Visualizing Literary Influence in The Paris Review Interviews

Alec Glassford
Stanford University
gla@stanford.edu

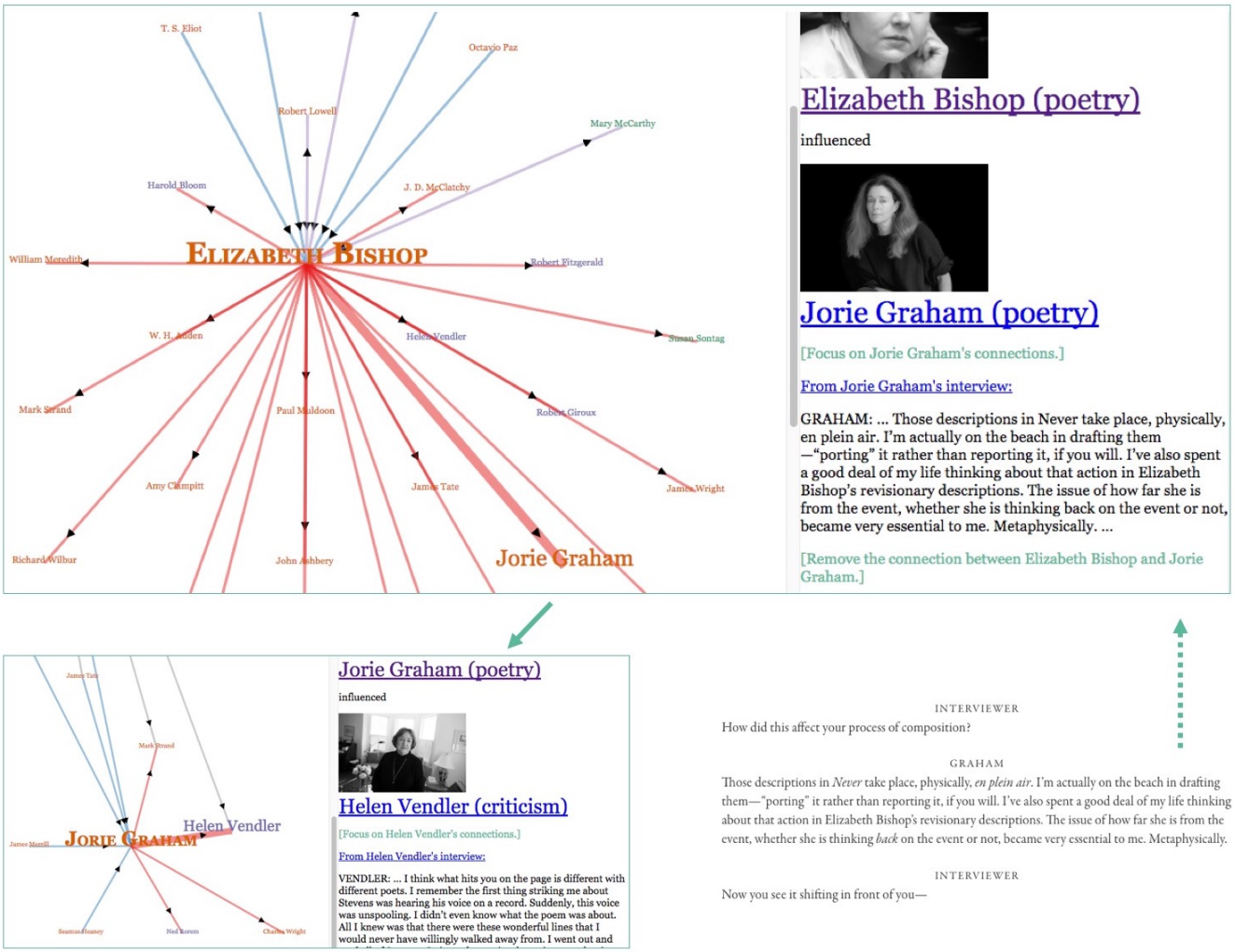

Figure 1. The original text of an interview with poet Jorie Graham (lower right) is transformed into an example backing a link between her and Elizabeth Bishop, which can be seen in the top image, represented both in the node-link diagram focused on Bishop and in the sidebar, where context is provided. Focusing on Graham (lower left) then allows the user to find new connections, such as Helen Vendler's writing on both poets.

ABSTRACT

This project extracts information about literary influence from a series of interviews with writers in *The Paris Review* and presents a tool for visualizing this influence in a meaningful way that allows users to explore the networks surrounding writers of interest. A central design choice is the presentation to the user of the analogous source text from which each edge of the graph is derived, in addition to a node-link visualization of influence. This contextualizes the graph and preserves meaning, even as the visualization gives a bird's eye view of connections.

Keywords

Visualization; network visualization; node-link diagrams; literary influence; digital humanities

1 INTRODUCTION

Since its first issue in 1953, renowned literary journal *The Paris Review* has published a long series of in-depth interviews [9] with the many of the world's foremost literary writers. The interviews are works of beauty in themselves, full of humor and thoughtful discourse, and they also provide insight into the minds of writers ranging

from T. S. Eliot and Ernest Hemingway to Toni Morrison and Elizabeth Bishop to Haruki Murakami and Elena Ferrante. Writers discuss their craft and contextualize their own work in the broader literary universe.

Through this last aspect, the interviews—with 377 different writers at the time of this research, up through issue 218 of the journal—provide a rich and unique dataset, wherein writers discuss their influences explicitly. Understanding artistic influence is a valuable task for anybody wishing to study trends in artistic and cultural change, add nuance and context to one's knowledge and interpretation of a particular author, or simply to find new art to engage with.

Visualizing the influence between writers as a graph provides easy exploration of these connections by users and gives a high-level view of how they relate to one another. At the same time, the unstructured natural language source of the data (i.e. the interviews) can provide nuance to this graph that a simple edge—"X is connected to Y"—may fail to make meaningful by itself.

This project aims to make this information easily navigable by a user with only limited previous knowledge of the writers in the dataset, yet maintain some of the important context of the interviews' collective 3.4 million words (see 3.1.1) and avoid abstracting its depth away to meaninglessness.

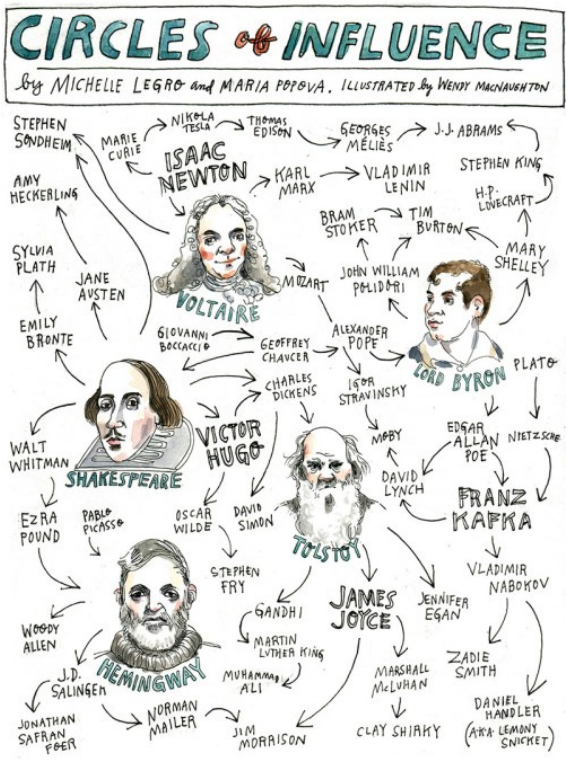

Figure 2. Legro et al's artistic network diagram, commissioned for Longshot Magazine, describes the connections of writers according the creators' perceptions of influence. It encourages viewers to explore, but provides little in the way of information density.

2 RELATED WORK

Past visualization research has explored different strategies for displaying social network graphs. Heer and Boyd's *Vizster* [5], for example, used a node-link diagram to visualize the connections between users of the social network Friendster, but also articulated the value of using additional techniques to better analyze large graph structures. They emphasized the importance of zooming, filtering, and clustering to explore aspects of their graph. But Heer and Boyd also notedin their paper that they moved away from prior visualization philosophies of providing an overview of the entire dataset before filtering, and instead built their tool with the mindset that users would "start with what you know, then grow."

"Given a lack of *a priori* knowledge of the user's familiarity with their extended network," they write, "starting from an egocentric perspective not only carries less perceptual and computational burden, but guarantees the presence of readily available landmarks for orienting the user … From this base, users can selectively expand the network to explore their greater community." This philosophy translates well to the task of exploring literary influence, as any given user may only be familiar with one or two of the writers in the *Paris Review* dataset.

Other research has further emphasized the insufficiency of node-link diagrams to visualize rich datasets fully, despite their strengths. For example, Henry et al.'s *NodeTrix* [6] visualized the co-authorship of research as a node-link diagram on the macro scale while also allowing visualization of arbitrary sections of the network as adjacency matrices to give greater resolution to clusters of interest.

Viégas and Donath [10] also found that using multiple visualizations of users' email-archive derived social networks proved more effective for users than either a clustered node-link view or a temporally focused visualization of relationships with specific contacts did alone.

Visualizing literary connections, in particular, has also been studied to some degree. Stanford's Center for Spatial and Textual Analysis has developed visualizations focused on the connections of specific Enlightenment thinkers through its *Mapping the Republic of Letters* projects [4]. Several of the project's studies have mapped and otherwise visualized the correspondences of writers like Voltaire and Benjamin Franklin. One key feature of many of these visualizations is the easy availability of the source material from the visualizations: for example, the map of Voltaire's correspondence [3] provides hyperlinks to the text of the original documents with each edge appearing in the visualization. This source context is especially significant for humanistic scholarship, where research ultimately centers on the texts.

It's worth noting that the concept of charting the influences of contemporary literary writers specifically has been explored to some degree in popular culture. Consider, for example, the illustration "Circles of Influence" by Legro et al. [7], which attempts at least as much to be a work of art as a visualization of information. While it is engaging to peruse, its design leaves many questions for the viewer—as intended. As one of its creators writees, "While some of the connections might be more obvious … others … may require some thinking, some Googling, and some general neuron flexing—and that's the point, to challenge yourself to examine how these creators might have influenced each other, tickling your curiosity with the urge to look something up, learn something new, and end up more attuned to creative cross-pollination as an agent of intellectual progress."

In this project, I hope to have worked in the same direction but with a little more foundational information for the user to launch their imagination from, with connections backed by specific textual evidence that is presented transparently as part of the visualization.

3 METHODS

The work of this project occurred in two general sections: data collection and transformation, and design of a visualization interface.

writer_type	# Interviews
fiction	239
poetry	81
theater	18
nonfiction	9
screenwriting	5
biography	5
translation	5
publishing	4
criticism	4
humor	3
comics	3
the-essay	2
editing	2
memoir	2
journalism	1
the-diary	1
the-musical	1

Table 1: writer_types extracted from the interviews and how many interviews there were of each type (for writers with multiple interviews, each is counted separately).

3.1 Data

The Paris Review has made all its interviews, excluding the most recent issue, available for free access on their website. At the time of data-gathering, this comprised 384 interviews with 377 authors (7 authors had two interviews about different aspects of their writing).

3.1.1 Data gathering

The journal provides listing pages of the interviews by decade (1950s to 2010s), and I wrote a Python script utilizing the third party "Requests" and "Beautiful Soup" packages to save and combine these listings into a master listing of the 384 interviews. The "writer_type" of each writer was extracted from the URL of their interview (the URLs are of the form "…/the-art-of-fiction-…" where "fiction" may be replaced by "poetry" or "theater" or "biography," etc.) and saved alongside the writer's name, the interview URL, and the URL to *The Paris Review*'s photo of the writer.

From this listing, the script downloaded the interview text from each interview, concatenating the interviews of authors who had conducted more than one interview. In total, the text of these interviews comprises 3,416,575 words. The mean length, then, of each interview is approximately 8,897 words.

3.1.2 Data Transformation: Finding Influence

The system operates off a very rudimentary operational definitional of influence: Writer X influenced writer Y if X's full name occurs in Y's interview. We'll consider some implications of this definition.

First, this means that the valence or meaning of the language surrounding X is not considered, which might be considered by many an important aspect of "influence" in a colloquial sense. Nevertheless, co-mentioning is evidence of some relationship between any two writers (negative relationships can be just as meaningful as positive ones), and context provided in the sidebar (3.2.1) and the ability to prune edges (3.2.2) offer users a way to handle this ambiguity.

Second, it does not distinguish between whether X was mentioned by Y or by Y's interviewer. This, too, has an impact on the meaning of the relationship, but the words of the interviewer (e.g. "Didn't you used to give Marianne Moore rhymes?") still convey information about X and Y's relationship. The ambiguities here are tackled by attaching speaker-tags to each example of text extracted and presenting them to users (3.2.1).

Third, this definition of influence fails to capture many connections between writers—there is low recall. Consider that Y may refer to X just by X's last name or by some otherwise oblique mention. Moreover, Y may refer to writers from outside the set of people who have given interviews to *The Paris Review* that would nevertheless be

Interview Listing

```
{
  "photo_url": "http://www.theparisreview.org/il/b0a15dc929/medium/Joan-Didion_Paris-Review.jpg",
  "writer_name": "Joan Didion",
  "writer_type": "fiction",
  "interview_url": "http://www.theparisreview.org/interviews/3439/the-art-of-fiction-no-71-joan-
didion"
},
{
  "photo_url": "http://www.theparisreview.org/il/b0a15dc929/medium/Joan-Didion_Paris-Review.jpg",
  "writer_name": "Joan Didion",
  "writer_type": "nonfiction",
  "interview_url": "http://www.theparisreview.org/interviews/5601/the-art-of-nonfiction-no-1-joan-
didion"
},
```

Interview Text

```
…
INTERVIEWER
It always turns into danger and apocalypse.
DIDION
Well, I grew up in a dangerous landscape. …
 Didion in her bedroom.

The last time this magazine spoke with Joan Didion, in August of 1977, …
Joan Didion was born in Sacramento, and both her parents, too, …
In 1973, Didion began writing for The New York Review of Books, …
In December of 2003, shortly before their fortieth anniversary, …
Our conversation took place over the course of two afternoons …

INTERVIEWER
By now you've written at least as much nonfiction as you have fiction. …
JOAN DIDION
Writing fiction is for me a fraught business, …
…
```

Final JSON

```
"Joan Didion": {
  "influencers": {
    "V. S. Naipaul": [
      "INTERVIEWER:  ... Since you write about yourself, /* ... */
    ],
    /* ... */
  },
  "extra": [
    "http://www.theparisreview.org/interviews/5601/the-art-of-nonfiction-no-1-joan-didion"
  ],
  "photo_url":  "http://www.theparisreview.org/il/b0a15dc929/medium/Joan-Didion_Paris-Review.jpg",
  "writer_name": "Joan Didion",
  "writer_type": "fiction, nonfiction",
  "influencees": {
    "John McPhee": [
      "MCPHEE:  ... No, it hasn't. /* ... */
    ],
    /* ... */
  },
  "interview_url": "http://www.theparisreview.org/interviews/3439/the-art-of-fiction-no-71-joan-
didion"
}
```

Table 2. Excerpts from each stage of data transformation that give examples of the schema used. In the first row, notice the two interviews recorded for Joan Didion. In the second excerpt, we have an excerpt of the raw text of Didion's two interviews, which ar concatenated together (each paragraph has been truncated to fit on one line). In the third row, we see the final JSON blob for Didion, with her list of influencers and influencees, each of whom has their list of example paragraphs backing their connection(all lists and paragraphs have been truncated with /* … */).

interesting to include in the visualization. These concerns are addressed in future work (5).

Nevertheless, despite the approach's limitations, it resulted in finding 2,340 edges between the 377 nodes of the dataset. The strategy for cataloging these edges was, for each writer, to search their interview with a regular expression for the names of each other writer. Then, for each match, record the paragraph of text that contained this name with the most recent speaker tag (e.g. "HEMINGWAY:" or "INTERVIEWER:") attached (ellipses were also concatenated for clarity if the speaker tag was a paragraph or more before the pattern-matching text). Thus for each writer, the transformation script saved a list of their influencers—with 1 or more of these examples attached.

These example-annotated lists of influencers were combined with information from the listing data and reverse-indexed to produce lists of those who had mentioned the writer, creating a final master JSON file of the 377 writers with all the listing information, as well as a dictionary of their influen**cers** and a dictionary of their influen**cees.**

Writers who had done multiple interviews were categorized as "fiction" or "poetry" first if they fell into one of these categories, as they were the most popular writer_typse; their other writer_type was then concatenated with the primary type for display, and the URL of their extra interview was attached as an extra field to their object.

It's worth noting that this pipeline of data gathering and transformation could easily be applied to a different set of texts that could feed into the same visualization tool. One

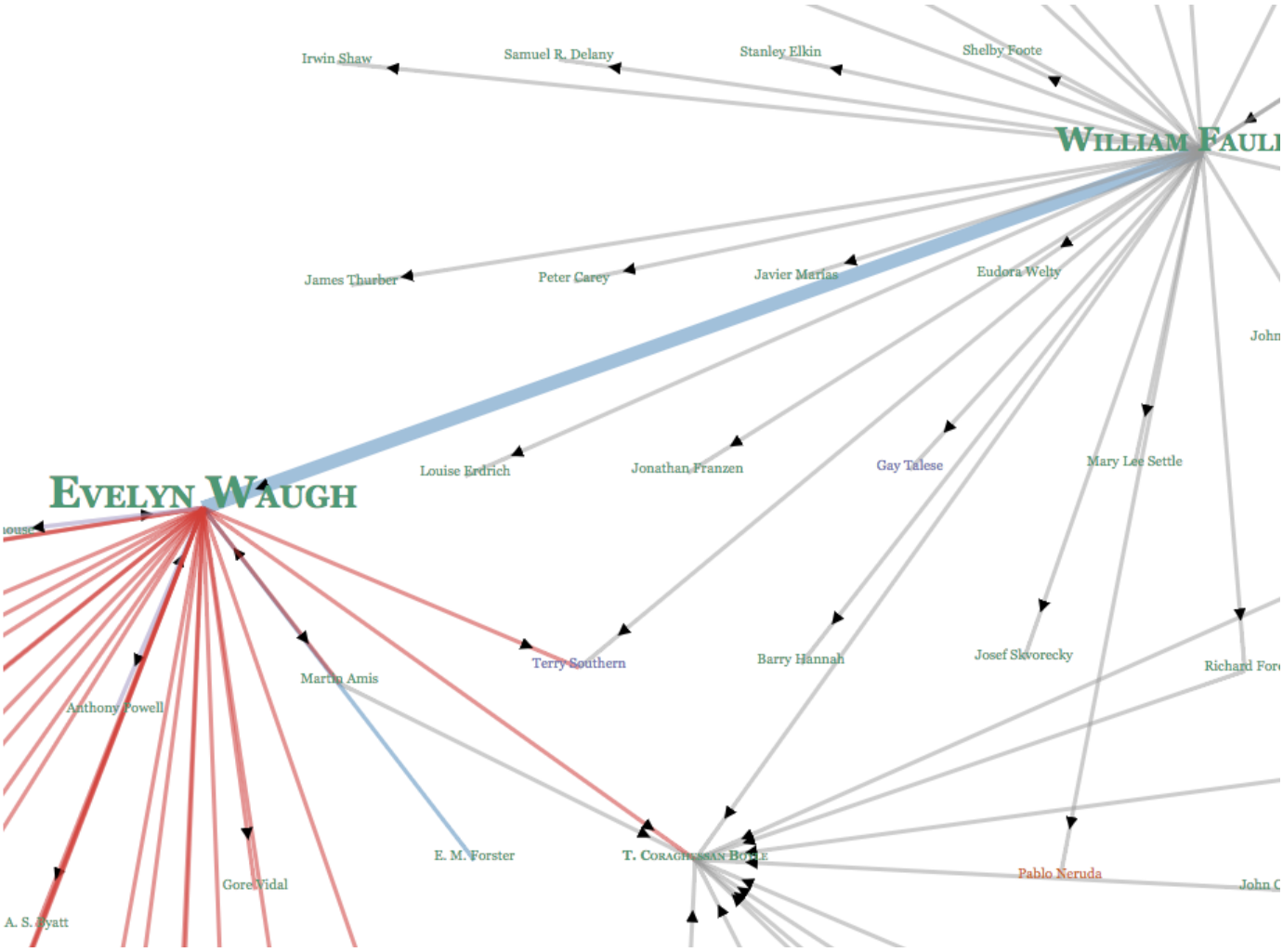

Figure 3. The node-link diagram with 3 foci added: T. Coraghessan Boyle, Evelyn Waugh (current focus), and William Faulkner). Note that the foci are arranged vertically according to influence (Faulkner influenced both others, Waugh influenced Boyle, Boyle influenced neither of the others). Other nodes are also vaguely arranged by this measure. I.e. arrows mostly point downward.

test user suggested the set of interviews with musicians published in *Rolling Stone*. Another obvious example is a set of scholarly research papers—this model of pipeline could go beyond typical graph visualizations by providing the context of paper mentions from within the paper body and not solely using the references section a la Google Scholar.

3.2 Visualization

A visualization tool was then built upon this dataset using D3.js and SVG.

3.2.1 Node-Link Force-Directed Graph

Initially, I considered visualizing the entire graph and navigating it through zooming and filtering, but while the graph is not maximally dense, it was dense enough to overwhelm. It quickly became clear that displaying the entire graph posed enormous difficulties for information comprehension due to occlusion and scale. Moreover, much like Heer and Boyd [5] described, for users with limited knowledge of the domain (interested in just one or two writers), much of the information presented in an overview would prove useless or distracting.

The use case that was most compelling, on the other hand, was for users to begin by seeing the links of a writer they were interested in and explore outward from there. This drove the focus of the rest of the design: Given a user with an interest in some writer, how can the tool help them deepen their understanding of that writer and discover other writers to read?

The user then begins by choosing one writer as "focus" of the graph, and then from that writer's neighbor nodes, can add another "focus" to add all their links to the graph, and so on.

One challenge of this dataset is its shape. At the micro level, there is a clear hierarchical structure. For a given writer X, there are 3 distinct sets of nodes related to that writer: X's influencers, X's influencees, and those who both influenced X and were influenced by X. From this

perspective, a hierarchical visualization method—something tree-like, or even table-like—would seem effective for understanding this data.

But for the sake of exploration, this methodology quickly creates difficulties. The dataset is incredibly cyclic and displaying all the links for even several nodes leads to a large amount of hierarchical ambiguity. Therefore I opted to use D3's force-directed graph package, which utilizes simulated physical forces to handle the distribution of nodes in the node-link diagram. While this sacrifices control over the clarity of the hierarchical structure—the x and y positioning of nodes, one of the strongest encodings for human perception, is somewhat sacrificed—it creates spacing that avoids occlusion and aids navigability more and, with my tweaks to the parameters, handles it effectively even as the graph grows.

To encourage as much hierarchy as possible, I adapted code by Bostock [1] to push nodes that are "influencers" upward in the simulation and nodes that are "influencees" downward. This code was scaled to underemphasize weighting based on the links between "foci" and their non-focus neighbors, in order to keep foci from getting stuck at the edges of the canvas due to all the weight pulling on them. On the other hand, it was scaled to overemphasize weighting based on the links between the foci: as these are the writers the user is most interested in, it is of utmost importance to keep them structured as hierarchically as possible.

Encodings were added to the graph to help convey as much of its information as possible. Foci are rendered with the small-caps font-variant, and the current focus at any given time is rendered in larger font, as is the node it is being compared to at any given moment (see 3.2.2). The names of all writers are color-coded by their writer_type: green for fiction, orange for poetry, and purple for all other types. The edges from the current focus is also color coded (blue for influencers, red for influencees, and purple for writers who are both). All these colors were chosen from ColorBrewer scales [2] (though some of the decorative text in the tool was colored according to schemes inspired by *The Paris Review*'s website). The direction of the influence links was also encoded with arrow markers (adapted from Maclean [8]).

3.2.2 Focus + Sidebar

The key feature of this tool is the sidebar that provides key contextual information to users. As previously discussed, seeing the link "X influenced Y" provides very little information on its own, especially if the user is not familiar with one or both of the writers.

The interface is designed, then, to always have two writers selected: the current focus and one other writer. These two writers are then displayed in the sidebar. Their names and images are displayed with their writer_type(s), and links are provided to the URLs of their original *Paris Review* interviews, should the user wish to learn by reading more about them. The direction of influence is also described in plain English.

Most importantly, the raw evidence behind the link will also be displayed. That is, if the user examines "X influenced Y" in the sidebar, they will see all the examples of paragraphs in Y's interview where X was mentioned. This contextual information is crucial for exploration of the graph, as it provides meaning to the graph. It is the best distilled information from the source text; the graph, on the other hand, though it also provides a higher-level view of multiple connections at once, acts primarily as a means of navigating between these pieces of evidence.

Toni Morrison (fiction)

influenced

Haruki Murakami (fiction)

[Focus on Haruki Murakami's connections.]

From Haruki Murakami's interview:

MURAKAMI: ... I'm a loner. I don't like groups, schools, literary circles. At Princeton, there was a luncheonette, or something like that, and I was invited to eat there. Joyce Carol Oates was there and Toni Morrison was there and I was so afraid, I couldn't eat anything at all! Mary Morris was there and she's a very nice person, almost the same age as I am, and we became friends, I would say. But in Japan I don't have any writer friends, because I just want to have . . . distance. ...

[Remove the connection between Toni Morrison and Haruki Murakami.]

Figure 4. The sidebar with Toni Morrison as the focus and her connection with Haruki Murakami selected. If the user does not think Murakami's discussion of Morrison provided in the example is compelling, they can remove the link by clicking the bottom button.

3.2.3 Interactivity

As previously described, the tool is designed for exploratory usage by users, so interactivity is central to its effectiveness. Given a starting graph of one writer as focus, with all that writer's links displayed, a user can then click on any node or link to display its relationship to the focus in the sidebar (see 3.2.2). The graph is responsive to this sort of navigation—links grow bold as you hover over them,

and nodes change color; upon selection, a writer's name and the pertinent link grow larger, as the writer appearing in the sidebar simultaneously.

Clicking again on a node after it has been selected (or clicking on a corresponding link in the sidebar) will make that node the current focus. When this happens, it switches positions in the sidebar and all its connections are added. If these are to writers not currently in the graph, they are added. The D3 force simulation restarts to optimally reorganize the new graph, and the canvas automatically scrolls to follow the new focus as things move around (though the user can break this autoscroll for their own navigation by moving the mouse). The user can then continue selecting other writers to see their relationship with the new focus. With each new focus added, the canvas expands by 500px in width and height to allow the new graph to distribute with minimal occlusion.

The edges of the old focus become greyed out, so the colors can be reset to correspond to the new focus, but at any point the user can reselect a previous focus to set it as focus once more (coloring its edges and throwing it into the sidebar). Moreover, a user can click on any edge (even greyed out ones, which maintain their arrow markers) to reselect its focus and throw both its nodes into the sidebar.

This cycle of growing the graph can be repeated for as long as the user wants to explore. As mentioned in 3.2.1, the force-direction combined with the weighting of influencers and influences reorganizes the graph with each increment of growth to arrange the nodes as hierarchically as possible, with the "oldest" influencers on top and the "newest" influences on the bottom.

Another key interactive element is the ability to prune edges that are unwanted by the user. As mentioned in 3.1.2, the operational definition for influence is very loose, so depending on the user's motivations, they may find some of the edges defined by the transformed data spurious or inappropriate. The sidebar provides a button to remove any selected edge; when this occurs, any nodes attached to the edge that are left without any edges are also pruned.

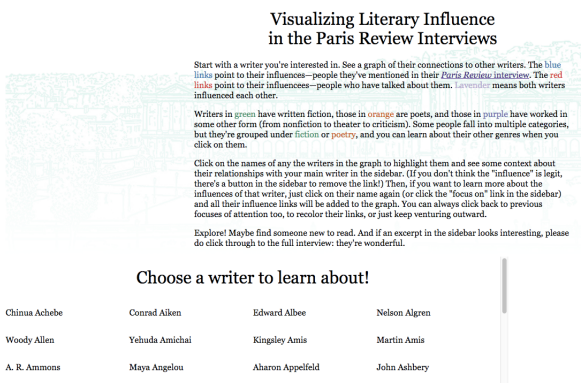

Figure 5. A screenshot of the application's homescreen, with user guide and key, as well as selection menu for user to choose the initial writer to focus on.

4 RESULTS

The resulting visualization is described thoroughly by the specifications of section 3 (especially 3.2.3). A working implementation is available at https://stanford.edu/~gla/paris and screenshots from it can be seen throughout this paper. This application provides an in-depth guide and key above the visualization and a selection tool for users to choose the first writer to set as focus.

The tool runs smoothly, without lag, and behaves as expected.

4.1 Informal User Feedback

Users who tested prototypes of the tool were excited to see the networks surrounding writers they were familiar with, and found it useful to read the context behind these connections.

However, many found it difficult to keep track of the entire graph as it grew (and the size of the canvas grew)—this led to the addition of the autoscroll on grow feature and adjustments to tighten the graph (while minimizing occlusion), which were implemented at the last minute. Further work could still be done in managing the shape and usability of the graph as it grows (see 5), but in general feedback surrounding local exploration was positive.

5 DISCUSSION AND FUTURE WORK

This project demonstrates the potential of extracting and visualizing structured graph data from unstructured source material (e.g. literary text/journalistic documents). It shows how this can be accomplished even without sophisticated natural language processing—regular expressions were the primary technology needed for extraction—and how this can be compensated for by transparently showing users how pieces of the graph are constructed from the source material, and giving them the opportunity to prune edges that don't satisfy them.

The use of the sidebar toward this end is what I consider the key innovation of the project's design. For this dataset—and for many others—the information of the network itself holds only a small fraction of the entire data's entire meaning. Annotation provides context without which the node-link diagram itself cannot be understood—especially in the cases like this one, where the edges of the graph ("influence") are not well-defined and are the output of a non-trivial transformation of the source material.

In this vein, a significant piece of this project's philosophy is providing the user with useful tools but also providing them with as much agency as possible: i.e. the user gets to see the decision-making behind the graph's structure, define what "influence" means to them, and decide how they want navigate the graph. This is all in line with the motivation of the tool to help users explore the pieces of this graph that they care most about (i.e. their favorite writer) and not be burdened with information that is extraneous to their interest (i.e. the network as a whole).

On the other hand, one can imagine future work with this dataset that branches in a different direction: there is also valuable information to be gleaned from the macro view of this graph. For example, are there clear clusters? Which writers are the most central? (Betweenness centrality could prove a particularly intriguing metric in this domain from an art historical perspective.) Or, applied on a slightly more micro-scale, what are the shortest paths of influence between any two arbitrary writers?

There are many other augmentations I would like to see (or attempt myself) in future work as well. The context provided in the sidebar could be further enriched—e.g. each writer could have the first paragraph of their Wikipedia biography attached to their name and photo. There could be tools to bulk-prune sections of the graph so that users could better focus on the most interesting groups of writers to them. And one extremely useful feature would be to autogenerate permalinks of the visualization as the graph is grown and pruned, so that users could save their explorations for later or share them with friends.

Finally, the data transformation itself could be improved, for example to use more sophisticated natural language processing to label the sentiment of sentences (Does Y like or dislike X?). Writers could be matched by just their first or last name instead of their full name—the most famous writers, Faulkner, for example, are often described like this—though care would have to be taken to minimize false positives. Or writers from outside the *Paris Review* set could be brought into the graph by using named entity recognition to match arbitrary names and an API like Google's Knowledge Graph to check if they are writers. At that point, however, the project would move move away from its current focus on exploring this particular dataset with specific richly-written source material backing it.

In conclusion, there are many directions that this work could inspire—including with entirely different domains (see 3.1.2) of anthologized natural language text. While my implementation focused on the use case of enthusiasts searching for new writers to read, it could be fruitful to collaborate with literary scholars with deeper domain knowledge to develop whatever aspects of the visualization would be most useful for scholarly work.

6 ACKNOWLEDGMENTS

Many thanks to Professor Maneesh Agrawala, Ludwig Schubert, and Peter Washington for their teaching and guidance in CS 448B. And thanks to *The Paris Review* for creating a wonderful series of interviews and providing them freely online.

7 REFERENCES

1. Bostock, M. 2016. Force-Directed Tree. Retrieved December 11, 2016 from http://bl.ocks.org/mbostock/1138500.
2. Brewer, C and Harrower, M. 2016. ColorBrewer. Retrieved December 11, 2016 from http://colorbrewer2.org/.
3. Edelstein, D., Comsa, M., Conroy, M., Kassabova, B., Wilan, C., Edmundson, C. and Nyaosi, B. 2010. Voltaire and the Enlightenment. In *Mapping the Republic of Letters*. Retrieved December 12, 2016 from http://republicofletters.stanford.edu/casestudies/voltaire.html.
4. Findlen, P., Edelstein, D. and Coleman, N. 2013. Mapping the Republic of Letters. Retrieved December 12, 2016 from http://republicofletters.stanford.edu/.
5. Heer, J. and Boyd, D., 2005, October. Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization*, 2005. INFOVIS 2005. (pp. 32-39). IEEE. Vancouver
6. Henry, N., Fekete, J.D. and McGuffin, M.J. 2007. NodeTrix: a hybrid visualization of social networks. *IEEE transactions on visualization and computer graphics*, *13*(6), pp.1302-1309.
7. Legro, M., Popova, M. and MacNaughton, W. 2011. Circles of Influence: Visualizing Creative Debt Throughout History. Retrieved December 12, 2016 from https://www.brainpickings.org/2011/08/01/circles-of-influence-longshot/.
8. Maclean, M. 2013. D3.js force directed graph example (basic). Retrieved December 11, 2016 from http://www.d3noob.org/2013/03/d3js-force-directed-graph-example-basic.html.
9. The Paris Review. 2016. Interviews. In *The Paris Review*. Retrieved December 12, 2016 from http://www.theparisreview.org/interviews.
10. Viégas, F.B. and Donath, J. 2004. November. Social network visualization: Can we go beyond the graph. In *Workshop on social networks, CSCW* (Vol. 4, pp. 6-10).