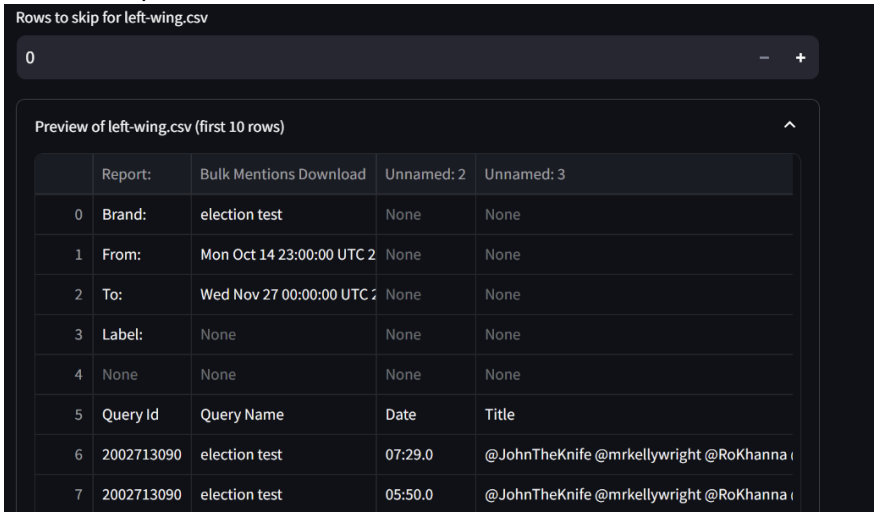
Overview

This tool enables you to:

1. Upload one or more CSV files and preview and process the data.
2. Select source, target, and optional attribute columns for network graph creation.
3. Apply processing to columns, such as extracting hashtags or domains.
4. Create and export network graphs in GEXF or CSV formats for further analysis in Gephi.

Step 1: Upload and Process CSVs

1. **Upload Your Files**:
    - Use the "Upload CSV files" button to select one or more CSV files from your computer. Multiple files can be uploaded at once, which can be merged later.
2. **Adjust Settings and Preview Data**:
    - For each uploaded file, you'll see an option to specify **rows to skip** (e.g., header rows or irrelevant rows).
    - This is due to how some CSV exports (namely BrandWatch) provide CSVs that are poorly formatted at the start. See the screenshot example below – in this case we'll need to skip the first 6 rows.



	Report:	Bulk Mentions Download	Unnamed: 2	Unnamed: 3
0	Brand:	election test	None	None
1	From:	Mon Oct 14 23:00:00 UTC 2	None	None
2	To:	Wed Nov 27 00:00:00 UTC 2	None	None
3	Label:	None	None	None
4	None	None	None	None
5	Query Id	Query Name	Date	Title
6	2002713090	election test	07:29.0	@JohnTheKnife @mrkellywright @RoKhanna
7	2002713090	election test	05:50.0	@JohnTheKnife @mrkellywright @RoKhanna

    - It is possible to click a preview of each CSV uploaded to show how many rows to skip.
3. **Process and Merge Files**:
    - Click the "Process CSV" button (or the "Process and Merge" button for multiple CSVs) once the right number of rows has been skipped and the files are ready to load.

Step 2: Select Columns and Processing Methods

1. Select Source and Target Columns:
    - Choose the Source column (the origin of relationships) from the dropdown menu.
    - Choose the Target column (the destination of relationships) from the dropdown menu.
    - For example, in an X dataset, the Source column could be "Author" (people who are initiating an interaction), and the Target column could be "Mentioned Users" (who they are interacting with).

- Each column can be processed using the following options:
  - **No Processing:** Use the data as-is.
  - **Free Text - Hashtags:** Extract hashtags from free text.
  - **Free Text - Domains:** Extract domain names from URLs in free text. Note: URLs are not included in BrandWatch's 'Full Text' column. Use 'Expanded URLs' instead with the 'Comma Separated List - Domains' value.
  - **Free Text - Mentioned Users:** Extract mentions of usernames (only when usernames are denoted with '@' – e.g. @elonmusk).
  - **Comma Separated List - Hashtags:** Extract hashtags from a comma-separated list (e.g. some datasets include a 'Hashtags' column with a list of hashtags used in each post like this: #electionfraud, #fraud, #rigged).
  - **Comma Separated List - Domains:** Extract Domains from a comma-separated list.
  - **Comma Separated List - Mentioned Users:** Extract mentioned users from a comma-separated list.

2. Optional Attributes:
   - Select additional columns to include as node attributes. These attributes will appear in the exported network graph and can be used for further analysis.
3. Process Columns:
   - Click "Process Columns" to apply the selected processing options and clean the data for graph creation.
   - A preview of the processed dataset will be displayed.

Step 3: Create and Export Network Graph

1. **Select Graph Type**:
   - Choose the type of network graph:
     - **Directed**: Relationships have a direction (A → B) – for example, user 'A' retweets 'B'.
     - **Undirected**: Relationships are mutual (A ↔ B) – for example, Telegram users 'A' and 'B' are in the same group together.
     - **Multi-Directed**: Allows multiple directed edges between the same nodes. While 'directed' will draw one edge from A → B, even if A has retweeted B hundreds of times, multi-directed will preserve multiple interactions.
       - In a dataset comprising users interacting with one another, a ***directed*** network graph is the best choice for simplicity and ***multi-directed*** is best if you want to capture and preserve multiple distinct interactions between the same pairs of entities.
     - **Multi-Undirected**: Allows multiple undirected edges between the same nodes.
       - For example, say we've downloaded group member lists for 100 Telegram groups. It may be interesting to note that users 'A' and 'B' are in 20 of the same groups, users 'A' and 'C' are in 5 of the same groups, and users 'B' and 'C' are in 8 of the same groups.
2. **Create Network Graph**:
   - Click **"Create Network Graph"** to generate the graph based on the processed data.
3. **Export Options**:
   - Select the export format:

        - **GEXF**: A format compatible with Gephi, which preserves node attributes and relationships (***recommended***)
        - **CSV (Nodes and Edges)** – useful for very complicated datasets, such as those incorporating time series data:
            - **Nodes CSV**: Contains a list of all nodes with their attributes.
            - **Edges CSV**: Contains a list of relationships (source and target).
4. **Download Files**:
    - Use the download buttons to save the graph as a GEXF file or separate Nodes and Edges CSV files.