Pareto distribution in virtual education: challenges and opportunities

La loi de Pareto appliquée à l'enseignement virtuel : défis et perspectives

Kristina Valkanas,[1] Phedias Diamandis[1]

[1]Princess Margaret Cancer Centre & Laboratory Medicine Program, University Health Network, Ontario, Canada

Correspondence to: Phedias Diamandis, MD, PhD, 12-308, Princess Margaret Cancer Centre, University Health Network, 101 College St, Toronto, M5G 1L7; email: p.diamandis@mail.utoronto.ca

In 1906, Italian Economist Vilfredo Pareto observed that 20% of the Italian population, he referred to as the "vital few," owned 80% of the country's wealth.[1,2] While much of our modern statistical training in the natural sciences emphasizes normally distributed phenomena, Pareto's "80/20" rule has continued to define many important aspects of human behavior such as trade, city structures, worker productivity, and relevant to our discussion; how on-demand knowledge is utilized.[2–6] In libraries, for example, roughly 20% of material is said to be responsible for 80% of loans.[5] In academia, 20% of research output receives approximately 80% of citations.[6,7] In July 2019, we, (the authors) started an educational neuroscience initiative on YouTube (NeuroscIQ; www.youtube.com/neuroscIQ) in an attempt to improve dissemination of neuroscience and neuropathology content in an open and scalable manner. During these past 24 months, most of which were during the COVID-19 pandemic, when many other modes of academic learning (e.g. conferences, university lectures) were also being largely transitioned online, we amassed close to 6,000 subscribers and 300,000 total views for content that spanned nearly 100 videos. Reflecting on this initial progress, we noted a Pareto-like distribution with only a handful of our generated content (three to four videos) being responsible for the majority (>80%) of our channel's key output metrics (views, subscribers, watch hours).

While this is only a single and personal experience, we observed some important patterns in this skewed "winner take most" distribution that raise important challenges and opportunities for the emerging field of virtual learning and education. Notably, despite attempts to cater to a neuroscience audience, our six most viewed episodes dealt with topics surrounding the COVID-19 pandemic. Our most viewed non-COVID-19 lectures (e.g. dopamine fasting, hypoxic ischemic injury) had 10-75 times fewer views than the top performing COVID-19-related content (e.g. viral (anosmia) and vaccine-related (autoimmune) complications). Importantly for this discussion, given that most videos were delivered by the same people, we could not attribute these massive discrepancies in community engagement to major differences in platform (YouTube), audience or the teaching styles of the educators. Instead, we believe our experience supports the view that by removing any scheduling, class size, or geographic restrictions to access, a unique feature of online/on-demand virtual education, "topical interest" may become an overwhelming driver of learning engagement and performance metrics using modern search tools. Indeed, even among our neuropathology lectures, massive differences in viewership appeared to be driven by the general public's interest in understanding and addressing conditions affecting close loved ones (e.g., hypoxic ischemic injury). These performance biases in knowledge dissemination and impact created important challenges for us, as a small and growing channel, and will likely create similar pressures as online education grows in our data- and performance-driven society and academic environment. How do educators (content creators) balance topic diversity and their experienced (albeit subjective)

view of what topics should be prioritized? What metrics do we use to design supply when the online demand is dynamically driven by global-sized classrooms? In our experience, the tendency of modern computational algorithms, such as those driving YouTube and Google searches, to optimize viewer retainment through the constant suggestion of similarly themed videos, perhaps at the expense of diversity, further exaggerated Pareto Principle of the "vital few" to the "vital very few" in our online education performance metrics. While powerful at engaging audiences, there are important ramifications that can subconsciously compromise education if not addressed.

Since making this observation, we have made numerous follow-up videos on neurologically-relevant topics relevant to the COVID-19 pandemic (e.g. venous sinus thrombosis, role of the hypothalamus in the fever response). These have invariably done relatively well; quickly and dramatically outperforming our other content. This exponentially positive feedback has been a strong coercive power, which at points, made us reconsider the overall theme of the channel from one focused on neuroscience (our passion and mission), to one optimized around general trending topics and known properties of online search algorithms. While the positive feedback and recognition may make it tempting to exploit these components of online search engines, we urge discipline and a continued focus on areas of interest and expertise to others. Similarly, considering these feedback loops, much effort needs to be invested into finding new ways to measure impact rather than views, citations, and popularity, given the exponential growth properties of the internet.

While we have wrestled with these challenges, there are also many unexpected positives that can be realized from understanding and managing the different properties of online knowledge sharing. The "vital very few" videos we produced, in addition to the views they received, also helped bring a disproportionate number of new subscribers to our educational channel. These individuals provided a baseline audience in which new content from our channel is reliably delivered to and helped bring additional attention to our somewhat very niche lectures that would have otherwise been lost online. Indeed, videos with similar content and presenters garnered four times the number of viewers when delivered just one year later.[8] While this can be partly attributed to improved designs of thumbnails and titles, we believe this largely also stemmed from our gradual growth of a subscriber base. Similarly, by

incorporating timely topics (e.g. Elon Musk's Neuralink) into tradition concepts and topics (e.g. spike sorting), we also observed improvements in interest, viewership and appearances in search. We believe this provides a healthy alternative to improving existing course material and lecture topics than traditional feedback such as course evaluations at the end of conventional courses. Similarly, in a world that requires more cross-pollination of ideas and concepts, databases that allow users to optimize recommendation strategies for distinct topics and content, rather than based on similarity, could be a powerful and transformative tool in education and innovation.

Education has undoubtedly been one of the most significant areas transformed by the COVID-19 pandemic. By accelerating the transition of educational content online, it is important to understand the positive and negative implications for educators. Some of our highly specialized lectures on neuropathology, delivered live to three to four local trainees, have continually and reliably garnered 10 views/day for almost an entire year. This amounts to thousand-fold more impact than ever possible with a single lecture and aims to empower and provide access to education for remote areas. Despite compelling positives, it is important to remember the hyper Pareto distribution-like properties of online education that has the potential to drive divide; with much content reaching almost no one, to others gaining exponential access and influence on a massive audience. This can create very narrow thinking paradigms that can destroy innovation and open mindedness in future generations of trainees and scientists. New approaches for content suggestions, search engine results and reward metrics need to be closely evaluated to ensure knowledge remains diverse and that social and world events do not disproportionately drive our educational framework.

References

1. Lipovetsky S. Pareto 80/20 law: Derivation via random partitioning. *Int J Math Educ Sci Technol*. 2009 Jan;40(2):271–7. https://doi.org/10.1080/00207390802213609

2. Tanabe K. Pareto's 80/20 rule and the Gaussian distribution. *Phys. A: Stat. Mech. Appl.* 2018 Nov 15;510:635–40. https://doi.org/10.1016/j.physa.2018.07.023

3. Néda Z, Varga L, Biró TS. Science and Facebook: the same popularity law! *PLoS ONE*. 2017 Jul 1;12(7). https://doi.org/10.1371/journal.pone.0179656

4. Middaugh DJ. Managing the 80/20 rule. Vol. 24, *MEDSURG Nursing*. Anthony J. Jannetti Publications, Inc.; 2015.

5. Nisonger TE. The 80/20 rule and core journals. *Ser Lib.* 2008;55(1–2):62–84.

6. Wang W, Wang Z, Yu T, Pak CM, Yu G. Research on citation mention times and contributions using a neural network. *Scientometrics.* 2020 Dec 1;125(3):2383–400. https://doi.org/10.1080/03615260801970774

7. Abramo G, D'Angelo CA, Soldatenkova A. The dispersion of the citation distribution of top scientists' publications. *arXiv*. 2018. https://doi.org/10.1007/s11192-016-2143-7

8. Faust K, Bala S, van Ommeren R, Portante A, Al Qawahmed R, Djuric U, et al. Intelligent feature engineering and ontological mapping of brain tumour histomorphologies by deep learning. *Nat Mach*. 2019 Jul 9;1(7):316–21. https://doi.org/10.1038/s42256-019-0068-6