



# Measuring centrality in film narratives using dynamic character interaction networks

**DOI:**

[10.1016/j.socnet.2020.03.003](https://doi.org/10.1016/j.socnet.2020.03.003)

**Document Version**

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**

Jones, P., Quinn, E., & Koskinen, J. (2020). Measuring centrality in film narratives using dynamic character interaction networks. *Social Networks*, 63. <https://doi.org/10.1016/j.socnet.2020.03.003>

**Published in:**

Social Networks

**Citing this paper**

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



# Measuring centrality in film narratives using dynamic character interaction networks<sup>1</sup>

Pete Jones<sup>1\*</sup>, Eithne Quinn<sup>2</sup>, and Johan Koskinen<sup>3</sup>

<sup>1</sup> School of Social Sciences, University of Manchester, United Kingdom.

<sup>2</sup> School of Arts, Languages and Cultures, University of Manchester, United Kingdom.

<sup>3</sup> Melbourne School of Psychological Sciences, University of Melbourne, Australia, and The Institute of Analytic Sociology, Linköping University, Sweden.

\*Corresponding author. E-mail: [pete.jones@manchester.ac.uk](mailto:pete.jones@manchester.ac.uk)<sup>2</sup>.

*Accepted for publication 26/3/2020.*

## Abstract

The tools of social network analysis offer a promising framework for studying fictional texts and the relational activity of the characters therein. The goal of this paper is to offer both a conceptual refinement of the project of measuring the centrality of characters within narratives using network tools, as well as the proposal of a novel measure with which to do so. Conceptually, we argue that as questions of time, order and sequence are central in narratives, measures of characters' narrative importance should be based on dynamic network representations which respect the time-ordering of narrative events. We suggest a directed dynamic measure of relative character importance based on character interactions and illustrate it through an examination of gender in the 2015 film *Star Wars: The Force Awakens*. We find that the measure helps illuminate important narrative dynamics which cannot be captured by static measures, and presents a platform on which future character network research can productively build.

Keywords: Character networks, Dialogue networks, Dynamic centrality, Narrative, Interactions, Gender

---

## 1. Introduction

This paper aims to develop a character interaction-based approach to investigating the positions of characters within film narratives. As we outline in the following section, several studies now exist which recognise the potential of network analysis for offering insights into the study of fictional narrative texts and the characters therein (e.g. Moretti 2011; Sack 2014; Waumans et al. 2015). These studies range from in-depth textual readings of particular texts to computational interventions in literary network extraction with little to no literary analysis. What the studies all share is the assumption that network tools might assist in the analysis of narrative texts, an assumption largely born out of the growing field of digital humanities scholarship. In this paper, we aim to contribute to the literature in this area by intervening in the discussion of how we relate node-level network measures to interpretative readings of characters' positions within narratives. Engaging with key ideas from narratology and narrative comprehension scholarship, we argue that the dominant approach of applying off-the-shelf centrality measures such as degree and betweenness to aggregated static character networks limits the usefulness of the character networks approach because of the amount of narrative

---

<sup>1</sup> Published version available from <https://www.journals.elsevier.com/social-networks>.

<sup>2</sup> Part of this work by the first author was completed as a Visiting Researcher to the Social Networks Lab in the Melbourne School of Psychological Sciences.

information that is lost through aggregation. Instead, we draw on recent developments in representing and analysing temporal network dynamics to develop a dynamic measure of the relative narrative positions of characters through time (Broccatelli et al. 2016; Everett et al. 2018; Moody 2002). Our contention is that dynamic approaches such as the one we pursue here are needed in order to capture the way the story is told through its characters. Thus, we contribute both a conceptual refinement to the project of measuring narrative centrality in character networks, as well as the proposal of a novel measure for doing so. These contributions are offered as exploratory steps towards a more analytically productive platform for future network-based study of narrative texts.

The structure of the paper is as follows. In the next section, we provide an outline of existing approaches to analysing fictional texts through character network models. Then, we consider the problem of how to say something meaningful about the relative importance of nodes within the narrative, focusing on how to appropriately represent narratives as network data given the importance of sequence in narratives. We suggest a number of principles for measuring narrative centrality, and outline a method for configuring film texts as character interaction networks based on the distribution of lines of dialogue between characters. Following this discussion, we propose a new approach to measuring the narrative positions of characters in fictional texts and define a dynamic relational measure based on the acts of speaking and being spoken to. As well as offering a conceptual modelling framework for analysing the relative importance of characters in a narrative, some formal properties of the proposed measure are further explored in the appendix. Because of the dynamic nature of the measure, only some properties are analytically tractable, and we cannot provide an exhaustive formal analysis within the scope of this paper. However, the analysis of idealised examples we present could be complemented by simulations in the future. Finally, we offer an application of this measure to the 2015 film *Star Wars: The Force Awakens*, drawing out the gendered dynamics in the film. This case study illustrates some of the things the measure allows us to observe about the ways in which characters relate to the narrative that cannot be inferred from static measures. In particular, through a focus on the vocal and relational disempowerment of women in film narratives, we show that the approach allows us to use character network data to engage with existing debates in the study of social representation in fictional narrative texts in a way which previous character network approaches cannot. We argue that by focusing on the conceptual links between narrative texts and the way we represent them as data, network-based measures such as this have the potential to contribute more to character-oriented scholarship on those texts than existing approaches have been able to offer.

## 2. Character networks

A character network is any network representation in which the nodes are characters in a fictional text. There now exists a number of character network-based studies of literary texts (e.g. Agarwal et al. 2012; Elson et al. 2010; Grayson et al. 2016; Kydros et al. 2015; Jayannavar et al 2015; Min and Park 2016; Moretti 2011; Sack 2014; Waumans et al. 2015). These literary character network studies have primarily been concerned with questions of network topology, the process of computational extraction of networks from literary texts, and validating existing literary and network theories using extracted character network data. Though these studies vary in tie definition, scope and audience, one of the common threads they share is their recognition that the analytical tools of social network analysis can prove fruitful for answering questions of substantive interest to scholars of fictional cultural texts. For example, Elson et al. (2010) use network analytical tools and concepts such as clique detection, density and network size to explore hypotheses related to literary theory on the relationship between urbanity/rurality and character interactions in Victorian novels. Network tools have also been applied to literary texts at the node level through the application of centrality measures to the characters in a text. In practice, this means calculating common measures such as degree, betweenness and closeness for nodes in the character network in order to make claims about the relationship between those characters and

the narrative (e.g. Kagan et al. 2019; Kydros et al. 2015; Moretti 2011), an approach we discuss further in the next section.

Despite the growing interest in this approach from a literary perspective, the use of character networks to study film has taken a different tack. Work in this area has been concerned with methods for large-scale computational extraction of networks from film corpora rather than questions of how a network-based approach can aid the reading and analysis of particular film texts. These studies are thus conducted from a primarily computational standpoint, aiming to develop tools that extract networks automatically from scripts (Agarwal et al. 2014; Gil et al. 2011; Mouchid et al. 2018), video (Lv et al. 2018; Weng et al. 2009) or subtitle files (Kagan et al. 2019). These automatic extraction methods require the construction of undirected ties based on some definition of co-presence within a scene rather than direct interaction between characters. As such, one of the compromises involved in pursuing a more computational character networks approach is that by defining ties based on scene-sharing we lose the capacity to answer many character-oriented questions such as how much the characters speak or to whom.

It is useful to contrast the character networks approach with the related but largely disjunct body of work concerning “narrative networks” (e.g. Abell 2004; Bearman and Stovel 2000; Padgett 2018; Smith 2007). Work on narrative networks has broadly aimed at mapping the temporal relations between narrative events to try to build a picture of how people construct identities and shared meanings through the assembly of historical events into narratives. As such, the narrative networks literature takes a different conceptual approach, as it is primarily concerned with understanding narrativity as a dimension of social life, rather than the study of fictional narrative texts. Some of the insights of this work, such as the importance of time for understanding narrative phenomena, are directly relevant to the development of character networks scholarship and will be drawn on in the next section. As there are some differences in aims, objectives and methods pursued in the two approaches, it might be useful to separate character networks and narrative networks as two distinct sub-fields of network research.

### 3. Relating network positions to narrative positions

One of the ways in which network tools can be brought to bear on the study of narrative texts is through the application of measures of node importance to the characters in a text in order to explore existing debates and theories concerning issues such as visibility, representation and the relationship between characters and narratives. The strategy that has been pursued in existing character networks approaches has been to apply standard network centrality measures (such as degree and betweenness) in order to make claims about the relative positions of characters within narratives. For example, Moretti (2011) looks at the degree centrality of characters in *Hamlet*, using binary and undirected data in which a tie represents the fact that any amount of words flowed between two characters. He uses this network data to claim, for example, that Claudius is the “second most central character in the play” because he interacts with 13 characters, one more than Horatio and two fewer than Hamlet (Moretti 2011, 88). Kydros et al. similarly present degree, betweenness, and closeness measures for characters in Homer’s *Iliad*, though they note that interpretation of the results is “not so easy” (2015, 125). In this section, we argue that the utility of this approach is limited because it is based on network representations that do not adequately capture narratives in ways that allow for meaningful conceptualisation and interpretation of the relationships between characters and narratives.

#### *Network representations*

Brandes et al (2013) note that there are two steps to obtaining a network model: first, there must be an appropriate abstraction of a phenomenon to a network concept; then, there must be an appropriate representation of that network concept in network data. In the case of character networks, the phenomenon in question is a narrative text. Thus, the first step is to abstract the narrative text to a network concept. As Moretti (2011) has noted, Alex Woloch (2003) offers a compelling basis for so doing in his work on major and minor

characters in the literary novel. Woloch argues that what we see as major and minor characters is the product of the “character-space” that those characters occupy within a broader “character-system” (Woloch 2003). In narratives, our attention is a finite resource which can only be allocated to certain narrative elements at once. Thus, we can think of narratives as a distribution of attention, as “[n]arrative meaning takes shape in the dynamic flux of attention and neglect toward the various characters who are locked within the same story but have radically different positions within the narrative” (Woloch 2003, 2).

Moreover, this conceptualisation of narrative texts as a dynamic distribution of attention through a character system is explicitly relational in Woloch’s theory, as he argues that “the emplacement of a character within the narrative form is largely comprised by his or her relative position vis-a-vis other characters” (Woloch 2003, 18, *italics in original*). Extending these ideas to more visual media such as film finds resonance with much of the theory on film narratives, where it has been noted that the technical practices of mainstream Hollywood film production were designed “to guide audience attention to salient narrative events from moment to moment” (Thompson 1999, 1). Work in this area has aimed to pair an analysis of film narratives with a deep understanding of narrative comprehension and how viewers process visual narratives, and has emphasised the role of relations between narrative elements (including characters) in the process. For example, David Bordwell’s theory of narrative states that “comprehending a narrative requires assigning it some coherence. At a local level, the viewer must grasp character relations, lines of dialogue, relations between shots, and so on” (Bordwell 1985, 34). Edward Branigan notes how these local events feed into a “global interpretation of changing data measured through sets of relationships” (Branigan 1992, 4). Moreover, Branigan argues that psychological experiments have shown that “an individual’s attention does not spread equally through a narrative text but works forward and backward in an uneven manner in constructing large-scale, hierarchical patterns which represent a particular story as an abstract grouping of knowledge” (Branigan 1992, 16). Thus, we can think of film narratives as dynamic patterns of character-driven events from which we construct these abstract relational groupings of knowledge.

These ideas suggest that thinking of narratives as systems of character relations wherein our attention as readers or viewers of the text is distributed unevenly into hierarchical patterns seems like a useful way of looking at a narrative text. However, the second step in obtaining the network model (the representation of this network concept as network data) is more fraught because existing character networks approaches typically represent the character system using static, aggregated network data. A brief consideration of how narratives have been defined and theorised will illustrate why this kind of network representation limits the utility of such approaches.

### *Narrative as dynamic*

Narratologists have long recognised that the narrative is by definition a dynamic and temporal phenomenon, and the importance of temporal sequence is fundamental to almost all definitions of narrative (Abbott 2008; Bal 2017; Chatman 1980; Genette 1980; Ricoeur 1980). One of the ways that narrative is frequently defined is through its relation to story. Story is usually defined as the events that take place, while narrative is defined as how that story is told. This opens up the question of story discourse versus narrative discourse and, crucially, story time versus narrative time (Chatman 1980; Genette 1980). Seymour Chatman (1980) theorised this as the “double time structuring” of narrativity – our engagement with a narrative asks us to understand the way in which story events are ordered even while those events may be told to us in a completely different order. For example, a simple story might involve 3 events: A = my alarm failed to go off; B = I missed the bus; and C = I was late to work. However, I might narrativise these story elements in a different order (C-A-B) by telling the story through the following narrative: “I was late to work because my alarm didn’t go off which made me miss the bus.” What is important here is the recognition that when we talk about narrativity, we are talking about sequence and time. Work on narrative comprehension has argued that “the appearance and ordering of story elements in the [narrative] discourse is critical to the process of comprehension” (Niehaus and Young 2014, 561), and that the decision to tell a story one way or another affects the ways in which we make sense of the story and the importance of the elements therein (Bordwell 1985; Branigan 1992). Therefore, network measures which are

invariant under permutations of sequence are missing a very central aspect of what makes a narrative a narrative.

The importance of sequence and time to the analysis of narratives has been integral in the literature on narrative networks, as noted in Section 2. For example, Padgett (2018) presents a complex and sophisticated model of how William Faulkner uses narration to build multiple dimensions of time and memory assembly into his novel *Absalom, Absalom!*, relating this to cognitive identity-construction processes. However, the character networks literature has paid much less attention to these temporal considerations, and has largely continued to treat narrative texts as static networks. For example, Moretti's *Hamlet* study (2011) defines a static tie between nodes indicating that any amount of dialogue passed between two characters, while Elson et al. (2010) and Waumans et al. (2015) define static ties based on the identification of adjacent attributed speech for two characters within the novel's text. However, as Branigan notes, "recognizing the complexity and dynamism of a text is usually more important than assigning a final, decisive label to it" (Branigan 1992, 8). Static character network representations are therefore unlikely to offer satisfactory models of fictional texts as they compress and flatten narratives through aggregation such that narrative time is lost in the representation<sup>3</sup> (Moody et al. 2005). Even when the frequency or amount of character interaction is retained through tie weights (e.g. as in Masías et al. 2017; Ruegg and Lee 2019), the temporal sequence of these interactions is lost. As narratologist Gérard Genette has argued, this elimination of time from the representation "is not only not sticking to the text, but is quite simply killing it" (Genette 1980, 35).

We argue that we should therefore think instead of narrative character networks as dynamic networks, as defined by Kontoleon et al. (2013). As with most other cases where we are interested in representing dynamic networks based on relational activity rather than fixed, stable relations, this has direct implications for our ability to capture centrality concepts using network analytical measures (Falzon et al. 2018; Moody 2002). Node-level measures based on a static network representation give us a static measure of that node's narrative importance. However, it does not make sense to talk about a character's position within a narrative as a fixed quality because character positioning evolves dynamically over the course of a narrative. A character can occupy many different narrative positions throughout the course of a narrative, and the dynamics of this ought to be of interest in any analysis of that character's narrative importance.

### *Principles for measuring relative narrative importance*

The previous discussion allows us to specify some principles which an appropriate measure of the relative importance of nodes in a narrative character network ought to satisfy. Firstly, such a measure ought to be dynamic and based on a temporally disaggregated network representation of the narrative text. Secondly, such a measure ought to be linked to a character's narrative activity, in order to satisfy the idea of narratives as a distributed field of attention between characters. When we focus on one character, we always do so at the expense of others, and so we assume that narrative activity is always an indicator of narrative importance; the more a character does something, the more attention they receive, the more character space they take up in the narrative, and the more information we have with which to "qualify" them as characters (Bal 2017). Thirdly, such a measure ought to take into account their position in the relational structure. That is, our sense of a node's relative importance (based dynamically on their narrative activity) should be dependent to some degree on the relative importance of the nodes with which they interact. David Bordwell has argued that when watching a film, one cognitive activity of the viewer is "looking for relevance, testing each event for its pertinence to the

---

<sup>3</sup> It should be noted that a few papers have analysed character networks from the perspective of longitudinal networks (Agarwal et al. 2012; Fischer et al 2017; Min and Park 2016; Prado et al. 2016). They analysed the networks as a series of snap-shots by aggregating character interactions within each chapter and then calculating static network statistics (including node-level measures such as degree) for each chapter. Thus, though such approaches aim to discuss narratives in terms of temporal dynamics, many of the points made here about the appropriateness of static network measures apply also to these papers as a great deal of information about the frequency and sequence of events within chapters is still lost through aggregation.

action which the film (or scene, or character action) seems to be basically setting forth" (1985, 34). Based on this, we assume that a character is more likely to be meaningfully engaging with the narrative if they are interacting with a character who has already occupied our attention than if they interact with a character we do not know yet. The third principle also ensures that such a measure will be relational, as we could easily satisfy the first two principles by tracking character activity (such as lines spoken) through time without using any relational information from the network.

## 4. Defining a measure of narrative positioning

### *Data representation*

In order to operationalise the ideas in the previous section, we assume a dialogue-based character interaction network model designed to investigate questions of vocal and relational (dis)empowerment in Hollywood narratives (see Jones 2018). Unlike existing approaches to character networks which tend to deploy automatic extraction methods based on textual proximity (discussed in Section 2), the manual approach proposed here constructs ties based on direct interaction between characters, as we cannot be sure from scene co-presence or speech adjacency that interaction between characters actually occurred, and thus we cannot learn anything about the pattern, frequency, direction or sequence of the interactions between characters in the narrative. We take the units of analysis in our representation to be the vocal interactions between characters, an approach which allows us to measure the amount of vocal activity of characters, as well as the interpersonal structures that emerge among characters as a result of these patterns of interaction.

Thus, for a given film, a network can be constructed by recording each line of dialogue between characters. A line of dialogue here is defined as a single continuous stream of speech; a line ends when the dialogue pauses to allow for a response, or the scene or topic changes notably. To be included in the network, a character must both speak and be named, and only intelligible lines of dialogue spoken by and to named speaking characters are recorded. The ties in such a network are directed in that each line of dialogue has a distinct sender and receiver(s). This is important as there are valuable distinctions in speaking and being spoken to.

These interactions can be represented as a static, aggregated network graph, with the frequency of interaction between characters being retained as a tie value or weight, allowing for simple visualisation of the network. However, for analytical purposes the basic data structure is the time-ordered and disaggregated list of interactions, the "raw materials" from which the static graph-based representation may be constructed (Amati et al. 2019). Thus, although this manual network extraction method is labour-intensive in comparison with computational approaches, it results in rich directed character interaction data which can serve as something of a gold standard when considering what other approaches are able to capture. In the conclusions, we discuss the ways in which other approaches based on automatic network extraction can satisfy the principles set forth here.

An additional distinction to make in establishing this data representation relates to the extent to which the dynamic network has concurrent relations or not. We can distinguish between concurrency across actors and within actors. For the cases considered by Moody (2002) and Kontoleon et al. (2013), ties may be concurrent across actors in the sense that there may be overlap between ties between distinct pairs, for example two distinct pairs may be in sexual relationships or in concurrent phone calls. An example where actions are not concurrent across actors is passing the ball in a football match. From a narrative perspective, concurrency across actors, while in principle possible, does not make narrative sense. Concurrency within actors has two separate dimensions that we could call *sender exclusivity* and *receiver exclusivity*. The call network of Kontoleon et al. (2013) has both sender and receiver exclusivity in the sense that it is assumed that an actor can only be in a phone conversation with one person at a time. In Moody's (2002) network there is neither sender nor receiver exclusivity, something which is deemed crucial for disease transmission. The dialogue-based character interaction network representation we describe here assumes sender exclusivity but not necessarily receiver

exclusivity. These narrative considerations taken together suggest a measure with a different aim than to summarise the temporal-topological structure as in Holme and Saramäki (2012) and Kontoleon et al. (2015).

Based on this data representation, we propose the cumulative measure  $C_{i,t}$  which represents the score  $C$  of node  $i$  at time  $t$ . The narrative activity referred to by the second principle specified in Section 3 can be approximated using the character interaction data outlined above such that vocal interactions are the basis of our understanding of a character's relative importance within the narrative. However, in order to distinguish between speaking and being spoken-to, we define two closely-related versions of the measure: the speaking measure  $C_{i,t}^{out}$ , which tells us the speaking score  $C^{out}$  of node  $i$  at time  $t$  and bases relative character importance on speaking activity; and the spoken-to measure  $C_{i,t}^{in}$ , which tells us the spoken-to score  $C^{in}$  of node  $i$  at time  $t$  and bases relative character importance on reception of dialogue. This allows for consideration that the narrative importance conveyed through speaking activity may be different from the narrative importance conveyed by being spoken to.

### *Definition of the speaking measure*

Defining the measure at each time  $t$  ensures satisfaction of the first principle, which dictates that the measure should be calculated at a level of temporal disaggregation so that character scores are specific to each measured point of time. Using the time-ordered interaction data, time  $t$  is equivalent to a line of dialogue, with lines being temporally ordered in a sequence of positive integers, such that the first line of dialogue takes place at  $t_1$ , the second line at  $t_2$ , and so on.

The second principle suggests that scores should increase upon narrative activity which, in the case of the dialogue network data, is a vocal interaction. The measure should therefore be cumulative, increasing monotonically when characters speak, and the third principle suggests that the scale of this increase should depend on the character(s) being spoken to. We can satisfy each of the three principles above by following a similar approach to that taken in Everett et al. (2018). These authors present a centrality-like measure (based on a dynamic two-mode network representation called the bi-dynamic line graph) in which nodes accumulate points through network activity, and the amount of points gained by a node in each temporally-situated activity is dependent in part on the points held by the other nodes also participating in that activity.

Following a similar logic, we propose letting speaking character  $i$ 's score increase by a fraction of the score of character  $j$  being spoken to at time  $t$ . This fraction can be defined using a weighting parameter  $\lambda$ , such that

$$C_{i,t}^{out} = C_{i,t-1}^{out} + C_{j,t-1}^{out}\lambda. \quad (1)$$

Thus, the amount that  $i$ 's score increases depends on the score of character  $j$ , ensuring that the measure is relational, as well as on the value of  $\lambda$ , allowing us to specify *how much* we want the relational aspect to contribute to our sense of  $i$ 's importance.

The measure has a flavour of how eigenvector centrality is defined on static graphs. Holme and Saramäki (2012) develop a dynamic node-level measure that respects the time-ordering of interactions explicitly to follow the logic of eigenvector centrality. However, their measure does not take into account the directionality of interactions and, because of its relation to eigenvector centrality, it is not non-decreasing, and thus does not satisfy the second principle above as some interactions can lower one's centrality score. This property could have interesting implications in other contexts where researchers may be interested in investigating how talking to the 'wrong' people could have adverse effects on your position in the dialogue network. However, this is not appropriate for narrative character networks given our argument that narrative activity always increases the amount of character space taken up by a character.



The measure as specified above would be sufficient for a case where a line of dialogue has only one recipient. However, as we do not require receiver exclusivity, the measure may be modified to account for multiple recipients. Simply adding a fraction of each spoken-to character  $j$ 's score to  $i$ 's score would mean that lines with multiple recipients have a larger effect on  $i$ 's score than lines with only one recipient. However, lines addressing multiple people are not inherently more narratively important than lines addressing one person, so we propose instead multiplying  $\lambda$  by the average score of the spoken-to characters. Thus, let  $x_{ij,t}$  be a variable that can take either the value 1 or 0 to indicate whether a character  $j$  is spoken to by character  $i$  at time  $t$ . Then the total number of characters  $j$  spoken to by speaking character  $i$  is given by  $\sum_j^n x_{ij,t}$ , where  $n$  is the total number of nodes in the network. The sum of all scores for spoken-to characters  $j$  is given by  $\sum_j^n C_{j,t-1}^{out} x_{ij,t}$ .

The speaking measure can then be more generally defined as

$$C_{i,t}^{out} = C_{i,t-1}^{out} + \frac{\lambda}{\max(\sum_j^n x_{ij,t}, 1)} \sum_j^n C_{j,t-1}^{out} x_{ij,t}, \quad (2)$$

which tells us that character  $i$ 's relative importance at line  $t$  is equal to whatever  $i$ 's importance was at line  $t - 1$  plus a fraction of the importance of the character(s) to whom they are speaking<sup>4</sup>. For characters not speaking at line  $t$ , (2) will simply imply that they retain their score from  $t - 1$  as  $x_{ij,t}$  will be equal to zero for these characters. Thus, so long as characters' scores at time  $t_0$  (the starting scores before a line has been spoken) take positive nonzero values, the measure is monotonically cumulative, increasing only at moments of narrative activity as expressed through character interactions. There are a number of options for how to specify scores at  $t_0$ , but giving all characters a score of 1 is a convenient choice for two reasons. Firstly, it allows us to assume that at the beginning of the narrative, no character has demanded any more attention than any other and thus has not occupied any more character space than any other. Secondly, the absolute values given by the measure are of less interest than the relative values such as those given by  $\hat{C}_{i,t}^{out} = \frac{C_{i,t}^{out}}{\sum_i^n C_{i,t}^{out}}$  (which will be the same for any value of  $C_{i,t_0}^{out}$  shared by all characters, see Appendix A.1), so taking the first positive integer is as good a starting point as any.

### Definition of the spoken-to measure

A similar measure can be defined in which importance is based on being spoken to rather than speaking. Defining the spoken-to measure is slightly simpler than the speaking measure as there is only ever one speaking character per line (by sender exclusivity). The spoken-to measure should let spoken-to character  $i$ 's score increase at line  $t$  by a fraction of the score of speaking character  $j$ , with this fraction being given by weighting parameter  $\lambda$ . In this case, let  $x_{ji,t}$  be a variable taking the value 1 if node  $i$  is being spoken to by  $j$  at line  $t$ . Then, for every node  $i$ , the spoken-to measure can be defined as  $C_{i,t}^{in} = C_{i,t-1}^{in} + \lambda \sum_j^n C_{j,t-1}^{in} x_{ji,t}$ , which tells us that character  $i$ 's relative importance at line  $t$  is equal to whatever  $i$ 's importance was at line  $t - 1$  plus a fraction of the importance of character  $j$  speaking to  $i$  at line  $t$ . Characters' scores will only increase when they are spoken to and will do so monotonically; characters not being spoken to retain their scores from  $t - 1$ .

Formally, it should be noted that both the spoken-to measure and the speaking measure share the property that they are not invariant under permutations of sequence. This is crucial as it allows the measure to reflect

---

<sup>4</sup> The specification of  $\max(\sum_j^n x_{ij,t}, 1)$  ensures that the equation cannot possibly entail division by zero for non-speaking actors.

the sequence of events which, as we have already noted, is at the heart of narrative questions relating to how the story is told. This property also distinguishes the measure from standard approaches to measuring node centrality in sequences, such as looking at the cumulative weighted degree of nodes after each interaction. Consider, for example, two sequences: (1)  $A \rightarrow B, B \rightarrow A, A \rightarrow B$ , and (2)  $A \rightarrow B, A \rightarrow B, B \rightarrow A$ . Table 1 calculates the spoken-to measure for both sequences, assuming that A begins each sequence with score  $C_A$  and B begins each sequence with score  $C_B$ . If we cumulatively calculate weighted in-degree for (1) and (2), A will end each sequence with a score of 1, and B will end each sequence with a score of 2. However, Table 1 shows that both nodes end with different scores in sequences (1) and (2) using the measure developed here. Thus, our sense of each character's importance is dependent on how the action unfolds.

Sequence	(1)		(2)	
Node	A	B	A	B
t=0	$C_A$	$C_B$	$C_A$	$C_B$
t=1	$C_A$	$C_B + \lambda C_A$	$C_A$	$C_B + \lambda C_A$
t=2	$C_A + \lambda C_B + \lambda^2 C_A$	$C_B + \lambda C_A$	$C_A$	$C_B + 2\lambda C_A$
t=3	$C_A + \lambda C_B + \lambda^2 C_A$	$C_B + 2\lambda C_A + \lambda^2 C_B + \lambda^3 C_A$	$C_A + \lambda C_B + 2\lambda^2 C_A$	$C_B + 2\lambda C_A$

Table 1. Spoken-to scores for sequences (1) and (2).

### Normalisation and properties

Both spoken to and speaking measures can be normalised by the total score across actors as outlined above. Here we consider some properties of the measures which generalise to all dynamic character interaction networks, in order to illustrate how the measure works in cases where narrative focus is distributed differently. Given the dynamic structure of the narrative network, it is not trivial to derive properties of the measures. The measures can be related to the achievable maximum and minimum scores for a number of stylised dynamic networks (details are provided in Appendix A.2). Although mainstream film narratives are typically more complex than the kinds of networks in the stylised graphs, exploring them helps us to understand how a key aspect of the measure, the “feedback” introduced by  $\lambda$ , works to either concentrate or diffuse narrative attention in the character system depending on how dialogue is distributed.

For example, we can consider two stylised dynamic graphs that have minimal narrative focus in terms of being spoken to. One would be a “game of telephone” chain of dialogue, where each actor being spoken to then speaks to someone who has not yet been spoken to. If the length of this chain  $r$  is less than  $n - 1$ , then the  $t^{\text{th}}$  actor in the chain will have a score of  $\sum_{s=0}^t \lambda^s$ . In other words, the character's importance will only depend on where in the timeline they appear. Another example of minimal narrative focus would be where only one character speaks and other characters are only spoken to once, for  $r$  acts of speaking,  $r \leq n - 1$ . For this star-like dynamic graph, actors would have a score of  $1 + \lambda$  if they are spoken to, and 1 otherwise. In the game of telephone there is more heterogeneity in scores as the narrative importance accumulates, but in the star-like network, the repeated speaking acts mean that the narrative importance does not accumulate as the importance gained by being spoken to is never passed on. In both of these cases, narrative focus is minimal because our sense of an actor's importance is never reinforced by an addressee responding to their speech, or by another speaker speaking to the same addressee.

On the other hand, we can consider two stylised dyadic dynamic networks which produce maximal narrative focus: one in which only one actor speaks and the other in which the actors take turns addressing each other. These can be thought of as graphs that maximise spoken-to scores as no act of speaking is wasted on addressing

someone who has not been spoken to before. For the first scenario, the score of the actor being spoken to increases linearly but for the second, the act of reinforcing each other's narrative importance means that the scores increase in a curvilinear way and, after a certain number of lines of dialogue, the scores of both actors overtakes that of the spoken-to actor in the first scenario. The maximal score after  $r$  lines of dialogue can thus be thought of as an achievable maximum spoken-to score.

## 5. Application: women in popular film

In this section we provide an example of how the measure can be applied to real questions about narrative texts using the example of gendered representation in film. In particular, we focus on the issue of the vocal and relational disempowerment of female characters in popular Hollywood cinema.

### *Relational and vocal disempowerment*

Feminist film theory has long theorised the way that a masculine perspective is privileged in Hollywood through the "male gaze", understood as the way that narrative cinema produces a masculine viewing position which renders male characters as active narrative forces and female characters as passive and defined primarily by their "to-be-looked-at-ness" (Mulvey 1975). While these ideas are based in the visual aspects of cinematic representation, we can see how the visual-narrative paradigm of the male gaze produces effects which also play out through dialogue. For example, feminist film scholars have discussed how the telling of stories from a male point of view has led to a silencing of women's voices, and a representational system in which women exist primarily through their relations to male protagonists (Haskell 1987; Kaplan 1983). This point has been popularised through the simple device known as the Bechdel test (based on a comic strip by Alison Bechdel), which asks whether a film has two female characters who speak to each other about something other than a man (Bechdel 1986, 22). Though originally intended as a joke, this intentionally low bar has proved too difficult to clear for an alarming number of Hollywood films. While the literature on women and film offers a theoretical basis for understanding why narratives might not give women the opportunity to interact with each other, there has been a conspicuous lack of empirical investigations of the gendered distribution of dialogue in contemporary Hollywood cinema, limiting our understanding of what Jennifer O'Meara calls the "vocal and verbal (dis)empowerment" of women on-screen (O'Meara 2016).

Most available empirical evidence on vocal representation comes from a number of content analyses published annually which look at mainstream films released that year and offer a quantitative report on the representation therein. From these yearly reports, we know that over the last decade or so (the period for which data is available), the proportion of speaking characters that are female in top-grossing Hollywood films is around 30% (Lauzen 2018; Smith et al. 2018). The consistency of these findings is remarkable: in the Smith et al. (2018) study, the prevalence of female speaking characters was never lower than 28.1% and never higher than 32.8% in a given year across their 11-year sample, while in the Lauzen (2018) study, the annual prevalence of women among speaking characters was never less than 30% and never more than 34%. This consistency suggests that the under-representation of women in speaking roles is part of the formula of the mainstream Hollywood screenplay. However, these figures cannot tell us very much about how vocal disempowerment actually operates in Hollywood narratives. For example, knowing that 30% of the speaking characters in a given film are female does not tell us whether 30% of the lines in the film were spoken by females, nor does it tell us to whom those lines were spoken (Jones 2018).

This latter point leads to the second aspect of the disempowerment of women in Hollywood narratives that we focus on here: relational disempowerment. By foregrounding female homosocial interaction, the Bechdel test insists on a relational understanding of the narrative effects of gendered representation in film narratives. In doing so, it chimes with scholarship on post-feminism and backlash discourses in popular media culture. This work has identified a trend in Hollywood towards "taking into account" feminism by representing strong female characters, while simultaneously individualising and depoliticising their narratives by isolating those female

characters among men rather than offering narrative opportunities for collective empowerment (Faludi 1992; McRobbie 2004; Sutherland and Feltey 2016).

However, much of the scholarship on women in film, especially in relation to active female heroes, has been geared towards identifying and theorising the strong female character at the level of individual characterisation. This is particularly the case in the literature on women in action cinema, where there has been a wealth of scholarship fleshing out and analysing the characteristics of what has been variously termed the “action babe”, “action chick” or “action heroine” (e.g. Brown 2011; Inness 2004; O’Day 2004; Schubart 2007). However, by focusing primarily on the characteristics of the individual female hero in discussions of gendered representation in action cinema, these studies overlook the ways in which such ostensibly strong female characters can be disempowered through their relational positioning within the overall narrative and lack of independence from men (Purse 2011). Moreover, such an approach risks losing sight of what E. Ann Kaplan refers to as “female positionings,” and “the larger structuring of the narrative and on the placement of the woman within that narrative” (Kaplan 1983, 2). Thus, there may be a lot to gain from developing an approach which re-embeds female characters into the context of the broader character system of the narrative. A network-based approach offers a way to situate the discussion of these female characters within the broader relational character system in which they play out their role in the story.

Elsewhere Jones (2018) has argued that character interaction networks offer an illuminating model for investigating both the vocal and relational elements of female representation identified here. In particular, the approach illustrates the limitations of knowing only the prevalence of women among speaking characters, as the proportion of dialogue spoken by women can differ significantly from this simple character distribution; it can highlight how the gendered distribution of dialogue can enable or constrain female characters from accessing a collective mode of empowerment in narratives; and it is able to identify important distinctions between speaking and being spoken-to. Some aspects of this relational disempowerment, such as the marginalisation of female-female relationships, can be further explored in more detail using dyadic approaches such as the relational event model (Butts 2008). However, node-level measures such as the one developed in this paper can shine further light on the problem by helping us to operationalise Kaplan’s notion of “female positionings” and consider where the strong female character stands in terms of the relative importance of the characters in the overall narrative, as defined by their relational activity.

### *The Force Awakens*

To explore these questions about the gendered distribution of character interactions in film narratives, we use the 2015 film *Star Wars: The Force Awakens* (henceforth simply *The Force Awakens*) as an example. *The Force Awakens* serves as an interesting case study for a couple of reasons. For one, the film is the highest grossing film of all time in the United States and the fourth highest grossing film of all time worldwide (Box Office Mojo 2019). This not only points to the immense popularity and cultural reach of the film, but also places the film clearly within the context of top-grossing or “blockbuster” films addressed by existing Hollywood content analyses. Furthermore, *The Force Awakens* sparked discussion about a more inclusive and diverse direction for representation in the hugely popular and culturally impactful *Star Wars* franchise through the film’s inclusion of female and non-white characters in central active roles. Much of this discussion centred around the prominence of the character Rey (Daisy Ridley) as a key protagonist in the film. For example, following the film’s release, Rey was credited with “carrying the heavily-mantled weight of the new series” (Kermode 2015); it was proclaimed that “Rey is clearly the main character and our destined-hero for this trilogy,” that “Star Wars has a female lead,” and that “the main male character is her sidekick” (Cox 2015); there were claims that, in the film, “the male-centric universe of the original *Star Wars* gives way to a woman warrior [Rey] and a female version of Yoda [Maz Kanata]” (Roddy 2015), and that “the plot of *The Force Awakens*, in fact, revolves around [Rey’s abilities]” (Garber 2015). This not only renders it an interesting example for exploring node positions through the lens of vocal and relational (dis)empowerment, but also provides a useful case for examining what

looking at characters' positions within the overall narrative character system can tell us that focusing only on the characterisation of the individual hero (Rey) cannot.

A brief summary of the plot of the film will help contextualise and aid understanding of the discussion of the positions of the characters therein. *The Force Awakens* is the seventh feature film in the *Star Wars* franchise and centres around a conflict between the tyrannical First Order and the Resistance. The film follows Poe Dameron (Oscar Isaac), a Resistance fighter pilot, and a turncoat First Order Stormtrooper named Finn (John Boyega) as they try to deliver a droid named BB-8 to the Resistance leaders as it contains information on the whereabouts of Luke Skywalker (Mark Hamill). Finn and Dameron crash and are separated on the planet Jakku, where a scavenger named Rey discovers BB-8 and agrees to help keep it safe. Finn encounters and joins Rey as the First Order arrive and the two escape in the Millennium Falcon. Han Solo (Harrison Ford), the ship's previous owner, finds the ship and boards it, wishing to reclaim his property. Rey and Finn explain their situation to Han, who agrees to help them deliver the droid to the Resistance, which is led by Han's estranged lover Leia Organa (Carrie Fisher). In the process of delivering BB-8 to the Resistance, Rey is captured by the First Order and delivered to Kylo Ren (Adam Driver), one of the First Order's chief commanders and the runaway son of Han and Leia. While interrogating Rey, Ren discovers that she, like Ren, has control of an energy-like power known as the Force. The Resistance fighters free Rey from the First Order base, destroy its main weapon and uncover the map to Skywalker's location.

In the dialogue data for *The Force Awakens*, there are 688 lines of dialogue distributed between 31 named speaking characters. 22 percent of named speaking characters in the film are female (well below the industry benchmark of around one-in-three), while 28.1 percent of the lines are spoken by female characters, who account for 26.5 percent of the recipients of lines of dialogue.

## Discussion

Figures 1 and 2 plot the speaking and spoken-to measures, respectively, using the  $\lambda$  value of 0.01. Lower  $\lambda$  values allow for smoother trajectories for the characters but make the measure less sensitive, which makes it harder for characters to establish their centrality later in the film. A second consequence of this insensitivity is that the score of characters who establish a high centrality earlier in the film but do not continue to contribute vocally to the film thereafter will decline at a much slower rate, and may end up with higher scores than characters who contributed more in the latter stages of the film. Using higher  $\lambda$  values on the other hand, has the opposite problem, being very sensitive to rises and falls in character centrality at each interaction, an effect that is compounded later in the film as the scores of characters are generally higher. Thus, the value of  $\lambda$  can be thought of as how many times a character would need to interact with another character to overtake their score. For a  $\lambda$  value of 0.1 on the spoken-to measure, for example, character  $i$  would receive all of  $j$ 's centrality after being spoken to ten times by  $j$  without reply. Thus, higher  $\lambda$  values allow characters to "steal the show" at later stages of the film, as the number of interactions required to overtake other characters is lower. Further illustration and discussion of some of these aspects of  $\lambda$  is provided in Appendix A.3. Setting  $\lambda$  at 0.01 here offers a balance between these sensitivity issues. In lay terms, using this  $\lambda$  value of 0.01 for the speaking measure means that each time a character speaks, their score becomes what it was prior to speaking plus one percent of the average score of the characters to whom they speak. For the spoken-to measure, it means that each time a

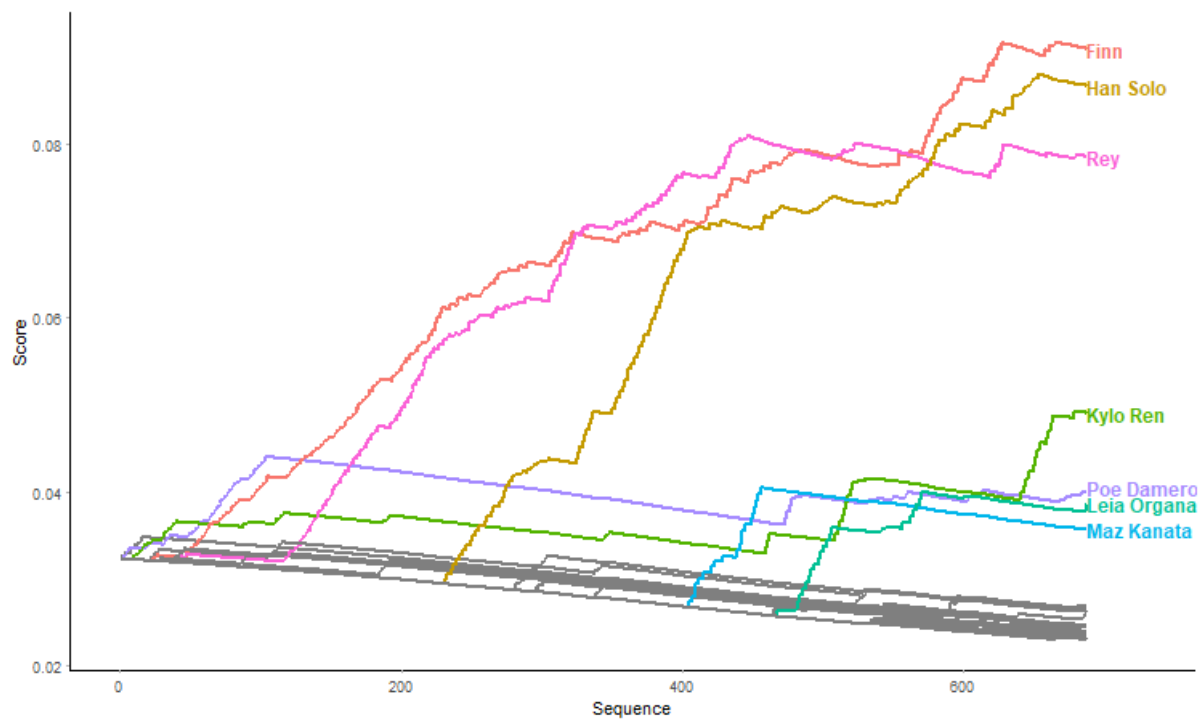


Figure 2. Normalised speaking measure scores in *The Force Awakens* ( $\lambda = 0.01$ ).

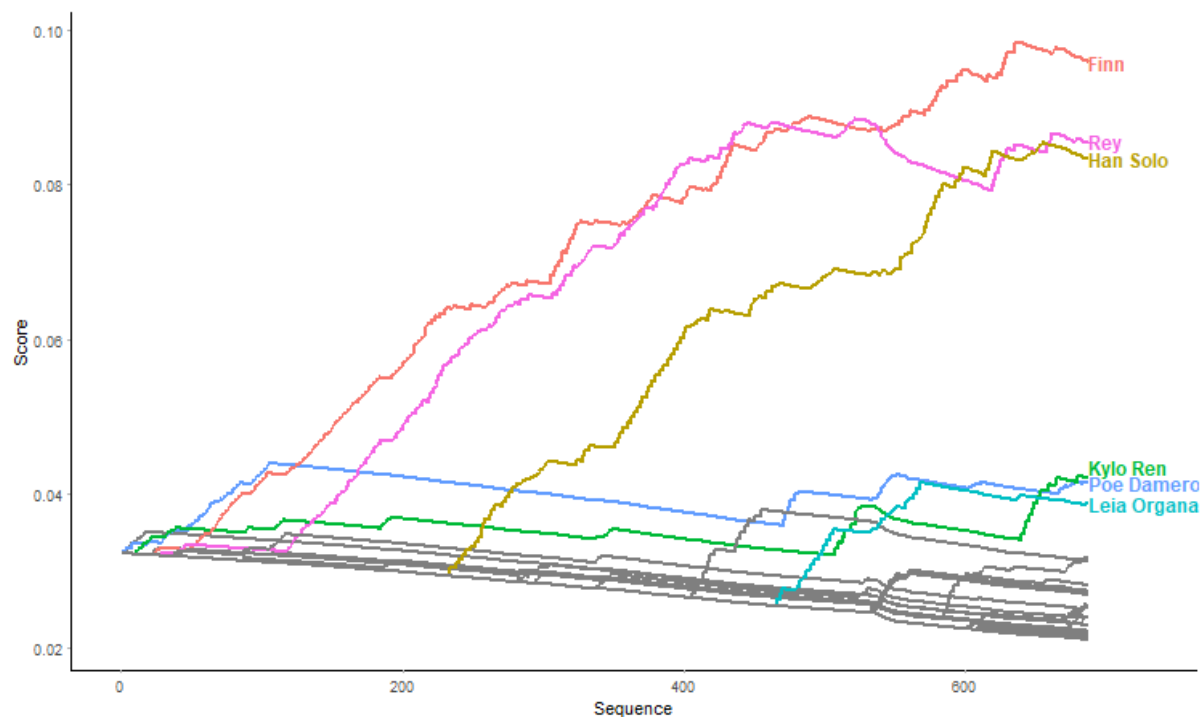


Figure 1. Normalised spoken-to measure scores in *The Force Awakens* ( $\lambda = 0.01$ ).

character is spoken to, their score becomes what it was prior to being spoken to plus one percent of the score of the character speaking to them.

The horizontal axis in Figures 1 and 2 indicates the sequence of lines, with all characters beginning with an equal score. Key characters are labelled in the graph based on whether their normalised score at  $t_{max}$  (the end of the film) is greater than their normalised score at  $t_0$  (the beginning of the film). In each case scores have

been normalised to give relative scores by dividing the scores by the sum of all characters' scores such that  $\hat{C}_{i,t}^{out} = \frac{C_{i,t}^{out}}{\sum_i C_{i,t}^{out}}$ . These relative scores effectively convey the share of total narrative importance, allowing scores to both increase and decrease. This operationalises the idea of the narrative as a distribution of finite attention and neglect between characters; for the normalised speaking scores, for example, when one character speaks their share of the total narrative importance increases, while each non-speaking character's share of the total narrative importance declines slightly.

The first thing to note from Figures 1 and 2 is that Finn spends the majority of the film (including, importantly, the final act) with the largest share of the scores on both the speaking and spoken-to measures. This is significant in its own right as Finn is a person of colour (in a historically very white franchise) and, although the focus here is on gender, this illustrates the broader appeal of being able to characterise characters' dynamic narrative positions for scholars of social representation in fictional texts. Moreover, while it is quite clear from the graphs that Rey is central to the plot of the film, it is also clear that from the perspective of vocal activity this centrality is secondary to Finn and, in the latter part of the film, Han. This is not to downplay the significance of her character's contribution to the narrative, but it should temper some of the claims made in the critical reception to the film about her role. If Rey is the main character in *The Force Awakens*, then we might question why Rey spends so little time holding the largest share of the scores on either of the measures. Moreover, when looking across scores for all the characters in the film, we can certainly call into question the idea that the male-centricity of the *Star Wars* universe has "given way," given how few female characters separate themselves from the pack throughout the narrative. The narrative is still driven vocally by men, for the most part.

There are also several things to note regarding the dynamic nature of the measure. We can see that all characters begin at the same point on the vertical axis, due to starting with an equal score. However, as lines are spoken, characters begin to delineate their narrative activity through their verbal interactions. As the narrative evolves, different characters become more and less central. For example, Finn and Rey occupy a prominent role in the narrative throughout the film, though Rey's activity declines somewhat in the third act due to her being imprisoned during this period. Han Solo emerges after the first act and becomes increasingly central to the plot thereafter, while Leia rises in importance in the third act as she orchestrates the Resistance's fightback against the First Order. Kylo Ren remains fairly central throughout the film, even as other characters dominate the narrative activity, though Ren's centrality rises towards the end of the film as he comes face to face with several of the film's other central characters for the first time. The measure thus allows us to say something about the relative narrative positions of these characters *within the context of the dynamics of the narrative*.

Consider, for example, the position of Maz Kanata in Figure 1. Over the course of a few key scenes in the second act of the film, Maz (Lupita Nyong'o) becomes very important to the plot, and her share of the scores increases significantly during this period. During these scenes, Han, Finn and Rey are dependent on Maz to help them contact the Resistance, and Maz is thus pivotal to the advancement of the narrative. Not only is she vocally active during this part of the film, but her vocal activity is with the most central characters in the film up to that point, thus rapidly increasing her relative importance to the narrative. However, her importance is defined temporally; she is important during those moments of the film, but she is not present at all in the final act of the film, and thus her share of the narrative centrality declines again. Likewise, we can see from Figures 1 and 2 that Poe (alongside Finn) is very central to the narrative in the first act of the film, as the two escape the First Order together, but that the narrative focuses more on Rey and Finn (and later Han) after this point. It would be impossible to adequately characterise these temporally-situated narrative positions using measures based on a static network representation.

These dynamic properties bring other advantages when relating this measure to more interpretive understandings of characters' positions within the narrative. Scholarship on the role of women in action cinema, for example, has discussed the ways in which women are often ejected from narratives (for example, killed or

kidnapped) in order to motivate a male hero's narrative actualisation (Purse 2011; Tasker 1993). In cases where this occurs, the scores of those characters who have ceased to be active narrative contributors will no longer increase, and thus their share of the character space will decline, reflecting this narrative ejection. Indeed, when Rey is captured and imprisoned in *The Force Awakens*, we see her share of the scores decrease as her role becomes temporarily more passive while others are motivated to act. Moreover, as the (non-normalised) measure is monotonically non-decreasing, interactions towards the end of the narrative will tend to be given more weight than interactions nearer the beginning. In substantive terms, this makes sense as we are more likely to have a sense of which characters are important after we have spent more time watching the narrative unfold (Bal 2017; Frermann et al. 2018). Thus, our sense of the importance of a character entering the narrative towards the end is informed by our accumulated understanding of the relative importance of the characters they interact with, though we have less information about this earlier in the film.

Finally, there are interesting differences between the speaking scores and spoken-to scores. Han Solo and Kylo Ren hold a greater share of the narrative importance on the speaking measure than they do on the spoken-to measure. This could imply simply that these characters speak more than they are spoken to in the film. However, looking at the number of lines spoken by and spoken to the characters reveals that Han speaks 124 lines and is spoken to 134 times, while Kylo Ren speaks 53 lines and is spoken to 48 times. From this we can infer that it is not only how much Han speaks that gives him higher speaking scores, but rather that his lines tend to also be delivered to more important characters (as defined by the measure). The measure suggests that when Han speaks, he tends to speak to other characters who speak a lot, and the characters he speaks to tend to be more important than the characters that speak to him. On the other hand, Rey's share of the narrative importance is greater when the measure is based on being spoken to than when it is based on speaking. Again, this suggests not only that Rey is spoken to a lot (Rey speaks 132 lines and is spoken to 160 times, both more than Han), but that the characters speaking to her tend to be more important than the characters she speaks to.

These differences between Han and Rey suggest two different kinds of importance - Han's tendency to speak to more important characters implies a more authoritative and teacherly role, while Rey's tendency to be spoken to by more important characters implies a more apprentice-like role. On the face of it, this makes sense in the context of the narrative, as Han is a much more senior and experienced character in the fictional universe while Rey is much more "green". However, a gendered reading of this discrepancy might problematise the relationship here between gender, seniority and vocal activity. Hollywood is known to allow older males to continue to feature prominently in narratives, while older women are underrepresented and negatively associated with seniority (Bazzini et al. 1997; Lauzen and Dozier 2005). This is likely to produce scenarios such as that between Han and Rey, where an older and more senior male figure is featured alongside a younger female figure. However, *The Force Awakens* does also feature Leia Organa, a senior female figure who, like Han, returns to the series from the original trilogy of films and holds an important role in the plot. Leia is the general of the Resistance and the mother of the film's antagonist Kylo Ren. One might assume that this would grant her narrative centrality comparable to that of Han, who is Kylo's father and not a senior Resistance figure. However, Figures 1 and 2 show that, despite her role, Leia's vocal activity pales in comparison with the likes of Han, Finn and Rey. This chimes with scholarship on the way that postfeminist media narratives merely "take into account" feminism (McRobbie 2004) as, unlike Han, Leia's leadership comes primarily from what we know of her position in the Resistance, rather than from us actually seeing her leading the protagonists. In this regard, we might question how much has changed since the original instalment of the saga in 1977. The directed nature of the measure thus allows for a deeper understanding of the ways in which the characters interact with each other over the course of the narrative.



## 6. Conclusions

In this paper we have aimed to develop a more integrated conceptual-methodological approach to the node-level analysis of character networks based on fictional texts. Thus, the paper contributes on two fronts: (1) a conceptual refinement of the character networks approach to analysing fictional texts, and (2) the development of a dynamic node-level measure for analysing the positions of characters within such texts. The conceptual refinement is found in the departure from the existing approach of applying off-the-shelf network analytical tools based on static network representations to narratives and instead pursuing a more bespoke approach to analysing narrative texts as dynamic networks. By doing so, we can relate the concept of centrality in the character network to a notion of narrative importance more in keeping with key ideas from the work on narratology and narrative comprehension. The measure presented in this paper is intended as a simple example of how such an approach might yield more useful, accurate and appropriate results for meaningful analysis of how characters are positioned in narratives. Through an application to *The Force Awakens*, we have shown that the dynamic node-level measure presented here can aid our understanding of the gendered dynamics in film texts in ways that existing measures cannot. Moreover, it allows us to describe the characters' evolving narrative positions within the context of the temporal dynamics of the narrative; it gives us a relational way of tracing character activity over the course of a narrative; and it can reveal relational differences in speaking and being spoken to.

It should be noted that the insistence on a dynamic network representation in this paper does not commit one to a labour-intensive manually coded network extraction method such as the one assumed here. All that is required for the principles set forth in this paper to be satisfied is that the basic structure from which the character network is constructed is a sequential list of relational events, and the question of how to represent the data is separate from the question of whether this process should be manual or computational. The limitation we identify with existing studies is that they have configured network extraction tools which move directly from the text to a static representation, usually either by increasing the cell values in an adjacency matrix for each interaction or by increasing the weight in an edgelist. Neither the adjacency matrix nor the weighted edgelist contain information about the temporal sequence of the events. However, those same extraction tools could just as easily store the interaction data in an event list format, with the same basic data structure as the one used in this paper. This event list can then be used to generate static weighted edgelists and adjacency matrices if desired, but if the event list step is skipped, it cannot be reverse-engineered later from the static representations.

For example, a screenplay parsing tool could add a row to an event list for each identified speech act, recording an event ID (incremented for each speech act, capturing the sequence), the scene ID (incremented after each scene boundary), the speaker ID (based on named entity recognition), and the "recipients" (which would not be direct recipients, but could be those characters who also spoke within the same scene). Then, for example, the dynamic speaking centrality measure developed here could be applied directly to this data structure without the need for modification. The spoken-to measure would no longer be appropriate, given that the tool can only accurately identify the speaker but not the recipients. However, for the speaking measure, we would still gain a sense of the vocal narrative activity of each character, with interactions weighted by the characters involved in the narrative at the time of the interaction, illustrating how attention is distributed among characters through the narrative. In this way, existing computational approaches can be brought much more in line with the dynamic approach which we have argued for in this paper. Thus, we suggest that one of the implications of this paper for character networks research is that the computational approach can be an effective tool for analysing how stories are told, with some minor adjustments.

Moreover, the utility of the approach is not limited to films and can add to existing character network studies of literary texts such as novels and plays. Whatever the media, the measure can help to enhance readings of how characters move through narratives by way of their activity in the character system, telling us much more than can be gleaned from measures such as degree or betweenness in a static or time-windowed network

representation. This is important as cultural texts engage in the reproduction of social meaning, and we need approaches that can accurately measure the important identificatory power relations at play in these texts. The stronger the conceptual links between narrative texts and the ways we represent them in data are developed, the more useful the character network approach becomes for illuminating these power relations.

The measure developed here is offered as a starting point for further thinking about how to measure the relationships between characters and narratives, and there are a number of ways in which the measure might be further developed. For one, dialogue is a crucial dimension of how film narratives are played out, but it is only one dimension. It would be useful to explore the extent to which the picture of the dynamic character system that emerges from dialogue-based network data would acquiesce with that derived from more visually-based data such as screen-sharing patterns. Moreover, the measure might be refined so that the content or type of an interaction is taken into account. It may be reasonable to assume that certain kinds of interaction carry more narrative weight. For example, talking about certain topics considered pivotal to the plot might be considered a greater signifier of the narrative importance of the characters. Alternatively, the measure might be refined so that a character can increase their score slightly if they are not directly involved in an interaction but they are the subject of that interaction. As currently defined, the measure assumes that a character's relative share of the narrative attention decreases when they are not involved in an interaction, but if characters A and B talk to each other about C, then some narrative attention is also being paid to C. Such modifications would require deeper consideration, but the measure developed here is presented as a platform on which more complex properties such as these might be developed.

## A. Appendix

### A.1. Scale invariance

The relative measure is invariant to multiplication of initial values  $\{C_{i0}\}$ , when  $C_{i0} = c$  for all  $i = 1, \dots, n$ . For spoken to, this is immediately obvious through observing that  $C_{it}^* = aC_{it}$  implies  $C_{it+1}^* = aC_{it+1}$ , given through direct calculation

$$C_{it+1}^* = C_{it}^* + \lambda \sum_{j \neq i} C_{jt}^* x_{ji,t} = aC_{it} + \lambda \sum_{j \neq i} aC_{jt} x_{ji,t} = aC_{it+1},$$

and  $C_{i0}^* = aC_{i0}$ . Consequently  $C_{it}^*/C_{jt}^* = C_{it}/C_{jt}$ , for all  $i, j \in \{1, \dots, n\}$  and it does not matter what initial value  $c > 0$  is used. The same relation holds trivially for speaking.

### A.2. Some minimal and maximal scores

We consider first two dyadic schematic dynamic graphs. To simplify the exposition, in the remainder we consider only the spoken to scores and dynamic graphs with both sender and receiver exclusivity. Consider two actors  $i = 1, 2$  speaking to each other, in which case we can define the sequence of speech in terms of a binary sequence  $\{y_t\}$  defined such that  $x_{12,t} = y_t$  and  $x_{21,t} = 1 - y_t$ , and assume  $C_{10} = C_{20} = 1$ .

**Dyadic Asymmetric Speech (DAS):** If  $y_t = 0$  for  $t \leq r$ , we get the dynamic graph for  $t = k, k + 1, \dots, r$  of Figure A1(a). For  $k = 1$ , the scores at the  $r^{\text{th}}$  round are  $C_{2r} = 1$  and  $C_{1r} = 1 + r\lambda$ . Thus if actor 2 speaks the first  $r$  lines, then the score of actor 1 increases linearly.

**Dyadic Turn-taking (DTT):** If  $y_t = 1 - y_{t-1}$ , the actors are taking turns speaking to each other. Assuming 2 speaks odd rounds, we get the dynamic graph in Figure A.1.(b)

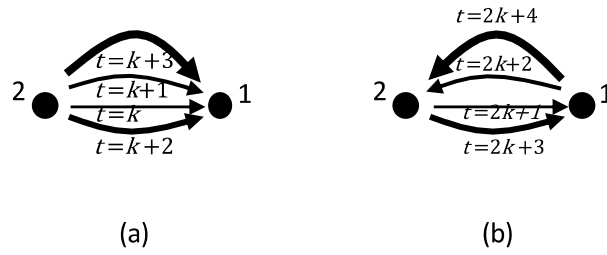


Figure A.1. The dynamic graph of DAS (a) and DTT (b) for 4 acts of speech.

To derive the scores for DTT let  $y_1 = 0$ , in which case

$$C_t = A_t C_{t-1}, \quad (4)$$

where

$$C_t = \begin{pmatrix} C_{1t} \\ C_{2t} \end{pmatrix}, \text{ and } A_t = \begin{pmatrix} 1 & \delta_t \lambda \\ (1 - \delta_t) \lambda & 1 \end{pmatrix},$$

for  $\delta_t$  equal to 1 or 0 according to whether  $t$  is odd or even, respectively. From (4) it also follows that

$$\begin{aligned} C_{2-\delta_r, r} &= C_{2-\delta_r, r-1} + \delta_r \lambda C_{1+\delta_r, r-1} \\ &= C_{2-\delta_r, r-2} + \delta_r \lambda C_{1+\delta_r, r-1}. \end{aligned} \quad (5)$$

Due to the recursion (4), the elements of  $C_t$  must be polynomials in  $\lambda$ .

*Lemma 1:* Defining  $a_{r,t} = 0$  for  $t > r$  and  $r < 0$ , the scores for DTT at the  $r^{\text{th}}$  line of dialogue, are given by

$$C_{2-\delta_r, r} = \sum_{t=0}^r a_{r,t} \lambda^t, \quad (6)$$

and

$$C_{1+\delta_r, r} = \sum_{t=0}^{r-1} a_{r-1,t} \lambda^t, \quad (7)$$

with coefficients  $\{a_{r,t} : r = 0, 1, \dots; t = 0, 1, \dots, r\}$  that satisfy the recursion

$$a_{r,t} = a_{r-2,t} + a_{r-1,t-1}. \quad (8)$$

*Proof:* For  $r=1, 2$ , it is trivial to check that (6), (7), and (8) holds. Assume that (6) holds for  $r = k-1, k$ , the firstly

$$\begin{aligned} C_{1+\delta_r, k+1} &= C_{1+\delta_r, k} \\ &= C_{2-\delta_{r-1}, k} \\ &= \sum_{t=0}^k a_{k,t} \lambda^t \end{aligned}$$

and by applying (5)

$$C_{2-\delta_r, k+1} = C_{2-\delta_r, k-1} + \delta_{k+1} \lambda C_{1+\delta_r, k}$$

$$\begin{aligned}
&= \sum_{t=0}^{k-1} a_{k-1,t} \lambda^t + \lambda \sum_{t=0}^k a_{k,t} \lambda^t \\
&= \sum_{t=0}^{k+1} (a_{k-1,t} + a_{k,t-1}) \lambda^t \\
&= \sum_{t=0}^{k+1} a_{k+1,t} \lambda^t. \square
\end{aligned}$$

*Theorem 1:* The coefficients that follow (8) are given by

$$a_{r,t} = \left( \left\lceil \frac{r+t-1}{2} \right\rceil \right)_t, \quad (9)$$

where  $\lceil a \rceil$  is the ceiling function.

Proof (David Schoch, personal communication, 2019): The equality (9) holds trivially for  $r=0,1,2$ . By direct calculation using Lemma 1,

$$\begin{aligned}
a_{r-2,t} + a_{r-1,t-1} &= \left( \left\lceil \frac{r+t-3}{2} \right\rceil \right)_t + \left( \left\lceil \frac{r+t-3}{2} \right\rceil \right)_{t-1} \\
&= \left( \left\lceil \frac{r+t-3}{2} \right\rceil + 1 \right)_t \\
&= \left( \left\lceil \frac{r+t-1}{2} \right\rceil \right)_t \\
&= a_{r,t}. \square
\end{aligned}$$

The coefficients  $\{a_{r,t}\}$  form a Pascal-like triangle, a flipped form of which was derived by Gould (1965). The triangle represented as a lower-triangular matrix, has as its first column 1s, second column repetitions of counting numbers, and third column repetitions of triangular numbers.

Analysing  $C_{2-\delta_r,r}$ , provides some insight into how  $\lambda$  controls the strength of feedback in accumulative narrative importance. Define the row-sums

$$a_r = \sum_{t=0}^r \left( \left\lceil \frac{r+t-1}{2} \right\rceil \right)_t.$$

*Corollary 1:* For  $\lambda = 1$ ,

$$C_{2-\delta_r,r} = F_r,$$

where  $F_r$  is the  $r^{\text{th}}$  Fibonacci number.

The proof follows easily from noting that  $C_{2-\delta_r,r} = a_r$ , and adding row-sums  $a_{r-1}$  and  $a_{r-2}$  as in Gould (1965).

Further insight is provided by comparing the highest score at  $t$  for DAS and DTT. In the following we will only consider the highest scores, i.e.  $1 + t\lambda$  and  $C_{2-\delta_r,r}$ , respectively. Firstly, the scores are equal for  $t = 0,1$ . The score  $1 + t\lambda$  for DAS is linear but the maximal score for DTT is a convex function of  $t$ . Clearly, if  $\lambda = 0$ , the scores for both remain 1. To investigate the conditions for  $C_{2-\delta_{t,t}} > 1 + t\lambda$ , define  $r_c = \min_{t>1} \{t: C_{2-\delta_{t,t}} > 1 + t\lambda\}$ .

Remark:  $r_c$  exists for  $\lambda > 0$

Proof: A very blunt lower bound is obtained from only using the first half of the coefficients

$$C_{2-\delta_t, r} = \sum_{t=0}^r \binom{\lceil \frac{r+t-1}{2} \rceil}{t} \lambda^t > \sum_{t=0}^{\lfloor \frac{r}{2} \rfloor} \binom{\lceil \frac{r}{2} \rceil}{t} \lambda^t = 1 + \lambda(1 + \lambda)^{\lfloor \frac{r}{2} \rfloor - 1}.$$

Consequently,  $r_c \geq r$  for an  $r$  that satisfies

$$(1 + \lambda)^{\lfloor \frac{r}{2} \rfloor - 1} > r,$$

where we can majorize the left-hand side using a known result

$$(1 + \lambda)^{\lfloor \frac{r}{2} \rfloor - 1} > \exp \left\{ \frac{\left( \left\lfloor \frac{r}{2} \right\rfloor - 1 \right)^2 \lambda}{\lambda \left( \left\lfloor \frac{r}{2} \right\rfloor - 1 \right) + \left( \left\lfloor \frac{r}{2} \right\rfloor - 1 \right)} \right\} = e^{\left( \left\lfloor \frac{r}{2} \right\rfloor - 1 \right) \lambda / (\lambda + 1)},$$

and thus  $r_c$  satisfies

$$\frac{1}{2} \frac{\lambda}{\lambda + 1} > \frac{\log(r_c)}{r_c - 3}.$$

Of course, the first part of the inequality can be sharpened considerably, for example

$$\sum_{t=0}^r \binom{\lceil \frac{r+t-1}{2} \rceil}{t} \lambda^t > 1 + \lambda \left[ \left\lfloor \frac{r}{2} \right\rfloor + \lambda \left( \left\lfloor \frac{r+1}{2} \right\rfloor \right) \right] + \lambda^{r-2} \left( F_r - \left( \left\lfloor \frac{r}{2} \right\rfloor + \left( \left\lfloor \frac{r+1}{2} \right\rfloor \right) + 3 \right) \right).$$

For  $\lambda = 1/2$ ,  $r_c = 5$ , and Figure 1 provides  $r_c$  as a function of  $\lambda$ .

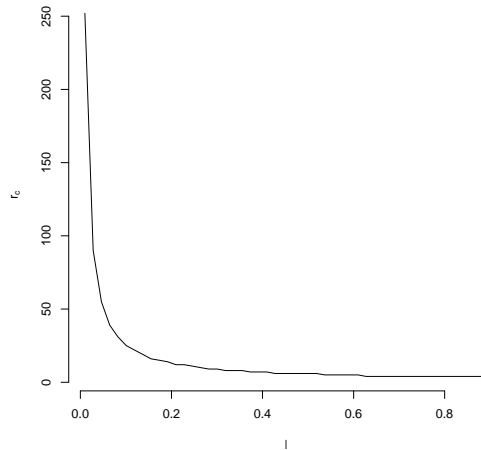


Figure A.2.  $r_c$  as a function of  $\lambda$ .

**Single Speaker Speech (SSS):** For  $r \leq n - 1$ , assume that only one actor speaks and other actors are only spoke to once. Without loss of generality assume that the speaking actor is  $i = 1$ , and speaking acts are  $(1, 2)$ ,  $(1, 3)$ , ...,  $(1, r + 1)$ , as in Figure A.3.(a). This yields scores of  $C_{ir} = 1$ , for  $i \in \{1, r, r + 1, \dots, n\}$ , and  $C_{ir} = 1 + \lambda$ , for  $i = 2, \dots, r - 1$

**Game of Telephone (GOT):** Figure A.3.(b); This type of graph has minimal narrative focus in that dialogue is simply passed on, not centring on any of the characters. Without loss of generality assume that the sequence is  $(1,2), (2,3), \dots, (t-1, t), \dots, (r, r+1)$ , in which case

$$C_{tr} = \sum_{s=0}^t \lambda^s,$$

for  $t \leq r+1$ .

**Game of Telephone with Coord (GTC):** Figure A.3.(c). Assume that at time  $r$ , speech is directed back to  $(r, i)$ , where  $i \leq r$ . The score for  $r$  is clearly  $C_{ir} = \sum_{t=0}^r \lambda^t$  from GOT, and

$$C_{ir+1} = \sum_{t=0}^i \lambda^t + \lambda \sum_{t=0}^r \lambda^t.$$

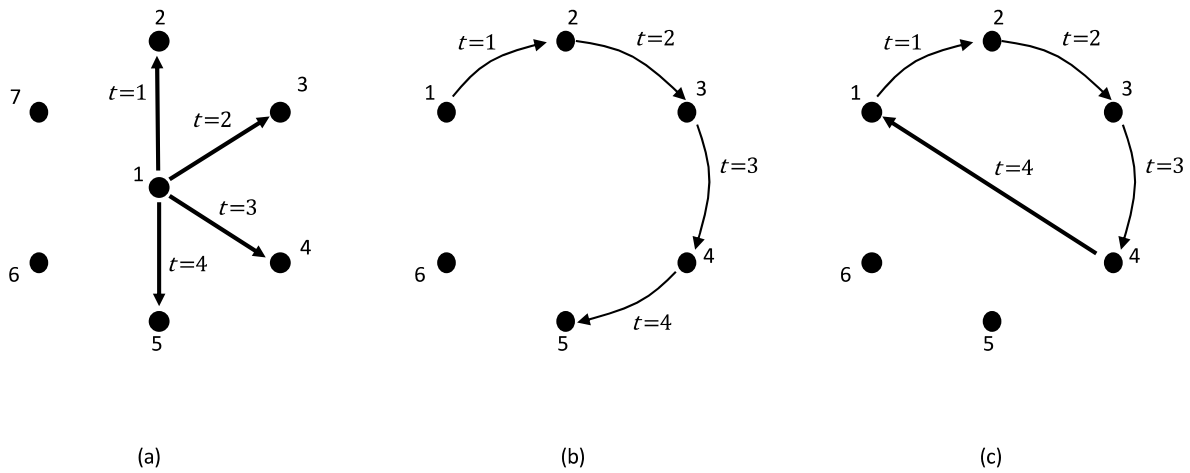


Figure A.3.: Distributed speech dynamic graphs SSS (a), GOT (b), and GTC (c).

Comparing SSS, GOT, and GTC with the two dyadic dynamic graphs DAS and DTT it becomes obvious that a dyadic graph will maximise the narrative measure. We offer the following without proof:

**Theorem 2:** Let actors be labelled so that  $C_{1k} \geq C_{2k} > C_{ik}$ , for  $i > 2$ , and let  $A = \{C_{jk+1} : j \in [n], (i_{k+1}, j_{k+i}) \in [2]^{(2)}\}$ ,  $B = \{C_{jk+1} : j \in [n], (i_{k+1}, j_{k+i}) \in \{1,2\} \times \{2, \dots, n\}\}$ , and  $C = \{C_{jk+1} : j \in [n], (i_{k+1}, j_{k+i}) \in [3, \dots, n]^{(2)}\}$ , then

$$\max(A) > \max(B) > \max(C).$$

The interpretation is that dyadic interactions maximise the narrative measure and that interactions involving 3 or more actors dilute the concentration of the narrative focus. Consequently,  $1 + r\lambda$  can be considered a maximum if  $r < r_c$  and  $\sum_{t=0}^r \binom{\lceil \frac{r+t-1}{2} \rceil}{t} \lambda^t$  if  $r_c \geq r$ .

### A.3. Choice of $\lambda$

All parametric centrality measures (e.g. Bonacich and Lloyd 2001; Bonacich and Lu 2012; Katz 1953; Opsahl et al. 2010; Taylor et al. 2017) offer the choice of how to set parameter values. The goal for the present measure is simplicity, with the effect of the parameter on output values being both clear and sensible. Like other parametric measures such as those cited above, different values emphasise different aspects of the

phenomenon, and there is no obvious or best value to choose. Thus, we provide in this section some more discussion of  $\lambda$ , explaining what the use of the value of 0.01 in our illustrations in Section 5 implies.

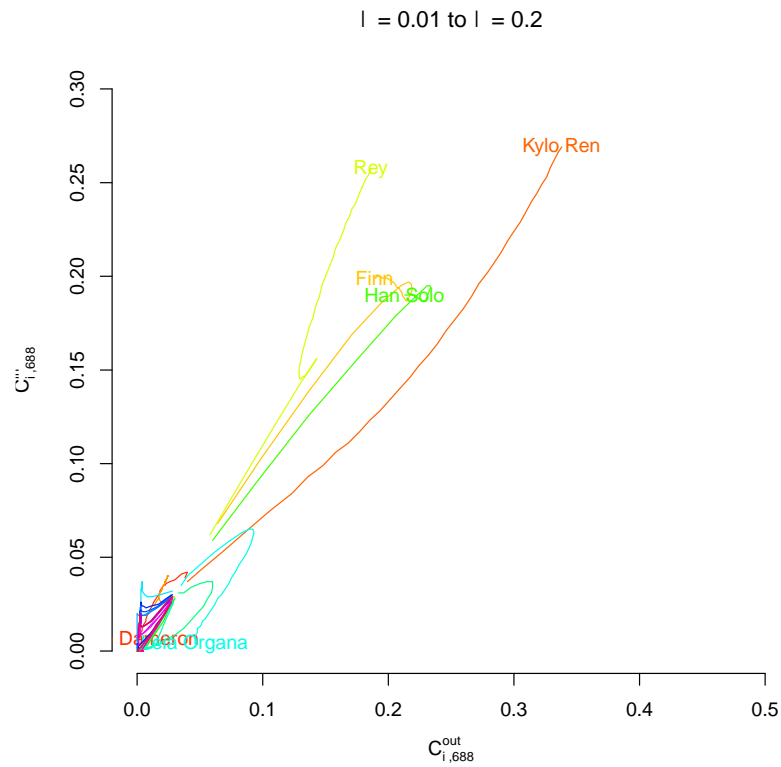


Figure A.4. Changes in measures as  $\lambda$  is increased.

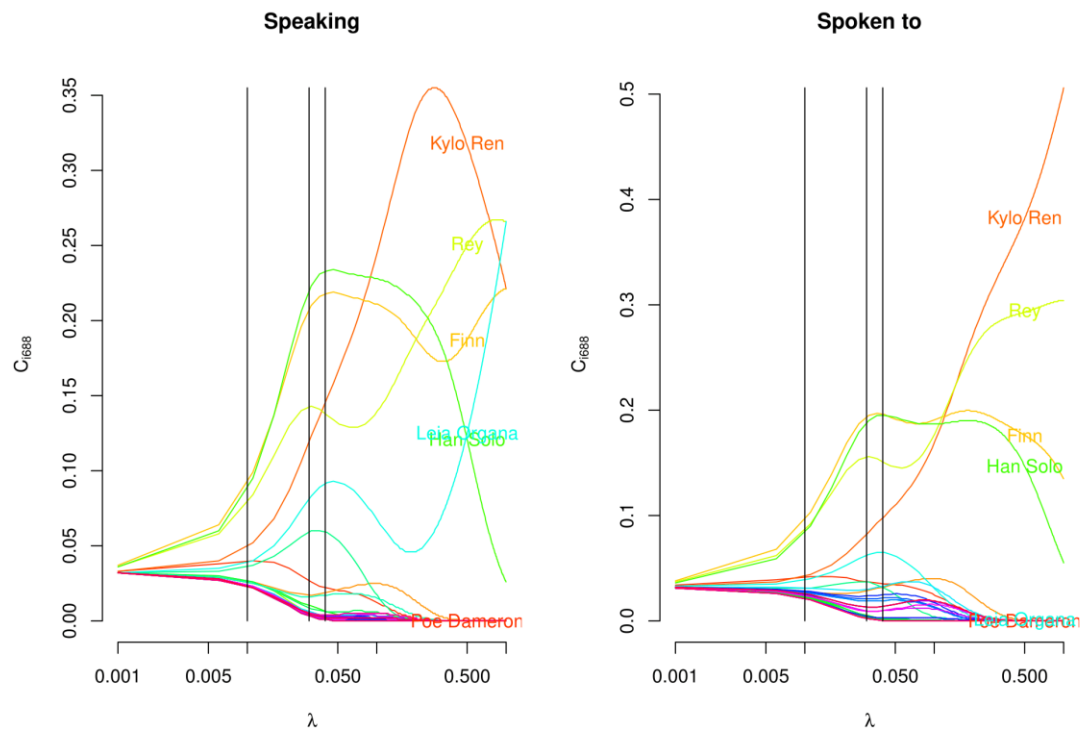


Figure A.5. Speaking and spoken to measures against value of  $\lambda$  (on log scale).

The parameter  $\lambda$  regulates the amount of feedback in the measure (as seen, for example, in comparison of DTT and DAS). For *The Force Awakens*, increasing  $\lambda$  amplifies dialogue with highly central actors, especially in later stages of the narrative. To illustrate this effect, Figure A.4 maps the trajectory of the measures when increasing from  $\lambda = 0.01$  to  $\lambda = 0.2$ .

The final scores for the characters in *The Force Awakens* are provided as a function of  $\lambda$  in Figure A.5. Figure A.5 shows that the range of output values for the measure remains fairly stable for changes in  $\lambda$  values at the lower end of the scale. The picture of relative character importance painted by the measure does not begin to change drastically until  $\lambda$  begins to reach values of around 0.03-0.04 (the area between the rightmost two reference lines in Figure A.5). At this point, increasing  $\lambda$  causes instability in the narrative, as illustrated by crossing lines. Thus, for  $\lambda$  values beyond this point, the scores become dominated by those characters involved in interactions at the conclusion of the film. This illustrates the “show-stealing” quality of higher  $\lambda$  values discussed in Section 5, most clearly seen in Kylo Ren’s rapid rise in relative centrality at higher  $\lambda$  values, and Han Solo’s corresponding decline. When identifying a useful  $\lambda$  value for the illustrations in Section 5, the range of  $\lambda$  values was narrowed to the area where the output values first begin to spread out in Figure A.5 (the area between 0.005 and 0.02) in order to preserve the importance of earlier narrative events, while still allowing for events to gain importance as the narrative progresses.

Character trajectories were then plotted using a selection of  $\lambda$  values in this range, as shown in Figure A.6. These plots were used to more closely consider the sensitivity of the measure in light of the goals of simplicity

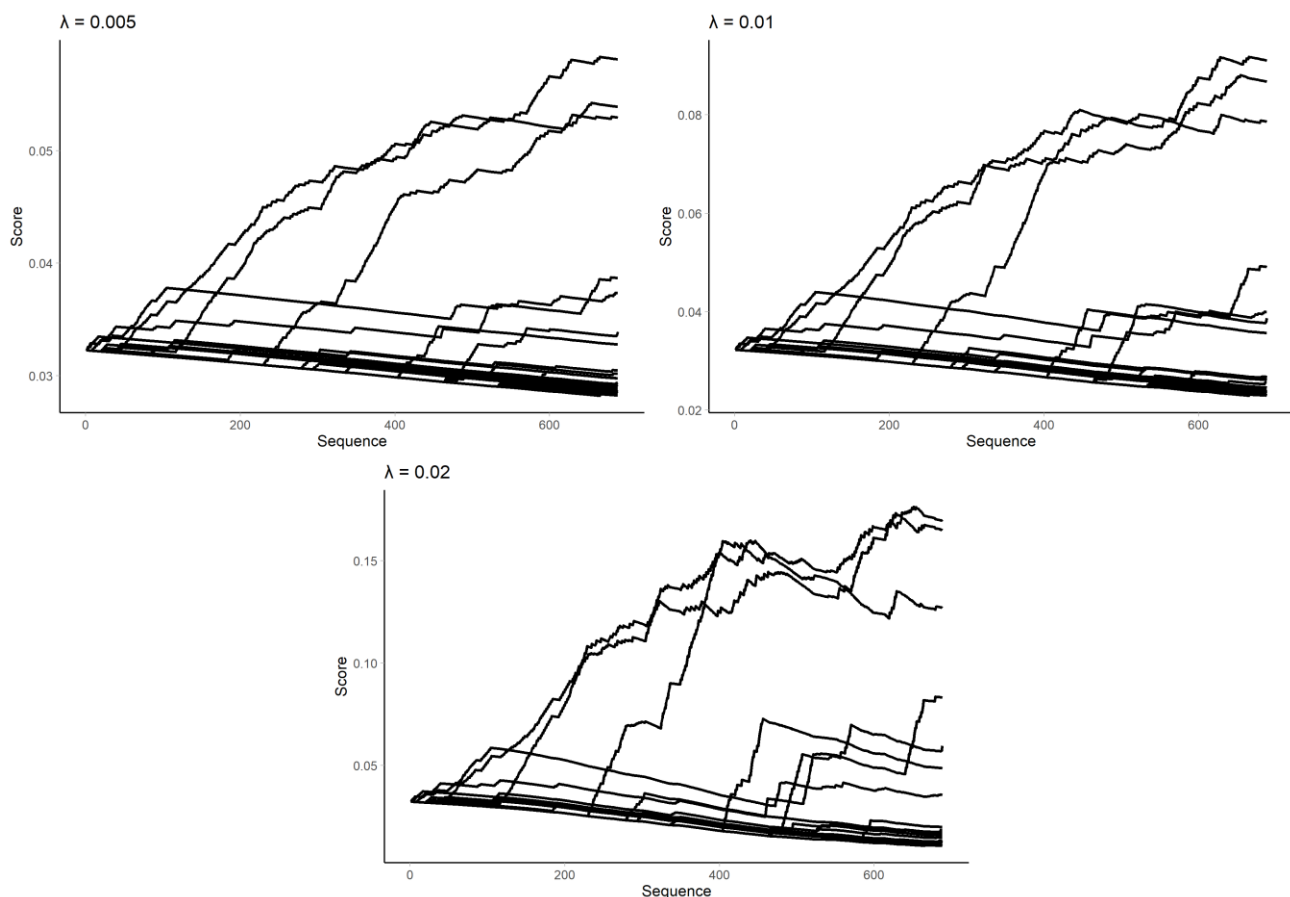


Figure A.6. Examining the effect of  $\lambda$  values between 0.005 and 0.02 on character trajectories.



and clarity. As discussed in Section 5, the  $\lambda$  value of 0.01 was chosen as a middle-ground between the insensitivity of 0.005 and the sensitivity of 0.02. Moreover, the trajectories mapped using the value of 0.01 appear to correspond with a sensible reading of how the narrative unfolds. This is not to say that 0.01 will be a useful value in all cases; for example, narratives with vastly fewer or more events may require more fine-tuning to ensure correspondence of output values to the narrative. However, for the present purposes, 0.01 was selected as a useful illustrative value.

To validate the results for *The Force Awakens* against a more extensive corpus, we calculate the proposed measures for each of the first 18 films of the Marvel Cinematic Universe (MCU) - from *Iron Man* (2008) to *Black Panther* (2018), each coded using exactly the same method as was used for *The Force Awakens*. This then gives us 19 different sets of character scores for each measure. Figures A.7a and A.7b provide a comparison of the final scores given by the proposed measures in these 19 films against two existing centrality indices for three different  $\lambda$  values (0.01, 0.05, and 0.1). Figure A.7a compares the proposed measure with the simple indegree and outdegree of each character. Figure A.7b compares the proposed measure with the total number of lines spoken and received. For all measures, scores have been normalised by dividing a character's score by the sum of the scores for all characters in that film. We again see the effect of higher  $\lambda$  values on the rank order of character importance: for  $\lambda$  values of 0.05 and 0.1, Kylo Ren is able to steal the show by overtaking Rey in importance despite speaking drastically fewer lines (Figure A.7b).

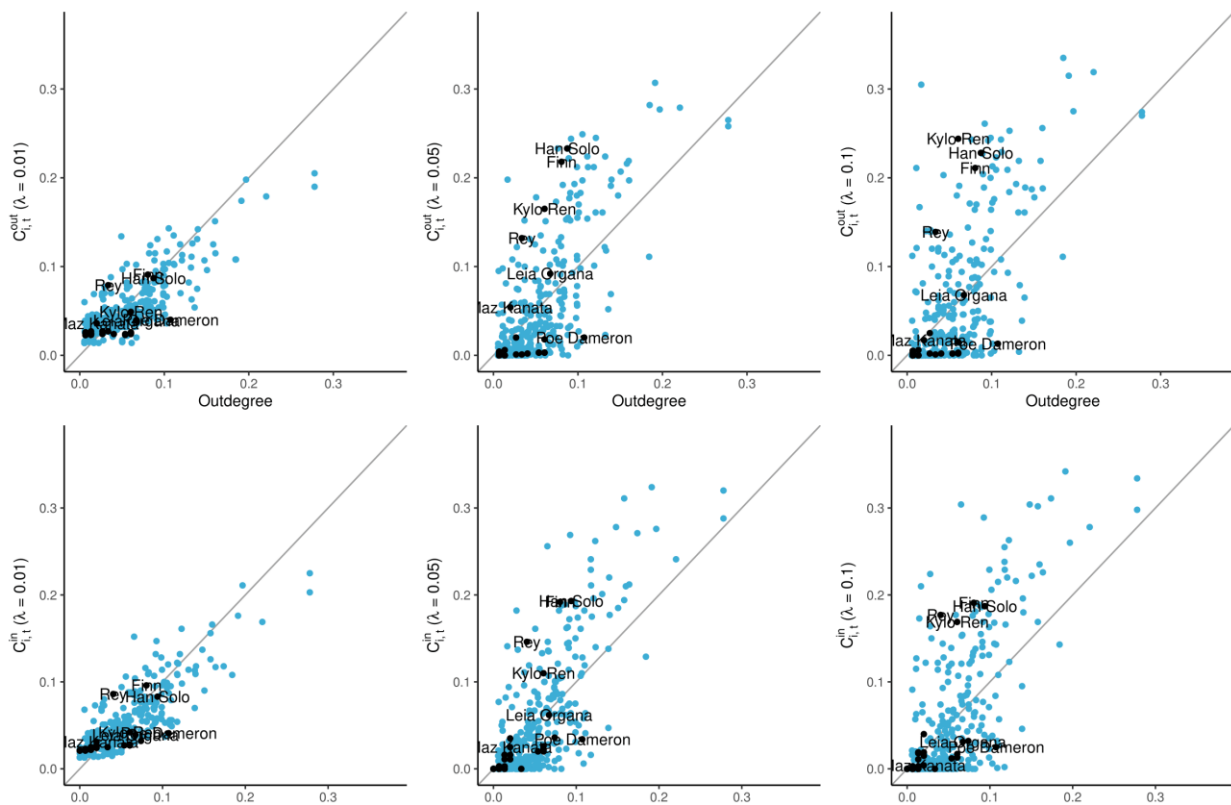


Figure A.7a. Speaking and spoken to measures against outdegree and indegree for  $\lambda = 0.01, 0.05$ , and  $0.1$ . Black points indicate characters from *The Force Awakens*, blue points indicate characters from 18 MCU films.

For each individual film we also calculate the Spearman rank correlation of the final normalised scores with the normalised outdegree/indegree and normalised number of lines spoken/received for a range of  $\lambda$  values. These correlations are plotted in Figure A.8. Figure A.8 shows that the final rank order of character importance given by the proposed measures is highly correlated with that given by the number of lines spoken/received, but that the association is weaker for higher  $\lambda$  values. The same cannot be said for the rank correlation between

the proposed measures and indegree/outdegree, where the association is weaker across each  $\lambda$  value. This illustrates that degree offers a picture of relative character importance that often corresponds only spuriously

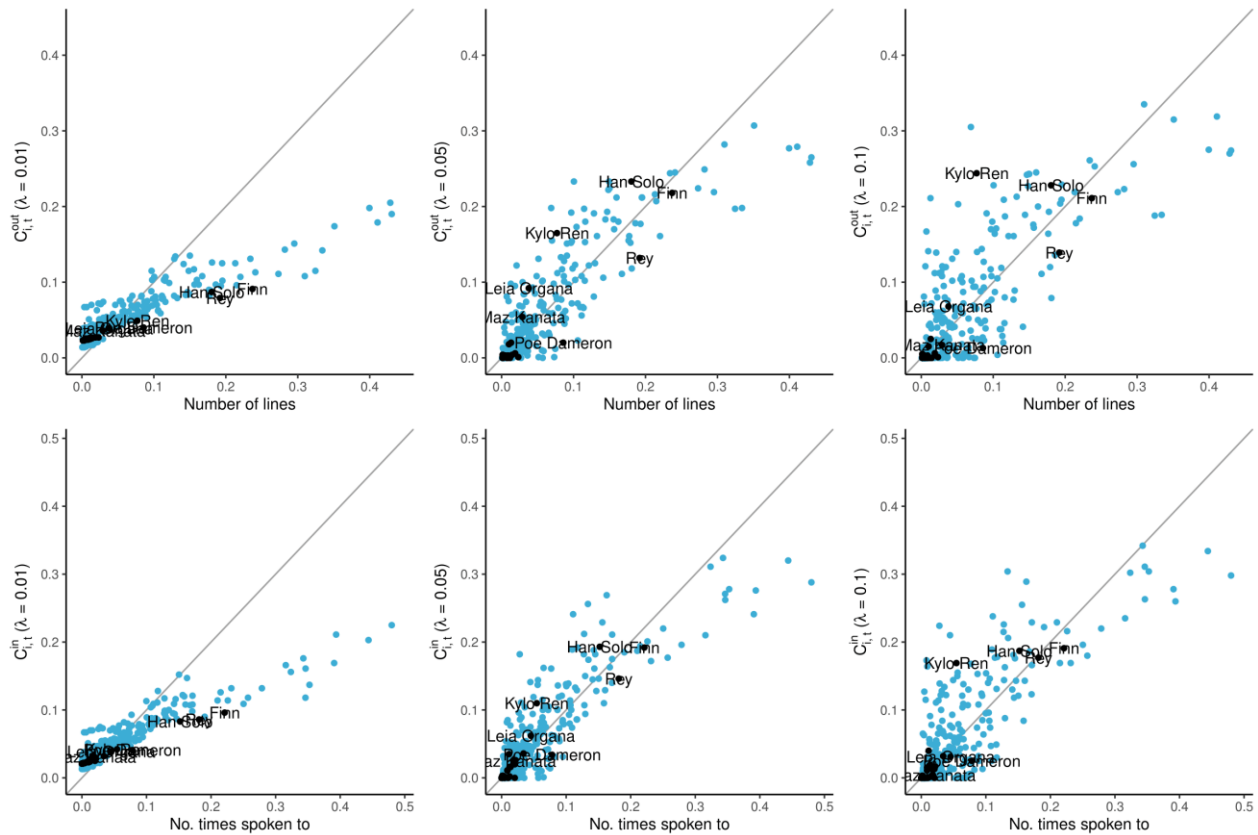


Figure A.7b. Speaking and spoken to measures against number of lines spoken and received for  $\lambda = 0.01, 0.05$ , and  $0.1$ . Black points indicate characters from *The Force Awakens*, blue points indicate characters from 18 MCU films.

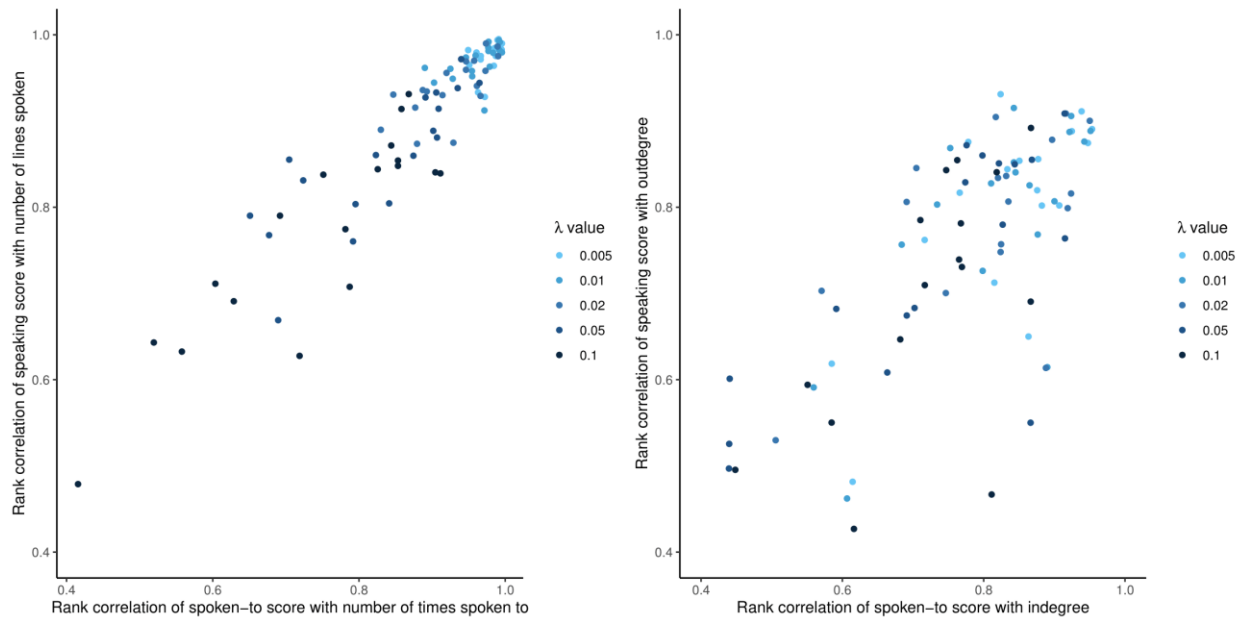


Figure A.8. Spearman's rank correlation coefficients for proposed measures against lines spoken/received and outdegree/indegree for five values of  $\lambda$ .

to the extent to which the narrative is told through those characters.

Finally, a summary measure across all  $\lambda$  is provided by area under the curve for the scores in Figure A.5. which may be obtained through numerically integrating the scores  $\int_0^1 C_i(\lambda) d\lambda$ . The total area under curve for each actor is provided in Figure A.9.

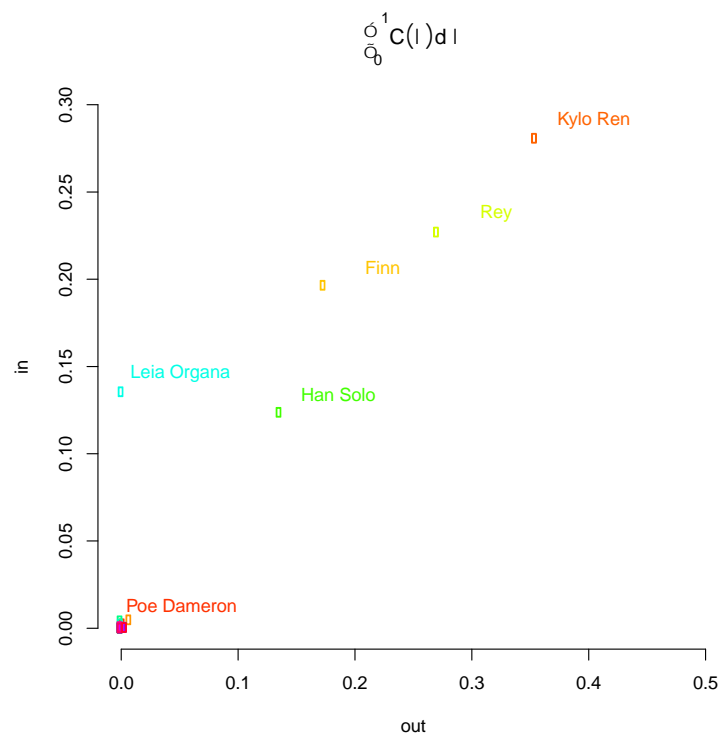


Figure A.9. Areas under the curves for speaking and spoken-to measures.

## References

- Abell, Peter. 2004. "Narrative Explanation: An Alternative to Variable-Centred Explanation?" *Annual Review of Sociology* 30: 287-310.
- Abbott, H. Porter. 2008. *The Cambridge Introduction to Narrative*. Cambridge: Cambridge University Press.
- Agarwal, Apoorv, Augusto Corvalan, Jacob Jensen and Owen Rambow. 2012. "Social Network Analysis of *Alice in Wonderland*." *Workshop on Computational Linguistics for Literature*, 88-96. Montréal, Canada, June 8 2012.
- Agarwal, Apoorv, Sriramkumar Balasubramanian, Jiehan Zheng and Sarthak Dash. 2014. "Parsing Screenplays for Extracting Social Networks from Movies." *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL)*, 50-58. Gothenburg, Sweden, April 27 2014.
- Amati, Viviana, Alessandro Lomi and Daniele Mascia. 2019. "Some days are better than others: Examining time-specific variation in the structuring of interorganizational relations." *Social Networks* 57: 18-33.
- Bal, Mieke. 2017. *Narratology: Introduction to the Theory of Narrative*. 4<sup>th</sup> ed. Toronto: University of Toronto Press.

- Bazzini, Doris G., William D. McIntosh, Stephen M. Smith, Sabrina Cook and Caleigh Harris. 1997. "The Aging Woman in Popular Film: Underrepresented, Unattractive, Unfriendly, and Unintelligent." *Sex Roles* 36 (7/8): 521-543.
- Bearman, Peter S. and Katherine Stovel. 2000. "Becoming a Nazi: A model for narrative networks." *Poetics* 27: 69-90.
- Bechdel, Alison. 1986. *Dykes to Watch Out For*. Ithaca: Firebrand Books.
- Bonacich, Phillip and Paulette Lloyd. 2001. "Eigenvector-like measures of centrality for asymmetric relations." *Social Networks* 23: 191-201.
- Bonacich, Phillip and Philip Lu. 2012. *Introduction to Mathematical Sociology*. Princeton: Princeton University Press.
- Bordwell, David. 1985. *Narration in the Fiction Film*. London: Routledge.
- Brandes, Ulrich, Garry Robins, Ann McRanie and Stanley Wasserman. 2013. "Editorial: What is network science?" *Network Science* 1 (1): 1-15.
- Branigan, Edward. 1992. *Narrative Comprehension and Film*. London: Routledge.
- Broccatelli, Chiara, Martin Everett and Johan Koskinen. 2016. "Temporal dynamics in covert networks." *Methodological Innovations* 9: 1-14.
- Chatman, Seymour. 1980. "What Novels Can Do That Films Can't (And Vice Versa)." *Critical Inquiry* 7 (1): 121-140.
- Cox, Susan. 2015. "Star Wars: *The Force Awakens* provides the kick-ass female lead we've been waiting for." *Feminist Current*, Dec, 27.
- Elson, David K., Nicholas Dames and Kathleen R. McKeown. 2010. "Extracting Social Networks from Literary Fiction". *ACL '10 Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*: 138-147.
- Everett, Martin G., Chiara Broccatelli, Stephen P. Borgatti and Johan Koskinen. 2018. "Measuring knowledge and experience in two mode temporal networks." *Social Networks* 55: 63-73.
- Faludi, Susan. 1992. *Backlash: The Undeclared War Against Women*. London: Vintage.
- Falzon, Lucia, Eric Quintane, John Dunn and Garry Robins. 2018. "Embedding time in positions: Temporal measures of centrality for social network analysis." *Social Networks* 54: 168-178.
- Fischer, Frank, Mathias Göbel, Dario Kampkaspar, Christopher Kittel and Peer Trilcke. 2017. "Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts." *Digital Humanities 2017* (Montréal, 8-11 August 2017). Book of Abstracts. Montréal: McGill University.
- Fermann, Lea, Shay B. Cohen and Mirella Lapata. 2018. "Whodunnit? Crime Drama as a Case for Natural Language Understanding." *Transactions of the Association for Computational Linguistics* 6: 1-15.
- Garber, Megan. 2015. "Star Wars: The Feminism Awakens." *The Atlantic*, Dec, 19.
- Genette, Gérard. 1980. *Narrative Discourse: An Essay in Method*. Ithaca: Cornell University Press.
- Gil, Sebastian, Laney Kuenzel, and Caroline Suen. 2011. "Extraction and analysis of character interaction networks from plays and movies." *Technical report*, Stanford University.
- Gould, Henry W. 1965. "A Variant of Pascal's Triangle." *Fibonacci Quarterly* 3 (4): 257-271.
- Grayson, Siobhán, Karen Wade, Gerardine Meaney and Derek Greene. 2016. "The Sense and Sensibility of Different Sliding Windows in Constructing Co-occurrence Networks from Literature." In *Computational History and Data-Driven Humanities*. Springer.

- Haskell, Molly. 1987. *From Reverence to Rape*, 2nd edn. Chicago: University of Chicago Press.
- Holme, Petter and Jari Saramäki. 2012. "Temporal networks." *Physics Reports* 519: 97-125.
- Inness, Sherrie A. 2004. *Action Chicks: New Images of Tough Women in Popular Culture*. New York: Palgrave Macmillan.
- Jayannavar, Prashant Arun, Melody Ju, Apoorv Agarwal and Owen Rambow. 2015. "Validating Literary Theories Using Automatic Network Analysis." *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*: 32-41. Denver, Colorado, June 4, 2015.
- Jones, Pete. 2018. "Diana in the World of Men: a character network approach to analysing gendered vocal representation in *Wonder Woman*." *Feminist Media Studies*, DOI: 10.1080/14680777.2018.1510846.
- Kagan, Dima, Thomas Chesney and Michael Fire. 2019. "Using Data Science to Understand the Film Industry's Gender Gap." *arXiv:1903.06469*.
- Kaplan, E. Ann. 1983. *Women and Film: Both Sides of the Camera*. New York and London: Methuen.
- Katz, Leo. 1953. "A new status index derived from sociometric analysis." *Psychometrika* 18 (1): 39-43.
- Kermode, Mark. 2015. "Star Wars: The Force Awakens review - consider me conquered." *The Guardian*, Dec, 20.
- Kontoleon, Nectarios, Lucia Falzon and Philippa Pattison. 2013. "Algebraic structures for dynamic networks." *Journal of Mathematical Psychology* 57 (6): 310-319.
- Kydros, Dimitrios, Panagiotis Notopolous and Georgios Exarchos. 2015. "Homer's Iliad - A Social Network Analytic Approach." *International Journal of Humanities and Arts Computing* 9 (1): 115-132.
- Lauzen, Martha. 2018. "It's a Man's (Celluloid) World: Portrayals of Female Characters in the Top 100 Films of 2017." *Center for the Study of Women in Television & Film* (San Diego State University). Accessed February 12, 2019. [https://womenintvfilm.sdsu.edu/wp-content/uploads/2018/03/2017\\_Its\\_a\\_Mans\\_Celluloid\\_World\\_Report\\_3.pdf](https://womenintvfilm.sdsu.edu/wp-content/uploads/2018/03/2017_Its_a_Mans_Celluloid_World_Report_3.pdf).
- Lauzen, Martha M. and David M. Dozier. 2005. "Maintaining the Double Standard: Portrayals of Age and Gender in Popular Films." *Sex Roles* 52 (7/8): 437-446.
- Ly, Jinna, Bin Wu, Lili Zhou and Han Wang. 2018. "StoryRoleNet: Social Network Construction of Role Relationship in Video." *IEEE Access*, DOI: 10.1109/ACCESS.2018.2832087.
- Masías, Víctor Hugo, Paula Baldwin, Sigifredo Laengle, Augusto Vargas and Fernando A. Crespo. 2017. "Exploring the prominence of Romeo and Juliet's characters using weighted centrality measures." *Digital Scholarship in the Humanities* 32 (4): 837-858.
- McRobbie, Angela. 2004. "Post-feminism and popular culture." *Feminist Media Studies* 4 (3): 255-264.
- Min, Semi and Juyong Park. 2016. "Mapping Out Narrative Structures and Dynamics Using Networks and Textual Information." *arXiv:1604.03029*.
- Moody, James. 2002. "The Importance of Relationship Timing for Diffusion." *Social Forces* 81 (1): 25-56.
- Moody, James, Daniel McFarland and Skye Bender-deMoll. 2005. "Dynamic Network Visualization." *American Journal of Sociology* 110 (4): 1206-1241.
- Moretti, Franco. 2011. "Network Theory, Plot Analysis." *New Left Review* 68: 80-102.
- Mourchid, Youssef, Benjamin Renoust, Hocine Cherifi and Mohammed El Hassouni. 2018. "Multilevel Network Model of Movie Script." *arXiv:1812.05718v1*.
- Mulvey, Laura. 1975. "Visual Pleasure and Narrative Cinema." *Screen* 16 (3): 6-18.

- Niehaus, James and R. Michael Young. 2014. "Cognitive models of discourse comprehension for narrative generation." *Literary and Linguistic Computing* 29 (4): 561-582.
- O'Day, Marc. 2004. "Beauty in Motion: Gender, Spectacle and Action Babe Cinema." In *Action and Adventure Cinema*, edited by Yvonne Tasker. Abingdon: Routledge.
- O'Meara, Jennifer. 2016. "What 'The Bechdel Test' doesn't tell us: examining women's verbal and vocal (dis)empowerment in cinema." *Feminist Media Studies*, DOI: 10.1080/14680777.2016.1234239.
- Opsahl, Tore, Filip Agneessens, John Skvoretz. 2010. "Node centrality in weighted networks: Generalizing degree and shortest paths." *Social Networks* 32 (3): 245-251.
- Padgett, John F. 2018. "Faulkner's Assembly of Memories into History: Narrative Networks in Multiple Times." *American Journal of Sociology* 124 (2): 406-478.
- Prado, Sandra D., Silvio R. Dahmen, Ana L. C. Bazzan, Pádraig Mac Carron and Ralph Kenna. 2016. "Temporal Network Analysis of Literary Texts." *arXiv:1602.07275*.
- Purse, Lisa. 2011. *Contemporary Action Cinema*. Edinburgh: Edinburgh University Press.
- Ricœur, Paul. 1980. "Narrative Time." *Critical Inquiry* 7 (1): 169-190.
- Roddy, Michael. 2015. "Review - New *Star Wars: The Force Awakens* lets women take the lead." *Reuters*, Dec, 16.
- Ruegg, Claire and James Jaehoon Lee. 2019. "Epic social networks and Eve's centrality in Milton's *Paradise Lost*." *Digital Scholarship in the Humanities*. DOI: 10.1093/llc/fqz001.
- Sack, Graham Alexander. 2014. "Character Networks for Narrative Generation: Structural Balance Theory and the Emergence of Proto-Narratives." In *Complexity and the Human Experience: Modeling Complexity in the Humanities and Social Sciences*, edited by Paul A. Youngman and Mirsad Hadzikadic. Boca Raton: CRC Press.
- Schoch, David. 2019. *Personal communication*.
- Schubart, Rikke. 2007. *Super Bitches and Action Babes: The Female Hero in Popular Cinema, 1970-2006*. Jefferson, NC: McFarland & Co.
- Smith, Stacy L., Marc Choueiti, Katherine Pieper, Ariana Case and Angel Choi. 2018. "Inequality in 1,100 Popular Films: Examining Portrayals of Gender, Race/ Ethnicity, LGBT, and Disability from 2007 to 2017." *Annenberg Inclusion Initiative Initiative*. Accessed February 12, 2019. <http://assets.uscannenberg.org/docs/inequality-in-1100-popular-films.pdf>.
- Smith, Tammy. 2007. "Narrative boundaries and the dynamics of ethnic conflict and conciliation." *Poetics* 35: 22-46.
- Sutherland, Jean-Anne and Kathryn M. Feltey. 2016. "Here's looking at her: an intersectional analysis of women, power and feminism in film." *Journal of Gender Studies*, DOI: 10.1080/09589236.2016.1152956.
- Tasker, Yvonne. 1993. *Spectacular Bodies: Gender, genre and the action cinema*. London: Routledge.
- Taylor, Dane, Sean A. Myers, Aaron Clauset, Mason A. Porter and Peter J. Mucha. 2017. "Eigenvector-Based Centrality Measures for Temporal Networks." *Multiscale Modeling & Simulation* 15 (1): 537-574.
- Waumans, Michaël C., Thibaut Nicodème, and Hugues Bersini. 2015. "Topology Analysis of Social Networks Extracted from Literature." *PLoS ONE* 10 (6), DOI: 10.1371/journal.pone.0126470.
- Weng, Chung-Yi, Wei-Ta Chu, and Ja-Ling Wu. 2009. "RoleNet: Movie Analysis from the Perspective of Social Networks." *IEEE Transactions on Multimedia* 11 (2): 256-271.
- Woloch, Alex. 2003. *The One vs. the Many: Minor Characters and the Space of the Protagonist in the Novel*. Princeton: Princeton University Press.