

Using Socio-Semantic Network Analysis for Assessing the Impact of Documentaries

Jana Diesner, Amirhossein Aleyasen, Jinseok Kim, Shubhanshu Mishra, Kiumars Soltani

The iSchool, Department of Computer Science, Illinois Informatics Institute

University of Illinois at Urbana Champaign

Contact: jdiesner@illinois.edu, <http://people.lis.illinois.edu/~jdiesner/>

Paper presented at Workshop of Information in Networks (WIN) 2013, NYU

Abstract

Documentaries are meant to tell a story, i.e. to create memory, imagination and sharing (Rose, 2012). More specifically, the goal with social justice documentaries is to motivate change in people's knowledge and/ or behavior (Barrett & Leddy, 2008). How can we know if a production has achieved these goals? And how early in the life cycle of a film project can we answer this question? These research questions are of high practical relevance: Given the scarce resources for financing documentaries, major funders such as the Sundance Documentary Fund, the Ford Foundation and BritDoc have a strong need for reliable, comprehensive and efficient assessments of the return of their investment, where their target function with these investments is to cause change on a group, organizational or societal level (Clark & Abrash, 2011; KnightFoundation, 2011). Additionally, filmmakers and (impact) producers are interested in tracking the effectiveness of the outreach and campaign work around a production from the earliest stages of a film project on. They can leverage this information to strategically allocate resources and tapping into existing social capital and awareness for an issue. We report on our research on developing, applying and evaluating a theoretically-grounded, computational solution for the practical assessment of the impact of documentaries in a scalable, empirical and systematic fashion. We approach this problem from a network analytical perspective: based on the assumption that documentaries are produced, screened and watched as part of larger and continuously changing ecosystems that involve multiple stakeholders and the (potential) flow of information between them, we track, map, fuse and analyze social and semantic networks that represent this information; using data from media, social media and focus group interviews. More specifically, we leverage techniques from social network analysis and natural language processing to detect and assess the structure, functioning and dynamics of the web of social agents and information associated with (the main issue of) a movie.

Introduction and Background

Prior work on assessing the impact of documentaries is limited in scope, depth and practical implementations (Barrett & Leddy, 2008; Figueroa, 2002): major media institutes have proposed systematic frameworks, which are mainly of theoretical and normative nature (Barrett & Leddy, 2008; Clark & Abrash, 2011; Figueroa, 2002; KnightFoundation, 2011). Some frameworks include network related indicators, but fail to implement and measure them. Scholarly work on this topic is mainly confined to studies of psychological effects of films on individuals, and conceptualizes documentaries as a subcategory of mass media. Overall, evaluation in this domain has typically been done by using (a) traditional, scalable and quantitative methods and metrics, such as the number of

visitors of a screening or webpage, or (b) conventional, qualitative and small-scale methods for in-depth analysis of the perception of a topic or product by small numbers of people, such as focus group interviews. We integrate these two levels by jointly considering (a) the social network of stakeholders involved with the main topic of a movie – regardless of whether they have anything to do with a particular film or not - and (b) the substance of the information produced and shared by these agents. The resulting socio-semantic networks allow for reasoning about two types of behavioral information - social relationships and information (Barthelemy, Chow, & Eliassi-Rad, 2005; Diesner, 2013; Diesner & Carley, 2005; Gloor & Zhao, 2006; Mika, Elfring, & Groenewegen, 2006; Roth & Cointet, 2010).

In close collaboration with the JustFilms division of the Ford Foundation, we have been applying the developed solution and technology to multiple social justice documentaries, e.g. about the US prison system, drug use and education of minorities. We recently presented our findings at the 2013 Sundance Creative Producing Summit, where we got valuable feedback from practitioners, which we are currently incorporating into our methodology and tool.

Theoretical Basis and Methodology

Our solution is based on a theoretical framework that we developed by synthesizing indicators of impact based on empirically tested theories from the fields of media effects, diffusion research, social and semantic network analysis, and collective action. The resulting CoMTI (content, medium, target, and impact) framework also incorporates indicators specific to documentary evaluation that we identified in discussions with subject matter experts. Overall, the CoMTI framework considers a variety of stimuli that have been associated with cognitive, attitudinal, and behavioral change on the individual, communal and societal over time.

Our methodology involves three steps, which are explained below in more detail: building and analyzing a baseline model, a ground truth model, and a change model, which ultimately get compared to each other. We start with mapping the discourse around the main issue(s) addressed in a movie prior to a film's initial public screenings or release. This results in a baseline model. The main purpose with this step is understand the given ecosystem of people and themes associated with the main issue of a film, and thereby helps to identify where impact is possible. The main issues of a film can be identified in a data driven way, e.g. by conducting topic modeling on the film transcript, and/or be elicited from the film maker, producer or funder. Based on our initial experience, the outcomes from both strategies do not necessarily align. We decided to go with the issues identified by subject matter experts on the film since these are the topics on which they want to motivate some change. Once these issues are identified, we use ConText (<http://context.lis.illinois.edu/>), a tool we built for this project, to a) collect social media data from Facebook and Twitter and b) processing media coverage that we collect from external sources, e.g. LexisNexis, as well as interview data that were collected by external project partners from focus groups. The same tool is then used to construct a) social networks of agents being mentioned in the bodies of unstructured, natural language text data, in structured meta-data and as account holders, and b) semantic networks of salient terms, themes and sentiments that explicitly or implicitly occur in the text data and meta-data. The respective outcomes and same technology are then used to conduct a) social network analysis – mainly to

identify (clusters of) key agents and organizations and b) text mining to identify main themes and trends in the discourse. Practitioners from the film domain can utilize this step to understand the given opportunity space for tying their campaign work to relevant stakeholders and themes; helping them to strategically allocate resources and tapping into existing social capital and public awareness.

Second, we extract the message of a documentary; mainly in order to understand the issues related to which a film *can* cause change. For this purpose, we apply the same text mining techniques as for step one, but to the film transcript. This results in a ground truth model, i.e. the message that a documentary can communicate. We understand that there is much more to a film than what is in the transcript, namely the cast, images, sound and other aesthetic elements, which are not yet considered with our methodology.

Third, we measure whether the film has moved the needle on an issue, i.e. trying to capture the measurable impact of a film. For this purpose, we reassess the key players and public discourse related to, i.e. co-mentioned with, the film since its release. For this step, the same types of data and analysis techniques as in the first step are being collected and used, respectively. This creates a model of the reality of change. We compare this model along the dimensions laid out in the CoMTI framework and via network comparison methods to a) the baseline model; looking for new (links between) stakeholders and themes that co-occur with the movie, and b) the ground- truth model; with the delta indicating the difference in the perception of movie by its makers versus its audience. Step three can be repeated several times throughout the life-cycle of a film, potentially detecting shifts from awareness (short term, right around (theatrical) release) to reputation (during period of public screenings, including around times for major film awards) to legacy (after main screening period).

To facilitate this process, we have been developing ConText, a publicly available tool that supports (a) the construction of network data from structured and unstructured natural language text data, a process also known as relation extraction (Diesner, 2013), and (b) the joint analysis of text data and network data; with the latter one being mainly realized by adapting Gephi (<https://gephi.org/>). The relation extraction routines are particularly crucial for integrating text analysis and network analysis, and entail a combination of entity extraction and co-occurrence based link formation. We built a graphical user interface for ConText and developed training material so that non-technical people, e.g. film makers, impact producers and funders can conduct this type of assessment on their own. However, the application contexts do not stop here: ConText is also designed to be of general applicability for conducting text mining and network analysis of data from other domains, with the limitation that the evaluation criteria for impact from the CoMTI framework might not apply.

Conclusions, Limitations and Next Steps

In summary, we are presenting a novel, theoretically grounded and computational solution for evaluating the impact of (social justice) documentaries by mapping and analyzing the web of stakeholders and information related to (the main issues of) a film in a systematic, empirical and scalable fashion. More specifically, we leverage techniques from social network analysis and natural language processing (NLP) to detect and assess the structure, functioning and dynamics of socio-technical networks. Combining network analysis and text mining applied to (social) media data and

interview data overcomes a main shortcomings of the aforementioned, prior approaches for impact assessment of films; namely a) the scalable integration of quantitative and qualitative, multimodal data and b) the implementation and empirical measurement of metrics indicative of impact. In our talk, we will provide illustrative examples of evaluations of documentaries that we have conducted so far.

With this work, we are bringing text mining and network analysis applied to the content of information disseminated by network participants to a domain where these methods have not yet been used for evaluation. Throughout the interaction and cooperation with practitioners and project partners from that domain, we have received valuable feedback that informs us about additional practical needs for additional capabilities of the developed methodology and tool. For example, we recently presented an assessment of the “The House I Live In”, a 2012 film by Eugene Jarecki, at the 2013 Sundance Creative Producing Summit. For this film, we found out that in classic media and on Twitter, the movie had not been successful in attracting attention related the substance matter of the film, namely mandatory minimum sentencing, but was discussed as an art product. A person involved with consulting on behalf of the makers of this film informed us that they often try to frame a movie as an art product first, and as a vehicle for communicating information about some issue second.

Another limitation that we plan to fix in the near future is our focus on types or targets of change: Currently, we only consider discourse and awareness as reflected in data from news coverage and social media as well as interviews with focus groups. This focus is inappropriate when a film aims to change political or corporate affairs. We plan to expand our framework, data sources and technology to cover these dimensions as well; e.g. by monitoring public databases of legal information and corporate reports.

Acknowledgement

This work is supported by the FORD Foundation, grant 0125-6162. We are also grateful to feedback and advice from Dr. Susie Pak from St. John’s University, Orlando Bagwell, former director of JustFilms at the Ford Foundation, and Joaquin Alvarado, Chief Strategy Officer of the Center for Investigative Reporting. We also thank graduate students Sean Wilner from the Illinois Informatics Institute at UIUC and Shameem Ahmed from the iSchool at UIUC for their contributions to ConText, and Sebastian Benthall from the iSchool at Berkeley for his feedback.

References

- Barrett, D., & Leddy, S. (2008). Assessing Creative Media's Social Impact. *The Fledgling Fund*.
- Barthelemy, M., Chow, E., & Eliassi-Rad, T. (2005). *Knowledge Representation. Issues in Semantic Graphs for Relationship Detection*. Paper presented at the AAAI Spring Symposium on AI Technologies for Homeland Security, Stanford, CA.
- Clark, J., & Abrash, B. (2011). Social justice documentary: Designing for impact: Center for Social Media.
- Diesner, J. (2013). From Texts to Networks: Detecting and Managing the Impact of Methodological Choices for Extracting Network Data from Text Data. *Künstliche Intelligenz/ Artificial Intelligence*, 27(1), 75-78. doi: 10.1007/s13218-012-0225-0

- Diesner, J., & Carley, K. M. (2005). Revealing social structure from texts: Meta-Matrix text analysis as a novel method for network text analysis. In V. K. Narayanan & D. J. Armstrong (Eds.), *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations* (pp. 81-108). Harrisburg, PA: Idea Group Publishing.
- Figueroa, M. E. (2002). *Communication for social change: An integrated model for measuring the process and its outcomes*: Rockefeller Foundation.
- Gloor, P., & Zhao, Y. (2006, July 2006). *Analyzing actors and their discussion topics by semantic social network analysis*. Paper presented at the 10th IEEE International Conference on Information Visualisation London, UK.
- KnightFoundation. (2011). *Impact: A Guide to Evaluating Community Information Projects*.
- Mika, P., Elfring, T., & Groenewegen, P. (2006). Application of semantic technology for social network analysis in the sciences. *Scientometrics*, *68*(1), 3-27.
- Rose, F. (2012). *The Art of Immersion: How the digital generation is remaking Hollywood, Madison Avenue, and the way we tell stories*: WW Norton.
- Roth, C., & Cointet, J. (2010). Social and semantic coevolution in knowledge networks. *Social Networks*, *32*(1), 16-29.