Erasmus University Rotterdam
M.A. in Cultural Economics and Entrepreneurship

Are We Watching More Diverse Movies in the Digital Age? A Quantitative Analysis on the Consumed Diversity on MovieLens

Huipeng Xu 575620hx

Master Thesis

Supervised by

Prof. Christian Handke

June, 2021

ABSTRACT

Changes in how users rate movies on MovieLens is a snapshot of an individual's lifetime movie consumption. This thesis studies the diversity of individual movie consumption by making use of the datasets about movies and users from MovieLens, a movie review website applying recommendation algorithms. Based on the Stirling Model and the Intra-List metric, a dissimilarity score is used as the indicator of diversity to measure the similarity between any two movies. The results show that users tend to watch less diverse movies, whether on an aggregate level or an individual level, even though we thought digitalization should have promoted a more diverse consumption. A decreasing trend in the overall diversity of individual movie consumption since 2017 is statistically identified with an interrupted time series analysis, which mainly results from a less diverse supply of new movies. Moreover, an individual's movie consumption is observed to be less diverse as experience grows due to a stronger preference for personalized movies and new releases. The methods and findings in this study could help players in the movie sector enhance product differentiation and evaluate the effectiveness of a cultural law or policy in promoting diversity.

Keywords: movie industry, cultural diversity, MovieLens, time series analysis, digitalization

Table of Contents

Acknowledgements

This master thesis is the curtain-raiser to my year-long academic journey. It was more meaningful to me than the one I did ten years ago at undergraduate level, as it opened my mind to both economics and programming. I would first like to thank my supervisor, Christian, for listening patiently to my ideas and building on them with enlightening advice on research methods and result interpretation. I would like to thank my husband, who taught me many programming skills during his busy work and PhD research. Also, I would like to thank my parents who, despite being far away in Shanghai, have been most supportive during the pandemic. This thesis is also a symbol of my love for film art.

Chapter 1. Introduction

Movies are one of the mainstream cultural products with both commercial and social value. Commercially, movie companies rely on producing and distributing movies that capture a wide audience, thereby realizing corporate profits. Socially speaking, movies enrich the entertainment life and the spiritual world of the general public. Meanwhile, through public intervention, government agencies and public institutions ensure that all citizens have access to a variety of movies. Moreover, both private and public sectors should not only focus on providing high-return, high-quality movies; increasing the diversity of movies available to everyone has become another important objective.

In today's digital age, consumer movie-watching behavior has changed greatly. They get used to searching movies with the help of various technologies and viewing movies on multiple medium. Whether the impact of digitalization has made individual movie consumption more diverse or accelerated the formation of each person's "filter bubble" is controversial (Hendrickx, 2018; Möller et al., 2018; Nguyen et al., 2014; Vrijenhoek et al., 2021). Learning diversity in movie consumption helps us solve this puzzle. If we know more about what factors affect the diversity in individual movie consumption, this could help corporates in the movie industry adjust product differentiation for a better commercial and social performance and help government educate the general public to respect different culture and engage in more diverse cultural expressions.

This thesis aims to study how the diversity of individual movie consumption has changed by making use of the datasets from MovieLens. The diversity in this paper

follows the Stirling model, while creatively converting it into measurable dissimilarity score with the ILS (Intra-List Similarity) metric. The paper proceeds as follows. Chapter 2 summarized the related literature on cultural demand and cultural diversity and presented hypotheses. Chapter 3 introduced the datasets collected from MovieLens and methods of measuring diversity. In Chapter 4 and 5, I estimated whether there is a consistent trend in the overall diversity of individual movie consumption with time series analysis and explained why individuals watch movies with such a trend. In Chapter 6 and 7, I analyzed whether the diversity of individual movie consumption correlates to the length of usage time on MovieLens and discussed what affect the consumed diversity with the increase of experience. Chapter 8 came to conclusions.

Chapter 2. Theoretical framework

2.1 Key concepts in cultural economics

Characteristics of cultural products

Unlike ordinary commodities, cultural goods and services have a number of special characteristics. Since cultural goods and services such as TV, radio and cultural heritage are non-rival and non-exclusive, interchangeably called public goods, and they are also considered merit goods (Musgrave, 1987), as each citizen benefits regardless of their willingness to pay so that this creates market failure. Due to this, governments need to correct such a market failure and maximize social welfare through taxes and subsidies. Experience goods is another characteristic of cultural goods and services (Towse, 2011), which makes the process of choosing a desirable cultural goods difficult, since consumers cannot have full information about the quality before consumption, and experience it once will exhaust the total future consumption. Suppliers thus communicate with consumers in various ways to signal quality information on products.

Digitalization refers to the use of digital technologies to create, change and improve business models, which has impacted the production, distribution and consumption of all manners of cultural goods and services (Bekar & Haswell, 2013). With the development of technology, digitalization has made many types of cultural products reproducible. Those cultural goods and services that were previously rival and excludable turned to quasi-public goods for Internet users (Handke et al., 2016). Since consumers can reach digital cultural products online, suppliers have much lower inventory costs than before. In addition, the prevalence of services applying big data

has also lowered the threshold for consumers to access a wider range of products. However, today's cultural sectors continue to exhibit radical uncertainty (Caves, 2000) and unpredictable changes in consumer demand due to the merit goods, public goods, and experience goods characteristics of cultural goods and services.

Cultural Diversity

Diversity refers to the richness of the typology in a system. It is employed in various disciplines, such as ecology (Odum, 1953;May, 1975; McCann, 2000), biology (Sugihara, 1982; Solow & Polasky, 1994), technology (Kauffman, 1992; Stirling, 2007), and sociology (Grabher & Stark, 1997; Johnson & Longmeyer, 1999). Scholars from various fields adopt the Stirling model (Stirling, 1998) to assess diversity.

The general definition of diversity is a combination of three properties: 'variety', 'balance' and 'disparity' (Stirling, 2007). Variety means the number of categories to which the system elements are apportioned, which corresponds to product variety in economics (Lancaster, 1979). The more categories in a system indicates a greater diversity. Balance means the evenness of apportionment of elements across categories. The more balanced the distribution of elements in each category, the greater the diversity. In economics, balance could refers to market concentration for suppliers (Finkelstein & Friedberg, 1967) and product differentiation for products (Lancaster, 1990). Disparity means the way and degree in which the elements are distinguished in the system. In economic terms, it could be heterogeneity (Kirman, 2006). The more different the elements are from each other, the greater the diversity. The Stirling model

not only unifies methods for measuring cultural diversity, but also brings clarity to what has been a vague concept of cultural diversity (Bonet & Négrier, 2011).

Cultural diversity is currently an increasingly important public value in cultural sectors (Christiansen, 2012). The concept of cultural diversity was first time popularized worldwide as a necessary objective for the long-term survival of humanity by UNESCO (Stenou, 2004). In general, it helps us to understand different perspectives in the world we live in and thus to eliminate negative stereotypes and personal prejudices against different groups. Furthermore, it is essential for the functioning of a democratic society where everyone is given the right to form their own views from diverse sources of information (Council of Europe, 2000). Nevertheless, as discussed earlier, the characteristics of cultural goods and services make it difficult for a free market to provide equal access to a wide variety of arts and culture. Therefore, governments need to intervene through laws and policies to create a culturally diverse environment. Measuring diversity with comprehensive data and scientific methods can better evaluate the effectiveness of a cultural law or policy in promoting diversity.

2.2 Cultural consumption in the movie industry

Feature film, shortened as movie, is a type of motion picture with 40 minutes or more in length for entertaining purposes (MPAA, 2020b). As an art form and cultural goods, movies have been developing for over a century. With the development of digital technology, the movie industry has changed in all aspects, including production, distribution and consumption. Today, movies are not only in cinemas, but also direct to home video and streaming services anywhere in the world. In 2019, the global movie

market including theatrical and home entertainment surpassed 100 billion dollars in revenue for the first time in history (MPAA, 2020a). The movie industry is still one of the most lucrative entertainment industries.

The movie industry is characterized by high fixed costs and economies of scale (Bourreau et al., 2002). As Caves stated (2000), art must meet commerce. Movie making is commonly going to be in a difficult business dilemma to make a profit. It requires a large capital investment, and its sunk costs only begin to be recovered months or even years before a movie is released (Towse, 2019). This entails that a movie could be profitable only when it has a market with a large enough consumer base and strong purchasing power to offset the high costs (Vogel, 1998; Wasko, 2003). Thus, economics of scale where the marginal cost per movie should be minimized as production increases is importance to firms. Nevertheless, nobody knows whether a movie will succeed or not commercially (De Vany, 2006). While many firms make blockbusters that cater to mass markets by investing heavily and casting superstars, this still does not guarantee that those movies will be a market hit (Chisholm, 2005; De Vany & Walls, 1999; Simonoff & Sparrow, 2000), as demand for specific movies is unpredictable. Neither consumers know whether they would enjoy a purchase before consumption (Nelson, 1970), nor suppliers could predict how many a certain type of cultural products they would sell (Caves, 2000).

The standard demand theory basically illustrates what influence consumers purchase one product per certain price level. First, consumer demand depends on the relative prices of goods and services. If the price is higher, fewer products are supposed

to be sold. Moreover, if a product has low price elasticity, reducing price will not increase much demand. Conversely, a product with high price elasticity will have a dramatic change in demand due to price changes. Second, the income of consumers affects consumer choice. A high-income consumer has more money to allocate on daily consumption. Third, the prices of substitutes affect consumer choice, as consumers would prefer alternative products when they are offered with a lower price. Fourth, consumers compare the opportunity costs of different choices. In other words, you gain some benefits from buying a product while you lose other benefits that could have enjoyed. Rational consumers make choice where the marginal benefits exceed the marginal costs, which maximizes their marginal utility. For instance, when the marginal utility of going to a cinema, where the marginal utility includes the money paid for a movie ticket, the time and travel expenses to the cinema, and the satisfaction of enjoying a movie, is already no more than the marginal utility of watching a DVD at home, they would give up the former choice.

Consumer choice is also affected by search costs. The more information consumers have, the more likely they are to be satisfied with their purchases. However, consumers cannot have full information about the quality of a cultural product before consumption because of its experience goods attribute. Thus, consumers try to reduce the search costs of cultural consumption in various ways. First, consumers observe the choices of others. As others' purchases are considered signals of good quality, demand for a cultural product may increase with the number of its consumers (Nelson, 1970). Second, consumers refer to expert opinion. Experts provide information on the quality of

products and services based on their professional knowledge and expertise (Tobias, 2004). Their opinion is especially valued in cultural sectors, as the quality of artistic production is difficult to evaluate. Third, individual experience of consuming a specific cultural product is by far the most reliable way to assess the utility of future cultural consumption. Consumers learn by continuous consuming to make better future choice (Lévy-Garboua & Montmarquette, 2002). When consumers invest money and time in accumulating knowledge on arts and culture, they build up 'consumption capital' (Towse, 2019).

Nevertheless, the lack of information on product quality among consumers persist. This leads to a special economic characteristic of demand in cultural sectors: uncertainty. The uncertainty of consumer demand for movies has led a group of economic scholars to explore consumers. Most of studies try to summarize the determinants of economic success of movies by analyzing the statistics of consumer spending on movie. Movie traits, such as genres and lineup, and marketing variables, such as pricing, timing of release and advertising costs, are two basic factors to box office performance (Hennig-Thurau et al., 2001). Besides, the behavior of audience and exhibitors are found to affects box office more directly (De Vany, 2000; Elberse & Eliashberg, 2003). For example, the numbers of screen arranged by exhibitors could contribute to the success of a movie in short term while the word-of-mouth of movies plays a more important role in the long-term success.

In economic terms, word-of-mouth is a common manifestation of information cascades, which refers to the phenomenon that individual decision is influenced by the

action of others regardless of his or her own information (Bikhchandani et al., 1992). Empirical research on information cascades found that the mass dissemination of box office performance such as evening news or rankings correlates with the increasing market share of movies (De Vany, 2000). With the advent of digitalization, consumers are able to watch movies online and further engage on movie review websites and applications where experts predict box office and users share ratings and reviews. These platforms are considered not only an electronic form of word-of-mouth (eWOM) but also an effective approach to promoting movies (Zufryden, 2000). A group of scholars used data on some movie review platforms to analyze how eWOM affects other consumers' choice and the sales performance of movies (Baek et al., 2014; Chiu et al., 2019; Duan et al., 2008; Liu, 2006; Moul, 2007).

Movie review platforms are also valuable source of data to track users' movie consumption. These data should be valued and applied. Still, less literature made use of those data to study cultural diversity in the movie industry. This is probably because the reliability and authenticity of the data from movie review platforms is still in doubt. Most scholars assess diversity of movies in different countries by using cinema statistics collected from relevant national or international institutions (Benhamou & Peltier, 2010; Moreau & Peltier, 2004). However, as more consumers do not watch movies through cinemas and DVDs, the data from official statistics departments are becoming less representative of the overall consumers.

2.3 Determinants of consumed diversity

Cultural diversity can be further distinguished between 'supplied diversity' and 'consumed diversity' (Eaton & Lipsey, 1989; Van Der Wurff & Van Cuilenburg, 2001), which are the diversity of what is made available and the diversity of what is actually consumed. In particular, consumed diversity depends on consumer tastes and supplied diversity (Napoli, 1999; Ranaivoson, 2007).

Supplied diversity

Consumed diversity relies on a diverse supply. Supplied diversity is mainly influenced by (1) degree of market competition, (2) costs of production and distribution, (3) selection by intermediaries, (4) digitalization, and (5) globalization.

Market competition affects supplied diversity. According to the theory of perfect competition, all firms sell identical products with no product differentiation. However, perfect competition exists in theory. Real world markets are monopolistic competitive, a mix of monopoly and perfect competition, where firms are price makers and they differentiate products to distinguish with competitors selling similar products (Chamberlin, 1933; Hotelling, 1929). In competitive markets with lower market concentration, firms are assumed to have stronger incentives to innovate, and more product innovation could lead to a more diverse range of products in the market. In oligopolistic markets where the majority of the market share is divided among a few firms, these firms have significant resources available for product innovation, which could likewise promote supplied diversity.

Costs affects supplied diversity, as it is a key factor for firms to adjust product differentiation in order to distinguish their own products from similar offerings on the market. Unlike a firm which only sells one product, firms using product differentiation decide how many different categories of products they should make when they could minimize the average product costs to reach economics of scope (Dixit & Stiglitz, 1977; Lancaster, 1979). It is apparently impossible for firms to provide a full collection of variants to satisfy consumers, because they have to balance the revenue gained from a more diverse supply and the less production costs resulted from a less diverse supply (Lancaster, 1990).

Intermediaries filter out a number of products before consumers have access to the selected ones, although they do play a very important role in gatekeeping. In cultural industries, due to the oversupply problems, the experience goods attributes, and the limited resources available to consumers, markets need filtering systems to control the number of products reaching the market (Peltoniemi, 2015). This more or less has lowered the diversity of goods and services available. Furthermore, gatekeepers are found to make biased decisions, which could impede supplied diversity. For instance, the creative works by female are naturally disadvantageous in terms of sales performance (Goldin & Rouse, 2000; Bocart et al., 2017).

Digitalization has an impact on supplied diversity, especially in cultural industries. Reduced costs due to digitization should promote supplied diversity. Production costs are lower than in the past. Firms now spend less on inventory costs, as many cultural goods and services are basically provided in digital form. Also, distribution costs have

been reduced as well. Consumers now have easier access to various cultural products than in the past, as they can freely search them online. Particularly, personalized online services are found to further lower consumers' search costs and still preserve the consumed diversity. Some empirical research on the news industry supports that the news selected by recommender systems can be more diverse than by human editors. By comparing the diversity of news content from editors, personalized and non-personalized recommender systems, some scholars found that personalized recommender systems actually yielded the highest topic diversity (Möller et al., 2018), and some other scholars are working to propose recommendation algorithms that enable higher diversity (Vrijenhoek et al., 2021). On the other hand, digitalization is believed to accelerate the snowball effect of super-popular products taking the lion's share of the market in terms of sales performance. Some research results show that the use of recommender systems have increased the sales concentration in movie industries (Fleder & Hosanagar, 2009; Wu et al., 2011).

Globalization has an impact on supplied diversity. People can access a vast array of goods and services from different countries and continents more easily than ever before. Such border-breaking exchange and communication allows people to have multi-cultural identities (Sotshangane, 2002). From this perspective, local products are influenced by foreign cultures and becomes more diverse. But from the perspective of cultural imperialism, content creation becomes more homogeneous globally, as local cultures move toward uniformization for a number of historical and economic reasons (Palmer, 2004).

Diverse tastes and taste for diversity

In cultural sectors, there is an extraordinary variety of goods and services available in the market, as diversity is valued by consumers for two reasons (Ranaivoson, 2012): (1) diverse tastes among consumers, and (2) consumers' taste for diversity.

Consumers have diverse tastes, especially in arts and culture. Hotelling (1929) implicitly suggested the diversity of consumer tastes and the positive externality of product diversity. To help consumers find a better match for their tastes, firms supply diverse goods and services. In economic terms, diversity of supply relates to product differentiation. Hotelling's model explains that firms produce segmented products based on their positioning to consumers, so that consumers are more satisfied. The process of forming tastes and preferences is cumulative (Blaug, 2001). This also suggests diverse tastes among consumers. With each new personal experience added, people gradually develop their personalized journey of consumption, which determines their unique tastes.

Consumers as a whole have a taste for diversity in cultural consumption. This could be explained by the "representative consumer" model (Dixit & Stiglitz, 1977; Spence, 1976) that consumers prefer to purchase different goods and services that are close substitutes rather than the same one. There are three main reasons for this assumption. First, taste for diversity relates to the law of diminishing marginal utility (Marshall, 1961). As consumers get more and more same products, they gain less satisfaction from every extra unit of the product, whereby the marginal utility of this product decline with the number consumed. This is especially true for movie viewers. Let's say that the

utility gained from watching a movie for the second time is normally less than from the first experience. Second, consumer tastes are changing over time through constantly accumulating experience of different cultural products (Blaug, 2001). The experience of watching a movie would subtly influence consumers' future choice of movies, as explained above for the process of 'learning by consuming'. For instance, a consumer has watched a movie and enjoyed it. He or she will choose the next movie to watch based on the elements he or she likes in that movie. The next movie is similar to the previous one, but also have some variant elements. Third, people are not only satisfied with consuming known goods and services available in the market, but also search for new or strange things (Scitovsky, 1976). This entails that consumers value diversity in terms of the balance between novelty and familiarity.

2.4 Hypotheses

Individual movie consumption has been undergoing a transformation in the past decade because of the Internet and digitalization. Digitalization has lowered the barriers for firms and creators to produce and distribute movies owing to lower costs, whereby more novel movies and more less mainstream movies could be supplied in the market. Additionally, as the long tail theory (Anderson & Andersson, 2004) suggests, the sales of all types of movies will be distributed in a more balanced way, so that the market share of superstar sellers would decrease conversely to that of niche sellers, just as some empirical research in other industries indicate (Benhamou & Peltier, 2007; Brynjolfsson et al., 2011). A more diverse supply of movies is thus expected.

Meanwhile, digitalization has lowered the price of watching any type of movies. Today's digital media services do not need spend large inventory costs storing massive media assets. If a consumer wants to watch a movie produced 50 years ago or a movie shot by an unknown director, all they need is to type the key words to search on Netflix or YouTube. Consumers thus have easier access to niche movies. Furthermore, online personalized services applying big data reduce consumers' search costs. Those services can predict movies that match consumers' unique tastes and preferences, even though consumers themselves don't exactly know what they like. A more diverse movie consumption is thus expected.

H1. Consumers overall watch more diverse movies recently than in the past.

In the movie industry, consumed diversity corresponds to the diversity of movies actually watched by individual consumers. As consumed diversity is determined by consumer tastes and supplied diversity, the above conditions should lead to an increasingly diverse movie consumption in general. Nevertheless, as the impacts of digitalization on consumers are intertwined with each other simultaneously, what has to be admitted is that it is difficult to measure its impact on individual movie consumption separately. Moreover, as digitalization has been taking effect progressively, it is also difficult to hypothesize a definite point of time when consumers' movie watching behavior has changed radically. It is thus hypothesized, in a slightly vague way, that with the increasing popularity of digitalization, the movies people watch should be more diverse year by year.

H2. On an individual level, the diversity in movies consumption slowly increases with experience

People have a taste for diversity of movies for several reasons as described previously. As age grows, people accumulate more experience of movie watching. Every new experience added, individual taste is refreshed accordingly. Every future choice of movies depends on their current tastes in movies. Thus, people tend to watch more diverse movies because of their changing tastes. Moreover, as people are being clearer about what they like over time, their tastes and preferences have been stabilized. People are less likely to explore movies that deviate from their previous experience, but to make small jumps from one movie to the next one. Therefore, it is hypothesized that the speed of increase of the diversity in individual movie consumption should be rather low.

Chapter 3. Data collection and methods

3.1 About the dataset

I use secondary data from MovieLens[1] — a website that asks its users to give movie ratings in order to improve personalized movie recommendations. The dataset on MovieLens (Maxwell & Konstan, 2015) is called 25M Dataset[2], which includes two parts: (1) ratings of movies by users with timestamps, and (2) a tagging data structure that contains tags to describe movies and tag relevance values to movies.

Rating

The user ratings are from the 25M Dataset. The ratings on MovieLens are expressed as a "half-star" value system, which is a standard user interface for users to input preferences (Maxwell & Konstan, 2015). The range of preference values is from 0.5 to 5.0 stars. The 25M Dataset contains the rating history of 162,541 users between January 9th, 1995 and November 21st, 2019. Those users were selected by MovieLens at random for inclusion. All selected users had rated at least 20 movies. There are totally 25,000,095 ratings across 62,423 movies. The information about movies and user rating history in the 25M Dataset is presented in the form of Table 3.1 and 3.2.

[1] MovieLens: www.movielens.org

[2] https://grouplens.org/datasets/movielens/25m/

MovieId	*Title*	*Genres*	*Released year*
1	Toy Story	Adventure|Animation|Children|Comedy|Fantasy	1995
2	Jumanji	Adventure|Children|Fantasy	1995
3	Grumpier Old Men	Comedy|Romance	1995
4	Waiting to Exhale	Comedy|Drama|Romance	1995
5	Father of the Bride PartII	Comedy	1995
...	...	...	...
209157	We	Drama	2018
209159	Window of the Soul	Documentary	2001
209163	Bad Poems	Comedy|Drama	2018
209169	A Girl Thing	(no genres listed)	2001
209171	Women of Devil's Island	Action|Adventure|Drama	1962

62011 rows × 4 columns

Table 3.1 Movie information (movieid, title, genres) in 25M Dataset[3]

UserId	*MovieId*	*Rating*	*Timestamp*	*Title*
1	296	5	2006-05-17 15:34:04	Pulp Fiction (1994)
1	306	3.5	2006-05-17 12:26:57	Three Colors: Red (1994)
1	307	5	2006-05-17 12:27:08	Three Colors: Blue (1993)
1	665	5	2006-05-17 15:13:40	Underground (1995)
1	899	3.5	2006-05-17 12:21:50	Singin' in the Rain (1952)
...	...	...	...	...
162541	50872	4.5	2009-04-28 21:16:12	Ratatouille (2007)
162541	55768	2.5	2009-04-28 20:53:18	Bee Movie (2007)
162541	56176	2	2009-04-28 20:31:37	Alvin and the Chipmunks (2007)
162541	58559	4	2009-04-28 21:17:14	Dark Knight, The (2008)
162541	63876	5	2009-04-28 21:01:55	Milk (2008)

25000095 rows × 8 columns

Table 3.2 Rating history (userid, movieid, rating, timestamp) in 25M Dataset

[3] "movieIds" are not numbered in numerical order.

Tagging

The tagging data structure is a sub-dataset called Tag Genome[4] from the 25M Dataset. To be more specific, it is a dense matrix: each movie has a relevance value for each tag in the genome. The tag relevance value represents the relevance of a tag to a movie on a continuous scale from 0 to 1[5]. It indicates how strongly a movie exhibits a particular property. Simply speaking, if a tag can well describe a movie, the relevance value of this tag to this movie is high, which is close to 1. The Tag Genome includes tag relevance values provided for 13,816 movies and 1,128 tags. The tag information in Tag Genome is shown like Table 3.3.

Tag Genome is a collective work by the users on MovieLens and the MovieLens team. Users on MovieLens are allowed to apply new tags to movies, that are words or short phrases. The MovieLens team computed relevance values through a machine learning algorithm and has normalized various factors[6] that affect the relevance between movies and tags (Vig & Riedl, 2012). According to my communication with the MovieLens team, they only run the algorithm on a selection of users, movies and movies tags. This makes sense from both a logistics and an accuracy standpoint, but this has limited me to calculating the dissimilarity score in the next section to only 13,816 movie included in Tag Genome.

---

[4] https://grouplens.org/datasets/movielens/tag-genome/

[5] It should be noted that each movie has a relevance value greater than 0 and less than 1 with respect to all 1128 tags. Even if a movie is extremely irrelevant to a certain tag, this value is approximately zero but not equal to zero.

[6] The factors to predict tag relevance include tag names, tag counts, tag share, similarity between tags, text of reviews on IMDB.com, text frequency, similarity between text, average ratings, similarity between ratings, etc.

tagId	tag
1	7
2	007 (series)
3	18th century
4	1920s
5	1930s
...	...
1124	writing
1125	wuxia
1126	wwii
1127	zombie
1128	zombies

1128 rows × 2 columns

movieId	tagId	relevance
1	1	0.02875
1	2	0.02375
1	3	0.0625
1	4	0.07575
1	5	0.14075
...	...	...
206499	1124	0.11
206499	1125	0.0485
206499	1126	0.01325
206499	1127	0.14025
206499	1128	0.0335

15584448 rows × 3 columns

Table 3.3 tag ID and relevance score in Tag Genome

3.2 Dissimilarity score (D-score)

Measurement of diversity

In this paper, the diversity specifically refers to the consumed diversity on MovieLens. In other words, it is about the diversity of the movies actually rated by individual users on MovieLens. According to the Stirling's model, diversity has three dimensions, namely 'variety', 'balance', and 'disparity' (Stirling, 2007). In the case of MovieLens, variety represents the number of movies that a user has watched. Balance represents

how those movies are evenly applied to different tags. Disparity represents to what extent those movies are different from one another based on their relevance to each tag.

Take *The Godfather (1972) and Interstellar (2014)* as examples. On MovieLens, *The Godfather (1972)* is tagged by 503 words and phrases, led by the most relevant tags such as ‘mafia’, ‘classic’, ‘Al Pacino’, ‘great acting’; *Interstellar (2014)* has 102 tags led by ‘confusing’, ‘Christopher Nolan’, ‘mindfuck’, ‘time travel’. If I measure the diversity of these two movies, there are three methods to measure diversity: (1) The variety corresponds to two movies in total; (2) The balance is more complex. The share of the tag application of ‘masterpiece’ among two movies is 100%, and the share for ‘Al Pacino’ is 50%, etc. The lower the mean of all the shares of tag application, the more even the distribution of tag application; (3) The disparity is about how many common tags can sufficiently apply to these two movies. The less common tags, the more disperse the two movies are.

Calculation of D-score

Nevertheless, the Stirling Model does not explicitly define to what extent the distribution is considered even. It does not help define to what extent tags are sufficiently applicable to a movie either, so that the disparity cannot be quantified with the Stirling Model. Since the Stirling’s Model does not provide a mathematical method to precisely measure diversity, I introduce the Intra-List Similarity (ILS) metric (Ziegler et al., 2005; Ekstrand et al., 2014; Nguyen et al., 2014), which is commonly used for a proxy as diversity in computer science. This ILS metric is calculated in a way that nicely consider all three diversity properties of the Stirling model. It is established in

this paper to measure to what extent two movies are different in terms of their relevance for all 1128 tags in the Tag Genome dataset. The measurable unit in the ILS metric is called ‘dissimilarity score’, abbreviated as ‘D-score’.

According Tag Genome, even though a movie is extremely irrelevant to a certain tag, the relevance value is approximately zero but not equal to zero, which is determined by a machine learning algorithm developed by MovieLens (Vig & Riedl, 2012). Thus, the relevance values of each movie to all the 1128 tags are available. According to the ILS metric, the similarity between two movies can be measured by measuring the Euclidean distance between two movie vectors. Therefore, the relevance value of any tag to any movie can be expressed as a vector, for example, *rel* (t_x, m_y). As shown in Figure 3.1, the first tag t_1’s relevance value to movie *i* (m_i) is subtracted from t_1’s relevance value to movie *j* (m_j) so that the difference is obtained. Then the differenced relevance value for the other tags are obtained until all tags are exhausted. Then the sum of squared difference values can be calculated. The square root of the sum is the D-score between m_j and m_j. The higher D-score indicates greater diversity.

$$d_{(m_i,m_j)} = \sqrt{\sum_{k=1}^{m}[rel(t_k, m_i) - rel(t_k, m_j)]^2}$$

Figure 3.1 The function of the Euclidean distance between two movie vectors (Nguyen et al., 2014)

To facilitate the calculation of D-score in the following sections, I have written two D-score generators with Python. The detailed code of the D-score generators can be

found in Appendix A. The first generator is used to calculate a D-score given any two movies, so that the score can measure how different the two movies are. Through this generator, I am able to produce D-scores for all the unique pair-wise movies among 13,816 movies from Tag Genome. The second generator is used to calculate a D-score given a list of movies, so that this D-score can measure how different the movies in this list are from each other. The second generator outputs the median score of all the D-scores of pair-wise movies from a list to reflect the degree of diversity of this movie list. If I randomly choose 100 movies[7] on MovieLens, the calculated D-scores each time shows a normal distribution. Since a set of values that exhibits a normal distribution can use the median to denote the average level of its values, a median D-score can reflect the degree of diversity of a list of movies.

Validity of D-score

If two movies are similar in terms of thematic content, directorial style, genre, storyline, etc., then this is in line with what we intuitively believe to be similar. Are the results of D-score calculation consistent with our intuition of the similarity of two movies? I test D-score with a number of genre movies and a number of art films that are hard to categorize by typical genres. I select 12 movies each belongs to 12 genres. Those genres are used by IMDB.com to categorize movies. Additionally, I select 10 movies among the best 25 arthouse movies picked by The Guardian (TheGuardian, 2010). For each

[7] A list of 100 movies has 4950 unique pairs of two movies in total, according to the Combination Formula: C (100, 2) = 100! / [2! (100−2)!] = 4950). Thus, there are 4950 pair-wise D-scores from a list of 100 movies.

movie *x*, I compare it with all the other movies on MovieLens by calculating the D-score of *x* and any other movies. Then I pick the movie that has the lowest D-score with *x*. This movie is thus the most similar movie to *x*.

As shown in Table 3.4 and 3.5, the results produced by the D-score generator conform to our intuition. Through the test, the results indicate that (1) the D-scores between a pair of movie series, for example, The Godfather and The Godfather II, are comparatively low, which is around 2.3; (2) the most similar movie to *x* can be a movie directed by the same director of *x*, for example, Psycho (1960) and Strangers on a Train (1951) by Alfred Hitchcock; (3) If *x* is a genre movie, the generated most similar movie shares a same genre with *x*; (4) If *x* is an art film, the generated most similar movie shares a similar plot or background with *x*.

Table 3.4 D-scores between 12 Genre movies and their most similar movies

Genre	***Movie x***	***Most similar to x among the dataset***	***D-score***
Comedy	Detroit Rock City (1999)	Empire Records (1995)	3.12
SCI-FI	Back to the Future (1985)	Ghostbusters (1984)	4.68
Horror	Psycho (1960)	Strangers on a Train (1951)	3.97
Romance	Annie Hall (1977)	Manhattan (1979)	2.82
Action	Bourne Identity (2002)	Bourne Ultimatum (2007)	2.27
Thriller	Hereditary (2018)	Us (2019)	2.34
Drama	Frances Ha (2012)	Paterson (2016)	3.54
Mystery	The Invisible Guest (2016)	Bad Times at the El Royale (2018)	4.41
Crime	Zodiac (2007)	The Jinx: The Life and Deaths of Robert Durst (2015)	4.21
Animation	Toy Story (1995)	Monster Inc. (2001)	2.66
Adventure	Spirited Away (2001)	Howl's Moving Castle (2004) Hayao Miyazaki	3.33
Fantasy	Groundhog Day (1993)	Defending Your Life (1991)	3.78

Table 3.5 D-scores between 10 arthouse movies in Top 25 best arthouse films of all time and their most similar movies

Ranking	***Movie x***	***Most similar to x among the dataset***	***D-score***
1	Andrei Rublev (1969)	Offret - Sacraficatio (1986)	2.98
3	L'Atalante (1934)	The Earrings of Madame de... (1953)	3.02
5	Citizen Kane (1941)	The Gold Rush (1925)	3.96
7	Days of Heaven (1978)	Sunrise: A Song of Two Humans (1927)	3.79
9	The White Ribbon (2009)	Phantom Thread (2017)	4.47
11	Aguirre, der Zorn Gottes (1972)	Black Narcissus (1947)	4.19
13	The Conformist (1970)	L'avventura (1960)	4.20
15	The Godfather (1972)	The Godfather: Part II (1974)	2.53
17	There Will Be Blood (2007)	Once Upon a Time in Hollywood (2019)	4.05
19	The Rules of the Game (1939)	The Earrings of Madame de... (1953)	3.54

Chapter 4. Results and analysis for Hypothesis 1

4.1 D-scores in 2015 - 2019

To compare the consumed diversity on MovieLens in consecutive years, I sample 10,000 users on MovieLens (Sample Group A), abbreviated as "users", and use 2015 to 2019 as the years for calculating D-scores. 10,000 users[8] are randomly sampled among those who have rating history in 2017 and had rated at least 20 movies throughout their usage. Each user had rated a number of movies in 2017 but not necessarily in each month.

The monthly D-scores of those sampling users in 2017 are computed as follows: Firstly, I calculate the median D-score of a list of all the movies watched by each user in one certain month, for example January, through the second D-score generator. So, each user who has rating history in January would have a D-score[9]. Secondly, among those D-scores of all the users, I choose the median D-score as the diversity of all the users' movie consumption in January, interchangeably called as "the overall user D-score in January". The overall user D-score in January 2017 is 6.84. Thirdly, I got the overall user D-scores for other 11 months in 2017 in the same way. The overall user D-score for each month in 2017 is shown in Figure 4.1, showing a downward trend. The only increase of the D-score during 2017 occurred in May, from 6.78 to 6.83. The

[8] Considering that the whole 25M Dataset is too large for Python to process, 10,000 is an amount that can be computed normally.

[9] If a user only rated one movie in certain month, this user would be automatically disregarded by the D-score generator, as at least two movies can have a monthly D-score. Additionally, if a user rated a movie that was not included in Tag Genome, this movie would be automatically disregarded as well.

diversity of movies rated by MovieLens users in 2017 per month exhibits a downward trend.

Based on the original 10,000 sampled users, I calculate the monthly overall user D-scores for 2015, 2016, 2018 and 2019 in the same way. As shown in Figure 4.1, the mean D-score in 2015 and 2016 are similar to in 2017, but there is no obvious trend in the monthly overall user D-score in 2015 and in 2016. The scores in 2015 and 2016 both vary randomly around their mean during each year. In contrast, 2018 saw an obvious downward trend in the monthly overall user D-score. The visual inspection of Figure 4.1 shows that the D-scores were stable around 6.6 to 6.8 from January 2015 to January 2017, then dropped from 6.8 to 5.6 during January 2017 and June 2019. The decline of D-score after 2017 needs statistical confirmation.

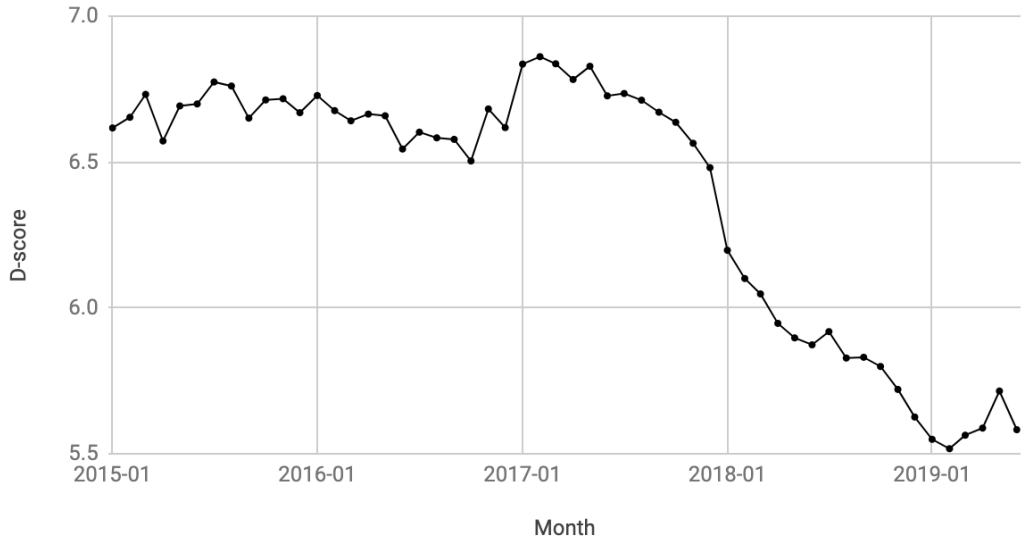


Figure 4.1 D-scores of the sampling MovieLens users in Jan 2015 - Jun 2019[10]

[10] Since the 25M Dataset was collected in November 2019, the data in 2019 is comparatively incomplete. Hence, I only use data for the first six months of 2019 in this paper.

4.2 Statistical analysis

A time series analysis for the entire series

A general time series analysis is conducted to statistically identify a downward trend over the period of 2015 to 2019. Three regression models, namely linear, quadratic and cubic regression, are estimated for the entire time series with 54 data points. As shown in Table 4.1, The R^2 of the linear regression is the smallest (0.676). The R^2s of the two polynomial regressions are larger, which means that the two polynomial models fit the observed data better. But it is hard to tell which one is better, as the quadratic and cubic model are virtually identical, as we see in Figure 4.2.

A T-test is used for each regression model to establish whether the effect of each coefficient on the dependent variable is significant. The test results of both linear and quadratic model, $p < 0.01$, are significant at the 0.01 level whereas the result of the cubic model, $p > 0.01$, is not significant. Thus, the cubic model should be rejected in the T-test. To establish whether a more complex regression model can better explain the population from which the data were sampled, I also use F-tests. The quadratic model has the largest F-statistic (195.4), followed by the cubic model (127.8) and the linear model (108.6). In addition, the probability values of the three F-tests are all exceedingly small number, which is smaller than α (0.01 level). Thus, there is much less than 1% chance that the F-statistic could have occurred by chance under the assumption of rejecting the three regression models. This confirms the quadratic regression model as the best fit line. To summarize, the entire time series is trending

upward in 2015, slowly decreasing from 2016 and decreasing at a high rate from 2017, as judged by the shape of the quadratic regression model plotted in Figure 4.2.

Table 4.1 A summary of the OLS regression results for the entire time series

	R^2	*coefficient*		*std err*	*t*	*P* > |t|	*F-statistic*
Linear	0.676	Intercept	7.02	0.07	100.08	0.000	108.6
		Slope(x)	-0.02	0.002	-10.42	0.000	Prob = 2.47e-14
Quadratic	0.885	Intercept	6.55	0.065	100.82	0.000	195.4
		x	0.03	0.005	5.06	0.000	Prob = 1.22e-24
		x^2	-0.0009	0.000	-9.60	0.000	
Cubic	0.885	Intercept	6.54	0.09	253.99	0.000	127.8
		x	0.03	0.01	-0.186	0.85	Prob = 1.91e-23
		x^2	-0.001	0.001	0.845	0.41	
		x^3	0.000	0.000	0.181	0.86	

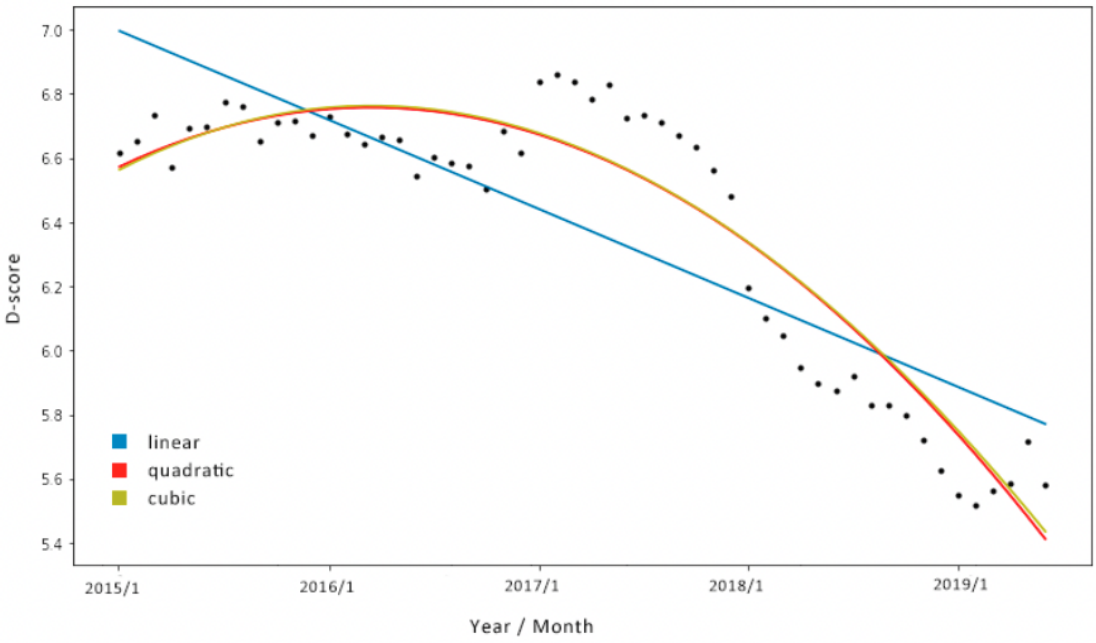


Figure 4.2 Visualization of the three regression models for the entire time series

Interrupted time series analysis

The visual inspection of Figure 4.2 indicates a consistent downward trend since January 2017, falling from 6.9 to 5.5, suggesting that an intervention event may have occurred in 2017. Possible external interventions could be that the impact of digitalization, such as recommendation systems and the other applications of big data, reached a tipping point in 2017, or that the supplied diversity of movies has become lower since 2017. Internally, the user patterns on MovieLens may have changed since 2017, or the dataset used to measure D-score is biased. These possible interventions will be discussed in the next chapter.

To determine whether changes before to after 2017 are statistically significant, I used the data point of January 2017 as a dividing point and further conduct an interrupted time series analysis, which is to assess the effect of intervention events on the observed behavior during a time series (Box et al., 2015). It is important to note that the intervention analysis conducted here is not a standard one, as I did not distinguish between prior period and post period by manipulating any independent variables beforehand but based on the apparent trend seen in the time series. I use the C statistic, an simplified intervention analysis developed by Tryon (1982). There are two reasons for employing the C statistic. First, the C statistic can be used on small data sets to evaluate the effects of an intervention (Tryon, 1982). Second, it can be applied no matter the previous phase contains a trend or not (Tryon, 1982). Table 4.2 presents the 5% and 1% critical values for testing the C statistic for samples of size 24 to ∞. The

observed data in both the prior period (January 2015 to January 2017) and the post period (February 2017 to June 2019), are displayed in Table 4.3.

Table 4.2 Critical values (Z) for testing the C statistic for selected sample sizes (N) at the 1% and 5% level of significance (Tryon, 1982; Young, 1941)

Sample sizes (N)	*1% Level of confidence (Z)*	*5% Level of confidence (Z)*
e24	2.27	1.64
25	2.27	1.64
∞	2.33	1.64

Table 4.3 The monthly D-scores of users on MovieLens (Jan 2015 - Jun 2019)

	Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
2015	6.62	6.65	6.73	6.57	6.69	6.70	6.77	6.76	6.65	6.71	6.72	6.67
2016	6.73	6.68	6.64	6.66	6.66	6.54	6.60	6.58	6.58	6.50	6.68	6.62
2017	6.84	6.86	6.84	6.78	6.83	6.73	6.74	6.71	6.67	6.64	6.56	6.48
2018	6.20	6.10	6.05	5.95	5.90	5.87	5.92	5.83	5.83	5.80	5.72	5.63
2019	5.55	5.52	5.56	5.59	5.71	5.58						

Although a visual inspection of Figure 4.2 does not suggest a consistent trend during the prior period, the resulting value of $Z = 1.84$, $p < 0.05$, for the prior period statistically indicates a trend. Therefore, detrending of the prior period is required to allow for an interrupted time series analysis based on the C statistic. Two methods of detrending are available for use. The first method is by differencing. The second one is by regression.

1. *Analysis based on first differenced time series*

Differencing is said to be the simplest method to remove a trend from a time series (Brownlee, 2017). Specifically, a new time series is created where each value is

calculated as the difference between the observed data of two consecutive time points in the original time series. Table 4.4a and 4.4b document the first differences of all the observed D-scores in the prior and post period, and also the C statistic of the two differenced time series. The resulting Z = -1.720, p > 0.05, for the prior period is not significant, so that the analysis based on first differenced time series is applicable. For the post period, the resulting Z = 1.903 is significant at 0.05 level but is not significant at 0.01 level. Then I go further by plotting the prior and post differenced time series to compare trends. As seen in Figure 4.3a, the differenced D-scores for the prior period vary randomly around zero, which means there is no significant trend as suggested by the C statistic. For the post period, most data points on the differenced time series are below zero (see Figure 4.3b), indicating the D-score was falling mostly during January 2017 and June 2019. This statistically confirms a significant downward trend in the monthly D-score in the post period.

Table 4.4a C-statistic for the prior period based on first difference

Year / Month	*D-score*	*Difference*	*Prior period*
2015/01	6.617	-	
2015/02	6.654	0.037	
2015/03	6.732	0.079	
2015/04	6.573	-0.160	
2015/05	6.693	0.120	
2015/06	6.700	0.007	
2015/07	6.775	0.075	
2015/08	6.761	-0.014	
2015/09	6.651	-0.110	

2015/10	6.713	0.062	
2015/11	6.717	0.004	C = -0.336
2015/12	6.670	-0.048	Sc = 0.196
2016/01	6.728	0.059	Z = -1.720
2016/02	6.677	-0.052	Not Significant
2016/03	6.642	-0.035	
2016/04	6.665	0.023	
2016/05	6.659	-0.006	
2016/06	6.545	-0.114	
2016/07	6.603	0.058	
2016/08	6.583	-0.019	
2016/09	6.578	-0.005	
2016/10	6.504	-0.074	
2016/11	6.682	0.178	
2016/12	6.618	-0.064	
2017/01	6.836	0.218	

Table 4.4b C-statistic for the post period based on first difference

Year / Month	***D-score***	***Difference***	***Post period***
2017/01	6.836	0.218	
2017/02	6.862	0.025	
2017/03	6.837	-0.024	
2017/04	6.783	-0.054	
2017/05	6.829	0.045	
2017/06	6.727	-0.101	
2017/07	6.735	0.008	
2017/08	6.712	-0.023	
2017/09	6.671	-0.041	
2017/10	6.637	-0.035	
2017/11	6.565	-0.072	
2017/12	6.481	-0.084	

2018/01	6.198	-0.283	C =	0.336
2018/02	6.101	-0.097	Sc =	0.176
2018/03	6.048	-0.053	Z =	1.903
2018/04	5.947	-0.101	Significant at 0.05	
2018/05	5.897	-0.050	Not significant at 0.01	
2018/06	5.873	-0.024		
2018/07	5.918	0.045		
2018/08	5.828	-0.091		
2018/09	5.831	0.003		
2018/10	5.799	-0.032		
2018/11	5.720	-0.079		
2018/12	5.626	-0.095		
2019/01	5.549	-0.077		
2019/02	5.517	-0.032		
2019/03	5.563	0.046		
2019/04	5.588	0.024		
2019/05	5.715	0.127		
2019/06	5.582	-0.133		

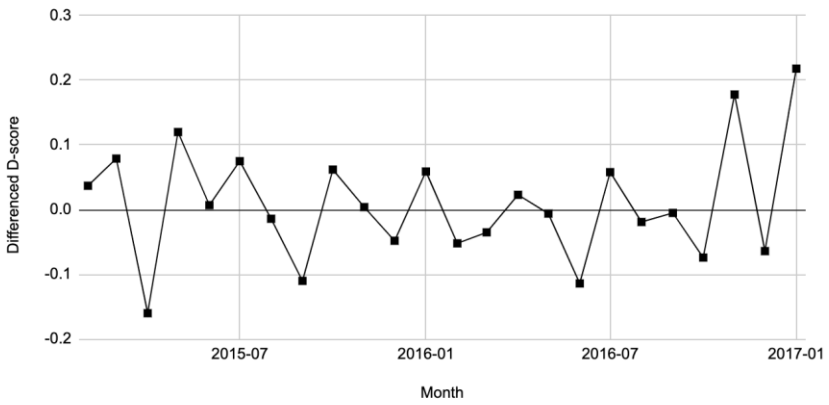


Figure 4.3a Visualization of the first differenced time series – prior period

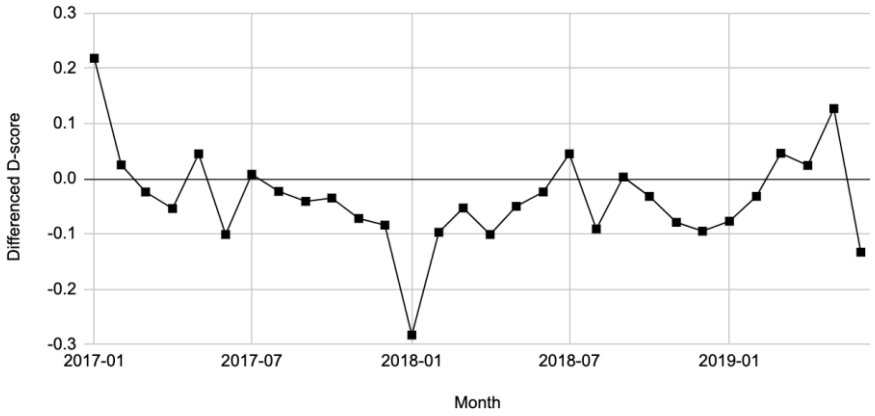


Figure 4.3b Visualization of the first differenced time series – post period

2. Analysis of residuals based on OLS regression

The second method is to create a comparison time series through standard regression techniques. A suitable regression model can be fit on the time index of the prior period to predict the observations in the post period. A comparison time series consists of residuals based on this regression model. Three regression models are estimated for the 25 observations in the prior period to quantify a trend. As shown in Table 4.5, the cubic regression model has the greatest R^2, followed by the quadratic and linear one in order. Also, I use a F-test to establish check whether the cubic model best fits the population from the sampled users. The cubic model has the largest F-statistic (5.826), followed by the linear model (1.188) and the quadratic model (0.588). In addition, the probability values of F-tests of the linear and quadratic model are both greater than α (0.01 level). The probability value of the cubic model is less than α (0.01 level), which means that there is much less than 1% chance that the F-statistic could have occurred by chance under the assumption of rejecting the cubic regression model. However, the plot in

Figure 4.4 suggests that the cubic regression may have over-fitting problem. If I choose the cubic regression model to predict values in 2019, the monthly D-score would be around 12, which is clearly far from the actual values. Thus, the linear regression model is optimal.

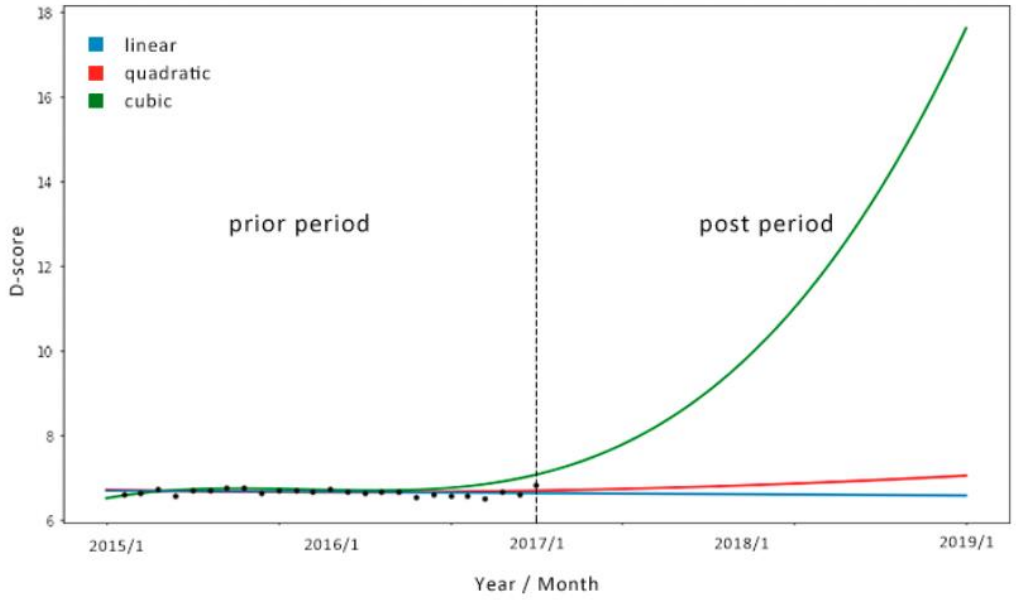


Figure 4.4 Plot of the three regression models

Table 4.5 A summary of the OLS regression results for the prior period

	R^2	coefficient		std err	t	P > |t|	F-statistic
Linear	0.049	Intercept	6.693	0.031	214.17	0.000	1.188
		Slope(x)	-0.0023	0.002	-1.090	0.287	Prob = 0.287
Quadratic	0.051	Intercept	6.713	0.050	122.28	0.000	0.588
		x	-0.0083	0.009	-0.888	0.384	Prob = 0.564
		x^2	0.0003	0.000	0.878	0.389	
Cubic	0.554	Intercept	6.514	0.055	119.13	0.000	9.095
		x	0.0723	0.017	4.204	0.000	Prob = 0.0004
		x^2	-0.0070	0.001	-4.797	0.001	
		x^3	0.0002	3.57e-05	5.064	0.000	

The comparison series is obtained by calculating residuals. Before calculating residuals for the post period, I take the predicted values for the prior period based on the above linear regression model and subtract them from the observed values in the prior period one by one. The Z value, 1.534, $p > 0.05$, for the residuals in the prior period is not significant (see Table 4.6a), which confirms that the residuals in the prior period vary randomly around their mean. This implies that the detrended prior period contains no significant trend. After detrending, the post period is eligible for testing with C statistics. Then I subtract the predicted values in the post period based on the linear regression model from the actual values in the post period in order. As shown in Table 4.6b, the Z resulting value, 5.578, $p < 0.01$, for the post period is significant, which means the observed data deviate significantly from the predicted data from the quadratic regression model. This result is consistent with the analysis of the first method. Again, there is a sharper decline in the monthly D-score during the post period in comparison to the prior period.

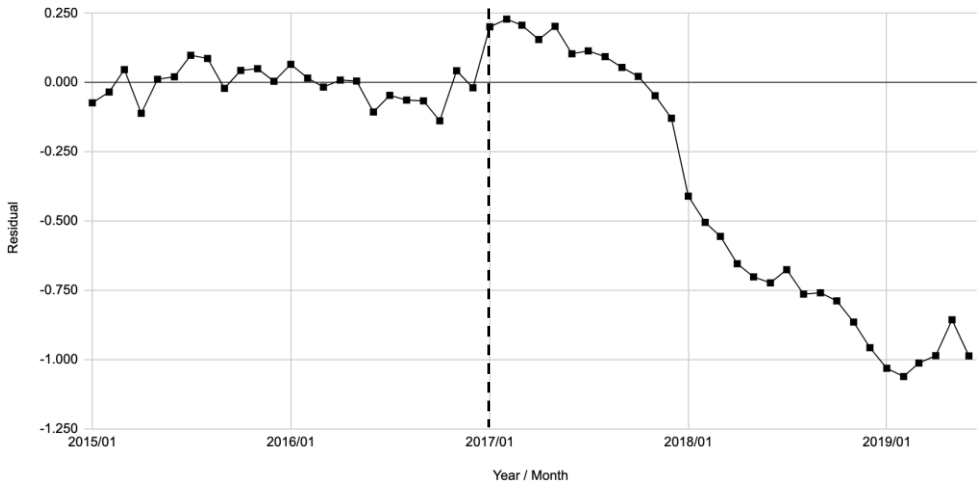


Figure 4.5 Visualization of the residuals for the linear regression

Table 4.6a C-statistic for the prior period based on the linear regression

Year / Month	*D-score*	*Model predicted D-score*	*Residual*	*Prior period*
2015/01	6.617	6.691	-0.074	
2015/02	6.654	6.688	-0.035	
2015/03	6.732	6.686	0.046	
2015/04	6.573	6.684	-0.111	
2015/05	6.693	6.682	0.011	
2015/06	6.700	6.679	0.020	
2015/07	6.775	6.677	0.098	
2015/08	6.761	6.675	0.086	
2015/09	6.651	6.672	-0.021	
2015/10	6.713	6.670	0.043	
2015/11	6.717	6.668	0.050	C = 0.295
2015/12	6.670	6.665	0.004	Sc = 0.192
2016/01	6.728	6.663	0.065	Z = 1.534
2016/02	6.677	6.661	0.016	Not Significant
2016/03	6.642	6.659	-0.017	
2016/04	6.665	6.656	0.009	
2016/05	6.659	6.654	0.005	
2016/06	6.545	6.652	-0.107	
2016/07	6.603	6.649	-0.047	
2016/08	6.583	6.647	-0.064	
2016/09	6.578	6.645	-0.067	
2016/10	6.504	6.642	-0.138	
2016/11	6.682	6.640	0.042	
2016/12	6.618	6.638	-0.019	
2017/01	6.836	6.636	0.201	

Table 4.6b C-statistic for the post period based on the linear regression

Year / Month	*D-score*	*Model predicted D-score*	*Residual*	*Post period*
2017/01	6.836	6.636	0.201	
2017/02	6.862	6.633	0.228	
2017/03	6.837	6.631	0.206	
2017/04	6.783	6.629	0.155	
2017/05	6.829	6.626	0.202	
2017/06	6.727	6.624	0.103	
2017/07	6.735	6.622	0.114	
2017/08	6.712	6.619	0.093	
2017/09	6.671	6.617	0.054	
2017/10	6.637	6.615	0.022	
2017/11	6.565	6.613	-0.048	C = 0.934
2017/12	6.481	6.610	-0.129	Sc = 0.176
2018/01	6.198	6.608	-0.410	Z = 5.578
2018/02	6.101	6.606	-0.505	Significant
2018/03	6.048	6.603	-0.555	
2018/04	5.947	6.601	-0.654	
2018/05	5.897	6.599	-0.702	
2018/06	5.873	6.596	-0.723	
2018/07	5.918	6.594	-0.676	
2018/08	5.828	6.592	-0.764	
2018/09	5.831	6.590	-0.759	
2018/10	5.799	6.587	-0.788	
2018/11	5.720	6.585	-0.865	
2018/12	5.626	6.583	-0.957	
2019/01	5.549	6.580	-1.031	
2019/02	5.517	6.578	-1.061	
2019/03	5.563	6.576	-1.013	
2019/04	5.588	6.573	-0.986	
2019/05	5.715	6.571	-0.856	
2019/06	5.582	6.569	-0.987	

4.3 Robustness check

It is important to clarify that the 10,000 research population were sampled among users having rating behavior in 2017. If the users sampled in 2017 are not representative of individual movie-watching behavior in recent years, then the findings in this chapter may be biased. To test whether the results are sensitive year on which the sample was based, another two groups of 10,000 users are sampled. Specifically, two groups of sampling users are selected among those who once rated movies on MovieLens in 2016 (Sample Group B) and 2015 (Sample Group C). The monthly D-scores from 2014 to 2019 for both the Sample Group B and C are calculated. The detailed sampling process and D-score calculation can be found in Appendix B.

As shown in Figure 4.6, all the three sample groups, including the original sample group, demonstrate a downward trend in the monthly D-score since 2017. Especially, their trends during 2018 and 2019 share a similar pattern. The only difference is the year when D-score peaked. The D-score in the original sample group reached the highest in 2017; The D-score in the Sample Group B reached the highest in 2016; The D-score in the Sample Group C reached the highest in 2015. This correlates to the fact that if research population are sampled among the users have rating history in a certain year, it is inevitable to include a large number of users joining MovieLens in that year, as shown in Figure 4.7. Such a difference is attributed to the influx of new users, which has a negligible impact on the overall trend of D-scores. Thus, the results generated from the original sample group is robust enough.

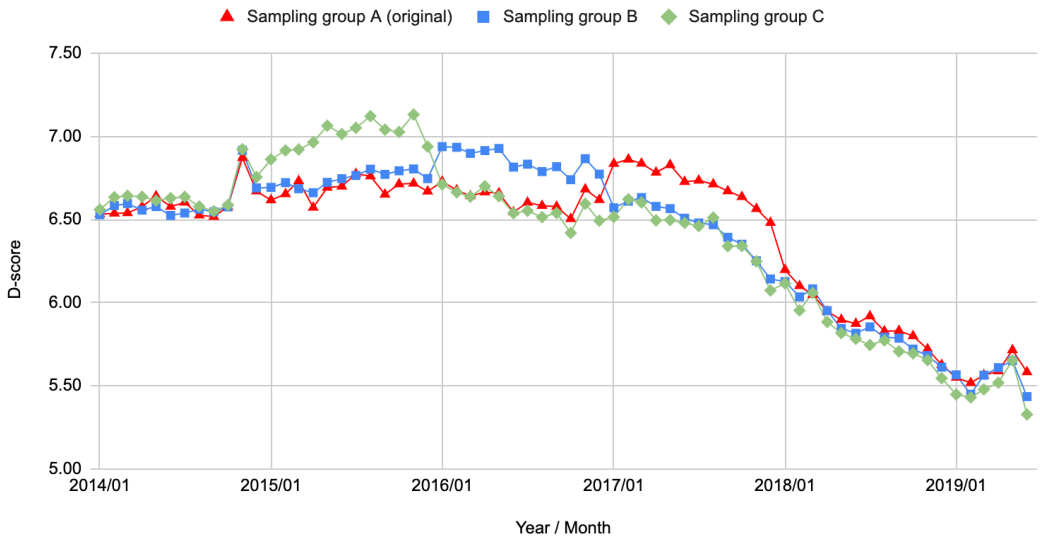


Figure 4.6 Monthly D-scores of three groups of sampling users in 2014-2019

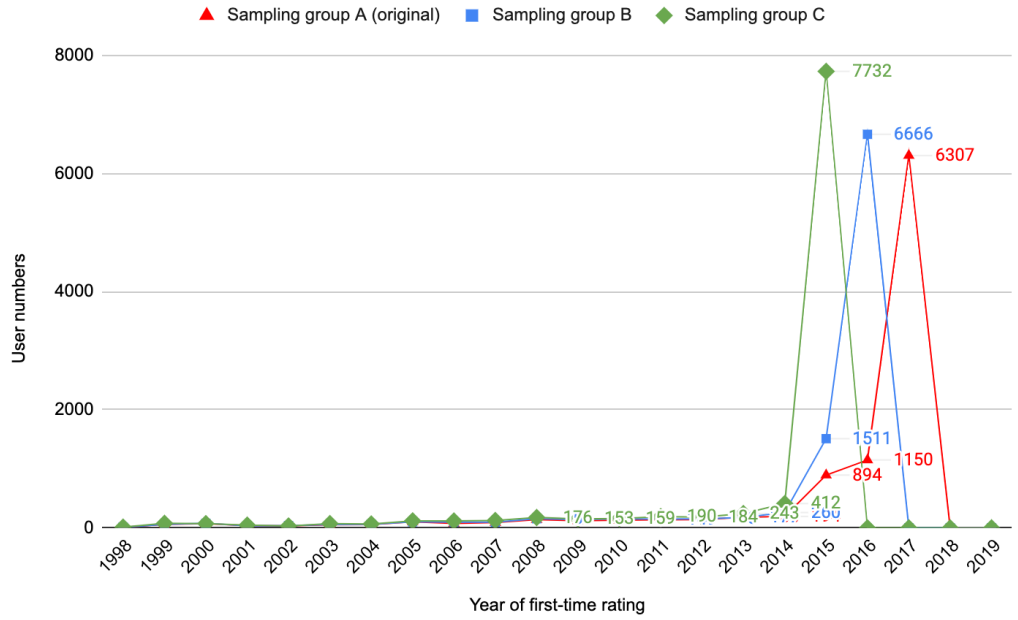


Figure 4.7 The number of new users joining in MovieLens in each sample group

Chapter 5. Discussion for Hypothesis 1

The above interrupted time series analysis confirmed a decline in the consumed diversity on MovieLens in recent years, which is exactly the opposite of the first hypothesis. The possible explanations of the declined consumed diversity in this study are multiple. It relates to supplied diversity and consumers' changing tastes.

Declining supplied diversity

The main contributor to a less diverse movie consumption in this study is the decreased supplied diversity. Consumers constantly watch a number of new movies for novel experience. Thus, if there are less diverse new movies available, consumers have less diverse choices. The consumed diversity would be accordingly lower. The supplied diversity of new movies can be estimated by calculating the diversity of the newly released movies based on the information from Tag Genome. I group all movies on MovieLens by year of release and calculate the D-scores for each group. The detailed calculation can be found in Appendix A. As shown in Figure 5.1, the annual median D-score during 2003 and 2016 stays around 5.9, and it drops to 5.3 in 2017. The D-score for 2018 and 2019 dramatically dropped to about 4.0. The D-scores show that the movies supplied since 2017 are less diverse year by year. The timing of the big drop in supplied diversity coincides with the timing of the decline in consumed diversity. This indicates a correlation between supplied diversity and consumed diversity.

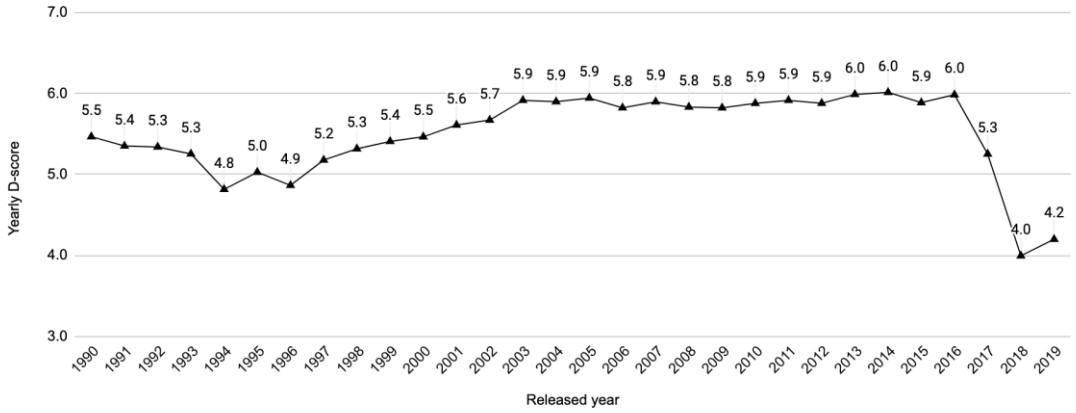


Figure 5.1 The annual median D-score of the movies released in 1990 - 2019

The less diverse supply of movie since 2017 may be due to the application of big data on movie production and the use of star power. Big data of user behavior and preferences are now great treasure to movie production companies. By applying artificial intelligence, firms such as Netflix and Amazon are able to analyze how audience react upon every clip of a video and thus to predict what content audience would enjoy better (Charles, 2018). Therefore, they make movies that deliberately cater to audience tastes and preferences based on a formula generated by big data. As more and more firms use this technique, they are sharing the same formula to create new movies at the same time. As more and more user data shared, firms are likely to produce increasingly similar movies.

In addition, as non-English movie markets especially Asia accounted for an increasingly larger share of the global revenue (MPAA, 2017), the financial incentive for world's major movie production companies to cater to Asian audience was greater than ever (Cruz et al., 2017). To avoid audience from non-English speaking countries to get lost in translation, those companies tend to produce movies with features that

appeal to a larger audience. Star power is proven to positively impact box-office in China (Peng et al., 2019). Production studios thus spent more money on inviting A-list movie stars than refining the scripts (Lin, 2019). This could lead to movies with big stars but a very boilerplate story, which may indirectly reduce the diversity of supplied movies. As some insiders in the movie industry criticize, film production companies are making films that are not culturally specific but share universal ideas and themes (Brook, 2014).

Another explanation of declined supplied diversity may be the inherent bias of the 25M Dataset provided by MovieLens. The D-score in this study is the quantitative indicator of diversity, and the D-score is derived based on Tag Genome in the 25M Dataset. The process of counting tag application and the mechanism of predicting tag relevance values are not fully transparent and open to researchers. All we know is that they are generated through a machine learning algorithm. According to the MovieLens team, they did not make any noteworthy changes or upgrades to the dataset around 2017. But we still have reasons to doubt the impartiality of the 25M Dataset.

"Filter bubbles" exist

A decline in the consumed diversity to some extent support the theory of "filter bubbles". While many sing the praises of efficiency and intelligence brought about by digital technology, Pariser (2011) developed the concept of "filter bubbles" to describe a tendency for personalized digital services to isolate consumers from a wider range of content. Consumers' movie watching behavior has undergone a transformation in the landmark year of 2017. As an alternative to going to cinemas or buying Blu-rays,

subscription video on demand services have emerged worldwide since 2017. The worldwide subscription numbers increased significantly by 33% from 2016 to 2017 and 27% from 2017 to 2018 (MPAA, 2018, 2019). It is true that digitalization has lowered the barriers for consumers to have access to a wider range of movies, but personalized services have mastered so much user behavior that they can accurately predict user tastes and preferences and thus probably filter out what they might dislike, which is so called algorithmic selection. As consumers have limited time and resources, they may choose a rather efficient way of finding movies. As they are surrounded by more and more recommender systems, they are likely to be busy bouncing from recommendation to recommendation without any thought to peek outside their filter bubbles. However, the above speculations need more empirical studies to confirm.

Chapter 6. Results and analysis for Hypothesis 2

6.1 D-scores during each month of using MovieLens

To test the second hypothesis, another 10,000 users[11] (Sample Group D) are randomly sampled among all the users who have rated movies on MovieLens since 2010. Each user has rated at least 20 movies. All the rating history of these 10,000 users are grouped by their usage time on MovieLens. Specifically, usage time is the difference between the time the movie was rated and the time the user rated movies for the first time, which is obtained by subtracting the time stamp of every user's first-time rating from the time stamp of the last-time rating. The unit of the usage time here is month(s). For instance, as shown in Table 6.1, "user 162533" rated *Senna (2010)* on Feb 17th 2012 and the first time he rated a movie on MovieLens was on Aug 3rd 2010. Thus, when he rated *Senna (2010)*, his usage time on MovieLens is 18 months[12].

UserId	*Rating*	*Titile*	*Time stamp*	*1st time stamp*	*Usage (month)*
34	3	Othello (1995)	2011-10-04 20:41:03	2011-10-04 20:31:46	0
34	3	Braveheart (1995)	2011-10-04 21:04:25	2011-10-04 20:31:46	0
34	4	Pulp Fiction (1994)	2011-10-04 21:06:40	2011-10-04 20:31:46	0
...	...	...	...	...	...

[11] The users rating over 3000 movies in one month were excluded, as such large amounts of ratings could make Python crash in computation and is suspicious regarding the source of ratings.

[12] When *x* is an integer, the usage time between *x* and *x.5* is grouped in *x* month; the exact usage time greater than or equal to *x.5* and less than *x+1* is grouped in *x+1* month.

162533	4	Senna (2010)	2012-02-17 21:29:33	2010-08-03 10:44:47	18
162533	4.5	Drive (2011)	2012-02-17 21:28:59	2010-08-03 10:44:47	18
162533	1.5	Rise of the Planet of the Apes (2011)	2012-02-17 21:32:50	2010-08-03 10:44:47	18
162533	2	Contagion (2011)	2012-02-17 21:32:33	2010-08-03 10:44:47	18
162533	3.5	Shame (2011)	2012-02-17 21:33:51	2010-08-03 10:44:47	18

Table 6.1 Partial rating history of users in Sample Group D

After categorizing all the user ratings, 9994 users are observed to have rated movies during the first two weeks of using MovieLens, whereas only about 1% of them continued to rate movies during the following four months after they joined MovieLens , and less than 1% rated movies in the fifth month. Thus, the first two weeks is listed as a separate usage period, as shown in Table 6.2. To summarize, almost all the new users are active on MovieLens in the first two weeks, and fewer users kept rating movies with the usage time on MovieLens.

	0.5 month[13]	*1st month*	*2nd month*	*3rd month*	*4th month*	*5th month*
# users	9,994	1,658	1,268	1,087	1,013	916
# movies	24,638	8,486	6,506	6,782	6,456	6,049

Table 6.2 Number of rating users and number of rated movies in each month of usage

[13] "0.5 month" represents the movies rated within 2 weeks after a user first time rated on MovieLens. The movies rated from the third to the sixth week are included in "1st month".

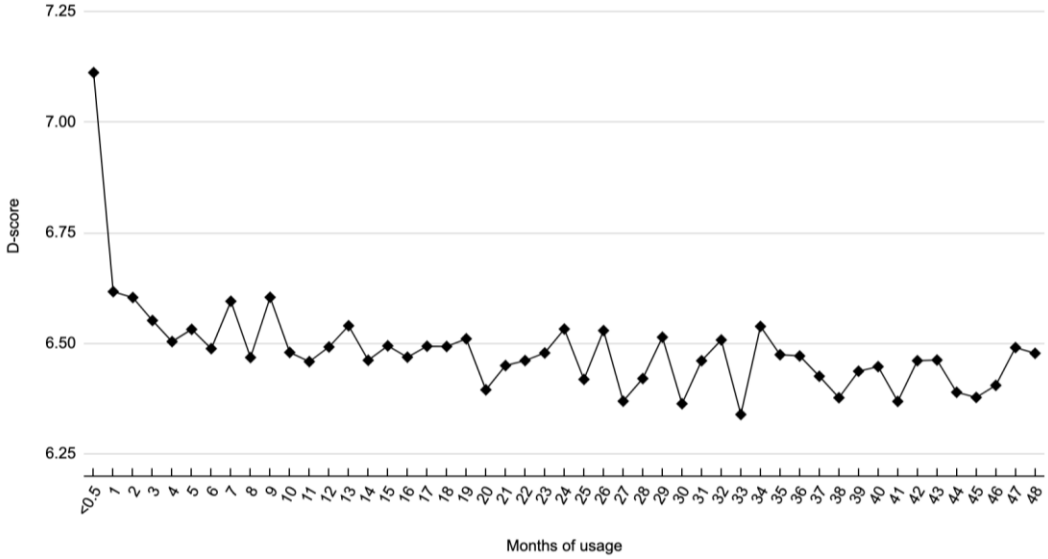


Figure 6.1 Median D-scores of the movies rated in each month of use on MovieLens

The D-score in every usage period is computed as follows. Firstly, I calculate the median D-score for all the movies rated by each user during their first half month of use on MovieLens. Then each user has a D-score. Secondly, among all the users, I choose the median D-score to represent the diversity of all the sampling users' movie ratings during the first half month. Lastly, the D-scores in the 1st to the 48th month are computed in the same way. All the detailed code can be found in Appendix B.

The D-score in every single usage month is displayed in Figure 6.1, showing a sudden decline in D-score from the first half month to the second half month and a continued decrease at a very low rate since the second month of usage. The big decrease during the first half month may be attributed to the changing user rating patterns. As shown in Table 6.2, 99% of sampling users were only actively rating during the first two weeks. They may have rated the most impressive movies they have watched previously and then never logged in. The small decrease in the later months may be

attributed to narrowing individual tastes. Only 1% of the sampling users are found to stay longer and keep rating movies on a regular basis in the later months. Those users are supposed to be movie lovers. It is assumed that movie lovers watch movies regularly. Once they have watched a movie, they rate it on MovieLens as a habit. Over time, their tastes in movies may have been narrowed through constant watching and rating. Such a taste narrowing process may be due to two reasons: decreasing consumption of popular movies and less diverse consumption in terms of age of movies. These two reasons will be further analyzed in the next two sections.

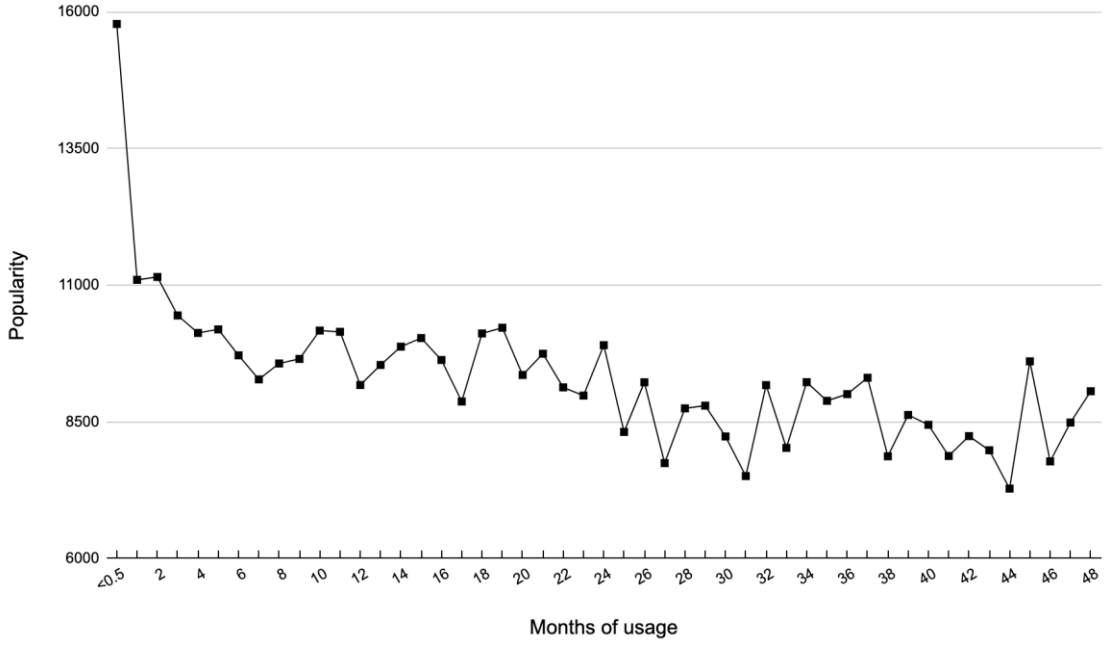
6.2 The effect of popularity of movies

The narrowing personal tastes in movies may be reflected on the changing popularity of movies that rated by users, because as experience grows, users are assumed to move from many popular movies to movies with certain unique commonalities. An index of popularity of movies is thus established. I count the number of ratings of every movie based on the 25M Dataset which contains 25 million ratings by 162 thousand users across 62 thousand movies on MovieLens. Thus, every movie has a total number of users who have rated this movie. This number is the popularity index of the movie.

The average popularity index of the movies rated in each usage month is presented in Figure 6.2. The results in the figure show that the movies rated within the first half month do have a much higher popularity index than in the later months, which echoes the speculation that when users just joined MovieLens they may have simply reported a large number of memorable movies that they had watched. Moreover, the popularity index for the later months still shows a downward trend with random fluctuations, but

the rate of decrease is low. This suggests that users tend to watch less popular movies with the usage time on MovieLens, and that this trend is very slow.

Figure 6.2 Popularity of movies rated by Sample Group D in each usage month



6.3 The effect of age of movies

To identify whether the decrease in D-score with the usage time on MovieLens is related to age of movies, I count all the movies that users in Sample Group D have rated by released year of movies. Specifically, for each month of usage, all the rated movies are divided into twelve age groups, namely ‘before 1900’, ‘1910s’, ‘1920s’ and etc. Movies released between 2000 and 2019 are defined as new movies, and movies released before 2000 are defined as old movies. For instance, *Pulp Fiction (1994)* belongs to ‘1990s’ and is considered old movies. The detailed list can be founded in Appendix C.

	0.5 month[14]	*1st month*	*2nd month*	*3rd month*	*4th month*	*5th month*
< 1900	0.0024	0.0006	0.0000	0.000	0.000	0.000
1910s	0.0024	0.0012	0.0008	0.001	0.000	0.000
1920s	0.030	0.018	0.028	0.019	0.016	0.014
1930s	0.072	0.042	0.036	0.064	0.064	0.062
1940s	0.10	0.07	0.06	0.09	0.09	0.09
1950s	0.14	0.10	0.10	0.14	0.16	0.15
1960s	0.04	0.12	0.16	0.22	0.21	0.24
1970s	0.34	0.26	0.23	0.27	0.31	0.35
1980s	0.48	0.53	0.51	0.69	1.11	0.76
1990s	0.70	1.19	1.01	1.04	1.10	1.37
2000s	1.30	1.48	1.58	2.09	1.80	2.01
2010s	0.46	1.23	1.41	1.61	1.51	1.57
sd	0.38	0.55	0.58	0.70	0.66	0.71

Table 6.3 Amounts of movies of different ages that an average user rated per month

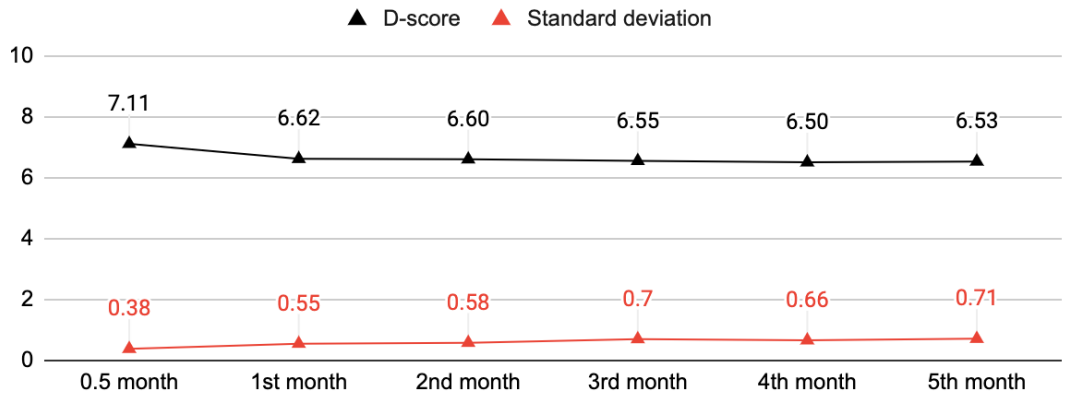


Figure 6.3 D-score of movies rated by users per month versus standard deviation of the average number of movies of different ages that were rated by users per month

[14] The numbers for the first half month are doubled in order to be comparable with other months.

The average number of movies of different ages that a user rated per month is presented in Table 6.3. To figure out the distribution of ratings in terms of the age of movies, the standard deviation (SD) for each month of usage is calculated. The SD for the first half month is the smallest compared with other months, indicating that in this period there is a relatively balanced distribution of consumption in terms of age of movies. During the later months, while the D-score shows a downward trend, the SD per month shows an upward trend, as seen in Figure 6.3, indicating that users were rating less diverse movies in terms of ages of movies. In addition, the numbers of ratings per month are concentrated in the row of 2000s and 2010s, and they were per month getting larger. This indicates that users rated more movies produced after 2000 in the later months of using MovieLens.

Chapter 7. Discussion for Hypothesis 2

The correlation between the decreased consumed diversity and the usage time on MovieLens has been identified. The results and analysis in Chapter 6 rejected the second hypothesis. The diversity of individual movie consumption suddenly declined from a high level in the first half month of using MovieLens and continued to slowly decrease in the later months.

The movies rated during the first two weeks are not necessarily what users actually watched during that period but probably the most impressive movies of all times to users. This speculation fits the behavior patterns of real-life users using a movie review website. After one month's usage, users' behavior patterns are assumed to change to real-time reporting movies that they have watched. Thus, the rating behavior after the first month of usage gives a more detailed picture of how users watch movies over time. In the later months, the diversity of individual movie consumption decreased at a quite low rate with the increase of experience. Such results have multiple explanations.

First, a user's taste in movies may change from relatively popular movies to more personalized ones as experience grows. After computing both the D-score and the popularity index for the movies rated in each month of use on MovieLens, the sampling users' consumed diversity is observed to correlate with the popularity of movies. This indicates that users tend to watch less popular movies over time and what they have watched tend to be less diverse. For new users who lack consumption experience, they watch movies mostly by observing choices of others or referring to expert opinion, such as movie rankings and awards. These movies represent a collection of most people's

tastes. Thus, these movies are more diverse than the movies only liked by a single person. After constant consuming and rating, new users grow to long-term users and have gradually learned what they like. In this study, the users who have stayed more than one month on MovieLens are assumed to choose next movie to watch by relying more on their current tastes and preferences and less on recommendations from others and experts. Long-term users thus tried more niche movies that match their specific tastes and watched less popular movies that recognized by the general public. Furthermore, when users focus on certain types of niche movies, they would explore similar movies rather than search movies in general. As the theory of "learning by consuming" states, the current experience of watching a movie influences the choice of next movie. The dissimilarity between the current movie and the next movie would be quite small. That is why the D-score decreased at a low rate in the later months of usage in our study, as the serial consumption of long-term users is basically narrowed within the movies that share some specific features.

Second, users seem to focus more on new movies rather than vintage ones as experience grows. When users first joined MovieLens, they are assumed to simply report what they have watched in the past. It is understandable that the distribution of movies across ages was relatively even among all their rated movies, as those movies reflect their past consumption so far. In the later months, users may have changed their rating patterns. They are assumed to rate each movie in real time. The newer movies began to make up a higher percentage of their monthly consumption. Most long-term sampling users are found to watch more new movies produced after 2000 than vintage

movies monthly. The movies rated by them tend to be more unevenly distributed in terms of age of movies. This echoes the theoretical assumption that people value diversity in terms of the balance between novelty and familiarity while in reality they were more in favour of novel experience. Additionally, further regression analysis is needed if we want to statistically confirm the correlation between the diversity of the age of consumed movies and the D-score.

Moreover, the application of big data may accelerate the formation of users' specific tastes as well as intensify users' exposure to newer movies, which leads to a decrease of D-scores. As users watch more movies online, various recommender systems will obtain more behavioural data about users, so that they could predict movies that users may like more efficiently. If users gain much utility from recommender systems, they are more likely to rely on such personalized media services and less refer to popular choices of movies. Meanwhile, more and more movie production and distribution companies choose precise marketing to targeted audience for their new releases by leveraging personalized media services, so that some targeted users can easily reach those new movies in the promotional period. Digitalization seems to have backfired in promoting diversity.

Chapter 8. Conclusions

The two hypotheses are both falsified, which is unexpected but somehow reasonable. Firstly, compared to the overall users on MovieLens in 2015 and 2016, the overall users after 2017 have per year watched less diverse movies. The consumed diversity decreased since the landmark year of 2017, a year around which coincided with the tipping point of the impact of digitalization on all aspects of the movie industry. Such a decrease directly stems from a less diverse supply of movies produced since 2017. The root cause may be that movie production and distribution have become more data-driven and have relied more on star power. Secondly, for an average user on MovieLens, the diversity of individual movie consumption decreases as individual experience grows. Such a decrease in diversity may be mainly attributed to the observations that long-term users watched less diverse movies in terms of movie ages and they watched more movies with some specific features than popular movies. Moreover, such a decrease shows a very slow trend as users stayed longer on MovieLens. This correlates with progressively narrowing personal tastes. As individual tastes become stabilized, users no longer search movies haphazardly, but explore for certain directions that they prefer.

Changes in how users rate movies on MovieLens is a snapshot of an individual's lifetime movie consumption. Studying the consumed diversity of MovieLens users shows us that people tend to watch less diverse movies, whether on an aggregate level or an individual level, even though digitalization is considered to have promoted a more diverse consumption. Still, not much empirical research has been done on the diversity

of movie-watching behavior. Academics should make more use of existing user data on movie-watching in future. As a continuation of this study, future research could further estimate whether the most popular movies in history are more diverse than movies in general. In addition, further research to investigate the link between the application of big data in movie industries and the diversity in movie supply as well as the link between age of consumed movies and consumed diversity could also be of interest. With more data-driven studies as reference, all sectors of the movie industry could pay more attention to the diversity of their cultural supply, thus preventing consumers from a relatively narrow-minded social values and cultural expressions.

It cannot be ignored that this paper has some limitations. First of all, the users' D-scores in this study may be inherently lower than the real consumed diversity, as not all the movies on MovieLens have been included. Tag Genome only provides tag relevance values to 13,816 movies that are well rated and tagged, as the rest of movies are almost untouched by enough users. This empirically demonstrates that the long tail of the movie industry is shockingly flat and long. More importantly, this excluded movies that are extremely niche, which would have raised the D-scores of the users who happened to have watched those movies. Second, ratings of movies are not equal to movie consumption, as it is difficult to make it clear whether a user marked a movie that they never watched or watched it in real life. Third, the research population on MovieLens does not necessarily represent consumers in general. Since the demographics of users on MovieLens are unknown, the variables such as age, gender, nationality, educational and occupational backgrounds were not controlled.

REFERENCES

Anderson, C., & Andersson, M. P. (2004). *Long tail*.

Baek, H., Oh, S., Yang, H.-D., & Ann, J. (2014). Chronological analysis of the electronic word-of-mouth effect of four social media channels on movie sales: comparing Twitter, Yahoo! Movies, Youtube, and Blogs. *PACIS 2014 Proceedings*. https://aisel.aisnet.org/pacis2014/65

Bekar, C., & Haswell, E. (2013). General purpose technologies. *Chapters*, 9–19. https://ideas.repec.org/h/elg/eechap/14906_1.html

Benhamou, F., & Peltier, S. (2007). How should cultural diversity be measured? An application using the French publishing industry. *Journal of Cultural Economics*, *31*(2), 85–107.

Benhamou, F., & Peltier, S. (2010). *Application of the Stirling Model to assess diversity using UIS cinema data*.

Blaug, M. (2001). Where Are We Now On Cultural Economics. *Journal of Economic Surveys*, *15*(2), 123–143. https://doi.org/10.1111/1467-6419.00134

Bocart, F., Gertsberg, M., & Pownall, R. A. J. (2017). Glass Ceilings in the Art Market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3079017

Bonet, L., & Négrier, E. (2011). The end(s) of national cultures? Cultural policy in the face of diversity. *International Journal of Cultural Policy*, *17*(5), 574–589. https://doi.org/10.1080/10286632.2010.550681

Bourreau, M., Gensollen, M., & Perani, J. (2002). *Economies of Scale in the Media Industry*.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*.

Brook, T. (2014). *How the global box office is changing Hollywood*. BBC Culture. https://www.bbc.com/culture/article/20130620-is-china-hollywoods-future

Brownlee, J. (2017). *Introduction to time series forecasting with python: how to prepare data and develop models to predict the future* (1.11). Machine Learning Mastery.

Brynjolfsson, E., Hu, Y. J., & Simester, D. (2011). Goodbye Pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, *57*(8), 1373–1386. https://doi.org/10.1287/mnsc.1110.1371

Caves, R. E. (2000). *Creative Industries: Contracts Between Art and Commerce*. Harvard University Press.

Chamberlin, E. (1933). The Theory of Monopolistic Competition. *The American Economic Review*, *23*(4), 683–685.

Charles, B. (2018). *How Big Data is Changing Filmmaking - Movie World Is Changing*. https://www.raindance.org/how-big-data-is-changing-filmmaking/

Chisholm, D. C. (2005). Hollywood Economics: How Extreme Uncertainty Shapes The Film Industry. *Journal of Cultural Economics*, *29*(3), 233–237. https://doi.org/10.1007/s10824-005-2858-4

Chiu, Y. L., Chen, K. H., Wang, J. N., & Hsu, Y. T. (2019). The impact of online movie word-of-mouth on consumer choice: A comparison of American and

Chinese consumers. *International Marketing Review*, *36*(6), 996–1025. https://doi.org/10.1108/IMR-06-2018-0190

Christiansen, S. (2012). Cultural diversity. In *Dairy Industries International* (Vol. 77, Issue 7, p. 24). https://doi.org/10.4337/9781788975803.00026

CouncilofEurope. (2000). *Recommendation Rec(2007)2 of the committee of ministers to memberstates on media pluralism and diversity of media content*.

Cruz, G. DeLa, Pedace, R., & Pinczower, Z. (2017). *Homogeneity in Hollywood: Discrimination in Motion Pictures*.

De Vany, A. (2000). *Private information, demand cascades, and the blockbuster*. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Private+information%2C+demand+cascades%2C+and+the+blockbuster+strategy&btnG=

De Vany, A. (2006). Chapter 19 The Movies. In *Handbook of the Economics of Art and Culture* (Vol. 1, pp. 615–665). Elsevier. https://doi.org/10.1016/S1574-0676(06)01019-2

De Vany, A., & Walls, W. D. (1999). Uncertainty in the movie industry: Does star power reduce the terror of the box office? *Journal of Cultural Economics*, *23*(4), 285–318. https://doi.org/10.1023/A:1007608125988

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, *67*(3), 297–308.

Duan, W., Gu, B., & Whinston, A. B. (2008). The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing*, *84*(2), 233–242. https://doi.org/10.1016/j.jretai.2008.04.005

Eaton, B., & Lipsey, R. (1989). Product differentiation. *Handbook of Industrial Organization*. https://www.sciencedirect.com/science/article/pii/S1573448X89010150

Ekstrand, M. D., Harper, F. M., Willemsen, M. C., & Konstan, J. A. (2014). *User Perception of Differences in Recommender Algorithms*. https://doi.org/10.1145/2645710.2645737

Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science*, *22*(3). https://doi.org/10.1287/mksc.22.3.329.17740

Finkelstein, M. O., & Friedberg, R. M. (1967). The Application of an Entropy Theory of Concentration to the Clayton Act. *The Yale Law Journal*, *76*(4), 677. https://doi.org/10.2307/795029

Fleder, D., & Hosanagar, K. (2009). Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science*, *55*(5), 697–712. https://doi.org/10.1287/mnsc.1080.0974

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, *90*(4), 715–741. https://doi.org/10.1257/aer.90.4.715

Grabher, G., & Stark, D. (1997). Organizing diversity: Evolutionary theory, network analysis and postsocialism. *Regional Studies*, *31*(5), 533–544. https://doi.org/10.1080/00343409750132315

Handke, C., Stepan, P., & Towse, R. (2016). Cultural economics and the Internet. In

Handbook on the Economics of the Internet (pp. 146–162). Edward Elgar Publishing. https://doi.org/10.4337/9780857939852.00014

Hendrickx, J. (2018). Is the filter bubble #fakenews? *Diamond*.

Hennig-Thurau, T., Walsh, G., & Wruck, O. (2001). *Wruck / Factors Determining the Success of Service Innovations*. http://www.amsreview.org/articles/henning06-2001/pdf

Hotelling, H. (1929). Stability in competition. *Economic Journal*, *39*, 41–57.

Johnson, N., & Longmeyer, S. (1999). *The science of social diversity*.

Kauffman, S. A. (1992). *The Origins of Order: Self-Organization and Selection in Evolution* (pp. 61–100). https://doi.org/10.1142/9789814415743_0003

Kirman, A. (2006). Heterogeneity in economics. *J Econ Interact Coord*, *1*, 89–117. https://doi.org/10.1007/s11403-006-0005-8

Lancaster, K. (1979). *Variety, equity, and efficiency: product variety in an industrial society*. Columbia University Press.

Lancaster, K. (1990). The Economics of Product Variety: A Survey. *Marketing Science*, *9*(3), 189–206. https://doi.org/10.1287/mksc.9.3.189

Lévy-Garboua, L., & Montmarquette, C. (2002). *The demand for the arts.*

Lin, W. (2019). *The international movie market is transforming Hollywood*. Salon.Com. https://www.salon.com/2019/07/13/the-international-movie-market-is-transforming-hollywood_partner/

Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, *70*(3), 74–89.

https://doi.org/10.1509/jmkg.70.3.074

Marshall, A. (1961). *Principles of economics: unabridged eighth edition* (C. Guillebaud (ed.)).

Maxwell, H. F., & Konstan, J. A. (2015). The MovieLens Datasets: History and Context. In *ACM Trans. Interact. Intell. Syst* (Vol. 20).

May, R. (1975). Patterns of species abundance and diversity. *Ecology and Evolution of Communities*. https://ci.nii.ac.jp/naid/10003520027/

McCann, K. (2000). The diversity–stability debate. *Nature,*.

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, *21*(7), 959–977. https://doi.org/10.1080/1369118X.2018.1444076

Moreau, F., & Peltier, S. (2004). Cultural Diversity in the Movie Industry: A Cross-National Study. *Journal of Media Economics*, *17*(2), 123–143. https://doi.org/10.1207/s15327736me1702_4

Moul, C. C. (2007). Measuring word of mouth's impact on theatrical movie admissions. *Journal of Economics and Management Strategy*, *16*(4), 859–892. https://doi.org/10.1111/j.1530-9134.2007.00160.x

MPAA. (2017). *Theatrical Market Statistics 2016*.

MPAA. (2018). *2017 Theatrical Home Entertainment Market Environment (THEME) Report*.

MPAA. (2019). *2018 Theatrical Home Entertainment Market Environment (THEME)*

Report.

MPAA. (2020a). *2019 Theatrical Home Entertainment Market Environment (THEME) Report*.

MPAA. (2020b). *SPECIAL RULES FOR THE INTERNATIONAL FEATURE FILM AWARD*.

Musgrave, R. A. (1987). Merit goods. *The New Palgrave: A Dictionary of Economics*, *3*, 452–453.

Napoli, P. M. (1999). Deconstructing the Diversity Principle. *Journal of Communication*, *49*(4), 7–34. https://doi.org/10.1111/j.1460-2466.1999.tb02815.x

Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy*, *78*(2), 311–329. https://doi.org/10.1086/259630

Nguyen, T. T., Hui, P. M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: The effect of using recommender systems on content diversity. *Proceedings of the 23rd International Conference on World Wide Web*, 677–686. https://doi.org/10.1145/2566486.2568012

Odum, E. (1953). *Fundamentals of ecology*.

Palmer, T. G. (2004). *Globalization and culture: Homogeneity, diversity, identity, liberty.* http://edoc.vifapol.de/opus/volltexte/2011/2414/pdf/OP2.pdf

Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*.

Peltoniemi, M. (2015). Cultural Industries: Product-Market Characteristics, Management Challenges and Industry Dynamics. *International Journal of*

Management Reviews, *17*(1), 41–68. https://doi.org/10.1111/ijmr.12036

Peng, F., Kang, L., Anwar, S., & Li, X. (2019). Star power and box office revenues: evidence from China. *Journal of Cultural Economics*, *43*(2), 247–278. https://doi.org/10.1007/s10824-018-9338-0

Ranaivoson, H. (2007). *Measuring cultural diversity: a review of existing definitions*.

Ranaivoson, H. (2012). Does the Consumer Value Diversity? How the Economists' Standard Hypothesis is Being Challenged. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2189425

Scitovsky, T. (1976). *The joyless economy: An inquiry into human satisfaction and consumer dissatisfaction*.

Simonoff, J. S., & Sparrow, I. R. (2000). Predicting Movie Grosses: Winners and Losers, Blockbusters and Sleepers. *CHANCE*, *13*(3), 15–24. https://doi.org/10.1080/09332480.2000.10542216

Solow, A. R., & Polasky, S. (1994). Measuring biological diversity. *Environmental and Ecological Statistics*, *1*(2), 95–103. https://doi.org/10.1007/BF02426650

Sotshangane, N. (2002). What Impact Globalization has on Cultural Diversity? *Alternatives: Turkish Journal of International Relations*, *1*(4).

Spence, M. (1976). Product selection, fixed costs, and monopolistic competition. *Review of Economic Studies*, *43*(2), 217–235. https://doi.org/10.2307/2297319

Stenou, K. (2004). *UNESCO AND THE ISSUE OF CULTURAL DIVERSITY Review and strategy, 1946-2004 A study based on officiel documents DIVISION OF CULTURAL POLICIES AND INTERCULTURAL DIALOGUE*.

Stirling, A. (1998). On the economics and analysis of diversity. In *Science Policy Research Unit (SPRU), Electronic Working Papers Series, Paper, 28, 1-156.* (No. 28; Electronic Working Papers Series).

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, *4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213

Sugihara, G. (1982). Diversity as a Concept and its Measurement: Comment. *Journal of the American Statistical Association*, *77*(379), 564. https://doi.org/10.2307/2287711

TheGuardian. (2010). *The 25 best arthouse films of all time: the full list*.

Tobias, S. (2004). Quality in the performing arts: Aggregating and rationalizing expert opinion. *Journal of Cultural Economics*, *28*(2), 109–124. https://doi.org/10.1023/B:JCEC.0000019472.97483.8c

Towse, R. (2011). A Handbook of Cultural Economics. In *Edward Elgar Publishing* (pp. 53–55).

Towse, R. (2019). *A Textbook of Cultural Economics*. Cambridge University Press.

Tryon, W. W. (1982). A Simplified Time-series Analysis for Evaluating Treatment Inverventions. *JOURNAL OF APPLIED BEHAVIOR ANALYSIS*.

Van Der Wurff, R., & Van Cuilenburg, J. (2001). Impact of moderate and ruinous competition on diversity: The dutch television market. *Journal of Media Economics*, *14*(4), 213–229. https://doi.org/10.1207/S15327736ME1404_2

Vig, J., & Riedl, J. (2012). The tag genome: Encoding community knowledge to

support novel interaction. *ACM Trans. Interact. Intell. Syst*, *2*(3). https://doi.org/10.1145/2362394.2362395

Vogel, H. L. (1998). *Entertainment Industry Economics A Guide for Financial Analysis*. Cambridge University Press. https://doi.org/10.1017/9781108675499

Vrijenhoek, S., Kaya Independent Researcher, M., Metoui Delft, N. T., Möller, J., Odijk, D., & Helberger, N. (2021). *Recommenders with a Mission: Assessing Diversity in News Recommendations*. *11*. https://doi.org/10.1145/3406522.3446019

Wasko, J. (2003). *How hollywood works*.

Wu, L.-L., Joung, Y.-J., & Chiang, T.-E. (2011). *Recommendation Systems and Sales Concentration: The Moderating Effects of Consumers' Product Awareness and Acceptance to Recommendations*.

Young, L. C. (1941). On Randomness in Ordered Sequences. *The Annals of Mathematical Statistics, 12*(3), 293–300.

Ziegler, C.-N., McNee, S. M., Nr, G., Konstan, J. A., & Lausen, G. (2005). Improving Recommendation Lists Through Topic Diversification. *Proceedings of the 14th International Conference on World Wide Web*, 22–32.

Zufryden, F. (2000). New film website promotion and box-office performance. *Journal of Advertising Research*, *40*(1–2), 55–64. https://doi.org/10.2501/JAR-40-1-2-55-64

APPENDICES

A. Code on Python Part 1, including the computation of two D-score generators, regression models and C statistics

B. Code on Python Part 2, including the computation of sampling process, D-scores by usage period and popularity index

C. All movies rated by users in Sample Group D, sorted by year of release