ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Hutter, Swen

Article — Published Version

Comment: Exploring the Full Conceptual Potential of Protest Event Analysis

Sociological Methodology

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Hutter, Swen (2019) : Comment: Exploring the Full Conceptual Potential of Protest Event Analysis, Sociological Methodology, ISSN 1467-9531, Sage, Thousand Oaks, CA, Vol. 49, Iss. 1, pp. 58-63, https://doi.org/10.1177/0081175019860239

This Version is available at:
https://hdl.handle.net/10419/222603

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Check for updates

Sociological Methodology
2019, Vol. 49(1) 58–82
© American Sociological Association 2019
http://sm.sagepub.com

COMMENT: EXPLORING THE FULL CONCEPTUAL POTENTIAL OF PROTEST EVENT ANALYSIS

Swen Hutter*

*Freie Universität & WZB Berlin Social Science Center, Berlin, Germany
Corresponding Author: Swen Hutter, swen.hutter@fu-berlin.de
DOI: 10.1177/0081175019860239

Zhang and Pan (this volume, pp. 1–57) should be applauded for the huge effort they put into building their new system to systematically detect in social media posts what they call "collective action events," specifically the effort that went into thoroughly cross-checking the validity of their new data set for China by means of comparisons with related Chinese and international data sets. The construction of their tools and data stands out with the careful combination of innovative tools from computational social sciences and extensive manual coding (at several points in the data collection process) and a thorough understanding of the social and political dynamics at work. Overall, the article makes an important contribution to protest and social movement studies as well as to methodological discussions in computational social sciences. Zhang and Pan have given the research community at least three gifts: (1) a two-stage classifier of text and images, (2) a large data set on offline protests reported online and taking place in China from 2010 to 2017, and (3) a very accessible text that explains the data collection steps as well as the chosen approach's potential pros and cons in an informative and balanced way. Therefore, this text is not only helpful for researchers who want to further push the frontiers of computational methods forward and build on Zhang and Pan's pipeline but also for a broader audience interested in protest mobilization and resistance in China and beyond.

As others are much better equipped to comment on the technical side of collective action from social media (CASM), I focus my short contribution on the conceptual side and potential avenues for further explorations of the data set and the tools Zhang and Pan provide. First, I sketch the history of protest event analysis (PEA) in social movement studies. The brief summary serves as a background in which I discuss the strengths and weak(er) spots of Zhang and Pan's work and ways to improve it. Specifically, I discuss the conceptualization of the coding unit (collective action events), the identification of duplicates, and relations between online and offline dynamics.

PEA has become a key method of social movement research over the past decades. The method gained significant ground in the 1980s and early 1990s. As Koopmans and Rucht (2002) stated some time ago, "PEA provides a solid ground in an area that is still often marked by more or less informed speculation" (p. 252). In bold strokes, I have identified four generations of PEA research (for details, see Hutter 2014).[1] The first generation, the initiators, consisted of researchers who were interested in various indicators for a large number of countries (e.g., the *World Handbook of Political and Social Indicators* by Taylor and Hudson 1972) or in long-term processes of social and political change (e.g., Tilly, Tilly, and Tilly 1975). For our context, it is important that these authors did not yet pay much attention to their sources' biases or the creation of fine-grained categories and well-documented procedures. These shortcomings were then addressed by a second generation that made more extensive use of protest event data by breaking it down according to various analytic criteria, which was possible because the categories used for data collection were far more sophisticated. Exemplary studies are McAdam's (1982) work on civil rights protests in the United States, Tarrow's (1989) work on the Italian protest cycle in the late 1960s and early 1970s, and Kriesi et al.'s (1995) book on new social movements in four Western European countries.

Although this second generation was more sophisticated in its coding procedures and source selection, the authors did not invest much in qualifying their sources' bias. A third generation assessed newspaper data's bias more systematically. Importantly, these authors focused on selection bias, specifically the fact that newspapers selectively report on protest events and do not provide a representative sample of all events taking place (for a review, see Earl et al. 2004). Among the third generation were scholars who tried to be more efficient by using automated approaches to select and precode protest events.[2] Unfortunately, the latter tended to fall back on the first generation's research in the selection of sources and coding procedures or their use for comparative research.

Finally, a fourth generation has developed recently. Like the second generation, its members are concerned with further expanding the usefulness of PEA, this time by expanding the coding unit beyond a strict focus on (aggregates of) protest events. On the one hand, scholars have unpacked single protest events or performances by focusing on action and interaction inside them (e.g., Tilly 2008). On the other hand, scholars broadened the unit of analysis to cover a larger group of public actions (including protest events). Examples are Koopmans and Statham's (1999) political-claims analysis or our recent suggestion of contentious episode analysis (Kriesi, Hutter, and Bojar 2019). These approaches share an attempt to capture the *relational* aspect of political contention better than traditional PEA and to collect data on covariates to move beyond the identification and mapping of protests to explanatory analysis.

The short history of PEA in social movement studies should highlight the good company of CASM and the considerable efforts that went into further improving application of the research technique by evaluating sources' biases and refining and broadening the coding unit. I say this because I think Zhang and Pan's work is focused on taking on broad aspects of selection and description bias, in particular by raising awareness of how technological innovations may affect the standing and usage of the selected source (Sina Weibo), how the content might be constrained by technical aspects (such as the limited number of characters) and social ones (such as repression and censorship), and by the extensive cross-validation of their new data with other existing data sets. In this respect, Zhang and Pan's CASM does not repeat the mistakes of early approaches to automatize PEA. At the same time, I think their work could profit from more thoroughly considering the conceptual advances in PEA over the past decades in the refining of subaspects of protest events and the expansion of coding units.

First, I think that by calling their coding unit a "collective action event" and not simply a "protest event," Zhang and Pan made it somewhat harder for themselves and their readers than necessary. Ultimately, one sees clearly from the third feature of their definition ("contentious event with a public physical presence"; p. 8) and the identified action categories (see p. 31) that the contents of their data set are no more extensive than what classical PEA covers. With their focus on collective action (operationalized as at least three participants) and physical presence, Zhang and Pan come closest to Tilly's early work and are even more restricted than what the second and third generations covered (which often also considered petitions or other actions that do not involve physical copresence). The notion of a "collective action event" invokes in my understanding a much broader repertoire of activities and may be misleading, particularly in the context of social media and Internet activism.

Having said this, if one sticks to classical protest event definitions, Zhang and Pan would be well advised to present earlier a complete operational definition of which specific action forms they cover in their research.[3]

Second, Zhang and Pan could profit from investing more in conceptualizing and measuring the subdimensions of protest events as the second PEA generation did. I am concerned that by not placing dimensions such as the action form, claims and issues, targets, or organizers of protests center stage throughout their work, they miss some opportunities to clarify their approach (what is covered?), increase its accuracy (e.g., what is a duplicate?), and for cross-validation (how large are the differences due to different sources or coding units?). To illustrate, by taking the dimension of action forms or issues more seriously throughout their work, Zhang and Pan would realize that the identification of duplicates would not just be better if one also considers these elements, but the conceptualization of their coding unit makes it essential that these elements are included (as people with different, at times even opposing, claims might be on the streets on the same day or stage different activities on the same day). To be fair, Zhang and Pan already hint at these elements in their text, and their analysis is much better than most automated protest event analyses in also automatically coding the forms and issues of protest. My point, however, is that by going back to the history of PEA, researchers may realize how much more they can gain by further investing here.

Finally, Zhang and Pan could easily take up the calls made in the fourth generation of PEA to broaden the coding units because a lot of what they aim to "eliminate" (e.g., related posts by government officials or posts about the same grievances but without mentions of collective actions) could be understood as social media covariates that might shape or be shaped by the identified protest events. In further work based on their material, for example, researchers could explore the evolution of these types of posts over time or the cross-regional variation in related posts with and without offline protest actions. Most important, in my opinion, broadening the perspective regarding the coding unit could help mitigate a main problem of the selected source: its varying popularity over time and related difficulties in inferring from the data the "real" ups and downs in protest mobilization in China. I think Zhang and Pan and CASM have much more to offer than worrying about this aspect. For example, I would be very interested in further explorations of what happens in the "online" world, for instance, for what types of protests people are more likely to use pictures with text to circumvent potential censorship and how that practice has potentially changed over the research period. What types of events get the most resonance in social media, in the sense of being posted more than once? Or how does the general social media activity of individuals who post "offline" protests differ

from the platform's average users or depend on the event types being posted?

As we do with any great invention, one is left wanting more of it. This is how my comments and questions based on the history of PEA in social movement studies should be understood. Together with other scholars who develop automated systems to accurately code protest events from social media and traditional news sources, Zhang and Pan are already shaping a new fifth generation of innovations in PEA.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author gratefully acknowledges funding from the Volkswagen Foundation (Lichtenberg grant).

Notes

1. On the development of PEA, see also Davenport (2009:25*ff*), Koopmans and Rucht (2002:232*ff*), and Tilly (2008:19*ff*).
2. The most prominent examples of the first half-automated procedures in social movement research are the European protest and coercion data collected by Francisco and colleagues (e.g., Francisco 1996) and the study on Europeanized protests by Imig and Tarrow (2001).
3. For example, the list on p. 8 contains boycotts, demonstrations, marches, sit-ins, and strikes. One is left wondering if this list is exhaustive or presents only (the most common?) examples.

References

Davenport, Christian. 2009. *Media Bias, Perspective, and State Repression*. Cambridge, UK: Cambridge University Press.

Earl, Jennifer, Andrew Martin, John D. McCarthy, and Sarah A. Soule. 2004. "The Use of Newspaper Data in the Study of Collective Action." *Annual Review of Sociology* 30:65–80.

Francisco, Ron. 1996. "Coercion and Protest: An Empirical Test in Two Democratic States." *American Journal of Political Science* 40(4):1179–1204.

Hutter, Swen. 2014. "Protest Event Analysis and Its Offspring." Pp. 335–67 in *Methodological Practices in Social Movement Research*, edited by D. della Porta. Oxford, UK: Oxford University Press.

Imig, Doug, and Sidney Tarrow. 2001. "Mapping the Europeanization of Contention: Evidence from a Quantitative Data Analysis." Pp. 27–49 in *Contentious Europeans: Protest and Politics in an Emerging Polity*, edited by D. Imig and S. Tarrow. Lanham, MD: Rowman & Littlefield.

Koopmans, Ruud, and Dieter Rucht. 2002. "Protest Event Analysis." Pp. 231–59 in *Methods of Social Movement Research*, edited by B. Klandermans and S. Staggenborg. Minneapolis: University of Minnesota Press.

Koopmans, Ruud, and Paul Statham. 1999. "Political Claims Analysis: Integrating Protest Event and Political Discourse Approaches." *Mobilization* 4(2):203–21.

Kriesi, Hanspeter, Swen Hutter, and Abel Bojar. 2019. "Contentious Episode Analysis." *Mobilization* 24(3).

Kriesi, Hanspeter, Ruud Koopmans, Jan Willem Duyvendak, and Marco Giugni. 1995. *New Social Movements in Western Europe*. Minneapolis: University of Minnesota Press.

McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency 1930–1970*. Chicago: University of Chicago Press.

Tarrow, Sidney. 1989. *Democracy and Disorder: Protest and Politics in Italy 1965–1974*. Oxford, UK: Oxford University Press.

Taylor, Charles L., and Michael C. Hudson. 1972. *World Handbook of Political and Social Indicators*. New Haven, CT: Yale University Press.

Tilly, Charles. 2008. *Contentious Performances*. Cambridge, UK: Cambridge University Press.

Tilly, Charles, Louise Tilly, and Richard Tilly. 1975. *The Rebellious Century, 1830–1930*. Cambridge, MA: Harvard University Press.

Author Biography

Swen Hutter is Lichtenberg Professor in Political Sociology at Freie Universität Berlin and vice director of the Center for Civil Society Research, a joint initiative of Freie Universität and WZB Berlin Social Science Center.