Genres, Subgenres, and Storytelling Tropes:
a Data Science Approach

What's the difference between a first-person shooter and an action-adventure video game? How are Korean dramas different from Japanese dramas? How are genres defined? Whether they are top-down impositions such as "Dark Suspenseful Gangster Dramas" (an actual genre on Netflix), or emerging musical categories such as hyperpop, genres are fundamental to how we navigate the world, and to the role that popular media plays in shaping and reflecting society.

Our research aims to answer the question: do the genres as defined by people match the genres emerging from the data? Through our data-driven analysis, we hope to provide new insights into the nature of genre and its role in shaping and reflecting culture. To do so, we draw from TVTropes and RAWG, sources of crowdsourced metadata on TV shows and videogames respectively, clustering artworks based on their associated tropes and tags and comparing these clusters to official genre classifications in sources such as IMDb. We then apply a combination of approaches from cultural evolution and stylometry, along with NLP methods such as topic modelling, sentence embedding, and stochastic block models. In particular, we use topSBM, a nonparametric stochastic block model-based topic model introduced by Gerlach et al. [1], to identify a hierarchy of genres and subgenres, as well as a hierarchy of storytelling tropes present in the data over time.

Our analysis reveals the similarities discrepancies between official genre classifications in IMDb and RAWG and the genres inferred from our data. Figure 1 shows the evolution of TVTropes and RAWG genres over time, in the top and bottom panels respectively. The analysis reveals discrepancies between official genre classifications in IMDb and RAWG and the genres inferred from our data. This data-driven approach suggests that official genre classifications often fall short in capturing the full range of subgenres in popular media. Our findings underline the potential of data science methods in understanding the evolution and nuances of cultural genres.
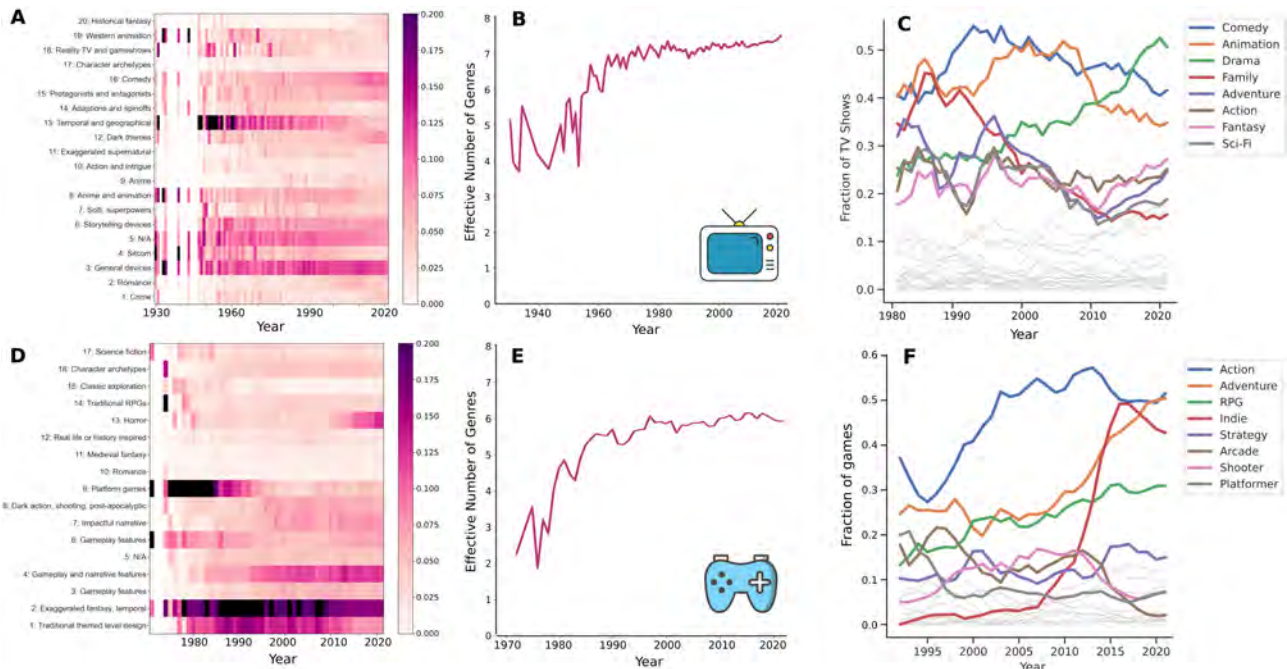

Figure 1: Genres over time, for TV (top) and video games (bottom). Panels (a) and (d) show heatmaps indicating topic frequency over time, panels (b) and (e) show the effective number of issues over time, and panels (c) and (f) show changes the frequencies of IMDb and RAWG genres.

References

[1] Martin Gerlach, Tiago P Peixoto, and Eduardo G Altmann. A network approach to topic models. *Science advances*, 4(7), 2018.