Analysis of a German Legal Citation Network

Tobias Milz and Jelena Mitrović (tobias.milz,jelena.mitrovic@uni-passau.de)

Chair of Data Science, University of Passau

Introduction

In recent years, investments into supporting technologies for the legal industry (also known as Legal Tech) have reached new record highs. NLP based information retrieval systems, in particular, have gained a lot of attention over the years as the legal domain is confronted with an ever-growing amount of text-based information. The CAROLL research group at the University of Passau is developing a German legal citation network in order to investigate ways of leveraging NLP and network analysis tools to support common legal research tasks and reveal new information about the German legal system.

Legal Citation Network

For the nodes of the network we used the Open Legal Data dataset [2], which contains the data of over 200,000 court decisions, 50,000 laws and 1,000 courts. The links between these nodes are created by extracting references from within the text of a court decision using regular expressions.
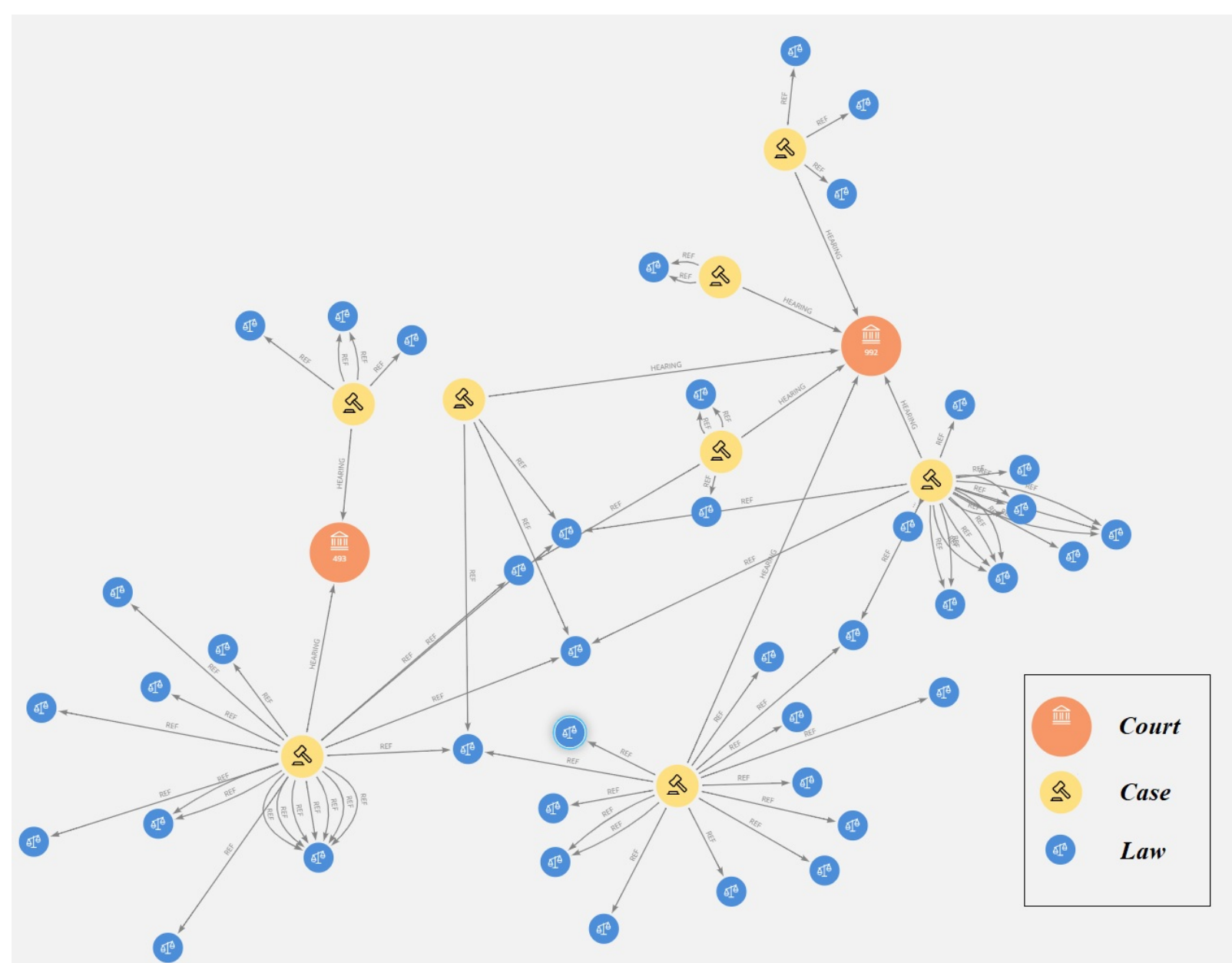

Figure 1:Snippet of the Legal Citation Network.

For example, a link exists between two case nodes (n) and (m) if (n) references (m) in the decision text. Similarly, if law (l) is cited in the decision text of case (n), an edge between a Case node (n) and Law node (l) is created. Additionally, an edge between two court nodes (c) and (d) is added whenever an indirect citation between courts is identified (i.e case (n) from Court (c) is referencing case (m) from Court (d)).

Scale-Free Network

Previous research has shown that case citation networks of the U.S. Supreme Court [3], Austrian Supreme Court [1] and European Court of Justice [4] exhibit scale-free characteristics. As evident by figure 2, the German citation network seems to display the same typical power-law degree distribution. More than 70% of court decisions are not cited at all and 92.6% are cited less than five times. This reveals that there is a very small number of court decisions (and laws) that receive most of the citations and consequently, hold a substantial amount of legal influence.
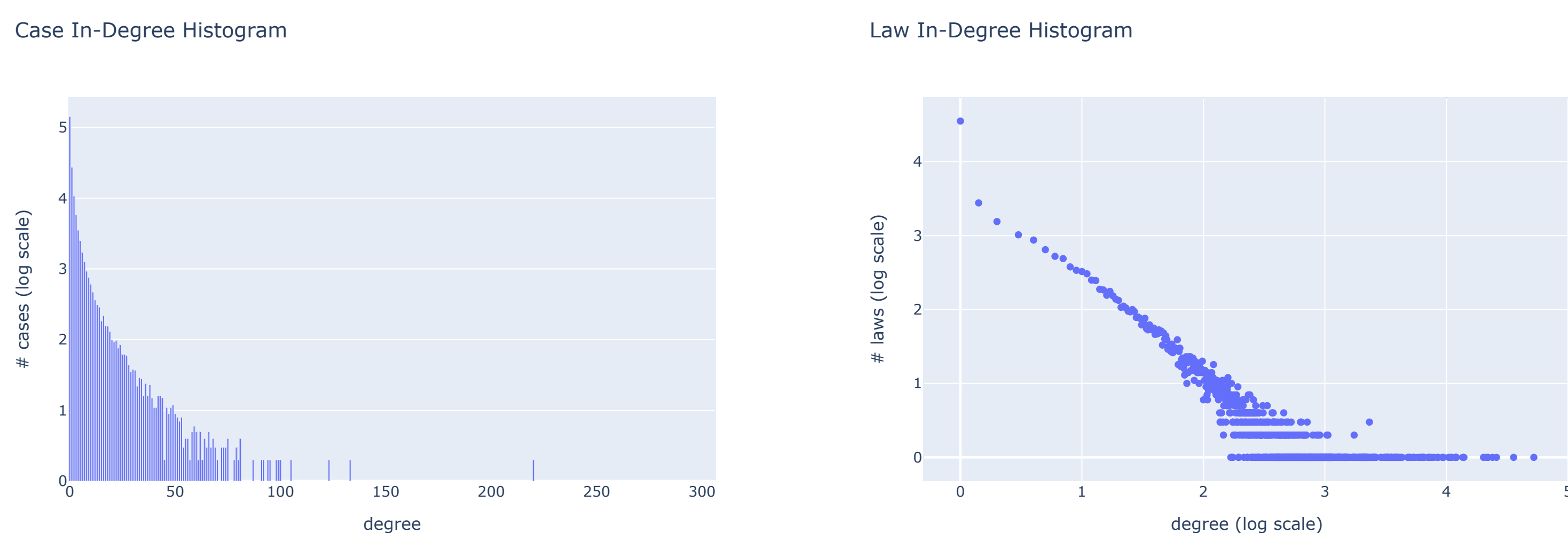

Figure 2:Distribution of incoming citations.

Centrality

Preceding analysis of ranking scores [1, 4] has proven that in-degree (number of incoming citations) and PageRank scores are strong indicators for identifying precedents or otherwise influential cases. Consequently, investigated the same metrics on laws and courts. As an example, figure 3 visualises the most cited laws of the German civil code.
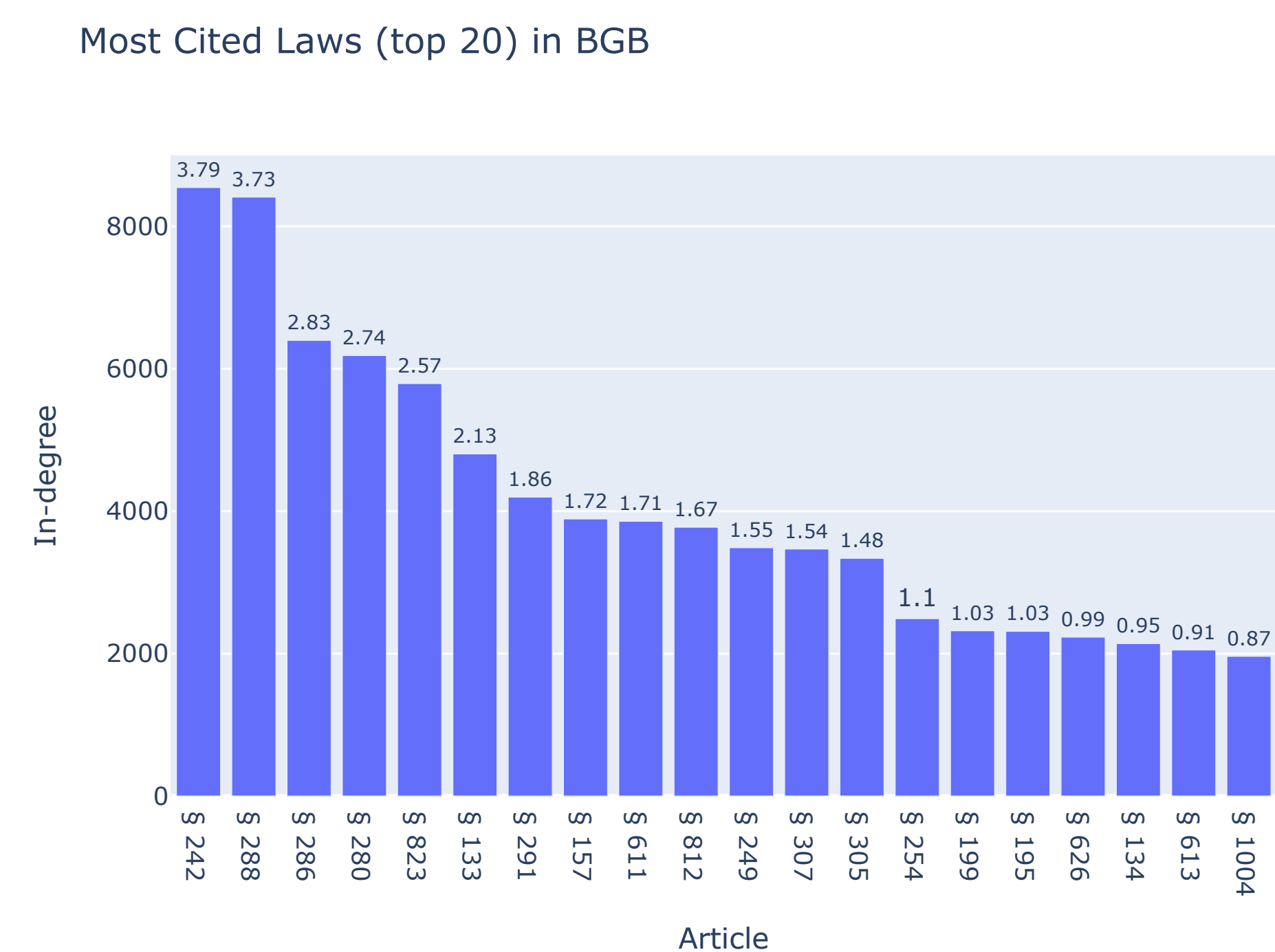

Figure 3:Distribution of citations towards the civil code.

In figure 4, we identified the most important court decisions based on their overall PageRank rating. The stacked bars show the change of their PageRank value over time. The graph indicates that most decisions become influential after three years.
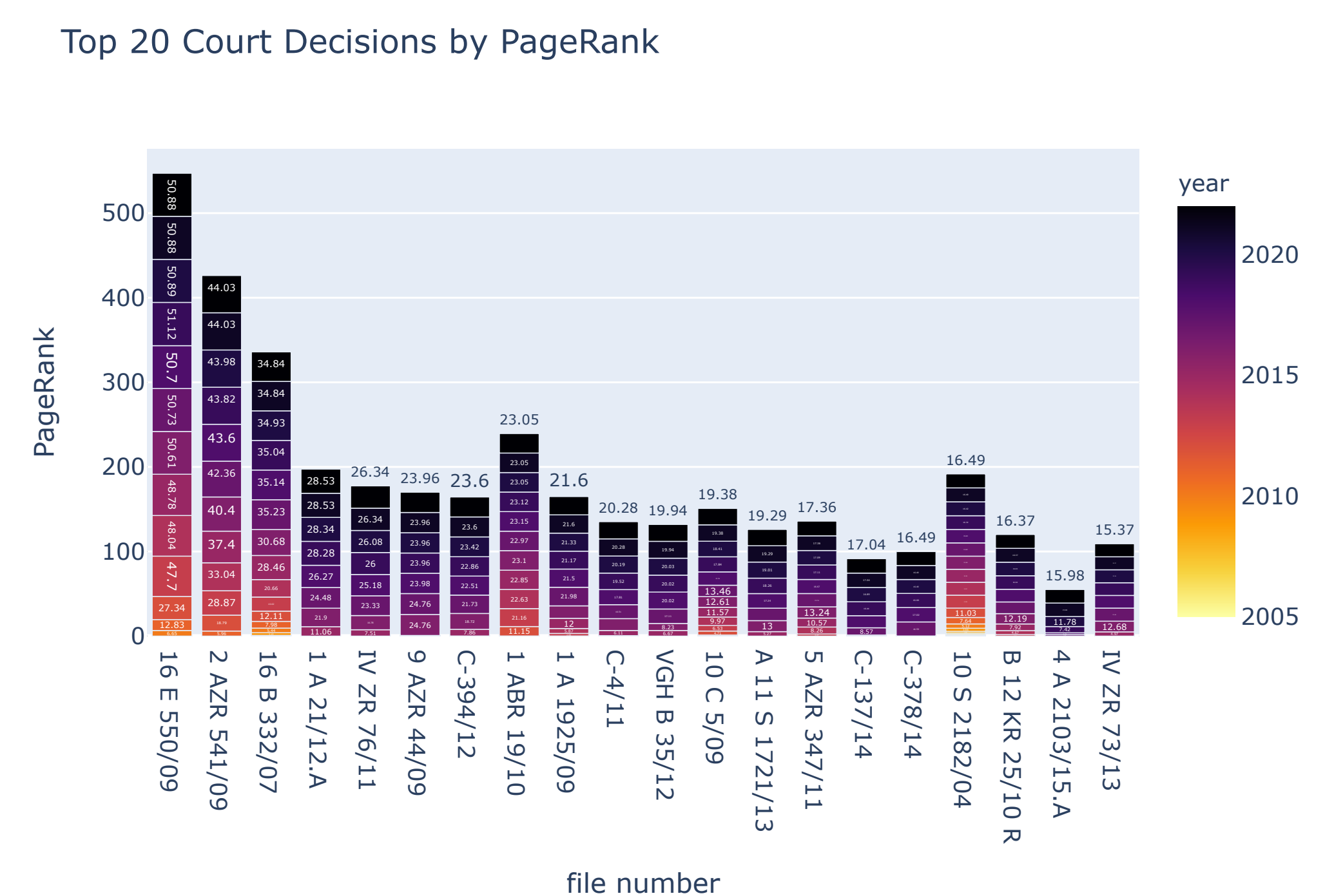

Figure 4:PageRank over time.

Node Similarity

An important task of legal research includes the discovery of similar cases to a specific topic or field of law. We computed the Jaccard similarity score between all pairs of nodes in the network and compared the results with a text-based, TF-IDF vector cosine similarity calculation. The Pearson correlation of 0.64 implies a positive correlation between the two measures, which suggests that network-based similarity approaches can provide an additional avenue for case-similarity identification tasks.

Acknowledgement & References

[1] Anton Geist. "Using Citation Analysis Techniques for Computer-Assisted Legal Research in Continental Jurisdictions". In: *SSRN Electronic Journal* (Mar. 2009). DOI: 10.2139/ssrn.1397674.

[2] Malte Ostendorff, Till Blume, and Saskia Ostendorff. "Towards an Open Platform for Legal Information". In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* (Aug. 2020). DOI: 10.1145/3383583.3398616. URL: http://dx.doi.org/10.1145/3383583.3398616.

[3] Thomas A.C. Smith. "The Web of Law". In: *SSRN Electronic Journal* (2005). ISSN: 1556-5068. DOI: 10.2139/ssrn.642863.

[4] Staffan Malmgren. "Towards a theory of jurisprudential relevance ranking Using link analysis on EU case law". PhD thesis. Stockholm University, 2011.