9th International Conference on Higher Education Advances (HEAd'23)
Universitat Politècnica de València, València, 2023
DOI: http://dx.doi.org/10.4995/HEAd23.2023.16175

Digital pedagogy for the present: an artificial intelligence methodology for curriculum development

Michelle Rufrano[1], Jean-Ezra Yeung[2]
[1]Department of Sociology & Anthropology, Fordham University, U.S.A., [2]Columbia Startup Lab, Columbia University, U.S.A.

Abstract

In an age of information, course creation at the university-level poses new challenges in that the constant influx of new data makes it difficult for instructors to update courses with the latest research. Using Artificial Intelligence (AI), we created a curriculum that matches the contemporary experiences of students. This approach was used to help design a course on Aging and Adulthood using the life course model. We were able to create a topological map of topics and subtopics that provided the instructor with the ability to quickly understand the shape of the data, thematic connections, and data voids. Choosing from a myriad of research articles from a derived skeleton of topics, the instructor was then able to design a unique course with the most relevant research and supplement the syllabus with seminal works in the field.

Keywords: *Artificial intelligence; topological data analysis; digital pedagogy; curriculum development; life course; epigenetics.*

1. Introduction

The first time I traveled to Spain I was 20 years old and when I left I cried all the way to the airport. My señora was my best Spanish teacher. We would pantomime and dance to make our meanings known. We needed no words when we parted, on her face was the sadness that can only follow the great joy of being forever changed by someone significant—linked lives.

Linked lives is one of the five characteristics of the life course model. A framework that articulates the cumulative effects of our life experiences over time (Elder Jr., 1998, p. 4). A tool that can help us identify vulnerable periods within the context of historical events and create opportunities for informed interventions (Elder Jr., 1998, p. 5). Inspired by my own life experiences, I endeavored to teach a course entitled *Coming of Age: Adulthood* grounded in the life course model. However, the life course model is a novel approach that has only become part of public health curricula as of the last decade (Begg et al., 2015, p. 1). Therefore, a meaningful course would not only need to cite seminal works in the field, and discuss the key innovations in epigenetics, but also be relevant to students. I have now successfully taught *Coming of Age: Adulthood* twice due to the application of artificial intelligence in outlining aging.

Using Artificial Intelligence (AI), the curriculum was up-to-date even given the COVID-19 pandemic and the recent explosion of research in the fields of the life course model, epigenetics, and exposome. Almost half (60,000+) of the articles (130,000+) containing the search term "epigenetics" were published in PubMed during the last five years. The preparation for this course was cut down to a few hours, only requiring a quick 15-minute update from semester to semester.

2. Methodology

Using PubMed, a biomedical literature database accessed at https://pubmed.ncbi.nlm.nih.gov/, the researchers extracted articles over 365 days ending on December 8th, 2021, using search terms from a life course glossary curated by the researchers. They used their domain expertise to determine the set of terms and boolean operators (and/or). Data were extracted using the Biopython API (Cock et al., 2009) and transformed data was loaded into a custom-built interactive Python Dash web app (Hossain, 2019) that leveraged topological data analysis. This included generating a topological network (Singh et al., 2007), computing a persistence diagram (Zomorodian & Carlsson, 2004), and identifying the representative cycle (Tralie et al., 2018). Additionally, a knowledge graph was constructed to supplement topological understanding by seeing the conceptual relationships involved in different areas of the topological network.

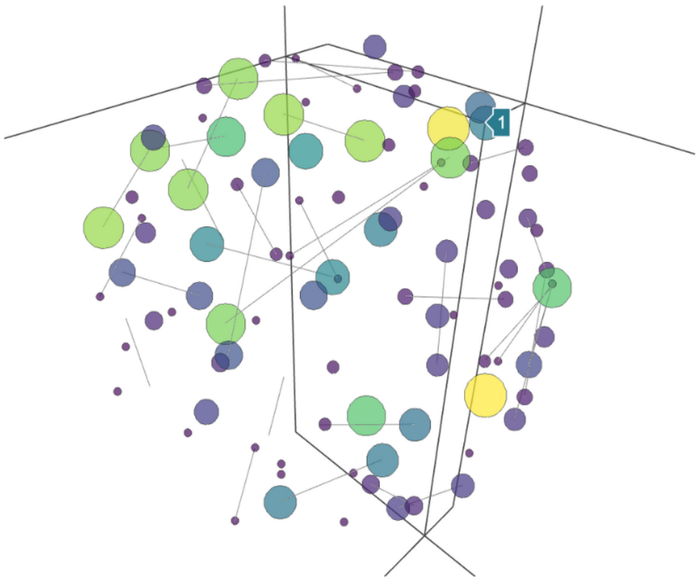

Figure 1. Topological network of Pubmed articles on the life course model

The procedure used to extract topics and sub-topics was a multi-step process. First, the knowledge graph was used to identify the key concepts and their relationships. Those concepts were mapped to the topological network's clusters, which provided a rough segmentation of articles into topic or sub-topic groups (Figure 1). In the topological network, lines between nodes represent connectivity and similarity, and each node represents a set of articles. The node color represents its similarity index, which is any function that summarizes its relative distances from other data points into a single relative distance value. To determine whether a cluster is a strong signal, the persistence diagram visualizes the overall distribution of hierarchical clustering, so that we can determine how many topics are likely and when a topic becomes saturated upon initial or repeated identification (Figure 2). In computing the persistence diagram, we also extracted the representative cycle to determine any persistent topological signal, as it highlights the most salient relationships between concepts across articles. Upon discussion and review of an exported comma separated file (.csv) or Excel file (.xlsx) containing article metadata and their cluster assignments, we determined topics and selected works to include in the curriculum. Topics were labeled, created, modified and/or collapsed to simplify the organization. Altogether, our AI tool extracts the shape of the research topics without having to manually sort through article titles and abstracts. Thus, this framework readily provides a skeleton of topics, subtopics, and their relationships.

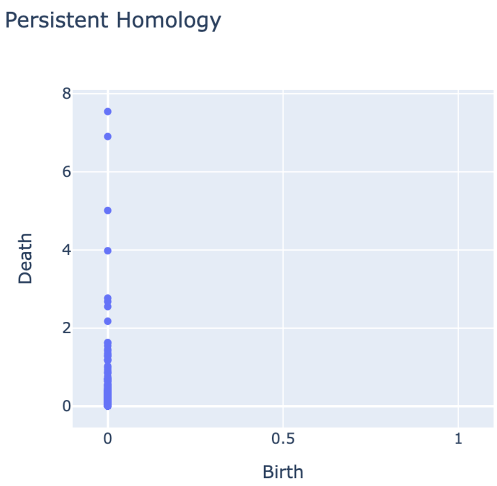

Figure 2. Persistent Diagram of PubMed articles

3. Findings

In the first semester teaching the course, there were 11 topics and some selected subtopics (Figure 3). As a semester is around 14 weeks at Fordham University, roughly one topic can be used for each week and larger or more complex topics are split over two weeks (e.g. Mental Health).

1. Disaster
   - History of pandemics
   - Syndemic
   - Sociology and aging: Adulthood
   - COVID-19 crisis effects
2. Epigenetics
   - DNA methylation
   - Exposome
3. Adverse experiences
   - Childhood
   - Adulthood
4. Mental Health
   - Digital mental health interventions
   - Suicide mortality
   - Opioid
     - Adverse childhood experiences, use behavior
     - Burden of substance use
5. Work
   - Among women
   - Work shape
   - School-to-work transition
   - Complex etiology
   - Employment transitions & weight gain
6. Lifestyle
   - Unpaid work Norwegian seniors
7. Women
   - Intimate partner violence
   - Sustained viremia
   - Maternal
     - Preterm births
8. Discrimination
   - Coping mechanisms
   - Racial
     - Stroke, mortality
     - Cardiovascular
     - Incivility, everyday life
     - Racial configuration, parental couples
     - Sexual minority adults across diverse racial
9. Life expectancy
   - Dementia status
10. Linked lives
    - Estrangement
11. Old age, Elderly, Frailty, Dementia, Alzheimer's, Sarcopenia

Figure 3. List of Topics and Subtopics

With the topic list set based on the topology of the life course, the instructor was then able to select readings from each topic, resulting in a comprehensive, up-to-date syllabus. The instructor supplemented a few seminal works to help guide the discussion of the significance of this research from a social context. Overall, the total time for the course creation was roughly two days.

The topological tools also allow researchers to recognize topic connectivity and circularity (identified as loops) as well as research voids. The instructor can use these patterns to discuss the relationships and limitations of data sources with students. The topological network of aging and the life course model clearly demonstrated that understudied groups such as women, racial and ethnic minorities, people who are differently abled or sexual minorities, etc. are often left out of the dominant discourse of research—a significant research void.

4. Discussion

The utilization of AI expedited the process of course development with the most updated research/ readings and positively impacted students. The individual articles covered a myriad of topics relevant to the contemporary experience of students. So much so that discussion easily flowed from week to week as we contended with studies on climate change, early onset ADHD, longitudinal effects of early life adversities (ELA), epigenetic processes, and maternal health from all over the world (Vergunst & Berry, 2022; Nigg et al., 2020; Suglia et al., 2021; Simons et al., 2021; Vedam et al., 2019).

Students' reactions were overwhelmingly positive, even saving articles to help them discuss the realities of mental health with parents or the longitudinal effects of pre-term births with personal doctors (Brenner & Bhugra 2020; Heikkilä et al., 2021).

Every semester students evaluate their courses and below are two anonymous statements from Fordham University Student Course Evaluations (2022) that specifically spoke to the course content:

> I really appreciated the way the course was setup–up– with us reading current studies that then relate to the following lecture. The studies having been applicable to current events made the class extremely engaging and interesting.
>
> I really liked the presentations, they were informative and opened my mind up to topics and issues I hadn't really learned about before.

Here are two direct emails from students after the course (2022):

> Thank you for such an incredible semester of thought-provoking discussions and readings. I truly learned so much

> Taking your class was one of my most fulfilling experiences at Fordham because the discussions and discourse we were able to have amongst peers was a chance to explore deeper thoughts about what shapes us as human beings and what effects are present amongst us in school communities, relationships, work environments, and the world in general.

As an instructor, I seek to teach the courses that inspire this kind of passion in my students, and while I may be an expert in my respective fields of sociology and public health—I do not know everything. However, with the help of this AI-powered methodology, I can enter every semester confident that the research studied in my courses can better prepare my students for life.

Novel data techniques and tools are no longer required just for data scientists but for higher education instructors seeking to maintain relevance and clarity in a rapidly evolving information landscape. Big data is ubiquitous—its overflow causes information fatigue syndrome, a weakening of our ability to think and discern (Han, 2017, p. 60). IBM defines characteristics of big data as having high levels of volume, velocity, and variety ("Big Data Analytics", 2023). The course taught required deconstructing text data from article titles and abstracts. Typically it would take months to manually extract topics and relationships with moderate success in identifying complex patterns. Furthermore, findings could become outdated during the months of preparation.

To maintain relevance in the present, the COVID-19 pandemic prompted us to integrate other sources of unstructured data (e.g. social media and government sources). Therefore, the methodology is adaptable to any text data source and can be used in other contexts requiring timely and significant research updates, such as the escalating impact of climate change, addressing evolving social crises in psychotherapy, and identifying misinformation in new media.

The limitation of this methodology is that the results are not repeatable as they require an individual to interpret the topological diagrams and knowledge graph to generate findings. However, the expectation is that the user is knowledgeable in the field and that they bring their expertise and lens to unveil an original research narrative.

References

Begg, M.D., Fried, L.P., Glover, J.W., Delva, M., Wiggin, M., Hooper, L., Saxena, R., De Pinho, H., Slomin, E., Walker, J.R. and Galea, S. (2015). Columbia public health core curriculum: short-term impact. *American journal of public health*, *105*(12), e7-e13. https://doi.org/10.2105/AJPH.2015.302879.

Big Data Analytics | IBM. (n.d.). Retrieved from https://www.ibm.com/analytics/big-data-analytics.

Brenner, M. H., & Bhugra, D. (2020). Acceleration of anxiety, depression, and suicide: secondary effects of economic disruption related to COVID-19. *Frontiers in psychiatry*, *11*, 1-21. https://doi.org/10.3389/fpsyt.2020.592467

Cock, P. J Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009).

Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422-1423. https://doi.org/10.1093/bioinformatics/btp163

Elder Jr, G. H. (1998). The life course as developmental theory. *Child development*, *69*(1), 1-12. https://doi.org/10.1111/j.1467-8624.1998.tb06128.x

Han, B. C. (2017). *In the swarm: digital prospects* (Vol. 3). MIT Press. https://doi.org/10.7551/mitpress/11055.001.0001

Heikkilä, K., Pulakka, A., Metsälä, J., Alenius, S., Hovi, P., Gissler, M., Sandin, S. & Kajantie, E. (2021). Preterm birth and the risk of chronic disease multimorbidity in adolescence and early adulthood: A population-based cohort study. *Plos one*, *16*(12), 1-15. https://doi.org/10.1371/journal.pone.0261952

Hossain, S. (2019). Visualization of bioinformatics data with dash bio. C. Calloway, D. Lippa, D. Niederhut, & D. Shupe (Eds.), In *Proceedings of the 18th Python in Science Conference*. Austin, Texas, 126–133. https://doi.org/10.25080/Majora-7ddc1dd1-012

Nigg, J. T., Sibley, M. H., Thapar, A., & Karalunas, S. L. (2020). Development of ADHD: Etiology, heterogeneity, and early life course. *Annual review of developmental psychology*, *2*, 559-583. https://doi.org/10.1146/annurev-devpsych-060320-093413

Simons, R. L., Lei, M. K., Klopach, E., Berg, M., Zhang, Y., & Beach, S. S. (2021). (Re) setting epigenetic clocks: An important avenue whereby social conditions become biologically embedded across the life course. *Journal of Health and Social Behavior*, *62*(3), 436-453. https://doi.org/10.1177/00221465211009309

Singh, G., Mémoli, F., & Carlsson, G. E. (2007). Topological methods for the analysis of high dimensional data sets and 3d object recognition. *PBG@ Eurographics*, *2*, 091-100. http://dx.doi.org/10.2312/SPBG/SPBG07/091-100

Suglia, S. F., Appleton, A. A., Bleil, M. E., Campo, R. A., Dube, S. R., Fagundes, C. P., Heard-Garris N.J., Johnson, S. B., Slopen, N., Stoney, C. M., & Watamura, S. E. (2021). Timing, duration, and differential susceptibility to early life adversities and cardiovascular disease risk across the lifespan: Implications for future research. *Preventive medicine*, *153*, 1-8. https://doi.org/10.1016/j.ypmed.2021.106736

Tralie, C., Saul, N., & Bar-On, R. (2018). Ripser.py: A lean persistent homology library for python. *Journal of Open Source Software*, *3*(29), 925. https://doi.org/10.21105/joss.00925

Vedam, S., Stoll, K., Taiwo, T. K., Rubashkin, N., Cheyney, M., Strauss, N., McLemore, M., Cadena, M., Nethery, E., Rushton, E. Schummers, L., Declercq, E., & the GVtM-US Steering Council (2019). The Giving Voice to Mothers study: inequity and mistreatment during pregnancy and childbirth in the United States. *Reproductive health*, *16*(1), 1-18. https://doi.org/10.1186/s12978-019-0729-2

Vergunst, F., & Berry, H. L. (2022). Climate change and children's mental health: a developmental perspective. *Clinical Psychological Science*, *10*(4), 767-785. https://doi.org/10.1177/21677026211040787

Zomorodian, A., & Carlsson, G. (2004, June). Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry* (347-356). https://doi.org/10.1145/997817.997870