Information Metrics.
Empirical Methods of Information Science

Inagural-Disseration
for the degree of Doctor of Philosophy (Dr. phil.)

submitted to the Faculty of Philosophy
Heinrich Heine University Düsseldorf
Germany

by
Isabelle Dorsch
from Hilden, Germany

Langenfeld, November 2020

Department of Information Science
Institute of Linguistics and Information Science
Heinrich Heine University Düsseldorf
D61

1st Supervisor: Prof. Dr. Wolfgang G. Stock
2nd Supervisor: Prof Dr. Ágnes Veszelszki

Oral examination: December 15, 2020

Acknowledgements

First of all, I would like to express my special appreciation and thanks to my advisor Prof. Dr. Wolfgang G. Stock. Thank you so much for motivating and encouraging me, for your always open door, all the things you taught me, and your overall guidance not only through this dissertation but rather since the very beginning of my study program. Likewise, I would like to express my special appreciation and thanks to my mentor and second advisor Prof Dr. Ágnes Veszelszki for your support and feedback.

Secondly, I would like to thank my co-authors for successfully working together on our projects and also the nice time we had. Furthermore, many thanks to my colleagues involved in any way or who have accompanied me on my way. Regardless of whether if it was a single inspirational talk or if you were one of my deep and close fellows of this journey.

A special thanks to my parents Marion and Manfred, Jörg, my family, my friends (especially to you Nadine for being so fabulous supportive), and everyone else who has supported me or was part of my journey.

I am very grateful to all of you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Information science as a young and highly interdisciplinary science comprises content, users, and systems (Stock & Stock, 2013). Jason Farradane proposed and coined the term "information science" in 1955 first (Shapiro, 1995). As its domain and scope were and are in steady development, over time serval definitions were proposed or updated (Saracevic, 2009; Stock & Stock, 2013). For example, Borko (1968) defines it as

> "[...] an interdisciplinary science that investigates the properties and behavior of information, the forces that govern the flow and use of information, and the techniques, both manual and [mechanical], of processing information for optimal storage, retrieval, and dissemination" (Borko, 1968, p. 5).

Likewise, and in the context of the digital development Stock and Stock (2013) propose the following definition:

> "Information Science studies the representation, storage and supply as well as the search for and retrieval of relevant (predominantly digital) documents and knowledge (including the environment of information)" (Stock & Stock, 2013, p. 3).

Its central point is information itself. Information can be a process, where one informs or is informed, as a transporter for knowledge or as a fixed thing in informative objects like a textual or non-textual document (Buckland, 1991). The establishment of the discipline in the 1950s and after the Second World War was not coincidental. It emerged as an answer to the information explosion at that time, utilizing technology as a solution to handle the massive amount of information. With the upcoming of more and more (information) technologies nowadays, information science deals with the same problems and also new aspects in the digital age (Saracevic, 2009). Thereby, the field always had a practical and a theoretical component (Borko, 1968). Likewise, it is often mentioned its two orientations, the one system-oriented, dealing with information retrieval techniques and systems and the other one user-oriented, dealing with information needs and users, i.e. with human information behavior (Saracevic, 2009). Nevertheless, it is apparent that these two branches cannot be strictly separate (also they were considered as such), but rather are strongly interchanging and go inherent with the content, the information itself. Stock and Stock (2013) propose a distinction between (1.) theoretical information science, (2.) empirical information science, and (3.) applied information science. More generalized, this covers (1.) fundaments and theories, for example, in information retrieval and knowledge representation, (2.) the systematic empirical study of information systems and users in informetrics, and (3.) the use of information practice for example in everyday life as information literacy or in information markets (Stock & Stock, 2013). In terms of the subjects' interdisciplinary and strong connections to other fields, information science "is by no means a mixture of other sciences, but a science on its own right" (Stock & Stock, 2013, p. 7), utilizing a variety of own and adapted concepts, models and methodologies.

Metrics studies always were a relevant area for information science. "What are the features and laws of the recorded information universe? While often connected with [systems], the emphasis in this area of information science is on information objects or artifacts rather than systems; these are the content of the systems. It is about [characterizing] content objects" (Saracevic, 2009, p. 11). Of course, metrics studies are not limited to the field of information science. Among others, there also exist biometrics, econometrics, or software metrics (Rousseau, Egghe, & Guns, 2018). Over time, several broad and narrow sub-fields dedicated to metrics in information science emerged and put in historical order and topical relation (Björneborn & Ingwersen, 2004; Rousseau et al., 2018; Stock & Weber, 2006). As Rousseau et al. (2018) point out, one of the first sub-field was bibliometrics, independently assigned from each other by Otlet and Pritchard. In 1934, Otlet designated the measurement of all aspects related to books and documents with the term bibliometrics or rather "la bibliométrie" since his proposal is in French (Otlet, 1934; Rousseau et al., 2018). In 1969, similarly, but unaware of Otlet's definition, Pritchard defined bibliometrics as "the application of mathematics and statistical methods to books and other media of communication" (Pritchard, 1969, p. 349) and in order to replace the term "statistical bibliometrics." Furthermore, Rousseau et al. (2018) state that Ranganathan coined out the term librametrics (librametry) that did not establish outside India. What followed is the emergence of various sub-fields, as can be obtained from the informetrics frameworks in Figure 1.1 and Figure 1.2.

Above all, informetrics as broadest discipline, and thus, includes all other subfields (Björneborn & Ingwersen, 2004). The tern "informetrics" (in German "Informetrie") was proposed by Blackert and Siegel (1979) as well as by Nacke (1979) for the first time. "Informetrics is the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists" (Tague-Sutcliffe, 1992, p. 1). Informetrics and bibliometrics were often used interchanged, however, according to the definitions bibliometrics is a sub-field of informetrics. Rousseau et al. shorten the informetrics definition to the following: "The study of the quantitative aspects of information in any form and in any social group" (Rousseau et al., 2018, p. 3).

> "Informetrics therefore includes all quantitative studies in information science. When

a researcher performs scientific investigations empirically, concerning for instance the behavior of information users, the scientific impact of academic journals, the development of a company's patent application activity, Web pages' links, the temporal distribution of blog posts discussing a given topic, the availability, recall and precision of retrieval systems, the usability of Web sites, etc., he is contributing to informetrics" (Stock & Stock, 2013, p. 445).

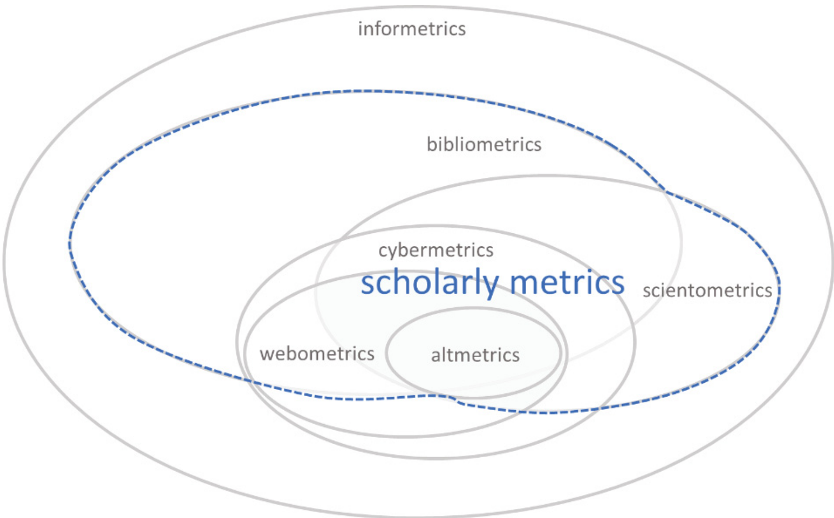


Figure 1.1: Informetrics and main sub-fields re-created from: Haustein (2016, p. 416) and adapted from Björneborn and Ingwersen (2004, p. 1217), "Sizes of the ellipses are not representative of field size but made for the sake of clarity only" (Haustein, 2016, p. 416).

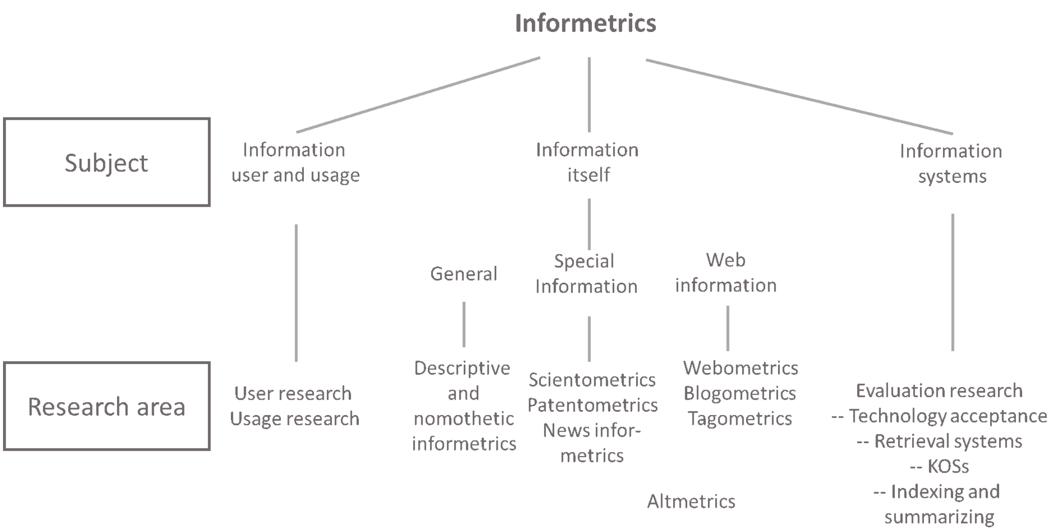


Figure 1.2: Informetrics subjects and research areas re-created from Stock & Stock (2013, p. 446).

The technical improvements within our digital age enable even more and more diversified, complex, and richer informetric analysis. At the same time, the development of new or the adjustment of existing methods faces new challenges. Information science (and thus also informetrics) is a field in a "constant flux" (Saracevic, 2009, p. 14) and new paths of research establish, like with new media like social media.

1.1 Social Media Research

With the emergence of social media and its large number of measurable items, it also allows for informetric analyses. Nowadays, social media is part of our everyday life. People all over the world are actively engaging with several applications daily and thus generating a vast amount of user-generated content. However, what is social media actually? Social media as a term has assigned multiple meanings, but relatively few formal definitions exist. It is not always clear what tools, platforms, and social phenomena belong to it (McCay-Peet & Quan-Haase, 2017). It is also argued that the term "social media" is misleading, as all media foster communication and are social (Papacharissi, 2015). Papacharissi (2015, p. 1) furthermore suggest:

> "Our understanding of social media is temporally, spatially, and technologically sensitive—informed but not restricted by the definitions, practices, and materialities of a single time period or locale. How we have defined social media in societies has changed, and will continue to change. Our use of the term social media is aimed at embracing the social character of media as it presents itself in media past, present, and future."

According to McCay-Peet and Quan-Haase (2017), social media studies often imply a definition by the websites and applications selected for investigation, rather than to state a formal definition. Moreover, there exist definitions of social media based on their type and field of application (e.g. social networking sites, bookmarking services, microblogging systems, etc.). However, although it is a constantly changing discipline a formal definition is possible and important. Based on six definitions for social media and the identification of common aspects, the following general and timely definition for social media is given by McCay-Peet and Quan-Haase (2017, p. 18):

> "Social media are web-based services that allow individuals, communities, and organizations to [collaborate,] connect, interact, and build community by enabling them to create, co-create, modifies, share, and engage with user-generated content that is easily accessible."

The vast amounts of data, also called "big data," address researchers from many disciplines to conduct research related to social media (Kitchin, 2014). The aspect of social media data creation, data collection, usage, storing, and analysis is thus a major topic for science and constantly offers new challenges for the quantitative but also qualitative social media analysis (McCay-Peet & Quan-Haase, 2017; Quan-Haase & Sloan, 2017). Like the general analysis of digital technology, the study of social media as part of it became a topic of interest in information science. From the perspective of content, users, and systems, social media provides a variety of units of assessment for informetric studies. And as always, with new research areas, this brings new opportunities and challenges. (Veszelszki, 2017, p. 12 f.) mentions the 2006 Time Magazine "Person of the Year," which was "you;" i.e. every internet user applying social media creating digital content:

> "[...] the person of the year in 2006 was the everyday internet user who achieved an unprecedental level of community and cooperation between people; the citizens of the 'new digital democracy' who had created things likethe 'cosmic knowledge base' of Wikipedia, the multu-million channel video sharing webstite YouTube, the (in the meantime outdated) online metropolis MySpace, or the social media site Facebook."

1.2 Scientometrics Research

> "Scientometric research is devoted to quantitative studies of science and technology. It aims at the advancement of knowledge on the development of science and technology, also in relation to societal and to policy questions. A special, but certainly not exclusive emphasis is placed on the role of quantitative, in particular bibliometric (i.e., based on data from scientific and technological literature) methods" (Van Raan, 1997, p. 205).

Already in 1969, Nalimov and Mulchenko (1969) coined out the term "scientometrics" ("naukometria" in Russian). As the name suggests, it is mainly used for the study of scientific literature (Hood & Wilson, 2001). Methodologies and indicators to assess productivity (based on publications) and impact (based on citations) are core topics within this sub-discipline. Although colleagues in the field use the term "scientometrics" similar to "bibliometrics" (Rousseau et al., 2018), they are not synonymous. As scientometrics also aims at the advances of knowledge on the development of science and technology, they may also be accompanied by qualitative approaches.

As scientific publications and citations are the subjects of scientometrics analyses, the question arises: what is a scientific or maybe better (because broad defined) a research publication? Furthermore, what is a citation? Research publications could be "bookstore media" (e.g. books, journals), invention documents (e.g. patents), grey literature (e.g. theses, reports, working papers), audiovisual media (e.g. broadcasting, scientific films), and internet documents (analog to articles, reports or books on the World Wide Web) (Stock, 2001) if they are formally published (Stock & Stock, 2013) in scientific, academic, or scholarly publishing sources (Dorsch, Askeridis, & Stock, 2018). More distinguished by document type a research publication can be a monograph, monograph as an edited work with several book sections, article in a peer-reviewed journal, article in the in-house journal of the own institute, habilitation thesis, dissertation, review, patent, note in a journal ("letter to the editor"), article in a daily newspaper, scientific film (Stock, 2001); conference proceeding, conference proceedings article and a scientific editorial. This list is quite long and in practice not all above might be included within a scientometrics analysis, although it would be part of the total research output of a scientist. All those documents form units of assessment. That can be on the micro (individuals), meso (departments, research groups), or macro level (countries, regions, universities) (Rousseau et al., 2018). However, the inclusion always depends on the respective evaluation criteria a scientometrician chose and is also determined by far more variables as the sole research publication type. Citations can be defined as follows:

> [C]itations are references to another textual element, from the perspective of the citing article. In order to have citations, there must be a cited-citing pair. From a formal perspective, cited-citing pairs are relations. By adding a dynamic perspective, these relations can be considered as relational operations" (Leydesdorff, 1998, p. 8).

Authors can cite a work in different ways, referring to the oeuvre (the complete works of an author), motif, opus (a single work like an authors' article), chunk (specific parts), or quantum (formula, phrase, chemical compound, method or result) (Cronin, 1994). In reference to the definition above, a citation is not limited to a textual element, but may also include, e.g., graphics, research models, research data, videos, etc..

Garfield (1955) introduced the *Science Citation Index* in 1955. It was the first citation index and led to various developments in the field. The idea was to provide a bibliographic tool that helps with the literature research. Further citation indices followed as the *Social Science Citation Index (SSCI)* or the *Arts & Humanities Citation Index*. This also had an impact on research. A new branch – research evaluation – opened up (Garfield, 1955; Mingers & Leydesdorff, 2015). Research evaluation is concerned with the measurement of productivity and impact of respective units of assessment in science. The allocation of resources, improvement of performance, enhancement of regional engagement, increase of visibility, stimulation of collaboration, promotion, or hiring are reasons for research evaluation (Rousseau et al., 2018). Derived numbers of scientific databases like *Web of Science (WoS)* or *Scopus* as well as "indicators based on recorded events of acts" like for example viewing, reading or citing "related to scholarly documents [...] or scholarly agents" are in short defined as "scholarly metrics" (Haustein, 2016, p. 416). They became main objects in research evaluation. The probably best-known indicators are the impact factor (Garfield, 1955, 1972, 2006) and the h-index (Hirsch, 2005). Nowadays, a sheer vast number of indicators and methods for research evaluation, and with them also controversial discussions and criticism within the scholarly community exist.

1.3 Information Metrics. Empirical Metrics of Information Science

This dissertation reports on content, users, and systems in empirical informetrics. It is thereby devoted to social media in informetrics as well as scientometrics. Based on this, the following research questions further frame this cumulative work:

RQ1: How does informetrics in social media research work? How can we measure information content of social media documents? How do social media users describe the content of their documents?

RQ2: How does scientometrics work? Is there a reliable data basis for scientometric studies? How is it possible to analyze research topics? How reliable are the data used for analysis in informetrics?

These two research questions serve as overall questions, where the presented studies are used to give an answer. Since informetrics investigates "all quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists" (Tague-Sutcliffe, 1992, p. 1) this works dedicates this to contrasting social media and scientometrics. Both branches can be analyzed in terms of their published content (e.g., social media postings, research publications) and likewise in terms of their impact (e.g., topics of interest, citations). Likewise, the same or at least similar informetric methodologies can be applied.

1.4 Methodology

Overall, a multi-method approach applies to the empirical informetric evaluation within this work. To answer the research questions, we worked with the following methodologies:

- Content analysis,
- Survey research,
- Research evaluation,
- Scientometric topic analysis.

The utilized content analysis (Krippendorff, 2004) for the studies in *Chapter 2, 4,* and *5* based on a direct (Hsieh & Shannon, 2005) also called deductive (Elo & Kyngäs, 2008) as well as conventional (Hsieh & Shannon, 2005) also called inductive (Elo & Kyngäs, 2008) approach. That means the developed codebook categories are derived from the literature (direct/deductive) but also from the content itself (conventional/inductive). Adopting the classification of micro, meso, and macro levels (Rousseau et al., 2018) for informetric social media research, the units of assessment could be classified into the meso level (*Chapter 2, 4, 5*) and the macro level (*Chapter 5*).

The two conducted online surveys (*Chapter 3 and 9*) based on nonprobability sampling (convenience/self-selected sample) (Silipigni Connaway & Radford, 2016). The questionnaire of the latter furthermore included one item being a multiple-choice knowledge test (Haladyna & Rodriguez, 2013). The units of assessment are on the micro (*Chapter 3*) and meso level (*Chapter 9*).

The studies in *Chapter 6* and *7* deal with quantitative research evaluation based on publication coverage obtained from scientific information services (Ingwersen, 2000; Miguel, Chinchilla-Rodriguez, & De Moya-Anegõn, 2011; Schlögl, 2013) and in contrast on personal publication lists (Gaillard, 1992; Hilbert et al., 2015; Krikwood, 2012) to propose a new empirical foundation in the field of scientometric indicators. Both analyses apply to information science researchers and are, therefore, on the micro level. *Chapter 8* provides a study analyzing scientometrically the publications' title terms according to a k-nearest neighbors topic clustering (Stock & Stock, 2013). Again, the foundation for the analysis are personal publication lists. This time, we studied two information science institutes, and therefore the research is on the meso level.

1.5 Topic Overview

The following paragraph introduces all following chapters with their respecting research questions and methods. The publications are equivalent to their published versions but slightly altered in this dissertation. The alterations include correction of typing mistakes or grammatical errors, formatting and design of text, tables or figures, and the unification of references to APA 6.

Part 1. Social Media in Informetrics is composed of four research studies. The study in *Chapter 2* analyzes how Instagram users tag their pictures regarding different kinds of pictures and hashtag categories. Picture categories considered are Food, Pets, Selfies, Friends, Activity, Art, Fashion, Quotes (captioned photos), Landscape, and Architecture. Hashtag categories distinguish between the content itself "Content-relatedness" (ofness, aboutness, and iconology), Emotiveness (referring to emotions), Isness (aspects referring to the metadata), Performativeness (aspects calling for an action), Fakeness (obvious false statements), "Insta"-Tags (everything related to the application), and Sentences as hashtag categories. Likewise, *Chapter 4* deals with the same topic, specialized in the analysis of gender-depended differences between women and men on Instagram.

Chapter 2: Dorsch, I. (2018). Content Description on a Mobile Image Sharing Service: Hashtags on Instagram. *Journal of Information Science in Theory and Practice, 6*(2), 46-61.

RQ1: Are there any differences in relative frequencies of hashtags in the picture categories?

RQ2: Given a picture category, what is the distribution of hashtag categories; and given a hashtag category, what is the distribution of picture categories?

RQ3: Is there any association between image categories and hashtag categories?

Chapter 4: Philipps, J., & Dorsch, I. (2019). Gender-specific Tagging of Images on Instagram. In G. Meiselwitz (Ed.), *Social Computing and Social Media. Design, Human Behavior and Analytics. HCII 2019* (pp. 396-413). Cham: Switzerland: Springer. (Lecture Notes in Computer Science, vol. 11578).

RQ1: Are there any gender-specific differences in the relative hashtag frequencies in the picture categories?

RQ2: Given a picture category, what is the gender-specific distribution of hashtag categories; and given a hashtag category, what is the gender-specific distribution of picture categories?

RQ3: Are there any gender-specific associations between picture categories and hashtag categories?

Since hashtags became an essential user-generated medium for social media communication, the exploratory study in *Chapter 3* faces the Instagram users' hashtagging creation behavior and selection process. It is distinguished between self-created hashtags, hashtags inspired, or generated through best practices or tools and the non-usage.

Chapter 3: Dorsch, I. (2020). Hashtags on Instagram: Self-created or Mediated by Best Practices and Tools? In *Proceedings of 53rd Hawaiian International Conference on System Sciences (HICSS53). January 7 – 10, 2020, Grand Wailea, Maui.* Honolulu, HI: HICSS (ScholarSpace).

RQ1: How many Instagrammers use hashtags and for what?

RQ2: How are best hashtagging practices and tools used on Instagram and to what extent do users create hashtags on their own?

RQ3: Do Instagram users intentionally assign false hashtags?

Likewise, concerned with the content production process of users in social live streaming services (SLSSs), the study in *Chapter 5* explores platform-, gender-, origin-, age-, and content-dependent differences of streamed content on Periscope, Ustream, and YouNow in respect to the streaming motivations fame or financial gain. Given different perspectives, social media content production and motivations are topics in *Part 1.* Instagram, Periscope, Ustream, and YouNow serve as a selection of social media platforms enabling the creation of multi visual content.

Chapter 5: Fietkiewicz, K. J., Dorsch, I., Scheibe, K., Zimmer, F., & Stock, W. G. (2018). Dreaming of Stardom and Money: Micro-celebrities and Influencers on Live Streaming Services. In G. Meiselwitz (Ed.), *Social Computing and Social Media. User Experience and Behavior. SCSM 2018* (pp. 240-253). Cham, Switzerland: Springer. (Lecture Notes in Computer Science; 10913).

RQ1: Which channels (Periscope, Ustream, YouNow) are preferred by users motivated by fame or financial gain?

RQ2: Are there gender-dependent differences regarding the streaming motivation being fame or financial gain?

RQ3: Are there origin-dependent differences (Germany, Japan, USA) regarding the streaming motivation being fame or financial gain?

RQ4: Are there age-dependent differences regarding the streaming motivation being fame or financial gain?

RQ5: What are the contents streamed by streamers whose motivation is fame or financial gain?

* * *

Following this, *Part 2 Scientometrics* composed of the following four research studies: The study in *Chapter 6* introduces a re-interpreted scholarly indicator "visibility." It is the share of the number of an author's publications on a certain information service relative to the author's entire oeuvre based upon his/her personal publication list. Based on the publication lists of the information scientists Blaise Cronin and Wolfgang G. Stock and their publication lists in scientific information services (ACM, ECONIS, Google Scholar, IEEE Xplore, Infodata eDepot, LISTA, Scopus, and Web of Science) as well as in the social media services Mendeley and ResearchGate the indicator is presented. *Chapter 7.* builds on this in connecting the visibility indicator to the new concepts of boundness for an author's scientific publication list. A list can be truebounded (exactly all publications), overbounded (more publications, not scientific/formally published), and underbounded (incomplete). The study generates and analyses personal publication lists and lists from information services (WoS, Scopus, Google Scholar) of nine International Society of Scientometrics and Informetrics (ISSI) Committee members.

Chapter 6: Dorsch, I. (2017). Relative Visibility of Authors Publications in Different Information Services. *Scientometrics, 112*(2), 917-925.

This article introduces a re-interpreted visibility indicator from a scientometrist's viewpoint.

Chapter 7: Dorsch, I., Askeridis, J., & Stock, W. G. (2018). Truebounded, Overbounded, or Underbounded? Scientists' Personal Publication Lists Versus Lists Generated Through Bibliographic Information Services. *Publications, 6*(1), 1-9.

RQ1: Are authors' personal publication lists, found on their personal sites on the Internet or oninstitutional repositories, truebounded, overbounded, or underbounded?

RQ2: Are the respective publication lists generated through bibliographic information services truebounded, overbounded, or underbounded?

Closer investigating the content, the study in *Chapter 8* focuses on a publication title term topic analysis at the Institute of Information Science and Information Systems at the University of Graz and the Department of Information Science at Heinrich Heine University in Düsseldorf.

Chapter 8: Dorsch, I., Schlögl, C., Stock, W. G., & Rauch, W. (2017). Forschungsthemen der Düsseldorfer und Grazer Informationswissenschaft (2010 bis 2016). *Information – Wissenschaft und Praxis, 68*(5-6), 320-328.

Finally, the study in *Chapter 9* addresses researchers' opinions on publication productivity, citation impact, and the h-index. Furthermore, their knowledge of the h-index is tested. In the context of scientific publication content, *Part 2* Scientometrics, therefore, addresses content, reliability of data, and opinion of the research "producers."

Chapter 9: Kamrani, P., Dorsch, I., & Stock, W. G. (2020). Publikationen, Zitationen und H-Index im Meinungsbild deutscher Universitätsprofessoren. *Beiträge zur Hochschulforschung, 42*(3), 78-98.

RQ1: How important are publications and citations to you?

RQ2: What importance do you attach to the visibility of our publications and your H-index in the respective information services?

RQ3: Do researchers know the H-index and its concrete formula?

RQ4: Are there any differences in the assessments and the level of knowledge regarding gender, subjects, and generations?

References

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, *55*(14), 1216–1227.

Blackert, L., & Siegel, S. (1979). Ist in der wissenschaftlich-technischen Information Platz für die Informetrie? *Wissenschaftliche Zeitschrift der TH Ilmenau*, *25*(6), 187–199.

Borko, H. (1968). Information science: What is it? *American Documentation*, *19*(1), 3–5.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, *42*(5), 351–360.

Cronin, B. (1994). Tiered citation and measures of document similarity. *Journal of the American Society for Information Science*, *45*(7), 537–538.

Dorsch, I., Askeridis, J., & Stock, W. G. (2018). Truebounded, overbounded, or underbounded? Scientists' personal publication lists versus lists generated through bibliographic information services. *Publications*, *6*(1), 1–9.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115.

Gaillard, J. (1992). Use of publication lists to study scientific production and strategies of scientists in developing countries. *Scientometrics*, *23*(1), 57–73.

Garfield, E. (1955). Citation indexes for science through association of ideas. *Science*, *122*(3159), 108–111.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*(4060), 471–479.

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, *295*(1), 90–93. Retrieved from `http://www.garfield.library.upenn.edu/papers/eval{\_}of{\_}science{\_}CBE(Utah).html`

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Routledge.

Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics*, *108*(1), 413–423.

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., … Stock, W. G. (2015). Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. *Online Information Review*, *39*(2), 255–264.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*(2), 291–314.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.

Ingwersen, P. (2000). The international visibility and citation impact of Scandinavian research articles in selected social science fields: The decay of a myth. *Scientometrics*, *49*(1), 39–61.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, *1*(1), 1–12.

Krikwood, P. E. (2012). Researcher publication checklist: A quantitative method to compare traditional databases to web search engines. In *119th american society for engineering education annual conference and exhibition.* San Antonio, TX.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks: Sage Publications Inc.

Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, *43*(1), 5–25.

McCay-Peet, L., & Quan-Haase, A. (2017). What is social media and what questions can social media research help us answer? *The SAGE Handbook of Social Media Research Methods*, 13–26.

Miguel, S., Chinchilla-Rodriguez, Z., & De Moya-Anegõn, F. (2011). Open access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the American Society for Information Science and Technology*, *62*(6), 1130–1145.

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, *256*(1), 1–19.

Nacke, O. (1979). Informetrie. Ein neuer Name für eine neue Disziplin. *Nachrichten für Dokumentation*, *30*(6), 219–226.

Nalimov, V. V., & Mulchenko, Z. M. (1969). *Scientometrics, the study of science as an information process [english translation. note: Original in russian].* Moscow: Nauka.

Otlet, P. (1934). *Traite de documentation, le livre sur le livre.* Brussels: D. Van Keerberghen & Sons.

Papacharissi, Z. (2015). We have always been social. *Social Media and Society*, *1*(1).

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, *25*(4), 348–349.

Quan-Haase, A., & Sloan, L. (2017). Introduction to the handbook of social media research methods: Goals, challenges and innovations. In *The sage handbook of social media research methods* (pp. 1–9).

Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise: A bibliometric guide for researchers.* Cambridge, MA: Chandos.

Saracevic, T. (2009). Information science. In M. J. Bates (Ed.), *Encyclopedia of library and information sciences* (3rd ed., pp. 2570–2585). NY: Taylor and Francis.

Schlögl, C. (2013). Internationale Sichtbarkeit der europäischen und insbesondere der deutschsprachigen Informationswissenschaft. *Information - Wissenschaft und Praxis*, *64*(1), 1–8.

Shapiro, F. R. (1995). Brief communication: Coinage of the term information science. *Journal of the American Society for Information Science*, *46*(5), 321–397.

Silipigni Connaway, L., & Radford, M. L. (2016). *Research methods in library and information science* (6th ed.). Santa Barbara, CA: ABC-CLIO.

Stock, W. G. (2001). *Publikation und Zitat. Die problematische Basis empirischer Wissenschaftsforschung.* Köln: Fachhochschule Köln; Fachbereich Bibliotheksund Informationswesen (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 29).

Stock, W. G., & Stock, M. (2013). *Handbook of information science*. Berlin, Boston: De Gruyter Saur.

Stock, W. G., & Weber, S. (2006). Facets of informetrics: Preface. *Information-Wissenschaft und Praxis*, *57*(8), 385–389.

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing and Management*, *28*(1), 1–3.

Van Raan, A. F. (1997). Scientometrics: State-of-the-art. *Scientometrics*, *38*(1), 205–218.

Veszelszki, Á. (2017). *Digilect. The impact of infocommunication technology on language*. Berlin, Boston, MA: De Gruyter Saur.

Part I

Social Media in Informetrics

Chapter 2

Content Description on a Mobile Image Sharing Service: Hashtags on Instagram

The mobile social networking application Instagram is a well-known platform for sharing photos and videos. Since it is folksonomy-oriented, it provides the possibility for image indexing and knowledge representation through the assignment of hashtags to posted content. The purpose of this study is to analyze how Instagram users tag their pictures regarding different kinds of picture and hashtag categories. For such a content analysis, a distinction is made between Food, Pets, Selfies, Friends, Activity, Art, Fashion, Quotes (captioned photos), Landscape, and Architecture image categories as well as Content-relatedness (ofness, aboutness, and iconology), Emotiveness, Isness, Performativeness, Fakeness, "Insta"-Tags, and Sentences as hashtag categories. Altogether, 14,649 hashtags of 1,000 Instagram images were intellectually analyzed (100 pictures for each image category). Research questions are stated as follows: RQ1: Are there any differences in relative frequencies of hashtags in the picture categories? On average the number of hashtags per picture is 15. Lowest average values received the categories Selfie (average 10.9 tags per picture) and Friends (average 11.7 tags per picture); for highest, the categories Pet (average 18.6 tags), Fashion (average 17.6 tags), and Landscape (average 16.8 tags). RQ2: Given a picture category, what is the distribution of hashtag categories; and given a hashtag category, what is the distribution of picture categories? 60.20% of all hashtags were classified into the category Content-relatedness. Categories Emotiveness (about 4.38%) and Sentences (0.99%) were less often frequent. RQ3: Is there any association between image categories and hashtag categories? A statistically significant association between hashtag categories and image categories on Instagram exists, as a chi-square test of independence shows. This study enables a first broad overview on the tagging behavior of Instagram users and is not limited to a specific hashtag or picture motive, like previous studies.

2.1 Introduction

"A picture is worth a thousand words"; but how would one describe a picture with a handful of terms? What elements of the picture would one want to represent? What kind of terms would one choose?

These questions always arise in the context of knowledge representation, or to be more precise, in indexing as an application of knowledge representation. In order to capture the content of a document like a picture, in professional information services indexing deals with the representation of single objects through controlled concepts (Stock & Stock, 2013). Besides a controlled keyword assignment by human or machine indexers, the free allocation of keywords by everyone has obtained a huge impact since the beginning of the Web 2.0. Nowadays, folksonomies (Peters, 2009) have become indispensable for the web. They can be found in many areas, like for example Twitter (in the field of microblogging), Mendeley (as a social reference management system), Flickr (as a photo management and sharing application), or Instagram (for sharing pictures and videos within a social networking mobile application). Instagram is becoming more and more popular as a social photo and video sharing application. Here, the free allocation of keywords is done by the assignment of hashtags.

How do users index a picture on Instagram with at most 30 hashtags?

What is new in this article? We found out for the first time that different picture categories (as, for instance, Architecture, Fashion, or Food) exhibit both, different numbers of hashtags as well as—more important— different kinds of hashtags (as Content-related tags or tags expressing Emotiveness, Fakeness, "Insta" aspects, Performativeness, and entire Sentences). Furthermore, it is demonstrated that there is an association between image categories and hashtag categories.

2.1.1 Indexing of Pictures by Folksonomies

Indexing of pictures can be conducted through a content-based or a concept-based approach (Rasmussen, 1997). Extracting features like color, shape, and texture belong to content-based indexing, whereas concept-based indexing requires a textual description (Lancaster, 2003; Rasmussen, 1997). Several methods and models were created for both approaches and the topic was widely discussed (Enser, 2008; Jörgensen, 2002, 2003). Above all, concept-based indexing refers to the ofness and aboutness of a picture (Shatford, 1986), which in turn are based on Panofsky's levels of meaning in art (Panofsky, 1955).

The first definition of the concept of "folksonomy" (being descended from "folk" and "taxonomy") goes back to Vander Wal and was stated in a blog entry of Smith (2004). Vander Wal (2007) chose this term for the free allocation of keywords—in a folksonomy called "tags"—by users in Web 2.0 services like Flickr or Del.icio.us. A folksonomy (Peters, 2009) is the result of the total quantity of all assigned tags in an information service.

Besides tags, the concept of "hashtags" exists. Hashtags are a composition of

and a character string, like for example #summer. On Instagram, it is even possible to use a hashtag emoji like #:). "[B]y using the # character to mark particular keywords, ... users communicate a desire to share particular keywords folksonomically" (Halavais, 2014, p. 36). As Halavais (2014) stated, some suggest the originator of the hashtag is Messina (2007), since he tweeted in August 2007 on Twitter "how do you feel about using # (pound) for groups. As in #barcamp [msg]?" (Messina, 2007). After this posting, the use of hashtags was established and was implemented on several platforms (Halavais, 2014). "Hashtags represent a way of indicating textually keywords or phrases especially worth indexing" (Halavais, 2014, p. 36). So hashtags are "user-generated metadata," too.

Tagging behavior in folksonomies was already observed several times (Daer, Hoffman, & Goodman, 2015; Golder & Huberman, 2006). Flickr, launched in February 2004 (Kremerskothen, 2012), applied one of the first well-known folksonomies for photos and therefore also for image and tag research (Beaudoin, 2007; Hollenstein & Purves, 2010; Nov, Naaman, & Ye, 2008; Rorissa, 2010; Stvilia & Jörgensen, 2010). For Facebook images, Denton, Weston, Paluri, Bourdev, and Fergus (2015) also developed and analyzed models for user conditional hashtag prediction. Besides Flickr, further photo services like Pinterest or Instagram were developed over time.

2.1.2 Instagram

During the last few years, Instagram has become more and more popular as a mobile social networking application for sharing photos and videos in various ways. It was launched in October 2010 with 25,000 signed-up users for the first day (Instagram, 2010). The number of users and functions has grown over time. The app's monthly active users numbered more than 800 million as of March 2018 (Instagram, 2018). The central point is still to share photos and videos.

Figure 2.1 shows a typical posting of a publicly posted Instagram picture. Beside the photo, the creator of the posting (in this case kajaf) has the possibility to add a description text and up to 30 distinct hashtags. The uploaded photo can be liked, commented on (also with hashtags), shared, or favored by Instagram users. The hashtags enable users to make the posting searchable under the respective chosen term. For example, the searching for #temple lists all content tagged with #temple.

Having a closer look at the hashtags in Figure 2.1, we are able to identify different sorts of hashtags. #temple or #bejing clearly describe the image's content. However, #Isurvivedchina is a complete sentence, #photography puts the document into a media category, #followme is a request to do something and, finally, #travelgram refers to Instagram. Obviously, there are different categories of hashtags on Instagram.

2.1.3 State of Research on Instagram

Research on Instagram lays its focus on different aspects like for example content analysis or hashtag use, which are summarized in the following paragraph.

Figure 2.1: Posted architecture picture by kajaf on Instagram.

Sheldon and Bryant (2016) surveyed 239 college students about their motives for using Instagram. Five distinct post types like advertising or information were analyzed by Coelho, de Oliveira, and de Almeida (2016) with respect to their likes and comments. The investigated posts depicted one of five business segments (food, hairdressing, ladies' footwear, body design, and fashion gym wear).

Giannoulakis and Tsapatsoulis (2015) investigated through an online survey whether other Instagram users would use the same hashtags like the picture owner (based on a set of 30 Instagram images and their most respectively descriptive hashtags). How hashtags are used to retrieve information in social media like Instagram was the subject of research by Buarki and Alkhateeb (2018).

Nashmi (2018) conducted a content analysis with 1,000 Instagram pictures which were posted up to four days after the Charlie Hebdo incident of January 2015 in order to investigate visual post changes. Two content analyses about posted Instagram photos by the 10 largest fast food companies as well as photos from users who posted about the companies were performed by Guidry, Messner, Jin, and Medina-Messner (2015). Holmberg, Chaplin, Hillman, and Berg (2016) categorized food images in order to analyze in which way and what kinds of food 14-year-old adolescents present on Instagram. To receive only pictures from 14-year-olds, they investigated appropriated user profiles by using the hashtag #14år ("14 years"). In order to characterize the dietary trend of cheat meals, Pila, Mond, Griffiths, Mitchison, and Murray (2017) conducted a thematic content analysis of 5,600 tagged #cheatmeal Instagram postings. Both photographic and textual elements of the cheat meal postings were analyzed. Also referring to the subject of food, Zhang, Hashim, Baghirov, and Murphy (2018) investigated the content of 1,382 Instagram postings tagged with #Malaysianfood. Therefore, the hashtag descriptions were placed into the categories informative or emotional as well as positive or negative.

Marcus (2016) performed an image content analysis about pro-anorexic and fat acceptance communities in Instagram. She identified several popular hashtags regarding those topics in order to obtain 400 suitable pictures for each community. 1,967 Instagram pictures tagged with #swisher (a popular cigar brand in the United States) were thematically analyzed by Allem, Escobedo, Chu, Cruz, and Unger (2017) to support future tobacco control efforts and education campaigns as well as to understand health behavior through social media data. Sensitive self-disclosures and their responses on Instagram were analyzed by Andalibi, Ozturk, and Forte (2017) through a three-phase methodology: "Phase I establishes the content of depression-tagged posts on Instagram; Phase II investigates the kinds of responses these posts attract; and Phase III examines relationships between the kinds of posts and the kinds of responses they attract" (Andalibi et al., 2017, p. 1489). Depression tagged postings were identified through the hashtag #depression. Their final sample consists of 788 images and captions. Souza et al. (2015) studied Instagram selfies in context of their characteristics (e.g., demographic information, distribution of post frequency, likes, comments, etc.). An examination of a variety of selfie hashtags with respect to their popularity was conducted, too. (Oh, Lee, Kim, Park, & Suh, 2016) investigated Instagram users' "participatory hashtag practices" regarding the Weekend Hashtag Project. "Participatory hashtag practices" describes a "hashtagging phenomenon, where a certain user account suggests a hashtag to its followers and promotes them to upload photos suitable to the hashtag" (Oh et al., 2016, p. 1281). Public Instagram photographs tagged with #funeral were analyzed by Gibbs, Meese, Arnold, Nansen, and Carter (2015). In the field of sports, Pegoraro, Comeau, and Frederick (2018) also investigated Instagram pictures which were tagged with certain hashtags. They conducted a content analysis for images tagged with #SheBelieves (n=629) or #FIFAWWC (n=706). Likewise in the context of sports, but with respect to postings tagged with the hashtag #fitspo, Carrotte, Prichard, and Lim (2017) analyzed 415 postings from Instagram, Tumblr, Facebook, and Twitter, whereas with 360 posts the majority were found in Instagram. Veszelszki (2016) investigated 400 Instagram postings tagged with #time, #truth, or #tradition in respect of the relationship between the image and text (hashtag).

The following study is one of the bases for this research project. In an empirical study about content and users on Instagram, Hu, Manikonda, and Kambhampati (2014) determined 8 popular photo categories (friends, food, gadget, captioned photo, pet, activity, selfie, and fashion) for Instagram and 5 different types of user clusters. Furthermore, they found that the number of a user's followers does not depend on the users' posted photos on Instagram Hu et al. (2014).

2.1.4 Tagging Behavior on Instagram

Previous studies on Instagram hashtags investigate one specific hashtag or some hashtags referring to a certain topic. This study analyzes the tagging behavior of Instagram users in a wider field by use of a content analysis (Krippendorff, 2004). Therefore the authors' hashtags, assigned to Instagram images of 10 different picture categories, were analyzed regarding different hashtag categories (Dorsch, Zimmer, & Stock, 2017). The main research questions are:

RQ1: Are there any differences in relative frequencies of hashtags in the picture categories?

RQ2: Given a picture category, what is the distribution of hashtag categories; and given a hashtag category, what is the distribution of picture categories?

RQ3: Is there any association between image categories and hashtag categories?

The research aspects of this study are represented in Figure 2.2. The tagging behavior will be investigated for Instagram users regarding their Instagram picture postings. The factors of a posting that are to be analyzed are the hashtag counts of a picture and the categories of the assigned hashtags, as well as the subject categories of the pictures. The study deepens our understanding about tagging behavior on social media with the example of Instagram. Additionally, we are going to gain insights regarding knowledge representation by layman indexers.

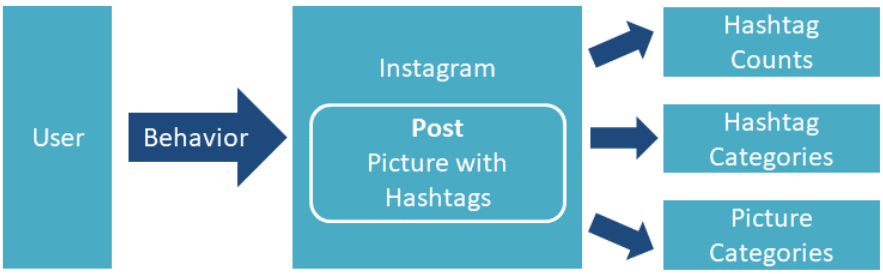


Figure 2.2: Research model.

2.2 Methods

Content analysis (Krippendorff, 2004) provides the option to analyze content systematically. To analyze the tagging behavior of Instagram users, pictures and hashtags were coded—through content analysis—into specific picture and hashtag categories.

2.2.1 Selection of Picture Categories

The collected 1,000 Instagram pictures conform to the following selected 10 image subject categories (Table 2.1). Thereby, each picture category contains 100 pictures.

Table 2.1: Analyzed picture categories with the descriptions of their image subjects.

Picture category	Description: The picture depicts…
Activity	… "both outdoor & indoor activities, places where activities happen, e.g., concert" (Hu et al., 2014, p. 597). This does not include landscape and architecture related pictures.
Architecture	… all architecture related content, except if places where activities happen are in the foreground. Art-related work is excluded.
Art	… drawings, paintings, sculptures, land art, crafted stuff, tattoos, and any other art-related content. Photo shots are not included, except where they show the art, like for example a photo of a sculpture. The original author of the artwork can be the profile owner of the posted picture or a foreign person.
Captioned Photo	… "pictures with embed text, memes, and so on" (Hu et al., 2014, p. 597). Besides this, the picture also has to show graphical content (e.g., persons, items, landscapes, etc.), because pure text does not allow analysis of the picture content in terms of the chosen hashtags.
Fashion	… "shoes, costumes, makeup, personal belongings, etc." (Hu et al., 2014, p. 597).
Food	… "food, recipes, cakes, drinks, etc." (Hu et al., 2014, p. 597).
Friends	… "users posing with others friends" (Hu et al., 2014, p. 597) or only the friends of a user. At least one person must be depicted in the picture. Contrary to the definition of (Hu et al., 2014, p. 597), faces do not have to be seen.
Landscape	… all nature related content, except if places where activities happen are in the foreground. Art-related work is excluded.
Pet	… "animals like cats and dogs which are the main objects in the picture" (Hu et al., 2014, p. 597).
Selfie	… "self-portraits; only one human face is present in the photo" (Hu et al., 2014, p. 597). It must be identifiable that the photo was taken by the person in the picture ("mirror pictures" are also allowed).

The categories are based on Hu et al. (2014) who detected empirically by cluster analysis and qualitative analysis (done by two human coders) 8 popular general Instagram photo categories in their study about Instagram photos and users. Categories 1, 4 to 7, and 9 to 10 originate from them, but their category descriptions were complemented or partly modified for this work after our pretests, as can be seen in Table 2.1.

It should be noted that the category Gadget from Hu et al. (2014) is missing, because it was omitted after the pretest of this study. The test demonstrated that it is too broadly defined for collecting pictures according to the chosen method of this study. Partially, too many pictures were tagged which had nothing to do with a gadget or had violated the rules of the codebook (especially general picture category codebook rule 2, R2) (Table 2.2). Since those #gadget pictures comprised more picture motifs, a noticeable amount of these pictures (compared to the other pretest picture categories) could not be sorted precisely into the picture category Gadget and therefore were not used for a pretest.

Table 2.2: General coding rules for categorizing Instagram posts.

#	Rule
R1	The basic requirement for each picture is that they contain their respective category hashtag. Without this hashtag, they are not admitted for the categorization.
R2	All category pictures were chosen so that they can be treated as prototypes for their respective categories and are therefore always classified into one category only. With respect to the diversity of picture motifs in social media like Instagram, it cannot always be excluded that motif borders blur. For example, a picture predominantly depicts the content of its respective category, but it can also contain some subsidiary further content (like for example a pet in the background of a selfie). For that reason preference rules were developed. They apply for those cases and ensure a clear picture categorization.

In particular, the categories Fashion and Gadget exhibit strong overlaps during the pretest. To select narrower (hashtag) terms of Gadget to serve as picture categories could be a possible solution for the stated problems. Such a solution was discarded for this study because this category approach would stand in contrast to the other broadly diversified categories.

Categories 2, 3, and 8 were created for this study and are also widespread subjects on Instagram, as their hashtag number in Table 2.3 displays. However, why had any categories to be chosen? The categories represent a selection of possible subjects; and the study does not claim to cover all possible types of picture subjects. Rather, they provide a balanced distribution of data. Furthermore, a strict picture segmentation into categories, with one (main) picture motif, enables the possibility to make clear statements about the categories and possible category relations. Due to limited coder capacities, only these 10 categories could be selected.

Table 2.3: Used top picture category hashtags and alternative hashtags with the total count of their assigned hashtags in Instagram (valid as of December 18, 2016).

Picture category	Top picture category hashtags	Number of tagged media for the top hashtag	Alternative hashtag(s) and the number of tagged media
Activity	#activity	0.7 m	#activities (0.4 m)
Architecture	#architecture	45.8 m	#architectures (0.2 m) #architectureporn (3.2 m) #cityscape (4.2 m) #cityscapes (0.4 m)
Art	#art	211.3 m	#artwork (26.6 m)
Captioned Photo	#quote	38.2 m	#quotes (33 m) #meme (14.7 m) #memes (10.4 m)
Fashion	#fashion	324 m	#fashions (0.6 m)
Food	#food	198 m	#foods (8.5 m) #foodporn (106.5 m)
Friends	#friends	247.7 m	#friend (47 m) #friendship (28.9 m)
Landscape	#landscape	46.5 m	#landscapes (3.1 m) #countryside (5.6 m) #countrysides (3.4 m)
Pet	#pet	39.7 m	#pets (25.5 m) #animal (29.1 m) #animals (25.8 m)
Selfie	#selfie	281.7 m	#selfies (17.1 m)

m, million.

2.2.2 Selection of Hashtag Categories

In knowledge representation, different concept categories can be derived for image indexing. Concepts describing aboutness and ofness are most popular in indexing non-textual documents. Those categories serve as a basis for the analysis of the Instagram hashtag tagging behavior in this study. The difference to the regular usage in indexing is that these categories were considered from a retro perspective. They do not help an indexer to decide whether a certain index term had to be chosen or not (and consequently to guarantee an appropriate capture of a document), but to categorize empirically the collected hashtags. Existing tagging categories influenced the creation of the categories, too. The following categories were developed for categorization: Content-relatedness (including ofness, aboutness, and iconology), Emotiveness, Isness, Performativeness, Fakeness, "Insta"-Tags, and Sentences.

2.2.3 Content-related Tags

Content-related tags involve everything a picture directly or abstractly depicts. This category refers to the definitions of aboutness and ofness in pictures which in turn are based on Panofsky's (1955) three levels of meaning in the visual arts. These levels are called per-iconographic, iconographic, and iconologic. They relate to different meaning aspects in an artwork. The pre-iconographic level refers to practical experience and can be factual or expressional. For example, factual would be the representation of natural objects like "human beings, animals, plants, houses, tools and so forth," or expressional "by identifying their mutual relations as events; and by perceiving such expressional qualities as the mournful character of a pose or gesture, or the homelike and peaceful atmosphere of an interior" (Panofsky, 1955, p. 28). In addition, the iconographic level comprises specific themes and concepts instead of basic objects and events. Here one is on the level of literary knowledge, as for instance

> "[...] by realizing that a male figure with a knife represents St. Bartholomew, that a female figure with a peach in her hand is a personification of veracity, that a group of figures seated at a dinner table in a certain arrangement and in certain poses represents the Last Supper, or that two figures fighting each other in a certain manner represent the Combat of Vice and Virtue" (Panofsky, 1955, pp. 28-29).

Finally, iconology declares the intrinsic meaning or content (e.g., Weltanschauung) and addresses therefore symbolic values: "It is apprehended by ascertaining those underlying principles which reveal the basic attitude of a nation, a period, a class, a religious or philosophical persuasion-qualified by one personality and condensed into one work" (Panofsky, 1955, p. 30). For example, to understand Leonardo da Vinci's fresco "Il Cenacolo"/"L'Ultima Cena" (The Lord's Supper) as culture of the Italian high renaissance, is an association to the level of iconology.

This classification of artworks became transferred to a general consideration of subject kinds in a picture (Shatford, 1986; Shatford Layne, 1994). A picture can be of something and also about something. In information science, this means that the first two of Panofsky's three levels (1955) correspond particularly to an ofness and aboutness. For the first level, the factual pre-iconographical objects comprise the ofness (also called "generic Of") whereas the expressional pre-iconography denotes the aboutness. Ofness of iconography (also called "specific Of") includes more "specifically what a picture is Of" (Shatford, 1986, p. 44) ; iconographical aboutness, on the other hand, denotes allegories, personifications, and symbols. Since Shatford (1986) discusses Panofsky's levels in the context of image indexing, iconology is excluded, because it addresses the interpretational aspects of a picture. It would not be possible to index it consequently. It is also possible to subdivide the different kinds of ofness and aboutness with respect to the facets time, space, activities, events, and objects to get a more specific classification (Shatford Layne, 1994).

In her explanations about the subject access to art images Shatford Lane's (2002) remarks that it is helpful to distinguish between ofness and aboutness in order to determine which index terms provide which kind of subject access (e.g., the difference between pictures that depicts death and are therefore of it versus pictures depicting death symbolically and thus are about death). Since the terms are already indexed and nobody has to decide whether an ofness or aboutness term is better, this is not necessary for an analysis of hashtags. Even so, it would be interesting to distinguish between them.

Furthermore, Shatford Layne (2002) explains that there is a wide reach between generic ofness and specific ofness. It is not possible to consider them dichotomously. Also, aboutness is not always determinable, obviously. In this case, she refers only to art images, but she already addresses this difficulty in earlier works for all kinds of images except abstract artworks (Shatford, 1986).

Subjectivity always occurs when pictures are indexed by ofness and aboutness concepts (Shatford, 1986). According to Turner (1995), the distinction between ofness and aboutness for the purpose of indexing is often impossible. Already Panofsky (1955) stated that the boundaries between pre-iconography and iconography could blur.

Besides, aboutness also depends on the point of view (Maron, 1977). For example, Maron (1977) distinguishes between S-about (subjective about; the relation between a document and its reader's experiences), O-about (objective about; "an external or observer's point of view" (Maron, 1977, p. 41)), and R-about (retrieval about; "the information searching behavior of a class of individuals" (Maron, 1977, p. 41)). Ingwersen (2002) also stated different kinds of aboutness: Author aboutness provides the content of an author's document, indexer aboutness is the interpretation of the content by an indexer, request aboutness is the information need formulated as a search argument, and user aboutness is the interpretation of the content by a user. This study (Rondeau, 2014) addresses the aboutness, assigned from different subjective users' perspectives.

The codebook was developed to be as objective as possible, so that coding is repeatable for anyone. Therefore we did not differentiate between all kinds of introduced ofness and aboutness. They were merged together into the category

Content-relatedness. Moreover, and as stated above, such a distinction is especially important for deciding whether a term should be indexed or not for a certain subject area and information need (Shatford, 1986). This difficulty is dropped when categorizing hashtags because the terms are already assigned. The fact that some of the hashtags refer (rather) to the author description of the picture is another reason for the combination to only one category.

Although the subjective interpretational level of iconology (Panofsky, 1955) is excluded for indexing in information science (Shatford, 1986), it can be tagged by users of a folksonomy (Stock & Stock, 2013). During the coding, we detected some hashtags that can be best described as iconological. Those hashtags were also sorted into the category Content-relatedness since they are relating to some "content." "User-specific tags [describing] or [evaluating] a document only from the user's very own perspective," so that some tags "are virtually meaningless to anybody except their creators" (Pluzhenskaia, 2006, p. 23) are likewise counted as content-related hashtags.

2.2.4 Emotiveness

The category Emotiveness comprises emotional hashtags. But how is the concept of emotions defined? As summarized by Knautz (2012) and Siebenlist (2013) there exists a variety of definitions. Several researchers tried to define a set of basic emotions (Ortony & Turner, 1990). Basic emotions used in this study are love, happiness, fun, surprise, aspiration, sadness, anger, disgust, fear, and shame, and were adopted from Siebenlist (2013). Besides this, all possible manifestations of emotions a user could formulate were considered as emotive hashtags.

Looking at the categories, one could argue against a separation of emotional hashtags. Formally, they belong to the content-related tags and as a consequence to the category Content-relatedness. Likewise Shatford (1986) assigns emotions as aboutness. However, in Folksonomies, those tags were analyzed separately (Beaudoin, 2007; Mohammad & Kiritchenko, 2015; Kipp, 2006; Schmidt & Stock, 2009; Yanbe, Jatowt, Nakamura, & Tanaka, 2007; Zhang et al., 2018). For Twitter, Mohammad and Kiritchenko (2015) also pointed out that emotional hashtags can function as labels of emotions in tweets. To treat emotional hashtags separately in order to get distinctive information and because such a distinction is not problematic for the objectivity of the codebook, emotional hashtags like #love, #happy, or #fun were separated to the category Emotiveness.

2.2.5 Sentences

If the separation of emotiveness tags from content-related hashtags is questionable, the same applies to sentences. Sentences represent content, but with the function of information condensation and not as information filters. They cannot be counted as part of an indexing process like single concepts and phrases, but are a part of abstracts (Stock & Stock, 2013). For that reason, they had to be separated from the content-related terms and phrases. Only whole sentences (containing subject, verb, and object) and their abbreviations (e.g., wiwt = what I wore today) were considered for this category. If a Sentences hashtag contains overlaps with other hashtag categories, the hashtag will always be categorized into the category Sentences, since the tag primarily represents a sentence.

2.2.6 Isness

The category Isness originates from Ingwersen (2002). It comprises all non-topical features of a document, e.g., for pictures all "technical aspects of the photograph (e.g., camera type, length of exposure, aperture)" (Stock & Stock, 2013, p. 616). Further specific examples for this study, corresponding to Ingwersen (2002), are the (user)name of the person who took the photo (only if she/he is not depicted), date, location (if not depicted), or type of the picture (e.g., selfie). Special consideration was given to art images. In contrast to isness features like whether the photo is a selfie, is a black and white picture, or is a photography overlayed with a photo filter, photos usually depicts artworks instead of being an artwork. According to Shatford Layne (2002), hashtags like #art or #artwork were assigned to the category aboutness, when they are related to a photo depicting art. If the entire image is an artwork (e.g., computer graphics, photographic art), such tags describe the isness. The same applies for people, objects, and places which could be—according to context— content-related or isness-related. For example, #polishgirl is content-related if the girl is depicted in the picture, but it has to be categorized into Isness if no person is visible and one had to assume a Polish girl is the author of the posting or is the photographer of the image.

2.2.7 Performativeness

"Performative tags" (Peters & Stock, 2007), also called "time and task related tags" (Kipp, 2006; Kipp & Campbell, 2006) or "signaling tags" (Dennis, 2006), call for an action like #followmearound, #discover, or #like4like. Performativeness goes back to Austin (1962), as there are sentences which do not describe a proposition but a call for or a promise of an action. They were classified into the category Performativeness. Especially for Instagram, so-called hashtag contests like "participatory hashtag projects" (Oh et al., 2016) or features exist. One participates by simply tagging a picture with a certain hashtag. In doing so, the chance of a price or feature could exist in some cases. Feature means an account posts (and thus promotes) a picture or username of someone, like for example the Instagram account hypetoronto displayed in Figure 2.3. Often, those tags are named similar to their related accounts like #hypetoronto for hypetoronto (Figure 2.3), so that the call to action is not always exactly recognizable. However, such hashtags still belong to the performative ones and were categorized to this category, too. Some cases received an exemption clause in the codebook.

Tags like #pictureoftheday or #petoftheday have a "contest character," because

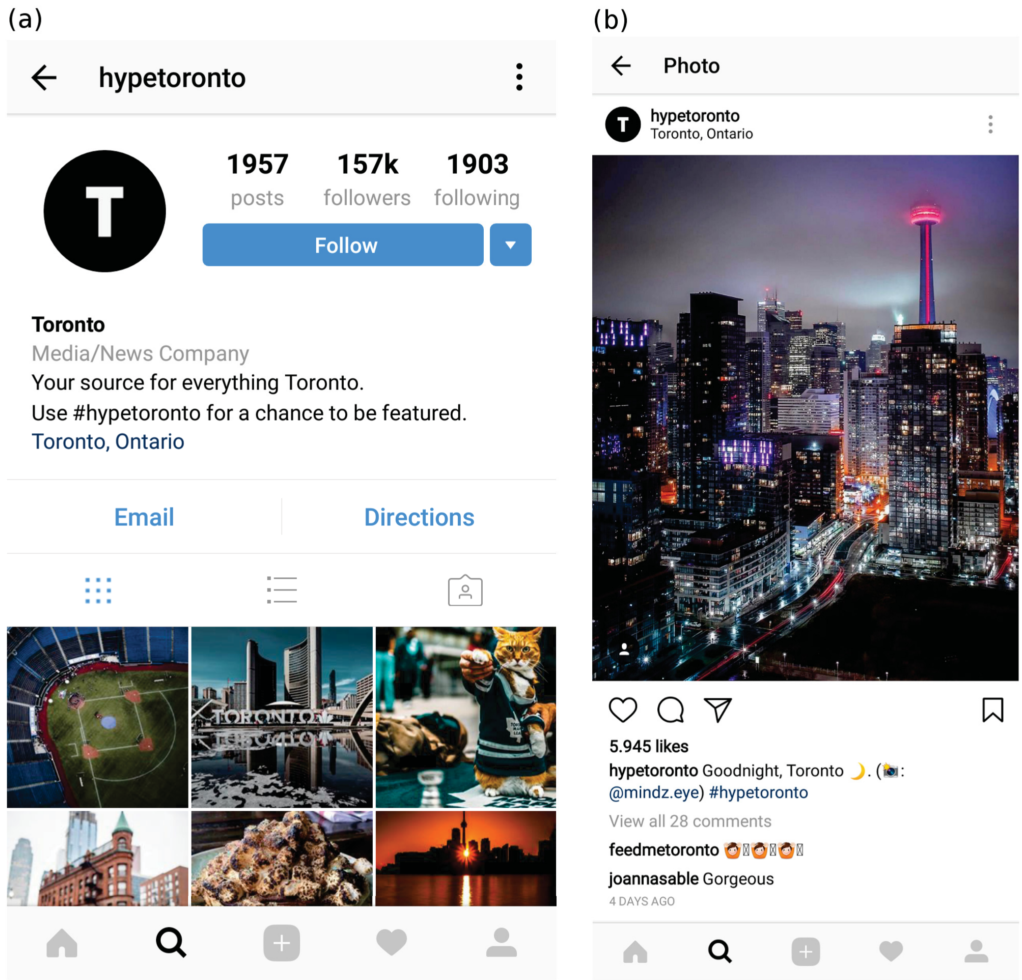


Figure 2.3: Featured performative hashtag. (a) Instagram account hypetoronto calls for using their hashtag #hypetoronto, to get the chance to be featured in their feed. (b) hypetoronto features a photograph by @ mindz.eye who previously used the hashtag #hypetoronto.

the tagged photos slightly compete with each other to be the best picture or the best depicted pet of the day. For that reason, they possibly could be categorized as performative tags. However, they cover up aboutness (e.g., “pet of the day,” when displayed in the picture) or isness (e.g., “photo of the day,” regarding aspects of the image and if no photo is depicted in the image). Since only one category could be selected, such general performative hashtags combined with “of the day” were not categorized into the category Performativeness. The categories Content-relatedness or Isness were chosen depending on the case.

2.2.8 Fakeness

The category Fakeness includes all stated hashtags that are not valid for the respective image or posting description in any way and are therefore deliberately incorrectly assigned, for instance, a picture tagged with #dog but depicting a single cat and nothing else. If the decision for a hashtag provides room for interpretation and it can not be clearly determined if the hashtag is true or false, it will not be categorized into Fakeness. The same applies to hashtags containing typos or the false singular/plural form. It was considered to categorize those latter hashtags into Fakeness or to create a category “Falseness.” However, it is well known that folksonomy vocabulary is not formally proved (Peters, 2009) and such mistakes may happen. Besides, some tags are more common in their singular or plural form. It may be not correct to assign a picture representing a single person with #friends, but it is not a false statement in total.

2.2.9 “Insta”-Tags

A special focus lays on the category “Insta”-Tags. This category contains the following hashtags or hashtag components: “Insta,” “gram,” “Instagram,” or an abbreviation of these expressions (like #IGers for Instagramers). Such components contain hashtags like for example #instagood, #instadaily, #instapic, #instaart, #instacat, #webstagram, or #webstagramers and would be classified into this hashtag category. Hashtags which include one or more stated components, but do not refer to Instagram, were excluded from this category.

Both content analyses for the stated picture and hashtag categories are predominantly based on a direct content analysis approach (Hsieh & Shannon, 2005) also known as deductive content analysis (Elo & Kyngäs, 2008). In those approaches, analysis and initial codes are based on previous knowledge of a theory or research findings (Elo & Kyngäs, 2008; Hsieh & Shannon, 2005). Additionally, new categories emerged from the observed data which are lean on conventional content analysis (Hsieh & Shannon, 2005), also known as inductive content analysis process (Elo & Kyngäs, 2008). For the hashtag category codebook, these newly created categories are Fakeness, “Insta”-Tags, and Sentences, where Content-relatedness (aboutness/ofness/ iconology), Emotiveness, Isness, and Performativeness were found in the literature. For the picture category codebook, we followed mainly deductive analysis and referred to the work by Hu et al. (2014). The new categories Architecture, Art, and Landscape are widespread on Instagram and thus inductively considered as categories.

2.2.10 Codebooks

In total, two codebooks were developed for this study: the "Instagram picture category codebook" for the selection of the 1,000 Instagram picture postings, and the "Instagram hashtag category codebook" for the analysis of the assigned picture posting hashtags. They contain a short introduction about their purpose and all common coding rules which apply on every investigated Instagram posting, as well as their specific picture or hashtag codes. The creation and structure of both codebooks are based on MacQueen, McLellan, Kay, and Milstein (2009).

The rules (R) in Table 2.2 apply to every investigated Instagram post and are taken from the Instagram picture category codebook. All pictures which do not predominantly show one category or suit the preference rules have been discarded. The rules (R) in Table 2.4 apply to every code and hashtag (and are taken from Instagram hashtag category codebook).

Based on MacQueen et al. (2009) every specific picture or hashtag code contains: a code name, a definition, and also a short definition for the coding category, information about when to apply or when not to use the code (as well as some preference rules), and coding examples.

It was necessary to continually update the books due to the pretest and coding process as well as to recode parts of data in the end, because some rules changed during the process.

2.2.11 Data Collection

All Instagram pictures were collected during November 2016 to January 2017. The final data comprises 50 Instagram pictures for a pretest and at least 1,000 Instagram pictures for the analysis. Pretest pictures were chosen manually. Category Architecture replaced category Gadget after the pretest and was thus not included into the pretest. The collection of the examined 1,000 Instagram pictures can be subdivided in an automatic and a manual part. The automatic part includes downloads of raw datasets via the Instagram web viewer Imgrum Application Program Interface (API)[1]. The images could not be downloaded directly from the Instagram API, because this API was not freely accessible at this time. Based on an adequate top hashtag (Table 2.3) for each picture category, a dataset was downloaded.

Criteria for selecting the top hashtags were the thematic reference to the respective picture category in combinatio nwith the number of tagged media. Regarding the thematic aspects, all hashtags were chosen in order to be very general and close to their picture category, such that they cover their category optimally. For instance, the category Fashion comprises not only clothes, but rather accessories and makeup

[1] `http://www.imgrum.net/`

Table 2.4: General coding rules for categorizing Instagram hashtags.

#	Rule
R1	Every hashtag can only be assigned to one category. Thereby, the category is chosen which (according to the coding rules and in combination with the depicted content on the picture) most closely applies to the hashtag.
R2	Only all author assigned hashtags in the picture description are considered for the coding. Hashtags in comments (also when the author of the picture commented on his/her posted picture) are excluded from the analysis.
R3	Analysis language of hashtags is English. Every hashtag will be excluded from analysis, if it is not formulated in English. Excluded from this rule are proper names (e.g. for cities, mountains, meals, names of regional holidays). They can be retained in their language. If the coder does not know them, he or she had to translate those proper names by himself or herself (only for understanding).
R4	The categorization of a hashtag primarily depends on the content of the associated posted Instagram picture. Additionally, further information can be obtained from the author's picture description. Similar as in R2, all other information in comments is excluded for the coding process, as well as the author's/commenter's profile information.
R5	Typos in hashtags are not considered.
R6	Special characters within hashtags or hashtags which consist only of one or more special characters are not taken into account for the analysis. The same applies for emojis.
R7	The coder has to figure out the meaning of hashtag abbreviations. Therefore hashtag meanings and definition websites2 can be used. It is important that the coder annotates all used definition websites. In addition, all coders keep a record of the abbreviations and their written-out form. If the meaning is ambiguous or vague, the coder has to choose the most appropriate category.
R8	"Divided tags" (e.g., #private #lessons; #black #cat) are always considered and categorized separately.
R9	Same hashtags, but double or multiple assigned tags are only considered as one (e. g. #food, #food, #food counts as one hashtag).

too. Therefore, it would not have been useful to choose a hashtag like #outfit, or #clothes. Relating to the count, the hashtag had to be one of the most common hashtags. Table 2.3 shows that every chosen hashtag had the highest number of tagged media in contrast to the alternative hashtags.

Such an approach enables a random picture selection which usually matches the selected hashtag aboutness and is adopted from Marcus (2016). However, it should be noted that not every picture subject conformed to the tagged hashtag, since the author could make false statements (e.g., using #selfie despite the picture not depicting a selfie). Besides, some appropriate subjects were not in accordance with the previously enumerated rules of the category codebook: for example, a pet picture subject tagged with #friends. The pet may be a friend in this case, but as the codebook denotes, only human friends are valid picture subjects for this category. Or, the picture shows elements of two or more categories in equal shares. In doing so, it fulfills not only the category the hashtag pretends, but rather more categories. For that reason, the automatically generated datasets covered a multiple of the respective required 100 category pictures, so that they could be cleaned up. It should be noted that some data had to be downloaded in several stages, since after the manual check only a few crawled pictures fulfilled the conditions for this study (language, category, etc.), so that the first data set did not always suffice. The datasets were collected in parallel to the coding process. Every final and adjusted downloaded data set contains the following information: user name of the posting author; URL to the posting on Imgrum; the full image description of the posting; all used hashtags in the image description; the total hashtag number of one image description; the picture category name (e.g., Activity, Architecture, Art, etc.); a unique picture ID; and a separate file, containing each picture file named with its ID.

As stated in coding rule R2 (in Table 2.4), further posting comments (whether from the author or the users) were not considered for this study.

2.2.12 Coding Process

The coding process involves the assignment of the picture description hashtags to the introduced hashtag categories and with regard to the code rules. In November 2016, the coding pretest took place, followed by the categorization of the 1,000 image dataset during December 2016 to February 2017. According to a 4-eyes principle, 2 coders were involved in the process. First, both of them had to code individually each data set in accordance with the hashtag codebook. What followed was a check as to whether the assigned codes matched. If they did not match, both coders had to discuss and to agree for one category (based on the codebook rules).

Coding is an iterative process. After the pretest and during the coding process some (special) cases appeared for the first time or have been defined as too vague, making it necessary to adjust or supplement some of the rules. Even if there was a pretest, coders had to adapt themselves to the coding rules at the beginning of the main coding process. This might cause additional reasons for recoding, too.

For instance, hashtags containing "of the day" like in #pictureoftheday or #photooftheday were widespread in Instagram. "Picture of the day" and similar hashtags generally refer to Isness. According to that and thus the former codebook rules, all kinds of "oftheday"-hashtags had to be coded into Isness at the beginning. During the main coding phase, this rule was revised in consequence of more and more "of the day"- hashtags addressing an element like a living being or object in the picture (e.g., "catoftheday" and a cat was depicted, "foodoftheday" and food was depicted). In this case and also in the other cases, the pretest and the pre-developed hashtag codebook were not sufficient to handle those cases correctly and the codebook had to be continually updated. In the end, a follow-up check and further recoding of coding data parts (regarding changed rules as well as the category content-relatedness) took place. As stated, the category content-relatedness comprises ofness, aboutness, and iconology in the end. At first, the coders had only distinguished between ofness and aboutness hashtags. After the entire coding process, both categories were merged into the category content-relatedness.

2.3 Results

The following paragraph presents the outcome of this study following the three research questions.

RQ1: Are there any differences in relative frequencies of hashtags in the picture categories?

Overall, 14,649 hashtags for a total 1,000 Instagram pictures were coded into their respective hashtag categories (Table 2.5). Thereby, the average number of hashtags per picture in the respective picture category (Figure 2.4 varies from nearly 11 to about 19 hashtags with an average of 15 hashtags. Especially, the person-related categories Selfie (average 10.9 tags per picture) and Friends (average 11.7 tags per picture) received the lowest average values. Pet (average 18.6 tags), Fashion (average 17.6 tags), and Landscape (average 16.8 tags) are the picture categories with the highest average hashtag count for their pictures. The standard deviation is similar for nearly every picture category (between 8 and 9).

RQ2: Given a picture category, what is the distribution of hashtag categories; and given a hashtag category, what is the distribution of picture categories?

With 60.20%, the majority of all hashtags were classified into the category Content-relatedness, followed by the hashtag category Isness with almost 14.87%. "Insta"-Tags (7.32%) were third most; tags of the category Performativeness (7.20%) were the fourth most assigned. A minority of hashtags was classified into the categories Fakeness (5.03%), Emotiveness (4.38%), and Sentences (0.99%).

Pictures of all 10 categories are also predominantly tagged with content-related hashtags, but high values for "Insta"-Tags in the category Pet (20.24%) and Fakeness

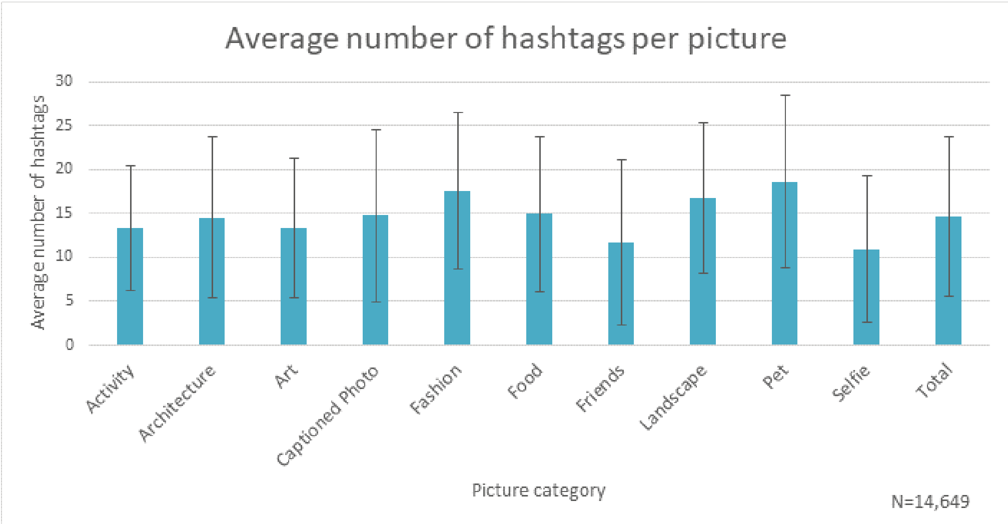


Figure 2.4: Average number of hashtags per picture and standard deviation within its respective picture category.

Table 2.5: Relative frequency of hashtag categories by picture categories (N=1,000 posts; 100 posts per picture category.)

	Content-related-ness	Emotive-ness	Fakeness	“Insta”-Tags	Isness	Performa-tiveness	Sentences	%
Activity	74.29	4.35	3.75	1.72	9.90	5.25	0.75	100.00
Architecture	63.20	2.54	2.54	6.67	15.41	9.35	0.28	100.00
Art	68.54	1.05	5.62	4.19	14.68	5.69	0.22	100.00
Captioned Photo	61.67	5.16	11.80	3.46	7.33	8.75	1.83	100.00
Fashion	59.51	3.52	7.38	5.91	16.70	6.02	0.97	100.00
Food	51.34	1.81	6.71	6.31	25.84	7.11	0.87	100.00
Friends	57.75	8.74	3.94	7.88	14.40	6.17	1.11	100.00
Landscape	61.68	3.98	1.66	5.47	16.16	10.75	0.30	100.00
Pet	53.93	8.02	3.82	20.24	7.59	4.31	2.10	100.00
Selfie	51.05	4.57	2.38	7.96	23.70	9.06	1.28	100.00
Total	60.20	4.38	5.03	7.32	14.87	7.20	0.99	100.00

for the category Captioned Photo (11.80%) are striking. In contradiction to the total values of all assigned hashtags, they are the second most assigned hashtag categories for their stated picture category. For all other picture categories, the second most assigned category is Isness. Even if the total amount of "Insta"-Tags is the third most assigned, this hashtag category is not the third most for any of the remaining 9 picture categories. For them, "Insta"-Tags are mostly fourth or fifth most, and for the categories Activity and Captioned Photos they were even sixth most assigned. The overall high distribution is therefore resulting due to the frequent occurrence in the category Pet. Considering the last ranks, for each picture category Sentences hashtags were tagged at least most, which is in accordance with the overall hashtag category distribution. It is noticeable, however, that the category Emotiveness is only three times in the sixth rank (for categories Art, Fashion, and Food).

The distribution of the hashtag category Content-relatedness varies between half and three quarters of all hashtags per each image category. With approximately 51%, Food and Selfie related pictures are tagged fewest of all with Content-related hashtags. The highest percentages record the categories Activity (74.29%) and Art (68.54%). Even though Food and Selfie have the lowest rate of content-related tags, the occurrence of Isness tags in these two categories is the highest. The majority of the other picture categories have around about 15%, apart from the categories Activity (9.90%), Pet (7.59%), and Captioned Photo (7.33%), with under 10%. The distribution for "Insta"-Tags amounts for all picture categories less than 8%, except for the category Pet (20.24%). With 1.72%, Activity pictures have an especially low frequency of "Insta"-Tags. Regarding Performativeness, the categories Landscape (10.75%), Architecture (9.35%), and Selfie (9.06%) have the most hashtags of this category. Generally, Fakeness tags are not strongly represented. Similar to "Insta"-Tags, there is only one category exceeding the 8% for Fakeness, namely the category Captioned Photos (11.80%). The fewest Fakeness tags were received by Landscape pictures (1.66%). Pictures about Friends (8.74%) or Pets (8.02%) were tagged most emotionally, whereas Art (1.05%) and Food (1.81%) images contained nearly no emotional hashtags. Within the hashtag category Sentences, Pet pictures (2.10%) record the highest percentage of those hashtags. It is notable that 35 out of 39 Pet Sentences hashtags are about emotional facts like for example #ilovemydog or #ihatemondays. So the category Pet received not only the most Sentences hashtags, but also the most emotional.

RQ3: Is there any association between image categories and hashtag categories?

A chi-square test of independence was conducted between hashtag categories and picture categories with the following hypotheses:

H_0: There is no association between hashtag categories and hashtag pictures.

H_A: An association exists between hashtag categories and hashtag pictures.

All expected cell frequencies were greater than five. There was a statistically significant association between hashtag categories and picture categories, $\chi 2(54) = 1557.860, P < .0005$. The association was small (Cohen, 1988), Cramer's $V = .133$. Therefore, we can reject the null hypothesis and accept the alternative hypothesis.

2.4 Discussion

This research study connects aspects of the research fields social media, knowledge representation, and image indexing in order to investigate the outcomes of tagging behavior of Instagram users. In respect of the research model, 1,000 Instagram user postings, consisting of a picture and a textual author description with hashtags, were analyzed in terms of their hashtag frequencies, hashtag categories, and picture categories.

On average, the Instagram users applied about 15 hashtags per picture. A closer consideration shows that the categories Pet, Fashion, and Landscape received between 1 and 4 more hashtags on average. Instagram users assigned the fewest tags to the person-related categories Selfie and Friends. The average hashtag number goes in line with previous research. An average 12.41 hashtags per picture (based on 400 pictures and 4,966 hashtags) was identified by Veszelszki (2016) in her study.

The representations of ofness and aboutness are important aspects for Instagram users. They tagged their pictures predominantly with Content-related hashtags (about 60%). Likewise as in our results, such picture describing hashtags (called thematizing context-marker hashtags) were the most frequent (83%) in the sample of Veszelszki's (2016) study. Beside ofness and aboutness, indexing refers in the classical sense of knowledge representation also to aspects regarding to the Isness of a document; however, in professional information services, there are specific bibliographic fields including Isness aspects. This is also reflected in this study since the relative frequency of Isness hashtags was the second most. It is noticeable that only a few emotional hashtags were assigned, although social media services such as Instagram are characterized by social and emotional interactions. The usage of Sentences hashtags is generally weak. Especially, "Insta"-Tags are often assigned to pictures of the category Pet. The same applies for Fakeness tags and pictures of the category Captioned Photo. A small statistical association between hashtag categories and picture categories was found.

Content analysis on Instagram was already realized in several studies as mentioned in the state of research. The key strength of this study is its diversity. It is the first study which evaluates Instagram not only with respect to a single topic or hashtag, but rather in a wider field of both picture categories and hashtags. In doing so, it reveals a general insight into the tagging behavior of Instagram users.

Data collection and sampling of social media data can be challenging. Therefore, the generalizability of these results is subject to the following limitations. Since Instagram offered only limited access to the data at the time of data collection, the data set is based on a non-probability sampling instead of a sample which represents the whole population. However, this problem is well-known in social media research and is not limited to this study. Furthermore, the study only investigated pictures and the respecting account owners' picture descriptions. Besides pictures, videos are part of Instagram content as well.

This study shows how Instagram users tag their photos. By the results of this work, further studies regarding users' intentions are enabled. At this point, we can draw some conclusions from the results. The proportion of Content-related tags is possibly so high since users finally want to find their photos by content. If someone wants to see something about, for example, travel places, he or she would use appropriate hashtags of this topic as search arguments. In order to make pictures detectable, the description of their content by hashtags is necessary. Which role do the rest of the hashtag categories play? An explanation for the generally weak hashtag use of Sentences could be that users consider those hashtags as too specific. Perhaps they think that nobody else searches for those tags and prefer more frequently used tags in order to get more attention for their postings. The phenomenon of "Insta"-Tags is platform-specific and—to the best of the authors' knowledge—not on any other platform as a similar phenomenon in use. Why and how was it established by users? Especially, "Insta"-Tags in the category Pet are frequent. Does this have particular reasons? Considering all 100 Pet postings, it is striking that for similar picture motives chosen hashtags were consimilar. Websites and apps listing top Instagram hashtag lists (in general and for specific topics in order to receive more interactions) are well known[2]. Those lists also include "Insta"-Tags. Maybe the account owners of the observed postings for the category Pet especially used those pre-built hashtag sets. This could explain the huge amount of similar hashtags as well as the huge amount of "Insta"-Tags in the category Pet. However, the use of "Insta" is not limited to pet related matters, so it is questionable what the exact reasons are. Why are Emotiveness hashtags so less frequently assigned, although social media is full of emotional content? For the category Art, this could be explained by the fact that the majority of analyzed artwork pictures are original works, created by the posting authors. Foreign art could evoke more emotions. Picture postings about Friends and Pets could have received the most emotional hashtags because they represent a relationship between entities causing emotions in us. However, the next step should be to prove these hypotheses.

To gather further insights about tagging behavior, the parameters of this study could be broadened. Besides pictures, videos could be analyzed, too. Are there any differences in tagging behaviors of pictures and videos on Instagram? Could it be interesting to distinguish between three main aspects; tags which only apply to the picture or video, tags which only apply to the descriptive content, and tags which apply to both? Are there any gender-specific or cultural differences concerning tagging behavior on Instagram?

Furthermore, it would be interesting to investigate how tagging behavior affects the success of an Instagram account. Do "successful" users (i.e., users with many

[2]e.g., `https://www.tagsforlikes.com/`

followers) tag their pictures differently than less successful users? To what extent does tagging behavior influence success on Instagram in general? Therefore, it is necessary to define which factors are responsible for success on Instagram. Possible factors could be the count of followers, likes per postings, comments, favored pictures, editing of pictures and videos, or social interaction with other users, stating only a few.

The findings of this paper show how Instagram users tag their pictures; however, the analysis of the reasons for their tagging behavior was not part of this study. In order to understand the tagging behavior more deeply, a holistic theory that also accounts for tagging motivations is required.

Acknowledgments

I would like to thank Anika Mehlem for creating and providing a program for the data collection as well as Franziska Zimmer for her work as coder during the content analysis.

References

Allem, J.-P., Escobedo, P., Chu, K. H., Cruz, T. B., & Unger, J. B. (2017). Images of little cigars and cigarillos on Instagram identified by the hashtag #Swisher: Thematic analysis. *Journal of Medical Internet Research*, *19*(7), e255.

Andalibi, N., Ozturk, P., & Forte, A. (2017). Sensitive self-disclosures, responses, and social support on Instagram: The case of #depression. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1485–1500). New York: ACM.

Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon.

Beaudoin, J. (2007). Folksonomies: Flickr image tagging: Patterns made visible. *Bulletin of the American Society for Information Science and Technology*, *34*(1), 7–11.

Buarki, H., & Alkhateeb, B. (2018). Use of hashtags to retrieve information on the web. *The Electronic Library*, *36*(2), 286–304.

Carrotte, E. R., Prichard, I., & Lim, M. S. C. (2017). "Fitspiration" on social media: A content analysis of gendered images. *Journal of Medical Internet Research*, *19*(3), e95.

Coelho, R. L. F., de Oliveira, D. S., & de Almeida, M. I. S. (2016). Does social media matter for post typology? Impact of post content on Facebook and Instagram metrics. *Online Information Review*, *40*(4), 458–471.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.

Daer, A. R., Hoffman, R., & Goodman, S. (2015). Rhetorical functions of hashtag forms across social media applications. *Communication Design Quarterly Review*, *3*(1), 12–16.

Dennis, B. M. (2006). Foragr: Collaborative tagged photographs and social information visualisation. In *Proceedings of the Collaborative Web Tagging Workshop at WWW 2006,* Edinburgh, UK.

Denton, E., Weston, J., Paluri, M., Bourdev, L., & Fergus, R. (2015). User conditional hashtag prediction for images. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1731–1740). New York: ACM.

Dorsch, I., Zimmer, F., & Stock, W. G. (2017). Image indexing through hashtags in Instagram. *Proceedings of the Association for Information Science and Technology*, *54*(1), 658–659.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115.

Enser, P. G. B. (2008). Visual image retrieval. *Annual Review of Information Science and Technology*, *42*(1), 1–42.

Giannoulakis, S., & Tsapatsoulis, N. (2015). Instagram hashtags as image annotation metadata. In R. Chbeir, Y. Manolopoulos, I. Maglogiannis, & R. Ralhajj (Eds.), *Artificial Intelligence Applications and Innovations IFIP Advances in Information and Communication Technology* (Vol. 458, pp. 206–220). Cham: Springer.

Gibbs, M., Meese, J., Arnold, M., Nansen, B., & Carter, M. (2015). #Funeral and Instagram: Death, social media, and platform vernacular. *Information Communication & Society*, *18*(3), 255–268.

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208.

Guidry, J. D., Messner, M., Jin, Y., & Medina-Messner, V. (2015). From #mcdonaldsfail to #dominossucks: An analysis of Instagram images about the 10 largest fast food companies. *Corporate Communications: An International Journal*, *20*(3), 344–359.

Halavais, A. (2014). Structure of Twitter: Social and technical. In K. Weller, A. Bruns, J. Burgess, N. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 29–41). New York: Peter Lang.

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, *1*(1), 21–48.

Holmberg, C., Chaplin, J. E., Hillman, T., & Berg, C. (2016). Adolescents' presentation of food in social media: An explorative study. *Appetite*, *99*, 121–129.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. In *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014* (pp. 595–598). Palo Alto, CA: AAAI Press.

Ingwersen, P. (2002). Cognitive perspectives of document representation. In H. Bruce, R. Fidel, P. Ingwersen, & V. P. (Eds.), *Proceedings of the 4th Inter-*

national Conference on Conceptions of Library and Information Science (pp. 285–300). Greenwood Village, CO: Libraries Unlimited.

Instagram. (2010). *Instagram launches.* Retrieved from https://instagram-press.com/blog/2010/10/06/instagram-launches-2/

Instagram. (2018). *Instagram statistics.* Retrieved from https://instagram-press.com/our-story

Jörgensen, C. (2002). Image access, the semantic gap, and social tagging as a paradigm shift. In J. Lussky (Ed.), *Proceedings of the 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research.* Milwaukee, WI.

Jörgensen, C. (2003). *Image retrieval: Theory and research.* Lanham, MD: Scarecrow Press.

Kipp, M. E. I. (2006). @toread and cool: Tagging for time, task, and emotion. *Proceedings of the 17th Annual ASIS&T SIG/CR Classification Research Workshop*.

Kipp, M. E. I., & Campbell, D. G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. *Proceedings of the ASIST Annual Meeting*, *43*.

Knautz, K. (2012). Emotion felt and depicted. consequences for multimedia retrieval. *Indexing and Retrieval of Non-Text Information*, 343–375.

Kremerskothen, K. (2012). *An amazing 8 years.* Retrieved from http://blog.flickr.net/en/2012/02/10/an-amazing-8-years/

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks: Sage Publications Inc.

Lancaster, F. W. (2003). *Indexing and abstracting in theory and practice.* Champaign, IL: University of Illinois, Graduate School of Library and Information Science.

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (2009). Codebook development for team-based qualitative analysis. In *The content analysis reader* (pp. 211–219).

Marcus, S.-R. (2016). Thinspiration vs. thicksperation: Comparing pro-anorexic and fat acceptance image posts on a photo-sharing site. *Cyberpsychology*, *10*(2).

Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, *28*(1), 38–43.

Messina, C. (2007). *How do you feel about using # (pound) for groups. As in #barcamp [msg]? chrismessina (Twitter).* Retrieved from https://twitter.com/chrismessina/status/223115412

Mohammad, S. M., & Kiritchenko, S. (2015). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, *31*(2), 301–326.

Nashmi, E. A. (2018). From Selfies to Media Events: How Instagram users interrupted their routines after the Charlie Hebdo shootings. *Digital Journalism*, *6*(1), 98–117.

Nov, O., Naaman, M., & Ye, C. (2008). What drives content tagging: The case of photos on flickr. In *Conference on Human Factors in Computing Systems - Proceedings* (pp. 1097–1100).

Oh, C., Lee, T., Kim, Y., Park, S., & Suh, B. (2016). Understanding participatory hashtag practices on Instagram: A case study of Weekend Hashtag Project. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1280–1287). New York: ACM.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological review*, *97*(3), 315–331.

Panofsky, E. (1955). *Meaning in the visual arts.* Garden City: Doubleday Anchor Books.

Pegoraro, A., Comeau, G. S., & Frederick, E. L. (2018). #SheBelieves: the use of Instagram to frame the US women's soccer team during #FIFAWWC. *Sport in Society*, *21*(7), 1063–1077.

Peters, I. (2009). *Folksonomies. Indexing and retrieval in web 2.0.* Berlin: De Gruyter Saur.

Peters, I., & Stock, W. G. (2007). Folksonomy and information retrieval. In *Joining Research and Practice: Social Computing and Information Science. Proceedings of the 70th ASIS&T Annual Meeting* (Vol. 44, pp. 1510–1542). Hoboken: Wiley.

Pila, E., Mond, J. M., Griffiths, S., Mitchison, D., & Murray, S. B. (2017). A thematic content analysis of #cheatmeal images on social media: Characterizing an emerging dietary trend. *International Journal of Eating Disorders*, *50*(6), 698–706.

Pluzhenskaia, M. (2006). Folksonomies or fauxsonomies: How social is social bookmarking? In *Advances in Classification Research Online* (Vol. 17, pp. 23–24).

Rasmussen, E. M. (1997). Indexing images. In *Annual Review of Information Science and Technology* (Vol. 32, pp. 169–196).

Rondeau, S. (2014). The life and times of aboutness: A review of the library and information science literature. *Evidence Based Library and Information Practice*, *9*(1), 14–35.

Rorissa, A. (2010). A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology*, *61*(11), 2230–2242.

Schmidt, S., & Stock, W. G. (2009). Collective indexing of emotions in images. A study in emotional information retrieval. *Journal of the American Society for Information Science and Technology*, *60*(5), 863–876.

Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging and Classification Quarterly*, *6*(3), 39–62.

Shatford Layne, S. (1994). Some issues in the indexing of images. *Journal of the American Society for Information Science*, *45*(8), 583–588.

Shatford Layne, S. (2002). Subject access to art images. In *Introduction to art image access: Issues, tools, standarts, strategies* (pp. 1–20). Los Angeles, CA: Getty Research Institute.

Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, *58*, 89–97.

Siebenlist, T. (2013). Emotionale Suche – Emotionales Information-Retrieval. In D. Lewandowski (Ed.), *Handbuch der Internet-Suchmaschinen 3: Suchmaschinen zwischen Technik und Gesellschaft* (pp. 299–327). Heidelberg: Akademische Verlagsgesellschaft.

Smith, G. (2004). *Folksonomy: Social classification.* Retrieved from http://atomiq.org/\%0Aarchives/2004/08/folksonomy_social_classification.html

Souza, F., De Las Casas, D., Flores, V., Youn, S., Cha, M., Quercia, D., & Almeida, V. (2015). Dawn of the selfie era: The whos, wheres, and hows of selfies on instagram. In *COSN 2015 - Proceedings of the 2015 ACM Conference on Online Social Networks* (pp. 221–231).

Stock, W. G., & Stock, M. (2013). *Handbook of information science.* Berlin, Boston: De Gruyter Saur.

Stvilia, B., & Jörgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in Flickr. *Journal of the American Society for Information Science and Technology*, *61*(12), 2477–2489.

Turner, J. M. (1995). Comparing user-assigned terms with indexer-assigned terms for storage and retrieval of moving images: Research results. *Proceedings of the 58th Annual Meeting of the American Society for Information Science*, *32*, 9–12.

Vander Wal, T. (2007). *No title.* Retrieved from http://www.vanderwal.net/folksonomy.html

Veszelszki, Á. (2016). Time, truth, tradition: An image-text relationship on instagram: Photo and hashtag. In *In the Beginning was the Image: The Omnipresence of Pictures: Time, Truth, Tradition* (pp. 139–150).

Yanbe, Y., Jatowt, A., Nakamura, S., & Tanaka, K. (2007). Can social bookmarking enhance search in the web? In *Proceedings of the ACM International Conference on Digital Libraries* (pp. 107–116).

Zhang, Y., Hashim, N. H., Baghirov, F., & Murphy, J. (2018). Gender differences in Instagram hashtag use. *Journal of Hospitality Marketing and Management*, *27*(4), 386–404.

Chapter 3

Hashtags on Instagram: Self-created or Mediated by Best Practices and Tools?

Social media enables conversations mediated through documents as texts, audio, images, or videos. Likewise, hashtags became an essential medium for social media communication. Instagram is well-known as one of the current platforms for hashtagging. This exploratory study investigates how hashtags used on Instagram became established in respect of self-creation and best practices or tools. The analysis is based on data obtained from an online survey (N = 1,006) of Instagram users. 55.7% of the respondents use hashtags on Instagram. Only self-created hashtags are assigned by 41.4%, whereas 58.6% are (sometimes) inspired by others. Best practices and tools based on friends/other users or Instagram functions are more frequently used in contrast to offers from influencers or third-parties (e.g. guides, hashtag-sets). Furthermore, the majority does not intentionally use false hashtags. This study enables a first overview of the Instagram users' hashtagging creation behavior and selection process.

3.1 Introduction

Conversation on social media is mediated by documents (e.g. text, audio, image, and video) and content-describing user-generated tags. On Instagram, documents are especially images, videos, and ephemeral moments, while their content is described by hashtags like for example #vacation if the posting is about vacations or shows vacation-related content. Therefore, it is not only important to know, for example, what kind of hashtags exist, when they are applied, or what motivations stand behind (hash)tagging, but also how they became established. Do Instagram users rather rely on self-created hashtags or do they apply best practices and tools for hashtagging? Little is known about the self-creation and creation by best practices and tools of hashtags, although such an analysis would add another important dimension to the overall hashtag and hashtagging analysis. Since hashtags serve as a further mean to communicate in social media it is important to follow this question. This exploratory study exactly intends to answer this question for hashtagging behavior in Instagram and sheds light on the hashtagging creation behavior and selection process on Instagram.

3.1.1 Background: Hashtagging on Instagram

Social media enables their users to communicate through a variety of channels using textual, audio, visual, or mixed media. Platforms as Twitter or Instagram also allow for using hashtags. A hashtag is composed of a #-sign and a character string (e.g. #happytobehear). "[By] using the # character to mark particular keywords, [...] users communicate a desire to share particular keywords folksonomically" (Halavais, 2014). Its idea originates from a Twitter Tweet by Chris Messina in 2007 (Messina, 2007). After this posting several platforms implemented a hashtag function and hashtagging on social media established (Halavais, 2014); still having its roots in the area of social tagging and user-generated knowledge representation (Peters, 2009; Peters & Stock, 2007; Vander Wal, 2007).

Nowadays, hashtags are integrated in our online and offline environment. As parts of online conversations or content captions, they play a role in understanding the interplay between media producers and media audiences. In such a computer-mediated discourse, hashtags indicate a shift "from online conversation to 'searchable talk'" (Dickinson, 2013; Sheldon, Herzfeldt, & Rauschnabel, 2019; Zappavigna, 2011). Moreover, they are also "an instrument for creative self-expression and language play" (Heyd & Puschmann, 2017, p. 51), started as a "rather peripheral typographic resource" and evolving "to an emblem of social media linguistic practice" (Heyd & Puschmann, 2017, p. 52).

The application Instagram is well-known for posting visual media, like for example picture postings, videos or so-called ephemeral moments known as "stories." The app's monthly active users numbered more than 1 billion as of June 2019 (Instagram, 2019b).

Research incorporating the analysis of hashtags in Instagram is diverse. In more general studies, hashtags serve as a filter to receive material for content analysis. For example, in order to receive access to political debates of Clinton/Trump supporter during the 2016 US presidential election, Schmidbauer et al. (2018) analyzed Instagram postings collected through 16 hashtags related to the topic.

Several studies have investigated motivational factors for using hashtags on Instagram based on the Uses and Gratifications (U&G) theory. The six motivations self-presentation, chronicling, inventiveness, information seeking, venting, and etiquette were identified by Erz et al. (2018).

Sheldon et al. (2017) compared American (individualist culture) and Croatian (collectivistic culture) undergraduate students' motivations of Instagram use. They point out self-promotion, social interaction, diversion, documenting, and creativity as gratifications influencing Instagram use for both cultures. According to the usage of hashtags, the American students use them mainly for documentation purpose.

Similarly, the Croatian students do, however they further use them for other reasons like self-promotion, social interaction, and creativity.

Motivational factors for hashtagging on Instagram, but also on other social media services as Twitter, Facebook, and LinkedIn, were the object of analysis in another study (composed of six empirical sub-studies), too. Here, the authors identified ten different motivations for the use of hashtags, namely amusing, organizing (especially on Twitter and Instagram), designing, conforming, trendgaging, bonding, inspiring, reaching, summarizing (especially on Instagram), and endorsing (Rauschnabel, Sheldon, & Herzfeldt, 2019).

Besides studies investigating motivational factors, further hashtagging behavior is also of interest. An analysis of social tagging dynamics in respect of tag adoption patterns and the variety of tagging usage on Instagram reveals that the examined media (based on a dataset containing 9 million hashtags) amounts in a power law distribution, with the majority of media tagged with a few tags (Ferrara, Interdonato, & Tagarelli, 2014).

The descriptive power of hashtags was observed in a further study. Based on 1,000 Instagram pictures, certain hashtags, assigned by the picture posting owners were compared with hashtags study participants should choose from a list to describe the picture. For 66% of the chosen hashtags study participants would choose the same hashtags as the picture owners did in order to describe the content of the picture (Giannoulakis & Tsapatsoulis, 2016).

A hashtag content analysis of 14,649 hashtags from 1,000 Instagram images (divided into ten picture categories) shows, with 60,2% the majority of analyzed hashtags refers to content-related aspects of the posting, visible in the pictures. In contrast, hashtags communicating emotions or being full sentences are less often assigned (Dorsch, 2018). Based on this a follow-up study on gender-specific hashtagging came to similar overall results and further found out that females tend to use more emotional hashtags, whereas males assign hashtags related to non-topical features, like technical aspects of the picture such as the camera type or Instagram related hashtags (e.g. #IGers, #instagood) (Philipps & Dorsch, 2019).

Focusing on gender hashtagging differences of Malaysian food-related postings, the analysis of 1,382 Instagram images tagged with the hashtag #Malaysianfood reveals women tend to use emotional and positive hashtags in comparison to male users (Zhang, Hashim, Baghirov, & Murphy, 2018).

The following studies consider more specific aspects of hashtagging. The hashtag #like4like, belongs to the group of performativeness hashtags which means those hashtags are calling for an action (Dorsch, 2018; Peters & Stock, 2007). Similarly, #goout or #followmearound, #like4like not only calls for the action to like a user's posting(s), but also promises to give (a) back-like(s) to those who liked the posting(s) (Peters & Stock, 2007; Zhang, Ni, Han, & Pang, 2017). According to the hashtagging behavior on Instagram, it has been found that #like4like does not really provoke more likes (after 2013) and that, contradictory to what the hashtag mediates, most users do not like back (Zhang et al., 2017).

An analysis of 18,366 Instagram picture postings tagged with #nofilter unveils that other as indicated 12% indeed used a filter for their image (Santarossa, Coyne, & Woodruff, 2017). Thus hashtag can mediate false information, too.

Using participatory hashtag practices (Oh, Lee, Kim, Park, & Suh, 2016) some people promote their followers to upload images which match a suggested hashtag. A well-known example is Instagram's Weekend Hashtag Project (#WHP).

There exist a variety of health- and disease-related hashtags on Instagram. For example, obvious hashtags addressing a certain disease (e.g. #endometriosis) or aspects such as symptoms/sequelae of a disease (e.g. #selfharm). When searching after #selfharm, Instagram currently pops up a "Can we help?" message, explaining that posts with this hashtag "often encourage behavior that can cause harm and even lead to death" and the option to show posts, cancel the request or receive help (Instagram, 2019a). Looking back, there were also times when – at some point – such hashtags were totally banned for the users. Consequently, variations (as for this case #selfharmm, #selfharmmm) emerged (Moreno, Ton, Selkie, & Evans, 2016). As a further result, a sub-group of *secret hashtags* established, especially used by adolescents (Metcalfe, Beesley, Watson, Sarwari, & Clayton, 2018; Moreno et al., 2016). They are ambiguous or encrypted, so that they are less often or less quickly blocked. Those hashtags function as an entry to find (mental) health communities or conversations and offers the opportunity of being part of them. Apart from that, their users have to fear to be banned by Instagram to a lesser extent (Dapper Goat Social Media, 2019). Those restrictions are not limited to health-related hashtags, but rather concerns a broader field of Instagram hashtags (Dapper Goat Social Media, 2019).

The previously stated cases have underlined in which ways hashtagging on Instagram is part of mediated online conversation. Likewise, they show which effects they can have and which consequences or restrictions might arise.

3.1.2 Best Practices and Tools

According to the latest definition of the Oxford dictionary (Oxford University Press, 2019) best practices are "[c]ommercial or professional procedures that are accepted or prescribed as being correct or most effective." Similarly, it is defined by the Cambridge dictionary (Cambridge University Press, 2019) as "a working method or set of working methods that is officially accepted as being the best to use in a particular business or industry." Consequently, best practices can be seen as the result of a comparative process of an action with any alternative courses, whereby the action is linked to some outcome or goal (Bretschneider, Marc-Aurele, & Wu, 2005). Following these definitions and considering its origin from the field of economics, the application of the concept expanded over time. Nowadays, best practices are adapted in a variety of areas.

With the rise of social media, best practices incorporating or even solely for this scope evolved. For instance, a short Google search reveals that there are plenty of best practice guides and advice for the usage of social networking services like Facebook or Instagram. Staying with Instagram, they exist for special aspects or functions of the app as well. Entire guides and tools exclusively address *how to hashtag* on Instagram. From stating more universal advice, over specific user or company centered best practices to providing “trending hashtags” or hashtag sets (focusing on a specific purpose or topic), the offer is manifold diverse. The information can be embedded in news or blog postings, but there also exist YouTube videos, entire websites or applications (also listing or generating hashtag-sets and hashtag statistics), all dedicated to hashtagging best practices and tools for this purpose.

Table 3.1: Pre-defined hashtag set for Instagram postings addressing dog-related topics by the app Tagstagram.

DOGS
#tagstagram #dog #dogs #dogsofinstagram #dogoftheday #dogstagram #dogsofinstaworld #dogs_of_instagram #doglover #doge #doggy #doglovers #dogs_of_world #doggie #doggo #dogsofinsta #instadog #dogsandpals #mansbestfriend #doggie

For example, Table 3.1 shows a pre-defined hashtag set for dog content postings provided by the iOS app Tagstagram. Within the app, users have the function “Copy Tags” to copy a whole hashtag set and to mark sets as “Favorite” (BYOApps LLC, 2019). Equivalent to this is the android app Tagify (Dev, 2019) or one of the many other applications or websites which come across with such a feature. One of the first hashtagging set providers might be TagsForLikes, but with the latest update in 2013 it is no longer up-to-date (TagsForLikes, 2013). Nevertheless, according to its founder, TagsForLikes were used by over 5 million users in more than 129 million pictures (Mohiuddin, n.d.).

As a second example and having a closer look on some up-to-date guides (e.g. Aynsley, 2019; Gilbert, 2019), they often provide some general reasons for using hashtags on Instagram, state what changed and what is currently important to know (like with these examples such guides often refer to a specific year for which they are “valid”), state concrete “tips and tricks” like “[m]ake sure that hashtag means what you think it means,” (Aynsley, 2019) or to better avoid popular Instagram hashtags like #like4like, #followme, or #tagsforlikes since they attract the wrong audience (“bots, spammers, and others users who have no intention of engaging with you in any meaningful way” (Aynsley, 2019)). Thereby, often is their main focus to increase the engagement on Instagram.

In our study, the term “Best Practices” means all suggested and well-established practices; “Tools” summarizes all available instruments for hashtagging. We distinguish between the following kinds of best practices and tools:

- Pre-defined hashtag-sets,

- Hashtag statistics,

- Hashtag guidelines,
- Direct hashtags recommendations by friends,
- Direct hashtag recommendations by influencers,
- Other users' or friends' hashtags seen on Instagram,
- Autocompletion on Instagram,
- Number of hashtags on Instagram,
- Instagram hashtag search function, and
- User's own practices, criteria, and experiences.

3.1.3 Best Hashtagging Practices and Tools on Instagram

Hashtags are essential parts of social media communication. As the literature review shows, there exist studies about hashtags on Instagram in general and specifically on certain hashtagging behaviors and motivations, but there is still need to examine how hashtags on Instagram are created by their users.

In order to get to know more about the general role of hashtagging in computer-mediated conversation it is also important to analyze how hashtags are inspired by other influential factors as best hashtagging practices and tools and to what extent users rely on their own when using hashtags. Therefore, this exploratory study concentrates on answering the following research questions:

RQ1: How many Instagrammers use hashtags and for what?

RQ2: How are best hashtagging practices and tools used on Instagram and to what extent do users create hashtags on their own?

RQ3: Do Instagram users intentionally assign false hashtags?

We focus on Instagram as platform to be analyzed since it is one of the current platforms which is, similar to Twitter, strongly used for hashtagging in social media (Pennington & Spiteri, 2019) and nowadays "the most prominent hashtagging platform" (Sheldon et al., 2019, p. 7). Furthermore, the possibility to annotate up to 30 hashtags as well as the establishment of new hashtags make the application especially interesting for this analysis. For example, hashtag phenomena as #like4like or even an entire platform-specific hashtag category called "Insta"-Tags (e.g. #IGers; #insta + any kind of term, like #instaartists) established on Instagram (Dorsch, 2018; Zhang et al., 2017).

Based on our research questions and the literature background the following research model (Figure 3.1) underlays our study. Referring to Instagram, a *user* has the possibility to *index content* to be published with hashtags (e.g. Instagram pictures, videos, etc.) or otherwise just to add a caption or post the content without

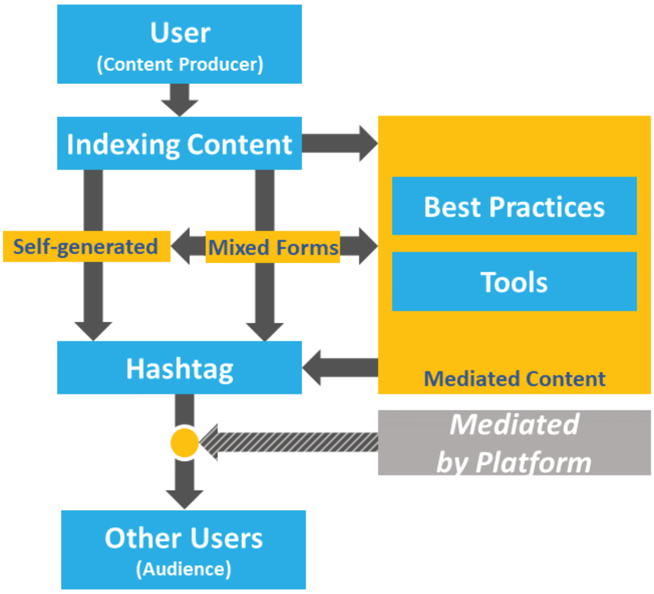


Figure 3.1: Research model.

any further textual information. If adding *hashtags* to the posting, they can be *self-generated*, based on *mediated content*, or based on *mixed forms* of both prior stated variations. The hashtagged content will then be published on Instagram and is available for its *audience* – other Instagram users. Thereby, the process of hashtagging contributions normally happens during the publishing process, however, for some content like pictures or videos, it is also possible that users add their hashtags past publishing in the comment section or modify, add, or delete them from their caption. Some hashtags are also *mediated* by the platform needing a confirmation to be displayed with their indexed content or being totally blocked for the users.

3.2 Methods

In order to answer the research questions for this exploratory study, a questionnaire for an online survey was developed. This approach was chosen to receive first general insights about this topic. Since the information questioned cannot be obtained directly from Instagram users' profiles, a survey provides a means of asking users about their preferences and behaviors. Although, experiments, field research, or interviews allow for gathering data on such user behavior, they were not suitable for this study. Experiments have the advantage to be able to exactly note down every behavioral aspect, but in the same way, the setting would have caused a man-made labor situation, not referring to the reality. Directly observing users within a field research context would involve certain problems. For example, people probably might not allow to observe them in such a context. Besides, both of these approaches would rather illustrate a current situation for specific cases as reveal more general statements. Interviews employ a solid starting point for exploratory studies. However, based on the existing literature and study structure, we choose an online survey setting. Furthermore, this approach allows us to reach a wide audience and to generate enough empirical data for the analysis.

3.2.1 Survey

Overall, the questionnaire is composed of 26 items (see Appendix). However, as can be obtained, each of these items does not apply to every survey participant. The first questions and notes (3.-9.) refer to a user's general Instagram usage and account settings. Questions 10. and 11. ask for the general usage and importance of hashtags. If participants stated to use hashtags the next question 12. concerns the area of application, followed by asking if one uses only self-created hashtags or if they are (sometimes) inspired by others (13.). Depending on the answer further questions concentrate on the application (14.-18.) or non-application (19.-22.) of Instagram hashtagging best practices or tools. The questionnaire concludes with items 23. to 26., surveying demographic statements and providing space for final comments.

The survey initially was composed in English language, followed by a German translation. Overall, it is based on our research model, the knowledge of the current state of research as well as the best practices and tools.

In total 7 persons (2 = male; 5 = female) pre-tested the survey and the received feedback where considered for or integrated in the survey. Umfrageonline[1] was chosen to host the survey.

3.2.2 Data Collection, Cleansing, & Analysis

The distribution of the survey took place during May 12, until June 5, 2019. According to a non-probability sampling, we digitally and physically promoted the survey participation. Therefore, the survey was spread via social media (Instagram, Reddit, Facebook, and Twitter), online forums, e-mail listservs, via flyers in Germany (mainly at places in Düsseldorf and Leverkusen), and Switzerland (Buchs and Chur).

During the data collection, we received the feedback that the gender denotation "divers" in the English questionnaire should be better changed into "other." This was done on May 22, 2019.

It was necessary to clean up the data gathered from our survey. Form the overall 1,338 participants, 1,156 completed the questionnaire. Since in question 3. 145 respondents selected not to use Instagram, 1,011 remain. However, five of them could still not be considered for the analysis. Two indicated in question 4. to have "0" Instagram account(s). Two further participants were removed as well because they selected one and two for their age and one due to problems with the software of

[1] `https://www.umfrageonline.com`

Umfrageonline. Therefore, the data cleansing results in a total of 1,006 participants, considered for the analysis.

Due to the inclusion of filter questions (see Appendix), the overall number of respondents varies for some questions. For a better understanding Tables 3.2 - 3.5 therefore include an “overall N” indicating how many persons answered this question overall. These also incorporate respondents which selected the statement “No experience” or “I don’t know” which is not further analyzed within these tables.

3.3 Results

In our study (N = 1,006), 63.1% of Instagram users identify as female, 33.2% as male, and 2.8% as other. The remaining 0.9% preferred to make no statement about their gender. Although the worldwide gender distribution on Instagram is more balanced (52% female and 48% male) (Statista, 2019), women generally are more likely to fill out online surveys (Smith, 2008).

Based on the age groups identified by Fietkiewicz (2017) 52% of our participants belong to Generation Y (born between 1980 and 1996), 42.7% to Generation Z (born after 1996), 5% to Generation X (born between 1960 and 1980), and 3% to the so-called “Silver Surfers” (born before 1960).

From a total of 55 countries, the majority come from Germany (37.7%) and the United States (35.1%), followed by Canada (5.8%), the United Kingdom (5.3%), Australia (2.3%), and the Netherlands (1.3%).

With 64.8% nearly two-thirds report to have one Instagram account, whereas 23.4% have two accounts and 11.7% three or up to seven accounts. However, one person indicated to have 18 accounts.

Considering the visibility settings for Instagram accounts the distribution is nearly fifty-fifty. This means 50.1% have their profile open to the public and conversely, 49.9% set their profiles on private mode, only allowing confirmed users to see their postings.

Having a closer look on the follower numbers, 61.8% respondents have between 100 and 999 Instagram follower, followed by 31.0% respondents with between 1 and 99; 5.2% respondents with between 1,000 and 9,999; 0,7% respondents with between 10,000 and 99,999 respondents; and 0.3% respondents with between 100,000 and 499,999 followers. To have 0 Instagram follower was selected by 1%.

Nearly all use their account only by themselves (98.4%) and just a very small amount shares it with further persons (1.6%).

RQ1: How many Instagrammers use hashtags and for what?

When analyzing the usage of hashtagging best practices and tools, it is also important to look at how many Instagram users make use of hashtags. More than half (55.7%) of the 1,006 respondents use hashtags on Instagram. Table 3.2 shows how important hashtags for searching postings as well as for using them in own contributions posted on Instagram are. 20.7% indicated they are rather important for

Table 3.2: Hashtag importance for searching contributions by other Instagram users and for the usage within own contributions.

7	6	5	4	3	2	1	Median	IQR	N
Searching contributions by others (e.g. pictures, videos, etc.)									
11.5%	16.0%	20.7%	13.4%	11.4%	10.9%	16.3%	4	4	958
Usage within your own contributions (e.g. pictures, videos, etc.)									
10.2%	11.2%	11.5%	13.8%	11.8%	13.0%	28.6%	3	4	934

7 = Very important, 6 = Important, 5 = Rather important, 4 = Neutral, 3 = Rather unimportant, 2 = Unimportant, 1 = Very unimportant

Overall N = 1,006

searching, whereas 28.6% find them very unimportant for their own contributions. Although, it becomes apparent that the opinions differ and have a width range, also demonstrated by the interquartile range (IQR) of 4 and the medians. The inclusion of all participants for this question might be a reason for such a broad distribution.

Table 3.3: Frequency of hashtag usage for Instagram activities.

7	6	5	4	3	2	1	Median	IQR	N
For my picture postings									
41.1%	23.3%	22.5%	4.8%	6.6%	0.9%	0.7%	6	2	559
For my video postings									
30.5%	15.2%	18.4%	8.6%	10.9%	1.4%	14.9%	5	4	348
For my story postings									
6.7%	6.7%	26.0 %	15.9%	17.1%	2.6%	25.0%	4	3,75	416
For my profile description									
1.4%	3.3%	6.7%	6.1%	5.5%	4.1%	72.9%	1	1	491
In comments to my own postings									
1.8%	4.5%	10.9%	12.1%	17.6%	1.4%	51.6%	1	3	494
In comments to other postings									
1.4%	2.6%	17.2%	12.0%	18.1%	2.0%	46.7%	3	3	507

7 = Very frequently, 6 = Frequently, 5 = Occasionally, 4 = Rarely, 3 = Very rarely, 2 = I only tried it once, 1 = Never

Overall N = 560

There is a clear trend for the frequency of hashtag usage for different activities on Instagram (Table 3.3). Especially own picture and video postings are (very) frequently hashtagged. In contrast, the clear majority of 72.9% never used hashtags for their profile description. Likewise, is this for hashtags in comments. Here, 51.6% never use them in own comments to their posting and with 46.7% this applies for comments to other postings.

RQ2: How are best hashtagging practices and tools used on Instagram and to what extent do users create hashtags on their own?

Considering all participants using hashtags (N = 560), 41.4% of them only use self-created hashtags. The remaining 58.6%, and thus the majority, reported to create their hashtags (sometimes) inspired by others and not solely by their own. Following this, participants using inspired hashtags were asked how important certain hashtagging best practices and tools are for them and how frequently they use them. The frequency of applying hashtagging best practices or tools can be obtained from Table 3.4. Overall, these results are quite similar to the ones found for the importance. Therefore, we focus in more detail only on one of these two aspects, namely the frequency.

For the usage of pre-defined hashtag-sets (47.9%), hashtag statistics from third-parties (65.2%), overall hashtag guidelines (54.5%), and direct hashtag recommendations or practices received from influencers (48.4%) the proportional majority stated to never use those. Direct hashtag recommendations or practices by friends (26.3%), users'/friends' hashtags directly seen on Instagram (30.3%), Instagram's hashtag autocompletion function (24.7%), the information about the number of times a hashtag is assigned on Instagram (23.4%), and the Instagram hashtag search function (24.1%) are mostly used occasionally (again based on the proportional majority). Especially for the last stated 4 aspects, Instagram hashtag functions and users'/friends' hashtags, the usage can be considered as rather frequent since many also stated to use it frequently or even very frequently. As indicated in Table 3.4, this is also supported by the medians and IQRs.

Besides, the respondents not only utilize best practices or tools, but rely frequently on own practices, criteria, and experiences (33.1%) as well, when using hashtags. Likewise, many use them occasionally (23.2%) or very frequently (21.2%).

Referring to the 232 participants who stated to create their hashtags only by themselves 70.3% answered for question 20. they don't know about the existence of best practices and tools for hashtagging and following 29.7% know about such practices and tools. Based on this, only a few participants answered in question 21. that they ever informed themselves about such practices (18.8%; N = 13). Here the majority (81.2%; N = 56) clearly does not inform themselves about such practices.

RQ3: Do Instagram users intentionally assign false hashtags?

Both, the Instagram users stated to be (sometimes) inspired by others when creating their hashtags and the users only using self-created hashtags relatively similar utilize intentionally false hashtags as can be seen in Table 3.5. Clearly, the majority don't use such hashtags (72.5% and 83.3%). Users inspired by others are slightly more including intentionally false hashtags in their hashtagging behavior.

Table 3.4: Frequency of using hashtagging best practices or tools for the own hashtagging behavior.

7	6	5	4	3	2	1	Median	IQR	N
Usage of pre-defined hashtag-sets (from web pages or apps)									
7.4%	11.6%	16.4%	7.1%	6.8%	2.9%	47.9%	2	4	311
Hashtag statistics (from third-parties)									
2.3%	5.6%	7.9%	7.2%	7.9%	3.9%	65.2%	1	2	305
Overall hashtag guidelines (provided on blogs, websites, Youtube, etc.)									
4.5%	3.5%	11.6%	10.6%	10.6%	4.5%	54.5%	1	3	310
Direct hashtag recommendations or practices by friends									
5.1%	12.7%	26.3%	10.8%	16.1%	4.4%	24.7%	4	3	316
Direct hashtag recommendations or practices by influencers									
5.4%	7.0%	15.2%	10.8%	8.9%	4.4%	48.4%	2	4	316
Based on users'/friends' hashtags directly seen on Instagram									
9.9%	14.6%	30.3%	10.5%	15.3%	1.6%	17.8%	5	2	314
Autocompletion hashtag recommendation on Instagram									
21.5%	21.8%	24.7%	8.5%	7.0%	2.5%	13.9%	5	2	316
Number of times a hashtag is assigned on Instagram									
14.9%	18.8%	23.4%	12.5%	6.9%	2.0%	21.5%	5	3	303
Instagram hashtag search function									
13.7%	21.0%	24.1%	14.9%	8.3%	1.9%	16.2%	5	3	315
Based on my own practices, criteria and experiences									
21.2%	33.1%	23.2%	10.3%	3.5%	0.6%	8.0%	6	1	311

7 = Very frequently, 6 = Frequently, 5 = Occasionally, 4 = Rarely, 3 = Very rarely, 2 = I only tried it once, 1 = Never

Overall N = 328

Table 3.5: Frequency of using intentionally false hashtags.

7	6	5	4	3	2	1	Median	IQR	N
Frequency of using intentionally false hashtags by users (sometimes) inspired by others for hashtag creation									
2.8%	4.1%	6.6%	6.9%	4.4%	2.8%	72.5%	1	1	320
Frequency of using intentionally false hashtags by users only self-creating hashtags									
0%	1.8%	3.5%	2.6%	6.6%	2.2%	83.3%	1	0	227

7 = Very frequently, 6 = Frequently, 5 = Occasionally, 4 = Rarely, 3 = Very rarely, 2 = I only tried it once, 1 = Never

Overall N = 232

3.4 Discussion

Our exploratory study investigated the application of hashtagging best practices and tools on Instagram in contrast to the self-creation of hashtags. In doing so, it shows how far best practices and tools influence the establishment of hashtags. For the study, we conducted an online survey with 1,006 participants.

In respect to **RQ1**, more than half of the respondents (55.7%) use hashtags on Instagram. The importance of searching after contributions with hashtags or using them in own contributions on Instagram generally varies but is slightly more important for searching with hashtags. Considering the frequency of application, hashtags are predominantly used in the picture and video postings, whereas profile descriptions, and comments to own or other postings are generally not or rarely assigned with hashtags.

The results for **RQ2** show that 41.4% of Instagram users use self-created hashtags only, whereas 58.6% are (sometimes) inspired by others. Interestingly, best practices or tools connected to an Instagram user's friend (or other users), and functions offered by Instagram itself are rather frequently used in contrast to Influencer or third-party offers.

Using intentionally false hashtags as questioned in **RQ3** is something the most have never done. This applies similarly to users (sometimes) inspired by best practices and tools as for those only using their own hashtags.

The overall results have shown that the hashtagging creation behavior and selection process is not only solely based on self-created hashtags or certain best-practices and that best practices appealing more personal or directly referring to Instagram are more frequently used. Instagram users might rely on more personal and individual inspiration in order to give their hashtags a more personal expression. One participant also gave us the feedback that his or her hashtags often serve as a medium to communicate jokes or to add ironic or humoristic elements to the posting (e.g. #hashtag) and that this is more important to him or her instead of making the contributions searchable. For the participant, such language plays function much better as hashtag than jokes within the normal text caption [personal communication]. Likewise, some other respondents stated to use them to make fun or jokes.

The data of this study is limited to a non-probability sample. It is not possible to create a sample based on the total Instagram population since such data is not freely accessible. However, this problem is – unfortunately – well-known in social media research and not limited to this study.

This study generated first insights about Instagram users' hashtag self-creation and the creation mediated by best practices and tools; it helps to analyze hashtags as part of social media communication. Our questionnaire also contained a few basic open questions. The conduction of further in-depth interviews could provide more detailed structured information for a better understanding of these first findings and broaden the study scope. For example, users could be asked about their motivations to apply or not apply best practices and tools and if they do so, looking back to their behavior, achieve to satisfy their motivations. Interviews also enable the possibility to exactly address the self-creation behavior or specific best practices and tools suggested by best practice guides, recommendations, etc. addressed in this study. Besides, a closer examination could identify if there exist gender-, age- or country-specific differences within the application and in how far account aspects like the number of followers, profile visibility or number of account administrators influence best hashtagging practices and tools.

Focusing on the platform, it is important as well to analysis in how far the hashtag mediation by Instagram affects the online communication through hashtags. Furthermore, there is also need to analyze how hashtags are perceived by their audience in more detail. For example, the best practice to hide hashtags in the comment section or caption of picture and video contributions. How is this perceived by the users and what does it communicate to them?

References

Aynsley, M. (2019). *The 2019 Instagram hashtag guide — How to use them and get results.* Retrieved from `https://blog.hootsuite.com/instagram-hashtags/`

Bretschneider, S., Marc-Aurele, F. J., & Wu, J. (2005). 'Best practices' research: A methodological guide for the perplexed. *Journal of Public Administration Research and Theory*, *15*(2), 307–323.

BYOApps LLC. (2019). *Tagstagram - Hashtag generator.* Retrieved from `https://apps.apple.com/us/app/tagstagram-hashtag-generator/id587756737`

Cambridge University Press. (2019). *Best practice.* Retrieved from `https://dictionary.cambridge.org/dictionary/english/best-practice`

Dapper Goat Social Media. (2019). *Banned Instagram hashtags 2019.* Retrieved from `https://www.dappergoat.com/blog/banned-instagram-hashtags-2019/`

Dev, G. (2019). *Tagify - Best hashtags for Instagram.* Retrieved from `https://play.google.com/store/apps/details?id=com.gongadev.hashtagram&hl=de`

Dickinson, P. (2013). "B/w U & me": The Functions of Formulaic Language in Interactional Discourse on Twitter. *The Linguistics Journal*, *7*(1), 7–38.

Dorsch, I. (2018). Content description on a mobile image sharing service: Hashtags on Instagram. *Journal of Information Science Theory and Practice*, *6*(2), 46–61.

Erz, A., Marder, B., & Osadchaya, E. (2018). Hashtags: Motivational drivers, their use, and differences between influencers and followers. *Computers in Human Behavior*, *89*(July), 48–60.

Ferrara, E., Interdonato, R., & Tagarelli, A. (2014). Online popularity and topical interests through the lens of Instagram. In *The Proceedings of the 25th ACM Conference on Hypertext and Social Media* (pp. 24–34). Santiago, Chile: ACM.

Fietkiewicz, K. (2017). Jumping the digital divide: How do 'silver surfers' and 'digital immigrants' use social media? *Networking Knowledge*, *10*(1), 5–26.

Giannoulakis, S., & Tsapatsoulis, N. (2016). Evaluating the descriptive power of Instagram hashtags. *Journal of Innovation in Digital Ecosystems*, *3*(2), 114–129.

Gilbert, S. (2019). *New rules for Instagram hashtags: How to find the right hashtags for your business.* Retrieved from `https://later.com/blog/instagram-hashtags/`

Halavais, A. (2014). Structure of Twitter: Social and technical. In K. Weller, A. Bruns, J. Burgess, N. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 29–41). New York: Peter Lang.

Heyd, T., & Puschmann, C. (2017). Hashtagging and functional shift: Adaptation and appropriation of the #. *Journal of Pragmatics*, *116*, 51–63.

Instagram. (2019a). *Can we help?* Retrieved from `https://www.instagram.com/explore/tags/selfharm/?hl=en`

Instagram. (2019b). *Instagram statistics.* Retrieved from `https://instagram-press.com/our-story/`

Messina, C. (2007). *How do you feel about using # (pound) for groups. As in #barcamp [msg]? chrismessina (Twitter).* Retrieved from `https://twitter.com/chrismessina/status/223115412`

Metcalfe, H., Beesley, A., Watson, C., Sarwari, H., & Clayton, J. (2018). #SpeakingTheirLanguage: Engaging with young people's mental health in a digital landscape. *Clinical Psychology Forum*(307), 9–14.

Mohiuddin, Z. (n.d.). *Products.* Retrieved from `https://www.zaheer.me/`

Moreno, M. A., Ton, A., Selkie, E., & Evans, Y. (2016). Secret Society 123: Understanding the language of self-harm on Instagram. *Journal of Adolescent Health*, *58*(1), 78–84.

Oh, C., Lee, T., Kim, Y., Park, S., & Suh, B. (2016). Understanding participatory hashtag practices on Instagram: A case study of Weekend Hashtag Project. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1280–1287). New York: ACM.

Oxford University Press. (2019). *Best practice.* Retrieved from `https://en.oxforddictionaries.com/definition/best{\_}practice`

Pennington, D. R., & Spiteri, L. (Eds.). (2019). *Social tagging in a linked data environment.* London, UK: Facet Publishing.

Peters, I. (2009). *Folksonomies. Indexing and retrieval in web 2.0.* Berlin: De Gruyter Saur.

Peters, I., & Stock, W. G. (2007). Folksonomy and information retrieval. In *Joining Research and Practice: Social Computing and Information Science. Proceedings of the 70th ASIS&T Annual Meeting* (Vol. 44, pp. 1510–1542). Hoboken: Wiley.

Philipps, J., & Dorsch, I. (2019). Gender-specific tagging of images on Instagram. In *Social Computing and Social Media. Design, Human Behavior and Analytics. HCII 2019. Lecture Notes in Computer Science, Vol. 11578.* (pp. 396–413). Springer.

Rauschnabel, P. A., Sheldon, P., & Herzfeldt, E. (2019). What motivates users to hashtag on social media? *Psychology and Marketing*, *36*(5), 473–488.

Santarossa, S., Coyne, P., & Woodruff, S. J. (2017). Exploring #nofilter images when a filter has been used. *International Journal of Virtual Communities and Social Networking*, *9*, 54–63.

Schmidbauer, H., Rösch, A., & Stieler, F. (2018). The 2016 US presidential election and media on Instagram: who was in the lead? *Computers in Human Behavior*, *81*, 148–160.

Sheldon, P., Herzfeldt, E., & Rauschnabel, P. A. (2019). Culture and social media: The relationship between cultural values and hashtagging styles. *Behaviour and Information Technology*(in press), 1–13.

Sheldon, P., Rauschnabel, P. A., Antony, M. G., & Car, S. (2017). A cross-cultural comparison of Croatian and American social network sites: Exploring cultural differences in motives for Instagram use. *Computers in Human Behavior*, *75*, 643–651.

Smith, W. G. (2008). Does gender influence online survey participation? A record-linkage analysis of university faculty online survey response behavior. *Eric Ed501717*.

Statista. (2019). *Distribution of Instagram users worldwide as of April 2019, by gender.* Retrieved from `https://www.statista.com/statistics/802776/distribution-of-users-on-instagram-worldwide-gender/`

TagsForLikes. (2013). *#TagsForLikes.* Retrieved from `https://www.tagsforlikes.com/`

Vander Wal, T. (2007). *No title.* Retrieved from `http://www.vanderwal.net/folksonomy.html`

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media and Society*, *13*(5), 788–806.

Zhang, Y., Hashim, N. H., Baghirov, F., & Murphy, J. (2018). Gender differences in Instagram hashtag use. *Journal of Hospitality Marketing and Management*, *27*(4), 386–404.

Zhang, Y., Ni, M., Han, W., & Pang, J. (2017). Does #like4like indeed provoke more likes? In *Proceedings of the International Conference on Web Intelligence (WI'17)* (pp. 179–186). Leipzig, Germany.

3.5 Appendix

#	Question/Note	Answer items	Likert
1.	Please choose a language.	English/German	
2.	Cover and survey note	Includes the following statements: Welcoming, topic, purpose, required time, data analysis and privacy, expression of thanks, contact information	
3.	Do you use Instagram? *	yes/no *[direction to the end page]*	
4.	How many Instagram accounts do you have? *	Dropdown menu: 0-20, 20+ *[0 = direction to the end page]*	
5.	Intermediate text page *[Appears only if selected "more than one account" in question 4.]*	"Please note, since you have **more than one Instagram** account, please focus on all further questions on the Instagram account you post content the **most**."	
6.	Is your Instagram account open to the public or private? *	open to the public/private (a request is necessary)	
7.	How many follower does your Instagram account have? *	Dropdown menu: 0/1-99/100-999/1,000-9,999/10,000-99,999/100,000-499,999/500,000-999,999/>=1,000,000	
8.	Please indicate, what applies to your Instagram account: *	I'm the only person using this account/It is a shared account with additional persons using this account	
9.	Intermediate text page *[Appears only if selected "shared account" in question 8.]*	"Please note, since you are **sharing the Instagram account**, please focus on all further questions on your **own** (inter)actions and behavior with this shared Instagram account."	
10.	Do you use hashtags on Instagram? * *Note: Hashtags are a composition of # and a character string, for example #summer*	yes/no *[direction to question 23.]*	
11.	How important are hashtags on Instagram to you? *	for searching contributions by others (e.g. pictures, videos, etc.) | for the usage within your own contributions (e.g. pictures, videos, etc.)	x
12.	For which of your activities on Instagram do you use hashtags and how frequently? *	for my picture postings | for my video postings | for my story postings | for my profile description | in comments to my own postings | in comments to other postings	x
13.	How do you create your hashtags? *	they are (sometimes) inspired by others (e.g. friends, recommendations, guides, etc.)/they are only self-created *[direction to question 19.]*	
14.	How important are best practices or tools for your own hashtagging behavior on Instagram? *	usage of pre-defined hashtag-sets (from web pages or apps) | hashtag statistics (from third-parties) | overall hashtag guidelines (provided on blogs, websites, Youtube, etc.) | direct hashtag recommendations or practices by friends | direct hashtag recommendations or practices by influencers | based on users'/friends' hashtags directly seen on Instagram | autocompletion hashtag recommendation on Instagram | number of times a hashtag is assigned on Instagram | Instagram hashtag search function | based on my own practices, criteria and experiences	x
15.	How frequently do you use the following best practices or tools within your own hashtagging behavior on Instagram? *	*[same items as in 14.]*	x
16.	How frequently do you intentionally use false hashtags on Instagram – for instance – to receive more views, likes, followers, etc.? *	very frequently/frequently/occasionally/rarely/very rarely/I only tried it once/never/I don't know	
17.	Which specific best hashtagging practices or tools do you use?	Blank box	
18.	Do you have a specific reason why you apply best hashtagging practices or tools?	Blank box *[direction to question 23.]*	
19.	How frequently do you intentionally use false hashtags on Instagram – for instance – to receive more views, likes, followers, etc.? *	very frequently/frequently/occasionally/rarely/very rarely/I only tried it once/never/I don't know	
20.	Do you know that there exist best hashtagging practices and tools (e.g. asking friends, recommendations, guides, etc.)? *	yes/no *[direction to question 23.]*	
21.	Did you ever inform yourself about such best practices or tools? *	yes/no *[direction to question 23.]*	
22.	Do you have a specific reason why you don't apply best hashtagging practices or tools?	Blank box	
23.	What is your age? *	Dropdown menu: 1-99	
24.	What is your gender? *	female/male/other/prefer not to say	
25.	Where are you currently located? *	Dropdown menu: Country list	
26.	Do you have any additional comment you want to share with us?	Blank box	

***Notes**: # in the same box were displayed at the same time; * = required questions; comments in [] contain information about the query sequence*

Chapter 4

Gender-Specific Tagging of Images on Instagram

Instagram is widely known and used as a social media application for visual content. In order to categorize and describe their posted content as well as to make it retrievable, users can assign hashtags to each posting. What kind of hashtags do female and male Instagram users assign to their picture postings? Which differences and similarities exist? This study analyzes genderspecific image tagging behavior on Instagram. Therefore, a content analysis of, in total, 14,951 hashtags from 1,000 Instagram pictures (respectively 500 pictures posted by female and male users) was performed. The subjects of the 1,000 Instagram pictures belong to overall ten picture categories (100 pictures per category): Activity, Architecture, Art, Captioned Photo, Fashion, Food, Friends, Landscape, Pet, and Selfie. Seven categories exist for the coding of the hashtags: Content-relatedness, Emotiveness, Fakeness, "Insta"-Tags, Isness, Performativeness, and Sentences. On average, women assigned 14 hashtags to their postings, whereas men used one hashtag more. For both genders, hashtags belonging to the category Content-relatedness were the most used (over 55% of assigned hashtags). Second most assigned (over 17%) were Isness related hashtags. Generally, females used slightly more emotional hashtags, whereas men assigned Isness and "Insta"-Tags in a higher frequency than females. "Insta"-Tags were assigned in high frequencies (over 22%) to Pet pictures by both genders. With under 2%, females and males did not use many Sentences hashtags. As a chi-square test of independence shows, there exists a small statistical association between hashtag and picture categories for male and female Instagram users, respectively.

4.1 Introduction

The well-known phrase "men are from Mars, women are from Venus" (Gray, 1992) exemplifies the impression that several people have regarding the degree of differences between both genders. Can distinct practices also be observed in a social media context with the primary focus on image tagging behavior? If yes, to what extent do these differences occur? What hashtags do men and women choose to describe their uploaded content and to what kind of content do they assign these hashtags?

The objective of this study is to answer these questions. In the field of knowledge representation, collaborative services allow the use of tags to index a document and make it retrievable. Tags can be applied by human indexers or automatically by the services and some need to follow rules (Stock & Stock, 2013). The ones that are freely assigned constitute folksonomies (I. Peters, 2009) which play an important role in the Web 2.0 and for platforms like Twitter and Instagram. Instagram is a popular mobile social photo and video sharing application which allows its users to apply up to 30 different hashtags to their postings. We found out that in several picture categories (e.g. Activity, Friends, or Pets) there are indeed significant differences, but similarities as well, in how male and female Instagram users apply different kinds of hashtags to their postings (for instance, tags related to the content, Emotiveness, or Fakeness).

4.1.1 Social Tagging and Gender-Specific Tagging on Social Media

When internet users advanced from consumers to prosumers of knowledge (Toffler, 1980), they started to "index" their produced content by means of social tagging to make it retrievable for other users in the respective collaborative online environment (Trant, 2009). These indexing terms are called "tags" which do not follow any guidelines (Stock & Stock, 2013) and form a folksonomy "[...] for each collaborative information service comprised of each individual user's tags" (I. Peters, 2009, p. 1). The concept "folksonomy" consists of "folk" as well as "taxonomy," and originates from Vander Wal who was quoted by Smith in a blog entry (Vander Wal, 2007). The term "taxonomy" is misleading though (Stock & Stock, 2013); unlike a taxonomy, folksonomies have no hierarchical structure (Laniado, Eynard, & Colombetti, 2007).

In addition to tags, hashtags exist which have the same function as tags. A hashtag begins with the # symbol, followed by a string of characters, e.g. #guineapig. "Initially, the hashtag was used within Internet chat rooms" (van den Berg, 2014, p. 4). On August 23th in 2007, Chris Messina suggested in a tweet: "how do you feel about using # (pound) for groups. As in #barcamp [msg]?" (Messina, 2007) and thus, for the first time, used the hashtag in a different context (van den Berg, 2014) which helped to establish it in various social network systems subsequently (Halavais, 2014). Both, tags as well as hashtags function as user-generated metadata for posted content (for example, visual media like photos).

Since the first tagging systems emerged in 2003 (M. E. I. Peters I.and Kipp et al., 2017), user tagging behavior and motivation was analyzed by various researchers (e.g. Daer, Hoffman, & Goodman, 2015; Dubinko et al., 2007; Golder & Huberman, 2006). Gender-specific tagging behavior in folksonomies was studied especially on Twitter. For instance, Cunha, Magno, Almeida, Gonçalves, and Benevenuto (2012) examined if male and female Twitter users applied different hashtags when talking about the same topic. Women tended to use more common hashtags for all topics chosen by the researchers. In the political debate, females chose more personal hashtags (first person singular), while males used more persuasive hashtags (third person imperative forms).

Holmberg and Hellsten (2015) researched gender tweeting behavior in the climate change debate. They noted that "[m]any of the hashtags represent[ed] a very general level of metadata describing the content or context of the tweet" (Holmberg & Hellsten, 2015, p. 816). Male twitter users employed politics-related hashtags, but rarely used tags related to climate change or to general environmental issues. Furthermore, they used more descriptive tags. Females employed hashtags connected to campaigns and online movements associated with climate change. They used more specific hashtags, e.g. by referring to a specific event, campaign, or person.

Shapp (2015, para. 1-2) differentiates two hashtag categories in Twitter: tag hashtags, which "are used to connect with a larger discussion and/or community" and commentary hashtags which "are not intended to affiliate widely, and are meant to be interpreted within the local context of the tweet." The results of the study show that female twitter users choose commentary hashtags more often than men. Contrarily, men use tag hashtags to a greater extent.

4.1.2 Instagram

Instagram is a mobile picture and video sharing application created by Kevin Systrom and Mike Krieger. Since April 2012 it is owned by Facebook (Instagram, 2012a). Launched in the beginning of October 2010 for iOS exclusively (Instagram, 2010), an Android version of the app was released in April 2012 (Instagram, 2012b). Since then, Instagram became a very popular social network, has over 1 billion users as of January 2019, and is still growing. Over 500 million active users are online daily (Instagram, 2019). Instagram enables its users to upload pictures and videos, as well as to share stories or create slideshows while using photo filters and applying hashtags and geotags to their posts.

4.1.3 Research on Instagram

Topics investigated on Instagram include for example hashtag utilization, content analyses, or motivation to use the social network. The following section presents an overview of works conducted by numerous researchers on Instagram.

Sheldon and Bryant (2016) surveyed 239 college students to figure out the main motivations for using Instagram. A comparison between Croatian and American students suggested that motivation for using Instagram is culture-independent, although the app was employed for different reasons (Sheldon, Rauschnabel, Antony, & Car, 2017). Evaluating 212 Instagram users, Lee, Lee, Moon, and Sung (2015) concluded that there are five social and psychological motives for using the app: social interaction, archiving, self-expression, escapism, and peeking.

Online expression of six emotions and their perceived appropriateness on Instagram and three other platforms was analyzed by Waterloo, Baumgartner, Peter, and Valkenburg (2018). Instagram scored as the least appropriate platform to express negative emotions and highly appropriate to express positive emotions. Men and women differed in their rating. Psychological subjects were investigated on Instagram as well. Holland and Tiggemann (2017) studied eating disorders and compulsive exercises in women. Lup, Trub, and Rosenthal (2015) pointed out that Instagram use may lead to psychologically feeling unwell because of negative social comparisons, amongst other things. Non-suicidal self-injury behavior was examined by Brown et al. (2018).

Olympic athletes' self-presentation by gender was investigated by Geurin- Egleman and Burch (2016) to research their use of Instagram as a communication and personal brand marketing tool. Analyzing 2,017 images of two football teams by categorizing them into product and non-product-related brand attributes, Anagnostopoulos, Parganas, Chadwick, and Fenton (2018) also investigated how Instagram is used as a tool for branding and how fans reacted in the comments to gain new insights for sport marketing. Lavoie (2015) examined the branding strategy of Dunkin' Donuts on Instagram. The text captions and the image content of 12 postings were analyzed by sorting them into categories like presence of a call to action in the textual data and color, products, or people information in the image data. Coelho, de Oliveira and de Almeida (2016) measured five post types against their likes and comments (advertising, fan, events, information, and promotion). They explored business profiles promoting food, hairdressing, ladies' footwear, body design, and fashion gym wear.

Fitness inspiration and body images are also Instagram research topics. Carrotte, Prichard, and Lim (2017) observed gender differences after a content analysis of 415 Instagram, Facebook, Twitter, and Tumblr postings, e.g. women were more likely to be sexualized than men. The postings derived predominantly from Instagram (360 posts in total). Talbot, Gavin, van Steen, and Morey (2017) inspected 734 pictures tagged with #thinspiration, #fitspiration, and #bonespiration as well as their top five alternative hashtags. 600 pictures were categorized in the context for body type, activity, objectification, and textual data in a content analysis conducted by Tiggemann and Zaccardo (2016). Santarossa, Coyne, Lisinski, and Woodruff (2016) analyzed 10,000 posts and 122 images tagged with #fitspo.

Researchers observed political debates on Instagram. For example, Schmidbauer, Rösch, and Stieler (2018) examined postings collected by crawling 16 hashtags related to Clinton/Trump supporters and opponents during the 2016 US presidential election. 9,000 multilingual hashtags were studied by Lee and Chau (2018) in the context of the Umbrella Movement in Hong Kong in 2014. Coding them into the categories language, fact, opinion, and emotion, the researchers found that, in addition to stating facts and opinions, Instagram users also were quite emotional about the political movement.

Santarossa, Coyne and Woodruff (2017) investigated 18,366 images with the hashtag #nofilter and concluded that 12 percent of these in fact used a filter. Mostly, women applied this hashtag and the subject of these pictures was mainly a person.

Giannoulakis and Tsapatsoulis (2016) studied the descriptive power of hashtags by analyzing 1,000 images. They concluded that 66% of the hashtags their study participants chose were identical with the ones the picture owner used. Moreover, half of these hashtags referred to the depicted image content. Oh, Lee, Kim, Park, and Suh (2016) explored the use of participatory hashtags on Instagram in context with the Weekend Hashtag Project (#WHP). Participatory hashtags are recommended by some users to their followers to promote uploading pictures using these certain hashtags.

A content analysis of 1,382 Instagram posts tagged with #Malaysianfood showed gender-specific utilization of hashtags (Zhang, Hashim, Baghirov, & Murphy, 2018). Female users employed more often than men emotional and positive hashtags, while male users showed higher use in informative and negative hashtags for pictures. Additionally, the researchers found a positive correlation between the number of hashtags and followers as well as likes.

The following studies form the base of this work. The first in-depth analysis about Instagram was conducted by Hu, Manikonda, and Kambhampati (2014). This empirical study's results encompassed eight popular image categories (friends, food, gadget, captioned photo, pet, activity, selfie, fashion) and five distinct types of users. Almost half of the analyzed photos belonged to the categories friends and fashion.

Dorsch (2018) researched tagging behavior of Instagram users in a wider scope. She distinguished between ten picture categories, namely Activity, Architecture, Art, Captioned Photo, Fashion, Food, Friends, Landscape, Pet, and Selfie, and seven hashtag categories (Content-relatedness, Emotiveness, Isness, Performativeness, Fakeness, "Insta"-Tags, and Sentences). The results showed that Instagram users predominantly tag images with hashtags relating to the content, followed by hashtags relating to Isness (Dorsch, 2018; Dorsch, Zimmer, & Stock, 2017).

4.1.4 Gender-Specific Tagging of Images on Instagram

Research that takes gender-specific image tagging on Instagram into account is still at its beginning. This study aims to contribute to the limited research on gender-specific Instagram tagging behavior in a broader scope. Conducting a content analysis (Krippendorff, 2004), hashtags of images from ten different picture categories are coded into seven distinct hashtag categories. The research questions of this work are:

RQ1: Are there any gender-specific differences in the relative hashtag frequencies in the picture categories?

RQ2: Given a picture category, what is the gender-specific distribution of hashtag categories; and given a hashtag category, what is the gender-specific distribution of picture categories?

RQ3: Are there any gender-specific associations between picture categories and hashtag categories?

4.2 Methods

Content analysis (Krippendorff, 2004) is a technique originating from the social sciences and investigates content of any form, e.g. texts, images (as in this study), recordings, or movies. It utilizes different procedures that conceptualize content depending on its context – a process called coding which produces new understanding and data to analyze. The content analysis in this study was performed on posts of Instagram images to analyze gender-specific tagging behavior. As a result, the pictures and hashtags were coded into specifically designed picture and hashtag categories, based on Dorsch's two codebooks (Dorsch, 2018).

4.2.1 Codebooks

Generally, codes map the content of the text to the model that the analyst constructed and therefore generate new information that can be analyzed. The process of coding needs to fulfill explicit guidelines that define text boundaries identifying with a specific code (MacQueen, McLellan, Kay, & Milstein, 2009). Such guidelines and codes form a codebook. Thereby, a codebook "[...] always reflects the analyst's implicit or explicit research questions" (Krippendorff, 2004, p. 2013). As mentioned, the two existing codebooks, one for categorizing pictures and one for classifying hashtags were used for this content analysis. The structure of the codebooks is simple and consists of six distinct components: the code, a brief description of the code, the full definition, guidelines when and when not to use it as well as examples. In addition to the six mentioned components, a seventh one exists in the picture categories codebook, namely the corresponding hashtag for the specific category. This format is based on MacQueen et al.'s (2009) recommendations.

Picture Categories

In total, ten Instagram picture categories were used, and each category has their respective hashtag. This study only considered pictures that were tagged with the corresponding picture category hashtags (4.1). Another requirement to be included in the content analysis was that the picture predominantly shows one category or fits the preference rules (Dorsch, 2018; Dorsch et al., 2017). The picture categories are based on research of Hu et al. (2014). The authors conducted a cluster analysis and used computer vision techniques in combination with two coders to identify popular Instagram photo categories and user types. Seven of the eight photo categories were chosen for the codebook, namely Activity, Captioned Photo, Fashion, Food, Friends, Pet, and Selfie. Four of these categories were modified as well, namely Activity, Captioned Photo, Friends, and Selfie. Three supplementary categories (Architecture, Art, and Landscape) were added. The category Gadget was omitted because pictures of this category overlapped with the other categories due to its broad definition (Dorsch, 2018).

The pictures in the category Activity depict indoor and outdoor activities (e.g. climbing, biking), as well as locations where activities take place (e.g. concerts). The

Table 4.1: Picture categories and their respective hashtags.

Category	Hashtag
Activity	#activity
Architecture	#architecture
Art	#art
Captioned photo	#quote
Fashion	#fashion
Food	#food
Friends	#friends
Landscape	#landscape
Pet	#pet
Selfie	#selfie

photos in architecture show subjects related to the categories' name like buildings, structures, and cityscapes. Art is the category for pictures showing art-related content in all forms, e.g. paintings, sculptures, tattoos, or crafted art. Pictures that show graphics with quotes belong to the Captioned Photo category. As the name indicates, the category Fashion depicts fashion-related matters like makeup or clothes. The category Food includes photos displaying food, drinks, or recipes. Images in the Friends category depict at least one person that is a friend of the user who uploaded the picture; group of friends fall into this category, too. Pictures with landscapes and nature-related content are sorted into the Landscape category. The images in the Pet category show cats, dogs, and similar animals. Self-taken pictures (selfies) belong to the category Selfie (Dorsch, 2018; Dorsch et al., 2017).

Hashtag Categories

Coding Instagram hashtags into categories which were designed to represent specific concepts (Dorsch, 2018; Dorsch et al., 2017), can lead to new results about gender-specific image tagging behavior. Overall, seven categories exist for the coding of hashtags: Content-relatedness, Emotiveness, Fakeness, "Insta"-Tags, Isness, Performativeness, and Sentences.

There are nine common coding rules: A hashtag can only be assigned to one of these categories. It must originate from the picture's caption and be in English (proper names are excluded from this rule). The picture's caption may be used to obtain additional information about the content of the picture. Spelling mistakes can be ignored. Hashtags with only one or more emojis and special characters are excluded from the analysis. Hashtag abbreviations need to be figured out by the coders. Detached tags, for example #Belgian #Shepherd, are categorized separately. If the same hashtag is assigned multiple times, it is counted only one time (Dorsch, 2018).

The category Content-relatedness is assigned to hashtags that refer "[. . .] in any form to content-related aspects in the picture" (Dorsch, 2018, p. 3). This includes the concepts ofness and aboutness (Shatford, 1986) which are based on Panofsky's (1955) three levels of art interpretation. For example, if the picture depicts a dog, the hashtags #pet, #dog, or #animal refer to the content and would therefore be coded into the category Contentrelatedness. Hashtags which are emotional or describe feelings, are coded into the category Emotiveness, e.g. #love or #sad. Considered basic emotions (love, happiness, fun, surprise, aspiration, sadness, anger, disgust, fear, and shame) originate from Siebenlist (2013). Fakeness is the category for "[...] intentionally wrongly chosen statements" (Dorsch, 2018, p. 7), e.g. when the tag #cat was assigned to a photo of a dog. Hashtags with any form of "insta," "gram," or other abbreviation of these terms, are sorted into the category "Insta"-tags, e.g. #instalike or #petstagram. These kinds of hashtags are a phenomenon specific to Instagram (Dorsch, 2018; Dorsch et al., 2017). Isness is the category for hashtags which represent technical aspects of the picture that are not depicted, like #landscapephotographer, #selfie, or #throwbackthursday. The concept of isness was introduced by Ingwersen (2002). Hashtags that call for actions are coded into the category Performativeness, e.g. #followforfollow, #explore, or #kickit (I. Peters & Stock, 2007). Performativeness derived from Austin (Austin, 1962). Hashtags with a participatory function are for example used in various projects on Instagram (Oh et al., 2016). Complete sentences are coded into the category Sentences, for example #lifeisgood or #thisislondon. Abbreviations of sentences like #tgif – thank God it's friday, are coded into this category as well (Dorsch, 2018; Dorsch et al., 2017).

4.2.2 Data Collection

Instagram pictures were mainly collected from November 2017 to January 2018. The dataset for the content analysis consists of 1,000 Instagram pictures, 500 from male users and 500 from female users, respectively. A multimethodological approach was chosen to collect these pictures. Due to the official Instagram API's access limits, the JSON response of the Instagram website was automatically downloaded and processed with a PHP-script. The script takes a top picture category hashtag and the number of pictures to be downloaded as parameters and saves the most recent pictures of the declared hashtag with metadata.

After the automatic data collection, the picture subject and gender of the picture's owner needed to be identified manually. For checking if the picture conformed to the category, the Instagram picture category codebook (Dorsch, 2018) was used (by two persons). Pictures that failed to comply with the codebook rules, e.g. a picture tagged with #selfie but depicting more than one person and therefore not matching the requirements, were deleted. The gender of the picture's owner was obtained with the help of the profile picture, the user biography, and the posts. The profile picture, full name, and biography are optional information a user can add to his or her profile. The biography can contain the user's full name and a statement about the gender. For example, in Fig. 4.1, Annika is the full name of the user rufinchenx3 and the expression Gamergirl in the biography, as well as the profile picture, indicate that this user is female.

Figure 4.1: Profile of Instagram user rufinchenx3.

There might be special cases a gender is difficult to identify or can't be identified. For example, the gender is harder to determine when pet owners create a profile for their pet and pretend to post and write as their pet. In such situations, it must be determined if the gender of the owner can be derived from the profiles' information. If not, the profile has to be excluded from the analysis.

The replacing of pictures cleaned the automatically crawled data, hence reducing the dataset. Due to the stated data cleansing process, more than 100 pictures for a category were downloaded first. If the manual verification of the data did not result in the number of needed pictures for this analysis (50 pictures of male users, 50 pictures of female users), more data was downloaded and processed in the described way.

All final datasets consist of the following metadata and information: the downloaded picture, the URL to the picture on Instagram, the picture owner's username, full name, biography and gender, all hashtags written in the picture description, the total number of these hashtags, the full caption of the post, the picture category name (e.g. Activity, Pet, etc.), and an ID for naming the picture file, containing the picture owner's gender, the picture category name, and the unique Instagram picture code.

4.2.3 Coding Process

Coding is the process of assigning specified codes to a text. In this work, the hashtags of the Instagram images were coded into seven different categories by using a codebook. After the data collection, the categorization of the 1,000 pictures took place from January to March 2018. Two coders classified the hashtags first separately, following a 4-eyes principle. After that, the intercoder-reliability was estimated using Krippendorff's alpha (Krippendorff, 2011). For every category, the alpha value was ¿.800 in the initial round and ¿.880 in a second round (which has to be performed due to interpretational inconsistencies in the category Fakeness).

After the calculation of the Krippendorff's alpha, the coders needed to recode the hashtags that they classified into different categories. They discussed these differences and agreed to one category. A PHP-script takes the file with the recoded hashtags and one of the coder's files as input parameters. For every picture that was tagged with hashtags which need to be corrected, all hashtag categories are scanned. If the hashtag is in the wrong category, it gets deleted from this category and is written to the right category.

4.3 Results

This chapter presents the results of the content analysis while answering the three research questions.

RQ1. Are there any gender-specific differences in the relative hashtag frequencies in the picture categories?

In total, 14,951 hashtags from 1,000 Instagram photos were coded into the seven hashtag categories. The average number of hashtags per pictures owned by female users was approximately 14 hashtags, whereas males applied over 15 hashtags (Fig. 4.2, Table 2 4.2). Men used significantly more hashtags in the categories Captioned Photo (∅ 21.8 to 12.74 hashtags) and Landscape (∅ 18.84 to 14.52 hashtags). A great difference of approximately nine hashtags was observed for captioned photos. For females, these images received the lowest number of hashtags. In the categories Food (∅ 16.18 to 12.34 tags) and Selfie (∅ 14.75 to 10.72 tags), female users tagged a picture with a significantly higher number of tags.

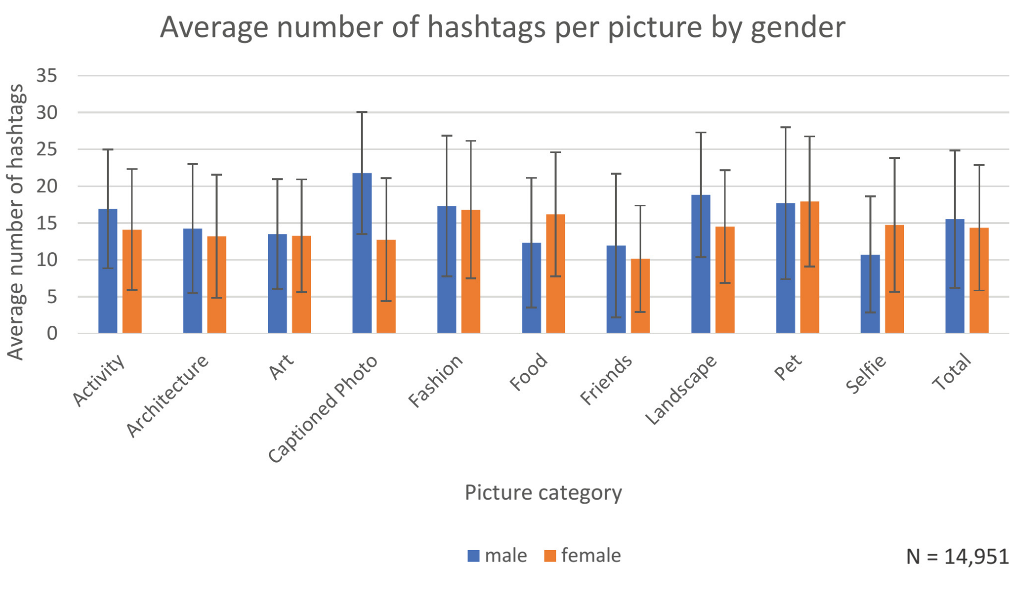


Figure 4.2: Average number of hashtags per picture by gender and standard deviation, sorted by picture category.

For male users, the category Selfie (∅ 10.72 tags) had the least hashtags. The mean values ranged from around 10 to almost 21 hashtags for males, and from over 10 to roughly 18 for females. This shows that the span of hashtag counts by gender varied noticeably in specific picture categories. The standard deviation lay between 7 and 10 for males, and between 7 and 9 for females. Even though this indicates

Table 4.2: Summary of the significant t-test results for the hashtag frequencies in the picture categories, $p \leq .05$.

	Hashtag category		Male (n = 50)	Female (n = 50)	P	t	df	Cohen's d
Captioned photo	Content-relatedness	M	12.64	7.12	$< .001$	5.166	98	1.03
		SD	5.914	4.702				
	"Insta"-Tags	M	2.36	.7	.001	3.583	98	0.72
		SD	2.926	1.474				
	Isness	M	2.28	1.32	.03	2.208	72.372	0.44
		SD	1.99	2.343				
	N	M	21.8	12.74	$< .001$	5.446	98	1.09
		SD	8.283	8.351				
Food	Content-relatedness	M	6.36	8.54	.039	-2.093	98	-0.42
		SD	4.733	5.643				
	N	M	12.34	16.18	.028	-2.23	98	-0.45
		SD	8.794	8.422				
Landscape	Fakeness	M	.72	.04	.02	2.404	49.969	0.48
		SD	1.99	.198				
	Isness	M	6.08	3.66	.001	3.355	71.285	0.67
		SD	4.58	2.246				
	N	M	18.84	14.52	.009	2.68	98	0.54
		SD	8.464	7.635				
Selfie	Emotiveness	M	.44	.86	.047	-2.016	98	-0.4
		SD	.861	1.195				
	N	M	10.72	14.78	.019	-2.387	96.092	-0.48
		SD	7.882	9.083				
Total	Emotiveness	M	.61	.75	.05	-1.964	998	-0.13
		SD	1.021	1.103				
	"Insta"-Tags	M	1.33	1.04	.029	2.191	969.422	0.14
		SD	2.218	1.865				
	Isness	M	2.94	2.52	.023	2.28	974.646	0.14
		SD	3.145	2.691				
	N	M	15.53	14.37	.04	2.056	990.239	0.53
		SD	9.324	8.532				

that the number of hashtags was almost the same for each picture and for both genders, it points out that the male data was more spread out than the female data, specifically in the Pet category with a standard deviation of over 10.

RQ2. Given a picture category, what is the gender-specific distribution of hashtag categories; and given a hashtag category, what is the gender-specific distribution of picture categories?

In total, male users applied 7,766 and female users 7,185 hashtags to their postings (Tables 4.2 and 4.3). Over half of these assigned hashtags were classified as Contentrelatedness tags for both genders (over 55%). The second most assigned category with about 18% was Isness for men and women alike, just like the least assigned category Sentences (under 1.5%). Females tagged pictures with Performativeness tags third most (7.85%) and with "Insta"-Tags (7.25%) fourth most, whereas males did the opposite (8.54% and 7.39%, respectively). Significant differences existed in the hashtag categories Emotiveness, Isness, and "Insta"-Tags. Emotiveness hashtags were assigned more often by females (5.19% in contrast to 3.95%), whereas Isness (18.93% to 17.52%) and "Insta"-Tags (8.54% to 7.25%) more often by men.

The distribution of picture categories to the hashtag category Content-relatedness was between almost 45% to over 70% for both genders. Female users assigned the lowest value of this hashtag category to the picture category Selfie (47.23%). With 44.69% of all hashtags, male users assigned Content-relatedness hashtags the least to

Landscape images which was the overall lowest assigned value of this hashtag category as well. Both genders tagged pictures related to art subjects the highest with those tags (over 70%). In the categories Captioned Photo, male users assigned Contentrelatedness tags in a significantly higher frequency than female users. On the other hand, women assigned significantly more of these tags to pictures depicting food (Table 4.2). Noticeably, Isness tags reached high values in the picture categories with the least assigned Content-relatedness tags by both genders. For example, the category Landscape scored with 32.27% the highest Isness value assigned by men, which is significantly different to the value assigned by women (25.21%). Furthermore, men assigned significantly more Isness hashtags to captioned photos. Males and females both added "Insta"-Tags to the category Pet the most and to the category Activity the least. Men (22.82%) were more likely to use those tags for Captioned Photo (10.83%). This likeness is statistically significant (Table 4.2). The highest percentage of the category Performativeness male users assigned to Architecture is the same value that female users assigned to the category Selfie for those hashtags (11.64%). Both genders distributed performative tags the least to images depicting art-related subjects. The picture categories Pet and Selfie were tagged with those tags more by women, whereas Friends obtained more of those tags by men.

Emotions were not expressed by men regarding the category Art (around 1%), while women assigned the least of these hashtags for Fashion (2.50%). Females tagged pictures of Selfies significantly more emotionally than men (5.82% and 4.1%, respectively). Usually, men got higher values for Fakeness hashtags. The category

Table 4.3: Relative frequencies of hashtag categories by picture categories and gender (N = 1,000 posts; 100 posts per picture category, and respectively 50 posts per gender for each picture category).

	Content-relatedness		Emotiveness		Fakeness		"Insta"-Tags		Isness		Performativeness		Sentences		%	
Gender	m	f	m	f	m	f	m	f	m	f	m	f	m	f	m	f
Activity	67.61	67.52	4.26	4.96	2.13	2.13	2.01	2.27	18.44	14.75	4.85	7.09	0.71	1.28	100.00	100.00
Architecture	49.37	51.97	1.68	2.88	1.54	0.60	6.45	6.36	28.75	26.06	11.64	11.21	0.56	0.91	100.00	100.00
Art	70.37	71.99	1.19	2.56	5.04	3.16	5.04	4.52	15.70	15.21	2.67	2.41	0.00	0.15	100.00	100.00
Captioned Photo	57.98	55.89	4.59	8.16	6.97	8.63	10.83	5.49	10.46	10.36	8.17	10.05	1.01	1.41	100.00	100.00
Fashion	58.03	57.31	1.85	2.50	7.63	6.90	6.24	6.54	15.95	18.79	8.09	6.42	2.20	1.55	100.00	100.00
Food	51.54	52.78	2.43	3.09	4.21	5.07	10.37	7.29	23.99	23.36	6.32	6.55	1.13	1.85	100.00	100.00
Friends	49.58	57.87	8.21	9.84	9.72	5.91	6.70	4.13	16.25	14.17	8.88	5.91	0.67	2.17	100.00	100.00
Landscape	44.69	55.37	2.76	3.72	3.82	0.28	5.10	4.55	32.27	25.21	10.62	10.33	0.74	0.55	100.00	100.00
Pet	49.72	53.46	8.25	9.38	4.52	2.34	22.82	20.31	7.57	5.25	3.95	7.03	3.16	2.23	100.00	100.00
Selfie	49.63	47.23	4.10	5.82	4.85	5.55	7.46	6.50	25.19	22.60	8.58	11.64	0.19	0.68	100.00	100.00
Total	55.03	56.87	3.95	5.19	5.03	4.01	8.54	7.25	18.93	17.52	7.39	7.86	1.12	1.29	100.00	100.00

with the highest percentage of Fakeness hashtags assigned by male users is Friends (9.72%). The category Captioned Photo (8.63%) received the highest frequency of fake tags by women. They tagged Landscape (0.28%) and Architecture (0.60%) pictures very rarely with fake tags, too, in contrast to men who added more of those tags to pictures of the category Landscape (3.82%). This difference is statistically significant (Table 4.2). Images showing Architecture (1.54%) got the least Fakeness hashtags assigned by men. The hashtag category Sentences was not distributed much among the picture categories. Both genders gave the lowest percentages of those tags to images about art. Mostly, Sentences tags were assigned to Pet and Captioned Photo by men and women alike. Women also tagged pictures depicting Friends with those tags. Nearly all of the Sentences hashtags from both genders in the Pet category were emotional and positive, like e.g. #ilovemypet.

RQ3. Are there any gender-specific associations between the picture categories and hashtag categories?

A chi-square test of independence was conducted to investigate if an association between the hashtag categories and picture categories for each gender (male and female) exists. The following hypotheses were formulated:

H0: No gender-specific association between picture categories and hashtag categories exists. **HA:** A Gender-specific association between picture categories and hashtag categories exists.

All expected cell frequencies were greater than five. A significant association between hashtag categories, picture categories, and both genders was found. For male users, the chi-square test result was v2(54, 500) = 970.993, p ¡ .001 with an effect size of Cramer's V = .144 which indicated a small association (Cohen, 1988). Likewise, female users (v2(54, 500) = 814.440, p ¡ .001) showed a small association (Cramer's V = .137) as well. The alternative hypothesis can be accepted, because the null hypothesis has been rejected.

4.4 Discussion

This research study examined the tagging of images by male and female Instagram users. The number of investigated Instagram picture postings was 1,000 (500 for each gender). The pictures and the text captions of the respective postings were analyzed to answer three research questions about the gender-specific distribution of hashtag and picture categories, as well as hashtag frequencies.

Usually, the number of hashtags differs from picture to picture and, for some image categories, by gender. Males assigned 15 hashtags on average to their postings, whereas women used one hashtag less. Notable categories where men tagged their images with at least 4 hashtags more on average were Captioned Photo and Landscape. A great difference of approximately 9 hashtags was observed for captioned photos. Women tagged Selfie and Food pictures with more hashtags on average. These differences are statistically significant. Both genders applied around 18 hashtags to the Pet category on average, but the male data was more dispersed.

When indexing non-textual content, internet users tend to mix together ofness, aboutness, as well as isness aspects (Stock & Stock, 2013). The same process has been observed in this Instagram study, where users indexed their postings with hashtags. Those tags were coded into Content-relatedness, Isness, and five other kinds of hashtag categories. Content-relatedness was the most important hashtag category for male and female Instagram users alike (over 55%). This category specified the subject of a picture by referring to its ofness and aboutness. Second most assigned category by both genders was Isness, indicating the importance of hashtags relating to isness elements of an image as well as further emphasizing the mentioned mixing together of ofness, aboutness, and isness aspects. Females assigned slightly more emotional hashtags to all image categories than men; this difference is statistically significant. Generally, men assigned Isness and "Insta"-Tags in a higher frequency than females. Furthermore, "Insta"-Tags were assigned in high frequencies by both genders to Pet. Generally, both genders did not use many Sentences as hashtags. A chi-square test of independence showed a small statistical association between hashtag and picture categories for male and female Instagram users, respectively.

The high frequency of Content-relatedness tags in this work are in accordance with two other studies which have shown that a high number of hashtags relate to the depicted content of the picture (Giannoulakis & Tsapatsoulis, 2016; Dorsch, 2018). Hashtags that refer mostly to the content or context of a document were also observed for tweets (Holmberg & Hellsten, 2015). It is possible that male and female Instagram users describe the content and context of their pictures to make them findable for other users. They want their images to be visible in the social network community. Men use informative hashtags more than women when tagging food pictures according to Zhang et al. (2018). Informative hashtags contain information and non-emotional descriptions of an image. Content-relatedness hashtags could be considered as informative and non-emotional, because emotions were classified as Emotiveness hashtags. The hashtag category Isness might be considered as informative as well, since it relates to technical aspects of the picture that are not depicted. In this study, the male Content-relatedness hashtag frequencies only exceed the female ones in four out of ten picture categories. Nevertheless, the difference in the total tag frequency is very small (about 1.5%). In the Isness category, men assign a higher percentage of tags to all image categories except Fashion (about 2.5% difference). The total frequency of Isness hashtags assigned by men is significantly higher when compared to women. If only the category Food would be considered, like in Zhang et al.'s work (2018), females tag food pictures significantly more with Content-relatedness tags. The results of this study would differ from Zhang et al.'s findings in this point when isolating this specific picture category.

Various studies concluded that women express more feelings and emotions than men in social networks (e.g. Zhang et al., 2018; Kivran-Swaine, Brody, Diakopoulos, & Naaman, 2012; Vikatos, Messias, Miranda, & Benevenuto, 2017). This study's results support those findings. Even though Emotiveness tags are not highly represented by both genders (¡10%), significant differences exist. Women assign more emotional hashtags than men to all picture categories. Women's selfies are tagged with more Emotiveness tags than men's. Pictures depicting friends and pets received the most emotional feedback by both genders, probably because of the bond that users can form with a person or an animal and thus, being able to feel emotions towards them than towards lifeless objects. Sheldon and Bryant (2016) found a positive relationship between regular hashtag use and the motive "Coolness" in a study with college students. It could be speculated that "Insta"-Tags imply, in a way, coolness due to their abbreviation or full inclusion of the application's name (Instagram). These tags are a phenomenon that occurs only on Instagram (Dorsch, 2018). Males make more frequently use of "Insta"-Tags, especially in the picture category Captioned Photo. Both genders assign "Insta"-Tags a lot when posting Pet pictures. Is there a specific reason for that? Instagram is especially favored by young adults aged 18 to 24 in the U.S. (Smith & Anderson, 2018). Is it young adult slang that introduced these tags to Instagram? Further studies are needed to investigate "Insta"-Tags and motivation for their application.

Another question that needs to be asked is why do male users assign Fakeness tags in significantly greater amounts to Landscape pictures? For this question, apps and websites[1] that provide top hashtags to gain more likes and followers on social media platforms could be relevant. Sometimes, top tags are assigned which only fit in the category Fakeness, e.g. when a picture of an older dog is tagged with #puppy. Nevertheless, these hypotheses require proof.

The tagging behavior of men and women on Instagram is described in this study. Future studies could investigate the motivation of male and female users for applying a specific hashtag to a specific picture motive. With help of uses and gratification theory (Katz, Blumler, & Gurevitch, 1974) for social media (Leung, 2013), the motivations behind tagging could be determined. Uses and gratification theory is about what people do with the media. In context of social media, it investigates motivations for content creation and how the activities in social media are affected by the gratifications of content generation. To the authors knowledge, only one study exists that applied uses and gratification theory to gender-specific hashtag application (Zhang et al., 2018). It noted that female users gratify needs of expressing feelings by using positive and emotional hashtags, whereas male users gratify needs of giving information by using informative hashtags. Questionnaires and qualitative interviews are possible ways to explain male and female tagging intentions and motivations further. This study could act as base to elaborate these surveys.

The picture categories in this investigation do not cover every potential picture motive. They exemplify possible motives in a broader scope and thus, make no claim to be a complete overview of all image motives on Instagram. Social media data sampling itself also has some limitations, as there often is a fixed time-frame for studies that should not be exceeded. The data for this study was also collected in a specific timeframe. Therefore, not all images that are conform to the picture categories have been included in the content analysis. Another related aspect is that not every Instagram user specifies his or her gender. In conclusion, this work is a

[1] e.g. `https://www.hashtagsforlikes.co/`.

case study, as it is impossible to generalize the results for the entire population.

In addition to analyzing pictures, videos could be examined. Do men and women show tagging differences in this media category as well? Would the differences be the same as the ones observed for the picture categories? Furthermore, a content analysis in regard to "hashtags for likes" websites could be conducted, to figure out if men and women make use of the provided hashtags in different frequencies for specific picture motives. In addition to gender, parameters like age, nationality, and lifestyle could be taken into account to research tagging behavior. Do younger Instagram users assign more "Insta"-Tags than older ones? Which lifestyle enables the use of performative tags? It would be interesting to determine if the opinion about a picture category is positive or negative by conducting a sentiment analysis on the Emotiveness hashtags. What kinds of emotions are shown only by men or by women?

The tagging behavior of certain user groups like for example influencers could be investigated as well. Influencers are individuals who can affect purchase decisions of others due to their high reputation and prestige on social media platforms. Do male and female influencers tag their postings differently, and which gender is more successful to which audience? Are there any differences when comparing influencer's and noninfluencer's tagging behaviors? Success could be measured by the number of followers and likes. Marketing strategies to target a specific customer group could be enabled. Finally, examining the image tagging behavior of men and women on a different social media platform may bring further insights to this subject.

This study analyzed gender-specific tagging of images on Instagram. It concludes that in several picture and hashtag categories there are indeed significant differences, but similarities as well, in the tagging behavior of male and female Instagram users.

References

Anagnostopoulos, C., Parganas, P., Chadwick, S., & Fenton, A. (2018). Branding in pictures: Using Instagram as a brand management tool in professional team sport organisations. *European Sport Management Quarterly*, *18*(4), 413–438.

Austin, J. L. (1962). *How to do things with words.* Oxford: Clarendon.

Brown, R. C., Fischer, T., Goldwich, A. D., Keller, F., Young, R., & Plener, P. L. (2018). #cutting: Nonsuicidal self-injury (NSSI) on Instagram. *Psychological Medicine*, *48*(2), 337–346.

Carrotte, E. R., Prichard, I., & Lim, M. S. C. (2017). "Fitspiration" on social media: A content analysis of gendered images. *Journal of Medical Internet Research*, *19*(3), e95.

Coelho, R. L. F., de Oliveira, D. S., & de Almeida, M. I. S. (2016). Does social media matter for post typology? Impact of post content on Facebook and Instagram metrics. *Online Information Review*, *40*(4), 458–471.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.

Cunha, E., Magno, G., Almeida, V., Gonçalves, M. A., & Benevenuto, F. (2012). A gender based study of tagging behavior in Twitter. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT 2012* (pp. 323–324). New York: ACM.

Daer, A. R., Hoffman, R., & Goodman, S. (2015). Rhetorical functions of hashtag forms across social media applications. *Communication Design Quarterly Review*, *3*(1), 12–16.

Dorsch, I. (2018). Content description on a mobile image sharing service: Hashtags on Instagram. *Journal of Information Science Theory and Practice*, *6*(2), 46–61.

Dorsch, I., Zimmer, F., & Stock, W. G. (2017). Image indexing through hashtags in Instagram. *Proceedings of the Association for Information Science and Technology*, *54*(1), 658–659.

Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., & Tomkins, A. (2007). Visualizing tags over time. *ACM Transactions on the Web*, *1*, 1–22.

Geurin-Eagleman, A. N., & Burch, L. M. (2016). Communicating via photographs: A gendered analysis of olympic athletes' visual self-presentation on Instagram. *Sport Management Review*, *19*(2), 133–145.

Giannoulakis, S., & Tsapatsoulis, N. (2016). Evaluating the descriptive power of Instagram hashtags. *Journal of Innovation in Digital Ecosystems*, *3*(2), 114–129.

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208.

Gray, J. (1992). *Men are from Mars, women are from Venus.* New York: Harper-Collins.

Halavais, A. (2014). Structure of Twitter: Social and technical. In K. Weller, A. Bruns, J. Burgess, N. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 29–41). New York: Peter Lang.

Holland, G., & Tiggemann, M. (2017). Strong beats skinny every time": Disordered eating and compulsive exercise in women who post fitspiration on Instagram. *International Journal of Eating Disorders*, *50*(1), 76–79.

Holmberg, K., & Hellsten, I. (2015). Gender differences in the climate change communication on Twitter. *Internet Research*, *25*, 811–828.

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. In *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014* (pp. 595–598). Palo Alto, CA: AAAI Press.

Ingwersen, P. (2002). Cognitive perspectives of document representation. In H. Bruce, R. Fidel, P. Ingwersen, & V. P. (Eds.), *Proceedings of the 4th International Conference on Conceptions of Library and Information Science* (pp. 285–300). Greenwood Village, CO: Libraries Unlimited.

Instagram. (2010). *Instagram launches.* Retrieved from `https://instagram-press.com/blog/2010/10/06/instagram-launches-2/`

Instagram. (2012a). *Instagram + Facebook (2012).* Retrieved from `https://`

instagram-press.com/blog/2012/04/09/

Instagram. (2012b). *Instagram for android.* Retrieved from https://instagram-press.com/blog/2012/04/03/instagram-for-android-available-now/

Instagram. (2019). *Instagram statistics.* Retrieved from https://instagram-press.com/our-story/

Katz, E., Blumler, J. G., & Gurevitch, M. (1974). Utilization of mass communication by the individual. In *The Uses of Mass Communications: Current Perspectives on Gratifications Research* (pp. 19–33). Beverly Hills: Sage Publications.

Kivran-Swaine, F., Brody, S., Diakopoulos, N., & Naaman, M. (2012). Of joy and gender: Emotional expression in online social networks. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (pp. 139–142). New York: ACM.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks: Sage Publications Inc.

Krippendorff, K. (2011). *Computing Krippendorff's alpha reliability. Departmental papers (ASC) (2011).* Retrieved from https://repository.upenn.edu/asc_papers/43

Laniado, D., Eynard, D., & Colombetti, M. (2007). Using WordNet to turn a folksonomy into a hierarchy of concepts. In G. Semeraro, C. Di Sciascio, H. Morbidoni, & H. Stoermer (Eds.), *Semantic Web Application and Perspectives: Proceedings of the 4th Italian Semantic Web Workshop (SWAP 2007). CEUR Workshop Proceedings* (Vol. 314, pp. 192–201). Bari: CEUR Workshop Proceedings.

Lavoie, K. A. (2015). Instagram and branding: A case study of Dunkin' Donuts. *Elon Journal of Undergraduate Research in Communications*, *6*(2), 79–90.

Lee, C., & Chau, D. (2018). Language as pride, love, and hate: Archiving emotions through multilingual Instagram hashtags. *Discourse Context Media*, *22*(0), 21–29.

Lee, E., Lee, J.-A., Moon, J. H., & Sung, Y. (2015). Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, Behavior, and Social Networking*, *18*(9), 552–556.

Leung, L. (2013). Generational differences in content generation in social media: The roles of the gratifications sought and of narcissism. *Computers in Human Behavior*, *29*(3), 997–1006.

Lup, K., Trub, L., & Rosenthal, L. (2015). Instagram #Instasad? Exploring associations among Instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking*, *18*, 247–252.

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (2009). Codebook development for team-based qualitative analysis. In *The content analysis reader* (pp. 211–219).

Messina, C. (2007). *How do you feel about using # (pound) for groups. As in #barcamp [msg]? chrismessina (Twitter).* Retrieved from https://twitter.com/chrismessina/status/223115412

Oh, C., Lee, T., Kim, Y., Park, S., & Suh, B. (2016). Understanding participatory hashtag practices on Instagram: A case study of Weekend Hashtag Project. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1280–1287). New York: ACM.

Panofsky, E. (1955). *Meaning in the visual arts.* Garden City: Doubleday Anchor Books.

Peters, I. (2009). *Folksonomies. Indexing and retrieval in web 2.0.* Berlin: De Gruyter Saur.

Peters, I., & Stock, W. G. (2007). Folksonomy and information retrieval. In *Joining Research and Practice: Social Computing and Information Science. Proceedings of the 70th ASIS&T Annual Meeting* (Vol. 44, pp. 1510–1542). Hoboken: Wiley.

Peters, M. E. I., I.and Kipp, Heck, T., Gwizdka, J., Lu, K., Neal, D. R., & Spiteri, L. (2017). Social tagging & folksonomies: Indexing, retrieving... and beyond? In *Bridging the Gulf: Communication and Information in Society, Technology, and Work. Proceedings of the 74th ASIS&T Annual Meeting* (pp. 1–4). Wiley, Hoboken.

Santarossa, S., Coyne, P., Lisinski, C., & Woodruff, S. J. (2016). #fitspo on Instagram: A mixedmethods approach using Netlytic and photo analysis, uncovering the online discussion and author/image characteristics. *Journal of Health Psychology*, *24*(3), 376–385.

Santarossa, S., Coyne, P., & Woodruff, S. J. (2017). Exploring #nofilter images when a filter has been used. *International Journal of Virtual Communities and Social Networking*, *9*, 54–63.

Schmidbauer, H., Rösch, A., & Stieler, F. (2018). The 2016 US presidential election and media on Instagram: who was in the lead? *Computers in Human Behavior*, *81*, 148–160.

Shapp, A. (2015). Gender variation in the pragmatic uses of Twitter hashtags. In *Poster session presented at the 89th Annual Meeting of the Linguistic Society of America, Portland, Oregon.* Retrieved from `https://www.nyu.edu/projects/shapp/Shapp_LSA2015.html`

Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging and Classification Quarterly*, *6*(3), 39–62.

Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, *58*, 89–97.

Sheldon, P., Rauschnabel, P. A., Antony, M. G., & Car, S. (2017). A cross-cultural comparison of Croatian and American social network sites: Exploring cultural differences in motives for Instagram use. *Computers in Human Behavior*, *75*, 643–651.

Siebenlist, T. (2013). Emotionale Suche – Emotionales Information-Retrieval. In D. Lewandowski (Ed.), *Handbuch der Internet-Suchmaschinen 3: Suchmaschinen zwischen Technik und Gesellschaft* (pp. 299–327). Heidelberg: Akademische Verlagsgesellschaft.

Smith, A., & Anderson, M. (2018). *Social media use in 2018.* Retrieved from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

Stock, W. G., & Stock, M. (2013). *Handbook of information science.* Berlin, Boston: De Gruyter Saur.

Talbot, C. V., Gavin, J., van Steen, T., & Morey, Y. (2017). A content analysis of thinspiration, fitspiration, and bonespiration imagery on social media. *Journal of Eating Disorders*, *5*(40), 1–8.

Tiggemann, M., & Zaccardo, M. (2016). “Strong is the new skinny”: A content analysis of #fitspiration images on Instagram. *Journal of Health Psychology*, *23*(8), 1003–1011.

Toffler, A. (1980). *The Third Wave.* New York: Morrow.

Trant, J. (2009). Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, *10*(1), 1–44.

van den Berg, J. A. (2014). The story of the hashtag (#): A practical theological tracing of the hashtag (#) symbol on Twitter. *HTS Teologiese Studies / Theological Studies*, *70*(1), 1–6.

Vander Wal, T. (2007). *No title.* Retrieved from http://www.vanderwal.net/folksonomy.html

Vikatos, P., Messias, J., Miranda, M., & Benevenuto, F. (2017). Linguistic diversities of demographic groups in Twitter. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media, HT* (pp. 275–284). New York: ACM.

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society*, *20*(5), 1813–1831.

Zhang, Y., Hashim, N. H., Baghirov, F., & Murphy, J. (2018). Gender differences in Instagram hashtag use. *Journal of Hospitality Marketing and Management*, *27*(4), 386–404.

Chapter 5

Dreaming of Stardom and Money: Micro-celebrities and Influencers on Live Streaming Services

Social live streaming services (SLSSs) are social media, which combine Live-TV with elements of Social Networking Services (SNSs). In social media and thus also in SLSSs, the so-called influencer and micro-celebrities play an important role, but to what extend are SLSSs' streamers motivated by fame or financial gain? We conducted a content analysis in order to investigate SLSSs' streamers (n = 7,667) on Periscope, Ustream and YouNow in respect to their general characteristics and streaming motivation being fame and financial gain. We have developed a research model referring to the platform used by the streamers, their gender, origin, age and streamed content (general characteristics), as well as the motivational aspects. Streamers of Ustream are mostly motivated by financial gain, whereas YouNow broadcasters seek to be famous. Considering the streamers age, older generations (Gen X, Silver Surfers) aspire after financial gain. With progressing age the motivation to become a star decreases. Mostly streamed content by streamers motivated by money is entertainment media. For streamers wanting to become a star chatting and making music are the preferred content categories.

5.1 Introduction

Since the turn of the Millennium and increasing usage of the Internet and its applications, research on people becoming "celebrities" or "micro-celebrities" thanks to the new technology is gaining on popularity and importance (e.g. Ingleton, 2014; Kane, 2010; Marwick, 2013; T. Senft, 2008). Now, ordinary social media users can become important players of the so-called attention economy (Marwick, 2015; Tufekci, 2013) with the help of self-branding and presentation strategies (e.g. Khamis, Ang, & Welling, 2017; Marwick, 2013; Page, 2012; T. M. Senft, 2013). We can find micro-celebrities and so-called social media influencers on YouTube, Instagram, or Snapchat. However, do users of a new kind of platforms like the social live streaming services also aspire "stardom and money"? This is an explorative study addressing this particular topic. First, we will shed light on the new form of social media – the social live streaming services as well as on the concepts of "micro-celebrity" and "influencer". Afterwards, we will elaborate on our applied methods and present results of our investigation based on observations of streamers on three different platforms. Finally, we will answer the question whether social live streaming users are indeed interested in fame and money.

5.1.1 Social Live Streaming Services

In recent years a new form of social media has established itself, the so-called Social Live Streaming Services (SLSSs). They combine Live-TV with elements of Social Networking Services (SNSs) as they include a backchannel between the viewers and the streamers as well as among the viewers. We can find such SLSSs as Periscope, Ustream, YouNow, YouTube Live, Facebook Live, Instagram Live, Snapchat Live Stories, niconico (in Japan), YiZhiBo, Xiandanjia, Yingke (all in China) or – for broadcasting e-sports or drawing – Twitch and Picarto, respectively. Such services allow their users to broadcast live anything they want and to everyone who is interested to watch.

The scientific research on SLSSs is gaining in importance as well as spectrum. In computer science, one can find studies on bandwidth (Bilal, Erbad, & Hefeeda, 2017), video quality (Stohr, Toteva, Wilk, Effelsberg, & Steinmetz, 2017) and the delay of comments' displays (Rodríguez-Gil, García-Zubia, Orduña, & López-Ipiña, 2017). SLSSs find application in private contexts (Scheibe, Fietkiewicz, & Stock, 2016), but also in more serious environments, e.g. in teaching neurosurgery (Maugeri, Giammalva, & Iacopino, 2016) or economics (Dowell & Duncan, 2016). They can also be applied in marketing (Keinänen, 2017). Furthermore, SLSSs are applied for live broadcasting sports events, however, this is also connected to some legal problems (Ainslie, 2015). Despite broadcasting sports events, also other general legal and ethical implications may arise (Alamiri & Blustein, 2016; Faklaris et al., 2016; Honka, Frommelius, Mehlem, Tolles, & Fietkiewicz, 2015; Zimmer, Fietkiewicz, & Stock, 2017). There are studies on topic-specific SLSSs, e.g. in e-sports context on Twitch (e.g. Bründl & Hess, 2016; Gros, Wanner, Hackenholt, Zawadzki, & Knautz, 2017), and on general SLSSs (without any thematic limitation) (Fietkiewicz & Scheibe, 2017; Friedländer, 2017a, 2017b; Scheibe et al., 2016; Stohr, Li, Wilk, Santini, & Effelsberg, 2015; Tang, Kivran-Swaine, Inkpen, & Van House, 2017). Studies found that general live streaming was appreciated for its authentic, uncurated, and interactive attributes (Tang, Venolia, & Inkpen, 2016) as well as for its role for sharing breaking news (Tang et al., 2017). However, we miss studies, which systematically investigate the motivation of streamers to become micro-celebrities or influencers on the general SLSSs. We aim to close this research gap with the following investigation.

5.1.2 Micro-celebrities and Influencers on Social Media

Media like television have been instrumental in generating new "celebrities" parallel to the "film-celebrities", who enjoy slightly more popularity (Kavka, 2015). With time and creation of new TV genres, a new kind of celebrity like "reality TV stars" attracted attention of the crowds (Kavka, 2015). With increasing popularity of social media further types of celebrities emerged, for example, YouTube stars (West, n.d.) or bloggers, usually reporting on both channels (Kavka, 2015; Marwick, 2013). This trending interest in uncensored (private) life of others and "Big Brother"-like shows is not unproblematic and became topic for many critical discourses, an example being the American movie "The Truman Show" (Fietkiewicz & Scheibe, 2017).

Still, media change together with the concept of celebrity—from celebrity focused solely on mass and broadcast media, to the one active on a diversified media landscape, and then further to participatory media (Kavka, 2012; Marwick, 2016). More interestingly, this development enables not only famous people (from TV or films), but also non-famous people "to generate vast quantities of personal media, manipulate and distribute this content widely, and reach out to (real or imagines) audiences" (Marwick, 2016). Hence, increasingly "ordinary people" are being transformed into celebrities (Kavka, 2012), or rather, thanks to social media and self-branding, they transform themselves into ones.

Marwick (2016) points out two major changes in celebrity culture due to the shift towards participatory media. First, the "traditional" celebrities are using "social media to create direct, unmediated relationship with fans, or at least the illusion of such" (Marwick, 2016). This illusion of a real face-to-face friendships with celebrities created through watching TV shows or listening to music is the so-called "para-social interaction" (Horton & Wohl, 1956; Marwick, 2016; Marwick & Boyd, 2010), however, with use of social media this interaction can become more "social" and "increase the emotional ties between celebrity and fan" (Marwick, 2016; Marwick & Boyd, 2011; Muntean & Petersen, 2009). The second change is related to the phenomenon of "micro-celebrity", a form of celebrity that may have a small audience, but is still "able to inhabit the celebrity subject position through the use of technologies" (Marwick, 2016). As opposed to the "broadcast era" where "celebrity was something a person was; in the Internet era, micro-celebrity is something people do" (Marwick, 2016). The phenomenon of micro-celebrity is strongly linked to the notions of self-branding and strategic self-presentation, and requires "viewing oneself as a consumer product", and "image" that needs to be sold to the right target group (Hearn, 2008; Lair, Sullivan, & Cheney, 2005; Marwick, 2016). Micro-celebrities view friends and followers on social media channels as their fanbase that needs to be managed by various affiliative techniques (Marwick & Boyd, 2011). These trends have empowered many participants in the newly emerging "online reputation economy, where the reputation generated by social media platforms functions as a new form of currency, and more generally, value" (Hearn & Schoenhoff, 2016, p. 203).

The emergence of online reputation economy has led to establishment of a new concept of the "micro-celebrity", namely the social media influencer (SMI). Such influencer "works to generate a form of 'celebrity' capital by cultivating as much attention as possible and crafting an authentic 'personal brand' via social networks, which can subsequently be used by companies and advertisers for consumer outreach" (Hearn & Schoenhoff, 2016). Businesses increasingly rely on social media influencers, on one hand "due to the sheer volume of advertising online, which drives down actual click-through rates and individual engagement levels", on the other hand, due to higher authenticity of claims made by "personal acquaintance" rather than by a rich celebrity (Hearn & Schoenhoff, 2016; Martin, 2018; Schaefer, 2012). Marketing strategists are looking for social media users with an extensive social network that is frequently used, as well as with "relevant or 'sticky' content about the product category, and whose personality 'resonates' with the tone and feel of the brand" (Hearn & Schoenhoff, 2016). This way ordinary social media users become social media influencers making money and their living by posting pictures, videos and blog posts—all the activities that other (non-influential) social media users do, but apparently not as good as the influencers.

Micro-celebrities and influencers will make money by advertising products or services. This also applies to social live streaming services. In addition to being paid by third parties, some of the services offer possibilities to make money by using the SLSS (of course, provided that the streamer attracts a considerable amount of viewers). Services like Facebook Live or Periscope allow pre-roll and mid-roll advertising as well as displaying overlay ads. Some of the gaming channels on YouTube also have access to sponsorships that are financed by the viewers who can purchase digital goods like badges and emojis and have access to "special perks" (YouTube, 2018). Very popular are also fan donations, for example, YouTube's Super Chat (viewers can get their chat message pinned to the top of the comments section by paying a small fee), or Bits on Twitch (viewers pay for affiliated streamers to receive a certain number of "Bits"). SLSSs as Twitch or Picarto offer monthly subscriptions (Kaser, 2017). On YouNow, streamers can earn money from tips and gifts. For this purpose, viewers can buy bars, with these they can buy gifts that they can give to a streamer who is a YouNow Partner (who in turn receives real money) (YouNow, 2018).

To sum up, with new forms of media, new forms of celebrities and "influential" people emerged—the micro-celebrities and social media influencers. They earn money doing advertising for products and services (with product placement or reviews), or on some of the platforms, especially on social live streaming services, by subscriptions, donations and gifts from the viewers. They also gain recognition and approval of their fan-base, which for some of them is as attractive and important as financial gain for others. In this study we are going to investigate whether general SLSSs users indeed aspire to become micro-celebrities and/or to earn money with the help of these service. This is an explorative study that is supposed to shed light on the general characteristics of streamers dreaming of "stardom and money".

5.1.3 Research Questions and Research Model

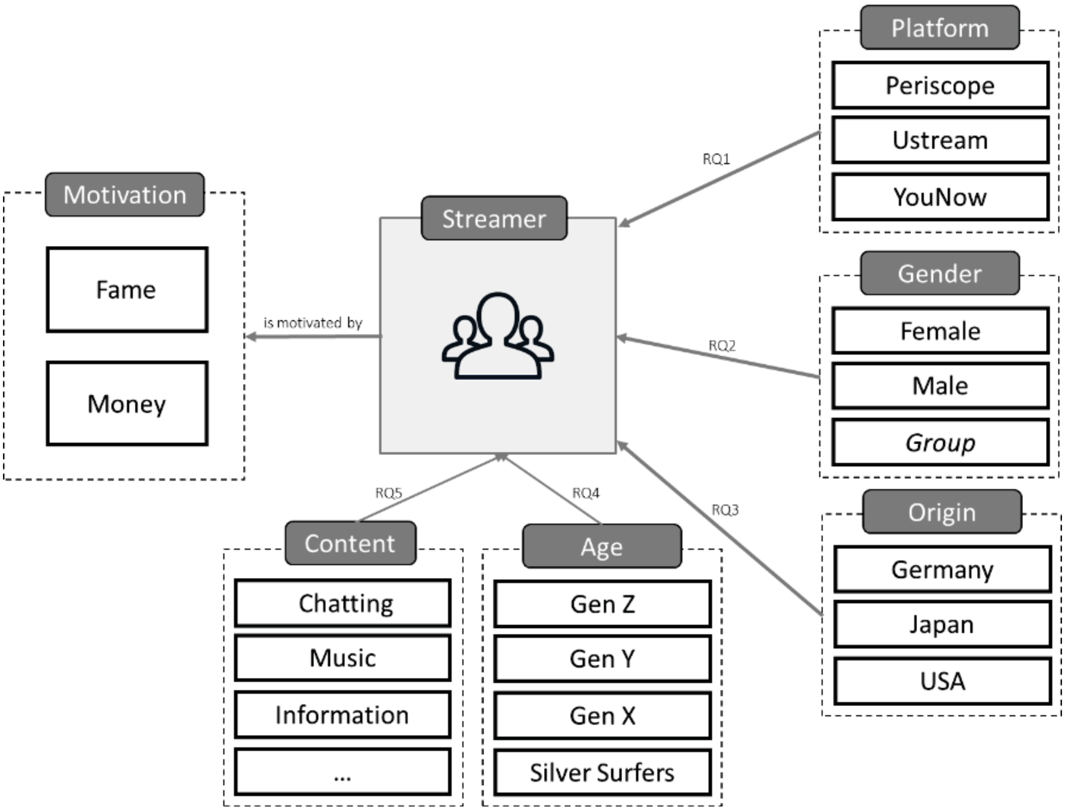


Figure 5.1: Our research model.

In order to explore the general characteristics of SLSSs users (in particular, streamers or producers) motivated by fame or financial gain we formulated the following research questions:

- RQ1: Which channels (Periscope, Ustream, YouNow) are preferred by users motivated by fame or financial gain?

- RQ2: Are there gender-dependent differences regarding the streaming motivation being fame or financial gain?

- RQ3: Are there origin-dependent differences (Germany, Japan, USA) regarding the streaming motivation being fame or financial gain?

- RQ4: Are there age-dependent differences regarding the streaming motivation being fame or financial gain?

- RQ5: What are the contents streamed by streamers whose motivation is fame or financial gain?

According to our research model (Figure 5.1), we focus on streamers that are either interested in financial gain or in becoming famous. These streamers will use a certain social live streaming platform. They will either stream by themselves (male or female streamer) or not (group of streamers). Furthermore, the streamers will have different origins (Germany, Japan or the USA). Moreover, there can be age-dependent differences between the streamers' motivations. And finally, they can stream different content types.

5.2 Methods

5.2.1 Systematic Observation of Live Streams

In order to answer our research questions, we have conducted observations of the streams. We evaluated and compared SLSSs' users' streaming behavior as well as the content of a stream and motives of a streamer to produce a live stream (Friedländer, 2017a; Zimmer et al., 2017). The empirical procedure of the content analysis included development of a codebook and a two-phased approach ensuring high reliability (Krippendorff, 2012; Lai & To, 2015; Recktenwald, 2017). First, the directed approach was implemented with help of literature on social media, in order to get guidance for the research categories. Second, the conventional approach via observation of live streams was used to get a general idea of what people stream about. This way we were able to define the categories of content of a stream and motivation of the steamer.

The content categories include: to chat; to make music; to share information; news; fitness; sport event; gaming; animals; entertainment media; spirituality; draw/paint a picture; 24/7; science, technology, and medicine (STM); comedy; advertisement; nothing; slice of life; politics; nature; food; and business information. The motivation categories include: entertainment (boredom, fun, hobby); information (to reach a specific group, exchange of views), social interaction (socializing, loneliness, relationship management, need to communicate, need to belong), and self-presentation (self-improvement, self-expression, sense of mission, to become a celebrity, to make money, trolling). "No comment" was marked if the streamer did not state a motivation or no person could be reached via chat, for example if an animal was shown or a 24/7 stream (e.g. from a webcam) was broadcasted. However, for this investigation we focus on two subcategories of the self-presentation category, namely "to become a celebrity" and "to make money". Hence, for the investigation only observations were selected, where streamer confirmed to be motivated by one of these two factors.

Norm entries were used for the socio-demographic data like gender (male, female, group) and age of the streamer. The data about the streams from three general SLSSs (YouNow, Periscope, and Ustream) were collected from three different countries, namely Germany, Japan, and the USA. To ensure that the streams originated from those countries the declaration of the country for a broadcast on each platform was checked for every stream. Additionally, the data collectors had the required language skills for those countries. Twelve research teams (each consisting of two people) were evenly distributed between the three countries. Every coder received a spread sheet to code the observed data. Each stream was observed simultaneously but independently by two people for two to a maximum of ten minutes. Usually the streams were observed in two phases. First, the stream was watched and the data were collected. In phase two, if some aspects were not clear, for example the motivation of the streamer, the streamer was asked via the chat system of the service. In the end, a data set of 7,667 different streams in a time span of four weeks, from April 26 to May 24, 2016, was collected.

5.2.2 Data Preparation and Analysis

Our dataset consisted of mostly nominal data. There were three categories of the variable *platform* (YouNow, Periscope, Ustream) as well as three categories of the variable *origin* (Germany, Japan, USA). The variable *gender* was not binary-coded, but included categories male, female, group (for streams with more than one streamer, where specification of one gender was not possible), and n/a (not available, for streams where no streamer could be seen; these cases were subsequently defined as missing values). Finally, the age of the streamers was coded on a metric scale.

In order to investigate the possible influence of the age of the streamers on their motivation, we have aggregated the data into generational groups. For this purpose, we have followed the categorization applied in studies on generational cohorts of social media users (Fietkiewicz, Lins, Baran, & Stock, 2016; Leung, 2013). According to these studies, there is the Silent Generation (born between 1925 and 1945), the Baby Boomers (1946–1960), Generation X (1961–1980), Generation Y (1981–1998) and Generation Z (born after 1998). Due to low observation numbers of older steamers, we have merged the "Baby Boomers" and "Silent Generation" into one group called Silver Surfers (N = 33).

For the investigation we have applied descriptive statistics including frequencies and Pearson Chi-Square test for association, since almost all of our variables were nominal with more than 2 categories. The chi-square test determines whether there is an association between two nominal variables (in our case, association between the general characteristics and the motivation for using SLSSs being "fame" or "money"). Furthermore, we have measured the effect size using Cramer's V to investigate the strength of the respective association. The magnitude of effect size can be interpreted as small (0.1), medium or moderate (0.3) and large (0.5) (Cohen, 1988; Laerd Statistics, 2018).

5.3 Results

In our study (observation of streams; N = 4,548 streams with single broadcasters; N = 1,082 of "groups"), we identified 61.2% male broadcasters and 38.8% females from Germany, Japan and the USA (Table 5.2). The results from (Tang et al., 2016, p. 4774) confirm this gender distribution: about three fifths of SLSSs' users are male. The observed streams were almost evenly distributed among the three platforms, with the highest number of observations for Periscope (38.5%) and the lowest one for YouNow (26.4%). As for the distribution by the country of origin, the most streams were from the USA (41%) and the fewest from Japan (25%). Finally, we have aggregated the age of the streamers into generational cohorts. The most represented generation is the youngest one—Gen Z with 37.2% followed by Gen Y with 33.5%. The older generational groups, Generation X and Silver Surfers, are much smaller as they represent only 6.4% and 0.4% of the observed streamers, respectively. Since we could not estimate the age of all observed streamers or the ones streaming in groups, the number of observations within the *Generation* category is accordingly lower.

Table 5.1: Demographic data of observed streamers.

		Frequency	Valid percent
Platform [N = 7,667]	Periscope	2,960	38.6%
	Ustream	2,686	35.0%
	YouNow	2,021	26.4%
Gender [N = 5,630]	Female	1,766	31.4%
	Male	2,782	49.4%
	Group	1,082	19.2%
Origin [N = 7,667]	Germany	2,604	34.0%
	Japan	1,920	25.0%
	USA	3,134	41.0%
Generation [N = 4,937]	Gen Z	1,839	37.2%
	Gen Z	1,839	37.2%
	Gen Y	2,572	33.5%
	Gen X	493	6.4%
	Silver Surfers	33	0.4%

5.3.1 Platform-Dependent Differences

Table 5.2: Platform-dependent differences in motivation to *make money* and *become a star*.

Platform	Making money	Becoming a star
Periscope (N = 2.960)	1.32%	2.53%
Ustream (N = 2.686)	12.92%	1.15%
YouNow (N = 2.021)	4.60%	9.65%
Pearson Chi^2	$p < 0.001$	$p < 0.001$

Platform-dependent differences regarding the motivational factor "making money" and "becoming a star" can be obtained from Table 5.2. With about 13%, the motivational factor money is highest for Ustream streamers, whereas becoming a star is of minor interest. Streamers of YouNow are mostly motivated by fame (9.65%). For Periscope streamers, neither factor plays a major role. A chi-square test for association was conducted for the platforms and the motivational factors. All expected cell frequencies were greater than five. There is a statistically significant association between the platforms and the motivational factors, however, the association is rather small (Cramer's V = 0.209 for money, and 0.179 for fame).

5.3.2 Gender-Dependent Differences

Regarding the gender-dependent differences, males are slightly more motivated by financial gain than women (Table 5.3). Nevertheless, both motivational factors are highest for the group-streams. With regard to the factor fame, there are no major gender-specific differences (p ¿ 0.05). The conducted chi-square test of independence between gender and the motivational factor money results in a very small (0.083) statistically significant association.

Table 5.3: Gender-dependent differences in motivation to *make money* and *become a star*.

Gender	Making money	Becoming a star
Female (N = 1,766)	2.49%	4.76%
Male (N = 2,782)	4.82%	5.14%
Group (N = 1,082)	7.49%	5.82%
Pearson Chi^2	$p < 0.001$	$p = 0.458$

5.3.3 Origin-Dependent Differences

Considering the streamers' origin and its influence on the motivations money and fame (Table 5.4), streamers located in USA are the ones most motivated by financial gain (7.45%), followed by Germans (5.91%) and Japanese (4.74%). In turn, fame is mostly aspired by German streamers (5.11%), followed by American (3.91%) and Japanese ones (2.34%). Even though there is a statistically significant association between origin and the motivational factors, the association is (similar to the gender-dependent differences) only small.

Table 5.4: Origin-dependent differences in motivation to *make money* and *become a star*.

Origin	Making money	Becoming a star
Germany (N = 2.604)	5.91%	5.11%
Japan (N = 1.920)	4.74%	2.34%
USA (N = 3.143)	7.45%	3.91%
Pearson Chi^2	$p < 0.001$	$p < 0.001$

5.3.4 Age-Dependent Differences

Differences in the streamers' motivation dependent on their age can be identified in Table 5.5. Unfortunately, we did not meet the assumption that all cells should have expected counts greater than five for one cell (12.5%), therefore, these result have to be interpreted with some caution. Apparently, with increasing age the intention to earn money rises, but simultaneously the goal to become a star decreases. 21.21% of the Silver Surfers seek financial gain, however, none of them wants to become a star. In contrast, 8.21% of the generation Z aim to become a star, whereas making money (2.39%) is a minor motivational factor. The chi-square test results in a statistically significant but small association (0.125 for money, and 0.092 for fame).

Table 5.5: Age-dependent differences in motivation to *make money* and *become a star*.

Generation	Making money	Becoming a star
Gen Z (N = 1.839)	2.39%	8.21%
Gen Y (N = 2.572)	4.12%	4.35%
Gen X (N = 493)	9.74%	2.64%
Silver Surfers (N = 33)	21.21%	0.00%
Pearson Chi^2	$p < 0.001$	$p < 0.001$

Table 5.6: Content of streamers motivated by *making money* (N = 479).

Content category		*Pearson* Chi^2
Entertainment media	40.29%	$p < 0.001$
Chatting	21.50%	$p < 0.001$
Share information	20.88%	$p < 0.05$
24/7	19.00%	$p < 0.05$
Make music	12.94%	$p < 0.05$
Advertising	11.90%	$p < 0.001$
News	11.27%	$p < 0.001$
Sport event	8.77%	$p < 0.001$
Slice of life	6.05%	$p < 0.001$
Business information	5.64%	$p < 0.001$
Comedy	4.18%	$p < 0.001$
Nothing	3.55%	$p < 0.001$
Food	3.55%	$p = 0.062$
Animals	3.34%	$p < 0.01$
Gaming	3.13%	$p < 0.05$
Politics	2.71%	$p < 0.05$
Nature	2.30%	$p < 0.01$
Draw/Paint picture	1.67%	$p < 0.05$
STM	1.46%	$p = 0.221$
Fitness	1.46%	$p = 0.96$
Spirituality	0.21%	$p < 0.001$

5.3.5 Differences in Streamed Content

Finally, we take a look at the potential differences in content streamed by broadcasters motivated by different aspirations. Streamers motivated by money (Table5.6) provide mostly content evolving around entertainment media (40.29%), chatting (21.50%), sharing information (20.88%) and 24/7 (19.00%). Especially entertainment media is important for them. Likewise, such content is also in the Top 5 content categories for streamers motivated by fame (Table 5.7), but with only 13.29%. Even if chatting is the second most streamed content for streamers motivated by making money, it is more important for streamers motivated by becoming a star. Altogether, 67.77% of those fame-oriented streamers cover such content. This is followed by making music (42.86%), which is more popular among fame-oriented streamers than the ones motivated by money (12.94%). To share information is chosen equally often by both groups. Further noticeable differences, above 10%, exist for the categories 24/7 and news. Both were more frequently found in streams aiming for financial gain. Finally, there also exist statistically significant association between the motivational factor and most of the content categories (p ¡ 0.05) as shown in Tables 5.6 and 5.7, however, all significant associations were only of small effect (¡0.3).

Table 5.7: Content of streamers motivated by *becoming a star* (N = 301).

Content category		*Pearson Chi*^{2}
Chatting	67.77%	$p < 0.001$
Make music	42.86%	$p < 0.001$
Share information	19.93%	$p = 0.196$
Entertainment media	13.29%	$p = 0.394$
Slice of life	7.31%	$p < 0.001$
Advertising	6.31%	$p < 0.001$
Comedy	5.65%	$p < 0.001$
Nothing	4.98%	$p < 0.001$
Sport event	2.99%	$p = 0.394$
Fitness	2.33%	$p = 0.185$
Food	1.99%	$p = 0.710$
24/7	1.33%	$p < 0.001$
Business information	1.33%	$p = 0.935$
Gaming	1.33%	$p < 0.01$
Draw/Paint picture	1.00%	$p = 0.689$
Nature	0.66%	$p < 0.001$
Animals	0.66%	$p < 0.001$
News	0.33%	$p < 0.01$
Politics	0.33%	$p = 0.111$
STM	0.00%	$p = 0.085$
Spirituality	0.00%	$p < 0.01$

5.4 Discussion

This study investigated general characteristics of SLSSs streamers (on Periscope, Ustream and YouNow) motivated by fame or financial gain. For this purpose, we developed a research model and explored platform-specific characteristics (RQ1), gender-dependent differences (RQ2), origin-dependent differences (RQ3), age-dependent differences (RQ4) and contents streamed by broadcasters (RQ5) whose motivation is financial gain or fame.

From total 7,667 observed streams, 4,548 showed individual broadcasters (61.2% male and 38.8% females). The results indicate that Ustream is mostly applied by streamers motivated by financial gain, whereas YouNow by streamers aiming at becoming famous. This could also be related to generational aspects, since YouNow is a platform mostly applied by the younger generations (Friedländer, 2017b). Interestingly, the older generations (Gen X, Silver Surfers) are motivated by monetary aspects, whereas the younger ones (Gen Z) by fame. One reason for this could be the fact that nowadays the Social Media landscape and the associated attention economy is increasingly ruled by so-called micro-celebrities and influencers. These "career-paths" might often be associated with quick success, fame, appreciation, interesting offers (such as product samples, gifts), travel opportunities, the freedomto do what one likes or is interested in and also financial gain. Furthermore, such influencers and micro-celebrities often belong to the younger generations. These reasons may make it attractive for younger streamers to follow a similar path. More mature streamers may be more settled, grounded and mainly interested in the financial aspects.

According to our results, no strong association between gender and the motivation being fame or money exists. Females and males are equally interested in these aspects. There are, however, differences in content streamed by the broadcasters whose motivation is either money or fame. For streamers wanting to make money, "entertainment media" is the preferred content. We defined "entertainment media" as every action involving some form of media, e.g. displaying digital pictures, streaming a TV show or playing music. For the streamers seeking fame, the most important content categories are chatting and making music.

Since this study explores and is limited to general characteristics of SLSSs streamers and their motivation regarding fame and financial gain, further research could include qualitative interviews in order to explain our results in more depth. Besides, it would be interesting to conduct a long-term study to analyze if the streamed content (depending on the motivation) really leads the streamers to becoming a star or making money. Finally, investigation of established micro-celebrities and influencers is the next important step for our research. This study focused only on users aiming at becoming a star or influencer, however, this dream will come true only for the chosen ones.

References

Ainslie, A. (2015). *The burden of protecting live sports telecasts: The real time problem of live streaming and app-based technology*. Los Angeles, CA: Southwestern Law School.

Alamiri, D., & Blustein, J. (2016). Privacy awareness and design for live video broadcasting apps. In C. Stephanidis (Ed.), *HCI 2016, Part I. CCIS* (Vol. 617, pp. 459–464). Cham: Springer.

Bilal, K., Erbad, A., & Hefeeda, M. (2017). Crowdsourced multi-view live video streaming using cloud computing. *IEEE Access*, *5*, 12635–12647.

Bründl, S., & Hess, T. (2016). Why do users broadcast? Examining individual motives and social capital on social live streaming platforms. In *Pacific Asia Conference on Information Systems Proceedings, PACIS* (Vol. 332).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.

Dowell, C. T., & Duncan, D. F. (2016). Periscoping economics through someone else's eyes: A real world (Twitter) app. *International Review of Economics Education*, *23*, 34–39.

Faklaris, C., Cafaro, F., Hook, S. A., Blevins, A., O'Haver, M., & Singhal, N. (2016). Legal and ethical implications of mobile live-streaming video apps. In *MobileHCI 2016* (pp. 722–729). New York: ACM.

Fietkiewicz, K. J., Lins, E., Baran, K. S., & Stock, W. G. (2016). Inter-generational comparison of social media use: Investigating the online behavior of different generational cohorts. In *Proceedings of the 49th Hawaii International Conference on System Sciences* (pp. 3829–3838). Washington, DC: IEEE.

Fietkiewicz, K. J., & Scheibe, K. (2017). Good morning... good afternoon, good evening and good night: Adoption, usage and impact of the live streaming platform YouNow. *Proceedings of the 3rd International Conference on Library and Information Science*, 92–115.

Friedländer, M. B. (2017a). And action! Live in front of the camera: An evaluation of the social live streaming service YouNow. *International Journal of Information Communication Technologies and Human Development*, *9*(1), 15–33.

Friedländer, M. B. (2017b). Streamer motives and user-generated content on social live-streaming services. *Journal of Information Science Theory and Practice*, *5*(1), 65–84.

Gros, D., Wanner, B., Hackenholt, A., Zawadzki, P., & Knautz, K. (2017). World of streaming. Motivation and gratification on Twitch. In *SCSM 2017, Part I. LNCS* (Vol. 10282, pp. 44–57). Cham: Springer.

Hearn, A. (2008). 'Meat, mask, burden': Probing the contours of the branded 'self'. *Journal of Consumer Culture*, *8*(2), 197–217.

Hearn, A., & Schoenhoff, S. (2016). From celebrity to influencer: Tracing the diffusion of celebrity value across the data stream. In P. Marshall & S. Redmond (Eds.), *A companion to celebrity* (pp. 194–211). Hoboken: John Wiley & Sons.

Honka, A., Frommelius, N., Mehlem, A., Tolles, J. N., & Fietkiewicz, K. J. (2015). How safe is YouNow? An empirical study on possible law infringements in Germany and the United States. *The Journal of MacroTrends in Social Science*, *1*(1), 1–17.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observation an intimacy at a distance. *Psychology*, *19*(2), 215–229.

Ingleton, P. (2014). Celebrity seeking micro-celebrity: 'New candour' and the everyday in the sad conversation. *Celebrity Studies*, *5*(4), 525–527.

Kane, B. (2010). Balancing anonymity, polularity & micro-celebrity: The Crossroads of social networking and privacy. *Albany Law Journal of Science and Technology*, *20*(2), 327–363.

Kaser, R. (2017). *How to make money livestreaming.* Retrieved from `https://thenextweb.com`

Kavka, M. (2012). *Reality tv.* Edinburgh: Edinburgh University Press.

Kavka, M. (2015). Celevision: Mobilizations of the television screen. In P. D. Marshall & S. Redmond (Eds.), *A Companion to Celebrity* (pp. 295–314). Edinburgh: Edinburgh University Press.

Keinänen, K. (2017). *The role of live streaming in marketing communications and corporate branding* (master's thesis). Lappeenranta University of Technology, LUT School of Business and Management, Lappeenranta, Finland.

Khamis, S., Ang, L., & Welling, R. (2017). Self-branding, 'micro-celebrity' and the rise of social media influencers. *Celebrity Studies*, *8*(2), 191–208.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks: Sage.

Laerd Statistics. (2018). *Chi-square test for independence using SPSS Statistics.* Retrieved from `https://statistics.laerd.com/`

Lai, L. S. L., & To, W. M. (2015). Content analysis of social media: A grounded theory approach. *Journal of Electronic Commerce Research*, *16*(2), 138–152.

Lair, D. J., Sullivan, K., & Cheney, G. (2005). Marketization and the recasting of the professional self: The rhetoric and ethics of personal branding. *Management Communication Quarterly*, *18*(3), 307–343.

Leung, L. (2013). Generational differences in content generation in social media: The roles of the gratifications sought and of narcissism. *Computers in Human Behavior*, *29*(3), 997–1006.

Martin, R. (2018). *Cult of influence. Marketing: Advertising, media and pr in Canada.* Retrieved from `http://www.marketingmag.ca`

Marwick, A. E. (2013). *Status update: Celebrity, publicity, and branding in the social media age.* New Haven: Yale University Press.

Marwick, A. E. (2015). Instafame: Luxury selfies in the attention economy. *Public Culture*, *27*(1), 137–350.

Marwick, A. E. (2016). You may know me from Youtube: (Micro-)Celebrities in social media. In P. D. Marshall & S. Redmond (Eds.), *A Companion to Celebrity* (pp. 194–212). Hoboken: John Wiley & Sons.

Marwick, A. E., & Boyd, D. (2010). I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*, *13*(12), 114–133.

Marwick, A. E., & Boyd, D. (2011). To see and be seen: Celebrity practice on Twitter. *Convergence*, *17*(2), 139–158.

Maugeri, R., Giammalva, R. G., & Iacopino, D. G. (2016). On the shoulders of giants, with a smartphone: Periscope in neurosurgery. *World Neurosurgery*, *92*, 569–570.

Muntean, N., & Petersen, A. H. (2009). Celebrity twitter: Strategies of intrusion and disclosure in the age of technoculture. *M/C Journal*, *12*(5).

Page, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse and Communication*, *6*(2), 181–201.

Recktenwald, D. (2017). Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics*, *115*, 68–81.

Rodríguez-Gil, L., García-Zubia, J., Orduña, P., & López-Ipiña, D. (2017). An open and scalable web-based interactive live-streaming architecture: The WILSP platform. *IEEE Access*, *5*, 9842–9856.

Schaefer, M. (2012). *Return on influence: The revolutionary power of klout, social scoring, and influence marketing.* New York: McGraw Hill.

Scheibe, K., Fietkiewicz, K., & Stock, W. G. (2016). Information behavior on social live streaming services. *Journal of Information Science Theory and Practice*, *4*(2), 6–20.

Senft, T. (2008). *Camgirls: Celebrity and community in the age of social networks.* New York: Peter Lang.

Senft, T. M. (2013). Microcelebrity and the branded self. In A. Burgess J. E.and Bruns (Ed.), *A Companion to New Media Dynamics* (pp. 346–354). Malden: Blackwell.

Stohr, D., Li, T., Wilk, S., Santini, S., & Effelsberg, W. (2015). An analysis of the YouNow live streaming platform. In *40th Local Computer Networks Conference Workshops* (pp. 673–679). Washington, DC: IEEE.

Stohr, D., Toteva, I., Wilk, S., Effelsberg, W., & Steinmetz, R. (2017). User-generated video composition based on device context measurements. *International Journal of Semantic Computing*, *11*(1), 65–84.

Tang, J. C., Kivran-Swaine, F., Inkpen, K., & Van House, N. (2017). Perspectives on live streaming: Apps, users, and research. In *Conference on Computer-Supported Cooperative Work and Social Computing, CSCW 2017* (pp. 123–126). New York: ACM.

Tang, J. C., Venolia, G., & Inkpen, K. M. (2016). Meerkat and Periscope: I stream, you stream, apps stream for live streams. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4770–4780).

Tufekci, Z. (2013). "Not this one": Social movements, the attention economy, and microcelebrity networked activism, journal=American Behavioral Scientist, volume=57, number=7, pages=848–870, language=English,.

West, L. (n.d.). *These YouTube stars you've never heard of have millions of teen fans [Blog post].* Retrieved from `http://jezebel.com/these-youtube-stars`

-youve-never-heard-of-have-millions-1493742066.s
YouNow. (2018). Retrieved from https://www.younow.com/partners
YouTube. (2018). Retrieved from https://support.google.com/youtubegaming/
Zimmer, F., Fietkiewicz, K. J., & Stock, W. G. (2017). Law infringements in social live streaming services. In T. Tryfonas (Ed.), *Human Aspects of Information Security, Privacy and Trust* (Vol. 10292 LNCS, pp. 567–585). Cham: Springer.

Part II

Scientometrics

Chapter 6

Relative Visibility of Authors' Publications in Different Information Services

Publication hit lists of authors, institutes, scientific disciplines etc. within scientific databases like Web of Science or Scopus are often used as a basis for scientometric analyses and evaluations of these authors, institutes etc. However, such information services do not necessarily cover all publications of an author. The purpose of this article is to introduce a re-interpreted scientometric indicator called "visibility," which is the share of the number of an author's publications on a certain information service relative to the author's entire oeuvre based upon his/her probably complete personal publication list. To demonstrate how the indicator works, scientific publications (from 2001 to 2015) of the information scientists Blaise Cronin (N = 167) and Wolfgang G. Stock (N = 152) were collected and compared with their publication counts in the scientific information services ACM, ECONIS, Google Scholar, IEEE Xplore, Infodata eDepot, LISTA, Scopus, and Web of Science, as well as the social media services Mendeley and ResearchGate. For almost all information services, the visibility amounts to less than 50%. The introduced indicator represents a more realistic view of an author's visibility in databases than the currently applied absolute number of hits in those databases.

6.1 Introduction

In informetrics and scientometrics it is known that all information services including general science databases as Web of Science (WoS) or Scopus as well as domain-specific services as Medline for medicine are incomplete (Hilbert et al., 2015). They are biased towards some languages, sources and scientific disciplines. However, all those information services are applied for empirical studies, for the description and evaluation of single scientists, institutes, scientific disciplines, cities, countries, etc. The fundaments of those endeavors are absolute numbers of publications found in the (biased) information services. Nowadays, this absolute number of publications is a picture of the "visibility" of a scientist, an institute and so on. Let us set an example: Two institutions, say A and B, sum up to 10 publications in an information service IS_1 each. In a first (absolute) interpretation of "visibility," both have a visibility of 10. Now we go to the institutional repositories and are able to identify 20 publications of A and 100 publications of B. With this additional information in mind, we can re-calculate the institutes' visibility as relative frequencies. Now, institute A has a relative visibility of 0.5, while B's value is only 0.1 in IS_1. Of course, we are able to repeat those calculations for further information services $IS_2, \ldots, IS_n$ in order to compare the respective values of relative visibility. While the absolute visibility is a well-known and aged metrics, the indicator of relative visibility (of an author, an institute, etc.) is a novel (and more realistic) metrics in scientometrics and informetrics.

For informetric and scientometric analyses, correct empirical data are fundamental to their accuracy. Many scientometric studies rely on publication counts of scientists, institutions, etc. According to the focus of research, different criteria and evaluation methods are selected and affect an author's (institute's, etc.) visibility. According to Stock (2000) there is no standardized system for these analyses. Depending on data selection and methods, the results can vary greatly, and hence, many analyses are not comparable. For a publication analysis of, for example, an author or an institution, the publication count is the basis for all further analyses regarding productivity and impact (Dorsch & Frommelius, 2015). Such analyses are often based on the content of information services, such as WoS or Scopus, which simply and rapidly provide the publications of a certain author or an institute. Especially WoS and Scopus are broadly recognized as authoritative indexes for publication and citation studies (Wildgaard, 2015). However, such a foundation could be criticized because the entire publication analysis is based only on the accessible content of the respective information service. An author can have additional publications that are not listed in the database or the researcher may not have full access to all collections within the database (e.g., WoS's Book Citation Index) and consequently, publications may be omitted. This makes it difficult to compare different publication counts, as well as citation counts and h-index analyses, because the completeness of the publication lists can vary from author to author and from information service to information service. "The coverage of journals in cited reference enhanced databases can be surprisingly uneven," Jacsó (2008, p. 278) states. An author's (institute's, etc.) visibility is dependent on the absolute number of publications within a database. Therefore, a "metric-wise" author tends to publish only in sources that are covered by certain information services (especially WoS and Scopus) in order to get or maintain high visibility (Rousseau & Rousseau, 2015).

Scientometric studies and database providers often start with the condition that the scope of the study or of the database is that of "quality papers" instead of all papers. Using WoS or Scopus as sources of the data would provide "quality papers"—and the definition of "quality papers" is that papers being listed in WoS are those of quality. "There is something obviously wrong with this circular argument," one of the reviewers stated. The reviewer continues, "Nevertheless, I believe some quality threshold should be used in any bibliometric research." Perhaps such a quality threshold is given by peer review. Only papers, which survived peer review, are able to be considered as "quality papers."

"Visibility" can be defined in many ways. For Cole and Cole (1968, p. 398), it is an indicator that "characterizes the men being looked at." They define visibility through "how well known" a scientist is and apply questionnaires as their methodology (Cole & Cole, 1968, p. 398). Ingwersen (2000) limits visibility to an author's publications, i.e., the absolute number of publications in the National Science Indicators (NSI) database. NSI is derived from WoS, and consequently, the visibility is dependent on an author's publication count within this information service. Schlögl (2013) also defines visibility as the absolute number of publications in an information service (again, in WoS). Miguel, Chinchilla-Rodriguez, and De Moya-Anegõn (2011, p. 1130) consider visibility as a valuing criterion that states "how avidly published work is received by the academic or scientific community." However, visibility may not only refer to publication and citation counts within established academic databases. Social media services (e.g., Mendeley or CiteULike) can also be used to study the visibility of an author (Bar-Ilan et al., 2012), and thus, the use of social media for scientific purposes can increase an author's visibility (Fitzgerald & Radmanesh, 2015).

This article introduces a re-interpreted visibility indicator from a scientometrist's viewpoint. Readers or researchers may expect also the accessibility of the examined databases as well as the access to full texts as criteria of visibility. This would be a further indicator; our indicator only settles for the existence of bibliographic data in certain information services.

We work with publication counts in information services and additionally with personal or institutional publication lists as the means of calibration (Hilbert et al., 2015; Kirkwood, 2012). Depending on the content of an information service, the visibility values are different. With our indicator, we are able to show comparable visibility rates concerning general databases (as WoS or Scopus), discipline-specific databases (e.g., LISTA for library and information science), and country-specific databases (e.g., the German repository Infodata eDepot for information science literature). Additionally, social media sources (as reference management services) cover scientific literatures. The indicator can be used for any kind of scientometric analysis of scientific institutions and their aggregates (single researchers, institutes, cities, regions, countries, etc.). Of course, there is a precondition— the metrics of relative visibility needs access to all personal or institutional publication lists.

6.2 Methods

To show how the metrics work, we performed a scientometric analysis of the information scientists Blaise Cronin and Wolfgang G. Stock. There are two sources of data: (1) the personal publication lists containing all publications of Cronin and Stock; and (2) the publications of both authors covered by a variety of information services.

(1) For the creation of the personal publication lists, Cronin and Stock were asked for information regarding their publications. It is also possible to receive the publication information from an author's personal or institutional website. However, since such publication lists are not necessarily complete, we checked the lists' completeness (through online searches in different databases) and scientific nature (e.g., identification and omission of not scientific literature like novels). Web/blog contributions and self-published novels were excluded from the analysis. For the purpose of simplification, all entries were counted as 1 (regardless of document type, length, and number of co-authors). (2) To determine the author's publication count within the mentioned information services, search arguments were formulated (Table 6.1).

With the help of the personal publication lists, they were manually controlled with respect to their correctness. With these two sources, the visibility can be determined by a comparison between the author's coverage in a certain information service IS and the complete publication count in the publication list. The information service specific author's visibility is calculated as follows:

$$Relative\ Visibility(IS) = \left(\frac{d}{r}\right) \times 100,$$

where d is the total number of an author's publications within the information service and r is the number of publications in the personal publication list of the same author.

This study refers to all scientific publications of Cronin[1] and Stock[2] within the time period from January 1, 2001, until December 31, 2015. Both authors were chosen because they publish in the same subject area and are of similar age and position (however, Cronin retired in 2014). The selected information services cover the subject areas of the authors as well as a wide spectrum of database types. Scopus and WoS are commercial multidisciplinary citation databases with a thematically widespread focus. ACM Guide to Computing Literature serves as a database for computer science, as does IEEE Xplore, but the former with a focus on more technical literature. LISTA focuses on library and information science while ECONIS covers economics including literature on information markets and information economics. Infodata eDepot is a repository that focuses on information science, but with special regard to German publications. Google Scholar is a web search engine indexing scholarly literature. Mendeley as academic social reference management system and ResearchGate as academic social networking site for scientists represent social media services. Since "Cronin B" is homonymous in several databases, we additionally look for the "right" Cronin (e.g. by adding affiliation details to the

[1] Personal communication (January 11, 2016).

[2] `http://www.isi.hhu.de/abteilungen/abteilung-fuer-informationswissenschaft/personal/professoren-dozierende/wolfgang-g-stock/publikationen-stock.html` and `http://www.isi.hhu.de/abteilungen/abteilung-fuerinformationswissenschaft/personal/professoren-dozierende/wolfgang-g-stock/publikationen-stock/archiv.html#c279043`.

Table 6.1: Relative visibility of Cronin and Stock in selected information services.

Information service	Search argument		Publication count		Visibility	
	Cronin	Stock	Cronin	Stock	Cronin (%)	Stock (%)
ACM Guide to Computing Literature	(+Blaise +Cronin)	(+Wolfgang +G. +Stock)	45	12	26.9	7.9
ECONIS	Cronin, Blaise	Stock, Wolfgang G.	0	9	0	5.9
Google Scholar	Blaise Cronin (User profile)	Wolfgang G. Stock (User profile)	119	83	71.3	54.6
IEEE Xplore	Blaise, Cronin	Wolfgang G., Stock	1	2	0.6	1.3
Infodata eDepot	Cronin	Stock, Wolfgang G.	0	49	0	32.2
LISTA	Cronin, Blaise	Stock, Wolfgang G.	69	16	41.3	10.5
Mendeley	"Blaise Cronin"	"Stock W G" "Wolfgang G Stock"	51	46	30.5	30.3
ResearchGate	Blaise Cronin	Wolfgang G. Stock	40	49	24.0	32.2
Scopus	Cronin, B Author IDs: 24351054500 55605719900 57093355600	Stock, Wolfgang G. Author ID: 8658221400	98	51	58.7	33.6
Web of Science (Core collection)	Cronin B*	Stock WG	115	20	68.9	13.2

N = 167 (Cronin) and N = 152 (Stock) publications;
* Truncation

search arguments). We searched for both authors in summer 2016 except for Google Scholar (February 2017).

6.3 Visibility of Cronin and Stock as Case Studies

Concerning their personal publication lists, Cronin has published 167 works during the specified time period of 15 years and Stock has published 152 works. Table 6.1 shows the publication count within the information services and the calculated relative visibility for each database. None of the information services list all publications from both authors. The highest visibility was found for Cronin and his publications listed in Google Scholar and WoS, but even in these information services, nearly a third of his publications are missing. With a visibility of 54.6%, the publications of Stock produced the highest visibility value in Google Scholar. Only about a half of his publications are covered. In the commercial academic database Scopus it is only a third (33.6%). The lower value compared to Cronin can be explained by the fact that 84 out of his 152 publications are written in German (Table 6.2). For scientific publications in information science, English is the *lingua franca*.

Table 6.2: Stock's language-specific visibility in selected information services (N = 152).

Information service	English (N = 67) (%)	German (N = 84) (%)	Hungarian (N = 1) (%)
ACM Guide to Computing Literature (N = 12)	17.9	0	0
Econis (N = 9)	1.5	9.5	0
Google Scholar (N = 83)	83.6	32.1	0
IEEEXplore (N = 2)	3.0	0	0
Infodata eDepot (N = 49)	26.9	36.9	0
LISTA (N = 16)	16.4	4.8	100.0
Mendeley (N = 46)	47.8	16.7	0
ResearchGate (N = 49)	58.2	11.9	0
Scopus (N = 51)	56.7	15.5	0
Web of Science (Core collection) (N = 20)	26.9	2.4	0
Average visibility over all information services	33.9	13.0	10.0

Other languages, including German, do not have the same standing within the scientific community and are less likely to be listed in international information services than publications in English (Gordon, 2012). Also the language-specific visibility for Stock (Table 6.2) confirms this with a higher relative visibility value for his English publications.

However, the visibility values of Cronin (Table 6.1)—whose publications are exclusively written in English—show that renowned English-writing authors are also not sufficiently covered in the databases. German information services such as In-

Table 6.3: Cronin's document-type specific visibility in selected information services (N = 167).

Information service	Articles in proceedings (N = 9) (%)	Book chapters (N = 18) (%)	Journal articles (N = 93) (%)	Editorials (N = 23) (%)	Other (N = 24) (%)
ACM Guide to Computing Literature (N = 45)	0	0	30.1	56.5	16.7
Econis (N = 0)	0	0	0	0	0
Google Scholar (N = 119)	33.3	16.7	82.8	95.7	58.3
IEEEXplore (N = 1)	0	0	0	0	4.2
Infodata eDepot (N = 0)	0	0	0	0	0
LISTA (N = 69)	11.1	0	54.8	52.2	20.8
Mendeley (N = 51)	33.3	5.6	45.2	21.7	0
ResearchGate (N = 40)	33.3	0	31.2	21.7	12.5
Scopus (N = 98)	33.3	5.6	78.5	91.3	0
Web of Science (Core collection) (N = 115)	77.8	5.6	83.9	87.0	37.5
Average visibility over all information services	22.2	3.3	40.6	42.6	15.0

Table 6.4: Stock's document-type specific visibility in selected information services (N = 152).

Information service	Articles in proceedings (N = 30) (%)	Book chapters (N = 15) (%)	Journal articles (N = 91) (%)	Other (N = 16) (%)
ACM Guide to Computing Literature (N = 12)	13.3	0	8.8	0
Econis (N = 9)	0	13.3	3.3	25.0
Google Scholar (N = 83)	80.0	33.3	50.5	50.0
IEEEXplore (N = 2)	6.7	0	0	0
Infodata eDepot (N = 49)	36.7	40.0	33.0	12.5
LISTA (N = 16)	0	6.7	16.5	0
Mendeley (N = 46)	36.7	6.7	33.0	25.0
ResearchGate (N = 49)	56.7	13.3	33.0	0
Scopus (N = 51)	46.7	6.7	39.6	0
Web of Science (Core collection) (N = 20)	16.7	0	16.5	0
Average visibility over all information services	29.3	12.0	23.4	11.3

fodata eDepot show a higher visibility for Stock than for Cronin who is not covered at all in this repository. Regarding the social media services, both authors receive visibility scores in the lower third. These authors could increase their visibility to 100% if they would upload all their publications but, obviously, both had neglected to do so, although some of their co-authors did.

A more detailed consideration of Cronin‘s publications shows that the document-type specific visibility in the analyzed information services is highest for editorials (especially in Google Scholar, WoS and Scopus) and journal articles (in WoS, Google Scholar and Scopus) (Table 6.3). For Stock this visibility is highest for proceedings, followed by journal articles (both in Google Scholar) (Table 6.4). Relative visibility also depends on a publication’s document type.

6.4 Conclusion and Outlook

This article introduces a re-interpreted visibility indicator and reveals the publication visibility misbalance in information services. In a case study, it demonstrates that the publication visibility of Cronin and Stock within academic databases and social media does not correspond to their real publication count as seen in their personal publication lists. It becomes apparent, that relative visibility stands in dependency with a publication’s document type and language. Especially journal publications got a high relative visibility in the considered information services. The publication language (here, English versus German) seems to decrease the visibility. At this point, the epistemic significance of culture and place (Cronin, 2003, 2008) enters the scene, but publications written in English are also affected by limited relative visibility. Therefore, it could be questioned how meaningful scientometric analyses are when they are solely based on the limited publication coverage of information services. Our re-interpreted visibility indicator provides, in combination with the use of personal publication lists, the possibility of including up to 100% of an author’s publications. The publication lists offer a high degree of reliability since they arise by or in cooperation with the respective authors. Only faulty declarations by the author have a negative impact on the reliability (e.g., the author forgets a publication, states wrong publication year, etc.), but such mistakes can happen even within the databases. With the metrics of relative visibility, researchers have means to know on the reliability of data derived from different information services.

This article has limitations insofar only two authors are analyzed as case studies. In future, more data on relative visibility on the levels of authors and institutions should be gathered.

In some cases one is unable to obtain the same number of hits indicated in the tables with the stated search queries. The reason is that some of the free sources (especially Google Scholar and the social media services) are adding older literature on occasion. The manual check of the search query results also reveals some database inaccuracies regarding to the obtained publications. Some publications were covered several times or were simply wrong. For instance, an editor activity was not denoted as such, but rather every contribution of the edited work was listed as the own publication of the editor. Some publications contained in the hit lists were not written by the author. For further studies the exact query date and a declaration of the total and adjusted hits and in the best case a full list of all queries with their results (additionally DOIs) placed in a repository could be helpful for a better transparency of the evaluated data.

It is also possible to use the total count of an author's publications in any digital information service (i.e., the union of all database-specific hit lists) relative to the personal publication list as a further indicator ("combined overall relative visibility").

There is an intermediate solution between complete personal publication lists and bibliographic selection lists from single databases such as WoS, which is based on the "combined overall relative visibility." The share of a specific database's results on an author's name relative to the total amount of results in any information service is another indicator.

The introduced visibility indicators refer to a bibliographic perspective. It could be expanded to full access of the databases and the full texts for everyone unimpeded by paywalls and commercial interests. However, it should be considered that not everyone has the same access conditions.

Nowadays, many personal publication lists can be found on the Web but they are not available in a standardized format. We see two solutions for this problem. First, as publication lists are open data, this could be a task for future research into linked open data (Xin, Hassanzadeh, Fritz, Sohrabi, & Miller, 2013). Second, an institution (e.g., a scholarly society) provides a platform where authors can upload and update their publication lists (Holl, Makara, Micsik, & Kovacs, 2014; Jones et al., 2011). However, both solutions of self-archiving papers could cause some inhomogeneity when authors chose different field schemata and different quality criteria for their personal publication lists. One list could contain only major research papers; another could contain a full bibliography. One author stresses complete bibliographic data, while another skips pagination and issue number. For both proposals, the inclusion of all publications (differenced by peer-reviewed "quality papers" and non-peer-reviewed "other papers") containing an accurate quality labeling is required.

Acknowledgements

The author would like to thank Stefanie Haustein for the access to the Web of Science core collection. Special thanks go to the reviewers. I am very grateful for your feedback and new insights for this study.

References

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social web. In *Proceedings of* *the 17th International Conference on Science and Technology Indicators* (pp. 98–109). Montreal: Science-Metrix et OST.

Cole, S., & Cole, J. R. (1968). Visibility and the structural bases of awareness of scientific research. *American Sociological Review*, *33*(3), 397–413.

Cronin, B. (2003). Scholarly communication and epistemic cultures. *The New Review of Academic Librarianship*, *9*(1), 1–24.

Cronin, B. (2008). On the epistemic significance of place. *Journal of the American Society for Information Science and Technology*, *59*(6), 1002–1006.

Dorsch, I., & Frommelius, N. (2015). A scientometric approach to determine and analyze productivity, impact and topics based upon personal publication lists. In F. Pehar, C. Schlögl, & C. Wolff (Eds.), *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th-21st May 2015* (pp. 578–580). Glückstadt, Germany: Hülsbusch. (Available online: `https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/heck/Bella_Nils.pdf`)

Fitzgerald, R. T., & Radmanesh, A. (2015). Social media and research visibility. *American Journal of Neuroradiology*, *36*(4), 637.

Gordon, A. (2012). The invisibility of science publications in hebrew: A comparative database study. *Journal of the American Society for Information Science and Technology*, *63*(3), 607–615.

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., . . . Stock, W. G. (2015). Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. *Online Information Review*, *39*(2), 255–264.

Holl, A., Makara, G., Micsik, A., & Kovacs, L. (2014). MTMT: The Hungarian scientific bibliography. In *Uses of open data within government for innovation and efficiency, 2014.06.30–2014.07.01* (p. 5). Samos.

Ingwersen, P. (2000). The international visibility and citation impact of Scandinavian research articles in selected social science fields: The decay of a myth. *Scientometrics*, *49*(1), 39–61.

Jacsó, P. (2008). The plausibility of computing the h-index of scholarly productivity and impact using reference-enhanced databases. *Online Information Review*, *32*(2), 266–283.

Jones, R., MacGillivray, M., Murray-Rost, P., Pitman, J., Sefton, P., O'Steen, B., & Waites, W. (2011). *Open bibliography for science, technology, and medicine.* Cambridge: University of Cambridge.

Kirkwood, P. E. (2012). Researcher publication checklist: A quantitative method to compare traditional databases to web search engines. In *119th American Society for Engineering Education Annual Conference and Exhibition, San Antonio, TX, June 10-13, 2012.* (Available online: `https://peer.asee.org/facultypublication-checklists-a-quantitative-method-to-compare-traditional-databases-to-web-search-engines`)

Miguel, S., Chinchilla-Rodriguez, Z., & De Moya-Anegõn, F. (2011). Open access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the American Society for Information Science and Technology*, *62*(6), 1130–1145.

Rousseau, S., & Rousseau, R. (2015). Metric-wiseness. *Journal of the American Society for Information Science and Technology*, *66*(11), 2389.

Schlögl, C. (2013). International visibility of European and in particular German-language publications in library and information science. In H.-C. Hobohm (Ed.), *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Tagungsband des 13. internationalen Symposiums für Informationswissenschaft (ISI 2013)* (pp. 51–62). Glückstadt: Hülsbusch. (Available online: `https://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/396`)

Stock, W. G. (2000). Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In K. Fuchs-Kittowski, H. Laitko, H. Parthey, & W. Umstätter (Eds.), *Wissenschaft und Digitale Bibliothek. Wissenschaftsforschung Jahrbuch 1998* (pp. 239–282). Berlin, Germany: Gesellschaft für Wissenschaftsforschung. (Available online: `http://www.wissenschaftsforschung.de/JB98_239-282.pdf`)

Wildgaard, L. (2015). A comparison of 17 author-level bibliometric indicators for researchers in Astronomy, Environmental Science, Philosophy and Public Health in Web of Science and Google Scholar. *Scientometrics*, *104*(3), 873–906.

Xin, R. S., Hassanzadeh, O., Fritz, C., Sohrabi, S., & Miller, R. J. (2013). Publishing bibliographic data on the semantic web using bibbase. *Semantic Web*, *4*(1), 15–22.

Chapter 7

Truebounded, Overbounded, or Underbounded? Scientists' Personal Publication Lists versus Lists generated through Bibliographic Information Services

A truebounded publication list of a scientific author consists of exactly all publications that meet two criteria: (1) they are formally published (e.g., journal article or proceeding paper); (2) they have scientific, scholarly, or academic content. A publication list is overbounded if it includes documents which do not meet the two criteria (such as novels); a publication list is underbounded if it is incomplete. Are authors' personal publication lists, found on their personal sites on the Internet or in institutional repositories, truebounded, overbounded, or underbounded? And are the respective publication lists generated through bibliographic information services truebounded, overbounded, or underbounded? As case studies, publications of nine International Society of Scientometrics and Informetrics (ISSI) Committee members (published between 2007 and 2016) were collected to create preferably complete personal publication lists according to the two criteria. We connect the "relative visibility of an author" with the concepts of truebounded, overbounded, and underbounded publication lists. The authors' relative visibility values were determined for the information services Web of Science (WoS), Scopus, and Google Scholar and compared to the relative visibility of the authors' personal publication lists. All results of the bibliographic information services are underbounded. Relative visibility is highest in Google Scholar, followed by Scopus and WoS.

7.1 Introduction

Many studies on the measurement of research outputs are based on publication counts of scientists, institutions, cities, countries, and topics. Common scientometric indicators are scientific productivity (by the measure of publication counts), scientific impact (by the measure of citation counts), and mentions in social media (by the measures provided by altmetrics). The basic indicator is the publication count which determines the scientist's, the institutionß's, etc. (in the following, "unit of assessment") visibility. A publication has to be visible in order to be cited; it has to be visible as well in order to be mentioned on any social media. The quality of all mentioned measures depends on the completeness of the unit of assessment's publication list. In this article, we introduce an indicator for the analysis and evaluation of publication lists: boundedness. The boundedness of publication lists consists of three manifestations:

- truebounded publication lists;
- overbounded publication lists;
- underbounded publication lists.

A truebounded publication list of a scientific author (or of any other unit of assessment) consists of all publications that meet two criteria: (1) they are formally published (e.g., as a journal article, an article in conference proceedings or anthologies, or as a book); (2) they have scientific, scholarly, or academic content. A publication list is overbounded if it includes documents that do not meet the two criteria (e.g., unpublished documents or novels); a publication list is underbounded if it is incomplete. It is possible that a publication list is both underbounded (missing publications) and overbounded (including documents not meeting our two criteria or publications which do not originate from the particular author).

We borrowed the terms of boundedness from city research (International Urban Research, 1959, pp. 6-7). Boundedness describes the relations between metropolitan areas and its administration unit(s). In an underbounded city, the administration unit is smaller than the entire metropolitan region. An example of an underbounded city is New York; as the metropolis covers not only the city of New York, NY, but additionally parts of New Jersey and large areas of Long Island. The overbounded city is larger than the core city and includes rural areas and further smaller towns, like Chóngqìng in China (an area with more than 82 k km2—as large as Austria). The ideal is the truebounded city, where administration unit and metropolitan region match (e.g., Singapore, SG).

To determine the state of boundedness, we address the measure of relative visibility (Dorsch, 2017). Relative visibility is the share of the number of the unit of assessment's publications on a certain information service or repository relative to the unit of assessment's entire oeuvre. It is possible to calculate relative visibility according to results lists of scientific information services (say, Web of Science, Scopus, or Google Scholar) or according to publication lists found on the Web which are provided by the author him- or herself or by institutional repositories. Personal publication lists apply different criteria for the arrangement of the documents. Among others, they could be arranged in inverse chronological order, by document type, or by subject (Kretschmer & Aguillo, 2004).

As we are going to work only on case studies on the author level, we are not going to consider other units of assessment in this paper. However, the lessons to learn for those levels will be probably the same. This article pursues answering two research questions (RQs).

- (RQ 1). Are authors' personal publication lists, found on their personal sites on the Internet or on institutional repositories, truebounded, overbounded, or underbounded?
- (RQ 2). Are the respective publication lists generated through bibliographic information services truebounded, overbounded, or underbounded?

7.1.1 Relative Visibility

In informetrics and scientometrics, it is well known that all information services, including general science databases such as Web of Science (WoS), Scopus, or Google Scholar as well as domain-specific bibliographic services such as for instance Medline or EMBASE for medicine, are incomplete (Hilbert et al., 2015). "The coverage of journals in cited reference enhanced databases can be surprisingly uneven," Jacsó (2008, p. 278) states. So completeness can vary from author to author as well as from database to database (Chen, 2010; Falagas, Pitsouni, Malietzis, & Pappas, 2008; Harzing, 2014; Mongeon & Paul-Hus, 2016). Furthermore, not every researcher has the same database collection access (e.g.,WoS with or without Book Citation Index). This makes it difficult to compare such studies and, based on them, the visibility of authors. Regarding the concept of "metric-wiseness" as proposed in an opinion letter by Rousseau and Rousseau, "metric-wise" authors may tend to publish only in sources that are covered by certain information services (especially WoS and Scopus) to achieve a high visibility (Rousseau & Rousseau, 2015).

Databases like WoS or Scopus are rated as "quality" information services. Every covered publication is called a "quality paper," so that bibliometric studies applying WoS or Scopus are always focused on "quality papers" of an author. What is a "quality paper"? A paper indexed in WoS or Scopus is considered a "quality" paper because the publishing journal has been assessed and was found to meet a series of quality thresholds. Why are WoS and Scopus "quality information services"? Because they include "quality papers." That is why some authors prefer to speak of "mainstream journals" (Gaillard, 1992) instead of "quality journals" that are included in such information services.

What is visibility? There exists a variety of definitions for this concept. For Cole and Cole (1968, p. 398), it indicates through questionnaires "how well known" a scientist is and "characterizes the men (nowadays, of course, women as well) being looked at." Ingwersen (2000) limits visibility to an author's publications and thus the absolute number of publications in the National Science Indicators (NSI) database. NSI is derived from WoS, so the visibility is dependent on an author's publication count within this information service. The same applies to Schlögl (2013) who defines visibility as the absolute number of publications in an information service (again, in WoS). However, visibility may not only refer to publication and citation counts within established academic databases. Social media services (like Twitter or Mendeley) can also be used to study the visibility of an author(Thelwall, Haustein, Larivière, & Sugimoto, 2013), so the use of social media for scientific purposes can increase an author's visibility. There is one further different approach. Dorsch (2017), following Gaillard (1992), Kirkwood (2012), and Hilbert et al. (2015), banks on personal publication lists published on authors' or institutions' websites. Complete personal publication lists can also allow for a comprehensive picture of a scientific institutions' research activities (Dorsch, Schlögl, Stock, & Rauch, 2017).

If we are able to collect 100 percent of an author's publications (in order to create a truebounded publication list), the indicator of relative visibility of the author in the different databases and in his or her personal publication list arises. With the total number of an author's publications within a database, homepage or repository (d) and the union of all publications in all databases and in the personal publication list of the same author (r), the database-specific author's visibility can be calculated:

$$Relative\ Visibility(Author,\ IS) = \left(\frac{d}{r}\right) \times 100$$

where IS is an information service (such as WoS or Scopus) as well as a personal publication list. If the visibility equals 100, the publication list is truebounded; if the visibility is below 100, the list is underbounded (or in other words, there are missing items); if it is above 100, the list is overbounded (thus, there are items which do not meet the criteria).

As it is possible that a database includes "false" hits (such as papers erroneously attributed to the author) and at the same time misses articles, the overall value of relative visibility may be misleading. Say, five articles are missing and simultaneously there are five "wrong" documents in the publication list, the relative visibility would be 100, which definitely does not reflect the whole story. Therefore, we have to take a closer look towards over- and underbounded publication lists.

7.1.2 What Is a Publication?

The crucial question for all publication lists of scientists is: What is a (scientific, academic, scholarly) publication? And what is not such a publication (Stock, 2000)? For the decision for or against the acceptance of a document as an author's publication, we propose two criteria, namely (1) the rule of formal publishing and (2) the rule of scientific content. Of course, there is an additional (rather self-evident) norm: in the publication's by-line, the name of the author is stated.

With the emergence of social media, the concept of "publication" changed. Tokar (2012) distinguishes between publications (also by scientists) on the social web (for instance on Facebook, Twitter, or Reddit) and academic or scholarly publications in "classical" (online as well as offline) media (like in journals, proceedings, or books). Only academic publications are formally published, while authors can publish their documents on the social web without any formal gatekeeping instance.

Also preprints on arXiv or other platforms are not formally published. Only formally published documents can be considered as publications in authors' or databases' publication lists.

What is a scientific (formally published) document? In philosophy of science, many authors have discussed criteria for demarcation between science and non-science or pseudo-science. For Carnap (1931), only reasonable (empirical as well as formal) sentences are able to be scientific; Popper (1976) calls only falsifiable propositions scientific; for Stegmüller (1973), science is the rational search for truth; for Haller (1989), it is adequacy in practice; and, finally, for OECD's Frascati Handbook (1963), new knowledge and new applications determine science (Stock, 2000, pp. 252–256). Chase (1970) found normative criteria for scientific publications such as logical rigor, replicability of research techniques, clarity and conciseness of the writing style and originality, among others. In scientometrics, it would be very problematic to check demarcation and normative criteria in every single case. We propose to be guided by the scientific, academic, or scholarly character of the publishing source and to exclude all documents, which are not published in such media. Applying this rule, Dorsch (2017) skipped all novels from the publication list of a distinguished scientist.

Following Tillett (2001), the translation of a document is an independent expression of a work, so that translated publications are to be considered as separate publications. Moreover, revised proceedings or journal publications that appear at a different place were counted as independent publications.

7.2 Materials and Methods

We investigated nine authors of the ISSI Scientific Committee in order to determine their relative visibility inWoS, Scopus, and Google Scholar as well as in their personal publication lists as case studies. (The goal of this article is not to present a visibility study on selected authors, but only to test if the proposed method will generally work). The author selection is based on the ISSI Scientific Committee List, included in the Proceedings of ISSI 2013 ("Scientific Committee", 2013). From the listed 200 members, we excluded authors with few publications and authors without personal publication lists. Finally, nine scholars that meet the above-mentioned criteria and that are well-known in the field of informetrics were selected. We chose ISSI Scientific Committee authors because they are researchers in the same subject area—the field of information science—and therefore widely comparable. Therefore, two sources of data are required, namely the personal publication list of each author as well as their publication lists in diverse information services.

A personal publication list consists of the publications' metadata (for example, title, author(s), document type, volume, and publisher for each publication). For the creation of the truebounded personal publication lists, we used the publication information on the authors' personal or institutional website. For authors who do not report on her/his publication online, it would be also possible to directly request this information from the author. We checked the lists' completeness through online searches in different databases (among others, WoS, Scopus, Google Scholar, ACM Digital Library, and LISTA). We skipped passages in the personal publication lists that obviously did not contain publications, such as the "Media Coverage and Reviews" paragraph in Haustein's publication list. We selected all scientific publications published between 1 January 2007 and 31 December 2016 from the personal publication list of each author and from information services corresponding to the stated selection criteria.

For Criteria 1, our lists include the following formally published documents: books chapters, monographs, proceedings (also including poster or workshop contributions if they were published in the proceedings), journal articles, editorials, and reviews. Edited material, web/blog contributions (not to be confused with formally, but only online published articles), first online/in-press articles, white papers, reports, preprints, lectures, talks, and all other informally or not-yet-published materials were excluded from analysis. For Criteria 2, we checked the scientific content applying title lists of scientific, academic, or scholarly recognized sources (e.g., journal title, title of proceedings, and publishing houses). We systematically excluded fiction or novels. To simplify the analysis, all entries were counted as 1 (regardless of document type, length, and number of co-authors). Additionally found publications (for instance, on Google Scholar) could not be considered for the truebounded publication list when there was no information as to whether the document had been published or not (such as the declaration of credits on the relevant journals' websites or in the relevant conference proceedings). However, we certainly marked such occurrences for the analysis of overbounded lists.

Based on the generated truebounded publication lists, the relative visibility in-WoS (Core Collection), Scopus, and Google Scholar could be determined. The databases represent a selection of widely used fee-based and free search services and were the subject of many visibility studies. Scopus and WoS are commercial multidisciplinary citation databases with a thematically widespread focus. However, WoS is a more selective index, since it includes about 18,200 journals and proceedings in its Core Collection. According to Mongeon and Paul-Hus (2016), there were exactly 13,605 journals in WoS in 2014. With about 22,800 serial titles, 20,346 of which are journals (Mongeon & Paul-Hus, 2016), Scopus is more inclusive. Compared to Ulrich's with around 63,013 active academic journals (Mongeon & Paul-Hus, 2016), both databases cover only minor parts of the entire scientific production. Google Scholar is a free web search engine indexing multidisciplinary scholarly literature.

We searched in every database by author (and limited to the fixed time period) in March 2017. Umlauts in author names were considered for searching in WoS (e.g., AU = (Schlogl C* OR Schlögl C* OR Schloegl C*)). In Scopus, we searched by author ID. For Google Scholar, existing author profiles were considered (excluding publications with no date, except if they were found through title term search). In all concerned information services, an additional search by title terms took place to ensure that all publications were found. Due to continuous updating in the databases, the publication counts can vary from our stated results.

7.3 Results

During the specified time window, the investigated ISSI Scientific Committee authors respectively published between 65 and 304 documents (Table 7.1). These publication counts include all publications of an author's personal publication list completed with additionally found publications in diverse databases; all listed documents were checked against the two criteria and thus the lists are recognized as truebounded publication lists.

Table 7.1: Test for underbounded and overbounded publication lists. Investigated authors of the ISSI Scientific Committee and their relative visibility in WoS, Scopus, and Google Scholar as well as in their personal publication lists.

Author	True-bounded Publication List[1]	Personal Publication List	Over-Bounded Documents	Relative Visibility		
				WoS	Scopus	Google Scholar
Judit Bar-Ilan	124	88.7%	8	47.6%	62.1%	84.7%
Katy Börner	100	91.0%	4	40.0%	65.0%	80.0%
Lutz Bornmann	271	95.2%	10	79.3%	83.4%	96.3%
Leo Egghe	107	99.1%	0	87.9%	88.8%	97.2%
Stefanie Haustein	72	97.2%	1	26.4%	44.4%	80.6%
Peter Ingwersen	71	83.1%	4	29.6%	57.7%	93.0%
Loet Leydesdorff	304	90.5%	18	65.1%	72.7%	92.4%
Christian Schlögl	65	96.9%	2	15.4%	29.2%	73.8%
Cassidy Sugimoto	129	91.5%	1	48.8%	67.4%	91.5%
Averages of all authors	138	92.6%	5	48.9%	63.4%	87.7%

[1] Number of publications from 2007 to 2016; overbounded documents: the number of items appearing in the personal publication lists that do not meet the two criteria.

In order to analyze the levels of underbounded publication lists, relative visibility values were calculated for each database, but also for the personal publication lists. No single database includes all publications from an author. Egghe and Bornmann receive the highest relative visibility values in all three databases. (It could be that both are notably metric-wise (Rousseau & Rousseau, 2015), but of course we cannot know and should ask the authors). Generally, the relative visibility varies from database to database, always highest in Google Scholar followed by Scopus and WoS. Eight of our nine case-study authors reach 80 or more percent of visibility in Google Scholar, whereas such high values in Scopus are achieved by three authors and in WoS by only one. Closer inspection of WoS values shows that, for seven out of nine authors, more than half of their publications are missing. For one third, not even 30 percent are covered in this database. In contrast, seven authors reach relative visibility values of more than 50 percent in Scopus.

The generally lower values for Schlögl can be explained by the fact that 46 percent of his publications (30 out of 65) are not published in the English language. For scientific publications in information science, English is the lingua franca. Other languages, including German, do not have the same standing within the scientific community and are less likely to be listed in international information services than publications in English (Gordon, 2012). Likewise, Harzing (2016) discussed a higher coverage of non-English publications in Google Scholar compared toWoS and Scopus. Furthermore, 12 of Schlögl's publications are German lexicon entries.

Although none of the personal publication lists is totally complete, they cover the majority of each author's publications. It always depends on the point of view on what the authors themselves consider as a publication and therefore state in their lists. Some authors did not count revisions as independent publications. Furthermore, we assume some authors simply forgot to list publications (or did not realize that one of their papers was published). In addition to this, a few authors had conference and journal publications with the same title (or almost the same title) but stated only the journal articles in their publication lists. Few items on the respective personal publication lists do not meet the two criteria of scientific publications. Therefore, most of the lists are slightly overbounded.

For WoS and Scopus, we did not find aspects of an overbounded results list; however, Google Scholar produces such unwanted results. Depending on the search argument (searching for names versus searching inside of author profiles), the lists can be strongly overbounded. There are two kinds of mistakes (both were found on author profiles), namely wrong author names (for instance, on Leydesdorff's profile, you can find on position no. 7 an article of Richard Rogers with no relation to Leydesdorff) and informally published documents found by Google anywhere on the web (such as "Wissenschaftliche Zeitschriften im Web 2.0" on Haustein's profile, which is an unpublished slide set).

7.4 Discussion and Conclusions

What are the objectives of this study? First of all, we intended to focus attention on the characteristics of publication lists. Therefore, we introduced the concepts of truebounded, underbounded, and overbounded publication lists. As a measure for the state of boundedness, we applied the authors' relative visibility on bibliographic databases and on their personal publication lists.

To avoid confusion, we clearly have to differentiate between visibility and coverage. "Relative visibility" is a property of the unit of assessment (e.g., an author), while coverage is a property of an information service. Relative visibility shows how visible a unit of assessment in a specific database is. The focus of relative visibility is the perspective of the unit of assessment. While the two concepts are inextricably linked and the distinction may be a semantic one, treating them as different may reveal some new insights. For example, in the ACM Digital Library, a full-text database covering all ACM publications and comprehensively covering the field of ACM-based computing, an author from a bordering discipline might check his/her relative visibility in the information service in order to assess his/her positioning in computing.

In contrast to the application of results lists of bibliographic information services (especially WoS and Scopus) in scientometrics, the use of personal publication lists for research output measurement is a promising alternative approach.

To answer RQ1, all personal publication lists of our case studies are slightly underbounded. However, with visibility values between 83 and 99 percent, those lists are relatively close to the truebounded lists. Nearly all personal publication lists of our case study authors are (again, slightly) overbounded. There are missing items (leading to underboundedness) and items which do not meet the criteria of publications (leading to overboundedness). The analyzed publication lists are both underbounded and overbounded. This clearly demonstrates the importance of boundedness in addition to the simple calculation of relative visibility. Concerning RQ2, all publication lists of our case studies in the bibliographic information services WoS, Scopus, and Google Scholar are underbounded.

The authors' relative visibility is slim in WoS and a little bit better in Scopus. The best visibility values are found on Google Scholar, which is remarkable since this information service is based largely on automated algorithms and crowdsourced editing. There are no overbounded results lists either on WoS or Scopus, but on Google Scholar. This information service is problematic for bibliometrics due to missing standardized data formats, poor metadata descriptions, and overbounded publication lists.

The main limitation of this article is the use of a small list of case studies of authors. This has to be massively broadened in further research to other scientific subjects and other authors (maybe not as metric-wise writers as scientometricians). Considering other authors and other disciplines, results could differ. For example, new authors might also publish in journals that are not indexed by traditional indexes like WoS and Scopus. As we have only considered the general scientific information services WoS, Scopus, and Google Scholar, it would be very interesting to extend the research to discipline-specific databases as Medline for the biosciences, ACM Digital Library for computer science or LISTA for library and information science. With respect to Google Scholar, it would be interesting to include Microsoft Academic as well. We also intend to invite other researchers to discuss the criteria for determining scientific publications.

This research study investigates relative visibility of selected ISSI Committee Authors in WoS, Scopus, and Google Scholar compared to the relative visibility of their personal publication lists and reveals a publication visibility imbalance in the observed information services. Personal publication lists provide a high coverage of an author's publications; they are only slightly underbounded and overbounded. Especially for some cases in WoS, publications are sparsely covered. "The use of personal publication lists are reliable calibration parameters to compare coverage of information scientists in academic citation databases with scientific social media", Hughes (2017, p. 126), following Hilbert et al. (2015), states.

There is need for scholarly information management to have the authors' personal publication lists online available. Dorsch (2017) discussed the application of linked open data techniques and the establishment of institutional, national, and scholarly society-based repositories. A further option could be the inclusion of such lists—free of charge—in commercial information services (such asWoS or Scopus) in order to add the authors' (more or less) truebounded publication lists to the underbounded so-called "quality paper" lists.

Acknowledgements

We would like to thank the anonymous reviewers for helpful information.

Author Contributions

Wolfgang G. Stock and Isabelle Dorsch conceived and designed the case study; Johanna M. Askeridis collected the publication data; Isabelle Dorsch and Wolfgang G. Stock analyzed the data; Isabelle Dorsch,Wolfgang G. Stock, and Johanna M. Askeridis wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

Carnap, R. (1931). Die Überwindung der Metaphysik durch logische Analysen der Sprache. *Erkenntnis*, *2*(1), 219–241.

Chase, J. M. (1970). Normative criteria for scientific publication. *The American Sociologist*, *5*(3), 262–265.

Chen, X. (2010). Google scholar's dramatic coverage improvement five years after debut. *Serials Review*, *36*(4), 221–226.

Cole, S., & Cole, J. R. (1968). Visibility and the structural bases of awareness of scientific research. *American Sociological Review*, *33*(3), 397–413.

Dorsch, I. (2017). Relative visibility of authors' publications in different information services. *Scientometrics*, *112*(2), 917–925.

Dorsch, I., Schlögl, C., Stock, W. G., & Rauch, W. (2017). Forschungsthemen der Düsseldorfer und Grazer Informationswissenschaft (2010 bis 2016). *Information - Wissenschaft und Praxis*, *68*(5-6), 320–328.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of Pubmed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, *22*(2), 338–342.

Gaillard, J. (1992). Use of publication lists to study scientific production and strategies of scientists in developing countries. *Scientometrics*, *23*(1), 57–73.

Gordon, A. (2012). The invisibility of science publications in hebrew: A comparative database study. *Journal of the American Society for Information Science and Technology*, *63*(3), 607–615.

Haller, R. (1989). Wandlung der Wissenschaftsauffassung. In E. Busek, W. Mantl, & M. Peterlik (Eds.), *Wissenschaft und Freiheit. Ideen zur Universität und Universalität* (pp. 46–58). München, Germany: Oldenbourg.

Harzing, A.-W. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics*, *98*(1), 565–575.

Harzing, A. W. (2016). *Do Google Scholar, Scopus and the Web of Science speak your language?* Retrieved from https://harzing.com/publications/white-papers/do-google-scholar-scopus-and-theweb-of-science-speak-your-language

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., ... Stock, W. G. (2015). Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. *Online Information Review*, *39*(2), 255–264.

Hughes, E. (2017). Personal publications lists serve as a reliable calibration parameter to compare coverage in academic citation databases with scientific social media. *Evidence Based Library and Information Practice*, *12*(1), 125–127.

Ingwersen, P. (2000). The international visibility and citation impact of Scandinavian research articles in selected social science fields: The decay of a myth. *Scientometrics*, *49*(1), 39–61.

International Urban Research. (1959). *The World's Metropolitan Areas*. Berkeley, CA, USA; Los Angeles, CA, USA: University of California Press. (Available online: https://books.google.de/books?id=p0-6Ej_Up3AC))

Jacsó, P. (2008). The plausibility of computing the h-index of scholarly productivity and impact using reference-enhanced databases. *Online Information Review*, *32*(2), 266–283.

Kirkwood, P. E. (2012). Researcher publication checklist: A quantitative method to compare traditional databases to web search engines. In *119th American Society for Engineering Education Annual Conference and Exhibition, San Antonio, TX, June 10-13, 2012.* (Available online: https://peer.asee.org/facultypublication-checklists-a-quantitative-method-to-compare-traditional-databases-to-web-search-engines)

Kretschmer, H., & Aguillo, I. F. (2004). Visibility of collaboration on the web. *Scientometrics*, *61*(3), 405–426.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228.

Popper, K. R. (1976). *Logik der Forschung* (6th ed.). Tübingen, Germany: JCB Mohr (Paul Siebeck).

Rousseau, S., & Rousseau, R. (2015). Metric-wiseness. *Journal of the American Society for Information Science and Technology*, *66*(11), 2389.

Schlögl, C. (2013). International visibility of European and in particular German-language publications in library and information science. In H.-C. Hobohm

(Ed.), *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Tagungsband des 13. internationalen Symposiums für Informationswissenschaft (ISI 2013)* (pp. 51–62). Glückstadt: Hülsbusch. (Available online: `https://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/396`)

Scientific committee. (2013). In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), *Proceedings of the ISSI 2013 Vienna: 14th International Society of Scientometrics and Informetrics Conference, Vienne, Austria, 15–20 July 2013.* Vienna, Austria: Austrian Institute of Technology.

Stegmüller, W. (1973). Band 4: Personelle und statistische Wahrscheinlichkeit. In *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie,* Studienausgabe Teil A. Berlin, Heidelberg, Germany; New York, NY, USA: Springer.

Stock, W. G. (2000). Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In K. Fuchs-Kittowski, H. Laitko, H. Parthey, & W. Umstätter (Eds.), *Wissenschaft und Digitale Bibliothek. Wissenschaftsforschung Jahrbuch 1998* (pp. 239–282). Berlin, Germany: Gesellschaft für Wissenschaftsforschung. (Available online: `http://www.wissenschaftsforschung.de/JB98_239-282.pdf`)

The Organisation for Economic Co-operation and Development (OECD). (1963). *The measurement of scientific and technical activities, proposed standard practice for surveys of research and experimental development (frascati-manual).* Paris, France: OECD.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS ONE*, *8*(5), e64841.

Tillett, B. B. (2001). Bibliographic relationships. In C. A. Bean & R. Green (Eds.), *Relationships in the Organization of Knowledge* (pp. 19–35). Boston, MA, USA: Kluwer.

Tokar, A. (2012). Semantic change of the publication concept? In A. Tokar et al. (Eds.), *Science and the Internet* (pp. 193–205). Düsseldorf, Germany: Düsseldorf University Press. (Available online: `http://dup.oa.hhu.de/151/1/Beitrag-Tokar-2.pdf`)

Chapter 8

Forschungsthemen der Düsseldorfer und Grazer Informationswissenschaft (2010 bis 2016)

Über eine szientometrische Erfassung der Titelterme der Publikationen gibt der Artikel einen Überblick zu den aktuellen Forschungsthemen des Instituts für Informationswissenschaft und Wirtschaftsinformatik der Karl-Franzens-Universität Graz sowie der Abteilung für Informationswissenschaft der Heinrich-Heine-Universität Düsseldorf. Für die Erscheinungsjahrgänge 2010 bis 2016 konnten 129 Publikationen aus Graz und 249 aus Düsseldorf identifiziert werden. Top-Themen in Graz sind Informationswissenschaft, Österreich, mobile Systeme, Kommunikation, Universität, Zitation und (wissenschaftliche) Zeitschrift; in Düsseldorf dominieren Informationskompetenz, Informationswissenschaft, Social Media, informationelle (smarte) Städte und Wissen.

8.1 Einleitung

Das vorliegende Heft 5–6/2017 und das Heft 1/2018 von „Information – Wissenschaft und Praxis“ sind einer Darstellung ausgewählter aktueller Forschungsthemen der beiden informationswissenschaftlichen Forschungsinstitutionen in Graz und Düsseldorf gewidmet. Ergänzend zu dem üblichen Vorwort geben wir hier einen szientometrisch gewonnenen quantitativen Überblick zu den Forschungsthemen in Graz und Düsseldorf in den Jahren 2010 bis 2016. Über die Sichtbarkeit von Wissenschaftlern und ihren Institutionen entscheiden deren wissenschaftliche Publikationen (Friedländer, 2014; Schlögl, 2013). Entsprechend haben wir eine Publikationsanalyse, genauer eine Titelanalyse der Publikationen durchgeführt.

Für das Institut für Informationswissenschaft und Wirtschaftsinformatik der Karl-Franzens-Universität Graz konnten in diesem Zeitraum 129 Publikationen, für die Abteilung für Informationswissenschaft der Heinrich-Heine- Universität Düsseldorf 249 Veröffentlichungen identifiziert werden. Zur Bestimmung der Publikationen setzen wir nicht auf Trefferlisten bibliographischer Datenbanken wie Web of Science oder Scopus, sondern gründen unsere Berechnungen wegen der höheren Sichtbarkeit (Dorsch, 2017; Dorsch & Frommelius, 2015) auf Publikationslisten der in Graz und Düsseldorf arbeitenden Informationswissenschaftlerinnen und Informationswissenschaftler. Die szientometrische Auswertung setzt für ihre Themenanalyse eine Beschreibung und Analyse der Titel aller Publikationen ein (Honka, Orszullok, Dorsch, & Frommelius, 2015). Wie sind wir konkret vorgegangen?

8.2 Methode

Die generierten Publikationslisten beider Forschungseinrichtungen umfassen alle veröffentlichten wissenschaftlichen Publikationen im Zeitraum vom 1. Januar 2010 bis zum 31. Dezember 2016. Die Publikationen entnahmen wir den persönlichen Publikationslisten der Autorinnen und Autoren (Institutionswebseite/Institutions-Repository, persönliche Webseite). Zu den berücksichtigten Autorinnen und Autoren zählen alle wissenschaftlich arbeitenden Mitarbeiterinnen und Mitarbeiter (beginnend bei studentischen Hilfskräften), die für den gesamten Analysezeitraum oder innerhalb des Analysezeitraums fest mit der jeweiligen Institution verbunden waren. Durch einen Lehrauftrag Finanzierte wurden nicht berücksichtigt, weil das primäre Ziel eines Lehrauftrags die Lehre und nicht die Forschung an einer Institution ist. Für das Institut für Informationswissenschaft und Wirtschaftsinformatik der Karl-Franzens-Universität Graz wurde zusätzlich die Gastprofessur von Rainer Kuhlen (ein Semester) mit einbezogen. Für szientometrische Analysen ist es wichtig zu definieren, was eine (wissenschaftliche) Publikation ist und was als eine Publikation gezählt wird (Stock, 2000). Bei den in den Publikationslisten enthaltenen Dokumenten handelt es sich um wissenschaftliche und formal publizierte Publikationen, die im Rahmen von Forschungsprojekten des Grazer oder Düsseldorfer Forschungsinstituts entstanden sind. Duplikate, die durch Ko-Autorschaften Grazer bzw. Düsseldorfer Autoren untereinander entstanden sind, sowie nicht formal publizierte Dokumente wurden hierbei entfernt. Lexikoneinträge fanden (wegen ihrer Kürze – manchmal nur wenige Zeilen) ebenfalls keine Berücksichtigung. Zu den in die Analyse einbezogenen Dokumenttypen der Publikationen gehören: Beiträge in Sammelbänden/Konferenzbänden, Herausgebertätigkeiten, Monographien, Reviews und Zeitschriftenbeiträge.

Den generierten Forschungsthemen liegt eine Themenanalyse zugrunde, die sich auf die Publikationstitel der zuvor erstellten Publikationslisten der Forschungseinrichtungen bezieht. Für die Themenanalyse fand eine intellektuelle Aufbereitung der Nomen- und Adjektiv-Titelterme statt. Stoppwörter und Verben wurden ausgeschlossen. Die Aufbereitung umfasst die Übersetzung von nicht-englischsprachigen Titeltermen ins Englische, die Grundformbildung der Titelterme in ihre jeweilige lexikalische Kategorie, die Auflösung von Abkürzungen sowie die Zusammenführung von Synonymen. Daneben wurden mehrmals vorkommende Terme innerhalb eines

Titels nur einmal gezählt, weil sonst eine indirekte Gewichtung der Terme stattgefunden hätte, die sich einzig auf das mehrfache Vorkommen im Titel bezieht und nicht auf die inhaltlichen Aspekte der Publikation. Die so generierten Titelterme zeigen aktuelle Forschungsthemen der Forschungseinrichtungen auf. Zusätzlich fand eine Themenclusterung nach dem k-nearest neighbors-Verfahren statt (Stock & Stock, 2013, S. 778).

8.3 Informationswissenschaft in Graz

Das Grazer Institut für Informationswissenschaft und Wirtschaftsinformatik ist eines von fünfzehn Instituten an der Sozial- und Wirtschaftswissenschaftlichen Fakultät der Karl-Franzens-Universität Graz. Eine Besonderheit besteht darin, dass das Grazer Institut kein eigenes Studium anbietet. Vielmehr können Studierende des Bachelor- und Masterstudiums aus Betriebswirtschaftslehre „Informationswissenschaft und Wirtschaftsinformatik“ als Spezialisierung wählen und/oder einzelne Lehrveranstaltungen im Rahmen der Wahlfächer belegen. Derzeit (Stand: 2017) sind am Institut fünf wissenschaftliche Mitarbeiter beschäftigt, davon vier Habilitierte.

Wie aus der Institutsbezeichnung hervorgeht, deckt das Institut zwei Disziplinen ab: Informationswissenschaft und Wirtschaftsinformatik (engl. Information Systems). Ähnlich breit ist auch die Institutsforschung ausgelegt, wobei von den Mitarbeitern jeweils unterschiedliche Forschungsinhalte abgedeckt werden. Die breite Forschungsausrichtung geht auch aus Tabelle 8.1 hervor.

Tabelle 8.1: Die Themen der Grazer informationswissenschaftlichen Forschung 2010 bis 2016; alle Themen mit mehr als drei Nennungen im Titel; N = 129 Publikationen; insgesamt 373 unterschiedliche sinntragende Titelterme.

INFORMATION	38
SCIENCE	19
AUSTRIA	18
MOBILE	15
SYSTEM	15
COMMUNICATION	14
UNIVERSITY	12
CITATION	11
ANALYSIS	10
Journal	10
APPLICATION	9
COMPARISON	9
TECHNOLOGY	9
BUSINESS	8
DIGITAL	8
ECONOMIC	8
KNOWLEDGE	8

Tabelle 8.1: (continued)

LIBRARY	8
READERSHIP	8
DATA	7
ENVIRONMENT	7
INTERNATIONAL	7
LEARNING	7
PERSONALIZATION	7
RESEARCH	7
SOCIAL	7
USAGE	7
CASE	6
DEVELOPMENT	6
DOWNLOAD	6
ACCEPTANCE	5
AREA	5
AWARENESS	5
DOCUMENTATION	5
GRAZ	5
LOGISTICS	5
MANAGEMENT	5
PRACTICE	5
REFORM	5
USER	5
BEHAVIOR	4
DIFFERENCE	4
EUROPEAN	4
FIELD	4
GERMAN	4
LAST	4
LITERACY	4
MILE	4
ONLINE	4
PROTECTION	4
SOCIETY	4

Wie zu sehen ist, zählen INFORMATION, SCIENCE und SYSTEMS zu den am öftesten vorkommenden Termen. Da das Institut das einzige seiner Art in Österreich ist, kann es mitunter vorkommen, dass dies auch in der Forschung Berücksichtigung findet. Beispielhaft seien zwei Publikationen von Wolf Rauch angeführt, die aus einem Projekt mit einem Kollegen aus Ungarn resultierten: „Austria and Hungary: Different Stages of Readiness to Create Added Value by Using Business Information Systems“, und „Old Monarchy in the New Cyberspace: Empirical Examination of

Information Security Awareness among Austrian and Hungarian Enterprises“.

Im Bereich der Wirtschaftsinformatik beschäftigt sich die Institutsforschung primär mit der Mobilkommunikation (MOBILE COMMUNICATION) bzw. mit mobilen Anwendungen (MOBILE APPLICATION). Ein informationswissenschaftliches Hauptforschungsgebiet befasst sich mit szientometrischen Analysen von wissenschaftlichen Fachzeitschriften (JOURNAL). Konkret wird der Zusammenhang (COMPARISON) von Zitaten (CITATION ANALYSIS), DOWNLOADS und sog. Lesehäufigkeiten (READERSHIP DATA) untersucht. In einem zweiten informationswissenschaftlichen Forschungsbereich wurden einige empirische Studien in Universitätsbibliotheken (UNIVERSITY LIBRARY) durchgeführt.

Das mit 38 Publikationen größte Grazer Themencluster zu INFORMATION (Tab. 8.2) zeigt im Wesentlichen eine Verdichtung von Tabelle 8.1. Wieder zu erkennen sind die zwei Hauptbereiche der Institutsforschung – INFORMATION SCIENCE und INFORMATION SYSTEMS – und deren teilweiser Bezug zu Österreich bzw. zu den Wirtschaftswissenschaften (ECONOMICS). Ebenfalls wieder sichtbar sind die Forschungsarbeiten zur Zeitschriften-Szientometrie. Dies ist darauf zurückzuführen, dass ein Teil der Analysen für Wirtschaftsinformatik-Zeitschriften durchgeführt wurde. Noch nicht erwähnt wurde der Forschungsbereich Inormationskompetenz (INFORMATION LITERACY). Hier gab es von Anfang an eine Zusammenarbeit mit der Düsseldorfer Informationswissenschaft. So wurde ein in Düsseldorf entwickeltes Testinstrument adaptiert, um einige Informationskompetenztests an der Karl-Franzens-Universität Graz durchzuführen. Die Zusammenarbeit äußert sich auch am gemeinsamen Überblicksbeitrag in diesem Schwerpunktheft. Aktuell leitet Stefan Dreisiebner ein EUProjekt, dessen Ziel die Entwicklung einer mehrsprachigen Informationskompetenz-MOOC ist.

Tabelle 8.2: Themencluster zu INFORMATION in der Grazer informationswissenschaftlichen Forschung 2010 bis 2016. k-Nearest Neighbors zu INFORMATION; N = 38 Publikationen; alle Themen mit mehr als drei Nennungen im Titel.

INFORMATION	
SCIENCE	15
SYSTEM	9
AUSTRIA	7
KNOWLEDGE	6
ECONOMIC	5
CITATION	4
DATA	4
DOCUMENTATION	4
DOWNLOAD	4
JOURNAL	4
LITERACY	4
READERSHIP	4

Tabelle 8.3: Themencluster zu CITATION in der Grazer informationswissenschaftlichen Forschung 2010 bis 2016. k-Nearest Neighbors zu CITATION; N = 11 Publikationen; alle Themen mit mehr als drei Nennungen im Titel.

CITATION	
JOURNAL	9
DATA	6
DOWNLOAD	6
READERSHIP	5
USAGE	5
CASE	4
COMPARISON	4
INFORMATION	4
SYSTEM	4

Tabelle 8.4: Themencluster zu LIBRARY in der Grazer informationswissenschaftlichen Forschung 2010 bis 2016. k-Nearest Neighbors zu LIBRARY; N = 8 Publikationen; alle Themen mit mehr als zwei Nennungen im Titel.

LIBRARY	
UNIVERSITY	7
HOUR	3
OPENING	3

Tabelle 8.5: Themencluster zu MOBILE in der Grazer informationswissenschaftlichen Forschung 2010 bis 2016. k-Nearest Neighbors zu MOBILE; N = 15 Publikationen; alle Themen mit mehr als zwei Nennungen im Titel.

MOBILE	
APPLICATION	7
ACCEPTANCE	4
ANALYSIS	3
COMMUNICATION	3
DEVELOPMENT	3
LEARNING	3

Tabelle 8.3 zeigt die häufigsten Titelterme zur Zeitschriften- Szientometrie. Ausgangspunkt war das Elsevier Bibliometric Research Project, in dessen Rahmen Zitations- (CITATION) und Downloaddaten (DOWNLOAD) für Elsevier-Zeitschriften (JOURNAL) bezogen werden konnten. Darüber hinaus wurden sogenannte Readership-Daten (READERSHIP DATA) des sozialen Literaturverwaltungssystems Mendeley verwendet. Mit diesen Daten konnten umfassende Analysen zu Gemeinsamkeiten und Unterschieden dieser drei Datenquellen angestellt werden. Am Grazer Institut war Christian Schlögl federführend an diesen Analysen beteiligt. Im IWP-Heft 1/2018 wird eine Fallstudie zu zwei Volkswirtschaftslehre-Zeitschriften erscheinen. Ein Teilaspekt in diesem Forschungsschwerpunkt war die Visualisierung von Co-Readership-Daten (Zusammengehörigkeit von Publikationen, die jeweils von denselben Personen gelesen wurden). Im Rahmen seiner Dissertation entwickelte Peter Kraker eine Anwendung, die ein Mapping von Teilbereichen der Mendeley-Datenbasis ermöglicht.

Das Cluster zu LIBRARY (s. Tab. 8.4) umfasst einen Teil der Forschung von Gerhard Reichmann. In diesen Arbeiten beschäftigte er sich mit Benutzerforschung an Universitätsbibliotheken, im Speziellen mit deren Öffnungszeiten.

Tabelle 8.5 enthält jene Titelterme, die öfter gemeinsam mit MOBILE aufgetreten sind. Konkret handelt es sich um die Publikationen im Bereich mobiler Anwendungen und Mobilkommunikation, die meist im Rahmen von Drittmittelprojekten entstanden sind. Üblicherweise wurden sie gemeinsam von Otto Petrovic und seinem Projektteam verfasst. Teilaspekte dieses Forschungsclusters beschäftigen sich mit der Akzeptanz (ACCEPTANCE) von mobilen Anwendungen und wie man diese testen kann sowie mit mobilem Lernen (LEARNING).

8.4 Informationswissenschaft in Düsseldorf

Organisatorisch gehört die Düsseldorfer Informationswissenschaft in die Philosophische Fakultät der Heinrich- Heine-Universität. Den Ergänzungsfachstudiengang Informationswissenschaft trägt sie alleine, die Bachelorund Masterstudiengänge Informationswissenschaft und Sprachtechnologie gemeinsam mit Kollegen aus Sprachwissenschaft und Computerlinguistik sowie aus der Informatik. Im Jahr 2017 sind insgesamt 17 Personen wissenschaftlich tätig, darunter zwei Professoren.

Die Düsseldorfer Informationswissenschaft bemüht sich, ihre Aktivitäten in Forschungsprogrammen zu bündeln (Gust von Loh & Stock, 2008), d. h. große Projekte mit diversen Einzelpublikationen werden innerhalb eines umfassenden Rahmens organisiert. Auch zwischen den Forschungsprogrammen ist Zusammenarbeit angesagt. Jede/r Wissenschaftler/in arbeitet demnach mit jeder/m anderen zusammen; zudem werden Studierende – soweit es sich anbietet – in die Forschungen miteinbezogen.

Die Düsseldorfer Themen (s. Tab. 6) folgen näherungsweise einem inversen Power Law: ein dominierendes Thema an der Spitze (für Informationswissenschaft nicht überraschend: ebenso wie in Graz INFORMATION), gefolgt von Themen mit recht großer Ausprägung (SOCIAL, CITY, LITERACY, WEB, KNOWLEDGE, INFOR-

MATIONAL, SERVICE, SCIENCE). Die Liste führt dann über Themen wie z. B. TWITTER, RETRIEVAL, EMOTION, FACEBOOK, EVALUATION, GOVERNMENT, CITATION, FOLKSONOMY und GAME in die lange Reihe der Verteilung mit hunderten verschiedenen sinntragenden Titeltermen.

Große aktuelle Forschungsprogramme sind derzeit (2017) Social Media-Forschung, informationelle (smarte) Städte sowie Informationskompetenz. Weitgehend abgeschlossen sind Projekte zum emotionalen Information Retrieval (federführend war Tobias Siebenlist), zum Wissensmanagement in kleinen und mittleren Unternehmen sowie im Krankenhaus (von Sonja Gust von Loh unter Mitarbeit von Laura Schumann) und zum Einsatz von Gamification in der Hochschullehre, geleitet von Kathrin Knautz (heute: DFG in Bonn) (mitgearbeitet haben u. a. Lisa Orszullok, Christine Meschede, Julia Göretz und Oliver Hanraths). Die bereits vor einigen Jahren von Katrin Weller (heute: GESIS in Köln) begonnenen Untersuchungen an Twitter werden heute im Rahmen der Social Media- Forschung von Aylin Ilhan und Kaja J. Fietkiewicz weitergeführt. In ähnlicher Weise wird die Arbeit von Isabella Peters (heute: ZBW und Christian-Albrechts-Universität zu Kiel) zu Social Tagging und Folksonomies beispielsweise von Isabelle Dorsch (im Rahmen des Forschungsprogramms zu den Social Media) weitergeführt, wobei sie Hashtags bei Instagram analysiert. Informetrische, szientometrische sowie altmetrische Studien gibt es in Düsseldorf seit ca. 2005. Grundlagenarbeit insbesondere zum Stellenwert wissenschaftlicher Zeitschriften hat Stefanie Haustein (heute: University of Ottawa, ON, Kanada) geleistet. Derzeit arbeiten auf diesem Gebiet Isabelle Dorsch und Johanna M. Askeridis (zu persönlichen Publikationslisten und zur Sichtbarkeit von Wissenschaftlern) sowie Christine Meschede (zur Effektivität und Effizienz von Altmetrics). In vielen Kontexten (u. a. in der Social Media-Forschung, bei selbst implementierten Retrievalsystemen wie MEMOSE oder QUESTLAB/Zyren, aber auch bei der Analyse ubiquitärer Städte) ist eine Evaluation von Informationssystemen notwendig. Hierzu hat Laura Schumann mit dem Information Service Evaluation (ISE)-Modell eine heuristische Basis geschaffen. Der Abteilung angeschlossen ist eine eigenständige Arbeitsgruppe, die zu Web Science forscht. Sergej Sizov und seine Mitarbeiter Sarah Piller, Kevin Jasberg und Mikel Bahn widmen sich unter anderem Empfehlungen unter Unsicherheit und dem ökonomischen Wert von Keyword-basierter Online-Werbung. Im Heft 1/2018 erscheint im Kontext von Web Science ein Artikel über menschliche Unsicherheit in Informationssystemen. Wenn wir von einigen eher theoretisch ausgerichteten Artikeln (etwa zu Begriffen und semantischen Relationen) absehen, geht die Düsseldorfer Informationswissenschaft bei ihren Arbeiten konsequent empirisch vor.

Tabelle 8.6: Themen der Düsseldorfer informationswissenschaftlichen Forschung 2010 bis 2016. Alle Themen mit mehr als vier Nennungen im Titel; N = 249 Publikationen; insgesamt 537 unterschiedliche sinntragende Titelterme.

INFORMATION	74
SOCIAL	57
CITY	32

Tabelle 8.6: (continued)

LITERACY	26
WEB	26
KNOWLEDGE	25
INFORMATIONAL	23
SERVICE	23
SCIENCE	21
MEDIA	19
ANALYSIS	18
NETWORK	17
TWITTER	14
ACADEMIC	12
TAG	12
LIBRARY	12
USER	12
LEARNING	11
RETRIEVAL	11
WORLD	11
EMOTION	10
USE	10
DIGITAL	10
FACEBOOK	10
SYSTEM	10
ACCEPTANCE	9
COMMUNICATION	9
MANAGEMENT	9
SEMANTICS	8
EVALUATION	8
RESEARCH	8
GERMANY	7
GOVERNMENT	7
CITATION	7
FOLKSONOMY	7
SCIENTIFIC	7
SEARCH	7
TWEET	7
BOOKMARK	6
INFORMETRIC	6
CASE	6
DATA	6
EDUCATION	6
GAME	6
INSTRUCTION	6

Tabelle 8.6: (continued)

RECOMMENDATION	6
SCHOOL	6
TECHNOLOGY	6
BEHAVIOR	5
ELECTRONIC	5
FRIEND	5
FUTURE	5
IMPACT	5
INFRASTRUCTURE	5
LIFE	5
OPEN	5
ORGANIZATION	5
PERCEPTION	5
PUBLIC	5
REPRESENTATION	5
SOCIETY	5
SOFTWARE	5
UBIQUITOUS	5
VIDEO	5

Das mit 74 Publikationen größte Düsseldorfer Themencluster zu INFORMATION (s. Tab. 8.7) umfasst mehrere durchaus unterschiedliche Einzelthemen. Betont wird bei der Thematisierung von Informationswissenschaft (INFORMATION + SCIENCE ist 14mal im Cluster vertreten) stets der Zusammenhang zwischen Wissen (statisch) und Information (dynamisch). Eine typische Buchpublikation dazu ist das „Handbook of Information Science“; aber auch die bei De Gruyter erscheinende Buchreihe „Knowledge & Information. Studies in Information Science“ betont die Beziehung zwischen Information und Wissen. Ein zweites Teilcluster zeigt die Düsseldorfer Retrievalforschung auf (INFORMATION + RETRIEVAL hat acht Nennungen). Hier sind Forschungen zum Einsatz von Folksonomies im Retrieval (Isabella Peters, Laura Schumann und Jens Terliesner), zu Experten-Recommendersystemen (Tamara Heck), zu Patentrecherchen (Jasmin Schmitz) sowiezumeigenen Retrievalsystem für emotional geladene Dokumente (MEMOSE) und zum emotionalen Retrieval allgemein zu finden. Das dritte und quantitativ größte Teilcluster mit 25 Nennungen verweist auf die Forschungen zur Informationskompetenz (INFORMATION + LITERACY). Wir haben einen Test für die Ermittlung von Informationskompetenz bei unterschiedlichen Bevölkerungsgruppen entwickelt (Lisa Beutelspacher), die Vermittlung von Informationskompetenz sowohl in der Schule (Sonja Gust von Loh) als auch in Öffentlichen wie Wissenschaftlichen Bibliotheken (Maria Henkel) diskutiert sowie Forschungen zur Informationskompetenz im Kindergarten (Sonja Gust von Loh, Maria Henkel) durchgeführt. Als Beispiel für das Cluster zu Information befindet sich in diesem Heft – als Graz-Düsseldorfer Kooperation – ein Überblicksartikel zur Informationskompetenz von Lisa Beutelspacher, Maria Henkel und Stefan Dreisiebner.

Tabelle 8.7: Themencluster zu INFORMATION in der Düsseldorfer informationswissenschaftlichen Forschung 2010 bis 2016. k-Nearest Neighbors zu INFORMATION; N = 74 Publikationen; alle Themen mit mehr als zwei Nennungen im Titel.

INFORMATION	
LITERACY	25
SCIENCE	14
RETRIEVAL	8
SOCIAL	8
CITY	7
KNOWLEDGE	7
ANALYSIS	6
SCHOOL	6
SERVICE	6
INSTRUCTION	5
MEDIA	4
RESEARCH	4
TECHNOLOGY	4
UBIQUITOUS	4
ACADEMIC	3
CASE	3
COMMUNICATION	3
DIGITAL	3
EVALUATION	3
FOLKSONOMY	3
FUTURE	3
INFORMATIONAL	3
LEARNING	3
LIBRARY	3
METHOD	3
SOCIETY	3
STUDENT	3
TWITTER	3
USE	3

Tabelle 8.8 listet alle Themen auf, die mehrfach mit CITY zusammenhängen. In den Projekten zur informationellen Stadtforschung befassen wir uns mit prototypischen Städten der aufkommenden Wissensgesellschaft, vor allem den informationellen (oder auch „smarten“) Weltstädten (Agnes Mainka) sowie den ubiquitären Städten wie beispielsweise Songdo (Aylin Ilhan und Rena Möhlmann) oder Oulu (Laura Schumann). Fallstudien (stets von Mainka begleitet) betrafen Singapur (Isabella Peters, Swiatlana Khveshchanka), London (Duwaraka Murugadas, Stefa-

nie Vieten, Janina Nikolic) und die Städte am Arabischen Golf. Spin-offs dieses Forschungsprogramms sind Analysen von Bibliotheken (Agnes Mainka, Maria Henkel, Lisa Orszullok, Anika Stallmann), e-Government, m-Government sowie Open Data (neben Mainka vor allem Kaja J. Fietkiewicz, Sarah Hartmann und Christine Meschede) und Citizen Relationship Management-Systemen (311-Systemen) (Sarah Hartmann) in solch smarten Städten. Ein umfassender Reviewartikel in diesem (Teil 1) und im nächsten Heft (Teil 2) widmet sich den Beziehungen zwischen Stadtforschung und Informationswissenschaft.

Tabelle 8.8: Themencluster zu CITY in der Düsseldorfer informationswissenschaftlichen Forschung 2010 bis 2016. k-Nearest Neighbors zu CITY; N = 32 Publikationen; alle Themen mit mehr als zwei Nennungen im Titel.

CITY	
INFORMATIONAL	21
WORLD	10
INFORMATION	7
SERVICE	7
LIBRARY	6
GOVERNMENT	5
KNOWLEDGE	5
UBIQUITOUS	5
SMART	4
CASE	3
DIGITAL	3
INFRASTRUCTURE	3
JAPAN	3
MEDIA	3
PUBLIC	3
SOCIAL	3
SOCIETY	3
SONGDO	3

Das Cluster zu SOCIAL (Tabelle 8.9) umfasst Aspekte der Düsseldorfer Bemühungen zu den Social Media. Welche Altersgruppen verwenden – wenn überhaupt – Informationsdienste der Social Media? fragen Kaja J. Fietkiewicz und Katsiaryna S. Baran. Auch und gerade den „Silver Surfers“ schenkt Fietkiewicz dabei Beachtung. Erforscht wurde zudem der Gebrauch der Sprache bei der Einwerbung von Mitteln über Crowdfunding (Fietkiewicz). Viele der Social Media Dienste haben eine quasi-monopolistische Stellung auf ihren Märkten. Wie ist dies im Sinne von Wettbewerbs- und Kartellrecht zu beurteilen (Fietkiewicz)? Wie diffundieren Nachrichten bei Twitter? Am Beispiel des Anschlags auf Charlie Hebdo und weiteren terroristischen Akten in Paris sowie in Brüssel analysieren Ilhan und Fietkiewicz die Tweets sowie die „Retweetability“ von Tweets von Nachrichtenagenturen und anderen Twitter- Nutzern. Ein von Kathrin Knautz und Katsiaryna S. Baran heraus-

Tabelle 8.9: Themencluster zu SOCIAL in der Düsseldorfer informationswissenschaftlichen Forschung 2010 bis 2016. k-Nearest Neighbors zu SOCIAL; N = 57 Publikationen; alle Themen mit mehr als zwei Nennungen im Titel.

SOCIAL	
MEDIA	17
NETWORK	14
WEB	13
SERVICE	11
INFORMATION	8
KNOWLEDGE	8
ACCEPTANCE	6
BOOKMARK	6
TAG	6
PERCEPTION	5
SOFTWARE	5
USE	5
ACADEMIC	4
MANAGEMENT	4
QUALITY	4
SCIENCE	4
SYSTEM	4
USER	4
ANALYSIS	3
BEHAVIOR	3
CITY	3
EXPERT	3
GOVERNMENT	3
INFORMATIONAL	3
JOURNAL	3
LIFE	3
RECOMMENDATION	3
REPRESENTATION	3
SEMANTICS	3
STANDARD	3
WORLD	3

gegebener Sammelband untersucht den Social Networking Service (SNS) Facebook sowie das Nutzerverhalten auf Facebook. Das Informationsverhalten von Nutzern ist auch Thema bezogen auf eine neue Art von Social Media, nämlich den Social Live Streaming Services, die eine Melange aus SNS und (Live-) Fernsehen darstellen. Der Beitrag von Kaja J. Fietkiewicz, Katrin Scheibe und Franziska Zimmer in diesem Heft widmet sich unseren Forschungen zu den Social Live Streaming Services.

8.5 Diskussion

Wo liegen Gemeinsamkeiten zwischen beiden Forschungseinrichtungen? Und wo die Unterschiede? Während sich die Düsseldorfer bevorzugt mit Informationsdiensten (SERVICE in Tabelle 8.6) befassen, geht es den Grazern eher um Informationssysteme (SYSTEM in Tabelle 8.1). Typische Grazer Themen sind die Mobilkommunikation (MOBILE), der Bezug auf Österreich (AUSTRIA) und die Analyse von Nutzungsdaten wissenschaftlicher Zeitschriften (JOURNAL, READERSHIP, DOWNLOAD). Als Teil der Sozial- und Wirtschaftswissenschaftlichen Fakultät der Uni Graz ist die Befassung mit wirtschaftswissenschaftlichen Themen (ECONOMIC, BUSINESS) für die Grazer Informationswissenschaft naheliegend.

Tabelle 8.10: Gemeinsame Forschungsthemen der Grazer und Düsseldorfer Informationswissenschaft. Schnittmenge der Themen aus Tabelle 8.1 und Tabelle 8.6.

Thema	Graz	Düsseldorf
INFORMATION	38	74
SCIENCE	19	21
COMMUNICATION	14	9
CITATION	11	7
ANALYSIS	10	18
TECHNOLOGY	9	6
DIGITAL	8	10
KNOWLEDGE	8	25
LIBRARY	8	12
DATA	7	6
LEARNING	7	11
RESEARCH	7	8
SOCIAL	7	57
CASE	6	6
ACCEPTANCE	5	9
MANAGEMENT	5	9
USER	5	12
BEHAVIOR	4	5
GERMAN/Y	4	7
LITERACY	4	26
SOCIETY	4	5

Mit der Forschung zu informationellen Städten (INFORMATIONAL + CITY) haben die Düsseldorfer ein Alleinstellungsmerkmal in der gesamten – auch internationalen – Informationswissenschaft. Die Studien zu den Social Media sind in Düsseldorf ausgeprägter als in Graz; empirische Erhebungen zu den Social Networking Services (SOCIAL + NETWORK) sowie zum Nutzerverhalten bei konkreten Diensten (TWITTER, TWEET, FACEBOOK) sind häufig anzutreffen. Klassische informationswissenschaftliche Themen wie Wissensrepräsentation (KNOWLEDGE + REPRESENTATION) sowie Information Retrieval (dieses auch verbunden mit EMOTION und RECOMMENDATION) finden allerdings auch Beachtung.

Tabelle 8.10 listet die gemeinsamen Titelterme auf. Sie ist als Schnittmenge aus den in Tabelle 8.1 genannten Grazer Themen und den in Tabelle 8.6 aufgeführten Düsseldorfer Themen entstanden. Die Betonung der Informationswissenschaft (INFORMATION + SCIENCE) als Gemeinsamkeit ist nicht überraschend. Auch die empirische Erfassung der Wissenschaftskommunikation (COMMUNICATION, CITATION) weist auf ähnliche Forschungsfelder innerhalb der Szientometrie hin. Das Thema DIGITAL zeigt die Orientierung beider Institutionen auf ein Kernthema heutiger Zeit, die Digitalisierung. Bezüge auf KNOWLEDGE und auf TECHNOLOGY sind sowohl in Graz als auch in Düsseldorf zu finden. Ebenso betreiben beide Einrichtungen empirische Bibliothekswissenschaft (LIBRARY), Forschungen zur Informationskompetenz (LITERACY) sowie zur Nutzerforschung bzw. zum Informationsverhalten von Nutzern (USER, BEHAVIOR).

Literatur

Dorsch, I. (2017). Relative visibility of authors' publications in different information services. *Scientometrics*, *112*(2), 917–925.

Dorsch, I., & Frommelius, N. (2015). A scientometric approach to determine and analyze productivity, impact and topics based upon personal publication lists. In F. Pehar, C. Schlögl, & C. Wolff (Eds.), *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th-21st May 2015* (pp. 578–580). Glückstadt, Germany: Hülsbusch. (Available online: `https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/heck/Bella_Nils.pdf`)

Friedländer, M. B. (2014). Informationswissenschaft an deutschsprachigen Universitäten - eine komparative informetrische Analyse. *Information - Wissenschaft und Praxis*, *65*(2), 109–119.

Gust von Loh, S., & Stock, W. G. (2008). Wissensrepräsentation - Information Retrieval - Wissensmanagement - Das Forschungsprogramm der Düsseldorfer Informationswissenschaft. *Information - Wissenschaft und Praxis*, *59*(2), 73–74. (Available online: `https://www.philfak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/1204543980editorial_.pdf`)

Honka, A., Orszullok, L., Dorsch, I., & Frommelius, N. (2015). Topical impact analysis. A new informetric indicator for the assessment of a scientific institution. In *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th-21st May 2015* (pp. 410–421). (Available online: `https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/heck/Axel_et_al.pdf`)

Schlögl, C. (2013). Internationale Sichtbarkeit der europäischen und insbesondere der deutschsprachigen Informationswissenschaft. *Information - Wissenschaft und Praxis*, *64*(1), 1–8.

Stock, W. G. (2000). Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In K. Fuchs-Kittowski, H. Laitko, H. Parthey, & W. Umstätter (Eds.), *Wissenschaft und Digitale Bibliothek. Wissenschaftsforschung Jahrbuch 1998* (pp. 239–282). Berlin, Germany: Gesellschaft für Wissenschaftsforschung. (Available online: `http://www.wissenschaftsforschung.de/JB98_239-282.pdf`)

Stock, W. G., & Stock, M. (2013). *Handbook of information science*. Berlin, Boston: De Gruyter Saur.

Chapter 9

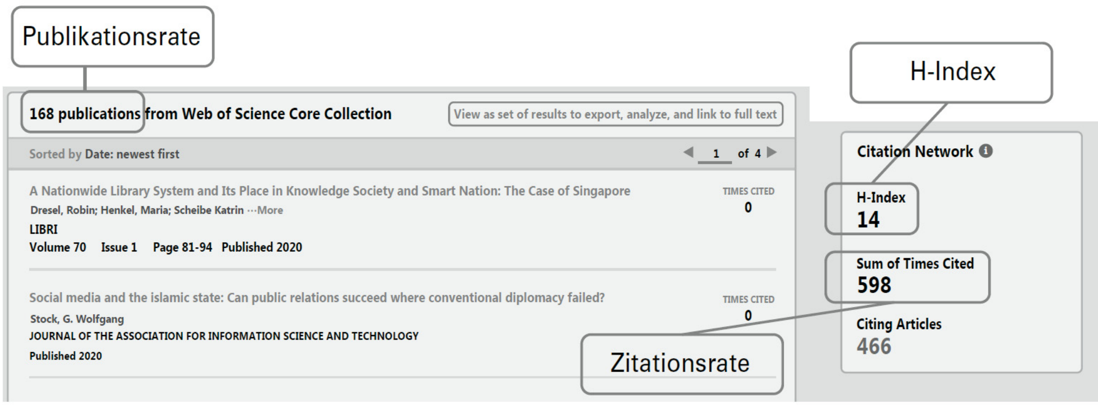
Publikationen, Zitationen und H-Index im Meinungsbild deutscher Universitätsprofessoren

Wie wichtig sind deutschen Universitätsprofessoren Publikations- und Zitationsraten? Haben sie Vorlieben für gewisse Datenbanken (wie Web of Science, Scopus oder Google Scholar)? Welche Bedeutung messen sie dem H-Index in den jeweiligen Informationsdiensten bei? Kennen sie Definition und Rechenweg beim H-Index? Es wurde mit einer Online-Umfrage (und einem Wissenstest zum H-Index) gearbeitet, die von mehr als 1000 Professoren ausgefüllt wurde. Dabei wurde zwischen den Ergebnissen für alle Teilnehmer und zusätzlich den Ergebnissen nach Geschlecht, Generation und Wissensgebiet unterschieden. Für die Mehrheit der befragten Forscher sind Publikationen wichtig, für Mediziner sind sie sogar sehr wichtig. Für Naturwissenschaftler und Mediziner sind Zitationen und H-Index bedeutsam, während Geistes- und Sozialwissenschaftler, Wirtschaftswissenschaftler und Juristen Zitationen und den H-Index (teilweise erheblich) weniger schätzen. Zwei Fünftel aller befragten Professoren kennen keine Details zum H-Index.

9.1 Einleitung

In der Forschung sind Publikationen, egal, ob als Zeitschriftenartikel, Bücher oder Beiträge in Sammelbänden und Proceedings, die Basis für die Kommunikation wissenschaftlichen Wissens. Mit den Zitationen bekommt man Aufschluss darüber, wie diese Publikationen in anderen Veröffentlichungen „angekommen“ sind. Sowohl Publikationsals auch Zitationsmaße gelten seit Jahrzehnten im Sinne meritokratischer Kriterien (Gross, Jungbauer-Gans, & Kriwy, 2008) als Basis für Evaluationen und Performancemessungen im Forschungsbereich (Jappe, 2020; Rassenhövel, 2010). Doch das sind zwei unterschiedliche Maße. Der von Jorge Hirsch (2005) entwickelte H-Index führt beide Maße zu einem einzigen Indikator zusammen, was jedoch gleichzeitig die Frage aufwirft, welches zugrundeliegende Konzept der H-Index exakt darstellt (Sugimoto & Larivière, 2013). Der H-Index ist die Anzahl h der Publikationen eines Forschers, die mindestens h-mal zitiert worden sind (W. G. Stock & Stock, 2013, S. 382). In den letzten 15 Jahren verbuchte der H-Index einen großen Popularitätszuwachs und gleichzeitig wurde er stark diskutiert und kritisiert. Inzwischen gibt es ganze Listen über seine Vor- und Nachteile (z. B. R. Rousseau, Egghe, & Guns, 2018). Ungeachtet seiner Nachteile ist der H-Index in unserem wissenschaftlichen System mittlerweile jedoch stark verdrahtet.

Den Markt für allgemeinwissenschaftliche bibliographische Datenbanken dominieren Web of Science (WoS), Scopus und Google Scholar, wobei die beiden erstgenannten kostenpflichtig sind und die dritte kostenlos zugänglich ist (Linde & Stock, 2011, S. 237). Sowohl die hier benutzten Informationsdienste Web of Science (Birkle, Pendlebury, Schnell, & Adams, 2020; M. Stock & Stock, 2003), Scopus (Baas, Schotten, Plume, Côté, & Karimi, 2020) und Google Scholar (Aguillo, 2011) als auch die weiteren, hier allerdings aus Gründen der Übersichtlichkeit nicht genutzten kostenfreien Datenbanken Microsoft Academic (Wang et al., 2020), Dimensions (mit der kommerziellen Variante Dimensions Plus) (Herzog, Hook, & Konkiel, 2020), Mendeley (Meschede & Siebenlist, 2018; Thelwall, 2018) und ResearchGate (Kraker & Lex, 2015) stellen gewisse Basisinformationen für bibliometrische Untersuchungen bereit. Alle drei in dieser Studie verwendeten Informationsdienste bieten Publikationszahlen, Zitationszahlen und den H-Index für Forscher an, deren Werte problemlos direkt nach einer Autorensuche ablesbar sind (Abbildung 9.1).



Quelle: Web of Science/Author Search

Abbildung 9.1: Angaben zur Anzahl von Publikationen und Zitationen sowie zum H-Index für einen der Autoren dieses Beitrags bei Web of Science.

Es geht in diesem Beitrag nicht um die nicht übersehbaren methodischen Probleme der Publikations- und Zitationsmaße (W. G. Stock, 2001) sowie des H-Index (Jan & Ahmad, 2020), sondern wir wollen die Forscher selbst befragen, wie sie dazu stehen. Wie wichtig sind ihnen Publikationen und Zitationen? Welche Bedeutung messen sie der Sichtbarkeit ihrer Publikationen und ihrem H-Index in den jeweiligen Informationsdiensten bei? Kennen Forscher den H-Index und seine konkrete Berechnungsformel überhaupt? Gibt es bei den Einschätzungen und dem Wissensstand Unterschiede beim Geschlecht, bei den Fächern und den Generationen? Der zugegebenermaßen stark zahlenlastige Artikel bringt erste Ergebnisse zu persönlichen

Einschätzungen deutscher Hochschullehrer zu standardmäßig eingesetzten szientometrischen Indikatoren.

9.2 Methoden

Um zu quantitativen Daten zu gelangen, haben wir uns für eine Online-Umfrage entschieden. Wir beschränkten uns auf Professoren, die an einer deutschen Universität arbeiten. Professor ist hierbei die Amts- bzw. Berufsbezeichnung, die sowohl männliche Professoren als auch Professorinnen umfasst. Als Testpersonen wurden ausschließlich Universitätsprofessoren ausgewählt (und andere akademische Mitarbeiterinnen und Mitarbeiter an Universitäten sowie Professoren an Fachhochschulen übersprungen), weil wir uns auf Personen konzentrieren wollten, die erstens bereits einen gefestigten Karriereweg haben bzw. hatten (im Gegensatz zu anderen akademischen Mitarbeitern) und zweitens in aller Regel darauf bedacht sind, ihre Forschungsergebnisse zu veröffentlichen (im Gegensatz zu Professoren an Fachhochschulen, die sich hauptsächlich an der Praxis orientieren).

Der Online-Fragebogen enthielt drei verschiedene Abschnitte. Im ersten Abschnitt haben wir nach persönlichen Daten gefragt (Geschlecht, Alter, akademische Disziplin und Universität). Abschnitt 2 befasste sich mit den persönlichen Einschätzungen der Professoren zur Bedeutung von Veröffentlichungen, Zitationen, ihrer Sichtbarkeit auf Web of Science, Scopus und Google Scholar, dem H-Index auf den drei Plattformen und der Bedeutung des H-Index in ihrer akademischen Disziplin. Für die Einschätzungen verwendeten wir eine 5-Punkt-Likert-Skala (von 1: „sehr wichtig“ über 3: „neutral“ bis 5: „sehr unwichtig“) (Likert, 1932). Es war für alle Fragen möglich, auch auf „keine Angabe“ zu klicken. Abschnitt 3 bestand aus zwei Fragen, nämlich einer subjektiven Einschätzung des eigenen Wissens über den H-Index und einem objektiven Wissenstest mit einem Multiple-Choice-Test (eine richtige Antwort: Nr. 3., vier falsche und die Option „Ich bin nicht sicher“), dessen Antwortmöglichkeiten wie folgt lauten:

1. H ist der Quotient aus der Anzahl der Zitationen von Beiträgen in Zeitschriften im Bezugszeitraum und der Anzahl veröffentlichter Beiträge in Zeitschriften im Bezugszeitraum.

2. H ist der Quotient aus der Anzahl der Zitationen auf Artikel (Zeitraum drei Jahre) und der Anzahl der Zitationen auf diese Artikel (in den vorherigen drei Jahren) für einen Wissenschaftler.

3. H ist die Anzahl der Artikel eines Wissenschaftlers, die mindestens H-mal zitiert worden sind.

4. H ist die Anzahl aller Zitationen zum H-Index, davon subtrahiert (H-Index)∧ 2.

5. H ist der Quotient der Anzahl der Zitationen einer wiss. Arbeit und des Alters dieser wiss. Arbeit.

Ein Antwortformat mit vorgefertigten Antworten wurde für den objektiven Wissenstest gewählt, da es als beste Wahl für die Wissensmessung empfohlen wird im Gegensatz zu z. B. Freitextfeldern (Haladyna & Rodriguez, 2013). Bei der Entwicklung der Wissenstestaufgaben folgten wir überwiegend den 22 Empfehlungen von Haladyna and Rodriguez (2013, in Abschnitt II).

Die Adressen der Universitätsprofessoren wurden nach dem Zufallsprinzip aus dem Hochschullehrerverzeichnis (Deutscher Hochschulverband, 2020) entnommen. Wir verteilten den Link zum Fragebogen (gehostet bei UmfrageOnline) per E-Mail an jeden einzelnen Professor. Wir starteten das Mailing im Juni 2019 und stoppten es im März 2020, als wir mehr als 1000 gültige (also bis zum Ende ausgefüllte) Fragebögen bekommen hatten. Insgesamt haben wir 5722 Professoren persönlich kontaktiert und sind zu 1081 ausgefüllten Fragebögen gekommen, was einer Rücklaufquote von 18,9 % entspricht. Die für Online-Umfragen vergleichsweise recht hohe Rücklaufquote lässt vermuten, dass das Thema bei einem großen Anteil der Professoren auf ein starkes Interesse gestoßen ist. Alle Fragebögen wurden vollständig anonymisiert.

Ein Vergleich zwischen unserer Stichprobe deutscher Professoren an Universitäten (Tabelle 9.1) und der Grundgesamtheit, wie sie in der amtlichen Statistik (Destatis, 2019) zu finden ist, ergibt nur geringfügige Unterschiede in Bezug auf die Geschlechterverteilung und auch nur wenige Unterschiede in Bezug auf die meisten Disziplinen. Zu vermerken sind jedoch zwei größere Differenzen. In unserer Stichprobe finden wir mehr Naturwissenschaftler als in der offiziellen Statistik und weniger Forscher aus den Geistes- und Sozialwissenschaften.

Tabelle 9.1: Zusammensetzung der Befragten in der Stichprobe (N = 1081).

		Relative Häufigkeit	Absolute Häufigkeit
Geschlecht	Männer	81,9%	864
	Frauen	18,1%	191
Fachbereich	Geowiss., Landwirtschaft usw.	7,9%	83
	Geistes- u. Sozialwissenschaften	23,7%	249
	Naturwissenschaften	46,7%	490
	Medizin	11,7%	123
	Recht	1,3%	14
	Wirtschaftswissenschaften	8,7%	91
Generation	Generation Y	9,2%	95
	Generation X	64,1%	663
	Baby Boomer	25,4%	263
	Silent Generation	1,4%	14

In unserer Analyse haben wir immer zwischen den Ergebnissen für alle Teilnehmer und zusätzlich den Ergebnissen nach Geschlecht, Generation und Wissensgebiet differenziert. Wir haben unterschieden nach zwei Geschlechtern (Männer, Frauen; der Fragebogen enthielt des Weiteren auch die Optionen „divers“ und „keine Angabe“), vier Generationen: Generation Y (geboren nach 1980), Generation X (geboren zwischen 1960 und 1980), Baby Boomer (die geburtenstarken Jahrgänge nach dem 2. Weltkrieg, geboren nach 1946 und vor 1960) und Silent Generation (die älteste Generation, geboren vor 1946) (Einteilung gemäß Fietkiewicz, Lins, Baran, and Stock, 2016) sowie sechs akademische Disziplingruppen: (1) Geowissenschaften, Umweltwissenschaften, Land- und Forstwirtschaft, (2) Geistes- und Sozialwissenschaften, (3) Naturwissenschaften (einschließlich Mathematik), (4) Medizin, (5) Rechtswissenschaft und (6) Wirtschaftswissenschaften. Diese Grobaufteilung der Wissensfelder entspricht der Fakultätsstruktur einiger deutscher Universitäten.

Da unsere Likert-Skala eine Ordinalskala ist, haben wir jeweils den Modus, den Median sowie den Interquartilsabstand (IQA) berechnet. Zur Analyse signifikanter Unterschiede verwendeten wir den Mann-Whitney-U-Test (Mann & Whitney, 1947) (für die beiden Werte des Geschlechts) und den Kruskall-Wallis-H-Test (Kruskal & Wallis, 1952) (für mehr als zwei Werte bei den Generationen und akademischen Disziplinen). Die Daten zum Wissensstand der Forscher über den H-Index liegen auf einer Nominalskala, daher haben wir die relativen Häufigkeiten für drei Werte berechnet (1: Forscher kennen den H-Index in ihrer Selbsteinschätzung und haben den Test bestanden; 2: Forscher kennen den H-Index in ihrer Selbsteinschätzung nicht; 3: Forscher meinen den H-Index in ihrer Selbsteinschätzung zu kennen, haben aber den Test nicht bestanden). Zur Analyse der Unterschiede zwischen Geschlecht, Wissensbereich und Generation haben wir hier den Chi-Quadrat-Test (Pearson, 1900) eingesetzt. Wir unterscheiden drei statistische Signifikanzniveaus, nämlich *: $p \leq 0,05$ (signifikant), **: $p \leq 0,01$ (sehr signifikant) und ***: $p \leq 0,001$ (extrem signifikant). Alle Berechnungen wurden mit Hilfe von SPSS durchgeführt.

9.3 Relevanz von Publikationen und Zitationen für die akademische Laufbahn

Publikationszahlen sind ein Indikator für die wissenschaftliche Leistung eines Wissenschaftlers, Zitationszahlen ein Indikator für seinen Einfluss auf andere Forscher (W. G. Stock, 2001). Tabelle 9.2 gibt an, für wie wichtig die befragten Professoren Publikationen und Zitationen für ihre wissenschaftliche Laufbahn einschätzen.

Für ungefähr 90 % aller Teilnehmer ist die Relevanz von Publikationen für die akademische Karriere auf der verwendeten 5-stufigen Skala wichtig (2) oder sogar sehr wichtig (1), Modus und Median liegen bei wichtig, der IQA von 1 deutet auf eine nur geringe Streuung hin. Es gibt mehr Professorinnen als männliche Kollegen, die Publikationen als sehr wichtig einstufen (48,7 % im Unterschied zu 41,3 %). Bei den Fächern gibt es einen Ausreißer nach oben: Bei Medizinern liegt der Median bei sehr wichtig, auch die Streuung ist gering; für fast 60 % sind Publikationen sehr wichtig. Für alle anderen Fächer sind Publikationen wichtig, bei der Einschätzung nach sehr wichtig ergibt sich eine Rangfolge von 48,2 % bei Geowissenschaften usw., Wirtschaftswissenschaften (45,1 %), Naturwissenschaften (41,8 %), Geistes- und Sozialwissenschaften (33,1 %) bis hin zu den Rechtswissenschaften (21,4 %). Bei den Generationen zeigt sich ein eindeutiges Bild: Je jünger Wissenschaftler sind, desto

Tabelle 9.2: Relevanz von Publikationen und Zitationen für die wissenschaftliche Laufbahn in der Einschätzung von Professoren in Deutschland.

a) Relevanz von Publikationen	(1)	(2)	(3)	(4)	(5)	Mo	Me	IQA	Sign.	N
Alle	42,4%	47,5%	7,9%	1,7%	0,5%	2	2	1	-	1070
Männer	41,3%	48,0%	8,4%	2,0%	0,3%	2	2	1	0,027 *	859
Frauen	48,7%	45,0%	5,2%	0,5%	0,5%	1	2	1		191
Geowiss., Landwirtschaft usw.	48,2%	44,6%	6,0%	1,2%	0,0%	1	2	1		83
Geistes- u. Sozialwissenschaften	33,1%	54,4%	9,3%	2,4%	0,8%	2	2	1		248
Naturwissenschaften	41,8%	48,1%	7,6%	1,9%	0,6%	2	2	1	0,000 ***	486
Medizin	58,5%	35,8%	5,7%	0,0%	0,0%	1	1	1		123
Recht	21,4%	71,4%	7,1%	0,0%	0,0%	2	2	0		14
Wirtschaftswissenschaften	45,1%	41,8%	12,1%	0,1%	1,1%	1	2	1		91
Generation Y	52,6%	42,1%	5,3%	0,0%	0,0%	1	1	1		95
Generation X	45,0%	47,0%	6,5%	1,2%	0,3%	2	2	1	0,000 ***	661
Baby Boomer	35,0%	49,0%	11,8%	3,8%	0,4%	2	2	1		263
Silent Generation	16,7%	66,7%	16,7%	0,0%	0,0%	2	2	0		12
b) Relevanz von Zitationen	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**Mo**	**Me**	**IQA**	**Sign.**	**N**
Alle	21,8%	42,2%	21,9%	10,0%	4,1%	2	2	1	-	1065
Männer	21,6%	42,5%	22,6%	9,4%	3,9%	2	2	1	0,027 *	855
Frauen	23,7%	38,4%	19,5%	13,7%	4,7%	2	2	1		190
Geowiss., Landwirtschaft usw.	29,3%	50,0%	15,9%	3,7%	1,2%	2	2	1		82
Geistes- u. Sozialwissenschaften	6,5%	27,8%	31,0%	24,1%	10,6%	3	3	2		245
Naturwissenschaften	28,5%	47,6%	17,2%	4,9%	1,6%	2	2	1	0,000 ***	487
Medizin	30,9%	38,2%	20,3%	7,3%	3,3%	2	2	2		123
Recht	0,0%	42,9%	28,6%	21,4%	7,1%	2	3	2		14
Wirtschaftswissenschaften	14,4%	50,0%	25,6%	6,7%	4,1%	2	2	1		90
Generation Y	22,1%	52,6%	18,9%	6,3%	0,0%	2	2	1		95
Generation X	22,7%	42,5%	21,2%	9,1%	4,4%	2	2	1	0,000 ***	656
Baby Boomer	21,4%	38,2%	22,5%	13,4%	4,6%	2	2	1		262
Silent Generation	7,7%	15,4%	30,8%	38,5%	7,7%	4	3	1		13

Frage: „Welche Bedeutung hat die Anzahl der Publikationen/Zitationen für Ihre wissenschaftliche Laufbahn?“
Skala: 1: sehr wichtig, 2: wichtig, 3: neutral, 4: unwichtig, 5: überhaupt nicht wichtig;
Mo: Modus; Me: Median; IQA: Interquartilsabstand; Sign.: Signifikanz (u-Test, h-Test); N: Anzahl der jeweiligen Teilnehmer.

häufiger schätzen sie Publikationen als sehr wichtig ein.

Die Relevanz der Zitationen ist für die Professoren weniger ausgeprägt als bei den Publikationen. Auch hier liegen für alle Befragten Modus und Median zwar bei 2 (IQA: 1), aber nur für rund ein Fünftel sind Zitationen sehr wichtig. Mehr Männer wählen im Vergleich zu den Frauen die Ausprägungen wichtig und neutral, während mehr Frauen sehr wichtig und unwichtig ankreuzten. Bei den Fächern springen zwei Ausreißer (diesmal jedoch nach unten) ins Auge: Geistes- und Sozialwissenschaftler sowie Juristen zeigen einen Median bei neutral (mit einem IQA von 2), sehr wichtig sind Zitationen nur für 6,5 % der Geistes- und Sozialwissenschaftler und für überhaupt keinen Juristen. Für die restlichen Disziplinen scheinen die Zitationen durchwegs wichtig zu sein, aber es zeigt sich eine Diskrepanz zwischen Naturwissenschaftlern und Medizinern auf der einen Seite (ca. 30 % stimmen für sehr wichtig) und Ökonomen auf der anderen (nur ca. 14 % halten Zitationen für sehr wichtig). Für die ältesten der Befragten haben Zitationen kaum noch Bedeutung (Median von 3 im Gegensatz zu 2 bei allen anderen); bei der Summe der Stimmen für sehr wichtig und wichtig ergibt sich wie bei der Relevanz der Publikationen eine Rangfolge nach Alter. Bei der Einschätzung der Relevanz von Publikationen wie Zitationen sind die Unterschiede bei den Geschlechtern signifikant, bei den Fächern und den Generationen sind sie sogar extrem signifikant.

9.4 Relevanz der Sichtbarkeit bei allgemeinwissenschaftlichen Informationsdiensten

Wie schätzen die Professoren die Abdeckung ihrer Publikationen bei den führenden allgemeinwissenschaftlichen Informationsanbietern Web of Science, Scopus und Google Scholar, mithin also ihre Sichtbarkeit in diesen Diensten (Schlögl, 2013; Dorsch, 2017) ein? Die Sichtbarkeit (V) – analog der H-Index – ist je nach Datenbank unterschiedlich, wobei meist die Ungleichheit

$$V(R)_{Web\ of\ Science} < V(R)_{Scopus} < V(R)_{Google\ Scholar}$$

für einen beliebigen Forscher R gilt (Dorsch, Askeridis, & Stock, 2018). Bei Web of Science müssen wir im Hinterkopf haben, dass dieser Informationsdienst aus vielen einzelnen Segmenten besteht (u. a. Science Citation Index Expanded, Social Science Citation Index, Arts & Humanities Citation Index, Emerging Sources Citation Index, Book Citation Index, Conference Proceedings Citation Index). Da die Bibliotheken nicht immer alle Segmente (und darin nicht alle Jahrgänge) abonniert haben, ist es keine Überraschung, dass die Sichtbarkeit (und ebenso der H-Index) in Abhängigkeit von der konkreten Subskription der jeweiligen Bibliothek steht (Hu, Wang, Ni, & Liu, 2020).

Für 36,7 % aller Forscher ist es sehr wichtig, dass ihre Publikationen bei Web of Science gelistet sind; bei Google Scholar sind dies 29,3 % und für Scopus 20,3 %, der Median liegt für Web of Science und Google Scholar bei 2, bei Scopus nur bei 3 (Tabelle 9.3). Bei den Geschlechtern sind bei dieser Einschätzung keine statistisch sichtbaren Unterschiede zu erkennen. Die Unterschiede bei den Disziplinen sind dagegen sehr ausgeprägt.

Tabelle 9.3: Relevanz der Sichtbarkeit bei Web of Science, Scopus und Google Scholar in der Einschätzung von Professoren in Deutschland.

a) Sichtbarkeit / Web of Science	(1)	(2)	(3)	(4)	(5)	Mo	Me	IQA	Sign.	N
Alle	36,7%	23,7%	15,9%	9,9%	13,8%	1	2	2	-	1024
Männer	37,2%	23,8%	15,9%	10,2%	13,0%	1	2	2	0,849	826
Frauen	37,1%	24,2%	14,6%	8,4%	15,7%	1	2	2		178
Geowiss., Landwirtschaft usw.	47,0%	33,7%	14,5%	4,8%	0,0%	1	2	1		83
Geistes- u. Sozialwissenschaften	8,6%	17,6%	23,1%	19,9%	6,8%	3	4	3		221
Naturwissenschaften	44,5%	23,9%	14,8%	6,8%	5,2%	1	2	2	0,000 ***	474
Medizin	58,2%	29,5%	6,6%	4,1%	0,2%	1	1	1		122
Recht	0,0%	0,0%	14,3%	14,3%	71,4%	5	5	1		14
Wirtschaftswissenschaften	26,7%	32,6%	19,8%	11,6%	9,3%	2	2	2		86
Generation Y	35,5%	28,0%	18,3%	6,5%	11,8%	1	2	2		93
Generation X	40,1%	22,2%	14,5%	9,0%	14,2%	1	2	2		634
Baby Boomer	30,9%	28,5%	16,3%	12,6%	11,8%	1	2	2	0,007 **	246
Silent Generation	0,0%	16,7%	33,3%	25,0%	25,0%	3	3	1		12
b) Sichtbarkeit / Scopus	(1)	(2)	(3)	(4)	(5)	Mo	Me	IQA	Sign.	N
Alle	20,3%	27,8%	24,5%	12,9%	14,6%	2	3	2	-	1009
Männer	19,7%	27,9%	25,2 %	13,1%	14,1%	2	3	2	0,297	814
Frauen	24,6%	28,6%	19,4%	12,0%	15,4%	2	2	2		175
Geowiss., Landwirtschaft usw.	30,1%	43,4%	18,1%	6,0%	2,4%	2	2	2		83
Geistes- u. Sozialwissenschaften	6,5%	17,5%	24,9%	21,7%	29,5%	5	4	2		217
Naturwissenschaften	26,0%	26,8%	23,8%	10,9%	12,6%	2	2	2	0,000 ***	470
Medizin	25,4%	39,0%	22,9%	9,3%	3,4%	2	2	2		118
Recht	0,0%	0,0%	21,4%	14,3%	64,3%	5	5	1		14
Wirtschaftswissenschaften	8,3%	34,5%	38,1%	10,7%	8,3%	3	3	1		84
Generation Y	29,8%	27,7%	24,5%	7,4%	10,6%	1	2	2		94
Generation X	20,7%	28,1%	23,6%	11,9%	15,6%	2	3	2		622
Baby Boomer	16,9%	29,3%	25,6%	15,7%	12,4%	2	3	2	0,003 ***	242
Silent Generation	0,0%	8,3%	33,3%	33,3%	25,0%	3;4	4	2		12
c) Sichtbarkeit / Google Scholar	(1)	(2)	(3)	(4)	(5)	Mo	Me	IQA	Sign.	N
Alle	29,3%	20,0%	21,4%	9,6%	10,8%	1	2	2	-	1035
Männer	28,7%	28,7%	21,8 %	10,2%	10,6%	1;2	2	2	0,181	833
Frauen	32,4%	30,8%	19,2%	7,1%	10,4%	1	2	2		182

Tabelle 9.3: (fortlaufend)

c) Sichtbarkeit / Google Scholar	(1)	(2)	(3)	(4)	(5)	Mo	Me	IQA	Sign.	N
Geowiss., Landwirtschaft usw.	31,3%	36,1%	25,3%	3,6%	3,6%	2	2	2		83
Geistes- u. Sozialwissenschaften	14,8%	27,0%	21,7%	14,3%	22,2%	2	3	2		230
Naturwissenschaften	36,5%	28,1%	19,7%	7,3%	8,4%	1	2	2	0,000 ***	477
Medizin	21,2%	35,6%	29,7%	11,0%	2,5%	2	2	1		118
Recht	0,0%	7,1%	28,6%	14,3%	50,0%	5	4	2		14
Wirtschaftswissenschaften	39,3%	30,3%	13,5%	10,1%	6,7%	1	2	2		89
Generation Y	46,7%	32,6%	13,0%	4,3%	3,3%	1	2	1		92
Generation X	30,0%	28,9%	20,8%	8,4%	11,8%	1	2	2	0,000 ***	643
Baby Boomer	22,2%	29,4%	24,2%	13,7%	10,5%	2	2	1		248
Silent Generation	0,0%	25,0%	25,0%	33,3%	16,7%	4	3	2		12

Frage: „Wie wichtig ist es Ihnen, dass Ihre Publikationen auf folgenden Plattformen erfasst/abgebildet werden:
Web of Science, Scopus, Google Scholar;"
Skala: 1: sehr wichtig, 2: wichtig, 3: neutral, 4: unwichtig, 5: überhaupt nicht wichtig; Mo: Modus; Me: Median; IQA: Interquartilsabstand; Sign.: Signifikanz (u-Test, h-Test); N: Anzahl der jeweiligen Teilnehmer.

Für Naturwissenschaftler und die Gruppe aus Geowissenschaftlern usw. ist die Sichtbarkeit bei Web of Science am wichtigsten (Median von 2; sehr wichtig für 47,0 % der Geowissenschaftler usw. und für 44,5 % der Naturwissenschaftler), gefolgt von der Sichtbarkeit bei Google Scholar (Median auch von 2; sehr wichtig jedoch nur für 31,3 % bzw. 36,5 %) und bei Scopus (Median von 2; sehr wichtig für 30,1 % bzw. 26,0 %). Die Mediziner haben mit einem Median von 1 und 58,2 % der Einschätzungen von sehr wichtig eine klare Präferenz für Web of Science. Scopus (Median von 2; sehr wichtig für 25,4 %) und Google Scholar (Median von 2, sehr wichtig für 21,2 %) spielen für diese Forscher nur eine untergeordnete Rolle. Ganz anders verhalten sich Geistesund Sozialwissenschaftler sowie Juristen. Für Geistes- und Sozialwissenschaftler ist ihre Sichtbarkeit allenfalls neutral (Median von 3 bei Google Scholar und Web of Science) oder sogar unwichtig (Median von 4 bei Scopus); für Juristen liegen die Werte noch niedriger (Median von 4 bei Google Scholar und von 5 bei Scopus und Web of Science). Die Wirtschaftswissenschaftler nehmen eine Mittelstellung zwischen den beiden Hauptgruppen ein: Ihnen ist ihre Sichtbarkeit bei den bibliographischen Datenbanken wichtiger als den Geistes-, Sozial- und Rechtswissenschaftlern, aber nicht so wichtig wie den Naturwissenschaftlern und Medizinern. Für sie ist die Sichtbarkeit bei Google Scholar am wichtigsten (Median von 2 und sehr wichtig-Antworten bei 39,3 %), gefolgt von Web of Science (Median von 2; sehr wichtig: 26,7 %) und abgeschlagen Scopus (Median von 3, sehr wichtig: 8,3 %). Wenn man von der Silent Generation absieht, schätzen die drei anderen Generationen ihre Sichtbarkeit bei Web of Science sehr ähnlich als wichtig ein (Median jeweils bei 2). An Scopus ist nur die jüngste Generation Y mit einem Median von 2 interessiert, für die Generation X und die Baby Boomer ist Scopus für ihre Sichtbarkeit neutral, für die Ältesten sogar unwichtig. Die Einschätzung der Relevanz von Google Scholar hängt eindeutig vom Alter ab: Je älter die Professoren sind, desto weniger ist für sie dieser Informationsdienst für die Sichtbarkeit ihrer Publikationen wichtig. Während 46,7 % der Generation Y Google Scholar als sehr wichtig einschätzen, stimmen diesem Votum nur 30,0 % der Generation X, 22,2 % der Baby Boomer und 0 % der Silent Generation zu.

9.5 Relevanz der H-Indices auf Web of Science, Scopus und Google Scholar in der Einschätzung der Professoren

Für Hirsch ist der H-Index eine Schätzung der Wichtigkeit, der Signifikanz und des breiten Einflusses des kumulierten Werks eines Wissenschaftlers („which gives an estimate of the importance, significance, and broad impact of a scientist's cumulative research contribution"; Hirsch, 2005, S. 16572) bzw. ein angedachter Indikator wissenschaftlichen Erfolgs bzw. wissenschaftlicher Leistung („measure of scientific achievement"; Hirsch, 2020, S. 4), der auch prognostische Aussagen zulässt (Hirsch, 2007). Der in Abbildung 9.1 genannte Autor hat bei Web of Science einen H-Index von 14, was heißt, dass 14 seiner (insgesamt 168) bei Web of Science gelisteten Publikationen mindestens 14-mal in ebenfalls bei Web of Science indexierten Publikationen zitiert worden sind. Die Gesamtzahl der Publikationen (hier 168) und der Zitationen (hier 598) spielt beim H-Index keine Rolle. Sichtbarkeit, Zitationen und H-Index hängen zusammen; schließlich kombiniert der H-Index Anzahlen von Publikationen und Zitationen. Hat ein Wissenschaftler auf einer Plattform nur eine geringe Sichtbarkeit (also wenige dort indexierte Publikationen), wird auch der H-Index niedrig ausfallen. Eine hohe Sichtbarkeit ist allerdings nur eine notwendige Bedingung für einen hohen H-Index. Hinzukommen müssen hohe Anzahlen an Zitationen, und diese sind abhängig von den Zitiergewohnheiten einer Disziplin, den Themen der Artikel, deren Alter und den bisherigen Zitationen der Werke des betreffenden Autors (Amancio, Oliveira, & Costa, 2012).

Es ist leicht, unterschiedliche Varianten des H-Index auszumachen, die einerseits von der zugrunde gelegten Datenbank und andererseits von modifizierten Berechnungsmethoden abhängen. Neben den H-Indices bei Web of Science, Scopus und Google Scholar (Bar-Ilan, 2008) gibt es zusätzlich Werte etwa bei ResearchGate. Nach Hirschs initialer Publikation zum H-Index wurden ähnliche, mathematisch nur unwesentlich modifizierte Formeln veröffentlicht, die allerdings kaum zu neuen Erkenntnissen führten (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009; Bornmann, Mutz, & Daniel, 2008; Jan & Ahmad, 2020), da zwischen den Werten der Varianten recht hohe Korrelationen bestehen (Bornmann, Mutz, Hug, & Daniel, 2011). Wir fragten die Professoren nur nach der ursprünglichen Variante des H-Index und nach ihrer Einschätzung der Relevanz dieses Indikators für die Darstellung ihrer Forschungsleistung auf den drei Plattformen Web of Science, Scopus und Google

Scholar (Tabelle 9.4).

Die Ergebnisse zur Einschätzung des H-Index sind denen zu den Publikationsraten bei Web of Science, Scopus und Google Scholar ähnlich, wobei sich die Werte zum H-Index im Gegensatz zur Sichtbarkeit etwas nach rechts (also in der Tendenz zu weniger wichtig) verschoben haben. Auch hier ist für alle Forscher der H-Index bei Web of Science (Modus: 2; Median: 2; sehr wichtig-Antworten: 25,2 %) der höchstgeschätzte Indikator, gefolgt von Google Scholar (Modus: 2; Median: 3, sehr wichtig- Antworten: 19,4 %) und Scopus (Modus: 3; Median: 3, sehr wichtig-Antworten: 13,7 %).

Tabelle 9.4: Relevanz der H-Indices auf Web of Science, Scopus und Google Scholar in der Einschätzung von Professoren in Deutschland.

a) h-Index / Web of Science	(1)	(2)	(3)	(4)	(5)	Mo	Me	IQA	Sign.	N
Alle	25,2%	27,7%	17,5%	13,3%	16,3%	2	2	3	-	961
Männer	24,3%	28,6%	18,2%	13,7%	15,2%	2	2	2	0,891	781
Frauen	30,9%	22,2%	13,6%	13,0%	20,4%	1	2	3		162
Geowiss., Landwirtschaft usw.	32,5%	40,0%	21,3%	2,5%	3,8%	2	2	2		80
Geistes- u. Sozialwissenschaften	6,3%	13,1%	16,8%	26,2%	37,7%	5	4	2		191
Naturwissenschaften	29,3%	30,2%	18,2%	9,8%	12,6%	2	2	2	0,000 ***	461
Medizin	41,5%	33,9%	11,0%	9,3%	4,2%	1	2	2		118
Recht	0,0%	0,0%	9,1%	18,2%	72,7%	5	5	1		11
Wirtschaftswissenschaften	19,7%	27,6%	23,7%	18,4%	10,5%	2	3	2		76
Generation Y	25,9%	31,8%	21,2%	9,4%	11,8%	2	2	2		85
Generation X	25,9%	28,7%	16,2%	11,4%	17,9%	2	2	3		599
Baby Boomer	25,3%	25,3%	17,5%	17,9%	14,0%	1;2	2	3	0,305	229
Silent Generation	11,1%	0,0%	44,4%	33,3%	11,1%	3	3	1		9
b) h-Index / Scopus	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**Mo**	**Me**	**IQA**	**Sign.**	**N**
Alle	13,7%	26,3%	27,0%	15,0%	18,0%	3	3	2	-	946
Männer	12,0%	27,2%	28,1 %	15,7%	17,0%	3	3	2	0,343	769
Frauen	22,6%	21,4%	21,4%	13,2%	21,4%	2;3	3	2		159
Geowiss., Landwirtschaft usw.	24,1%	35,4%	32,9%	2,5%	5,1%	2	2	1		79
Geistes- u. Sozialwissenschaften	4,2%	13,2%	19,0%	26,5%	37,0%	5	4	2		189
Naturwissenschaften	15,4%	27,5%	29,5%	12,5%	15,2%	3	3	2	0,000 ***	455
Medizin	18,4%	39,5%	25,4%	11,1%	5,3%	2	2	1		114
Recht	0,0%	0,0%	9,1%	18,2%	72,7%	5	5	1		11
Wirtschaftswissenschaften	12,2%	23,0%	35,1%	18,9%	10,8%	3	3	2		74
Generation Y	19,8%	31,4%	30,2%	8,1%	10,5%	2	2	1		86

0,015
*

Tabelle 9.4: (fortlaufend)

b) h-Index / Scopus	(1)	(2)	(3)	(4)	(5)	Mo	Me	IQA	Sign.	N
Generation X	13,9%	27,5%	25,8%	12,9%	19,9%	2	3	2		589
Baby Boomer	12,6%	23,3%	27,8%	21,1%	15,2%	3	3	2		223
Silent Generation	0,0%	11,1%	33,3%	44,4%	11,1%	4	4	1		9
c) h-Index / Google Scholar	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**Mo**	**Me**	**IQA**	**Sign.**	**N**
Alle	19,4%	27,7%	24,3%	12,8%	15,7%	2	3	2	-	966
Männer	18,3%	28,5%	25,0 %	13,4%	14,8%	2	3	2	0,876	783
Frauen	24,2%	23,0%	21,8%	11,5%	19,4%	1	3	2		165
Geowiss., Landwirtschaft usw.	20,3%	34,2%	34,2%	6,3%	5,1%	2;3	2	1	0,000 ***	79
Geistes- u. Sozialwissenschaften	9,7%	17,4%	20,0%	20,5%	32,3%	5	4	3		195
Naturwissenschaften	23,0%	31,3%	22,5%	10,3%	12,9%	2	2	1		466
Medizin	16,7%	32,5%	34,2%	12,4%	4,4%	3	3	1		114
Recht	0,0%	0,0%	9,1%	19,2%	72,7%	5	5	1		11
Wirtschaftswissenschaften	29,9%	22,1%	23,4%	13,0%	11,7%	1	2	2		77
Generation Y	34,9%	27,9%	20,9%	7,0%	9,3%	1	2	2	0,000 ***	86
Generation X	19,2%	29,8%	22,8%	11,1%	17,1%	2	3	2		604
Baby Boomer	15,4%	22,9%	28,6%	18,9%	14,1%	3	3	2		227
Silent Generation	0,0%	11,1%	44,4%	33,3%	11,1%	3	3	1		9

Frage: „Wie wichtig ist es Ihnen, auf folgenden Plattformen einen hohen H-Index zu erlangen: Web of Science, Scopus, Google Scholar?“
Skala: 1: sehr wichtig, 2: wichtig, 3: neutral, 4: unwichtig, 5: überhaupt nicht wichtig;
Mo: Modus; Me: Median; IQA: Interquartilsabstand; Sign.: Signifikanz (u-Test, h-Test); N: Anzahl der jeweiligen Teilnehmer.

Es gibt keine signifikanten Unterschiede zwischen Frauen und Männern, wohl aber bei den Fächern und den Generationen. Mediziner schätzen den H-Index bei Web of Science, akzeptieren den bei Scopus und verhalten sich gegenüber dem H-Index bei Google Scholar eher neutral. Für Naturwissenschaftler und Geowissenschaftler usw. ist der H-Index auf allen drei Plattformen wichtig (Median jeweils 2), für Geistes- und Sozialwissenschaftler unwichtig (Median jeweils 4) und für Rechtswissenschaftler sogar völlig irrelevant (Median jeweils 5). Den Wirtschaftswissenschaftlern ist der H-Index bei Google Scholar am wichtigsten (Median: 2, sehr wichtig-Antworten: 29,9 %), während die H-Indices bei Web of Science und Scopus im Schnitt neutral eingestuft werden. Analog zur Sichtbarkeit ist für alle Generationen (außer der Silent Generation) der H-Index bei Web of Science wichtig (Median: 2), während die Einschätzung der Relevanz der H-Indices bei Scopus und Google Scholar mit zunehmendem Alter geringer ausfällt.

9.6 Wissensstand der Professoren zum H-Index

Abschließend kommen wir zum Wissensstand der Professoren zum H-Index. Bei dieser Forschungsfrage geht es nicht nur um die persönliche Einschätzung der Umfrageteilnehmer zu ihrem Wissen bezüglich des H-Index, sondern zusätzlich um einen kleinen Wissenstest (Tabelle 9.5). 60,5 % der Professoren kennen den H-Index und haben ihr Wissen darüber korrekt eingeschätzt, 32,4 % kennen nach eigener Einschätzung diesen Indikator nicht und 7,2 % meinen, ihn zu kennen, fallen aber beim Wissenstest durch. Zwischen Männern und Frauen bestehen keine statistisch signifikanten Unterschiede, obgleich die Hälfte der Professorinnen nach eigener Einschätzung Definition und Berechnungsweg des H-Index nicht kennt (Männer: 28,7 %). Die Unterschiede bei den Fächern und den Generationen sind auch bei dieser Untersuchung extrem signifikant. Bei Naturwissenschaftlern (79,1 %), Geowissenschaftlern usw. (74,4 %) und Medizinern (70,6 %) ist das Wissen über den H-Index weit verbreitet, bei Geistes- und Sozialwissenschaftlern (21,1 %) und Rechtswissenschaftlern (7,1 %) weiß nur eine Minderheit, wie sich der H-Index berechnen lässt. Eine besondere Stellung nehmen die Wirtschaftswissenschaftler ein: Hier kennen zwar 48,3 % den H-Index, aber 13,8 % meinen fälschlicherweise, ihn zu kennen. Bei den Generationen gilt, dass mit zunehmendem Alter die Wahrscheinlichkeit sinkt, dass ein Professor den H-Index kennt.

9.7 Wesentliche Ergebnisse

Unsere Hauptergebnisse sind Daten zur Einschätzung der Forscher über die Wichtigkeit von Publikationen und Zitationen, zu ihrer eigenen Sichtbarkeit auf Web of Science, Scopus und Google Scholar, zu den H-Indices bei diesen Informationsdiensten sowie ihrem Wissensstand zum H-Index. Für nahezu alle befragten Professoren sind Publikationen wichtig, für Mediziner sogar sehr wichtig. Bei allen anderen Fragen haben wir hochsignifikante Unterschiede zwischen den Fächern festgestellt. Für die Naturwissenschaftler (einschließlich Geowissenschaftler, Landwirtschaftswissenschaftler usw.) und Mediziner sind ihre Zitationen, ihre Sichtbarkeit und ihr H-Index wichtig, während für die Geistes- und Sozialwissenschaftler, Wirtschaftswissenschaftler und Juristen Zitationen, Sichtbarkeit wie H-Index wesentlich weniger wichtig sind. Für die Befragten aus Naturwissenschaft und Medizin sind Sichtbarkeit und H-Index bei Web of Science am wichtigsten, gefolgt vom H-Index bei Google Scholar und Scopus. Überraschenderweise ist der H-Index von Google Scholar für Ökonomen sehr attraktiv. Wir fanden kaum signifikante Unterschiede zwischen den Ergebnissen von Männern und Frauen, es gibt jedoch erhebliche Unterschiede in Bezug auf die Generationen: Je älter die Professoren sind, desto weniger wichtig schätzen sie für sich Sichtbarkeit und H-Index ein.

Zwei Fünftel der Professoren kennen keine Details zum H-Index oder – was schon ziemlich befremdlich ist – glauben fälschlicherweise zu wissen, wie der H-Index definiert ist und berechnet wird, haben aber unseren einfachen Wissenstest nicht bestanden. Je älter die Generation ist, desto höher ist der Anteil der Teilnehmer, die die

Tabelle 9.5: Wissensstand von Professoren in Deutschland zum H-Index.

Wissensstand zum H-Index	Forscher kennt H-Index (1)	Forscher kennt H-Index nicht (2)	Forscher schätzt Wissensstand falsch ein (3)	Sign.	N
Alle	60,5%	32,4%	7,2%	-	1017
Männer	64,6%	28,7%	6,9%	0,284	837
Frauen	41,6 %	50,0%	8,3%		180
Geowiss., Landwirtschaft usw.	74,4%	16,7%	9,0%	0,002 ***	78
Geistes- u. Sozialwissenschaften	21,1%	71,7%	7,2%		237
Naturwissenschaften	79,1%	15,0%	6,0%		479
Medizin	70,6%	21,8%	7,6%		119
Recht	7,1%	92,9%	-		14
Wirtschaftswissenschaften	48,3%	37,9%	13,8%		87
Generation Y	64,9%	26,6%	8,5%	0,000 ***	94
Generation X	62,9%	32,0%	5,1%		644
Baby Boomer	53,2%	35,6%	11,2%		250
Silent Generation	16,7%	75,0%	8,3%		12

Wissensstand:
(1): Forscher meint, Definition und Berechnung des H-Index zu kennen und besteht den Wissenstest;
(2): Forscher meint, Definition und Berechnung des H-Index nicht zu kennen;
(3) Forscher meint, Definition und Berechnung des H-Index zu kennen und besteht den Wissenstest nicht;
Sign.: Signifikanz (Chi-Quadrate); N: Anzahl der jeweiligen Teilnehmer.

Definition und Berechnung dieses szientometrischen Indikators nicht kennen. Konkretes Wissen der Forscher über den H-Index ist in den akademischen Disziplinen der Naturwissenschaften und der Medizin mehrheitlich verbreitet, in den Geistes- und Sozialwissenschaften ist es dagegen viel weniger zu finden.

9.8 Warum gibt es diese Unterschiede bei den Wissenschaftsdisziplinen?

Wie können wir die Unterschiede bei den Fachgruppen erklären? Die durchaus sehr großen allgemeinwissenschaftlichen Informationsdienste Web of Science und Scopus sind, verglichen mit den persönlichen Literaturlisten von Forschern, recht unvollständig (Hilbert et al., 2015). Ebenso ist dort eine ausgeprägte Ungleichbehandlung gewisser Disziplinen (Mongeon & Paul-Hus, 2016) und vieler Sprachen (außer Englisch) zu beobachten (Vera-Baceta, Thelwall, & Kousha, 2019). Vielleicht halten diese Fakten insbesondere Vertreter der benachteiligten Disziplinen und Sprachen (darunter auch der deutschen) davon ab, die Relevanz ihrer Sichtbarkeit und ihren H-Index auf diesen Plattformen als wichtig einzustufen. Dann verwundert aber die ebenso zu sehende Ablehnung der Kennwerte bei Google Scholar, denn dieser Informationsdienst ist der mit Abstand vollständigste (Martin-Martin, Orduna-Malea, Thelwall, & Lopez-Cozar, 2018). Hier verhalten sich jedoch die Wirtschaftswissenschaftler sehr informiert, da sie – als einzige akademische Vertreter – ihre Sichtbarkeit und ihren H-Index bei Google Scholar hoch schätzen. Die Verwendung von Google Scholar zur Forschungsevaluierung wird allgemein und kritisch diskutiert (Halevi, Moed, & Bar-Ilan, 2017). Je nach der eigenen Meinung eines Forschers zu diesem Thema könnte dies ein Grund für die hohe Ablehnung sein.

Ein weiterer Erklärungsversuch könnten die unterschiedlichen Forschungskulturen in den verschiedenen akademischen Disziplinen sein. Gemäß Kagan (2009, S. 4) sehen Naturwissenschaftler und Mediziner ihr Hauptforschungsinteresse in Erklärung und Prognose, während es für Geisteswissenschaftler eher das Verstehen ist (ähnlich argumentierten Snow, 1959 und bereits Dilthey, 1895, S. 10). Der H-Index ist durchaus typisch für die Forschungskultur der Naturwissenschaften. Forscher aus der Naturwissenschaft und der Medizin sind an Zahlen gewöhnt, während Geisteswissenschaftler selten quantitativ arbeiten. Nach Kagan (2009, S. 5) sind Geisteswissenschaftler nur minimal auf externe Unterstützung angewiesen, Naturwissenschaftler und Mediziner sind dagegen in hohem Maße von externen Finanzierungsquellen abhängig. Sichtbarkeit und H-Index können als Argumente bei der Allokation externer Unterstützung dienen. Für Naturwissenschaftler und Mediziner sind Sichtbarkeit und H-Index sehr verbreitete Gebilde, die sie offenbar für ihr akademisches Überleben benötigen. Geisteswissenschaftler sind mit numerischen Indikatoren nicht so vertraut, und für sie wird der H-Index nicht so dringend benötigt wie für ihre Kollegen aus den Fakultäten für Naturwissenschaft und Medizin. Diese dichotome Klassifizierung von Forschungskulturen mag jedoch eine zu vereinfachende Lösung sein (Kowalski & Mrdjenovich, 2016) und auch in den Geistes- und

Sozialwissenschaften gibt es einen Trend für die Verwendung solcher Indikatoren zur Forschungsevaluation. Für die Erstellung einer zufriedenstellenden Theorie des Verhaltens von Forschern in Bezug auf Sichtbarkeit in Informationsdiensten und auf den H-Index (oder im Allgemeinen in Bezug auf szientometrische Indikatoren) – und dies auch in Abhängigkeit von dem Hintergrund einer akademischen Disziplin – ist noch viel mehr empirische Arbeit in der Hochschulforschung erforderlich. Neben der vorgestellten Situation für deutsche Professoren und ihr Wissen über den H-Index bilden weitere Fallstudien bereits eine (Aksnes & Rip, 2009; Buela-Casal & Zych, 2012; Chen & Lin, 2018; Derrick & Gillespie, 2013; Haddow & Hammarfelt, 2019; Hammarfelt & Haddow, 2018; Lemke, Mehrazar, Mazarakis, & Peters, 2019; Ma & Ladisch, 2019; S. Rousseau & Rousseau, 2017). Da Sichtbarkeit, Publikations- und Zitationszahlen sowie der H-Index immer noch einen wichtigen Einfluss auf die Evaluation von Wissenschaftlern haben und nicht alle Forscher über diese szientometrischen Forschungsindikatoren sehr gut informiert sind, scheint es eine gute Idee für die Hochschulpraxis zu sein, das Wissen der Hochschullehrer im breiteren Bereich der „Metrik-Weisheit" („metrics wiseness"; R. Rousseau et al., 2018) zu vertiefen. Analog zur Ausbildung in Informationskompetenz schlägt Haustein (2018) Lehrmaterialien zur Erlangung von Metrikkompetenzen („metrics literacies") vor, um über korrekte Interpretationen von Indikatoren zur Forschungsevaluation aufzuklären, adäquate Einsatzgebiete bestimmter Indikatoren zu benennen und um Missbrauch (Haustein & Larivière, 2015) zu minimieren.

Literatur

Aguillo, I. F. (2011). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics*, *91*, 343–351.

Aksnes, D. W., & Rip, A. (2009). Researchers' perceptions of citations. *Research Policy*, *38*(6), 895–905.

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). H-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*(4), 273–289.

Amancio, D. R., Oliveira, O. N., & Costa, L. D. (2012). Three-feature model to reproduce the topology of citation networks and the effects from authors' visibility on their h-index. *Journal of Informetrics*, *6*(3), 427–434.

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386.

Bar-Ilan, J. (2008). Which h-index? – A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, *74*(2), 257–271.

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, *1*(1), 363–375.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, *59*(5), 830–837.

Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H.-D. (2011). A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics*, *5*(3), 346–359.

Buela-Casal, G., & Zych, I. (2012). What do the scientists think about the impact factor? *Scientometrics*, *92*, 281–292.

Chen, C. M.-L., & Lin, W.-Y. C. (2018). What indicators matter? The analysis of perception toward research assessment indicators and Leiden Manifesto. The case study of Taiwan. In *Proceedings of the 23rd International Conference on Science and Technology Indicators (STI 2018)* (pp. 688–698). Leiden, NL: Centre for Science and Technology Studies (CWTS).

Derrick, G. E., & Gillespie, J. (2013). "A number you just can't get away from": Characteristics of adoption and the social construction of metric use by researchers. In *Proceedings of the 18th International Conference on Science and Technology Indicators* (pp. 104–116). Berlin: Institute for Research Information and Quality Assurance.

Destatis. (2019). Bildung und Kultur. *Personal an Hochschulen (Fachserie 11, Reihe 4.4)*. Wiesbaden: Statistisches Bundesamt.

Deutscher Hochschulverband. (2020). Universitäten Deutschland. *Hochschullehrer-Verzeichnis 2020,* Band 1. 28. Ausgabe. Berlin, New York: De Gruyter Saur.

Dilthey, W. (1895). Ideen über eine beschreibende und zergliedernde Psychologie. *Sitzungsberichte der königlich preussischen Akademie der Wissenschaften zu Berlin, 7. Juni 1894, Ausgabe XXVI, Sitzung der philosophisch historischen Classe*, 1–88.

Dorsch, I. (2017). Relative visibility of authors' publications in different information services. *Scientometrics*, *112*(2), 917–925.

Dorsch, I., Askeridis, J., & Stock, W. G. (2018). Truebounded, overbounded, or underbounded? Scientists' personal publication lists versus lists generated through bibliographic information services. *Publications*, *6*(1), 1–9.

Fietkiewicz, K. J., Lins, E., Baran, K. S., & Stock, W. G. (2016). Inter-generational comparison of social media use: Investigating the online behavior of different generational cohorts. In *Proceedings of the 49th Hawaii International Conference on System Sciences* (pp. 3829–3838). Washington, DC: IEEE.

Gross, C., Jungbauer-Gans, M., & Kriwy, P. (2008). Die Bedeutung meritokratischer und sozialer Kriterien für wissenschaftliche Karrieren – Ergebnisse von Expertengesprächen in ausgewählten Disziplinen. *Beiträge zur Hochschulforschung*, *30*(4), 8–32.

Haddow, G., & Hammarfelt, B. (2019). Quality, impact, and quantification: Indicators and metrics use by social scientists. *Journal of the Association for Information Science and Technology*, *70*(1), 16–26.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items.*

New York: Routledge.

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation. Review of the literature. *Journal of Informetrics*, *11*(3), 823–834.

Hammarfelt, B., & Haddow, G. (2018). Conflicting measures and values: How humanities scholars in Australia and Sweden use and react to bibliometric indicators. *Journal of the Association for Information Science and Technology*, *69*(7), 924–935.

Haustein, S. (2018). *Metrics literacy [Blog post].* Retrieved from `https://stefaniehaustein.com/metrics-literacy/`

Haustein, S., & Larivière, V. (2015). The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects. In I. Welpe, J. Wollersheim, S. Ringelhan, & M. Osterloh (Eds.), *Incentives and Performance: Governance of Research Organizations* (pp. 121–139). Cham, CH: Springer.

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, *1*(1), 387–395.

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., … Stock, W. G. (2015). Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. *Online Information Review*, *39*(2), 255–264.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Hirsch, J. E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19193–19198.

Hirsch, J. E. (2020). Superconductivity, What the h? The emperor has no clothes. *Physics and Society*, *49*(1), 4–9.

Hu, G. Y., Wang, L., Ni, R., & Liu, W. S. (2020). Which h-index? An exploration within the Web of Science. *Scientometrics*, *123*, 1225–1233.

Jan, R., & Ahmad, R. (2020). H-index and its variants: Which variant fairly assess author's achievements. *Journal of Information Technology Research*, *13*(1), 68–76.

Jappe, A. (2020). Professional standards in bibliometric research evaluation? A metaevaluation of European assessment practice 2005 – 2019. *PLoSONE*, *15*(4), e0231735, 1–233.

Kagan, J. (2009). *The Three Cultures. Natural sciences, social sciences, and the humanities in the 21st century.* Cambridge, MA: Cambridge University Press.

Kowalski, C. J., & Mrdjenovich, A. J. (2016). Beware dichotomies. *Perspectives in Biology and Medicine*, *59*(4), 517–535.

Kraker, P., & Lex, E. (2015). A critical look at the ResearchGate score as a measure of scientific reputation. In *Quantifying and Analysing Scholarly Communication on the Web (ASCW'15).* Oxford, 30 June 2015 (3 S.).

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621.

Lemke, S., Mehrazar, M., Mazarakis, A., & Peters, I. (2019). "When you use social media you are not working": Barriers for the use of metrics in social sciences. *Frontiers in Research Metrics and Analytics*, *3*(39), 1–18.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 5–55.

Linde, F., & Stock, W. G. (2011). *Information markets. A strategic guideline for the i-commerce*. Berlin, New York: De Gruyter Saur.

Ma, L., & Ladisch, M. (2019). Evaluation complacency or evaluation inertia? A study of evaluative metrics and research practices in Irish universities. *Research Evaluation*, *28*(3), 209–217.

Mann, H., & Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*(1), 50–60.

Martin-Martin, A., Orduna-Malea, E., Thelwall, M., & Lopez-Cozar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177.

Meschede, C., & Siebenlist, T. (2018). Cross-metric compatibility and inconsistencies of altmetrics. *Scientometrics*, *115*, 283–297.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *Series 5, 50*(302), 157–175.

Rassenhövel, S. (2010). *Performancemessung im Hochschulbereich. Theoretische Grundlagen und empirische Befunde*. Wiesbaden: Gabler.

Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise: A bibliometric guide for researchers*. Cambridge, MA: Chandos.

Rousseau, S., & Rousseau, R. (2017). Being metric-wise: Heterogeneity in bibliometric knowledge. *El Profesional de la Información*, *26*(3), 480–487.

Schlögl, C. (2013). Internationale Sichtbarkeit der europäischen und insbesondere der deutschsprachigen Informationswissenschaft. *Information - Wissenschaft und Praxis*, *64*(1), 1–8.

Snow, C. P. (1959). *The two cultures and the scientific revolution*. Cambridge, UK: Cambridge University Press.

Stock, M., & Stock, W. G. (2003). Web of Knowledge. Wissenschaftliche Artikel, Patente und deren Zitationen: Der Wissenschaftsmarkt im Fokus. *Password*, *10*, 30–37.

Stock, W. G. (2001). *Publikation und Zitat. Die problematische Basis empirischer Wissenschaftsforschung*. Köln: Fachhochschule Köln; Fachbereich Bibliotheks- und Informationswesen (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 29).

Stock, W. G., & Stock, M. (2013). *Handbook of information science.* Berlin, Boston: De Gruyter Saur.

Sugimoto, C. R., & Larivière, V. (2013). *Measuring research: What everyone needs to know.* New York: Oxford University Press.

Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, *115*(3), 1231–1240.

Vera-Baceta, M. A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, *121*(3), 1803–1813.

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, *1*(1), 396–413.

Chapter 10

Discussion and Conclusion

This cumulative dissertation reports on social media and scientometrics in empirical informetrics. The focus lay in both fields and with this, on their intersections in terms of information content, users, and systems in information science. On the one hand, scientometrics is a well-established information science field. Considering its roots within bibliometrics and the field of documentation, it is a discipline continuously developing since the beginnings of information science (Saracevic, 2009). On the other hand, social media, as a young research area, arose with the upcoming of digital technologies, user-generated content, and the so-called web 2.0 (O'Reilly, 2005). Researchers from various disciplines already heavily investigate social media. It became an inherent part of the study of information science. How does research in both branches in terms of their published content (e.g., social media postings, research publications) and likewise in terms of their impact (e.g., topics of interest, citations) work? Following RQ1, the results and implications of *Chapter 2-5* in *Part 1* are discussed.

Part 1: Social Media in Informetrics

RQ1: How does informetrics in social media research work? How can we measure information content of social media documents? How do social media users describe the content of their documents?

Social media platforms like Instagram or YouNow allow their users not only the creation of textual-content but rather multi-visual media. Photos, (live-streamed) videos, and ephemeral moments became an indispensable part of the applications. For textual-content, besides pure text, hashtags for the representation of knowledge exist. Hashtags are user-generated, enable social media communication, and are often not only metadata to describe the content but rather a part of the posted content, e.g., within an Instagram picture posting denoting a further dimension or in relation with other elements of the posting (e.g., picture content, caption) (Laestadius, 2017). Thus, they are a rich resource for social media (content) analysis. *Chapter 2* and *4* focus on hashtagged Instagram picture postings in general (*Chapter 2*) and also with respect to gender dependent differences (*Chapter 4*). One picture posting contains 15 hashtags on average (*Chapter 2* and *4*), whereas women use one hashtag less than men (*Chapter 4*) (RQ1). A further key finding of both studies is that the majority of assigned hashtags describes the posted content itself, followed by hashtags relating to technical aspects not depicted (Isness, e.g., picture author). Considering gender-differences, women tend to use slightly more emotional hashtags. Men use slightly more "Insta"- and "Isness"-tags, but overall, both genders do not tag so differently. Furthermore, both studies report on a small statistical association between hashtag categories and picture categories (for male and female Instagram users).

Following this, *Chapter 3* thematizes the analysis of Instagram users' hashtag creation behavior and selection process in terms of solely self-created hashtags and the utilization of best practices. More than half (55.7%) of the surveyed responded to use hashtags on Instagram. Further, they state to make use of both self-created hashtags (41.4%) and hashtags inspired by others (58.6%). More personal sources, like friends, other users, or Instagram functions, serve as frequently used hashtag inspiration. The utilization of false hashtags does not play a role.

Chapter 5 lays its focus on social live streaming users and their motivation for financial gain or fame. The content analysis of streams on the platforms Periscope, Ustream, and YouNow reveals, users motivated by financial gain mostly stream entertainment media. For streamers with the wish to become a star or "micro-celebrity", chatting and making music are most important. In regards to the platform and generational dependencies on YouNow and for users of generation Z (born after 1998), the motivation for fame is highest. Older generations (Gen X, Silver Surfers) are the strongest motivated ones to become rich. Likewise, Ustream is the platform with streamers highest motivated for financial gain.

There exist several ways to analyze social media data. Methodology and required data should be selected in dependency of the research question(s) (Sloan & Quan-Haase, 2017). Informetric analyses may also be applied for social media but they differ depending on the format of the analyzed pieces of information. As we have seen in the conducted studies at hand, for multi-visual media like in Instagram, text and hashtags add further information to media. Likewise, previously ephemeral moments become more persistent and enduring on the platform (Laestadius, 2017). As Laestadius (2017) points out based on (Boyd, 2010, 2014), Instagram provides persistence, visibility (public accounts, besides private), replicability (possibility of screenshots), searchability (especially through hashtags), and a high degree of interpretability of content for research studies (rich resource of data, not only text). That leads us to the discussion of *Chapter 6-9* within *Part 2*.

* * *

Part 2: Scientometrics

RQ2: How does scientometrics work? Is there a reliable data basis for scientometric studies? How is it possible to analyze research topics? How reliable are the data used for analysis in informetrics?

Within the analysis of scholarly research output, the analysis of research publications lays in the area of scientometrics. Research publication evaluation of productivity and impact can take place on different units of assessment (Rousseau, Egghe, & Guns, 2018), like on the output of individual researchers (micro level) or institutions (meso level). Digital scientific information services, for example, databases as Web of Science (WoS) or social media platforms like ResearchGate are part of many studies. One aspect is the coverage of publications and citations in such services. The re-interpreted scholarly indicator "visibility" (*Chapter 6*) addresses research publications and their coverage in information services in comparison to the full research output of a unit of assessment. In this case, individual information science researchers Blaise Cronin and Wolfgang G. Stock (*Chapter 6*) and nine International Society of Scientometrics and Informetrics (ISSI) Committee members (*Chapter 7*) were analyzed in detail. What is important to mention, visibility focuses on the perspective of the unit of assessment. That means, how visible a unit of assessment within an information service is. Several information services were part of my analyses. These case studies on researchers' visibility reveal a misbalance in information services. None of the researchers had full visibility in any of the investigate information services and thus a truebounded publication list. The highest visibility is given in Google Scholar. However, this is also the database often criticized for issues like low accuracy (Halevi, Moed, & Bar-Ilan, 2017). For scientometric analysis, a truebounded publication list but also accuracy within the information service are important. The results further show a publications' document type and language can influence a researchers' visibility. In general, the publication medium (and with this the document type) plays a role since some information services like Web of Science and Scopus are limited to certain publishers and document types. In social media services like Research Gate and Mendeley, but also for the free accessible database Google Scholar, researchers can influence their visibility. If they have an author profile, they can update their publication lists within these information services. More universal, this is also possible with an own ORCID ID (ORCID, 2020a). This 16-digit digital number serves as a career-long identifier for a researchers' professional activities (including publications) (ORCID, 2020c, 2020b).

Further concerning the research productivity, *Chapter 8* reveals what research topics are relevant at the Institute of Information Science and Information Systems at the University of Graz and the Department of Information Science at Heinrich Heine University in Düsseldorf during the years 2010 to 2016. Both have an expected overall focus on INFORMATION and SCIENCE. This is not surprising since both institutes are located in this discipline, although with different foci. The institution in Düsseldorf is more specialized in the analysis of SERVICEs, whereas it is SYSTEMs for the institution in Graz. MOBILE, AUSTRIA, JOURNAL, READERSHIP, DOWNLOAD, ECONOMIC, and BUSINESS are topics related to research in Graz as well. Düsseldorf studies INFORMATIONAL CITY, social media (SOCIAL + NETWORK; TWITTER, TWEET, FACEBOOK), KNOWLEDGE + REPRESENTATION, and information retrieval related to EMOTIONs and RECOMMENDATIONs. Other topics of both institutions are within the branch of science communication (COMMUNICATION, CITATION) and digitalization (DIGITAL). Empirical library science (LIBRARY), research on information literacy (LITERACY), user research, or the information behavior of users (USER, BEHAVIOR) are points of contact too.

With respect to productivity and impact, *Chapter 9* identifies German university professors' opinions about publications, citations, and the h-index, as well as their knowledge on the latter. For the majority, publications are important. For medical professors, they are even very important. Likewise, professors in the fields of natural sciences and medicine assess citations, their visibility, and their h-index as important. These aspects are less important for professors in the field of humanities and social sciences, economy, and law. In general, the importance of the h-index decreases with the increasing age of our participants. The visibility and the h-index within the information service Web of Science are most important for natural and medical scientists, followed by Google Scholar and Scopus. Economists value the h-index calculated in Google Scholar. Two-fifths of all professors surveyed do not know details about the H-index. Scholars in the field of natural sciences and medicine know more about the h-index than their colleagues in humanities and social sciences. Similar to the importance, the knowledge of the h-index decreases with the increasing age of our participants.

The studies show that the solely analysis based on information services is not always a reliable data basis in terms of the completeness of the content. Scholarly indicators referring to publications and citations mostly provide a quantitative analysis. Furthermore, and like in the social media part of this work, the direct survey of researchers adds more information of the general perception of such analyses but also on their issues. During time, the scientometric' toolbox had also led to several misuse (Haustein & Larivière, 2015). As the study in *Chapter 9* provides first insights about the opinion and knowledge of the h-index, as one scholarly indicator heavily discussed and misused within research evaluation, there is more need for change within the academic system. One promising possibility seems to be the education of metrics literacies (Haustein, 2018).

Overall and in conclusion, studies on social media and scientometrics are subjects of analyses in informetrics. Information content, users, and systems are the objects of study. This work builds on the analysis of social media in the area of information science. New insights in this area specific to the content and users of the multi-visual platforms Instagram, Periscope, Ustream, and YouNow are given. Previous content analysis on Instagram focused on a specific topic or event. Hashtags or user profiles build the foundation of analysis. However, the studies in this work set the content analysis in a broader context and connected it to established information science concepts and theories. In traditional information science, more specifically in the field of knowledge representation, metadata is used to describe the content of documents. Research focuses on the representation organization and condensation of knowledge in digital systems. For example, how can knowledge in such systems be represented (Stock & Stock, 2013)? With digital information and in digital information services like research databases, indexers and authors are responsible for this task. For instance, when an author writes an abstract about their study (information condensation) and chose suitable keywords (information filtering). Besides the content, a document contains further formal information (Argamon et al., 2007). In social media, users and the respective system create such information, and the *Chapters 2* to *5* not only demonstrate this but rather show how this picture is drawn for multimedia content. In respect to scientometrics, this work provides research critical investigating applied methodologies of scientometrics and the scholarly indicator use. It is shown that visibility is a dependent factor and that methodologies for scientometric analysis further have to be adjusted.

Within information science, informetrics is empirical information science. Objects of study are systems, users, and information, thus the content itself. Conducting an informetric analysis means to generate new information about the objects of study in information science. It brings the obtained information into new relations (Stock & Stock, 2013). Following this, without informetrics, there would not exist such a thing as empirical information science. That emphasizes how important the role of informetrics is within the entire discipline of information science. The upcoming of more and more digital technologies furthermore, facilitates more and more possibilities for more diversified, complex, and richer informetric analyses. Informetrics is also utilized by related or other disciplines. However, this entails the development and adjustment of new or existing methods and this in turn not only brings new chances but also new challenges into the field. The central issue is always the data to be collected. An important factor is the completeness and integrity of the data. As the scientometric studies in this work have shown, data completeness and integrity are nothing for granted, for example, in (scientific) information services like Web of Science or Scopus. The same might apply to the field of social media, but with a slightly different context. Scientific information services are devoted to providing information. Thus, missing data can be attributed to the fact that for some objects not all information is available. Even though, the system - ideally - should provide it. Social media and their stakeholders are not primarily aligned to provide information and to be analyzed. Rather the rapid and dynamic evolution of social media brought up the analysis of social media data from different perspectives. McCay-Peet and Quan-Haase (2017) point out methodological questions, ethical questions, questions of scale (the examination of a phenomenon from different angels), social media use itself, the information or understanding of a phenomenon, social media as a toolkit, the online-offline gap, discipline orientations, and elements of social media engagement (in total seven) as perspectives for analysis (McCay-Peet & Quan-Haase, 2017). That provides a vast number of perspectives and points for data gathering. However, as mentioned, there might be "missing data" as well. It already lays in the nature of social media. For instance, when postings or part of postings are deleted or modified. Or, as another example, the methodological aspect that not everybody uses social media. Some data and analyses thus are limited to the population of social media users (which may not have the same scaled representation of our entire world population in certain kinds of topics). McCay-Peet and Quan-Haase (2017) highlight the need for more critical research to understand biases within social media research to support researchers and protect social media users.

> "Of particular importance is the need to determine best practices around ethical considerations. For instance, can scholars make use of social media data without the consent of authors of user-generated content? If they make use of the data, should this be done only in aggregate form? What repercussions, for example, imprisonment, stigma, ridicule, and harm to reputation can participants suffer from scholars [making] tweets and blog text searchable, even if anonymized? There is much work to be done not only around the social phenomena under investigation on social media platforms, but also concerning how scholars are procuring, storing, interpreting, and making use of social media data" (McCay-Peet & Quan-Haase, 2017, p. 13).

Indeed, these are significant claims and highly relevant for informetric social media analyses. At the same time, relation to scientometric methodologies can be drawn. This discipline is more established than social media research. In terms of its existence, there already exist research and programs to support scientometric researchers, but also to protect them (e.g., San Francisco Declaration on Research Assessment (DORA) for research evaluation (DORA, 2020)). Likewise, the study in *Chapter 9* contributed to this issue. However, there is more need to improve the use of scientometric data. As stated at the beginning, information science (and thus also informetrics) is a field in a "constant flux" (Saracevic, 2009) and informetrics in social media and scientometrics will continue to develop. Thereby, both branches should further learn and benefit from each other.

References

Argamon, S., Whitelaw, C., Chase, P., Hota, S. R., Garg, N., & Levitan, S. (2007). Stylistic text classification using functional lexical features. *Journal of the American Society for Information Science and Technology*, *58*(6), 802–822.

Boyd, D. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. *Networked Self: Identity, Community, and Culture on Social Network Sites*, 39–58.

Boyd, D. (2014). *It's complicated: The social lives of networked teens.* New Heaven: Yale University Press.

DORA. (2020). *About.* Retrieved from `https://sfdora.org/`

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation. Review of the literature. *Journal of Informetrics*, *11*(3), 823–834.

Haustein, S. (2018). *Metrics Literacy: Developing online material to educate researchers and research administrators about bibliometrics, usage metric, altmetrics, and technometrics.* Retrieved from `http://stefaniehaustein.com/metrics-literacy/{\%}0A`

Haustein, S., & Larivière, V. (2015). The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects. In I. Welpe, J. Wollersheim, S. Ringelhan, & M. Osterloh (Eds.), *Incentives and Performance: Governance of Research Organizations* (pp. 121–139). Cham, CH: Springer.

Laestadius, L. (2017). Instagram. In *The SAGE Handbook of Social Media Research Methods.* SAGE Publications Ltd.

McCay-Peet, L., & Quan-Haase, A. (2017). What is social media and what questions can social media research help us answer? *The SAGE Handbook of Social Media Research Methods*, 13–26.

ORCID. (2020a). *About.* Retrieved from `https://orcid.org/`

ORCID. (2020b). *Structure of the ORCID identifier.* Retrieved from `https://support.orcid.org/hc/en-us/articles/360006897674-Structure-of-the-ORCID-Identifier`

ORCID. (2020c). *What is ORCID?* Retrieved from `https://support.orcid.org/hc/en-us/articles/360006973993-What-is-ORCID-`

O'Reilly, T. (2005). *What Is Web 2.0: Design patterns and business models for the next generation of software.* Retrieved from `https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html`

Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise: A bibliometric guide for researchers.* Cambridge, MA: Chandos.

Saracevic, T. (2009). Information science. In M. J. Bates (Ed.), *Encyclopedia of library and information sciences* (3rd ed., pp. 2570–2585). NY: Taylor and Francis.

Sloan, L., & Quan-Haase, A. (Eds.). (2017). *The SAGE handbook of social media research methods* (1st ed.). SAGE Publications Ltd.

Stock, W. G., & Stock, M. (2013). *Handbook of information science.* Berlin, Boston: De Gruyter Saur.

References

Aguillo, I. F. (2011). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics*, *91*, 343–351.

Ainslie, A. (2015). *The burden of protecting live sports telecasts: The real time problem of live streaming and app-based technology.* Los Angeles, CA: Southwestern Law School.

Aksnes, D. W., & Rip, A. (2009). Researchers' perceptions of citations. *Research Policy*, *38*(6), 895–905.

Alamiri, D., & Blustein, J. (2016). Privacy awareness and design for live video broadcasting apps. In C. Stephanidis (Ed.), *HCI 2016, Part I. CCIS* (Vol. 617, pp. 459–464). Cham: Springer.

Allem, J.-P., Escobedo, P., Chu, K. H., Cruz, T. B., & Unger, J. B. (2017). Images of little cigars and cigarillos on Instagram identified by the hashtag #Swisher: Thematic analysis. *Journal of Medical Internet Research*, *19*(7), e255.

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). H-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*(4), 273–289.

Amancio, D. R., Oliveira, O. N., & Costa, L. D. (2012). Three-feature model to reproduce the topology of citation networks and the effects from authors' visibility on their h-index. *Journal of Informetrics*, *6*(3), 427–434.

Anagnostopoulos, C., Parganas, P., Chadwick, S., & Fenton, A. (2018). Branding in pictures: Using Instagram as a brand management tool in professional team sport organisations. *European Sport Management Quarterly*, *18*(4), 413–438.

Andalibi, N., Ozturk, P., & Forte, A. (2017). Sensitive self-disclosures, responses, and social support on Instagram: The case of #depression. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1485–1500). New York: ACM.

Argamon, S., Whitelaw, C., Chase, P., Hota, S. R., Garg, N., & Levitan, S. (2007). Stylistic text classification using functional lexical features. *Journal of the American Society for Information Science and Technology*, *58*(6), 802–822.

Austin, J. L. (1962). *How to do things with words.* Oxford: Clarendon.

Aynsley, M. (2019). *The 2019 Instagram hashtag guide — How to use them and get results.* Retrieved from `https://blog.hootsuite.com/instagram-hashtags/`

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386.

Bar-Ilan, J. (2008). Which h-index? – A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, *74*(2), 257–271.

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social web. In *Proceedings of the 17th International Conference on Science and Technology Indicators* (pp. 98–109). Montreal: Science-Metrix et OST.

Beaudoin, J. (2007). Folksonomies: Flickr image tagging: Patterns made visible. *Bulletin of the American Society for Information Science and Technology*, *34*(1), 7–11.

Bilal, K., Erbad, A., & Hefeeda, M. (2017). Crowdsourced multi-view live video streaming using cloud computing. *IEEE Access*, *5*, 12635–12647.

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, *1*(1), 363–375.

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, *55*(14), 1216–1227.

Blackert, L., & Siegel, S. (1979). Ist in der wissenschaftlich-technischen Information Platz für die Informetrie? *Wissenschaftliche Zeitschrift der TH Ilmenau*, *25*(6), 187–199.

Borko, H. (1968). Information science: What is it? *American Documentation*, *19*(1), 3–5.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, *59*(5), 830–837.

Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H.-D. (2011). A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics*, *5*(3), 346–359.

Boyd, D. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. *Networked Self: Identity, Community, and Culture on Social Network Sites*, 39–58.

Boyd, D. (2014). *It's complicated: The social lives of networked teens*. New Heaven: Yale University Press.

Bretschneider, S., Marc-Aurele, F. J., & Wu, J. (2005). 'Best practices' research: A methodological guide for the perplexed. *Journal of Public Administration Research and Theory*, *15*(2), 307–323.

Brown, R. C., Fischer, T., Goldwich, A. D., Keller, F., Young, R., & Plener, P. L. (2018). #cutting: Nonsuicidal self-injury (NSSI) on Instagram. *Psychological Medicine*, *48*(2), 337–346.

Bründl, S., & Hess, T. (2016). Why do users broadcast? Examining individual motives and social capital on social live streaming platforms. In *Pacific Asia Conference on Information Systems Proceedings, PACIS* (Vol. 332).

Buarki, H., & Alkhateeb, B. (2018). Use of hashtags to retrieve information on the web. *The Electronic Library*, *36*(2), 286–304.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, *42*(5), 351–360.

Buela-Casal, G., & Zych, I. (2012). What do the scientists think about the impact factor? *Scientometrics*, *92*, 281–292.

BYOApps LLC. (2019). *Tagstagram - Hashtag generator.* Retrieved from `https://apps.apple.com/us/app/tagstagram-hashtag-generator/id587756737`

Cambridge University Press. (2019). *Best practice.* Retrieved from `https://dictionary.cambridge.org/dictionary/english/best-practice`

Carnap, R. (1931). Die Überwindung der Metaphysik durch logische Analysen der Sprache. *Erkenntnis*, *2*(1), 219–241.

Carrotte, E. R., Prichard, I., & Lim, M. S. C. (2017). "Fitspiration" on social media: A content analysis of gendered images. *Journal of Medical Internet Research*, *19*(3), e95.

Chase, J. M. (1970). Normative criteria for scientific publication. *The American Sociologist*, *5*(3), 262–265.

Chen, C. M.-L., & Lin, W.-Y. C. (2018). What indicators matter? The analysis of perception toward research assessment indicators and Leiden Manifesto. The case study of Taiwan. In *Proceedings of the 23rd International Conference on Science and Technology Indicators (STI 2018)* (pp. 688–698). Leiden, NL: Centre for Science and Technology Studies (CWTS).

Chen, X. (2010). Google scholar's dramatic coverage improvement five years after debut. *Serials Review*, *36*(4), 221–226.

Coelho, R. L. F., de Oliveira, D. S., & de Almeida, M. I. S. (2016). Does social media matter for post typology? Impact of post content on Facebook and Instagram metrics. *Online Information Review*, *40*(4), 458–471.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.

Cole, S., & Cole, J. R. (1968). Visibility and the structural bases of awareness of scientific research. *American Sociological Review*, *33*(3), 397–413.

Cronin, B. (1994). Tiered citation and measures of document similarity. *Journal of the American Society for Information Science*, *45*(7), 537–538.

Cronin, B. (2003). Scholarly communication and epistemic cultures. *The New Review of Academic Librarianship*, *9*(1), 1–24.

Cronin, B. (2008). On the epistemic significance of place. *Journal of the American Society for Information Science and Technology*, *59*(6), 1002–1006.

Cunha, E., Magno, G., Almeida, V., Gonçalves, M. A., & Benevenuto, F. (2012). A gender based study of tagging behavior in Twitter. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT 2012* (pp. 323–324). New York: ACM.

Daer, A. R., Hoffman, R., & Goodman, S. (2015). Rhetorical functions of hashtag forms across social media applications. *Communication Design Quarterly Review*, *3*(1), 12–16.

Dapper Goat Social Media. (2019). *Banned Instagram hashtags 2019.* Retrieved from https://www.dappergoat.com/blog/banned-instagram-hashtags-2019/

Dennis, B. M. (2006). Foragr: Collaborative tagged photographs and social information visualisation. In *Proceedings of the Collaborative Web Tagging Workshop at WWW 2006,* Edinburgh, UK.

Denton, E., Weston, J., Paluri, M., Bourdev, L., & Fergus, R. (2015). User conditional hashtag prediction for images. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1731–1740). New York: ACM.

Derrick, G. E., & Gillespie, J. (2013). "A number you just can't get away from": Characteristics of adoption and the social construction of metric use by researchers. In *Proceedings of the 18th International Conference on Science and Technology Indicators* (pp. 104–116). Berlin: Institute for Research Information and Quality Assurance.

Destatis. (2019). Bildung und Kultur. *Personal an Hochschulen (Fachserie 11, Reihe 4.4).* Wiesbaden: Statistisches Bundesamt.

Deutscher Hochschulverband. (2020). Universitäten Deutschland. *Hochschullehrer-Verzeichnis 2020,* Band 1. 28. Ausgabe. Berlin, New York: De Gruyter Saur.

Dev, G. (2019). *Tagify - Best hashtags for Instagram.* Retrieved from https://play.google.com/store/apps/details?id=com.gongadev.hashtagram&hl=de

Dickinson, P. (2013). "B/w U & me": The Functions of Formulaic Language in Interactional Discourse on Twitter. *The Linguistics Journal*, *7*(1), 7–38.

Dilthey, W. (1895). Ideen über eine beschreibende und zergliedernde Psychologie. *Sitzungsberichte der königlich preussischen Akademie der Wissenschaften zu Berlin, 7. Juni 1894, Ausgabe XXVI, Sitzung der philosophisch historischen Classe*, 1–88.

DORA. (2020). *About.* Retrieved from https://sfdora.org/

Dorsch, I. (2017). Relative visibility of authors' publications in different information services. *Scientometrics*, *112*(2), 917–925.

Dorsch, I. (2018). Content description on a mobile image sharing service: Hashtags on Instagram. *Journal of Information Science Theory and Practice*, *6*(2), 46–61.

Dorsch, I., Askeridis, J., & Stock, W. G. (2018). Truebounded, overbounded, or underbounded? Scientists' personal publication lists versus lists generated through bibliographic information services. *Publications*, *6*(1), 1–9.

Dorsch, I., & Frommelius, N. (2015). A scientometric approach to determine and analyze productivity, impact and topics based upon personal publication lists. In F. Pehar, C. Schlögl, & C. Wolff (Eds.), *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th-21st May 2015* (pp. 578–580). Glückstadt, Germany: Hülsbusch. (Available online: https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/heck/Bella_Nils.pdf)

Dorsch, I., Schlögl, C., Stock, W. G., & Rauch, W. (2017). Forschungsthemen der

Düsseldorfer und Grazer Informationswissenschaft (2010 bis 2016). *Information - Wissenschaft und Praxis*, *68*(5-6), 320–328.
Dorsch, I., Zimmer, F., & Stock, W. G. (2017). Image indexing through hashtags in Instagram. *Proceedings of the Association for Information Science and Technology*, *54*(1), 658–659.
Dowell, C. T., & Duncan, D. F. (2016). Periscoping economics through someone else's eyes: A real world (Twitter) app. *International Review of Economics Education*, *23*, 34–39.
Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., & Tomkins, A. (2007). Visualizing tags over time. *ACM Transactions on the Web*, *1*, 1–22.
Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115.
Enser, P. G. B. (2008). Visual image retrieval. *Annual Review of Information Science and Technology*, *42*(1), 1–42.
Erz, A., Marder, B., & Osadchaya, E. (2018). Hashtags: Motivational drivers, their use, and differences between influencers and followers. *Computers in Human Behavior*, *89*(July), 48–60.
Faklaris, C., Cafaro, F., Hook, S. A., Blevins, A., O'Haver, M., & Singhal, N. (2016). Legal and ethical implications of mobile live-streaming video apps. In *MobileHCI 2016* (pp. 722–729). New York: ACM.
Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of Pubmed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, *22*(2), 338–342.
Ferrara, E., Interdonato, R., & Tagarelli, A. (2014). Online popularity and topical interests through the lens of Instagram. In *The Proceedings of the 25th ACM Conference on Hypertext and Social Media* (pp. 24–34). Santiago, Chile: ACM.
Fietkiewicz, K. (2017). Jumping the digital divide: How do 'silver surfers' and 'digital immigrants' use social media? *Networking Knowledge*, *10*(1), 5–26.
Fietkiewicz, K. J., Lins, E., Baran, K. S., & Stock, W. G. (2016). Inter-generational comparison of social media use: Investigating the online behavior of different generational cohorts. In *Proceedings of the 49th Hawaii International Conference on System Sciences* (pp. 3829–3838). Washington, DC: IEEE.
Fietkiewicz, K. J., & Scheibe, K. (2017). Good morning... good afternoon, good evening and good night: Adoption, usage and impact of the live streaming platform YouNow. *Proceedings of the 3rd International Conference on Library and Information Science*, 92–115.
Fitzgerald, R. T., & Radmanesh, A. (2015). Social media and research visibility. *American Journal of Neuroradiology*, *36*(4), 637.
Friedländer, M. B. (2014). Informationswissenschaft an deutschsprachigen Universitäten - eine komparative informetrische Analyse. *Information - Wissenschaft und Praxis*, *65*(2), 109–119.
Friedländer, M. B. (2017a). And action! Live in front of the camera: An evaluation of the social live streaming service YouNow. *International Journal of Information*

Communication Technologies and Human Development, *9*(1), 15–33.
Friedländer, M. B. (2017b). Streamer motives and user-generated content on social live-streaming services. *Journal of Information Science Theory and Practice*, *5*(1), 65–84.
Gaillard, J. (1992). Use of publication lists to study scientific production and strategies of scientists in developing countries. *Scientometrics*, *23*(1), 57–73.
Garfield, E. (1955). Citation indexes for science through association of ideas. *Science*, *122*(3159), 108–111.
Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*(4060), 471–479.
Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, *295*(1), 90–93. Retrieved from http://www.garfield.library.upenn.edu/papers/eval{_}of{_}science{_}CBE(Utah).html
Geurin-Eagleman, A. N., & Burch, L. M. (2016). Communicating via photographs: A gendered analysis of olympic athletes' visual self-presentation on Instagram. *Sport Management Review*, *19*(2), 133–145.
Giannoulakis, S., & Tsapatsoulis, N. (2015). Instagram hashtags as image annotation metadata. In R. Chbeir, Y. Manolopoulos, I. Maglogiannis, & R. Ralhajj (Eds.), *Artificial Intelligence Applications and Innovations IFIP Advances in Information and Communication Technology* (Vol. 458, pp. 206–220). Cham: Springer.
Giannoulakis, S., & Tsapatsoulis, N. (2016). Evaluating the descriptive power of Instagram hashtags. *Journal of Innovation in Digital Ecosystems*, *3*(2), 114–129.
Gibbs, M., Meese, J., Arnold, M., Nansen, B., & Carter, M. (2015). #Funeral and Instagram: Death, social media, and platform vernacular. *Information Communication & Society*, *18*(3), 255–268.
Gilbert, S. (2019). *New rules for Instagram hashtags: How to find the right hashtags for your business.* Retrieved from https://later.com/blog/instagram-hashtags/
Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208.
Gordon, A. (2012). The invisibility of science publications in hebrew: A comparative database study. *Journal of the American Society for Information Science and Technology*, *63*(3), 607–615.
Gray, J. (1992). *Men are from Mars, women are from Venus.* New York: HarperCollins.
Gros, D., Wanner, B., Hackenholt, A., Zawadzki, P., & Knautz, K. (2017). World of streaming. Motivation and gratification on Twitch. In *SCSM 2017, Part I. LNCS* (Vol. 10282, pp. 44–57). Cham: Springer.
Gross, C., Jungbauer-Gans, M., & Kriwy, P. (2008). Die Bedeutung meritokratischer und sozialer Kriterien für wissenschaftliche Karrieren – Ergebnisse von Expertengesprächen in ausgewählten Disziplinen. *Beiträge zur Hochschul-*

forschung, *30*(4), 8–32.

Guidry, J. D., Messner, M., Jin, Y., & Medina-Messner, V. (2015). From #mcdonaldsfail to #dominossucks: An analysis of Instagram images about the 10 largest fast food companies. *Corporate Communications: An International Journal*, *20*(3), 344–359.

Gust von Loh, S., & Stock, W. G. (2008). Wissensrepräsentation - Information Retrieval - Wissensmanagement - Das Forschungsprogramm der Düsseldorfer Informationswissenschaft. *Information - Wissenschaft und Praxis*, *59*(2), 73–74. (Available online: `https://www.philfak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/1204543980editorial_.pdf`)

Haddow, G., & Hammarfelt, B. (2019). Quality, impact, and quantification: Indicators and metrics use by social scientists. *Journal of the Association for Information Science and Technology*, *70*(1), 16–26.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items.* New York: Routledge.

Halavais, A. (2014). Structure of Twitter: Social and technical. In K. Weller, A. Bruns, J. Burgess, N. Mahrt, & C. Puschmann (Eds.), *Twitter and society* (pp. 29–41). New York: Peter Lang.

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation. Review of the literature. *Journal of Informetrics*, *11*(3), 823–834.

Haller, R. (1989). Wandlung der Wissenschaftsauffassung. In E. Busek, W. Mantl, & M. Peterlik (Eds.), *Wissenschaft und Freiheit. Ideen zur Universität und Universalität* (pp. 46–58). München, Germany: Oldenbourg.

Hammarfelt, B., & Haddow, G. (2018). Conflicting measures and values: How humanities scholars in Australia and Sweden use and react to bibliometric indicators. *Journal of the Association for Information Science and Technology*, *69*(7), 924–935.

Harzing, A.-W. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics*, *98*(1), 565–575.

Harzing, A. W. (2016). *Do Google Scholar, Scopus and the Web of Science speak your language?* Retrieved from `https://harzing.com/publications/white-papers/do-google-scholar-scopus-and-theweb-of-science-speak-your-language`

Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics*, *108*(1), 413–423.

Haustein, S. (2018a). *Metrics literacy [Blog post].* Retrieved from `https://stefaniehaustein.com/metrics-literacy/`

Haustein, S. (2018b). *Metrics Literacy: Developing online material to educate researchers and research administrators about bibliometrics, usage metric, altmetrics, and technometrics.* Retrieved from `http://stefaniehaustein.com/metrics-literacy/{\%}0A`

Haustein, S., & Larivière, V. (2015). The use of bibliometrics for assessing research:

Possibilities, limitations and adverse effects. In I. Welpe, J. Wollersheim, S. Ringelhan, & M. Osterloh (Eds.), *Incentives and Performance: Governance of Research Organizations* (pp. 121–139). Cham, CH: Springer.

Hearn, A. (2008). 'Meat, mask, burden': Probing the contours of the branded 'self'. *Journal of Consumer Culture*, *8*(2), 197–217.

Hearn, A., & Schoenhoff, S. (2016). From celebrity to influencer: Tracing the diffusion of celebrity value across the data stream. In P. Marshall & S. Redmond (Eds.), *A companion to celebrity* (pp. 194–211). Hoboken: John Wiley & Sons.

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, *1*(1), 387–395.

Heyd, T., & Puschmann, C. (2017). Hashtagging and functional shift: Adaptation and appropriation of the #. *Journal of Pragmatics*, *116*, 51–63.

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., ... Stock, W. G. (2015). Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. *Online Information Review*, *39*(2), 255–264.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Hirsch, J. E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19193–19198.

Hirsch, J. E. (2020). Superconductivity, What the h? The emperor has no clothes. *Physics and Society*, *49*(1), 4–9.

Holl, A., Makara, G., Micsik, A., & Kovacs, L. (2014). MTMT: The Hungarian scientific bibliography. In *Uses of open data within government for innovation and efficiency, 2014.06.30–2014.07.01* (p. 5). Samos.

Holland, G., & Tiggemann, M. (2017). Strong beats skinny every time": Disordered eating and compulsive exercise in women who post fitspiration on Instagram. *International Journal of Eating Disorders*, *50*(1), 76–79.

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, *1*(1), 21–48.

Holmberg, C., Chaplin, J. E., Hillman, T., & Berg, C. (2016). Adolescents' presentation of food in social media: An explorative study. *Appetite*, *99*, 121–129.

Holmberg, K., & Hellsten, I. (2015). Gender differences in the climate change communication on Twitter. *Internet Research*, *25*, 811–828.

Honka, A., Frommelius, N., Mehlem, A., Tolles, J. N., & Fietkiewicz, K. J. (2015). How safe is YouNow? An empirical study on possible law infringements in Germany and the United States. *The Journal of MacroTrends in Social Science*, *1*(1), 1–17.

Honka, A., Orszullok, L., Dorsch, I., & Frommelius, N. (2015). Topical impact analysis. A new informetric indicator for the assessment of a scientific institution. In *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th-21st May 2015* (pp. 410–421). (Available online: https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Informationswissenschaft/heck/Axel_et_al.pdf)

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*(2), 291–314.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observation an intimacy at a distance. *Psychology*, *19*(2), 215–229.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.

Hu, G. Y., Wang, L., Ni, R., & Liu, W. S. (2020). Which h-index? An exploration within the Web of Science. *Scientometrics*, *123*, 1225–1233.

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. In *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014* (pp. 595–598). Palo Alto, CA: AAAI Press.

Hughes, E. (2017). Personal publications lists serve as a reliable calibration parameter to compare coverage in academic citation databases with scientific social media. *Evidence Based Library and Information Practice*, *12*(1), 125–127.

Ingleton, P. (2014). Celebrity seeking micro-celebrity: 'New candour' and the everyday in the sad conversation. *Celebrity Studies*, *5*(4), 525–527.

Ingwersen, P. (2000). The international visibility and citation impact of Scandinavian research articles in selected social science fields: The decay of a myth. *Scientometrics*, *49*(1), 39–61.

Ingwersen, P. (2002). Cognitive perspectives of document representation. In H. Bruce, R. Fidel, P. Ingwersen, & V. P. (Eds.), *Proceedings of the 4th International Conference on Conceptions of Library and Information Science* (pp. 285–300). Greenwood Village, CO: Libraries Unlimited.

Instagram. (2010). *Instagram launches.* Retrieved from https://instagram-press.com/blog/2010/10/06/instagram-launches-2/

Instagram. (2012a). *Instagram + Facebook (2012).* Retrieved from https://instagram-press.com/blog/2012/04/09/

Instagram. (2012b). *Instagram for android.* Retrieved from https://instagram-press.com/blog/2012/04/03/instagram-for-android-available-now/

Instagram. (2018). *Instagram statistics.* Retrieved from https://instagram-press.com/our-story

Instagram. (2019a). *Can we help?* Retrieved from https://www.instagram.com/explore/tags/selfharm/?hl=en

Instagram. (2019b). *Instagram statistics.* Retrieved from https://instagram-press.com/our-story/

International Urban Research. (1959). *The World's Metropolitan Areas.* Berkeley, CA, USA; Los Angeles, CA, USA: University of California Press. (Available

online: https://books.google.de/books?id=p0-6Ej_Up3AC))
Jacsó, P. (2008). The plausibility of computing the h-index of scholarly productivity and impact using reference-enhanced databases. *Online Information Review*, *32*(2), 266–283.
Jan, R., & Ahmad, R. (2020). H-index and its variants: Which variant fairly assess author's achievements. *Journal of Information Technology Research*, *13*(1), 68–76.
Jappe, A. (2020). Professional standards in bibliometric research evaluation? A metaevaluation of European assessment practice 2005 – 2019. *PLoSONE*, *15*(4), e0231735, 1–233.
Jones, R., MacGillivray, M., Murray-Rost, P., Pitman, J., Sefton, P., O'Steen, B., & Waites, W. (2011). *Open bibliography for science, technology, and medicine.* Cambridge: University of Cambridge.
Jörgensen, C. (2002). Image access, the semantic gap, and social tagging as a paradigm shift. In J. Lussky (Ed.), *Proceedings of the 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research.* Milwaukee, WI.
Jörgensen, C. (2003). *Image retrieval: Theory and research.* Lanham, MD: Scarecrow Press.
Kagan, J. (2009). *The Three Cultures. Natural sciences, social sciences, and the humanities in the 21st century.* Cambridge, MA: Cambridge University Press.
Kane, B. (2010). Balancing anonymity, polularity & micro-celebrity: The Crossroads of social networking and privacy. *Albany Law Journal of Science and Technology*, *20*(2), 327–363.
Kaser, R. (2017). *How to make money livestreaming.* Retrieved from https://thenextweb.com
Katz, E., Blumler, J. G., & Gurevitch, M. (1974). Utilization of mass communication by the individual. In *The Uses of Mass Communications: Current Perspectives on Gratifications Research* (pp. 19–33). Beverly Hills: Sage Publications.
Kavka, M. (2012). *Reality tv.* Edinburgh: Edinburgh University Press.
Kavka, M. (2015). Celevision: Mobilizations of the television screen. In P. D. Marshall & S. Redmond (Eds.), *A Companion to Celebrity* (pp. 295–314). Edinburgh: Edinburgh University Press.
Keinänen, K. (2017). *The role of live streaming in marketing communications and corporate branding* (master's thesis). Lappeenranta University of Technology, LUT School of Business and Management, Lappeenranta, Finland.
Khamis, S., Ang, L., & Welling, R. (2017). Self-branding, 'micro-celebrity' and the rise of social media influencers. *Celebrity Studies*, *8*(2), 191–208.
Kipp, M. E. I. (2006). @toread and cool: Tagging for time, task, and emotion. *Proceedings of the 17th Annual ASIS&T SIG/CR Classification Research Workshop*.
Kipp, M. E. I., & Campbell, D. G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. *Proceedings of the ASIST Annual Meeting*, *43*.

Kirkwood, P. E. (2012). Researcher publication checklist: A quantitative method to compare traditional databases to web search engines. In *119th American Society for Engineering Education Annual Conference and Exhibition, San Antonio, TX, June 10-13, 2012.* (Available online: `https://peer.asee.org/facultypublication-checklists-a-quantitative-method-to-compare-traditional-databases-to-web-search-engines`)

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, *1*(1), 1–12.

Kivran-Swaine, F., Brody, S., Diakopoulos, N., & Naaman, M. (2012). Of joy and gender: Emotional expression in online social networks. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (pp. 139–142). New York: ACM.

Knautz, K. (2012). Emotion felt and depicted. consequences for multimedia retrieval. *Indexing and Retrieval of Non-Text Information*, 343–375.

Kowalski, C. J., & Mrdjenovich, A. J. (2016). Beware dichotomies. *Perspectives in Biology and Medicine*, *59*(4), 517–535.

Kraker, P., & Lex, E. (2015). A critical look at the ResearchGate score as a measure of scientific reputation. In *Quantifying and Analysing Scholarly Communication on the Web (ASCW'15).* Oxford, 30 June 2015 (3 S.).

Kremerskothen, K. (2012). *An amazing 8 years.* Retrieved from `http://blog.flickr.net/en/2012/02/10/an-amazing-8-years/`

Kretschmer, H., & Aguillo, I. F. (2004). Visibility of collaboration on the web. *Scientometrics*, *61*(3), 405–426.

Krikwood, P. E. (2012). Researcher publication checklist: A quantitative method to compare traditional databases to web search engines. In *119th american society for engineering education annual conference and exhibition.* San Antonio, TX.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks: Sage Publications Inc.

Krippendorff, K. (2011). *Computing Krippendorff's alpha reliability. Departmental papers (ASC) (2011).* Retrieved from `https://repository.upenn.edu/asc_papers/43`

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology* (3rd ed.). Thousand Oaks: Sage.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621.

Laerd Statistics. (2018). *Chi-square test for independence using SPSS Statistics.* Retrieved from `https://statistics.laerd.com/`

Laestadius, L. (2017). Instagram. In *The SAGE Handbook of Social Media Research Methods.* SAGE Publications Ltd.

Lai, L. S. L., & To, W. M. (2015). Content analysis of social media: A grounded theory approach. *Journal of Electronic Commerce Research*, *16*(2), 138–152.

Lair, D. J., Sullivan, K., & Cheney, G. (2005). Marketization and the recasting of the professional self: The rhetoric and ethics of personal branding. *Management Communication Quarterly*, *18*(3), 307–343.

Lancaster, F. W. (2003). *Indexing and abstracting in theory and practice*. Champaign, IL: University of Illinois, Graduate School of Library and Information Science.

Laniado, D., Eynard, D., & Colombetti, M. (2007). Using WordNet to turn a folksonomy into a hierarchy of concepts. In G. Semeraro, C. Di Sciascio, H. Morbidoni, & H. Stoermer (Eds.), *Semantic Web Application and Perspectives: Proceedings of the 4th Italian Semantic Web Workshop (SWAP 2007). CEUR Workshop Proceedings* (Vol. 314, pp. 192–201). Bari: CEUR Workshop Proceedings.

Lavoie, K. A. (2015). Instagram and branding: A case study of Dunkin' Donuts. *Elon Journal of Undergraduate Research in Communications*, *6*(2), 79–90.

Lee, C., & Chau, D. (2018). Language as pride, love, and hate: Archiving emotions through multilingual Instagram hashtags. *Discourse Context Media*, *22*(0), 21–29.

Lee, E., Lee, J.-A., Moon, J. H., & Sung, Y. (2015). Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, Behavior, and Social Networking*, *18*(9), 552–556.

Lemke, S., Mehrazar, M., Mazarakis, A., & Peters, I. (2019). "When you use social media you are not working": Barriers for the use of metrics in social sciences. *Frontiers in Research Metrics and Analytics*, *3*(39), 1–18.

Leung, L. (2013). Generational differences in content generation in social media: The roles of the gratifications sought and of narcissism. *Computers in Human Behavior*, *29*(3), 997–1006.

Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, *43*(1), 5–25.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 5–55.

Linde, F., & Stock, W. G. (2011). *Information markets. A strategic guideline for the i-commerce*. Berlin, New York: De Gruyter Saur.

Lup, K., Trub, L., & Rosenthal, L. (2015). Instagram #Instasad? Exploring associations among Instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking*, *18*, 247–252.

Ma, L., & Ladisch, M. (2019). Evaluation complacency or evaluation inertia? A study of evaluative metrics and research practices in Irish universities. *Research Evaluation*, *28*(3), 209–217.

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (2009). Codebook development for team-based qualitative analysis. In *The content analysis reader* (pp. 211–219).

Mann, H., & Whitney, D. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*(1), 50–60.

Marcus, S.-R. (2016). Thinspiration vs. thicksperation: Comparing pro-anorexic and fat acceptance image posts on a photo-sharing site. *Cyberpsychology*, *10*(2).

Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, *28*(1), 38–43.

Martin, R. (2018). *Cult of influence. Marketing: Advertising, media and pr in Canada.* Retrieved from `http://www.marketingmag.ca`

Martin-Martin, A., Orduna-Malea, E., Thelwall, M., & Lopez-Cozar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177.

Marwick, A. E. (2013). *Status update: Celebrity, publicity, and branding in the social media age.* New Haven: Yale University Press.

Marwick, A. E. (2015). Instafame: Luxury selfies in the attention economy. *Public Culture*, *27*(1), 137–350.

Marwick, A. E. (2016). You may know me from Youtube: (Micro-)Celebrities in social media. In P. D. Marshall & S. Redmond (Eds.), *A Companion to Celebrity* (pp. 194–212). Hoboken: John Wiley & Sons.

Marwick, A. E., & Boyd, D. (2010). I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*, *13*(12), 114–133.

Marwick, A. E., & Boyd, D. (2011). To see and be seen: Celebrity practice on Twitter. *Convergence*, *17*(2), 139–158.

Maugeri, R., Giammalva, R. G., & Iacopino, D. G. (2016). On the shoulders of giants, with a smartphone: Periscope in neurosurgery. *World Neurosurgery*, *92*, 569–570.

McCay-Peet, L., & Quan-Haase, A. (2017). What is social media and what questions can social media research help us answer? *The SAGE Handbook of Social Media Research Methods*, 13–26.

Meschede, C., & Siebenlist, T. (2018). Cross-metric compatibility and inconsistencies of altmetrics. *Scientometrics*, *115*, 283–297.

Messina, C. (2007). *How do you feel about using # (pound) for groups. As in #barcamp [msg]? chrismessina (Twitter).* Retrieved from `https://twitter.com/chrismessina/status/223115412`

Metcalfe, H., Beesley, A., Watson, C., Sarwari, H., & Clayton, J. (2018). #SpeakingTheirLanguage: Engaging with young people's mental health in a digital landscape. *Clinical Psychology Forum*(307), 9–14.

Miguel, S., Chinchilla-Rodriguez, Z., & De Moya-Anegõn, F. (2011). Open access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the American Society for Information Science and Technology*, *62*(6), 1130–1145.

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, *256*(1), 1–19.

Mohammad, S. M., & Kiritchenko, S. (2015). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, *31*(2), 301–326.

Mohiuddin, Z. (n.d.). *Products.* Retrieved from `https://www.zaheer.me/`

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228.

Moreno, M. A., Ton, A., Selkie, E., & Evans, Y. (2016). Secret Society 123: Understanding the language of self-harm on Instagram. *Journal of Adolescent Health*, *58*(1), 78–84.

Muntean, N., & Petersen, A. H. (2009). Celebrity twitter: Strategies of intrusion and disclosure in the age of technoculture. *M/C Journal*, *12*(5).

Nacke, O. (1979). Informetrie. Ein neuer Name für eine neue Disziplin. *Nachrichten für Dokumentation*, *30*(6), 219–226.

Nalimov, V. V., & Mulchenko, Z. M. (1969). *Scientometrics, the study of science as an information process [english translation. note: Original in russian].* Moscow: Nauka.

Nashmi, E. A. (2018). From Selfies to Media Events: How Instagram users interrupted their routines after the Charlie Hebdo shootings. *Digital Journalism*, *6*(1), 98–117.

Nov, O., Naaman, M., & Ye, C. (2008). What drives content tagging: The case of photos on flickr. In *Conference on Human Factors in Computing Systems - Proceedings* (pp. 1097–1100).

Oh, C., Lee, T., Kim, Y., Park, S., & Suh, B. (2016). Understanding participatory hashtag practices on Instagram: A case study of Weekend Hashtag Project. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1280–1287). New York: ACM.

ORCID. (2020a). *About.* Retrieved from `https://orcid.org/`

ORCID. (2020b). *Structure of the ORCID identifier.* Retrieved from `https://support.orcid.org/hc/en-us/articles/360006897674-Structure-of-the-ORCID-Identifier`

ORCID. (2020c). *What is ORCID?* Retrieved from `https://support.orcid.org/hc/en-us/articles/360006973993-What-is-ORCID-`

O'Reilly, T. (2005). *What Is Web 2.0: Design patterns and business models for the next generation of software.* Retrieved from `https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html`

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological review*, *97*(3), 315–331.

Otlet, P. (1934). *Traite de documentation, le livre sur le livre.* Brussels: D. Van Keerberghen & Sons.

Oxford University Press. (2019). *Best practice.* Retrieved from `https://en.oxforddictionaries.com/definition/best{\_}practice`

Page, R. (2012). The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse and Communication*, *6*(2), 181–201.

Panofsky, E. (1955). *Meaning in the visual arts.* Garden City: Doubleday Anchor Books.

Papacharissi, Z. (2015). We have always been social. *Social Media and Society*, *1*(1).

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London,* *Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *Series 5, 50*(302), 157–175.

Pegoraro, A., Comeau, G. S., & Frederick, E. L. (2018). #SheBelieves: the use of Instagram to frame the US women's soccer team during #FIFAWWC. *Sport in Society*, *21*(7), 1063–1077.

Pennington, D. R., & Spiteri, L. (Eds.). (2019). *Social tagging in a linked data environment.* London, UK: Facet Publishing.

Peters, I. (2009). *Folksonomies. Indexing and retrieval in web 2.0.* Berlin: De Gruyter Saur.

Peters, I., & Stock, W. G. (2007). Folksonomy and information retrieval. In *Joining Research and Practice: Social Computing and Information Science. Proceedings of the 70th ASIS&T Annual Meeting* (Vol. 44, pp. 1510–1542). Hoboken: Wiley.

Peters, M. E. I., I.and Kipp, Heck, T., Gwizdka, J., Lu, K., Neal, D. R., & Spiteri, L. (2017). Social tagging & folksonomies: Indexing, retrieving… and beyond? In *Bridging the Gulf: Communication and Information in Society, Technology, and Work. Proceedings of the 74th ASIS&T Annual Meeting* (pp. 1–4). Wiley, Hoboken.

Philipps, J., & Dorsch, I. (2019). Gender-specific tagging of images on Instagram. In *Social Computing and Social Media. Design, Human Behavior and Analytics. HCII 2019. Lecture Notes in Computer Science, Vol. 11578.* (pp. 396–413). Springer.

Pila, E., Mond, J. M., Griffiths, S., Mitchison, D., & Murray, S. B. (2017). A thematic content analysis of #cheatmeal images on social media: Characterizing an emerging dietary trend. *International Journal of Eating Disorders*, *50*(6), 698–706.

Pluzhenskaia, M. (2006). Folksonomies or fauxsonomies: How social is social bookmarking? In *Advances in Classification Research Online* (Vol. 17, pp. 23–24).

Popper, K. R. (1976). *Logik der Forschung* (6th ed.). Tübingen, Germany: JCB Mohr (Paul Siebeck).

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, *25*(4), 348–349.

Quan-Haase, A., & Sloan, L. (2017). Introduction to the handbook of social media research methods: Goals, challenges and innovations. In *The sage handbook of social media research methods* (pp. 1–9).

Rasmussen, E. M. (1997). Indexing images. In *Annual Review of Information Science and Technology* (Vol. 32, pp. 169–196).

Rassenhövel, S. (2010). *Performancemessung im Hochschulbereich. Theoretische Grundlagen und empirische Befunde.* Wiesbaden: Gabler.

Rauschnabel, P. A., Sheldon, P., & Herzfeldt, E. (2019). What motivates users to hashtag on social media? *Psychology and Marketing*, *36*(5), 473–488.

Recktenwald, D. (2017). Toward a transcription and analysis of live streaming on Twitch. *Journal of Pragmatics*, *115*, 68–81.

Rodríguez-Gil, L., García-Zubia, J., Orduña, P., & López-Ipiña, D. (2017). An open and scalable web-based interactive live-streaming architecture: The WILSP platform. *IEEE Access*, *5*, 9842–9856.

Rondeau, S. (2014). The life and times of aboutness: A review of the library and information science literature. *Evidence Based Library and Information Practice*, *9*(1), 14–35.

Rorissa, A. (2010). A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology*, *61*(11), 2230–2242.

Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise: A bibliometric guide for researchers.* Cambridge, MA: Chandos.

Rousseau, S., & Rousseau, R. (2015). Metric-wiseness. *Journal of the American Society for Information Science and Technology*, *66*(11), 2389.

Rousseau, S., & Rousseau, R. (2017). Being metric-wise: Heterogeneity in bibliometric knowledge. *El Profesional de la Información*, *26*(3), 480–487.

Santarossa, S., Coyne, P., Lisinski, C., & Woodruff, S. J. (2016). #fitspo on Instagram: A mixedmethods approach using Netlytic and photo analysis, uncovering the online discussion and author/image characteristics. *Journal of Health Psychology*, *24*(3), 376–385.

Santarossa, S., Coyne, P., & Woodruff, S. J. (2017). Exploring #nofilter images when a filter has been used. *International Journal of Virtual Communities and Social Networking*, *9*, 54–63.

Saracevic, T. (2009). Information science. In M. J. Bates (Ed.), *Encyclopedia of library and information sciences* (3rd ed., pp. 2570–2585). NY: Taylor and Francis.

Schaefer, M. (2012). *Return on influence: The revolutionary power of klout, social scoring, and influence marketing.* New York: McGraw Hill.

Scheibe, K., Fietkiewicz, K., & Stock, W. G. (2016). Information behavior on social live streaming services. *Journal of Information Science Theory and Practice*, *4*(2), 6–20.

Schlögl, C. (2013a). Internationale Sichtbarkeit der europäischen und insbesondere der deutschsprachigen Informationswissenschaft. *Information - Wissenschaft und Praxis*, *64*(1), 1–8.

Schlögl, C. (2013b). International visibility of European and in particular German-language publications in library and information science. In H.-C. Hobohm (Ed.), *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Tagungsband des 13. internationalen Symposiums für Informationswissenschaft (ISI 2013)* (pp. 51–62). Glückstadt: Hülsbusch. (Available online: `https://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/396`)

Schmidbauer, H., Rösch, A., & Stieler, F. (2018). The 2016 US presidential election and media on Instagram: who was in the lead? *Computers in Human Behavior*, *81*, 148–160.

Schmidt, S., & Stock, W. G. (2009). Collective indexing of emotions in images. A study in emotional information retrieval. *Journal of the American Society for*

Information Science and Technology, *60*(5), 863–876.
Scientific committee. (2013). In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), *Proceedings of the ISSI 2013 Vienna: 14th International Society of Scientometrics and Informetrics Conference, Vienne, Austria, 15–20 July 2013.* Vienna, Austria: Austrian Institute of Technology.
Senft, T. (2008). *Camgirls: Celebrity and community in the age of social networks.* New York: Peter Lang.
Senft, T. M. (2013). Microcelebrity and the branded self. In A. Burgess J. E.and Bruns (Ed.), *A Companion to New Media Dynamics* (pp. 346–354). Malden: Blackwell.
Shapiro, F. R. (1995). Brief communication: Coinage of the term information science. *Journal of the American Society for Information Science*, *46*(5), 321–397.
Shapp, A. (2015). Gender variation in the pragmatic uses of Twitter hashtags. In *Poster session presented at the 89th Annual Meeting of the Linguistic Society of America, Portland, Oregon.* Retrieved from `https://www.nyu.edu/projects/shapp/Shapp_LSA2015.html`
Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging and Classification Quarterly*, *6*(3), 39–62.
Shatford Layne, S. (1994). Some issues in the indexing of images. *Journal of the American Society for Information Science*, *45*(8), 583–588.
Shatford Layne, S. (2002). Subject access to art images. In *Introduction to art image access: Issues, tools, standarts, strategies* (pp. 1–20). Los Angeles, CA: Getty Research Institute.
Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, *58*, 89–97.
Sheldon, P., Herzfeldt, E., & Rauschnabel, P. A. (2019). Culture and social media: The relationship between cultural values and hashtagging styles. *Behaviour and Information Technology*(in press), 1–13.
Sheldon, P., Rauschnabel, P. A., Antony, M. G., & Car, S. (2017). A cross-cultural comparison of Croatian and American social network sites: Exploring cultural differences in motives for Instagram use. *Computers in Human Behavior*, *75*, 643–651.
Siebenlist, T. (2013). Emotionale Suche – Emotionales Information-Retrieval. In D. Lewandowski (Ed.), *Handbuch der Internet-Suchmaschinen 3: Suchmaschinen zwischen Technik und Gesellschaft* (pp. 299–327). Heidelberg: Akademische Verlagsgesellschaft.
Silipigni Connaway, L., & Radford, M. L. (2016). *Research methods in library and information science* (6th ed.). Santa Barbara, CA: ABC-CLIO.
Sloan, L., & Quan-Haase, A. (Eds.). (2017). *The SAGE handbook of social media research methods* (1st ed.). SAGE Publications Ltd.
Smith, A., & Anderson, M. (2018). *Social media use in 2018.* Retrieved from `http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/`
Smith, G. (2004). *Folksonomy: Social classification.* Retrieved from `http://atomiq`

.org/\%0Aarchives/2004/08/folksonomy_social_classification.html

Smith, W. G. (2008). Does gender influence online survey participation? A record-linkage analysis of university faculty online survey response behavior. *Eric Ed501717*.

Snow, C. P. (1959). *The two cultures and the scientific revolution.* Cambridge, UK: Cambridge University Press.

Souza, F., De Las Casas, D., Flores, V., Youn, S., Cha, M., Quercia, D., & Almeida, V. (2015). Dawn of the selfie era: The whos, wheres, and hows of selfies on instagram. In *COSN 2015 - Proceedings of the 2015 ACM Conference on Online Social Networks* (pp. 221–231).

Statista. (2019). *Distribution of Instagram users worldwide as of April 2019, by gender.* Retrieved from https://www.statista.com/statistics/802776/distribution-of-users-on-instagram-worldwide-gender/

Stegmüller, W. (1973). Band 4: Personelle und statistische Wahrscheinlichkeit. In *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie,* Studienausgabe Teil A. Berlin, Heidelberg, Germany; New York, NY, USA: Springer.

Stock, M., & Stock, W. G. (2003). Web of Knowledge. Wissenschaftliche Artikel, Patente und deren Zitationen: Der Wissenschaftsmarkt im Fokus. *Password*, *10*, 30–37.

Stock, W. G. (2000). Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In K. Fuchs-Kittowski, H. Laitko, H. Parthey, & W. Umstätter (Eds.), *Wissenschaft und Digitale Bibliothek. Wissenschaftsforschung Jahrbuch 1998* (pp. 239–282). Berlin, Germany: Gesellschaft für Wissenschaftsforschung. (Available online: http://www.wissenschaftsforschung.de/JB98_239-282.pdf)

Stock, W. G. (2001). *Publikation und Zitat. Die problematische Basis empirischer Wissenschaftsforschung.* Köln: Fachhochschule Köln; Fachbereich Bibliotheks- und Informationswesen (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 29).

Stock, W. G., & Stock, M. (2013). *Handbook of information science.* Berlin, Boston: De Gruyter Saur.

Stock, W. G., & Weber, S. (2006). Facets of informetrics: Preface. *Information-Wissenschaft und Praxis*, *57*(8), 385–389.

Stohr, D., Li, T., Wilk, S., Santini, S., & Effelsberg, W. (2015). An analysis of the YouNow live streaming platform. In *40th Local Computer Networks Conference Workshops* (pp. 673–679). Washington, DC: IEEE.

Stohr, D., Toteva, I., Wilk, S., Effelsberg, W., & Steinmetz, R. (2017). User-generated video composition based on device context measurements. *International Journal of Semantic Computing*, *11*(1), 65–84.

Stvilia, B., & Jörgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in Flickr. *Journal of the American Society for Information Science and Technology*, *61*(12), 2477–2489.

Sugimoto, C. R., & Larivière, V. (2013). *Measuring research: What everyone needs* *to know*. New York: Oxford University Press.

TagsForLikes. (2013). *#TagsForLikes.* Retrieved from https://www.tagsforlikes.com/

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing and Management*, *28*(1), 1–3.

Talbot, C. V., Gavin, J., van Steen, T., & Morey, Y. (2017). A content analysis of thinspiration, fitspiration, and bonespiration imagery on social media. *Journal of Eating Disorders*, *5*(40), 1–8.

Tang, J. C., Kivran-Swaine, F., Inkpen, K., & Van House, N. (2017). Perspectives on live streaming: Apps, users, and research. In *Conference on Computer-Supported Cooperative Work and Social Computing, CSCW 2017* (pp. 123–126). New York: ACM.

Tang, J. C., Venolia, G., & Inkpen, K. M. (2016). Meerkat and Periscope: I stream, you stream, apps stream for live streams. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4770–4780).

The Organisation for Economic Co-operation and Development (OECD). (1963). *The measurement of scientific and technical activities, proposed standard practice for surveys of research and experimental development (frascati-manual).* Paris, France: OECD.

Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, *115*(3), 1231–1240.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS ONE*, *8*(5), e64841.

Tiggemann, M., & Zaccardo, M. (2016). "Strong is the new skinny": A content analysis of #fitspiration images on Instagram. *Journal of Health Psychology*, *23*(8), 1003–1011.

Tillett, B. B. (2001). Bibliographic relationships. In C. A. Bean & R. Green (Eds.), *Relationships in the Organization of Knowledge* (pp. 19–35). Boston, MA, USA: Kluwer.

Toffler, A. (1980). *The Third Wave.* New York: Morrow.

Tokar, A. (2012). Semantic change of the publication concept? In A. Tokar et al. (Eds.), *Science and the Internet* (pp. 193–205). Düsseldorf, Germany: Düsseldorf University Press. (Available online: http://dup.oa.hhu.de/151/1/Beitrag-Tokar-2.pdf)

Trant, J. (2009). Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, *10*(1), 1–44.

Tufekci, Z. (2013). "Not this one": Social movements, the attention economy, and microcelebrity networked activism, journal=American Behavioral Scientist, volume=57, number=7, pages=848–870, language=English,.

Turner, J. M. (1995). Comparing user-assigned terms with indexer-assigned terms for storage and retrieval of moving images: Research results. *Proceedings of the 58th Annual Meeting of the American Society for Information Science*, *32*, 9–12.

Van Raan, A. F. (1997). Scientometrics: State-of-the-art. *Scientometrics*, *38*(1), 205–218.

van den Berg, J. A. (2014). The story of the hashtag (#): A practical theological tracing of the hashtag (#) symbol on Twitter. *HTS Teologiese Studies / Theological Studies*, *70*(1), 1–6.

Vander Wal, T. (2007). *No title.* Retrieved from `http://www.vanderwal.net/folksonomy.html`

Vera-Baceta, M. A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, *121*(3), 1803–1813.

Veszelszki, Á. (2016). Time, truth, tradition: An image-text relationship on instagram: Photo and hashtag. In *In the Beginning was the Image: The Omnipresence of Pictures: Time, Truth, Tradition* (pp. 139–150).

Veszelszki, Á. (2017). *Digilect. The impact of infocommunication technology on language*. Berlin, Boston, MA: De Gruyter Saur.

Vikatos, P., Messias, J., Miranda, M., & Benevenuto, F. (2017). Linguistic diversities of demographic groups in Twitter. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media, HT* (pp. 275–284). New York: ACM.

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, *1*(1), 396–413.

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society*, *20*(5), 1813–1831.

West, L. (n.d.). *These YouTube stars you've never heard of have millions of teen fans [Blog post].* Retrieved from `http://jezebel.com/these-youtube-stars-youve-never-heard-of-have-millions-1493742066.s`

Wildgaard, L. (2015). A comparison of 17 author-level bibliometric indicators for researchers in Astronomy, Environmental Science, Philosophy and Public Health in Web of Science and Google Scholar. *Scientometrics*, *104*(3), 873–906.

Xin, R. S., Hassanzadeh, O., Fritz, C., Sohrabi, S., & Miller, R. J. (2013). Publishing bibliographic data on the semantic web using bibbase. *Semantic Web*, *4*(1), 15–22.

Yanbe, Y., Jatowt, A., Nakamura, S., & Tanaka, K. (2007). Can social bookmarking enhance search in the web? In *Proceedings of the ACM International Conference on Digital Libraries* (pp. 107–116).

YouNow. (2018). Retrieved from `https://www.younow.com/partners`

YouTube. (2018). Retrieved from `https://support.google.com/youtubegaming/`

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media and Society*, *13*(5), 788–806.

Zhang, Y., Hashim, N. H., Baghirov, F., & Murphy, J. (2018). Gender differences in Instagram hashtag use. *Journal of Hospitality Marketing and Management*, *27*(4), 386–404.

Zhang, Y., Ni, M., Han, W., & Pang, J. (2017). Does #like4like indeed provoke more likes? In *Proceedings of the International Conference on Web Intelligence (WI'17)* (pp. 179–186). Leipzig, Germany.

Zimmer, F., Fietkiewicz, K. J., & Stock, W. G. (2017). Law infringements in social live streaming services. In T. Tryfonas (Ed.), *Human Aspects of Information Security, Privacy and Trust* (Vol. 10292 LNCS, pp. 567–585). Cham: Springer.

Workshare of Co-authored Publications

Chapter 2: Dorsch, I. (2018). Content Description on a Mobile Image Sharing Service: Hashtags on Instagram. *Journal of Information Science in Theory and Practice, 6*(2), 46-61.

Workshare: 100%
* Article based on own master thesis

Chapter 3: Dorsch, I. (2020). Hashtags on Instagram: Self-created or Mediated by Best Practices and Tools? In *Proceedings of 53rd Hawaiian International Conference on System Sciences (HICSS53). January 7 – 10, 2020, Grand Wailea, Maui.* Honolulu, HI: HICSS (ScholarSpace).

Workshare: 100%

Chapter 4: Philipps, J., & Dorsch, I. (2019). Gender-specific Tagging of Images on Instagram. In G. Meiselwitz (Ed.), *Social Computing and Social Media. Design, Human Behavior and Analytics. HCII 2019* (pp. 396-413). Cham: Switzerland: Springer. (Lecture Notes in Computer Science, vol. 11578).

Workshare: 20%
*Article based on supervised master thesis by Julia Philipps

Contribution:

- conceptualization,
- methodology (partially),
- data curation (feedback/supervision),
- manuscript preparation (abstract),
- review and editing

Chapter 5: Fietkiewicz, K. J., Dorsch, I., Scheibe, K., Zimmer, F., & Stock, W. G. (2018). Dreaming of Stardom and Money: Micro-celebrities and Influencers on Live Streaming Services. In G. Meiselwitz (Ed.), *Social Computing and Social Media. User Experience and Behavior. SCSM 2018* (pp. 240-253). Cham, Switzerland: Springer. (Lecture Notes in Computer Science; 10913).

Workshare: 35%

Contribution:

- manuscript preparation (abstract (together with Fietkiewicz), results paragraphs 3.1, 3.2, 3.3, 3.4, 3.5, discussion (together with Fietkiewicz)),
- review and editing

Chapter 6: Dorsch, I. (2017). Relative Visibility of Authors Publications in Different Information Services. *Scientometrics, 112*(2), 917-925.

Workshare: 100%

Chapter 7: Dorsch, I., Askeridis, J., & Stock, W. G. (2018). Truebounded, Overbounded, or Underbounded? Scientists' Personal Publication Lists Versus Lists Generated Through Bibliographic Information Services. *Publications, 6*(1), 1-9.

Workshare: 40%

Contribution:

- conceptualization (partially),
- methodology (partially),
- investigation (partially),
- data curation (partially, predominantly in a supervising position),
- manuscript preparation (parts of paragraph 2, paragraph 3 and 4),
- review and editing

Chapter 8: Dorsch, I., Schlögl, C., Stock, W. G., & Rauch, W. (2017). Forschungsthemen der Düsseldorfer und Grazer Informationswissenschaft (2010 bis 2016). *Information – Wissenschaft und Praxis, 68*(5-6), 320-328.

Workshare: 25%

Contribution:

- methodology (partially),
- investigation (partially),
- data curation,
- manuscript preparation (paragraph: 2 Methods),
- visualization (Tables 1-10),
- review and editing

Chapter 9: Kamrani, P., Dorsch, I., & Stock, W. G. (2020). Publikationen, Zitationen und H-Index im Meinungsbild deutscher Universitätsprofessoren. *Beiträge zur Hochschulforschung, 42*(3), 78-98.

Workshare: 20%
*Article based on co-supervised bachelor thesis by Pantea Kamrani

Contribution:

- conceptualization (partially),
- methodology (partially),
- manuscript preparation (inclusion of the text-passages referring to the references: p. 78 (Sugimoto/Larivière 2018), p. 78 (Rousseau/Egghe/Guns 2018), knowledge test question and p. 80-81 (Haladyna/Rodriguez 2013), p. 93 (Halevi/Moed/Bar-Ilan 2017), (Aksnes/Rip 2009; Buela-Casal/Zych 2012; Derrick/Gillespie 2013; Rousseau/Rousseau 2017; Chen/Lin 2018; Hammarfelt/Haddow 2018; Ma/Ladisch 2019; Haddow/Hammarfelt 2019; Lemke/Mehrazar/Mazarakis/Peters 2019); inclusion of reference p. 93 (Haustein/Larivière 2015); inclusion of some phrases and deleting other),
- review and editing

All Publications (2014-2020)

Publications marked with (*) are part of this dissertation.

*(022) Kamrani, P., Dorsch, I., & Stock, W. G. (2020). Publikationen, Zitationen und H-Index im Meinungsbild deutscher Universitätsprofessoren. *Beiträge zur Hochschulforschung, 42*(3), 78-98.

(021) Huvila, I., Ilhan, A., & Dorsch, I. (2020). Guest editorial. *Aslib Journal of Information Management, 72*(4), 441-444. DOI: https://doi.org/10.1108/AJIM-07-2020-372

(020) Dorsch, I., Ebrahimzadeh, S., Jeffrey, A., & Haustein, S. (2020). Metrics Literacies: Introduction of researcher personas for the understanding and use of scholarly metrics in academia. (Version v1). DOI: https://doi.org/10.5281/zenodo.4046019

*(019) Dorsch, I. (2020). Hashtags on Instagram: Self-created or Mediated by Best Practices and Tools? In *Proceedings of the 53rd Hawaii International Conference on System Sciences, January 7–10, 2020, Grand Wailea, Maui* (p. 2771–2780). Honolulu, HI: HICSS (ScholarSpace). DOI: https://doi.org/10.24251/HICSS.2020.338

(018) Dorsch, I. (2019). Trends der Informationswissenschaft: Suchmaschinen und Information Retrieval. *Information - Wissenschaft & Praxis, Band 70*(4), 203–205. DOI: https://doi.org/10.1515/iwp-2019-2025

*(017) Philipps J., & Dorsch I. (2019). Gender-Specific Tagging of Images on Instagram. In: Meiselwitz G. (Ed.), *Social Computing and Social Media. Design, Human Behavior and Analytics. HCII 2019. Lecture Notes in Computer Science*, Vol. 11578. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-21902-4_29

(016) Dorsch, I. (2018). Becoming Metric-Wise: A Bibliometric Guide for Researchers. Ronald Rousseau, Leo Egghe, and Raf Guns. Cambridge, MA: Chandos Publishing, 2018. 402 pp. $125.00 (paperback). (ISBN 9780081024744). *Journal of the Association for Information Science and Technology, 70*(5), 530-532. DOI: https://doi.org/10.1002/asi.24108

(015) Nikolic, J., Dorsch, I., Scheibe, K., Zimmer, F., & Stock, W. G. (2019). Country-Specific Sentiment on Microblogs. In *Proceedings of the 2019 Hawaii University International Conferences on Arts, Humanities, Social Sciences & Education (AHSE)* (18 pp.). Honolulu, HI: Hawaii University.

(014) Alperin, J. P., Dumas, C., Karami, A., Moscrop, D., Singh, V., Zamir, H., Ilhan, A. & Dorsch, I. (2018). Politicians & the Public: The Analysis of Political Communication in Social Media. In *Proceedings of the 81st Annual Meeting of the Association for Information Science & Technology, Nov 10-14, 2018, Vancouver, Canada* (pp. 646-649). Silver Spring, MD: Association for Information Science and Technology. DOI: https://doi.org/10.1002/pra2.2018.14505501071

(013) DePaula, N., Fietkiewicz, K. J., Froehlich T. J., Million, A. J., Dorsch, I. & Ilhan, A. (2018). Challenges for Social Media: Misinformation, Free Speech, Civic Engagement, and Data Regulations. In *Proceedings of the 81st Annual Meeting of the Association for Information Science & Technology, Nov 10-14, 2018, Vancouver, Canada* (pp. 665-668). Silver Spring, MD: Association for Information Science and Technology. DOI: https://doi.org/10.1002/pra2.2018.14505501076

*(012) Fietkiewicz, K. J., Dorsch, I., Scheibe, K., Zimmer, F., & Stock, W. G. (2018). Dreaming of Stardom and Money: Micro-celebrities and Influencers on Live Streaming Services. In Meiselwitz G. (Eds.) *Social Computing and Social Media. User Experience and Behavior. SCSM 2018,* (pp. 240-253). Cham, Switzerland: Springer. (Lecture Notes in Computer Science; 10913). DOI: https://doi.org/10.1007/978-3-319-91521-0_18

*(011) Dorsch, I. (2018). Content Description on a Mobile Image Sharing Service: Hashtags on Instagram. *Journal of Information Science in Theory and Practice, 6*(2), 46-61. DOI: https://doi.org/10.1633/JISTAP.2018.6.2.4

*(010) Dorsch, I., Askeridis, J., & Stock, W. G. (2018). Truebounded, Overbounded, or Underbounded? Scientists' Personal Publication Lists versus Lists generated through Bibliographic Information Services. *Publications, 6*(1), 1-9. DOI: https://doi.org/10.3390/publications6010007

(009) Ilhan, A., Henkel, M., Dorsch, I., & Meschede, C. (2017). Diversifizierung der Informationswissenschaft: Social Media, Open Innovation, Altmetrics. *Open Password 312*(1).

*(008) Dorsch, I., Schlögl, C., Stock, W. G., & Rauch, W. (2017). Forschungsthemen der Düsseldorfer und Grazer Informationswissenschaft (2010 bis 2016). *Information – Wissenschaft und Praxis, 68*(5-6), 320-328. DOI: https://doi.org/10.1515/iwp-2017-0060

(007) Dorsch, I., Zimmer, F., & Stock, W. G. (2017). Image Indexing through Hashtags in Instagram. In *80th Annual Meeting of the Association for Information Science & Technology, Washington, DC, Oct. 27 – Nov. 1, 2017* (pp. 658-659). DOI: https://doi.org/10.1002/pra2.2017.14505401105

(006) Ilhan, A., & Dorsch, I. (2017). Alcohol, Sex and Smoking: Adolescents on Facebook and their Self-Presentation Behavior. In *Proceedings of the 3rd International Conference on Library and Information Science (LIS)* (pp. 62-81). Taipeh, Taiwan: International Business Academics Consortium.

*(005) Dorsch, I. (2017). Relative Visibility of Authors' Publications in Different Information Services. *Scientometrics, 112*(2), 917-925. DOI: https://doi.org/10.1007/s11192-017-2416-9

(004) Dorsch, I., & Ilhan, A. (2016). Photo Publication Behavior of Adolescents on Facebook. In K. Knautz, & K. S. Baran (Eds.), *Facets of Facebook: Use and Users* (pp. 45-71). Berlin, Germany, Boston, MA: De Gruyter Saur. (Knowledge & Information. Studies in Information Science).

(003) Dorsch, I., & Frommelius, N. (2015). A Scientometric Approach to Determine and Analyze Productivity, Impact and Topics Based upon Personal Publication Lists. In F. Pehar, C. Schlögl, C. Wolff (Eds.), Re:inventing Information Science in the Networked Society. *Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th-21st May 2015* (pp. 578-580). Glückstadt, Germany: Verlag Werner Hülsbusch.

(002) Honka, A., Orszullok, L., Dorsch, I., & Frommelius, N. (2015). Topical Impact Analysis. A New Informetric Indicator for the Assessment of a Scientific Institution. In F. Pehar, C. Schlögl, C. Wolff (Eds.), *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th-21st May 2015* (pp. 410-421). Glückstadt, Germany: Verlag Werner Hülsbusch.

(001) Friedländer, M. B. (2014). Informationswissenschaft an deutschsprachigen Universitäten – eine komparative informetrische Analyse. *Information –Wissenschaft und Praxis, 65*(2), 109-119. DOI: https://doi.org/10.1515/iwp-2014-0018
Note: Contact person for Mathilde B. Friedländer (2014)

Résumé

Isabelle Dorsch
isabelle.dorsch@hhu.de || www.isi.hhu.de/dorsch || ORCID: 0000-0001-7391-5189

HIGHER EDUCATION

2014-12 – 2017-08	Master, Information Science and Language Technology, Heinrich Heine University Düsseldorf, Germany
2011-10 – 2014-11	Bachelor, Information Science and Language Technology, Heinrich Heine University Düsseldorf, Germany

WORK & PROFESSIONAL EXPERIENCE

Since 2017	Research associate and lecturer, Department of Information Science, Heinrich Heine University Düsseldorf, Germany
2015-04 – 2016-12	Research assistant and lecturer, Department of Information Science, Heinrich Heine University Düsseldorf, Germany
2014-10 – 2015-03	Student assistant and lecturer, Department of Information Science, Heinrich Heine University Düsseldorf, Germany
2019-02 – 2019-03	Visiting scholar, ScholCommLab Ottawa, ON, Canada
2013-08 – 2013-09	Internship in social media marketing, Infospeed GmbH, Cologne, Germany
2010-09 – 2011-08	Voluntary social year in culture, Department of Education, Kunstsammlung NRW, Düsseldorf, Germany

ACADEMIC SERVICE

Reviewer activity (conferences/journals), teaching and mentoring, erasmus coordinator, co-guest editor, minitrack co-chair, co-workshop organizer, ASIS&T SIG Social Media: co-founder and first co-chair (2017-2019)

Declaration of Academic Honesty

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der ‚Ordnung über die Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf' erstellt worden ist.

Ort, Datum