Structure and Dynamics of Signed Citation Networks

Srijan Kumar
University of Maryland
srijan@cs.umd.edu

ABSTRACT

Citations are important to track and understand the evolution of human knowledge. At the same time, it is widely accepted that all the citations made in a paper are not equal. However, there is no thorough understanding of how citations are created that explicitly criticize or endorse others. In this paper, we do a detailed study of such citations made within the NLP community by differentiating citations into endorsement (positive), criticism (negative) and neutral categories. We analyse this signed network created between papers and between authors for the first time from a social networks perspective. We make many observations – we find that the citations follow a heavy-tailed distribution and they are created in a way that follows weak balance theory and status theories. Moreover, we find that authors do not change their opinion towards others over time and rarely reciprocate the opinion that they receive. Overall, the paper builds the understanding of the structure and dynamics of positive, negative and neutral citations.

1. INTRODUCTION

Citations help to connect and contrast new research with the already known information. It provides an easy way to track the progress and evolution of the knowledge in a particular domain. While citations a provide measure to quantify the impact of research and researchers, it is also widely accepted that all the citations made in a paper are not equal [8]. However, there is a lack of understanding about the dynamics of creation of various type of citations over time, and how social factors affect it. Existing research looks at each citation in isolation. However, this does not happen in practice as there are many social and scholarly constraints that play a factor while creating any citation. Therefore, modeling these factors efficiently is important to better understand the working of the scientific community in general.

In this paper, we take a look at the dynamics behind creation of various citation sentiments - endorsement or "positive" citations, criticism or "negative" citations and neutral citations, from both the papers' and authors' perspective. We model them as directed and signed networks. Particularly, we look at the distribution of these citations and the triads they create in the paper citation network, and also at the change and reciprocity of sentiment of authors towards each other in the author citation network.

A very recent paper has studied the role of negative citations in the Immunology literature [3], but they do not look at the dynamics of the creation of such citations and the evolution of these networks. Researchers have previously studied dynamics of citation networks [6], but never from signed networks perspective.

2. CREATING SIGNED CITATION NETWORK

In this section, we will describe how we create the signed paper and author citation networks. We use the full text of 3877 articles from the field of Computational Linguistics, that are part of ACL Anthology Reference Corpus (ACL ARC) [2]. Since finding the relevant context of each citation is still an open problem, we only use the citing sentence in order to determine each citation sentiment. After this, we find 25,354 citations between the articles, and only consider these citations for our analysis.

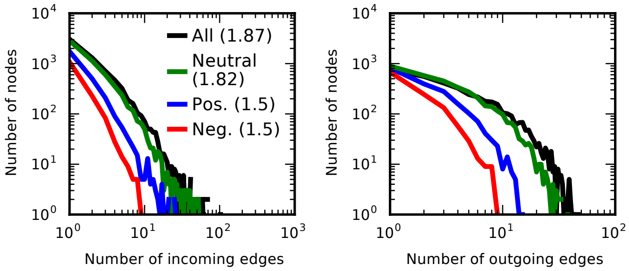

Figure 1: Histograms for the in- and out-edge distributions for the paper citation network, drawn on log-log scale. All types of in-edge distributions follow a power-law distribution (alpha coefficient in bracket), while outgoing edge distributions do not.

To classify the citation sentiment into positive, negative or neutral, we use a simple, yet powerful, keyword-based technique to classify the citation sentiment.[1] We use the list of positive and negative words from [10], which was used as a part of opinion finding system. We manually filtered 173 words that resulted in lots of misclassification of the ground truth citation sentiments, as given in [1]. Each citing sentence is then classified as positively (negatively) citing if it has atleast one positive (negative) word and no negative (positive) words. All sentences that are not in either of these categories are assigned a neutral category. Among the 25,354 citations, 2182 (8.6%) are classified as negative, 5258 (20.7%) positive and remaining 17,914 (70.5%) as neutral. The labelled dataset with the keywords is available for download at the project website.[2]

We also create an author citation network from the paper citation network, which represents the sentiment of authors towards each other. The nodes in this network are authors. The edges are created from each author of the citing paper to each author of the cited paper, with the same sign as the sign between the papers. A pair of authors can have multiple edges between them, one for each paper by one that cites the other. This author citation network has 7,495 authors and 174,448 edges. It has 37,047 (21.2%) positive, 14,812 (8.5%) negative and 122,635 (70.3%) neutral edges.

3. ANALYSIS OF SIGNED NETWORKS

In this section, we look at a few key observations from the signed paper citation and author citation networks.

Power-law distribution of citations: As we observe in both the networks, negative citations form a very small fraction of the total number of citations. This may be because of explicitly criticizing another paper may have negative impact, and therefore, mentioned more subtly. Positive citations occur twice as frequently as negative ones, indicating that papers tend to explicitly praise prior research that they build on. We also observe that majority of the citations are neutral.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889391.

[1] We first implemented the supervised learning framework based on features derived from citation text as described in [1], but it classified all citations into the same class, so we could not use it.

[2] Project website: `http://cs.umd.edu/~srijan/citations/`

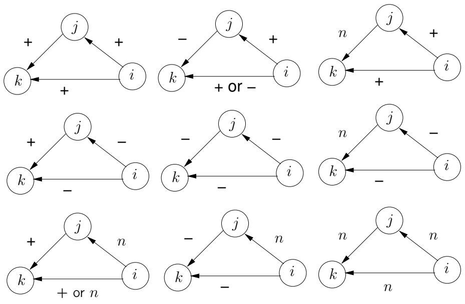

Figure 2: This figure shows all the prominent triads in the paper citation network, when a paper i cites two papers j and k, after j already cites k. The labels $+, -$ and n represent positive, negative and neutral citations, respectively. The figure shows that in most cases, i creates the same citation towards both papers and that weak balance and status theories are followed.

Let us look at the distribution of the edges in the paper citation network. Figures 1(a) and 1(b) show the histogram of the total (without sign), positive, negative and neutral incoming and outgoing edges of all the nodes in the network. It is very interesting to note that the distributions for all the four types of incoming edges individually follow a power-law distribution. The alpha coefficient of the fit is shown within the figure. Similar observation was made for total incoming edges in the DBLP network [4], and for incoming positive and negative edges in other signed networks such as Slashdot and Epinion [9]. The distribution for the negative and positive citations have the most negative slope, indicating a decreased tendency of papers to receive many explicit negative or positive citations. At the same time, out-edges do not follow power-law distribution, possibly due to the limited number of references that are given in each paper.

Citation network follows weak balance and status theories: Here we look at triads, one of the major building blocks of networks, to understand how citations are created over time. Since citations can only be formed to papers published earlier chronologically, only transitive triads can be formed between three papers, i, j and k, as follows – first j cites another paper k, and then i cites both j and k. There are 8,313 triads out of a total of 60,454 possible triads in the paper citation network.

There are many possible configurations of the triads when the citation sentiment is considered. However, only a few of them happen to be created prominently in practice. To find them, we compare the frequency of occurrence of each type of triad with its random baseline distribution – the ones that occur more (less) frequently in practice than random are stable (unstable) in the network and social factors bias towards (against) creating such triads. The random baseline frequencies are found by randomly distributing the signs in the citation network multiple times while keeping the endpoint of edges fixed (10,000 times in our case) and calculating the 95% confidence interval of the frequency of each triad. When the actual number of triads of a particular type lies above the interval, then the triad is over-represented and it is prominent in the network. Figure 2 shows these triads in the paper citation network.

Looking at the prominent triads, we observe that whenever paper i creates an edge to both papers j and k, its sentiment towards both of them is same, in most cases. Moreover, all the triads satisfy weak balance theory [5], which states that triads other than the ones with one negative edge are stable and would occur very frequently in the network. Also, most of the triads satisfy status theory [7] where a positive edge is considered to be pointing towards a node of higher status than the edge originating node (similarly, negative and neutral indicate lower and equal status, respectively). Then according to status theory, triads are created such that edge signs conform to the status. We find that citations are created such that balance and status theories are followed.

Authors rarely reciprocate sentiment: Now, we briefly look at a couple of key observations from the signed author citation network. We want to understand whether authors express similar sentiment towards each other. So, we calculate the average author sentiment by averaging the multiple sentiment from one to other. There are 13,383 author pairs with that cite each other. We find that the Pearson Correlation Coefficient ρ of the sentiment value of author A towards B and from B to A is only 0.277 (p-value = 1.82×10^{-131}). This value is very low, indicating that authors do not reciprocate the sentiment they receive from another author.

First opinion is the average opinion: Here we try to understand the relation between the first sentiment created by an author towards other and their average sentiment. The intuition is to infer how much the opinion of an author changes towards another compared to his/her first opinion towards them. For all 26,641 authors that cite another one more than once, we find that the opinion does not change much (ρ=0.82, p-value=0.0). Therefore, one would expect consistent citations between two authors over the years.

4. CONCLUSION

In this paper, we try to understand the dynamics of the formation of positive, negative and neutral citations in citation networks. We create and analyse paper and author citation networks for the first time as directed signed network. We make many interesting observations. We find that the incoming edges follow power law distribution, while outgoing edges do not. Citations are created in a way that follows weak balance theory and status theory. Moreover, we find that authors do not change their opinion towards others over time and rarely reciprocate the sentiment they receive. Overall, the paper builds the understanding of how positive, negative and neutral citations are created and its social dynamics.

The current study can be improved and extended in many directions. Advanced NLP techniques can be used to infer the sentiment of citations, instead of the simplistic keyword based technique currently used, and citation context can be incorporated. The results of this paper can be used in various tasks, such as to create robust citation metrics, enhanced citation classification functions and realistic citation network evolution model.

Acknowledgements. Parts of this work were supported by ARO grants W911NF11103, W911NF1410358, and W911NF09102. We thank Prof. Amol Deshpande for the helpful discussions.

5. REFERENCES

[1] A. Athar. Sentiment analysis of citations using sentence structure-based features. In *Proceedings of the ACL 2011 student session*, 2011.

[2] S. Bird. The acl anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. *LREC*, 2008.

[3] C. Catalini, N. Lacetera, and A. Oettl. The incidence and role of negative citations in science. *PNAS*, 112(45):13823–13826, 2015.

[4] T. Chakraborty, S. Sikdar, N. Ganguly, and A. Mukherjee. Citation interactions among computer science fields: a quantitative route to the rise and fall of scientific research. *Social Network Analysis and Mining*, 4(1):1–18, 2014.

[5] J. A. Davis. Structural balance, mechanical solidarity, and interpersonal relations. *American Journal of Sociology*, pages 444–462, 1963.

[6] Y.-H. Eom and S. Fortunato. Characterizing and modeling citation dynamics. *PLoS One*, 2011.

[7] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *ACM SIGCHI*, 2010.

[8] M. J. Moravcsik and P. Murugesan. Some results on the function and quality of citations. *Social studies of science*, 5(1):86–92, 1975.

[9] J. Tang, X. Hu, and H. Liu. Is distrust the negation of trust?: the value of distrust in social media. In *25th ACM conference on Hypertext and social media*, 2014.

[10] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *ACL HLT-EMNLP*, 2005.