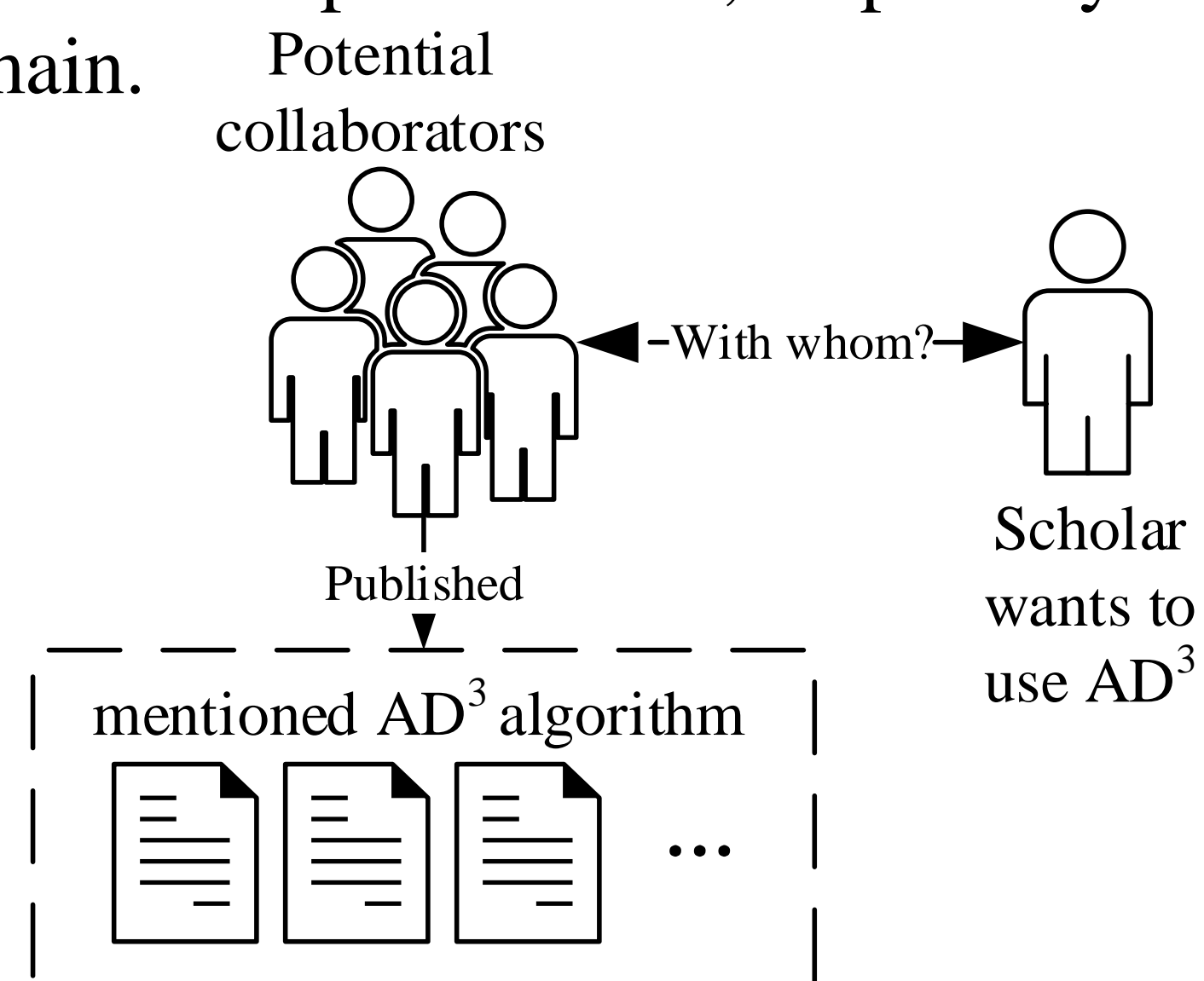# A Bipartite Author-algorithm Network for Exploring Versatile Scholars

Yi Zhao, Yuzhuo Wang, Chengzhi Zhang

Department of Information Management, Nanjing University of Science and Technology, Nanjing, China

Introduction

☐ The need for particular skills is one of the factors that motivate scholars to find collaborators.

☐ It's difficult to know directly what fine-grained skills (e.g., algorithm) a scholar masters, but this kind of information is deeply embedded in scholars' publications, especially in the artificial intelligence domain.

☐ It is essential to build a connection between authors and algorithms. In this paper, we briefly
  1) construct a bipartite author-algorithm network
  2) explore the scholars with the most diverse skills (i.e., versatile scholars) in the NLP domain.

Methodology

Dataset

☐ Our dataset consists of two parts:

✓ Algorithm entities: we used the in-house dataset annotated by Wang and Zhang (2020).

✓ Author information: we first downloaded and parsed the metadata from the dataset (https://github.com/lingo-iitgn/NLPExplorer). Then, we extracted authors' affiliations, affiliation locations from each paper manually.

✓ We integrated the two parts through paper IDs provided by the ACL Anthology (https://www.aclweb.org/anthology/).

☐ The ACL Anthology volunteer team has used multiple approaches to address name ambiguities (Mohammad 2020).

Method

☐ effective partners(EPs):

$$H_j = -\sum_{i=1}^{s} \frac{a_{ij}}{a_{.j}} log_2 \frac{a_{ij}}{a_{.j}}$$

$$EPs_j = \begin{cases} 2^{H_j} \\ 0, if\ a_{.j} = 0 \end{cases}$$

a_{ij} is the number of co-occurrences between author j and algorithm i, $a_{.j}$ is the total number of co-occurrences between author j and all algorithms. Eps_j is the effective number of algorithms mentioned by author j.

☐ Bipartite network:

$$G = (U, V, E)$$

U and V are two types of nodes in network G. U and V denote the authors and algorithms respectively. E is the set of edges in G.

Results

Statistics of Dataset

Table 1. presents the basic statistical description of our dataset.

Table 1. Statistics of the dataset

	Author	Affiliation	Country	Algorithm
Total	12,142	12,711	12,711	59,277
Unique	5,619	1,034	65	891

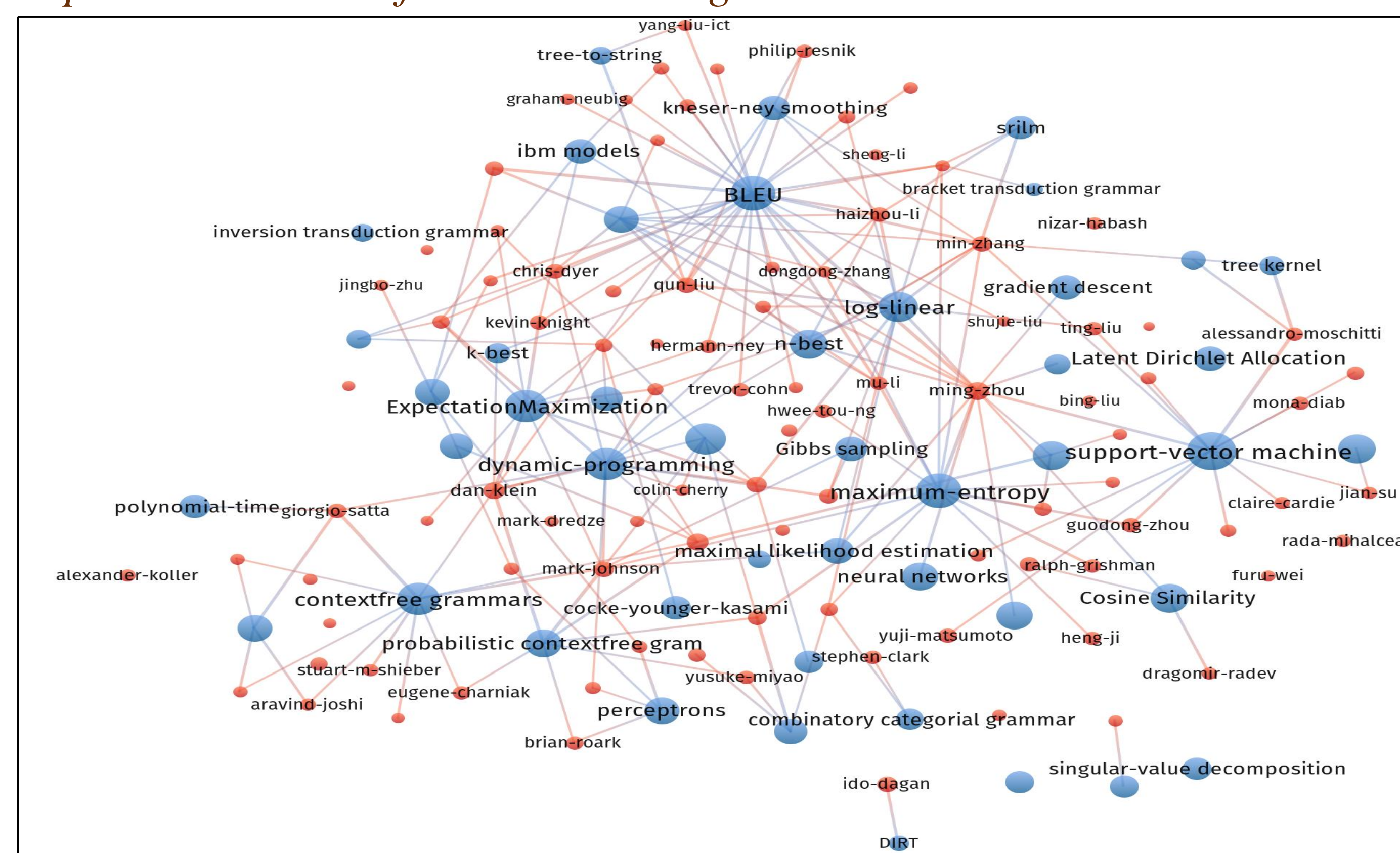### *Bipartite network of authors and algorithm entities in the NLP domain*

Figure 1. Bipartite network of authors and algorithm entities

☐ The author node is red, and the algorithm node is blue. A link between the author and the algorithm exists if the author has published a paper that mentioned the algorithm. The thickness of the link represents the number of papers an author published that mentioned an algorithm. The size of the node denotes the number of papers an author published(red) or the number of papers published that mentioned an algorithm(blue).

☐ Support vector machine, BLEU and maximum entropy are widely mentioned by authors.

☐ 4.52 types of algorithms are mentioned per paper, whereas the average number of publications per author is 2.14.

☐ For Ming Zhou, a famous NLP scholar, 117 types of algorithms are extracted from his papers.

The most versatile scholars in the NLP domain

☐ Eps provides interesting rankings for authors who have mentioned or even used a variety of algorithms in their papers.

☐ Most of the top-5 versatile scholars are influential scholars in the NLP domain: ACL Fellows (i.e., Christopher Manning, Noah Smith), ACL president (i.e., Ming Zhou).

☐ Four of the five authors are affiliated with elite universities in the USA, while Ming Zhou is affiliated with an excellent firm that is located in China.

Table 2. The top-5 versatile scholars

Author	EPs	Affiliation	Location
Christopher Manning	83.34	Stanford University	USA
Ming Zhou	77.76	Microsoft Research Asia	China
Noah Smith	76.96	University of Washington	USA
Daniel Klein	75.54	University of California at Berkeley	USA
Chris Dyer	65.03	Carnegie Mellon University	USA

Conclusion

☐ The author-algorithm network displays that each author has mentioned what types of algorithms in their papers.

☐ We find that 4 of the 5 versatile scholars are from the US

☐ The validity of the author-algorithm network to facilitate the scientific collaboration will be demonstrated in the future.

References

Bersier, L. F., Banassek-Richter, C., & Cattin, M. F. (2002). Quantitative descriptors of food-web matrices. *Ecology,* 83(9), 2394-2407.

Mohammad, S. M. (2020). Gender Gap in Natural Language Processing Research: Disparities in Authorship and Citations. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*(pp.7860–7870). Online: ACL.

Wang, Y., & Zhang, C. (2020). Using the full-text content of academic articles to identify and evaluate algorithm entities in the domain of natural language processing. *Journal of Informetrics,* 14(4), 101091.

Contact

Yi Zhao (yizhao93@njust.edu.cn)

Chengzhi Zhang (Corresponding author, zhangcz@njust.edu.cn)