Character network analysis of Émile Zola's *Les Rougon-Macquart*

DH 2015 – Abstract – Long Paper

Yannick Rochat, DHLAB

École Polytechnique Fédérale de Lausanne, Switzerland

yannick.rochat@epfl.ch

Abstract

In this work, we use network analysis methods to sketch a typology of fiction novels based on characters and their proximity in the narration. We construct character networks modelling the twenty novels composing *Les Rougon-Macquart*, written by Émile Zola. To categorise them, we rely on methods that track down major and minor characters relative to the character-systems. For that matter, we use centrality measures such as degree and eigenvector centrality. Eventually, with this analysis of a small corpus, we open the stage for a large-scale analysis of novels through their character networks.

1 Character Network Analysis

A character network is a model of a novel's plot focusing on a single dimension among the different types of narrative entities, that is the *character* or, at the level of the whole novel, the character-system:

> [. . .] the arrangement of multiple and differentiated character-spaces–differentiated configurations and manipulations of the human figure–into a unified narrative structure. [Woloch, 2003, p. 14]

Characters are represented in the network by nodes. The relations among them are determined on the basis of their proximity in the narration: if two characters appear side-by-side more often than a given threshold, then a link (i.e. an edge) is created between them in the network [Rochat, 2014]. If two characters never appear close together, or not significantly enough according to the defined threshold, then they are not linked in the character network.

As examples of existing research, Franco Moretti explored the narrative importance of a character by comparing some features of a character network before and after deletion of the said character [Moretti, 2011]. [Mac Carron and Kenna, 2012] extracted the structures of three mythological works (*Beowulf*, *Iliad* and *Táin*) and compared them one to one another and to real social networks, concluding that they were "discernable from real social networks" [p. 5] and eventually proposing to rank them "from the real to the fictitious" [p. 5].

2 *Les Rougon-Macquart*

The novels constituting *Les Rougon-Macquart* were published between 1871 and 1893, starting with *La Fortune des Rougon* and ending with *Le Docteur Pascal*. They cover a historical period going from 1852 to 1870. In these, Zola arranged a society of fictional and real characters in dissimilar ways, once focusing on a single character, and at other times dividing the attention between a few complementary protagonists, along with other characters recurring from one novel to another:

> I wish to explain how a family [...] conducts itself in a given so-cial system [...] I shall endeavour to discover and follow the thread of connection which leads mathematically from one man to another. [Zola, 1967, translation by E. A. V. Merton]

In his study of *Les Rougon-Macquart*'s character-systems, Philippe Hamon writes that some novels have one main protagonist, while others have more than one protagonist:

> Polyfocalisation of the system on a few heroes—rather than unfo-calisation—, which alternately shares the "hero spots" of the sys-tem, polyfocalisation of which *Pot-Bouille*, *La Bête Humaine* and *La Débâcle* are the best examples, processes issued from a network made of marked "nodes" and interstitial light layers, which take distance from a fixed "pyramid-like" hierarchy (a hero, secondary and marginal characters, etc. according to a non-adjustable scale) of classic works. [Hamon, 1998, p. 320, own translation]

We propose a mathematical formalism to study these questions in section 5. The index of *centralisation* measures how centralised the network is, i.e. how much more central the most central character is compared to all the other characters, "central" being an open concept thus far. Then, *coreness* highlights who the characters at the center of the narration are.

3 The Index

In order to construct the character networks, we consider an index built on the whole series [Zola, 1967, pp. 1795–1884], for which the indexer details his/her choices. It is a table compiling the occurrences of characters, from which we extract the co-occurrences that lead to the determination of the sets of edges. Contrary to an automatic extraction process, here we can rely on the professional work of scholars, which provides exact positions at a page-level by disambiguating characters cited by nicknames, pronouns or multiple names.

The index contains supplementary information from which we use the novel names (characters frequently appear in more than one novel) and characters'

descriptions to distinguish characters with the same name: for example, the six different characters named *Rose*.

Eventually, we transformed the index into a table composed of 40768 entries, each one of them having three attributes: name of character, name of novel and page. The table contains 1343 unique characters and 7290 unique pages.

4 The Networks

The table is then divided into twenty smaller tables, each one corresponding to a novel. We apply the method developed in [Rochat, 2014] to include co-occurrences on overlapping pairs of pages in order to take characters appearing in the same sentence but on different pages into account when creating the edges, since they need to be linked together. We build bipartite networks from these tables, with one set of nodes composed of the characters, and the other set composed of the pages. Then, we compute the graph projections on the sets of characters to obtain the character networks shown in figure 1 (see [Fruchterman et al., 1991] for the layout algorithm).

The character networks show significant diversity (table 1). The number of nodes (i.e. the *order*) varies from 16 to 88 and the number of edges (i.e. the *size*) from 68 to 1181. Works like *Le Rêve* and *La Faute de l'Abbé Mouret* feature few characters and relations: this is consistent with their intimate subjects. In comparison, *Pot-Bouille*, *Au Bonheur des Dames* and *Germinal* feature many characters and relations: they are composed of a rich crowd along with narrative events involving many characters.

The *density* of a network is the ratio of the number of existing edges by the number of all possible edges. Low density implies that the characters are sparsely connected, while high density means that the characters are more intricately connected to each other. In our case, this property can be used for categorisation, since large (*La Débâcle*) and rather small (*La Fortune des Rougon*) character networks obtain small density values. However, large density values can also be attained by large (*Germinal*) as well as small (*Le Rêve*) character networks.

5 Typology based on major vs. minor characters

In this section, we develop two ways to categorise character networks by exploiting the distributions of major and minor characters. The first one consists of studying *centralisation*, a global measure based on the centrality of all the characters, while the second one measures the *coreness* of the network, that is the size of a particularly dense subgraph that we view as a core of protagonists of the network.
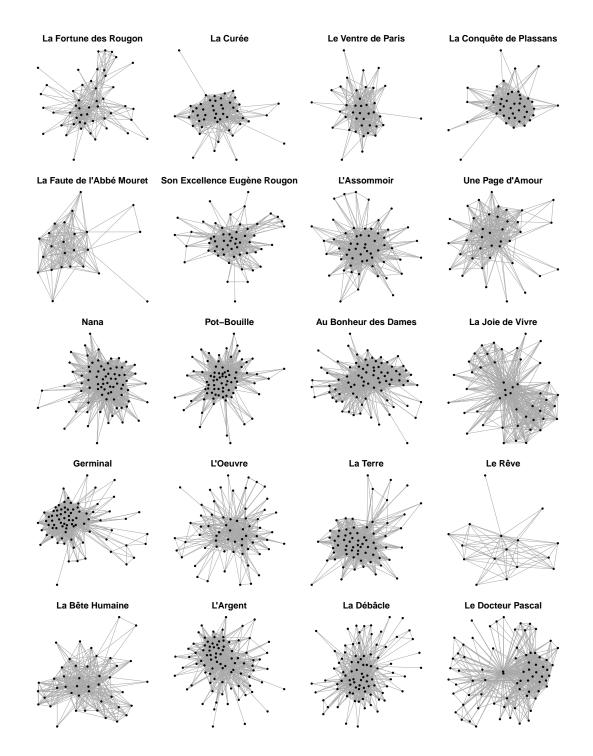
La Fortune des Rougon **La Curée** **Le Ventre de Paris** **La Conquête de Plassans**

La Faute de l'Abbé Mouret **Son Excellence Eugène Rougon** **L'Assommoir** **Une Page d'Amour**

Nana **Pot–Bouille** **Au Bonheur des Dames** **La Joie de Vivre**

Germinal **L'Oeuvre** **La Terre** **Le Rêve**

La Bête Humaine **L'Argent** **La Débâcle** **Le Docteur Pascal**

Figure 1: The character networks of the *Rougon-Maquart*'s twenty novels.

Novel	Order	Size	Dens.
La Fortune des Rougon	49	273	0.23
La Curée	49	528	0.45
Le Ventre de Paris	39	313	0.42
La Conquête de Plassans	45	576	0.58
La Faute de l'Abbé Mouret	24	169	0.61
Son Excellence Eugène Rougon	57	521	0.33
L'Assommoir	58	616	0.37
Une Page d'Amour	39	310	0.42
Nana	77	1181	0.40
Pot-Bouille	78	1160	0.39
Au Bonheur des Dames	82	905	0.27
La Joie de Vivre	47	458	0.42
Germinal	80	1104	0.35
L'Oeuvre	64	373	0.19
La Terre	68	964	0.42
Le Rêve	16	68	0.57
La Bête Humaine	44	339	0.36
L'Argent	88	844	0.22
La Débâcle	88	660	0.17
Le Docteur Pascal	74	634	0.23

Table 1: Basic network properties.

5.1 Centralisation

Centrality is a wide concept mathematically expressed by families of measures reflecting particular properties of the network under study. For example, degree is one of them. Here, we use in particular betweenness centrality: it measures how much a character acts as an intermediary at the level of the network. Betweenness centralisation is the global network measure based on betweenness centrality: we sum the differences between the maximal betweenness score and each node's betweenness score, and then divide it by the theoretical maximal sum [Freeman, 1979]. A centralisation index returns a value located between 0 and 1: a value close to 0 means that there is no node playing a central role (e.g. a ring graph), while a value close to 1 implies that there is a centralised structure (e.g. a star graph).

We observe the scores in table 2: most of the networks have low betweenness centralisation. However, those who rank first are significantly more centralised: *L'Oeuvre, L'Argent, Le Docteur Pascal* and *Son Excellence Eugène Rougon* have one and only one protagonist (the main character of *L'Argent* appears on every page) and *La Débâcle* is the story of two men at the front and their strong friendship.

Novel	c_{betw}
La Fortune des Rougon	0.14
La Curée	0.10
Le Ventre de Paris	0.17
La Conquête de Plassans	0.13
La Faute de l'Abbé Mouret	0.19
Son Excellence Eugène Rougon	0.27
L'Assommoir	0.19
Une Page d'Amour	0.16
Nana	0.13
Pot-Bouille	0.13
Au Bonheur des Dames	0.16
La Joie de Vivre	0.09
Germinal	0.24
L'Oeuvre	0.41
La Terre	0.10
Le Rêve	0.21
La Bête Humaine	0.14
L'Argent	0.36
La Débâcle	0.32
Le Docteur Pascal	0.28

Table 2: Centralisation scores.

5.2 Coreness

In order to delimit the *core* of the network (in opposition to the *periphery*), we consider the notion of *k-core* [Seidman, 1983; Csardi et al., 2006], that is the maximal induced subgraph with all its nodes having a degree equal or superior to k. Normalised by its respective network order, the highest possible k value in a network is a measure of how compact the main group of characters is. We call it *coreness*.

Results are shown in figure 2, plotted with the networks' orders. *La Faute de l'Abbé Mouret*'s character network is composed of a very dense component consisting of more than half the total number of the characters. We remark that among the three "polyfocalised" novels noticed earlier by Hamon, two of them (*Pot-Bouille* and *Germinal*) have high values of coreness, meaning that the central and prominent characters are well connected among themselves and act as interchangeable figures. However, for the third one, *La Débâcle*, the coreness is low, suggesting that having strong protagonists in a sparser network diminish the strength of the core of protagonists.
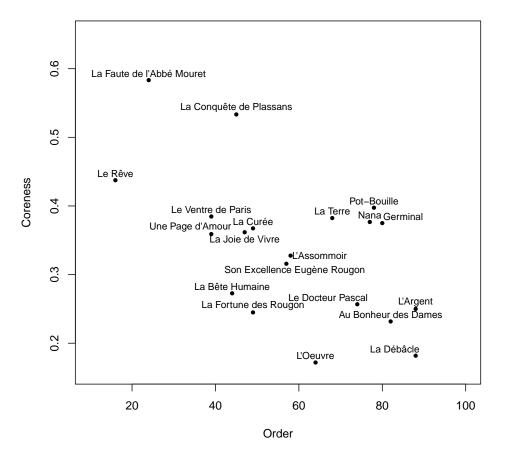
Figure 2: Coreness.

6 Conclusion

In this work, we have shown a descriptive approach to compare character networks. Our results show that it is possible to discriminate them. By iteration, the comparison of character networks leads to the analysis of large numbers of character networks.

7 Bibliography

Linton C. FREEMAN, *Centrality in Social Networks I: Conceptual Clarification.* Social Networks, 1(3):215–239, 1979.

Thomas M. J. FRUCHTERMAN and Edward M. REINGOLD, *Graph drawing by force-directed placement.* Software: Practice and experience, 21(11):1129–1164, 1991.

Philippe HAMON, *Le personnel du roman : le système des personnages dans les Rougon-Macquart d'Émile Zola.* Librairie Droz, 1998.

Gabor CSARDI, Tamas NEPUSZ, *The igraph software package for complex network research.* InterJournal, Complex Systems, 1695, 2006.

Pádraig MAC CARRON and Ralph KENNA, *Universal properties of mythological networks.* Europhysics Letters, 99(2):28002, July 2012.

Franco MORETTI, *Network Theory, Plot Analysis.* New Left Review, 68, April 2011.

Yannick ROCHAT, *Character Networks and Centrality.* Ph.D. thesis, University of Lausanne, 2014.

Stephen B. SEIDMAN, *Network Structure and Minimum Degree.* Social Networks, 5(3):269–287, 1983.

Alex WOLOCH, *The One Vs. the Many: Minor Characters and the Space of the Protagonist in the Novel.* Princeton University Press, 2003.

Émile ZOLA, *Les Rougon-Macquart.* Gallimard, 5, 1967.