

Character Centrality in Present-Day Dutch Literary Fiction

Roel Smeets¹, Eric Sanders², and Antal van den Bosch³

¹Radboud University, Department of Literary and Cultural Studies

²Radboud University, Centre for Language and Speech Technology

³KNAW Meertens Institute

In the critique of literary representation, the depiction of literary characters has been studied from ideological perspectives. Hierarchies can be exposed by determining the centrality of a character relative to other characters. As an addition to such close reading methods, the present paper proposes an approach to character centrality that combines network analysis with narratology. This explorative study is based on a dataset of demographic metadata on 2,137 characters from a corpus of 170 contemporary Dutch novels. We extract social networks of characters from each novel, and rank all characters according to five centrality metrics. Then, we perform a multiple linear regression to test which of the demographic variables predicts a character's position in the rankings. Our results suggest that immigrant and female characters score higher on two of the five centrality metrics. As a narratological evaluation, we contextualise this observed pattern in relation to a close reading of Özcan Akyol's *Eus* (2012), a novel from the corpus that thematises both descent and gender. We demonstrate that our data-driven and empirically informed approach to character centrality lays bare surprising patterns of representation which only gain relevance in light of close readings of specific cases.

Keywords: Dutch literature, character representation, social network analysis

1 Quantifying the Critique of Literary Representation

Literary studies have a rich tradition of critically analysing hierarchies in literary texts from an ideological perspective. In the wake of the poststructuralist turn (J.Culler (1983)), the critique of literary and cultural representation has evolved into an engaging field with roots in various ideological strands, such as Marxism, postcolonialism and feminism. These so-called 'hermeneutics of suspicion' (Ricoeur (1979), Felski (2009)) have focused on a variety of topics, including the hierarchical representation of story

characters, the inhabitants of fictional story worlds. In Dutch literary studies, this is illustrated by a range of studies that critique the representation of characters of a certain gender, descent or class (e.g. Pattynama (1994), Meijer (1996a), Meijer (1996b), Pattynama (1998), Minnaard (2010), Meijer (2011)). For the sake of the present research, it suffices to say that this field of study is concerned with hierarchies between characters and their identities, although this is of course not their only focus. These ideological approaches to character representation are commonly (and/or implicitly) concerned with how important, influential, dominant or central a character in a narrative is as opposed to other characters. 'Centrality' will be used in this article as an umbrella term to refer to abstract notions such as importance, dominance, influence and power. When a character is central, it means that he or she is important, dominant, influential or powerful in a specific way.

A number of quantitative studies on character representation in Dutch literature have been conducted only recently (van der Deijl et al. (2016), Van der Deijl and Smeets (2018), Koolen (2018): 162-244). These studies make use of new applications of social network analysis and other quantitative methods to reconstruct and analyse story worlds in a data-driven way, following earlier research on non-Dutch texts (e.g. Alberich et al. (2002), Stiller et al. (2003), Elson et al. (2010), Lee and Yeung (2012), Karsdorp et al. (2012), Agarwal et al. (2013), Jayannavar et al. (2015), Karsdorp et al. (2015b), Lee and Wong (2016)). For the study of hierarchies between characters, these methods provide the means for a formalisation and quantification of the concept of 'character centrality'. This is potentially interesting for the study of character representation as practised in the critique of literary representation. Except for these recent examples, studies on character representation mainly use close reading methods. This can lead to powerful interpretations for one or a few cases, but such qualitative readings do not result in general insights into the centrality of characters at a larger scale. However, the importance of a character in a narrative might very well be expressed by the numerical frequency with which he/she features in the narrative. While being fully aware that data-driven approaches are not ideologically neutral either, the present study aims to bridge this gap by considering the centrality of characters in both narratological and statistical terms.

In this contribution, we try to answer the following question: To what extent can a data-driven and empirically informed approach to character centrality contribute to the ideological critique of literary representation? First, we provide a succinct overview of how character centrality has been understood in the narratological tradition of analysing literary texts. Second, we confront these narratological considerations with a computational approach to identifying networks of characters. Third, we describe our data and the method we used to rank all 2,137 characters in a corpus of 170 contemporary Dutch novels on the basis of five statistical metrics. Fourth, we analyse and interpret the results of a multiple regression analysis that tested which demographic feature (gender, descent) was the best predictor for a character's place in the rankings. Based on simple descriptive statistics of the gender and descent distributions among characters in the corpus, we hypothesise that male and non-immigrant characters will score higher than female and immigrant characters. Fifth, in order to evaluate the output of the statistical model narratologically, we confront these findings with a close reading of one novel from the corpus. We conclude with the argument that our approach to character centrality lays bare surprising patterns of representation, although they only start to make sense when contextualised through a qualitative reading of a specific case.

2 Centrality in Narratology

Narratology is one of the traditional methodological toolkits for the study of narratives. This toolkit offers various instruments to analyse the centrality of characters in literature, of which we will mention two of the most straightforward. A character's position in the story world is already predetermined by some basic structural features of a literary text. The mode in which a novel is narrated is commonly a first indicator of how important a character is in the storyline. Some narrative layers in a text are embedded in others, which is particularly relevant for the position of narrating characters in first-person novels. As narrating characters belong to the highest narrative layer, they consequently 'produce text that is not perceived by the characters' that belong to more embedded narrative layers (Van Boven and Dorleijn (2013): 33).¹ As such, 'the narrating instance is located on a higher textual level' and is 'above the world of the characters' who do not have a narrating role (ibid.). In this sense, it is logical to ascribe a more central role to a narrating character in a first-person novel than to the other characters, as the narrator is the one who is in the best position to control the flow of information.

Focalisation is a narratological concept that is also applicable to the centrality of characters. It was coined by the French structuralist Gerard Genette to distinguish between who narrates and who perceives in a text (Genette (1972)). Others have suggested revisions of the concept (e.g. Bal (1977), Nelles (1990), Jahn (1996)); the revision that has become most popular is that of the Dutch scholar Mieke Bal. She defined focalisation as 'the relation between the vision and that which is "seen", perceived' (Bal (2009): 145-146), which made it possible to discern hierarchical relations between characters who occupy active focalising roles and characters who are mainly in a passive position in which they are being focalised by other characters. The extent to which a character features in active focalising roles is thus another indicator of his/her place in the character hierarchy.

3 Centrality in Network Theory

Network theory has been occupied with the question of how to measure the centrality of nodes in a network. Relational, networked structures are interesting in this respect as they can yield insights into the centrality of certain actors as opposed to others. The centrality of a node can be measured in a number of ways to consider different aspects of the network structure. In 1978, the American sociologist Linton Freeman observed that there is 'certainly no unanimity on exactly what centrality is or on its conceptual foundations, and there is very little agreement on the proper procedure for its measurement' (Freeman (1978): 217). He conceptualised three basic centrality measures – degree, betweenness and closeness – which are still being used today, albeit frequently in revised form, and which are thought to 'cover the intuitive range of the concept of centrality' (idem: 237). It is worth mentioning that Freeman's intent was not 'to "lock in" to any sort of ultimate centrality measure' (idem: 217), as centrality is a rather abstract concept and therefore hard to pinpoint statistically. Existing measures as those used by Freeman at best help to clarify what might be understood as central, but they do not necessarily give any definitive answers on which actors are most important in a network.

Before Freeman's innovative proposition, centrality was mainly viewed in terms of

¹ Unless otherwise indicated, all translations are the authors' own.

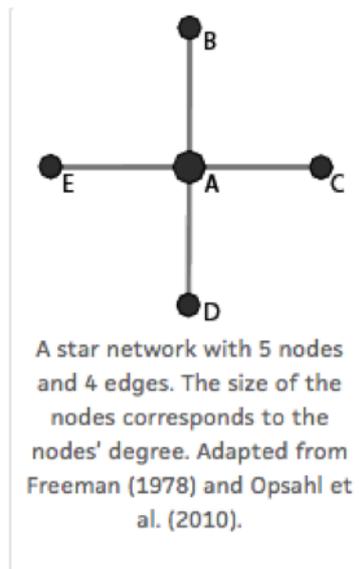


Figure 1: Adapted from Opsahl and Others (2010).

degree, which is the most straightforward measure of centrality. In Figure 1, node A has an advantage over B, C, D and E because it has more relations to others in the network: A has a degree of 4, whereas B, C, D and E each have a degree of 1. The main limitation of *degree centrality*, however, is that it does not take into account the overall structure of the network. A node can be related to many other nodes but located in the periphery of the network, which results in a situation where the node is far removed from the opposite side of the network.

As an alternative to degree centrality, *closeness centrality* is defined as the sum of distances to all other nodes in the network. An advantage of closeness is that it takes into account the relative access that a node has to other nodes in the network. In Figure 1, node A has a higher closeness than B, C, D and E, as it is directly connected with its neighbours, whereas B, C, D, and E need to cross through A to reach a node other than A. The disadvantage of closeness centrality, however, is that it cannot properly be applied to networks that are not fully connected. By definition, nodes in two disconnected components of a network are unable to reach one another, and therefore closeness cannot be computed for the overall structure of a network with disconnected components.²

Freeman was the first to propose *betweenness centrality*, which computes the extent to which a node lies on the shortest path between two other nodes. In Figure 1, node A has a high betweenness centrality because it connects all four nodes with each other. As it is applicable to networks with disconnected components, betweenness has an advantage over closeness. However, as a metric, it is limited because nodes are often not located on the shortest path between two other nodes. Because of that, B, C, D and E, in Figure 1, all have a betweenness centrality of 0.

Network theory makes a distinction between unweighted and weighted graphs. In a weighted graph, the edges represent the intensity with which two nodes are connected. As the basic centrality measures of degree, closeness and betweenness are devised for application to unweighted, binary networks, alternative metrics have been

² In case a network has a lot of disconnected components, a convenient approach would be to only compute closeness centrality for the largest component of the network.

proposed. Degree centrality has been redefined for weighted graphs by not focusing on the number of relations but on the sum of the weights of those relations (Barrat et al. (2004)). Dijkstra's algorithm (Dijkstra (1959)), named after the Dutch computer scientist Edsger W. Dijkstra, has been used to redefine closeness and betweenness centrality by looking at the shortest paths in terms of distances (Newman (2001), Brandes (2001)). As these new proposed metrics target primarily the weights and are less reliant on the number of relations, a second redefinition was needed to take into account both weight and number of relations (Opsahl and Others (2010)).

Every network thus demands a specific approach; there is no general method that applies to every network. The first question should be which elements constitute the network, the second how those elements are related. Then, it should be decided if the network is binary and unweighted, or if the elements are gradually related to one another. The appropriate centrality measures should be derived from the specific nature of the network (weighted/unweighted, unipartite/bipartite³) and the question through which it is approached, as not every centrality measure is relevant in all possible instances.

In the following we will explore to what extent it is useful to make a synthesis between the narratological and the network analytic approach to centrality. We do so by considering literary texts as social networks made up of characters which can be ranked according to the centrality measures described above. The structure of these character networks will be adjusted to the mode of narration and focalisation of the novels. In the context of character representation, narratology thus informs a quantitative and statistical conceptualisation of centrality. Conversely, the study of social networks of fictional characters is informed by some basic narratological insights.

4 Data & Method

In order to test how a data-driven approach to character centrality might contribute to the critique of literary representation, we devised a computational model that takes into account both network theoretical and narratological considerations on mode of narration and focalisation. There is an emerging branch of studies that apply network analysis to fictional populations of characters (e.g. Alberich et al. (2002), Stiller et al. (2003), Elson et al. (2010), Lee and Yeung (2012), Agarwal et al. (2013), Jayannavar et al. (2015), Karsdorp et al. (2015b), Lee and Wong (2016), Moretti (2013):211-240, Rydberg-Cox (2011), MacCarron and Kenna (2012)), though all use different methods for their purposes. One of the main challenges for this type of analysis is the conceptual issue of how to define and automatically identify characters in texts (Dekker et al. (2019)). This can be done manually (e.g. Moretti (2013):211-240) or automatically (e.g. Elson et al. (2010)). Vala et al. (2015) have shown that automatic detection is a difficult task due to the poor performance of existing pronominal and coreference resolution techniques.⁴ Because of this poor performance, we do not aim for full coreference resolution, but instead use a semi-automatic method that departs from a predefined

³ Unipartite networks exclusively consist of elements from the same category, e.g. people connected to people. Bipartite networks consist of elements from different categories, e.g. people connected to organisations. The number of elements in multipartite networks can be extended endlessly in theory, but it is usually restricted to three different categories (tripartite).

⁴ An alternative approach to character detection is automatically classifying animacy in texts (Karsdorp et al. (2015a)).

set of characters. Inspired by narratological theories on what constitutes a character (Herman and Vervaeck (2005): 60-61; Van Boven and Dorleijn (2013): 335), we define characters as *people or creatures which to a greater or lesser extent are presented as human, existing of not more than a few linguistic features including one or more names*. For each novel, a list of names is created with Named Entity Recognition (NER); characters whose name frequency is above a normalised threshold value (based on the number of words of the text) will be regarded as characters.⁵

The other challenge is how to define and automatically identify relational ties between those characters. One of the most used definitions of character relation frames connections between characters in terms of conversations or dialogues (Stiller et al. (2003), Elson et al. (2010), Lee and Yeung (2012), Moretti (2013): 211-240, Jayannavar et al. (2015), Lee and Wong (2016)). The quantifiable unit of the conversation is, however, not the best indication for character interactions, as there are plenty of characters that do not enter into a conversation but are related to one another in some other way. For instance, two characters with family ties might never speak to each other, but such a relation should definitely be regarded as a character relation. Another way to define relational ties is in terms of co-occurrence in the same window of *N* words, sentences, paragraphs or chapters (Alberich et al. (2002), Grayson et al. (2016)). Defining character relations in terms of adjacency in the text will be able to capture more instances of character interaction than when it is defined in conversational terms. This is the most bottom-up definition of character relations, as characters do not have to communicate in a literal sense (as is the case in conversation networks) to be considered as having some form of interaction.

Based on these considerations, we operationalise the strength of character relations through *co-occurrences of character name variants in a window of N tokens*. We experimented with different window units and sizes for different types of novels to find the 'sweet spot' where not too many and not too few character interactions are detected (cf. Grayson et al. (2016)). However, such a sweet spot is different for every novel. In order to be able to compare the novels, we decided to use the same window unit and window size for every novel. As sentences are the smallest linguistic structures which are semantically meaningful in themselves (cf. Mann and Thompson (1988)), we used sentences as the window unit, which we tokenized using Ucto.⁶ The window size was set to two sentences, as semantic relations are known to extend over two sentences through connectives (cf. Bluhdorn (2010)).⁷ We devised a customised co-occurrence approach for each narrative mode, which we describe in detail below.

We have used a sample corpus of 170 contemporary Dutch novels, consisting of all submissions to the 2013 Libris Literatuur Prijs, one of the most prestigious literary prizes in the Dutch language area. This prize is awarded to novels published in the year before, in this case in 2012. In that year, 1,397 Dutch novels were published; our sample of 170 novels thus makes up 12.2 percent of the total number of novels

⁵ There are several NER-tools available, but not all are suitable for the same task. NER-tools have to be trained for specific languages, and their accuracy depends on the nature of the training data (e.g. a tool trained on newspaper articles performs badly on literary fiction). For the current research the Namespace-tagger is used, which is trained on Dutch literary fiction and which is demonstrated to be the most accurate for the present purposes, although it is still not perfect as a F1-score of 0.72 was reported (Smeets (2017)).

⁶ <https://github.com/proycon/python-ucto>, last accessed: 3-7-2018.

⁷ As the plain texts of the novels in our corpus are unstructured, we could not rule out the possibility that characters co-occur in two sentence windows that transcend the boundaries of a paragraph or chapter. We are aware that this creates noise, as in those cases it could be argued that there is no meaningful interaction between characters.

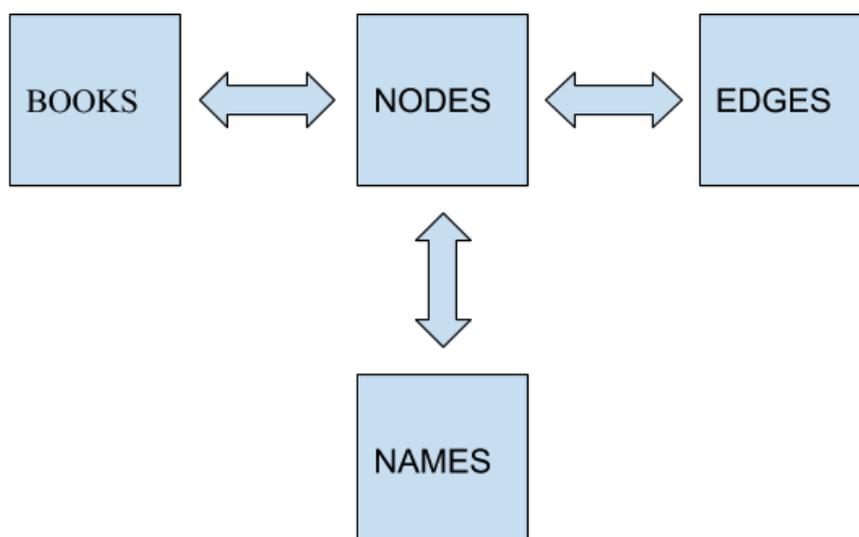


Figure 2: Visualisation of database linkage

published in that year.⁸ With a few exceptions, these novels belong to the genre of literary fiction.⁹

In earlier research (van der Deijl et al. (2016)), the following demographic information for 1,176 characters in the corpus was manually gathered, if known: gender, age, country of descent, city of descent, country of residence, city of residence, profession. In more recent research (Volker and Smeets (2019)), the number of characters was increased to a total of 2,137 characters and semi-automatically enriched with 4,459 relational ties between each character. These relational ties were grouped in one of the following five categories: family, friend, lover, colleague, enemy.¹⁰ These character metadata were stored in four interrelated database tables (see Figure 2).

Lists of all variants of a character's name were automatically generated with named-entity recognition and were stored in a table called NAMES. BOOKS contains all relevant metadata of the novels, such as title, the name, gender and age of the author, the publisher and the filename of the digital version of the novel. NODES contains all relevant metadata of the characters, such as name, gender, country of descent, city of descent, country of residency, city of residency, education and profession. EDGES contains all relevant metadata on the character relations, such as the specific nature of the relation (friend, family, enemy, lover, colleague). All tables are linked to one another through a unique book id. NAMES, NODES and EDGES are also connected through a character id. The character networks are computed through an Object-Oriented model written in the Python programming language, consisting of three main classes: Book,

⁸ This number is based on all Dutch language novels published in 2012 with NUR-code 301 (literary fiction), that is, 1,780 in total. These include 383 duplicates or reissues, which were subtracted from the total number. Thus, the total number of 'original' Dutch literary fiction published in 2012 is 1,397.

⁹ For a list of all 170 novels, see: <http://www.librisliteratuurprijs.nl/2013-groslijst>.

¹⁰ We used top-down relational labels assigned to characters by two expert annotators. These annotators based their annotations on rather narrow definitions: e.g. the label 'enemy' was only assigned when the relation is clearly hostile, the label 'friend' when the relation is clearly friendly. Differences in annotations between the annotators were resolved through discussion. The annotators also accounted for changing relations between characters. In those cases double labels were assigned, such as Colleague_Enemy. Double labels were also assigned when the nature of the relation changed over time, such as friends becoming enemies.

Character and Network.¹¹

Each book in the corpus has a unique id from 1 to 170. Every character in the corpus has a unique character id that corresponds to a book id stored in database BOOKS. For instance, *De lichtekooi van Loven* by Ineke van der Aa is represented by the book id 1. In database NODES, character 'Louise' is represented by this same book id followed by character id 1 and her name (1_1_Louise). In database NAMES, this same unique identifier is followed by every name variant of the character. The name variants for this character are 'Louise', 'Louisje' and 'Louiseke', which is represented in NAMES as 1_1_Louise_Louise, 1_1_Louise_Louisje and 1_1_Louise_Louiseke. Each novel's text was then searched for each of these name variants, after which these variants were replaced by the unique character identifier.¹² As such, the locations of each character in the text were automatically identified.

The corpus was divided into three sub-corpora based on their mode of narration:

third-person, first-person, multi-perspective. Third-person novels are narrated by an anonymous narrator who follows one main character. First-person novels are narrated by an I-narrator. Multi-perspective novels are narrated by multiple narrators, either in third or first person. For every subcorpus a slightly different co-occurrence approach was used based on the specific mode of narration. For all novels, irrespective of their mode of narration, relations between characters were pre-established when they were annotated with one of the relational labels stored in EDGES (friend, family, lover, enemy, colleague). In all cases, the procedure below was used to establish the weight of the relations.¹³

Third-person novels (63 novels) For every character in the novel, a sliding window approach was used in which co-occurrences of two characters were mapped in a window of two sentences. Whenever two characters occur in the range of the same two sentences, a relation between those characters was established. The more often such co-occurrence takes place, the stronger their relation becomes.

First-person novels (73 novels) *a)* As the first-person narrator has by definition high centrality in narratological terms, the relations of the first-person narrator with all other characters were simply defined by counting every occurrence in the novel of characters other than the first-person narrator. As every character is embedded in the narration of the first-person narrator, it can be argued that every character occurrence represents a relational tie with the first-person narrator. The more often a character occurs in the novel, the stronger its relation with the first-person narrator.

b) For every character other than the first-person narrator, a sliding window approach was used in which co-occurrences of two characters were mapped in a window of two sentences. Whenever two characters occur in the range of the same two sentences, a relation between those characters was established.

¹¹ All software and data are accessible through the following open access GitHub repository: <https://github.com/roelsmeets/character-networks>. This repository does not contain the corpora because of copyright issues.

¹² A similar approach is used by Grayson et al. (2016): 4, who replace character aliases with a character's name.

¹³ In some cases, two characters have a relational label such as "family" assigned to them while the weight of their relation is 0. This is possible as characters do not have to be adjacent in the text to have a family tie, just as people in real-world networks can be family without being in each other's physical presence or without talking about each other.

The more often such a co-occurrence takes place, the stronger their relation becomes.

Note that this approach will in most cases rightfully lead to relatively strong relations between the first-person narrator and all other characters, whereas this is not the case for the relations between and among all other characters.

Multi-perspective novels (34 novels) For each of these novels, student assistants annotated where a character perspective begins and ends in the text. These annotations also contain information on the narrative mode and focalisation: a first-person or third-person narration was annotated as such, and for third-person narration the main focaliser was annotated. On the basis of those annotations, each novel was divided into separate sections. For sections narrated in first or third person, the first- or third-person method was applied. After that, the co-occurrence counts between characters were aggregated for all the separate sections.

All these relations are symmetrical, and thus undirected. This means that the character relations are not regarded in terms of directionality, which is a logical consequence of the co-occurrence approach, as adjacency is a priori a symmetrical issue. Furthermore, the resulting network, with characters as nodes and character relations as edges, will both be undirected and weighted. Not every relation between any two characters will have the same status, as the strength of a relation is increased when two characters co-occur more often in the novel.

With Python's software package *networkx*,¹⁴ the resulting networks for each individual novel were used to rank the characters on the basis of five centrality metrics. Among those metrics are the above described degree, betweenness and closeness centrality, as well as eigenvector and Katz centrality, two metrics on which Google's PageRank algorithm is based. PageRank is used by Google's search engine to rank web pages by relevance. PageRank, eigenvector and Katz are all based on the same, seemingly circular assumption that a node in a network becomes more important when it is connected to other important nodes (Page et al. (1998)). The computation of all these metrics was based on the *weighted* edges. Then, a regression analysis was carried out to see which of the demographic variables (gender, descent) is the best predictor for a character's place in the rankings.

5 Results regression analysis

Because of the exploratory nature of the present research and the absence of prior research on this topic, we did not have any formal hypothesis about which demographic factors would possibly determine a character's place in the rankings. However, we preferred to not just enter all possible variables into the regression equation as this would have possibly obscured the results of the analysis. Therefore, we formulated a non-formal hypothesis based on traditional, non-statistical research in the critique of literary representation. Several studies suggest that female characters and/or characters of mainly non-Western descent are often represented in a stereotypical manner and are therefore likely to be staged in less central, more marginal positions in literary texts (e.g. Pattynama (1994), Meijer (1996a), Meijer (1996b), Pattynama (1998), Minnaard (2010), Meijer (2011)). Gender and descent might therefore be possible

¹⁴ <https://networkx.github.io/>, last accessed 7-5-2018.

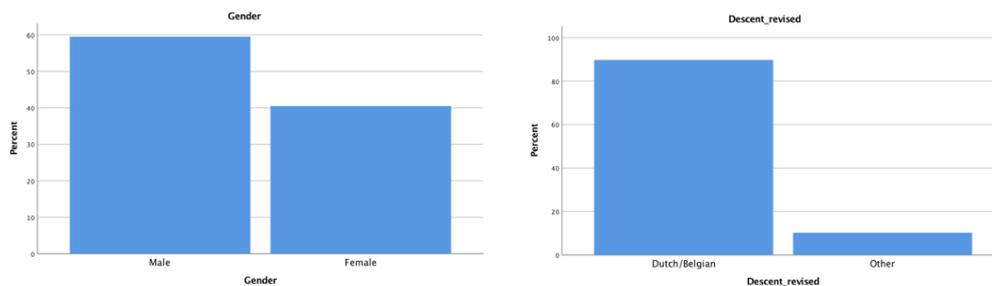


Figure 3: Gender and descent distributions among characters in the corpus ($N = 2137$). The numbers are based on types, not on tokens.

predictors of a character's position in the rankings. Based on descriptive statistics of our data, we suspected that male and non-immigrant characters would end up as more central,¹⁵ as these types of characters are simply more present in the dataset (see fig. 3).¹⁶ More precisely, we hypothesise that both male characters and non-immigrant characters will score higher on the centrality metrics than female characters and characters with a migrant background.

For each of the five centrality metrics (degree, betweenness, closeness, eigenvector, Katz), a multiple linear regression was conducted to predict characters' centrality scores based on their gender and descent. Gender is coded as 0 for male and 1 for female. Descent was coded as 0 for non-immigrant and 1 for immigrant. As our aim was to generalise across all novels, we did not include the division into sub-corpora (third-person, first-person, multi-perspective) in the statistical model.

No significant results were found for betweenness, closeness and eigenvector centrality. Gender and ethnicity are thus no predictors for characters' scores on betweenness, closeness and eigenvector centrality.

However, significant results were found for degree and Katz centrality. First, for degree centrality, a significant regression equation was found ($F(2, 2128) = 6.424, p < 0.01$), with an R^2 of 0.006. Characters' predicted degree centrality is equal to a B value of $0.428 + 0.024 (\text{GENDER}) + 0.059 (\text{DESCENT})$ (see fig. 4). This means that, on degree centrality (on a scale from 0 to 1), female characters scored 0.024 higher than male characters, and immigrant characters scored 0.059 higher than non-immigrant characters.

Secondly, for Katz centrality a significant regression equation was found ($F(2, 2128) = 6.124, p < 0.01$), with an R^2 of 0.006. Characters' predicted Katz centrality is equal to a B value of $0.272 + 0.009 (\text{DESCENT}) + 0.007 (\text{GENDER})$ (see fig. 5). This means that,

¹⁵ Referring to characters who either have or do not have a migrant background, the terms 'immigrant' and 'non-immigrant' are used in a loose sense. In this article, immigrant characters can also refer to characters who are born in the Netherlands or Belgium but whose parents migrated to the Netherlands or Belgium. In this broad definition, immigrant characters are considered to have some sort of bond with a socio-cultural tradition that is not the same as their current country of residence. We chose the Netherlands and Belgium as a point of departure as the books in the corpus are either written by Dutch or Flemish authors who operate in a shared literary field of Dutch literature.

¹⁶ A chi-square goodness of fit test was calculated comparing the occurrence of male and female characters with the hypothesized occurrence of a 50-50 gender distribution. Significant deviation from the hypothesized values was found ($\chi^2(1) = 82,030, p < .001$). Also, a chi-square goodness of fit test was calculated comparing the occurrence of characters with a Dutch/Belgian and Other country descent with the hypothesized occurrence of an equal distribution among those categories. Significant deviation from the hypothesized values was found ($\chi^2(1) = 1350,773, p < .001$). This means that the 40-60 gender divide and the 89,8-10,2 divide in descent are not due to chance, but is a statistically significant difference.

Table 1: Linear model of predictors of degree centrality, with unstandardized coefficients (columns 2 and 3) and standardized coefficients (column 4)

Model	Unstandardized		Standardized	Sig.
	B	Std. Error	Beta	
1 (Constant)	0.438	0.006		0.000
Descent_revised	0.058	0.019	0.065	0.003
2 (Constant)	0.428	0.008		0.000
Descent_revised	0.059	0.019	0.066	0.002
Gender	0.024	0.012	0.043	0.048

Table 2: Linear model of predictors of Katz centrality, with unstandardized coefficients (columns 2 and 3) and standardized coefficients (column 4)

Model	Unstandardized		Standardized	Sig.
	B	Std. Error	Beta	
1 (Constant)	0.273	0.002		0.000
Gender	0.007	0.003	0.061	0.005
2 (Constant)	0.272	0.002		0.000
Gender	0.007	0.003	0.062	0.004
Descent_revised	0.009	0.004	0.045	0.038

on Katz centrality (on a scale from 0 to 1), immigrant characters scored 0.009 higher than non-immigrant characters, and female characters scored 0.007 higher than male characters.

These findings suggest that our initial hypothesis, based on traditional critiques of literary representation, should be rejected. Contrary to what we expected, female characters and immigrant characters scored higher, at least on two of the five centrality metrics used in the analysis. Furthermore, it should be noted that a higher frequency distribution of a character type does not necessarily lead to a more central position in a character network, as the results of the regression analysis has shown. Although male and non-immigrant characters are more present in the corpus, they do not end up as more central in network analytic terms. The question remains as to how these results can be interpreted in a close reading context.

6 Narratological evaluation

The quantitative representational patterns suggested by the outcome of the multiple linear regression require a narratological evaluation, as it is unclear what their significance is for the critique of literary representation. In concrete terms, the outcome for degree centrality is that female and immigrant characters have significantly more relations than male and non-immigrant characters. More specifically, women and characters with a migrant background often co-occur with a wider range of fellow characters in the novels. The higher scores of female and immigrant characters on Katz centrality indicate that they often co-occur with characters who also have relatively

high Katz centrality. In sum, female and immigrant characters have both *more* relations in general and more relations with *important* characters.

In order to make sense of this pattern, we subsequently conducted a small narratological exploration of character centrality in one novel from the corpus and confronted it with the results of the statistical analysis. As a case study, we used a novel that thematises both gender and ethnicity, as these issues of representation were taken as points of departure for the regression analysis. For the sake of brevity, we only used the two concepts of narrative mode and focalisation as points of departure. Note that there is a wide variety of other narratological concepts and perspectives that might potentially lead to alternative insights.

Eus (2012) by Özcan Akyol is a semi-autobiographical, first-person novel, in which the reader follows the life of the first-person narrator Eus, the son of Turkish immigrants living in Deventer, a small city in the Netherlands. Eus gets involved in criminal activities and ends up in jail, where he starts a writing career. This plotline foregrounds the theme of upward social mobility: a character with a migrant background who initially has a hard time finding his way in Dutch society eventually finds his creative ambition and becomes a successful author.

This theme is underscored by Eus's foregrounding of the economic and social hierarchies that exist between Dutch people and people with a migrant background. At the beginning of the novel, Eus states that he and his friends '... didn't dare to go to the better neighbourhoods', although they 'knew that they existed' (Akyol (2012): 24). An implicit opposition is thus postulated between 'better neighbourhoods' and Eus's own rough neighbourhood. Later on, Eus is more explicit when he characterises the 'indigenous youth, rich kids' as 'white scum' (idem: 120).

Another less prevalent, but latently present theme in the novel is the way men with a migrant background engage with (Dutch) women. Throughout the novel, women are treated with little respect by Eus and his friends. Female characters are either objects of sexual desire or considered a man's possession. They are repeatedly referred to as 'whores' (idem: 36, 58, 85, 145) and 'sluts', or variants of the term (idem: 43, 57, 62, 86, 145, 157, 176, 253). The male characters seem mostly interested in whether or not a woman is 'fuckable' (163). In general, women are constantly objectified. The following quote is a good example:

Sometimes I stared out of the window for hours, in search of the hottest girls in school, about whom I then started fantasising. How beautiful they were! Nice tits! Nice ass! (idem: 50)

On the basis of such thematic cues, one could argue that two hierarchical oppositions take shape in the narrative:

1. immigrants \longleftrightarrow non-immigrants
2. male \longleftrightarrow female

Although there are some indications in the quotes given above, it is not evident what the specific hierarchy of these relations might be. The two basic concepts of narrative mode and focalisation might help to clarify this. First of all, the novel is narrated by Eus, which means that he controls the flow of information in the narrative. It is a logical consequence of the I-narration that Eus decides which events to report and which to leave out. When he, e.g., reports that 'I was born and raised in Koekstad, a small town by the IJssel, exactly on the border of two eastern provinces' (idem: 13),

he chooses to use an alias ('Koekstad') for a town the reader might know as Deventer. As an I-narrator, Eus is able to manipulate the narrative at will.

Furthermore, he is also the main focaliser: the narrative events are filtered through his perceptions. This means that the description of events is not neutral but colored by the vision and judgement of Eus. The following two sentences are illustrative:

The next two years I went to the Hegius school, where I was surrounded by the beautiful, posh girls who pursued the highest level of education. Rumour had it that these girls had an above-average interest in foreign boys because they never saw those types of boys (idem: 37)

A whole group of girls is lumped together by Eus and characterised as 'beautiful' and 'posh'. Eus thus foregrounds their physical appearance and highlights their poshness, thereby suggesting that these girls are spoiled rich kids. In the second sentence he repeats a rumour relating to their supposed sexual interest in boys with a migrant background. These two sentences show an extremely biased representation of a specific type of characters (in this case: female, highly educated).

These basic narratological observations are of major importance for the interpretation of character hierarchies in the novel. As first-person narrator and main focaliser Eus is both a character with a migrant background and male, so the non-Dutch and male perspectives are a priori more dominant than the Dutch and female perspectives. Taking into account that Eus's friends and fellow criminals (Kosta, Ata, Meltem, Mahir) are also predominantly of a migrant background and male, one could argue that the centre of gravity lies with non-Dutch and male characters.

But although this interpretation of character centrality is based on plain narratological insights, it relies heavily on qualitative evidence (i.e. the few quotes we used to illustrate our point). A more data-driven and quantitative approach might potentially shed a different light on the question of character centrality in this novel. How does our network analytic approach relate to this narratological approach? Table 3 shows the characters in the novel ranked by their scores on degree centrality.

First of all, this novel conforms to the general pattern as observed in the regression equation only with regards to descent. 12 of the 21 identified characters have a migrant background, and they are higher in the rankings than the Dutch characters, which is in line with the general pattern according to which characters with a migrant background have significantly higher degree centrality.

However, with regards to gender, the novel deviates from the pattern. 14 of the 21 identified characters are male, and they occupy higher positions in the rankings on degree centrality, indicating that the male characters in *Eus* have *more* relations than female characters.

For Katz centrality, a similar pattern emerges. Table 4 lists the characters in the novel ranked by their scores on Katz centrality. Here, too, both characters with a migrant background and male characters occupy higher positions in the rankings than non-immigrant and female characters. The first types of characters are thus connected to more important characters than the latter.

Interestingly, these rankings are in line with the findings of our narratological approach to character centrality in the novel. Our narratological argument that the perspectives of male characters with a migrant background are dominant in terms of narrative mode and focalisation is backed up by our quantitative argument that these types of characters have higher scores on degree and Katz centrality. In this specific case, our narratological argument that the male and immigrant perspectives are dominant does not conflict with the character rankings for this particular novel.

Table 3: Characters in Eus (2012) ranked by degree centrality score

	Name	Gender	Descent	Degree
1	Kosta	male	immigrant	0.65
2	Kareltje	male	non-immigrant	0.55
3	Eus	male	immigrant	0.50
4	Turis	male	immigrant	0.40
5	Meltem	male	immigrant	0.40
6	Ata	male	immigrant	0.40
7	Mahir	male	immigrant	0.35
8	Selma	female	immigrant	0.30
9	Metin	male	immigrant	0.30
10	Haakneus	female	immigrant	0.30
11	Levine	female	non-immigrant	0.30
12	Theo	male	non-immigrant	0.15
13	Nathan	male	non-immigrant	0.15
14	Eef	female	non-immigrant	0.15
15	Inez	female	non-immigrant	0.1
16	Ömer	male	immigrant	0.1
17	Angelo	male	non-immigrant	0.1
18	Vinny	male	non-immigrant	0.1
19	Osman	male	immigrant	0.05
20	Daphne	female	non-immigrant	0.05
21	moeder Eus	female	immigrant	0.00

Table 4: Characters in Eus (2012) ranked by Katz centrality score

	Name	Gender	Descent	Katz
1	Kosta	male	immigrant	0.218218982823223
2	Mahir	male	immigrant	0.21821840342212764
3	Eus	male	immigrant	0.2182184034212662
4	Kareltje	male	non-immigrant	0.2182182875401764
5	Turis	male	immigrant	0.21821822960063142
6	Ata	male	immigrant	0.218218113719957
7	Meltem	male	immigrant	0.2182179398985992
8	Selma	female	immigrant	0.21821782401818632
9	Haakneus	female	immigrant	0.218217824018094
10	Levine	female	non-immigrant	0.21821782401787862
11	Metin	male	immigrant	0.2182178240177709
12	Theo	male	non-immigrant	0.2182177660783798
13	Nathan	male	non-immigrant	0.21821770813789643
14	Eef	female	non-immigrant	0.21821765019709
15	Inez	female	non-immigrant	0.21821759225711432
16	Angelo	male	non-immigrant	0.218217592257022
17	Vinny	male	non-immigrant	0.218217592257022
18	Ömer	male	immigrant	0.21821759225692972
19	Daphne	female	non-immigrant	0.21821753431670787
20	Osman	male	immigrant	0.21821753431661559
21	moeder Eus	female	immigrant	0.21821747637639374

In sum, our narratological evaluation highlights two important points with regards to the interpretability of our regression analysis. 1) A narratologically oriented analysis of a case can provide a qualitative contextualisation of a statistical argument. The mode of narration and focalisation in *Eus* illustrate the dominance of the male, immigrant perspective, which is supported by the characters rankings for the novel. A narratological argument may also give nuance to or conflict with a statistical argument, but that is not the case for this specific novel. 2) A specific novel (in this case, *Eus*) can very well deviate from an observed general statistical pattern. Characters with a migrant background score higher than Dutch characters in the novel in terms of network centrality, which is in line with our regression model. But contrary to the general pattern, the male characters score higher than the female characters. This highlights the importance of a qualitative contextualisation of the statistical analysis.

7 Conclusion

In this contribution, we have demonstrated that a data-driven and empirically informed approach to character centrality informs the ideological critique of literary representation in at least three ways.

First, instead of focusing on a limited number of cases, general patterns can be discerned in a larger corpus that represent a specific literary-historical period, which can lead to new, surprising insights. Contrary to what is suggested by the wide range of ideologically oriented close readings of character representation in literature, our results suggest that female and immigrant characters take up a more central position in the social networks of present-day Dutch literary fiction than non-immigrant and male characters, statistically speaking. This remarkable outcome requires an explanation, particularly in light of the highly imbalanced frequency distribution of immigrant and non-immigrant characters in the corpus. Almost 90% of the characters in the corpus are non-immigrants, but our regression model suggests that immigrants are more central in the networks than non-immigrants. Possibly, these higher centrality scores might be explained by the probability that novels that thematise descent, and stage a higher number of immigrants, also ascribe more central roles to them. Overall, the novels have fewer immigrant characters (only around 10% of all characters have an immigrant background), but these immigrants score higher on degree and Katz centrality. Something similar holds for female characters: there are fewer female characters than male characters in the corpus (almost a 40-60 ratio), but they have relatively high centrality values. In order for immigrant or female characters to be central in network theoretical terms, a high frequency of occurrence is not a necessary prerequisite as long as they interact with a high number of other (central) characters. This is a possible explanation of the discrepancy between the descriptive statistics and the outcome of the regression equation.

Second, combining a narratological close reading with network analysis enables a formalisation of abstract terms as importance, influence or power that are typically used in a strict metaphorical sense. Our case study on *Eus* illustrates that ideologically oriented interpretations regarding gender or ethnicity can be either backed up or nuanced by network statistics which make such interpretations less susceptible to unarticulated and implicit presuppositions.

Third, this study demonstrates that quantitative statistical patterns only make sense when confronted and contextualised with close readings of specific cases. Statistical trends might indicate general patterns of literary representation, but they can only

serve as an analytic backdrop for the individual analysis of particular novels. The novel *Eus*, for instance, lives up to the pattern only with regards to descent but not with regards to gender. The extent to which a single novel either conforms to or deviates from the general pattern can then be used to determine the particularity of a certain aspect of representation.

Our main contribution to the field of Digital Literary Studies lies in bringing together the methodological toolkits of narratology and social network analysis, which often seems to be lacking in data-driven approaches to networks in literary texts. As opposed to other network extraction methods, we propose a method that departs from domain knowledge and combines this with a bottom-up operationalisation of character interactions. As such, this article is situated within the mixed-method framework of the text-oriented Digital Humanities, and provides an argument for more strongly connecting qualitative and quantitative strands of research in the field.

8 Acknowledgements

The authors would like to thank annotators Maartje Weenink and Lisa Rooijackers for their precise annotations and Lucas van der Deijl and the anonymous reviewers for commenting on earlier versions of this article.

References

- Agarwal et al.
2013. Automatic extraction of social networks from literary text: A case study on alice in wonderland. *IJCNLP*.
- Akyol, O.
2012. *Eus*. Prometheus.
- Alberich et al.
2002. Marvel universe looks almost like a real social network. *arXiv*.
- Bal, M.
1977. *Narratologie: Essais sur la signification narrative dans quatre romans modernes*. Klincksieck.
- Bal, M.
2009. *Narratology. Introduction to the Theory of Narrative*. Toronto/Buffalo/London.
- Barrat, A., M. Barthelemy, R. Pastor-Satorras, and A. Vespignani
2004. The architecture of complex weighted networks. *Proceedings of the national academy of sciences*, 101(11):3747–3752.
- Bluhdorn, H.
2010. *40 Jahre Partikelforschung*, chapter A semantic typology of sentence connectives. Stauffenburg Linguistik.
- Brandes, U.
2001. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177.

- Dekker, N., T. Kuhn, and M. Van Erp
2019. Evaluating named entity recognition tools for extracting social networks from novels. *PeerJ Computer Science*.
- Dijkstra, E. W.
1959. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271.
- Elson et al.
2010. Extracting social networks from literary fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Pp. 138–147.
- Felski, R.
2009. After suspicion. *Profession*, Pp. 28–35.
- Freeman, L. C.
1978. Centrality in social networks: Conceptual clarification. *Social Networks*.
- Genette, G.
1972. *Narrative Discourse. An Essay in Method*. Oxford.
- Grayson, S., K. Wade, G. Meaney, and D. Greene
2016. The sense and sensibility of different sliding windows in constructing co-occurrence networks from literature. *International Workshop on Computational History and Data-Driven Humanities*.
- Herman, L. and B. Vervaeck
2005. *Vertelduivels. Handboek verhaalanalyse*. Vantilt.
- Jahn, M.
1996. Windows of focalization: Deconstructing and reconstructing a narratological concept. *Style*.
- Jayannavar et al.
2015. Validating literary theories using automatic social network extraction. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*.
- J.Culler
1983. *On Deconstruction. Theory and Criticism after Structuralism*. Routledge.
- Karsdorp et al.
2012. Casting a spell: Identification and ranking of actors in folktales. In *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH-2)*.
- Karsdorp et al.
2015a. Animacy detection in stories. In *6th Workshop on Computational Models of Narrative*.
- Karsdorp et al.
2015b. The love equation: Computational modeling of romantic relationships in french classical drama. In *Proceedings of the Sixth International Workshop on Computational Models of Narrative*.

- Koolen, C.
2018. *Reading Beyond the Female. The Relationship Between Perception of Author Gender and Literary Quality*. ILLC Dissertation Series.
- Lee and Wong
2016. Hierarchy of characters in the chinese buddhist canon. In *Proceedings of the Twenty-Ninth International Florida Artificial Intelligence Research Society Conference*.
- Lee and Yeung
2012. Extracting networks of people and places from literary texts. In *Proceedings of the 26th Pacific Asia Conference on Language, Information and Computation*.
- MacCarron, P. and R. Kenna
2012. Universal properties of mythological networks. *EPL*.
- Mann, W. and S. Thompson
1988. Rhetorical structure theory: toward a functional theory of text organization. *Text: Interdisciplinary Journal for the Study of Discourse*.
- Meijer, M.
1996a. *Het omstreden slachtoffer: geweld van vrouwen en mannen*, chapter De verschrikkelijke sneeuwman: projectie, geweld en nieuwe mannelijkheid in het werk van Jan Wolkers. Ambo.
- Meijer, M.
1996b. *In tekst gevat. Inleiding tot de kritiek van de representatie*. Amsterdam University Press.
- Meijer, M.
2011. *Diversiteit*, chapter Zwartheid in de witte verbeelding. Printart Press Kft.
- Minnaard, L.
2010. The spectacle of an intercultural love affair. exoticism in van deyssel's blank en geel. *Journal of Dutch Literature*.
- Moretti, F.
2013. *Distant reading*. Verso.
- Nelles, W.
1990. Getting focalization into focus. *Poetics Today*.
- Newman, M. E.
2001. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E*, 64(1):016132.
- Opsahl and Others
2010. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245 – 251.
- Page et al.
1998. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

- Pattynama, P.
1994. Oorden en woorden. over rassenvermenging, interetniciteit, en een indisch meisje. *Tijdschrift voor Vrouwenstudies*.
- Pattynama, P.
1998. *Domesticating the Empire. Race, Gender, and Family Life in French and Dutch Colonialism*, chapter Secrets and Danger: Interracial Sexuality in Louis Couperus's The Hidden Force and Dutch Colonial Culture around 1900. University Press of Virginia.
- Ricoeur, P.
1979. *Freud and Philosophy: An Essay on Interpretation*. Yale University Press.
- Rydberg-Cox, J.
2011. Social networks and the language of greek tragedy. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*.
- Smeets, R.
2017. Finding characters: An evaluation of named entity recognition tools for dutch literary fiction. In *European Alliance for Digital Humanities (EADH) day*.
- Stiller et al.
2003. The small world of shakespeare's plays. *Hum Nat*.
- Vala et al.
2015. Mr. bennet, his coachman, and the archbishop walk into a bar but only one of them gets recognized: On the difficulty of detecting characters in literary texts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*.
- Van Boven, E. and G. Dorleijn
2013. *Literair mechaniek*. Coutinho.
- van der Deijl, L., S. Pieterse, M. Prinse, and R. Smeets
2016. Mapping the demographic landscape of characters in recent dutch prose: A quantitative approach to literary representation. *Journal of Dutch Literature*, 7(1):20–42.
- Van der Deijl, L. and R. Smeets
2018. Tussen close en distant. personage-hiërarchieën in peter buwalda's bonita avenue. *Tijdschrift voor Nederlandse Taal -en Letterkunde*, 134(2):123–145.
- Volker, B. and R. Smeets
2019. Mirrors or alternative worlds? comparing ego networks of characters in contemporary dutch literature with the population in the netherlands. *Poetics (In press)*.