A Novel Application of Bayesian Network Structure Learning to a Dataset of Fictional Exoplanets

Emma Johanna Puranen[1] **Emily Finer**[1] **Christiane Helling**[1] **V Anne Smith**[1]

[1]Centre for Exoplanet Science, University of St Andrews, St Andrews, United Kingdom

Abstract

A novel interdisciplinary approach applies Bayesian network analysis to a database of fictional exoplanets, or planets outside our solar system. The aim is to find connections between fictional characteristics of the planets, such as whether they support life, and astronomically-relevant pieces of information, such as whether the media featuring the exoplanet was published before or after the discovery of real exoplanets in 1995, to map the relationships between science and science fiction world-building. Banjo is used for network structural analysis, and the network is visualised with GraphViz and BayesPiles. Influence scores show that fictional exoplanets designed after the discovery of real exoplanets are slightly less likely to be in the liquid-water habitable zone, and moderately less likely to host intelligent life. This mirrors the real-life discoveries of a diverse range of exoplanets, many of which are far from human-habitable.

1 EXTENDED ABSTRACT

1.1 INTRODUCTION

Since the first discovery in 1995 [Mayor and Queloz, 1995], more than 4000 exoplanets have been discovered by astronomers. Information such as mass, radius, and length of year are available for many of these worlds, and exoplanet discoveries have changed long-standing models of solar system formation [NASA Exoplanet Archive, 2021]. For example, it had previously been thought that gas giants could not orbit close to their stars, but exoplanet data show that such "hot Jupiters" are actually common. Exoplanet science is now moving from discovery to characterisation, with plans to use upcoming Extremely Large Telescopes to determine the chemical makeup of exoplanet atmospheres [Kaltenegger, 2017].

Science fiction (SF) creators have been imagining fictional exoplanets since at least the 1600s [Cavendish, 1668], and in great volume since the 20th century. As a result, the general public is more familiar with fictional exoplanets than with real ones. Fictional exoplanets therefore present an interesting topic of study in the fields of science communication and science literacy. SF creators and scientists take inspiration from each other in a continuous dialogue; just as current science provides ideas for SF creators, SF creators inspire people to become scientists. They also model *what-if* scenarios through their stories, addressing how scientific ideas and inventions and discoveries might affect humanity. However, SF scholarship usually takes the form of case studies and close readings. Instead, we present a big data, digital humanities approach to SF scholarship, deliberately abstracting a large amount of data to find patterns and understand how the interplay between SF and science affects the construction of fictional exoplanets.

1.2 METHODS

Bayesian network structure learning with Banjo [Yu et al., 2004] was performed on a dataset of 8 variables (characteristics of fictional exoplanets) with 63 observations (fictional exoplanets), representing whether the planet:

BeforeAfter: first appeared in fiction before (0) or after (1) the discovery of real-life exoplanets in 1995. ***HabZone***: is in the liquid-water habitable zone (1) or outside it (0). ***RealStar***: is portrayed as being part of a real star system (1) or not (0). ***Life***: is home to native life (1) or not (0). ***Intelligent***: is home to intelligent native life (1) or not (0). ***HumansBreathe***: has an atmosphere human characters can breathe without ill effect (1) or not (0). ***MediaType***: originally appeared in a film (0), book (1), TV show (2), video game (3), or podcast (4). ***PlanetType***: is Earth-like and rocky/terrestrial (0) or Jupiter-like and gaseous (1).

Accepted for 15th Bayesian Modeling Applications Workshop at UAI 2021 (BMAW2021).

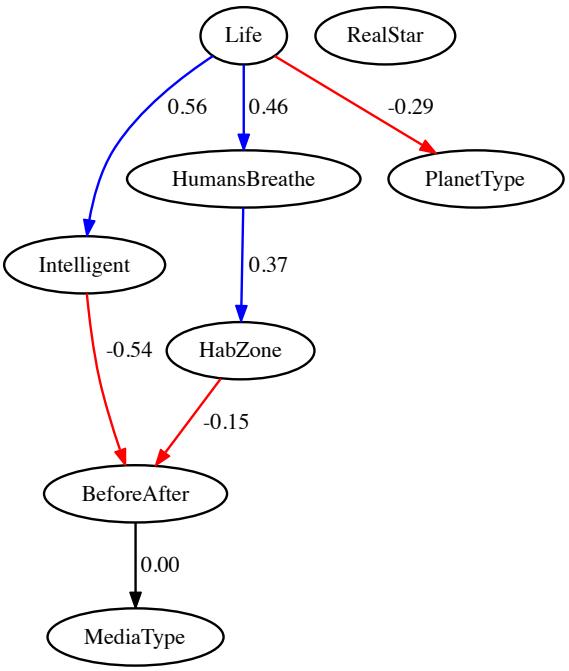

Figure 1: Top-scoring network in GraphViz [Gansner and North, 2000]. Links are labelled with influence scores; blue links are positive, red negative, and black non-monotonic.

Visualisation via BayesPiles [Vogogias et al., 2018] showed multiple 10-minute Greedy searches produced top networks of the same equivalence class, thus a single top network was chosen. Banjo provides influence scores showing magnitude (0 to 1) and sign $(+/-)$ of influence between variables, with 0.0 representing a non-monotonic influence (e.g., medial categories related to high/low states) [Yu et al., 2004].

1.3 RESULTS

The Bayesian network in Fig. 1 shows which variables were found to influence each other. As expected, Life, HumansBreathe, Intelligent, and HabZone all form a chain of connections – a planet cannot have intelligent life if it has no life at all.

It had been hypothesised that whether or not an SF creator used a real star system for their exoplanet might reflect their familiarity with current scientific literature, and perhaps be negatively related to HumansBreathe as a proxy for lower scientific accuracy in a work. However, this is not the case, as RealStar is not connected to any other variables.

We are most interested in the variable BeforeAfter, which moderately (-0.54) negatively influences Intelligent, meaning that there have been fewer portrayals of intelligent native life on fictional exoplanets created since 1995. BeforeAfter has a weaker negative influence (-0.15) to HabZone, meaning fewer fictional exoplanets have been in the habitable zone since 1995.

1.4 CONCLUSION

These preliminary results begin to paint a picture of how the portrayal of fictional exoplanets has changed with the discovery of real exoplanets. Before the discovery of real exoplanets, astronomers had very little idea what to expect, having only our own solar system as an example. Creators of SF, telling stories that naturally tend to centre the human experience, portray very Earth-like exoplanets, with familiar gravities and atmospheres and lengths of day, and recognisable life forms. These worlds dominate our database of fictional exoplanets, both before and after the discovery of exoplanets. Certainly, there are some very inhospitable fictional worlds, including some developed before 1995, like Larry Niven's Jinx, for example. Overall, however, the negative influences between BeforeAfter and HabZone and BeforeAfter and Intelligent show that fictional exoplanets have become less human-hospitable and less Earth-like over time. This has mirrored the huge plethora of real exoplanet discoveries. In recent decades, the discoveries of increasingly weird worlds – hot Jupiters trailing comet tails of atmosphere, worlds that rain diamonds, and worlds that orbit multiple stars – have become common headline material in popular science articles, and more familiar to SF creators. This is shown in their more prominent fictional portrayals.

References

M Cavendish. *The Description of a New World, Called the Blazing-World*. 1668. URL `https://digital.library.upenn.edu/women/newcastle/blazing/blazing.html`.

ER Gansner and SC North. An open graph visualization system and its applications to software engineering. *Sofware–Practice and Experience*, 30:1203–1233, 2000.

L Kaltenegger. How to characterize habitable worlds and signs of life. *Annual Review of Astronomy and Astrophysics*, 55:433–485, 2017.

M Mayor and D Queloz. A jupiter-mass companion to a solar-type star. *Nature*, 378:355–359, 1995.

NASA Exoplanet Archive, 2021. URL `https://exoplanetarchive.ipac.caltech.edu/`.

A Vogogias, J Kennedy, D Archambault, B Bach, VA Smith, and H Currant. Bayespiles: Visualisation support for Bayesian network structure learning. *ACM TIST*, 10:5, 2018.

J Yu, VA Smith, PP Wang, AJ Hartemink, and ED Jarvis. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, 20:3594–3603, 2004.