Modeling Conflict: Representations of Social Groups in Present-Day Dutch Literature

Roel Smeets, Maarten De Pourcq, Antal van den Bosch

Roel Smeets, Maarten De Pourcq, Radboud University Nijmegen
Antal van den Bosch, University of Amsterdam
Peer-Reviewer: Rachel Buurma, Hoyt Long
Data Repository: 10.7910/DVN/JSKPQV

ABSTRACT

This essay responds to a lack of scholarly attention for conflict as a narrative mechanism since the formalist models of Vladimir Propp and Algirdas Julien Greimas. Building on recent developments within cultural analytics, the essay argues for a new understanding of narrative conflict by integrating classic narratological models with data-driven, statistical methods. It does so by (a) proposing two computational models of conflict based on theoretical insights from narratology, conflict studies, and network theory, (b) applying those models to a sample corpus of 170 present-day Dutch novels, and (c) briefly illustrating the narratological value of the results by interpreting the representation of social groups in two novels from the corpus – Bart Koubaa's *De Brooklynclub* (2012) and Leon de Winter's *VSV* (2012) – in light of the statistical patterns found for the corpus as a whole. The analyses of dyadic (two characters) and triadic (three characters) conflict leads to two central conclusions: 1) lower educated characters are more dominant in dyadic conflicts and 2) the majority of triadic conflicts exist in a state of social balance.

'Can you name a single great work of art which is not about conflict?'
Yuval Noah Harari, *Sapiens. A Brief History of Humankind* (2011: 184)

Antagonism in narrative fiction sets the story in motion. Conflict as a narrative mechanism has been famously studied by formalists such as Vladimir Propp in the 1920s and Algirdas Julien Greimas in the 1960s, both of whom devised models of narrative action in which conflicts have a vital function.[1] Although these classic narratological models have made the idea that conflict is a driving force behind narrative action common knowledge,[2] an in-depth conceptualization and a replicable operationalization of conflict in narrative fiction is lacking from contemporary literary scholarship.[3] In this contribution, we aim to fill this gap while building forth

on Propp and Greimas. In order to do so, we bring together typologies, concepts, terms, and tools from a wide variety of research traditions such as conflict studies, social psychology, theatre studies, narratology, and network theory. Our research is situated within the broader methodological framework of computational literary criticism, distant reading, and cultural analytics, and in particular within the epistemological shift from measurement to modeling as recently proposed by Ted Underwood.[4] We present two models of conflict and use them to test two hypotheses on the role of conflict in literary texts, the first focusing on dyadic conflict (between two characters) and the second on triadic conflict (between three characters). To narrow down the scope of our models, we focus on the social status that is often, but not always, an important dimension of conflict. We demonstrate how these models help to gain insight into the representation of social groups in literary texts. We work with a heavily annotated corpus of 170 present-day Dutch novels for which extensive metadata on all 2,137 identified characters was semi-automatically collected in earlier research,[5] but the models can be applied to texts written in any language and from any period.[6]

What is conflict?

We turn to the field of peace and conflict studies to pinpoint some key features of the concept. For a first working definition, we draw upon the work by Johan Galtung, who defines conflict in strong association with violence:

> Whenever there is violence there is an unresolved conflict. Unresolved conflict means that there is an incompatibility of goals, including means, that has not been resolved, superseded, transformed, or transcended. That conflict can be direct, between actors who have conscious goals, or structural, between parties that have their interests. In other words, if you don't like violence solve the conflict.[7]

According to this definition, violence is a marker of conflict; there cannot be violence without conflict. It makes sense, therefore, to integrate the notion of violence into our conceptualization of narrative conflict. Furthermore, this association with violence highlights the hierarchical nature of conflict. From an ideological point of view, conflicts between characters generally indicate

representational hierarchies through which a form of physical, verbal or ideological violence is expressed. Such hierarchies between (groups of) characters, then, determine the dynamics and outcome of conflicts: arguably, there is a tendency for more powerful actors or parties to be the dominant party in the conflicts. Often, this is exemplified by a dominance of one representational category over another – e.g. male over female, Western over 'exotic', higher over lower class.

A second characteristic of conflict is its fundamental relational mechanism; it takes (at least) two to conflict. In the field of social psychology, a distinction is made between interpersonal and intergroup conflict.[8] For *interpersonal* conflicts, theories have primarily focused on the nature of e.g. frustration or aggression (ibidem). The realistic group conflict theory focuses on *intergroup* conflict:

> Opposed group interests in obtaining scarce resources promote competition, and positively interdependent (superordinate) goals facilitate cooperation. Conflicting interests develop through competition into overt social conflict. It appears, too, that intergroup competition enhances intragroup morale, cohesiveness, and cooperation [...] Thus, the real conflict of group interests not only create antagonistic intergroup relations but also heightens identification with, and positive attachment to, the in-group.[9]

The 'scarce resources' can be anything from material sources such as food and money, to immaterial sources such as prestige and respectability. In works of narrative fiction, a competition for scarce resources can be represented explicitly through e.g. the opposing armies in *Game of Thrones* trying to reign over a geographical area, or implicitly through e.g. the different ethnic groups in Özcan Akyol's *Eus* (2012) struggling for respect in their social environment.

People in intergroup conflict will behave as a function of their group as it 'heightens identification with, and positive attachment to, the in-group'. A sense of *intra*-group cohesion may arise when social groups are in conflict with each other. If this holds true for narrative fiction, then characters can be expected to act according to their social identity as they 'will not interact *as* individuals, on the basis of their individual characteristics or interpersonal relationships, but as members of their groups standing in certain defined relationships to members of other groups'.[10]

Conflict thus has the potential to co-shape social identity. Hypothesizing that conflict has a similar function in narrative fiction, we examine the ways in which conflicts are co-constitutive of the literary representation of (groups) of characters. To ensure that our models are applicable (and generalizable) to a broad range of narrative fiction, we combine basic narratological insights with the formal tools of social network theory. Social network analysis provides some practical instruments to formalize and analyze conflicts between characters on a larger scale.

Conflict in narratology

For a theoretical understanding of conflict in narrative texts, a starting point is provided by a variety of interrelated lemma's in the seminal analytic dictionary of semiotics by Greimas and Courtés.[11] As in most handbooks and encyclopedias of literary analysis, the lemma 'Conflict' is absent, but the lemmas 'Confrontation, 'Polemical' and 'Constraint' cover elements that are directly related to it.

What stands out in the description of these lemmata are multiple references to the 'subject' and its relation to the 'anti-subject'. Confrontation is said to occur 'when the goal of the narrative programme [of the subject] is contrary to the goal of the narrative programme of the anti-subject'.[12] This clash of two narrative programmes can result in three situations: 1) a domination of the subject or anti-subject over the other, 2) an exchange between the subject and the anti-subject, or 3) a (mutual) contract between the subject and the anti-subject. In case the clash is of a polemical nature, the narrative typically contains 'the figure of the [...] opponent as a metonymic manifestation of the anti-subject'.[13] Characters who oppose one another can be a manifestation of subject and anti-subject in a direct sense, but the subject/anti-subject opposition is broader as two clashing narrative programmes in a novel might also take shape in the form of two opposing political ideologies, e.g. the confrontation in Orwell's *1984* between the ruling totalitarian regime and the more liberal conviction of its protagonist. Furthermore, clashing narrative programmes can also occur within a single character. In the slipstream of experiments by modernist writers, inner conflicts have arguably become one of the characteristic features of modern literature.[14]

What stands out most obviously in Greimas and Courtés's lemmata is that the subject and the anti-subject are performed by two or more characters who are engaged in a hostile relation. It is important to stress that such a relation is hierarchical in case the conflict is resolved by a domination of the subject over the anti-subject, as the anti-subject is then subjugated to the narrative programme of the subject. This can be characterized as a situation of semiotic constraint, which is defined as 'a range of voluntary or involuntary, conscious or unconscious, obligations which the individual takes up through its involvement in a semiotic practice'.[15] It is comparable to the subject or the anti-subject accepting certain 'rules of play' (ibidem). In a concrete sense, this can be thought of as a character – either willingly or unwillingly – accepting the norms and values of another character. In Orwell's *1984*, this is illustrated by Winston Smith's inescapable submission to Big Brother's totalitarian ideology: the protagonist has no choice but accepting his rules of play. Constraint, in all its manifestations, indicates a hierarchical opposition expressed through conflict.

For more practical applications of the concept of conflict, theatre studies offer some points of departure. In his *Dictionnaire du théâtre* (2004), leading theatre scholar Patrice Pavis provides a typology of five different forms of dramatic conflict:

1. Rivalry between two characters because of money, love, morality, politics, etc.
2. Two conflicting worldviews or irreconcilable moral conceptions.
3. (Inner) conflict between e.g. passion and reason.
4. Conflict between the particular and the general, e.g. between the individual and society.
5. A moral or metaphysical conflict between a character and a principle such as God or an ideal. (66)

This typology ranges from the most concrete, direct type of conflict (type 1) to the most abstract type (type 5). Each type defines two elements that are in conflict: two characters (type 1), two worldviews (type 2), two conflicting feelings of thoughts (type 3), the particular and the general (4), a character and a principle (type 5). In the analyses that will follow, this typology will be used to specify the type of conflict at stake. As there is a wide variety of different types of conflict, it will prove useful

in the analyses to make explicit what type of conflict we are talking about. This is not only relevant for means of clarification, but also to keep track of the shifts between types of conflict that can take place within a narrative, such as an inner conflict of a single character leading to a rivalry between two characters.

As mentioned in the introduction to this article, conflict has an important function in two closely related classic narratological models, first described in the 1920s by Vladimir Propp in *Morphology of the folktale* (1928), and further developed in the 1960s by A.J. Greimas in *Sémantique structurale* (1966). Propp was the first to analyze narrative structures using a 'morphological' method referring to the analysis of all constituent elements that comprise a narrative. His method is based on the assumption that 'it is possible to make an examination of the forms of the tale which will be as exact as the morphology of organic formations'.[16] Although his model is based on a distinct collection of 100 Russian fairy-tales and therefore is not necessarily generalizable to fictional narratives in general, Propp's conviction that 'the labyrinth of the tale's multiformity' can be reduced to 'an amazing uniformity' (ibidem) still seems rather universalistic.

From a contemporary point of view, such a formalistic, universalistic approach seems outdated and of little use for analyzing the diversity and complexity of present-day (Dutch) novels. Propp's model is clearly modeled after the action that takes place on the level of plot. This works probably better for fairy tales than for novels, as the first are generally more plot-oriented, whereas style is typically a more central asset of the latter. However, Propp's first axiom – 'Functions of characters serve as stable, constant elements in a tale, independent of how and by whom they are fulfilled'[17] – is relevant for the conceptualization of conflict in novels. The building blocks of novels are typically characters who perform certain functions, although these characters are obviously not always 'stable, constant elements'. More precisely, conflict situations often take the form of a function in which the villain and the hero are confronted with one another.

Propp also claimed that fairy tales revolve around seven general character types, 'dramatis personae'[18] which served as an inspiration for Greimas, who turned it into a more general abstraction commonly known as the actantial model (Greimas; see Figure 1). Just as in Propp's morphology, this model focuses primarily on action

taking place at the level of plot. It can be used to reduce every narrative action to a set of six components: the subject, the object, the helper, the opponent, the sender, and the receiver. Each of these actants revolve around three axes. The subject and object are juxtaposed at the axis of desire. This relationship between subject and object is called a junction, and can take the form of a conjunction when the subject and object are brought together, and the form of a disjunction when the subject is being freed of the object. The axis of power is where the helper and the opponent are positioned. At the axis of knowledge, also known as the axis of transmission, resides the sender who asks for the junction between subject and object, and the receiver who profits from this junction.
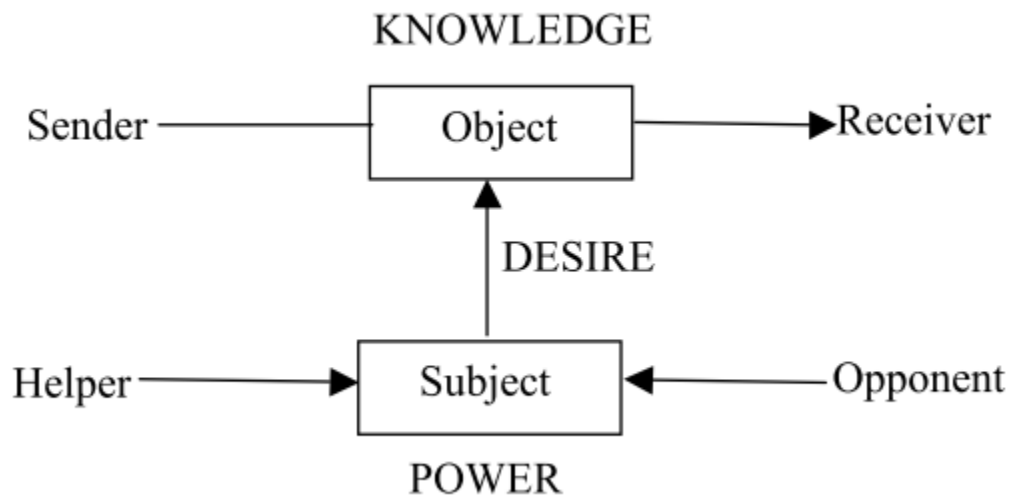

Figure 1. Greimas's actantial model. Adopted from Greimas, A.J. *Sémantique structurale*. Paris: Presse universitaires de France, 1966.

Point of view is a crucial aspect in establishing an actantial model for narrative actions. Greimas underwrites that there is no final, definite model, but that a range of different models can be devised for the same sequence of actions as different points of view are taken into account.[19] Furthermore, multiple roles can be played by the same character.

The axis of power is of particular relevance for this article, as conflicts center around this axis. For that reason, it will be reframed here as the axis of conflict, although

the original name also makes sense as conflicts typically denote hierarchical power relations in which a form of semiotic constraint is present. We reframe the helper and opponent roles as respectively friends and enemies of the subject. More generally, the helpers and opponents might be conceived as being each other's enemies, as they have conflicting goals, i.e. either helping or opposing the subject. An inner conflict can arise when a character is either subject and opponent at the same time (e.g. a drug addict (subject/opponent) wishing to be clean (object) but keeps taking drugs), or helper and opponent at the same time (e.g. the character who is a sobriety coach and a drug dealer at the same time). Finally, it is noteworthy that actants do not necessarily have to be characters, as values, principles, belief systems, and ideologies can also take up the role of an actant.

Conflict in network theory

In network theory, complex structures are composed of objects called 'nodes' and the connections between those objects called 'edges'. Georg Simmel's formalistic relational sociology has inspired the application of network theory to social structures. Conflicts in social networks are studied from a variety of different angles. Most obviously, conflicts are expressed through negative edges between nodes in a network.[20] As in every network, relations between characters in a novel can have either positive or negative valence. Positively connoted edges are represented by e.g. friendship relations, whereas negatively connoted edges are represented by e.g. enemy relations. In general, each edge exists on a negative-positive spectrum; the positivity/negativity of the edge is always relative to the overall social dynamics in the network. Furthermore, it should be noted that the nature of relations can change over time, as friends can become enemies and vice versa.

A network in which both positive and negative edges are ascribed to social relations is called a signed network.[21] Unsigned networks pay no attention to features of signed relations such as like/dislike, respect/disrespect, and love/hate. Signed networks are often associated with (structural or social) balance theory, of which social psychologist Fritz Heider laid the foundation in the 1940s. In a seminal essay he sets out the basis of the theory:

Attitudes towards persons and causal unit formations influence each other. An attitude towards an event can alter the attitude towards the person who caused the event, and, if the attitudes towards a person and an event are similar, the event is easily ascribed to the person. A balanced configuration exists if the attitudes towards the parts of a causal unit are similar.[22]

This densely formulated theory is best explained by the common expression 'The enemy of my enemy is my friend'. In Figure 2, the rationale behind this maxim is visualized in what became known as Heider's model of social balance. Imagine that a person P is in a hostile relation with a person O. There is a state of social balance, as both have the same negative attitude towards each other. Then, a third person X enters the scene, who also happens to dislike O. In order to maintain social balance, P and X should become friends, based on the fact that they share a negative attitude towards O. Heider's theory asserts that there is social *balance* whenever a triadic relationship consists of either two negative relationships and one positive relationship as in this case, or when all relations between P, O and X are positive.
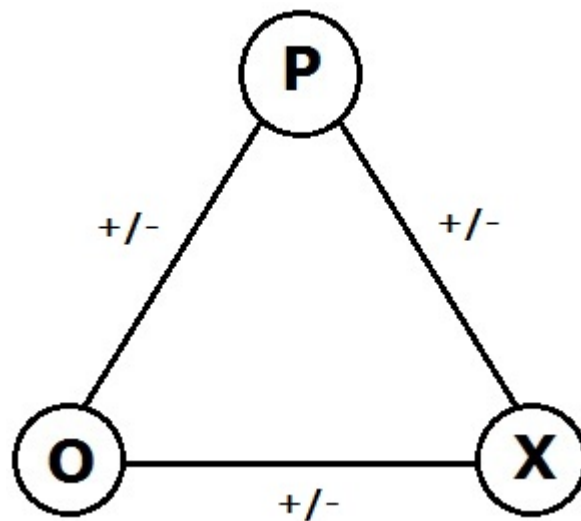

Figure 2. Heider's model of social balance. Adopted from Khanafiah & Situngkir 2.

Conversely, a state of social *imbalance* arises when P likes O and X, but finds out that O and X dislike each other. There is social imbalance as there are now two positive relations and one negative relation in this triad. This imbalance is resolved

if P either starts to dislike X just as O does, or when X and O become friends. Social imbalance occurs when there are two positive relations and one negative relation in a triad, or when all relations are negative.

Following Heider, balance theory has served as an inspiration for research on signed networks. It has been formalized and further refined in the 1950s and 1960s.[23] However, although balance theory sparks one's imagination and has proved to be a fruitful point of departure for formal network methodologies, it does not necessarily hold empirically. The simple premise that signed networks strive to balanced structures is simply not always true, as there are numerous signed networks containing imbalanced triads in the real world.[24] For the present research, it remains to be seen to what extent the balance theory applies to the fictional worlds of characters in novels, which will be studied in section 6. To consider narrative conflict in terms of social balance enables us to go beyond the classic protagonist-antagonist or hero-villain duality. While our first model departs from this duality, our second model transitions to a triadic conception of narrative conflict by testing Heider's balance theory.

Corpus and dataset

Both the dyadic and the triadic models of conflict developed in this article were based on a heavily annotated dataset of a representative corpus of 170 present-day Dutch books in the genre of literary fiction. This corpus consists of all 170 submissions to one year (2013) of the Libris Literatuur Prijs,[25] one of the most prestigious literary prizes in the Dutch language area.[26] The year of 2013 is a randomly chosen sample year of the contemporary production of Dutch literary fiction. As the Libris prize targets novels in the previous year, all novels in the corpus were published in 2012. Constituting a fair amount of the production of literary fiction in 2012, the sample is representative of the Dutch literary fiction within that year. According to the database of the Koninklijke Bibliotheek, the national library of the Netherlands,[27] 1,475 books with NUR-code 301 (literary fiction) were published in 2012. Subtracting duplicates, reissues, and exclusively online publications from this list results in a total amount of originally published Dutch literary fiction of 460. The sample of 170 novels thus represents 37% of the books of Dutch literary fiction published in this year.

The major benefit of the current sample size is that it enables manual annotation of character features. Up to the present, it is only possible to automatically detect the gender of characters by using programming pipelines such as BookNLP,[28] but this is done on the basis of predictions and is not as accurate as manual annotation. Demographic features such as country of descent, level of education, age, and professional occupation are probably never fully automatically detectable, as they often remain implicit in the text and thus require a fair amount of interpretation. Furthermore, a sample size of 170 also enables a more efficient qualitative evaluation of algorithmic and statistical analyses than would be the case with a larger sample size. It requires less effort to check for, say, 10% of the 170 novels if the output of the algorithms match up human intuitions than it would be to do this for 10% or even 1% of 7,500 novels.

The analyses in the following sections are based on an extensive collection of metadata – gender, country and place of descent, country and place of residence, age, level of education, profession or daily activity – on 2,137 characters, as identified by human annotators. The data collection was carried out in several phases in a period roughly between 2014 and 2018. In 2014, a group of annotators contributed to the first phase of data collection, resulting in metadata on 1,176 characters.[29] As the guidelines for annotation were not as precisely defined in the first iteration, the annotation process was repeated twice with more clearly defined guidelines. Most importantly, a formal threshold based on the relative occurrence of character names in the text was introduced to be able to distinguish which entities in a text should or should not be adopted in the database as characters. In the period of 2017-2018, two student assistants incremented the amount of characters to a total of 2,137 characters and complemented some of the missing metadata in the earlier version of the database. Furthermore, these research assistants added relational information between characters (if known) – friend, enemy, colleague, lover, family – to the database.[30] Disagreements between annotators were resolved through discussion.

Quite obviously, a deliberate reductionism is inherent to this process of data collection. First, not all demographic features are deducible from the narrative worlds presented in the texts. Contrary to most sociological survey data collections,

a considerable amount of features is unknown for the 2,137 identified characters in the corpus.[31] Second, literary texts have the possibility to (de)construct various definitions of, for instance, gender, as an author can play with notions of femininity and masculinity. In most cases, the annotations of the demographic categories cannot account for such artistic strategies. Third, the annotations are inevitably binary: this either/or logic does obviously not cover the variety of identities within one demographic category.

Model I: conflict between two characters

At the axis of conflict, hierarchies in character representation take shape. According to Galtung, violence is the result of unresolved conflict, and violent practices commonly lead to a domination of one of the involved parties over the other (see section 2). Our first model operationalizes hierarchies between characters by automatically determining for every pair of characters annotated as enemies who the more powerful party in the conflict is. Then, we test statistically whether one of three identity categories – gender, descent, level of education, – is a predictor of a character's dominance or subordination in the conflict. Are, for instance, characters *without* a migration background or male characters statistically more dominant in conflict situations than characters *with* a migration background or female characters? In order to assess the narratological value of the resulting statistical pattern, it will be confronted with how one novel from the corpus represents social conflict.

The type of conflict under consideration in this section is exclusively focused on two characters and thus relates to the first type offered by Pavis's typology (see section 3): Rivalry between two characters because of money, love, morality, politics, etc.

Conflict scores

A first challenge is to define under which specific conditions characters are in conflict. As these conditions may vary endlessly in nature and intensity, it was decided to not settle this in a data-driven way, but to rather use the top-down relational labels assigned to characters by expert annotators, which are stored in a database.[32] These annotators adopted the following annotation strategy. From the five relational labels – friend, enemy, lover, family, colleague –, enemy is the only

label that explicitly points at a hostile relation between characters. Sometimes friends, lovers, family and colleagues are in conflict with one another; 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. The labels friend and enemy are the only mutually exclusive labels. Two characters can be enemies and colleagues, lovers or family, but they cannot be enemies and friends at the same time. In case a double label friend_enemy or enemy_friend was assigned, the order of the labels reflect the change in those relations, e.g. the relational label 'friend_enemy' denotes that the relation was initially friendly, but later became hostile.

How to establish which party is the most powerful one in the conflict? In order to tackle this problem, the results from earlier research were taken into account. In this study a method was developed to extract character networks based on their co-occurrence in the same window of two sentences.[33] The resulting 170 networks were then used to compute character centrality: all 2,137 characters in the corpus were ranked on the basis of five centrality metrics, each of them indicating a specific form of centrality. As demonstrated there, a character's position in the rankings is a sign of its importance in the narrative, and possibly of the power it exerts over other characters. In this line of thinking, the higher-ranked character in a conflict can be perceived as the dominant party in that conflict. Framed in terms of Greimas's actantial model (see section 3), characters with higher centrality scores than their enemies arguably have better chances of fulfilling their goals than their enemies.

For every two characters annotated as enemies, it was automatically established who of them has a higher degree, betweenness, closeness, eigenvector and katz centrality. Each of these metrics represent a specific form of centrality or importance of the characters as nodes in their fictional social networks: degree centrality is based on the amount of relations a character has, betweenness and closeness on the network position of the characters, and eigenvector and katz on the strength and amount of relations a character has with other central characters.[34] The 'conflict score' of a character is incremented by one in case that character has a higher score on one of these centrality metrics. Some characters have more conflicts than others (i.e. have more enemies), and the likelihood of a higher conflict score therefore potentially increases for characters with multiple enemies. An example: in the novel *Heldhaftig*

by Britta Bolt, there are nine characters who show enmity to others characters, but not every character has the same amount of enemies. The character named Najib has six enemies, whereas the character named Posthumus has only two enemies. This means that Najib's conflict score for each of the centrality metrics can be 6 at a maximum, as there are potentially 6 points he can 'earn'. Conversely, Posthumus's maximum conflict score for each centrality metric is only 2. Figure 3 shows the computed conflict scores for both Najib and Posthumus.[35]

	Conflict score degree	Conflict score betweenness	Conflict score closeness	Conflict score eigenvector	Conflict score katz	Composite conflict score (mean)
Najib	5	6	4	5	5	5.0
Posthumus	2	1	1	2	2	1.6

Figure 3. Example of computation of conflict scores for characters Najib and Posthumus from the novel Heldhaftig *by Britta Bolt*

Najib's conflict score for betweenness centrality is 6, which means that he was ranked higher than all his six enemies on this particular centrality metric. For closeness centrality, Najib's conflict score is 4, which means that two of his enemies were ranked higher than him on this metric. Finally, all five conflict scores are collapsed in a mean composite conflict score. Based on this composite score, Najib (score: 5.0) is more central than Posthumus (score: 1.6) in terms of conflict. Whereas each individual conflict score represents the extent to which characters are more important in terms of a specific form of centrality (degree, betweenness, closeness, eigenvector and katz), this composite conflict score more generally represents the extent to which characters are the dominant party in one-on-one conflicts.

Multiple linear regression

Are gender, descent and/or education predictors of characters' conflict scores? If characters with a certain demographic profile have higher conflict scores than other types of characters, that would indicate a more powerful position in conflict situations of e.g. the male as opposed to the female, the Dutch as opposed to the non-Dutch, or the higher educated as opposed to the lower educated. As there is no previous statistical research on this topic, there are no strong reasons to formulate

hypotheses about which identity categories will likely have an effect on a character's position in conflict situations. Nevertheless, qualitative close reading-based studies provide a vantage point for a non-formal hypothesis.

Ideological approaches to (Dutch) literature have repeatedly suggested that male, Dutch and higher class characters have favorable positions in representational hierarchies due to mechanisms of othering and stereotyping.[36] Against the background of this strand of research, we hypothesize that these types of characters have better chances to end up high in the scores. This non-formal, cultural-critical hypothesis fits in a general scheme of binary oppositions of which the first known example in Western culture is the Pythagorean Table of Opposites referenced in Aristotle's *Metaphysics A*. The table contains ten opposites, among which male-female, of which Aristotle prefers the left part over the right, such as male over female. Our current hypothesis takes its cue from the observation that Western culture generally and implicitly prioritizes one part of those binary oppositions.[37]

A multiple linear regression was conducted to predict the composite conflict scores of characters based on their gender, descent and education.[38] Gender is coded as 0 for male and 1 for female. Descent is coded as 0 for characters *without* a migration background and 1 for characters *with* a migration background.[39] Education is coded as 0 for higher educated and 1 for lower educated. With the composite conflict score as a dependent variable, this resulted in a regression model in which only education was featured as an effective independent variable. Gender and descent were excluded as independent variables by the model as they do not produce significant effects. A significant regression equation was found ($F(1, 363) = 7.362, p < 0.01$), with an R^2 of 0.020. Characters' predicted conflict score is equal to a B value of 0.933+ 0.405 (Education) (see Figure 4). This means that lower educated characters scored 0.405 higher than higher educated characters on their composite conflict scores. Education is thus a predictor of characters' conflict scores. Characters who are lower educated have significantly higher conflict scores than characters who are higher educated.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,933	,083		11,246	,000
	education	,405	,149	,141	2,713	,007

Figure 4. Linear model with the composite conflict score as dependent variable. Gender, descent and education were entered as independent variables. Only education generated statistically significant effects.

What is the relevance of this observed pattern in terms of the representation of social groups? If education is considered as a proxy for a character's socio-economic class (without suggesting that education and class always coincide), then a possible, Marxist-oriented interpretation is that conflict situations are a place where characters from lower classes effectively rebel against their fixed socio-economic position. In the fictional worlds in which they are depicted, lower class characters have less socioeconomic status, but more central roles are ascribed to them in situations where they are in conflict with characters who are higher up the socioeconomic ladder. In this line of thinking, the dominance of the lower classes over the higher classes can be framed as a protest of the underdog.

An alternative interpretation of the statistical pattern, and one that is diametrically opposed to the former, is that lower educated characters have limited social and intellectual skills increasing the likelihood of them using violence quicker than higher educated characters. This is in line with research suggesting that people who are more educated tend to be less involved in violent and/or criminal activities.[40] The inclination of lower educated characters towards violence could subsequently lead to a more dominant, and thus central, position in the network. In this interpretation, the lower educated characters conform to a rather stereotypical image of the lower social classes. As opposed to the former interpretation, these characters do not resist their fixed socio-economic position but rather reinforce the rigidity of that position.

It is only at the level of the individual text that one of these interpretations might demonstrate their relevance. In order to illustrate the narratological value of these statistical results and their interpretations, we briefly demonstrate below how dyadic conflict between characters from different social classes in one novel from the corpus can be read in light of our findings.

Dyadic conflict in Bart Koubaa's **De Brooklynclub**[41]

Cult-novel *Fight Club* (1996) by American author Chuck Palahniuk clearly served as an inspiration for Bart Koubaa's *De Brooklynclub* (2012). Similar to Palahniuk's novel, Koubaa's story centers around a secret club where people from all layers of society get together to fight. Co-founder of this so-called 'Brooklynclub' is the novel's nameless first person narrator (hereafter: the protagonist) who is locked up in prison and unfolds the story of the events that led to his imprisonment. Next to the protagonist, the most prominent characters are the protagonist's (former) girlfriend Lauretta and real-estate giant Mayer. A central event in the story is when Mayer rapes Lauretta in front of all the people present at the Brooklynclub at that moment, after which Lauretta then paradoxically leaves the protagonist to marry Mayer.

Violent conflict is the central tenet of the novel, and is symbolized through the metaphor of the fight club. More specifically, the novel centers around a conflict between the protagonist and Mayer, which can be seen as a metonymic expression of a clash between the higher and lower socio-economic classes. Being employed in environments such as an abattoir, a restaurant, a bar, and as an elevator operator, respectively, the protagonist is undoubtedly part of the lower socio-economic classes of society. As a successful businessman who is one of the most powerful persons in the world of real estate (the young Donald Trump comes to mind), Mayer is undoubtedly upper class. Furthermore, 'Mayer' is a speaking name: alluding to the word 'mayor', it emphasizes his influential societal status.

The conflict between the protagonist and Mayer symbolizes the class struggle of those who occupy a lower place on the social ladder. The protagonist represents the underdog who is dominated by his societal superior Mayer. The raping of Lauretta by Mayer serves as an illustration of that fact: Mayer is so powerful that he can take

away the most precious 'possession' of the protagonist. By murdering Mayer, the protagonist not only takes revenge on his enemy, but also breaks the power hierarchies that are a result of his and Mayer's socio-economic positions.

This reading of the novel fits in the overall statistical pattern in which lower class characters are more central in conflict situations than higher class characters. In this particular novel, the dyadic conflict between the lower class protagonist and the higher class Mayer is indeed reflected by their conflict scores: the protagonist has a composite conflict score of 2.75, Mayer has a composite conflict score of 1.5. At the level of plot, the discrepancy in these scores are exemplified by the protagonist's murdering of Mayer, which is the ultimate victory over one's enemy. In line with our Marxist-oriented interpretation of this finding, the dyadic conflict between Mayer and the protagonist can thus be interpreted as a rebellion of the lower class to their fixed socio-economic position. Central events within the plot of the novel seem to perfectly illustrate that statistical pattern: by murdering his higher-class enemy Mayer, the lower-class protagonist becomes the central party in their mutual clash.

However, in a more resistant reading of the narrative conflict, *De Brooklynclub* escapes this observed statistical pattern subtly. There is, for instance, a recurring 'Doppelgänger'-motif in the novel that suggests that the protagonist and Mayor are actually one and the same person and that Mayer is a product of the protagonist's imagination. Read from this point of view, the conflict at stake is not a socio-economic clash between the high and the low class, but rather an inner conflict taking place within the boundaries of the protagonist's own psyche (and thus conforming to the third type of Pavis's typology: (Inner) conflict between e.g. passion and reason, see section 3). In this second reading, *De Brooklynclub* is not primarily a story about a lower class character taking revenge on a higher class character, but rather a novel about a confused underdog fighting against the person in the mirror.

While our statistics-based model of dyadic conflict presents the rather clear-cut finding that lower educated characters are more dominant in such antagonistic relations, this brief qualitative exploration at the level of one individual text shows that social conflict can be read in many ways depending on which of the possible readings is granted priority.

Model II: conflict between three characters

Now that there is a clearer image of the nature of hierarchies exposed through conflicts between two single characters, a closer look on broader network structures in conflict situations is warranted. In the statistical analyses carried in the former section, the conflicts at stake encompass only the first type offered by Pavis's typology (see section 3): Rivalry between two characters because of money, love, morality, politics, etc. This is a logical consequence of a focus on enemy-pairs consisting of only two characters. In theory, the second type of conflict from Pavis's typology – Two conflicting worldviews or irreconcilable moral conceptions – can also be expressed through dyadic relations, as two characters could have conflicting worldviews resulting in them becoming enemies. But this more abstract form of conflict generally involves more than two characters, as worldviews and moral conceptions can be expressed metonymically by multiple characters belonging to a group with shared values.

A focus on more than two characters potentially lays bare broader network dynamics that are at play in conflict situations. Such dynamics are also relatable to the fourth type of Pavis's typology: Conflict between the particular and the general, e.g. between the individual and society. It is imaginable that the intragroup morale leads to a clash with an individual belonging to that group. Furthermore, it might also relate to the fifth type: A moral or metaphysical conflict between a character and a principle such as God or an ideal. For instance, a group can have ideals that an individual has trouble accepting or refuses to adopt.

In order to limit this potentially large and heterogeneous subject of conflict in which more than two characters are involved, this section only considers subnetworks of three characters, also called triads. Triads are the second smallest network structure after dyads (two nodes linked by one edge). According to sociologist Georg Simmel, the transformation from a dyadic to a triadic network is the most radical, phase-shifting relational change.[42] Following Simmel, the shift of focus from dyads to triads makes it possible to research character conflicts in two fundamentally different – dyadic and triadic – contexts.

An empirical testing of Heider's social balance theory (see section 4) can gain insight into the effect that conflicts have on relationships between more than two characters. As explained above, the social balance theory posits that triadic signed networks are either balanced or unbalanced depending on the composition of the nodes' positive and negative attitudes towards the other nodes (see Figure 5 for a visualization of the theory). In the analysis below, positive relations are represented by friendship relations, negative relations by enemy relations.
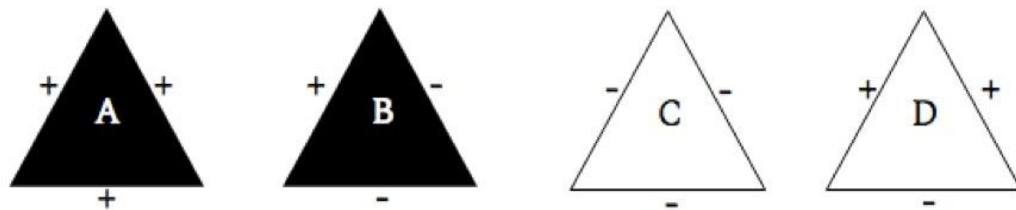

Figure 5. Visualization of Heider's social balance theory, adopted from Eve Kraicer and Andrew Piper. "Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction." Journal of Cultural Analytics (2019). Triangles A and B are balanced, triangles C and D are imbalanced.

To what extent are triadic relationships in the corpus balanced (triangle A or B in Figure 5) or imbalanced (triangle C or D in Figure 5), and what are the consequences of social (im)balance for the representation of social groups? According to Heider, imbalanced structures strife to balanced states because of people's urge for cognitive consistency. While it has been shown that the theory does not always hold empirically,[43] it is still generally considered as an established theory within the social sciences. Hypothesizing that the theory applies to character networks in literary fiction as well, we expect that the corpus contains relatively few imbalanced triads. In the following, social balance theory will serve as the framework for the analysis of character triads. We present a statistical pattern for the corpus as a whole, and we briefly illustrate the narratological value for one novel from the corpus that contains both balanced and imbalanced triadic subnetworks.

Automatic modeling of social balance in enemy/friend triads

For every triad consisting of either friends and/or enemies, it was automatically established whether or not it is balanced or imbalanced.[44] The following steps were taken to determine whether or not a triadic relation of either positive (friends) or

negative (enemies) relations is either balanced or imbalanced. First, from our database only those characters that have either an enemy or a friend relation were isolated.[45] Second, for each of these characters (N= 560) all their triadic subnetworks consisting of all friendly relations, all hostile relations, or a combination of friendly and hostile relations were singled out. Third, it was automatically determined whether or not the observed triads fall into the balanced or the imbalanced category (see Figure 5). The relative distribution of social (im)balanced states shows that 65% of the observed triads is balanced, as opposed to 35% imbalanced (N = 560). The absolute distributions of the (im)balanced categories friend-friend-friend (balance), friend-enemy-enemy (balance), enemy-enemy-enemy (imbalance), and enemy-friend-friend (imbalance) are shown in Figure 6.
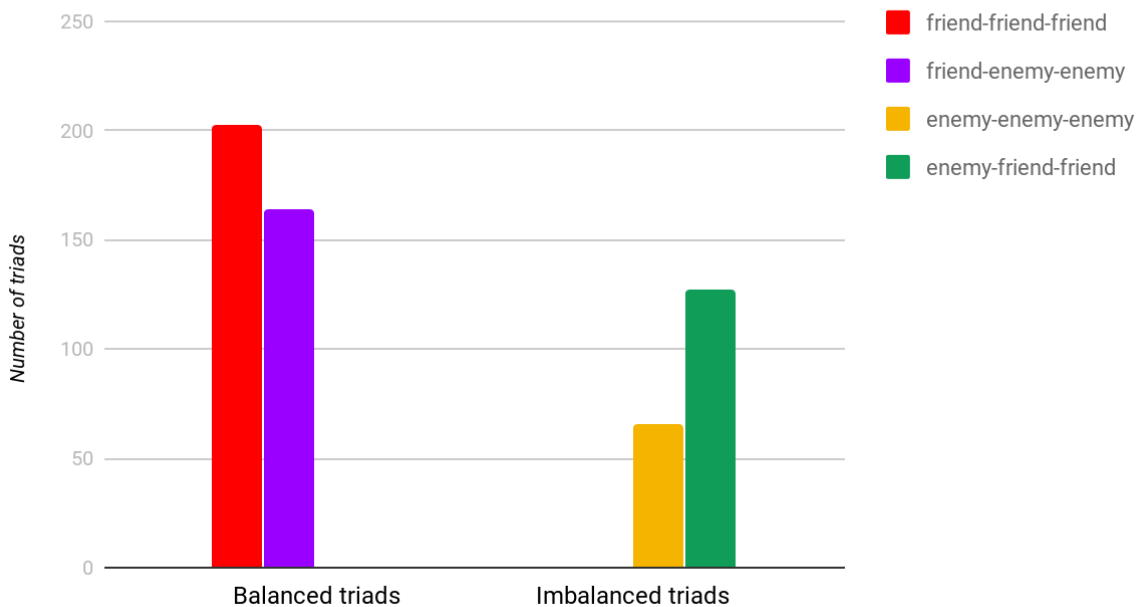

Figure 6. Absolute distribution of social (im)balance for all enemy/friend-triads in the corpus divided by type (N =560).

These results broadly confirm Heider's theory, but with counterexamples. The novelists in the corpus tend to prefer balanced triads over imbalanced triads, which might be interpreted as a literary reflection of Heider's assertion that people prefer balance over imbalance. To be clear, Figure 6 represents the distribution of

enemy/friend-triads across the corpus as a whole, it does not show what the process of striving for social balance looks like within individual novels. At the level of the individual novel, social balance might very well be a continuous process taking place within the linear time frame of the narrative – for instance, imbalanced triads at the start of the narrative that have been transformed into balanced triads by the end. We cannot, unfortunately, capture the continuity of this process based on our annotated data.

Furthermore, it remains the question how generalizable these results are beyond our corpus of 170 novels. Up to the present, as far as we are aware, there is only one other study by Kraicer and Piper that uses social balance theory for framing interactions between characters in literary texts.[46] This study focuses on a corpus of 1,333 contemporary English novels and reports a distribution of 53% balance as opposed to 47% imbalance. However, their method for detecting negative and positive relations deviates from the method used here. Kraicer and Piper automatically detect negativity/positivity of character relations bottom-up by using sentiment analysis, which results in a fairly rough and potentially inaccurate estimate of the nature of those relations. The present research uses top-down expert annotations of negative and positive relations between characters and comes arguably closer to how readers would frame those relations. In order to make generalizable statements on social balance in literature, comparative research has to be conducted on several corpora, preferably from other language fields and other time periods, using the same method.

Although there has been some critique on the generalizability of the theory, as imbalanced triadic structures have been reported in social networks,[47] this is not the place to make assertions on the general tenability of social balance theory. Presuming that Heider's theory has some truth in it, it is remarkable that 65% of the character triads in the corpus conform to it. In the context of Greimas's actantial model (see section 3), it would be more likely if fictional networks do not strive for social balance in the same way as real-world networks do. In fictional social networks, an alternative organizing principle might be in place that does not prefer social balance over social imbalance. This intuition appeals to the idea that disparity and divergence are driving factors behind the fictional worlds of novels, which can be associated with the axis of conflict in Greimas's actantial model. Action in

fictional narratives is considered to revolve around three axes (knowledge, desire, power/conflict), but the roles of the helpers and opponents, or friends and enemies, at the axis of conflict is of particular importance to keep the story interesting for the reader. In case the desires of the subject are not frustrated by any opponents at all, a tedious story would potentially be the result. More generally, an over-representation of imbalanced triads in the corpus might be an indication that stories do not tend to evolve smoothly but rather evolve with a necessary amount of squirming and irritation.

However, this appears not to be the case for the present data, as there is an over-representation of balanced triads as opposed to imbalanced triads. One interpretation of this unexpected outcome is that social structures in fictional worlds tend to resemble real-world structures. Assuming that in actual societies social balance is more present than social imbalance, these results can be used as an argument for the mimetic powers of literary worlds, which can be traced back all the way to Plato's and Aristotle's reflections on how literature imitates the world in which it is produced. Also in modern times, the idea that literature is a medium in which societal tendencies are reflected has been put forward.[48] In order to make a solid case for this interpretation, social (im)balance should have to be measured in contemporary Dutch society and compared with the present results.[49]

An alternative interpretation of the overrepresentation of social balance is that only a minimum of action can take place at the axis of conflict. One could argue that, in order for the story to evolve at all, the subject should have ample opportunities to reach his goal. In case there is too much social imbalance, too much activity at the axis of conflict, the story could not gain any momentum. Referring back to Propp's example of fairy tales: too much social imbalance would possibly result in the hero not even leaving his castle to save the princess, as he would be stuck in socially imbalanced structures at home.

In order to illustrate the narratological relevance of this finding, we briefly turn to a novel from the corpus in which the social (im)balance seems to create a rather schematic opposition between the social groups of Dutch characters and a group of characters with a Moroccan and Muslim background.

Social balance in Leon de Winter's VSV, of daden van onbaatzuchtigheid[50]

Conflict is definitely a driving force behind Leon de Winter's *VSV of daden van onbaatzuchtigheid* (2002). The novel's general setting is the polarized Dutch political climate that came into being after the murders of politician Pim Fortuyn (2002) and film director Theo van Gogh (2004). Both of them were outspoken critics of Islamic fundamentalism and were assassinated by people who were offended by their criticism. *VSV* is set in a country still recovering from the aftermath of these tragic events. Through the narrative perspectives of eleven different characters, a plot is set in motion that recalls the ideological conflict between religious fundamentalism and freedom of speech that lay at the heart of the Dutch public debate in the early 2000s. This is done specifically through the staging of actual people who played a role in sparking those debates, such as Theo van Gogh and his assassinator Mohammed Boujeri.

Generally speaking, the ideological conflict is represented by two opposing groups of characters, which activates a conflict of the second category of Pavis's typology: Two conflicting worldviews or irreconcilable moral conceptions. On the one hand, Dutch liberalism is embodied by a group of Dutch characters (Theo van Gogh, Max Kohn, Geert Wilders, Job Cohen, Piet Hein Donner, Leon de Winter). On the other hand, Islamic fundamentalism is embodied by a group of characters with a Morrocan migration background (Mohammed Boujeri, Sallie, Frits (Firas), Karel (Kareef), Kichie). Although a binary moral scheme is already installed by the very fact that each of the characters in this latter group is involved in criminal activity, reading the conflict between these groups in terms of social balance reveals the concrete social dynamics of this moral opposition.

There is a pattern of balanced triadic relationships in which two befriended characters from the latter group stand a negative relation towards one character from the first group. Figure 7 shows an example of such a balanced triad.
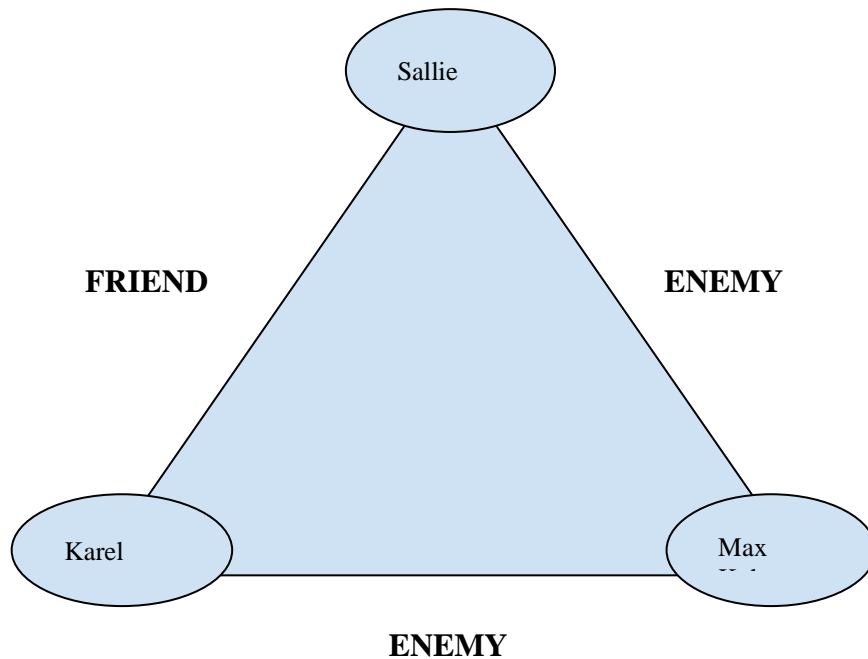
Figure 7. Balanced triad in Leon de Winter's VSV, *with characters Sallie (a.k.a. Salheddine Ouaziz), Karel (a.k.a. Kareef) and Max Kohn.*

This exemplifies the intergroup conflict between characters with and without a migration background. Being part of the group of Dutch characters, Max Kohn has a hostile relationship with Sallie and Karel who are part of the group of Moroccan characters with a criminal background. 'The enemy of my enemy is my friend' holds for this example: the enemy of Sallie's/Karel's enemy, i.e. Kohn, is Sallie's/Karel's friend, i.e. Karel/Sallie.

Most importantly, the Sallie-Karel-Kohn triad does not stand on its own; it is representative of a pattern. There is a range of other balanced triads in the novel for which the same argument can be made. Sallie and Karel not only feature in a balanced triad with Kohn, they also feature in similar balanced triads with each of the following Dutch characters: Nathan Verstraete (Kohn's son), Sonja Verstraete (Kohn's ex-lover), Job Cohen, Marijke Hogeveld (Cohen's lover), and Geert Wilders. On top of that, Sallie and Frits – another member of Sallie's group of criminal friends with a Morrocan migration background – feature in the exact same balanced triads as Sallie and Karel. In total, twelve balanced triads occur that are all indicative of the same two conflicting worldviews.

These twelve observed balanced triads fit perfectly in two different actantial models in which characters from each of the two conflicting groups function as each other's opponents on the two axes of conflict. In other words, each of these balanced triads stand in a metonymic relation to a general actantial model with either the Dutch or the Morrocan characters as a subject. No imbalanced triads are observed that potentially complicate these highly schematic moral oppositions between both groups. Instead, the intra-group morale in these twelve balanced triadic subnetworks confirm an overall moral scheme in which the Dutch characters are good and the Morrocan characters are bad. Subsequently, this is not only an opposition in terms of cultural background, it is also an opposition in terms of belief systems. The fact that these balanced triads strongly conform to such broadly defined actantial models suggests a bias towards the cultural values of Dutch liberalism.[51] This brief illustration shows how social balance at the level of the individual text can be read against the background of a general observation that the triads in the corpus as a whole tend to be balanced.

Conclusion

Conflict is a key element of narrative fiction, but literary theory up to this point lacks instruments to study this structural narrative mechanism in a data-driven way. Propp published *Morphology of the folktale* in the 1920s, almost 40 years later Greimas published *Sémantique structurale*. More than fifty years later, their classic models still serve as an inspiration for any structural analysis of narrative action. Recent developments within cultural analytics have opened up possibilities to take the ideas of Propp and Greimas to the next level. Computational and statistical analysis have the potential to fill up the gap left since Propp and Greimas, and enables a conceptualization and an operationalization of (the analysis) of conflict at the corpus level. Synthesizing typologies, concepts, terms and tools from a variety of research domains, we thus presented a new method to study conflict in literary texts. Given the often hierarchical aspects of conflict, we used this method to analyze the extent to which narrative conflict shapes the representation of social groups. Two models of conflict were presented testing two hypotheses related to the literary representation of genders, ethnicities and classes.

We modeled dyadic conflict by computing a conflict score for every pair of hostile characters in the corpus indicating the dominance of character in this two-way conflict, and hypothesized that male, characters without a migration background, and higher educated characters would have higher conflict scores. Surprisingly, gender and descent turned out to have no effect on a character's importance in a conflict. Education is the only statistically significant predictor of the height of these scores, as lower educated characters scored significantly higher than higher educated characters. Going beyond this classic protagonist-antagonist conflict, we studied triadic conflict by modeling the extent of social balance between all possible triangular configurations between enemies and friends in the corpus. 65% of these triads turned out to be socially balanced.

Both our dyadic and our triadic models of conflict have provided data-driven insights into the mechanism of narrative conflict. For the representation of social groups, the models have proven to be useful in exploring and mapping out hierarchies between characters with different demographic profiles. The main contribution of our research is, then, the introduction of such empirical patterns to the analysis of a key narratological concept such as conflict. For future research, we envision data-driven studies focusing on other narratological concepts, such as perspective, focalization, or dialogism, making use of similarly extensively annotated datasets.

References

[1] Propp, V. *Morphology of the Folktale* [2nd edition]. Austin/London: University of Texas Press, 1968 [originally published in 1928]. First edition translated by Laurence Scott with and introduction by Svatava Prokova-Jakobson, second edition revised and edited with a preface by Louis. A. Wagner and new introduction by Alan Dudes. Greimas, A.J. *Sémantique structurale*. Paris: Presse universitaires de France, 1966.

[2] Some handbooks of creative writing also incorporate conflict as an indispensable component of narrative fiction, see for instance Colin. *Creative writing: A guide and glossary to fiction writing*. Polity, 2007.

[3] This is not to say that there has been no critical engagement with the works of Propp and Greimas in the decades following their publication. See for instance Fredric Jameson, "Magical narratives: Romance as Genre." *New Literary History* 7, no. 1 (1975): 135-163. In postclassical narratology, the notions of narrative conflict and the related notion of problem-solving have also been thematized, e.g. in Herman, David. "Narratology as a cognitive science." *Image and Narrative* 1, no. 1 (2000). For a computational operationalization of Propp, see Mark Alan Finlayson, "Inferring Propp's functions from semantically annotated text." *The Journal of American Folklore* 129, no. 511 (2016): 55-77.

[4] Underwood, T. *Distant Horizons. Digital Evidence and Literary Change*. Chicago: University of Chicago Press, 2019. Underwood reverses the typical distant reading process as described by pioneer Franco Moretti: 'Instead of

measuring things, finding patterns, and then finally asking what they mean, we need to start with an interpretive hypothesis (a "meaning" to investigate) and invent a way to test it' (17). A model is defined by Underwood as nothing more than 'a relation between variables' (19).

[5] Deijl, L. van der, S. Pieterse, M. Prinse, and R. Smeets. "Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation." *Journal of Dutch Literature* 7 (2016) 1: 20-42. Volker, B., and R. Smeets. "Imagined social structures: Mirrors or alternatives? A comparison between networks of characters in contemporary Dutch literature and networks of the population in the Netherlands." *Poetics* (2019) https://doi.org/10.1016/j.poetic.2019.101379. Smeets, R., E. Sanders, and A. van den Bosch. "Character Centrality in Present-Day Dutch Literary Fiction." *Digital Humanities Benelux Journal* 1 (2019) 1.

[6] However, some basic criteria have to be met: the story needs to contain characters, and there has to be a certain extent of antagonism. Although there are literary texts in which characters and antagonism are less important or even absent, most literary texts meet these criteria. Furthermore, there are some practical issues: in order to apply our models to a specific national literature, one needs to be able to read the particular language as there is manual annotation involved.

[7] Galtung, J. (2010). Breaking the cycle of violent conflict. Retrieved September 26, 2018, from Public Speech at University of California Television website: https://www.youtube.com/watch?v=16YiLqftppo

[8] Tajfel, Henri, John C. Turner, William G. Austin, and Stephen Worchel. "An integrative theory of intergroup conflict." *Organizational identity: A reader* 56 (1979): 65, p.33. Besides interpersonal and intergroup conflict, we distinguish *intra*personal conflict taking place within the psyche of a single character. A more extensive typology of conflict in narrative fiction is introduced in the next section.

[9] Tajfel, Henri, John C. Turner, William G. Austin, and Stephen Worchel. "An integrative theory of intergroup conflict." *Organizational identity: A reader* 56 (1979): 65, p.33

[10] Tajfel, Henri, John C. Turner, William G. Austin, and Stephen Worchel. "An integrative theory of intergroup conflict." *Organizational identity: A reader* 56 (1979): 65, p.35

[11] Greimas, A.J., and J. Courtés. *Analytisch woordenboek van de semiotiek*. Tilburg: Tilburg University Press, 1987. Translated by Ko Joosse and Paul de Maat.

[12] Greimas, A.J., and J. Courtés. *Analytisch woordenboek van de semiotiek*. Translated by Ko Joosse and Paul de Maat. Tilburg: Tilburg University Press (1987), p.70

[13] Greimas, A.J., and J. Courtés. *Analytisch woordenboek van de semiotiek*. Translated by Ko Joosse and Paul de Maat. Tilburg: Tilburg University Press (1987), p.324

[14] Katz, Tamar. "Modernism, Subjectivity, and Narrative Form: Abstraction in" The Waves"." *Narrative* 3, no. 3 (1995): 232-251.

[15] Greimas, A.J., and J. Courtés. *Analytisch woordenboek van de semiotiek*. Translated by Ko Joosse and Paul de Maat. Tilburg: Tilburg University Press (1987), p.200

[16] Foreword in Propp, V. *Morphology of the Folktale* [2nd edition]. Austin/London: University of Texas Press, 1968 [originally published in 1928]. First edition translated by Laurence Scott with and introduction by Svatava Prokova-Jakobson, second edition revised and edited with a preface by Louis. A. Wagner and new introduction by Alan Dudes.

[17] Idem, p. 21-23.

[18] Ide, 79-80.

[19] Greimas, A.J. *Sémantique structurale*. Paris: Presse universitaires de France, 1966, p.172-191.

[20] For studies taking these into account see Bohn, A., C. Buchta, K. Hornik, and P. Mair. "Making friends and communicating on Facebook: implications for the access to social capital." *Social Networks* 37 (2014): 29-41. Box-Steffensmeier, J.M, and D.P. Christenson. "The evolution and formation of amicus curiae networks." *Social Networks* 36 (2014): 82-96. Camp, M. van de, and A. van den Bosch. "The socialist network." *Decision Support Systems* 53 (2012) 4: 761-769.

[21] Doreian, P. "Signed Networks." In *Encyclopedia of Social Networks*, edited by George A. Barnett, 777. Thousand Oaks, CA: SAGE Publications, Inc., 2011. doi: 10.4135/9781412994170.n320.

[22] Heider, F. "Attitudes and Cognitive Organization." *The Journal of Psychology* 21 (1946): 107.

[23] Cartwright, D., and F. Harary. "Structural Balance: A Generalization of Heider's Theory." *Psychological Review* 63 (1956). Davis, J. "Clustering and Structural Balance in Graphs." *Human Relations* 20 (1967).

[24] Doreian, P. "Signed Networks." In *Encyclopedia of Social Networks*, edited by George A. Barnett, 777. Thousand Oaks, CA: SAGE Publications, Inc., 2011. doi: 10.4135/9781412994170.n320.

[25] The prize roughly follows the system used by the Man Booker Prize for Fiction. There is no restriction to the amount of novels publishers can annually submit. From the aggregated list of all submitted novels, members of an annually changing professional jury first select a longlist of 18 titles, and then a shortlist of 6 titles. As publishers submit novels that they hope to be possible winners, the full list of submissions reflects what they see as high quality literary fiction. Since 2010, the prize targets literary novels for adults exclusively. The consequence is that other forms of prose cannot be submitted: the bulk list of submitted novels does not contain young adult novels, children's books, fantasy, or (literary) thrillers.

[26] As all new novels written in the Dutch language are eligible for the prize, the list of submissions consists of both Dutch and Flemish authors. In this study, 'Dutch literature' thus refers to both Dutch and Flemish literature as is common in the Dutch language area.

[27] Many thanks to Erik Vos for providing us with up-to-date information.

[28] For studies using this BookNLP see, for instance, Bamman, D., Underwood, T., & Smith, N. A. (2014). A bayesian mixed effects model of literary character. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 370–379. Underwood, T., Bamman, D., & Lee, S. (2018). 'The transformation of gender in English language fiction.' *Journal of Cultural Analytics*. DOI: 10.31235/osf.io/fr9bk. Piper, Andrew. *Enumerations: data and literary study*. University of Chicago Press, 2018.

[29] These results were published in Lucas van der Deijl, Saskia Pieterse, Marion Prinse, and Roel Smeets. "Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation." *Journal of Dutch Literature* 7, no. 1 (2016).

[30] All data and code are available via an open-access GitHub repository (https://github.com/roelsmeets/character-networks), which, however, does not contain the digital versions of the 170 texts of the novels due to copyright issues.

[31] For a detailed overview of the descriptive statistics of each of the variables, see the chapter 'Data' in Roel Smeets, *Character Constellations*. Leuven University Press, to appear.

[32] See section 5, Corpus and Dataset. The software and data used in this section can be found in the following open access GitHub repository (see the file conflict.py for the Python code, and the file character-rankings_conflictscore.csv for the output of the code): https://github.com/roelsmeets/character-networks.

[33] For details, see Smeets, R., E. Sanders, and A. van den Bosch. "Character Centrality in Present-Day Dutch Literary Fiction." *Digital Humanities Benelux Journal* 1 (2019) 1. In general, this approach is based on the idea that characters co-occurring more often have stronger edges than characters co-occurring less often. Characters were automatically identified in the texts through the names they are referred to with. This approach resulted in 170 networks, one for each novel, with an average of 12 identified characters per novel (and a minimum of 3 and a maximum of 29 identified characters).

[34] For a detailed description of these centrality metrics in the context of literary fiction, see R.J.H. Smeets, E. P. Sanders, and A. P. J. van den Bosch. "Character Centrality in Present-Day Dutch Literary Fiction." *DH Benelux Journal*, 2019.

[35] The conflict score of a character was *not* normalized by dividing the resulting number through the amount of enemies a character has, as this would level out relevant differences in conflict scores between characters with varying amounts of enemies. E.g a character that has 10 enemies and is higher ranked than 8 of them, should have a higher conflict score than a character that has only 1 enemy to which he is higher ranked. When these scores are normalized, both of these characters would have a conflict score of 1, which does not do justice to the fact that one character is more often higher ranked than the other one.

[36] Meijer, M., and E. van Alphen (eds.). *De canon onder vuur*. Amsterdam: Van Gennep, 1991. Meijer, M. "De verschrikkelijke sneeuwman: projectie, geweld en nieuwe mannelijkheid in het werk van Jan Wolkers." In *Het omstreden slachtoffer: geweld van vrouwen en mannen*. edited by. R. Romkens, and S. Dijkstra. Baarn: Ambo, 1996. 39-58. Meijer, M. *In tekst gevat. Inleiding tot de kritiek van de representatie*. Amsterdam: Amsterdam University Press, 1996. Meijer, M. "Zwartheid in de witte verbeelding." In *Diversiteit*, edited by Judit Gera. Boedapest: Printart Press Kft., 2011. 47-74. Minnaard, L. "The Spectacle of an Intercultural Love Affair. Exoticm in Van Deyssel's Blank en Geel." *Journal of Dutch Literature* 1 (2010) 1: 74-90. Pattynama , P. "Oorden en woorden. Over rassenvermenging, interetniciteit, en een Indisch meisje." *Tijdschrift voor Vrouwenstudies* 15 (1994) 1: 30-45. Pattynama, P. "Secrets and Danger: Interracial Sexuality in Louis Couperus's *The Hidden Force* and Dutch Colonial Culture around 1900." In *Domesticating the Empire. Race, Gender, and Family Life in French and Dutch Colonialism*, edited by J. Clancy-Smith, and F. Gouda. Charlottesville/London: University Press of Virginia, 1998. 84-107.

[37] Kristeva, J. *Séméiôtiké: recherches pour une sémanalyse*. Paris: Edition du Seuil, 1969, p. 183. Cassin, B. (ed.). *Nos Grecs et leurs modernes*. In *L'antiquité classique*, Tome 63, 1994, 151-276

[38] The dataset of 2,137 characters roughly has a 60:40 ratio of male and female characters, a 90:10 ratio of characters with and without a migration background – a relatively small portion of these features are unknown. For more than 40% of the characters, their educational background is unknown, whereas 42% is higher educated and 18% is lower educated. For a detailed overview of the descriptive statistics of each of the variables, see the chapter 'Data' in Roel Smeets, *Character Constellations*. Leuven University Press, to appear.

[39] The labels 'with a migration background' and 'without a migration background' are used in a loose sense in the present study. Characters with a migration background can also refer to characters whose parents migrated to the Netherlands or Belgium, but who were themselves born in the Netherlands or Belgium. In this broad definition, characters with a migration background are considered to have some sort of bond with a socio-cultural tradition that is not the same as their current country of residence. The Netherlands and Belgium were chosen as a point of reference as the books in the corpus are either written by Dutch or Flemish authors who arguably operate in a shared literary field of Dutch literature.

[40] Lochner, L., and E. Moretti. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests and Self-reports."*American Economic Review* 94 (2004): 155-89. Barrera, F., and Ibánez, A.M., "Does violence reduce investment in education?: A theoretical and empirical approach. ". *No. 002382*. Universidad de los Andes-CEDE, 2004.

[41] For a more detailed and nuanced analysis of this novel, see Roel Smeets, *Character Constellations*. Leuven University Press, to appear.

[42] 'Relations between two people have strictly specific characteristics – this is not only demonstrated by the fact that the relation changes fundamentally when a third one comes in, but even more by the following frequently stated fact: if the network is extended to four or more people, this will by no means change the essence of the network any further.' Quote from page 92 in Simmel, G. *Een keuze uit het werk van Georg Simmel*. Deventer: Van Loghum Slaterus B.V., 1976. Translated by Han Israëls.

[43] Doreian, P. "Signed Networks." In *Encyclopedia of Social Networks*, edited by George A. Barnett, 777. Thousand Oaks, CA: SAGE Publications, Inc., 2011. doi: 10.4135/9781412994170.n320.

[44] The software and data used in this section can be found in the following open access GitHub repository (see the file named 'conflict.py'): https://github.com/roelsmeets/character-networks

[45] As mentioned before, characters can have multiple types of relations with the same character, e.g. Colleague_Friend. The analysis exclusively focused on the categories of Friend and Enemy, and ignored the other categories in case of double, triple or more relational annotations.

[46] Eve Kraicer and Andrew Piper, "Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction," *Journal of Cultural Analytics* (2019). DOI: 10.22148/16.032.

[47] Doreian, P. "Signed Networks." In *Encyclopedia of Social Networks*, edited by George A. Barnett, 777. Thousand Oaks, CA: SAGE Publications, Inc., 2011. doi: 10.4135/9781412994170.n320.

[48] See, for instance, Auerbach, E. *Mimesis: The representation of reality in Western Literature*. Princeton: Princeton University Press, 2003 [1946]. Hoggart, R. "Literature and society." *The American Scholar* (1966): 277-289. Anjana, B. K., and R. L. Bhambhra. "Is literature the mirror of society." *International Journal of English Language, Literature and humanities* (2016).

[49] See Volker, Beate, and Roel Smeets. "Imagined social structures: Mirrors or alternatives? A comparison between networks of characters in contemporary Dutch literature and networks of the population in the Netherlands." *Poetics* 79 (2020): 101379. This study compares the actual networks of people living in the Netherlands with the network of characters in our corpus. The research does not focus on social balance theory, but the dataset of actual personal networks does seem to conform to an overrepresentation of social balance as opposed to social imbalance. In this dataset (SSND) on personal networks of people in the Netherlands, networks were, however, delineated according to name generators that are likely to tap into structurally balanced network ties, such as the question 'with whom do you talk about important personal matters?'.

[50] For a more detailed and nuanced analysis of this novel, see Roel Smeets, *Character Constellations*. Leuven University Press, to appear.

[51] Obviously, there are examples of novels in which the occurrence of either balanced or imbalanced triads complicates the construction of too broadly defined ideological schemes rather than accommodating them such as the balanced triads in *VSV*. One such a novel in the corpus is Tommy Wieringa *Dit zijn de namen* [These are the names] (2012), in which a range of imbalanced triads obscures the tendency to divide characters in two opposed social and moral groups.