Comparing the relational structure of the Gospels. Network Analysis as a tool for biblical sciences

Martin Grandjean

Comparing the Relational Structure of the Gospels
Network Analysis as a Tool for Biblical Sciences

Martin GRANDJEAN

Society of Biblical Literature, University of St. Andrews, July 8 2013

Introduction: character network analysis

Network analysis allows the literary scholar to take a step back from a given text, by providing both an overview to approach the object differently, and a new structural and mathematical angle. With this paper, we wish to present a process, some elements of which lay the basis of an approach that we think biblical studies should seize.

This paper is divided into four main parts: after a short introduction, it addresses, both generally and technically, the stakes of network analysis for narratology. Secondly, it presents the results provided by a network analysis of characters in the four Gospels. Then, a third part examines several levels of analysis: visual, quantitative and mathematical. Before a conclusion, a fourth section discusses the benefits of scaling in network analysis, focusing on particular groups within the narrative field.

While character network analysis and visualization has evolved considerably over the past ten years, it remains a stammering field. This contribution can be seen as a continuation of the debate launched by Cohen (2007) which questions some uses of network analysis and visualization in literary studies. Commenting on Wattenberg and Viégas (2007) promoting *Many Eyes*, the new IBM online data visualization service, he rightly shows the limits of the example chosen to prove the merits of this tool: an analysis of the New Testament presenting a so-called "social network of the New Testament" based on the co-occurrences of characters in chapters. Cohen's criticism is directed to the wonder of researchers towards a tool merely highlighting what anyone could have established: the centrality of the person of Christ in the New Testament. He concludes "Make sure your visualizations expose something new, hidden, non-obvious". In fact, the authors of this questionable visualization do not draw such conclusion since they keep to the technical exercise. But it shows very well that data visualization - and network visualization in particular - should not be limited to a nice visual result.

Our goal here is to bring other elements to this discussion and to show that the centrality of the Christ in the Gospel can be questioned through network analysis. The narrative situation is not as simple as it appears, especially since Jesus is not only present through the occurrences of his own name. We will firstly show how such a simple automated distant reading is not sufficient, and secondly that the visual aspect of a network isn't by far its only reading key. It is also to show that in the world of network graphics, the interest is sometimes actually in the details rather than in what is immediately obvious, at the center of the graph.

A text is a network

Characters analysis and narratology

The shape of a given narrative influences its analysis. More specifically, the point of view of the narrator (see the typology in Abrams 1999:231-236) determines the axis of the quantitative study based on the characters in a novel. Various studies, following Barthes' (1975:238) advice not to be reluctant to structural analysis inspired by experimental sciences have focused on the interactions between characters (e.g. Batagelj et al. 2002, Stiller and Hudson 2005, Elson et al. 2010, or Sack 2013). This allows the partial reconstruction and visualization of the latter's social network, although it is limited by the gaze of a more or less omniscient narrator. There is no "social" network in our study, because our goal is to avoid these social approaches by analyzing the simultaneous presence of characters in the scenes of the story. Such an analysis is made possible in the Gospels by the omniscient narrator's position. It is therefore not a mapping of social interactions but a comprehensive and interpretative vision of the facts that are told, a focus on the content of the narrative and not its form. Thus, by mapping the presence of the characters, we analyze the space created by the narration, we reconstruct a fictional space based on character appearance metadata. In this respect, it is interesting to recall how Rimmon-Kenan (1983:29-42 cited by Hunt et al. 2013:5) describes the difference between the "level of the story" and the "level of the text".

Analyses that automate terms, nouns and pronouns indexing (e.g. Keim et al 2008 or Mott et al. 2006), seem to forget that characters are built by the reader, as Chatman (1978:121-130) defends. It is not enough to identify their strict occurrences; one must also focus on their implicit appearances (in a group, in the background of an action, etc.). Note also that the analysis of co-occurrences of characters (their identification in a common scene) is one of six aspects of Finnerrn (2010:162-164) methodology for the study of Biblical characters, without forgetting that "the guiding principle behind literary network analysis is that narratives are not merely depictions of individual experience in language but are also artificial societies whose imaginary social forms can be quantified and analysed" (Sack 2013:185).

Of course, the analysis and visualization of the actors in a text is only one way among others to understand the content of a narrative, often studied through its succession of topics (e.g. Jacquemin 1997, Budanitsky and Hirst 2001, Plaisant et al. 2006, Cui et al. 2011, or Kim et al. 2011), language specifications (Rydberg-Cox 2011) or simply words themselves (Burrows 2004 or Vuillemot et al. 2009), not to mention traditional qualitative analyzes (Hamon 1998 and Bennema 2009).

Watching a text as a network

To extract a network out of a text, we must first divide the text into parts. Compared to many literary works that do not have a well-defined structure, the biblical texts[1] are particularly easy to divide due to their very precise and studied internal organization, although the choice of the unit is large. We have chosen a division into sections of several verses, the pericopes. It has the great advantage of being a coherent thematic, temporal and spatial unit.

In less structured texts, researchers rely on other types of divisions: ten-words frame in Sack (2013:186) on Cervantes, Dickens, Woolf or page by page in Rochat et al. (2013) on Rousseau, but these approaches only consider the mentioned characters, not the characters present in the scene without being cited (which of course would be impossible for many novels). Analysis by ten-words frames or pages postulates that spatial proximity of proper names in a text means a relationship between them (this measure is still expertly weighted to minimize the effect of page turning). Analysis by unity of action - as the text sections we have chosen, but also scenes in a theatre play or a movie script - ensures that meaning. Taking the section as a unit goes against traditional practices, for this division is not dependent on the original text (which raises the question of what an original Biblical text is) but rather on the given edition. We will see how this division is in fact unproblematic (it does not bias the analysis, especially since this division is interpretative and still respects the unity of place) but also that it allows the analyst to leave the textual reality and focus solely on the content and action reality, which is beneficial for our characters analysis.

The advantage of focusing on the presence and not the mere naming of a character is to leave a blind automated *distant reading* for an interpretative qualitative approach. Obviously, this requires a very careful reading of the text and a meticulous tracking of all "presences"; something which is made easy, in the case of the Gospels, by a small, well studied corpus.

One must first identify all the actors in the section, i.e. those involved in the action which takes place. Those cited as references (the Lord, Moses, Isaiah, etc.) and the crowd[2] are therefore excluded. Furthermore, one can extrapolate the presence of specific characters from global references (the twelve disciples become Peter, John, James, etc.).

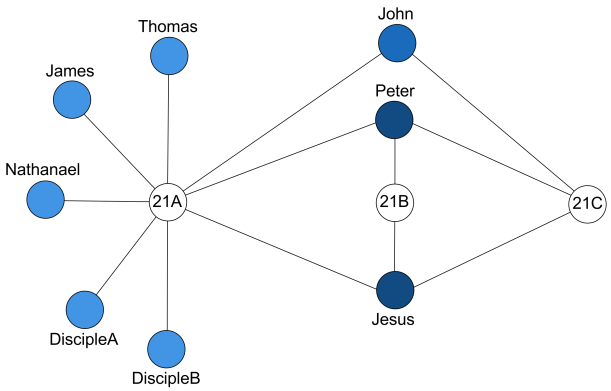


Figure 1. John 21 is divided into three sections A, B and C (sets of verses that forms coherent units).

In our first example from John 21 (the miraculous catch), randomly selected for its diversity, we consider three sections: Jesus meets seven disciples (21A), then speaks with Peter (21B), then with Peter in the presence of John (21C).

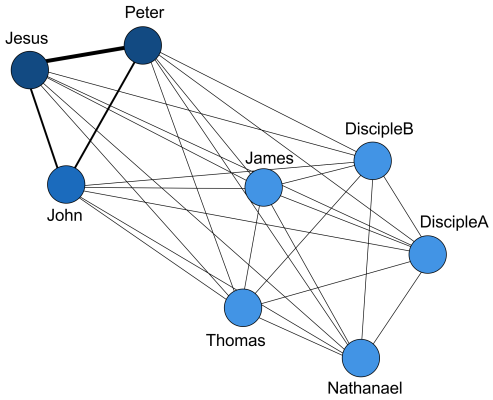


Figure 2. Co-presence graph (projection of fig.1).

We can represent the situation in a simple graph (fig.1): eight characters are listed in the first subdivision, two are present in the second, and three are present in the third. Here, an edge connects each character to the fragment(s) of text in which he appears. This is visually interesting for such a short excerpt. However, this type of bipartite graph complicates the analysis if the volume of data to visualize increases. The data is therefore visualized as a projected graph (fig.2), where an edge materializes the presence of the two characters in the same section, eliminating the "references" nodes to keep only the "characters" nodes.

As all the characters were present in the first section, there is an edge between each of them. Some edges are thicker because of several (two or three) co-occurrences. We see a denser group consisting of Jesus, John and Peter, leaving the other disciples after the miraculous catch.

[1] In this paper, the division of the Gospels into chapters, sections (pericopes) and verses is taken from the Nouvelle Bible Segond (NBS 2002) in French.

[2] The crowd can rightly be considered as a "character", but the fact that this group changes composition in every scene is problematic.

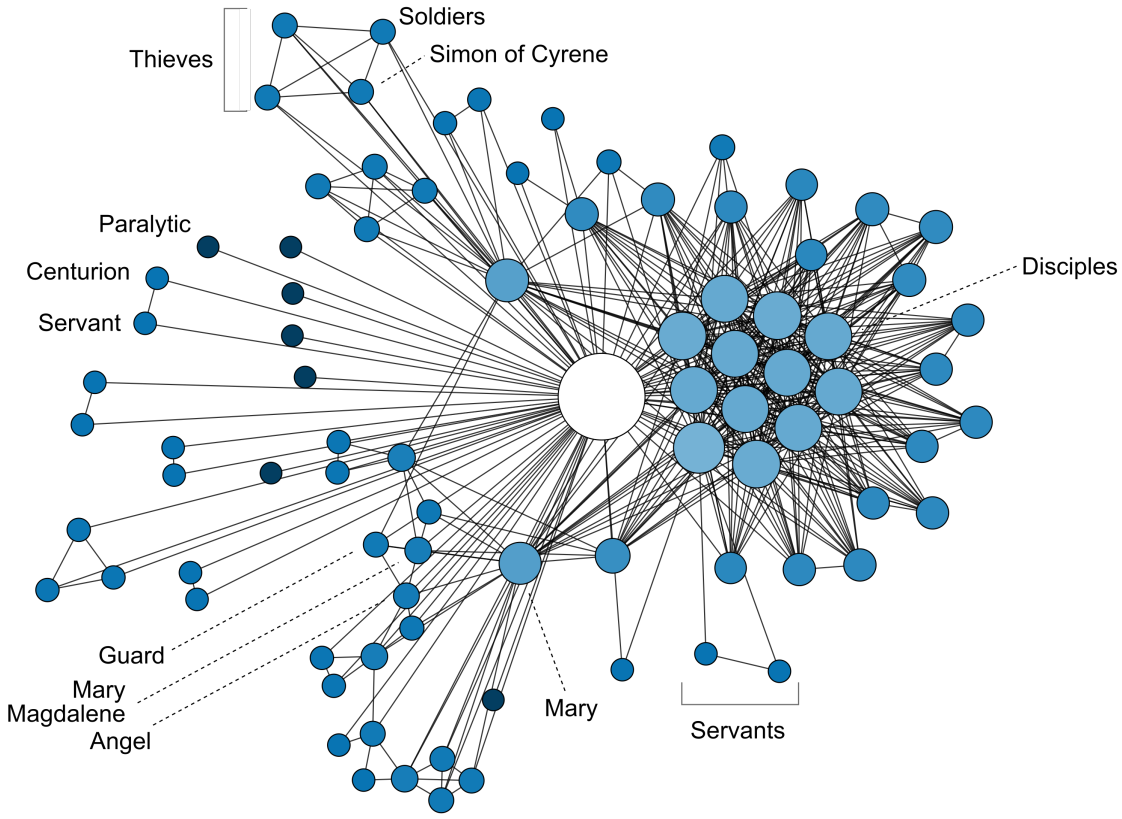


Figure 3. Gospel of Matthew

Four Gospels

Matthew

Fig.3 shows the result of this procedure for the Gospel of Matthew. The size of the nodes and their color varies according to their *degree centrality* (Koschützki et al. 2005), the number of connections that link them to other characters (we always speak about co-occurrences in the sections/pericopes). Their spatial distribution is obtained using a *force-directed* algorithm that moves nodes in the manner of magnets seeking to reject others but being attracted to those they are connected with.[3]

Jesus logically occupies the central place, here in white.[4] Secondly, one can clearly see the twelve disciples, which are densely intertwined to each other as they often appear simultaneously. In this graph, Jesus is connected to every single other character, except the two servants because of whom Peter denied him.

We can differentiate the characters connected both to Jesus and the disciples (above right) and those who appear in events where only Jesus is present (left). At first glance, we also clearly distinguish identified groups such as the cluster composed by Mary, Mary Magdalene, the guards and the angel of the tomb scene. Another cluster represents a paralytic who met Jesus alone. By zooming in on the image, we observe that the situation is not as obvious as it first appears: the twelve disciples are not all equally connected with the rest of the entourage of Christ. We will focus on these differences in the last section of this paper, but we can already see how character network analysis does not only produce quantitative output but also structural information about the characters' involvement in the plot.

Mark

As can be expected, the network created from the Gospel of Mark (fig.4) is very similar to that of Matthew's, with a few more characters. One can clearly observe the ministry above the graph and the "Easter" cluster at the bottom. A very small cluster (left) composed by the paralytic and his 4 carriers (plus the scribes criticizing Jesus) can also be found on the left.

Luke

Because of its high number of characters, the Luke graph (fig.5) is particularly interesting. One can clearly see all the peripheral scenes and actions. It's also the only graph where some characters are disconnected from the whole network. These are Luke and Theophilus, because of the very first verse in which the evangelist addresses the recipient of the Gospel. Since this situation fundamentally changes the graph's properties, we shall ignore these 2 people for the sheer sake of our analysis. According to their meta status, they do not partake in the narrative as the other characters do.

[3] Network graphs produced using Gephi (Bastian et al. 2009).

[4] As in the graph commented by Cohen (2007), yet that graph depicted the whole New Testament divided into chapters.

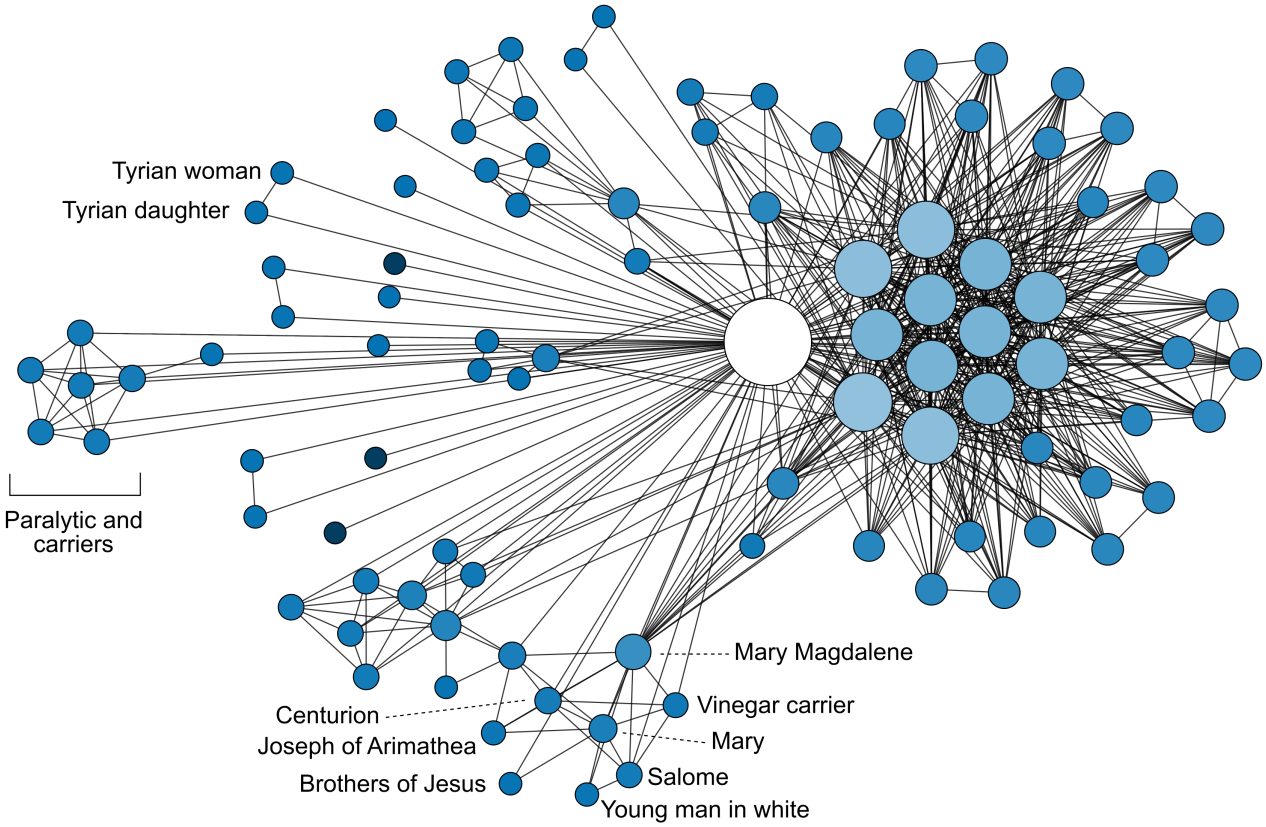


Figure 4. Gospel of Mark

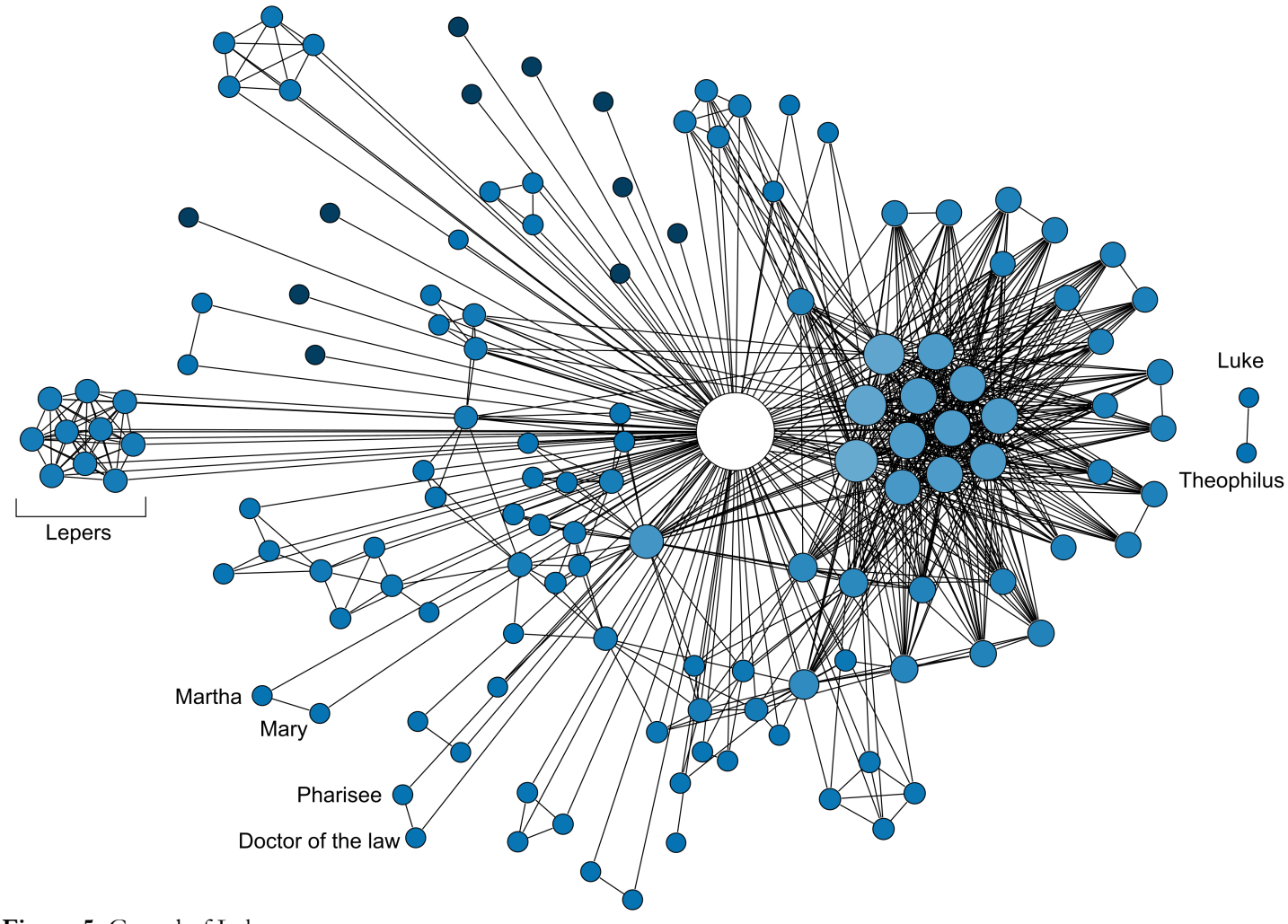
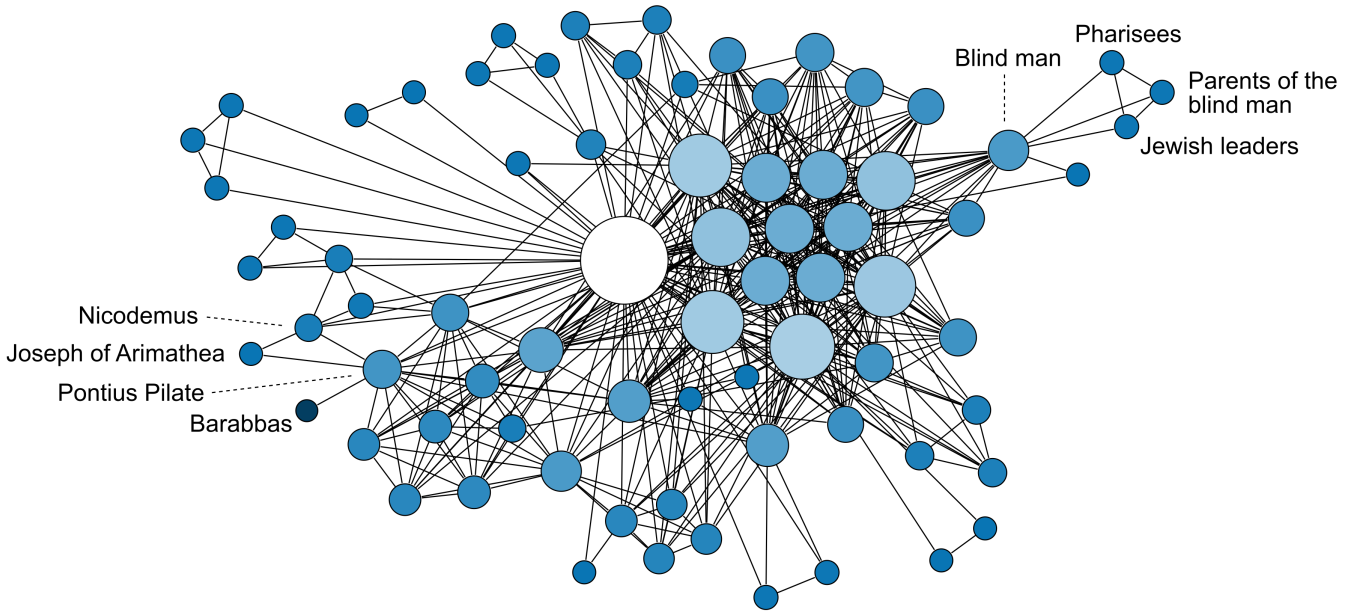


Figure 5. Gospel of Luke

John
Because it is entirely different from the three synoptic Gospels on a narrative level, this Gospel's graph is also visually very dissimilar (fig.6). The twelve apostles are less closely interwoven, and the side stories are less visible (on the left) in a graph where everything seems to be part of a unique big cluster.

Even though it is not possible to make a precise analysis of the differences in these four networks, a global panorama such as this already provides a fresh outlook and interrogate our assumptions. Indeed, using the same spatialization algorithm highlights the structural differences.

Figure 6. Gospel of John



Three levels of analysis

Graphs and visual aspects

The first level of analysis is the visual one. It should not be depreciated because of its simplicity. In the analysis of very large corpora, it is sometimes very informative for the researcher to look at global patterns (e.g. Ryan 2007). Seeing the object from up close provides a global outlook which can then feed into a more precise analysis. In our case, we can already see that the visual comparison of the four graphs obtained above allows us to make some assumptions.

General properties

The second level of analysis is the quantitative aspect of the corpus: the number of chapters (fig.7) is an obvious information, as is the number of sections (fig.8) or words (fig.9). The number of characters (fig.10) however, along with that of relations between those characters (fig.11) are not trivial and give insights about the network graphs.
Unsurprisingly, the number of relations seems to be proportional to the number of characters. But this proportionality is not always observed: for instance, the number of relations is very high in Mark, considering its length (few text sections, so fewer possibilities to contain relations).

This is a first quantitative indicator of a qualitative aspect of the text, that is, the fact that it perhaps contains more action and dialogue scenes and less teaching and descriptive scenes than other Gospels.

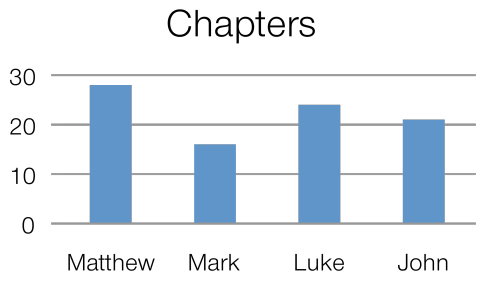


Figure 7. Number of chapters

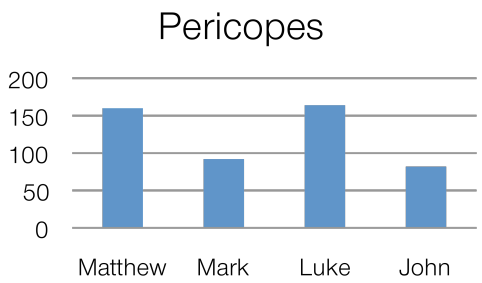


Figure 8. Number of sections

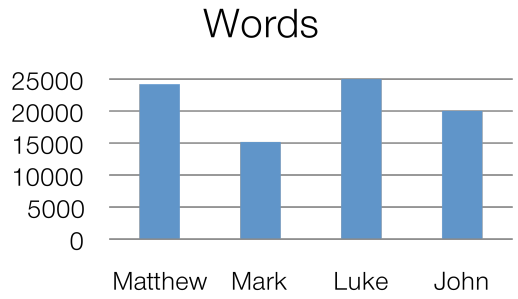


Figure 9. Number of Words

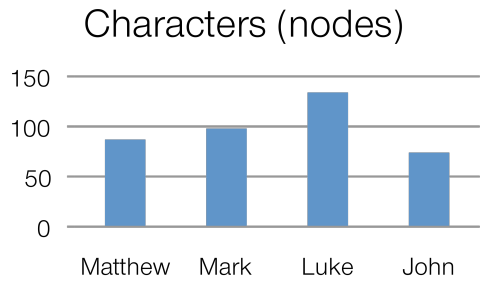


Figure 10. Number of characters

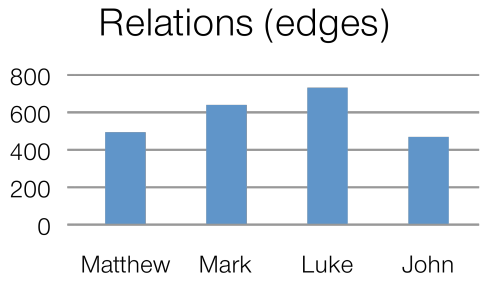


Figure 11. Number of relations

Thus, to derive information through a quantitative approach restores the qualitative aspect of the object, as Moretti argues when he writes, speaking about the number of significant characters present in a narrative space, that "the difference between five and fifteen is not just a matter of quantity, here: it's a qualitative, morphological one" (Moretti 1999:68). This also puts qualitative information in an intelligible context (Moretti 1999:149).

Network properties

The third level of analysis concerns the network mathematical properties themselves. Indeed, the humanities and social sciences can take advantage of network analysis, for it provides them with many useful metrics to categorize graph properties and well-described algorithms. The analysis of these metrics differs from visual analysis: the results do not need to be subjectively interpreted to be usable (the visual result of a network analysis may be displayed differently depending on the spatial algorithm selected).

In our example, we'll have a look at three of them, and use them to conduct a comparison of the characteristics of Luke and John.

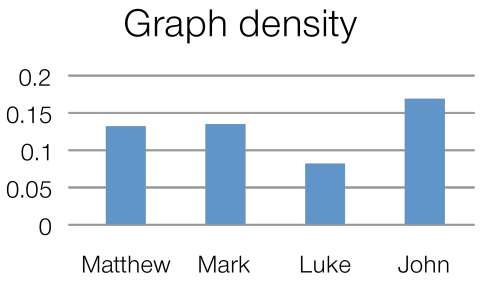


Figure 12. Graph density

The *density* (fig.12) describes the completeness of the network. It measures how close the network is to being complete. A complete network has all possible edges (an edge between each node) and a density equal to 1. Here, because of it high number of nodes, Luke has a much lower density (0.082) than John (0.169).

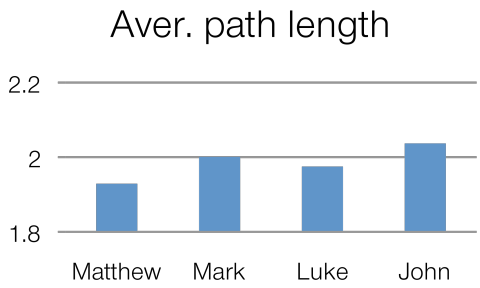


Figure 13. Average path length

The *path length* (fig.13) is a value obtained by testing the distance between all pairs of nodes. Adjacent nodes have a graph distance of 1. The average of all these paths provides information about the structure of the network. Although John was the most dense graph, its average path length is very high, which is not obvious at first. This is due to the position of Jesus. While he is connected to almost all characters in the synoptic Gospels, it is not the case with John (take, for example, the position of this blind man, connected with Pharisees and his parents, who are themselves not connected with Jesus). If Jesus had been connected to everyone, then the path length would always be 2 because everybody could go through Jesus to reach another character. In this hypothetical case, the value would obviously be less than 2, because of the connections between the characters themselves.

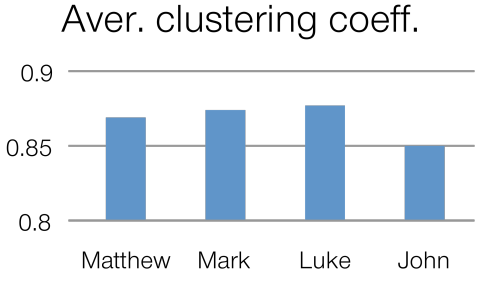


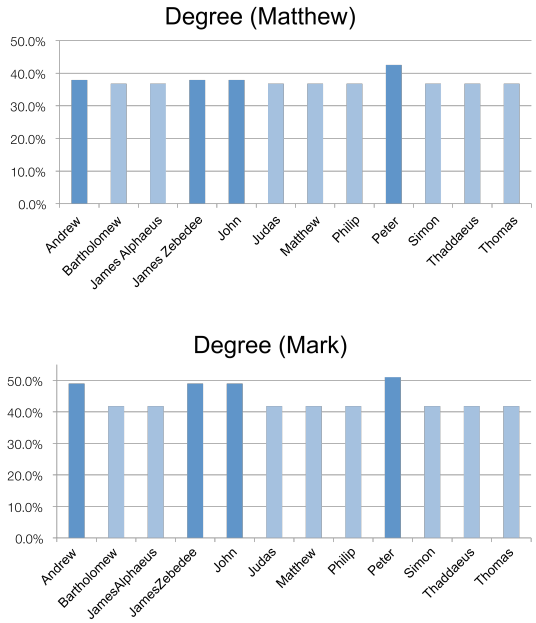
Figure 14. Average clustering coefficient

The *clustering coefficient* (fig. 14) is a way of measuring the nodes' tendency to appear as groups of various sizes. It indicates how densely connected the neighborhood is.
A very heterogeneous network will have a high coefficient. It appears that John doesn't contain as many little communities as Luke and the other synoptic Gospels (this is a confirmation of the "visual" analysis, shown above). One can also observe that the disciples cluster is less interconnected than in the other texts: the disciples appear less united in John.

Changing scale

Focusing on a structured group: the Apostles

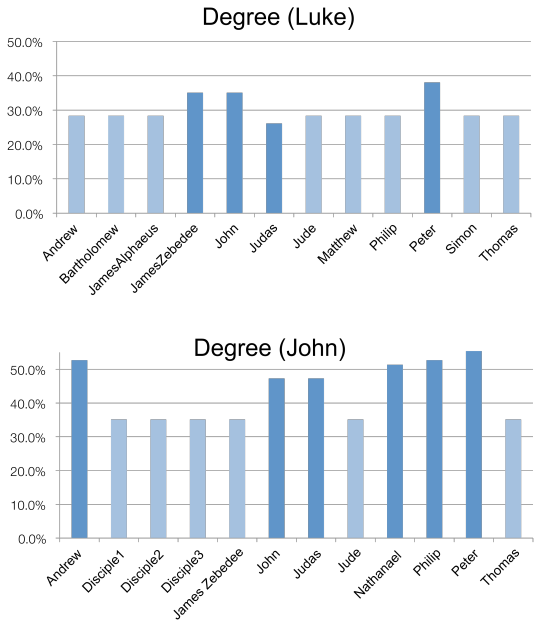
Until now, we have tried to obtain a global view of the network, but the most interesting observations are generally possible when focusing on specific parts of networks. This explains why the "magical" heuristic overviews of the data visualization should not be sought at any price: the most important is to clearly define the area to be analyzed.
One way to problematize this approach is to focus on an already formed group, such as the twelve disciples. In the Gospel of Mark, four disciples stand out: the two sons of Zebedee, along with Peter and Andrew. As they appear earlier in the text, they are more often cited along with other characters, giving them a degree (number of connections) above average. Peter also has a higher degree than the three others.



Figures 15 and 16. Degree centrality of the 12 Apostles in Matthew and Mark.

In the three Synoptic Gospels, the first followers of Jesus have a superior degree to the others. They are four in Matthew and Mark, and three in Luke. Although Judas appear in the lower middle in the Synoptics (above all because of his leaving the group after the Last Supper), he holds an important position in John.

In John, the most connected disciples exceed by far the number of connections of the other disciples, which is quite remarkable and shows that the passages where the disciples are all together are very occasional in this text.

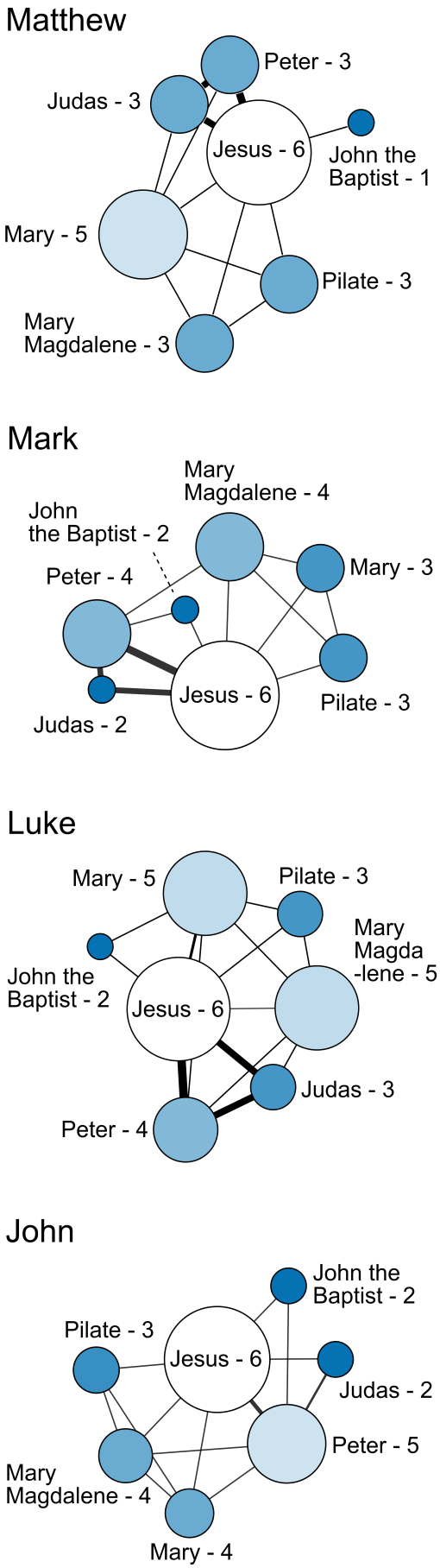


Figures 17 and 18. Degree centrality of the 12 Apostles in Luke and John.

These charts, and therefore the data behind them can become signatures of the networks underlying the four texts. Their four distinguishable properties make them instantly recognizable to the reader able to decipher them.

Focusing on an isolated core network

Another way of trying to develop a graph signature is to isolate a nucleus of well-defined characters. Consider Jesus, Mary, Mary Magdalene, Judas, Peter, Pilate and John the Baptist, and focus on the internal connections of this small group. The strong relationship between Jesus and his disciples is represented by thick edges in fig. 19-22 because they very often appear in the same sections of the text. Less involved in this core network, Mary, Pilate and Mary Magdalene are connected because of their appearance in the section in which Joseph of Arimathea asks for permission to take Jesus down from the cross. In a very visual way, we see four new signings. Again, it is not worth looking at the position of Jesus, but at the secondary characters.
It is to be noted for example that John the Baptist changes "partners" in each Gospels: he only is with Jesus in Matthew (fig.19) at the baptism. He is with Peter in Mark (fig.20), because his capture is placed in the same section as the first calling of the disciples. He is connected to Mary in Luke (fig.21), as he is concretely present when Mary visits Elizabeth. Finally, he stands with Peter in John (fig.22), as Peter is one of the two disciples of John the Baptist who begin to follow Jesus.



Figures 19-22. 'Core network' around the same set of characters. The number that appears next to the name indicates the degree centrality of the character.

Again, beyond the visual aspect, the degree centrality of the characters is an element that allows recognizing patterns. Without forgetting that a pattern is not only to be found in regular forms: the lack of regularity is also meaningful. Similarly, we should note that network analysis is not the only way to search for patterns in such a text, as evidenced by Oelke et al. (2013) about fingerprint matrices, or Don et al. (2007).

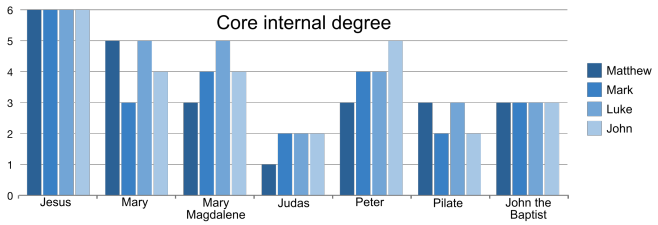


Figure 23. Comparison of the degree of the main characters in the four texts.

As we have done above, we can apply to this core network the same mathematical calculation (fig.23), or vary the composition of the core with recurrent characters. This kind of focus may sometimes be more interesting in the least explored corners of the network than in the core network (see Williams 1996 and Focant 2003 on secondary characters in Marc, Hunt et al. 2013 on John, or Woloch 2003 who questions the role of minor characters in novels, in a rather inspiring way), as far as the data is rich enough to avoid drawing conclusion on exceptional or completely marginal cases.

Perspectives

In this study, we showed that network analysis and visualization can lead to a new understanding of an already well studied object. We can easily distinguish the four Gospels in the three selected observation levels. Visually, John is very different from the three Synoptics. Quantitatively, although much shorter than the other three, Mark has a very high density of relationships. Finally, in terms of network properties, John is the most fragmented. This article also shows that focusing on small selections of characters, makes it possible to distinguish four patterns, four different signatures.

Definitely, network analysis carries interesting potential for biblical studies. When applied to this kind of textual corpus, it complements the research on authorship attribution – although the latter does not always need this kind of visual aspect, e.g. the "Delta" proposed by Burrows (2002) – since it can be used both syntactically and semantically. Without revolutionizing the discipline, this tool illustrates and provides additional insight to the question of the synoptic problem. In particular, opportunities exist in the "horizontal" comparison of the texts, which can also be extended to all the apocryphal manuscripts.
This method can also be of great interest in the "vertical" comparison of the many versions of the same text, mapping

syntactic units instead of characters. In both cases, horizontal and vertical, these means do not replace the expert eye of the researcher but speed up the process. They also allow the systematic comparing of information, offering a new look at this well explored object.

This approach can be completed by a dynamic analysis (see Van de Bunt et al. 1999, Brandes and Corman 2003, Carley 2003, Berger-Wolf and Jared 2006, Goldenberg and Zheng 2006 and Stein et al. 2010) to compensate for the fact that the display of a comprehensive network does not describe all aspects of the narrative, thus avoiding some characters that appear very infrequently, but in a context where there are many stakeholders, taking center stage. Very important people may be thrown to the periphery of the graph if their presence is irregular, or if they are only present along, and overshadowed by the main heroes (the opposite is also true: according to Agarwal et al. 2012:94 the *Mouse* in *Alice in Wonderland* is very central in the complete graph despite appearing only once in a crowded chapter). However, a dynamic network is sometimes hard to read, especially because it involves the risk for the reader of losing track of the "mental map", when facing a graph whose points change space, colour or size too drastically (Purchase et al. 2007). This problem can be partially offset by an adjacent matrix visualization (Stein et al. 2010), yet the latter raises further issues of large datasets.

Bibliography

- Abrams 1999. *A Glossary of Literary Terms*. 7th ed. Boston.
- Agarwal, Corvalan, Jensen and Rambow 2012. "Social Network Analysis of Alice in Wonderland." *NAACL-HLT*. 88-96.
- Agarwal, Rambow and Passonneau 2010. "Annotation scheme for social network extraction from text." *Proceedings of the fourth Linguistic Annotation Workshop*. 20-28.
- Barthes 1975. "An introduction to the structural analysis of narrative." *New Literary History*, 6:237-272. (translation Duisit)
- Bastian, Heymann and Jacomy 2009. "Gephi: an open source software for exploring and manipulating networks." *International AAAI Conference on Weblogs and Social Media*. 361-362.
- Batagelj, Mrvar and Zaversnik 2002. "Network Analysis of Texts." *Network Technologies*. 143-148.
- Bennema 2009. *Encountering Jesus: Character Studies in the Gospel of John*. Milton Keynes.
- Berger-Wolf and Jared 2006. "A frame-work for analysis of dynamic social networks." *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York:523-528.
- Brandes and Corman 2003. "Visual Unrolling of Network Evolution and the Analysis of Dynamic Discourse." *Information Visualization* 2(1):40-50.
- Budanitsky and Hirst 2001. "Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures." *Proceedings of the NAACL 2001 Workshop on WordNet and other lexical resources*. Pittsburgh:29-34.
- Burrows 1987. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. New York.
- Burrows 2002. "Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17(3):267-287.
- Burrows 2004. "Textual analysis." in Schreibman, Siemens and Unsworth *A Companion to Digital Humanities*. Oxford.
- Carley 2003. "Dynamic network analysis.", In Breiger, Carley and Pattison 2003. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. Washington:133-145.
- Celikyilmaz, Hakkani-Tur, He, Kondrak and Barbosa 2010. "The actor-topic model for extracting social networks in literary narrative." *Proceedings of the NIPS 2010 Workshop Machine Learning for Social Computing*.
- Chatman 1978. *Story and Discourse: Narrative Structure in Fiction and Film*, New York.
- Cohen 2007. "It's about Russia." dancohen.org/2007/03/06/its-about-russia.
- Cui, liu, Tan, Shi, Song, Gao, Qu and Tong 2011. "Textflow: Towards better understanding of evolving topics in text." *IEEE Transactions on Visualization and Computer Graphics* 17 12:2412-2421.
- Don, Zheleva, Gregory, Tarkan, Auvil, Clement, Shneiderman and Plaisant 2007. "Discovering interesting usage patterns in text collections: integrating text mining with visualization." *Proceedings of the 16th ACM Conference on Information and Knowledge Management*. 213-222.
- Eder, Jannidis and Schneider, eds. 2010. *Characters in Fictional Worlds: Understanding imaginary beings in Literature, films and other media*. Berlin.
- Elson 2012. *Modeling Narrative Discourse*. Thesis Columbia University.
- Elson and McKeown 2009. "Extending and Evaluating a Platform for Story Understanding." *Proceedings of the AAAI 2009 Spring Symposium on Intelligent Narrative Technologies II*:32-35.
- Elson, Dames and McKeown 2010. "Extracting Social Networks from Literary Fiction." *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala:138-147.
- Fekete and Dufournaud 2000. "Compus: visualization and analysis of structured documents for understanding social life in the 16th century." *Proceedings of the 5th ACM Conference on Digital librairies*. 47-55.
- Ferrara 1974. "Theory and Model for the Structural Analysis of Fiction." *New Literary History* 5(2):245-268.

- Finnern 2010. *Narratologie und Biblische Exegese: Eine integrative Methode der Erzählanalyse und ihr Ertrag am Beispiel von Matthäus*, Tübingen.
- Focant 2003. "Le rôle des personnages secondaires en Marc: L'exemple des récits de guérison et d'exorcisme." In Steffeck and Bourquin, eds. 2003. *Raconter, interpréter, annoncer. Parcours de Nouveau Testament*, Genève. 115-126.
- Goldenberg and Zheng, 2006. "Exporatory study of a new model for evolving networks." *ICML'06 Proceedings of the 2006 conference on Statistical network analysis*. 75-89.
- Graesser, Golding and Long 1991. "Narrative representation and comprehension." In Barr, Kamil and Mosenthal, *Handbook of reading research*. NewYork vol. 2:171-205.
- Hamon 1998. *Le personnel du roman: le système des personnages dans les Rougon-Macquart d'Emile Zola*. Genève.
- Hunt, Tolmie and Zimmermann 2013. *Character Studies in the Fourth Gospel, Narrative Approaches to Seventy Figures in John*. Tübingen.
- Jacquemin 1997. "Guessing morphology from terms and corpora." *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval*. 156-165.
- Keim, Mansmann, Oelke and Ziegler 2008. "Visual Analytics: Combining Automated Discovery with Interactive Visualizations." *Proceedings of the 11th International Conference on Discovery Science* 5254:2-14.
- Kim, Ko, Elmqvist and Ebert 2011. "Word-Bridge: Using Composite Tag Clouds in Node-Link Diagrams for Visualizing Content and Relations in Text Corpora." *Proceedings of the Hawaii International Conference on System Sciences*. 1-8.
- Koschützki, Lehmann, Peeters, Richter, Tenfelde-Podehl and Zlotowski 2005. "Centrality Indices." In Brandes and Erlebach 2005. *Network Analysis*. Berlin. 16-61.
- Lee 2007. "A Computational Model of Text Reuse in Ancient Literary Text." *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*:472-479.
- MacCarron and Kenna 2012. "Universal properties of mythological networks." *EuroPhysics Letters* 99(2):28002
- Moretti 1999. *Atlas of the European Novel, 1800-1900*. London.
- Moretti 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London, New York.
- Mott, Lee and Lester 2006. "Probabilistic goal recognition in interactive narrative environments." *Proceedings of the twenty-first National Conference on Artificial Intelligence*.
- Mueller 2003. "Story understanding through multi-representation model construction." In Hirst and Nirenburg 2003. *Text Meaning: Proceedings of the HLT-NAACL 2003 Workshop*, East Stroudsburg:46-53.
- Newman and Girvan 2004. "Finding and evaluating community structure in networks." *Physical Review E* 69(2):026113
- NBS 2002. *La Nouvelle Bible Segond*. Paris.
- Oelke, Kokkinakis and Keim 2013. "Fingerprint Matrices: Uncovering the dynamics of social networks in prose literature." *Eurographics Conference on Visualization* 32(3):371-380.
- Oelke, Kokkinakis and Malm 2012. "Advanced Visual Analytics Methods for Literature Analysis." *Language Technology for Cultural Heritage, Social Sciences and Humanities*. 35-44.
- Plaisant, Rose, Yu, Auvil, Kirschenbaum, Smith, Clement and Lord 2006. "Exploring erotics in Emily Dickinson's correspondence with text mining and visual interface." *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital librairies*. 141-150.
- Purchase, Hoggan and Görg 2007. "How Important is the 'Mental Map'? – an Empirical Investigation of a Dynamic Graph Layout Algorithm." *Proceedings of the 14th International Conference on Graph Drawing*. 184-195.
- Rimmon-Kenan 1983. *Narrative Fiction: Contemporary Poetics*. London.
- Rochat, Kaplan and Bornet 2013. "A social network analysis of Rousseau's autobiography Les Confessions", *Digital Humanities 2013*, Lincoln.
- Rorhrer, Ebert and Sibert 1998. "The Shape of Shakespeare: Visualizing Text using Implicit Surfaces." *Proceedings of the 1998 IEEE Symposium on Information Visualization*. 121-129.
- Ryan 2007. "Diagramming narrative." *Semiotica* 165:11-40.
- Rydberg-Cox 2011. "Social Networks and the Language of Greek Tragedy." *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1(3).
- Sack 2013. "Character Networks for Narrative Generation: Structural Balance Theory and the Emergence of Proto-Narratives." *2013 Workshop on Computational Models of Narrative* 32:183-197.
- Stein, Wegener and Schlieder 2010. "Pixel-Oriented Visualization of Change in Social Networks." *Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining*. 233-240.
- Stiller and Hudson 2005. "Weak Links and Scene Cliques Within the Small World of Shakespeare." *Journal of Cultural and Evolutionary Psychology* 3(1):57-73.
- Stiller, Nettle and Dunbar 2003. "The Small World of Shakespeare's Plays." *Human Nature* 14(4)397-408.
- Van de Bunt, Van Dujin and Snijders 1999. "Friendship networks through time: An actor-oriented dynamic statistical network model." *Computational and Mathematical Organization Theory* 5 2:167-192.
- Vuillemot, Clement, Plaisant and Kumar 2009. "What's being said near 'Martha'? Exploring name entities in literary text collections." *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*. 107-114.

- Wattenberg and Viégas 2007. "IBM's Many Eyes App After One Month." readwrite.com/2007/03/04/ibm_many_eyes_after_one_month.
- Williams 1996. "Discipleship and Minor Characters in Mark's Gospel." *BS* 153/611:332-343.
- Woloch 2003. *The One Vs. The Many, Minor characters and the space of the protagonist in the novel*. Princeton.

Note: this paper is uploaded very long after it was written. In the meantime, more research on the subject has been published. See the examples below to go further:

- Rochat 2014. *Character Networks and Centrality*, University of Lausanne.
- Rochat 2015. "Character network analysis of Emile Zola's Les Rougon-Macquart", *Digital Humanities 2015*.
- Grandjean 2015. "Network visualization: mapping Shakespeare's tragedies", www.martingrandjean.ch/network-visualization-shakespeare/
- Xanthos, Pante, Rochat and Grandjean 2016. "Visualising the dynamics of character networks", *Digital Humanities 2016*, 417-419.
- Fischer, Göbel, Kampkaspar, Kittel and Trilcke 2017. "Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts", *Digital Humanities 2017*.
- Labatut and Bost 2019. "Extraction and analysis of fictional character networks: a survey", *ACM Computing Surveys*, 52, 5, 89.