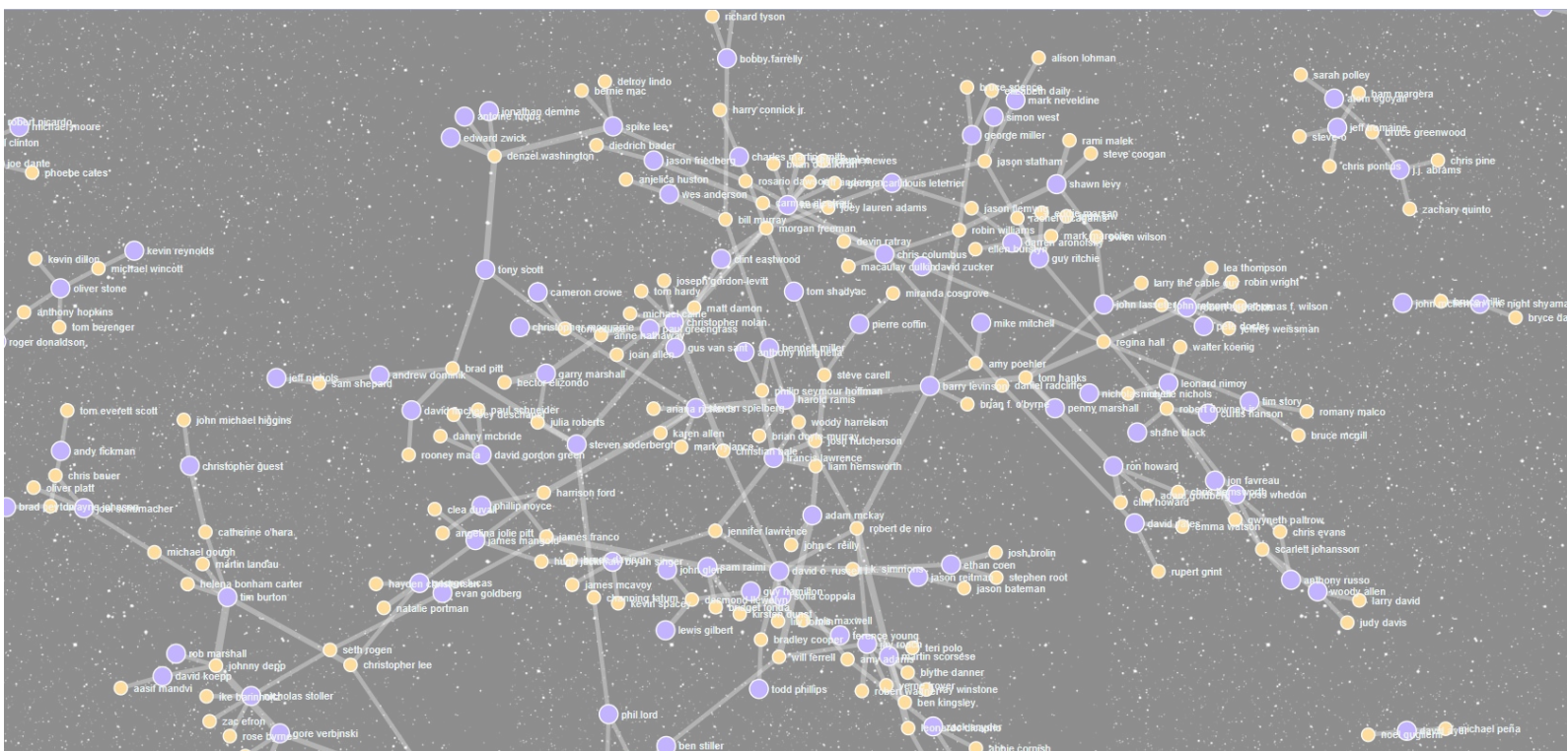# Actors / Directors Constellation

Process book

Kushagra Shah   David Dieulivol   Bruno Schmitt

Project Goal

The relationship between a director and an actor might indicate an artistic understanding, a functional routine, or even a marketing strategy. With our project, we propose a visual investigation of the success of such partnerships by analyzing the results for multiple factors involved in the movie industry. These factors include the number of movies made together, their budget, revenue, and IMDb rating.

Through our [visualization](), enthusiasts, critics, and students of the movie industry can learn more about the relationships among the stars (actors and directors) of that universe.

Evolution of the Constellation

We will now walk through the steps that got us to our visualization, along with the various decisions we made along the way.

Birth of the Idea

After many brainstorming sessions, discussions, and dataset hunting, we decided to go with the IMDb movies dataset. The chosen topic interests all three of us, like most of our generation, who love watching movies. There exist movies that everyone loves, movies that make record-breaking money, and our personal favourites. Maybe an expert in the field can explain this phenomenon, but it may not be directly intuitive for a student filmmaker, movie critic, or a curious enthusiast.

Of course, the filmmaking team's efforts and creativity play a significant role in this, but it also matters how this team comes together. Some teams work well together. Naturally, others might not. Moreover, the director-actor relationship can play a critical role in the success of a movie. And this is what we want to explore!

Choosing the Stars

We found multiple datasets which could help us visualize our idea. Many people have tried to do something similar, studying and visualizing movie data, but never this specific goal. However, we managed to find the perfect dataset for our visualization.

Among our many choices, we initially chose the [dataset-1000](#) as our primary dataset, another [dataset-5000](#) as the secondary one. However, we later decided to swap the roles as the latter contained additional 'budget' fields of great interest to us. We also had a [massive dataset](#) as our backup in case we ran out of data points or needed to fill in any missing fields. Luckily, we did not have to resort to it.

Next, we needed to tailor the data to our needs, which went as follows:
1. First, we mainly worked with the [dataset-1000](#). We performed some basic data cleaning and filtering, followed by extracting the director-actor pairs into a JSON file. We used the results to generate the constellation graph proof-of-concept for Milestone 2.
2. We decided to switch to the [dataset-5000](#) for the next phase, which had a slightly different format. After adapting to the new format and re-doing the previous processing steps, we also incorporated extra information from [dataset-1000](#). Even though it lacked the budget column, we could still use the additional actors for any movies pre-existing in our dataset. This helped us increase our dataset size.
3. We appended the associated movie metadata for each director-actor pair into the JSON file for the final phase. This part was crucial for our visualization as it helped us avoid expensive computation on-site by pre-loading the data. Furthermore, we took care of some irrelevant or conflicting entries. For example, if the same person directing the movie also starred in it, we removed that field as we are interested in finding how the director-actor relationship works, not on individual performance.

Filling the Sky with Stars

After having some clarity in our idea, finding an appropriate dataset, and acquiring the necessary skills from the course, we moved on to create the visualization. Inspired by the [Directors and their Stars web page](#) presented in class, we were already interested in visualizing the data as a graph network. The inevitable cluster of points created by relationships between actors and directors formed structures that can be seen as constellations.

Finally, each of us came up with exciting ideas for the website (we have brainstormed individually and later on regrouped our ideas). We went from a vague idea about the network and its animations to a concrete design. For further interaction and details, an on-click side window was a perfect choice. Initially, we thought of the content of this window to be a tabular and monotonous display of information. Soon enough, we realized the aim of the

project, *visualization*, and came up with a more intuitive way of presenting the information, which led to our two main visualizations seen on the website.

Once our data representation was in place, the next task of adding features was pretty straightforward. Keeping our goal in mind, we added features that would make the visualization more functional, intuitive, and appealing. Since this process can be endless, we first stuck to the essential must-haves followed by extra nice-to-haves. Finally, we see great potential in our visualization, and at the end of the report, we present more details and features which could be added to the visualization.

Website Prototyping

With a concrete idea in our heads, printing out the discussed design onto paper was simple enough. The following is our first pen-paper sketch:
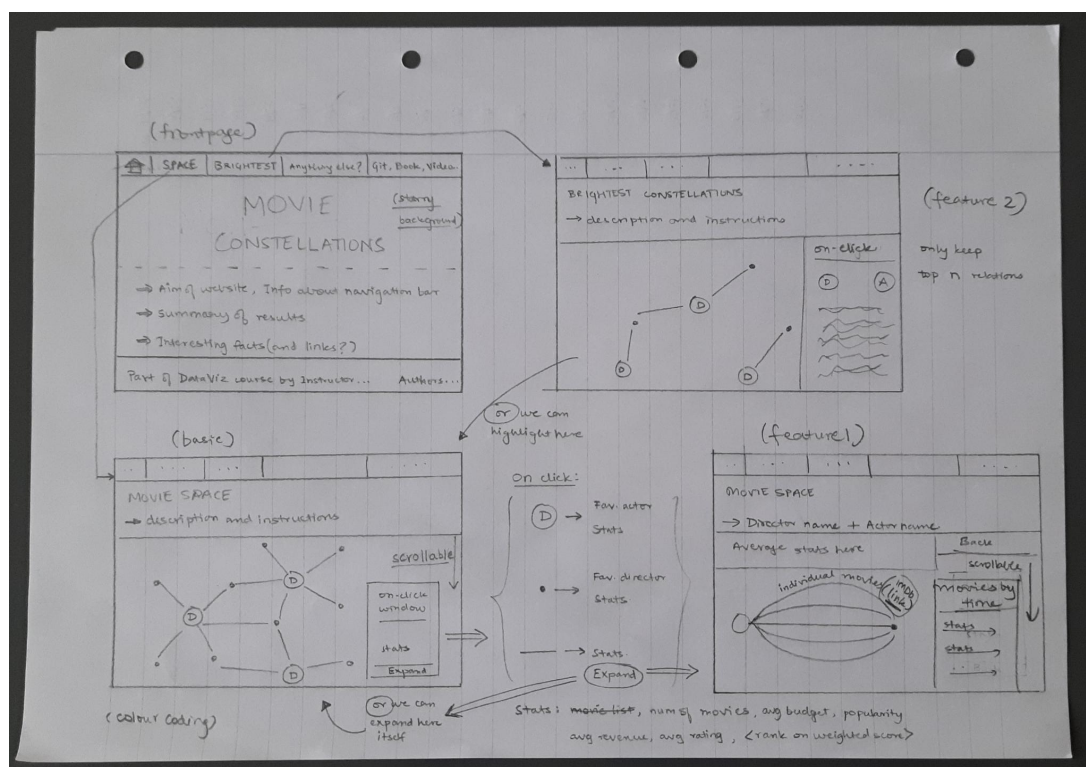
Our initial design consisted of a front-page to welcome the user (top left window) with the necessary links and text. The next page (bottom left) is our main visualization with the graph network and the side window. We did not have a clear idea about the side window's content, but it was decided for our next sketch. The two windows on the right (top and bottom) were the extra features - top N director-actor pair filters and expanded edges on-click - originally envisioned in separate pages. However, later, we thought better.
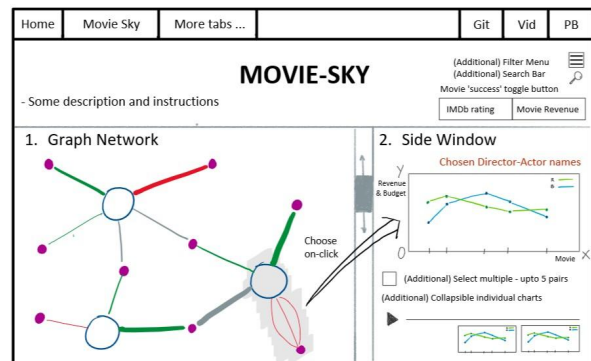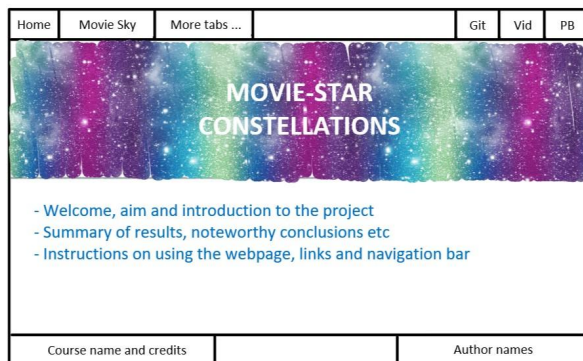
Finally, with OneNote tools, a stylus pen, and better ideas, we made the next version of the sketch, submitted as a part of Milestone 2. The welcome page is the same (but fancier), and the main visualization is improved. The graph network shows the director and actor nodes distinctly, and the edges represent movie metadata through physical properties. However, we later decided against using gradient edge colors to represent a movie's success for a cleaner design.

The sketch shows our vision of the expanded edges on click and the side window. The side window features our idea of the line chart for the movie's success, represented by the IMDb rating or the movie revenue. This was controlled using a toggle button on the top. The top right of the page also contains additional buttons for the filter and search features. The bottom right part was saved for more features such as comparison and related charts.

Upon working on the prototype design, we realized that the sketch should not be accomplished literally. The previous sketches show a very zoomed-in view, which is not very navigable and shows a tiny picture. With the data available, very complicated links can be formed, and a better constellation was born, as shown above. As a part of the prototype design for Milestone 2, only a primary network was created, with simple features and animations.
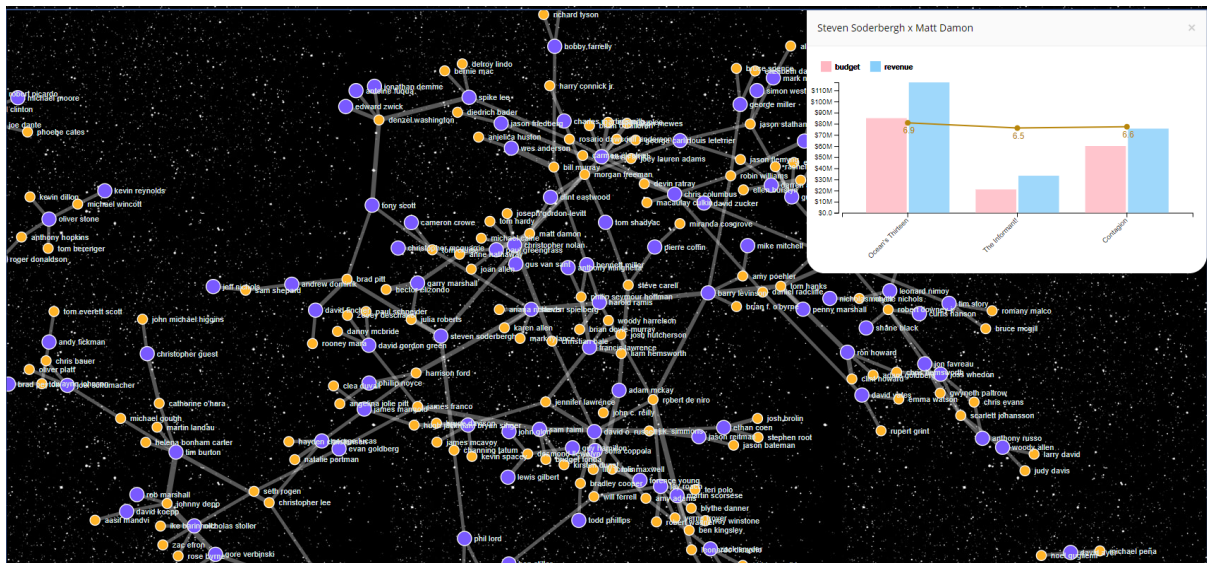
As we went further ahead, we modified a few more aspects of the visualization. It included adding the movie metadata as physical attributes of the network and creating the side window design. However, for director-actor pairs who have done fewer movies together, a single line chart was not intuitive. Hence, we switched to a bar plot, with the IMDb rating information included as a line plot. This enabled us to display the movie's success parameters in a more compact way, and hence, we removed the toggle button for the IMDb rating and movie revenue.

Next, we decided to modify our initial design goals according to our new discoveries. We realised that some of our goals were not realistic. The on-click edge expansion was a nice-to-have feature but didn't add much to the visualization's goal. And the filtering options required deletion, addition and transitioning individual nodes and links in the network, which was much more complicated than expected. Hence, we opted out of these advanced features. However, we added an advanced search bar, which highlights any director or actor's name upon search. Finally, we performed minor tweaks and improved our visualization aesthetically, leading to the final design as shown here.

Implementation

Constellation Graph

As described previously, we aim to study and present the relationship between the directors and actors of various movies. We represent this relationship using a graph in which the vertices represent directors and actors, and the edges represent their relationship, i.e., if they worked together. Our implementation uses D3's [force-directed graph](#) to create the visualization.

- The vertices represent actors and directors, which differ by color and size depending on which it's representing. Vertices representing actors are smaller and colored orange. In contrast, vertices representing directors are colored purple.
- An edge between a director and an actor indicates that they have worked on at least two movies together. Its thickness will represent the strength of the relationship, i.e., proportional to the number of movies done together.

Side window

Clicking on an edge of the constellation graph triggers an event that opens a detail window on the web page's right side. This window features a chart that compares the budget, the revenue, and the IMDb rating of all movies the pair worked on together. Pink and light blue bars represent the budget and the revenue, respectively. A line with points represents the IMDb rating. The movies are shown in ascending year order from left to right.

Challenges

We came from different backgrounds and different prior knowledge in Javascript/web development. Therefore we had to either adapt or drop some of our ideas. For example, most director-actor pairs have worked on only two movies together. If we followed our original idea of presenting a line chart on the detailed window, the appearance of that chart would have been aesthetically displeasing. (We implemented it and it looked horrible.) Therefore we opted to change how the chart on the detail window presents the data.

Overall initial constellation graph and the chart in the side window were straightforward to implement. We found it more challenging to implement the transitions that happen when a user changes the selected edge. The transition

requires new data to be bound to the elements of the chart (bars, points, and the line). These elements need to be updated to reflect the values of the new data points. This transition is straightforward if we transition from pairs that have worked on the same number of movies together. However, often there is a mismatch between the number of data points and the number of chart elements. In this case, we need to add or remove new elements while rearranging those that we keep.

The search bar also gave us a lot of headaches. We have been confronted with the [General Update Pattern](#) of D3, and understanding how it would work with the constellation. After several failed attempts, we have finally found a way to present it nicely while keeping all the graph nodes intact.

Finally, the visualization can be improved further, given more time and work. Advanced features such as filter and edge expansion, and mobile browser support can be helpful.

Peer assessment

Kushagra

- Evaluation of datasets
- Sketches of website/components
- Data preprocessing + JSON generation
- Refactoring of the constellation graph

David

- First and second version of the website (without D3 animations)
- Search bar component + Filtering in the constellation graph
- Screencast video

Bruno

- Data preprocessing + JSON generation
- The constellation graph (Proof-of-Concept and initial implementation)
- The side window

Thank you for reading our process book!