

A Understanding document modification behaviors

In this section, we focus on understanding document modification behaviors E_{on} . Such behaviors could be inferred by comparing consecutive document versions via the Myers difference algorithm [66].

A.1 Task Analysis

In our interviews with communication researchers, there are two tasks about analyzing document modification behaviors:

T1: identify editing dynamics. Communication researchers hypothesize that NS and NNS contribute differently to the joint document and would like to know the difference reflected by the evolution of documents.

T2: analyze specific content. Communication researchers are interested in finding frequently co-edited content and seeing how NS and NNS edit them during collaborative writing.

A.2 Understanding editing dynamics by branched time curves

To tackle T1, we draw inspirations from the folded time curve proposed by Bach et al. [5], where each dot represents a document version, the order on the curve denotes the temporal information, and the pairwise distance depends on the text similarity. We proposed a modified version as depicted in Figure 8: since authors write individually before collaborating on the same document, we adopted a two-head start, indicated by the gray arrows, which refer to the initial draft written by NS and NNS. Then, we use red and blue to refer to NNS and NS' contributions, respectively. The communication researchers suggest highlighting the differences between intermediate document versions V_{inter} and end-of-the-day versions V_{end} , so we apply big filled dots for V_{end} and small hollow dots for V_{inter} .

The branched time curves give an overview of the document's evolution, especially the "V" shape part, which shows how the merged version differs from NNS and NS' initial draft. Inspired by the proximity of the merged version to NS' version, in a recently published paper at a premier social computing conference (not cited here for anonymity), our collaborators conducted a statistical significance test and found that the merged document's lexical distance was indeed significantly closer to NS' initial writing. Meanwhile, in subsequent versions, NS' edits always result in a larger lexical distance than NNS.

A.3 Tracking specific content with Sentence Flow

Though the time curves provide an overview on the document-level contributions from NS and NNS, the coarse granularity limits us from uncovering more content-specific insights (T2). Inspired by history flow [96, 97], we propose a revised version called Sentence Flow, as depicted in Figure 9. The x-axis is the author; the y-axis are individual sentences in the document. The edits of a sentence is color-coded: white for no activity, blue and red for word-level add and remove, and black for deleting the entire sentence. The darkness of red and blue indicates the edit distance, the darker the larger. On top of the sentence flow, there are also bar charts

quantify the total edit distance. If users click on a sentence, they can see the content and change logs.

Our collaborators identified two dominant editing patterns in sentence flow. One is within-author editing, where authors primarily revise their own sentences and rarely modify others', e.g., in Team 1, adjacent colored areas are uncommon, with only one instance shown in Figure 9, where the NS made minor modifications to NNS' sentences. The second pattern is between-author editing, as seen in Team 9 in Figure 9, where several sentences were consecutively revised by different authors, resulting in a more balanced distribution.

Though the modified versions of existing visualizations reveal document modification behaviors of NS and NNS, it does not answer questions about how content are generated in collaborative writing. To address these questions, we conduct and present an event sequence analysis of content generation behaviors in the next section.

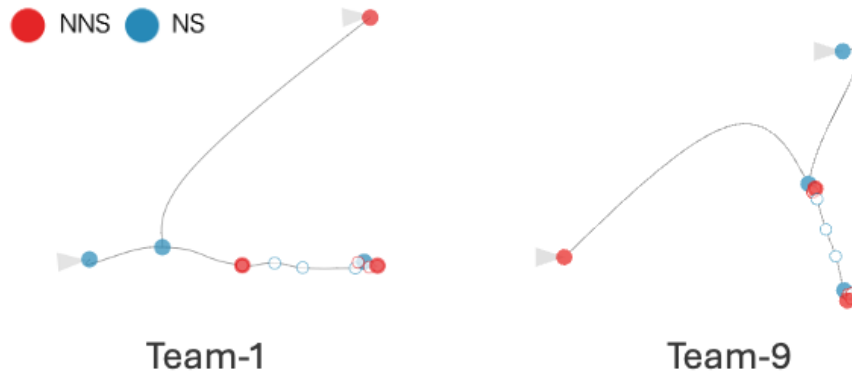


Figure 8: Branched time curves of team-1 and team-9. The gray arrows indicate the first versions written by NS and NNS. Big and filled dots are major versions, and small and hollow dots are intermediate versions.

add remove

*****3-NNS to 4-NS*****

- Who knows they may return the favor too, either with a genuine thank you or a comment on your own social media **accounts**.
- Who knows they may return the favor too, either with a genuine thank you or a **like**/comment on your own social media **feed**.

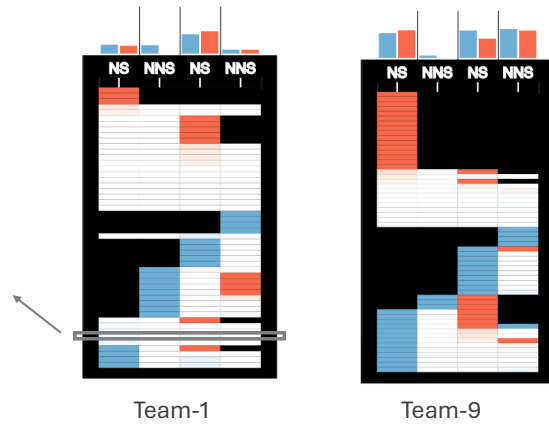


Figure 9: Sentence Flow of team-1 and team-9. Red: remove, blue: add, gray: edited by both authors.