The impact of multi-studio franchises on brand-associations in the motion-picture industry

David Dornekott
Universität Osnabrück

Cite as:
Dornekott David (2020), The impact of multi-studio franchises on brand-associations in the motion-picture industry. *Proceedings of the European Marketing Academy*, 49th, (60911)

Paper from the 49th Annual EMAC Conference, Budapest, May 26-29, 2020.

The impact of multi-studio franchises on brand-associations in the motion-picture industry

Abstract:

Using an extensive dataset of more than four million comments submitted to social news aggregator reddit.com, collected over the course of a full calendar year, brand associations within the motion picture industry are investigated. A network model of brands is proposed, wherein production studios are connected to individual products via intermediate brands such as actors, franchises and sub-brands. It can be shown that consumers strongly associate titles with their producing companies, as well as other titles produced by those companies, indicating a high salience of producer's brands in the minds of consumers. Similarly, actors are highly associated with titles they appear in, signifying their importance to the marketing of motion pictures. Evidence of spill-over between the brands of competing studios engaged in joint productions of franchises is found, implying that a new challenge to brand management has been introduced by the recent emergence of multi-studio film franchises.

Keywords: *user-generated content, brand-associative networks, brand alliances*

Track: Digital Marketing & Social Media

1. Introduction

Branding has long been considered a key aspect in the marketing of motion pictures, as brands play a crucial role in setting consumer expectations and signaling quality and familiarity. Motion pictures can be thought of as composite products that involve multiple cooperating brands, wherein aspects of the product itself - such as its placement within a franchise or the characters it features - and the companies and people involved in its production - such as studios, labels, actors or directors - can themselves be considered independent brands that together inform the product's overall identity. (O'Reilly & Kerrigan, 2013) More broadly, they can thus be considered special cases of brand alliances (Kupfer, Pähler vor der Holte, Kübler, & Hennig-Thurau, 2018). Recent years have seen the emergence of cinematic universes, franchises-of-franchises, wherein each property is connected to other properties via shared themes, characters or locations and is marketed under an overarching brand. Their production and distribution routinely involves not only brands on different levels of the value chain, such as studios and actors, but also brands that are otherwise considered competitors. This can be a result of direct cooperation (e.g. Disney and Sony cooperating in the production of titles within the Spider-Man franchise) or indirect affiliation through secondary brands (e.g. Disney and Fox's indirect affiliation through the Disney-owned Marvel Comics, which is the overarching brand of the Fox-produced X-Men franchise). As a result, a studio's brand may be inadvertently tied to products controlled by competing studios, emphasizing just how prominent and complex brand interactions are in the contemporary motion picture industry. The goal of this study is to shine a light on these complex arrangements by empirically measuring how brands in the motion picture industry are associated in the minds of consumers. More specifically, the existence of spillover between competing studio's brands is to be investigated. To do so, a brand-associative network model (John, Loken, Kim, & Monga, 2006) of the contemporary film industry is inferred from a large corpus of movie-related discussions gathered from the social-news aggregation platform *reddit*. The resulting model is subsequently internally validated and analyzed.

While a number of extant studies concern themselves with the mining of brand and product-related information from user generated content (such as, for example, Netzer, Feldman, Goldenberg, & Fresko (2012), as well as Culotta & Cutler (2016) and Klostermann, Plumeyer, Böger, & Decker (2018)), this study is, to the knowledge of its author, the first to do so within the context of the contemporary motion picture industry. Additionally, while the impact and dynamics of online user generated content has been intensively studied, often in

the context of movies (see Babić Rosario, Sotgiu, De Valck, & Bijmolt (2016) for a recent meta-analysis), none of them have used data gathered from reddit, which has a number of unique properties that make it attractive for marketing research.

2. Theoretical framework and research questions

Motion pictures as a form of experience goods and the branding aspects surrounding them have been intensively studied by marketing researchers. (Kupfer et al., 2018; O'Reilly & Kerrigan, 2013; Vujić & Zhang, 2018) What, as of now, remains an open question is whether the complex arrangements of brands involved in the production of motion pictures can lead to spillover between brands engaged in such cooperative efforts, via their (direct or indirect) association in the minds of consumers. Such brand associations must first be quantified in some way. One approach to this comes in the form of brand-concept maps as described by John et al. (2006), who used a survey-based approach to identify qualities associated with brands. However, such survey-based methods are fairly involved and expensive to apply on an industry-encompassing scale. Other approaches take advantage of the growing amount of digital user generated content to infer associations between keywords based on their co-occurrence within documents. Online media enable consumers to publically communicate about products and brands and have become increasingly important platforms for brand management, both as a source of information and as a tool to engage consumers (Gensler, Völckner, Egger, Fischbach, & Schoder, 2016). Netzer et al. (2012) use text-mining techniques to infer brand- and product-characteristic-related associations from online discussion boards for automobiles and diabetes drugs. This study builds upon their approach and applies it to the motion picture industry, while differing in some key ways. Most notably, the structure of the model is not inferred from the data, but pre-identified based on industry data. Furthermore, the focus of the analysis is squarely put on associations between brands, as opposed to those between product characteristics and producer's brands.

Once the model is constructed and the associations between brands have been quantified, the model can be used to investigate a variety of questions on brand-interactions within the motion picture industry. Actors, for example, routinely partake in the marketing of titles they appear in and actively manage their own brands by associating themselves with certain roles (Kupfer et al., 2018). Given this prominence within marketing campaigns, it would follow that consumers should more strongly associate actors with those titles they appear in, compared to titles they do not appear in. By analogous argument, titles should be more strongly associated with the studios involved in their production as compared to other studios.

Such conjectures can serve as baseline hypotheses, which can be used to validate the model insofar as ensuring it produces the a priori expected outcomes. Contingent on the model behaving as expected, further research questions may then be investigated. Of particular interest is whether the aforementioned multi-studio production arrangements underlying cinematic universes lead to spillover between the brands of studios engaged in them. In order to provide evidence for or against the existence of such spillover-effects, the relationship between studios, sub-brands and franchises is analyzed in depth.

3. Data and methodology

3.1 Reddit as a source of user-generated content

User generated content has become a staple source of data for marketing research over the past two decades. Popular targets for data collection have been twitter (Vujić & Zhang, 2018), Instagram (Klostermann et al., 2018) and static web forums (Netzer et al., 2012). A platform that has so far been entirely absent in the marketing literature, is reddit. As a social news aggregation and discussion platform, it allows registered users to submit content - such as links to other websites, images or videos, as well as self-authored text - which is subsequently voted on and discussed by other users. Users can *upvote* or *downvote* a given submission, affecting its ranking relative to other submissions. The ranking is determined by the score of the submission (downvotes subtracted from upvotes) weighted by its age as measured by the time in minutes since submission it was submitted (Stoddard, 2015). Each submission to reddit includes its own comments section, wherein users can discuss the submission's content. The top 50 ranked submissions at any given moment appear on the *frontpage* of reddit, which commonly results in a large boost in exposure for submission that are featured this way. Apart from the general frontpage of reddit, users self-organize into specialized communities referred to as *subreddits*. This makes reddit very conductive to researchers investigating specific topics, such as motion pictures.

3.2 Data collection and aggregation

Reddit data was collected over a period covering the full calendar year of 2018. The dataset consists of a total of 438,725 observations of 76,231 unique submissions to the frontpage of reddit, as well as the movie-focussed subreddit */r/movies*, collected via hourly snapshots. These snapshots provide information on all submissions visible at the time of collection and their ranking relative to each other. The postings' unique identifiers (*permalinks*)

were subsequently used (via the official Python Reddit API Wrapper) to retroactively retrieve all comments pertaining to each movie-related post on January 28th, 2019, totaling 4,116,944 unique comments by 546,110 authors. Daily North American box office data starting from January 1st 1993 were collected on January 23rd 2019 from boxofficemojo.com. Additional information was taken from the IMDB data repository (https://datasets.imdbws.com/) on October 24th, 2019. This included information on all titles, actors and their associated roles in IMDB's database.

3.3 Identification of relevant keywords and construction of association measures

Based on the box office data, the top 30 movie studios of 2018, as measured by total theatrical gross, were first identified, along with all titles released by these top 30 studios. The resulting list of titles was then matched with the IMDB data to connect associated characters and actors. The relationship between characters and the titles they appear in enabled the inference of franchises, as movies that share (non-ambiguously named) characters can be considered likely to be part of a common franchise. This narrowed down the list of potential franchise-entries to a more manageable 135, for which connections to franchises and production labels were subsequently researched. The described approach minimizes the amount of manual research needed to construct a comprehensive set of relevant keywords, while providing a pre-structured model of the investigated industry to test the previously formulated hypotheses against. The structure chosen for the model at hand features studios (which finance and distribute movies), sub-brands (usually direct subsidiaries of studios), titles (the motion picture products), franchises (overarching brands of multiple connected titles), as well as actors and the characters they portray.

Overall, 2070 unique keywords were identified, constituting 29 studios, 6 sub-brands, 28 franchises, 376 characters, 589 actors, and 1042 titles. As a first step in analyzing the comment dataset for mentions of these keywords, the unit of observation that serves as the basis for analysis had to be chosen. Netzer et al. (2012), who obtained data from static discussion forums, identify discussion threads (sets of messages, ideally about a shared topic), messages (sets of sentences) and sentences (sets of words) as possible observational units, all of which have their functional equivalent within the reddit data. Thus, analogous to Netzer et al. (2012), the message level was chosen as the observational unit for analysis. All previously identified keywords were subsequently matched against all collected comments' texts using case insensitive direct matching. The list was then manually checked for and cleaned of key-

words that were likely to have triggered a high number of false positives due to their ambiguity or use as common idioms (i.e. “Anything”, “Wonder”, “Bigger”, “Zero”). Only keywords that were mentioned more than 50 times overall were used to detect co-occurrences, in order to reduce computational time and model complexity. Analogous to Netzer et al. (2012), the Jaccard Index (henceforth multiplied by 1000 for increased readability and abbreviated as JI) was chosen as a measure of association, which was subsequently computed for all dyadic keyword relationships with more than ten co-occurrences. The left half of Figure 1 shows a high-level overview of the network structure derived from these associations, pruned to only show nodes connected to studio nodes and visualized using a force-directed graph algorithm using the measured JI as edge weights. The right half of Figure 1 shows a magnified view of the central cluster. Red and green fill colors denote studio and sub-brand nodes, respectively. Titles are generally clustered around their producing studios, with franchises and sub-brands acting as bridges between studios. Actors are connected to titles they appear in, as well as other actors connected to those titles and the characters they portray. The major studios are heavily clustered, with Disney standing out as a major stand-alone structure. Streaming services (Netflix and Amazon) and arthouse-oriented studios (A24 and Annapurna Pictures) similarly form their own distinct clusters.

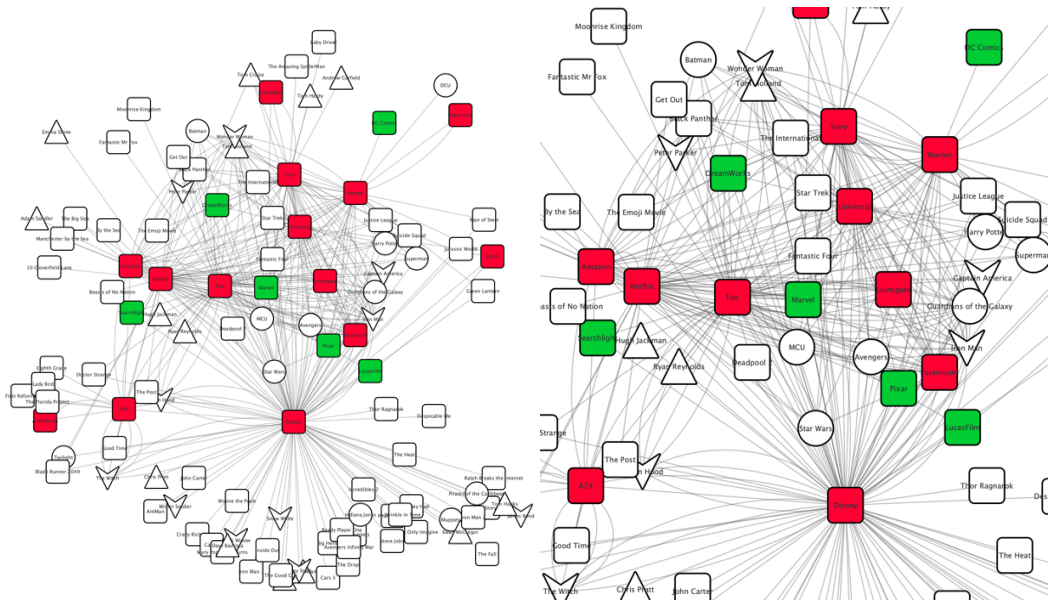

Figure 1: Network graph of the brand model. Left: Overall topology. Right: Zoomed portion of the core cluster

4. Analysis of the brand-associative model

In order to validate that the model is fundamentally qualified to capture associations between keywords, it is first investigated whether it exhibits certain behaviors that would be

expected on a baseline level. To do so, associations are first grouped by the type of relationship they signify (i.e. studios and titles they produced, titles and titles produced by the same studio, actors and characters they portray and so forth). Subsequently, the distribution of the association measures within these group-conditions is compared to the distribution of associations within the group of relationships forming the logical opposite condition (i.e. associations between actors and titles they appear in are compared to associations between actors and titles they do not appear in). It is then tested whether both sets of association measures share the same distribution. Since the association measures cannot be assumed to be normally distributed, the paired Mann-Whitney-Wilcoxon test is used comparing the null-hypothesis of the test-condition group stemming from a distribution with a mean smaller or equal the mean of the comparison group (Bauer, 1972). Comparing the set of all associations between actors and the titles they act in, averaged over each actor (mean JI of 4.62, standard deviation of 4.8 based on 115 observations), to the set of all averaged associations between actors and the titles they do not act in (mean JI of 0.06, standard deviation of 0.045), unsurprisingly results in a rejection of the null hypothesis at the highest level of confidence ($p < 0.001$), with an estimated location shift of 3.87. Similarly, comparing the set of associations between all studios and titles produced by them (mean JI of 2.69, standard deviation of 2.37, based on 11 observations), to the set of all associations between the same studios and titles they were not involved in producing (mean JI of 0.14, standard deviation of 0.05), results in an estimated location shift of 2.14, which is again statistically significant at the highest level of confidence ($p < 0.001$). Analogously, associations between titles and titles that were produced by the same studio (mean JI of 1.11, 240 observations) are significantly stronger compared to those between titles and titles produced by other studios (mean JI of 0.25), with an estimated location shift of 0.25 ($p<0.001$).

Table 1 gives an overview of the top ten dyadic relationships, as well as the ten most highly connected nodes in the network as denoted by their degree (the number of unique nodes they are connected to). The most commonly observed keyword-pairs are between the characters of Batman and Superman, followed by the studio Disney and its subsidiary Marvel. The dominance of Marvel-related properties is immediately apparent, although it should be noted that the simplicity and lack of specificity of the keywords (most of which are both characters or franchises, as well as subsets of specific titles) likely plays an important role in this. When aggregating by type of relationship (i.e. by the classifications of keywords), it can be seen that associations between sub-brands, franchises, as well as franchises and sub-brands are particularly strong, as are those between studios and sub-brands and studios. The top ten

highest-degree nodes in the Network are again dominated by comic-book properties, with Iron Man and Wonder Woman as the most highly connected nodes.

Top co-occurrences				Strongest relationship types			Highest degrees	
keyword one	**keyword two**	**n**	**Jaccard**	**type one**	**type two**	**mean Jacc.**	**keyword**	**degree**
Batman	Superman	5936	128.64	sub-brand	sub-brand	6.40	Iron Man	992
Disney	Star Wars	4022	52.54	franchise	subbrand	5.02	Wonder Woman	796
Iron Man	Marvel	2816	51.89	franchise	franchise	4.56	Guardians of the Galaxy	652
Marvel	MCU	2654	47.50	studio	studio	3.93	Marvel	380
Disney	Marvel	2547	31.42	studio	subbrand	3.74	Star Wars	355
Marvel	Avengers	2464	44.92	character	character	3.43	Disney	342
Marvel	Star Wars	2010	25.95	character	subbrand	2.92	Robin Hood	332
Marvel	Black Panther	1960	37.36	character	franchise	2.81	The Witch	324
Iron Man	Avengers	1912	63.80	title	title	2.76	Batman	322
Disney	Fox	1893	39.06	actor	actor	2.64	Netflix	316

Table 1: Top 10 co-occurrences of keyword-pairs, relationships by node type and degrees by node.

Taking steps to answer the underlying research question of whether spillover between competing studios can be observed, the relationships between studios, sub-brands and franchises are more thoroughly investigated. To build upon the previously given example, Sony is strongly associated with Disney (Sony-Disney JI = 10.5, compared to an average association with other studios of 3.8 for Sony and 5.4 for Disney). It is also indirectly associated with Disney through the Marvel sub-brand (Marvel-Sony JI = 26.8, Marvel-Disney JI = 31.4), as well as the Spider-Man property (SM-Sony JI = 39.6, SM-Disney = 9.5). This finding can be considered evidence in favor of the hypothesis that such spillovers were a driving factor in Disney seeking increased creative control over the Spider-Man franchise, as postulated in industry publications (The Hollywood Reporter, 2019). Similarly, Fox is strongly associated with Disney (Fox-Disney JI = 39.1, compared to an average association with other studios of 5.9 for Fox), likely due to discussion of the former's acquisition by the latter, which was announced over the course of data collection. However, it is also heavily associated with the Marvel sub-brand (Marvel-Fox JI = 15.5, Marvel-Disney JI = 31.4), as well as the X-Men (XM-Fox JI = 44.6, XM-Disney = 26.8, XM-Marvel JI = 18.9) and Deadpool (DP-Fox JI = 13.4, DP-Disney JI = 8.6, DP-Marvel JI = 9.8) properties. It can thus be seen that the X-Men and Deadpool franchises are both highly associated with Disney, though less so compared to Fox, which is their producing studio and Marvel, which is their overarching sub-brand.

In summary, studios that are indirectly affiliated via shared franchises are significantly more strongly associated compared to studios they do not share such affiliations with. Furthermore, the properties at the heart of these relationships are strongly associated with these

studios, even when they had no active part in their production.

5. Discussion and future research

This study has shown that the combination of a pre-defined model with weights determined via text-mining of a large corpus of user generated content can be a valuable tool for marketing research. It was found that consumers associate movies with their producing studios and actors with titles they appear in, providing confirmatory evidence for common assumptions about the roles of actors and studio brands in the marketing of motion pictures. Regarding the existence of spillover-effects between brands engaged in multi-studio production deals, it was observed that consumers more strongly associate studios that are connected via shared franchises, providing evidence for the existence of the phenomenon. This evidence must, however, still be considered anecdotal due to the low number of cases and lack of control variables, leaving plenty of room for future research on this topic.

As the model used in this paper is built upon a fairly rudimentary text-mining approach, wherein only direct co-mentions of pre-identified keywords were measured, it does not take into account the context in which the co-mentions occur (see also Culotta & Cutler (2016), who further elaborate on why this is problematic). The next step to extend the present model would thus be to provide more context to the associations, for example by including their valence. The model could then be further extended with by including contextual keywords, which might be identified via topic analysis and association rule learning. This would transform the model from one of brand associations into one of brand perceptions. In future research, the question of what determines the strength of associations between brands should be addressed. A possible approach to this could be to gather data on how the specific titles used in the present sample were branded in their official marketing communication and whether this has a measurable effect on the strength of their association in the minds of consumers. The data this study is based on, as well as the R and Python code used in the construction and analysis of the model, are available upon request.

References

Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of Marketing Research*, *53*(3), 297–318. https://doi.org/10.1509/jmr.14.0380

Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, *67*(339), 687–690.

https://doi.org/10.1080/01621459.1972.10481279

Culotta, A., & Cutler, J. (2016). Mining Brand Perceptions from Twitter Social Networks. *Marketing Science*, *35*(3), 343–362. https://doi.org/10.1287/mksc.2015.0968

Gensler, S., Völckner, F., Egger, M., Fischbach, K., & Schoder, D. (2016). Listen to Your Customers: Insights into Brand Image Using Online Consumer-Generated Product Reviews. *International Journal of Electronic Commerce*, *20*(1), 112–141. https://doi.org/10.1080/10864415.2016.1061792

John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, *43*(4), 549–563. https://doi.org/10.1509/jmkr.43.4.549

Klostermann, J., Plumeyer, A., Böger, D., & Decker, R. (2018). Extracting brand information from social networks: Integrating image, text, and social tagging data. *International Journal of Research in Marketing*, *35*(4), 538–556. https://doi.org/10.1016/j.ijresmar.2018.08.002

Kupfer, A.-K., Pähler vor der Holte, N., Kübler, R. V., & Hennig-Thurau, T. (2018). The Role of the Partner Brand's Social Media Power in Brand Alliances. *Journal of Marketing*, jm.15.0536. https://doi.org/10.1509/jm.15.0536

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, *31*(3), 521–543. https://doi.org/10.1287/mksc.1120.0713

O'Reilly, D., & Kerrigan, F. (2013). A view to a brand: Introducing the film brandscape. *European Journal of Marketing*, *47*(5), 769–789. https://doi.org/10.1108/03090561311306868

Stoddard, G. (2015). Popularity and Quality in Social News Aggregators: A Study of Reddit and Hacker News. *Www*, 815–818. https://doi.org/10.1145/2740908.2742470

The Hollywood Reporter. (2019). "Spider-Man" Shocker: Disney, Sony Striking a Deal for One More Movie. Retrieved November 27, 2019, from https://www.hollywoodreporter.com/heat-vision/spider-man-shocker-disney-sony-striking-deal-new-movie-1243777

Vujić, S., & Zhang, X. (2018). Does Twitter chatter matter? Online reviews and box office revenues. *Applied Economics*, *50*(34–35), 3702–3717. https://doi.org/10.1080/00036846.2018.1436148