Natalie M. Houston

Meaningful Aesthetics: A Comparison of Open Source Network Analysis Tools

Abstract: As a contribution to a critical hermeneutics of data visualization, this chapter presents a critical examination of the default aesthetics used in three open-source network analysis software packages: Pajek, Cytoscape and Gephi. The aesthetics of network graphs are produced by the selection of the algorithm, algorithm-specific parameters, the visualization of statistical measures and the elements of line, shape and color that are applied. In this study, visualizations of the character interaction network from Victor Hugo's novel *Les Misérables* are generated using the default parameters available in each software package, in order to explore the assumptions embedded in the tool and contribute to a humanist analysis of network analysis software and its outputs. Because network visualizations assist in understanding data structures at a variety of scales, critical awareness of how network graphs are produced contributes to a more situated, self-reflective data visualization practice that recognizes how aesthetics are always creating meaning.

Keywords: network analysis, data visualization, aesthetics

1 Introduction

Today we live in a world marked by physical, technological, political and social networks that operate at a scale beyond human comprehension. The rapid growth of the internet as both the dominant popular concept of a network and an important material, infrastructural network is but one example, and as Fredric Jameson suggests, contemporary technology is "mesmerizing and fascinating not so much in its own right but because it seems to offer some privileged representational shorthand for grasping a network of power and control even more difficult for our minds and imaginations to grasp: the whole new decentered global network of the third stage of capital itself" (Jameson 1991: 37). Within this "multinational and decentered" complexity, human perception struggles, according to Jameson, "to locate itself, to organize its immediate surroundings perceptually, and cognitively to map its position" (Jameson 1991: 44). Drawing on theories of urban space and Althusser's definition of ideology, Jameson theorizes cognitive mapping as an attempt to locate the subject in the "vaster and properly unrepresentable totality" of postmodern society, which he

https://doi.org/10.1515/9783111317779-014

describes as marked by “spatial as well as [. . .] social confusion” (Jameson 1991: 51, 54). In the decades since Jameson first explored these ideas, their relevance has only become clearer. What he identified as the literature of “high tech paranoia” in spy novels and early cyberpunk fiction has become quite resonant with reality: “the circuits and networks of some putative global computer hookup are narratively mobilized by labyrinthine conspiracies of autonomous but deadly interlocking and competing information agencies” (Jameson 1991: 38). Ruth Ahnert et al suggest that competing information organizations became visible in popular culture with the investigation into global terrorist networks following 9/11 and the founding of Facebook in 2004, which brought once-obscure concepts and methods of network science into the popular vocabulary. They identify “the network turn” at the beginning of the twenty-first century as “a whole host of converging thoughts and practices around the turn of the new millennium — the zeitgeist of the networked age” (Ahnert et al. 2020: 3). Kieran Healy also points to the expansion of both material and symbolic networks during this period: “The rapid development of computing power, the infrastructure of the Internet, and the protocols of the World Wide Web, together transformed the capacity to construct, visualize, analyze and build networked systems in practice. They were also accompanied by a major shift in the cultural salience of network imagery” (Healy 2015: 186). Network theory has been extensively operationalized in recent decades so as to substantially modify the world we inhabit and our everyday experience of it (Healy 2015: 195–198). Thus, as Jameson suggests, “cognitive mapping cannot (at least in our time) involve anything so easy as a map [. . .] mapping has ceased to be achievable by means of maps themselves” (Jameson 1991: 409). Instead we turn to networks, and particularly to network graphs, to understand our position, whether at a subway stop in an unfamiliar city or within a professional discourse community on social media (Derrible and Kennedy 2009: 17–25; Grandjean 2016). Network visualizations enable the viewer to perceive relationships at a variety of scales at the same time: we can holistically perceive the overall structures that connect an enormous set of data points and we can also zoom in on certain nodes, either with the assistance of technological tools or simply by focusing our attention on a portion of the graph.

Not surprisingly, the visualization and analysis of network data have become an important method for research in the digital humanities during recent decades. Projects like Six Degrees of Francis Bacon, Kindred Britain, and The Viral Texts Project have demonstrated how network visualizations can help researchers and students understand the scale, scope, and spread of historical relationships among people and texts.[1] Since the early 2000s, the availability of large-scale cultural data has increased with the development of mass digitization libraries like the Internet Archive, Google Books, and the Hathi Trust Research Library. Data visualization of all kinds has become more widely used in humanities scholarship as well as in journalism. Digital humanities training courses and workshops often focus on the use of software tools for data visualization and analysis, including network analysis tools like Gephi and NodeXL. These developments have encouraged increasing numbers of humanities researchers to use network visualizations for exploratory analysis of many different kinds of large-scale data, such as contemporary or historical social networks; relationships among characters in books, television or film; and economic or political relationships among individuals, organizations and institutions. However, the implications of the aesthetic choices produced by network visualization tools are not yet widely understood.

Traditional scientific approaches to data visualization focus on accuracy and clarity, assuming that graphs and charts can represent a set of data and therefore the phenomena to which it refers (Tufte 1983). However, as Johanna Drucker suggests, both the data itself and the methods for its visualization need to be examined through "humanistic inquiry," which "acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, taken, not simply given as a natural representation of preexisting fact" (Drucker 2011: para. 3). There are no natural representations: the recognition, presentation, and shaping of knowledge into information is an ideological act that reflects many interpretive choices on the part of its creators that are too often hidden from view to end users. As Lev Manovich says, "data does not just exist — it has to be generated" (Manovich 2001: 224).

Interpretive choices are also embedded in the tools used to create visualizations. As Ahnert et al point out, "we do need to be aware of the assumptions encoded in the tools we use so that we can bend them to our own needs" (Ahnert et al 2020: 64). This is especially relevant for interdisciplinary researchers who often use methodologies and tools from other disciplines and apply them to humanities contexts. Data visualization tools are often designed for specific purposes and research communities and their affordances reflect the values and activities of those communities.

Current network analysis tools make it possible for researchers to visually explore, filter and manipulate data at a variety of scales. The results of those op-

1 "Six Degrees of Francis Bacon," accessed December 15, 2021, http://www.sixdegreesoffrancisbacon.com; "Kindred Britain," accessed December 15, 2021, https://kindred.stanford.edu/; and "The Viral Texts Project," accessed December 15, 2021, https://viraltexts.org/.

erations can be output in structured data tables as well as in graphical images. Because network visualization software provides for the creation and comparison of multiple views of the data, it can support humanistic research, which "consists not of converging toward a single interpretation that cannot be challenged but rather of examining the objects of study from as many reasonable and original perspectives as possible to develop convincing interpretations" (Sinclair et al. 2013: para. 1). However, the outputs from such software should be presented within a critical framework that also reflects on the creation of the data, the layout algorithm and specific choices made within the software. Ignoring these elements leads Alexander Galloway to claim that the visual similarities among "four different maps of the internet, produced by different methods and sources, selected from numerous examples available via a normal web search" mean that "only one visualization has ever been made of an information network, for there can be only one" (Galloway 2011: 89–90). By presenting these graphs at very small scale and without citation to their source, Galloway limits deeper inquiry into the techniques and technology of their creation. However, it appears as though at least three of the four were generated with the same layout algorithm and with community detection applied as colors to the nodes and edges. Galloway ignores these visualization choices and describes these networks as endlessly repetitive because he sees them as part of a "positivistic dominant of reductive, systemic efficiency and expediency" (Galloway 2011: 100). Yet as Drucker suggests, critically examining the methods that produce data visualization can help to dismantle its ideological mystification: "The apparently neutral declarative statements of interface and data display share an ideological agenda *to simply appear to be what is*. Taking apart the pseudo-transparency by showing the workings and apparatus of the interface and graphical display of data is a crucial act of hermeneutics applied to information displays and systems" (Drucker 2020: 132).

As a contribution to this critical hermeneutics of data visualization, this chapter presents a critical examination of the default aesthetics used in three open-source network analysis software packages: Pajek, Cytoscape and Gephi. Many different choices affect the aesthetics of network graphs, including the selection of the algorithm, algorithm-specific parameters, the visualization of statistical measures, and the elements of line, shape and color that are applied. In network analysis software tools, some of these choices are implemented by default and others are available for researchers to choose from. In this study, visualizations of the classic *Les Misérables* character interaction network are generated using the default parameters available in each software package, in order to explore the assumptions embedded in the tool and contribute to a humanist analysis of network analysis software and its outputs.

2 Contexts

2.1 Tool criticism

Along with Drucker and Ahnert, a number of scholars have called for the critical examination of digital tools. Karin van Es suggests that tool criticism "reflects on how the tool (e.g., its data source, working mechanisms, anticipated use, interface, and embedded assumptions) affects the user, the research process and output, and its reliance on the user's training" (van Es et al. 2021: 52). Tools must be understood in the context of how they are being used, especially as they are adapted by researchers in the humanities. The incredible capabilities of today's software bring network visualization of large-scale data within reach of many users, but as Bernhard Rieder and Theo Röhle suggest, this power can mean that users "happily produce network diagrams without having acquired robust understanding of the concepts and techniques the software mobilizes" (Rieder and Röhle 2017: 118). Scott Weingart offers a simple example: although a user can load a bimodal network (containing nodes that represent two different types of entities, like authors and books) in Gephi and run the centrality calculation on it, the implementation of node centrality that is built into the tool is designed only for unimodal networks. Thus "although the network loads into Gephi perfectly fine, and although the centrality algorithm runs smoothly, the resulting numbers *do not mean what they usually mean*" (Weingart 2011, emphasis in original). Understanding how a given tool works and thinking critically about it can assist in working with "humanistic data," which Weingart characterizes as "uncertain, open to interpretation, flexible, and not easily definable" (Weingart 2011). Marijn Koolen argues that even if researchers lack the mathematical knowledge to critically review the algorithms behind the tools they use, "tool criticism should analyze and discuss tools at the level of data transformations [. . .] how inputs and outputs differ and what this means for interpreting the transformed data" (Koolen 2019: 382). This is particularly relevant for network visualization, which transforms a mathematical matrix into a spatial representation.

2.2 Network visualization

Network analysis is predicated on the assumption that understanding the connections between entities leads to greater understanding of a dataset. Although the statistical analysis of networks originated with studies of physical systems, such as the roadways within a city or machines on a computer network, network analysis was soon also applied to figurative connections between people as seen in club memberships or correspondence networks (Wilson 1986; Zachary 1977). Visualizations of such networks represent entities as nodes, or vertices, and the connections between them as edges. Network graphs are thus a subset of the larger category of node-link diagrams, which use a common “set of abstractions [. . .] so close to ubiquitous that it can be called a visual grammar. Entities are almost always shown using outline boxes, circles or small symbols. The connecting lines generally represent different kinds of relationships, transitions or communication paths between nodes” (Ware 2013: 222). A variety of statistical measures are used to describe network graphs, including node degree, edge weight, modularity and measures of centrality. These measures help identify important nodes and communities within a network and the ways that information, power or prestige flow between them. However, the visualization of a network “can allow users to see relationships, such as patterns and outliers, that would not be apparent through a metrics-based analysis alone” (Gibson et al. 2012: 325). This is especially important in visualizations of large-scale datasets that are difficult to comprehend in numerical terms.

Since 1984, a number of different algorithms have been developed to generate force-directed network graphs, which are widely used today especially for very large datasets. Force-directed graphs visualize the nodes in a network as if they were powered by spring or electrical forces of attraction and repulsion (Kobourov 2013). Nodes that are more closely interconnected are displayed closer together, near the center of the graph and nodes that are less connected are dispersed outwards to its margins. It is important to recognize, however, that these spatial representations do not have a direct relationship to the data, as Ahnert et al point out: “Networks express an internal logic of relationships between entities that is inherently intuitive. They also lack an explicit external spatial referent, whether the latitude and longitude of cartography, the scale and sequence of a timeline, or the categories and measures that mark the x-y axis of a statistical graph” (Ahnert et al. 2020: 57). The meaning of a network’s layout can only be understood within its own visual codes.

Because most force-directed layouts draw the layout iteratively, starting from randomly selected nodes, the resulting graph will look somewhat different each time a user runs a specific algorithm, even with the same parameters selected. Tommaso Venturini suggests that the term “spatialization” is more appropriate than “visualization”, since “Force-directed layouts do not just project networks in space—they create a space that would not exist without them [. . .] In a force-spatialized visualization there are no axes and no coordinates, and yet the relative positioning of nodes is significant” (Venturini et al. 2021: 3). Within this space, looking for indications of polarization, density, and clustering help the researcher to develop interpretations of the data (Venturini et al 2021: 4; Gibson et al. 2012:

345). However, users are likely to interpret node position as having more significance than it does: in a layout of a network containing three clusters, sometimes clusters 1 and 2 will be closer together, and sometimes clusters 1 and 3 will be closer (Gibson et al. 2012: 324). To counteract this mistaken perception, a key practice of exploratory data visualization is to produce multiple drawings of a network in order to see which aspects of its structure persist; this process can be enhanced or constrained by the visualization's aesthetics.

2.3 Meaningful aesthetics

Within the context of data visualization, visual elements such as color, shape, symmetry and size contribute to the viewer's perception of both the graph's meaning and its overall beauty or aesthetic impact. Aesthetics are considered to be deeply related to the communication of meaning within graph drawing: "Creating aesthetically appealing graphs is more than a quest for the beautiful – it has the practical aim of revealing underlying meaning and structure. In general, researchers associate aesthetics with readability, and readability with understanding" (Bennet et al. 2007: 57). Although contemporary neurobiology has offered different explanations of the underlying mechanisms of perception, the Gestalt principles of pattern perception continue to be relevant for understanding how we perceive visual designs. These principles include proximity, similarity, relative size and continuity, which explain that viewers are likely to perceive objects that are similar and/or near each other as constituting a group; to correlate differences in size with differences in quantity or strength; and to perceive connection from continuous lines (Ware 2013: 181–186). Many guidelines and tools for data visualization incorporate these fundamental principles of perception. A recent investigation into viewers' subjective rating of the beauty and interest evoked by randomly generated network graphs suggests that curved shapes are more likely to be perceived as beautiful, which corresponds with prior research in consumer design focused on material objects. More complex structures were rated as more interesting as visual stimuli. The researchers suggest that the combination of both beauty and interest are important in designing network graphs, as interest is required to engage attention for longer periods of time (Carbon et al. 2018).

The aesthetics of network graphs are produced by the combination of the network layout algorithm and the visual design choices selected in the visualization software. Within computer science, "aesthetically pleasing" layouts have been an explicit goal of force-directed algorithms since at least 1984, when Peter Eades specified that edges should be the same length and that the graph layout should be symmetrical (Kobourov 2013: 385). Later algorithms would add other criteria, like the even distribution of nodes, minimizing edge crossings and node separation and non-overlap (Gibson et al 2012: 22). These aesthetic criteria are thought to "ensure that a graph is displayed more effectively and allow the user to easily perceive the topological structure of a graph", but as Helen Gibson points out, in practice these principles sometimes conflict with one another and with users' subjective perceptions of the graphs (Gibson et al. 2012: 326–330).[2] A variety of visual design choices are available in network visualization tools, including node and edge color; node size and shape; and line styles and widths. These choices can be used to represent different features of the nodes and edges, whether present in the original data or calculated through statistical analysis of the network. These meaningful aesthetics help users gain insight about the relationship between node attributes and the overall topology of the network (Gibson et al. 2012: 341).

Even before a user interprets these visual cues in relation to the underlying data, they are likely impacted by the overall aesthetic appearance of any graph, as Helen Kennedy and Martin Engebretsen suggest: "Our encounters with form, colour, and composition are informed by bodily experience as well as aesthetic judgement [. . .] Data visualizations thus create meanings through visual and other codes. But they also generate feelings, by which we mean the emotional responses that are connected to human encounters with data visualizations. Meanings and feelings are inseparable in our situated interactions with texts" (Kennedy and Engebretsen 2020: 23–24).[3] Recent art exhibits featuring network diagrams foreground these emotional responses to network aesthetics, but they should also be explored as part of humanist knowledge creation.[4]

3 Method

This paper presents a study of network visualization tools and their outputs.[5] Because understanding the design history of visualization tools and the creation of data used for analysis is integral to the humanist critique of the apparently objective appearance of traditional data visualization, this section describes the tools and data used in this study.

2 For examples of empirical user studies, see Purchase 2002 and Purchase et al. 2002.

3 See Kennedy and Hill 2018 for a user study focused on these emotional responses.

4 See, for example, *The Art of Networks III*, held at https://www.barabasilab.com/art/exhibitions.

5 A number of comparative studies exist, but most focus on technical, rather than aesthetic comparisons. See, for example Broasca et al. 2019; Combe et al. 2010; and Majeed et al. 2020.

3.1 Network analysis tools

Three network analysis tools (Pajek, Cytoscape, and Gephi) were selected because they are cross-platform, open-access tools capable of visualizing very large networks. These are all mature tools in continuous development, with active user communities and wide adoption by researchers. Additionally, these tools are considered more comparable with each other from the user perspective because each provides a graphical user interface (GUI) and thus do not require knowledge of a programming language like Python or R. All of these tools will help users seeking to understand the relationships among entities in a dataset, but they will each do that work somewhat differently.

3.1.1 Pajek

Pajek was first released in January 1997 and remains in continued development to the present time. Pajek provides for a large number of analytic operations on six distinct data objects: network matrices or graphs; and five data objects that contain information about the nodes: "partitions" of nominal or ordinal properties; "vectors" of numerical properties; "clusters," or subsets from a partition; "permutations" containing ranked properties; and "hierarchies" which represent nodes in a tree diagram. As its developers admit, "Pajek is not '*a one click program*', some users call it '*the network calculator*'. That means that for obtaining some result several basic operations must be executed in a sequence" (Mrvar and Batagelj 2016: 2). Although this design means that Pajek has a steeper learning curve than some other programs, it is very powerful: the main program can handle networks of up to one billion nodes, and there are two versions designed for enhanced memory optimization for processing even larger networks. Version 5.15a of Pajek64 (for 64-bit Windows systems) released in May 2022 was used in this study.

3.1.2 Cytoscape

First released in July 2002, Cytoscape was specifically designed for "integrating biomolecular interaction networks with high-throughput expression data and other molecular states into a unified conceptual framework" (Shannon et al. 2003: 2498). It provides for the integration of data from scientific databases of gene information and a number of analyses specific to biological research. Over time it has developed into "a general platform for complex network analysis and visualization" and the project website highlights uses of the software in the social sciences and general study of networks.[6] Many additional capabilities can be installed into the framework with apps developed by the Cytoscape team and the user community. Cytoscape is designed for optimal rendering of networks with up to 100,000 nodes. Cytoscape version 3.9.1 released in January 2022 was used in this study.

3.1.3 Gephi

First released in July 2008, Gephi was designed as a "flexible, scalable and user-friendly software" that would provide "better network visualization to both experts and [an] uninitiated audience" (Bastian et al. 2009: 361). According to the Gephi documentation site, "Gephi is a tool for people that have to explore and understand graphs. Like Photoshop but for graphs, the user interacts with the representation, [to] manipulate the structures, shapes and colors to reveal hidden properties".[7] This design for user interactivity includes displaying the network while the layout algorithm is running, direct manipulation of the shape of the graph and options for changing its aesthetics. Although it includes statistical analysis components, Gephi is explicitly described as a "a software for Exploratory Data Analysis" that can help users to hypothesize and "intuitively discover patterns" in networks.[8] Along with continued developments to the primary software, a large number of optional plugins designed by the Gephi team and other users extend the capability of the platform. Gephi is capable of handling networks with up to 100,000 nodes and 1 million edges. Gephi version 0.9.5, released in May 2022, was used in this study.

3.2 *Les Misérables* character interaction dataset

Computer scientist Donald Knuth's 1994 book *The Stanford GraphBase: A Platform for Combinatorial Computing* included a number of datasets Knuth created from literary works (Knuth 1994: 12–14, 45–46, 180–191). One of these datasets lists 80 characters from Victor Hugo's 1862 novel *Les Misérables* and their interactions in each chapter (Knuth 1994: 14). Although his dataset encompassed 80 characters, he documented interactions among only 77 of them, creating a network containing 77 nodes and 254 edges.[9] This dataset was selected for this study in part be-

---

6 "What is Cytoscape?," accessed May 1, 2022, https://cytoscape.org/what_is_cytoscape.html.
7 "Gephi documentation wiki," accessed May 1, 2022, https://github.com/gephi/gephi/wiki.
8 Ibid.
9 The Stanford GraphBase data files are available from Skiena 2008. Knuth's personal website notes that he realized later that he omitted an interaction between Fantine and Cosette, but

cause it has been widely used in network analysis scholarship, beginning with Newman and Girvan's community detection algorithm paper in 2004 and is freely available online in a number of network data repositories (Newman and Girvan 2004). A graph of this network is also included as an example within Gephi's installation files.

However, it is important to note that the collection of data, particularly from cultural works like Hugo's novel, involves a number of decisions that are unfortunately not well documented for Knuth's dataset: he says only that "the vertices of the graphs represent the characters that appear in well-known novels, and the edges represent encounters between those characters" (Knuth 1994: 12). For literary scholars, neither characters nor encounters are self-evident: key questions include how a fictional character is defined for the purposes of the research (i.e., do they have to be named within the novel to count) and how their interactions would be defined and documented (via direct and/or indirect speech, direct physical actions, or simply being described as present in the same scene of the novel) (Moretti 2011). For example, Michal P. Ginsburg's dataset of character interactions in *Les Misérables*, which includes inferred encounters and unnamed characters, is much larger than Knuth's, comprising 181 characters and 500 interactions.[10] However, since the purpose of this study is to evaluate network visualization tools, Knuth's dataset was selected because it has become a standard reference point in network analysis scholarship.

3.3 Approach

As discussed above, the appearance of a network graph depends upon both the layout algorithm and aesthetic choices applied. Each of the network visualization tools discussed here implements several different force-directed layout algorithms, as noted in Table 1. Each of these tools also allow users to apply many different aesthetics to a network graph, such as node color, size and shape; line color, width and style; label shape, size and font; overall image zoom; and background color. Within each of these aesthetics there are frequently at least ten different options, and Gephi and Cytoscape provide for user selection of colors using

notes that his original data files should be considered "forever-frozen examples of typical data that is more or less accurate." See Knuth, "The Stanford GraphBase," accessed May 7, 2022, https://www-cs-faculty.stanford.edu/~knuth/sgb.html.

10 Michal P. Ginsburg, "*Visualizing Les Misérables*," accessed May 7, 2022, http://lesmiserables.mla.hcommons.org/.

RGB or six-digit hexanumeric notation, which encompasses 16.7 million colors. (Pajek offers 96 defined colors.) Thus a comprehensive examination of all possible combinations is beyond the scope of this chapter. Instead, this chapter examines the assumptions and effects of the default settings provided in the tools, which are likely to influence many users, especially those with limited background knowledge or technical expertise in graph drawing. Beyond this direct influence, understanding the assumptions that are built into these tools contributes to the critical awareness of how software shapes interpretation. In this study, the *Les Misérables* dataset was imported into Pajek, Cytoscape and Gephi, and images were generated of the initial presentation of the data and the different force-directed layout options available in the tool, using the default display settings.

Table 1: Graph Layout Implementation in Network Visualization Tools.

Tool	Force-Directed Layouts	Other Layouts
Pajek 5.15a	Kamada-Kawai Fruchterman-Reingold	Circular Pivot MDS VOS Mapping EigenValues Tile Components
Cytoscape 3.9.1	Edge-Weighted Force Directed Layout (Biolayout) Edge-Weighted Spring-Embedded Layout (Kamada-Kawai) Prefuse Force-Directed Layout Compound Spring-Embedder Layout	Grid Layout Attribute Circle Layout Group Attributes Layout Circular Layout Hierarchical Layout
Gephi 0.9.5	Force Atlas Force Atlas 2 Fruchterman Reingold OpenOrd Yifan Hu Yifan Hu Proportional	Circular Layout Contraction Dual Circle Layout Expansion Label Adjust Noverlap Radial Axis Layout Random Layout Rotate

4 Results

As noted above, the *Les Misérables* character interaction network is of modest size, containing 77 nodes and 254 edges. As such, it is often used in studies of network layouts and is even included within Gephi as a sample file. It thus serves as a good example for examining how the three tools initially visualize the topography of the network. This is a unimodal network, in which all nodes represent the same kind of entity, in this case characters in the novel. Because force-directed algorithms display the more highly-connected nodes towards the center of the diagram, the distribution of node degree within a dataset strongly influences the visualization of the network. Within the Knuth dataset, the node degree, or the number of different nodes a given node is directly connected to, ranges from 1 to 36, with Jean Valjean, the novel's protagonist, interacting with the greatest number of other characters (36). Other high degree nodes include the urchin Gavroche (22), the student Marius (19), the police inspector Javert (17) and the innkeeper Thenardier (16). Figure 1 displays a histogram of node degree in the dataset. The majority (78%) of the novel's characters are connected to ten or fewer other characters, with 35% of the nodes in the dataset connected to only one or two others.

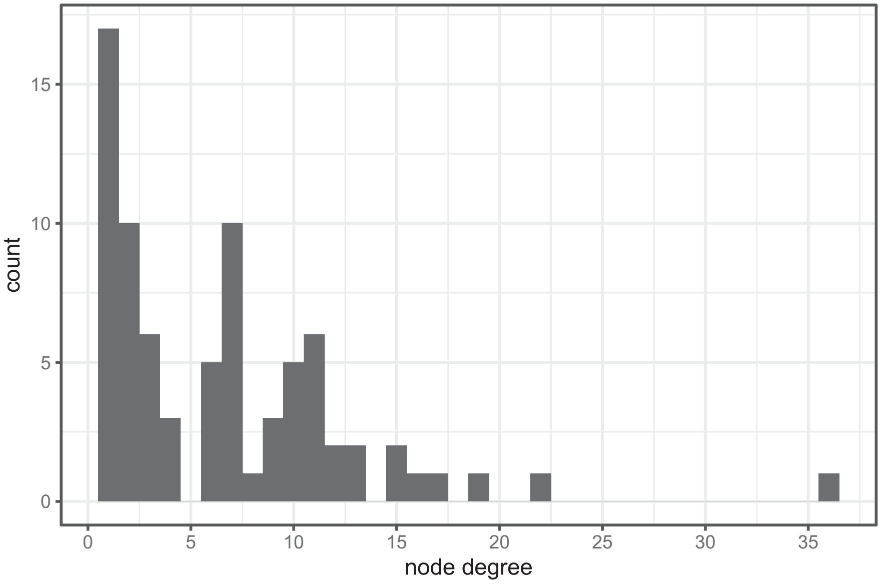

Figure 1: Node degree frequency in the *Les Misérables* dataset.

Another fundamental measure of a network's structure is edge weight, the count of how many times a specific connection is repeated; in this dataset, characters who repeatedly interact within the novel's chapters have edges with higher weights than those who only interact once or twice. Most (87%) of the character interactions recorded by Knuth occur five or fewer times and 97 of the 254 (38%) interactions recorded in Knuth's dataset occur only once. Notably, there are 14 characters with a node degree of one who interact with that other character only one time; six of these interact with Bishop Myriel and four with Valjean. With such low node degree, the Myriel cluster, which also includes a seventh character with node degree of only two, tends to be positioned near the edges of the graph in most visualizations. Within the novel, these upper-class characters exist in a very separate world than do the working- and middle-class characters.

There are four pairs of characters who interact repeatedly throughout the novel and thus have very high edge weights: the protagonist Valjean and his adoptive daughter Cosette (31); Cosette and her suitor Marius (21); Valjean and Marius (19); and Valjean and his enemy Javert (17). These four characters are central to the novel's plot and themes. As the novel's protagonist, Valjean has the highest node degree, meaning he is connected to the greatest number of other characters (36), and he also appears in the novel the greatest number of times (158, or 62% of the interactions in the dataset). However, an important character does not necessarily have to interact with a lot of different characters: Cosette's node degree is only 11, but she appears very frequently in the novel's pages: 68, or 27% of the interactions in the dataset. In network visualization, edge weight is conventionally visualized through line thickness, but other aesthetics such as color or line type (i.e., solid, dotted or dashed) can also enable the viewer to visually compare edges based on their weight, in order to understand the frequency of that interaction in the data. Node degree and edge weight are fundamental measurements that greatly impact the visualization of the network.

4.1 Pajek

In Pajek, a user's first encounter with a set of network data is through a shaped, defined form because the software automatically applies a circular network graph layout to the data in the Draw window. The low resolution Draw window generates the working view of the graph along with any aesthetics that have been applied, either as part of an analysis or as a manual selection. Figure 2 shows the initial visualization of the *Les Misérables* character interaction network in Pajek's Draw window, before any layout algorithm has been applied by the user. (This view can also be generated under Draw/Layout/Circular/Original.) Although this layout is labeled a circular layout within the software, the shape of the graph depends on the window used for the Draw view and on a typical computer screen it tends to be more elliptical in shape rather than an exact circle.

In this default working view of the network, all nodes in the graph are represented by equally-sized yellow circles which are evenly distributed in an elliptical layout against a tan background with dark red node labels. Even in a relatively

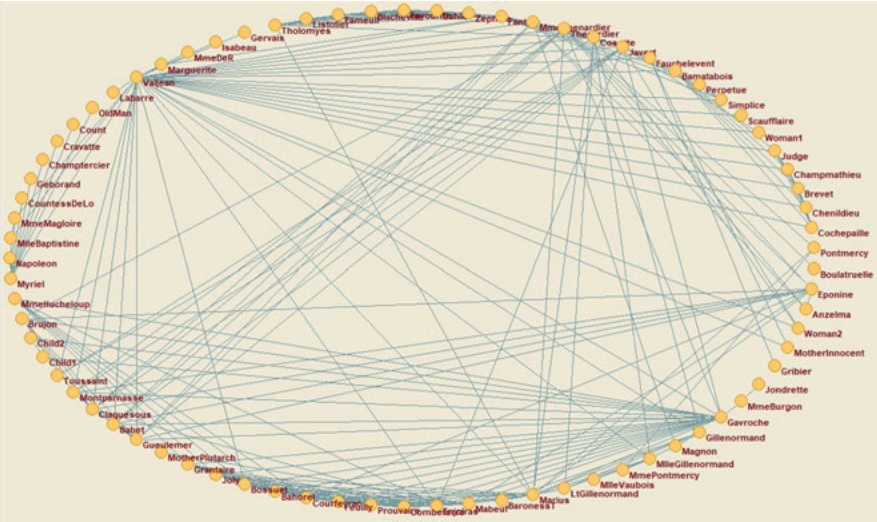

Figure 2: Pajek: initial view of the network in the low resolution Draw window.

small network such as this one, many of the node labels are illegible because they overlap with nodes and edges. The display of node labels supports exploratory understanding of network information, but they can interfere with the perception of the overall topology of the network. Although the display of node labels is a default feature of graph drawing in Pajek, labels have been removed in the remaining graph images so as to facilitate comparisons among the different layouts and tools.

The colors, shapes and line styles used by default in the low resolution Draw window emphasize the nodes more than the edges and the tan background minimizes contrast between the visual elements. Higher resolution Encapsulated PostScript (EPS) images can also be exported from Pajek, with the colors for nodes and edges set from a defined list of 96 named colors in the Export/Options screen. An exported EPS image thus uses different aesthetics than what is displayed in the working view of the graph. Figure 3 displays the initial view of the data with the default color settings used for EPS file export. Here the darker blue lines used for edges shift the visual balance a bit away from the nodes, which are displayed in a light salmon color. These EPS file settings are used in the remaining images generated from Pajek.

Although the Pajek reference manual simply states that nodes in this initial layout are positioned by default "in order determined by the network," examination of the graph along with the node weights reveals that this layout is designed to reveal clusters of connected nodes (Mrvar and Batagelj 2022: 69). High degree

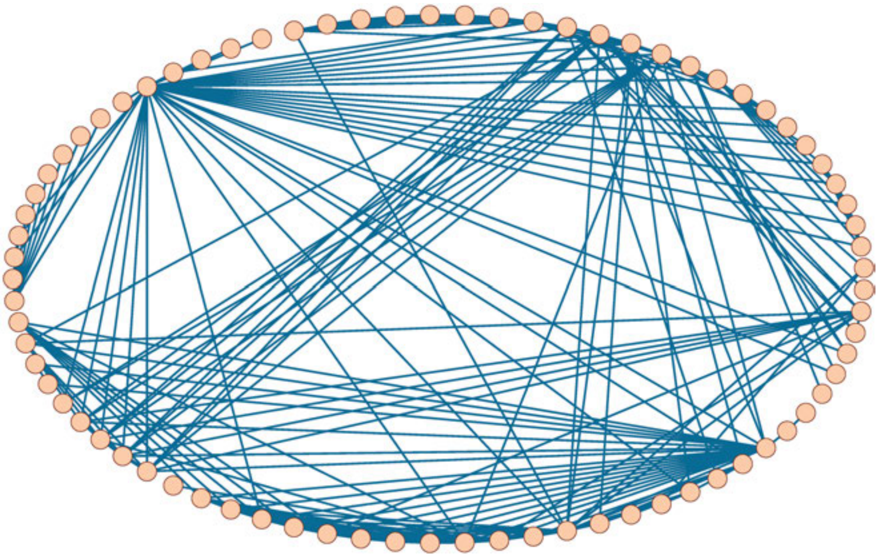

Figure 3: Pajek: initial view of the network with default color settings for EPS file export.

nodes (those with a lot of connections), such as the one representing the novel's protagonist Valjean, are positioned so that the edges connecting them to low degree nodes cross the diameter of the ellipse. By default in Pajek, all edges are displayed as equal width, drawn with thin blue lines, rather than visually indicating edge weight. These visual choices place more emphasis on node degree, or the number of edges connected to a given node, rather than on significant relationships between node pairs.

Because this layout is the first view presented of the data in Pajek, the software promotes the assumption that the connections to high degree nodes are the most meaningful. Although this is a standard approach to analyzing networks, examining the data in other visualizations can highlight other ways of seeing the nodes in the network. Figure 4 shows the same elliptical ("circular") layout with random positioning of the nodes. As discussed above, low-degree nodes are predominant in this dataset, which points to Hugo's representation of class, gender and professional distinctions within an urban setting. Viewing the network with random positioning of nodes shifts the focus from the plotting around the novel's central characters to the socio-historical view offered in the novel.

Pajek includes two classic force-directed algorithms in its layout options, which it calls "Energy" layouts: Kamada-Kawai (Kamada and Kawai 1989) and Fruchterman-Reingold (Fruchterman and Reingold 1991). The default setting for these layouts selects a random node as the starting point for drawing the graph, although a

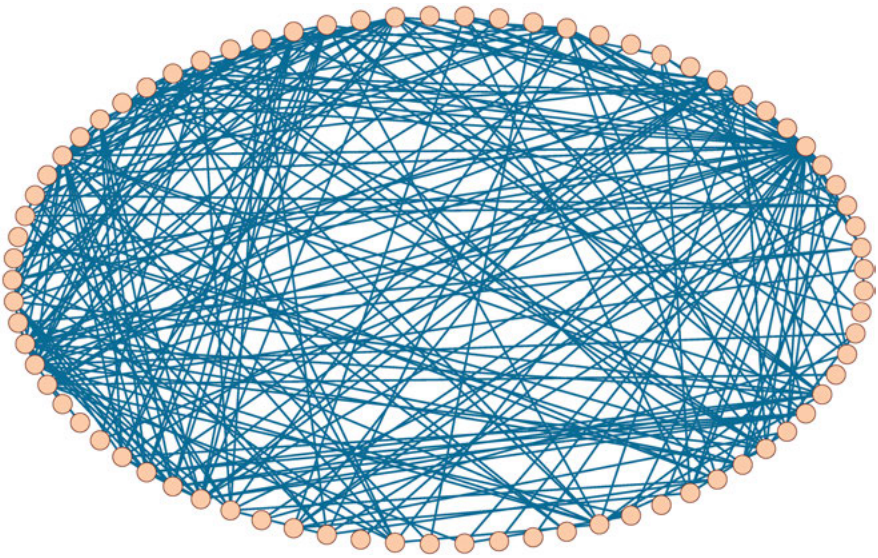

Figure 4: Pajek: elliptical (circular) graph layout with random node ordering.

specific starting point may be selected by the user. Pajek offers several layout parameters for the Kamada-Kawai algorithm, which is based on spring forces, including options to fix the position of specific nodes, to optimize subclusters or to optimize components (Mrvar and Bagatelj 2022: 69). This layout tends to emphasize symmetry and produces a boxy effect in the display of node clusters (Gibson et al. 2012: 330). Figure 5 shows the Kamada-Kawai layout with separated components, which reduces visual overlap of the nodes (Kobourov 2013: 383). This layout visually distinguishes four groups among the more connected nodes and the default aesthetics applied in Pajek represent higher edge weights with thicker lines. Because of the strong attractive force applied in the visualization, it is difficult to distinguish the individual nodes for Valjean, Marius, Javert and other key characters at the center of the graph. As in all force-directed layouts, this algorithm places highly-connected nodes towards the center of the graph and the very low degree nodes are clearly visible around the perimeter. Because these nodes are evenly arranged at some distance from the center of the graph, they are strongly distinguished from the more highly connected nodes, but it is difficult to perceive how they relate to the network as a whole. For example, the group of seven very low-degree nodes connected to Myriel are located in the upper right quadrant of the graph. Because they are spread out so far from each other and their edges cross others before connecting with the Myriel node, these nodes are not visually distinct as a group.

Figure 5: Pajek: Kamada-Kawai layout with separated components.

Figure 6 displays the *Les Misérables* character network in Pajek using Fruchterman and Reingold's algorithm, which is based on the attractive and repulsive forces exerted between "atomic particles or celestial bodies" combined with annealing or cooling processes (Kobourov 2013: 385). This layout spreads out both the highly weighted and less weighted nodes to form a rounded shape to the graph overall. Pajek's implementation of the algorithm tends to draw edges closely alongside each other, creating an elongated visual effect. The intricacy of connections is minimized in this layout in favor of displaying node groups, which is helpful for understanding the structures within Hugo's *Les Misérables*: the high degree nodes for the key characters Valjean, Javert, Marius and Cosette are located at the very center of the graph; the nodes for the political revolutionaries connected to Courfeyrac and Enjolas are arranged in an overlapping wedge in the upper left quadrant; and the cluster of upper-class characters connected to Bishop Myriel is clearly visible on the right side of the graph. In this layout even small differences in node degree are visually distinguished because of the way the algorithm calculates the attractive force at the graph's center. Although node placement does not have inherent meaning, the placement of the nodes relative to one another and to the center of the graph is meaningful.

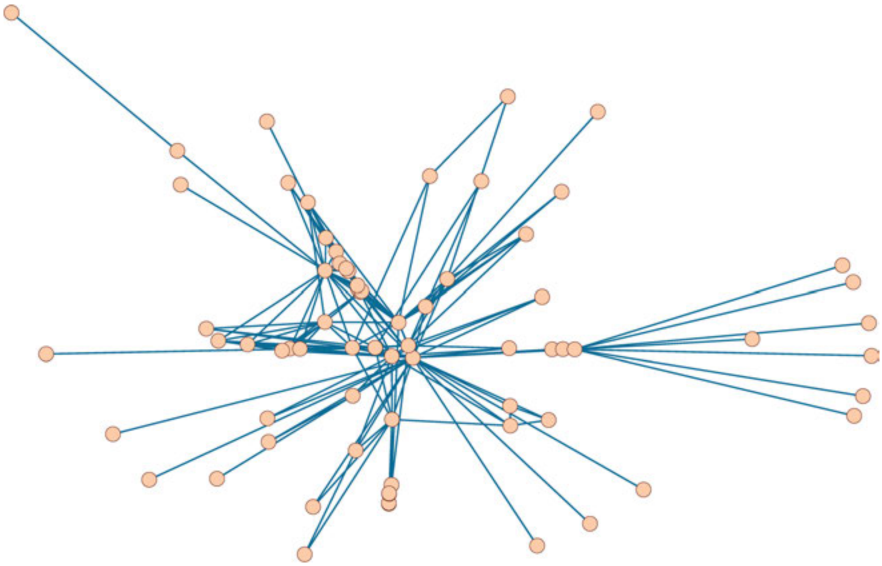

Figure 6: Pajek: Fruchterman-Reingold layout.

4.2 Cytoscape

Cytoscape provides four force-directed layouts in the main program: the Prefuse Force-Directed Layout; the Edge-Weighted Force Directed Layout based on the Biolayout algorithm, which is specifically designed for similarity analysis in biological research and does not work well for other kinds of network data; the Edge-Weighted Spring-Embedded Layout, which is an implementation of the Kamada-Kawai algorithm; and the Compound Spring-Embedder Layout, which is optimized for use with compound graphs as well as other networks (Heer et al. 2005; Enright and Ouzounis 2001; Kamada and Kawai 1989; Dogrusoz et al. 2009). Cytoscape manages aesthetics for edge and node style, color, shape and size, along with the data features that these aesthetics represent, through a palette of 18 predefined "styles." Users can also modify these existing styles or create their own.

The current version of Cytoscape initially displays network data using the Prefuse force-directed layout, although the software documentation states that the Grid Layout is the default view of the data. As shown in Figure 7, this view of the data is initially displayed in the "default" palette style, which represents nodes with light blue rectangles containing the node label in black and edges as thin grey lines on a white background. As with the default aesthetics applied in Pajek, the emphasis here is on the nodes, more than the edges: the large rectangular nodes provide information but also obscure some of the edge intersections, making it difficult to perceive the relationships between the nodes.

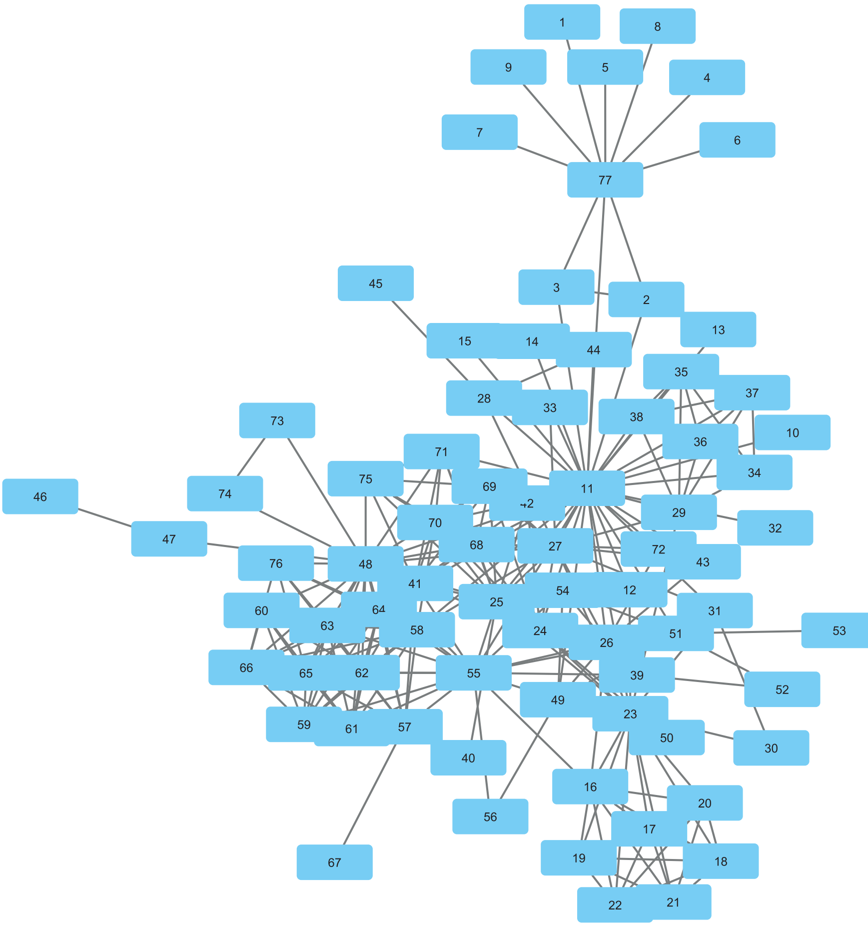

Figure 7: Cytoscape: initial view of the network (Prefuse layout with default style).

Figure 8 shows the same Prefuse layout with the “default black” aesthetic style applied and node labels removed. This style represents the nodes with small white circles and edges with thin green lines on a black background, which provides a better topological understanding of the network than the “default” style. Both of these styles display node labels by default and do not visually represent edge weights. These aesthetics are user-modifiable, but the default settings reflect the overall focus in Cytoscape on node information, which derives from its original development for biological research. Nevertheless, the layout algorithms in the software provide for the exploration of network structures as well. Nodes are evenly spaced in the Prefuse layout and clusters are arranged in a circular pattern that clearly shows the interactions among the nodes. Even the many edges connecting the high-degree nodes at the graph's center are visually distinct in this figure. To facilitate comparisons of the layouts and tools, the "default black" style is used in the remaining images generated from Cytoscape.

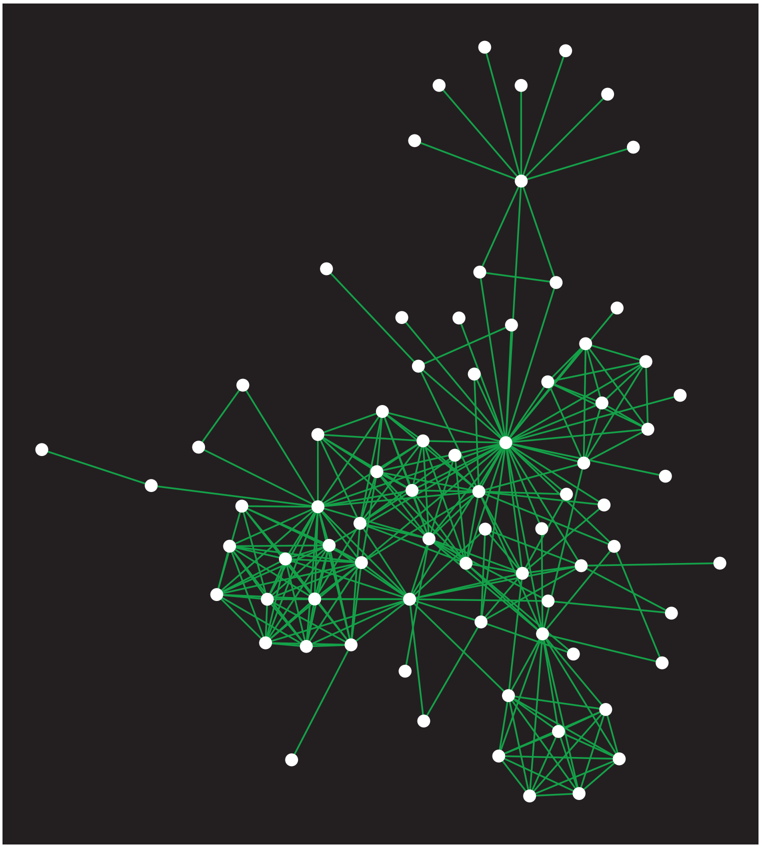

Figure 8: Cytoscape: Prefuse layout with default black style.

Figure 9 shows the *Les Misérables* network in Cytoscape's Edge-Weighted Spring Embedded Layout, which is an implementation of the Kamada-Kawai algorithm. Despite the layout's title and the fact that it includes a weighting parameter, the default aesthetic styles do not visualize the edge weights unless that option is selected by the user. Comparing Cytoscape's implementation of Kamada-Kawai with that in Pajek (Figure 5) shows overall topological similarities, as would be expected, but individual nodes and groups within the network are more clearly separated. The high-degree nodes for Valjean, Javert, Marius and Cosette near the center of the graph can be clearly distinguished from one another; the two groups of political revolutionaries are shown as overlapping but distinct wedge shapes on the left; and the Myriel group is placed in the upper right quadrant of the graph.

Figure 9: Cytoscape: Edge-Weighted Spring-Embedded layout (Kamada-Kawai).

Figure 10 shows the *Les Misérables* network in the Compound Spring-Embedder layout in Cytoscape, which evenly spaces the nodes in the graph and thus tends to arrange nodes in a circular fashion. Clusters of nodes are less clearly distinguished from one another and the node spacing makes the edges in the graph more visible. This layout is thus useful for exploring the paths of connection among individual nodes and clusters. As noted previously, the fact that the Myriel cluster of nodes appears on the left side of this graph rather than on the right, as it did in other figures, bears no significance. What is significant, and represented in each algorithm and tool, is that the characters connected to Myriel are strongly distinguished from those connected to the revolutionaries Courfeyrac and Enjolras, as shown in the annotations in Figure 11.

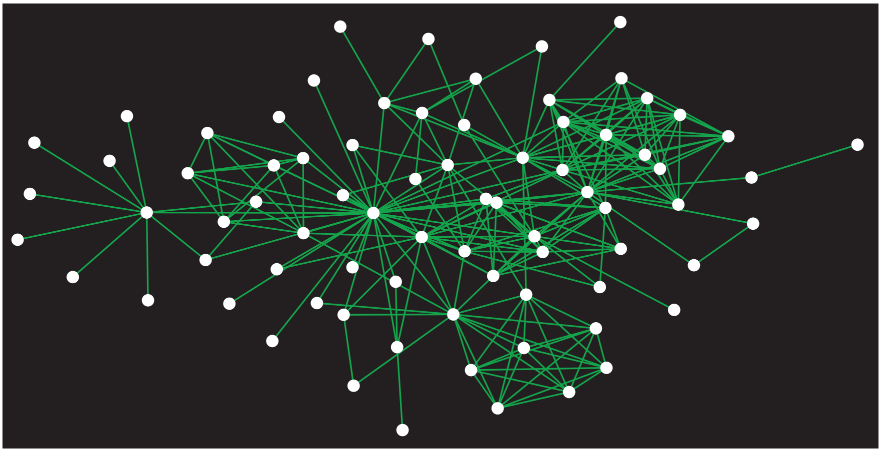

Figure 10: Cytoscape: Compound Spring-Embedder layout.

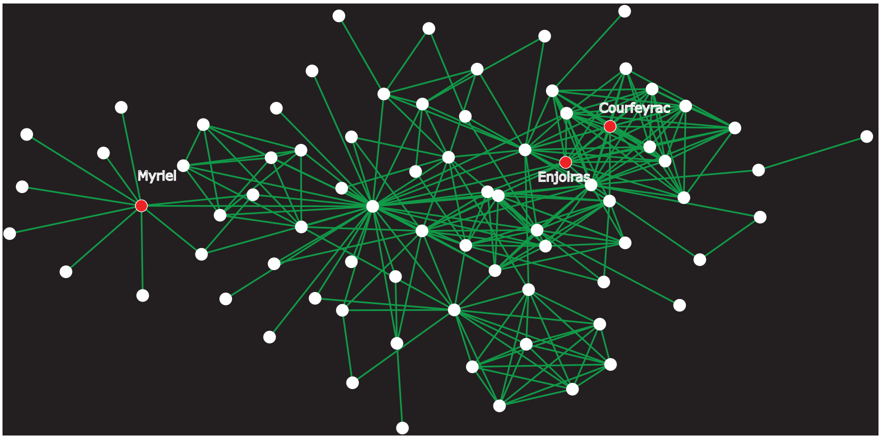

Figure 11: Cytoscape: Compound Spring-Embedder layout with select nodes annotated.

4.3 Gephi

The Gephi interface is divided into three main windows: the Data Laboratory, which displays the node and edge data tables; the Graph window, which allows for analysis and manipulation of the network; and the Preview window, in which the image is fine-tuned for exporting to an image file. Aesthetic choices, such as color and size for nodes and edges, can be applied in both the Graph and Preview windows. Figure 12 displays the initial view of the *Les Misérables* network in the Graph window, before any layout algorithm has been applied. Nodes are displayed as small black circles and edge weights are visually indicated with proportional edge widths. Of the three tools surveyed here, Gephi is the only one that privileges a random view of the dataset by making it the default view of a new network. But this random view is not without an implied interpretive approach: by displaying edge weights in this initial view of the data, certain connections are visually distinguished from the rest, such as the frequent interactions between Valjean and Cosette. Figure 13 displays the same initial view of the data, before any layout or aesthetics have been applied, in the Preview window. Even though networks are displayed with straight edges in the Graph window, the default setting in the Preview window is to display curved edges. In the figures that follow, straight edges are used to facilitate comparison with earlier figures.

Gephi provides four force-directed layouts in the main program: Fruchterman Reingold, Yifan Hu, Force Atlas and OpenOrd (Fruchterman and Reingold 1991; Hu 2006; Jacomy et al. 2014; Martin et al. 2011). As noted previously, the Fruchterman-Reingold algorithm is based on the forces between celestial bodies and the algorithm was designed to produce "even node distribution, few edge crossings, uniform edge length, symmetry and fitting the drawing to the frame" (Gibson et al. 2012: 331). Fruchterman-Reingold layouts tend to produce an overall rounded shape with evenly spaced nodes throughout the graph (Gibson et al. 2012: 332). Figure 14 displays the *Les Misérables* network using the default settings for speed and gravity in Gephi's Fruchterman-Reingold implementation. Because Gephi visually encodes edge weight by default, this graph emphasizes strong connections between particular nodes, rather than node degree. Key groups of nodes are visible in the graph, but are more difficult to distinguish than in some other layouts because of the way low-degree nodes are arranged around the perimeter with long edges connecting them to nodes in the center of the graph. As shown in Figure 15, the Myriel group is on the left side of the graph, but is difficult to discern because it overlaps with the group connected to Fantine. The Gestalt principles of similarity and figure-ground patterns explain why the Myriel group, with some of the lowest edge weights in the network, appears as though it lies behind the Fantine group: darker lines are perceived as though they are in the foreground and thus the Fantine group, with edge

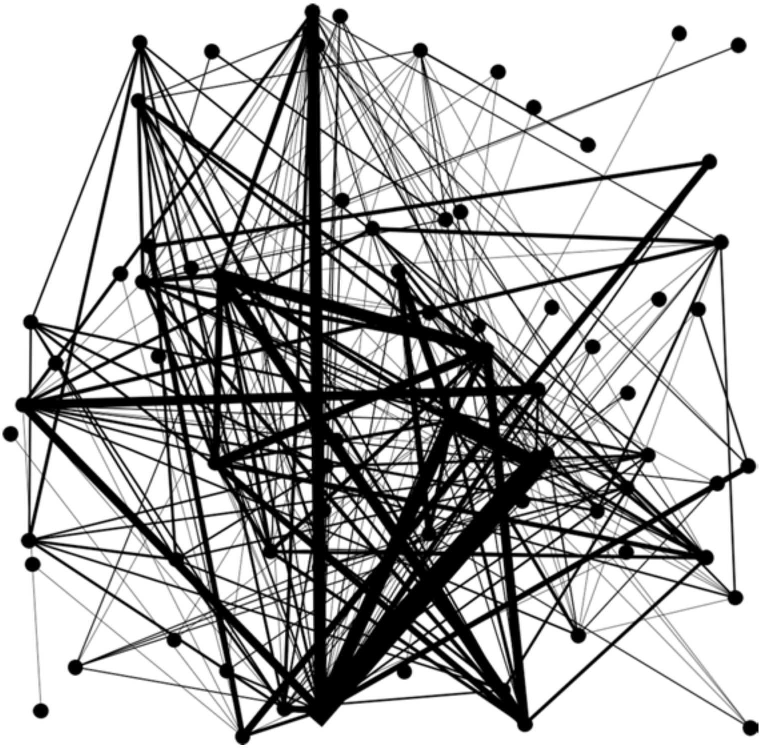

Figure 12: Gephi: initial view of the data in the Graph window.

weights ranging from 9–15, is easier to see (Ware 2013: 189–191). The uniform node separation in this layout creates a strongly rounded shape to the graph as a whole with triangular shapes between connected nodes. This layout thus emphasizes relationships between individual nodes, rather than the structure of groups or clusters within the graph.

The Yifan Hu layout algorithm uses the attraction and repulsion forces produced by pairs or groups of nodes to first produce a version of the graph at a coarse resolution. By iteratively filling in the rest of the graph structure, this algorithm reduces the power required to process very large graphs (Gibson et al. 2012: 336–337). The current version of Gephi offers both the initial Yifan Hu algorithm and the Yifan Hu Proportional layout, which adds more distance between central and outer nodes (Cherven 2015: 76). Figure 16 shows the *Les Misérables* network in the Yifan Hu layout with default settings applied. Compared to Fruchterman-Reingold, the Yifan Hu layout visually separates groups of nodes and distinguishes between tightly interconnected clusters and groups of nodes that fan

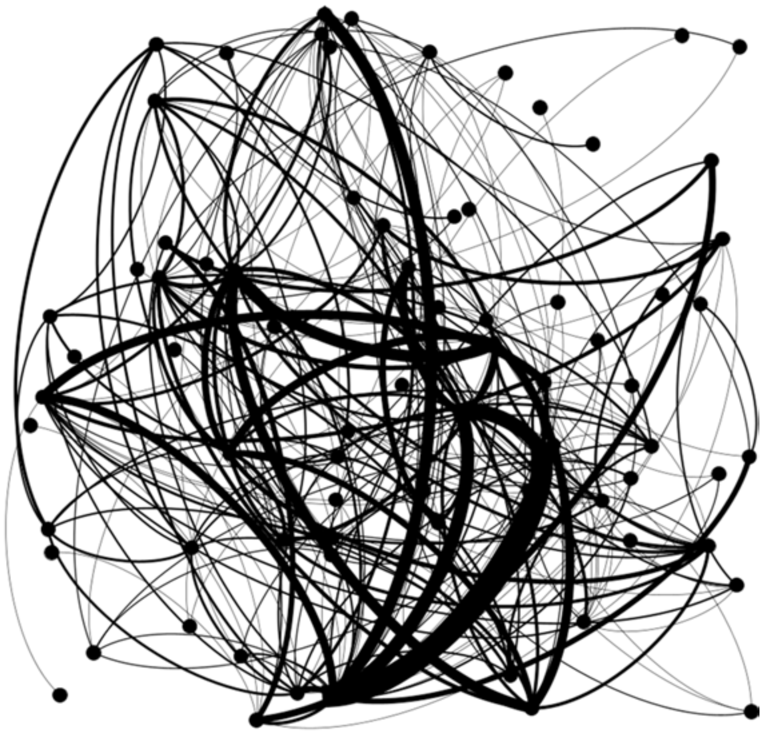

Figure 13: Gephi: initial view of the data in the Preview window.

out from a single shared connection. The Myriel group is clearly visible at the bottom of the graph. However, Gephi's default coloring of black nodes and edges, along with the scaling used to represent edge weights, can make it difficult to see all of the nodes in a network, especially with heavily weighted connections such as those between Valjean, Javert and Marius, shown here with very thick edge lines.

The OpenOrd algorithm, based on simulated annealing processes, is another multi-level approach designed for showing global structure in very large networks; however, with a network like this one with under 100 nodes, it can be less effective than other force-directed layouts (Martin et al. 2011; Gibson et al. 2012: 338). By design, OpenOrd produces longer edges in a graph than Fruchterman-Reingold to help distinguish clusters within the network's structure. OpenOrd offers a number of user-customizable parameters including edge cutting, which affects edge lengths in the graph. Figure 17 shows the *Les Misérables* network in OpenOrd with the Gephi default parameter settings and default aesthetics. Figure 18 shows the same

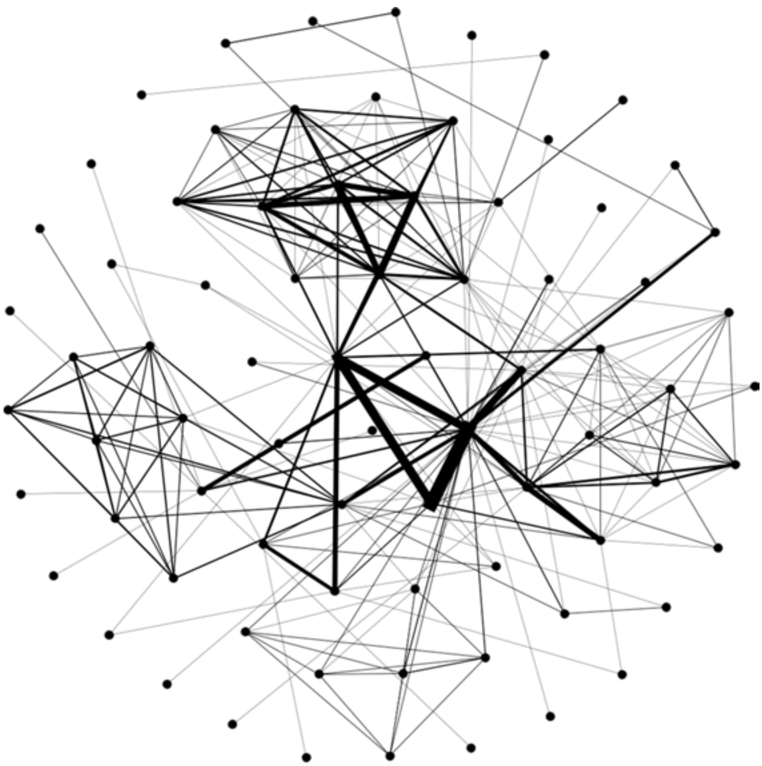

Figure 14: Gephi: Fruchterman-Reingold layout.

layout colored by modularity, which highlights four main clusters in the network: the cluster of Valjean and other high-degree main characters in the center of the graph; the revolutionary cluster at the lower right, which because of the edge width scaling looks almost as prominent; the group of seamstresses connected to Fantine at the upper right and the Myriel group at lower left. This layout emphasizes the highly weighted connections across and within clusters.

The ForceAtlas algorithm was designed by the Gephi team with the release of the software in 2009 to make possible real-time continuous network visualization with user-customizable parameters for attraction, repulsion and gravity (Bastian et al. 2009: 361). Since the 2014 release of ForceAtlas2, ForceAtlas is considered obsolete, but is still included in the program (Jacomy et al. 2014: 1–2). ForceAtlas2 is a continuous algorithm: "As long as it runs, the nodes repulse and the edges attract. This push for simplicity comes from a need for transparency. Social scientists cannot use black boxes, because any processing has to be evaluated in the perspective of the methodology" (Jacomy et al. 2014: 2). By design, Gephi users

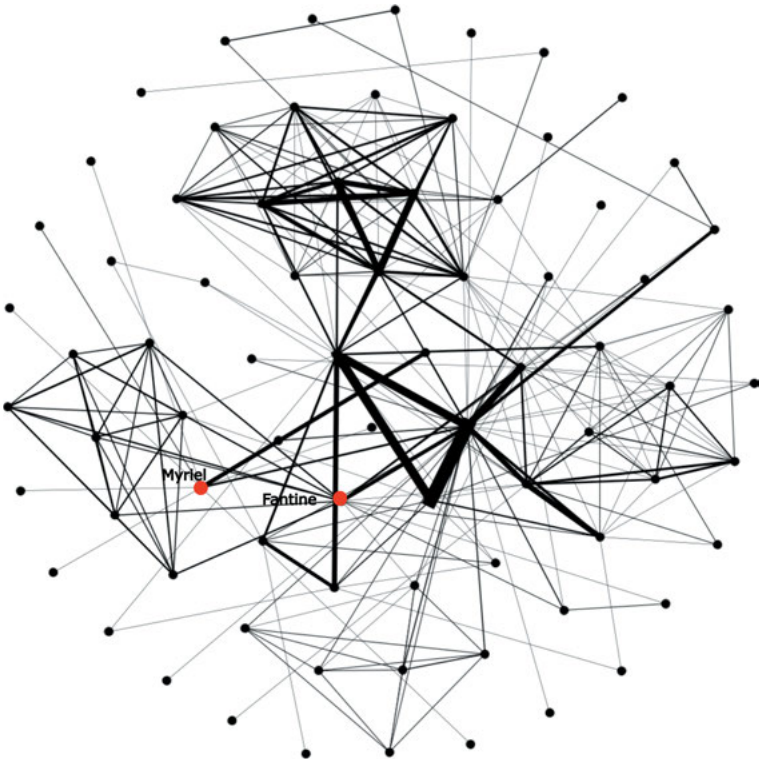

Figure 15: Gephi: Fruchterman-Reingold layout, detail of Myriel and Fantine groups.

can experiment to see the effects of changing the algorithm's parameters. Figure 19 shows the *Les Misérables* network in the ForceAtlas2 layout with default settings applied. Because the algorithm tends to overlap the nodes unless the default parameters are changed, the node clusters are difficult to distinguish in Gephi's default aesthetic, which for this algorithm displays the nodes at a large size relative to the edges. Figure 20 shows the same graph with grey, rather than black, nodes and edges, which provides a better view of the network's structure. Like Yifan Hu, ForceAtlas2 separates smaller node groups from the central highly connected nodes of the network.

As noted above, Gephi's default aesthetics scale the width of the edge lines according to edge weight, and color both nodes and edges black on a white background. As with Pajek and Cytoscape, Gephi makes it possible to change the aesthetics of node shape, size and color according to features of the data, such as node degree, statistical measures of centrality, modularity or categorical partitions in the data. However, Gephi also provides an interactive interface that al-

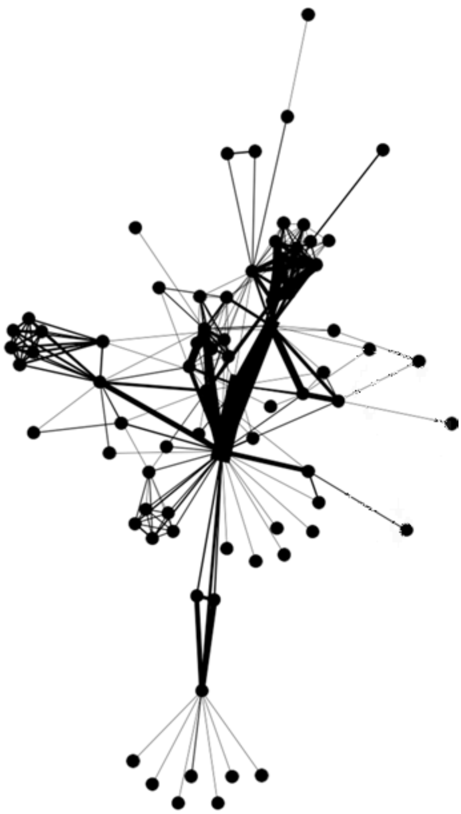

Figure 16: Gephi: Yifan Hu layout.

lows a user to click directly on nodes to move them or to apply specific aesthetics independent of the data features. Edge color, width and style can also be changed to reflect features of the data or other aesthetic preferences. Aesthetic changes are immediately visible in the Graph window and contribute to the exploratory environment for which Gephi is designed. Additional aesthetics for node and edge label styles, node opacity and borders and line shape are applied in the Preview window. With so many aesthetic possibilities, nearly infinite aesthetic transformations are possible for any given network graph, even without adjusting the layout.

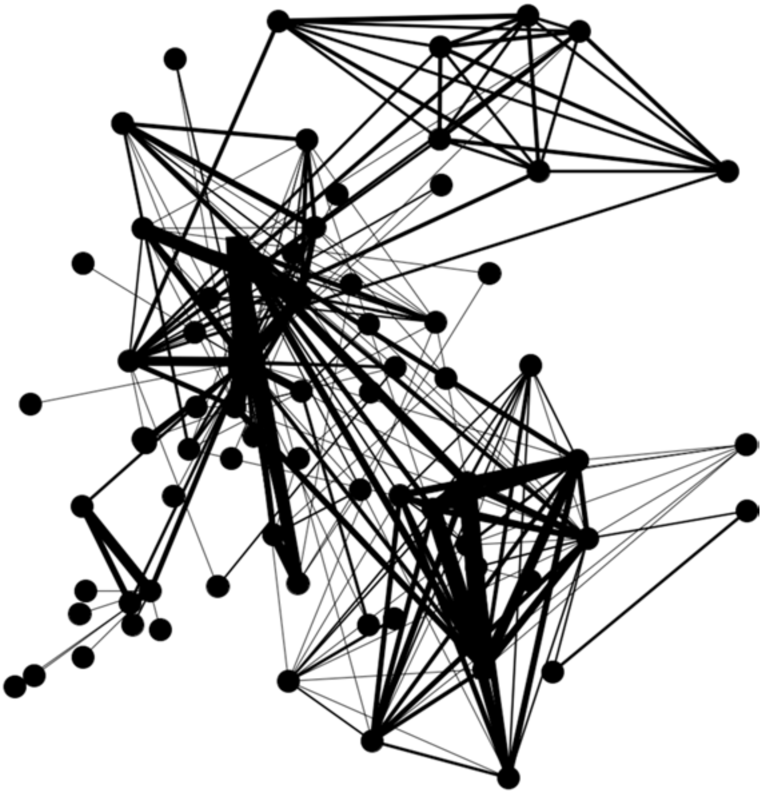

Figure 17: Gephi: OpenOrd layout.

5 Discussion

Because network visualization relies on aesthetic properties to communicate interpretations of data, understanding how different tools implement key layout algorithms and aesthetic styles is important, not only in terms of selecting the best tool for a particular project, but also for critically analyzing the products of these tools. As suggested here, the aesthetic dimension of network graphs produces meaning far beyond their explicit data mappings. Symmetry, contour, shape and color influence our attention and perception and can serve to highlight or obscure certain aspects of the network. Any network visualization is the product of numerous decisions which are too often not only undocumented, but even deliberately hidden from view. This section examines the most widely circulated visualization of the Knuth *Les Misérables* network in order to reveal the technological and aesthetic mystification it embodies.

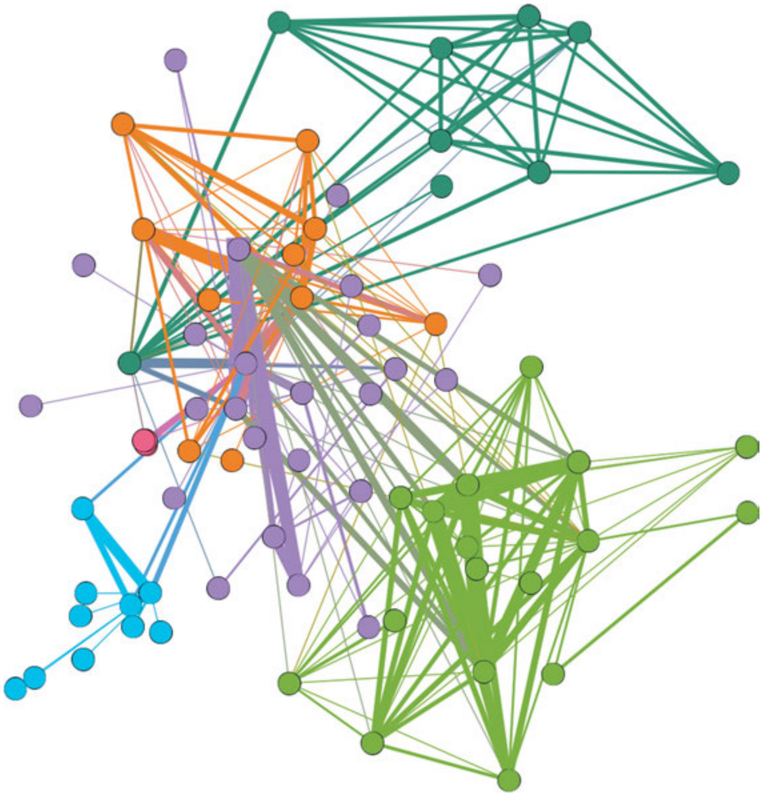

Figure 18: Gephi: OpenOrd layout with modularity.

When a user first opens the Gephi program, a welcome screen offers the user shortcuts to open recent files, to start a new project, or to open one of three network files that are included with the basic installation. These sample files offer insight into some of the assumptions and expectations that were built into Gephi's design. When a user opens the included file of Knuth's *Les Misérables* network, it appears in the program's Graph window not as unformatted, randomly ordered nodes, but with a layout algorithm and numerous aesthetic features already applied, as shown in Figure 21. These aesthetics are presumably intended to do double duty, serving not only to elucidate the character interactions that structure Hugo's novel, but also to exemplify good network visualization practices.

This network graph is built into the visualization software itself and presented without comment or explanation, as if its design makes it fully self-explanatory. However, like all network visualizations, this graph is a construction, so examining how it was made is important. It demonstrates several conventional practices in

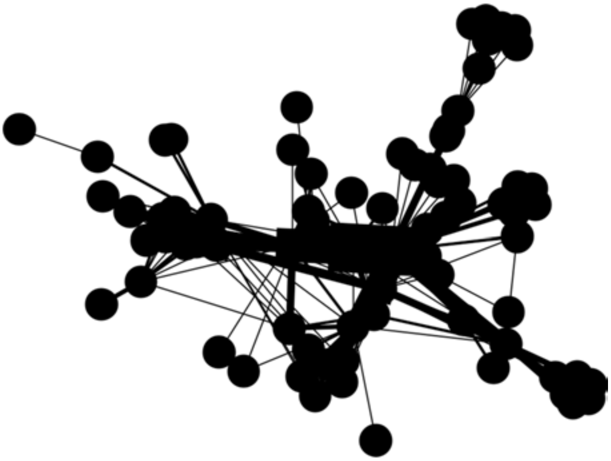

Figure 19: Gephi: ForceAtlas2 layout.

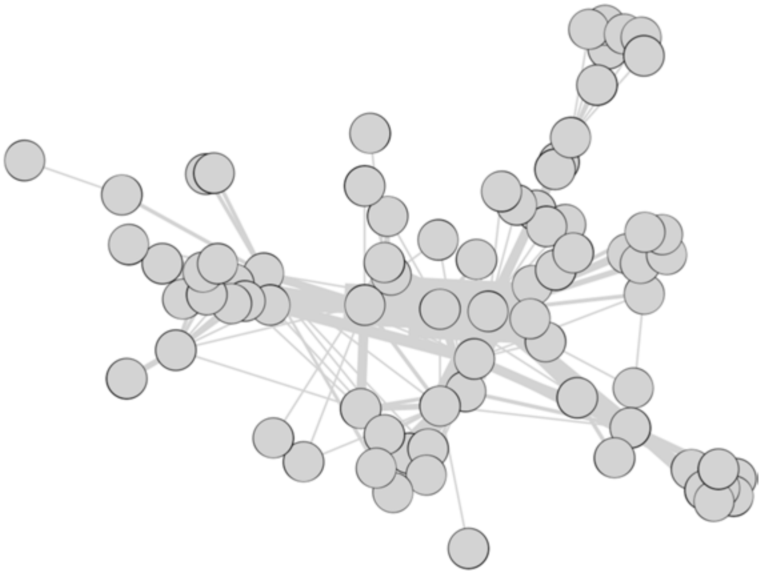

Figure 20: Gephi: ForceAtlas2 layout with grey color applied.

network visualization: nodes are sized according to node degree, edge widths are sized according to edge weight, and nodes are colored by modularity groupings. In addition, this graph embeds a variety of assumptions in its aesthetics. Bright, appealing colors and large node sizes simplify the appearance of the

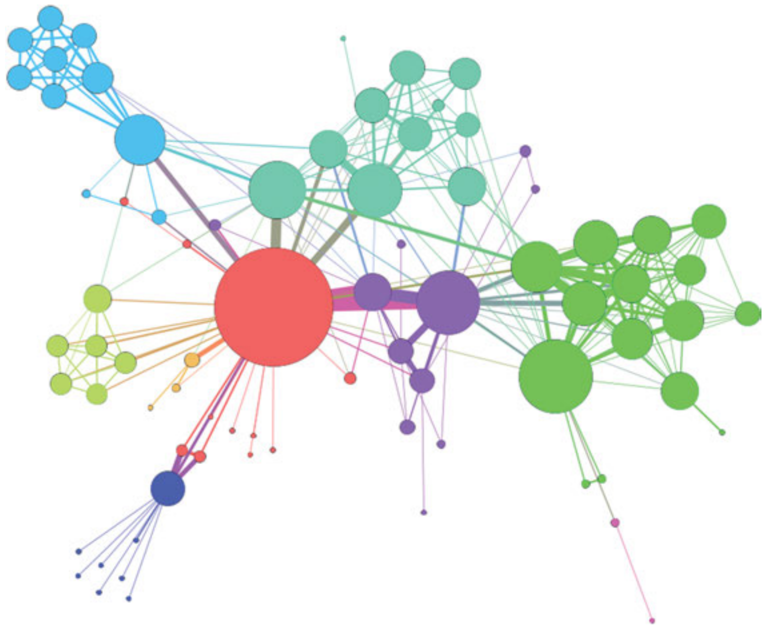

Figure 21: *Les Misérables* network graph included with Gephi installation.

graph. In this visualization, Hugo's multi-character saga is simplified into a story focused on the large red node representing Valjean. The complexity of the characters' connections are visually minimized by using very fine edge lines relative to the node sizes, which modify one's perception of the scale of the full dataset.

Not only is this graph not produced with the default settings in the software, but most of its aesthetics appear to have been manually created to produce certain desired effects. Gephi provides numerous ways for users to alter the aesthetics applied to a graph in order to reveal or enhance certain interpretations of the network; unless these adjustments are documented, which the software itself does not do, viewers may not be able to recognize their effect on the final graph. For example, this visualization uses a bright palette that is not among the standard color sets included with the program; to use this palette, each color would have to be individually specified with RGB or hexadecimal color values. Not including this example's color palette in the software almost seems like a deliberate attempt to frustrate users who might wish to emulate it.

An attempt at recreating the Gephi sample visualization using the ForceAtlas2 layout algorithm is shown in Figure 22. The process of recreating that visualization reveals several key aesthetic manipulations that were used to produce the sample visualization. Several experiments with this layout and with Yifan Hu, which ForceAtlas2 most closely resembles, suggests that the length and placement of the edges towards the periphery of the network may have been manually adjusted to create tighter clusters. In addition, the node sizes have been adjusted to produce the dramatic size differences shown in this sample visualization, rather than using a mathematically scaled approach to representing the node degree range in the network. These adjustments significantly alter the appearance of the graph and the way it guides interpretation of the novel's data.

Figure 22: Reconstruction of the Gephi sample network graph.

As shown in Figure 21, nodes in the sample visualization are colored according to modularity groups. Gephi's modularity analysis tool implements the Louvain method for community detection, which discovers small communities and then iteratively groups them into larger ones within the network (Blondel et al. 2008). When this modularity analysis is run on the *Les Misérables* network using the default settings, the program produces six modularity classes rather than the nine that are shown in Figure 21, suggesting some manual adjustments were made to produce that sample visualization. These adjustments have important effects on the visualization, because the modularity group colors and node sizings used in the Gephi sample visualization sharply distinguish Valjean from other main characters in the novel, including Javert, Thenardier, Marius and Fantine, as shown in the detail in Figure 23. In the sample file, the Valjean cluster consists only of Valjean and a few minor characters with low node degree, and the other high degree characters are clearly separated into other modularity groups.

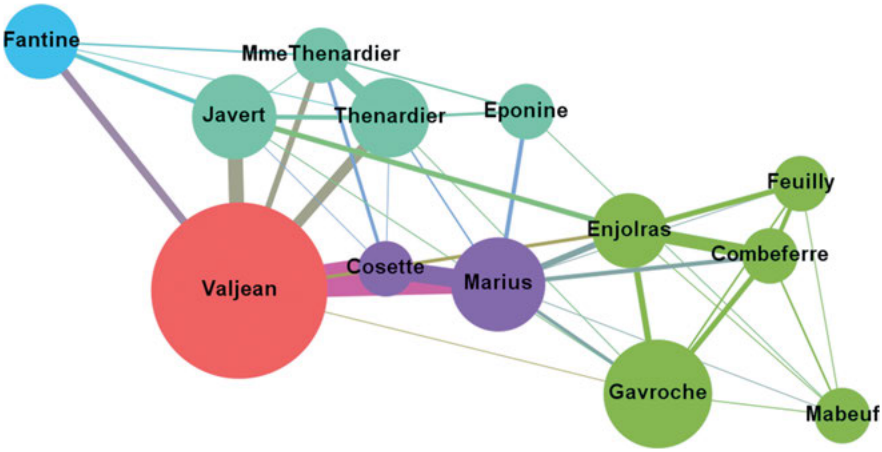

Figure 23: Detail of high degree nodes in the Gephi sample graph shown in Figure 21.

While recreating the sample visualization for Figure 22, multiple experiments were made with adjusting the modularity resolution setting, which affects how many groups are produced. However, Valjean was always grouped with one or more of the high degree characters he is highly connected to. For example, in the reconstruction shown in Figure 22, Valjean is grouped with his antagonist, police inspector Javert, as shown in the detail in Figure 24. These experiments, like the other visualizations presented in this chapter, suggest that Valjean's interconnections with a wide range of characters may be of equal or greater significance in the novel. By setting this graph as an example to learn from, but not providing a full explanation of how it was created, Gephi's designers set unrealistic expectations for simplicity and clarity in network visualizations. Even more problematic

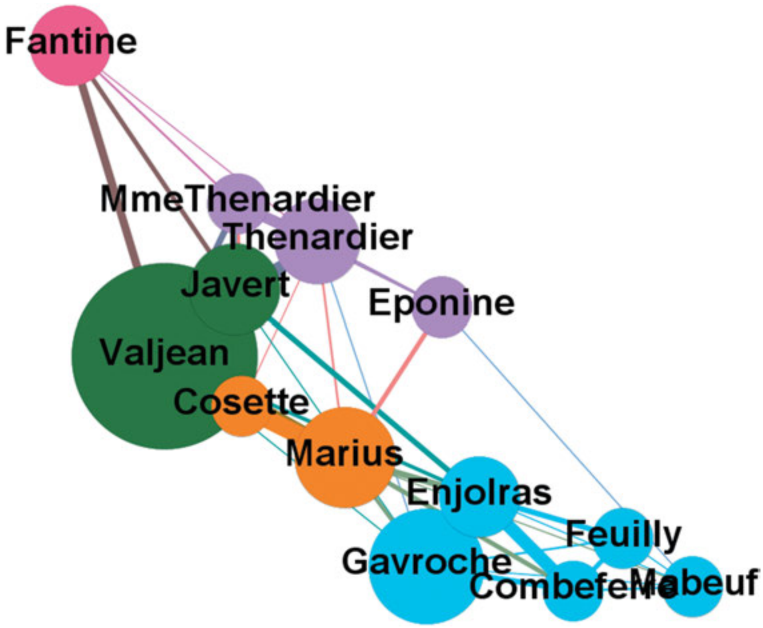

Figure 24: Detail of high degree nodes in the reconstruction of the Gephi sample graph shown in Figure 22.

is the fact that numerous aesthetics in this image were manipulated to create certain effects that shape viewers' perceptions of the underlying data.

This critical account of the default settings in three open source network visualization tools network reveals how visualization aesthetics encode meaning both explicitly and implicitly. Each of these tools offers numerous ways that the user can customize the appearance of the graph so as to promote interpretations of the data. Learning more than one tool allows the researcher to select the software that will be most appropriate for a given project. Although it works very well on large graphs, Pajek's visualization capabilities are somewhat less flexible than the other tools, due to the limited color selection, low resolution images in the Draw screen and differences between the working view and the exported image. The Cytoscape style palettes offer simple one-click adjustments to the appearance of the graph that alter multiple aesthetics at once, but most of these are designed for specific scientific purposes. Although the Gephi sample visualization demonstrates some of the rich aesthetic possibilities in the tool, its default aesthetics are very spare and learning to manipulate the many different aesthetic controls can take some time. No single tool or algorithm is necessarily better than the others; rather, understanding the differences among them and exploring a given network in multiple layouts and in different tools can provide new insights. Because network visualizations assist in understanding data structures at a variety of scales, critical awareness of how network graphs are produced contributes to a more situated, self-reflective data visualization practice that recognizes how aesthetics are always creating meaning.

References

Ahnert, Ruth, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. *The Network Turn: Changing Perspectives in the Humanities*. Cambridge: Cambridge University Press, 2020.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. "Gephi: An Open Source Software for Exploring and Manipulating Networks." *Proceedings of the International AAAI Conference on Web and Social Media* 3.1 (2009): 361–362.

Bennett, Chris, Jody Wyall, Leo Spalteholtz, and Amy Gooch. "The Aesthetics of Graph Visualization." In *Computational Aesthetics in Graphics, Visualization, and Imaging*. Ed. Douglas W. Cunningham, Gary Meyer, and Laszlo Neumann, 57–64. The Eurographics Association, 2007.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* P10008 (2008): 1–12.

Broasca, Laura, Versavia-Maria Ancusa, and Horia Ciocarlie. "A qualitative analysis on force directed network visualization tools in the context of large complex networks." In 2019 *23rd International Conference on System Theory, Control and Computing* (ICSTCC). IEEE, 2019. 656–661.

Carbon, Claus-Christian, Tamara Mchedlidze, Marius Hans Raab, and Hannes Wächter. "The Power of Shape: How Shape of Node-Link Diagrams Impacts Aesthetic Appreciation and Triggers Interest." *i-Perception* 9.5 (2018): 1–18.

Cherven, Ken. *Mastering Gephi Network Visualization*. Birmingham: Packt Publishing, 2015.

Combe, David, Christine Largeron, Elod Egyed-Zsigmond, and Mathias Géry. "A comparative study of social network analysis tools." In *Web Intelligence & Virtual Enterprises*. 2010. Accessed July 2, 2023. https://hal.archives-ouvertes.fr/hal-00531447.

Derrible, Sybil and Christopher Kennedy. "Network Analysis of World Subway Systems Using Updated Graph Theory." *Transportation Research Record* 2112.1 (2009): 17–25.

Dogrusoz, Ugur, Erhan Giral, Ahmet Cetintas, Ali Civril, and Emek Demir. "A layout algorithm for undirected compound graphs." *Information Sciences* 179.7 (2009): 980–994.

Drucker, Johanna. "Humanities Approaches to Graphical Display." *DHQ: Digital Humanities Quarterly* 5.1 (2011). Accessed July 2, 2023. http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Drucker, Johanna. *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, MA: The MIT Press, 2020.

Enright, Anton J. and Christos A. Ouzounis. "BioLayout—an automatic graph layout algorithm for similarity visualization." *Bioinformatics* 17.9 (2001): 853–854.

Fruchterman, Thomas MJ, and Edward M. Reingold. "Graph drawing by force directed placement." *Software: Practice and experience* 21.11 (1991): 1129–1164.

Galloway, Alexander. “Are Some Things Unrepresentable?” *Theory, Culture & Society* 28.7–8 (2011): 85–102.

“Gephi documentation wiki.” Accessed May 1, 2022. https://github.com/gephi/gephi/wiki.

Gibson, Helen, Joe Faith, and Paul Vickers. “A Survey of Two-Dimensional Graph Layout Techniques for Information Visualisation.” *Information Visualization* 12.3–4 (2012): 324–357.

Ginsburg, Michal P. “*Visualizing Les Misérables*.” Accessed May 1, 2022. http://lesmiserables.mla.hcommons.org/.

Grandjean, Martin. “A Social Network Analysis of Twitter: Mapping the Digital Humanities Community.” *Cogent Arts and Humanities* 3.1 (2016), Accessed July 2, 2023. DOI: 10.1080/23311983.2016.1171458.

Healy, Kieran. “The Performativity of Networks.” *European Journal of Sociology* 56.2 (2015): 175–205.

Heer, Jeffrey, Stuart K. Card, and James Landay. “Prefuse: A Toolkit for Interactive Information Visualization.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ‘05)*, 421–430. New York: Association for Computing Machinery, 2005.

Hu, Yifan. “Efficient, High-Quality, Force-Directed Graph Drawing.” *The Mathematica Journal* 10 (2006): 37–71.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. “ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.” *PloS ONE* 9.6 (2014): 1–12.

Jameson, Fredric. *Postmodernism, or, The Cultural Logic of Late Capitalism*. Durham: Duke University Press, 1991.

Kamada, Tomihisa and Satoru Kawai. “An algorithm for drawing general undirected graphs.” *Information processing letters* 31.1 (1989): 7–15.

Kennedy, Helen and Martin Engebretsen. “Introduction: The relationships between graphs, charts, maps and meanings, feelings, engagements.” In *Data Visualization in Society*, edited by Martin Engebretsen and Helen Kennedy, 19–32. Amsterdam: Amsterdam University Press, 2020.

Kennedy, Helen and Rosemary Lucy Hill. “The Feeling of Numbers: Emotions in Everyday Engagements with Data and Their Visualisation.” *Sociology* 52.4 (2018): 830–848.

“Kindred Britain.” Accessed December 15, 2021. https://kindred.stanford.edu/.

Kobourov, Stephen G. “Force-Directed Drawing Algorithms.” In *Handbook of Graph Drawing and Visualization*, edited by Roberto Tamassia, 383–408. New York: Chapman and Hall/CRC, 2013.

Koolen, Marijn, Jasmijn van Gorp, and Jacco van Ossenbruggen. “Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice.” *Digital Scholarship in the Humanities* 34.2 (2019): 368–385.

Knuth, Donald E. “The Stanford GraphBase.” Accessed May 7, 2022. https://www-cs-faculty.stanford.edu/~knuth/sgb.html.

Knuth, Donald E. *The Stanford GraphBase: A Platform for Combinatorial Computing*. New York: ACM Press, 1994.

Majeed, Sadia, Muhammad Uzair, Usman Qamar, and Aftab Farooq. “Social Network Analysis Visualization Tools: A Comparative Review.” In 2020 IEEE 23rd *International Multitopic Conference* (INMIC). IEEE, 2020. 1–6.

Manovich, Lev. *The Language of New Media*. Cambridge, MA: The MIT Press, 2001.

Martin, Shawn, W. Michael Brown, Richard Klavans, and Kevin W. Boyack. “OpenOrd: an Open-Source Toolbox for Large Graph Layout.” *Proceedings of SPIE Conference on Visualization and Data Analysis* 7868 (2011): 786806.1–786806.11. https://doi.org/10.1117/12.871402.

Moretti, Franco. “Network Theory, Plot Analysis.” *New Left Review* 68 (2011): 80–102.

Mrvar, Andrej, and Vladimir Batagelj. “Analysis and visualization of large networks with program package Pajek.” *Complex Adaptive Systems Modeling* 4.1 (2016): 1–8.

Mrvar, Andrej and Vladimir Batagelj. *Pajek, PajekXXL, Pajek 3XL Programs for Analysis and Visualization of Very Large Networks: Reference Manual*. Ljubljana: A. Mrvar and V. Batagelj, 2022. Accessed July 30, 2023. http://mrvar.fdv.uni-lj.si/pajek/pajekman.pdf.

Newman, M.E.J. and M. Girvan, “Finding and Evaluating Community Structure in Networks.” *Physical Review* 69.026113 (2004): 1–16.

Purchase, Helen C. “Metrics for Graph Drawing Aesthetics.” *Journal of Visual Languages and Computing* 13 (2002): 501–516.

Purchase, Helen C., David Carrington, and Jo-Anne Allder. “Empirical Evaluation of Aesthetics-Based Graph Layout.” *Empirical Software Engineering* 7 (2002): 233–255.

Rieder, Bernhard and Theo Röhle. “Digital Methods: From Challenges to *Bildung.”* In *The Datafied Society*, edited by Mirko Tobias Schäfer and Karin van Es, 109–124. Amsterdam: Amsterdam University Press, 2017.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. “Cytoscape: a software environment for integrated models of biomolecular interaction networks.” *Genome Research* 13.11 (2003): 2498–2504.

Sinclair, Stéfan, Stan Ruecker, and Milena Radzikowska. “Information Visualization for Humanities Scholars.” In *Literary Studies in the Digital Age: An Evolving Anthology*, edited by Kenneth M. Price, Ray Siemens, Dene Grigar, and Elizabeth M. Lorang. New York: Modern Language Association, 2013. Accessed July 2, 2023. https://dlsanthology.mla.hcommons.org/information-visualization-for-humanities-scholars/.

“Six Degrees of Francis Bacon.” Accessed December 15, 2021. http://www.sixdegreesoffrancisbacon.com.

Skiena, Steven. “The Stanford GraphBase.” The Stony Brook Algorithm Repository. Last modified July 10, 2008. Accessed July 20, 2023. https://www3.cs.stonybrook.edu/~algorith/implement/graphbase/implement.shtml.

“The Viral Texts Project.” Accessed December 15, 2021. https://viraltexts.org/.

Tufte, Edward. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.

Van Es, Karin, Mirko Tobias Schäfer, and Maranke Wieringa, “Tool Criticism and the Computational Turn: A ‘Methodological Moment’ in Media and Communication Studies.” *M&K Medien & Kommunikationswissenschaft* 69.1 (2021): 46–64.

Venturini, Tommaso, Mathieu Jacomy, and Pablo Jensen, “What Do We See When We Look at Networks: Visual Network Analysis, Relational Ambiguity, and Force-Directed Layouts.” *Big Data and Society* 8.1 (2021): 1–16.

Ware, Colin. *Information Visualization: Perception for Design*. 3rd ed. Boston: Morgan Kaufmann, 2013.

Weingart, Scott B. “Demystifying Networks, Parts I & II.” *Journal of Digital Humanities* 1.1 (2011). Accessed July 2, 2023. http://journalofdigitalhumanities.org/1-1/demystifying-networks-by-scott-weingart/.

“What is Cytoscape?” Accessed May 1, 2022. https://cytoscape.org/what_is_cytoscape.html.

Wilson, Robin J. “An Eulerian Trail Through Königsberg.” *Journal of Graph Theory* 10.3 (1986): 265–275.

Zachary, Wayne W. “An Information Flow Model for Conflict and Fission in Small Groups.” *Journal of Anthropological Research* 33.4 (1977): 452–473.