of popularity from the dynamics of collective choice. *Econophysics and Sociophysics: Trends and Perspectives*, pages 417–447, 2006.

[29] Paul Stoneman. Soft innovation: changes in product aesthetics and aesthetic products. In *Royal Economic Society Annual Conference, University of Warwick. March*, pages 17–19, 2008.

[30] UNCTAD. Creative economy report 2008. Technical Report UNCTAD/DITC/2008/2, United Nations, 2008.

[31] UNESCO. Creative economy report 2013 special edition. Technical report, United Nations Development Programme (UNDP) and UNESCO, 2013.

Appendices

Appendix 1

I sampled networks for six different points in time; 05/2009, 01/2010, 01/2011, 01/2012, 01/2013, 01/2014 and 31/12/2014. For each point in time, I used 100.000 samples. I used the facilities of Warwick University to use a supercomputer to run the sampling algorithms. The code was ran in parallel, with twenty 'tasks' each producing 5000 samples, per point in time.

To sample random networks with the same size and degree sequence of my data, I used the algorithm developed by Hyunju Kim, Charo del Genio et al. [14]. This algorithm has several main advantages. Firstly, it is a fast algorithm. Secondly, the created sample networks are statistically independent of each other and thirdly it provides the user with an appropriate probability for all sample configurations of the network. This is why error bars can subsequently be determined to the average statistics of the sample networks.

The reason the algorithm provides sample probabilities is the following. The algorithm builds the network step by step by creating edges. In stead of creating an edge by connecting two nodes randomly, the algorithm carefully checks for each suggested edge if there is any network satisfying the given degree sequence and size that has the suggested edge in it. In case there isn't this means that creating the edge would fail the building of a sample network. More specifically, fixing one node and going through all possible other nodes to create edges, the algorithm checks for each other possible node whether creating the edge would fail the building of the sample network. From all nodes that do not fail the build, one is chosen at random[15] and hence for each step the probability is stored. This way, when the network is fully build the final probability of obtaining that specific network can be calculated.

[15] The node is chosen not uniformly on the possible nodes, but uniformly on the degrees of all possible nodes.

After obtaining the 100.000 sample networks I calculated averages by means of West's Algorithm and logarithmic sums. Using logarithms sums was done to prevent round-offs errors due to working with number of very different order of magnitude. Notice that the power-law data has both very small numbers and large numbers.

Appendix 2

The Figures 14 and 15 show the cumulative distribution of the number of likes tracks obtained (14) and the number of Witch House tracks users like (15), plotted on a log-log scale. Both variables appear to follow a power-law with cutoff, with exponent 1.9 and 2.97 respectively. The exponents are estimated using the framework and code developed by Clauset & Newmann [7], and it cannot be rejected they come from the fitted distribution on a significance level of 1%.

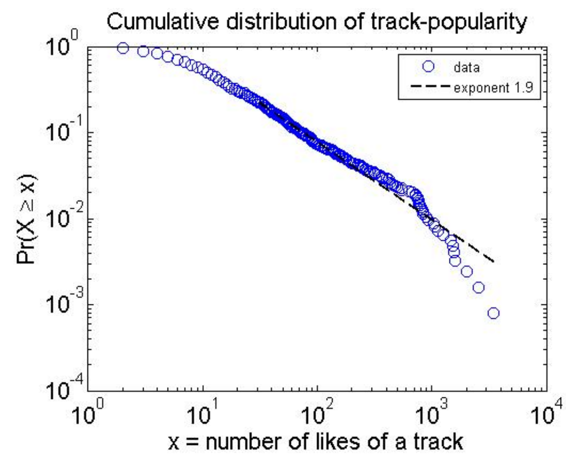


Figure 14. *This figures shows the cumulative distribution for 'likes' that tracks received (blue). It is plotted on a log-log scale. The data seems to be consistent with a power law with cut-off and exponent 1.9*

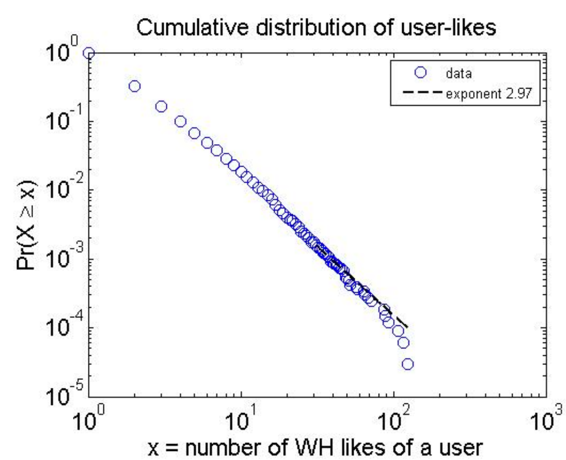
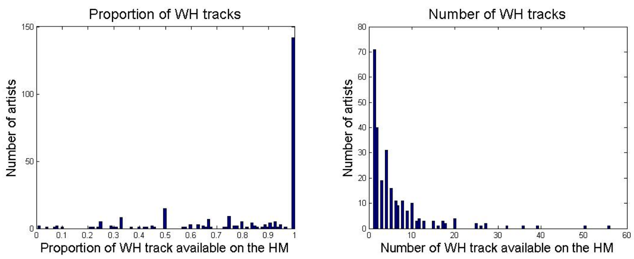


Figure 15. *This figures shows the cumulative distribution for 'likes' that tracks received (blue). It is plotted on a log-log scale. The data seems to be consistent with a power law with cut-off and exponent 1.9*

Appendix 3

This appendix shows some extra statistics of the data on artists. Firstly, Figure 16a shows a histogram of the number of Witch House tracks released as a proportion of their total track available on the Hypemachine, with the number of artists on the y-axis. Figure 16b shows a histogram of the number of of Witch House tracks artist released, with the number of artists on the y-axis.



(a) ***(b)***

Figure 16

Figure 17 shows that, on average, the more likes a track obtained the higher its percentage of 'one-likers'. One-likers are here defined as users that solely like one track in the Witch House genre.

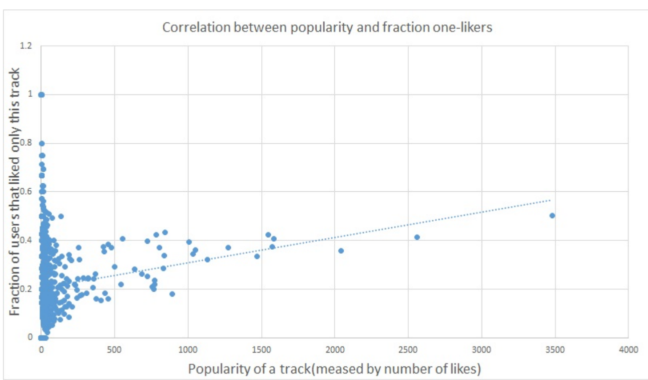


Figure 17. *This figure (scatterplot) shows the relation between popularity and the fraction of users that liked only this track of the track total users. Each dot corresponds to one track and the x-axis and y-axis give the value of its total likes and fraction 'one-likers' likes respectively.*