

Visualising the dynamics of character networks

DH 2016, Kraków

Aris Xanthos¹, Isaac Pante¹, Yannick Rochat², Martin Grandjean¹

¹University of Lausanne (UNIL)

²Swiss Federal Institute of Technology in Lausanne (EPFL)

ONLINE v.1: <https://maladesimaginaires.github.io/intnetviz/>

Introduction

The character network of a given narrative (novel, play, film, graphic novel, etc.) models the structure formed by the relations in its character-system (Woloch, 2003). A relation between two characters symbolises their co-presence in parts of the narrative; the entire set of relations between all characters constitutes a formal model of this character-system and lends itself to display and analysis. For example, Moretti (2011) used network modelling to compare the importance of protagonists from Shakespeare's *Hamlet*, while Trilcke *et al.* (2015) created character networks for 465 German plays and used them to initiate a wider study of German Theatre.

Most applications of character network analysis have disregarded temporality, possibly because of its representational complexity. Consequently, all relations in the system are considered as happening at the same time: one cannot distinguish if a given edge symbolises a relation at the start, at the end, or in several parts of the work under study. Furthermore, because temporality is not being accounted for, there is usually no way of relating the network visualisation with the unfolding of the source narrative.

Based on these observations, we set out to develop an open source web application which models the character-system of theatre plays as a sequence of network states synchronised with the actual narrative content.¹ This paper proposes a high-level overview of our application, successively focusing on the underlying structure extraction process, the conception of the graphical interface, and the range of uses envisioned for it. In the conclusion, we evoke the ways in which we intend to develop it and reflect on the potential significance of this development at a more epistemological level.

¹ <https://github.com/maladesimaginaires/intnetviz>

Tool overview

The underlying workflow has been divided into two parts. First, the structure of a play is extracted using *Orange Textable*² (Xanthos 2014), an open source text analysis add-on for the *Orange Canvas*³ visual programming environment. Textable essentially enables us to divide the play into its component parts (acts, scenes, lines) and identify the characters present in each scene and associated with each line. These data are then imported into a web interface based on the open source *D3*⁴ JavaScript library (Bostock et al., 2011), which allows the user (author, researcher, teacher, etc.) to manipulate the character network without prerequisite installation.

Structure extraction

The workflow is designed to facilitate the later inclusion of new data. The play used in the initial development phase was Molière's *L'Ecole des femmes*, as found in raw text format on the Project Gutenberg website.⁵ Theatre plays constitute a privileged input for the automatic extraction of structural information: such information is in general explicitly encoded in these data, as illustrated by the excerpt of Molière's play reproduced on Figure 1.

```
ACTE V

SCÈNE I.--ARNOLPHE, ALAIN, GEORGETTE.

ARNOLPHE.

Traîtres! qu'avez-vous fait par cette violence?

ALAIN.

Nous vous avons rendu, monsieur, obéissance.
```

Figure 1. Excerpt of *L'Ecole des femmes* illustrating the explicit structuration of the data.

Structure extraction is performed by a mostly linear chaining of segmentation and recoding operations based on regular expressions which gradually transform the raw text of the play into the tables that will be later used for controlling the web interface. Each table has the same number of rows, corresponding to the extracted lines of the play. The first table gives the text of each line (lightly annotated in XML) along with the associated character and stage directions (Table 1). The second table indicates the presence or absence of each character when each line is spoken (see Table 2).

² <http://langtech.ch/textable>

³ <http://orange.biolab.si/>

⁴ <https://github.com/mbostock/d3>

⁵ <http://www.gutenberg.org/files/43535/43535-0.txt>

play	act	scene	line	character	text	stage direction
l_ecole_des_femmes	1	1	1	CHRYSLALDE	< >Vous venez, dites-vous, pour lui donner la main?</ >	
l_ecole_des_femmes	1	1	2	ARNOLPHE	< >Oui. Je veux terminer la chose dans demain.</ >	
l_ecole_des_femmes	1	1	3	CHRYSLALDE	< >Nous sommes ici seuls; et l'on peut, ce me semble,</ > < >Sans craindre d'être ouïs, y discourir ensemble.</ > < >Voulez-vous qu'en ami je vous ouvre mon coeur?</ > < >Votre dessein, pour vous, me fait trembler de peur;</ > < >Et, de quelque façon que vous tourniez l'affaire,</ > < >Prendre femme est à vous un coup bien téméraire.</ >	
...

Table 1. First rows of the table containing the extracted text and stage directions.

...	CHRYSLALDE	ARNOLPHE	ALAIN	GEORGETTE	AGNES	HORACE	LE NOTAIRE	...
...	1	1	0	0	0	0	0	...
...	1	1	0	0	0	0	0	...
...	1	1	0	0	0	0	0	...
...

Table 2. First rows of the table indicating the presence or absence of each character when each line is spoken (the first four columns are identical to Table 1 and have been omitted, along with the last few character columns).

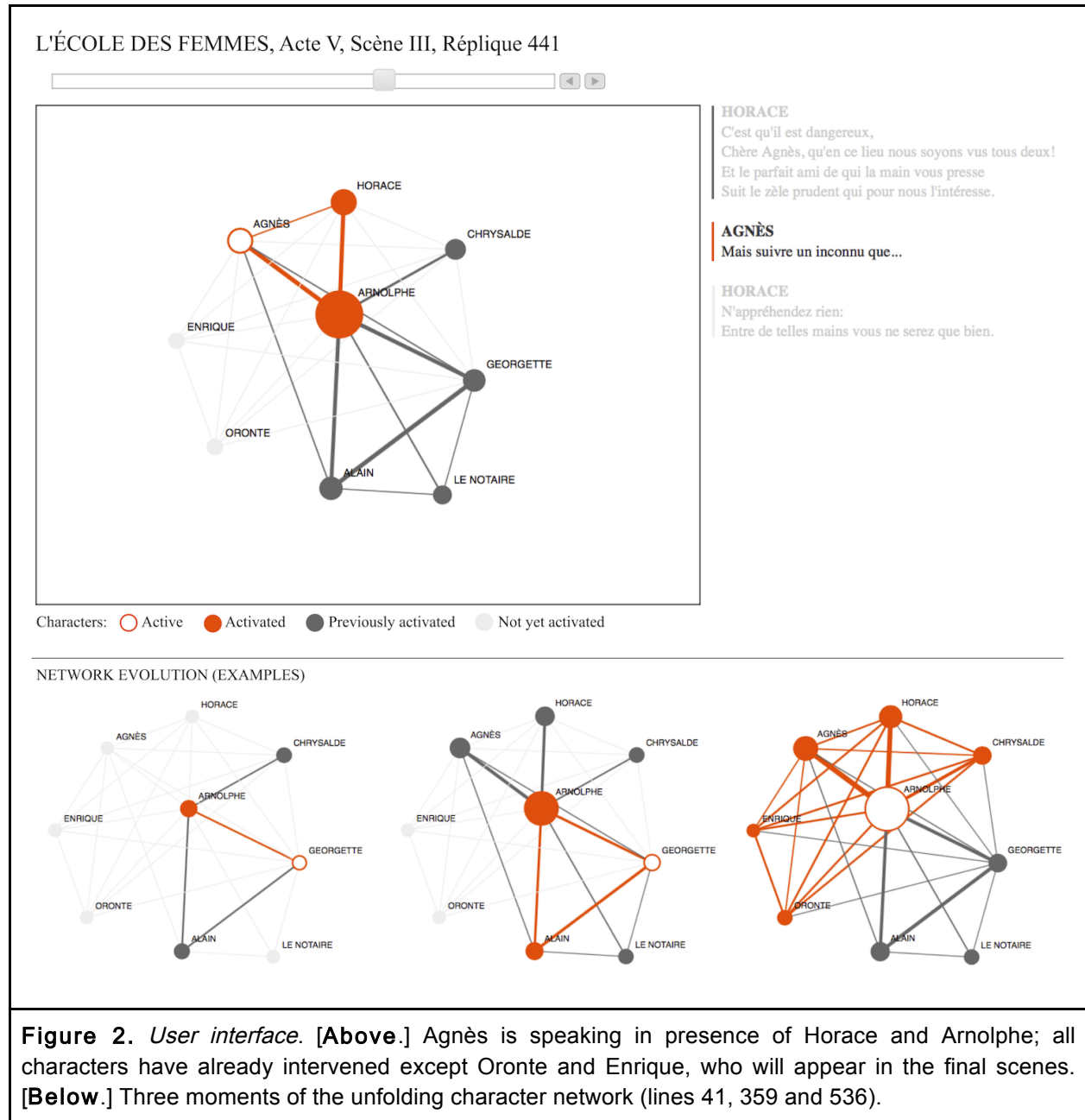
Applied to *L'Ecole des femmes*, the process was found to give fairly reliable results: about 98% of the lines were correctly extracted. Most errors could be fixed by determining *a priori* the set of acceptable character names in a given play. An unexpected source of errors was the assumption that characters could be consistently identified by a single string; even without taking into account situations where a character is explicitly “renamed” as part of a surprise effect, linguistic processes such as the succession of indefinite and definite articles in French discourse lead to formal variations in the designation of characters.

Interface

Data files are then imported into a browser and D3 is used to build an interactive visualisation of the character-system⁶ where the state of the network is synchronised with the current line (Figure 2). Character nodes are positioned using a force-based layout. Browsing the text simultaneously updates the network, line, and position indicator (act, scene, line). At each step, the weight of edges (expressed by their width) increases with the number of co-presences

⁶ A demo is available at <http://bit.ly/network-demo>.

between the characters up to this point. Similarly, the weight of nodes (their radius) increases with the number of lines spoken by this character.



A character node can be in one of four states. It is:

- “active” when the character is speaking.
- “activated” when it is present in the scene but not active.
- “previously activated” if it was present in the play but is not currently active nor activated.
- “not yet activated” if its first appearance is in a later scene.

The latter state makes it possible to offer a view of the final network state, highlighting absences as well as presences in the earlier stages.

Uses

We display each line in front of the matching network state for a reason: the network does not exhaust the richness of the play. Since co-presence analysis does not consider the actual content of lines, it cannot—nor aims to—account for discursive references to characters. However, by extracting an actantial model from speech turns, the tool provides a concise and dynamic representation of the enunciative structure of the play. Paired with background knowledge about theatrical narration, such a visualisation offers new ways of reading.

From a philological perspective, relating sociological variables (gender, age, social status) with structural properties of the network (dynamical statistics, centrality measures, word and line counts) helps clarifying which profiles lead the interactions, which ones merely react to them, and which ones are excluded from them, revealing potential social stigma. By expliciting not only the presence but also the absence of links between certain characters, the proposed visualisation may contribute to a visual and interactive deconstruction of the plays (Derrida, 1967).

By allowing the user to browse through successive pictures of the interactions, the interface provides a unique opportunity to "play" the play and visualise the flow of speech between characters. We believe that this playful dimension is particularly interesting for pedagogic uses, especially when discovering a new narrative work with students.

Last but not least, our method can also be applied to any text in the course of the writing process. In this context, disposing of a visualisation of existing interactions at each moment of the text may help the writer distance herself from an impressionistic representation.

Perspectives and conclusion

One of the most significant benefits of a digital humanities approach is the dialectic relation created between the development of a prototype and the verification of scientific hypotheses. Discussing seemingly trivial design issues like the number of colors requested to map the network relations has frequently led us to expliciting and fruitfully questioning differences in our epistemological backgrounds. Thus we consider the following perspectives not only as potential improvements to our tool but also, more importantly, as opportunities to challenge our own theoretical preconceptions:

- Further enriching the visualisation (e.g. by adding line count histograms or highlighting of the first intervention of a character) would benefit the back-and-forth movement between the content and the structure of a given play.
- The dynamical nature of the interface would enable us to compare the distinctive features of different plays not only in terms of the final state of character networks, but also and more importantly, of their evolution; this should help bring out "interactional styles" and discover *Familienähnlichkeit* (Wittgenstein, 1953) by author or period.
- Other kinds of texts will benefit from such a dynamical analysis: other fictional texts such as screenplays, but also linguistic transcriptions as used in conversation analysis.

By allowing the user to browse the content of a narrative and manipulate an interactive character network, our motivation is ultimately to contribute to a better integration of distant and close reading practices.

References

Bostock, Michael, Ogievetsky, Vadim and Heer, Jeffrey (2011). *D3: Data-Driven Documents*. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis). <http://vis.stanford.edu/papers/d3> (Accessed on Oct. 28, 2015.)

Derrida, Jacques (1967). *Of Grammatology*. Baltimore: The Johns Hopkins University Press.

Moretti, Franco (2011). *Network theory, plot analysis*. New Left Review, 68:80–102.

Trilcke, Peer, Fischer, Frank and Kampkaspar, Dario (2015). *Digital Network Analysis of Dramatic Texts*. Digital Humanities 2015. http://dh2015.org/abstracts/xml/FISCHER_Frank_Digital_Network_Analysis_of_Dramati/FISCHER_Frank_Digital_Network_Analysis_of_Dramatic_Text.html (Accessed on Oct. 28, 2015.)

Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Blackwell Publishing.

Xanthos, Aris (2014). *Textable: programmation visuelle pour l'analyse de données textuelles*. In Actes des 12èmes Journées internationales d'analyse statistique des données textuelles (JADT 2014), pp. 691-703. <http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/57-JADT2014.pdf> (Accessed on Nov. 1, 2015.)