Kalin Hershey
Yuhan Liu
ECS 272
3/19/18

Final Project Report - Untangling Euler Diagrams

1 Introduction

1.1 Goal

Our ultimate goal for our project was to recreate the two diagrams shown in the paper along with all the additional features shown in the accompanying video. Our main focus though was to create both the Compact Rectangular Euler Diagram (ComED) and Euler Diagram with Duplications (DupED) using a new set of data in order to make observations about relationships between countries.

1.2 Approach

Our approach was to start by creating ComED (which we assumed would be the more difficult of the two to implement), using the hierarchy sorting algorithm from the paper. We planned to implement DupED afterwards along with as many of the extra features as possible. We also decided that instead of creating new diagrams for each of the features like in the video, we would try to implement the features for our ComED and DupED visualizations.

As the project came along, the general approach was to create a feature for ComED, and once it worked properly, do the same for DupED if applicable. We also divided the work so that while one person worked on ComED, the other could be fixing or implementing features for DupED and vice versa, which allowed us to make more progress without needing to worry about merge conflicts.

2 Implementation

2.1 Similarities and Differences

For our project we were able to implement the basic ComED and DupED visualizations for our chosen dataset. The general format of the diagrams is the same. We used the same rounded rectangles with different bright colors to easily differentiate between sets. We also added the feature of thickening the border of an element's border as shown in the demo video.

There are however several noticeable differences between our project and the visualizations shown in the demo video that accompanied the paper. The primary differences are the features that were in the demo but that we were not able to implement

in our project. These features are the reshaping of set blocks by dragging elements as well as the creation and merging of duplicates by moving set blocks. Another noticeable difference is that our sets and elements appear in lines moving from left to right and top to bottom (especially within the hierarchy) and overall appear in a very rectangular shape while the diagrams shown in the paper appear clustered around a central point.

We also have a different version of link bundling since we were unable to implement the version discussed in the paper. While the link bundling from the demo and paper uses the barycenter of the set blocks, we found our center by simply averaging the top left corners of the set blocks. Our links are also formed using simple lines instead of the arcs shown in the paper.

We did however add a couple extra features to our implementation. We allow the user to move elements from one set to another by dragging the element to a new set, which will trigger an update and redrawing of the diagram. In addition, hovering over a set box will show the name of it, which we found useful given the number of links that can cause clutter in the diagrams.

2.2 Challenges

We had several challenges while working on the implementation of the visualization. Many of these stemmed from the lack of information provided by the paper on how to implement the two diagrams. The paper primarily focused on the diagrams as visualizations and user experience with these new types of diagrams, so sections devoted to implementation were few and short.

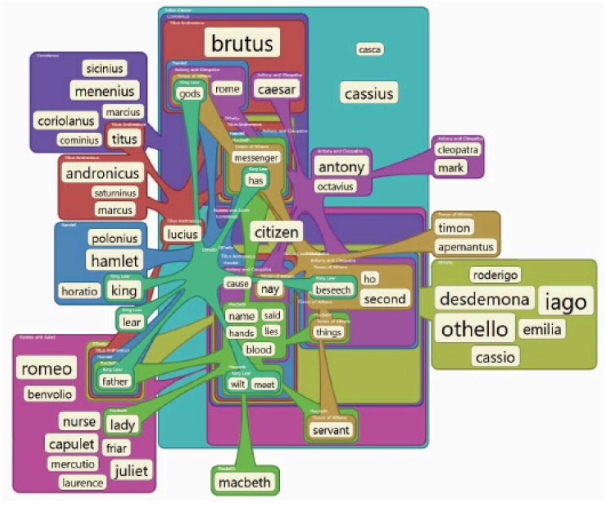
The paper provided the algorithm that was used to create the hierarchy used for ComED, and explained in general terms how the link bundling was achieved, but otherwise gave no other information on how to implement anything shown in the demo video. In addition, features like merging and creating duplicates as well as the lasso selection were only briefly mentioned while nothing about the implementation was explained. These made it difficult to achieve several of these features and in fact led to us being unable to complete them all in time.

2.3 Strengths and Limitations

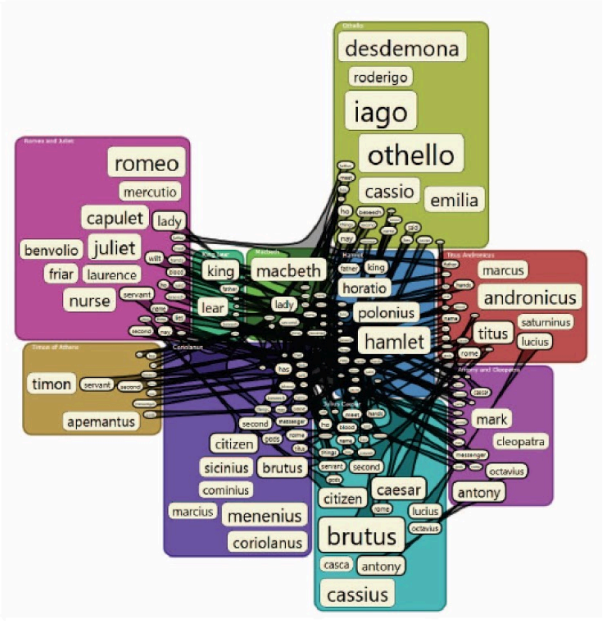
Our implementation of ComED and DupED has a few strengths. Our use of tooltips allows the user to easily see which set block represents which set regardless of how many links are overlaid over the block. This was an issue that we noticed in the paper's examples that used larger datasets.

Of course, our implementation also has many limitations. One of the most important ones is the inability to rearrange and resize the set blocks interactively. The lack of the implementation of the merging and creation of duplicates is also a limitation.

There are other extra features not shown in the demo that we believed could also be beneficial, which would be the ability to hover or click on a set and see more information about it if the name does not give enough information. In ComED, the ability to see a simple list of the elements in that set might also be helpful if the sets are very large. In DupED, a list of the all the sets to which an element belongs could be useful. Again, these would be useful for when there are many duplicates, causing a lot of clutter, such as in the example below (from the paper):



(a) ComED showing frequency across all plays.



(b) DupED where, since the words appear once per group, we can see the frequency each word is used within each play

3 Analysis

3.1 Dataset

For our dataset, we decided to pull the memberships of various international organizations. Other than our chosen constraint not to pick organizations that had more than 30 member countries, our choices were generally random (except for our decision to include the European Union). We also made the decision to cap the number of organizations at 10 since we felt that the diagrams were getting too large with more than 10 sets.

Our dataset is a bit different from many of those used in the examples in the paper. For most of the paper, the dataset appeared to be just names or numbers sorted randomly into sets that didn't have names. In other words, most of the example datasets from the paper didn't have any meaning beyond showing how the diagram should look. At the end of the paper though, their datasets included actors in the top 100 ranked movies (where the actors were the elements and movies were sets), the top 20 movies sorted into genre sets, and the word frequency across Shakespeare's plays.

Our dataset would be most similar to that of the actors in the top 100 movies in that it gives an idea of how often certain people (or countries) come together to work on something. Although, of course our dataset is much smaller, with only 10 sets and 91 elements, compared to the 100 sets and 1174 elements of the top 100 movies dataset.

3.2 Observations and Insights

When looking at our visualization, we expected to see divides based on region or alliance. For example, we expected these smaller organizations to include only countries from the same region or only known allies.

However, what we found was that there was a lot of overlap of memberships and that unless the organization was in some way defined by the region (e.g. the European Union), the member countries could be from very different areas of the world. For example, Digital 7 (or D7) has only 7 member countries, but has members from every continent except Africa and Antarctica. It was also surprising to see how often countries that otherwise do not get along are in the same organizations, such as China and Japan which belonged to many of the same organizations.

4 Conclusion

This project was definitely a challenge for the group. We unfortunately were unable to achieve our goals after running into more issues than we had anticipated at the start of the project. However, we felt that we were still able to deliver an implementation that covered the basics of the paper as well as a couple extra features.

By working with our own implementation and changing the data, it helped make clear how the features mentioned in the paper would be helpful as well as how the additional features we mentioned, such as tooltips for the sets and elements, would also be beneficial when dealing with larger, more complicated data. It would be interesting to see what information a user could gain from a full implementation of the diagrams along with tooltips or other extra features. Perhaps it would be helpful for detecting patterns among large groups or for our data, the level involvement of certain countries in international organizations.