I'll Have What She's Having: Network Formation and Social Spillovers in Film Consumption on Letterboxd.com

Johnny Ma[*]

Under the Direction of Professor Leonardo Bursztyn[†]

May 20, 2018

Abstract

Why do we watch the films we watch? How large of a role do peer effects play in watching behavior? Using individual film watching behavior and friend networks from 9000 users scraped from film diary and social network website *letterboxd.com*, and 200 films' box office runs scraped from *boxofficemojo.com*, we quantify the effect of social learning and social utility on trips to the cinema. We adapt Moretti's 2011 Bayesian model of film consumption to include both social learning based on rating feedback from friends and social utility based on unexpected popularity of specific films. We fit a probit regression using data from a film's theatrical run to quantify the effect of social learning versus social utility. We find that while both are positively and significantly correlated with higher probabilty of watching, the effect of social utility (11%) is an order of magnitude greater than that of social learning (0.7%). Future work on the social nature of film consumption in social media networks can be informative for film studios in viral marketing campaigns at the box office.

[*] I would like to thank Professors Elena Manresa, Amir Goldberg, Emir Kamenica, and Stefano DellaVigna for general advice on this paper. Thanks to Professors Victor Lima and Yoshida Kotaro as well as Julie Wong for help in the B.A thesis process. Special thanks to my friends, Angela Sun, Paul Beckman, William Jones, Sylvia Klosin, Thomas Yu, and Lindsey Currier for great discussions and coming to my movie nights! Finally, I would like to thank *letterboxd.com* for letting me scrape their data and for being such a great platform. Data and code are available at `http://johnnyma.info`.

[†] Assistant Professor, University of Chicago Kenneth C. Griffin Department of Economics. 1126 E 59th Street, Chicago, IL 60637.

Contents

1 Introduction

Why do people watch the films they watch? While taste and aesthetic preferences factor into the enjoyability of a movie, movies are undeniably social and cultural goods. People often receive film recommendations from their friends, go to the theater with friends, and share their ratings and feelings about the movie afterwards. As friendships are often formed on the psychological or demographic traits that correlate with taste in movies, a friend's recommendation may be more personalized and valuable than the plethora of signals sent from marketing campaigns. The movie industry, and its movie star drama, also holds a prominent part of America's cultural zeitgeist, with ample media coverage of both popular franchises such as *Star Wars* and *Marvel's Avengers* and critically acclaimed films such as *Get Out* and *Moonlight*. The rise of social media and memes have amplified the reach and power of movie references, with these inside jokes and shared conversation permeating all aspects of life [1]. In the context of opening weekend box office predictions, where ”nobody knows anything,”[2] these social forces must be incorporated into film consumption models to better understanding audience behavior.

Within the economics literature, these forces are called ”peer effects,” where others' outcomes and behavior enter an individual's utilty function. There have been many models [2] and empirical examples [8] that quantify the role of peer effects on behavior. Notably, economists have identified two channels through which peer effects occur: social learning and social utility [5]. Social learning occurs when the decisions and ratings of peers provide information on the quality of the film (”movie X was amazing, I like you'll like it!”), while social utility is when the number of peers who have consumed a good shows up directly in a movie-goer's utility function (”you haven't seen movie X? Everyone is talking about it!”). These peer effects are notoriously difficult to disentangle due to the unidentified direction of

[1] On the political side, the #MeToo movement emerged out of scandals involving film producer Harvey Weinstein and Oscar winning actor Kevin Spacey. On the fun side, visit `https://www.facebook.com/photo.php?fbid=1857301980988186`

[2] A famous quote from prolific screenwriter William Goldman, shared by most in the film industry.

casuality [13] and the fact that social learning and social utility can occur at the same time and are likely correlated (is it popular because its good, or is it good because its popular?).

We model the social nature of the choice to watch a given film using parts of Moretti's 2011 Bayesian model of a movie-goer's decision making process. We say that a film's utility comes from both the quality of the movie and the popularity of the film, as previously discussed. Therefore, those that watch the movie during opening weekend (henceforth referred to as "OW") have decided that the estimated utility from priors on a film's quality and eventual popularity is greater than the opportunity cost of the next best option. We then consider individuals who did not watch the movie during OW, assuming that all watching behavior in subsequent weeks is driven solely by social signals from their friends and the site as a whole. Namely, they update their beliefs on the quality of the film using rating feedback from their network of friends who watched the film in previous weeks (social learning), as well as on the popularity of a film using information from the size or prevelence of site wide discussion (social utility). This model has considerable assumptions, such as holding outside options fixed and assuming similarly distributed priors, but we believe it generally captures the real life social elements behind movie consumption decisions.

While this paper does not as rigorously identity the separate effects of these two channels as previous papers have attempted to, it overcomes the literature's reliance on aggregate viewing behavior, analyzing a social environment where it is likely that these two channels act largely separately: online social media. Many social media websites show content based on both personal friend networks and a general population of users [3], a slight overlap in groups notwithstanding. One such social media website for film is *letterboxd.com*, a movie diary website designed for "sharing your taste in film."[4] The website consists of a community of film fanatics who enter time-stamped watching behavior into their film diary, find and friend other users with similar tastes, and get information from film pages that show aggregate

[3] Facebook and Twitter show both content from friends as well as site-wide "top trending" stories and hashtags.

[4] https://letterboxd.com/about/-questions/

information alongside their friends' ratings (and reviews). We scrape *letterboxd.com* to obtain around 9000 users' time-stamped watching and rating behavior, each user's friend network, and film information for thousands of popular films. More details on *letterboxd.com* is provided in the Data section. While this website is certainly not representative of the average American movie-goer[5], the paper is more interested in analyzing the social dynamics that play out in a contained online social network such as *letterboxd*, which sheds light into what might be the future of human social interaction.

In this Bayesian box office environment, we measure social information by calculating the share of OW friends who rated the film above the film's average rating (a value from 0 to 1) [6], and measure unexpected social utility using the residual from the regression of number of opening screens on first week attendance [7]. While higher quality movies are oftentimes more popular [8], in the world of social media sites it is more likely that a user is driven to want to watch a film to join a general site-wide conversation or understand the references and memes they see, rather than interact specificially with the users in their friend network who they likely do not know in person and do not have personal conversations with [9].

We empircally estimate this model using a probit model of the role of social learning and social utility on the probabilty of a user watching a film during its theatrical run, traditionally defined as the first six weeks a movie runs in the box office. We first remove users who watched the film during OW and therefore did not receive any social signals from their peers or the general populace. Despite the richness of the data and the growing prevalence of on-demand digital streaming, we restrict our analysis to theatrical runs as they provide a clean environment where individuals start with similar, noisy beliefs (from marketing campaigns) and can easily observe weekly outcomes and quickly update their beliefs. Using box office

[5] Opening Weekend watching volume on letterboxd is highly correlated with nationwide watching volume, see Appendix.

[6] A measure borrowed from Udry and Conley's [6] analysis of social learning in networks exposed to new technology.

[7] A unique measure that Moretti argues captures the "surprise in the appeal of a movie, given that movie theaters are incentivized to correctly predict OW demand"

[8] Not always true. See Twilight series, Suicide Squad, and many other summer flicks.

[9] The most popular reviews for the most popular films have thousands of likes, but less than 30 comments.

data from the industry standard *boxofficemojo.com*, we construct and validate [10] Moretti's measure of unexpected popularity. We then show that users on *letterboxd.com* form friend connections based on similarity in taste, an index calculated using the standard practice of correlation between film ratings[11]. We then construct the OW "friend surprise" measurement unique to each individuals' friend network using rating data from *letterboxd.com*.

Finally, we estimate the probit model using viewing behavior from *letterboxd.com* using movie fixed effects. While both social learning and social utility, as well as movie quality, are positively correlated with the decision to watch a film after its OW, the coefficent of social learning is an order of magnitude smaller than that of social utility. This suggests that an unexpected explosion of popularity surrounding a film has a much greater role in the decision making process than more information about the quality of said film. This is not an entirely unexpected finding; while people obviously prefer movies that play to their tastes, the overwhelming social and cultural role that films play in American society may push people to watch films simply to get a reference or be a part of the conversation. This also supports the idea of an "underlying demand" for weekend trips to the movie theater found by economists [9], where movie-goers are mostly concerned with doing something with friends or family over a weekend and are relatively more elastic to the actual quality of the film they watch. Film studios and marketing firms might want to focus more on campaigns that convey the social importance of the film ("THE summer blockbuster hit that everyone is talking about!" or "A film you DON'T want to miss.") rather than the quality of the film (Rotton Tomatos score, plot information, etc.), or ideally both for a certain "social multipler" effect.

While there are many empirical and theoretical concerns with this type of analysis, the richness of the data and the unique nature of social media networks can give us a better look at actual film consumption behavior and the role social information and utility play. Future

[10] Validation is largely intuitive. For example, the films among the highest residuals, *Black Panther* (2018) and *Deadpool* (2017), are widely regarded as massive box office surprises.

[11] A part of "nearest neighbors" algorithms that are used by Netflix and others to build film recommendation systems. See Appendix.

studies can work on either rigorously identify these effects using exogenous shocks such as weather, or estimate these effects in different environments beyond a movie's theatrical run, analyzing future rental or on-demand streaming. More information about users (who is more affected by peer effects, women or men? Teenagers or families?) and more consideration of the content of the films (sequel, genre, year released, lead actors, plot and script, etc.) would give a cleaner picture of heterogeneity in magnitude and direction of social effects. Perhaps these results can shed more light into the multi-trillion dollar film industry where ”nobody knows anything,” help studios understand movie going behavior, and ultimately predict the next big hit.

2 Literature Review

While there have been plenty of models that explain the importance of social information in decision making [3], the idea of ”social utility” spillovers from others' behavior was formalized by Gary Becker [1] , who suggested that ”the pleasure from some goods is greater when many people want to consume it.” While there have been many papers focusing on the effect of either social learning or social utilty, the paper that best distinguishes between these two channels is an experiment run by Bursztyn *et al.* [5] that randomizes uptake and revealed information in the purchasing of financial assets. While our paper does not have the power to randomize outcomes and information[12], the narrative behind the separation of these channels is an attempt to follow this line of inquiry in peer effects literature.

This paper is not the first to look at the social nature of film consumption. The seminal paper in this topic is Enrico Moretti's 2011 paper [14] that sets up a model of information based social learning based on peer consumption decisions, supporting their empirical predictions using box office dynamics. This paper is of considerable use in setting up our model and providing the measurement of aggregate ”surprise” in estimates of a film's overall popularity, as well as providing a system of empirical predictions based on social learning.

[12] Though one can imagine that *letterboxd* or Netflix could run a similar experiment

Simiarly, Gilchrist *et al.* [10] attempt to identify the role of social utility using exogenous weather shocks, arguing that these shocks only affect OW attendance and are orthogonal to film quality or social information, with the conclusion that shared experience plays a role in film consumption. These papers cannot, however, rule out the other side of the story[13] and rely too much on oblique assumptions[14].

The more pressing problem of these papers are empirical data problems. These papers rely heavily on proxies for actual film consumption, such as aggregate box office returns or MSA-level google searchs, and therefore must assume that peer effects exert their influence at an aggregate level. Since friendships are likely polarized and correlated with taste, information from MSA viewing are likely hard to estimate and are as informative as aggregate point estimates such as Rotton Tomatos. Certainly these rough aggregates are less informative than an actual friend's personalized reccomendation. Information on the unique structure of each individual's watching behavior and friend network provides the necessary heterogeneity to quantify the relative size of the channels of peer effects.

This paper's main contribution is that it addresses these data problems, providing both panel-level viewing behavior and the actual structure of each users' friend networks. To this end, it also contributes to the growing literature studying social media networks [12]. This paper also draws many best pratices for handling box office and film consumption data from previous papers, such as a film matching algorithm from Dellavigna *et al.* [7] and controls for the cyclical nature of film consumption from Liran Einav [9], among others [11].

3 A Simple Model of Social Learning and Social Utility

In this section, I outline a highly stylized and simplistic model that provides empirical predictions for the effect of social learning and social utility on post-OW film consumption. This model is based on a combination of Moretti's simple model of social learning and an

[13] Social utility for Moretti, social learning for Gilchrist.

[14] Moretti analyzes the effect of network size by assuming that teenagers have larger networks, for example.

idea borrowed from Gilchrist *et al.*; namely the addition of "cumulative prior viewership" in an individual's utility function. The following set of equations are a workable model that gives the probabilty of a user watching during OW:

$$U_{ij} = \alpha_j^* * + CV_j + \epsilon_{ij}$$

$$\alpha_j^* \sim \mathcal{N}(X_j'\beta, \frac{1}{m_j}), \qquad CV_j \sim \mathcal{N}(f(X_j'\beta), \frac{1}{f(m_j)}), \qquad \epsilon_{ij} \sim \mathcal{N}(0, \frac{1}{k_j})$$

$$P_1 = Pr(\mathbb{E}_1[U_{ij1}|X_j'\beta]) = Pr(\omega_j X_j'\beta + (1-\omega_j)f(X_j'\beta) > q_{i1})$$

The idea is relatively straightfoward. Consumer i get relatively more utlity from watching a higher quality film and/or a more popular film in comparison to a lower quality and/or less popular film. During OW, consumers observes the set of film playing in the box office j, uses a personal prior $X_j'\beta$ to estimate the quality α_{ij} of the film based on its observable qualities X_j', such as information from marketing campaigns (with m_j as the precision of the prior, not used). All consumers also share a prior on how popular they believe the film will be CV_j, also as a function of these observable qualities. An individual movie-goer makes the decision to go during OW if the film specific weights ω_j on the importance of quality and popularity lead to an expected utility greater than the utilty from the best outside option q_{i1} (which in practice we hold fixed between weeks).

In the absence of weekly social learning and social utility, the probabilty of watching a film is constant, as this model does not include changes in utility for rewatches. Thus, we should see constant viewership if the outside option is the same for each week[15]. In a world with peer effects, users in week t use information from previous weeks $t-1$ to update their beliefs on film quality and popularity. With S_{ij} quality "surprise" signals from f peers in i's friend network k and popularity signals from residuals from the OW screens-gross sales

[15] Or, with exit after one watch, rapidly decaying watches in the subsequent weeks.

regression, RES_j, the probabilty of watching in week t is:

$$P_t = Pr(\mathbb{E}_t[U_{ijt}|X_j'\beta]) = Pr(\omega_{j1t}X_j'\beta + \omega_{j2t}f(X_j'\beta) + \sum_{f\in k}\omega_{j3f}S_{ijf} + \omega_{j4t}RES_{jt} > q_{it})$$

The major take away for post-OW watching is that each individual i gets heterogeneos quality information from their unique friend network f, while all individuals share the effect of a surprise in the film's popularity. This leads to two predictions:

1. In the presence of strong social utility, stronger (weaker) than expected OW demand increases (decreases) probabilty of watching.
2. In the presence of strong social learning, high (low) share of OW above average reviews increases (decreases) probabilty of watching.

The places where the empirical model strays from the estimation of P_t will be specified in the data and empirics section.

4 Data

There are two main datasets used in this paper: box office data from *boxofficemojo.com* and individual-level watching behavior and friend networks from *letterboxd.com*. Both of these data sources were web scraped using html hypertext and Ajax tables with the ”rvest” package in R[16]. One concern is that many of the metrics collected from *letterboxd.com* are calculated whenever the server is queried. For exampe, the average rating for a film changes from day to day based on arriving user input. We do our best to estimate these metrics for each day to simluate the acutal page each user views, especially for the week after OW, but this generally should not be a major problem as the average rating of a film rarely deviates after the first week[17]. This iteration of the data was collected from May 5th - May 12th, 2018.

[16] `https://github.com/hadley/rvest`

[17] This might be a problem for friend network formations, but data trends show us that most friends are made close to account creation in short bursts rather than each day.

4.1 Box Office Data

Our national box office data come from Box Office Mojo, a reporting service owned by Internet Movie Database (IMDb). For each day, we obtain the following information for 15 of the top grossing movies currently in theaters: title, daily US box office gross, total gross to date, days in running, and number of screens. In the weeks just following release (when a movie can generally be viewed exclusively in theaters), box office data provide an excellent measure of a movies audience size, as movie tickets are price standarized. We track audience sizes during the 6 weeks following the date of wide release. We focus throughout on weekend (Friday, Saturday, and Sunday) audiences since they account for the vast majority of ticket sales. Our ticket sales sample comprises all movies wide-released in US theaters between October 11th, 2011 and April 25th, 2018. A wide-release movie is defined as a movie that opens with its absolute maximum number of screens. This includes most commercial films such as *Avenger's: Infinity War* and *Star Wars: The Force Awakens* but excludes many art-house and indie films such as *La La Land* and *Lady Bird* as they open first in a select number of theaters in New York and Los Angeles. After doing string merging of film titles from our other data source, we end up with 212 films in our dataset. Figure 1 shows the per weekend gross of films in our data set.

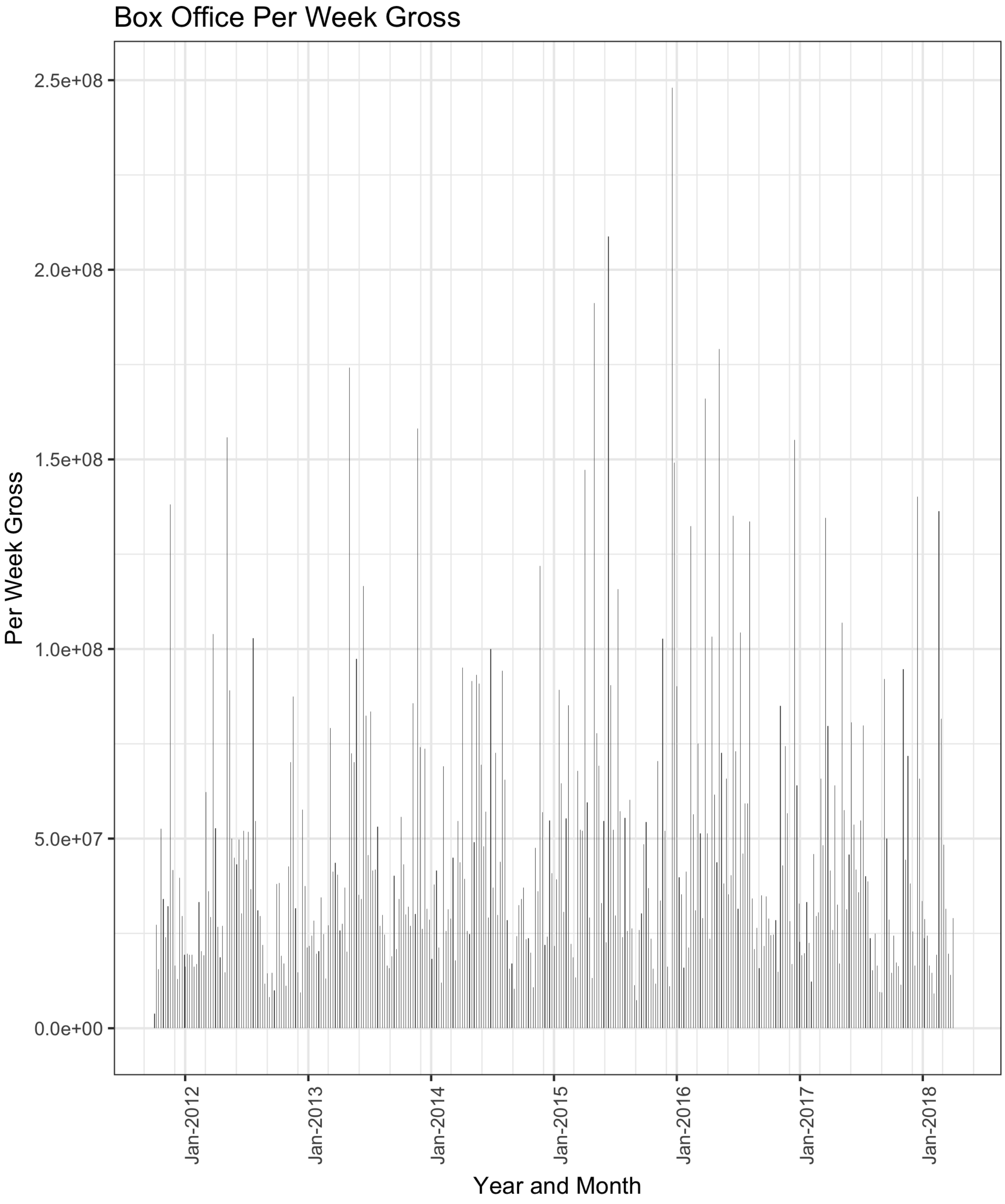


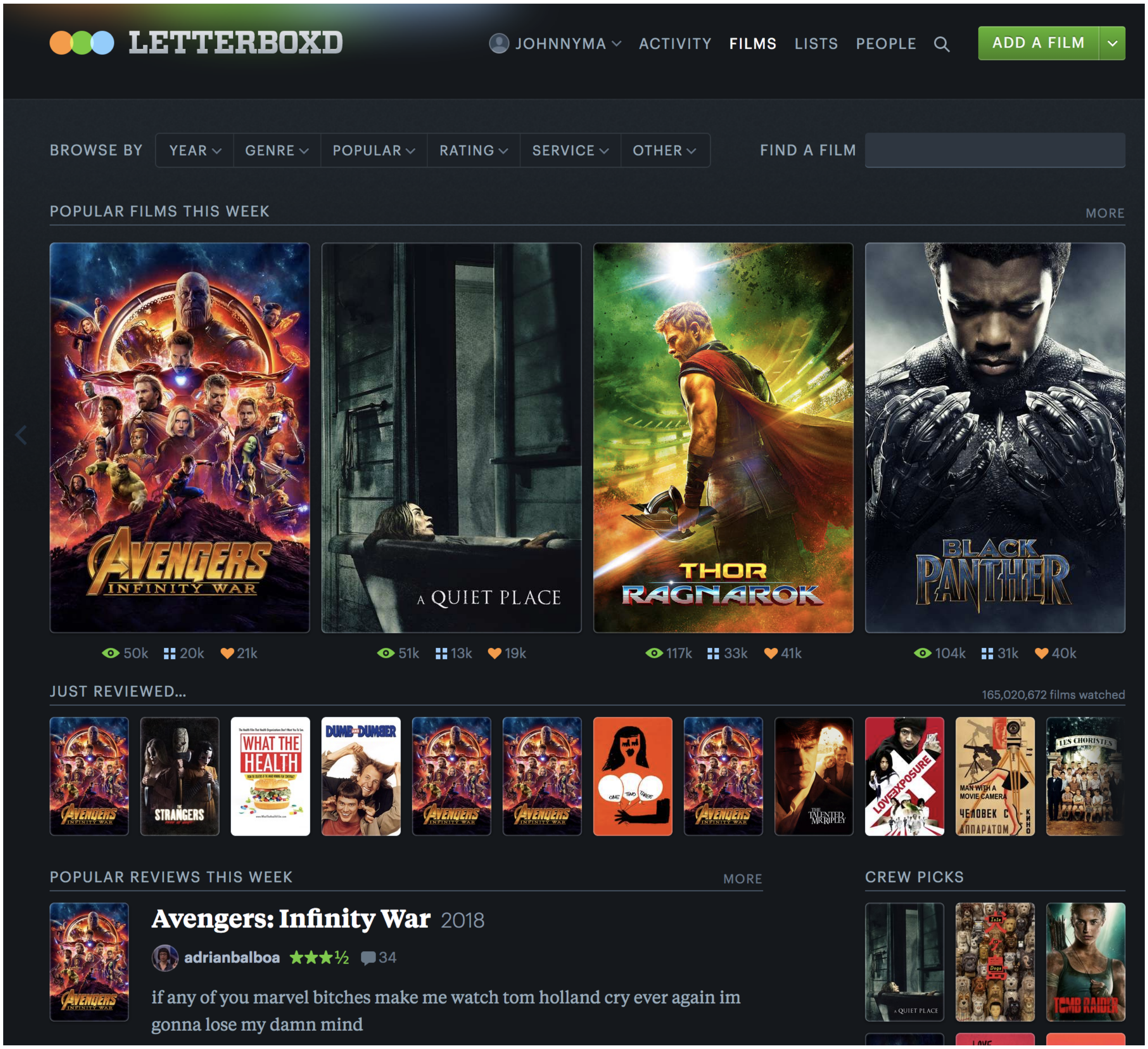
Figure 1: Per weekend gross returns of all films in the box office, 2011 - 2018.

4.2 Letterboxd.com Data

4.2.1 User Level Data

Our panel-level watching and friend network data come from the website, `letterboxd.com`. Letterboxd was founded in October 2011 as a film diary and social network website. The website has at least 200,000 active users as of 2017[18], with steady growth in users. The types of users range from college-aged film buffs to critics for the NYT to CEOs of movie studios. Users have the option to look up films they have watched, record the date they watched it, give it a rating (0-10, half star increments), and write a review. All reviews and ratings are publically availabile, and each user can comment and like other's activites. Figure 2 and 3 shows the "film" page and the front page seen by a registered user. Figure 4 shows a user's "info" page that visitors see when they look up your account. A visitor can then see every film the user has registered as "watched" (Figure 5), along with the associated ratings and date they were entered in (Figure 6). Figure 7 shows the user's film diary, where users can specify the day they watched a film, their rating, if they "liked" it, if it was a "rewatch," and their written review, if they have one. This is the panel-data that we obtain for each user in our dataset. Figure 8 is the other part of our dataset, namely which other users an individual is following. We select our list of users from the list of people in the "People" tab as seen in Figure 9.

[18] The most watched film on the website, *Mad Max: Fury Road*, has 233,000 recorded watches. The recently released *Avenger's: Infinity War* currently sits at 90,000 watches. More stats are available at `https://letterboxd.com/2017/#title-page`

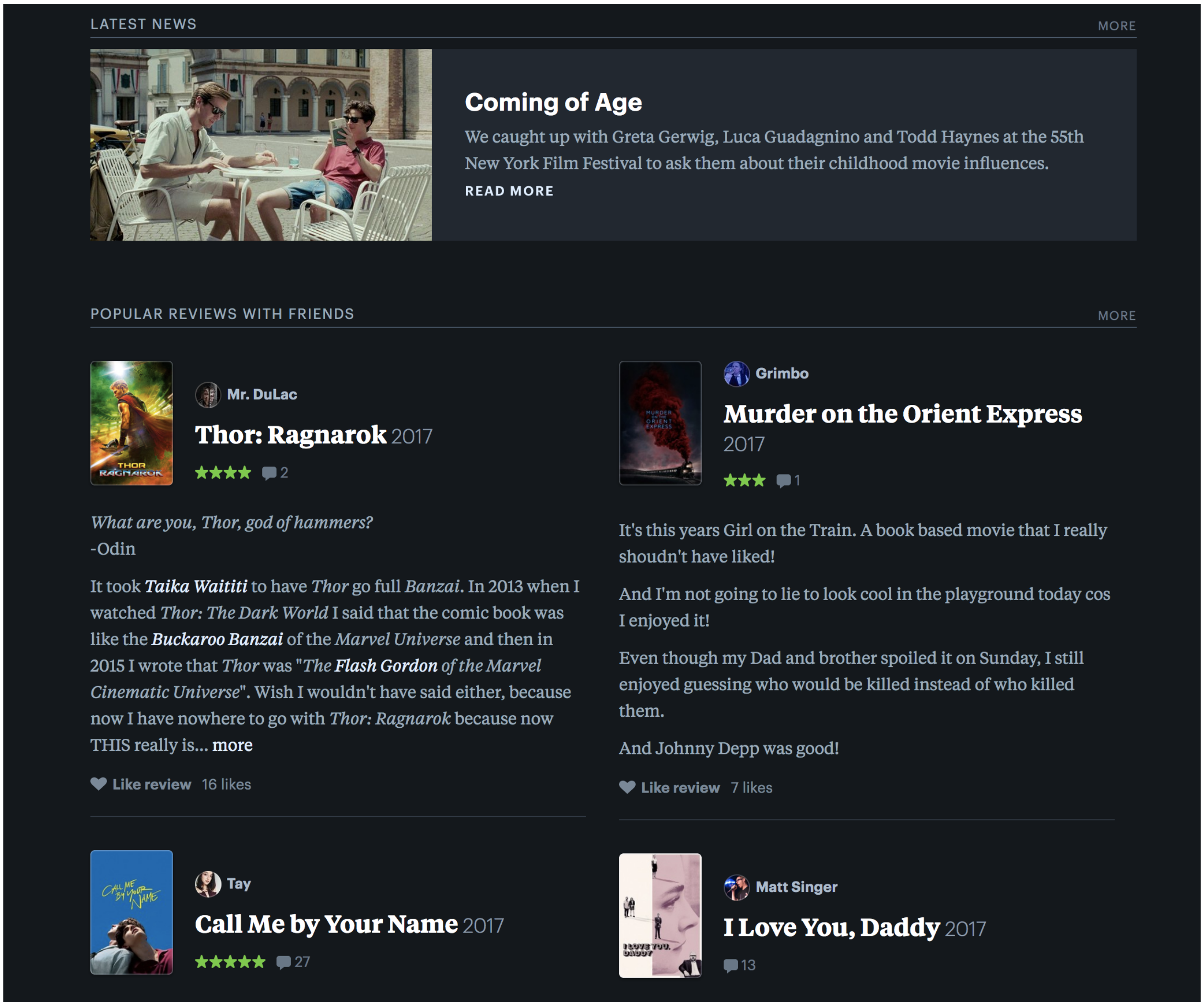


Avengers: Infinity War 2018

adrianbalboa ★★★½ 34

if any of you marvel bitches make me watch tom holland cry ever again im gonna lose my damn mind

Figure 2: Film page. Notice the prominence of what is popular on the site.



LATEST NEWS

MORE

Coming of Age

We caught up with Greta Gerwig, Luca Guadagnino and Todd Haynes at the 55th New York Film Festival to ask them about their childhood movie influences.

READ MORE

POPULAR REVIEWS WITH FRIENDS

MORE

Mr. DuLac

Thor: Ragnarok 2017

★★★★ 2

What are you, Thor, god of hammers?
-Odin

It took ***Taika Waititi*** to have *Thor* go full *Banzai*. In 2013 when I watched *Thor: The Dark World* I said that the comic book was like the ***Buckaroo Banzai*** of the *Marvel Universe* and then in 2015 I wrote that *Thor* was "*The **Flash Gordon** of the Marvel Cinematic Universe*". Wish I wouldn't have said either, because now I have nowhere to go with *Thor: Ragnarok* because now THIS really is... **more**

♥ Like review 16 likes

Grimbo

Murder on the Orient Express 2017

★★★ 1

It's this years Girl on the Train. A book based movie that I really shoudn't have liked!

And I'm not going to lie to look cool in the playground today cos I enjoyed it!

Even though my Dad and brother spoiled it on Sunday, I still enjoyed guessing who would be killed instead of who killed them.

And Johnny Depp was good!

♥ Like review 7 likes

Tay

Call Me by Your Name 2017

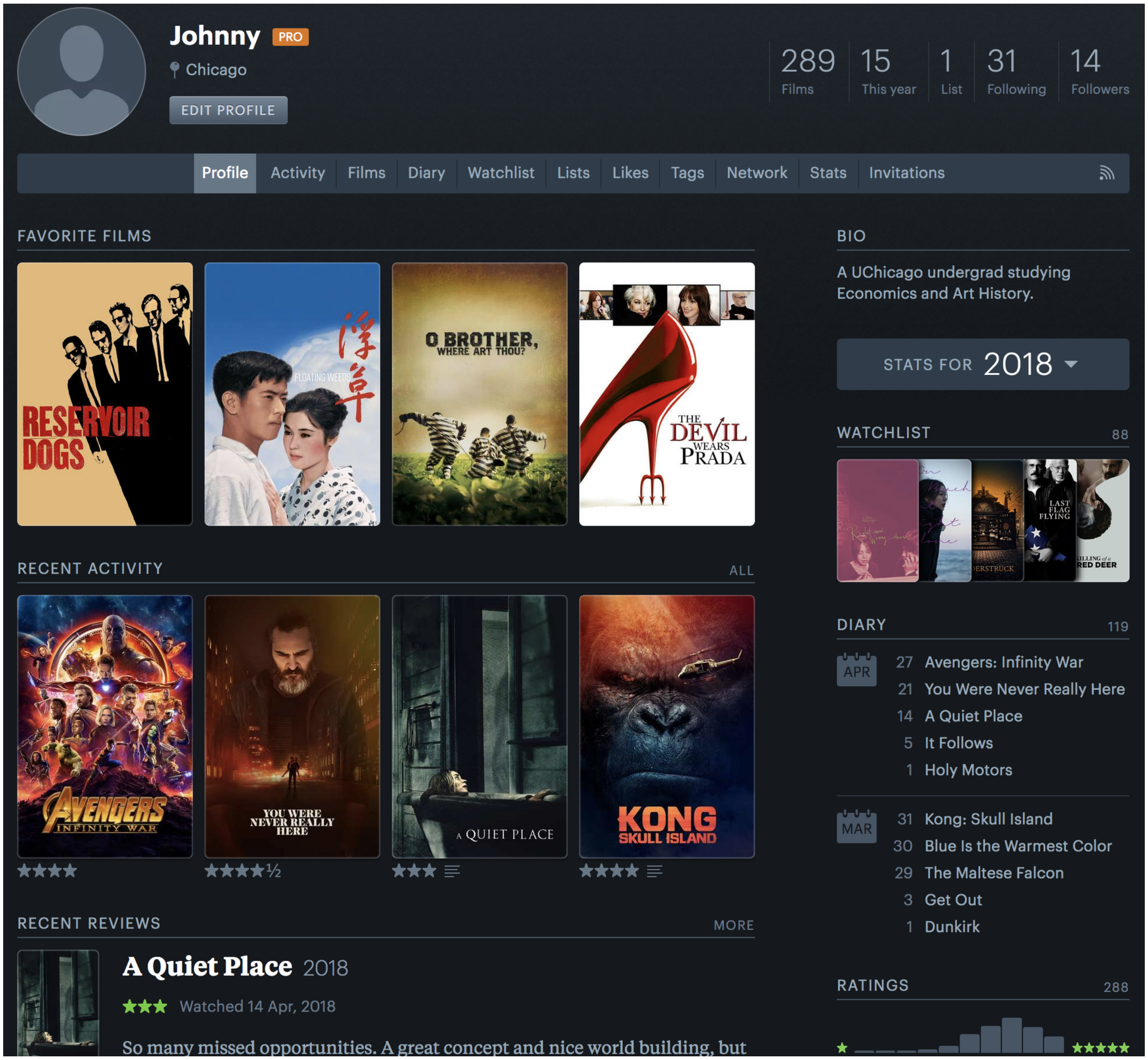
★★★★★ 27

Matt Singer

I Love You, Daddy 2017

13

Figure 3: Second half of front page, with reviews and recent behavior of your friends.



BIO

A UChicago undergrad studying Economics and Art History.

DIARY 119

APR	27	Avengers: Infinity War
	21	You Were Never Really Here
	14	A Quiet Place
	5	It Follows
	1	Holy Motors
MAR	31	Kong: Skull Island
	30	Blue Is the Warmest Color
	29	The Maltese Falcon
	3	Get Out
	1	Dunkirk

A Quiet Place 2018

★★★ Watched 14 Apr, 2018

So many missed opportunities. A great concept and nice world building, but

Figure 4: A user's page. A number of stats are easily available, as well as a brief bio and the user's favorite films.

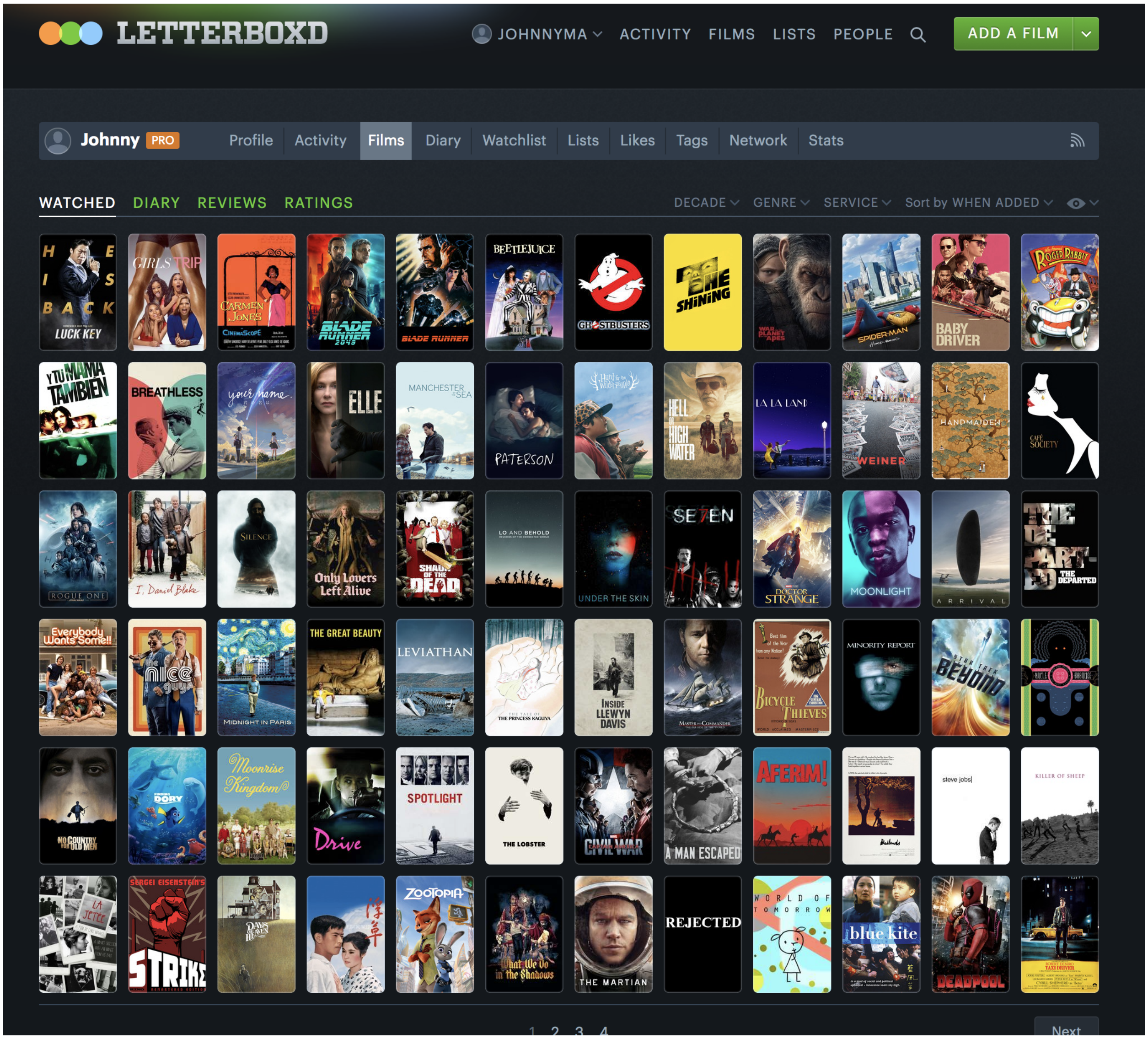


Figure 5: The list of films a user has marked as "watched."

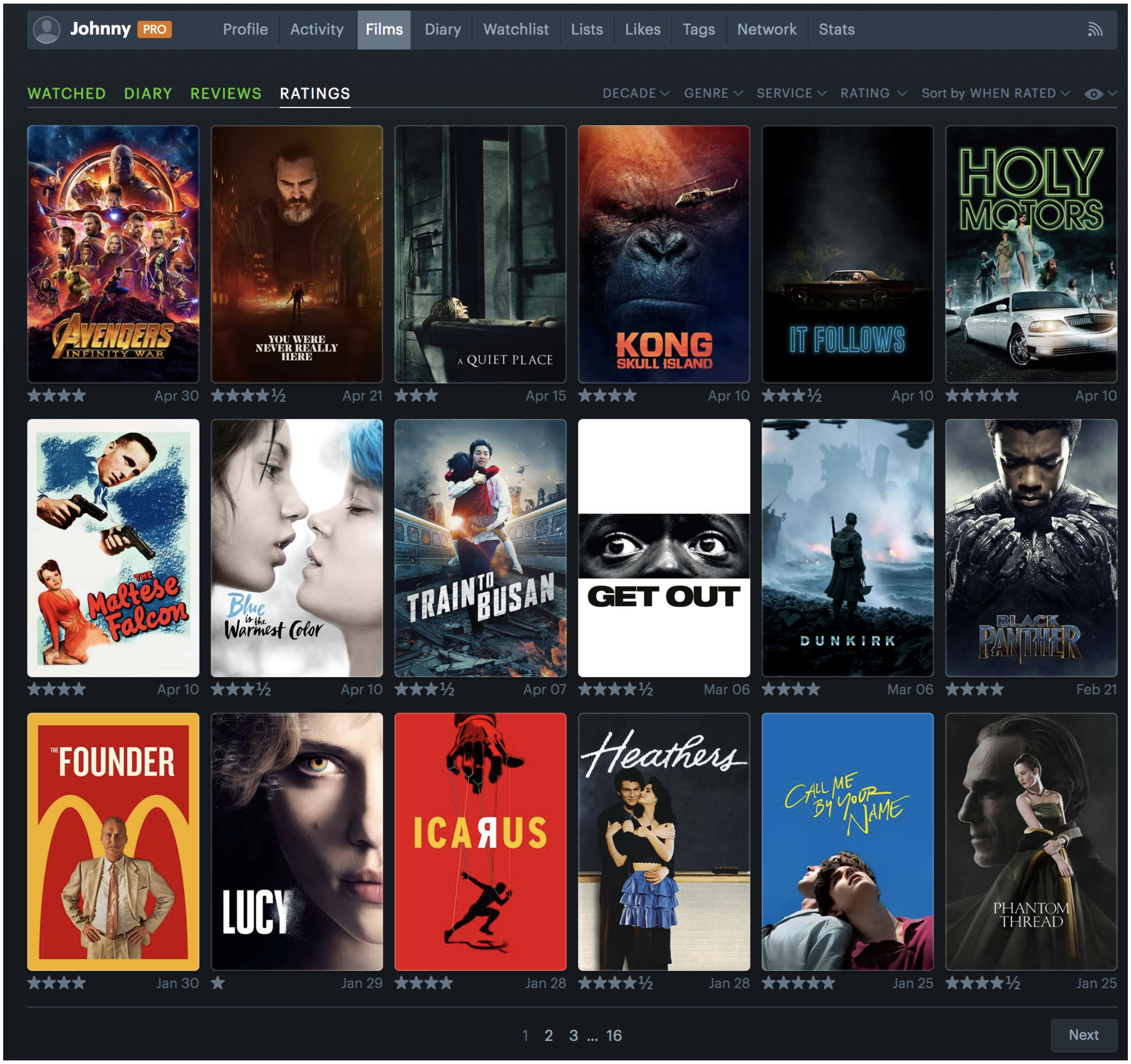
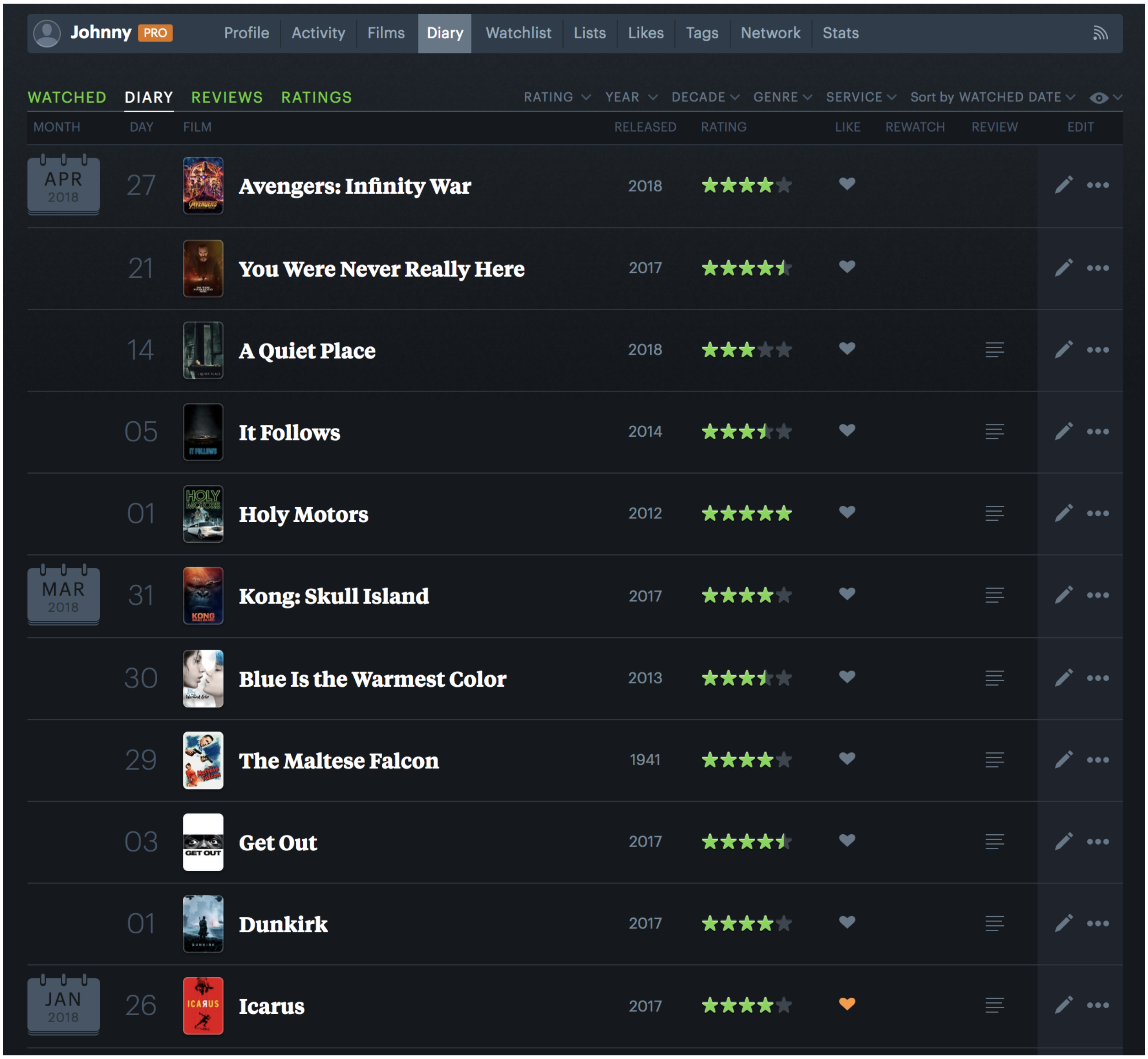
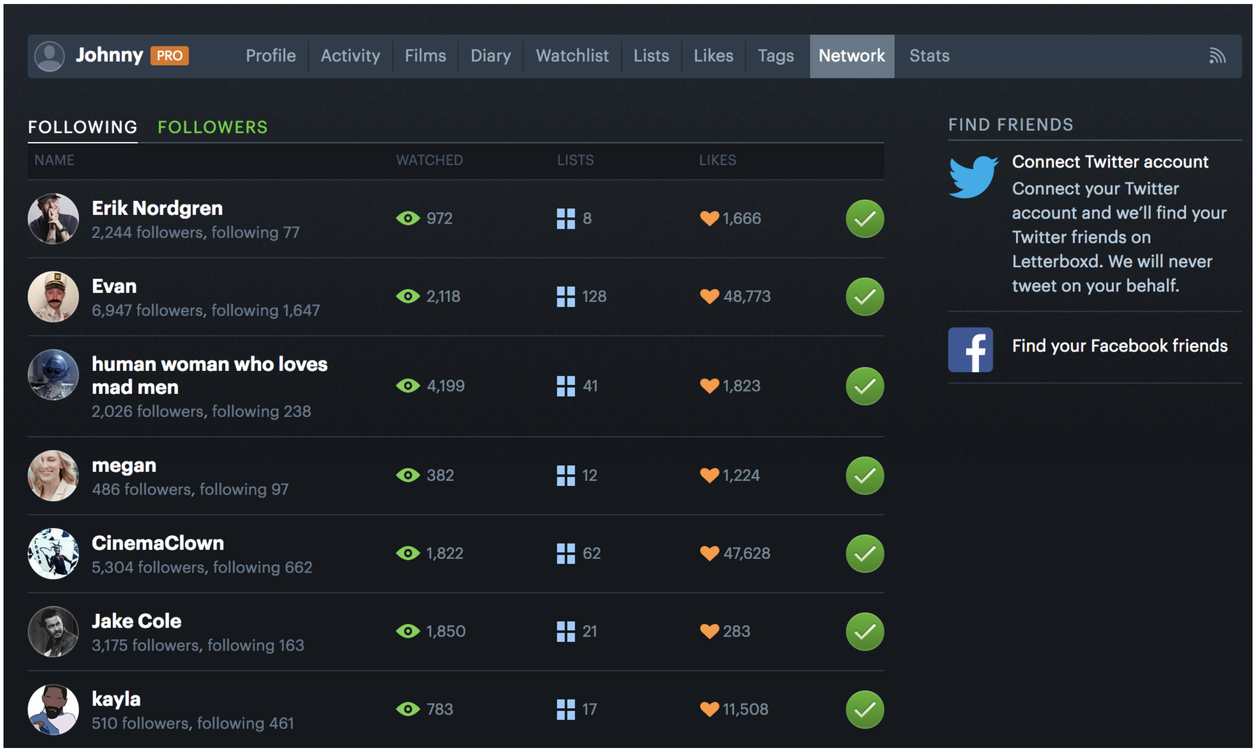


Figure 6: A more detailed list.



MONTH	DAY	FILM	RELEASED	RATING	LIKE	REWATCH	REVIEW	EDIT
APR 2018	27	Avengers: Infinity War	2018	★★★★	♥			✎ •••
	21	You Were Never Really Here	2017	★★★★½	♥			✎ •••
	14	A Quiet Place	2018	★★★	♥		≡	✎ •••
	05	It Follows	2014	★★★½	♥		≡	✎ •••
	01	Holy Motors	2012	★★★★★	♥		≡	✎ •••
MAR 2018	31	Kong: Skull Island	2017	★★★★	♥		≡	✎ •••
	30	Blue Is the Warmest Color	2013	★★★½	♥		≡	✎ •••
	29	The Maltese Falcon	1941	★★★★	♥		≡	✎ •••
	03	Get Out	2017	★★★★½	♥		≡	✎ •••
	01	Dunkirk	2017	★★★★	♥		≡	✎ •••
JAN 2018	26	Icarus	2017	★★★★	♥		≡	✎ •••

Figure 7: The user's diary. This is our panel level data.



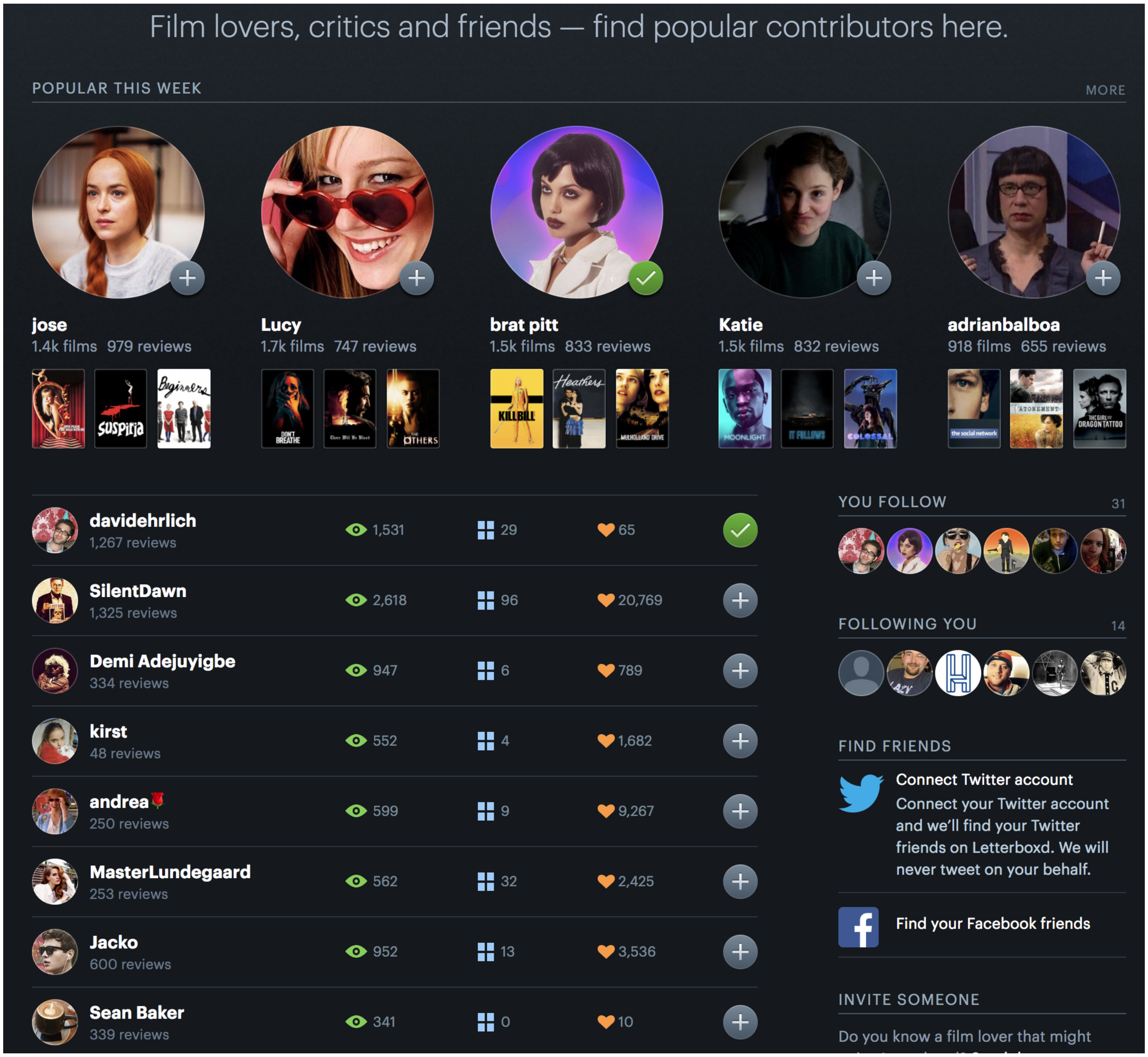
NAME	WATCHED	LISTS	LIKES	
Erik Nordgren 2,244 followers, following 77	972	8	1,666	✓
Evan 6,947 followers, following 1,647	2,118	128	48,773	✓
human woman who loves mad men 2,026 followers, following 238	4,199	41	1,823	✓
megan 486 followers, following 97	382	12	1,224	✓
CinemaClown 5,304 followers, following 662	1,822	62	47,628	✓
Jake Cole 3,175 followers, following 163	1,850	21	283	✓
kayla 510 followers, following 461	783	17	11,508	✓

Connect Twitter account

Connect your Twitter account and we'll find your Twitter friends on Letterboxd. We will never tweet on your behalf.

Find your Facebook friends

Figure 8: A list of other users the user is friends with. This determines some of the content a user sees on the website.



Film lovers, critics and friends — find popular contributors here.

POPULAR THIS WEEK

MORE

jose	1.4k films	979 reviews
Lucy	1.7k films	747 reviews
brat pitt	1.5k films	833 reviews
Katie	1.5k films	832 reviews
adrianbalboa	918 films	655 reviews

davidehrlich 1,267 reviews	1,531	29	♥ 65	✓
SilentDawn 1,325 reviews	2,618	96	♥ 20,769	+
Demi Adejuyigbe 334 reviews	947	6	♥ 789	+
kirst 48 reviews	552	4	♥ 1,682	+
andrea🌹 250 reviews	599	9	♥ 9,267	+
MasterLundegaard 253 reviews	562	32	♥ 2,425	+
Jacko 600 reviews	952	13	♥ 3,536	+
Sean Baker 339 reviews	341	0	♥ 10	+

YOU FOLLOW 31

FOLLOWING YOU 14

FIND FRIENDS

Connect Twitter account

Connect your Twitter account and we'll find your Twitter friends on Letterboxd. We will never tweet on your behalf.

Find your Facebook friends

INVITE SOMEONE

Do you know a film lover that might

Figure 9: A list of people on the website, roughly ordered by popularity and activity on the site.

From the list of users whose information we scrape from the website, we exclude users that have fewer than half of their films watched recorded in their diary[19], have at least 50 films in their diaries, and have more than 5 friends. This leaves us with about 9,000 users with panel-level diary and friend networks[20]. We also collect the popularity of each user's review at the time the data is scraped, which is roughly stable as the more popular reviews

[19] Many users import viewing data from previous diaries such as IMDB that are not time stamped. We exclude these movies from our selection criteria.

[20] I also include myself in the dataset.

tend only to get more popular. A few summary statistics are calculated using this data, as shown in the following two Figures 10 and 11.

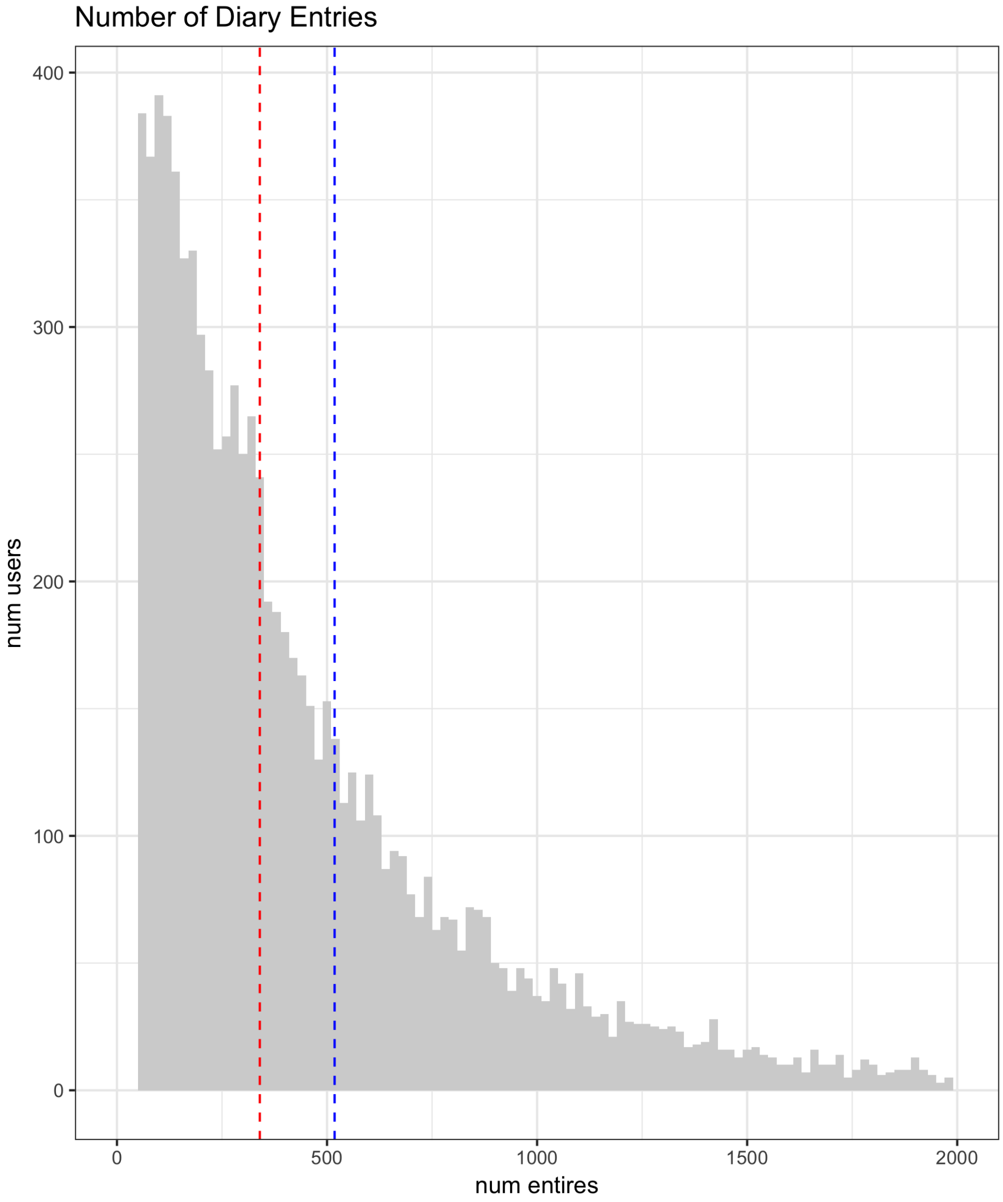


Figure 10: A histogram of number of diary entries. The median is 323 and the mean is 505. While these are unusually high numbers, *letterboxd* users show highly correlated box office behavior to that of the wider market. See Appendix.

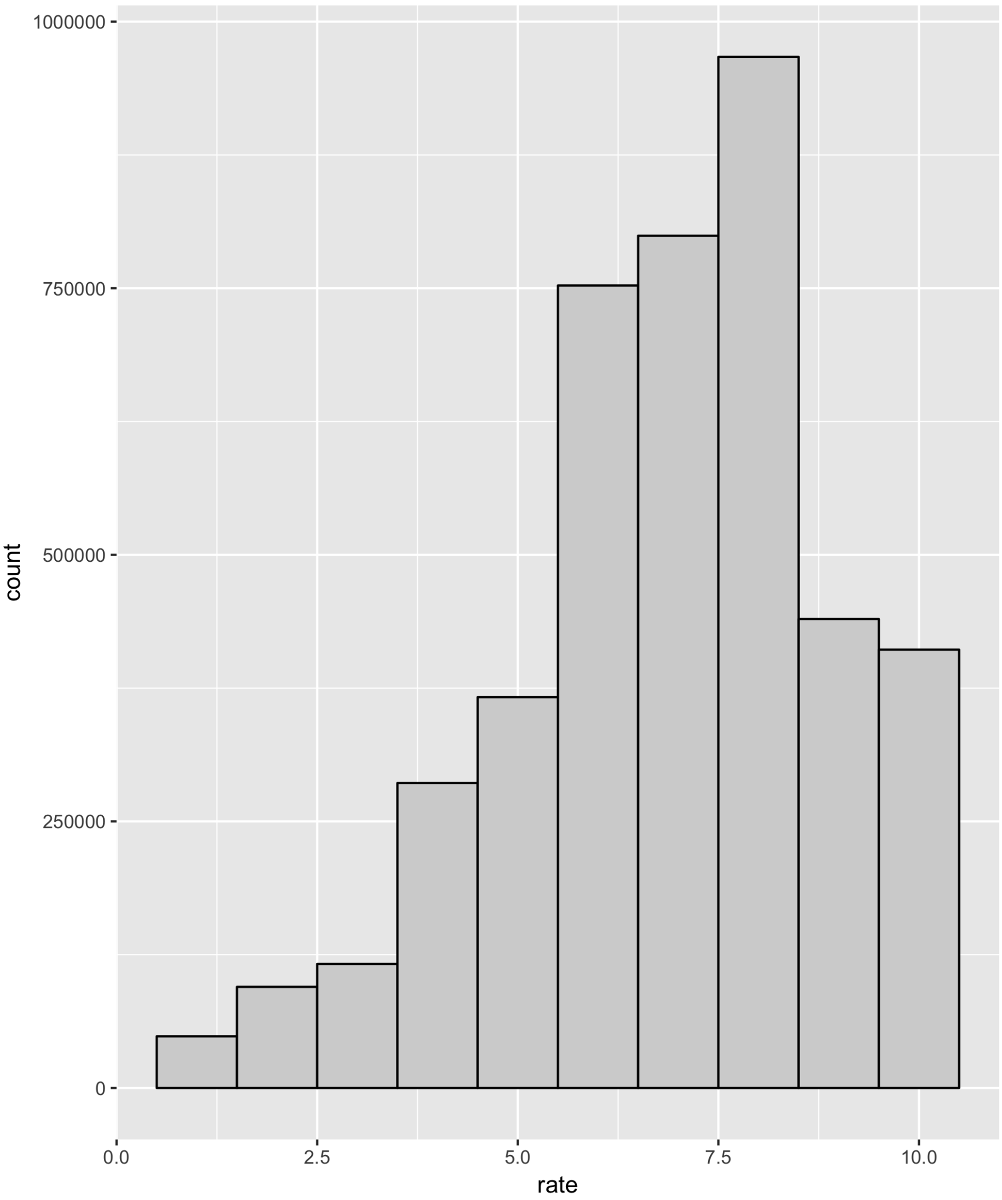


Figure 11: A histogram of the average rating for a film. The distribution is roughly a left skewed normal, as most users sample movies they are likely to enjoy.

4.2.2 Film Level Data

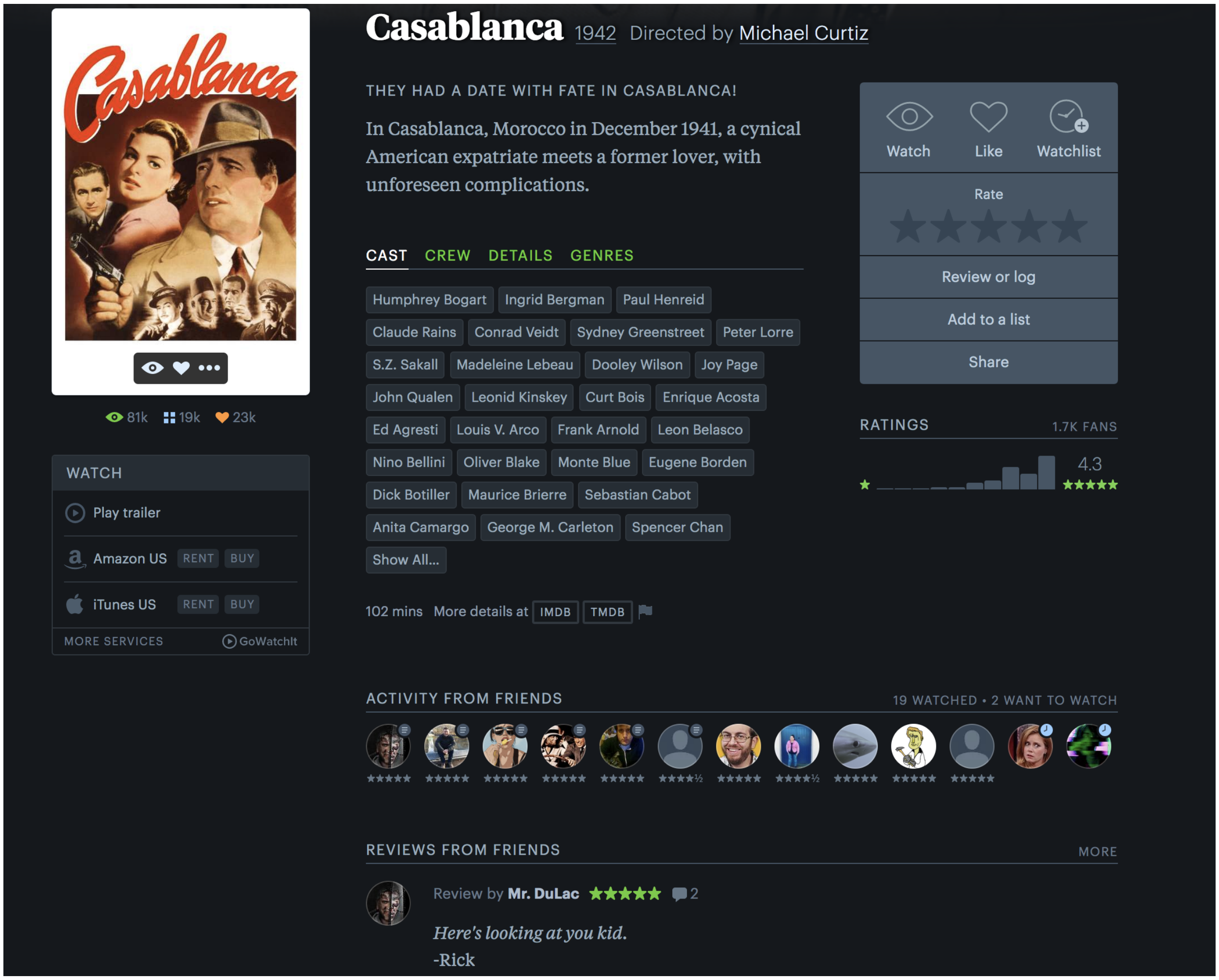
The website is also useful as a aggregator of film information, as well as a crowdsourced pool of ratings and reviews. Users will navigate to a specific film's page to check out the characteristics of the film as well as various reviews from their friends and the larger *letterboxd* community. Figure 12 shows a typical page for a film, in this case *Casablanca* [21]. The page contains cast, crew, genres, availability, and other information on the film. Along these publically available details, the page also shows a crowdsourced histogram of ratings from all users on the website, along with an average star rating (out of 5). Underneath this, the page shows the top 13 reviews of friends in your network, ordered by popularity of the review (which is calculated by number of likes on that review). Immediately after, letterboxd provides the text and rating of the top three most popular reviews by those in your network[22], followed by the three most popular reviews by those NOT in your network[23], as seen in Figure 13 for the film *The English Patient*[24]. We collect all the information on these pages, most notably the average rating that we use to construct our ”social information.”

[21] The favorite film of B.A. preceptor Kotaro Yoshida.

[22] Including re-reviews.

[23] An interesting level of heterogeneity that can be exploited in future papers.

[24] The favorite film of B.A. preceptor Victor Lima.


Casablanca 1942 Directed by Michael Curtiz

THEY HAD A DATE WITH FATE IN CASABLANCA!

In Casablanca, Morocco in December 1941, a cynical American expatriate meets a former lover, with unforeseen complications.

CAST CREW DETAILS GENRES

Humphrey Bogart, Ingrid Bergman, Paul Henreid, Claude Rains, Conrad Veidt, Sydney Greenstreet, Peter Lorre, S.Z. Sakall, Madeleine Lebeau, Dooley Wilson, Joy Page, John Qualen, Leonid Kinskey, Curt Bois, Enrique Acosta, Ed Agresti, Louis V. Arco, Frank Arnold, Leon Belasco, Nino Bellini, Oliver Blake, Monte Blue, Eugene Borden, Dick Botiller, Maurice Brierre, Sebastian Cabot, Anita Camargo, George M. Carleton, Spencer Chan, Show All...

102 mins More details at IMDB TMDB

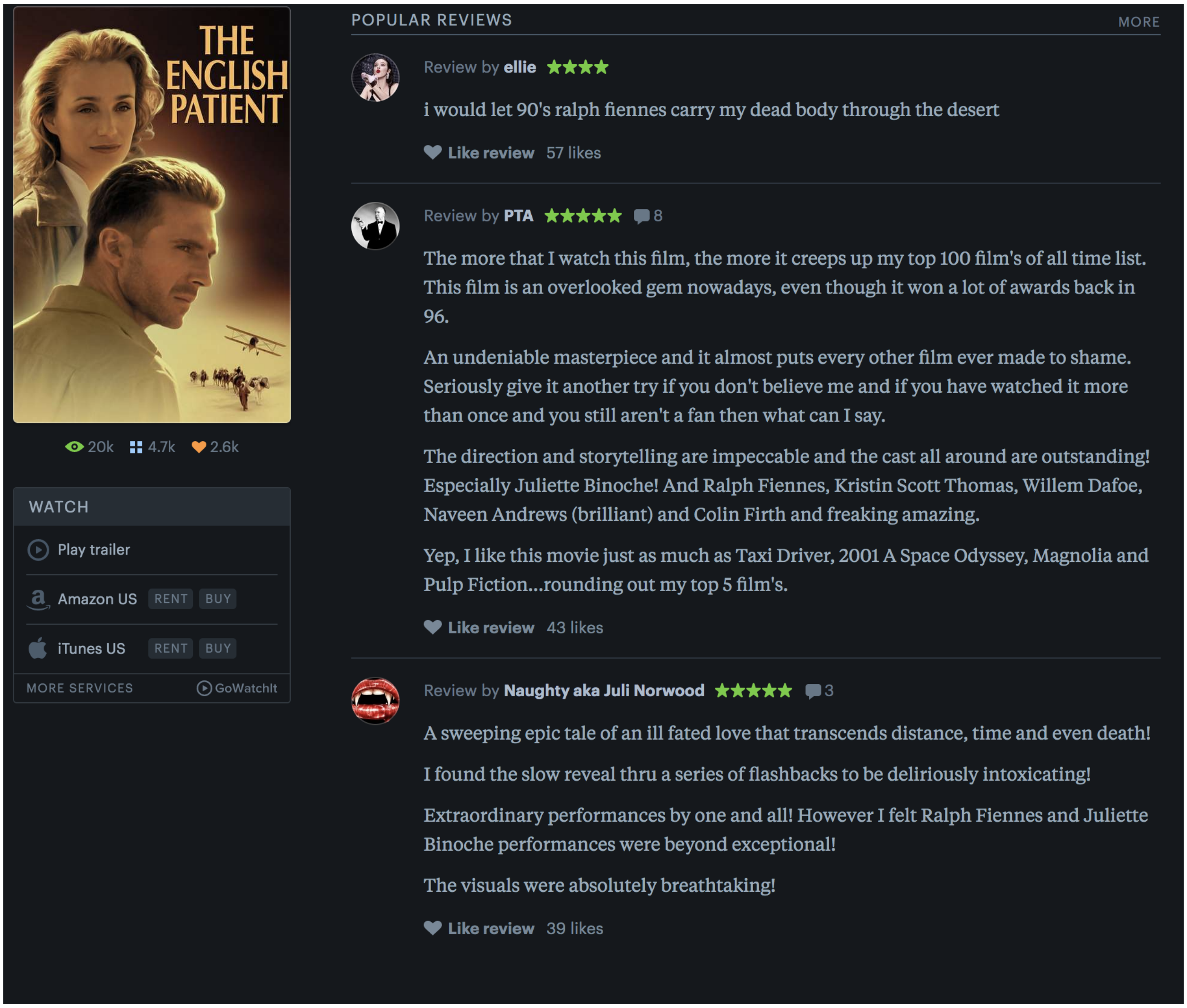
ACTIVITY FROM FRIENDS 19 WATCHED • 2 WANT TO WATCH

REVIEWS FROM FRIENDS MORE

Review by Mr. DuLac ★★★★★ 2

Here's looking at you kid.
-Rick

Figure 12: The sample page of a film. Note the heavy incoporation of information from friends, especially the list of ratings from friends ranked by review popularity.



Review by **ellie** ★★★★

i would let 90's ralph fiennes carry my dead body through the desert

♥ **Like review** 57 likes

Review by **PTA** ★★★★★ 8

The more that I watch this film, the more it creeps up my top 100 film's of all time list. This film is an overlooked gem nowadays, even though it won a lot of awards back in 96.

An undeniable masterpiece and it almost puts every other film ever made to shame. Seriously give it another try if you don't believe me and if you have watched it more than once and you still aren't a fan then what can I say.

The direction and storytelling are impeccable and the cast all around are outstanding! Especially Juliette Binoche! And Ralph Fiennes, Kristin Scott Thomas, Willem Dafoe, Naveen Andrews (brilliant) and Colin Firth and freaking amazing.

Yep, I like this movie just as much as Taxi Driver, 2001 A Space Odyssey, Magnolia and Pulp Fiction...rounding out my top 5 film's.

♥ **Like review** 43 likes

Review by **Naughty aka Juli Norwood** ★★★★★ 3

A sweeping epic tale of an ill fated love that transcends distance, time and even death!

I found the slow reveal thru a series of flashbacks to be deliriously intoxicating!

Extraordinary performances by one and all! However I felt Ralph Fiennes and Juliette Binoche performances were beyond exceptional!

The visuals were absolutely breathtaking!

♥ **Like review** 39 likes

Figure 13: A few sample reviews ranked by popularity. Most popular reviews are somewhere between one-liner jokes and a detailed analysis of characters and themes.

5 Empirical Model

The following is the main regression specification of the paper:

$$Pr(Watch_{ij})_{t+1} = \alpha_i + \beta_1 * s(friend)_{k,t} + \beta_2 * (residual)_{j,t=1} + \beta_3 * user_i + \beta_4 * quality_j + \epsilon_{ijt}$$

This is probit model for the probabilty of watching film j any week t after OW $t = 1$ for an individual i who did not watch j during OW, with user fixed effects and a control for the quality of film, defined as the average rating. The β coefficient in a probit model is interpretable as the marginal effect on the z-score of the probabilty $Watch_{ij} = 1$. We will report only the standard OLS results, which are very similar to the probit model.

The s(friend) term is $\in [0, 1]$ and is defined as the share of friends in your private network k that liked the film above the film's average rating. This is the idea of "social information" that comes from each users' unique network of friends[25]. For example, if two out of five of your friends gave a 10/10 rating to a film with an average rating of eight, and the other three friends gave a rating of 6/10, the s(friend) term would be 2/5. To best simulate the information a user would see on a film page, we exploit the algorithm that *letterboxd* uses to display your "friend activity." The site shows up to 13 of your friends, ordered by the user with the highest number of "likes" on their review, and so on. To calculate this "share" term we find the rating of up to 13 friends in each user's network that would be shown the weeks following OW, based on review popularity information obtained on the day the data was scraped. While this is at best a rough approximation, it gets us much closer to what a user would actually see post-OW.

The (residual) term is $\in [0, 1]$ and is defined as the residual from the regression of OW number of screen on OW gross for a film j, with controls for studio, rating, genre, week, holiday, and year as is typical in the literature. This is the OW "surprise in appeal" defined in Moretti's paper, with a positive value equating to unexpected demand. This is the idea of "social utility" that comes from the popularity of the film. In our model, an unexpected boost in popularity would cause each user to reevaluate the value that comes from watching the film and being "in" on the conversation.

[25] We follow Udry and Conley (2011) in measuring social information using a share ratio, giving equal weights to all friends.

6 Results

6.1 Moretti's Measure of Surprise in OW Demand

The first step to testing the predictions of the model is to empirically replicate Moretti's "surprise" using our box office data. The regression of number of OW screens on OW total gross for all films released in our time frame is reported below in Table 1.

Table 1: OW regression from Moretti

	Moretti Regression:
	OW Total Gross
Theaters Opening	1.278^{***} (0.039)
Observations	956
Controls?	✓
R^2	0.841
Adjusted R^2	0.816
Residual Std. Error	0.654 (df = 824)
Note:	*p<0.1; **p<0.05; ***p<0.01

We find the same positive and highly significant predictive power of number of opening screens on OW total gross, with a high R^2 value of 0.841. The remaining normally distributed residuals are normalized to $[0, 1]$ and are captured for the 212 films in both datasets. This is used in our final regression as the unit of social utility.

6.2 Social Model of Film Consumption

The main predictions of the social model of post-OW consumption is that positive feedback from friends on the quality of the film should increase the estimated quality and therefore the probability of watching, and higher than anticipated demand should increase the estimated utility of joining the site-wide conversation and therefore the probability of watching.The results of the OLS model of film consumption post-OW is reported below in Table 2.

Table 2: Social Model of Film Consumption

	OLS Model:			
	Probability of Watching After OW			
	(1)	(2)	(3)	(4)
share of friends above average	0.0138*** (0.0011)		0.0146*** (0.0011)	0.0073*** (0.0004)
residual box office surprise		0.1240 *** (0.0024)	0.1243 *** (0.0024)	0.1143 *** (0.0024)
movie quality				0.0368 *** (0.0004)
Fixed effcts?	User	User	User	User
Observations	359983	359983	359983	359983
R^2	0.1516	0.1575	0.1579	0.1721
Residual Std. Error	0.270 (df = 352071)	0.272 (df = 352071)	0.272 (df = 352070)	0.2705 (df = 352069)

*p<0.1; **p<0.05; ***p<0.01

Table 2: Share of friends above average is calculated using the ratings of the 13 friends with the most popular reviews and saw the movie during OW. Residual box office surprise is calculated from the Moretti regression. Movie quality is the average rating of the film out of 10.

As the results do not change much when moving from column (1) to (4), the full specification, we will focus the discussion on column (4)[26]. We can quickly see that the coefficient on "share of friends above average," the mechanism of social learning, and "residual box office surprise," the mechanism of social utility, are both positive and highly significant. This supports the notion that these measures of peer effects have something to do with the decision to watch a movie after OW, despite the low R^2 value of 0.1721. Unsurprisingly, the quality of the movie is also highly significant, with a movie with an average rating of 10 (highest) is associated with a 36% increase in the probability of post-OW watches compared to a movie with an average rating of 1 (lowest), *ceteris paribus*. Interestingly, the magnitude of the coefficent on the social information variable is 0.0073, an order of magnitude smaller than the coefficient on the social utility variable 0.1143. Since both variables are coded the same from 0 to 1, they can be compared by their magnitude. If all 13 friends rated a film

[26] This finding gives some backing to the separabliity of the two channels of social spillover, though is far from cleanly identified.

better than average during OW, this is only associated with an increase in 0.7% probability of watching compared to all disliking the film, whereas the residual box office surprise is asociated with a 11.4% increase in probability. While these two channels are not fully identified, the evidence seems to suggest that for these 9,000 users and 212 films, social utility plays a much larger role than social information does in determining post-OW trips to the box office.

7 Conclusion

This paper set out to quantify the role of peer effects on film watching behavior. The contribution of this paper to the literature is the usage of micro level watching data scraped from the newest realm of social interactions: online social media. Using data from *letterboxd.com* and *boxofficemojo.com*, we were able to look at watching behavior during a commercial film's theatrical run. Applying a Bayesian model of film consumption with updating of priors on film quality and popularity, we ran regressions looking at the role of social spillovers in increasing the probability of post-OW theater visits. Though the two channels are not cleanly identified, the paper largely supported previous literature that found positive and significant effects for both social learning and social utility, with social utility's effect on watching behavior dwarfing that of social information.

There is much future work to be done. FIrst, one can better identify these two channels using a friend-of-a-friend-but-not-my-friend IV approach to identify social learning [4] or use exogenous shocks on OW attendance to identify social utility. Second, we can look at heterogeneity from user and film level characteristics and quantify the effect of social learning versus social utility for these groups. We could potentially uncover those with reversed direction of effect for social learning (contrarians) or social utility (hipsters). Third, we can extend the "social spillover" nature of this analysis to event studies that affect the social stigma and identity that certain films hold. The Oscars and other awards ceremonies endow a certain status on films, perhaps affecting both social information and social utility. For better identification we could use exogenous events, such as the disgrace of Kevin Spacey or the death of a celebrity such as David Bowie. Fourth, we can add a bevy of robustness checks such as permutation tests to validate our findings for both the network formations and sampling behavior.

With these results, we could get closer to quantifying how how social forces and peer effects play a role in film consumption and taste formation. The landscape of digital content is constantly changing, and these methods and way of thinking as are applicable to films as they are to Youtube personalities, video game markets, Twitch streaming, and other new content found on social media websites. As industries evolve around these new arts, it is more important than ever to quantify and build products or campaigns around the absolutely vital role of social networks and peer effects. For now, film remains in its special place in the social milleu of Americans, with movie references and memes infiltrating every corner of social life. While cinema is at once industry and art form, these motion pictures, characters, and stories undoubtably capture our collective imagination. A clearer understanding film consumption can ultimately help us better understand the nature of human interactions and the shared human condition.

References

[1] Becker, Gary S. ”A note on restaurant pricing and other examples of social influences on price.” *Journal of Political Economy* 99, no. 5 (1991): 1109-1116.

[2] Bikhchandani, Sushil, David Hirshleifer, and IvoWelch. 1992. ”ATheory of Fads, Fashion, Custom, and CulturaL Change as Informational Cascades.” *Journal of Political Economy* 100 (5): 9921026.

[3] Bikhchandani, Sushil, and Sunil Sharma. ”Herd behavior in financial markets.” IMF Staff papers 47, no. 3 (2000): 279-310.

[4] Bramoull, Yann, Habiba Djebbari, and Bernard Fortin. ”Identification of peer effects through social networks.” *Journal of Econometrics* 150, no. 1 (2009): 41-55.

[5] Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. ”Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions.” *Econometrica* 82, no. 4 (2014): 1273-1301.

[6] Conley, Timothy G., and Christopher R. Udry. ”Learning about a new technology: Pineapple in Ghana.” *The American Economic Review* 100, no. 1 (2010): 35-69.

[7] DellaVigna, Stefano, and Johannes Hermle. ”Does Conflict of Interest Lead to Biased Coverage? Evidence from Movie Reviews.” *Review of Economic Studies* 84, no. 4 (2017): 1510-1550.

[8] Durlauf, Steven N., and H. Peyton Young, eds. *Social dynamics.* Vol. 4. Mit Press, 2004.

[9] Einav, Liran. ”Seasonality in the US motion picture industry.” *The Rand Journal of Economics* 38, no. 1 (2007): 127-145.

[10] Gilchrist, Duncan Sheppard, and Emily Glassberg Sands. ”Something to talk about: Social spillovers in movie consumption.” *Journal of Political Economy* 124, no. 5 (2016): 1339-1382.

[11] Goldberg, Amir, and Tony Vashevko. ”Social Boundedness as Market Identity Evidence from the Film Industry.” (2013).

[12] Jackson, Matthew O. Social and economic networks. Princeton university press, (2010).

[13] Manski, Charles F. ”Identification of endogenous social effects: The reflection problem.” *The Review of Economic Studies* 60, no. 3 (1993): 531-542.

[14] Moretti, Enrico. ”Social learning and peer effects in consumption: Evidence from movie sales.” *The Review of Economic Studies* 78, no. 1 (2011): 356-393.

Appendix A

7.1 Are Letterboxd Diaries a Good Proxy for Box Office Audience?

A major concern with using social network data from a specialized site is that the users on the site are not typical of the larger population we are trying to model. While it is true that the users of letterboxd.com likely spend far more time watching and thinking about movies than your average weekend cinema goer, the volume of watches from letterboxd users is highly correlated with box office returns (a proxy for audience size) during a film's theatrical run. In Figure A1, we plot both weekly aggregate number of diary entries from users on letterboxd and the weekly box office gross of the Disney movie *Black Panther* (2018). The correlation betweeen these two measures of audience size is 0.84, higher than previous literature's usage of google search results, which had a total correlation of 0.74. Taking all 212 films in our dataset, we calculate an overall correlation of 0.78, implying that letterboxd watching data is a good proxy for national viewership.

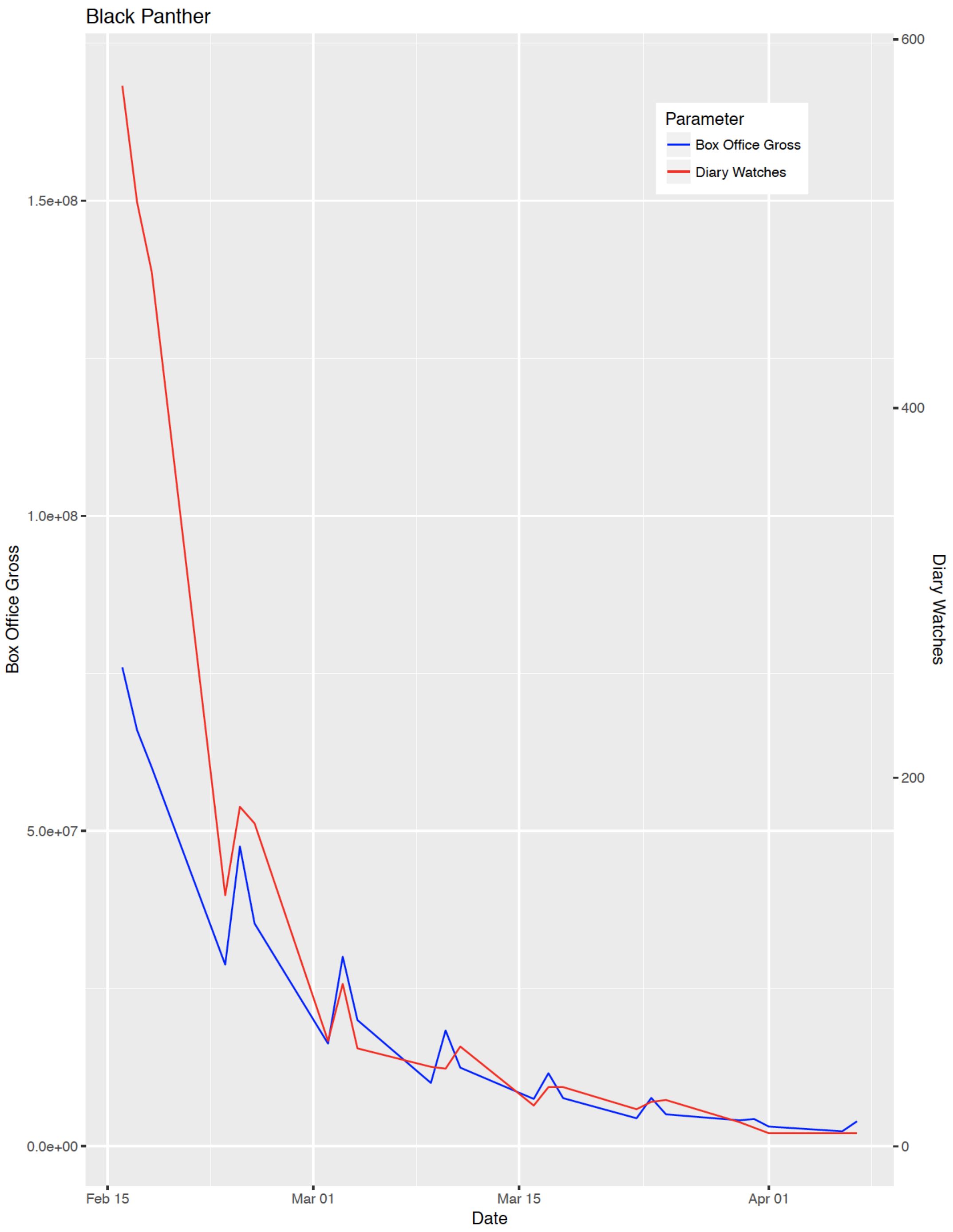


Figure A1: The aggregate watches and gross of the movie *Black Panther* over its theatrical run.

7.2 Are Friends Formed Through Similar Taste?

A key part of our analysis is the assumption that people form friends based on similar taste. This implies that a friend's rating would be more informative than a randomly selected individuals, as tastes are correlated. We attempt to empirically validate this logic by creating a measure of taste similiarity. Taking each user's set of film ratings, we calculate the cosine similarity between the long vector of ratings for a user and all other users, (excluding NAs) giving us an $i \times i$ covariance matrix of taste simliarity, ranging from -1 to 1. This is similar in nature to how recommendation systems for films, TV, and other goods are built through crowdsourcing. We then regress this measurement of taste against a binary variable of friend links, 1 if two users are linked and 0 if not. This should capture the effect of having similar taste on probability of forming a friendship. The results are reported below.

Table 3: Model of Network Formation based on Taste

	Is Friend Binary:
	linked
Cosine Similarity of Taste Vector	0.031 ***
	(0.0058)
Observations	159913
Note:	*p<0.1; **p<0.05; ***p<0.01

As the results are significant and positive, we can confidently say that friend networks are formed at least somewhat along taste lines. A more rigorous measurement of taste may be useful to build for future projects.