



# Number and Quality of Diagrams in Scholarly Publications is Associated with Number of Citations

Guy Clarke Marshall<sup>1</sup>(✉)(iD), Caroline Jay<sup>1</sup>(iD), and André Freitas<sup>1,2</sup>(iD)

<sup>1</sup> Department of Computer Science, University of Manchester, Manchester, UK

guy.marshall@postgrad.manchester.ac.uk,  
{caroline.jay, andre.freitas}@manchester.ac.uk

<sup>2</sup> Idiap Research Institute, Martigny, Switzerland

**Abstract.** Diagrams are often used in scholarly communication. We analyse a corpus of diagrams found in scholarly computational linguistics conference proceedings (ACL 2017), and find inclusion of a system diagram to be correlated with higher numbers of citations after three years. Inclusion of more than three diagrams in this 8-page limit conference was found to correlate with a lower citation count. Focusing on neural network system diagrams, we find a correlation between highly cited papers and “good diagramming practice” quantified by level of compliance with a set of diagramming guidelines. This study suggests that diagrams may be a useful source of quality data for predicting citations, and that “graphicacy” is a key skill for scholars with insufficient support at present.

**Keywords:** Neural networks · Scholarly diagrams · Corpus analysis · Bibliometrics · Graphicacy

## 1 Introduction

Diagrams form a part of communications about Artificial Intelligence (AI) systems, such as papers published at the Association of Computational Linguistics (ACL), a top natural language processing (NLP) conference. We argue that system diagrams are an important source of data about scholarly authorship practices in computer science, specifically neural networks (NN) for natural language processing, and have insufficient attention in many academic writing guides. Using “Transactions of ACL 2017” as a corpus, we show that system diagrams are prevalent. We find that papers containing a system diagram are more likely to have a higher number of citations, perhaps indicating that their authors are effective science communicators, or that they write papers about systems, which are more highly cited. Further, papers containing more than two diagrams are found to be more likely to have a lower number of citations, and possible reasons for this are explored.

Corpus analysis of diagrams is nascent, with recent analysis into connecting lines in data visualisations [8]. We use a corpus-based approach to examine diagrams within a wider social context, and have designed our approach to leverage existing document-level citation metrics, allowing quantitative analysis. Our main contribution is to test for compliance with an existing set of neural network system diagram guidelines, using a corpus-based approach. In summary, we find system diagrams are prevalent, occurring in 82% of papers at ACL 2017, and that diagrams in highly cited papers are more likely to contain “good diagrams” in the sense of conforming to an existing set of guidelines [10].

## 2 Background

Natural Language Processing is a discipline within Computer Science, and is concerned with creating systems that solve tasks relating to natural language interpretation. NLP systems take a text input, go through data manipulation steps, and create an output that is usually a classification, ranking, regression or prediction, such as what the next word in a sequence is likely to be. The state-of-the-art systems are technically complex, requiring application of mathematical and algorithmic techniques. These NLP systems are often described through diagrams. We have chosen to examine scholarly neural network systems, described in diagrams within NLP conference proceedings.

Contemporary NLP systems are often based on neural networks, and it is these systems we focus on. A neural network takes an input (in NLP, text), and then processes this via a series of *layers*, to arrive at an output (classification/prediction). Within each layer are a number of *nodes* which are attributes of the representation, and there is a connection between them. Specific mathematical functions or operations are also used in these systems, such as sigmoid, concatenate, softmax, max pooling, and loss. The *system architecture* describes the way in which the components are arranged. Different architectures are used for different types of activities. For example Convolutional Neural Networks (CNN), inspired by the human visual system, are commonly used for processing images. Long Short Term Memory networks (LSTM), a type of Recurrent Neural Network (RNN) which are designed for processing sequences, are often used for text.

These neural networks “learn” a function, but have to be trained to do so. Training consists of providing inputs and expected outputs, allowing the system to develop a representation which can be used for interpretation. The system is then tested with unseen inputs, to verify for generalisation. Diagrams almost always depict the training process. A more detailed introduction to LSTM architectures, including schematics, is provided by Olah [16].

## 3 Method

We use ACL 2017 scholarly papers as a corpus from which to extract diagrams, because it is an appropriate size for analysis (195 long papers), is distributed with

a CC-BY licence, and is recent enough to be relevant whilst allowing for short-term (3 year) citation analysis. Web of Science contains statistics of peer reviewed citations which provides additional robustness to the measure which we use (“Times cited, WoS Core”). Using a chi-squared test we found this metric highly correlated with the less curated “Times cited, All”. Our method follows Lee et al. [9], adapted to use a manual extraction process in order to reduce systematic omission and make use of the validity ensuring method of Lechner [8].

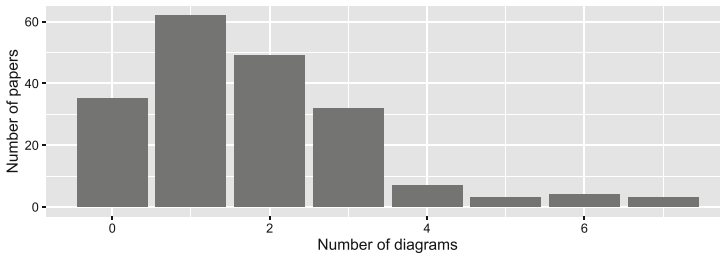
1. Using Web of Science, publication metadata was manually extracted from all long papers from ACL 2017, including number of citations.
2. Every figure that displayed a diagram was manually extracted, except figures in the Results section. We added diagram count as additional paper metadata. The term “diagram” is used to describe a conceptual diagram, usually a figure, which is not reporting results, displayed as a table, nor describing an algorithm. In practice, it encompasses system diagrams, parts of systems, graphical representations of algorithms, concept maps, flow charts of methods or systems and parse trees.
3. Diagrams were stored as separate image files, labelled according to which paper they were extracted from.
4. In each paper, at most one diagram was identified as the primary system diagram. Where multiple system diagrams were found, the one with the largest number of graphical elements was used. Additional metadata was captured, including conformity to each individual guideline, and whether the diagram was colour or monochrome.
5. Following the method of Lechner [8], inter-rater reliability was measured to validate scoring of guidelines compliance, on a subset of 15% of the resulting NN system diagram (17/119). This resulted in 204 pairs of pieces of metadata scored as “true”, “false”, or “not applicable”. Using this, Gwet’s  $AC_1$  coefficient was calculated [7], finding “good” reliability when considering the guidelines as a set. Individual guideline conformity was variable, with Guidelines 2, 4, 7, 10 and 12 (in the ordering presented in [10]) scoring a less than “good” Gwet’s  $AC_1$ , and required further clarification beyond the guideline text alone to agree scoring. Subsequent assessment was done with a single coder. This manual coding resulted in the addition of over 1,600 pieces of diagram metadata, together with 400 additional paper metadata items (diagram count, and system diagram inclusion).
6. The conference area of each paper was manually extracted, as defined by ACL organisers [3].
7. Data were analysed in R [17], using ggplot2 to create graphics [19].

## 4 Results

### 4.1 Diagrams in Context

Figure 1 shows the frequency of diagrams in ACL 2017 proceedings. The large number of papers, particularly highly cited papers, which include system

diagrams demonstrates the importance of system diagrams in communicating at ACL 2017.



**Fig. 1.** Number of (non-results) diagrams in ACL 2017 papers is normally distributed, with inclusion of one or two diagrams most common. Most papers include a diagram.

To summarise the key insights, with correlations using chi squared test:

- 160/195 (82%) of all ACL 2017 papers included diagrams to represent system conceptualisations (not including results or algorithms).
- 124/195 (64%) of all ACL 2017 papers included at least one system diagram.
- Including 1–2 diagrams, of which at least one is a system diagram, is correlated with a 250% higher number of citations.
- Having more than two diagrams is correlated with lower number of citations. In a linear model each additional diagram is correlated with 5.6 fewer citations ( $p = 0.02$ ). In the subset of papers which include a system diagram, this effect increases to 7 fewer citations per additional diagram.
- 82/119 (69%) of NN diagrams used colour, which may affect accessibility.
- Diagrams may be a valuable source of data for modeling number of citations. See Sect. 6.2.

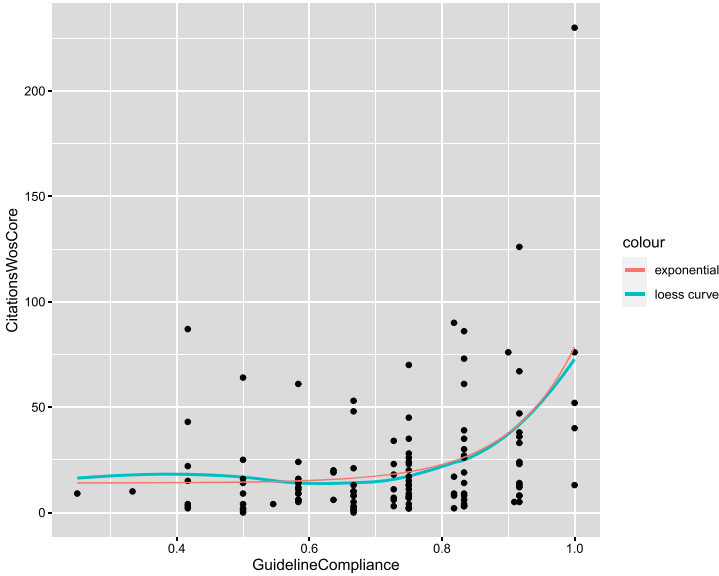
## 4.2 Conference Areas

In an attempt to remove some of the effect of the content of paper, we analysed whether there was a relationship between the 17 conference areas [3] and (i) citations (ii) inclusion of a system diagram (iii) number of diagrams (iv) usage of examples. We found no significant difference between pairs of these attributes using chi-squared tests, using the entire dataset.

To further investigate any potential paper-content-related cause, we found 21 papers contain the word “architecture” and 18 of those contain a system diagram. Number of citations and the abstract containing the word “architecture” are correlated ( $p < 0.01$ ), with those containing “architecture” having on average 20.4 more citations (than 15.9). As would be expected, the abstract containing the word “architecture” and including a system diagram are not independent: There is a significant relationship ( $p < 0.05$ ). Causality is therefore ambiguous, as to whether architectural papers are more likely to be highly cited, or whether it is due to the presence of the diagrams.

### 4.3 Neural Network System Diagram Guideline Conformity

119 of 124 system diagrams described neural network systems (the others being diagrams of an embedding only, or not a neural system). These 119 diagrams were assessed against each of the 12 guidelines established in an interview study [10]. These guidelines were chosen in favour of other diagramming guidelines due to their domain specificity.



**Fig. 2.** Scatter plot of number of citations versus NN system diagram guideline compliance (as a quantitative proxy for “how good the diagram is”). LOESS curve for locally weighted smoothing is in blue, and the function  $y = e^{10(x-7/12)} + 14$  is in red. (Color figure online)

In this exploratory analysis, we found a correlation between number of citations and “specific” ( $p < 0.05$ ), and also “self contained” ( $p < 0.05$ ) guidelines. The other guidelines alone did not correlate with a significant difference in number of citations. However, the best correlation was found with an average of the guideline compliance. The LOESS curve in Fig. 2 can be approximated by an exponential function,  $citations = e^{10(compliance-7/12)} + 14$ , where “7/12” captures the increase in citations observed from higher levels of compliance, “14” captures the asymptotic average number of citations for low compliance papers, and the multiplier “10” fits the curve. The only independent variable is average guideline compliance of each diagram. We do not aim to model citations accurately, arguing instead that the guidelines capture aspects of diagramming behaviours of effective communicators. Data has been made available [11].

## 5 Related Work

Much attention is given to the automated extraction of information from scholarly figures, including the classification of charts into bar charts, pie charts, etc. Roy et al. [18] recently created a classification system for neural network system diagrams. Their system classifies deep learning architectures into six categories, (e.g. 2D boxes, pipeline) based on how the layers are visually represented. This, and many other scholarly processing systems, rely on pdffigures 2.0 [4] for diagram extraction, which has known limitations and edge-cases. In particular, some types of figure (such as those with an L-shape) are systematically omitted [5]. Manual classification of NN system diagrams has been conducted based on mental model categories [13] and semiotics [14], and VisDNA has been applied to neural network system diagrams [15]. Marshall et al. [10] conducted an interview study on the role of diagrams in scholarly AI papers, which reported 12/12 participants using diagrams to get a summary of the paper, and found some participants (3/12) used the diagram before any text in the paper. The potential role of NN diagrams in improving scholarly communication has also been explored [12].

## 6 Discussion

### 6.1 Limitations

- Our corpus analysis is based on one year of one venue, and cannot be generalised.
- The manual data extraction process does not scale well.
- Number of citations can be affected by many other factors, including author institution, author name, twitter presence, and so on. We mitigate venue by restricting to one venue. We do not take action to reduce the impact of other factors, focusing analysis on features of diagrams.
- Our inter-rater reliability covered only guideline conformity, not the diagram extraction or classification of figures.
- Whilst using the guidelines alone provided “good” inter-rater reliability, raters needed to make subjective judgements, and required more than the guidelines alone to ensure replicability.
- Unlike Lee et al. [9], we examine only diagrams, not all figures.

### 6.2 Using Diagrams in Models to Predict Number of Citations

A simple linear model for number of citations in the ACL 2017 corpus can be made using only (i) whether the abstract contains the word “architecture” and (ii) the level of conformity of the system diagram to a set of guidelines (zero if a system diagram is absent). This model has an R-squared of 0.132, suggesting that 13.2% of the variation in number of citations can be explained by these two factors alone. This simple model performs comparatively to existing

citation predictions based on the entire text of the paper [20], which report 0.13 R-squared on a different scholarly corpus. Richer state-of-the-art models using social variables have R-squared around 0.4 at the 3rd year, again on a different scholarly corpus [1]. This supports the claim of the utility of diagrams in the medical-centric “viziometrics” research agenda of Lee et al. [9] and suggests figures may be a underutilised data source more broadly for scientometrics.

### 6.3 The Potential of Scholarly Diagrams

Our results provide further motivation for improved scientific scholarly graphicacy, the benefits of which to are often pedagogically focused. “Drawing to learn” is an active research area [2], and studies have been conducted concerning benefits of drawing for scientific thinking specifically (see Fan [6]). Our findings support the centrality of diagrams in scholarly communication previously identified in Medical Science [21], and lends weight to the reported primacy for some users of diagrams within the AI scholarly context [10].

## 7 Conclusion

Diagrams are an important, prevalent, and neglected component of scholarly communication about neural network systems, and diagramming is not proportionately discussed in many scholarly writing guides. At ACL 2017, high quantities of diagrams were found to be correlated with lower numbers of citations. Usage of system diagrams was found to be correlated with higher numbers of citations, suggesting this is a good scholarly communication practice in this domain. We have shown good domain-specific diagramming practices, quantified by compliance with a set of guidelines, to be correlated with a higher number of citations for ACL 2017 papers. This study demonstrates diagrams are important for communicating about scholarly neural network systems, and may be an underutilised tool for understanding and improving scholarly communication.

## References

1. Abrishami, A., Aliakbary, S.: Predicting citation counts based on deep neural network learning techniques. *J. Informetr.* **13**(2), 485–499 (2019)
2. Ainsworth, S.E., Scheiter, K.: Learning by drawing visual representations: potential, purposes, and practical implications. *Curr. Dir. Psychol. Sci.* **30**(1), 61–67 (2021)
3. Barzilay, R., Kan, M.Y.: Accepted papers, demonstrations and TACL articles for ACL 2017 (2017). <https://acl2017.wordpress.com/2017/04/05/accepted-papers-and-demonstrations/>. Accessed 08 Jan 2021
4. Clark, C., Divvala, S.: Pdffigures 2.0: mining figures from research papers. In: 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL), pp. 143–152. IEEE (2016)
5. Clark, C., Divvala, S.: Pdffigures 2.0 readme (2016). <https://github.com/allenai/pdffigures2/blob/master/README.md>. Accessed 05 Mar 2021

6. Fan, J.E.: Drawing to learn: how producing graphical representations enhances scientific thinking. *Transl. Issues Psychol. Sci.* **1**(2), 170 (2015)
7. Gwet, K.L.: *Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC (2014)
8. Lechner, V.E.: Modality and uncertainty in data visualizations: a corpus approach to the use of connecting lines. In: Pietarinen, A.-V., Chapman, P., Bosveld-de Smet, L., Giardino, V., Corter, J., Linker, S. (eds.) *Diagrams 2020*. LNCS (LNAI), vol. 12169, pp. 110–127. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-54249-8\\_9](https://doi.org/10.1007/978-3-030-54249-8_9)
9. Lee, P.S., West, J.D., Howe, B.: Viziometrics: analyzing visual information in the scientific literature. *IEEE Trans. Big Data* **4**(1), 117–129 (2017)
10. Marshall, G., Freitas, A., Jay, C.: How researchers use diagrams in communicating neural network systems. arXiv preprint [arXiv:2008.12566](https://arxiv.org/abs/2008.12566) (2020)
11. Marshall, G., Jay, C., Freitas, A.: Data supporting “number and quality of diagrams in scholarly publications is associated with number of citations” (2021). <http://dx.doi.org/10.6084/m9.figshare.14812959>
12. Marshall, G.C., Jay, C., Freitas, A.: Diagrammatic summaries for neural architectures. In: *Beyond Static Papers: Rethinking how we Share Scientific Understanding in ML-ICLR 2021 Workshop* (2021)
13. Marshall, G.C., Jay, C., Freitas, A.: Scholarly AI system diagrams as an access point to mental models. arXiv preprint [arXiv:2104.14811](https://arxiv.org/abs/2104.14811) (2021)
14. Marshall, G.C., Jay, C., Freitas, A.: Structuralist analysis for neural network system diagrams. In: *International Conference on Theory and Application of Diagrams*. Springer (2021)
15. Marshall, G.C., Jay, C., Freitas, A.: Understanding scholarly neural network system diagrams through application of VisDNA. In: *International Conference on Theory and Application of Diagrams*. Springer (2021)
16. Olah, C.: Understanding LSTM networks (2015). <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Accessed 22 May 2020
17. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2020). <https://www.R-project.org/>
18. Roy, A., et al.: Diag2graph: representing deep learning diagrams in research papers as knowledge graphs. In: *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 2581–2585. IEEE (2020)
19. Wickham, H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York (2016). <https://doi.org/10.1007/978-0-387-98141-3>, <https://ggplot2.tidyverse.org>
20. Yan, R., Huang, C., Tang, J., Zhang, Y., Li, X.: To better stand on the shoulder of giants. In: *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 51–60 (2012)
21. Yang, S.T., et al.: Identifying the central figure of a scientific paper. In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1063–1070. IEEE (2019)