

Twitter's Response to Shang-Chi and the Legend of the Ten Rings

Written by Fynn Schmitt-Ulms,¹

Lin Xiao Zheng,²

Nicole Zhou³

fynn.schmitt-ulms@mail.mcgill.ca¹, lin.x.zheng@mail.mcgill.ca², jingyi.zhou2@mail.mcgill.ca³

Introduction

Shang-Chi and the Legend of the Ten Rings, is a recently-released movie directed by Destin Daniel Cretton and produced by Marvel Studio. In this study, we investigate which aspects of the movie are most often discussed and analyze the sentiment associated with each topic. To do so, we use a sample of 1,000 tweets in our analysis. We found the discussions around the movie are most often focused on the audiences' opinions or states, Popular culture and characters or story arc in the movie. The topics we selected for annotation are Actions, Representation, Cast, Characters / story arc, Pop culture, Disney plus, Watching / Opinion. The overall response to this movie was positive since all topics selected received more positive sentiment than others.

This report summarizes each phase of the process, from data scraping to statistical modeling to visualization and the analysis results associated with the salient topics and sentiments discussed around the movie. The purpose of this report is to document both topics design, discuss relative engagement with these topics and capture the responses to the movie.

The remainder of this report is organized as follows. The Data section presents the dataset collection methods, including the filtering decisions made. In the Methods section, we describe our annotation and analysis approach. Results defines our selected topics and presents figures for topic characterization and engagement. In the Discussion section, we give our interpretation of the results and provide feedback on the movie.

Data

The data was collected from the Twitter API using the *tweepy* python package. In total, 1000 tweets were scraped using the keyword "shang-chi" as a search query. Retweets were filtered during the collection process and results were limited to English tweets created between Nov 12, 2021 and Nov 15, 2021. The API was also instructed to return the full tweet content, instead of the truncated content that it defaults to.

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Once the tweets had been collected, the *selenium* python package was used to collect screenshots of each tweet. These screenshots were then stored on an AWS bucket, allowing them to be accessed and displayed by our annotation site (See Methods).

For each tweet the following information was collected and stored:

Value	Description
created_at	The date the tweet was created
id	A unique integer id for each tweet
text	The full text content of the tweet
user	The author's twitter handle
user_id	A unique integer id for each user
twitter_url	The url of the tweet
im_url	The url of the tweet screenshot

Table 1: Tweet Properties Stored for Annotation

All of this information was then made available during the annotation phase, providing annotators with the context behind each tweet.

Methods

During the data annotation phase, we decided that each group member should perform both the open coding and the manual annotation individually. For the open coding part, we each came up with a list of potential topics and organized a meeting to discuss a unified topic list with definitions as specific as possible. A group session was held with all teammates to go through 25 tweets and make sure the guidelines were clear. Then, we each individually annotated the complete dataset using the finalized guidelines.

A web app in the form of a Doccano annotation site was deployed to an EC2 to streamline our annotation process¹. Doccano is an open source platform for data annotation that allows data to be displayed clearly during the annotation phase and provides annotators with shortcuts to speed up

¹To test out the Annotation site, visit <http://ec2-3-137-172-183.us-east-2.compute.amazonaws.com:8000> and click "Get Started" and login with Username: Grader and Password: comp598!

annotations. The platform also allowed us to have a unified label system which prevented mislabelling caused by typos. Furthermore, this tool considerably improved annotation efficiency which allowed each of us to annotate the whole dataset despite our limited time.

Having three sets of annotation allowed us to improve the robustness of our annotation process and reduce the impact of human error and individual bias. We reconciled any disagreement using a majority wins rule. In the case of a tie, the document was manually reconciled by a member of the team.

For computing the tf-idf score, we define:

$$\text{td-idf}(w, \text{category}, \text{script}) = \text{tf}(w, \text{category}) \times \text{idf}(w, \text{script})$$

where $\text{tf}(w, \text{category})$ is the number of times the word w is used under this category and

$$\text{idf}(w, \text{script}) = \frac{\log(\text{total number of categories})}{\text{number of categories that use the word } w}$$

To avoid meaningless result, we decided to delete frequent words based on a stop word list¹ and words that occur less than 5 times across all original posts. This threshold is determined by the average frequency of each word. Furthermore, the words contained in the title of the movie, *Shang-Chi and the Legend of the Ten Rings*, were removed given that they do not offer any meaningful information on the topics. We also treat all words encountered as case insensitive.

Results

In this projects, we have 3 possible sentiments and 7 topics selected to classify the discussions around the movie. Each tweet was labeled with one sentiment and one topic.

1. Topic Definitions

The following topics are listed with hierarchy from highest to lowest. Thus, if a tweet mentions multiple topics, we choose the topic with higher priority. Figure 1 shows overall topic distribution over all data collected.

Action: A tweet mentions a specific action or fight scene appeared in the movie, including commenting about the martial arts or Kung Fu.

Representation: A tweet mentions the language, usually Mandarin or Cantonese, used in the movie or spoken by the actors during the interviews. Additionally, it can discuss about relative Asian or Asian American culture.

Cast: A tweet mentions any actors, the director, the writers or the staff behind the scenes, but not necessarily mention by specific name. This includes the comments on the acting or directing.

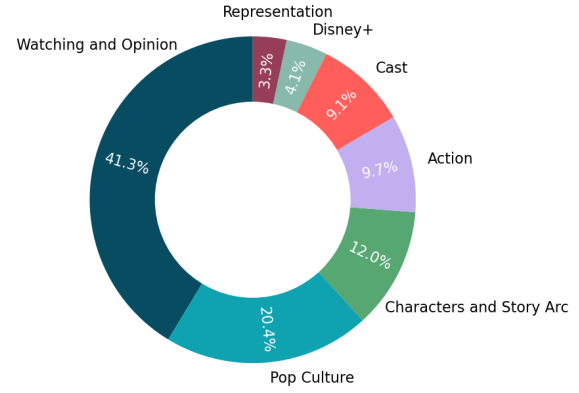


Figure 1: Overall Topic Distribution

Characters and story arc: A tweet mentions about any characters or animal protagonists appear in the movie. Also, it may mention the story arc or the plot of this movie.

Pop culture: A tweet refers to characters in the other pop culture or Marvel Cinematic Universe. It can either mention or compare to other movies.

Disney+: A tweet describes IMAX experience of watching this movie on Disney plus or mentions this platform.

Watching and Opinion: A tweets presents the opinions of the movie or expresses the feeling towards the movie. It also includes the states of audience, for instance, currently watching or re-watching the movie.

2. Topic Characterization

We characterized the topics selected by computing 10 words in each category with highest tf-idf scores. See Tables 2, 3, 4, and 5.

Ranking	Positive	Neutral	Negative
1	Choreography	Bus	Dem
2	Fight	San	Subtitles
3	Letterboxd	Francisco	Half
4	gt ²	Trakt	Trying
5	Review	Nowwatching	Fighting
6	Tony	Falcon	Called
7	Leung	Operator	Shit
8	Action	Red	Liked
9	MCU	Notice	Look
10	Favourite	Destin	April

Table 2: Top TF-IDF Terms by Tweet Sentiment

¹Freely available at <http://www.lextek.com/manuals/onix/stopwords1.html>

²The "gt" value comes from ">" which is how twitter stores the ">" symbol.

Ranking	Action	Representation	Cast
1	Choreography	Subtitles	Michelle
2	Fight	Asian	Yeoh
3	Bus	Chinese	Tony
4	Action	Cast	Leung
5	Scenes	Representation	Liu
6	Operator	Heroes	Zhang
7	Fights	Culture	Awkwafina
8	Sequences	Form	Simu
9	San	Badass	Ben
10	Francisco	America	Cast

Table 3: Top TF-IDF Terms in Action, Representation, and Cast Categories

Ranking	Characters	Pop Culture	Disney+
1	Cute	gt	Plus
2	Morris	Phase	IMAX
3	Called	Winter	Somehow
4	Katy	Favourite	Disneyplus
5	Dad	Red	Disney
6	Xialing	Notice	Version
7	Feet	MCU	Theater
8	Sister	Falcon	Audience
9	Character	Soldier	Loving
10	Dragon	Loki	Review

Table 4: Top TF-IDF Terms in Characters and Story Arc, Pop Culture, and Disney+ Categories

Ranking	Watching and Opinion
1	Review
2	Letterboxd
3	Trakt
4	Nowwatching
5	Watching
6	Happy
7	Obsessed
8	nw
9	Holy
10	Discord

Table 5: Top TF-IDF Terms in the Watching and Opinion Category

3. Topic Engagement

In general, the percentage of positive sentiment is significantly higher than the percentage of neutral and negative sentiment over all topics. See Figure 2.

Positive:

From Table 2, we can observe that most tweets with positive sentiment mention "choreography", "fight" or "action" in the movie. Many twitter users also discuss and review this movie using Letterboxd or by comparing it to the other movies, and some of them think this is their favourite Marvel Cinematic Universe movie. In addition, Tony Leung who acts protagonist's father receives many

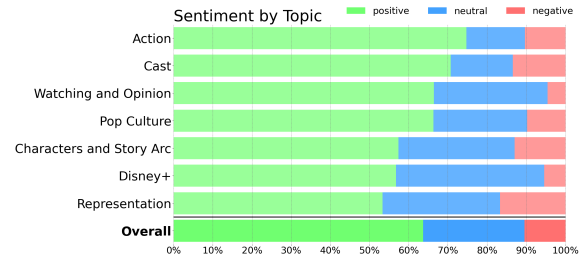


Figure 2: Sentiment distribution by topics

positive feedback.

Neutral:

The neutral sentiment are mostly regarding to a San Francisco operator reviews on the bus fight scene in the movie. For tweets that only narrate they are watching it right now.

Negative:

Since the amount of tweets express negative sentiment is the least, the trend is not as obvious as the other. However, we can find that some of them complain about the subtitles or the fighting in the movie.

Action:

A large amount of discussion under this category focus on the choreography or the action and the fighting scenes. The bus fight scenes are discussed mostly and passionately. From Figure 2, we know the audiences are mostly positive towards actions in the movie.

Representation:

The portion of positive sentiment of this topic is the smallest among all while there are many discussions express negative feeling towards representations in the movie. The subtitles of characters are mentioned mostly, and there are many tweets discuss about the Asian cast, culture, heroes produced by Marval Studio.

Cast:

From the second bar in Figure 2, the discussions around the cast are positive in general. However, some tweets express their dislike towards a specific actor or actress. The tweets under this category are widely talked about the top cast: Simu Liu, Tony Leung, Awkwafina, Michelle Yeoh, Ben Kingsley, Meng'er Zhang. The audience either comment the acting of cast or the characters acted by these cast.

Characters/story arc:

There are many discussions about a cute and popular character named Morris. Also, the sister of Shang-chi, Katy, acted by Awkwafina is often mentioned, and the father of Shang-chi who is acted by Tony Leung is a controversial character. Since there is a dragon riding scenes in the movie, Dragon is mentioned a lot.

Pop culture:

Most tweets compare *Shang-Chi and the Legend of the Ten Rings* to other movies produced by Marvel Studio, such as, *Red Notice*, *Falcon*, *Loki*. Moreover, it is considered as the favourite MCU movie for many people.

Disney plus:

As Shang-chi is on Disney plus, many tweets describe experience of watching this movie on Disney plus or mentions this platform in their reviews.

Watching/Opinion:

Most discussions state their opinions of this movie or present their reviews by stars or scores on Letterboxd. Some audience feel "happy" or "obsessed" about watching this movie. Note that there is a great proportion of neutral responses about *Watching* and *Disney+* since the audiences usually do not express their sentiments towards movie while commenting on these topics.

Discussion

Let's break the discussion down by topic. Our first topic is the one for action. This topic is considered the most important because *Shang-Chi* is primarily an action movie. Given the positive responses on this subject, it seems that the film delivered successfully on that part. The fighting choreography especially received praise, and a specific fight scene involving a bus was extremely memorable to the viewers.

Second on our hierarchy is the topic of representation, more specifically Asian American representation. Although this subject is overall positive, it is the most divisive topic in our list, with the highest ratio of negative tweets at over 15%. The subtitles were highly discussed, both positively and negatively, given that many parts of the movie are in Mandarin. However, discussions around the casting and the representation of Chinese (American) culture were more overall positive. This is extremely important since this is the first Marvel superhero movie with an Asian lead. Despite some divisiveness in this topic, the overall positive sentiment shows that people were satisfied with the representation in *Shang-Chi*.

As for the topic of the cast, separate from the issue of representation, it is interesting to note that there are as many female actresses present in the most mentioned names as there are male actors. This demonstrates that the casting in this movie was well performed, for actors of either gender. In fact, the most mentioned person, according to the topic characterization, is Michelle Yeoh, followed by Tony Leung, both of whom play supporting roles to the lead, Simu Liu. In *Shang-Chi*, the supporting actors were just as memorable to the viewers, if not more, than the lead.

This brings us to the characters played by the cast and the story arc that they are involved in. An overwhelming

number of tweets in this topic were about Morris, a CGI animal sidekick inspired by Chinese mythology, whom people found very "cute". This implies that the design of the character was very well executed, especially taking into account that Morris does not have a head. There was also a lot of discussion around *Shang-Chi's* friendship with a character named Katy. A lot of positive emphasis was put on this friendship which shows movie producers that platonic relationships can be just as important as romantic ones. In fact, two female characters made it to the topic characterization by name. Echoing findings for the cast, these female characters were very memorable for the viewers. Given the positive sentiment surrounding these discussion, it can be concluded that the female characters were written and portrayed in this movie. Finally, the topic of *Shang-Chi's* dad in relation to the character's backstory was also highly discussed. Marvel clearly left a mark on its viewers with this story arc.

Now, onto the topic of pop culture, which was the second most populated topic, with 18.4% of the tweets going under this label. Many people were happily discussing any pop culture reference found in the film. Any effort from the producers to insert these references paid off. However, the bulk of the content in this topic were comparisons between *Shang-Chi and the Legend of the Ten Rings* and other movies from the Marvel Cinematic Universe (MCU), e.g. *Captain America: Winter Soldier*. Given the cult following that Marvel Studios has garnered over its movies, it's unsurprising that so many people are talking about these movie in relation to one another. Most people seem to have accepted *Shang-Chi* as part of the MCU.

The topic of *Disney+* does not have much representation with only 3.7% of the tweets under its label. Given the relative novelty of the streaming service and the recently shortened theatrical window allowing people to watch *Shang-Chi* for the first time in the comfort of their own homes shortly after its release, we thought it would be interesting and meaningful to gather opinions on *Disney+*. However, a relatively large number of the tweets under this subject are neutral, only mentioning *Disney+* in passing. One important point that were brought up in many tweets is how people were glad to be able to watch the movie without having to go to theaters, given COVID. Importantly, people are definitely using the streaming service, and there is a very low ratio of negative comments about it.

Finally, the most populous category we have: watching and opinions. This topic was overwhelmingly positive with over 65% positive tweets and only about 5% negative. Given that this is at the bottom of our topic hierarchy, this captures tweets with no detail about the movie that overlaps with the other topics. We can infer that these tweets can reflect the opinions of an average viewer. An average viewer could just be someone watching the movie without much background information on it, in contrast to someone who has extensive knowledge about the MCU, or someone who is watching the movie for the actors. This topic gives a somewhat contained

look into how the movie is perceived as an independent entity. In the end, Shang-Chi was clearly very well received by the general public.

Contributions

Fynn Schmitt-Ulms

- data collection and creation of data annotation tool
- participation in open coding and manual annotation
- writing of data section
- generating visualizations from the results

Nicole Zhou

- participation in open coding and manual annotation
- computing topic characterization
- writing of introduction and results sections

Lin Xiao Zheng

- participation in open coding and manual annotation
- merging and reconciling labels after multiple annotation
- writing of methods and discussion sections