9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Computational Framework of AST Model Reveals Mechanism of Knowledge Accumulation

Keywords: Knowledge production, AST model, coordination, empirical analysis, Wikipedia

Extended Abstract

Knowledge has been a vital driver of our modern society and has been accumulated through collaborations. The advantage of collaboration in knowledge accumulation is its mutual complementarity; When the information added by an individual is incomplete, others can modify that incomplete information by subtracting or adding information. Through this additive and subtractive transformation (AST), we collaborate with others to create and accumulate knowledge. Recent studies find that AST can capture the cognitive state of participants in creative tasks and points out human have an imbalance in their cognitive perception of the subtractive and additive that "people People systematically overlook subtractive changes" [1].

However, even though it is obvious that knowledge accumulation is done through AST deeply related to some cognitive biases, we little know about the contribution of both additive and subtractive transformation and the imbalance between them to the final output (knowledge) in a real-world situation due to the lack of a computational tool for empirical analysis. More specifically, the interplay between the additive and subtractive transformations in the collaboration for knowledge production remains unclear. The literature investigates the role of collaboration in knowledge accumulation, studying its vital drivers [2, 3], its embedded conflicts or biases [4], or its dynamics [5], but we know a little about how collaborators build the knowledge sharing their AST tasks.

To obtain empirical evidence of the role of AST, this study proposes a computational model of AST that allows us to conduct an empirical analysis with massive data on human behavior observed in digital platforms. We apply our model to the Wikipedia edit history that records the additive and subtractive transformation (edit) of Wikipedia articles. Figure 1 illustrates that. The proposed framework first constructs a bipartite graph to model the AST in which participants contribute additive and subtractive transformations of Wikipedia article edits (Fig. 1(a)). We then utilize the word embedding algorithm (SGNS) to obtain the representation of additive and subtractive transformation from such bipartite graph for each article, meaning that we obtain two embedding representations on Additive-side (A-side) and Subtractive-side (S-side) for each article (Fig. 1(b)).

With the proposed computational framework, we investigate the AST of popular Wikipedia articles. To understand the behavior of the participants (users) in the AST, we construct the network where each node represents the transformation of the article (either A or S side). In the network, we connect nodes when they share the participants calculating the similarity between nodes based on the obtained embedding vectors. We analyze the constructed network and find that more people contribute on the A-side than the S-side, and plot the degree distribution of the graph (Fig. 2 (a-b)). The longer tail of the degree of subtractive nodes (orange) suggests that the transformation (edit) on S-side is more collective than A-side. The connectivity of the S-side is also stronger. Figure 2 (c) demonstrates that removing the subtractive nodes increases the diameter of the network more than the additive nodes.

The above two findings suggest that the A and S sides are different in their collectiveness and it is raising an intriguing question: *is this difference connected to the quality of output?* To answer this question, we measure the disparity between the A and S sides by inverse cosine similarity between the embedding vectors of the A and S sides for the same article. To refer to this inverse cosine similarity, we use "Division of Labor Index (DLI)". A large DLI of a given article means that the users who did the additive transformation and subtractive transformation even though they transform the same article.

We then study the relationship between DLIs and the quality of articles by looking at high-quality articles on Wikipedia. To select high-quality articles, we use the annotation from Wikipedia that provides "Featured articles" and "Good articles"[1] annotated by the editors of Wikipedia following the objective and taught criteria[2]. Using those annotations, we will compare the differences in DLI: Featured articles, Good articles, and the others. Figure 3 demonstrates that the Featured articles and Good articles have high DLIs than the Normal articles, suggesting that the high-quality articles tend to have different users in their A and S sides. This result implies that the division of labor plays a pivotal role in generating high-quality articles. On the other hand, we do not see a meaningful difference in DLI between Featured and Good articles. Given that the articles annotated as feature articles are top-notch articles in Wikipedia, our finding suggests that the division of labor is vital to generate high-quality articles, but other factors make them the best of the best articles.

Our computational framework of the AST model allowed us to reveal the mechanism and consequence of AST in knowledge production. Our analysis contributes to our knowledge of how AST works in a real-world setting and advances this emerging important realm of interdisciplinary research. Importantly, our empirical analysis confirmed in the imbalances between the additive and subtractive transformation found in the laboratory experiments conducted by [1]. In this study, we found i) different user groups contribute to A and S sides, and ii) the editing of the high-quality articles is carried out through a high degree of division of labor compared to the normal quality articles. Our investigation also identifies iii) the users who divide their labor in editing to have different interests.

References

[1] Gabrielle S Adams, Benjamin A Converse, Andrew H Hales, and Leidy E Klotz. People systematically overlook subtractive changes. *Nature*, 592(7853):258–261, 2021.

[2] Goran Murić, Andres Abeliuk, Kristina Lerman, and Emilio Ferrara. Collaboration drives individual productivity. In *CSCW*, 2019.

[3] Edward L Platt and Daniel M Romero. Network structure, efficiency, and performance in wikiprojects. In *ICWSM*, 2018.

[4] Feng Shi, Misha Teplitskiy, Eamon Duede, and James A Evans. The wisdom of polarized crowds. *Nature human behaviour*, 3(4):329–336, 2019.

[5] Khandaker Tasnim Huq and Giovanni Luca Ciampaglia. Characterizing opinion dynamics and group decision making in wikipedia content discussions. In *WWW*, pages 632–639. ACM, 2021.

[1] `https://en.wikipedia.org/wiki/Wikipedia:Quality_articles`

[2] `https://en.wikipedia.org/wiki/Wikipedia:Good_article_criteria;https://en.wikipedia.org/wiki/Wikipedia:Featured_article_criteria`

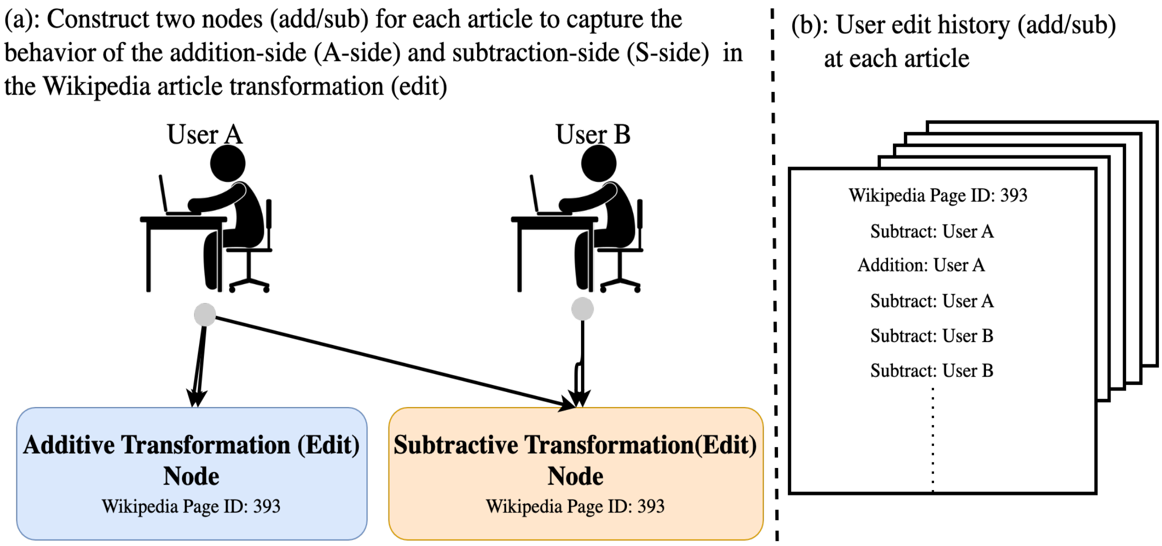

Figure 1: Schematic illustration of modeling "two-sided" nature in editing Wikipedia articles

Note: This figure illustrates the schematic of our model to obtain embedding vectors of Wikipedia articles.

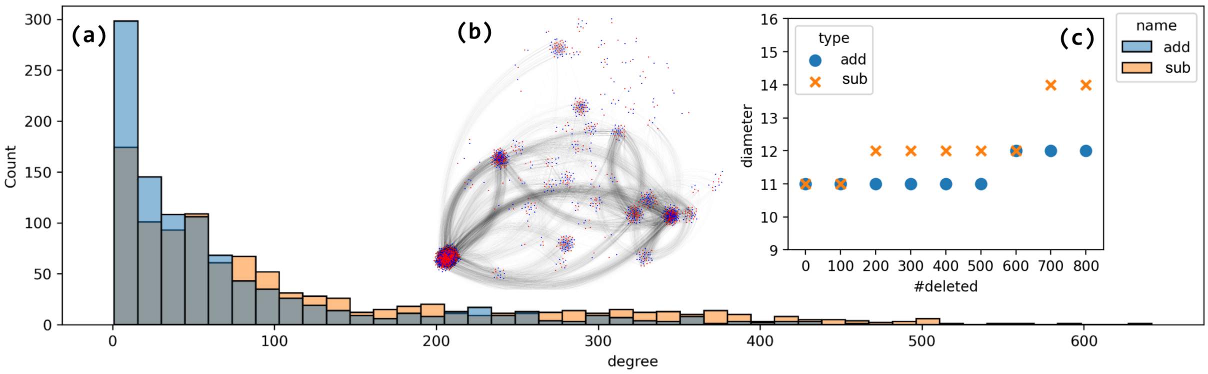

Figure 2: Anatomy of the similarity graph of addition and subtractive nodes in Wikipedia

Note: Analyzing the similarity graph by the Wikipedia article editing vectors to reveal the characteristic differences of the users engaged in different roles **(a-b)**. **(c)** shows the result of the anatomy of the network by the dismantling procedure that removes the nodes with a high degree from the graph and calculates the diameter of that graph.

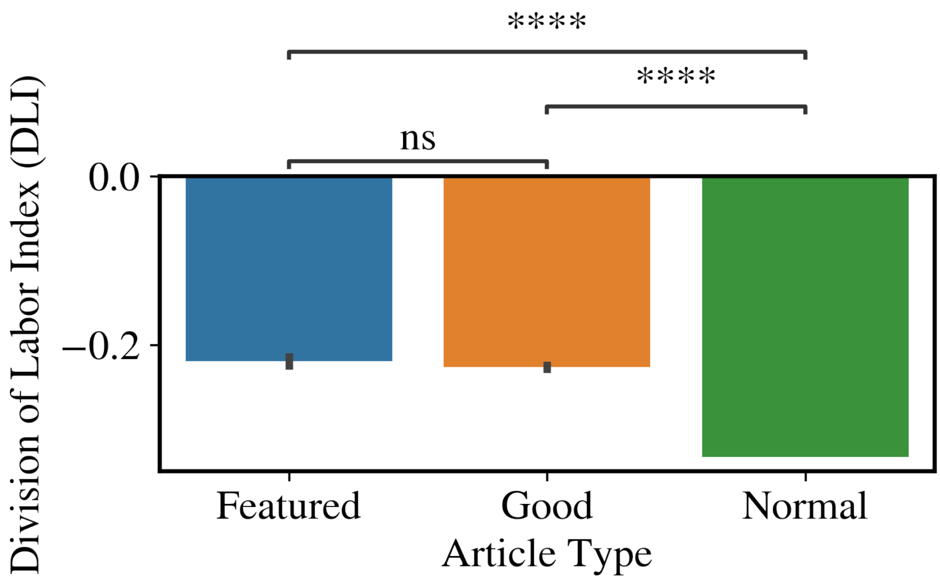

Figure 3: Imbalance of editing and its quality

Note: We calculate the DLI for each article. ; ****: p-val$<$ 0.0001, ns: p-val $>$ 0.05 (Welch's t-test)