

# Linguistic analysis of differences in portrayal of movie characters

Anil Ramakrishna<sup>1</sup>, Victor R. Martínez<sup>1</sup>, Nikolaos Malandrakis<sup>1</sup>, Karan Singla<sup>1</sup>, and Shrikanth Narayanan<sup>1,2</sup>

<sup>1</sup>Department of Computer Science

<sup>2</sup>Department of Electrical Engineering

University of Southern California, Los Angeles, USA

{akramakr, victorrm, malandra, singlak}@usc.edu, shri@sipi.usc.edu

## Abstract

We examine differences in portrayal of characters in movies using psycholinguistic and graph theoretic measures computed directly from screenplays. Differences are examined with respect to characters' gender, race, age and other metadata. Psycholinguistic metrics are extrapolated to dialogues in movies using a linear regression model built on a set of manually annotated seed words. Interesting patterns are revealed about relationships between genders of production team and the gender ratio of characters. Several correlations are noted between gender, race, age of characters and the linguistic metrics.

## 1 Introduction

Movies are often described as having the power to influence individual beliefs and values. In (Cape, 2003), the authors assert movies' influence in both creating new thinking patterns in previously unexplored social phenomena, especially in children, as well as their ability to update an individual's existing social boundaries based on what is shown on screen as the "norm". Some authors claim the inverse (Wedding and Boyd, 1999): that movies reflect existing cultural values of the society, adding weight to their ability in influencing individual beliefs of what is accepted as the norm. As a result, they are studied in multiple disciplines to analyze their influence.

Movies are particularly scrutinized in aspects involving negative stereotyping (Cape, 2003; Dimnik and Felton, 2006; Ter Bogt et al., 2010; Hedley, 1994) since this may introduce questionable beliefs in viewers. Negative stereotyping is believed to impact society in multiple aspects such as self-induced undermining of ability (Davies

et al., 2005) as well as causing forms of prejudice that can impact leadership or employment prospects (Eagly and Karau, 2002; Niven, 2006). Studies in analyzing stereotyping in movies typically rely on collecting manual annotations on a small set of movies on which hypotheses tests are conducted (Behm-Morawitz and Mastro, 2008; Benshoff and Griffin, 2011; Hooks, 2009). In this work, we present large scale automated analyses of movie characters using language used in dialogues to study stereotyping along factors such as gender, race and age.

Language use has been long known as a strong indicator of the speaker's psychological and emotional state (Gottschalk and Gleser, 1969) and is well studied in a number of applications such as automatic personality detection (Mairesse et al., 2007) and psychotherapy (Xiao et al., 2015; Pennebaker et al., 2003). Computational analysis of language has been particularly popular thanks to advancements in computing and the ease of conducting large scale analysis of text on computers (Pennebaker et al., 2015).

To perform our analysis, we construct a new movie screenplay corpus<sup>1</sup> that includes nearly 1000 movie scripts obtained from the Internet. For each movie in the corpus, we obtain additional metadata such as cast, genre, writers and directors, and also collect actor level demographic information such as gender, race and age.

We use two kinds of measures in our analyses: (i) linguistic metrics that capture various psychological constructs and behaviors, estimated using dialogues from the screenplay; and (ii) graph theoretic metrics estimated from character network graphs, which are constructed to model inter-character interactions in the movie. The linguistic metrics include psycholinguistic normatives,

<sup>1</sup>[http://sail.usc.edu/mica/text\\_corpus\\_release.php](http://sail.usc.edu/mica/text_corpus_release.php)

which provide word level scores on a numeric scale which are then aggregated at the dialog level, and metrics from the Linguistic Inquiry and Word Counts tool (LIWC) which capture usage of well studied stereotyping dimensions such as sexuality. We estimate centrality metrics from the character network graphs to measure relative importance of the different characters, which are analyzed with respect to the different factors of gender, race and age.

The main contributions of this work are as follows: (i) we present a scalable analysis of differences in portrayal of various character subgroups in movies using their language use, (ii) we construct a new corpus with detailed annotations for our analysis and (iii) we highlight several differences in the portrayal of characters along factors such as race, age and gender.

The rest of the paper is organized as follows: in section 2 we describe related work. We explain the data collection process in section 3 and experimental procedure in section 4. We explain results in section 5 and conclude in section 6.

## 2 Related work

Previous works in studying representation in movies largely focus on relative frequencies, particularly on character gender. In (Smith et al., 2014), the authors studied 120 movies from around the globe which were manually annotated to capture information about character gender, age, careers, writer gender and director gender. However, since the annotations are done manually, collecting information on new movies is a laborious process. We avoided this by estimating the metadata computationally, enabling us to scale up efficiently.

Automated analyses of movies using computational techniques to analyze representation has recently gained some attention. In (NYFA, 2013; Polygraph, 2016), the authors examine differences in relative frequency of female characters and note considerable disparities in gender ratio in these movies. However, the analyses there too are limited to comparing relative frequencies. Our work is closest to (Ramakrishna et al., 2015) where the authors study difference in language used in movies across genders, but their analysis is one dimensional. In our work we perform fine grained comparisons of character portrayal using multiple language based metrics along factors such as gen-

der, race and age on a newly created corpus.

## 3 Data

### 3.1 Raw screenplay

We fetch movie screenplay files from two primary sources: *imsdb* (IMSDb, 2017) and *daily scripts* (DailyScript, 2017). In total, we retrieved 1547 movies. After removing duplicates we retain 1434 raw screenplay files, of which 489 were corrupted or empty leaving us with 945 usable screenplays. Tables 1, 3 and 4 list statistics about the corpus.

### 3.2 Script parser

The screenplay files are formatted in human readable format and include dialogues tagged with character names along with auxiliary information of the scene such as shot location (interior/exterior), character placement and scene context. The screenplays are from a diverse set of writers and include a significant amount of noise and inconsistencies in their structure. To extract the relevant information, we developed a text parser<sup>2</sup> that accepts raw script files and outputs utterances along with character names. We ignore scene context information and primarily focus on spoken dialogues to study language usage in the movies.

### 3.3 Movie and character meta-data

For each parsed movie, we fetch relevant meta-data such as year of release, directors, writers, and producers from the Internet Movie Database (IMDb, 2017).

Since most screenplays are drafts and subject to revisions such as changes in character names, matching them to an entry from IMDb is non-trivial. We first start with a list of all movies that have a close match with the screenplay name; given this list of potential matches we compute name alignment scores for each entry as the percentage of character names from the script found online. The character names are mapped using term frequency-inverse document frequency (TFIDF) to compute the name alignment score following (Cohen et al., 2003). Finally, the entry with highest alignment score is chosen. For all actors listed in the aligned result, we collect their age, gender and race as detailed below.

<sup>2</sup>[https://bitbucket.org/anil\\_ramakrishna/scriptparser](https://bitbucket.org/anil_ramakrishna/scriptparser)

### 3.3.1 Gender

Given the names of actors and other members of production team found in a movie, we use a name based gender classifier to predict their gender information. Table 4 lists statistics on gender ratios for the production team in the corpus. Female-to-male ratios were found in close agreement with previous works (Smith et al., 2014).

As mentioned above, several screenplays get revised during production. In particular character names get changed, sometimes even gender. As a result, some characters may not be aligned to the correct entry from IMDb. In addition, digitized screenplays sometime include significant noise thanks to optical character recognition errors, leading to character names failing to align with entries from IMDb. To correct these, we perform manual cleanup of all the movie alignments, fix incorrect gender maps, and manually force match movies if they’re mapped to the wrong IMDb entry.

### 3.3.2 Age

We also extract age for each actor to study possible age related biases in movies. We include age in our analysis since studies report preferential biases with age in employment particularly when combined with gender (Lincoln and Allen, 2004). In addition, there may be biases in portrayal of specific age groups when combined with gender and race.

For each actor in the mapped IMDb entry, we collect his/her birthday information. We subtract the movie production year obtained also from IMDb from the actor’s birthday to get an estimate of the actor’s age during the movie’s production. We note however that the age obtained in this manner may be different from the portrayed age of the character. To account for this we bin the actors into fifteen year age groups before our analysis, since its generally unlikely to have actors further than fifteen years from their portrayed age.

### 3.3.3 Race

We parse ethnicity information from the website (ethnicebs.com, 2017), which includes ethnicity for approximately 8000 different actors. The information obtained from this site is primarily submitted by independent users, and exhibits significant amount of variation among the possible ethnicities with about 750 different unique ethnicity types. Since we are more specifically interested in

<i>Race</i>	<i># Actors</i>	<i>Percentage</i>
African	585	7.44%
Caucasian	6539	83.24%
East Asian	73	0.93%
Latino/Hispanic	161	2.05%
Native American	15	0.19%
Pacific Islander	5	0.063%
South Asian	43	0.547%
Mixed	434	5.52%

Table 1: Racial categories

racial representations, we map the ethnicity types to race using Amazon Mechanical Turk (MTurk). We use a modified version of the racial categories from the US census which are listed in Table 1 along with frequency of actors from each racial category in our corpus.

The ethnicities obtained from the site above primarily cover major actors with a fan base with no information for several actors who play minor roles. We annotate racial information for nearly 2000 such actors using MTurk with two annotations for each actor, manually correcting nearly 400 cases in which the annotators disagreed.

## 4 Experiments

### 4.1 Character portrayal using language

To study differences in portrayal of characters, we use two different metrics: psycholinguistic normatives, which are designed to capture the underlying emotional state of the speaker; and LIWC metrics, which provide a measure of the speaker’s affinity to different social and physical constructs such as religion and death. We explain these two metrics in detail below.

#### 4.1.1 Psycholinguistic normatives

Psycholinguistic normatives provide a measure of various emotional and psychological constructs of the speaker, such as arousal, valence, concreteness, intelligibility, etc. and are computed entirely from language usage. They are relatively easy to compute, provide reliable indicators of the above constructs, and have been used in a variety of tasks in natural language processing such as information retrieval (Tanaka et al., 2013), sentiment analysis (Nielsen, 2011), text based personality prediction (Mairesse et al., 2007) and opinion mining.

The numeric ratings are typically extrapolated from a small set of keywords which are annotated

by psychologists. Manual annotations of word ratings is a laborious process and is hence limited to a few thousand words (Clark and Paivio, 2004). Automatic extrapolation of these ratings to words not covered by the manual annotations can be done using structured databases which provide relationships between words such as synonymy and hyponymy (Liu et al., 2014), or using context based semantic similarity.

In this work, we use the model described in (Malandrakis and Narayanan, 2015) where the authors use linear regression to compute normative scores for an input word  $w$  based on its similarity to a set of concept words  $s_i$ .

$$r(w) = \theta_0 + \sum_i \theta_i \cdot sim(w, s_i) \quad (1)$$

where,  $r(w)$  is the computed normative score for word  $w$ ,  $\theta_0$  and  $\theta_i$  are regression coefficients and  $sim$  is similarity between the given word  $w$  and concept words  $s_i$ .

The concept words can either be hand crafted suitably for the domain or chosen automatically from data. Similar to (Malandrakis and Narayanan, 2015), we create training data by posing queries on the Yahoo search engine from words of the aspell spell checker of which top 500 previews are collected from each query. From this corpus, the top 10000 most frequent words with atleast 3 characters were used as concept words in extrapolation of all the norms. The linear regression model is trained using normative ratings for the manually annotated words by computing their similarity to the concept words. The similarity function  $sim$  is the cosine of binary context vectors with window size 1. The computed normatives are in the range  $[-1, 1]$ .

The psycholinguistic normatives used in this work are listed in Table 2. *Valence* is the degree of positive or negative emotion evoked by the word. *Arousal* is a measure of excitement in the speaker. Valence and arousal combined are common indicators used to map emotions. *Age of Acquisition* refers to the average age at which the word is learned and it denotes sophistication of language use. *Gender Ladenness* is a measure of masculine or feminine association of a word. 10 fold Cross Validation tests are performed on the normative scores predicted by the regression model given by equation 1. Correlation coefficients of the selected normatives with the manual annota-

tions are as follows: Arousal (0.7), Valence (0.88), Age of Acquisition (0.86) and Gender Ladenness (0.8). The high correlations render confidence in the psycholinguistic models.

In our experiments, the normative scores are computed on content words from each dialog. We filter out all words other than nouns, verbs, adjectives and adverbs. Word level scores are aggregated at the dialog level using arithmetic mean.

#### 4.1.2 Linguistic inquiry and word counts (LIWC)

LIWC is a text processing application that processes raw text and outputs percentage of words from the text that belong to linguistic, affective, perceptual and other dimensions. It operates by maintaining a diverse set of dictionaries of words each belonging to a unique dimension. Input texts are processed word by word; each word is searched in the internal dictionaries and the corresponding counter is incremented if a word is found in that dictionary. Finally, percentage of words from the input text belonging to the different dimensions are returned.

For our experiments, we treat each utterance in the movie as a unique document and obtain values for the LIWC metrics. Table 2 lists the metrics used in our experiments.

#### 4.2 Character network analytics

In order to study representation of the different subgroups as major characters in movies, we construct a network of interaction between characters using which we compute importance measures for each character. From each movie script, we construct an undirected and unweighted graph where nodes represent characters. We place an edge  $e_{ab}$  if two characters A and B interact at least once in the movie. For our experiments we assume interaction between A and B if there is at least one scene in which one speaks right after another. This graph creation method based on scene co-occurrence is similar to the approach used in (Beveridge and Shan, 2016).

We estimate different measures of a node’s importance within the character network and use it as proxy for the character’s importance. We employ two types of centralities: *betweenness centrality*, the number of shortest paths that go through the node, and *degree centrality*, which is the number of edges incident on a node. These centrality measurements have been previously used in the con-



Psycholinguistic norms	Valence, Arousal, Age of Acquisition, Gender Ladeness
LIWC metrics	Achievement, Religion, Death, Sexual, Swear

Table 2: Psycholinguistic Normatives and LIWC metrics used in analysis

	<i>male</i>	<i>female</i>	<i>total</i>
# Characters	4899	2008	6907
# Dialogues	375711	154897	530608
Number of movies			945

Table 3: Character statistics

<i>role</i>	<i>male</i>	<i>female</i>	<i>total</i>
Writers	1326	169	1495
Directors	544	46	590
Producers	2866	870	3736
Casting Directors	135	275	410
Distributing Companies			2701

Table 4: Production team statistics

text of books, films and comics (Beveridge and Shan, 2016; Bonato et al., 2016; Alberich et al., 2002; Ribeiro et al., 2016).

## 5 Results

We study differences in various subgroups along multiple facets. We first report results on differences in character ratios from each subgroup since this has implications on employment and can have social-economic effects (Niven, 2006). We next use psycholinguistic normatives and LIWC metrics described in the previous section to study differences in character portrayal along the primary markers: age, gender and race. We finally use the graph theoretic centrality measures to estimate characters' importance and analyze differences among the different subgroups.

Since we are interested in character level analytics, we treat all utterances from the character as a single document to compute the aggregate language metrics. We perform all our experiments using non-parametric statistical tests since the data fails to satisfy preconditions such as normality and homoscedasticity required for parametric tests such as ANOVA.

### 5.1 Difference in relative frequency of subgroups

We first filter our characters with unknown gender/race/age leaving us with 6907 characters in to-

<i>character genders</i>		
	<i>f</i> (28.9%)	<i>m</i> (71.1%)
<i>f</i>	249 (41.2%)	356 (58.8%)
<i>m</i>	1541 (27.6%)	4040 (72.4%)

(a) writers gender

<i>f</i>	114 (39.3%)	176 (60.7%)
<i>m</i>	1676 (28.4%)	4220 (71.6%)

(b) directors gender

<i>f</i>	1374 (29.1%)	3350 (70.9%)
<i>m</i>	416 (28.5%)	1046 (71.5%)

(c) casting directors gender

Table 5: Contingency tables for character gender v/s writers, directors and casting directors' gender; f: female and m: male; each cell gives frequency of character gender for that column and production member gender for that row, numbers in braces indicate row wise proportion of character gender

tal. Table 3 lists the number of characters and dialogues from each gender. As noted in previous studies, the ratio is considerably skewed with male actors having nearly twice as many roles and dialogues compared to female actors. Table 4 lists relative frequency among male and female members of the production team. Table 1 lists the percentage of actors belonging to different racial categories in the corpus.

We perform chi-squared tests between character gender and gender of production team members who are most likely to influence characters gender: writers, directors and casting directors. Table 5 shows contingency tables with gender frequencies for each of these cases along with percentages. Note we filter out nearly 100 movies for this test in which the gender of the production team members was unknown. Of the three tests we perform, character gender distributions for writer and director genders are significantly different from the overall character gender distribution ( $p < 10^{-10}$  and  $p < 10^{-4}$  respectively;  $\alpha = 0.05$ ). In particular, female writers and directors appear to produce movies with relatively balanced gender proportions (still slightly skewed towards the male side) compared to male writers

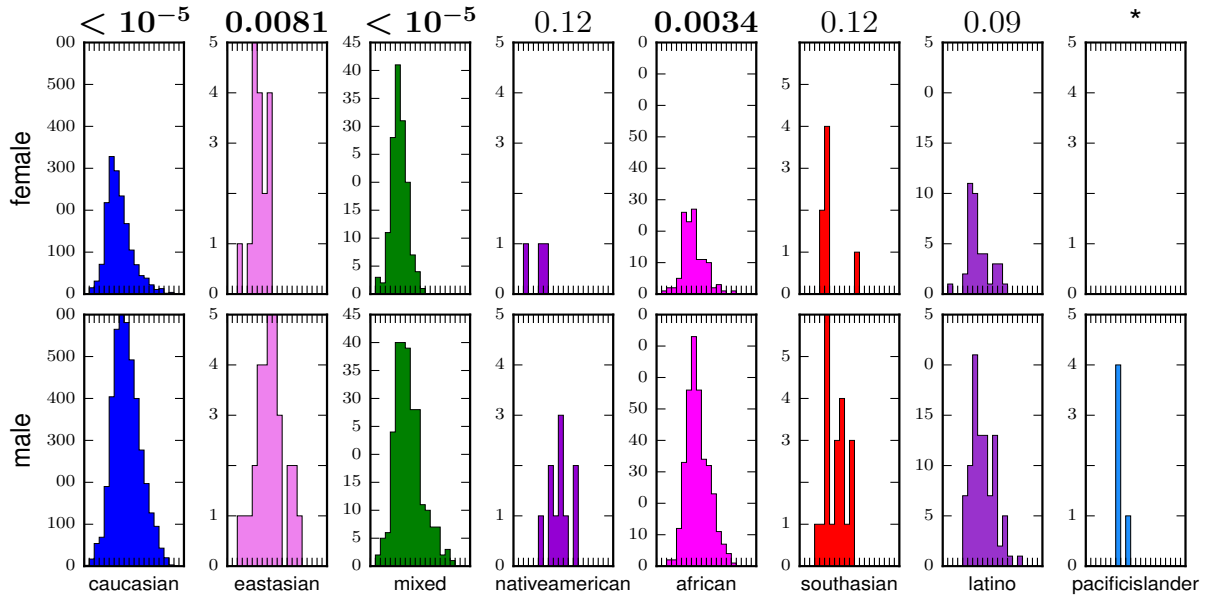


Figure 1: Histogram of age for actors belonging to different gender and racial categories with p-values on top; significant values at  $\alpha = 0.05$  are highlighted; \*: no test performed since the female group is empty

and directors. Casting directors however appear to have no influence on gender of the characters.

Studies report potential biases in actor employment with age (Lincoln and Allen, 2004), particularly in female actors. To evaluate this, we plot histograms of age for male and female characters for each of the racial categories in Figure 1. The distribution of age for each category appears approximately normal, except for the *nativeamerican* and *pacificislander* character groups which have a small sample size. For most categories of race, the mode of the distribution for female actors appears to be at least five years less than the mode for male actors. To check for significance in this difference we conduct Mann-Whitney U tests on male and female age groups for each race with the resulting p-values shown in the figure. We ignore characters belonging to the *pacificislander* racial group since there are no female actors from this race in our corpus. The difference in age groups is significant in most categories with large sample sizes, suggesting possible preferences towards casting younger people when casting female actors.

## 5.2 Character portrayal using language

To analyze differences in portrayal of subgroups, we compute psycholinguistic normatives and LIWC metrics as described before. For each of the metrics listed in Table 2, we conduct non-

	<i>m</i> (4894)	<i>f</i> (2008)	<i>p</i>
<i>age of acq.</i>	<b>-0.1590</b>	<b>-0.1715</b>	$< 10^{-5}$
<i>arousal</i>	0.0253	0.0246	0.41
<i>gender</i>	<b>-0.0312</b>	<b>-0.0055</b>	$< 10^{-5}$
<i>valence</i>	<b>0.2284</b>	<b>0.2421</b>	$< 10^{-5}$
<i>sex</i>	0.00015	0.0000	0.08
<i>achieve</i>	<b>0.0087</b>	<b>0.0080</b>	$< 10^{-5}$
<i>religion</i>	0.0025	0.0022	0.10
<i>death</i>	<b>0.0025</b>	<b>0.0016</b>	$< 10^{-5}$
<i>swear</i>	<b>0.0037</b>	<b>0.0015</b>	$< 10^{-5}$

Table 6: Median values for male and female characters along with p values obtained by comparing the two groups using Mann-Whitney U test; highlighted differences are significant at  $\alpha = 0.05$

parametric hypothesis tests to look for differences in samples from the subgroups. We treat the different metrics independently, performing statistical tests along each separately. We avoid statistical tests combining two or more factors since some of the resulting groups would be empty due to the skewed group sizes along race. We defer such analyses to future work.

### 5.2.1 Gender

We perform Mann-Whitney U tests between male and female characters along the nine dimensions and the results are shown in Table 6. In all of

the cases, higher values imply higher degree of the corresponding dimension, except for valence in which higher values imply positive valence (attractiveness) and lower values imply negative valence (averseness). The difference between male and female characters are statistically significant along six of the nine dimensions. The results indicate slightly higher age of acquisition scores for male characters. Regarding gender ladenness, male characters appear to be closer to the masculine side than female characters on average, agreeing with previous results.

Our results also indicate that female character utterances tend to be more positive in valence compared to male characters while male characters seem to have higher percentage of words related to achievement. In addition, male characters appear to be more frequent in using words related to death as well as swear words compared to female characters.

### 5.2.2 Race

To study differences in portrayal of the racial categories, we perform Kruskal-Wallis test (a generalization of Mann-Whitney U test for more than two groups) on each of the nine metrics with race as the independent variable. We found significant differences in distribution of samples for gender ladenness, sexuality, religion and swear words. For gender ladenness, *caucasian* and *mixed* race characters have significantly higher medians than *african* and *nativeamerican* characters. In sexuality, *latino* and *mixed* race characters were found to have higher median than at least one other racial group with significance indicating a higher degree of sexualization in these characters. *Eastasian* characters were found to be significantly lower than medians of three other races (*caucasian*, *african* and *mixed*) in using words with religious connotations. In swear word usage, the only significant difference found is between *caucasian* and *african* characters with *african* characters using higher percentage of swear words. In all of the above cases, significance was tested at  $\alpha = 0.05$ .

### 5.2.3 Age

To examine the relationship between age and the different metrics, we build separate linear regression models with each dimension as the dependent variable and character age as the independent variable. Table 7 reports regression coefficients for age along with p values for each dimension. The

	$\beta_1 (\times 10^{-3})$	$p$ -value
<i>age of acq.</i>	<b>3.9</b>	$< 10^{-10}$
<i>arousal</i>	<b>-1.1</b>	$< 10^{-10}$
<i>gender</i>	<b>-2.5</b>	$< 10^{-10}$
<i>valence</i>	0.078	0.7
<i>sex</i>	<b>-0.25</b>	$< 10^{-5}$
<i>achieve</i>	<b>0.26</b>	$< 10^{-10}$
<i>religion</i>	<b>0.12</b>	0.001
<i>death</i>	-0.039	0.2
<i>swear</i>	<b>-0.34</b>	$< 10^{-5}$

Table 7: Coefficients of age for linear regression models along each dimension along with p-values; highlighted cells are significant at  $\alpha = 0.05$

positive coefficient for age of acquisition indicates an increase in sophistication of word usage with age. Arousal, on the other hand, has a significant negative coefficient indicating a decrease in activation, on average, as character age increases. Gender ladenness also has a significant negative coefficient indicating that as age increases, the average gender ladenness value decreases. Similar trends are observed for sexuality and swear word usage. Usage of words related to achievement and religion however, seem to increase with age.

## 5.3 Character network analytics

To study differences in major roles assigned to the different subgroups, we compute two centrality metrics from the character network graph constructed for each movie: *degree* centrality measures the number of unique characters that interact with a given character, *betweenness* centrality measures how much would the plot be disrupted if said character was to disappear completely, i.e., how important is a character to the overall plot. Similar to the language analyses from previous section, we test differences in these metrics along the three factors of gender, race and age. All statistical tests reported below are conducted at  $\alpha = 0.05$ .

### 5.3.1 Gender

Male characters were found to have higher values in the two metrics compared to female characters but the differences were not statistically significant. Motivated by studies (Sapolsky et al., 2003; Linz et al., 1984) which report interactions between genre and gender, we performed Mann-Whitney U tests between male and female char-

acters given different genres. To avoid type I errors we corrected for multiple comparisons using the Holm-Bonferroni correction. Significant differences were found only in horror movies where the median *degree* centrality for females (0.221) was higher than the median *degree* centrality of males (0.166). This is in agreement with prior studies which report female characters to have a more prominent presence in horror movies, particularly as victims of violent scenes (Welsh and Brantford, 2009).

### 5.3.2 Race

To examine differences in major roles across the racial categories, we perform Kruskal-Wallis tests similar to previous subsection. Significant differences were found with both *degree* and *betweenness* centrality measures ( $p < 0.001$ ;  $\alpha = 0.05$ ).

*Latino* characters were found to have significantly lower *degree* centralities compared to *caucasian* and *southasian* races suggesting non-central roles in these characters. *Caucasian* characters were found to have median *betweenness* centralities significantly higher than at least one other race. Characters from the *nativeamerican* race exhibit significantly lower medians in both *degree* and *betweenness* centralities than *caucasian*, *african* and *mixed* characters, which agrees with (Rosenthal, 2012).

### 5.3.3 Age

We investigate the effects of age on importance of character roles by building a linear regression model on the two centralities with age as the independent variable. In both cases, age was found to be significant ( $p < 0.001$ ;  $\alpha = 0.05$ ). With *degree* centrality, the regression coefficient  $\beta$  was found to be equal to 0.003. In *betweenness* centrality, the regression coefficient was also positive, given by  $\beta = 8.41 \times 10^{-4}$ . Both these metrics indicate a positive correlation for character importance with age, i.e. as characters age, there is an increased interaction with other characters in the movie as well as higher prominence in the movie plot.

## 6 Conclusion

We present a scalable automated analyses of differences in character portrayal along multiple factors such as gender, race and age using word usage, psycholinguistic and graph theoretic measures. Several interesting patterns are revealed

in the analysis. In particular, movies with female writers and directors in the production team are observed to have balanced gender ratios in characters compared to male writers/directors. Across several races, female actors are found to be younger than male actors on average.

Female characters appear to be more positive in language use with fewer references to death and fewer swear words compared to male characters. Female characters also appear to be more prominent in horror movies compared to male characters. Latino and mixed race characters appear to have higher usage of sexual words. Eastasian characters seem to use significantly fewer religious words. As characters aged, their word sophistication seems to increase along with usage of words related to achievement and religion; there was also a significant reduction in word activation, usage of sexual and swear words as character age increases.

Future work includes expanding the analyses to non-English movies and combining the linguistic metrics with character networks. Specifically, character network edges can be weighted using the psycholinguistic metrics to analyze the emotional patterns in inter-character interactions.

## 7 Acknowledgments

We acknowledge support from NSF and our partnership with Google and the Geena Davis Institute on Gender in Media.

We thank Naveen Kumar for all the helpful discussions and feedback during this work.

## References

- Ricardo Alberich, Joe Miro-Julia, and Francesc Rosselló. 2002. Marvel universe looks almost like a real social network. *arXiv preprint cond-mat/0202174*.
- Elizabeth Behm-Morawitz and Dana E Mastro. 2008. Mean girls? the influence of gender portrayals in teen movies on emerging adults' gender-based attitudes and beliefs. *Journalism & Mass Communication Quarterly* 85(1):131–146.
- Harry M Benshoff and Sean Griffin. 2011. *America on film: Representing race, class, gender, and sexuality at the movies*. John Wiley & Sons.
- Andrew Beveridge and Jie Shan. 2016. Network of thrones. *Math Horizons* 23(4):18–22.
- Anthony Bonato, David Ryan D'Angelo, Ethan R Elenberg, David F Gleich, and Yangyang Hou. 2016.



- Mining and modeling character networks. In *Algorithms and Models for the Web Graph: 13th International Workshop, WAW 2016, Montreal, QC, Canada, December 14–15, 2016, Proceedings 13*. Springer, pages 100–114.
- Gavin S Cape. 2003. Addiction, stigma and movies. *Acta Psychiatrica Scandinavica* 107(3):163–169.
- James M Clark and Allan Paivio. 2004. Extensions of the paivio, yuille, and madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers* 36(3):371–383.
- William Cohen, Pradeep Ravikumar, and Stephen Fienberg. 2003. A comparison of string metrics for matching names and records. In *Kdd workshop on data cleaning and object consolidation*. volume 3, pages 73–78.
- DailyScript. 2017. *The daily script*. [Online; accessed 1-February-2017]. <http://dailyscript.com/>.
- Paul G Davies, Steven J Spencer, and Claude M Steele. 2005. Clearing the air: identity safety moderates the effects of stereotype threat on women’s leadership aspirations. *Journal of personality and social psychology* 88(2):276.
- Tony Dimnik and Sandra Felton. 2006. Accountant stereotypes in movies distributed in north america in the twentieth century. *Accounting, Organizations and Society* 31(2):129–155.
- Alice H Eagly and Steven J Karau. 2002. Role congruity theory of prejudice toward female leaders. *Psychological review* 109(3):573.
- ethnicebs.com. 2017. *Celebrity ethnicity*. [Online; accessed 1-February-2017]. <http://ethnicebs.com>.
- Louis August Gottschalk and Goldine C Gleser. 1969. *The measurement of psychological states through the content analysis of verbal behavior*. Univ of California Press.
- Mark Hedley. 1994. The presentation of gendered conflict in popular movies: Affective stereotypes, cultural sentiments, and men’s motivation. *Sex Roles* 31(11-12):721–740.
- Bell Hooks. 2009. *Reel to real: race, class and sex at the movies*. Routledge.
- IMDb. 2017. *Internet movie database*. [Online; accessed 1-February-2017]. <http://www.imdb.com/>.
- IMSDb. 2017. *Internet movie script database*. [Online; accessed 1-February-2017]. <http://www.imsdb.com/>.
- Anne E Lincoln and Michael Patrick Allen. 2004. Double jeopardy in hollywood: Age and gender in the careers of film actors, 1926–1999. In *Sociological Forum*. Springer, volume 19, pages 611–631.
- Daniel Linz, Edward Donnerstein, and Steven Penrod. 1984. The effects of multiple exposures to filmed violence against women. *Journal of Communication* 34(3):130–147.
- Ting Liu, Kit Cho, George Aaron Broadwell, Samira Shaikh, Tomek Strzalkowski, John Lien, Sarah M Taylor, Laurie Feldman, Boris Yamrom, Nick Webb, et al. 2014. Automatic expansion of the mrc psycholinguistic database imageability ratings. In *LREC*. pages 2800–2805.
- François Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research* 30:457–500.
- Nikolaos Malandrakis and Shrikanth S Narayanan. 2015. Therapy language analysis using automatically generated psycholinguistic norms. In *INTER-SPEECH*. pages 1952–1956.
- Finn Årup Nielsen. 2011. A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.
- David Niven. 2006. Throwing your hat out of the ring: Negative recruitment and the gender imbalance in state legislative candidacy. *Politics & Gender* 2(04):473–489.
- NYFA. 2013. *Gender inequality in film*. [Online; accessed 1-February-2017]. <https://www.nyfa.edu/film-school-blog/gender-inequality-in-film/>.
- James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report.
- James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54(1):547–577.
- Polygraph. 2016. *Film dialogue from 2,000 screenplays, broken down by gender and age*. [Online; accessed 1-February-2017]. <http://polygraph.cool/films/>.
- Anil Ramakrishna, Nikolaos Malandrakis, Elizabeth Staruk, and Shrikanth S Narayanan. 2015. A quantitative analysis of gender differences in movies using psycholinguistic normatives. In *EMNLP*. pages 1996–2001.
- Mauricio Aparecido Ribeiro, Roberto Antonio Vosgerau, Maria Larissa Pereira Andruchiw, and Sandro Ely de Souza Pinto. 2016. The complex social network of the lord of rings. *Revista Brasileira de Ensino de Física* 38(1).
- Nicolas G Rosenthal. 2012. *Reimagining Indian country: native American migration and identity in twentieth-century Los Angeles*. Univ of North Carolina Press.

- Burry S Sapolsky, Fred Molitor, and Sarah Luque. 2003. Sex and violence in slasher films: Re-examining the assumptions. *Journalism & Mass Communication Quarterly* 80(1):28–38.
- Stacy L Smith, Marc Choueiti, and Katherine Pieper. 2014. Gender bias without borders: An investigation of female characters in popular films across 11 countries. *USC Annenberg* 5.
- Shinya Tanaka, Adam Jatowt, Makoto P Kato, and Katsumi Tanaka. 2013. Estimating content concreteness for finding comprehensible documents. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, pages 475–484.
- Tom FM Ter Bogt, Rutger CME Engels, Sanne Bogers, and Monique Kloosterman. 2010. “shake it baby, shake it”: Media preferences, sexual attitudes and gender stereotypes among adolescents. *Sex Roles* 63(11-12):844–859.
- Danny Wedding and Mary Ann Boyd. 1999. Movies & mental illness: Using films to understand psychopathology. .
- Andrew Welsh and Laurier Brantford. 2009. Sex and violence in the slasher horror film: A content analysis of gender differences in the depiction of violence. *Journal of Criminal Justice and Popular Culture* 16(1):1–25.
- Bo Xiao, Zac E Imel, Panayiotis G Georgiou, David C Atkins, and Shrikanth S Narayanan. 2015. “rate my therapist”: Automated detection of empathy in drug and alcohol counseling via speech and language processing. *PloS one* 10(12):e0143055.