Some thoughts on predicting citations

(Just a discussion, really)

The Idea

Case law is where abstractly formulated statutory law meets the world of facts, events, and social practices. In this sense, case law is the frontier of law, where it is hashed out how statutory law should be interpreted. In certain cases, case law even supplements the law, when no statutes apply immediately. Legal precedent is important precisely as the mechanism through which case law becomes a source of law. When the court decides a case, all future cases should be decided the same way. New legal practices thus arise only when there is something substantially new in the facts of a case (or when the social practice relating to the facts) changes.

Here we investigate the development of case law through the citation patterns of The Court of Justice of the European Union (CJEU) in order to illuminate underlying factors which may shape the decision-making process. We consider the citations occurring in the period between 1955 and 2014. These form a network of individual cases (nodes) connected by citations (directed links). As time goes by, new cases become part

Case law is where abstractly formulated statutory law meets the world of facts, events, and social practices. In this sense, case law is the frontier of law, where it is hashed out how statutory law should be interpreted. In certain cases, case law even supplements the law, when no statutes apply immediately. Legal precedent is important precisely as the mechanism through which case law becomes a source of law. When the court decides a case, all future cases should be decided the same way. New legal practices thus arise only when there is something substantially new in the facts of a case (or when the social practice relating to the facts) changes.

Here we investigate the development of case law through the citation patterns of The Court of Justice of the European Union (CJEU) in order to illuminate underlying factors which may shape the decision-making process. We consider the citations occurring in the period between 1955 and 2014. These form a network of individual cases (nodes) connected by citations (directed links). As time goes by, new cases become part

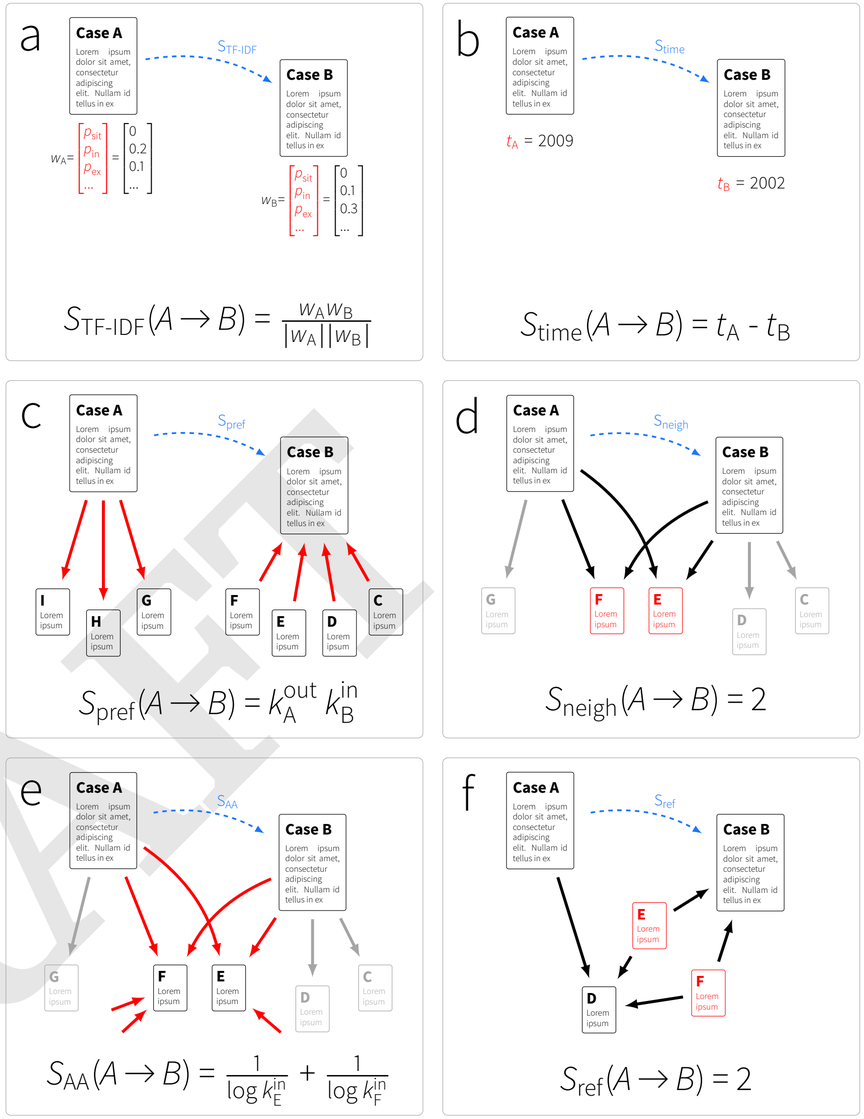
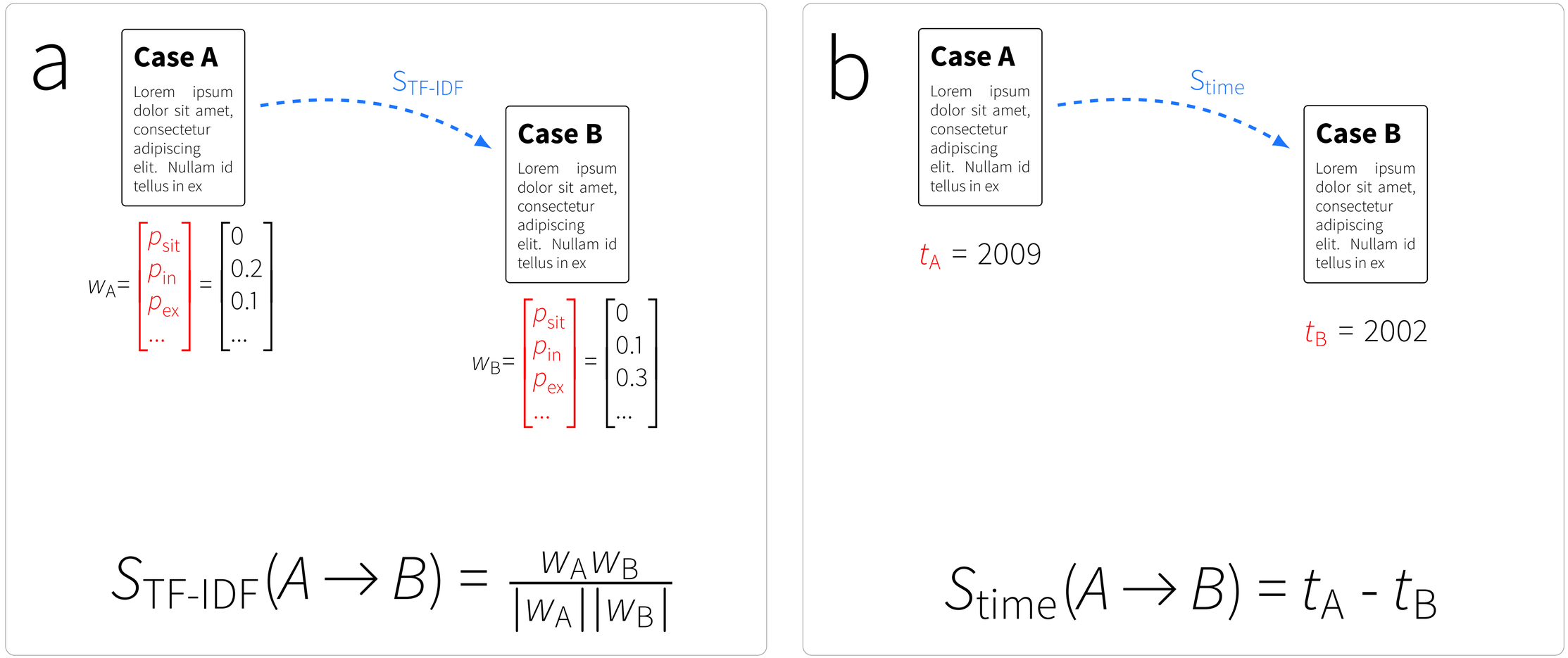


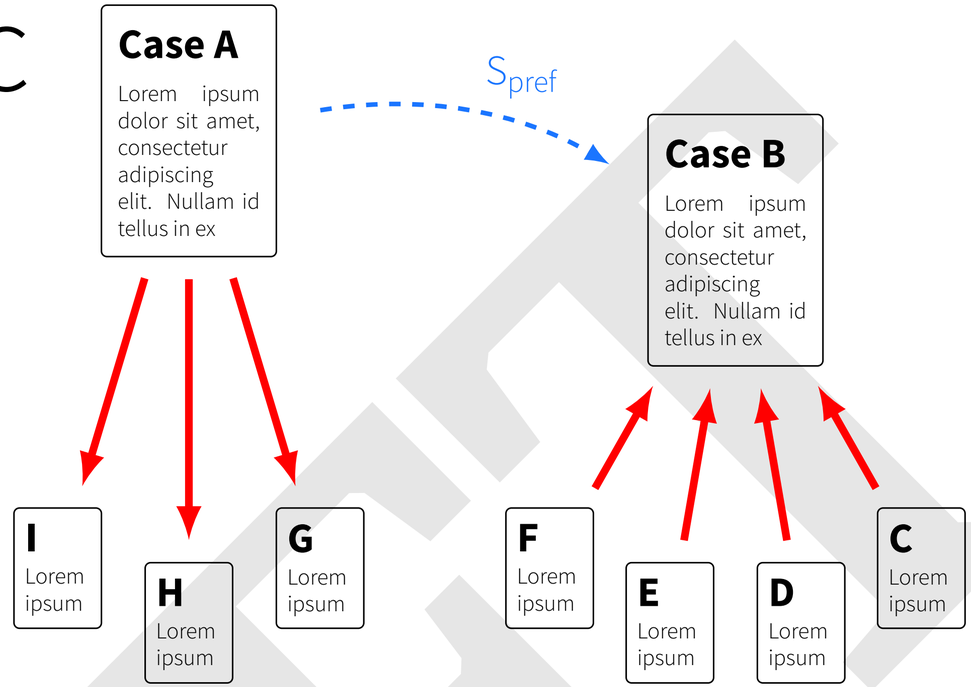
Fig. 1. Features used in the inspection of the CJEU court. Each feature can be described as nodal or structural. Nodal features: a) TF-IDF ($S_{\mathrm{TF-IDF}}$), b) time difference (S_{time}), c) preferential attachment (S_{pref}). Structural features: d) Adamic–Adar (S_{AA}), e) common neighbors (S_{neigh}), f) common referrers (S_{ref}). Documents and links that are included in the calculation of the feature are highlighted in red.


a
Case A
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam id tellus in ex
STF-IDF
Case B
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam id tellus in ex
wA= psit pin pex ... = 0 0.2 0.1 ...
wB= psit pin pex ... = 0 0.1 0.3 ...
STF-IDF(A → B) = wA wB / |wA||wB|
b
Case A
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam id tellus in ex
Stime
Case B
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam id tellus in ex
tA = 2009
tB = 2002
Stime(A → B) = tA - tB
c
Case A
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam id tellus in ex
Spref
Case B
Lorem ipsum dolor sit amet, consectetur
d
Case A
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam id tellus in ex
Sneigh
Case B
Lorem ipsum dolor sit amet, consectetur

$$S_{\text{TF-IDF}}(A \to B) = \frac{w_A w_B}{|w_A||w_B|}$$

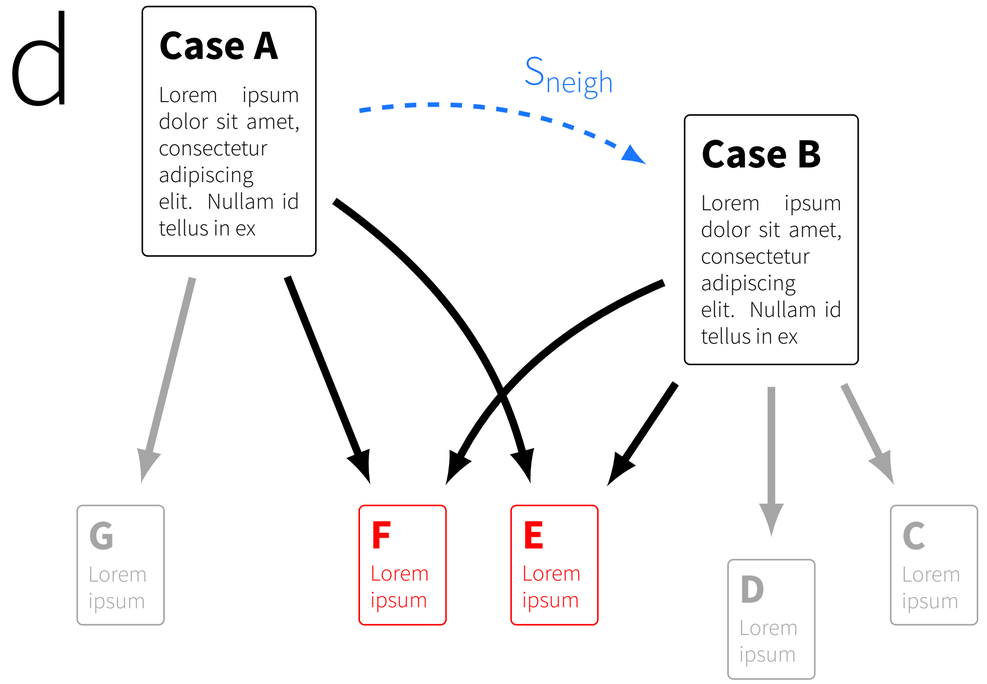
$$S_{\text{time}}(A \to B) = t_A - t_B$$

c



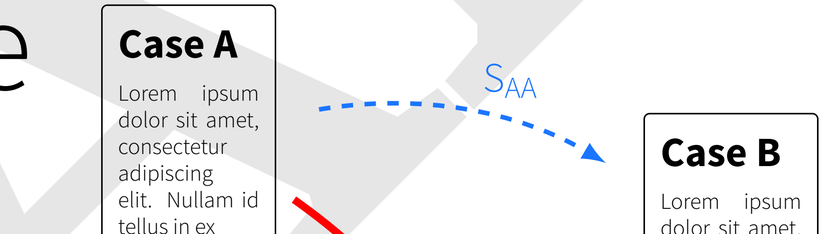
$$S_{\text{pref}}(A \to B) = k_A^{\text{out}}\, k_B^{\text{in}}$$

d



$$S_{\text{neigh}}(A \to B) = 2$$

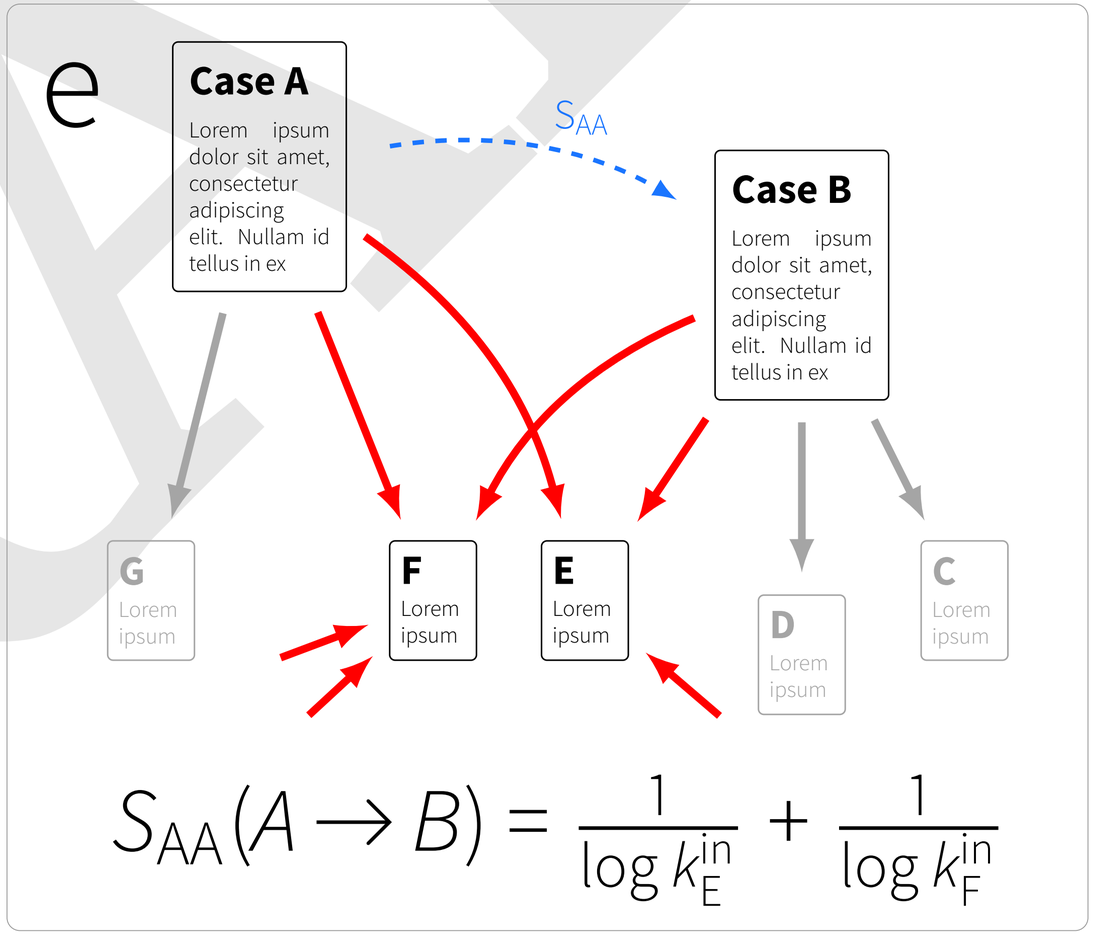
e



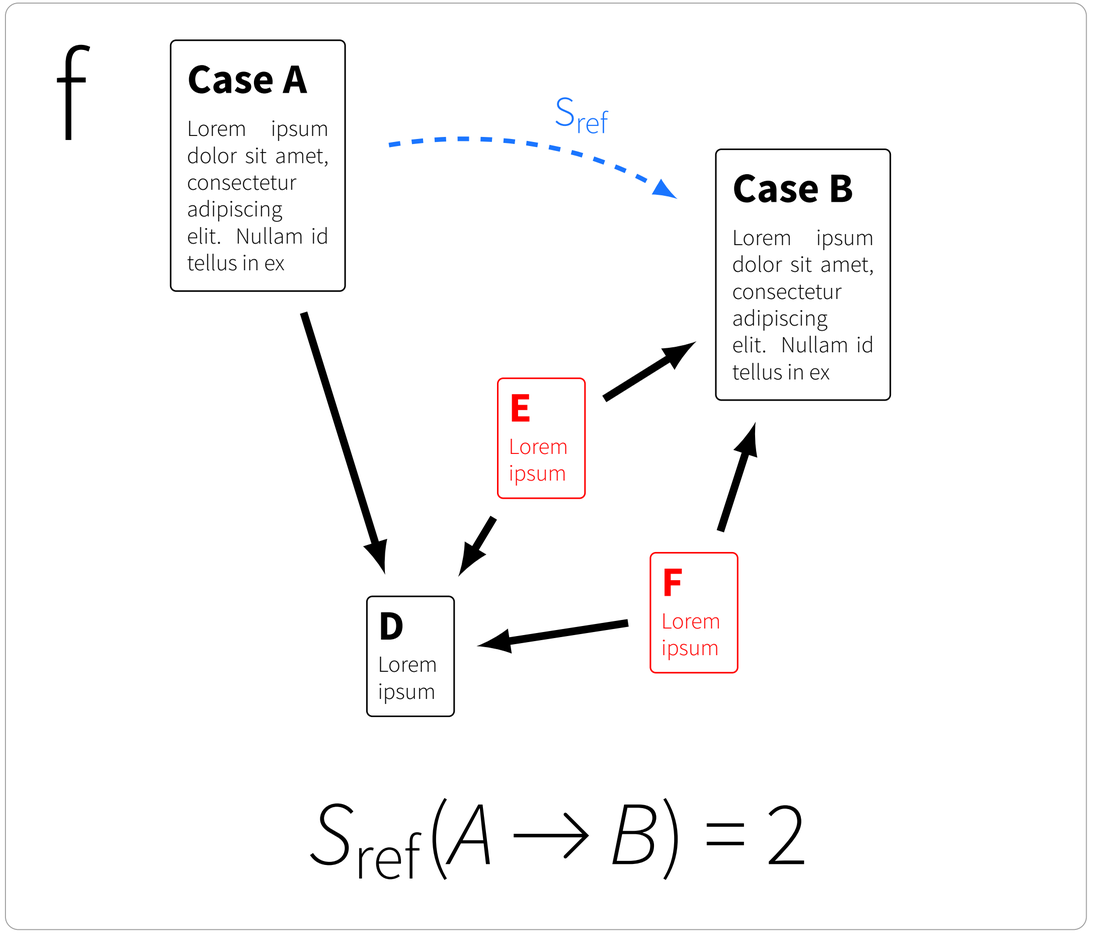
f

$$S_{\text{pref}}(A \to B) = k_{\text{A}}^{\text{out}}\, k_{\text{B}}^{\text{in}}$$

$$S_{\text{neigh}}(A \to B) = 2$$



$$S_{\text{AA}}(A \to B) = \frac{1}{\log k_{\text{E}}^{\text{in}}} + \frac{1}{\log k_{\text{F}}^{\text{in}}}$$



$$S_{\text{ref}}(A \to B) = 2$$

Fig. 1. Features used in the inspection of the CJEU court. Each feature can be described as nodal or structural. Nodal features: a) TF-IDF ($S_{\text{TF-IDF}}$), b) time

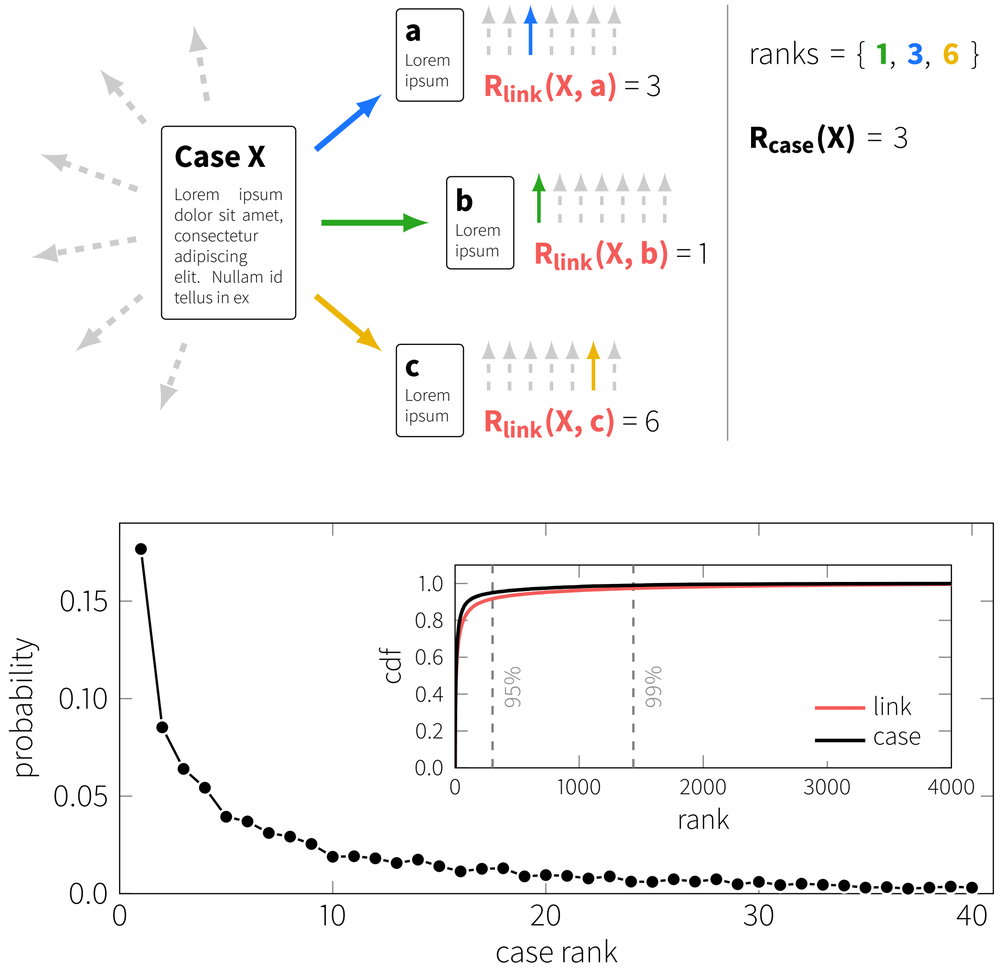
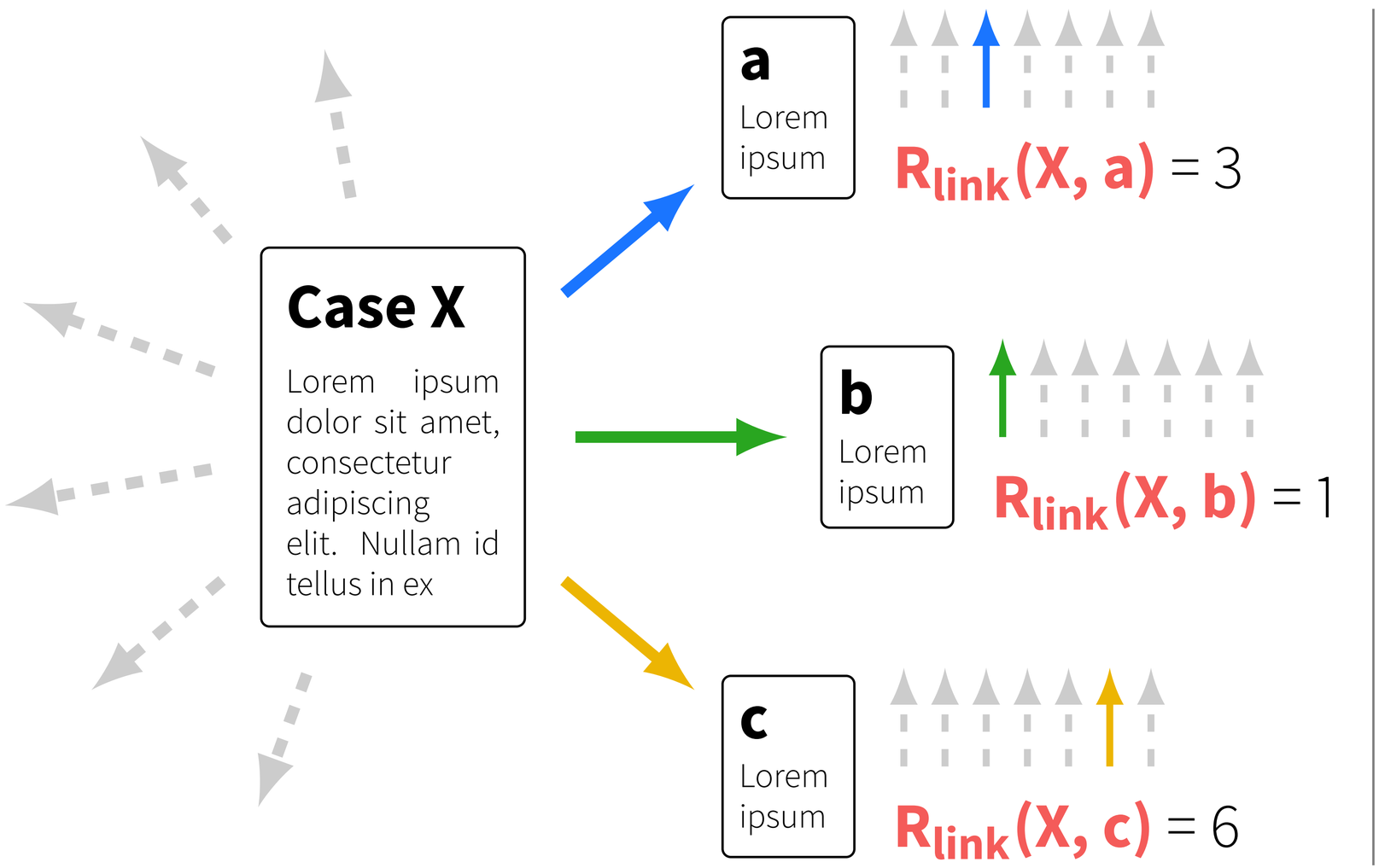


Fig. 2. Global performance of the link prediction. Top: definition of the median prediction rank performance measure: each link of a case (colored arrows) is compared against all non-existent links (dashed gray) providing the link ranks (bold red). The full median prediction rank (calculated for each case)is defined as the median rank of outward links corresponding to the case (bold black). Main plot shows the probability mass function low case ranks, inset shows the cumulative distribution function of the link (red) and case ranks (black). Dashed lines mark the 95% and 99% percentage of the total number of cases.


Case X
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam id tellus in ex
a
Lorem ipsum
$R_{link}(X, a) = 3$
b
Lorem ipsum
$R_{link}(X, b) = 1$
c
Lorem ipsum
$R_{link}(X, c) = 6$
ranks = { 1, 3, 6 }
$R_{case}(X) = 3$

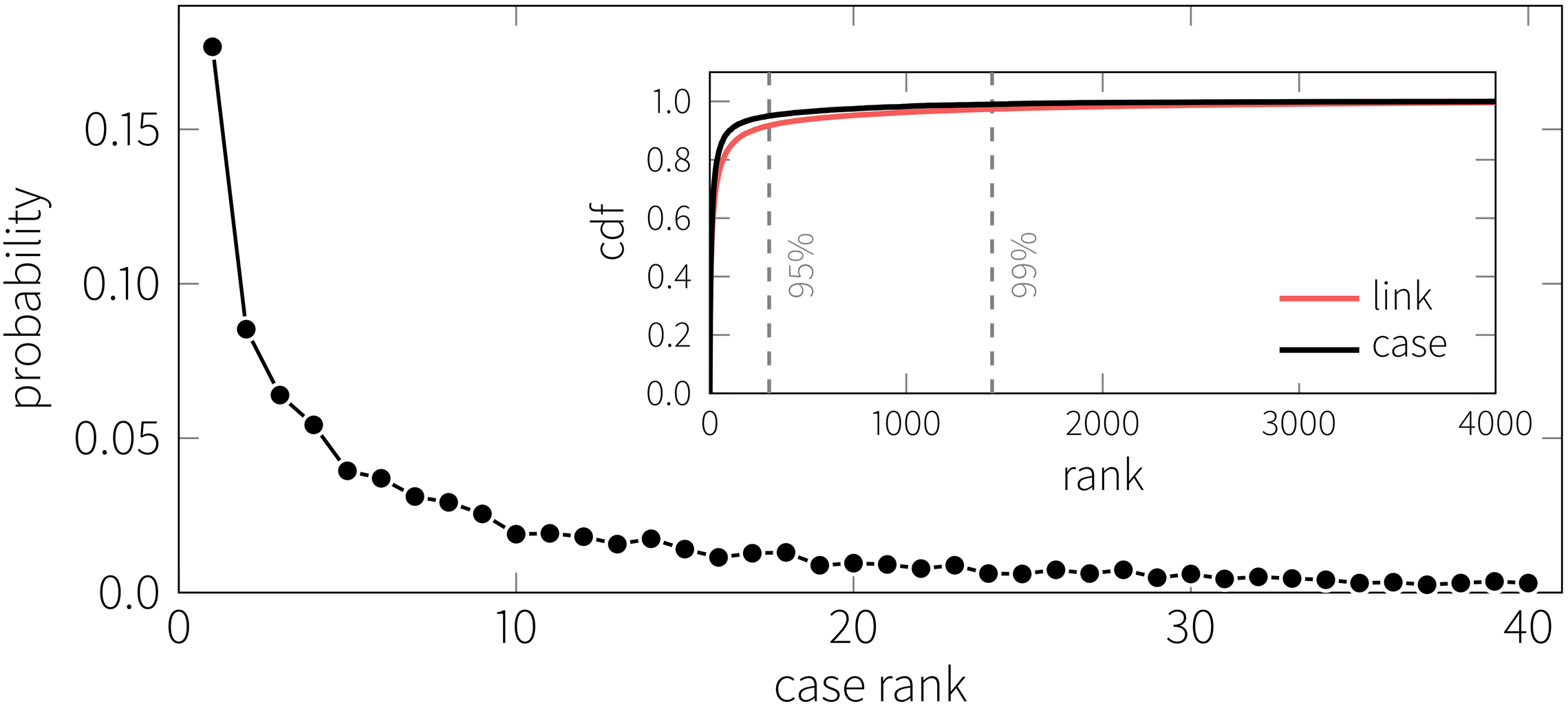


Fig. 2. Global performance of the link prediction. Top: definition of the median

And BTW, the mistakes are meaningful

Interpreting model errors. Finally, it is interesting to study the situations where the model makes mistakes. Here we focus on false positives and false negatives. In the case of false positives, our model recommends a reference between two cases, which in turn does not exist in the citation network of the court. Empirically, these cases that, according to the algorithm, are 'supposed' to be cited, discuss similar legal topics, but with subtle differences in the specific details of the legal issues (see Supplementary Information for details). An illustrative example is the potential citation of Case C-412/05 P in Case C-304/06 P. Both of these cases deal with Community Trade Mark Law, including the distinctive characteristics of the mark. However, the potential cited case is concerned primarily with the admissibility of the appeal. In other words, they discuss highly similar topics but there are slight differences in the particular focus of the cases which the algorithm was not able to discover.

On the other hand, false negatives are the references that were not found by our model but are observed in the court. There are several reasons for false negatives. First, it is common to cite previous cases to provide an example for a type of argument even though the example itself is not related to the legal issue in question. An example is Case 124/83 citing joined cases 94-63 and 96-63. The citation emerged as evidence of a general principle that an authority which adopts measures affecting the persons concerned or which withdrawn a favorable decision must bear the burden of proof itself. Second, false negatives also include clerical errors – the citations are mistakes as another case should have been cited. For full details, see Tables S1-S3. in the Supplementary Information.

So what’s going on?

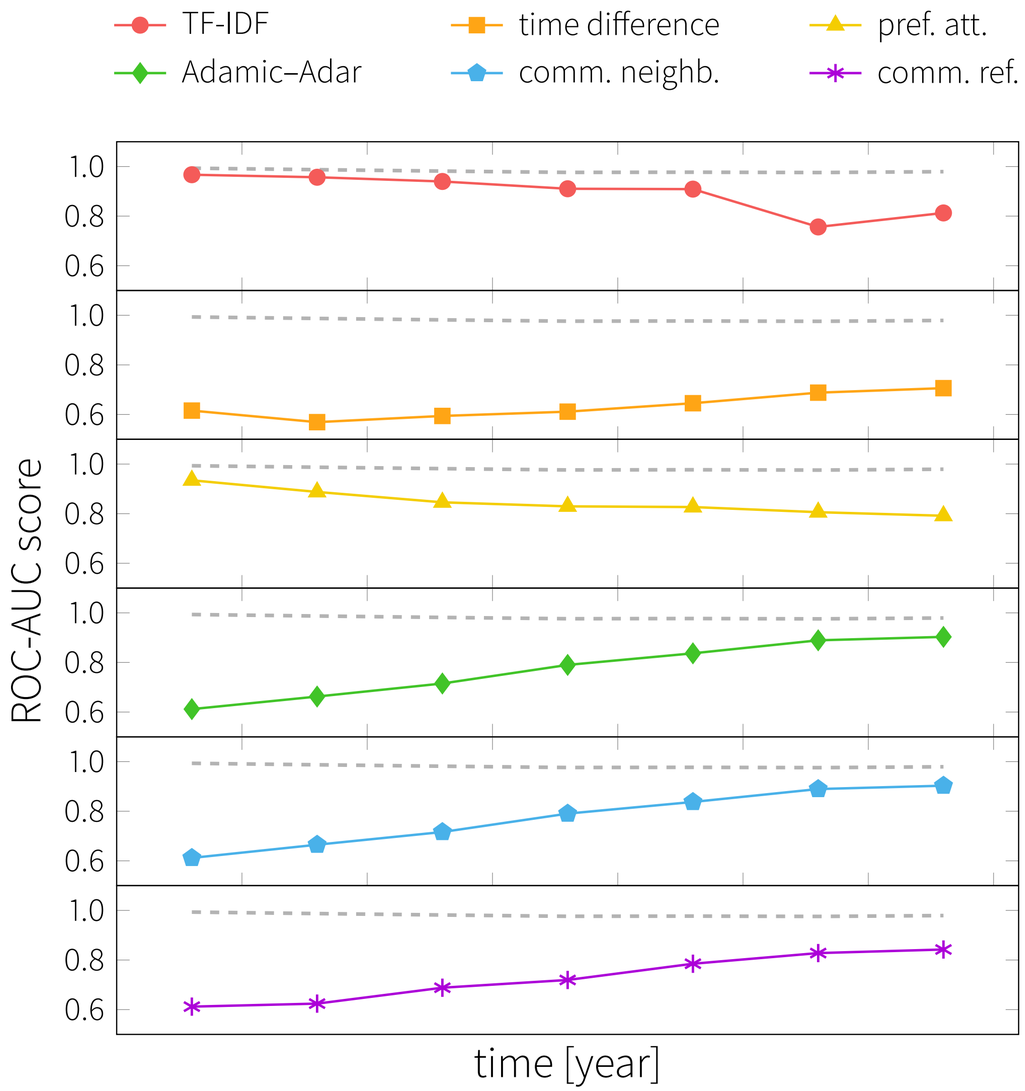


Fig. 3. Predictive power of individual features. Lines show the change in ROC-AUC for each feature calculated from the raw feature values. The dashed gray lines show the value 1.0 as a guide to the eye.

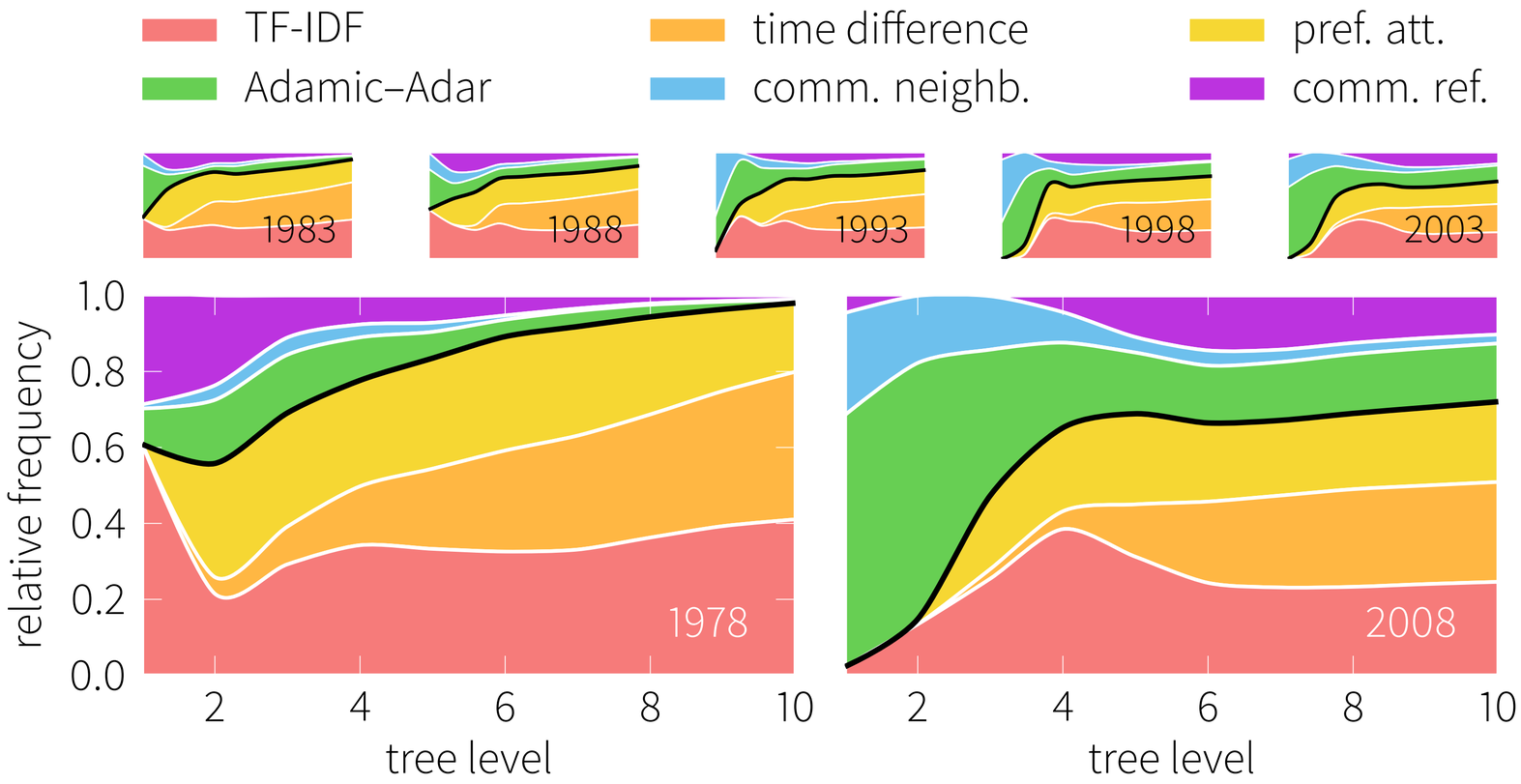


Fig. 3. Details of feature usage inside the decision trees. The curves show the fraction of decision nodes in the decision trees that use a specific feature in different levels of the trees (they add up to one). For each feature, we calculate the number of (internal) decision nodes that make the split based on the value of that feature, normalized by the total number of nodes in that level. Results are averaged over all trees in a random forest and over 5 independent forests.

Things to think about

Predictability is desirable

Importance of understanding empirical patterns of case law usage. A deeper understanding of the principles that shape the usage is key. Since consistency in how cases are treated, not only supports equality before the law, but also enhances predictability and effectiveness. Predictability is desirable because when those who are subject to the law know that new cases will be treated consistently with previous decisions, they can use earlier cases as a legal compass, to navigating their behavior in accordance with the law. Currently, maintaining consistency and predictability is expensive. It requires an insight into and overview of previous case law, which is difficult for a single human being to achieve.

Better recommendations can help professionals and non-professionals alike

Moreover, while many cases are unimportant (e.g. regarding uncommon scenarios, or trivial repeat cases) a few key cases have proven cardinal in understanding unwritten legal principles and explaining statutory law, and others are important for very specific fact situations. Important cases are currently identified by scholars and lawyers simply reading court cases. We posit that it may helpful to introduce information managing tools based on methods such as the ones proposed there in order to help agents to navigate the case law. Case law databases make it possible to index cases by specific categories, but the cases in the database must first be categorized in a way that makes sense for supporting legal reasoning. Improvements in search engines have made it possible to do full text search, making the job of finding applicable case law considerably easier. However, knowing what to look for, given a particular problem, remains a skill reserved for legal professionals and is susceptible to these professionals' own biases and other human limitations. [SUNE: henrik add citations to this section]

Some use-cases

Applications. This paper shows that there is a possibility of predicting which piece of case law is applicable given the content (text and citations) of an already existing case. Algorithmically identifying relevant cases may have several advantages, among which improving the reproducibility of doctrinal legal studies and reducing human bias is one important gain. Introducing a technology that is capable of interacting with the insight of expert humans has potential to bring several advantages to the legal sector overall. In the following we list the most obvious applications of our link prediction system (see SI for full discussion).

- Transferring information from CJEU to domestic settings. E.g. making it easier for administrative agencies to make informed decisions about rights of citizens.

- Make it easier for legal service providers to find relevant CJEU case law to support arguments made for clients.

- Support to the CJEU itself. A link prediction system could help the court navigate its own case law when preparing new cases and may even be used to check whether new judgements by the court sufficiently cites relevant former cases.

- The link prediction system could be implemented in legal research and teaching settings. Allowing students and legal scholars to navigate case law in the more advanced ways could potentially allow for new insights by legal scholars and law students.

Errors

Limitations. The link prediction system discussed here is not perfect. And while our predictions have respectable performance from a machine learning perspective (see Fig. 2), the performance is less impressive from a legal point where accuracy is essential [citation]. Finding most of the right cases most of the time is simply not good enough [citation]. A link prediction system, however, can still deliver the advantages mentioned above. But we argue that any recommender system should be seen as a support for humans in order to deliver the advantages discussed above. In this way, critical legal reasoning can be applied to the process of selecting those cases that are relevant to the individual case at hand.

It’s not enough to just reproduce the past

As we rely on humans to mitigate the limitations of a recommender systems, we highlight the potential dangers of relying too blindly on systems like the one we have developed here. Law is an organic system of normativity that adapts to evolving social life by making concrete judgments about how generally formulated laws should be applied in individual specific circumstances [citation]. The justice system is supposed to render justice and this in turn implies a holistic view of the overall case [citation]. Relying mechanically on former cases and imitating those former decisions when making new ones is not an ideal model to follow. If every new decision is framed as a copy of a former decision, the system will risk freezing up and becoming a parody of justice. Predictability and the principle of treating like cases alike is one important part of rendering justice, and our link prediction system helps to advance this in a more effective and systematic manner. But predictability and equality is not the only component of justice – not the only value that Courts should promote. The courts must also somehow reflect prevailing social normativity and fulfill its role as a bulwark against abuse of power. This requires more than relying on former case law – it requires that every case is judged anew so that every person who appears before the court will be served ‘fresh’ justice.

[SUNE: Henrik, maybe also something about false negatives. What happens if our system neglects to find the right case for some reason]

[SUNE: Henrik (important!), can we think of something to write about the increasing importance of structure and how it connects to limitations? A connection to legal precedent?]

[SUNE: Henrik what does it mean to legal scholars that the network becomes more important for the algorithm]

Communities

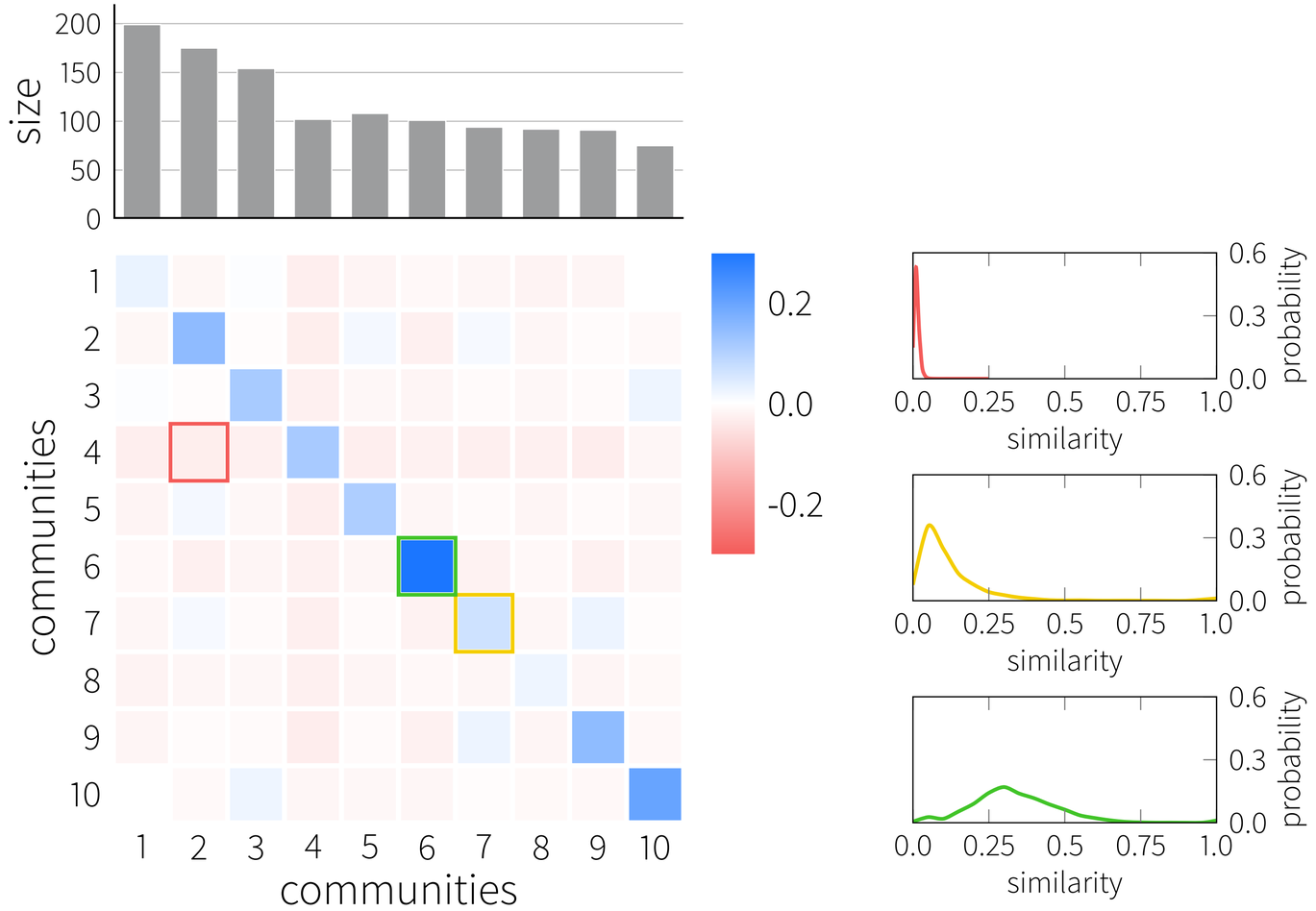


Fig. 4. Nodal feature similarity within and between communities. Heat map similarity matrix shows the median cosine similarity of TF-IDF vectors between cases of indicated communities, normalized by the average value of the tiles. The size of each communities is shown at the top of the heat map. Figures on the left display three typical distribution of similarity measures highlighted in the heat map.

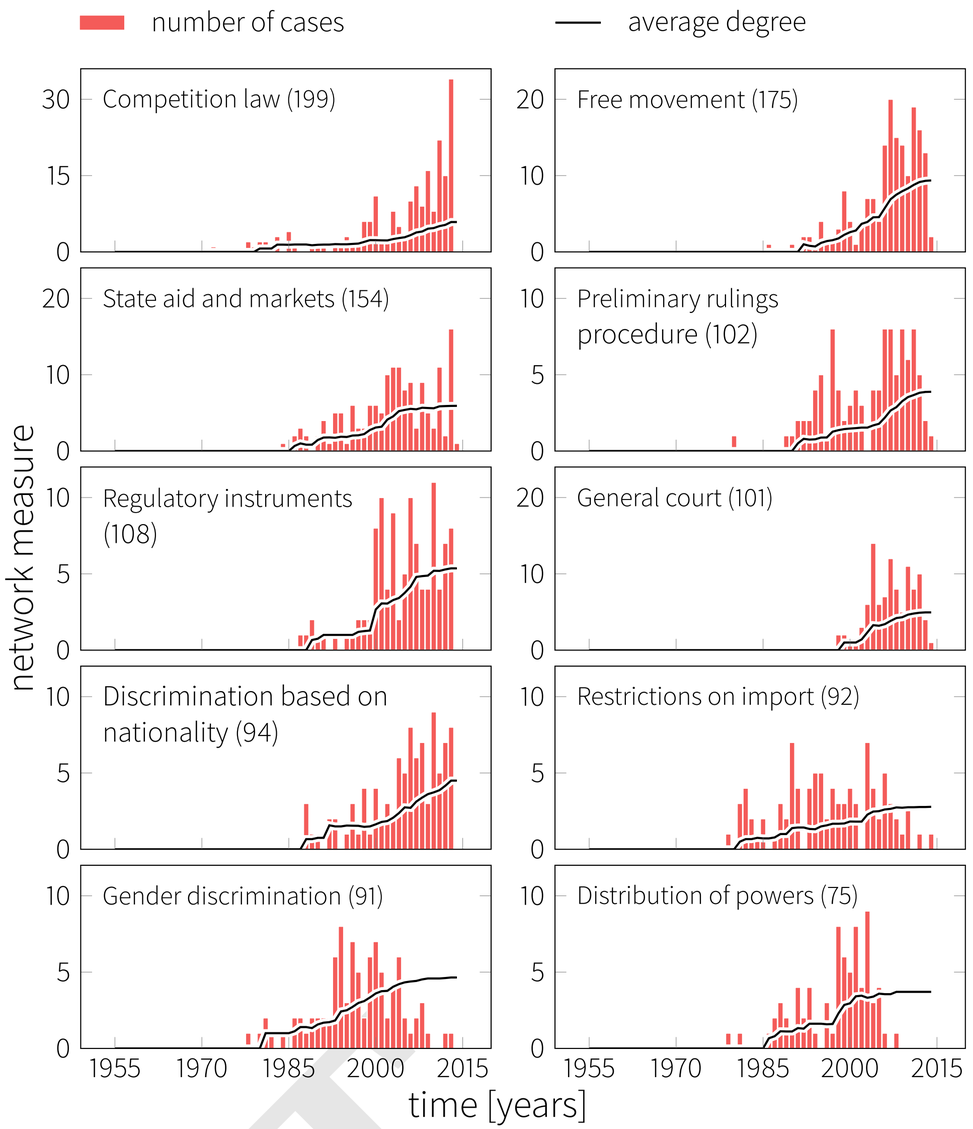


Fig. 5. Evolution of communities. For each of the ten communities, we report the number of cases (red bars) as well as the average degree (black line). The numbers next to the communities' name denote the number of cases in the group.

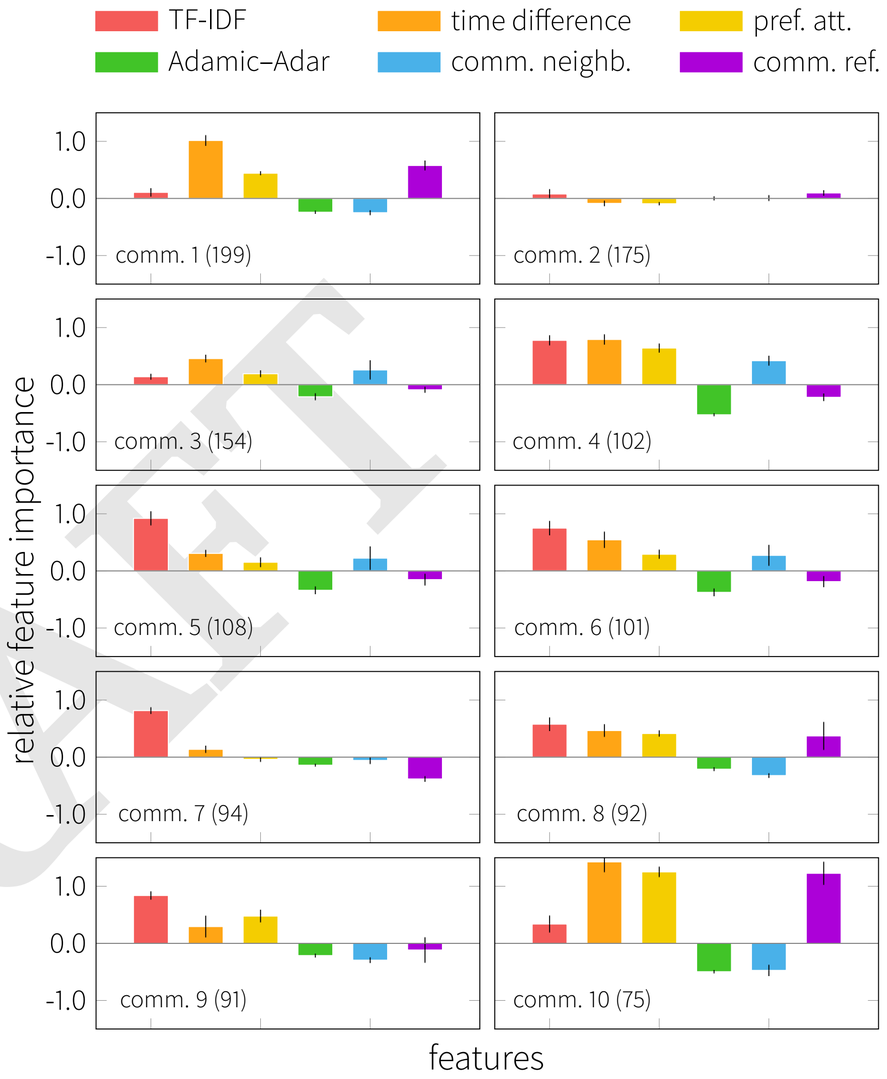


Fig. 6. Community fingerprints. For each community, histograms show the difference in feature importance between the community trained classifier and the globally trained one. Black strokes mark the error of the importance values defined by the standard deviation over the cases in the community.