Characters and Characters’ Networks in *Les Misérables*

In 1994 the computer scientist Donald E. Knuth published his seminal Stanford GraphBase: A Platform for Combinatorial Computing. Among the databases he used for demonstrating various methods of combinatorial computing were the characters in Homer’s *Illiad* and four nineteenth-century novels, including Hugo’s *Les Misérables*. Knuth’s database for *Les Misérables* features 54 characters and represents the plot of the novel by listing the encounters among these characters, chapter by chapter. This data has been the basis of several graphs, each using a different algorithm to group the characters into clusters or communities.[[1]](#endnote-1)

Though Knuth and the computer scientists who used his data were not at all interested in *Les Misérables* as a literary text, their work can be useful to literary critics: as will become clear—I hope—from what follows, it offers an innovative and productive way of studying characters in the novel. However, the very fact that the interest of literary scholars is in the interpretation of a novel (rather than in combinatorial computing) means that they look at the data (and the graphs) differently. It is because of this difference in goal and purpose that the original database had to be revised dramatically, graphing could not be left entirely to the algorithm, and the distortion caused by visualization had to be contended with.

The very decision to methodically list characters, chapter by chapter, indicates that the goal is to list all of them. But even a quick look at Knuth’s database shows that while it includes some very obscure characters it omits many others: it lists a Mlle Vaubois, friend of Mlle Gillenormand (herself a rather minor character), but does not list Gavroche’s friend, Navet; it includes the Marquis de Champtercier, an ultra-royalist in Digne, and Géborand, a retired merchant in Digne (with each of whom Bishop Myriel is shown in one, short exchange), but omits the anonymous man condemned to death, whom Myriel accompanies to the guillotine. The latter example suggests that the presence or absence of a name may be an important factor in the decision to include or exclude a character. In the case of nameless characters (some of whom appear in Knuth’s database), it seems as though appearing in isolation, rather than in close proximity to others, makes characters more visible or memorable. A clear example is the two chapters’ sequence telling of M. Madeleine trip to Arras (I, 7, v and I, 7, vii), where over the short span of 18 pages we have an aggregation of as many as 16, mostly nameless, characters—none of whom made it to the original database.[[2]](#endnote-2)

It seems, therefore, that some notion of character as “a person” or “an individual” (rather than as a textual element contributing to the plot and to thematic content) was applied, consciously or unconsciously, in compiling the data; hence the importance of the name as well as of appearing in isolation (that is, appearing at least formally as an individual).

A second methodological decision, this time clearly deliberate[[3]](#endnote-3), was the use of the distinction between “witnessed” and “inferred” encounters as a criterion for what encounters (and hence also what characters) would be included or excluded. This distinction is between what the analysis of narrative calls “showing” vs “telling.” There is no doubt that this difference is *rhetorically* an important one and when looking at encounters we may want to ask ourselves why in one case Hugo chose to “show” us (i.e. dramatize) an encounter and in another to “tell” it. But since the whole novel is told, and much of it is told in retrospect relative to another point in time, to exclude an encounter because it is told rather than shown is problematic. Thus, to take just one example, it does not make *literary* sense to exclude Valjean’s sister and her children, who were the reason for his stealing the bread and hence for the whole story, because the part of the novel that deals with their relations is told as a flashback rather than dramatized (and I should note that Knuth’s database itself is not consistent in excluding inferred encounters).[[4]](#endnote-4)

For these two reasons, the data constituted by Knuth needed to be revised if it were to be used for literary analysis. Small errors in the data (which all subsequent graphs ignored and reproduced) were corrected: Jondrette, father of Gavroche, is not actually a character but merely the alias Thénardier uses while living in the Gorbeau house; the name of Thénardier’s young daughter is Azelma and not Anzelma, etc. The way many of the nameless characters were referred to in previous graphs was also changed—Woman 1 and Woman 2 in the previous graphs did not make for easy recognition—to the way they are referred to in the novel.

But, mostly, the problem with the old database was that of omission. My goal in revising it was to include all the characters that appear in the novel, no matter how insignificant they seem, and all encounters, whether they are shown or told. The result was the discovery that the novel has many, many more characters than readers (myself included) remember or even notice while reading. Most of these forgotten, unrecognized characters are nameless, play a marginal role in the novel’s plot, appear only briefly before disappearing without leaving a trace; they therefore may not be considered to be “characters” at all. I would argue, however, that their presence is of the utmost importance since they stand precisely for “les misérables” of the novel’s title. Thus our habitual reading practices demonstrate the problem Hugo’s sought to bring to our attention: the invisibility of the miserable ones to the social world we, the readers, represent. Recognizing these characters means countering our tendency (to which filmic adaptations of the novel contribute) to think of the novel as the story of a small number of unforgettable individuals, rather than as the story of “les misérables.”

The discovery that the novel has many, many more characters than we remember or even notice while reading is well worth reflecting upon. The process of reading is always that of sorting out information, separating what seems to us important from what seems to be less important or not important at all. However, this sorting out depends, at least in part, on cultural assumptions, which it is the purpose of critical reading to reveal and question. Listing each and every character mentioned, no matter how trivial that character appears, no matter how many of them there are, goes against the grain of normal reading and is therefore not a simple task. But the commitment to consider every character and every encounter as equally important—in the sense of equally deserving to be noted—has the advantage of freeing us from these presuppositions and allowing us to see all there is in the text. Since Hugo’s novel deliberately challenges our assumptions and defies our expectations as novel readers, especially by including huge amount of material that seems irrelevant (especially but not exclusively in the so called “digressions”) it is not surprising that the gap between what we read and what we find in the text is so huge.

The decision to list barely noticeable characters and do away with the distinction between “real” and “inferred” encounters changed the database dramatically: It now lists 181 characters instead of 54 and many, many more encounters. Doing away with the distinction between “witnessed” and “inferred” encounters simplified the task of defining encounters but other issues regarding encounters remain: for example, when Marius spies on the ambush at the Thénardiers, is he part of the encounter? When Eponine enters Marius’s room while he hides under the bed, when Javert recognizes the body of the dead Eponine—are these encounters? The answer to these and similar questions is not cut and dry and so an element of arbitrariness, and hence possible inconsistency, remains. And of course, it is always possible that there are still some characters missing.[[5]](#endnote-5)

But though it is not flawless, the new database shows us important things about the novel: thematically, as I have suggested, it forces us to face the miserable ones: 93 of the 181 characters appear in only one chapter and 67of them interact with only one other character (these two groups partially overlap). Formally it highlights the extent to which the plot itself (and not only the digressions that surround it) is dilatory rather than teleological. A good example is the chapter telling about the complicated plotting of the Patron-Minette criminals, imprisoned in La Force prison, and their friends outside, to rob the house inhabited by Valjean and Cosette—a complicated plotting, involving a large number of characters, that ends with the verdict: “rien à faire.”

In constituting the database I followed Knuth’s method of identifying each character by a two letter key and listing characters and encounters by chapter (see database below); like him I did not include historical characters unless they encountered at least one of the novel’s fictive characters so that the digressions are for the most part left out; chapters that are exclusively discursive are also left out. As a result, data is extracted from 285 of the novel’s 365 chapters. The revised, expanded database has been turned into two interconnected tables that communicate and can be sorted according to different criteria; in addition to being a useful tool for retrieving information it provides insights into the novel, other than those generated by the graphs—which, unfortunately I don’t have to discuss here.

As mentioned above, the data compiled by Knuth was used to generate graphs where the vertices are characters and the edges represent encounters between these characters. It received several graphical forms that differ both in the algorithm used and in the mode of visualization. The graph we produced is different from all previous ones in that it uses the updated data of 181 characters and their encounters.

Different graphs use different algorithms that produce different articulations into clusters and we have run the data using several of them. Every algorithm produces a somewhat different interpretation of the data, depending on its set of assumptions. We therefore had to evaluate the different algorithms we considered according to how satisfying their interpretation of the data was. Broadly speaking we were looking for an algorithm that was articulated enough to be helpful (did not lump all 181 characters in a small number of clusters). We were also looking for an algorithm that would require the smallest number of manual adjustments we felt were necessary to make interpretive sense.

One reason for manual changes is that the algorithm cannot factor in the relative importance of encounters in terms of their role in the plot and therefore cannot judge properly, in some cases, to what cluster a character should belong. Thus the character Bamatabois appears in the novel twice: he is the dandy who puts snow into Fantine’s dress as she walks the streets of Montreuil-sur-Mer, which brings about the scuffle between them, her arrest, Valjean’s encounter with her, his promise to rescue Cosette—that is, much of the novel’s plot. Hugo also decided, for reasons that are not clear (at least to me) to place Bamatabois on the jury in the trial of Champmathieu, where M. Madeleine’s denounces himself as Jean Valjean; his presence on the jury, while curious and intriguing, has absolutely no effect on what happens in this scene. But the algorithms we examined all chose to place him in a cluster that consists of all the characters that appear in the trial. From the point of view of his role in the plot this does not make sense and we therefore moved him manually to Fantine’s cluster. There are other characters who appear only twice and therefore are hard to placed. Napoleon encounters both Bishop Myriel and Colonel Pontmercy. To which group does he belong? Here different algorithms gave different answers. After going back and forth we finally put him in the cluster around Bishop Myriel, primarily because of the astonishing fact that Myriel’s ecnounter with *Napoleon* is the very first encounter in the novel!

Other manual changes involved integrating some 2-character clusters the algorithm produced into other clusters. It seemed sensible to posit as a minimal number for a community three characters, especially since most of these 2-character clusters involve one totally passive character who is there only to allow the other character to externalize his thoughts. It is always important to remember that while the encounters between characters constitute the novel’s plot, they sometimes serve a formal need rather than playing a role in the plot—which is the case with the passive characters in the 2-character “false” clusters.

Increasing the number of characters from 54 to 181 has not increased the number of clusters proportionately. What has increased is the size of some clusters, especially the one around Jean Valjean. As a result we now see very clearly how large the periphery is in proportion to the center: In Valjean’s network, the majority of the characters are marginal ones; his “community” is thus predominantly of transient characters, a non-community. This modifies our understanding of Valjean’s “centrality” since most of his interactions happen in the periphery of the plot.

The graph for the whole novel is clearly multi-centered, with Valjean certainly at the center of a large network but also with Fantine and Myriel at the center of smaller ones. And, even more importantly, Valjean’s centripetal/peripheral network is quite different from the very complicated, uncentered network around Valjean’s three opponents, Javert, Thénardier, and Marius. Javert’s position especially is complicated: though his role in the plot is crucial, he does not have his own cluster because unlike Valjean, Myriel, Fantine, Marius, Thénardier and Gavroche he is not our entry way into, or point of contact with, any specific set of characters; but unlike the marginal characters in the periphery, he is related to many of the main characters—Valjean, Fantine, Thénardier, Marius. He too has a paradoxical status as both central and marginal—and yet even in this he is different from Valjean. THIS PART WILL NEED TO BE REWRITTEN ONCE WE HAVE CHOSEN THE ALGORITHM FOR THE WHOLE NOVEL AND I HAD A CHANCE TO ANALAZE IT

The characters’ graph inevitably gives a static picture of the novel’s world of characters. To mitigate this we produced 5 graphs, one for each of the novel’s five parts. They show how the novel’s world changes over time. Each of these partial graphs is also a static image, though on a smaller scale. One should, logically, continue by creating a graph for each of the novel’s 48 books; and then, for each of its chapters—at which point the graphs will catch up with and duplicate the process of reading, losing much of the their function.

One of the peculiarities of *Les Misérables* is that some of its major characters disappear early (Myriel, Fantine) while others appear late into the novel (Marius, Gavroche, Gillenormand). But the partial graphs refine this perception. Part I is dominated by Jean Valjean’s cluster but by Part II most of the characters in this cluster disappear and only very few new ones enter the plot. Once Valjean rescues Cosette from the Thénardiers, his world shrinks dramatically, which explains why when he loses her there is nothing left for him. In Part III there is no longer properly speaking a cluster around him since his cluster is equally that of Thénardier or of Eponine. Cosette now is part of the cluster around Marius—she is already lost to him, even before he explicitly gives her up to Marius in Part V. In Part IV the plot involving Valjean, Javert, Marius, and Cosette, is shrunk to one group, eclipsed by the revolutionaries (including Gavroche), the criminals, and Gillenormand’s cluster. Part V shows the approach of the denouement as all the major characters—Valjean, Cosette, Javert, Thénardier, Marius, Eponine, Gillenormand—are now combined in one cluster and the groups that dominated the previous part, shrink (revolutionaries) or disappear altogether (criminals). We also notice that while Fantine, Myriel, and Gavroche are at different times the center of a cluster, Cosette never is; she moves from Thénardier’s (Part I) to Valjean (Part II) to Marius (Part III). Her centrality to the novel—she is the universal object of desire—is thus quite problematic. We thus come to the conclusion that no character is the novel is truly central.

Visualization

Independently of the algorithm chosen there is the question of visualization. Though the graphs produced here do not use frequency to determine clusters, the do give frequency a visual representation: the size of the dot that stands for a character’s place in the network indicates the relative importance of the character, judged by the number of characters with whom he/she interacts.The graph also shows the frequency with which characters interact: the more frequently two characters meet, the thicker is the line that connects them.

The visualization of the data inevitably produces a certain distortion: in order to keep the nodes from overlapping, for example, we need to stretch some of the edges. Therefore the length of the edges is not proportional to their weight, even when we take edge weight into account. However, sometimes visualization insinuates a meaning that does not come from either the algorithm or the data. The following graph presents Jean Valjean as the center of the novel, which, as I have argued, is far from the case: the novel is equally centripetal and centrifugal and all centers are problematic. But though we can avoid deliberate misrepresentations, it seems impossible to avoid producing a visualization that does not lead to our attributing meaning to what is meaningless. All we can do is make clear that not everything we see necessarily signifies.

1. List of graphs [↑](#endnote-ref-1)
2. If one were to summarize the episode of the trip to Arras one would say: M. Madeleine left Montreuil-sur-mer and arrived in Arras where he attended the trial and denounced himself. Everything that happens between his departure and his arrival and self denunciation would be only a “filler.” But of course the point of this sequence is not what this summary tells us. The point, rather, is that the voyage was full of delays but that in spite of these delays he ended up making it on time. This tension between delay and arriving on time is the correlative of the tension in Jean Valjean between a desire not to go to Arras at all, not to denounce himself, and his conscience, which drives him to go nevertheless; it is also the correlative of the tension between his feeling of being free to do as he wishes and his feeling under the control of an external force, as well as between chance and fate. Practically every one of the 16 characters with whom he interacts during this voyage to Arras represents this tension and duality—between going fast, being in a hurry to get there, and being stuck or delayed. Thus, for example, the road mender tells him: “Pas moyen d’aller plus loin” (337) and then “Le garçon du cheval vous guidera dans la traverse” (337). The Postilion, guiding him: “Voilà le palonnier cassé” (338), but also: “Nous serons à Arras à huit heures” (338). A resident of Arras he questions: “Si c’est un process que monsieur veut voir, il est un peu tard” (349); And then, when he sees the lit court: “Ma foi, monsieur, vous arrivez à temps, vous avez du bonheur” (349). The Lawyer he addresses: “C’est fini, dit l’avocat” (350); and then: “[La lumière] c’est pour l’autre affaire qu’on a commencé il y a à peu près deux heures” (350).) What pushes Jean Valjean forward and holds him back, what responds to his conflicted desire to arrive on time and to be late, are human beings, agents of fate and of chance, helpers as well as adversaries, without whom the entire sequence would have been impossible. [↑](#endnote-ref-2)
3. In an email dated 1/20/14 Knuth states: “Encounters are generally those that we witness, not that we infer.” [↑](#endnote-ref-3)
4. For example: Knuth’s database includes Mme Pontmercy, Marius’s mother, even though we never “witness” any encounter with her; the two chapters in which she is discussed (3.2.8 and 3.3.2) are in the “telling” mode and we can only “infer” that she encountered her sister, Mlle Gillenormand, her husband, or her son. [↑](#endnote-ref-4)
5. It is partially because of the lingering arbitrariness in defining and delimiting encounters that we have decided not to use frequency of encounters to determine communities. The other reason is that in many of the algorithm we considered there was no difference in the grouping between the weighted and non weighted version. [↑](#endnote-ref-5)