Sleeping Beauties in Case Law

Pedro V. HERNANDEZ SERRANO [a1], Kody MOODLEY[a,b], Gijs VAN DIJCK [b]
Michel DUMONTIER[a]

[a] *Institute of Data Science at Maastricht University*
[b] *Faculty of Law, Maastricht University*
{p.hernandezserrano;kody.moodley;gijs.vandijck;michel.dumontier}@maastrichtuniversity.nl

Abstract. A challenge in computational legal research is the quantitative assessment of "relevance" in a network of court decisions. The term "sleeping beauty" (SB) was coined to denote an article that received almost no attention immediately after publication, but suddenly received multiple citations many years later. These publications can be identified by calculating their *Beauty coefficient (B-coefficient)*. In this contribution, we apply approaches used for identifying SBs to decisions arising from the Court of Justice of the European Union (CJEU). We compared B-coefficients of CJEU cases with their centrality scores from classical algorithms from network analysis, finding that these measures tend to correlate. We discuss the implications of this that are interesting for legal scholars, acknowledging that future work is required to calibrate the scale of the time variable in the B-coefficient formula for finer-grained application to case law. Our study's setup provides a foundation for new case law analytics methodologies that extends the power of traditional network analysis techniques for answering questions about the behavior of European courts.

Keywords. Citation Networks, Network Analysis, Sleeping Beauties in Science, Empirical Legal Research, Computational Legal Research

1. Introduction

A 'Sleeping Beauty' (SB) is a publication that goes unnoticed ('sleeps') for a long time, before it suddenly attracts a lot of attention and is cited frequently ('is awakened'). Van Raan [1], who first introduced SB terminology in a computational setting, provided a mathematics for describing the phenomenon. This mathematics offered a way to compare SB significances by balancing their number of citations once awakened and their age. However, van Raan's method was not generalizable outside of the particular corpus of scientific articles he studied. Other approaches, such as that of Redner [2], were focused on particular scientific domains e.g. Physics. In 2015, Ke et al. built upon van Raan and Redner's works by proposing a formula for identifying them in general scientific articles [3]. The central mathematical measure used in this study is the *beauty coefficient* or *B-coefficient* for any publication's "*beauty*": a term denoting the publication's significance as indicated by its age and degree of influence after awakening. Importantly, their mathematical foundations are a priori applicable outside the context of scientific articles, to any corpus of authored documents in which the documents reference each other over time. Since court decisions by the Court of Justice of the European Union (CJEU) have that characteristic, B-coefficients can be calculated for them and the associated citation behaviour can be studied. Previous network analyses of case law have largely overlooked the speed or delay in citations. In classical and doctrinal legal research literature there are also many references to the term "sleeping

---

[1] Corresponding Author: Institute of Data Science at Maastricht University, Paul-Henri Spaaklaan 1 6229 EN, Maastricht, The Netherlands; E-mail: p.hernandezserrano@maastrichtuniversity.nl

beauty" when describing legal articles, directives and court decisions with delayed recognition and significance [4, 5 and 6]. However, the term is used qualitatively to describe the significance of these texts with no acknowledgement of the potential to *computationally* define, identify and compare them. There has not been, to date, an application of SB metrics to study court decisions and this work seeks to fill this gap.

In this contribution, we apply the B-coefficient measure to study citation behavior of court decisions by the Court of Justice of the European Union (CJEU). Our objectives are to understand the influencing factors of SB cases, learn the prevalence of SBs in case law, and calculate the correlation between the B-coefficient and legal network analysis measures such as Degree centrality, Closeness centrality, PageRank and Hyperlink-Induced Topic Search (HITS). By doing so, we explore whether applying a SB methodology provides additional value to other (centrality) measures in determining the relevance of court decisions, offering insight into whether relevant cases are overlooked. From a more legal perspective, we are interested in whether SBs can be detected in time for them to be used when building and presenting a case. Our research questions are: RQ1: What is the prevalence and proportion of SBs in different areas of EU law? RQ2: Does the B-coefficient correlate with classic centrality metrics? RQ3: Do the most relevant cases have a slow awakening? RQ4: Does the duration of a case have an effect on sleep time?

Awakenings can be relevant because they may indicate importance of a certain court decision, potentially signaling a different direction by the court or 'new' arguments in old decisions, showing that certain legal questions have become salient again. To ensure similar cases are treated equally [7], it is important to recognize such developments and further analyze SBs, computationally and non-computationally.

2. Methodology

Ke et al.'s mathematical formula for the Beauty Coefficient *B* is used to quantify and capture both the incoming citations and duration of the "sleeping" period. It was implemented on Web of Science articles (384,649 papers) and American Physical Society (APS) articles (22,379,244 papers) and has power-law behavior. Our Python implementation of the B-coefficient formula is available in a public repository[2]. In previous work we gathered the CJEU data [8], consisting of metadata and citations in CSV format for 13,358 judgements (all published until December 2019)[3]. For this study we selected the subset of CJEU cases cited at least once in the citation network, resulting in 8,979 out of the 13,358 cases (67.21%), and then computed the B-coefficient and variables for all selected cases, resulting in a 13358 by 13 matrix of values, which served as the core dataset for our analyses.

Assuming a case with many incoming citations (*in-degree*) is a proxy for relevance [9], we split the dataset into two groups: *Top cited* cases (cases in the upper 10th percentile) and *Non-top cited* cases (the remainder of cases). The average citation count per case and the standard deviation (SD) for Top cited cases are 18.27 and 9.89 respectively, while for Non-top cited cases they are 3.31 and 2.41 respectively. The average sleeping time in years of all cases is 3.55 (SD 6.22). The maximum value of *sleeping time* is 55 years and the corresponding case has the highest B-coefficient of 4709.50 and an awakening time of 2 years – meaning that it slept for 55 years, was awakened, and then took 2 years to receive its maximum number of annual citations. On

---

[2] https://github.com/maastrichtlawtech/sleeping-beauties-case-law
[3] https://github.com/MaastrichtU-IDS/case-law-analysis

average, the expected time to gain the maximum number of annual citations in a year for all cases is 0.94 (SD 0.58) years.

A case's B-coefficient has to be above a certain threshold for the case to be a SB. This threshold is not fixed because it depends on factors specific to the citation behavior in the corpus. We adopted Ke et al.'s proposed baseline [3] of taking the cases with the top 0.1% of the B-coefficient values as the SBs. We therefore identified 90 unique SBs from the CJEU rulings dealing with a variety of legal topics.

3. Results and Findings

RQ1: What is the prevalence and proportion of SBs in the different areas of EU law? We aggregated the SBs according to legal topics as specified on the EUR-Lex website. *Approximation of Laws* has the most SBs (8 cases). *Concerned Practices* cases have, on average, the highest B-coefficients (93.93). *Copyright and related rights* have the highest proportion of SBs (1.47%). We also counted the number of cases which cite a particular piece of legislation. Then, for each legislative topic, we identified which of the cases are SBs (according to our threshold). We found that the *Regulation (EU) 2017/1001 of the European Parliament of 14 June 2017 on the European Union trademark* is the legal topic addressed by the most (six) SB cases.

RQ2: Does the B-coefficient correlate with classic centrality metrics? We selected the 90 SBs discovered and calculated classical centrality measures (Degree centrality, Closeness centrality, PageRank, and HITS algorithm[4]) with the whole citation network. Degree centrality and PageRank showed significance ($p < .05$), whereas, Closeness centrality and HITS algorithm showed only marginal significance ($p < .10$). All four metrics showed positive correlations, but no more than .30. A difference of means test compared the average sleeping time between the two case groups. Sleeping time has a right skewed distribution (Sk = 3.04); thus, a log transformation was calculated, before a T-test was performed on the *log sleeping* with a one-sided distribution and equal variance, resulting in a null hypothesis rejection: Top-cited cases sleep longer than the Non-top cited ones (0.86 years – almost 11 months).

RQ3: Do the most relevant cases have a slow awakening? The average awakening time was compared between the Non-top cited and the Top-cited groups using a T-test performed on a one-sided distribution with unequal variance. The null hypothesis was rejected, meaning that the mean awakening time of the Top-cited cases is larger than that of the Non-top cited cases (0.7 years – around 8 months).

RQ4: Does the duration of a case have an effect on sleep time? Case duration is the quantification of the days between the lodge and publication dates, potentially indicating case complexity. *Case duration* has a normal distribution with a mean and median of 1.5 years (SD 0.74). However, *sleeping time* is right skewed; therefore, a *Spearman* coefficient is used to calculate a correlation. The test's results were non-significant: there are no signs of correlation between a case's *sleep time* and the *case duration* in years.

4. Discussion and Conclusion

We applied the *B-coefficient* to identify publications with delayed citations (sleeping beauties) from CJEU case law. We found the methodology to be complementary to traditional network analysis metrics, adding novel analytical power to computational

---

[4] Degree centrality, Closeness centrality, PageRank and HITS algorithm were calculated using the open source implementations of NetworkX Python package. https://networkx.github.io/

studies of case law. By applying these metrics to CJEU cases, we observed the following main results: Top-cited cases usually sleep longer and tend to awaken slower than other cases. There are no signs that case duration has any effect on sleep time. While the B-coefficient generally correlates with well-known centrality measures when identifying significant cases, it does identify new ones missed by traditional metrics.

If a relevant case is dormant for a while, once it is cited the awareness of this case increases rapidly, and it is swiftly used in other decisions. On average, decisions reach the point of highest citations within the first year, but often the time between the lodge date and decision date is longer than a year, meaning parties may not know which decisions are relevant to their cases. Future research should analyze the extent to which different case types cite previous decisions that are similar. B-coefficient's positive correlation with four traditional centrality metrics, but no more than 0.30, suggests that the *B-coefficient* partly captures the information provided by traditional measures, yet produces complementary information. However, this requires further analysis as the *sleep–peak ratio* could be indicative of cases being (or becoming) landmark cases. Furthermore, Top-cited cases tend to sleep longer than the Non-top cited ones; thus, the relationship between SBs and B-coefficients with precedents and landmark cases should be explored further.

In future work we intend to address the unweighted influence of case age on the B-coefficient. The frequency and volume of citations in case law follows a different behavior from scientific literature, which is published and cited more frequently than case law [1, 2, 3]. Therefore, the hypothesis is that we need to calibrate the B-coefficient formula for case law by tuning the intensity of the timescales in the formula to match the case law context. We will also perform a deeper study of the legal topics that SBs tend to address, as well as enlist the aid of legal researchers to corroborate our findings and comment on their significance, especially comparing them across subfields of law.

References

[1] van Raan, A. F. J. (2004). Sleeping Beauties in science. *Scientometrics*, *59*(3), 467–472. https://doi.org/10.1023/B:SCIE.0000018543.82441.fl

[2] Redner, S (2005) Citation statistics from 110 years of physical review. Phys Today 58(6):49–54.

[3] Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying Sleeping Beauties in science. *Proceedings of the National Academy of Sciences*, *112*(24), 7426–7431. https://doi.org/10.1073/pnas.1424329112

[4] Clancy, Pearce (2018). Arise, Sleeping Beauty: What PESCO Means for Ireland. Irish Yearbook of International Law, Vol. 13, pp. 79-98

[5] von der Dunk, Frans G. (2008). A Sleeping Beauty Awakens: The 1968 Rescue Agreement after Forty Years. Journal of Space Law, Vol. 34, Issue 2 (Winter 2008), pp. 411-434

[6] Stedman, John C. (1957-1958). Merger Statute: Sleeping Giant or Sleeping Beauty. Northwestern University Law Review , Vol. 52, Issue 5 , pp. 567-617.

[7] Maltz, E. The nature of precedent. New California Law Review 1987, vol. 66, 367.

[8] Moodley, K., Hernandez Serrano, P. V., van Dijck, G., & Dumontier, M. (n.d.). *IOS Press Ebooks—Similarity and Relevance of Court Decisions: A Computational Study on CJEU Cases*. Retrieved 7 August 2020, from http://ebooks.iospress.nl/volumearticle/53654

[9] M. van Opijnen (2012). Citation Analysis and Beyond: in Search of Indicators Measuring Case Law Importance. In JURIX, 250, 95-104, IOS Press