Editorial

New Formats, New Methods: Computational Approaches as a Way Forward for Media Entertainment Research

Johannes Breuer [1,*], Tim Wulf [2] and M. Rohangis Mohseni [3]

[1] Data Archive for the Social Sciences, GESIS—Leibniz Institute for the Social Sciences, 50667 Cologne, Germany;
E-Mail: johannes.breuer@gesis.org
[2] Department of Media and Communication, LMU Munich, 80538 Munich, Germany; E-Mail: tim.wulf@ifkw.lmu.de
[3] Institute of Media and Communication Science, Ilmenau University of Technology, 98693 Ilmenau, Germany;
E-Mail: rohangis.mohseni@tu-ilmenau.de

* Corresponding author

Abstract
The rise of new technologies and platforms, such as mobile devices and streaming services, has substantially changed the media entertainment landscape and continues to do so. Since its subject of study is changing constantly and rapidly, research on media entertainment has to be quick to adapt. This need to quickly react and adapt not only relates to the questions researchers need to ask but also to the methods they need to employ to answer those questions. Over the last few years, the field of computational social science has been developing and using methods for the collection and analysis of data that can be used to study the use, content, and effects of entertainment media. These methods provide ample opportunities for this area of research and can help in overcoming some of the limitations of self-report data and manual content analyses that most of the research on media entertainment is based on. However, they also have their own set of challenges that researchers need to be aware of and address to make (full) use of them. This thematic issue brings together studies employing computational methods to investigate different types and facets of media entertainment. These studies cover a wide range of entertainment media, data types, and analysis methods, and clearly highlight the potential of computational approaches to media entertainment research. At the same time, the articles also include a critical perspective, openly discuss the challenges and limitations of computational methods, and provide useful suggestions for moving this nascent field forward.

With the rapid development of technology and the growing competition for the attention (and money) of the audience, the entertainment media landscape is constantly changing. The global spread of broadband internet, mobile devices, streaming platforms, and communication tools with which people can, for example, discuss entertainment content, have had an immense impact on the structure and use of entertainment media. These fundamental changes in the media entertainment landscape not only affect the everyday lives of people worldwide, but also create opportunities and challenges for research that looks at the use and effects of these media.

New media entertainment formats entail new research questions about, for example, the motivations and experiences of their users, and the effects that the use of these new formats can have on them. Answering

these questions may require revisions to existing theories or even completely novel theoretical approaches. In addition to that, studying new media entertainment formats may also necessitate the development of new methods for collecting and analyzing data. Besides the development or refinement of theories and methods, another important aspect associated with the emergence of new (digital) media formats and platforms is the huge amount of data that their usage generates, which is both a challenge and an opportunity for entertainment research.

For several decades now, most quantitative research on the content, use, and effects of media entertainment has been based on data from surveys, manual content analyses, or lab experiments. While there is no doubt that these studies have produced many important insights into media entertainment, the data they are based on have certain limitations. For example, several recent studies have shown that self-reports of media use tend to be unreliable (e.g., Araujo, Wonneberger, Neijens, & de Vreese, 2017; Scharkow, 2016). This is especially problematic if researchers are interested in very specific, rare, or socially undesirable forms of media entertainment. Experimental lab studies, on the other hand, tend to have relatively small samples and often occur in somewhat unnatural settings. Moreover, manual content analyses are not suitable for the large amounts of data that users of media entertainment generate (e.g., discussion threads on Reddit or tweets about a show, movie, or video game).

Parallel to the largely technology-driven developments in the entertainment landscape, the methodological portfolio of social-scientific research has also been substantially extended by the rise of computational social science which "leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale" (Lazer et al., 2009, p. 722). According to Hox (2017), two key identifying features of computational methodology are the use of "big data" (although the term is often defined differently and tends to be underdefined) and the use of analysis techniques that are suited for these kinds of data. These analysis methods typically belong to the areas of text mining and natural language processing, machine learning, and network analysis. Regarding the type of data used, especially for computational communication research, it is typically more precise to speak of digital trace data which can be roughly defined as "records of activity (trace data) undertaken through an online information system" (Howison, Wiggins, & Crowston, 2011) and can originate from various sources, including social media platforms, websites, or smartphone apps. These traces can be intentional, such as tweets or Reddit comments, or unintentional, such as information about users or their activity collected by a streaming platform (Hox, 2017). Given their expertise in analyzing the use, content, and effects of digital media, "communication scholars are in a uniquely strategic position to lead the development of the computational approaches that promise to offer novel and exciting insights" (Hilbert et al., 2019, p. 3932). Indeed, computational communication science is a distinct subdiscipline "that investigates the use of computational algorithms to gather and analyze big and often semi- or unstructured data sets to develop and test communication science theories" (Van Atteveldt, Margolin, Shen, Trilling, & Weber, 2019, p. 1; also see Domahidi, Yang, Niemann-Lenz, & Reinecke, 2019; Van Atteveldt & Peng, 2018). Computational communication research has seen a rapid growth over the last few years. One clear indicator of this is that the former interest group Computational Methods has become a full division of the International Communication Association in 2020. While most studies in this area have looked at topics related to information seeking, news consumption, or political communication, there has been relatively little research on entertainment media. This thematic issue seeks to address this gap.

The characteristics identified by Hox (2017) also apply to the articles included in this thematic issue: They use (big) digital trace data and advanced analysis methods to study various phenomena related to the use of different kinds of entertainment media. In addition, they combine different analysis methods and types of data, which is also typical of computational communication research (and computational social science in general). To illustrate the diversity of topics and approaches, we provide an overview of the media and data types as well as the analysis methods in Table 1. Interestingly, there is a striking overlap in the types of data and analysis methods that almost all large entertainment companies nowadays use to evaluate and improve their products (as well as to profile and better target users) and the new computational methods that researchers have started to use for entertainment research. This further highlights the practical relevance of computational approaches in entertainment research.

A key challenge for computational entertainment and communication research, and even computational social science in general, is the question of how to access digital trace data and what can be done with them. Researchers not only have to consider the (privacy) interests of the people whose data they collect and use, but also those of commercial companies as typically specified in their Terms of Service (Van Atteveldt, Strycharz, Trilling, & Welbers, 2019). Especially when it comes to the ideals of open science, the interests of the researchers who use the data and the commercial companies who control it can be conflicting (Breuer, Bishop, & Kinder-Kurlanda, in press). Against this background, we are especially excited that for several of the articles included in this thematic issue, the authors were able to make their analysis code and data available (see Table 1). Of course, using digital trace data also has other limitations and potential pitfalls. These include the common lack of individual level information about the users and relevant outcome variables (Stier, Breuer, Siegers, & Thorson, 2019) or potential biases (Sen, Flöck, Weller,

Table 1. Overview of the articles in this thematic issue, including focus, data types, and analysis methods.

Article title	Author(s)	Entertainment media focus	Data types	Analysis method(s)	Open materials
What Is Important When We Evaluate Movies? Insights from Computational Analysis of Online Reviews	Schneider, Domahidi, and Dietrich	Movies	Online movie reviews, self-reports from online surveys	Correlated topic models, qualitative content analysis	https://osf.io/pqnk6
A Graph-Learning Approach for Detecting Moral Conflict in Movie Scripts	Hopp, Fisher, and Weber	Movies	Movie scripts	Social network analysis, natural language processing	Script: https://github.com/EdinburghNLP/scriptbase Data: https://osf.io/rbdws
(A)synchronous Communication about TV Series on Social Media: A Multi-Method Investigation of Reddit Discussions	Unkel and Kümpel	Series	Reddit threads, self-reports from online surveys	Automated content analysis	https://osf.io/7v49t
Popular Music as Entertainment Communication: How Perceived Semantic Expression Explains Liking of Previously Unknown Music	Lepa, Steffens, Herzog, and Egermann	Music	Self-reports from surveys/experiments, audio data (music)/music information retrieval data	Regression, factor analysis, machine learning, algorithmic audio signal analysis	
A Computational Approach to Analyzing the Twitter Debate on Gaming Disorder	Schatto-Eckrodt, Janzik, Reer, Boberg, and Quandt	Video games	Tweets	Sentiment analysis, network analysis, topic models	https://osf.io/vzymj
Exploring the Effect of In-Game Purchases on Mobile Game Use with Smartphone Trace Data	Boghe, Herrewijn, De Grove, Van Gaeveren, and De Marez	Video games	Smartphone log data	Survival analysis	
Open-Source's Inspirations for Computational Social Science: Lessons from a Failed Analysis	Poor	Not applicable	Not applicable	Not applicable	

Weiss, & Wagner, 2019). It is reassuring and promising for the future of this young field to see that all contributions in this thematic issue are aware of these issues and explicitly address them.

The study by Schneider, Domahidi, and Dietrich (2020) compares insights from self-report measures with online movie reviews to capture how viewers evaluate movies. They used subjective movie evaluation criteria (SMEC), identified based on self-report data from online surveys, and related those to a correlated topic model that explores the underlying topics of openly available user reviews. The study found correspondences for three major SMEC categories (hedonism, narrative, and actors' performance) in the online reviews, with additional qualitative analyses revealing further occurrence of SMEC categories in the review texts.

The study by Hopp, Fisher, and Weber (2020) also looks at movies. Using a combination of social network analysis and natural language processing techniques, they were able to develop a method for detecting moral conflict in scripts of more than 80,000 movie scenes. Among other things, they found that moral conflict can be identified by changes in the structures of social networks of movie characters.

Unkel and Kümpel (2020) also used a combination of computational and traditional methodological approaches to study synchronous and asynchronous communication about a TV series on Reddit. Specifically, they examined the motives of using Reddit forums for communication before, while, and after watching new episodes of the final season of *Game of Thrones*. Combining automated content analyses of these threads with a survey among thread users, they found that different motives lead to using these thread types, and different thread types are associated with different forms of interactions.

The contribution by Lepa, Steffens, Herzog, and Egermann (2020) employed a set of computational and other methods to study popular music as entertainment communication. Using an existing dataset, they developed a model for predicting listener liking ratings for previously unknown songs and found that unknown music is liked more, the more it is perceived as emotionally and semantically expressive. In a second study, the authors developed and tested a machine learning model drawing on automatic audio signal analysis and found that it can predict significant proportions of variance in musical meaning decoding.

Schatto-Eckrodt, Janzik, Reer, Boberg, and Quandt (2020) made use of computational approaches for analyzing the debate about gaming disorder on Twitter around the time in 2018 when the World Health Organization (WHO) decided to include the addictive use of digital games (gaming disorder) as a diagnosis in the International Classification of Diseases. The authors used a combination of sentiment, network, and automated content analysis (topic models), and found that the debate was largely organic (i.e., not driven by spam accounts) and heavily impacted by the WHO decision.

The article by Boghe, Herrewijn, De Grove, Van Gaeveren, and De Marez (2020) also looks at digital games, although with a very different research question and methodological approach. They used smartphone data to explore the effect of in-game purchases on continual mobile game use. In a survival analysis with the log data, they found that, while making an in-game purchase initially decreases the risk of stopping to play a game, there is a reversal effect in the sense that previous in-game purchases negatively affect the chance of continued play at a later point in time.

Unlike the other articles, the final contribution to this thematic issue by Poor (2020) does not present empirical results but offers a critical meta-perspective on computational approaches to media entertainment research. Building on his own experiences, the author discusses how and why computational research can fail and what the young field of computational social science can learn from the long history of the open source (software) movement.

Overall, the articles in this thematic issue cover different topics and employ different (methodological) approaches to study media entertainment. Despite their differences, they all show the potential of computational approaches for media entertainment research, while at the same time also highlighting some of the challenges and potential limitations. What all of the articles clearly illustrate is that combinations of different methods (including computational as well as more traditional approaches) and data types (including digital trace data as well as other types, such as self-reports) represent a promising way of moving entertainment research forward. Hence, we believe that with this thematic issue we offer researchers in the field of (entertainment) communication a diverse portfolio of applications of computational methods for various research questions. We hope that this work will inspire entertainment research and guide the way to a more nuanced triangulation and diversity of methods used in this research area.

Acknowledgments

Conflict of Interests

The authors declare no conflict of interests.

References

Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures*, *11*(3), 173–190. https://doi.org/10.1080/19312458.2017.1317337

Boghe, K., Herrewijn, L., De Grove, F., Van Gaeveren, K., & De Marez, L. (2020). Exploring the effect of in-game purchases on mobile game use with smartphone trace data. *Media and Communication*, *8*(3), 219–230.

Breuer, J., Bishop, L., & Kinder-Kurlanda, K. (in press). The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships. *New Media & Society*.

Domahidi, E., Yang, J., Niemann-Lenz, J., & Reinecke, L. (2019). Outlining the way ahead in computational communication science: An introduction to the IJoC Special Section on "computational methods for communication science: Toward a strategic roadmap." *International Journal of Communication*, *13*, 3876–3884.

Hilbert, M., Barnett, G., Blumenstock, J., Contractor, N., Diesner, J., Frey, S., . . . Zhu, J. J. H. (2019). Computational communication science: A methodological catalyzer for a maturing discipline. *International Journal of Communication*, *13*, 3912–3934.

Hopp, F. R., Fisher, J. T., & Weber, R. (2020). A graph-learning approach for detecting moral conflict in movie scripts. *Media and Communication*, *8*(3), 164–179.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, *12*(12), 767–797. https://doi.org/10.17705/1jais.00282

Hox, J. J. (2017). Computational social science methodology, anyone? *Methodology*, *13*(Supp. 1), 3–12. https://doi.org/10.1027/1614-2241/a000127

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., . . . Van Alstyne, M. (2009). Life in the network: The coming age of computational science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Lepa, S., Steffens, J., Herzog, M., & Egermann, H. (2020). Popular music as entertainment communication: How perceived semantic expression explains liking of previously unknown music. *Media and Communication*, *8*(3), 191–204.

Poor, N. (2020). Open-source's inspirations for computational social science: Lessons from a failed analysis. *Media and Communication*, *8*(3), 231–238.

Scharkow, M. (2016). The accuracy of self-reported internet use: A validation study using client log data. *Communication Methods and Measures*, *10*(1), 13–27. https://doi.org/10.1080/19312458.2015.1118446

Schatto-Eckrodt, T., Janzik, R., Reer, F., Boberg, S., & Quandt, T. (2020). A computational approach to analyzing the Twitter debate on gaming disorder. *Media and Communication*, *8*(3), 205–218.

Schneider, F. M., Domahidi, E., & Dietrich, F. (2020). What is important when we evaluate movies? Insights from computational analysis of online reviews. *Media and Communication*, *8*(3), 153–163.

Sen, I., Flöck, F., Weller, K., Weiss, B., & Wagner, C. (2019). A total error framework for digital traces of humans. *arXiv.org*. Retrieved from https://arxiv.org/abs/1907.08228

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2019). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review.* Advance online publication. https://doi.org/10.1177/0894439319843669

Unkel, J., & Kümpel, A. S. (2020). (A)synchronous communication about tv series on social media: A multimethod investigation of Reddit discussions. *Media and Communication*, *8*(3), 180–190.

Van Atteveldt, W., Margolin, D., Shen, C., Trilling, D., & Weber, R. (2019). A roadmap for computational communication research. *Computational Communication Research*, *1*(1), 1–11. https://doi.org/10.5117/CCR2019.1.001.VANA

Van Atteveldt, W., & Peng, T.-Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, 12(2/3), 81–92. https://doi.org/10.1080/19312458.2018.1458084

Van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019). Towards open computational communication science: A practical roadmap for reusable data and code. *International Journal of Communication*, *13*, 3935–3954.

About the Authors

Johannes Breuer (PhD) is a Senior Researcher at the Data Archive for the Social Sciences at GESIS—Leibniz Institute for the Social Sciences in Cologne, Germany, where his work focuses on the topics of data linking and using digital trace data for social science research. His other research interests include the use and effects of digital media, computational methods, and open science. More information: https://www.johannesbreuer.com

Tim Wulf (PhD) is a Post-Doctoral Researcher at the Department of Media and Communication at LMU Munich in Germany. His research interests include the effects of media-induced nostalgia, media psychological perspectives on video games and video game streaming, and persuasion through narrative media. More information: https://www.tim-wulf.de

M. Rohangis Mohseni (Dr.rer.nat) is a Post-Doctoral Researcher of the Media Psychology and Media Design research group at Ilmenau University of Technology in Germany. His research interests include electronic media effects and moral behavior. His latest publications address gendered hate speech on YouTube (*SCM*, 2020), digital aggression (*merz*, 2020), and mobile learning (*Handbuch Bildungstechnologie*, 2020). More information: http://www.rmohseni.de