

A Distant Reading of Cervantes' *Rinconete y Cortadillo* Using Network Analysis Techniques

Rebecca Beattie, Digital Humanities and Digital Knowledge

Context

This paper uses network analysis techniques to explore speech interactions between characters in Miguel de Cervantes' short story *Rinconete y Cortadillo*. Therefore, the general context of the project is literature, the specific application is literary analysis, and the specific analytic approach is that of distant reading. In contrast to close reading – an analytic approach which focuses on textual detail – distant reading abstracts from the particularity of the text and instead focuses on global and transcendent attributes. Constructing a text as a network is one such way of carrying out distant reading: in doing so, we set aside textual detail, instead choosing to construct and to analyse an atemporal and static model of the text. It should be noted that the technique of distant reading proper ought to be carried out on large quantities of textual data, collected either diachronically or synchronically (e.g. the entire *oeuvre* of a particular writer) (Moretti 2013: 211, 213, 240). Nevertheless, as this paper will presently demonstrate, illuminating insights can also result from distant reading carried out on short, discrete texts, proving it to be a useful tool in the armour of literary criticism.

Problem and Motivation

Critics have noted that *Rinconete y Cortadillo* is heavily dominated by speech and dialogue, with very little narrator intervention (Montero Reguera 300; Paterson 107, 110–1; Saffar 37). Casaldueiro even notes that the narrator is presented “no ya como creador, sino como mero agente de estas figuras que tiene adelante” (“no longer as a creator, but as a mere agent of these beings that stand before him”) (Casaldueiro 61). For example, the story's opening dialogue between the two title characters is only interrupted by the occasional “dijo Rinconete” (“said Rinconete”) or “respondió Cortadillo” (“responded Cortadillo”) on the part of the narrator. Unlike the typical narrator of a prose text, the narrator in this story is more like a stage-manager, and the story itself is more reminiscent of a play than a narrative in prose. Furthermore, critics have also noted that the two title characters themselves, like the narrator, become side-lined as the narrative progresses (more specifically, when they enter the house of Monipodio), turning into observers of the “Monipodio show” and of the events that unfold in his brotherhood. For example, at one point, the characters there theatrically break out into song, comically using a clog, a broom, and some broken plates as musical instruments, and reminding the reader of the *entremés* tradition common in seventeenth-century drama (in which a musical interlude would take place between the acts of a play). This creates a *mise-en-abîme* in which the theatre of Monipodio's house is inserted into the already theatrical dialogue of the two title characters. Analysis of the story's networks of speech interactions will aim to ascertain, quantifiably, the veracity of critics' intuitions regarding the gradual “peripheralisation” of the characters of Rinconete and Cortadillo and the centralisation of the character of Monipodio.

Datasets

This paper follows Moretti’s assertion that “two characters are linked if some words have passed between them: an interaction is a speech act” (2011: 81). It also heeds his warning that, due to the limited set of characters within a literary text, “the mere existence of a connection [between characters] is seldom sufficient to establish a hierarchy, and must be integrated with other measurements” (82). It does this by integrating weight and direction, as also encouraged by Sparavigna (21). Finally, this paper also utilises Ardanuy and Sporleder’s idea of the “dynamic network” (32) which, in contrast to the “static network”, allows us to add a temporal dimension to network representation, by producing independent networks for each part into which the text is divided. In order to examine a *process* of character “peripherilisation”/centralisation within *Rinconete y Cortadillo*, the construction of two networks was necessary: one for the first half of the story featuring the exploits of the title characters in the streets of Seville, and one for the second half of the story covering events after the boys enter the house of Monipodio. The data required to construct these networks was gathered by carrying out a series of XQuery searches on a version of the story marked up in TEI/XML, which included mark-up denoting the speaker and recipient of each speech act. This data was stored in DL files for use and analysis on UCINET.¹

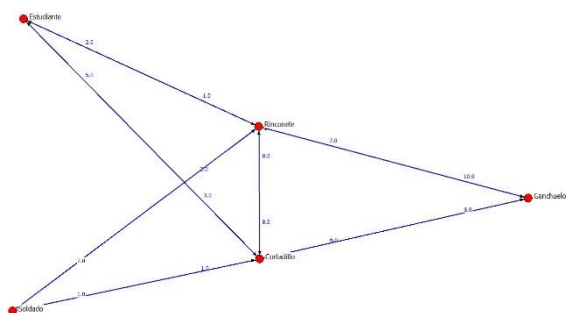


Figure 1: Network of character interactions in *Rinconete y Cortadillo* Part 1

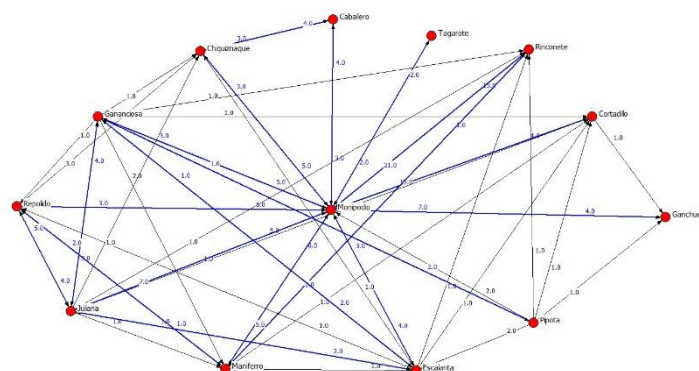


Figure 2: Network of character interactions in *Rinconete y Cortadillo* Part 2

Validity and Reliability

In terms of validity (i.e. how closely the network model represents reality), this paper has built upon Moretti’s assertion that an interaction between characters is a speech act. However, Moretti nuances this approach when he considers prose narratives, in contrast to drama, stating that, in prose writing, “much of what characters say and do is not uttered, but narrated, and direct discourse only covers a part of the plot – at times, a very small part” (2011: 94). In other words, when modelling the network of a prose narrative, building the model upon speech acts is often insufficient because narrated interactions are not incorporated into it (these could include interactions beyond the relational event of verbal communication. For example, relational states, such as co-occurrences, relational roles, or relational cognition). Such a model, therefore, does not represent the reality of character

¹ The TEI/XML file, DL files, and high-resolution image files of the two networks can be found here: <https://github.com/RebeccaJillianBeattie/NetworkAnalysis>.

interactions within the text. However, given the dominance of speech in *Rinconete y Cortadillo*, its theatricality, and the sparsity of narrator intervention, this paper builds upon Moretti's initial premise that a speech act is a proxy for character interaction, and it takes for granted that this is a valid approach *for this particular prose narrative*.

In terms of reliability (i.e. how consistent the network model is), it must be noted that the data which forms the basis of the two networks is not from an objective source: TEI/XML represents an attempt to standardise the mark-up of literary texts but, in any specific application, the choice of tags and attributes is a subjective one. To give a concrete example: during the aforementioned musical interlude, Escalanta, Gananciosa, Monipodio, and Juliana la Cariharta sing to the room at large, but to no character in particular. A choice must be made here about who the intended recipient(s) of these speech acts is/are. It was decided that they are addressed to all characters physically present in the scene, but it is conceded that this may distort the network through commission (for example: Escalanta, Gananciosa, and Juliana la Cariharta do not otherwise address Rinconete and Cortadillo at all). Nevertheless, this choice is a reasonable one and omitting these ties would break the consistency of the mark-up more generally. Although other mark-up choices are possible, the ones upon which this study has been built have been made with internal consistency, and reliability, in mind.²

Measures

The network visualisation of the first half of the story (figure 1) allows us to make some intuitive deductions regarding the structure of character relationships: for example, Rinconete and Cortadillo are clearly the core characters, connected to all others, who are not at all connected to each other. However, the network visualisation of the second half (figure 2) does not lend itself to extensive assessment by the naked eye. Therefore, the application of certain measures is required in order to glean useful information from it, and in order to facilitate comparison between the networks.

a) Analysis at a global level (cohesion, connectedness, compactness, reciprocity, core/periphery)

These measures allow us to gain an overview of the two networks' general characteristics, providing us with an initial, sweeping insight into the structure of character interactions in the two halves of the story. Measures such as cohesion, connectedness, and compactness allow us to assess, in different ways, the extent to which nodes in the networks are connected to each other. Reciprocity tells us the extent to which the edges between nodes are reciprocal (i.e. a character may address another character, but does the latter return the speech act or is the interaction primarily one-way?). And a core/periphery analysis indicates to us variations in the density of interaction between characters (i.e. is there a core group of characters amongst whom interaction is particularly dense? And is there, consequently, a set of characters who are on the periphery of this core?).

² Indeed, this musical scene represents the story's only moment of ambiguity in this respect; the speaker and recipient(s) of all other speech acts are clear.

b) Analysis of central nodes (degree, Eigenvector, closeness, betweenness)

These measures enable us to ascertain which nodes can be considered as “central” in various senses, allowing us to assess who the most dominant characters are in each network. Degree centrality, in a directed network, is a simple measure which tells us the in- and out-degree of each node (i.e. how many speech interactions these characters receive and how many they produce). Eigenvector centrality nuances degree centrality by considering the centrality of the nodes to which a particular node is connected (i.e. a character may have many interactions, but are these interactions with central characters? If so, a greater centrality weighting is awarded to the initial node. In other words, Eigenvector centrality, unlike degree centrality, takes into account the quality of connections, and not just the quantity).³ Closeness and betweenness concern themselves with “physical” centrality: the mean “distance” from a node to all other nodes, and the extent to which a node lies on the path between other nodes as a kind of mediator, respectively.

c) Analysis of node groupings (cliques, kcores, transitivity, triad census, clustering coefficient)

These measures allow us to determine the nature of groupings between nodes within the networks, providing further insight into the centrality or “peripherality” of particular characters. Cliques are sets of nodes in which each member of the set is connected to every other (i.e. each character in a particular set interacts with every other). Kcores represent a more flexible version of cliques, where each member of a set of nodes is connected to *at least* k others. This allows us to conceptualise the network in an onion-like structure in which nodes that lie in the highest kcores are core nodes, and nodes outside these are peripheral. Transitivity, at a whole network level, indicates the extent to which triads of nodes tend to be closed (i.e. if node A is connected to node B, and node B is connected to node C, then node A is also connected to node C). Triad census allows us to analyse further the nature of the networks’ triads, whether closed or not. And the clustering coefficient tells us the likelihood that the nodes to which a particular node is connected are themselves connected (i.e. do the characters with whom a particular character interacts interact with each other?).

d) Analysis of node relations (structural equivalence)

This measure allows us to assess how similar particular nodes are, allowing us to extrapolate whether or not characters can be considered to play similar or different roles within the narrative. Structural equivalence can be applied using different techniques: by counting the number of neighbours two nodes have in common; by measuring the Euclidean distance between nodes (i.e. the distance between their vectors in an n-dimensional space); or by using the Pearson correlation coefficient (i.e. the degree of linear association between nodes), amongst others.

³ When using Eigenvector centrality, careful attention must be paid to trail-end nodes that have in-degree zero (i.e. no incoming edges), and that therefore make no contribution to the centrality of its neighbour nodes. In these cases Katz centrality can be used. No nodes in the two networks which form the basis of this study have in-degree zero, meaning that Eigenvector centrality is sufficient.

Results⁴

a) Analysis at a global level (cohesion, connectedness, compactness, reciprocity, core/periphery)

The cohesion measure of the networks for part one and part two of the story are 0.7 and 0.385, respectively, meaning that it is much more likely for nodes to be connected in the former network than in the latter. However, the connectedness of both networks is a perfect 1: despite the variation in cohesion level, in reality, all characters in both networks are connected. The compactness measure for each network is 0.850 and 0.684, respectively, revealing that, despite the perfect connectivity score of both networks, the connection paths in the first network are slightly shorter. The reciprocity measure for the first network is a perfect 1 (for both arc and dyad reciprocity), meaning that all interactions between characters are symmetric. However, the same measure for the second network results in arc and dyad reciprocity scores of 0.633 and 0.463, respectively, proving the network for the second half of the story to be a much less reciprocal one. Finally, the core/periphery measure revealed, unsurprisingly, that the core characters of the first network are Rinconete and Cortadillo, with all others in the periphery. However, surprisingly, the core characters of the second network were revealed to be Monipodio *and* Rinconete, with all others relegated to the periphery (including Cortadillo).

b) Analysis of central nodes (degree, Eigenvector, closeness, betweenness)

Measuring degree centrality of all characters in the network for the first half of the story shows us that the title characters are perfectly matched in both in- and out-degree (18 and 22, respectively). The other three characters in this half play clearly subordinate roles. Measuring degree centrality in the network for the second half of the story reveals a very different picture: Rinconete and Cortadillo's degree centralities are very different to each other (with in- and out- degrees of 27 and 16 (R), and 18 and 5 (C), respectively). However, Monipodio clearly dominates in both in- and out- degree (56 and 77, respectively). These results are corroborated by the measure of Eigenvector centrality: in the first network Rinconete and Cortadillo have well-balanced Eigenvector centralities of 0.576 and 0.554, respectively. In the second network, Rinconete has an Eigenvector centrality of 0.458, Cortadillo of 0.294, and Monipodio of 0.664 (the highest in the network, by far). The measure of closeness applied to the first network confirms that the title characters are equally "close" to all others; applied to the second network it reveals that Rinconete and Cortadillo are, surprisingly, also similarly "close" to all characters (with in- and out- closeness of 0.667 and 0.522, respectively). The closeness measure reveals Monipodio to be a "hub" of the network with a perfect in-closeness of 1 and a nearly perfect out-closeness of 0.923. The betweenness measure confirms these results: in the first network Rinconete and Cortadillo have normalised betweenness of 25%, while in the second network Monipodio has a normalised betweenness of 61.553% (52.273% above the second in the list). Interestingly, Cortadillo has a higher normalised betweenness score (1.389%) than Rinconete (0.189%).

⁴ Complete results for all measures can be found at: <https://github.com/RebeccaJillianBeattie/NetworkAnalysis>.

c) Analysis of node groupings (cliques, kcores, transitivity, triad census, clustering coefficient)

Analysing the two networks for cliques reveals that the network for the first half of the story contains 3 cliques, each of which contain Rinconete and Cortadillo. The network for the second half of the story contains 8 cliques: Monipodio is a member of all, Rinconete is a member of 2, Cortadillo is a member of 3, and Rinconete and Cortadillo do not share membership of any clique. An examination of the kcores of the first network shows us that all characters are connected to at least 2 others. The same examination of the second network shows us that all characters are connected to at least 5 others, except the very peripheral Ganchuelo, Tagarote, and the Caballero. The transitivity level of the first network is 0.6 and of the second network is 0.597, showing them to be, at a general level, similarly transitive. A triad census analysis of the first network reveals that the most common type of triad is the transitive one, while in the second network the most common type of triad is the one in which 2 characters are connected, leaving a spare character unconnected to this grouping; the second most common type of triad is the one in which 3 characters have zero connections between them; and the third most common type of triad is the one in which 2 characters point towards a common third character, but not to each other. Finally, the clustering coefficient measure tells us that, in the first network, the probability that Rinconete and Cortadillo's neighbours are closed is equal (with a value of 2) and, in the second network, that the probability of Monipodio's neighbours being closed is very low (0.477) (with all other characters having a similarly higher probability). This follows the common network pattern that nodes with the highest degree have the lowest clustering coefficient (i.e. for characters with a high number of interactions, it is less probable that the characters with whom they interact interact with each other).

d) Analysis of node relations (structural equivalence)

Analysing the structural equivalence, using the Pearson correlation coefficient, of nodes in the network of the first half of the story reveals, unsurprisingly, that Rinconete and Cortadillo are most similar to each other (with a score of 0.89). The same measure applied to the network of the second half of the story reveals that Rinconete and Cortadillo are again similar (with a score of 0.94). Monipodio is unlike any other character, except Rinconete, with whom he shares a structural equivalence of 0.79 (in comparison to the very low level of structural equivalence he shares with Cortadillo, of only 0.27, or with any other character).

Conclusion

The results of this study confirm certain aspects of critics' intuition regarding the "peripheralisation" of Rinconete and Cortadillo and the centralisation of Monipodio between the first and second halves of Cervantes' *Rinconete y Cortadillo*. However, the results also reveal that critics' tendency to consider Rinconete and Cortadillo as a single unit has been misguided: between the first and second halves of the story the two title characters become increasingly individualised, and Rinconete plays a *less* peripheral role than Cortadillo.

Firstly, considering the centralisation of Monipodio, the applied measures confirm that, between the first and second halves of the story, the network of character interactions

depletes in reciprocity/symmetry; becomes less coherent (despite consistent connectedness); and is marked by triad clusters of characters which do *not* interact with each other (with 2 interacting characters and a spare, 3 unconnected characters, and 2 characters interacting with another character in common but not with each other, as the three most common triad structures). This evidence confirms a fundamental shift in network structure between the two halves which points towards the insertion of a central, “hub” character who dominates the network’s interactions, as both the major recipient and producer of speech acts. This vision of the network is bolstered by Monipodio’s high closeness, betweenness, degree centrality, and Eigenvector centrality scores, and by his low clustering coefficient score.

Secondly, considering the “peripheralisation” of Rinconete and Cortadillo, alongside the gradual individualisation of these two characters, the applied measures confirm that, in the first half of the story, Rinconete and Cortadillo’s roles are almost perfectly balanced and central, shown by their equal or similar degree and Eigenvector centrality scores, and by their mutual presence in all cliques. However, the measures also confirm that, in the second half of the story, Rinconete and Cortadillo’s roles begin to diverge: despite their (surprisingly) equal or similar closeness and structural equivalence scores, their degree and Eigenvector centrality scores are markedly different, in Rinconete’s favour, and they no longer appear in any cliques in common (for the reason that they do not interact with each other in the second half). It is at this point that critics’ intuitions begin to be proven somewhat erroneous due to their reluctance to consider Rinconete and Cortadillo as two, individualised characters: the measures applied show us that, as expected, Rinconete and Cortadillo compose the core of the network for the first half of the story; however, they also show that the core of the network for the second half of the story is composed by Monipodio *and* Rinconete, but not by Cortadillo, supported by the former two characters’ relatively high level of structural equivalence.

In sum, this paper has validated, quantifiably, certain facets of critics’ intuition regarding the “peripheralisation” of the story’s title characters and the centralisation of Monipodio, but has also highlighted the need for a more nuanced consideration of the text’s protagonists as *individuals*. As Moretti argues, the measures we apply to literary networks are marked by the power-law distribution characteristic of all networks, in contrast to the normal distribution of the Gaussian curve: in other words, there is no character that represents the “average” character. As Moretti puts it, “characters-in-general do not exist” (2011: 85–6). With this in mind, critics ought to pay renewed attention to the *differences* between the story’s title characters, in order to imbue them with an individuality which has hitherto been lacking in the critical discourse surrounding *Rinconete y Cortadillo*.

Critique

This paper has been constructed upon the assumption that a speech act between characters serves as a valid proxy for character interaction. Despite the aforementioned pitfalls of using this approach on prose narratives, it is this paper’s belief that, given the dominance of speech within the story, it is warranted and reasonable. The presumption behind this view – that *Rinconete y Cortadillo* contains a preponderance of speech and dialogue – could be bolstered

by an analysis of Cervantes' other *Novelas ejemplares* (*Exemplary Novels*) in order to confirm, quantifiably, that it is indeed the case that this story is markedly dominated by speech acts, in comparison to Cervantes' other tales.

Furthermore, in order to heed Moretti's warning that networks of prose texts must take character interactions described by the narrator into account, amplifying the text's TEI/XML mark-up to incorporate other types of interactions is a possibility (although such narrator descriptions are admittedly sparse in *Rinconete y Cortadillo*). For example: mark-up could be added to denote the subject and object of other kinds of interactions, such as "looks at", "watches", "touches", "mocks", "lies to", "judges" etc., meaning that the network would incorporate ties representing a variety of relational events and semantic values.

Finally, given the conclusion that this paper reached – regarding the gradual individualisation and unequal "peripheralisation" of *Rinconete* and *Cortadillo* between the two halves of the story – insightful findings may result by combining the two nodes representing these characters into one node and by repeating the analysis. Somewhat paradoxically, considering *Rinconete* and *Cortadillo* as one character could produce results which alter the network dynamics to such an extent that the argument in favour of considering the two characters as individuals is strengthened.

Bibliography

- Ardanuy, Mariona Coll and Sporleder, Caroline. (2014). "Structure-Based Clustering of Novels". *Proceedings of the 3rd Workshop on Computational Linguistics for Literature*, pp. 31–9.
- Casalduero, Joaquín. (1969). *Sentido y forma de las 'Novelas ejemplares'*.
- Montero Reguera, José. (2005). "'Entre parejas anda el juego' / 'All a Matter of Pairs': Reflections on some characters in the *Novelas ejemplares*". *A Companion to Cervantes' "Novelas ejemplares"*, edited by Stephen Boyd.
- Moretti, Franco. (2000). "Conjectures on World Literature". *New Left Review, Volume 1*, pp. 54–68.
- (2011). "Network Theory, Plot Analysis". *New Left Review, Volume 68*, pp. 80–102.
- (2013). *Distant Reading*.
- Paterson, A.K.G. (2005). "Language as Object of Representation in *Rinconete y Cortadillo*". *A Companion to Cervantes' "Novelas ejemplares"*, edited by Stephen Boyd.
- Saffar, Ruth El. (1974). *Novel to Romance: A Study of Cervantes' "Novelas ejemplares"*.
- Sparavigna, Amelia Carolina. (2013). "On Social Networks in Plays and Novels". *International Journal of Sciences, Volume 2.10*, pp. 20–5.