Finding Relevant Indian Judgments using Dispersion of Citation Network

Akshay Minocha*
IIIT-Hyderabad
Hyderabad, India.

Navjyoti Singh
IIIT-Hyderabad
Hyderabad, India.

Arjit Srivastava
IIIT-Hyderabad
Hyderabad, India.

ABSTRACT

We construct a complex citation network of a subset of Indian Constitutional Articles and the legal judgments that invoke them. We describe, how this dataset is constructed and also introduce the term of *dispersion* [1] from network science related to social networks, in the context of legal relevance. Our research shows that *dispersion* is a decisive structural feature to show the importance of relevant legal judgments and landmark decisions. Our method provides similarity information about the document in question, which otherwise remains undetected by standard citation metrics.

Categories and Subject Descriptors

E.2 [**Data**]: Data Storage Representations—*Linked representations*; H.2.8 [**Database Management**]: Database Applications—*Data Mining*; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Information networks*

General Terms

Legal Networks, Experimentation.

Keywords

Networks in Law, Network Analysis, Social Networks, Database Applications, Information retrieval.

1. INTRODUCTION

Law in India, has reached its current phase after crossing various stages which comprise of influence from religious beliefs, secular legal systems, common law, to the constitutional and legal system which we see today. The number of legal judgments which are readily available for us to use on the web is increasing day by day. In the last couple of years, the Indian judiciary system has undergone an extremely remarkable process of modernization by digitizing cases and developing a platform for storage and management of cases. However, the process is not yet complete, which is to say that a lot of digital reforms need to take place before everyone gets eased up with the complexity of the Indian legal judgments, and details regarding them present on the web. A lot of times, Indian judiciary is criticized for delays with respect to the judgments of cases, but when we take into consideration the ratio of number of cases which are filed in India every year to the number of people in the judiciary department, we realise that the figure is impressive. By 2013, with the help of 1000 fast track courts, 3.2 million cases were decided in 11 years that led up to the time, but even then, there are more than 32 million cases which remain pending, increasing by the year [1].

Very few studies have been done which enquire into the confluence of artificial intelligence and law in the Indian context. In [7] and [6] the authors use statistical measures and connective properties in text to predict the similarity of legal judgments, and on the other front [8], [9] talk about a novel method of legal document summarising and effective retrieval by suggesting that we approach the problem with an ontological perspective.

This paper aims to analyse the structural uniqueness of the legal and regulatory framework in our dataset to look at the problem of legal relevance. By creating a network structure of the references, and by using the basic links/relationships and nodes/judgments, we find empirical and quantifiable indicators of statistical similarities using techniques which were applied on real social networks. We believe that investigating this technique will allow scholars and other legal practitioners, to be able to measure the many features of law and model legal developmental framework as well as to find out legally relevant and fundamentally similar important cases, in practice.

In Section 2, we talk about the related work in the fields of network analysis in law and also introduce the concept of dispersion that we plan to apply to our dataset. Section 3 talks about the methodology which went into the selection and the creation of the dataset that we chose for our experiment, and also the technique of relating structural information from social networks to our legal context. In Section 4, we choose some interesting results and explain how a high value of dispersion between the pairs shows a relationship between them which is otherwise not possible by standard degree and structural methods, and finally in Section 5, we

*akshay.minocha@research.iiit.ac.in

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2744717 .

[1] http://www.newindianexpress.com/nation/Despite-1000-Fast-Track-Courts-32-Millions-Cases-Still-Pending/2013/12/23/article1961278.ece

outline important points from the discussion in brief and also mention some of our plans for future work.

2. RELATED WORK

In this section, we would like to talk about research from two areas of interest: first, the part of network science which deals with the legal network, and second, literature that looks at the structural features of the network to infer properties which are more interesting than the conventional centrality or embeddedness measures that have been adopted in the past.

In [2] the authors focus on Italian regulatory frameworks and the extensive use of references in the same. The authors talk about the need for developing technological tools which are capable of helping a legal professional to handle the regulatory complexity by cross references. Further discussion is about the structural features of the network in terms of the community structure and how we can achieve that by iteratively looking at the edge betweenness.

In [4], the authors try to find out the legal importance of precedents and thus figure out the most legally central and relevant decisions of the Supreme Court of U.S.A. at a given point of time. The authors also pointed out a shortcoming of in-degree centrality which was being used up until that point as an important feature - as a variable which treats all inbound citations equally, i.e, no difference in ranks were taken into account when dealing with a citation from a higher and a lower court, and thus it implies that not all cases are equally positioned to act as a precedent. They, also devise new techniques of inward and outward relevance which are a newer measure of centrality and gives a more clearer picture of individual cases embedded in the network.

Now coming to the second part of our topic of interest we would like to explore the possibility of this interesting research [1], where the authors develop a new measure of tie strength that is termed as *dispersion*. This research was carried out on Facebook networks to predict the link of romantic partnerships on a small subset of individuals of the social media giant. *Dispersion* in simple terms means the, extent to which two individual's mutual friends are not themselves very well connected. Using the network neighborhood information of two linked nodes, dispersion proves to be better in predicting the mentioned result compared to embeddedness which takes into account in total, the common or shared links in the network.

3. METHODOLOGY

3.1 Dataset

In this work, we have considered cases which in some manner invoke the Articles from 264 to 300. These Articles are the XIIth part of the Indian constitution and deal with the subject of finance, property, contracts and suits.

The way the nodes were populated in our dataset involved a three level hierarchy system - where the upper level comprises of the judgments or articles which have been cited by or which cite the documents in the immediately lower level in hierarchy. The citations are verified from a larger pool of data. We construct a directed graph of the network, where an arrow is a citation link, pointing from a citing case to a cited case. The total number of judgments are 35999, and the degree of the graph is 91661. Figure 1, shows a strongly connected subgraph from the whole network in our dataset, the complexity of the network as well as a few nodes with high inbound links are seen, following the fact that real networks follow a type of power law where there are many nodes with low degrees and a few nodes with larger degrees [2].

Figure 1: A Network representation of a strongly connected directed subgraph representing the citation network, from the dataset in 3.1.

This approach based on network analytics also relates to how scholars generally think of law as a web of connected and authoritative set of legal rules and cases, which are the results of the continuous interpretation of the legal rules.

3.2 Dispersion in a Legal Network

As mentioned earlier in [1], the idea was to identify the romantic relationship of a specific user using the structural attributes. Many individuals have large number of connections with other individuals corressponding to a well-defined foci of interaction in their lives. For example, a person is connected to family members, friends from school, co-workers and so on, but it is to be noted that these different groups are well connected amongst themselves showing high embeddedness but they do not necessarily correspond to strong ties in reality. The authors improve their method by using a lot of structural as well as interactional attributes, but in the legal environment there are no interactional features that we could use as such. The research [1] also states that in contrast, this particular individual's relationship partner or other closest friends may have a lower embeddedness, but they will have mutual neighbors from several different foci and clusters, reflecting the fact that the social orbits of these close friends are not bounded by any one focus - consider for example a wife who is connected to some of her husband's coworkers, family and school friends, even though these people belong to different groups and do not know each other. Thus, we are looking towards a dispersed structure to identify the partner instead of prioritizing embeddedness. The limitations of legal embeddedness forced us to think in lines with the theory of social foci [3] and we were able to come up with the analogy.

Suppose a judgment cites many other precedent judgments or articles or legal documents, most of them would occur in clusters or *communities* which individually deal specifically with one area of legal interest. Another Judgment which is connected to and is similar to this node in question, should have similar connections to relevant cases in the same *communities* too, and that is where *dispersion* plays an important role, to identify such documents.

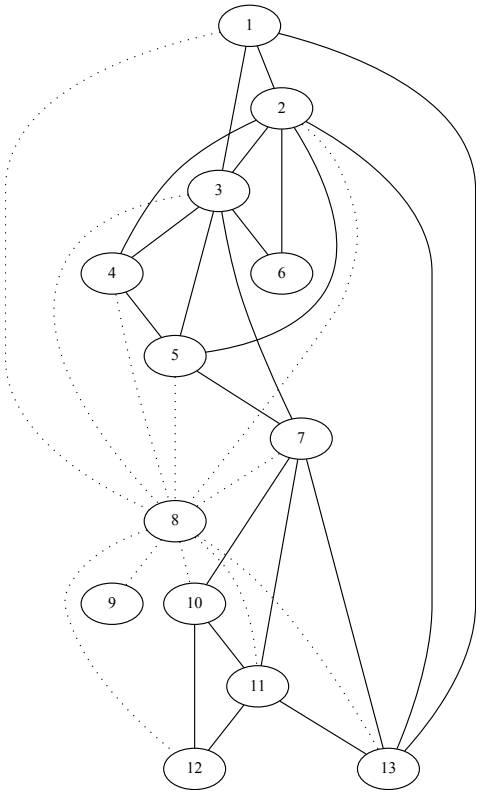

Figure 2: An Example Network representation to explain the concept of dispersion

Lets take an example and relate it with legal factors. In Figure 2 Case1(Node 8) is known to us initially and is our point of enquiry and the source of our *dispersion* score. The figure represents a network of the nodes related to $Case_1$ (G_{C_1}), and the connections between them. Let us say the node that is to be found out in subgraph G_{C_1} is v, which is the target of our *dispersion* score. $C_{C_1 v}$ be the common neighbors for the two nodes. Absolute dispersion would show that the nodes in $C_{C_1 v}$ should be far apart in G_{C_1} ;

$$disp(C_1, v) = \sum_{s,t \in C_{C_1 v}} d_v(s,t) \quad (1)$$

In Equation 1 above, d_v is the distance function, and as in [1] it can be anything, from literal distance to a complex function, for simplicity in absolute dispersion lets assume the value of $d_v(s,t)$ equal to 1 when s and t are not directly connected and 0 otherwise. According to this distance function, in Figure 2, $disp(C_1, Node_{13}) = 2$, since there are two pairs of nodes in $C_{C_1 Node_{13}}$ that are not directly linked and also have no neighbors common in G_{C_1} - $C_1, Nodes_{13}$, and the highest value is for $Node_7$ for which $disp = 4$.

In our results we use the value of *normalized dispersion* which is absolute dispersion normalized over embeddedness of the two nodes.

4. RESULTS

We took the help of the Networkx implementation of dispersion, and calculated the values for our network, and later made a list of maximum valued pairs. In Table 1, a list of five pair of cases are compiled and their analysis is done below. *Disp.* is the normalized dispersion value which is used to make the results compatible for graph-size-independent scale networks. *Emb.* is the embeddedness or the mutual common neighbors of both the cases as mentioned earlier, *tdidf* measures the tf-idf cosine similarity of the two pair which includes an all term formulation. *Indegree and Outdegree* for both cases are represented as well.

The five pair of cases that we randomly selected in Table 1, have comparatively high normalized dispersion in the dataset but represent different conditions where some statistical factors fail to recognize the similarity whereas dispersion does. Instead of selecting a case beforehand, we ran the calculations on the dataset, and discovered pairs which are similar and relevant but would be otherwise hard to find. In the first pair, both revolve along the idea that after reorganisation of the states in India, a lot of properties and businesses ended up being a part of a different state, so rules specific to new states were to act on them. However, the point to be noticed here is the existence of a high dispersion value, inspite of the fact that the value of embeddedness is only 4.

In the Second pair we see that only RF 1973 SC1461 (Decision of Supreme Court) is common to both of them - one case deals with Sales Tax disputes, while in the other Articles regarding income tax are invoked. However, what we also can infer from the structure of both the cases is the fact that they challenge the appeals of the officers. A comparative low score by tf-idf all term similarity score, shows possibily less related content. In the third pair, the cases discuss the Presidential proclamation of emergency and their ambit of judicial review. Article 356 of the Indian constitution which confers power on the President to impose emergency in case of failure of constitutional machinery in State is majorly discussed in the cases. The dispersion value of 22, justifies what we know about those cases and how their structure is inherently the same. The common nodes represent the common outward links, as the newer (2006) case had no inward links, following the fact that law develops while having some attachment to history [5]. In the fourth comparison pair : Others vs. State of Kerala, is the case that saved our democracy from crumbling down by preserving the supremacy of the Indian Constitution. It is also known as the case that laid down the *Doctrine of Basic Structure*. Thereby, anything that violates the doctrine of basic structure shall be held unconstitutional and is threatens the basic structure of the Constitution. This case also laid down that the Preamble is an integral part of the Constitution. The fifth pair of case for analysis is: B. Archana Reddy and others vs. State Of A.P., and K.C. Vasanth Kumar & another vs. State Of

Table 1: Comparing Structural scores across a few pairs

Case 1.	Case 2.	Disp.	Emb.	In degree C1.	In Degree C2.	Out Degree C1	Out Degree C2	tfidf
Kunnathat Thathunni Moopil Nair vs The State Of Kerala in 1960	East India Tobacco Co. vs State Of Andhra Pradesh in 1962	21.10	4	268	227	35	21	0.78
Smt. Ujjam Bai vs State Of Uttar Pradesh in 1961	Basheshar Nath vs The Commissioner Of Income Tax Delhi and Rajasthan in 1958	12.56	27	148	245	147	92	0.93
Rameshwar Prasad & Others vs Union Of India in 2006	S.R. Bommai vs Union Of India in 1994	22	45	0	178	125	185	0.97
Kuldip Nayar vs Union Of India & Others in 2006	Kesavananda Bharati vs State Of Kerala in 1973	20.31	44	55	397	182	414	0.92
K.C. Vasanth Kumar vs State Of Karnataka in 1985	B. Archana Reddy And Others. vs State Of A.P, 2005	21	43	43	0	93	132	0.91

Karnataka. The dispersion value of 21, shows strong base for both of the cases - and both the cases deal with inclusion-exclusion of backward classes in a society in two different states.

Table 2: Coverage of Mean normalized dispersion and Mean tf-idf similarity scores on 3714 pairs.

Set No.	Mean Disp.	Mean tf-idf sim.
1	0.216	0.753
2	0.463	0.675
3	0.573	0.665
4	0.795	0.778
5	1	0.749
6	1.33	0.762
7	1.71	0.762
8	2.3	0.791
9	3.36	0.788
10	15.36	0.799

In Table 2, the coverage of the Mean normalized dispersion and Mean tf-idf similarity scores are given on equally divided and sorted (dispersion) sets derrived from 3712 pairs in total, which had non-zero dispersion values and are at Level 1 of the dataset. It can be clearly seen that a non-zero dispersion implies that a moderately high tf-idf similarity is ensured. The variations though increasing mainly, are not constant over the values of changing dispersion as they take into account alot of structural features which are left out of consideration in other methods.

5. CONCLUSION

In this article we offer to look at the problem of finding relevant judgments, with our understanding of the Indian legal procedure. With the digitization of proceedings in India, it is wise to think that our findings try to answer a face of one of the most long standing problems of legal practitioners, which is to figure out appropriate precedents. Also the applications of multi-disciplinary network science has potential in bibliometric based and citation networks. In future we would like to construct many such attributes and opt for a supervised technique which would more efficiently point out the potential of finding similar proceedings and whose results are comparable to ontological methods.

6. REFERENCES

[1] L. Backstrom and J. Kleinberg. Romantic partnerships and the dispersion of social ties: A network analysis of relationship status on facebook. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 831–841. ACM, 2014.

[2] G. Boella, L. Di Caro, A. Violato, and L. Robaldo. Exploiting networks in law.

[3] S. L. Feld. The focused organization of social ties. *American journal of sociology*, pages 1015–1035, 1981.

[4] J. H. Fowler, T. R. Johnson, J. F. Spriggs, S. Jeon, and P. J. Wahlbeck. Network analysis and the law: Measuring the legal importance of precedents at the us supreme court. *Political Analysis*, 15(3):324–346, 2007.

[5] O. W. Holmes. *The common law*. Harvard University Press, 2009.

[6] S. Kumar. *Similarity Analysis of Legal Judgments and applying Paragraph-linkto Find Similar Legal Judgments.* PhD thesis, International Institute of Information Technology Hyderabad, 2014.

[7] S. Kumar, P. K. Reddy, V. B. Reddy, and A. Singh. Similarity analysis of legal judgments. In *Proceedings of the Fourth Annual ACM Bangalore Conference*, page 17. ACM, 2011.

[8] M. Saravanan, B. Ravindran, and S. Raman. Using legal ontology for query enhancement in generating a document summary. *Frontiers In Artificial Intelligence and Applications*, 165:171, 2007.

[9] M. Saravanan, B. Ravindran, and S. Raman. Improving legal information retrieval using an ontological framework. *Artificial Intelligence and Law*, 17(2):101–124, 2009.