Journal of Business Research 117 (2020) 322–330

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

TUTORIAL: AI research without coding: The art of fighting without fighting: Data science for qualitative researchers

Leon Ciechanowski[a], Dariusz Jemielniak[a,*], Peter A. Gloor[b]

[a] Kozminski University, Management in Networked and Digital Societies (MINDS) Department, Jagiellonska 59, 03-301 Warszawa, Poland
[b] Massachusetts Institute of Technology, Center for Collective Intelligence, 245 First Street, E94, Cambridge, MA 02142, USA

ARTICLE INFO

Keywords:
Twitter
Data scraping
Sentiment analysis
Tribe finding
Wikidata

ABSTRACT

In this tutorial, we show how to scrape and collect online data, perform sentiment analysis, social network analysis, tribe finding, and Wikidata cross-checks, all without using a single line of programming code. In a step-by-step example, we use self-collected data to perform several analyses of the glass ceiling. Our tutorial can serve as a standalone introduction to data science for qualitative researchers and business researchers, who have avoided learning to program. It should also be useful for experienced data scientists who want to learn about the tools that will allow them to collect and analyze data more easily and effectively.

1. Introduction

In the 1973 classic, *Enter the Dragon*, Bruce Lee plays a formidable Shaolin kung-fu master, who describes his style as "the art of fighting without fighting," and defeats an adversary by tricking him into entering a dinghy to go to an isolated island for a duel, only to leave him adrift behind a larger boat. We will use this philosophy in our tutorial. Even though some of us are proficient in coding, and we prepared a standalone software program for the readers of this issue, we will also show how to use free tools to collect and analyze data without needing to know any programming languages.

The social sciences in general and business studies in particular have undergone revolutionary changes because of the rapid growth of ICT platforms (Caputo & Walletzký, 2017) and big data (Lazer & Radford, 2017). Troves of data resulting from a vast expansion of social media (Kaplan, 2015) make it possible to answer questions previously beyond the reach of science and make the performance of digital analysis a necessity for most social studies. Publicly accessible open data allows the granular study of collaborative software development (Chełkowski, Gloor, & Jemielniak, 2016), the discovery of cultural differences in the perceptions of what constitutes knowledge (Jemielniak & Wilamowski, 2017), the identification of top performers through email pattern analysis (Wen, Gloor, Fronzetti Colladon, Tickoo, & Joshi, 2019), and even the prediction of crude oil prices (Elshendy, Colladon, Battistoni, & Gloor, 2018).

Tapping into the treasure chest of digital datasets is important, as online data and social media not only reflect the social imaginary but also have a significant impact on people's beliefs and behaviors, whether in politics (Chmielewska-Szlajfer, 2018), personal relationships (Das & Hodkinson, 2019), local activism (Stasik, 2018), smart cities management (Caputo, Walletzky, et al., 2019), organizational performance (Malone, 2018), or medical knowledge (Smith & Graham, 2019). It is difficult to imagine a sphere of life that does not have some important digital component that cannot be studied by the use of online data. Social media, digital work and digital private life have revolutionized the way we function in business and society (Jemielniak & Przegalińska, 2020).

It is not a surprise that we can observe an advancing datafication of social sciences (Millington & Millington, 2015), and of management, business, and marketing (Erevelles, Fukawa, & Swayne, 2016; Vanhala et al., 2020; Wamba et al., 2017). The data, however, never speak for themselves (Dourish & Cruz, 2018), and applying Big Data methods poses many challenges (Sivarajah, Kamal, Irani, & Weerakkody, 2017). The question about the required conditions for effective Big Data management and research remains open and timely (Caputo, Mazzoleni, et al., 2019). Interpreting big datasets requires contextualization and reflection, typical for qualitative research, but interpretive studies alone are becoming less insightful in the shadow of big datasets.

In fact, drawing from the vast resources of online data is becoming more useful than ever for qualitative researchers. Qualitative research has a long tradition in business studies, dating back to the early 20th

* Corresponding author at: Kozminski University, Management in Networked and Digital Societies (MINDS) Department, Jagiellonska 59, 03-301 Warszawa, Poland.
E-mail addresses: darekj@kozminski.edu.pl, dariuszj@mit.edu (D. Jemielniak).

https://doi.org/10.1016/j.jbusres.2020.06.012
Received 17 January 2020; Received in revised form 3 June 2020; Accepted 5 June 2020

century and the famous Elton Mayo experiments (Mahoney & Baker, 2002). Qualitative studies have the unique advantage of getting insight into the internal logic of a group, organization or culture (Ciesielska & Jemielniak, 2018a, 2018b). In short, while quantitative studies can plausibly show what is happening and seek correlations and measurable effects, qualitative studies are better at dealing with social backgrounds and explaining their reasons and rationales (Ravitch & Carl, 2019).

Even though qualitative studies in business research were treated with suspicion in the past, they have found an equal place in the methodological repertoire of management scholars (Eriksson & Kovalainen, 2015; Ghauri, Grønhaug, & Strange, 2020; Myers, 2019). However, the weaknesses of qualitative studies are also widely recognized; these include questions of reliability, validity, generalizability, and objectivity (Sinkovics, Penz, & Ghauri, 2008; Symon & Cassell, 2012). They also rely on the researcher's own interpretation and processing of fieldwork material (Queirós, Faria, & Almeida, 2017). Grounding and framing the scope of the study then is inherently subjective. For this reason, taking advantage of quantitative approaches to inform qualitative studies has been advocated as particularly useful (Creswell & Plano Clark, 2017; Ivankova & Wingo, 2018)

2. Aim and objectives: Big data meets thick data

Although the advantages of using mixed methods for business research have already been recognized (Harrison, 2013), and qualitative-quantitative crossovers are known to bring many benefits (Polsa, 2013), allowing qualitative researchers to perform pilot studies and reconnaissance by the use of data science allows the quick mapping of areas that merit deeper, long-term analysis, to frame the qualitative part more accurately. This is why combining big data with thick data, or a *Thich Big Data* approach, make so much sense (Charles & Gherman, 2019; Jemielniak, 2020), and why even diehard non-coding researchers could benefit from the use of data science and machine learning.

Luckily, these fields are rapidly changing (Haenlein & Kaplan, 2019). Many tasks that once required advanced programming skills now do not. However, knowledge about the novel and evolving tools is not readily available.

In this paper we introduce several powerful tools for data analysis to researchers, especially qualitative researchers, who have no experience in collecting large digital datasets. In this tutorial, we show how someone with no knowledge of coding can conduct research with the use of data science and AI, and then, by acquiring primary data, processing and analyzing it. In the spirit of open science, aimed at eliminating financial and knowledge-access barriers in academia (Vicente-Saez & Martinez-Fuentes, 2018), we are going to demonstrate five powerful and free tools. Together, they give business research scholars a strong arsenal. We believe that just by relying on these tools, academics who so far have not been doing digital quantitative studies, should be able to embark on their data science adventure. Creating such tutorials and beginner's guides is crucial for advancing business research (Haenlein & Kaplan, 2004; Watson, 2014). It is also an important step in the direction of closing the gap between qualitative and quantitative researchers.

3. Theoretical framework

In the tutorial—as an example of our methodology—we will collect data on the glass ceiling, a topic of vital importance for research in social media and business. It is still not fully understood in the context of rapid digital transformations, and has a multiplier effect on many other social processes online (Humprecht & Esser, 2017; Jemielniak, 2016; Yarchi & Samuel-Azran, 2018).

In the following section, we present an accessible way to scrape (download) tweets containing the phrase "glass ceiling." Then, we perform sentiment analysis (assessment of emotional content) of the tweets, analyze their "tribes" and study them by collecting information about the most influential authors with Wikidata tools.

Sentiment analysis or opinion mining is a method of automatic assessment of the emotional charge of text (e.g., tweets) based on algorithms using Artificial Intelligence (Liu & Zhang, 2012). Its task is to classify emotionally charged texts, both those that indicate the author's emotional state and those that may indicate the emotional effect of the text on the recipient. It can be used to analyze people's opinions, feelings or attitudes about products, services, organizations, events or individuals; this makes it potentially useful in the social and economic sciences (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011; Liu, 2012; Medhat, Hassan, & Korashy, 2014) in addition to business (Aldahawi & Allen, 2013; Saura, Palos-Sanchez, & Grilo, 2019; Wang & Zhou, 2009). Sentiment analysis tools are built using neural networks, which is a part of AI, especially through various convolutional and recurrent neural networks (Ain et al., 2017).

In sum, our theoretical framework relies on a methodology of exploring "glass ceiling" tweets and encompasses the following steps:

1. Retrieval of tweets with Twitter Archiver
2. Sentiment extraction with MeaningCloud
3. Twitter and sentiment analysis using our original and free software FWF: Fighting Without Fighting:
   a. Sentiment analysis
   b. Tags Cloud generation
4. Wikidata scraping with Wikipedia and Wikidata Tools
5. Tribe analysis with TribeFinder

4. Methodology: A complex quantitative exploration of a research problem without complex tools

1. Tweet retrieval with Twitter Archiver

We collected tweets containing the phrase "glass ceiling" using Twitter Archiver, a Google Sheets add-on, which even in its free version is fully functional. The setup is user-friendly: after registering and authorizing our Twitter account (we recommend using a prop one for data harvesting) it is enough to type in the requested parameters in the form (Fig. 1).

After approving the choice of topics ("update search rule") the script starts collecting tweets directly into Google Sheets and keeps adding newly found tweets every hour until turned off.

From 26 October through 16 December 2019, we collected 6651 tweets containing the exact phrase "glass ceiling." The database that was generated contains well-structured information on the tweets with that phrase, including date, screen name, full name, tweet text, tweet ID (linking to the original tweet), link(s) from the tweet, media, location, the number of retweets, the number of likes, app used, the number of followers, or the number of accounts followed. There are also several other useful cells, such as informing whether the account is verified, the user location and time zone, the date of registering the account, or a link to the user's website and a profile image.

The Google Sheet we were working on is available at bit.ly/GlassCeilingTweets. We added the collected and processed data in separate sheets.

Of the tweets, 2757 received a retweet or a like; 1431 were retweeted, 1252 were both retweeted and liked at least once; One hundred twenty six were retweeted at least 10 times, and 389 were liked at least 10 times. The most-often liked tweet (1042 likes) was the following, supporting the breaking of the glass ceiling, authored by @jtLOL (Fig. 2).

The second most-often liked tweet (731 likes) ridiculed Hillary Clinton, authored by @DFBHarvard (Fig. 3).

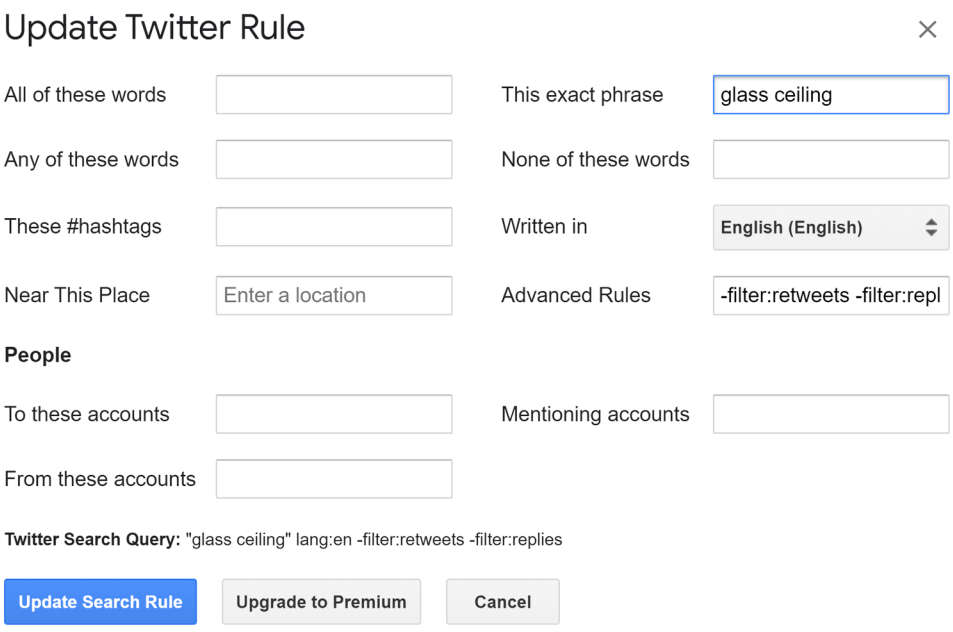


Fig. 1. Twitter Archiver main window.

Fig. 2. The most-often liked tweet about the glass ceiling.

Among the 10 most-liked tweets, two criticized Hillary Clinton. The remaining eight supported gender equality and breaking the glass ceiling.

2. Sentiment extraction with MeaningCloud

For the purpose of sentiment extraction from the collected tweets, we used MeaningCloud (meaningcloud.com), which also offers a fully functional free Google Sheets add-on (Fig. 4). Processing more than 6000 tweets was not possible in one go: the add-on got stuck a couple of times and we had to restart it on the remaining cell range and combine the results manually. Nevertheless, the process was easy, and batches of 400–500 tweets were analyzed.

MeaningCloud attempts to determine whether the analyzed content was very positive, positive, neutral, negative, very negative, or undetermined. It also estimates whether the content expresses agreement or disagreement, whether it is objective or subjective, ironic or non-ironic, and gives algorithmic confidence in the estimation (Fig. 5).

From our database, 2038 tweets had content classified as negative and 222 as very negative. In addition, 2176 were positive and 382 very positive. Surprisingly, positive comments noticeably outnumbered negative ones, and very positive ones significantly outnumbered very negative ones. All results are added as a separate sheet in our demo link.

3. Twitter and Sentiment analysis using our software

After using Twitter Archiver and MeaningCloud, we ended up with a database that can be easily analyzed with the usage of any statistical program like IBM SPSS, Statistica, or free and open-source ones like Jamovi or Jasp. One can also use a Python-based program FWF: Fighting Without Fighting[1] that we have developed for this paper, for readers who are not well versed in programming or inferential statistics. The program functionalities allow for automatic statistical analyses and visualization of the data generated with Twitter Archiver and MeaningCloud. The program is available for free unrestricted use, as a

Fig. 3. The second most-often liked tweet.

[1] http://bit.ly/JBRsoftware.

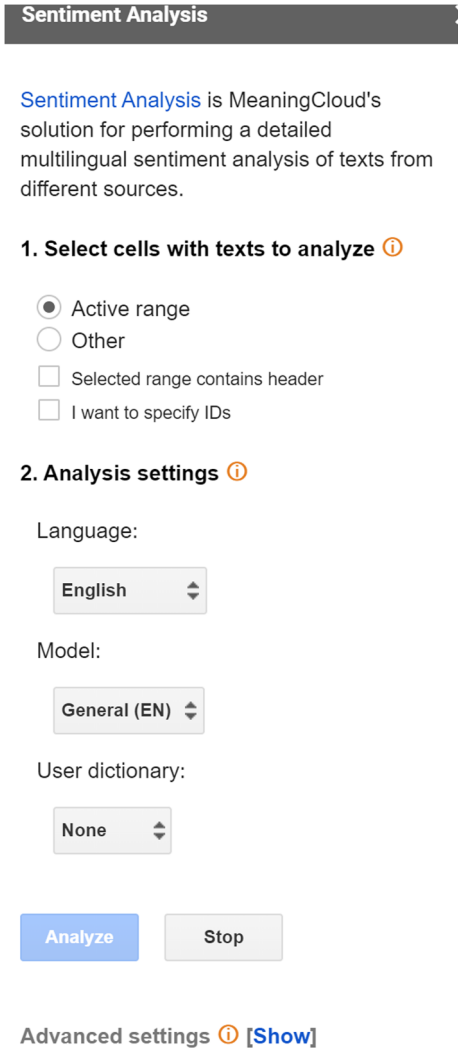


Fig. 4. MeaningCloud, a free Google Sheets add-on for automatic sentiment extraction.

license requirement imposing only the obligation to cite this source article. Below we present an exemplary use of FWF.

We intended our software to be intuitive and self-explanatory (Fig. 6).

After loading the user's file, the dataset can undergo the following analyses.

4.1. Sentiment analysis

Users can pick the desired columns containing data produced by MeaningCloud, preprocessing the data and generating visualizations (Fig. 7) and statistical analyses (Fig. 8). It is worth mentioning that FWF software can analyze and visualize any data relationship chosen by the user, provided that he or she picks the columns containing categorical and continuous values.

We can see that the most positive tweets were most often retweeted. The most negative ones were most rarely retweeted. The program

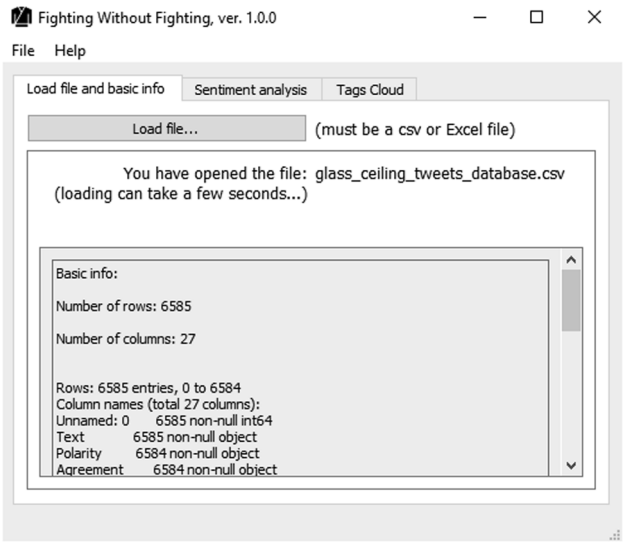


Fig. 6. Main screen from FWF: Fighting Without Fighting after loading an exemplary dataset. Source: http://bit.ly/JBRsoftware.

performs non-parametric analysis, since the distributions of each group of tweets with various sentiments are not normal, and the number of tweets in each group is different. A similar situation will likely take place in most Twitter projects; it is unlikely that researchers will find the same number of tweets in five different sentiments. The Kruskal–Wallis test for independent groups was statistically significant ($H = 26.006$, $p < 0.001$). The post-hoc analysis, with the use of Mann-Whitney test with Bonferroni adjustment for multiple comparisons, revealed that positive (P) and negative (N) tweets are significantly different in the number of retweets ($p < 0.001$), and there was also a statistically significant difference between neutral (NEU), and negative (N) tweets ($p = 0.001$).

4.2. Tags cloud

Users may pick here any column from the data that contains text, and this tab will produce Word Clouds or Tag Clouds (Fig. 9). Word Clouds present the words that appear most frequently in a set of analyzed sentences, tweets and data. The more often the word or phrase was used, the bigger it is displayed in the picture.

We can observe that the most visible difference (except for obvious differences in positive and negative word choice) is the repetition of terms like "Hillary" or "Hillary Clinton." It is, therefore, possible that commenters' negative sentiments were directed more at the person than at the topic of glass ceiling.

4. WikiData check

We also used the Wikipedia and WIkidata Tools Google Sheets add-on to check which of the people whose tweets led to stronger reactions have biographical articles on Wikipedia. It is worth remembering that we were able to only the declared combinations of full first and last names given on Twitter, and not verify the actual identity of a person.

We analyzed a subset of 712 tweets that were retweeted at least twice. We copied a list of full names of authors of these tweets,

Don't be afraid of the glass ceiling. It's just an illusion. We'll guide you through. Click here for more: #bu	P	AGREEMENT	OBJECTIVE	92	NONIRONIC
Shout to @trenchmag for including @ChildrenOfZeus' 'Ghost' (from the 'Excess Baggage' EP) in their r	P+	AGREEMENT	SUBJECTIVE	100	NONIRONIC
How Over-40 Female Writers Are Trying to Shatter Hollywood's Last Glass Ceiling via @thr	NONE	AGREEMENT	OBJECTIVE	100	NONIRONIC
"I can't say how many times I've been up for jobs for books that I should adapt. Just very clearly, 'I'm th	N+	AGREEMENT	OBJECTIVE	92	NONIRONIC
I fall asleep from the sound of rain and the glass ceiling at my job doesn't help	N	AGREEMENT	OBJECTIVE	92	NONIRONIC

Fig. 5. Part of a Google Spreadsheet with a sentiment of each tweet.

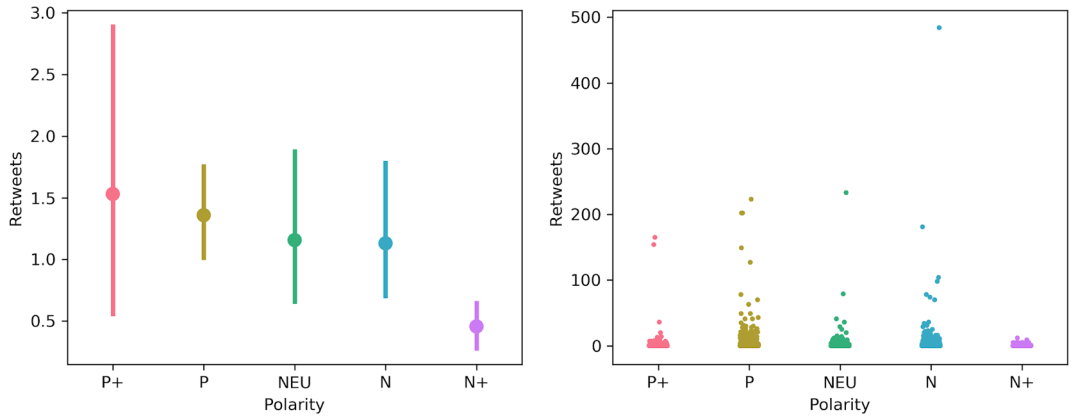
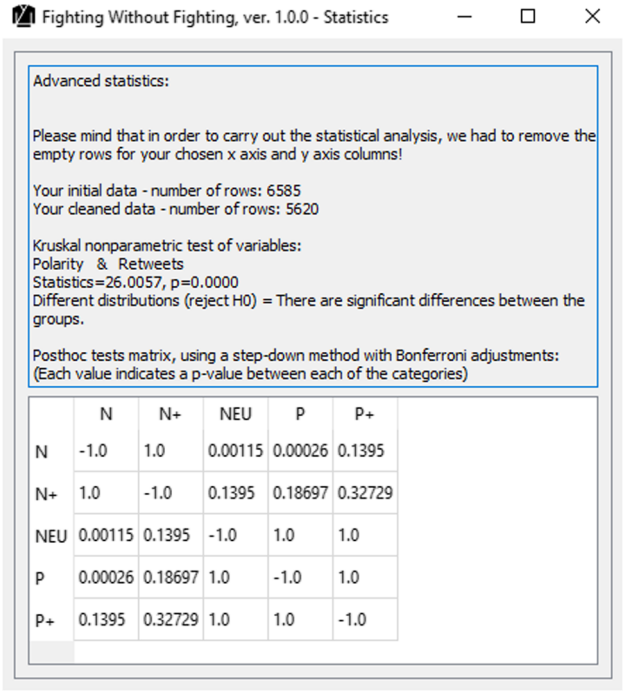


Fig. 7. The figure presents the number of retweets for various sentiments (P + – very positive tweets, P – positive, NEU – neutral, N – negative, N + – very negative). The graph on the left presents means and confidence intervals of the number of retweets. The graph on the right presents all tweets (each dot is one tweet) and their number of retweets, divided into five categories of sentiment.



	N	N+	NEU	P	P+
N	-1.0	1.0	0.00115	0.00026	0.1395
N+	1.0	-1.0	0.1395	0.18697	0.32729
NEU	0.00115	0.1395	-1.0	1.0	1.0
P	0.00026	0.18697	1.0	-1.0	1.0
P+	0.1395	0.32729	1.0	1.0	-1.0

Fig. 8. Statistical analyses of the variables (columns) chosen by the user.

deduplicated the column (as some people had more than one tweet in it), and ended up with 668 unique accounts. We manually removed honorifics, such as "Dr.," so that only first and last names were left in the cells. We used a `=WIKIDATAQID("en:" & A1)` (where A1 was the cell containing the first and last name) formula to check if this combination was covered in English Wikipedia and retrieve its Wikidata identifier. Identifier values were returned for 193 account names, indicating that they are covered. We then used a `=WIKIDATAFACTS(C1, "first," "P21")` (where C1 was the cell containing the Wikidata identifier), to check the gender of the described person or character. Other properties available from Wikidata can be explored at https://tools.wmflabs.org/prop-explorer/. As a result of our query, we found that 61 accounts had their gender described in WikiData. Thirty were described as men and 31 as women, which shows a striking gender balance.

It is worth reiterating that we identified gender for the account names, not for actual people. In some cases, the query worked in an unexpected way. For instance, for "Captain Beefy" the script identified the account as belonging to a man, even though the captain is a fictional character. However, on English Wikipedia, "Captain Beefy" is a redirect to George Lowe, an American voice actor, and comedian which triggered this output. For more qualitative analysis, we also used `=WIKIDATADESCRIPTIONS(WIKIDATAQID(A1), {"en"})` to determine the listed people's profession. This way we established that as many as 20 of the accounts could be attributed to politicians, eight to professional athletes, five to journalists, and others to different professions. We include these results in our demo Google Sheet.

5. Tribe analysis

A tribe in business studies is "a network of heterogeneous persons linked by a shared passion or emotion" (Cova & Cova, 2002, p. 602). Tribe members usually share some beliefs, and thus they can display unique behavioral and communication patterns, which may be studied through big data analysis of internet fora, groups, and social networks

Fig. 9. The figure presents word clouds for positive (P; on the left), and negative (N; on the right) tweets.

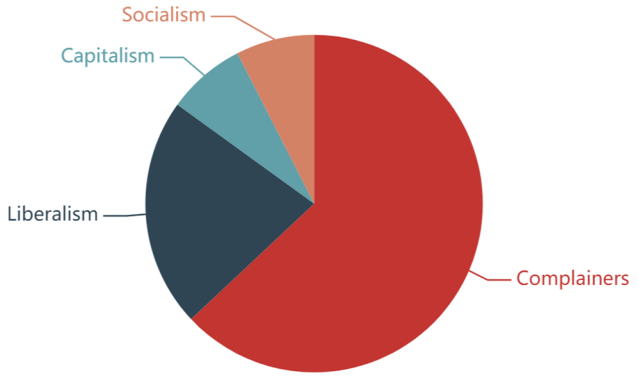


Fig. 10. Tribe analysis of @jtLOL followers – analysis of dominating ideologies.

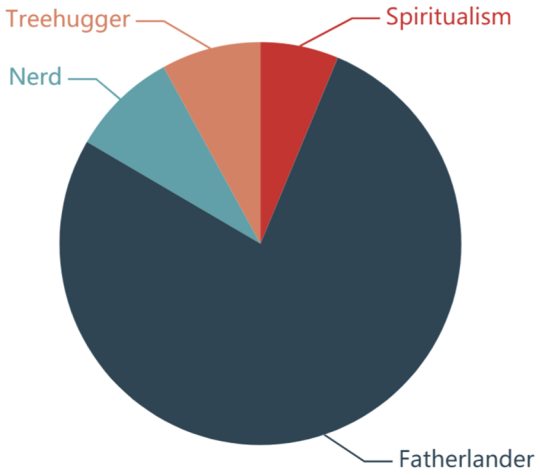


Fig. 11. Tribe analysis of @DFBHarvard followers – analysis of dominating lifestyle.

(De Oliveira & Gloor, 2018). TribeFinding is a new method that one of us developed to approximate users' tribal affiliations by a deep learning analysis of their vocabulary (Gloor, Fronzetti Colladon, de Oliveira, & Rovelli, 2019).

To show a simple application of a TribeFinder tool, we created two tribes, based on the followers of authors of two most-liked tweets from our database: @jtLOL and @DFBHarvard. After registering on https://galaxyscope.galaxyadvisors.com/ and authorizing access with a Twitter account, we were able to point-and-click "Tribe Creator," add the Twitter profiles and do the Follower search, adding 200 recent followers to the tribe. We repeated the procedure for both accounts separately, so as to create two distinct groups.

TribeFinder relies on machine learning categorizations of analyzed text into tribes. These tribes are described by one-word names reflecting a more complex life choice and philosophy (for a list of available classifications, see e.g., Przegalinska, Ciechanowski, Stroz, Gloor, & Mazurek, 2019). What we learned from the analysis of @jtLOL followers was that they are predominantly "Complainers" (people not strongly affiliated with any of the three dominant ideologies) (Fig. 10).

At the same time, we also found that @DFBHarvard followers are predominantly "Fatherlanders" (conservatives, extreme patriots, traditionals, and xenophobes) (Fig. 11).

TribeFinder also allows the easy study of the tribe network, hashtags over time, hashtag clouds, influencer network, or even emotions expressed in tweets over time. It is also very useful for studying hashtag networks interactively. For instance, among @DFBHarvard followers "MAGA" ("Make America Great Again," Donald Trump's 2016 campaign slogan) hashtag dominated with 63 instances. Below you can see the same hashtag network for with a hover on "trump" hashtag (Fig. 12).

It is possible to delete certain nodes, and to search specific hashtags in the network analysis directly from within the system.

We also used the Tribe Comparator tool to compare the tribes across criteria and check the in which they differ most on scales of archetypal typologies. One finding was that the followers of @DFBHarvard (in red) were much less "Sedentary" (people who get less exercise), but also

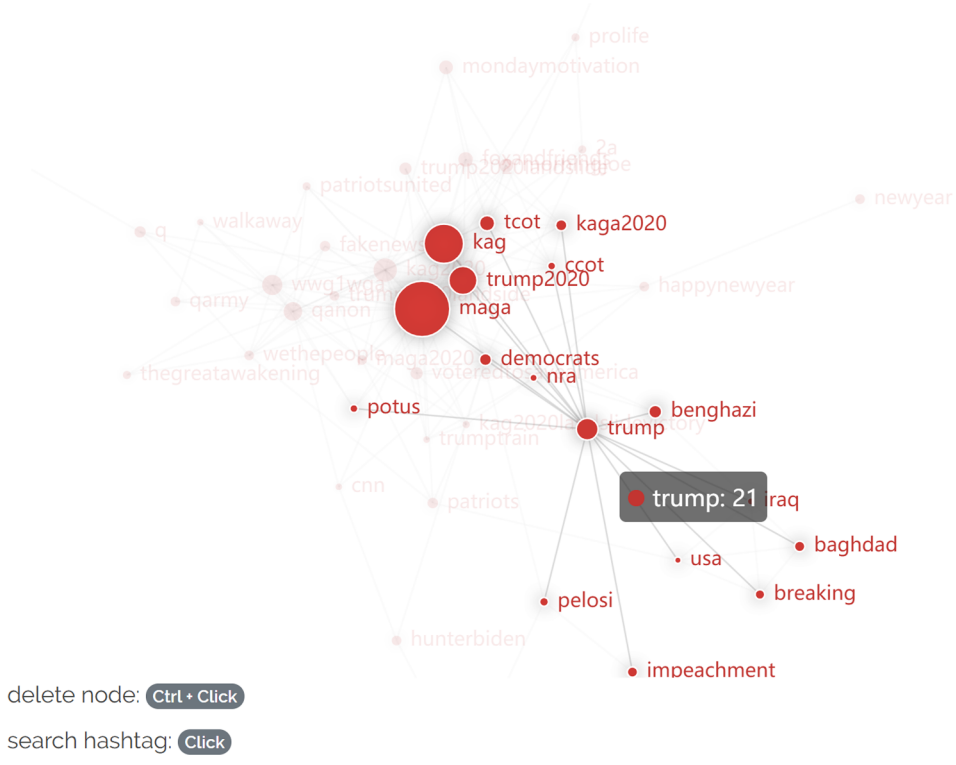


Fig. 12. "Trump" Twitter hashtag displayed in the software.

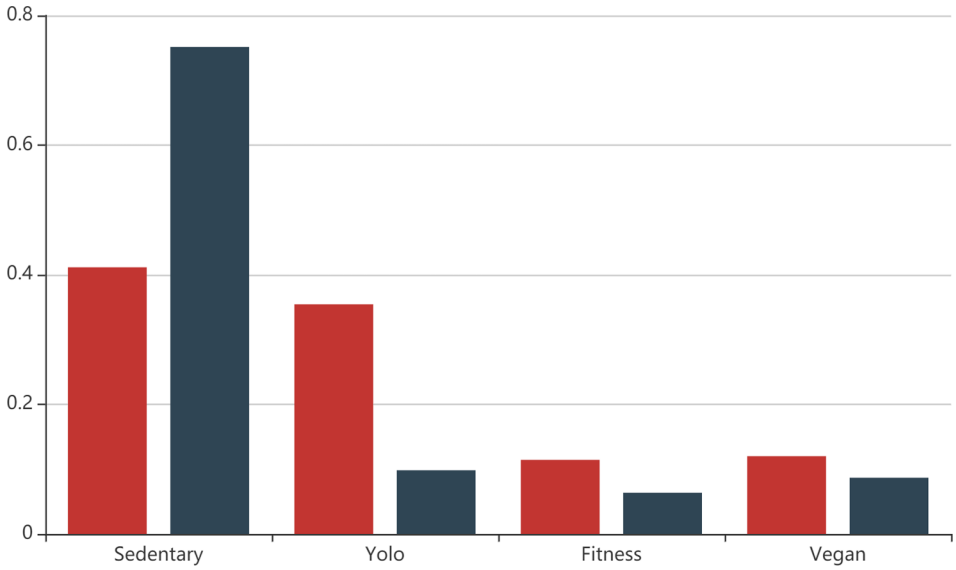


Fig. 13. Tribe Comparator in use. Comparison of @DFBHarvard (red) and @jtLOL (blue) followers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

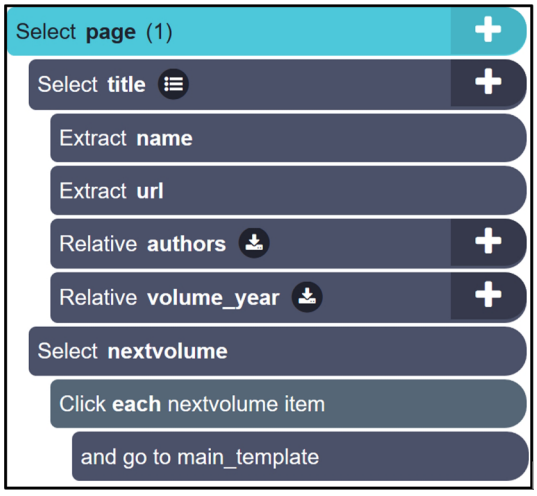


Fig. 14. Commands from ParseHub application.

Fig. 15. Word cloud for all article titles in the *Journal of Business Research*.

much more “Yolo” (“you only live once”: people who may display hedonistic, impulsive, or reckless behavior) than the followers of @jtLOL (in blue) (Fig. 13).

6. Additional tutorial: Scraping any website with ParseHub

As a final part of our tutorial, we introduce ParseHub, a point-and-click WYSIWYG scraper, which allows us to acquire data from most websites, including those based on Ajax (with dynamic loading of content), and those requiring a login. It can prove useful for extracting data from sites such as Reddit, Quora, or Amazon. ParseHub is extremely intuitive, and in its base form, it only requires clicking on the data, which we want to acquire, and clicking again to show the program in which the next section of the site is the same category of data. The rest is automated and works well.

Here again we relied only the free version of the tool, which is limited to scraping 200 pages, (which is still very useful). The paid version has a higher data limit and works faster, can rotate IPs, and be run on a schedule. For our proof-of-concept, we scraped all article titles and author names for papers published in the *Journal of Business Research* from 1973 to 2019. Even though acquiring data directly from the publisher would be possible with ParseHub, we simplified our showcase by relying on EconPapers, as it has a simpler site structure. As a result, we were able to collect 6156 article titles with authors and links to abstracts by creating a point-and-click algorithm with only five commands. The whole process took less than 20 min, including the script running. The commands are shown below. You can import our project into ParseHub from here for practicing, although the system is very intuitive (Fig. 14).

We were able to retrieve our results in both .csv or .json, and we added them as a separate sheet in the Google Sheets file linked above. Other similar tools include OctoParse, and some are even fully functional as browser add-ons, like Instant Data Scraper.

We have also processed the data in our Fighting Without Fighting software. Using the Word Cloud generation functionality we have come up with a graph presenting a cloud made of the journal articles’ titles (Fig. 15). We may observe that the most frequent words among the titles were “effect,” “consumer,” “marketing,” “role,” and “brand.”

After a couple of easy steps of data sorting in Excel, we created a word cloud for article titles from the journal’s first two years, 1973–1974 (85 titles), and for 2019 (482 titles), with the use of FWF software (Fig. 16). We may notice that scholars today are interested in slightly different topics in business research than they were 50 years ago.

5. Additional guidelines and recommendations

The tools we presented and the software we developed for this article should be sufficient for preliminary research and mapping the

Fig. 16. Word cloud for article titles in *Journal of Business Research* from 1973 to 1974 (left), and 2019 (right).

Table 1
Advantages and Disadvantages of Fighting Without Fighting.

Advantages	Disadvantages
Acquisition of rich initial material in which to ground qualitative work	Cannot gradually build coding skills by performing analyses
Effortless acquisition of large datasets	More restricted than relying on Python or R modules
Codeless scraping and analyzing data, including sentiment analysis	Less flexible, tailored and data-intensive than relying on specialized libraries
Ready-made analyses and graphical representations available in an app developed specifically for this article	Much fewer analytical and graphical representation options than in Python or R modules

most interesting topics and issues for further qualitative study. Researchers relying on these tools should keep in mind that even though the easy and wide access to big data are amazingly helpful, they can be deceptive in the sense of suggesting nonexistent causations, and need to be treated with caution, just like all other digital quantitative insights not followed up by additional checks.

While we enthusiastically encourage everyone to use the Fighting Without Fighting in conjunction with the proposed tools, we also suggest caution in relying on these methods alone for social science. We presented five methods because using all of them mitigates the bias of using a single kind of data and analysis, which researchers should keep in mind if just selecting one of the proposed tools for their study.

In terms of collecting data from Twitter or through scraping, quite often the larger the dataset, the better. Planning a study ahead of time allows for the gradual collection of data, the correction of mistakes in the initial phases, and the minimization of the risk of creating a large dataset that will not be optimally suited to answer the research question. Also, given the fact that both tools, APIs, and the terms of use change dynamically, we strongly advise starting collecting data early and adjusting often in the initial phases, so that even if changes in the terms of use along the way preclude finishing the whole project, one still generates enough data for a smaller project.

6. Summary

Our aim in this tutorial was to show the diverse ways of acquiring and analyzing data without the need for coding. Based on the arsenal of tools described here, including the FWF software we developed for this article, you can conduct robust and powerful analyses in business research without programming. We hope that the described methods will be useful and lead to new discoveries, as well as propagate the idea of doing data science in business studies. The table summarizes the key advantages and disadvantages of our approach for business research (see Table 1):

Funding

Working on this article was possible thanks to grant No. PPN/BEK/2018/1/00009 from the Polish National Agency for Academic Exchange.

References

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data in the proceedings of workshop on language in social media.

Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment analysis using deep learning techniques: A review. *International Journal of Advanced Computer Science & Applications, 8*(6), 424.

Aldahawi, H. A., & Allen, S. M. (2013). Twitter mining in the oil business: A sentiment analysis approach. *2013 International conference on cloud and green computing* (pp. 581–586). .

Caputo, F., Mazzoleni, A., Pellicelli, A. C., & Muller, J. (2019). Over the mask of innovation management in the world of Big Data. *Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2019.03.040.

Caputo, F., & Walletzký, L. (2017). Investigating the users' approach to ICT platforms in the city management. *Systems, 5*(1), 1.

Caputo, F., Walletzky, L., & Štepánek, P. (2019). Towards a systems thinking based view for the governance of a smart city's ecosystem: A bridge to link Smart Technologies and Big Data. *Kybernetes. The International Journal of Cybernetics, Systems and Management Sciences, 48*(1), 108–123.

Charles, V., & Gherman, T. (2019). Big data analytics and ethnography: Together for the greater good. In A. Emrouznejad, & V. Charles (Eds.). *Big data for the greater good* (pp. 19–33). Springer International Publishing.

Chełkowski, T., Gloor, P. A., & Jemielniak, D. (2016). Inequalities in open source software development: Analysis of contributor's commits in apache software foundation projects. *PLoS ONE, 11*(4), e0152976.

Chmielewska-Szlajfer, H. (2018). Opinion dailies versus Facebook fan pages: The case of Poland's surprising 2015 presidential elections. *Media, Culture & Society, 40*(6), 938–950. https://doi.org/10.1177/0163443718756065.

Ciesielska, M., & Jemielniak, D. (Eds.). (2018). *Qualitative methodologies in organization studies: Volume II: Methods and possibilities*. Palgrave Macmillan.

Ciesielska, M., & Jemielniak, D. (Eds.). (2018). *Qualitative methodologies in organization studies: Theories and new approaches*. Palgrave Macmillan.

Cova, B., & Cova, V. (2002). Tribal marketing: The tribalisation of society and its impact on the conduct of marketing. *European Journal of Marketing, 36*(5/6), 595–620.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research.* SAGE Publications.

Das, R., & Hodkinson, P. (2019). Tapestries of intimacy: Networked intimacies and new fathers' emotional self-disclosure of mental health struggles. *Social Media + Society, 5*(2) 2056305119846488.

De Oliveira, J. M., & Gloor, P. A. (2018). GalaxyScope: Finding the "Truth of Tribes" on social media. In F. Grippa, J. Leitão, J. Gluesing, K. Riopelle, & P. A. Gloor (Eds.). *Collaborative innovation networks: Building adaptive and resilient organizations* (pp. 153–164). Springer International Publishing.

Dourish, P., & Cruz, E. G. (2018). Datafication and data fiction: Narrating data and narrating with data. In Big data & society (Vol. 5, Issue 2, p. 205395171878408). https://doi.org/10.1177/2053951718784083.

Elshendy, M., Colladon, A. F., Battistoni, E., & Gloor, P. A. (2018). Using four different online media sources to forecast the crude oil price. *Journal of Information Science and Engineering, 44*(3), 408–421.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 897–904.

Eriksson, P., & Kovalainen, A. (2015). *Qualitative methods in business research: A practical*

guide to social research. SAGE.
Ghauri, P., Grønhaug, K., & Strange, R. (2020). *Research methods in business studies.* Cambridge University Press.
Gloor, P. A., Fronzetti Colladon, A., de Oliveira, J. M., & Rovelli, P. (2019). Put your money where your mouth is: Using deep learning to identify consumer tribes from word usage. *International Journal of Information Management.* https://doi.org/10.1016/j.ijinfomgt.2019.03.011.
Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding Statistics, 3*(4), 283–297.
Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review, 61*(4), 5–14.
Harrison, R. L. (2013). Using mixed methods designs in the Journal of Business Research, 1990–2010. *Journal of Business Research, 66*(11), 2153–2162.
Humprecht, E., & Esser, F. (2017). A glass ceiling in the online age? Explaining the underrepresentation of women in online political news. *European Journal of Disorders of Communication: The Journal of the College of Speech and Language Therapists, London, 32*(5), 439–456.
Ivankova, N., & Wingo, N. (2018). Applying mixed methods in action research: Methodological potentials and advantages. *The American Behavioral Scientist, 62*(7), 978–997.
Jemielniak, D. (2016). breaking the glass ceiling on Wikipedia. *Feminist Review, 113*(1), 103–108.
Jemielniak, D. (2020). *Thick big data: Doing digital social sciences.* Oxford University Press.
Jemielniak, D., & Przegalińska, A. (2020). *Collaborative society.* MIT Press.
Jemielniak, D., & Wilamowski, M. (2017). Cultural diversity of quality of information on Wikipedias. *Journal of the Association for Information Science and Technology.*. http://onlinelibrary.wiley.com/doi/10.1002/asi.23901/full.
Kaplan, A. M. (2015). Social media, the digital revolution, and the business of media. *International Journal on Media Management, 17*(4), 197–199.
Lazer, D. M. J., & Radford, J. (2017). Data ex Machina: Introduction to big data. *Annual Review of Sociology, 43*(1), 19–39. https://doi.org/10.1146/annurev-soc-060116-053457.
Liu, B. (2012). *Sentiment analysis and opinion mining.* Morgan & Claypool Publishers.
Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining text data (pp. 415–463). https://doi.org/10.1007/978-1-4614-3223-4_13.
Mahoney, K. T., & Baker, D. B. (2002). Elton Mayo and Carl Rogers: A tale of two techniques. *Journal of Vocational Behavior, 60*(3), 437–450.
Malone, T. W. (2018). *Superminds: The surprising power of people and computers thinking together.* Brown: Little.
Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093–1113.
Millington, B., & Millington, R. (2015). "The datafication of everything": Toward a sociology of sport and Big Data. *Sociology of Sport Journal, 32*(2), 140–160.
Myers, M. D. (2019). *Qualitative research in business and management.* Sage Publications Limited.
Polsa, P. (2013). The crossover-dialog approach: The importance of multiple methods for international business. *Journal of Business Research, 66*(3), 288–297.
Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P., & Mazurek, G. (2019). In bot we trust: A new methodology of chatbot performance measures. *Business Horizons, 62*(6), 785–797.
Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies.* https://oapub.org/edu/index.php/ejes/article/view/1017.
Ravitch, S. M., & Carl, N. M. (2019). *Qualitative research: Bridging the conceptual, theoretical, and methodological.* SAGE Publications.
Saura, J. R., Palos-Sanchez, P., & Grilo, A. (2019). Detecting indicators for startup business success: Sentiment analysis using text data mining. *Sustainability: Science Practice and Policy, 11*(3), 917.
Sinkovics, R. R., Penz, E., & Ghauri, P. N. (2008). Enhancing the trustworthiness of qualitative research in international business. *Management International Review, 48*(6), 689–714.
Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research, 70*, 263–286.
Smith, N., & Graham, T. (2019). Mapping the anti-vaccination movement on Facebook. *Information, Communication and Society, 22*(9), 1310–1327.
Stasik, A. (2018). Global controversies in local settings: Anti-fracking activism in the era of Web 2.0. *Journal of Risk Research, 21*(12), 1562–1578.
Symon, G., & Cassell, C. (2012). *Qualitative organizational research: Core methods and current challenges.* SAGE.
Vanhala, M., Lu, C., Peltonen, J., Sundqvist, S., Nummenmaa, J., & Järvelin, K. (2020). The usage of large data sets in online consumer behaviour: A bibliometric and computational text-mining–driven analysis of previous research. *Journal of Business Research, 106*, 46–59.
Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research, 88*, 428–436.
Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356–365.
Wang, W., & Zhou, Y. (2009). E-business websites evaluation based on opinion mining. *International Conference on Electronic Commerce and Business Intelligence, 2009*, 87–90.
Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems, 34*(1), 65.
Wen, Q., Gloor, P. A., Fronzetti Colladon, A., Tickoo, P., & Joshi, T. (2019). Finding top performers through email patterns analysis. *Journal of Information Science and Engineering* 0165551519849519.
Yarchi, M., & Samuel-Azran, T. (2018). Women politicians are more engaging: Male versus female politicians' ability to generate users' engagement on social media during an election campaign. *Information, Communication and Society, 21*(7), 978–995.

Leon Ciechanowski is a research assistant at Management in Networked and Digital Environments (MINDS) department, Kozminski University.

Dariusz Jemielniak is Full Professor and head of Management in Networked and Digital Environments (MINDS) department, Kozminski University, visiting scholar at he Center for Collective Intelligence at MIT's Sloan School of Management, and associate faculty at Berkman-Klein Center for Internet and Society, Harvard University. He is a corresponding member of the Polish Academy of Sciences. His recent books include Collaborative Society (2020, MIT Press, with A. Przegalinska), Thick Big Data (2020, Oxford University Press), Common Knowlege? (2014, Stanford University Press).

Peter A. Gloor is a Research Scientist at the Center for Collective Intelligence at MIT's Sloan School of Management where he leads a project exploring Collaborative Innovation Networks (COIN). He is also Founder and Chief Creative Officer of software company galaxyadvisors where he puts his academic insights to practical use, helping clients to coolhunt by analyzing social networking patterns on the Internet – spot the next big thing by finding the trendsetters, and to coolfarm – increase organizational happiness, creativity and performance through workforce analytics. In addition Peter is a Honorary Professor at University of Cologne and a Honorary Professor at Jilin University, Changchun China.