Addressing Scale (In-)Variance in Novelty Metrics

Nils Rochowicz [*,†,‡]

Keywords: Novelty Metrics, Scaling, Networks

Extended Abstract

Measuring novelty has been an important research strand in the literature on innovation and science of science. Although significant progress has been made, a thorough understanding of novelty in artifacts and knowledge remains elusive. This is particularly evident when it comes to the measurement of novelty: While there are many novelty metrics proposed, there is surprisingly little systematic study of different novelty metrics and their properties. This partially explains the lack of unifying frameworks for novelty assessment (with Foster et al. [2021] being a notable exception), and makes the novelty metrics difficult to compare. This paper attempts to fill part of this gap by analysing the mathematical properties of the most-used novelty metric, introduced by Uzzi et al. [2013] and applied to patents by Kim et al. [2016]. It was already shown by Fontana et al. [2020] that this metric is sensitive to the unit of analysis, and not systematically predictive of highly novel discoveries. This paper takes a similar, but different angle, and shows that this novelty metric exhibits problematic scaling of novelty with the number of combinatorial elements (like technology codes on patents, or citations in scientific publications): The more technology codes there are on the patent, the higher the measured novelty of the patent in expectation. This is due to the fact that the novelty metric is defined at the level of binary combinations of technology codes on the patent. On patents with more technology codes, the number of novelty values measured increases almost quadratically. When aggregating these values to a metric at the patent level, the aggregating function should incorporate the average novelty of the combination, but also the most novel addition of the combination. However, as more values are measured, the discrepancy between average and extreme values in novelty values measured will increase in expectation. This scaling is proven in two different ways: First, through methods from order statistics, it will be shown that any reasonable aggregation function will exhibit such scaling, as long as any quantiles or extreme values of the sample are included (which is very reasonable for any aggregation function from an innovation theory standpoint, since the extreme values are the most novel parts of the invention). Second, anti-concentration bounds (see Wainwright [2019]) will be used to bound the deviation of the minimum from the median from below, dependent on the sample size. This will be used to derive a lower bound specifically for the null model and aggregation function chosen by Uzzi et al. [2013] and Kim et al. [2016]. Simulation results show that the actual scaling is stronger than the lower bound, showing an approximately logarithmic scaling behaviour. Data analysis on the universe of US patents from 1900 to 2010 shows that the scaling is also found in the data, both in the aggregate and for individual primary technology codes. When controlling for the number of codes that a patent is classified with, the correlation between the novelty metric and the number of forward citations gathers is shown to be weaker, and the predictive power of the novelty metric is barely significant. To remedy the scaling, this paper suggests a class of novelty metrics based on random walks on hypergraphs that are scale-invariant.

*Mathematical Institute, University of Oxford
†Institute for New Economic Thinking, University of Oxford
‡Department of Economics, Chemnitz University of Technology

Bibliography

References

Magda Fontana, Martina Iori, Fabio Montobbio, and Roberta Sinatra. New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy*, 49(7):104063, 2020.

Jacob G Foster, Feng Shi, and James Evans. Surprise! measuring novelty as expectation violation. 2021.

Daniel Kim, Daniel Burkhardt Cerigo, Hawoong Jeong, and Hyejin Youn. Technological novelty profile and invention's future impact. *EPJ Data Science*, 5(8), 2016.

Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. Atypical combinations and scientific impact. *Science*, 342(6157):468–472, 2013.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.