International Journal of Communication 13(2019), 3885–3893 1932–8036/20190005

# Crafting a Strategic Roadmap for Computational Methods in Communication Science: Learnings From the CCS 2018 Conference in Hanover

*Commentary*

JULIA NIEMANN-LENZ[1]
SOPHIE BRUNS
DOROTHÉE HEFNER
KATHARINA KNOP-HUELSS
DANIEL POSSLER
SABINE REICH
Hanover University of Music, Drama and Media, Germany

LEONARD REINECKE
Johannes-Gutenberg University Mainz, Germany

JULE SCHEPER
CHRISTOPH KLIMMT
Hanover University of Music, Drama and Media, Germany

The rise of digital communication brings about dramatic challenges for theories of communication (and social sciences in general), but also opportunities and problems regarding the expansion of empirical methods employed to study human communication and its manifold digital traces. A team of German communication scholars at Hanover University of Music, Drama and Media and at the University of Mainz held an international

Julia Niemann-Lenz: Julia.Niemann-Lenz@ijk.hmtm-hannover.de
Sophie Bruns: Sophie.Bruns@ijk.hmtm-hannover.de
Dorothée Hefner: Dorothee.Hefner@ijk.hmtm-hannover.de
Katharina Knop-Hülß: Katharina.Knop-Huelss@ijk.hmtm-hannover.de
Daniel Possler: Daniel.Possler@ijk.hmtm-hannover.de
Sabine Reich: Sabine.Reich@ijk.hmtm-hannover.de
Leonard Reinecke: leonard.reinecke@uni-mainz.de
Jule Scheper: jule.scheper@ijk.hmtm-hannover.de
Christoph Klimmt: christoph.klimmt@ijk.hmtm-hannover.de
Date submitted: 2019–05–16

[1] The authors of this contribution were the organizers of the CCS 2018 conference in Hanover.

conference in Hanover with both young and experienced scholars in February 2018 to promote the development of collective, cross-institutional strategies for computational communication science (CCS). The key findings and learnings from the conference are documented in this conference report. We intend to stimulate organizational and collaborative efforts in establishing infrastructures, knowledge bases, standards of best practice, and a spirit of solidarity among interested scholars to push the field's digital-methodological progress and to compensate for different starting conditions that scholars and institutions are facing on their journey into the digital future.

*Keywords: computational communication science, research strategies, research infrastructure, best practice, ethical standards, future agenda*

The Internet infrastructure of mass, interpersonal, and hybrid modes of communication has gained tremendous importance around the globe. Large parts of the world population habitually access online platforms, services, and media, and they perform a broad diversity of activities, including, but not limited to, actual communication. As contemporary life is becoming increasingly digital, the social sciences—and communication in particular—need to adapt to this development: The rise of digital communication brings about dramatic challenges for theories of communication (and social sciences in general), but also opportunities and problems regarding the expansion of empirical methods employed to study human communication and its manifold digital traces (boyd & Crawford, 2012; Choi, 2018; Lazer et al., 2009).

Given that Internet-based communication is penetrating more and more domains of daily life, including business, intimate relationships, (semi-) public deliberation of politics, and education, the analysis of "big data" traces is rapidly turning from an interesting opportunity into a pressing necessity for communication, psychology, and other social sciences. At the same time, the hopes that are associated with computational methods for communication science are that they allow researchers to (a) understand theoretically new phenomena (such as human communication through social media) and (b) overcome limitations of conventional methods of the social sciences. One example for this hope is the notorious problem of limited resources that force communication scholars to operate with samples (of media messages or media users) instead of investigating all single incidents or cases of interest that occur in the real world. With computational methods, it does (technically) not make a big difference whether a scholar wants to examine 100 or 1,000,000,000 acts of human communication ("posts" in social media, for instance). A second important example of the positive expectations about computational methods is that many of them are nonreactive and may therefore be capable of overcoming measurement problems that scholars are facing when employing conventional methods such as self-report instruments.

Against this backdrop, the field of communication is enjoying many initiatives and innovative approaches to computational methods that are often pursued by teams of young scholars. Software frameworks and best-practice procedures are being developed and applied to expand the capabilities of the field in the digital realm (e.g., Trilling & Jonkman, 2018; Weber, 2018). However, many researchers are also facing severe problems and great obstacles in meeting the diverse preconditions of successful computational research. Thus, there is substantial variation among scholars and institutions of

communication science with regard to the velocity, depth, and effectiveness of adopting computational methods. The fierce competition within the field might also contribute to a lack of collective action in creating a shared methodological knowledge base for departing toward the computational future of empirical communication scholarship.

Therefore, a team of German communication scholars at Hanover University of Music, Drama and Media and at the University of Mainz held an international conference in Hanover with both young and experienced scholars in February 2018 to promote knowledge sharing and the development of collective, cross-institutional strategies for computational communication science (CCS). Generous funding of the Volkswagen Foundation enabled us to invite about 80 junior and senior researchers and involve them in a two-day highly interactive workshop setting. The mission of the conference was to generate building blocks for collective strategies in developing capabilities in digital methods: a strategic roadmap for computational communication science. As part of the closing session of the conference, all attendees participated in a panel of semistructured, moderated group discussions of up to eight participants to summarize the central insights gained over the course of the conference and to reflect on the current challenges faced by CSS researchers. The key findings and learnings from these moderated discussions are documented in the following sections of this conference report. We intend to stimulate organizational and collaborative efforts in establishing infrastructures, knowledge bases, standards of best practice, and a spirit of solidarity among interested scholars to push the field's digital-methodological progress and to compensate for different starting conditions that scholars and institutions are facing on their journey into the digital future.

We organize the many learnings of the conference into four sections: building new competencies for applying computational methods; institutional-organizational readiness for the digital methods of the future; ethical issues in computational research; and implications of big data research for scholarly dissemination and publication. Our hope is that the emerging dynamic community of computational communication researchers will benefit from this report in their innovation strategies and collaborative efforts. We wish to thank all those who attended the CCS conference in Hanover in February 2018—participants, speakers, and contributors to the discussions summarized hereafter—for their engagement and valuable input.

## Building New Competencies for Applying Computational Methods

During the discussion on the most essential competencies to enable researchers to productively participate in CCS, two major topics emerged: technical literacy and collaboration skills. Gaining a basic understanding of the logic of at least one programming language was agreed on as key to participating in CCS. As a first step toward building this competency, participants suggested switching from a proprietary software solution, such as Excel and SPSS, to a more transparent framework, such as the open-source programming language and software environment R or Python. Participants reported that some universities—for example, in Germany and the United States—have already started teaching statistics and data analysis using R starting at the BA level. This way, students acquire basic statistics while also learning a programming language, including knowledge on how to document code and how to customize or debug scripted code. However, discussants also underlined that priority should continue to be given to teaching methods (both traditional and computational) at a level that enables students to ask the right questions and

then to apply those methods intelligently. To provide researchers with technical literacy, participants suggested that those university departments with experience and knowledge should serve as lighthouse projects that provide training opportunities for all academic levels above a master's degree.

The second key competency to effectively participating in CCS identified by the conference attendants was developing collaborative skills. Developing these skills is especially important given that CCS projects are often interdisciplinary. Furthermore, technological possibilities are constantly evolving—for example, R might eventually be replaced by a more advanced framework, which would prompt new questions and challenges. Participants therefore agreed that it is invaluable to know how to work together as a research team with experts on computational questions, which includes respecting and appreciating each other's expertise and knowing when, where, and how to ask for help. To build up such collaboration skills, some conference participants envisioned joint research classes at the graduate level in which students from communication and computer science would learn, practice CCS methods, and develop entire studies together.

Last, participants agreed that in general, communication scholars should obtain a basic knowledge of what computational methods are available and which research problems can be solved using those methods. Just like communication scholars learn when to employ a survey study and when to conduct a content analysis as part of their basic training, it is of increasing importance for communication researchers from all parts of our discipline to also acquire a basic understanding of the scope and limitations of computational methods. The discussants suggested including this new methods expertise in the curricula of communication programs at both the undergraduate and graduate levels.

### Institutional-Organizational Readiness for the Digital Methods of the Future

The conference participants also discussed how CCS should be organized and institutionalized within individual universities and communication science in general. The debate focused on the question of whether CCS can be integrated into existing institutional structures or whether it requires the development of new solutions. Some participants argued that specialized CCS professorships within existing department structures should be established. These professorships could advance the application of computational methods in the department and would signalize that CCS is institutionally defined as part of communication science. However, other discussants rejected this idea because specialized professorships might increase the division between CCS scholars and other areas of the field (e.g., media effects research). Instead, they argued that computational methods constitute a cross-cutting topic relevant for many communication research questions.

Similarly, participants presented different ideas on how the technological resources required for CCS (e.g., server capacities) could be provided. While some suggested that existing IT services at universities should be in charge, others proposed that academic associations should develop new inter-institutional (even inter-national) solutions to provide resources such as server infrastructure (i.e., virtual machines) or centralized solutions for data collection and storage. In addition, the importance of funding agencies was highlighted in this context. Some participants suggested that funding guidelines need to be adjusted to allow scholars to apply for equipment and resources required for CCS methods and procedures.

With regard to the institutionalization of CCS in communication science, participants agreed that some mode of formally organizing the CCS community is required to advance the field, particularly to establish methodological standards and improve training options (e.g., online courses tailored for the needs, practices, and standards of communication science). While some argued for the benefits of founding a new CCS-related community, interest group, or network (such as the Computational Methods Special Interest Group at the International Communication Association ICA), others suggested that computational methods should be included in broader focused method groups, which have already been established in some academic associations (e.g., the German Communication Association). Moreover, the participants were in favor of institutionalizing meetings of scholars interested in CSS in the context of existing conferences (e.g., a regular preconference before the ICA conference). However, the development of new session formats might be necessary for such meetings because many participants expressed that they are interested not only in presenting their findings but also in method training. Moreover, it was suggested that these meetings should also enhance interdisciplinary networking, given that scholars outside the field of communication often use similar computational methods and deal with closely related research questions. Finally, some participants noted that these efforts for institutionalizing CCS and promoting interdisciplinary exchange place heavy burdens on scholars' limited resources. Therefore, they voted for changing the incentive structure of communication sciences to better acknowledge services provided to the community.

## Ethical Issues in Computational Research

A third stream of discussion during the conference addressed ethical issues in CCS. In this context, discussants named sensitive data handling as a central ethical challenge. CCS often operates with publicly available data, which frequently include sensitive data such as information on gender, health, or sexual orientation. This raised the question regarding to what extent publicly available data can be collected, processed, and published without ethical concern. Participants called for a clear policy for CCS researchers on how to handle publicly available personal data in a sensitive way.

In the context of sensitive data handling, participants assumed that Internet users are not fully aware of how their publicly available personal data can be used by third parties. However, for researchers to be able to use this data with fewer ethical concerns, it is necessary for users to master this sophisticated knowledge. In the eyes of the participants, a possible solution included strengthening citizen awareness of data use, implementation of informed consent, and the resulting transparency of data use.

In addition to the lack of awareness among Internet users, a missing awareness among CCS researchers was debated as an ethical challenge in itself. Participants agreed that many scientists are not properly informed about ethical issues in CCS (e.g., in the context of crawling data). Consequently, ethically problematic research approaches may occur. For participants, a possible solution included strengthening the general awareness of ethical challenges of both scholars and institutions. Therefore, participants called for closer cross-university and transnational cooperation as well as the consideration of regulations and policies of other research institutions and research disciplines. This should result in new guidelines for ethical CCS practices that are cross-university and transnational on the one hand, but flexible for different research areas on the other. As a challenge for these guidelines, participants noted the internationally varying possibilities for data collection in CCS. On the question of how guidelines should be monitored, some

participants called for scientists themselves to assume more responsibility; others demanded that universities or academic journals assume more responsibility for the ethical compliance of CCS studies. Further, local ethics commissions, which are already in place in the field of psychology, were proposed as a supervisory body.

As a last ethical challenge, a possible conflict between the current open science movement and ethical requirements of CCS research was discussed. While data sharing in the context of open science could improve the reproducibility of results, it also increases the risk of data misuse by third parties. Furthermore, participants agreed that the costs for compliance with the necessary ethical standards of open CCS research (e.g., complete anonymization of very large amounts of data) could potentially be greater than the benefit (e.g., reproducibility). As a solution, participants proposed establishing more platforms that make data accessible not to everyone, but only on request. This way, researchers remain in control of their data and avoid unnecessary preparatory effort before data sharing is actually requested by other scholars.

**Implications of Big Data Research for Scholarly Dissemination and Publication**

With the growing use of computational methods in communication science, authors, reviewers, and editors face the challenge of adapting the publication process to the requirements and possibilities of these methods. Topics of this stream of discussion at the CCS conference related to the current availability of outlets for publication, necessary adjustments of the review and publication process, and legal and ethical considerations related to publishing big data in the context of human communication.

With regard to possibilities of publishing studies using computational methods, two options were considered: improving the integration of computational methods into already existing communication science journals, and establishing specific computational methods and research journals. To overcome the status quo of computational methods as unconventional and peripheral to the field and to integrate them into the standard methods repertoire of communication research, editors of traditional academic journals were named responsible. They need to take measures to facilitate the publication of studies using computational methods. The proposed measures ranged from introducing computational methods in special issues and revising author guidelines to allow other forms of publication beyond the printed version (e.g., interactive visualizations), to providing reviewer instructions for the evaluation of computational methods. However, a computational methods journal—such as the recently launched journal *Computational Communication Research*—is regarded as a valuable outlet for methods research and the development of methodological standards in CCS.

Given that the growing field of computational methods research comprises a wide range of different approaches, a main challenge for the publication of individual studies was seen in finding knowledgeable and competent reviewers. Recruiting experts as "statistical or methods reviewers" who are exclusively responsible for evaluating the methods and results section of an article and therefore do not necessarily need to be familiar with the applied theory (e.g., experts from other disciplines) was considered crucial for establishing high quality standards within this rising body of research. Furthermore, common standards for reporting methods and results, citing code and software, and providing data sets, scripts, and documentation need to be developed, required by journals as mandatory, and consistently applied by authors.

In the context of increasing demands for transparency and replicability through data sharing and new possibilities for accessing and analyzing big data in the context of human communication, careful consideration of legal (e.g., copyright) and ethical aspects (e.g., privacy concerns) in the publication process is required. Therefore, scientists are confronted with the challenge of answering many partially complex questions before their study can be published. For example: How many resources need to be invested to make a data set sharable (e.g., cleaning, anonymization)? Does sharing a data set violate copyrights, potentially in different countries? Is technical infrastructure available to make large data files accessible for reviewers and readers? What information do other researchers need to review and work with the provided data (e.g., documentation, programming scripts, software)? Obviously, ethical questions in CCS research (mentioned earlier) also manifest themselves in the context of publishing.

## Synthesis: Drafting a Roadmap for the CCS of the Future

The moderated debates at the Hanover CCS conference revealed a broad array of issues that need intellectual reflection, careful addressing in practical research, and organized efforts for setting standards of computational communication research. Young scholars particularly are eager to acquire skills, to pilot implementation opportunities, and to generate entirely new kinds of insights into human communication through CCS. At the same time, the emergence of computational methods brings about challenges that are similar to those that the conventional social sciences have been facing since their early years as well: How do we ensure validity of empirical findings through optimal choice, setup, and implementation of methods? How do we balance the (possibly conflicting) goals of producing relevant research results and of ensuring ethical principles of science in society? How do we agree on common practices (standards of the field) without suppressing method innovation and unique approaches that expand our knowledge beyond the state of the art?

The answers that our discussants have given suggest rethinking key elements of the process of empirical communication research for the CCS era. We synthesize their contributions into four central building blocks of our CCS roadmap: partnerships, training, active debate over standards, and solidarity of sharing knowledge and resources.

**Partnerships.** To successfully harmonize the technological-digital ramifications with social-scientific principles, communication scholars and their institutions need to practice a continuous spirit of partnerships. Partnering with scholars from computer science or with other social research teams who bring missing expertise into joint projects will be a necessary and powerful ingredient of future CCS. Such partnerships may be quite asymmetrical: Some technical procedures may look immensely challenging to a communication scholar, but can easily be implemented by a master's student in computer science, as one of our discussants illustrated at the Hanover conference. Alternatively, partnerships may turn out to be highly symmetrical, as computer scientists can benefit from communication research goals and theory in finding a real-world "case" for applying their innovative technologies, demonstrating their functionality, and justifying the relevance of their academic investments. Partnerships may last for single studies or projects, but they may also relate to enduring configurations such as collaborative interdisciplinary programs of study (communication plus informatics) or joint-appointment professorships of computational communication research (methods) that rest on agreements among different university departments.

**Training**. Many interesting training opportunities exist for scholars who want to apply available computational methods, and making use of such opportunities clearly is a worthwhile investment for communication scholars. Finding orientation among the many options is, however, difficult as long as there are no well-defined "standard procedures" or "best-practice" methods in CCS. Because many current training services stem from scholars and experts outside the field of communication, an important developmental pathway for the emerging CCS community will be the elaboration and provision of field-specific training in computational *communication* research methods. One interesting observation that we made at the Hanover conference with regard to training in CCS pertains to assisting senior scholars in acquiring knowledge about CCS even if they do not intend to apply the methods themselves. Senior scholars make decisions in the peer review system about funding proposals and about article publication, and they file evaluations of (young) scholars' performance. It is thus imperative to also familiarize senior scholars in empirical communication research with the principles and (evolving) quality standards of CCS to enable them to fulfill their role as decision makers and evaluators; only then will CCS be sustainably adopted by the field as a whole. Training senior personnel may require different instruments from those courses offered for young scholars; seniors may need overview knowledge and an understanding of basic principles, as opposed to acquiring the technical details of a programming tool, for example.

**Active debate over standards.** With computational methods, a huge diversity of options for accessing, collecting, analyzing, and interpreting data becomes available to communication scholars. Not all these options will meet quality criteria such as reliability, generalizability, and validity. To establish comparability and criticizability of computational methods, standards and criteria of excellence need to be developed, debated, and agreed on among the research community. Such standards would refer not only to technical parameters in implementing computational methods, but also to research ethics and documentation principles in research dissemination. The open and friendly debates we hosted at the Hanover conference inspired us to suggest adopting the "request for comments" (RFC) principle of stimulating discussions about standards that Steven Crocker (1969) introduced in 1969 to the Internet development community. RFC documents are essentially proposals on specific procedures, methods, scripts, or other methodologically relevant concepts that are disseminated among community members who are invited to provide feedback, suggest improvements, signal (dis)agreement, and suggest alternatives. The RFC "culture" that evolved during the early days of the Internet was enjoyed for its democratic, open spirit and its capacity to drive forward consensus building among a relatively large expert community. We believe that if communication scholars who hold experience with applying computational methods to "classic" communication research questions formalized their methods propositions as RFCs, this would help to shape community by fostering debate and agreement over excellence (dos and don'ts) in CCS.

**Solidarity**. Some types of CCS research require tremendous resources, such as access to data repositories, server capacity, or specialized software tools. Some scholars may enjoy the benefits of powerful host institutions that can provide these resources (including the technical experts to run and customize them), whereas many other scholars may face severe barriers to participating in such types of CCS. If communication as a field is to adopt CCS into its portfolio of common methods and practices, the community will need to find ways of sharing scarce resources in a solidary manner in order to avoid new asymmetries between "rich" and "poor" institutions and countries and to enable the best scholarly ideas to flourish. The

sharing ideology of the open-source software community (e.g., Hars & Ou, 2002) may provide inspiration to the CCS community in terms of how to practice, organize, and enjoy such solidarity.

## Conclusion

The current summary of the conversations held at the 2018 Hanover CCS conference has identified major challenges that communication as a field is facing on its way into the future of digital methods. An impressive set of suggestions and solutions has been generated by our participants; we hope that the resulting elements of our proposed roadmap for the field will turn out to be as helpful in supporting active CCS scholars in their efforts to apply, improve, and institutionalize computational methods as powerful, valid, and ethical additions to the empirical toolbox for studying human communication. The contributions that the editors of the current Special Section of *IJoC* have collected certainly converge with this mission: They showcase effective connections between communication scholarship and data science and further illustrate the challenges that are lying ahead for the next big step in digitalization of the field.

## References

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662–679. http://dx.doi.org/10.1080/1369118X.2012.678878

Choi, S. (2018). When digital trace data meet traditional communication theory: Theoretical/methodological directions. *Social Science Computer Review.* https://doi.org/10.1177/0894439318788618

Crocker, S. (1969). *Host software*. RFC No. 1. Retrieved from http://www.rfc-editor.org/rfc/pdfrfc/rfc1.txt.pdf

Hars, A., & Ou, S. (2002). Working for free? Motivations for participating in open-source projects. *International Journal of Electronic Commerce*, *6*(3), 25–39. https://www.jstor.org/stable/27751021

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., . . . Van Alstyne, M. (2009). Computational social science. *Science, 323*(5915), 721 LP–723. https://doi.org/10.1126/science.1167742

Trilling, D., & Jonkman, J. G. (2018). Scaling up content analysis. *Communication Methods and Measures*, *12*(2–3), 158–174. https://doi.org/10.1080/19312458.2018.1447655

Weber, M. S. (2018). Methods and approaches to using Web archives in computational communication research. *Communication Methods and Measures*, *12*(2–3), 200–215.