Asian Journal of Communication

ISSN: 0129-2986 (Print) 1742-0911 (Online) Journal homepage: https://www.tandfonline.com/loi/rajc20

Introducing computational social science for Asia-Pacific communication research

Tai-Quan Peng, Hai Liang & Jonathan J. H. Zhu

ASIAN JOURNAL OF COMMUNICATION
2019, VOL. 29, NO. 3, 205–216
https://doi.org/10.1080/01292986.2019.1602911

EDITORIAL

Introducing computational social science for Asia-Pacific communication research

Tai-Quan Peng[a], Hai Liang[b] and Jonathan J. H. Zhu[c,d]

[a]Department of Communication, Michigan State University, East Lansing, MI, USA; [b]School of Journalism and Communication, Chinese University of Hong Kong, Shatin, Hong Kong; [c]Department of Media and Communication, City University of Hong Kong, Shatin, Hong Kong; [d]School of Data Science, City University of Hong Kong, Shatin, Hong Kong

ABSTRACT
Computational social science has caused a shift of research paradigm in social science in general and communication in particular. The special issue brings together a community of active researchers to introduce computational social science for Asia-Pacific communication research. The special issue outlines major computational methods closely related to communication research and demonstrates how computational methods can be applied to address theoretical and practical questions in Asia-Pacific societies. The advantages and limitations of computational methods have been conceptually discussed and/or empirically illustrated. Finally, the special issue provides a guideline of conducting computational research for communication researchers in Asia-Pacific societies and beyond.

KEYWORDS
Computational social science; communication research; research paradigm; big data; social media

It has been a decade since the publication of the manifesto of computational social science (CSS) (Lazer et al., 2009). Almost all subject fields in social science have embraced the emergence and prosperity of CSS either actively or reactively. CSS is not a hype triggered by increasingly bigger data on the Internet. CSS is not a myth caused by increasingly sophisticated computing algorithms. CSS is not a mechanical piece-together of social science and computer science. CSS represents a paradigm shift in social science that pose equally, if not more, rigorous demands for theorization, research design, statistical analysis, and results presentation.

To respond to this paradigm shift, several influential journals have published special issues on challenges and opportunities that CSS has brought to social science in general and communication in particular, for example, *Communication Methods and Measures*' special issue on 'Computational methods for communication science' in 2018, *Annals of the American Academy of Political and Social Science*'s special issue on 'Toward computational social science: big data in digital environments' in 2015, and *Journal of Communication*'s special issue on 'Big data in communication research' in 2014.

While pioneering, these publications did not directly address the specific conceptual, methodological, and technological challenges and opportunities in the application of computational methods in Asia-Pacific communication research. Such challenges and

CONTACT Tai-Quan Peng winsonpeng@gmail.com

opportunities are deeply rooted in the linguistic, cultural, and political diversity in Asia-Pacific societies. To fill the gap, the current special issue brings together a community of active researchers to introduce CSS for Asia-Pacific communication research. In this introduction, we will first outline what are major computational methods closely related to communication research. Then we will discuss what are the advantages and limitations of computational methods to communication research in general and to Asia-Pacific communication research in particular. We will further discuss how communication researchers can adopt computational methods in their own research. Finally, we will provide a brief summary of all the papers included in the special issue.

Computational methods for communication research: what are they?

Communication research deals with the study of communication phenomena as a social process, which is concerned with *Who Says What to Whom in Which Channel with What effects*. As an emerging paradigm of social science, the CSS paradigm can inform empirical research in all 5Ws domains of communication research. The CSS paradigm is of particular relevance in the following areas: user analytics in audience research, text mining in content analysis, and experiment research in effects research.

User analytics

User analytics focuses on unraveling and understanding behavioral patterns with digital traces on social media. Such behavioral patterns can be about *how* users will use social media in certain ways, while it can be about *how* users make behavioral responses to social stimuli (e.g. campaign messages, advertising messages) accessible on social media. Moreover, user analytics aims to uncover psychological and social factors associated with the formation, maintenance, and change of behavioral patterns on social media. User analytics in computational communication research differs from traditional audience analysis in terms of its emphasis on *dynamic* and *structural* perspectives.

With the increasing proliferation of behavioral options available on social and mobile media, time and timing in human behavior have moved from the background to the foreground, which calls for a dynamic perspective in the empirical examination of behavioral patterns in user analytics. The dynamic perspective explicitly puts human behaviors under a temporal lens and contextualizes human behaviors from their historical occurrences. It enables researchers to detect intricate temporal regularities, such as circadian rhythm and weekly cycle, underlying human behaviors. Moreover, it empowers researchers to re-conceptualize the unfolding of human behaviors as a sequential process. The sequence conceptualization of human behaviors offers innovative ways to 'understand emergence and change as well as stability, and they incorporate understandings of causality as constituted through chains of events rather than through abstract correlations' (Langley, Smallman, Tsoukas, & Van de Ven, 2013, p. 10).

Users on social media are empowered to connect and communicate with one another via multitudinous behaviors (e.g. befriending other users, retweeting others' messages, commenting on others messages), which facilitates their networking and interaction at an unprecedented scale. The networking and interactive features on social media call for a structural perspective in the empirical examination of behavioral patterns in user analytics. The structural perspective, arguing for the inter-connection among individuals as building blocks of social systems, has gained increasing prominence in CSS. By contextualizing a user from the network of other users, the structural perspective can offer new insights on users' behavioral patterns that are difficult, if not impossible, to capture in traditional attribute-based audience analysis.

User analytics with a structural perspective enables empirical examination of topological characteristics of social and communication networks induced by users' behavioral interactions on social media. Moreover, the structural perspective can help researchers reveal and explain the differences on users' structural positions in such networks, understand how users' behavioral decisions (e.g. diffusion, participation, and engagement) are influenced by their local network positions in addition to personal characteristics, and discover how behavioral interactions between dyads or triads of users can lead to the formation of a network with certain global-level topology.

Text mining

Content produced in communication processes is social artifacts that carry important information about its creators and audiences. Content analysis, a mainstream research method in the toolkit of communication researchers, has been widely adopted to analyze social artifacts in an objective, systematic, and quantitative way. In the past, only social artifacts produced by professional organizations (e.g. mass media and commercial companies) and elite social groups (e.g. political leaders) are well archived, which has been the major data source for content analysis in communication research. With the increasing popularity of social media in the past decades, ordinary users are entitled to produce and share messages at an unprecedented scale. Most of the user-generated content are well documented on various social media platforms, which has triggered the curiosity of communication researchers to understand what has been said on various topics by different groups of users there. However, the large-scale, unstructured and streaming nature of user-generated content on social media requires an innovative and efficient way to detect latent semantic patterns (e.g. topics, themes, and sentiment) underlying manifest content.

Text mining, theoretically and technically rooted in linguistics and computer science, has offered a viable approach to extract hidden concepts from a large collection of texts and detect semantic relations among extracted concepts. Text mining has gained a great deal of attention from communication researchers in recent years because of the tremendous amount of text data available on social media, the improved efficiency and ease-of-use of analytical algorithms and tools, and the emergence of powerful and affordable computational infrastructures (van Atteveldt & Peng, 2018).

Text mining includes a class of methods and techniques that can be generally grouped into two approaches: unsupervised approach and supervised approach. Unsupervised text mining aims to find latent structure out of unlabeled text data, which does not require any prior inputs from researchers. The most common unsupervised text mining techniques include cluster analysis and topic modeling. Supervised text mining requires an adequate amount of manual effort in developing labeled training dataset. The training dataset will then be used to infer a function or learn a classifier in order to perform predictions on unlabeled data. Supervised text mining techniques include but are not limited to logistic regression, naïve Bayes, support vector machine, artificial neural networks, and random forests.

The unsupervised approach outperforms the supervised approach in terms of its generality. As the unsupervised approach requires very limited input from manual coding, it can be applied to any text data in any domains of interest. Nevertheless, the supervised approach is tuned to the training data and thus needs retraining if it is to be applied in other contexts. The unsupervised approach falls behind the supervised approach in terms of its accuracy. The supervised approach is likely to produce more accurate classification result than the unsupervised approach given the quality of the training data is adequately sound, while the validity and interpretability of the results from unsupervised text mining have become a shared concern in empirical applications.

Digital filed experiment

Experiments have proven to be the most powerful method to test causal relationships in scientific research. Many experimental studies were run in offline settings where a group of ‘unusual’ subjects (i.e. college students) is recruited to perform certain ‘unusual’ tasks (i.e. experimental stimulus) in an ‘unusual’ environment (i.e. artificial laboratories) (Salganik, 2018). The ‘unusualness’ in offline lab experiments seriously constrains the external validity of the causal conclusions drawn from them. The development of social and mobile media in the past decades has drastically changed how experiments can be designed and implemented in social science in general and communication research in particular. Among all the changes, three are arguably the most disruptive ones: (1) the recruitment of sizable and heterogenous research participants, (2) enhanced measurements of subjects’ responses to experimental stimuli, and (3) enriched strategies in running online experiments.

Social media platforms like Facebook and crowdsourced marketplace like Amazon Mechanical Turk create an online labor pool (Golder & Macy, 2014) where researchers can recruit fairly sizeable number of participants. It is not unusual now to run an experiment with thousands of online participants. Some online experiments (e.g. Bond et al., 2012; Kramer, Guillory, & Hancock, 2014) recruited millions of participants, which was completely unimaginable in prior research. More importantly, in comparison with the college student participants in most offline lab experiments, participants recruited in online platforms can have greater heterogeneity in their demographic characteristics (e.g. age, gender, race, cultural background, etc.) and other key variables relevant to specific researcher contexts, which can allow researchers to explicitly control the ‘unobserved heterogeneity’ (Hutchinson, Kamakura, & Lynch, 2000) and improve the external validity of experimental research.

Digital traces on social and mobile media significantly enhance the measurements of subjects’ responses to experimental controls in online experiments. In offline lab experiments, self-reported measures have been widely adopted to observe the effects of experimental stimuli. On social media, user-generated content and user-initiated behaviors (e.g. liking, commenting) can be creatively employed to observe users’ cognitive, emotional, and behavioral reactions to experimental stimuli. Moreover, the advancement of wireless and mobile technologies, ranging from smartphones to wireless sensors, has enabled researchers to recruit and track participants in a natural environment and to collect their real-time, continuous biological, behavioral, and environmental data (Kaplan & Stone, 2013). These data have the potential to contribute to our understanding of how the causal effects in an experiment can evolve over time and how the causal effects can be conditioned or confounded by natural environments.

The virtual space offers enriched strategies to run online experiments. Salganik (2018) provided a detailed discussion of these strategies as well as their pros and cons. Researchers can run online experiments by leveraging existing platforms (e.g. Facebook, Twitter), which is the most cost-effective and low-risk strategy. Second, researchers, with adequate funding and technical support, can develop customized platforms or products to run online experiments. Third, researchers can collaborate with IT companies that can enable them to run experiments that researchers cannot do by themselves.

Computational methods for communication research: advantages and limitations

Computational methods demonstrate several obvious advantages over traditional quantitative research methods. Foremost is the nature of the data used in CSS research: a major data source is digital traces on the Internet. Such digital traces include multi-dimensional information, such as social relations, social interactions, text, and audio/video, most of which are recorded with fine-grained timestamps. Moreover, most of the digital traces are produced by ordinary users in a natural environment, which diminishes the obtrusiveness of measurement instruments in empirical research. The user-generated, multi-dimensional, and time-stamped digital traces offer enriched perspectives to explicate human communication phenomena in various domains (e.g. political, health, organization, and advertising) across multiple levels (e.g. individual and aggregate).

The second advantage is the efficiency and effectiveness in data collection. The increasing power and ease-of-use of computational algorithms improve the speed and lower the cost in retrieving digital traces. It is not a luxury for researchers to obtain thousands or millions of digital records for a single research project. Moreover, the digital traces can be accessed in near-realtime, which make it possible to track the emergence and evolution of social phenomena or human behaviors over a much longer period of time.

The third advantage of CSS research is the enhanced modeling techniques in data analysis. The time-stamped, multi-dimensional, and large-scale digital traces in CSS research facilitate the adoption of enhanced modeling techniques to uncover intricate mechanisms and subtle patterns in human communication. Such modeling techniques include but are not limited to time series analysis, sequential modeling, mixed effects modeling, network analysis, and spatial modeling. These enhanced modeling techniques allow communication researchers to revisit classical theoretical questions with powerful research design or explore new research concerns with refreshed perspectives and instruments.

Despite the aforementioned advantages of computational methods, it is reasonable to acknowledge and discuss the limitations of computational methods, which can prevent communication researchers from abusing or misusing computational methods in empirical settings. The first limitation is tied to the 'found' nature of the data. Digital traces on social and mobile media platforms are owned by commercial IT companies. They are incidentally found rather than intentionally made by third-party researchers for their empirical studies with different concerns. The 'found' nature of the data imply that researchers have very limited control over the representativeness of the research subjects involved in the found data. It is an empirical question whether users who generate digital traces on the Internet can represent a population of a society. The 'found' nature of the data also imply that researchers need to be very cautious about measurement validity of the variables constructed from the found data.

The second limitation is caused by the behavioral nature of the data. Most of digital traces on the Internet are about human behavior or the artefacts of human behavior. They provide very limited information about psychological traits underlying human behaviors, such as attitude, affect, and motivation. Although computational algorithms have been developed to infer psychological traits from behavioral traces, it calls for rigorous assessment of the measurement validity of derived psychological traits based on certain ground truth.

Last but not least, it has become an increasingly eminent issue in CSS research on how to achieve a balance between the protection of users' privacy on media platforms, the protection of the commercial interest of media platforms, and the protection of researchers' right. There is no doubt that users' privacy and the commercial interest of media platforms should be well respected and adequately protected. However, they should not hurt real public research. To find such a balance is not an easy task, which requires a deliberative and constructive discussion among all stakeholders, including the general public, scholars, IT companies, and government agencies.

Computational methods for communication research: why Asia-Pacific?

Asia-Pacific societies have the largest population in the world. However, human communication in certain Asia-Pacific societies, such as Cambodia, Myanmar, and Bangladesh, has been under-studied mostly because of the unavailability of research datasets with sound quality. With the increasing penetration of social and mobile media in these societies, it is possible for researchers to utilize user-generated digital traces to explore the uniqueness and commonness of human communication in these societies. For instances, social and communication relations among users in these under-studied societies can be analyzed with user analytics methods, public attention and public sentiment towards various social issues in these under-studied societies can be mined with text mining approach, and the persuasiveness of health or political campaigns can be empirically tested with online experiments in these under-studied societies. These studies can advance our factual knowledge about human communication in these societies.

Asia-Pacific societies vary substantially in the cultural, linguistic, political and economic systems, which imply both opportunities and challenges in introducing computational methods for communication research in Asia-Pacific region. Such variations mean more heterogeneity in online expressions, in behavioral patterns, and in the backgrounds of research subjects. When such variations are properly handled, they can substantially enhance the external validity of empirical findings and expand the boundaries of existing communication theories.

Nevertheless, cultural and linguistic diversities in Asia-Pacific societies imply technical challenges in applying computational methods in Asia-Pacific communication research. Most of the text mining algorithms are developed with English and other western languages as the testbed. To apply them to non-western languages requires a huge amount of revisions, customization, or tests of the original algorithms. To correctly interpret the outcomes of text mining algorithms also requires a domain knowledge about local social and cultural systems.

Computational methods for communication research: how to do?

Successful application of computational methods in communication research requires a well-planned research design, deliberate data gathering, sophisticated data analytics, and critical evaluation. Sound application of computational methods in communication research also hinges on interdisciplinary collaborations between communication scholars and their peers in other disciplines.

Research design

There are two types of research designs in social science – deductive and inductive –, both of which are applicable in CSS research. The overarching research questions in empirical studies determine which design is more appropriate, what kind of data to be collected, and what statistical models to be estimated. Sometimes the inductive approach is more useful and straightforward in big data research, whereas the deductive approach can play better roles in online surveys and experiments. In some instances, the deductive and inductive designs can be well integrated in one study. Salganik (2018) provided a summary of various design strategies in CSS research. Here we focus on two issues that are of particular importance in computational communication research: conceptualization and causal inference.

Conceptualization is the foremost procedure in both inductive and deductive research designs. In online experiments, the conceptualization procedure is similar to conventional research designs. However, it requires much effort in CSS studies with behavioral data. For instance, the following relationships on Facebook and Twitter might be conceptualized in different ways. On Facebook, the following relationship represents a virtual representation of friendship. Users are willing to engage with their 'friends' in various aspects. However, the following relationship is likely to represent a relationship of information seeking on Twitter where users follow other users for informational purposes. It is of vital importance for researchers to understand if the behavioral data collected from social media platforms can really measure what theoretical variables are conceptualized to be.

Due to cultural diversity in Asia-Pacific societies, researchers need to pay special attention to conceptualization of emerging phenomenon on social media platforms. The same or similar functions on social media could be perceived differently by ordinary users with different cultural background. For example, swearing comments on Chinese social media may not necessarily be associated with incivility, even in political contexts. Some previously well-known swearwords are considered appropriate and acceptable in online conversations of Chinese social media. Mutual following relationships on Chinese Weibo may be a more valid indicator of social capital than friendship. Privacy and self-disclosure behavior on social media platforms are also found to be culturally specific (Liang, Shen, & Fu, 2017).

Causal inference is a common goal purused by many, if not all, empirical studies. Generally, it is quite difficult to make such causal inferences in CSS research with behavioral data than online experiments. In online experiments, researchers have greater flexibility to develop valid instruments to measure the causes and consequences. The popularity of social media platforms provides new opportunities for online experiments. On one hand, researchers could conduct digital field experiments in a real-world context, which can improve the external validity of controlled lab experiments. On the other hand, it is straightforward to incorporate big data analytics into online field experiments. Muise and Pan in this special issue have provided a detail discussion on these issues.

Although causal inference in CSS research with behavioral data is not an easy task, some solutions have been developed and employed to make causal inference using observational behavioral data. First, benefiting from the always-on data streams online, natural experimental design is possible by finding random (or as if random) events (Salganik, 2018). A commonly used technique is an interrupted time series design (e.g. Hobbs & Roberts, 2018). Second, other causal inferences techniques are also useful for big data analysis too, such as matching, instrumental variables, and panel designs. These solutions, which have been rarely employed in existing communication studies, offer great potential for communication researchers to make casual inferences within the CSS paradigm.

Data gathering

In some Asia-Pacific societies (e.g. China, Cambodia, and Myanmar), systematic data collection using surveys or other conventional methods are usually difficult. Computational methods provide new opportunities to collect big data and conduct digital field experiments. Data can be collected both passively (data archives) and actively (digital field experiments). Interested readers can refer to Zhu et al. in the special issue for detailed description of four major ways in collecting passive data and refer to Muise and Pan in the special issue for a detailed description of collecting data in field experiments.

Data analytics & evaluation

There are unique technical challenges in big data analytics, such as data storage, retrieving, and parallel computing. However, these technical challenges are less relevant to communication scholars. Usually, it is not the primary research objective of communication researchers to develop new algorithms or models. Instead, it is critically important for communication scholars to consider the selection, interpretation, evaluation of computational models.

Computational models could be existing statistical models in social sciences, such as time series analysis, regression models, and social network analysis. Computational models can also be recently developed models in computer science, such as natural language processing and machine learning models, which are quite new to communication scholars but have been proven very useful in communication studies.

Two strategies are available to incorporate computational methods in communication studies. The first strategy is to use computational methods to develop new measurements of theoretical concepts in communication research. For example, text mining and machine learning have been used to classify communication documents into different categories, whose rationale is similar to the traditional content analysis. Beyond classification, some innovative measures could be developed, such as information similarity and redundancy (Liang & Fu, 2017, 2019). Moreover, social network analysis provides a set of network metrics, such as centrality, clustering, modularity, for communication studies.

The second strategy is to use computational methods for analyzing relationships among theoretical variables. Computational methods could provide unique models for communication studies to test existing theories and propose new theories. For example, social network analysis provides a set of statistical models to test network formation and network effects. Exponential random graph models are very popular in explaining the formation of communication networks (e.g. Peng, Liu, Wu, & Liu, 2016). Stochastic actor-oriented models are also used to test the coevolution of network and behaviors (e.g. Liang, 2014). Although text mining is heavily used for constructing new measures, formal explanatory models have been developed for hypothesis testing. The structured topic modeling has been incorporating covariates in predicting topic prevalence. Fong and Grimmer (2016) developed a new method to test text effect by combing text mining and experimental design. Combining text mining with social network models can produce some unexpected findings (e.g. Liang, 2014).

Interpretation of the modeling outputs is more challenging than the model selection. It requires researchers to understand both their research questions and the rationale of the selected models. It is a central question to consider whether the outputs produced by computational models provide supporting or falsifying evidence for research questions proposed in a study. For example, cosine similarity between documents in text mining can provide empirical support for the similarity between documents under study, while they cannot provide any evidence about the similarity between frames or opinions mined from the documents.

As discussed earlier, computational methods have inherent disadvantages that must be handled carefully and cautiously. First, passively collected data online are not representative of a population in a society. It is usually inappropriate to use this kind of data for descriptive purposes. However, it still possible to collected simple random samples on Twitter and Weibo. Even when drawing a random sample is unlikely, the heterogeneity included in big data observations can be maximized. In order to investigate the censorship in China, King, Pan, and Roberts (2013) collected the data from various websites systematically. The diversity of the data can also help researchers to infer population by matching or weighting the data properly. In addition, non-probability samples are usually good enough for causal inferences in specific research designs (e.g. digital field experiments), which are less vulnerable to the sampling bias.

Second, cross-validation and robustness check are important practices in the assessment the outputs of computational models. There are different data sources and equivalent computational models that can be employed to address a research question. It is very important to empirically assess if the outputs of a CSS study are not methodological artefact caused by the data and the model employed in a study. To minimize such cherry-picking bias in CSS research, it is critically important to cross-validate the outputs with out-of-sample data and empirically evaluate if the outputs are robust across multiple equivalent models and across different sources of data.

Interdisciplinary research

As CSS research involves many techniques developed in computer science and many other disciplines, it is always desirable to form an interdisciplinary team in order to conduct a sound CSS study. A common collaboration scenario between computer scientists and communication scholars is that communication scholars produce research ideas and computer scientists help collect and analyze the data. However, this way of collaboration is not mutually beneficial and usually unnecessary. Most communication studies only involve minimal computing skills. There are many commercial and user-friendly software that can collect social media data automatically. If computational algorithms are really needed, such as machine learning and clustering algorithms, most standard algorithms have been well incorporated into standard statistics or data mining packages in R or Python. It is much more convenient for communication scholars to learn some basic skills to use the software or packages than seeking help from their colleagues in other disciplines.

Besides the technical issues, there are fundamental differences between communication or other social science disciplines and computer science. Social questions are not necessarily social science questions. It is similar to that engineering questions are not necessarily science questions. For example, monitoring online public opinion is a social question that is important for policymakers. However, monitoring public opinion *per se* might not be sufficient to be a social science question. In communication studies, monitoring is the very first step of data collection and measurement. The more critical and challenging steps in the next is to connect public opinion with other theoretically meaningful variables, such as demographics, media use, and participation. In this case, monitoring is akin to an engineering question: its aim is to solve some concrete problems.

It does not mean that there is no meaningful way to collaborate with scholars in other disciplines (e.g. computer science). First, both computer scientists and communication researchers need to learn from one another about their research traditions and major research paradigms. They need to understand each other's strength, uniqueness, and needs. Second, we need to understand that both engineering and science questions are important for society. For computational communication studies, its primary purpose is to answer more fundamental and theoretical questions, which usually involves advanced engineering questions. In this process, communication scholars can contribute to the design of the whole study, while computer scientists can play an important role in solving relevant engineering questions.

Summary of the special issue

The special issue covers the major CSS approaches that are of particular relevance to Asia-Pacific communication research. Zhu et al. argue that user analytics is not a single method but rather a set of diverse methods for studies of use and effects of various media. Based on a review of user analytics research history evolving from TV ratings data to online user logs, the authors discuss what data sources, measures, and analyses can be used for what types of user analytics questions. To help explain the strengths and weaknesses of user analytics, they compare it with both traditional method (e.g. survey) and other computational methods (content mining and online experiment). A case study of Chinese blog users is presented to further illustrate how to apply user analytics for both theoretical and practical purposes.

Muise and Pan discuss the opportunities and challenges of online field experiments for communication studies in Asia. Communication studies in Asian societies face both conceptual and methodological obstacles. Macro-level data are lacking and micro-level data collection is difficult. The popularity of social media and other information communication technologies provides new research opportunities facilitated by CSS methods. Online field experiment is one such promising approach to studying psychological states and behavioral responses of individuals, organizations, and governments. It can uncover previously unobservable phenomenon like censorship and make causal inference possible. In the paper, the authors also outline some methodological considerations in the design of online field experiments: ethical and legal considerations, construct validity, randomization and spillover effect, statistical significance, and external validity. The paper provides a useful guideline for communication scholars interested in online field experiments.

Using Twitter data collected from Australia and Korea, Ackland, O'Neil, and Park compare the patterns of news consumption and engagement across political affiliations and cultures. The paper demonstrates how CSS methods and global social media data could advance cross-national comparative studies. Moreover, it illustrates how basic statistics, such as correspondence analysis and information entropy, can be effectively and innovatively applied to computational communication research.

Jaidka, Ahmed, Skoric, and Hilbert compare election prediction models using Twitter data in three Asian countries: Malaysia, India, and Pakistan. Although cross-country comparative studies are common in communication studies, it is rare to compare different approaches at the same time. The study adopts a systematically comparative design to address a popular research question: election prediction using social media data. Methodologically, it combines text mining (sentiment analysis in particular) with social network analysis and integrates them into a prediction framework (machine learning).

Kobayashi, Ogawa, Suzuki, and Yamamoto show how social scientists and computer scientist can work together meaningfully to address a social science question. The study combines big data with survey data to test the generalizability of a well-known communication problem – news audience fragmentation – in Japan. Instead of using unsupervised machine learning to classify users' political ideologies, the study collects self-reported political ideologies through an online survey, which is used as the ground truth to train machine-learning models. In addition, both structural and semantic features are included in the model.

To conclude, communication researchers in Asia-Pacific societies should and can embrace CSS as an emerging research paradigm in social science. This special issue shows that computational thinking and methods can offer communication researchers enriched opportunities to explore and can stimulate communication researchers to think outside the box. This special issue also shows that it is not beyond the reach of communication researchers to incorporate computational thinking and methods in their own research. This special issue also reminds communication researchers in Asia-Pacific societies that some old-school principles in social science research (e.g. measurement validity, causal inference, unit of analysis) still rule in the age of CSS.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Tai-Quan Peng is an Associate Professor at the Department of Communication, Michigan State University. His current research focuses on computational social science, health communication, and mobile analytics.

Hai Liang is an Assistant Professor at the School of Journalism and Communication, Chinese University of Hong Kong. His research interests include computational social science, political communication, and public health.

Jonathan J. H. Zhu is a Chair Professor of Computational Social Science at City University of Hong Kong with a joint-appointment between Department of Media & Communication and School of

Data Science. His current research focuses on the structure, content, use, and effects of social and mobile media.

References

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*, 295–298. doi:10.1038/nature11421

Fong, C., & Grimmer, J. (2016). Discovery of treatments from text corpora. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)* (Vol. 1, pp. 1600–1609). Berlin: Associations for Computational Linguistics.

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*(1), 129–152. doi:10.1146/annurev-soc-071913-043145

Hobbs, W. R., & Roberts, M. E. (2018). How sudden censorship can increase access to information. *American Political Science Review*, *112*(3), 621–636. doi:10.1017/S0003055418000084

Hutchinson, J. W., Kamakura, W. A., & Lynch, J. J. G. (2000). Unobserved heterogeneity as an alternative explanation for "reversal" effects in behavioral research. *Journal of Consumer Research*, *27*(3), 324–344. doi:10.1086/317588

Kaplan, R. M., & Stone, A. A. (2013). Bringing the laboratory and clinic to the community: Mobile technologies for health promotion and disease prevention. *Annual Review of Psychology*, *64*(1), 471–498. doi:10.1146/annurev-psych-113011-143736

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343. doi:10.1017/S0003055413000014

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. doi:10.1073/pnas.1320040111

Langley, A., Smallman, C., Tsoukas, H., & Van de Ven, A. H. (2013). Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal*, *56*(1), 1–13. doi:10.5465/amj.2013.4001

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

Liang, H. (2014). Coevolution of political discussion and common ground in web discussion forum. *Social Science Computer Review*, *32*(2), 155–169. doi:10.1177/0894439313506844

Liang, H., & Fu, K. W. (2017). Information similarity, overload, and redundancy: Unsubscribing information sources on Twitter. *Journal of Computer-Mediated Communication*, *22*(1), 1–17. doi:10.1111/jcc4.12178

Liang, H., & Fu, K. W. (2019). Network redundancy and information diffusion: The impacts of information redundancy, similarity, and tie strength. *Communication Research*, *46*(2), 250–272. doi:10.1177/0093650216682900

Liang, H., Shen, F., & Fu, K. W. (2017). Privacy protection and self-disclosure across societies: A study of global Twitter users. *New Media & Society*, *19*(9), 1476–1497. doi:10.1177/1461444816642210

Peng, T. Q., Liu, M., Wu, Y., & Liu, S. (2016). Follower-followee network, communication networks, and vote agreement of the U.S. members of congress. *Communication Research*, *43*(7), 996–1024. doi:10.1177/0093650214559601

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton, NJ: Princeton University Press.

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