Is Cinema Becoming Less and Less Innovative With Time? Using neural network text embedding model to measure cultural innovation ⋆

Edgar Dubourg*, Andrei Mogoutov and Nicolas Baumard

Institut Jean Nicod, Département d'études cognitives, Ecole normale supérieure, Université PSL, EHESS, CNRS, 75005 Paris, France

Abstract

Current discourse reflects a growing skepticism towards contemporary popular culture, specifically the realm of cinema, with an emerging consensus that its creative capacity is on a waning trajectory. This study introduces a novel approach which employs natural language processing techniques and embedding methods to measure semantic novelty of cultural items' descriptions. We apply this methodology to cinema, analyzing plot summaries of over 19,000 movies from the United-States spanning more than a century. Our measure's robustness is validated through a series of tests, including a fit with a genre-based novelty score, a manual inspection of films identified as highly innovative, and correlations with award recognitions. The application of our Innovation Score reveals a compelling pattern: an increase in the rate of cinematic innovation throughout the 20th century, followed by a stabilization in the rate of innovation in the 21st, despite an ever-growing production of films. Contrary to the often-voiced lament that cinema is losing its innovative edge, our study suggests that the level of innovativeness in cinema is not in decline.

Keywords

innovation, creativity, culture, text embedding

1. Introduction

The film industry has been the subject of numerous debates regarding its perceived decline in innovation. Critics and audiences alike have voiced concerns about the increasing prevalence of sequels, remakes, and franchise films, arguing that these trends would reflect a lack of originality and creativity. These apprehensions have been amplified by the publicized sentiments of esteemed filmmakers such as Francis Ford Coppola, who likened popular fictions to "prototypes made over and over and over again," Ken Loach, who compared them to "commodities like hamburgers," or Alejandro Iñárritu, who dismissed them as "basic and simple.". The apprehensions over the diminishing creative vigor in cinema are not confined to academic or elite circles; they resonate deeply with the broader public. A quick glance at social media,

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France

*Corresponding author.
edgar.dubourg@gmail.com (E. Dubourg); mogoutov@gmail.com (A. Mogoutov); nbaumard@gmail.com (N. Baumard)
https://edgardubourg.fr/ (E. Dubourg); http://nicolasbaumard.org (N. Baumard)
0000-0002-2162-6526 (E. Dubourg)

film forums, and audience reviews reveals a torrent of sentiments expressing disappointment with the perceived stagnation of storytelling, reliance on formulaic plotlines, and the increasing tendency to prioritize profit over artistic innovation. But is it the case? Is innovation in cinema on decline?

The challenge in defining innovation, and in determining whether a product is innovative, lies in its subjective nature. We define innovation as what is novel for humanity at large, as opposed to novelty, which is what is contextually novel for an individual [18, 19]. This distinction is crucial because it shifts the focus from individual perceptions to a collective level. Prior research has attempted to quantify innovation in film by analyzing the unique combinations of IMDb genres [8] or IMDb plot keywords [17]. However, these methods have inherent limitations in tracking innovation over time. They heavily rely on metadata that tends to be more abundant and precise for recent movies. This can result in a skewed perspective, as older films often lack comprehensive metadata. Other measures that aimed at quantifying innovation in other creative domains such as the arts, technology, or science, relied on creative *individuals* (e.g., their interaction, see [15]; their number and productivity, see [2]; their place of birth and movement, see [16]; their reputation, see [4]).

In this paper, we develop and apply a computational measure of innovation to movies based on summaries, which are standardized and rather homogenous in both IMDb and Wikipedia. This new measure is straightforwardly applicable, not to individuals, but to cultural products and aims at measuring their *objective level of innovativeness.* It could in principle be applied in different cultural domains, in different periods and countries, on different human productions such as scientific papers, patented technologies, or literary novels—or any other products with textual descriptive metadata.

2. Methodology: The computation of the cultural innovation score

The Sentence-BERT (SBERT) algorithm [13, 3, 7, 20] is a robust tool for natural language processing, widely utilized in applications ranging from text classification to information retrieval. SBERT is built upon pre-trained transformer models like BERT and RoBERTa [3, 7], which have been trained on vast text datasets and have found applications in literary text comprehension [5]. Transformer models, with their self-attention mechanism, are adept at weighting the significance of different parts of input data, making them highly effective for tasks such as language translation, text summarization, and question-answering [12, 14].

SBERT excels at computing semantic proximity between words, phrases, or even entire paragraphs. It learns the contextual relationships between words (i.e., word embedding; [9, 10]) and can calculate the semantic similarity between new and existing text based on shared context. This allows SBERT to accurately determine which words, sentences, or paragraphs are most similar to each other, even if they are not exact matches. For instance, in the context of cinema, SBERT could compute the semantic distance between the plot summaries of *Star Wars IV* (1977) and *Star Trek: The Motion Picture* (1979), recognizing their relatedness and scoring them accordingly.

In our measure of cultural innovation, we utilize SBERT to encode descriptions of cultural products into fixed-length vectors that encapsulate the semantic meaning of the descriptive text. These vectors are then used to compute the pairwise cosine similarity between them, serving as a measure of their semantic similarity. The advantage of using SBERT over traditional bag-of-words models is its ability to capture the meaning of the text, rather than just the frequency of words, which is particularly useful when dealing with short texts and texts coming from different periods or different contexts—where words differ although meaning is similar.

To compute our measure of cultural innovation for each individual cultural product, we first encode their description (here, movie plots) into high-dimensional vectors using SBERT. We then compute the cosine similarity between each vector and all previous vectors. That is, we compute the similarity between each product and all previous products. By reversing this score, we transform the resulting similarity values into distance scores. Therefore, the final score, for each product, is computed as *the average of its distance scores from all previous products.* Building on this methodology, it is important to note that our measure inherently captures the increasing difficulty of innovation as more products are released within the same domain. As the number of preceding products grows, the space for unique, unexplored ideas naturally shrinks, making it increasingly challenging to create something truly innovative.

In the domain of movies, here, the innovation score for a given movie thus quantifies how different its summary is from all previous movies in the dataset. This approach allows us to capture the level of novelty of a movie plot compared to the plot of all movies having been previously released.

Our measure of innovation is conceptually similar to other methods that compute innovation by assessing semantic distances between individual cultural products and their predecessors within the same domain. For example, in the realm of French theater, Cafiero and Gabay [1] have employed a similar approach, as have Kelly and colleagues [6] in the analysis of technological patents (see [11], for an application).

We can further formalize this measure. Given a set of cultural products, each with a description D_i, the Innovation Score (IS) for the i-th product is calculated as:

$$IS_i = \frac{1}{i-1}\sum_{j<i}\left(1 - \frac{V_i \cdot V_j}{||V_i|| \cdot ||V_j||}\right)$$

where $V_i = \text{SBERT}(D_i)$ is the high-dimensional vector representation of the description D_i obtained using the Siamese-BERT (SBERT) algorithm, $\cdot$ denotes the dot product, and $||V||$ denotes the Euclidean norm of vector V. IS essentially measures how distinct or innovative a cultural product's description is compared to the descriptions of all the previous products in the set. A higher IS would indicate that the product's description is more unique and innovative within the given set. Conversely, a lower IS would suggest that the product's description shares more similarities with the descriptions of previously seen products.

3. Validity Check: Evaluation of our Measure of Innovation

Our study utilizes a comprehensive dataset compiled from IMDb and Wikipedia, encompassing metadata for 19,254 movies produced in the United States. The IMDb data, obtained directly

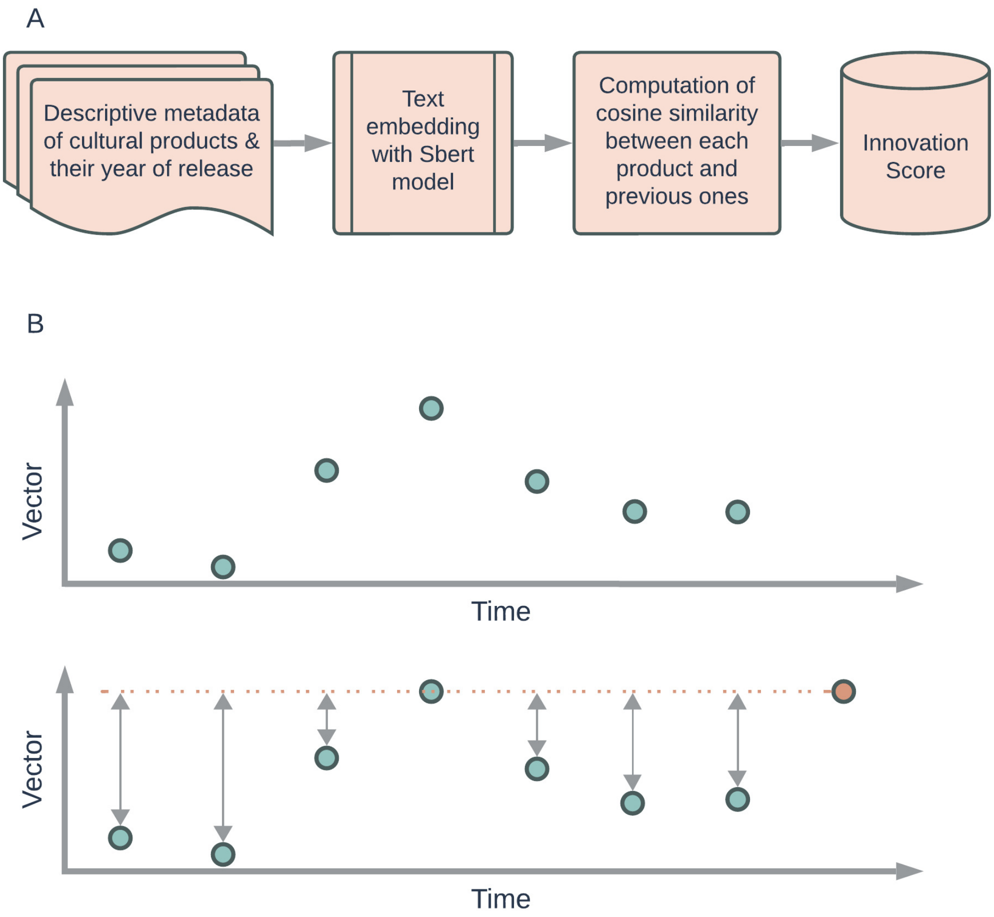

Figure 1: A. Steps for the computation of the innovation score. B. Schematic representation of the computation of the innovation scores for two products. Here, each dot represents an embedded description, with the y-axis being a 1-dimensional vector and the x-axis the time. When we add a new subsequent cultural product, its Innovation Score is computed as the mean of the distances to all previous ones.

from the website, was supplemented with information scraped from individual Wikipedia pages and corresponding Wikidata entries. The consolidated dataset includes a wide array of details such as movie title, release year, director, writer, genre, plot summary, and cast information, providing a robust foundation for our in-depth analysis of cultural innovation in cinema. Here, we compute the Innovation Score for all 19,254 movies, as outlined in the Methodology section.

We conduct a series of tests to check the validity of this measure:

3.1. Robustness Across Different Sources of Description

We turned to IMDb plot summaries, which differ markedly from their Wikipedia counterparts in length and standardization. Despite these differences, our Innovation Score remained consistent, demonstrating a significant positive correlation between the scores derived from both sources ($\beta = .28$, $p < .001$). This result underscores the adaptability of our measure, capable of capturing innovation irrespective of the descriptive metadata available.

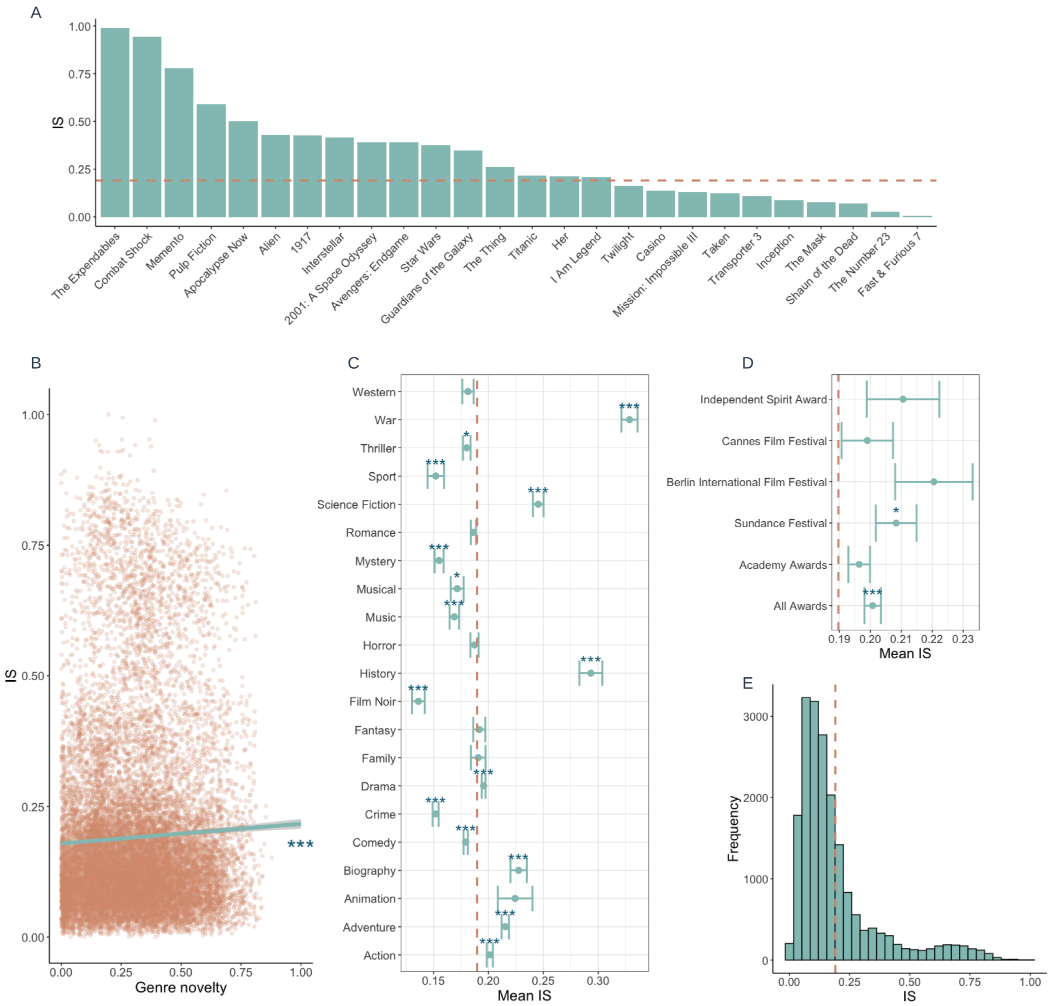

Figure 2: A. Innovation Score (IS) for selected movies. B. Correlation between the Innovation Score and another measure of innovation. C. Mean Innovation Score by genre. D. The mean Innovation Scores of awarded movies. E. Distribution of the Innovation Scores in the dataset. In all figures but B., the orange dotted line represents the average Innovation Scores of all movies.

3.2. Robustness Across Different Random Seeds

To further assess the robustness of our Innovation Score, we altered the random seed used in the calculation process. This analysis consistently revealed a strong positive correlation ($\beta = 0.89$, $p < .001$) between the Innovation Scores obtained using two different random seeds. This substantial correlation reinforces the reliability and stability of our Innovation Scores, demonstrating its consistency even when varying the initial randomization.

3.3. Robustness Across Different Timeframes

The notion of cultural forgetting would suggest that a movie can appear innovative even if its narrative resembles older films, if it is at least different from *recent* ones. To examine this, we calculated three new Innovation Scores, progressively considering a narrower temporal window: we computed the average distance of each movie from movies released 10 years before, 5 years before, and just 1 year before. Strikingly, our analysis revealed that the level of innovation remained nearly identical across these different timescales (for all correlations, $\beta > .98$, $p < .001$). This suggests that a movie's innovativeness is a consistent trait, irrespective of the specific timeframe under scrutiny.

3.4. Qualitative Observation

Upon manual inspection, we found that our algorithm indeed identified movies widely acclaimed as innovative, such as *2001: A Space Odyssey*, *Pulp Fiction*, and *Interstellar*, as highly innovative (see Figure 2.A.). However, it is important to note that these are cherry-picked examples, and we could have chosen others that would not have aligned with our intuition. For instance, while *Inception* is also widely considered innovative, it received a relatively low Innovation Score in our analysis. This qualitative examination serves as an initial validation to check that some scores fit our intuitions.

3.5. Correlation with Another Measure of Innovation

To ensure the robustness of our measure beyond qualitative inspection, we compare our Innovation Score with a Novelty Score derived from a method proposed by Luan and Kim [8], which gauges the uniqueness of a movie's genre combination relative to preceding films. This Novelty Score is genre-based: it rewards films introducing rare combinations of genres. We used this Novelty Score as a benchmark to evaluate the effectiveness of our Innovation Score in capturing a movie's deviation from genre conventions. As anticipated, we found a significant positive correlation between our Innovation Score and the genre-based Novelty Score ($\beta = .04$, $p < .001$), bolstering the external validity of our measure (see Figure 2.B.).

3.6. Correlation with Movie Genres

We conducted multiple two-sample t-tests comparing the aggregated Innovation Scores of movies within a specific genre to those outside of it (with Bonferroni correction for multiple testing). This test allowed us to discern whether there was a significant difference in the mean level of innovation between movies belonging to a genre and all the other ones (see Figure 2.C.). Genres that encompass formulaic narrative plots, such as Film Noir, Mystery, Crime, Sport, and Thriller, exhibit lower average Innovation Scores. This phenomenon arguably occurs because, to belong to these genres, a given movie needs to adhere to specific narrative conventions. In contrast, genres like Adventure, Action, and History, characterized by non-specific themes, allow for greater innovation because they accommodate a broader spectrum of narratives that can deviate from traditional storytelling structures. While it may seem counterintuitive at first, the high average Innovation Scores in War films can be attributed to their ability to draw from various historical events and kinds of warfare. Science Fiction's high average Innovation Scores can be attributed to its futuristic focus and the inherent audience expectation for novelty in Science Fiction movies.

3.7. Correlation with Awarded Movies

Awarded movies have higher average Innovation Scores. Building on the common intuition that award juries, who are cinema experts, tend to reward innovation in cinema, we sought to investigate the relationship between our Innovation Scores and movie awards. We created a binary variable indicating whether a movie had won at least one award, based on mentions in Wikipedia pages, and conducted a two-sample t-test to compare the innovation levels of award-winning movies and those without awards. This analysis was extended to specific awards like the Academy Awards, Sundance Awards, and the Palme d'Or (with Bonferroni correction for multiple testing). Our findings revealed that, in general, award-winning movies tend to have higher Innovation Scores (see Figure 2.D.). Notably, movies recognized at the Sundance Film Festival, known for its focus on innovative independent films, were associated with higher innovation levels. However, this correlation was not observed for other specific awards, such as the Independent Spirit Awards. This discrepancy could underlie the idea that our Innovation Score captures a specific type of *narrative* innovation, which may not align with the criteria used by all award bodies. Nevertheless, the overall trend supports the validity of our Innovation Score as a measure of innovation in cinema.

4. Results: The Evolution of Innovation in Cinema

In our exploration of the temporal dynamics of innovation in cinema, we aggregated the Innovation Scores by year and conducted a series of regression analyses. We fitted three models: a linear model, a quadratic model, and a logarithmic model, each with the mean Innovation Score as the dependent variable and the year as the independent variable.

The linear model, which assumes a constant rate of change in innovation over time, accounted for approximately 18% of the variance in the data. The logarithmic model, which posits a decelerating rate of innovation, performed similarly to the linear model, explaining approximately 18% of the variance. However, the quadratic model, which allows for a changing rate of innovation, performed significantly better, explaining about 28% of the variance. To further compare these models, we calculated the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), both of which balance the goodness-of-fit of a model with its complexity. Lower values of AIC and BIC indicate a better model. The quadratic model outperformed the other two models on both criteria, further supporting its superiority (Figure 3.A.).

The quadratic model suggests a non-linear relationship between time and innovation in cinema. The positive linear term in the model indicates an overall increase in innovation over the years, while the negative quadratic term (lower in magnitude) suggests a slowing down of this increase (Figure 3.B.). Specifically, by observing the plot of the fitted quadratic model, we can infer that innovation in cinema experienced a surge throughout the 20th century. However, this rate of increase appears to have decelerated and reached a plateau in recent years, indicating a stabilization of innovation levels in the cinematic landscape.

In addition to our regression analyses, we wanted to explore to what extent the production of films over the course of history deviates from what it would have been if the films had appeared randomly over time. We used a Monte Carlo simulation approach. This simulation involved generating 1000 datasets, each with movies randomly shuffled across different years while keeping the number of movies per year constant. The results of this Monte Carlo simulation are striking. They reveal a decrease in the average Innovation Score during the initial years, followed by a relatively constant, lower level of innovation per year. The reason is straightforward: the more films already exist, the more difficult it is to innovate on a purely random basis. This pattern contrasts sharply with the actual data, which showed an increase in innovation throughout the last century.

This suggests that if movies were randomly distributed in time, without consideration of what came before, they would produce movies that share similarities with previous eras purely by chance, leading to a flatter and lower average Innovation Score. In contrast, in the real dataset, they seem to actively strive to innovate and differentiate their work from what has come before, leading to an overall increase in innovation over time. This observation, therefore, highlights the strong connection between movie production, creativity, and the influence of past cinematic trends. Filmmakers draw from the past while attempting to break away from prevailing storylines, resulting in the observed increase in innovation in the real dataset as opposed to the simulated ones.

5. Conclusion

Our analysis of over 19,000 movies spanning more than a century has yielded fascinating insights into the trajectory of cinematic innovation. We observed a significant increase in innovation throughout the 20th century, underscoring the era's reputation as a period of rapid creative evolution. Thus, contrary to the often-voiced lament that cinema is losing its innovative edge, our study suggests that the level of innovativeness in cinema is not in decline. In fact, according to our model, the level of innovation today is as high as it was during the golden era of cinema in the 1950s. This implies that the use of formulaic plots is not more prevalent now than it was in the past.

A

	R2 (adj.)	AIC	BIC
Linear model	.18	-470.1614	-462.4684
Logarithmic model	.18	-470.3755	-462.6824
Quadratic model	**.28**	**-480.9902**	**-471.7328**

B

	OLS regression with the mean Innovation Score per year as the outcome variable		
	Estimate	**Std. Error**	**P-value**
Intercept	-42.370289009	11.028485216	< .001 ***
Year	0.042829573	0.011189696	< .001 ***
Year^2	-0.000010773	0.000002838	< .001 ***

C

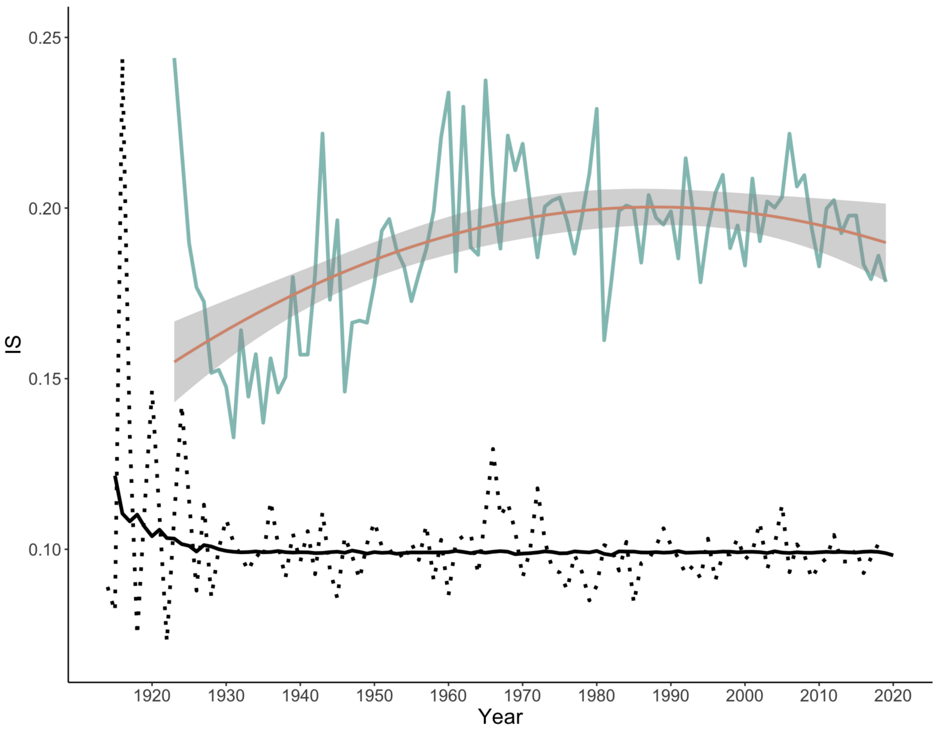

Figure 3: A. Adjusted R2, AIC, and BIC for the three models. B. Output of the quadratic model that showed the best fit to the data C. Evolution of the Innovation Score of movies across time, in blue, with the fitted model in orange. The black dotted line represents one of the 1000 simulations with movies randomly shuffled across different years while keeping the number of movies per year constant. The black straight lines represent the average of all 1000 simulations.

References

[1] F. Cafiero and S. Gabay. “Rise and Fall of Theatrical Genres in Early Modern France: a Centroid-Based Approach”. In: *Publisher: arXiv* (2023).

[2] B. d. Courson, V. Thouzeau, and N. Baumard. “Quantifying the scientific revolution”. In: *Evolutionary Human Sciences* 5 (2023), e19. doi: 10.1017/ehs.2023.6. url: https://www.cambridge.org/core/journals/evolutionary-human-sciences/article/quantifying-the-scientific-revolution/60249C6B9DF636D2EC8446F6B7E454F8.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: (2018). doi: 10.48550/arxiv.1810.04805. url: https://arxiv.org/abs/1810.04805.

[4] S. P. Fraiberger, R. Sinatra, M. Resch, C. Riedl, and A.-L. Barabási. “Quantifying reputation and success in art”. In: *Science* 362.6416 (2018), pp. 825–829. doi: 10.1126/science.aau7224. url: https://www.sciencemag.org/lookup/doi/10.1126/science.aau7224.

[5] T. He, F. Breithaupt, S. Kübler, and T. T. Hills. “Quantifying the retention of emotions across story retellings”. In: *Scientific Reports* 13.1 (2023), p. 2448. doi: 10.1038/s41598-023-29178-8. url: https://www.nature.com/articles/s41598-023-29178-8.

[6] B. Kelly, D. Papanikolaou, A. Seru, and M. Taddy. “Measuring Technological Innovation over the Long Run”. In: *American Economic Review: Insights* (2021).

[7] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. “RoBERTa: A Robustly Optimized BERT Pretraining Approach”. In: *arXiv:1907.11692 [cs]* (2019). eprint: 1907.11692. url: http://arxiv.org/abs/1907.11692.

[8] Y. Luan and Y. J. Kim. “An integrative model of new product evaluation: A systematic investigation of perceived novelty and product evaluation in the movie industry”. In: *PloS One* 17.3 (2022), e0265193. doi: 10.1371/journal.pone.0265193.

[9] T. Mikolov, K. Chen, G. Corrado, and J. Dean. “Efficient Estimation of Word Representations in Vector Space”. In: *Publisher: arXiv Version Number: 3* (2013). doi: 10.48550/arxiv.1301.3781. url: https://arxiv.org/abs/1301.3781.

[10] J. Pennington, R. Socher, and C. Manning. “Glove: Global Vectors for Word Representation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1532–1543. doi: 10.3115/v1/D14-1162. url: http://aclweb.org/anthology/D14-1162.

[11] M. Posch, J. Schulz, and J. Henrich. “Surname Diversity, Social Ties and Innovation”. In: *SSRN Electronic Journal* (2023).

[12] Y. Qi, D. Sachan, M. Felix, S. Padmanabhan, and G. Neubig. “When and Why Are Pre-Trained Word Embeddings Useful for Neural Machine Translation?” In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). New Orleans, Louisiana: Association for Computational Linguistics, 2018, pp. 529–535. DOI: 10.18653/v1/N18-2084. URL: http://aclweb.org/anthology/N18-2084.

[13] N. Reimers and I. Gurevych. “Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, 2020, pp. 4512–4525. DOI: 10.18653/v1/2020.emnlp-main.365. URL: https://www.aclweb.org/anthology/2020.emnlp-main.365.

[14] N. Reimers and I. Gurevych. *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. 2019. DOI: 10.48550/arXiv.1908.10084. arXiv: 1908.10084[cs]. URL: http://arxiv.org/abs/1908.10084.

[15] M. Schich, C. Song, Y.-Y. Ahn, A. Mirsky, M. Martino, A.-L. Barabási, and D. Helbing. “Quantitative social science. A network framework of cultural history”. In: *Science (New York, N.Y.)* 345.6196 (2014), pp. 558–562. DOI: 10.1126/science.1240064.

[16] M. Serafinelli and G. Tabellini. “Creativity over time and space”. In: *Journal of Economic Growth* 27.1 (2022), pp. 1–43. DOI: 10.1007/s10887-021-09199-6. URL: https://doi.org/10.1007/s10887-021-09199-6.

[17] S. Sreenivasan. “Quantitative analysis of the evolution of novelty in cinema through crowdsourced keywords”. In: *Scientific Reports* 3.1 (2013), p. 2758. DOI: 10.1038/srep02758. URL: http://www.nature.com/articles/srep02758.

[18] A. Tacchella, A. Napoletano, and L. Pietronero. “The Language of Innovation”. In: *Plos One* 15.4 (2020), e0230107. DOI: 10.1371/journal.pone.0230107. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230107.

[19] F. Tria, V. Loreto, V. D. P. Servedio, and S. H. Strogatz. “The dynamics of correlated novelties”. In: *Scientific Reports* 4.1 (2014), p. 5890. DOI: 10.1038/srep05890. URL: https://www.nature.com/articles/srep05890.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. “Attention Is All You Need”. In: *Publisher: arXiv Version Number: 5* (2017). DOI: 10.48550/arxiv.1706.03762. URL: https://arxiv.org/abs/1706.03762.