The Sense and Sensibility of Different Sliding Windows in Constructing Co-occurrence Networks from Literature

Siobhán Grayson, Karen Wade, Gerardine Meaney, Derek Greene

The Sense and Sensibility of Different Sliding Windows in Constructing Co-occurrence Networks from Literature

Siobhán Grayson[1], Karen Wade[2], Gerardine Meaney[2], and Derek Greene[1]

[1] School of Computer Science, University College Dublin, Ireland
{siobhan.grayson, derek.greene}@insight-centre.org
[2] Humanities Institute, University College Dublin, Ireland
{karen.wade, gerardine.meaney}@ucd.ie

Abstract. In this paper, we explore the design and effects of applying different sliding window methodologies to capture character co-occurrences within literature in order to build social networks. In particular, we focus our analysis on several works of 19th century fiction by Jane Austen and Charles Dickens. We define three different sliding window techniques that can be applied: collinear, coplanar, and combination. Through simple statistical analysis of each novel's underlying textual properties we derive tailored window sizes for each case. We find that the selection of such parameters can significantly affect the underlying structure of the resulting networks, demonstrated through the application of different social network metrics on each of our novels. We also examine how the choice of window strategy can help address specific problems in current critical understanding of the novel.

Keywords: social network analysis, data-driven research, modeling and analysis, literary analysis

1 Introduction

Computational approaches are being increasingly adopted by humanities scholars to explore questions in the field of literature from new perspectives [11]. In particular, social network analysis (SNA) provides researchers with an array of existing analysis techniques, together with a unique level of abstraction (*i.e.* a network of nodes and edges), whilst still maintaining the social structure of the novels and the societies they depict. The application of SNA in a literary context often involves the construction of *character networks* from a digital text, where each node in the network represents a character and each edge indicates some kind of relation between characters. In practice, such associations are identified by analysing the *co-occurrences* of pairs of characters within the text. Using these networks, methods from SNA potentially allow humanities scholars to test existing or new literary hypotheses from a quantitative perspective, in conjunction with existing close reading strategies. However, the success of these methods is intrinsically dependent on the quality of the underlying networks themselves.

The extraction of character networks from 19th-century texts is non-trivial, due to the fact that characters often share names and aliases and are frequently referenced in implicit or ambiguous ways.

To date, most literary social networks have been extracted automatically, with authors making allowances for their incompleteness or inconsistencies [2, 4, 8]. In this paper, we describe three different *character network* construction strategies to detect co-occurrences, based on sliding a context window over the text of each chapter in a novel. These co-occurrences are then used to construct a weighted undirected network representation of the novel. In Section 4 we demonstrate the impact of the choice of method and associated window size parameter using the texts of nine popular 19th century novels written by the British authors Jane Austen and Charles Dickens, available from Project Gutenberg[1]. These texts have been manually annotated in order to include as many character entities as possible, including minor and collective-presenting characters. We illustrate how altering the network construction method affects the structure and density of the resulting character networks.

In Section 4.2, we discuss in detail the application of different window strategies to the construction of networks in a specific novel, *Sense and Sensibility*, and give examples of ways in which the differing types of character networks have implications for current literary scholarship. We find that character centrality for the overall novel is closely associated with wealth where the top highest centrality scoring characters represent an elite subsection of society. This provides a new perspective on the depiction of financial stability in the works of Austen. Furthermore, new insights can be gained by analysing individual chapter networks generated using each of our window strategies. In particular, we find that the collinear method is more reflective of an author's narrative technique and useful for identifying narrative divergences and asides, while combination strategies illuminate the connectivity between and across social classes.

2 Related Work

2.1 Social Networks in Literature

A range of different approaches have been considered to identify meaningful interactions between characters in fictional texts. One of the first studies, conducted by Alberich [3] *et al.*, assembled the Marvel Universe collaboration network by identifying connections between characters on whether they occurred in the same comic, independent of the type of interaction itself. Gleiser *et al.* [5] modified this method by introducing weights to account for the possibility of stronger collaborative ties existing between characters that repeatedly co-occur throughout the same text. Taking a different approach, Moretti [11] analysed the works of Shakespeare by constructing networks defined on the basis of dialogue alone demonstrating that interactions in dramatic works can be readily

[1] `https://www.gutenberg.org`

converted to a network representation. However, when dealing with prose, limiting interactions to quoted speech will exclude large amounts of non-quoted dialogue, observations, and thoughts [1]. Even when focusing on quoted speech, the construction of conversational networks from classical literary texts is not straightforward.

Elson *et al.* [4] constructs social networks from 19th century literature by detecting conversations from sets of dialogue acts, which involves character name clustering followed by automated speech attribution. While this approach achieves a high level of precision (96%), the level of recall for conversational interactions is low (57%), even before other types of character interactions are considered. In an attempt to overcome the limitations of using dialogue alone, Agarwal *et al.* [1] examine two distinct types of social events involving characters in Lewis Carroll's *Alice in Wonderland* (1865): interactions and observations. The authors construct a weighted undirected social network from instances of the former, and a weighted directed network from instances of the latter, where edge direction is based on who is observing whom.

More recently, Jayannavar *et al.* [8] also apply an extraction technique which goes beyond dialogue, looking at the network of general character interactions, as well as considering specific cases of conversational interactions and observations. The edges in these networks were subsequently used to test a set of literary hypotheses. Other authors have also looked at general interaction networks extracted from fictional texts. Rydberg-Cox *et al.* [13] employ SNA to visualise and explore the interactions between characters in Greek tragedies aiming to meld distance and close reading.

2.2 Term Co-occurrence Analysis

Beyond the study of literature, co-occurrence analysis has often been used to identify the linkages between words in unstructured texts. For instance, the relationship between pairs of terms occurring within a constant-sized context window is a key component of popular word embedding methods such as *word2vec* [10]. In topic modeling, the frequent co-occurrence of a pair of terms within a sliding window of fixed size moving over a corpus is used to measure topic coherence [12]. In both applications, the choice of context window size is often not considered in detail. However, Zadeh and Handschuh [15] demonstrated the importance of context window sizes when identifying co-occurring terms for the purpose of classification and characterised the use of context windows based on their size and the direction in which they are extended. For instance, Traag *et al.* [14] examined networks of public figures extracted from media articles, where edges were created between pairs of disambiguated occurring in the same sentence. Such approaches, while suitable for contemporary factual texts which are carefully structured and formatted, will not be applicable to poorly digitised or inconsistently formatted literature from previous centuries.

3 Methods

3.1 Data Preparation

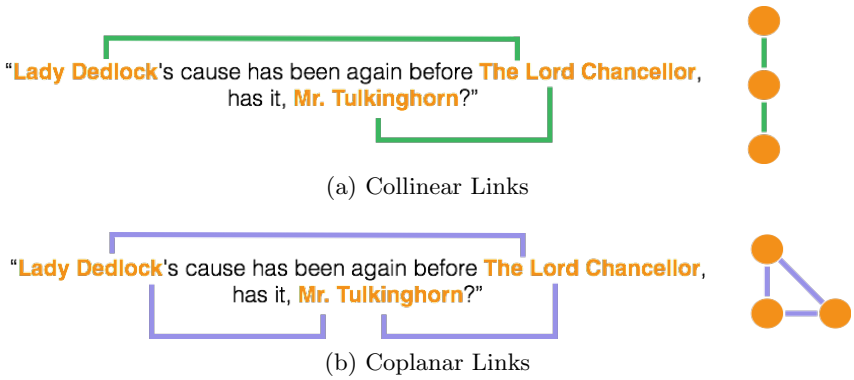
In this paper we consider a collection of nine novels from two 19th century British novelists - six by Jane Austen and three by Charles Dickens - sourced from Project Gutenberg. Initial data preparation involves the manual annotation of the novels, where literary scholars identify all character references in the text of each novel. The annotation process itself consists of a number of steps. Firstly, a *character dictionary* is constructed, which includes a single entry for each unique character in the novel (identified by their *definitive name*) and the corresponding *aliases* for that character which appear in that novel (*i.e.* all names used to refer to them). For instance, Elizabeth Bennet in *Pride and Prejudice* is referred to by a number of aliases, including Elizabeth, Lizzy, and Eliza. Once the dictionary has been compiled, all instances of a character's aliases in the novel text are replaced with their definitive name. For the six Austen novels in our study the average dictionary size was 153, while for Dickens the average size was considerably larger at 288 characters.

3.2 Character Networks

Once a novel has been annotated, we can construct a corresponding network representation. Formally, the *character networks* described in this study are defined as undirected, weighted graphs, denoted by $G = (N, E)$ where N is a set of nodes representing the cast of all unique characters, and E a set of edges representing all associations between unordered pairs of characters. The numeric weight on an edge indicates the strength of the association. In practice, we construct a detailed character network from an annotated novel by first creating a node for each character in the novel's character dictionary. Each chapter of the annotated text is then tokenised and an appropriate strategy is applied to identify and count all co-occurrences of character mentions. For each chapter, we count the number of co-occurrences for every pair of characters. We then create a weighted character network for the chapter, where edges are weighted to reflect multiple co-occurrences. Finally, we construct an overall network for the novel by aggregating the individual networks from all chapters.

3.3 Collinear Co-occurrence Window Strategy

In this strategy, a sliding window of size w_l tokens moves over the text of each chapter. A co-occurrence between characters X and Y is identified when Y appears after X within this window. The strategy is collinear in that only consecutive pairs of characters are counted, and it is conservative in the sense that a co-occurrence between Y and another character appearing prior to X is not counted. This can be viewed as a variant of the left-hand context window approach described for term co-occurrence in [15]. The size of the sliding window w_l is identified independently for each novel. Firstly, we construct an overall



(a) Collinear Links

(b) Coplanar Links

Fig. 1: Examples of the collinear (a) and the coplanar (b) strategies used for creating edges between character co-occurrences demonstrated using an except of text from Chapter 2 of *Bleak House* by Charles Dickens.

character network for each window size $w_l \in [20, 300]$ words. We then calculate the weighted edge density as w_l increases and plot these values. Finally, we automatically identify the point at which this plot plateaus. This indicates that increasing the window size further will not capture any additional unique character interactions. An example of this strategy is shown in Fig. 1(a).

3.4 Coplanar Co-occurrence Window Strategy

Our second strategy is less conservative in that it aims to capture associations beyond pairs of consecutive mentions, as illustrated in Fig. 1(b). In both cases, window size is important in establishing which characters are considered connected. However, due to the nature of coplanar connections, the method used to derive collinear window sizes is not applicable. This is because as the window size increases, rather than plateauing, the weighted edge density continues to increase until every character is connected to each other. Instead, the number of tokens between characters, referred to as "gaps", are analysed. The theory being that as the number of tokens increases between characters, the probability of an interaction decreases. Thus, treating gaps as the boundaries of character interaction events, window sizes are generated by exploring the most probable upper limits derived by applying simple, non-parametric statistical analysis on each text's gaps distribution (D_g). In particular, we take advantage of the interquartile range ($ICR = Q_3 - Q_1$) to define $inf(D_g) = Q_1 - 1.5 \times IQR$ and $sup(D_g) = Q_3 + 1.5 \times IQR$ where Q_1 is the first quartile, and Q_3 is the third quartile. Any elements which lie outside these limits are considered suspected outliers and are trimmed. Three window sizes are then considered: $w_{p1} = Q_3$, $w_{p2} = (sup(D_g) + Q_3)/2$, and $w_{p3} = sup(D_g)$.

3.5 Combined Sliding Window Strategy

As described above, the coplanar strategy captures associations beyond pairs of consecutive mentions, however, this is at the expense of seizing potential interactions which are further spaced out, and which would be naturally accommodated for by the larger window sizes enjoyed by collinear methods. Thus, the combined strategy consists of executing both the collinear and coplanar methods to identify character interaction pairs. The resulting co-occurrence pair sets are then merged, where pairs present in the collinear method, but not the coplanar, are added to the coplanar pair set. Thus, combination networks not only represent coplanar associations but also capture the further spaced interactions that collinear accounts for.

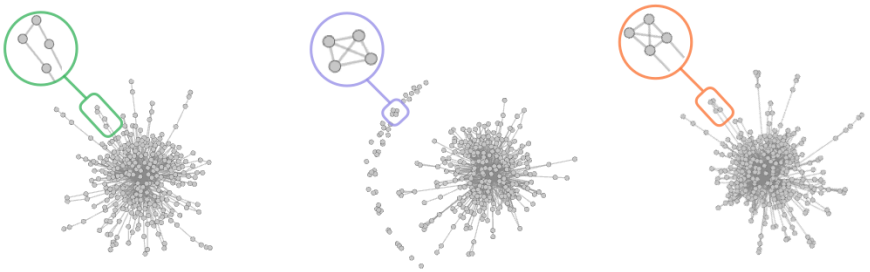
4 Results

4.1 Network Analysis

A summary of each novel's properties and the resulting window sizes for the different network construction strategies is given in Table 1. Interestingly, Austen's *Northanger Abbey* has the largest collinear sliding window with $w_l = 130$, despite having the least amount of tokens $T = 57153$. It also has the highest coplanar window sizes indicating that a larger amount of text passes between character mentions within the plot. However, this correlation is not observed elsewhere. For instance, *Oliver Twist* has the second highest collinear window size ($w_l = 120$) but generated the lowest coplanar window sizes. To quantify the effect of each window strategy in terms of network topology, we have applied a number of common SNA metrics which we now discuss.

Table 1: Summary of overall character network properties for the novels in our study (6 from Austen, 3 from Dickens) and selected window sizes. Here N is number the of characters, $\#T$ is the number of tokens (including character mentions), w_l is the collinear window size, and w_{p1}, w_{p2}, w_{p3} are coplanar window sizes. All window sizes are in unit tokens (words).

Novel	$\#N$	$\#T$	$\#Chap$	w_s	w_{m1}	w_{m2}	w_{m3}
Northanger Abbey	94	75153	31	130	45	72	99
Pride and Prejudice	117	120262	61	90	34	54	74
Persuasion	136	81809	24	90	37	60	83
Sense and Sensibility	158	118149	50	70	34	54	74
Emma	193	156364	55	100	37	59	80
Mansfield Park	218	157800	48	90	39	62	85
Oliver Twist	286	153990	53	120	32	51	69
Great Expectations	288	177043	59	110	39	63	87
Bleak House	516	341441	67	100	36	58	79



(a) (b) (c)

Fig. 2: Overall network of *Oliver Twist* with the same group of four characters highlighted and focused on in each case where (a) is collinear, (b) is coplanar $w_p = 32$, and (c) is combination $w_l = 120$.

As expected, each graph's weighted edge density, d_w, increases as we move from collinear, to coplanar, through to combination, demonstrating that the collinear methods is more robust against reaching the upper limits of graph density, and that a high density is a natural consequence of the coplanar strategy. We also measured the average node disconnect within each graph and found it decreases from collinear to coplanar and combination. In Fig. 2, the overall network of *Oliver Twist* is visualised for each window strategy where the same group of four characters have been highlighted and focused on in each case. Fig. 2 (a) represents the collinear network which on closer inspection shows the group of four character interactions occurring in a chain. Fig. 2 (b) depicts the coplanar-32 network; in contrast with Fig. 2 (a), this shows a large number of characters and groups which are now disconnected, including the previous group of four, although further interactions have been established between the characters within this subgroup. Finally, Fig. 2 (c) reconciles both approaches. In this combination-32 network, not only is the subgroup of four members reattached to the remainder of the network by way of links originally established by the collinear approach, but the associations between members of this group are also preserved.

Another way of illustrating the effects of each window strategy is to compare the average clustering coefficient (C) and average betweenness (B) of all characters. We found both C and B decrease as we move from collinear to coplanar through to combination. These results highlight how the collinear strategy primarily forms edges in a chaining succession (see Fig 3 (a)), causing the same characters within chapter 12 of *Pride and Prejudice* to be linked in such a manner as to have inflated clustering and betweenness values in comparison to coplanar (Fig 3 (b,c,d)) and combinational models (Fig 3 (d)) for the same text.

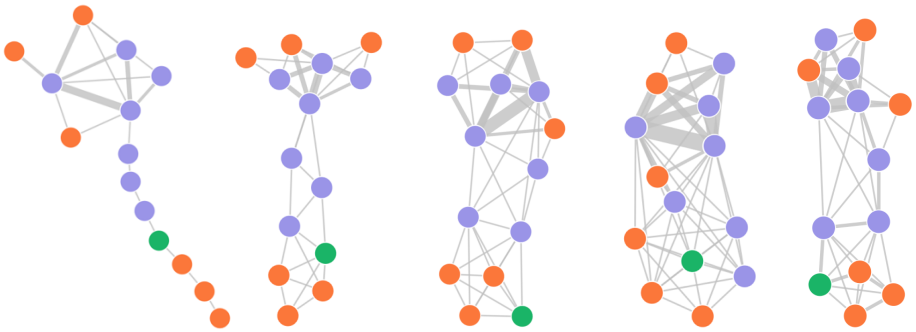


Fig. 3: Chapter 12 networks from Austen's *Pride and Prejudice* using four different sliding windows. (a) is collinear, (b,c,d) are coplanar, and (e) is combination with $w_p = 54$. Nodes coloured according to gender: purple is female; orange is male; and green is collective or NA.

To examine the effect of using different window strategies and sizes on the character rankings, we focus on *Oliver Twist* by Charles Dickens. The difference is quickly apparent when we consider only the top five characters ranked by degree. Strikingly, not one of the coplanar networks replicates the ordering of the five characters, highlighting the influence that window size can have upon the centrality of even "major" nodes within a network. Interestingly, collinear ($w_l = 120$) and coplanar-32 are most comparable, despite their completely different methodologies and sizes. When extended to view the top ten degree characters of *Oliver Twist*, not only do characters change ranking, but the characters that appear can also differ. For instance, The Artful Dodger replaces Mrs. Maylie within the top ten degree ranking for the collinear network, and supplants Nancy from the top ten degree ranking within the coplanar-32 network. We will now examine in detail, with reference to Jane Austen's *Sense and Sensibility*, how the choice of window strategy can influence whether a novel network is more illustrative of the society depicted within the novel, or of the narrative technique utilised by its author.

4.2 Discussion: Literary Implications for Character Networks

Character Centrality, Societal Status and Wealth. Examining the networks generated for *Sense and Sensibility* using collinear and combination window strategies, we find that certain elements in the different networks correspond to issues that have been raised within ongoing debates in literary scholarship. One such discussion, pertaining to the social exclusivity of the world of Austen's novels and focusing upon the financial status of her characters[2]. Robert D. Hume

[2] An extended discussion of the social exclusivity of the world of Austen's novels has been ongoing, originating with the publication of Copeland's *Women Writing About*

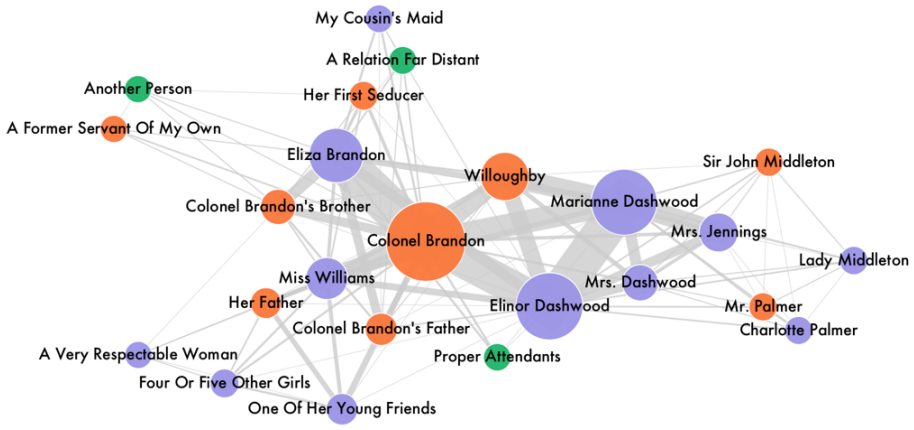


Fig. 4: Overall combination network of Chapter 31 from *Sense and Sensibility* using the largest window size $w = 74$. Nodes are sized according to weighted degree where larger nodes reflect higher weighted degree values and visa versa. The node colour reflects the gender of the character: purple for female; orange for male; and green for collective or NA.

[7] has argued that we cannot read Austen with any real clarity if we do not understand the economic circumstances of her character: "Sense and Sensibility poses a blunt question: what is a satisfactory competence on which a family may live decently?" Drawing on an analysis of the 1801 census, Hume [7, p. 293] points out that the modest and sensible £850 a year which Elinor and Edward consider adequate to embark on married life together would in fact place them in the top 1.8% of society in terms of income, while the £2000 to which Marianne aspires in order to satisfy her refined sensibility and passion would put her in the top 0.17%. Upon examination of the networks, it becomes apparent that character centrality is closely associated with wealth; the top ten characters in the novel (calculated by betweenness, eigenvector and weighted centrality measures) are all undoubtedly members of an elite subsection of society. This provides a new perspective on the depiction of financial (in)stability in Austen's works; the position of her female characters in particular is financially precarious, and the appearance of gentility is expensive to maintain, but nonetheless their perception of themselves as dispossessed and afflicted needs to be understood in the context of an era in which 95% of families would have subsisted upon less than £250 a year.

Money (1995) and was galvanised in the last decade by the 2005-2008 Cambridge Edition of the Works of Jane Austen (general editor, Janet Todd), which systematically interprets for the modern reader the financial information which Austen provides in remarkable detail.

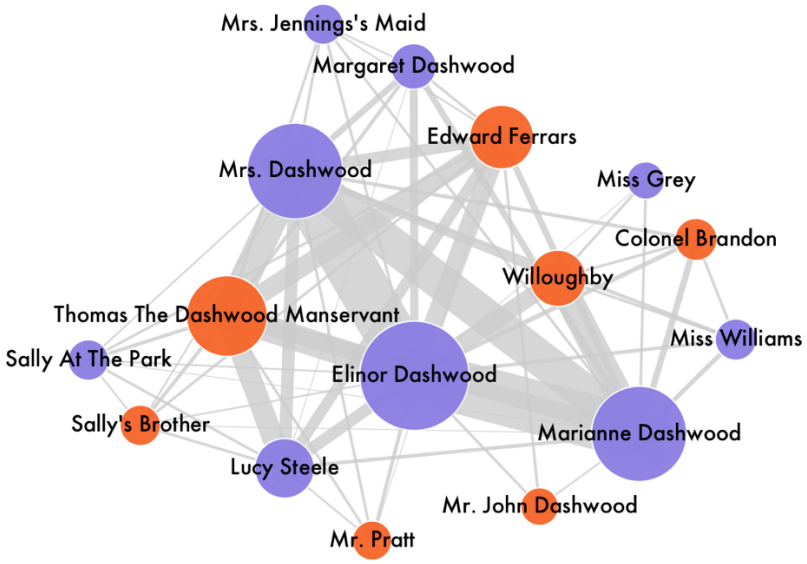


Fig. 5: Overall combination network of Chapter 47 from *Sense and Sensibility* using the largest window size $w = 74$. Nodes are sized according to weighted degree where larger nodes reflect higher weighted degree values and visa versa. The node colour reflects the gender of the character: purple for female; orange for male; and green for collective or NA.

Combination Networks of *Sense and Sensibility*. Analysing centrality at the level of *Sense and Sensibility* as a whole, social network analysis therefore strongly supports Hume's assertion that the world of the novels is not a bourgeois world, but an aristocratic one. However, at the chapter level, the combination method, together with the principle of radical inclusivity in compiling character dictionaries (in which all possible characters and character collectives, rather than just the main characters, are identified, tagged and counted), allows us to qualify and question the extreme social exclusivity which recent critics have attributed to the world of Austen's novels. Upon close examination of two chapters where collinear and coplanar analysis produce minor divergences (chapters 31 and 47) in *Sense and Sensibility*, we can see that although the principal protagonists fall into the top 5% of income groups as identified by Hume, this is not true of all of the nodes in the network. The lives of the Dashwoods are deeply enmeshed with the lives of others - not only of their servants, although these are the most prominent representatives of the world of the 95%, but through them with a wider world that includes the post-boy and his family, random encounters on the street, respectable women, and seducers and their victims.

Two crucial reports - one, in chapter 47, from the Dashwoods' manservant Thomas about the possibility of Edward Ferrars's marriage, and the second from Colonel Brandon in chapter 31, on the history of his ward, Miss Williams - de-

pend on information passing between social classes. In chapter 31, information also passes across the deep abyss between propriety and impropriety into which Marianne is in danger of falling, when following her sensibility. In both cases, both narrative and character development is dependent on a wider social net than that which is typically identified by either traditional literary scholarship (which tends to concentrate on the main characters) or by the social network analysis of fiction, which has so far focused upon statistically central characters. Moreover, the use of the combination window strategy to construct networks of individual chapters provides a more accurate picture of the exchanges of information in broader social networks, which connect characters across disparate groups and are crucial to dramatic development. For example, in Chapter 31 (Fig.4), Eleanor is exposed (via their manservant Thomas) to false rumours about Edward Ferrars's marriage, while in chapter 47 (Fig. 5), a micronarrative reveals both Colonel Brandon's worthiness and Willoughby's corrupt and fickle nature.

Collinear Networks of *Sense and Sensibility*. While Austen's fiction is characterised by being exceptionally well integrated and symmetrical in terms of plot, the perspective provided by a collinear network is nonetheless particularly useful for identifying narrative divergences and asides, such as the features that we have elsewhere termed *micronarratives* [6]. These are less likely to attract comment in more traditional scholarly approaches to Austen's novels, being frequently concerned with less prominent characters; they usually illuminate some aspect of a character's personality, but can also create or strengthen

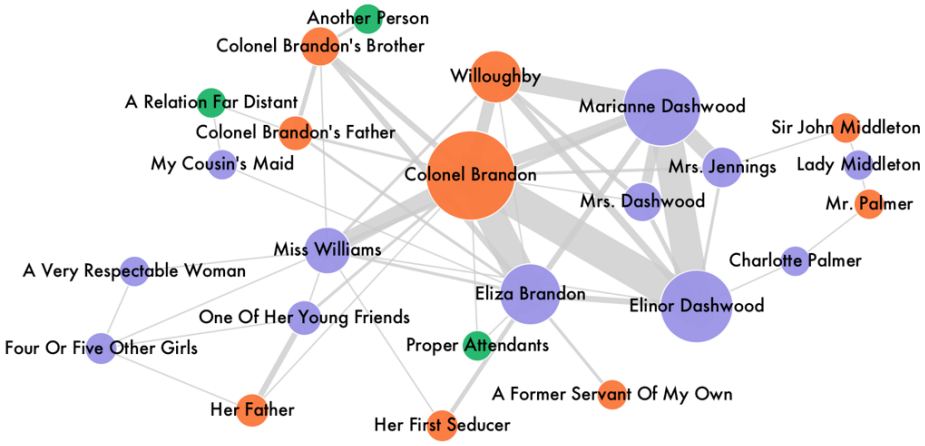


Fig. 6: Overall collinear network of Chapter 31 from *Sense and Sensibility*. Nodes are sized according to weighted degree where larger nodes reflect higher weighted degree values and visa versa. The node colour reflects the gender of the character: purple for female; orange for male; and green for collective or NA.

associations between characters. In Chapter 31, for example, while the combination strategy is useful for illustrating the connectivity between and across social classes, the collinear visualisation is much more illuminating with respect to narrative technique. As can be seen in Fig. 6, Colonel Brandon and Marianne become an adjacent pair in this visualisation as the novel begins to acclimatise us to the prospect of their May-December marriage. In the micronarrative of Brandon's unfortunate ward, Miss Williams, the combination method is more accurate in terms of identifying her social set as distinct from but also connected to the general narrative - and, significantly, the morally dubious Willoughby. By contrast, the collinear approach has long-term potential for analysing the much-commented-upon chain of narrative cause and effect in Austen. In this instance, Colonel Brandon's brother dies, resulting in his going to visit Miss Williams; the brother in this case is not a social link but a narrative one, but in the overdetermined world of Austen's novels, all causality has a moral dimension. The moral inadequacies of the older Brandon brother have contributed to the precarious position in which Miss Williams finds herself; the loyal and virtuous Colonel sets about rescuing her and is vilified and suspected of being her father out of wedlock, but his actions are rewarded when this story gets him closer to marrying Marianne. The collinear chain, then, is at once narrative and moral cause and effect.

The roles of both Thomas and Colonel Brandon in the collinear networks for these chapters is illuminative of an unexpected aspect of gender in the novel: Austen's use of male characters to bring news and hidden histories to the drawing-rooms of her more socially confined heroines, with major narrative consequences. This is an area which is promising for further investigation.

5 Conclusions

In this paper, we have presented three different sliding window strategies that can be employed to capture character associations and generate character networks from literary texts. The text sources that were examined consisted of nine novels from Project Gutenberg, six by the author Jane Austen and three by Charles Dickens. In particular, we focused on collinear, coplanar, and combination methodologies, applying different window sizes in the later two cases to investigate their dependency on this size parameter. Our findings suggest that the choice of strategy is non-trivial, and can have a considerable impact on the resulting character networks. However, it is important to remember that character networks provide an abstract model to be used in conjunction with, rather than in lieu of, more traditional close reading approaches. While computational approaches to the novel to date have tended towards "macroanalysis" [9], microanalysis of two chapters in *Sense and Sensibility* provides a case study demonstrating how combination and collinear approaches can illuminate specific areas of interest in current critical understandings of the novel.

Acknowledgments. The authors would like to sincerely thank and acknowledge the contribution of Dr. Maria Mulvany and Dr. Jennie Rothwell of the Humanities Institute, University College Dublin, in helping to annotate the vast array of characters used in this study.

This research was partly supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289, in collaboration with the Nation, Genre and Gender project funded by the Irish Research Council.

References

1. A. Agarwal, A. Corvalan, J. Jensen, and O. Rambow. Social network analysis of Alice in Wonderland. In *Proc. Workshop on Comp. Linguistics for Literature*, pages 88–96, 2012.
2. A. Agarwal, O. Rambow, and R. J. Passonneau. Annotation scheme for social network extraction from text. In *Proc. 4th Linguistics Annotation Workshop*, pages 20–28, 2010.
3. R. Alberich, J. Miro-Julia, and F. Rossello. Marvel Universe looks almost like a real social network. *arXiv:cond-mat/0202174*, (February 2008):14, 2002.
4. D. K. Elson, N. Dames, and K. R. McKeown. Extracting social networks from literary fiction. In *Proc. 48th Meeting of Assoc. Comp. Ling.*, pages 138–147, 2010.
5. P. M. Gleiser. How to become a superhero. *Journal of Statistical Mechanics: Theory and Experiment*, (09):P09020–, 2007.
6. S. Grayson, J. Rothwell, M. Mulvany, K. Wade, G. Meaney, and D. Greene. Discovering structure in social networks of 19th century fiction. In *Proc. ACM Web Science 2016*, 2016.
7. R. D. Hume. Money in jane austen. *The Review of English Studies*, 64(264):289–310, 2013.
8. P. A. Jayannavar, A. Agarwal, M. Ju, and O. Rambo. Validating literary theories using automatic social network extraction. *Proc. 4th Workshop on Comp. Linguistics for Literature*, pages 32–41, 2015.
9. M. L. Jockers and D. Mimno. Significant themes in 19th-century literature. *Poetics*, 41(6):750–769, 2013.
10. T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.
11. F. Moretti. Network Theory, Plot Analysis. *New Left Review*, 68:80–102, 2011.
12. M. Röder, A. Both, and A. Hinneburg. Exploring the space of topic coherence measures. In *Proc. 8th Int. Conf. Web Search & Data Mining*, pages 399–408, 2015.
13. J. Rydberg-Cox. Social Networks and the Language of Greek Tragedy. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(3):11, 2011.
14. V. A. Traag, R. Reinanda, and G. van Klinken. Structure of an elite co-occurrence network. *arXiv preprint arXiv:1409.1744*, 2014.
15. B. Q. Zadeh and S. Handschuh. Evaluation of technology term recognition with random indexing. In *Proc. 9th Int. Conf. on Language Resources and Evaluation*, pages 4027–4032, 2014.