Social Network Analysis of *Alice in Wonderland*

Apoorv Agarwal[1]* **Augusto Corvalan[1]**** **Jacob Jensen[1]†** **Owen Rambow[2]‡**

[1] Department of Computer Science, Columbia University, New York, NY, USA
[2] Center for Computational Learning Systems, Columbia University, New York, NY, USA
* apoorv@cs.columbia.edu ** ac3096@columbia.edu
†jej2120@columbia.edu ‡rambow@ccls.columbia.edu

Abstract

We present a network analysis of a literary text, *Alice in Wonderland*. We build novel types of networks in which links between characters are different types of *social events*. We show that analyzing networks based on these social events gives us insight into the roles of characters in the story. Also, static network analysis has limitations which become apparent from our analysis. We propose the use of dynamic network analysis to overcome these limitations.

1 Introduction

In recent years, the wide availability of digitized literary works has given rise to a computational approach to analyzing these texts. This approach has been used, sometimes in conjunction with more traditional literary analysis techniques, to better grasp the intricacies of several literary works. As the field matured, new approaches and ideas gave rise to the use of techniques, like social networks, usually reserved for quantitive fields in order to gain new insights into the works. Recently, Elson et al. (2010) extracted networks from a corpus of 19th century texts in order to debunk long standing hypotheses from comparative literature (Elson et al., 2010). Moretti (2011) examined a social event network constructed from Hamlet in order to delve deeper into its infamously dense character network.

While this approach is clearly powerful, it is not without drawbacks. As Moretti (2011) points out, undirected and unweighted networks are blunt instruments and limited in their use. While, as discussed below, some researchers have sought to rectify these limitations, few have done so with a strict and specific rubric for categorizing interactions.

In this paper, we annotate Lewis Carroll's *Alice in Wonderland* using a well-defined annotation scheme which we have previously developed on newswire text Agarwal et al. (2010). It is well suited to deal with the aforementioned limitations. We show that using different types of networks can be useful by allowing us to provide a model for determining point-of-view. We also show that social networks allow characters to be categorized into roles based on how they function in the text, but that this approach is limited when using static social networks. We then build and visualize dynamic networks and show that static networks can distort the importance of characters. By using dynamic networks, we can build a fuller picture of how each character works in a literary text.

Our paper uses an annotation scheme that is well-defined and has been used in previous computational models that extract social events from news articles (Agarwal and Rambow, 2010). This computational model may be adapted to extract these events from literary texts. However, the focus of this paper is not to adapt the previously proposed computational model to a new domain or genre, but to first demonstrate the usefulness of this annotation scheme for the analysis of literary texts, and the social networks derived from it. All results reported in this paper are based on hand annotation of the text. Furthermore, we are investigating a single text, so that we do cannot draw conclusions about the usefulness of our methods for validating theories of literature.

We summarize the contributions of this paper:

- We manually extract a social network from *Al-*

ice in Wonderland based on the definition of *social events* as proposed by us in (Agarwal et al., 2010).

- We use static network analysis (in a bottom-up approach) for creating character sketches. We show that exploiting the distinction between different types of social events (*interaction* and *observation*), we are able to gain insights into the roles characters play in this novel.

- We point certain limitations of the static network analysis and propose the use of dynamic network analysis for literary texts.

The rest of the paper is organized as follows. In Section 2, we present previous work. In Section 3, we present a brief overview of social events. In Section 4, we discuss the data and annotation scheme. In Section 6, we present results on static network analysis, and results on dynamic network analysis in Section 7. We conclude and present future direction of research in Section 8.

2 Literature Review

The power of network analysis in the field of literature is evidenced by the rapid rise of work and interest in the field in recent years. Network extraction and analysis has been performed on subjects as varied as the Marvel universe (Alberich et al., 2002), *Les Misérables* (Newman and Girvan, 2004), and ancient Greek tragedies (Rydberg-Cox, 2011). Elson et al. (2010) has looked at debunking comparative literature theories by examining networks for sixty 19th-century novels. Elson et al. (2010) used natural language processing techniques to attribute quoted speech to characters in the novels, and then used this data to create networks that allowed the researchers to make novel observations about the correlation between setting and the number of characters. Because the study was limited to quoted speech, however, a large chunk of interactions (such as non-quoted dialog, observations and thoughts) were missing from the network and subsequent analysis. Our work specifically addresses these missed cases, and in that sense our technique for creating social networks is complementary to that of Elson et al. (2010).

Several other researchers have found network theory to be useful in the study of literature. In his study of Dicken‘s *Bleak House*, Sack refines the granularity of interaction types by breaking down links by the purpose of the interaction, differentiating between conversations meant, for example, for legal investigation vs. philanthropy. Sack (2006) also expands on the definition of ties, including face-to-face interaction as well as what he terms “weak ties”, which includes interactions like being involved in the same legal suit. His links are a hybrid of quantitative and qualitative. Characters are linked by interaction, but how these interactions are then classified are subjective according to Sack (2006). Thus, they do not follow a strictly defined rubric. Celikyilmaz et al. (2010) have also worked along a similar track, analyzing networks built based on topical similarity in actor speech.

A theorist who has grappled with the limitations of network analysis is Franco Moretti. In *Network Theory Plot Analysis*, Moretti (2011) takes a similar path as Elson et al. (2010), where the act of speech signifies interaction. Moretti (2011) points out that his close reading of the network extracted from *Hamlet* is limited by several factors. First, edges are unweighted, giving equal importance to interactions that are a few words and long, more involved conversations. Second, edges have no direction, which eliminates who initiated each interaction. Moretti (2011) concludes that more rigorous network analysis tools are needed in order to make further headway in the field. In this paper we extract two types networks from *Alice in Wonderland*, one directed and the other undirected, both of which are weighted. We show that indeed discriminating between uni-directional and bi-directional linkages gives us insight into the character profiles and their role in the novel.

Overall, the previous work has primarily focused on turning time into space, flattening out the action in order to bring to light something that was obfuscated previously. However, time and its passage plays a crucial role in literature. Literature is, after all, built in layers, with successive scenes stacking up on each other. Texts reveal information not all at once, like a network, but in spurts. This is not merely an unfortunate side-effect of the medium, but a central element that is manipulated by authors and is central in extracting “meaning” (Perry, 1979).

However, the static social network (SSN) medium itself is not suited to clearly reveal these changes. Dynamic social networks (DSN), on the other hand, can go beyond the summary statistics of SSN. Moreover, because of their flattening effect, SSNs can lead to inaccurate or inexact information (Berger-Wolf et al., 2006). The DSN approach has many applications, from analyzing how terrorist cells evolve over time (Carley, 2003), to mapping the interactions in the writing community (Perry-Smith and Shalley, 2003). One of the obstacles to using DSNs is that they are not as straight-forward to visualize as SSNs. In this paper, we use a visualization outlined in Moody et al. (2005). While the visualization may not be novel, to the best of our knowledge, DSNs have not yet been used to observe networks extracted from literary texts. Our goal is to push beyond the limitations of static network analysis of literature by adding the crucial element it lacks: dynamism.

3 Social Events

A text may describe a social network in two ways: explicitly, by stating the type of relationship between two individuals (e.g. *Mary is John’s wife*), or implicitly, by describing an event whose repeated instantiation may lead to a stronger social relationship (e.g. *John talked to Mary*). These latter types of events are called *social events* (Agarwal et al., 2010). Agarwal et al. (2010) defined two broad types of social events: **interaction** (**INR**), in which both parties are aware of each other and of the social event, e.g., a conversation, and **observation** (**OBS**), in which only one party is aware of the other and of the interaction, e.g., thinking of or talking about someone. An important aspect of annotating social events is taking into consideration the intention of the author: does the author want us to notice an event between characters or is he/she simply describing a setting of a plot? Since our definition of social events is based on cognitive states of characters, as described by the author, we do not annotate a social event in Example (2) below since there is no evidence that either *Alice* or the *Rabbit* are aware of each other. However, in Example (1), there is clear evidence that *Alice* notices the *Rabbit* but there is no evidence that the *Rabbit* notices *Alice* as well. Therefore, there is only a one-directional social event between these entities called the **observation** (**OBS**) event.

1. (1) Then [Alice] {saw} the [White Rabbit] run by her. OBS

2. (2) The [White Rabbit] ran by [Alice]. No social event

Agarwal et al. (2010) have defined finer sub-types of these two coarse types of events. These sub-types include recording physical proximity of characters, verbal and non-verbal interactions, recording if the thought process of thinking about the other entity is initiated by a previous event or by reading a magazine or other social medium. Many of these sub-types are irrelevant for this literary text simply because it does not describe use of technology. There are no emails being sent (which would be a verbal interaction which does not happen in close physical proximity), no one is watching the other on television etc. Therefore, for this paper, we only focus on two broad social event types: **interaction** versus **observation**. For details and examples of other sub-categories please refer to (Agarwal et al., 2010).

4 Data

We annotate an abridged version of *Alice in Wonderland* from project Gutenberg.[1] This version has ten chapters, 270 paragraphs and 9611 words.

Agarwal et al. (2010) trained two annotators to annotate social events in a well known news corpus – Automated Content Extraction (ACE2005, (Walker, 2005)). Once trained, we used one of the annotators to annotate the same events in *Alice in Wonderland*. Unlike the ACE corpus, we did not have previous gold annotations for entity mentions or mention resolution. However, since we are primarily interested only in social events, we instructed the annotator to all and only record entity mentions that participate in a social event.

Since the text is fairly short, the authors of this paper checked the quality of annotations during the annotation process. After the annotation process was complete, one of the authors went over the annotations as an adjudicator. He did not propose deletion of any annotation. However, he proposed adding a

[1] http://www.gutenberg.org/ebooks/19551

couple of annotations for chapter 3 for the *mouse drying ceremony*. In this scene, the *mouse* instructs a group of birds to dry themselves. Lewis Carroll refers to groups of birds using *them, they*. Our annotation manual does not handle such group formations. Do we introduce a *part-of* relation and associate each bird in the group with the group mention (marking the group mention as a separate entity) or not? If yes, and if the group loses one entity (bird in this case), do we mark another group entity and associate the remaining birds with this new group or not? In general, the problem of such groups is hard and, to the best of our knowledge, not handled in current entity recognition manuals. We postpone handling the annotation of such groups for future work.

Another point that the adjudicator raised, which is out of scope for our current annotation manual, is the way of handling cases where one entity interacts with the other but mistakenly thinking that the entity is someone else. For example, the *Rabbit* interacts with *Alice* thinking that she is *Mary Ann*.

5 Social Network Analysis (SNA) metrics

In this section we briefly describe some of the widely used SNA metrics that we use throughout the paper for drawing conclusions about the social network of *Alice in Wonderland*.

Notation: A network or graph, $G = (N, E)$ is given by a set of nodes in the network, N and a set of edges, E. G can be represented as an adjacency matrix A such that $A_{i,j} = I((i, j) \in E)$. Following are the metrics we use:

Degree centrality (Newman, 2010): A node's degree centrality is equal to the total number of its incoming and outgoing edges. The number of connections is often a good proxy for a node's importance.

In-degree centrality (Newman, 2010): Degree centrality, but summing only a node's incoming edges. In the undirected case, this reduces to Degree centrality.

Out-Degree centrality (Newman, 2010): Degree centrality, but summing only a node's outgoing edges. In the undirected case, this reduces to Degree centrality.

Hubs (Kleinberg, 1999): A node's hub score is its element in the largest eigenvector of AA'. This quantifies how well it reliably points to high-scoring authorities. Intuitively, a high Hub score means a good directory of important nodes.

Authorities (Kleinberg, 1999): A node's authority score is its element in the largest Eigenvector of $A'A$. This quantifies how much attention it gets from high-scoring hubs. Intuitively, a high authority score means a node of importance.

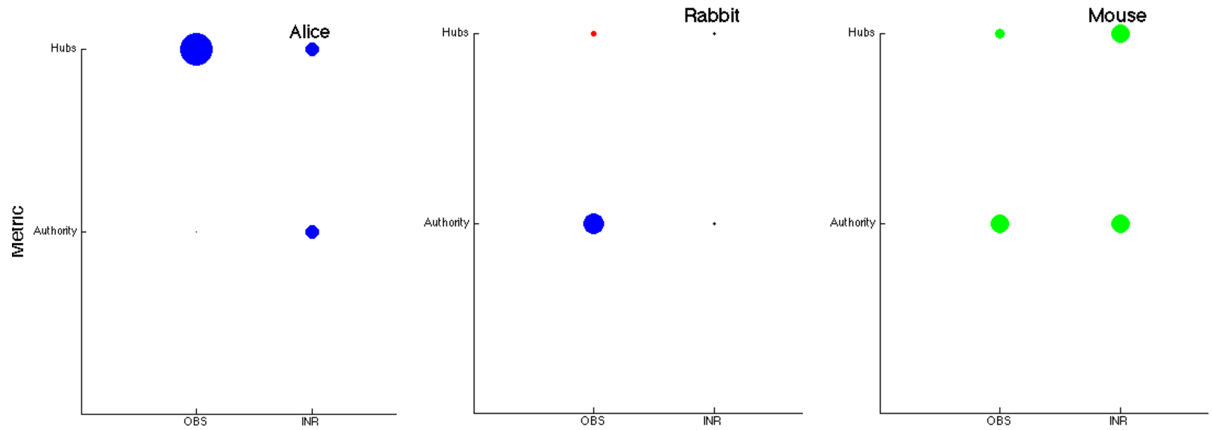
6 Static Network Analysis

In this section we present results for static network analysis of the different types of networks extracted from *Alice in Wonderland*. We use a bottom-up approach. We extract different types of social networks and look at the profiles of characters based on these networks and network analysis metrics. We observe that the profiles of some characters are strikingly different. In this paper, we discuss three characters whose profiles we found most interesting. We are able to show that making a distinction between types of networks based on directionality (who is observing whom) is indeed useful.

6.1 Data Visualization

We calculate hubs and authority weights of all the characters in *Alice in Wonderland*. Since we are using a bottom-up approach, there is a lot of data to look at along different dimensions. We develop a data visualization scheme that makes it easy for us to compare profiles of characters along different dimensions and to compare their profiles with each other.

Following are the different dimensions that we are interested in: 1) type of network, denoted by set $N = \{\text{OBS, INR}\}$, 2) network analysis metric, denoted by the set $M = \{\text{Hub weight, Authority weight}\}$, 3) rank of a character based on type of network and network analysis metric used, denoted by the set $R = \{1, 2, 3, \ldots 52\}$, and 4) absolute separation of consecutively ranked characters for a particular network analysis metric, denoted by a continuous set $S = [0, 1]$. We need this last dimension since one character may be ranked higher than another, yet the separation between the absolute values of the network analysis metric is fairly small. We treat characters with such small separations in absolute values as having the same rank. There are a to-



(a) Alice (b) White Rabbit (c) Mouse

Figure 1: Static networks analysis plots for three characters of *Alice in Wonderland*. X-axis denotes network types, OBS, INR, Verbal and Non-verbal (inorder), Y-axis denotes network analysis metrics, Authority weight and Hub weight. Color coding: Blue = rank 1, Green = rank 2, Red = rank 3 and all other ranks are color Black. Size of the dot is directly proportional to separation from next lower rank, in terms of the network analysis value.

tal of four dimensions for each character, and a total of $2 * 2 * 52 = 208$ data points to look at (ignoring the last dimension, absolute separation from the consecutively ranked character). We represent these four dimensions dimensions in a 2-D scatter plot as follows:

X-axis: We plot the network types along the X-axis.

Y-axis: We plot the network analysis metric along the Y-axis.

Color: Color of a dot denotes the rank of the character. We choose the following color coding. Blue denotes rank one, Green denotes rank two, Red denotes rank three and all the remaining ranks are denoted by color Black. After rank three the absolute value of the metrics plummet and are very close to one another i.e. the separation between absolute values (of network analysis metrics) for consecutively ranked characters is less than 0.001.

Size: The size of a dot denotes the fourth dimension i.e. the absolute separation in network analysis metric of the character under consideration to the next lower ranked character. For example, in Figure 1, rank of the *Rabbit* for network type OBS when looking at the authority weight is 1 and the separation from ranked 2 character, the *Mouse*, is high, as denoted by the larger circle. Alternatively, when looking at rank for *Rabbit* as a hub for network type OBS, he is ranked 3, but there is very little separation between him and the next lowest ranked character.

This visualization enables us to compare a lot of numbers conveniently, out of which arise three interesting character profiles. These profiles yield information as to how each character functions in the story.

6.2 Point-of-View

Alice: Alice has the highest centrality for every network which, using the definition of protagonist given by Moretti (2011), makes her the protagonist of the text. However, from our analysis we are also able to conclude that the story is being told from Alice's perspective. Note that protagonist and perspective-holder are not always the same. For example, *The Great Gatsby* is narrated by Nick Carraway, but the protagonist is Jay Gatsby. Even though to a reader of the text, the perspective holder(s) might be easy to identify, to the best of our knowledge there are no network analysis approaches that can do this. We show that by treating interaction and observation events in isolation, we are able to conclude that *Alice* is the only perspective holder in the story.

The perspective, or point of view, is the "mode (or modes) established by an author by means of which the reader is presented with the characters, dialog, actions, setting and events" (Abrams, 1999). There are four of these:

1. First-Person: The story is being told from the perspective of a narrator that refers to itself as "I" (or "we").

2. Second-Person: Similar to first-person, but the narrator refers to a character(s) in the story as "you". This form of narration is not common.

3. Third-Person Limited: Here, the narrator is not a character in the story, but an outside entity that refers to other characters as "he/she/it/they". However, in limited, this entity is limited to one focal character that the narrator follows.

4. Third-Person Omniscient: A type of third-person narration where the narrator has access to the thoughts and actions of multiple characters.

For first, second and third-person limited, it is expected that the character who is observing other characters is the perspective holder. In order to isolate observations from mentions, the OBS network should be built ignoring quoted speech. Computationally, we believe this would be a fairly easy task. In terms of the terminology we introduce, the perspective holder will have *observation* links pointing to other characters but will not receive *observation* links. In a first-person narration, this character will be an "I" or a name if the "I" is named. The same case for second-person and "you." In third-person limited, while an entity is narrating the story, there is one focal character whose perspective limits and sometimes colors the narration. Thus, that character will still be the one with *observation* links emanating but not receiving. In third-person omniscient, since the narrator has access to every character's thoughts and actions, it is expected that many characters would receive and emanate *observation* links, while there would be an absence of characters who are emanating *observation* links but not receiving any. Therefore, the behavior of perspective holding character is consistent across different types of narrations – it is the character that emanates *observation* type of links but does not receive any. This analysis extends to the case where there are multiple character perspectives being used by seeing which characters are sending but not receiving OBS links and which are not. However, in the rare case where an actor whose point-of-view is being received overhears himself being mentioned, this will be annotated as having him receive a OBS link, thereby throwing off the categorization. We ignore this rare case for now.

Looking at hub and authority weights of *Alice's* **OBS** network (Figure 1(a)), it is apparent that all the *observation* links are pointing outwards from *Alice*. *Alice* is ranked one (color of the dot is blue) and has a high separation from the second ranked entity (size of the dot) for Hub-weight metric. A high hub-weight rank means that most of the links are emanating from this character. In comparison, *Alice's* authority-weight of **OBS** network is low. This means that other characters are not talking about *Alice*. Thus, the story must be being told from the point-of-view of *Alice*.

It should be noted that for concluding who is the perspective holder, it is important to only look at the **OBS** network. The same conclusion cannot be made if we look at the **INR** network. This supports our effort to make a distinction between uni-directional versus bi-directional links.

6.3 Character Sketch for Minor Characters

<u>White Rabbit</u>: The *White Rabbit* has a very different profile when we look at its **OBS** network in comparison to *Alice* (figure 1(b)). *Rabbit* is ranked one but as an authority, instead of as a hub, in the **OBS** network. This means that most of the *observation* links are leading to *Rabbit* i.e. *Rabbit* is being observed or talked about by other characters. On the other hand *Rabbit* is ranked third in **INR** (for which hub and authority have the same value, since INR is non-directional). Thus, *Rabbit* is frequently observed and talked about, yet remains insular in his interactions with other characters. This suggests that *Rabbit* is playing some sort of unique role in the text, where importance is being placed on his being observed rather than his interactions.

<u>Mouse</u>: *Mouse* has yet another kind of profile. For *Mouse*, both hub and authority weights are ranked two and have a clear separation from the next ranked character. We may observe that *Mouse* not only interacts with many characters, but mentions and is

mentioned in abundance as well. This makes him a very important and well-connected character in the story, behind only *Alice*. Thus, we can suggest that his role in the text is as a connector between many characters. *Mouse* mentions many characters to other characters, interacts with them and is in turn mentioned by them.

6.4 Need for Dynamic Analysis

The need for a dynamic analysis model is made clear in the case of *Mouse*. His huge importance (overshadowing more traditionally popular characters such as *the Queen* and *Mad Hatter*) was an unexpected result. However, this is not the whole story: *Mouse* actually only appears in one scene in chapters 2-3. In the scene, Alice has created a large lake with her tears and meets *Mouse*, who introduces her to many minor characters during a drying ceremony. Outside of this ceremony, *Mouse* does not reappear in the text. This one scene, while important, should not be enough to overshadow characters such as *the Queen*, who is responsible for Alice's life or death during the climax of the text. Thus, it is clear from the formation of these character profiles that certain information is being skewed by static network analysis. Most notably, the importance of time as it flows in text is being lost. This observation is the impetus for a new model that addresses these issues, as outlined in the following section.

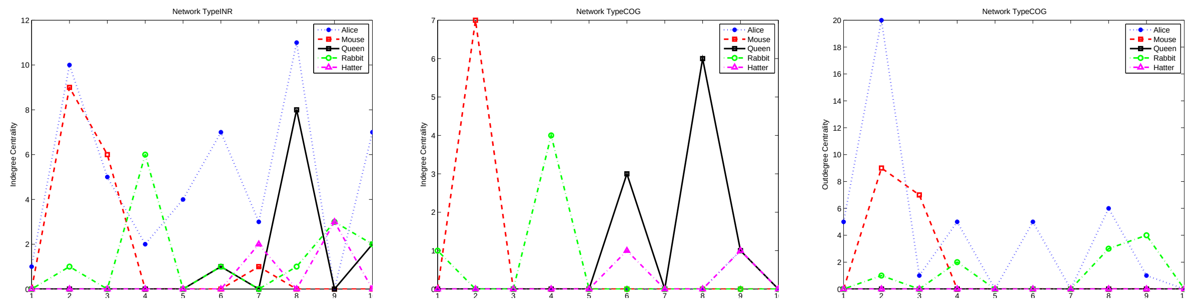
7 Dynamic Network Analysis

Figure 2 presents plots for dynamic network analysis of the different types of networks extracted from *Alice in Wonderland*. We look at interaction (INR) and observation (OBS) networks, as we did for the previous section, except we do this for each of the 10 chapters independently of all other chapters. The social network metrics we consider are: degree, in-degree and out-degree centrality. Note that for an undirected network (i.e. INR), all three network analysis metrics are the same. In this section we present insights about the three characters considered in the previous section (Alice, Mouse and Rabbit), that are lost in static network analysis.

From Figure 2, it is clear that *Alice* (dotted blue line) is *not* the most central character in every chapter, something that is lost in the static network. Consider figure 2(a) i.e. degree centrality of INR network. *Alice* ranks 2 in chapters 3, 4 (the drying ceremony mentioned above) and 9. In chapter 9, *Alice* is overshadowed by *The Hatter* and *Rabbit*. This makes sense, as this chapter concerns *Rabbit* and *The Hatter* being witnesses at *Alice's* trial. By breaking the story down chapter by chapter like this, it becomes evident that although *Alice* is a very active character throughout, there are moments, such as the trial, where she is inactive, indeed powerless. Yet as soon as the trial is over and *Alice* is back in her own world in chapter 10, we see a spike as she again takes an active role in her fate.

Figure 2(b) shows in-degree centrality for the OBS network. This represents how often a character is thought about or talked about by another character. Notice that *Alice* is completely absent in this network: no one thinks about or mentions her. This is to be expected, as *Alice* is our guide through Wonderland. No one mentions her because she is present in every scene, thus any dialog about her will become an interaction. Likewise, no one thinks of her because the reader is not presented with other character's thoughts, only *Alice's*. This is consistent with earlier observations made in the static network. Interestingly, *Queen* (solid black line) comes to dominate the later chapters, as she becomes the focus of *Alice's* thoughts and mentions. Again, this spike in *Queen's* influence (Figure 2(b)) is lost in the static network. But it is *Queen* who ultimately has the power to decide the final punishment for *Alice* at the end of the trial, so it is fitting that *Alice's* thoughts are fixated with her.

Figure 2(c) shows the out-degree centrality of the OBS network, a starkly different picture. Here, we see why *Mouse* (dashed red line) has such importance in the static network. Over the course of the drying ceremony in chapter 2 and 3, he mentions a very large number of characters. The dynamic network allows us to see that while *Mouse* does play a key role at one point of the story, his influence is largely limited to that one section. Other characters overshadow him for the rest of the text. Comparing *Mouse's* role in the in-degree centrality graph (figure 2(b)) vs. out-degree centrality (figure 2(c)), we can see that much of *Mouse's* influence comes not from entities referring to him (in-degree), but rather the number of entities he mentions. His importance



(a) Degree centrality measure for INR network (b) In-degree centrality measure for OBS network (c) Out-degree centrality measure for OBS network

Figure 2: Dynamic network analysis plots for all 10 chapters of *Alice in Wonderland*. Each plot presents the change of centrality values (Degree, In-degree, Out-degree) in different types of network (INR and OBS). X-axis has the chapter numbers (one through ten) and Y-axis has the value of the relevant centrality measure.

in the piece, then, appears to be isolated to a key chapter where he acts as a guide to introduce many entities to the reader.

Likewise, tracing *Rabbit* (dash-dotted green line) across in- and out-degree centrality of the OBS network (figure 2(b) and 2(c)) gives a more fine-grained view of how he works in the text. He is the most mentioned in chapters 1 and 4, chapters that sandwich a big event, the drying ceremony of chapters 2 and 3. Likewise, he reemerges for another big event, Alice's trial (chapter 8, 9, 10). As previously mentioned, *Queen* is the primary concern in *Alice's* mind during the length of the trial. However, *Queen* is absent from the out-degree graph–she makes no reference to off-screen characters. *Rabbit*, who has a large spike in out-degree links during these chapters, is the one who actually mentions a large number of characters, while *Queen* focuses on interacting with those already present. Thus, *Rabbit* is a character that concerns Alice during large set-pieces, one whose primary purpose comes in noticing and being noticed.

We see that using a dynamic network can provide a more subtle view than using a static network. Characters who are key in certain sections are no longer overshadowed, like *Queen*, nor are their importance exaggerated, like *Mouse*. It can also provide us with a better view of when and how a protagonist is most important throughout the text. Finally, analyzing across data dimensions can provide a very specific idea of how a character is functioning, as seen with *Rabbit*.

8 Conclusion

In this paper we have motivated a computational approach to dynamic network analysis. We have hand-annotated Lewis Carrol's *Alice in Wonderland* using a strict and well-defined annotation scheme and created social event networks from these annotations. From these, we have shown the usefulness of using different types of networks to analyze different aspects of a text. We derive point-of-view from a social network. We also break down important characters into certain roles that describe how they function in the text. Ultimately, we find that these roles are limited by the static nature of social networks and create dynamic networks. From these, we extract a clearer picture of how these roles work, as well as other characters overshadowed in the static network. Having shown the value of such analysis, future work will focus on adapting our computational model (Agarwal and Rambow, 2010) for extracting social events from a different domain (news articles) to this new domain (literary text). We will then investigate a large number of literary texts and investigate how we can use our machinery to empirically validate theories about literature.

Acknowledgments

We would like to thank three anonymous reviewers for very useful comments and suggestions, some of which we intend to pursue in future work. This work is supported by NSF grant IIS-0713548.

References

M.H. Abrams. 1999. *A Glossary of Literary Terms*. Harcourt Brace College Publisher.

Apoorv Agarwal and Owen Rambow. 2010. Automatic detection and classification of social events. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*.

Apoorv Agarwal, Owen C. Rambow, and Rebecca J. Passonneau. 2010. Annotation scheme for social network extraction from text. In *Proceedings of the Fourth Linguistic Annotation Workshop*.

R. Alberich, J. Miro-Julia, and F. Rossello. 2002. Marvel universe looks almost like a real social network. *eprint arXiv:cond-mat/0202174*, February.

Berger-Wolf, Tanya Y., and Jared Saia. 2006. A framework for analysis of dynamic social networks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pages 523–528, New York, NY, USA. ACM.

K. M. Carley. 2003. Dynamic network analysis. In R. Breiger, K. M. Carley, and P. Pattison, editors, *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pages 133–145, Washington, DC.

Asli Celikyilmaz, Dilek Hakkani-Tur, Hua He, Greg Kondrak, and Denilson Barbosa. 2010. The actor-topic model for extracting social networks in literary narrative. *Proceedings of the NIPS 2010 Workshop – Machine Learning for Social Computing*.

David K. Elson, Nicholas Dames, and Kathleen R. McKeown. 2010. Extracting social networks from literary fiction. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 138–147.

Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, September.

James Moody, Daniel McFarland, and Skye Benderde-Moll. 2005. Dynamic network visualization. *American Journal of Sociology*, 110(4):1206–1241, January.

Franco Moretti. 2011. Network theory, plot analysis. *New Left Review*.

M. E. J. Newman and M. Girvan. 2004. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(2), February.

Mark Newman. 2010. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA.

Jill E. Perry-Smith and Christina E. Shalley. 2003. The social side of creativity: A static and dynamic social network perspective. *The Academy of Management Review*, 28(1):89–106.

Menakhem Perry. 1979. Literary dynamics: How the order of a text creates its meanings [with an analysis of faulkner's "a rose for emily"]. *Poetics Today*, 1(1/2):35–361, October.

Jeff Rydberg-Cox. 2011. Social networks and the language of greek tragedy. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(3).

Alexander Graham Sack. 2006. Bleak house and weak social networks. *unpublished thesis, Columbia University*.

Christopher R Walker, 2005. *ACE (Automatic Content Extraction) English Annotation Guidelines for Events Version 5.4.3 2005.07.01*. Linguistic Data Consortium.