

Expert Insights into Analyzing Images from Political Online Communication

An Interview with Cody Buntain

Publication date: January 17, 2024

Most analyses of social online behavior are text-based. However, the impact visual materials such as images, videos, and graphics have on informing, convincing, and manipulating people is undisputed, let alone the simple fact that social media by default operate visually and on screens. It is worthwhile considering how we can leverage visual elements for (computational) social science research if we do not want to undervalue the very properties of the media we study.

We talked with Cody Buntain, Assistant Professor at the iSchool at the University of Maryland about the work with visual contents. Cody has several years of experience in studying online communication and is especially interested in online behavior during disasters and times of unrest. His work also covers political engagement and information quality. In this interview he is sharing his experience in shifting from text-based social media analysis to broader approaches that would also include image and video data collected from online platforms.

The interview was conducted by Indira Sen and Leon Fröhling on June 7, 2023, who met Cody during the International Conference on Web and Social Media (ICWSM-23) in Limassol, Cyprus. The interview transcript was edited for length and clarity.

Keywords: social media, political campaigning, image data, multimedia analysis, crisis informatics, data sharing, YouTube, TikTok, Instagram

GESIS: Hello, Cody, and thank you very much for providing insights into your current research. What are you working on at the moment, especially with regard to the analysis of visual media and multi-modal data from social media platforms?

Cody Buntain: From a very broad context, I am interested in understanding and studying online political behavior. That includes political discourse around voting, elections, but also around online manipulation. In this vein, we want to understand both the vulnerability of actors and who are – potentially – the relevant actors that are – potentially

– trying to manipulate them. A lot of work has been done on text understanding, on social network analysis, on the “who's interacting with whom”, but less work has been done on the visual aspects. Yet we know that visual media is way more popular. It is even growing in popularity with the explosion of TikTok and static visual media platforms like Instagram. It is a major vector of how people get exposed to information and is more impactful in how they might be creating or changing beliefs and integrating new information into their behavior.

I am looking at visual media, understanding what it tells us about people's behavior and interests, and then comparing it to their behavior in other modalities.

I am very interested in bringing this together, bringing our understanding of social networks, text sharing, image sharing, link sharing into one broader understanding of individual behavior. To do that, first we need to get a better understanding of image sharing in general, specific to that kind of modality. How are people sharing images in a political context? What does that tell us about their behavior, about their political interests? How aligned is that with the kind of content, the textual content they share? Much of what I am working on is focused on those questions. I am looking at visual media, understanding what it tells us about people's behavior and interests, and then comparing it to their behavior in other modalities.

GESIS: You gave us a bit of context about how politics was the lens you used to study images. Could you tell us how you first entered this field?

Cody Buntain: Originally, my dissertation work was on crisis informatics, how people use social media in the aftermath of disaster. In particular, we wanted to understand how we can find high quality information on social media *rapidly*. This work started when I was sitting in a bar at the University of Maryland, watching news coverage of the Boston Marathon bombing. The news coverage from one of the major mainstream news platforms was doing an interview with the Boston chief of police. What he was saying was directly contradicting the scroll at the bottom. The scroll at the bottom was from Twitter. These news organizations were using Twitter data to understand events on the ground, but their understanding was wrong. That is not to say that social media is bad. It is extremely useful for collective sense-making, for people to figure out what is going on. But when we use it from a news context to make decisions, especially if you are a first responder or a politician, it can have major impact and consequences.

I started doing a lot of work on that context, crisis informatics, high quality information, rapid response. Turns out, though, around the 2016 election, all the kinds of questions we ask about crises were: how do we get good quality information? How do we get it fast? How do we assess it? These questions are also massively important in the context of elections and campaigns and anytime that you have social unrest. Between the 2014

protests and demonstrations in Ferguson, Missouri, to the 2016 election and the concerns around online manipulation that came out there, I moved more onto the politics side, dealing more with how people engage around protests, around elections, and with our knowledge about malevolent actors who are trying to actively manipulate and influence our information spaces.

GESIS: You just started to talk a bit about the development that you took since starting in the field. Were there any major turning or inflection points in your development? Especially in this field, there has been such a rapid development of technology. Is there anything that stands out to you?

Cody Buntain: The first thing that comes to mind is our reliance on Twitter. When I was getting my PhD in 2015, everything I was doing was based on Twitter data. I was asking myself if I should even focus so much on Twitter because it could be gone the next year. The 2016 election happened and proved that totally wrong. Then I said, well, I can continue to do Twitter work. Now we are in a world where we have changed back again to the uncertainty if Twitter will be gone or our access to Twitter will be gone very soon. This overreliance on Twitter is a concern [1].

Partially, that was driven by an understanding that Russian influence efforts in the 2016 election relied heavily on visual content, memetic content, and visual memes.

Since 2015/16, though, I have been actively trying to look at more platforms. We know that Twitter is one sample of the population; it is not a representative sample; it is not the most popular in many low- and middle-income countries – unlike Facebook. How do we deal with those kinds of questions? That is one big data source question. Second from that is this move to multimodality, i.e., from just looking at text and online interaction we are now trying to look at visual media. Partially, that was driven by an understanding that Russian influence efforts in the 2016 election relied heavily on visual content, memetic content, and visual memes. There is some work about how incidental exposure and first exposure to a piece of information is massively impactful for the perspectives on that piece of information. That and the goal of looking at other platforms than Twitter pushed me towards looking at other modes of sharing, not just text-based interactions but also visual ones.

GESIS: Could you give us a high-level walkthrough of the pipeline that you use for visual media collection and analysis?

Cody Buntain: There are two main pipelines we use for this. In 2018, when I was looking back at the 2016 election, we were looking at what known Russian influence agents were sharing on Twitter, and they were sharing a huge volume of YouTube content. We wanted

to understand what kind of content it was – was it political? Was it left leaning? Was it right leaning? For that, we worked with and recruited political science students and asked them to watch these videos and then describe what these videos were. From there, we built a codebook to assess from a qualitative standpoint whether a video could be considered a political or politically oriented video and who it is supporting.

The results were insightful and showed that accounts that were sharing liberal news sources were sharing a larger than expected amount of conservative YouTube videos, which is not what we expected. And it is not symmetric: if you are sharing conservative news, you are not sharing any liberal videos, you are *only* sharing conservative videos. We did not know if that was because YouTube is – or at the time was – primarily full of conservative content, and if you pull *any* video, you get a conservative one, or if it was a strategic element of “I’m going to try and pull the left to the right by dropping in these YouTube videos”. That was a very manual process – it is high quality, but not necessarily scalable. It is good for video, though.

[We get] something like a dense feature vector describing the image. From there we can do a lot of analysis for clustering, figuring out similar kinds of images or identical images.

More recently, we have done much more work in using machine learning and computer vision models to characterize images. We collect images, out of the millions of images that have been shared in political contexts. They are too many to code, but we can push them through existing computer vision models that are pretrained on object detection tasks, like ImageNet [2].

We have a number of different models that we can choose from, because there are always new computer vision models coming out. That gives us a pipeline that could be used to analyze what kind of objects are being shared. But we actually do not use it for that. We take one layer off from the computer vision model and get out an embedding, a display, something like a dense feature vector describing the image. From there we can do a lot of analysis for clustering, figuring out similar kinds of images or identical images. This works well for images of infographics and not as well for images of groups of people. Looking at two images of groups of people, an object-identification model may say these two images are the same or very similar while a human would say they are very different. That is because there is much more context around the group and around how these people are engaged in conversation in these images.

These validity issues are a reasonable cost to pay for being able to process millions of images. One thing that already became quite clear, and I think that is also something that many people who until now predominantly work with text data are waking up to, is that there is this need to engage with multimedia content, images, or videos.

The most straightforward and actionable piece of advice that I could give is to try multiple computer vision models.

GESIS: Do you have any practical hands-on recommendations for people who want to start changing their mindsets from looking at text only to images and videos?

Cody Buntain: The most straightforward and actionable piece of advice that I could give is to try multiple computer vision models. Once you have a pipeline going, it is easy to drop in different models. If you are going to use VGG-19 [3] or ResNet-50 [4] or some other model, it is very easy to pull those in. One of the things that we have found is looking across five different models released between 2016 and 2022, the models are mostly consistent except for one model that gives us totally different results. So, if you were unlucky and you chose that one model as the model to use, you might get very different results than somebody else who used a different model. Picking three or four different models, you will have to invest additional cost and time, but you can do some good robustness checks to see how consistent the results are. That is the easiest, most straightforward piece of advice that you do not read about a lot.

GESIS: What about the combination of images and text, like in memes. Could you talk a bit about the analysis that you do there? Would you use a text classifier and an image classifier separately, combinations of them, or rather multimodal models?

Cody Buntain: There are a couple of pieces to disentangle here. When we think of memes, we generally think of a visual image with some text overlay. And there are approaches to pulling the text out using optical character recognition. We have not done that. We tend to look at all the text or images that a particular account shares and we build some characterization of that, like a topic model or an embedding. And we tend to put those two things together. Now it is unclear how well this works and there are newer models out there. Savvas Zannettou showed one during this conference called CLIP, Contrastive Language Image Pipeline [5], I think it is out of OpenAI, where it is both language and visual images. It is a unified model that you could use here. And that is where we are trying to go now.

GESIS: Were there any findings from some of the image multimedia analysis that you did, that were especially surprising, maybe even contradicting what you might have observed in text-based data?

Liberals tend to share a wider variety of imagery, conservatives share a particular kind of imagery, and this results in asymmetric results for ideology.

Cody Buntain: Nothing that was incredibly surprising. I was actually somewhat more surprised by how well the image analysis worked for the couple of tasks we were doing. But one of the major things that came up was very consistent: looking across political ideology, which is one of the things we are doing right now, we find that in the language that is being shared there is more diversity of language on the left, on liberal audiences, and a wider variety of liberal news sources and liberal news sharing than there is on the right in the United States. This also appears to be the case for imagery. Liberals tend to share a wider variety of imagery, conservatives share a particular kind of imagery, and this results in asymmetric results for ideology. We are getting better at predicting or placing liberals or liberal politicians based on the images they share, in the correct order of how liberal they are. And we are not very good at that for conservatives. This is consistent across a number of different image models that we tried, and it is also consistent across a number of different modalities.

Sharing of image data is ethically fraught and logistically hard.

GESIS: Working with visual media-based datasets from social media, could you tell us a bit about your experience in sharing these datasets. What would be your tips and guidelines for people who also work in this area and who are interested in sharing these types of data with the research community?

Cody Buntain: Sharing of image data is ethically fraught and logistically hard. There are a couple of issues about this. The logistics problem is that images are larger compared to text. A great example of this is the Twitter electoral integrity datasets [6] and the Twitter moderation research consortium datasets [7]. These are influence campaigns Twitter has identified. They share information or they release information about accounts and their messaging associated with these campaigns. The archive itself is around 10 terabytes in size across a number of different influence campaigns, but legitimately less than half of 1 percent of that is text, the rest is images and video.

The size makes image data very difficult to store, to work with, and especially makes it difficult to share.

The size makes image data very difficult to store, to work with, and especially makes it difficult to share. One of the things that we are trying to do to alleviate this problem is rather than sharing the images themselves, we share a small sample of the images so you can get a sense of what is in the data. And we share the image embeddings for all the images. The image embeddings are typically 256 or maybe 1280 dimensions, but not nearly as large as the actual images themselves, and you can share them much more easily. That helps with some of the ethical aspects as well because we are not sharing potentially copyright claimed images or personal pictures.

It is unclear how useful that is, though, considering the fact that there are new image models coming out constantly and we still do not have a good way to go from embedding in one model to embedding in a different model. However, we are trying to address some issues with image sharing by working on building infrastructure specifically to make search in images and doing research and analysis on images for researchers, who do not have these resources, much easier since they then can do searches using these embeddings. If there is a need for a specific detail of images, we might be able to provide those.

GESIS: The next question is going in the direction of representativity; for tweets and social media studies in general, we oftentimes talk about the issue that there are very active users that are close to the platform versus people who just linger around and observe what is going on. Is this issue even more pronounced for images? Do people feel that images are more personal than texts, does it take more commitment to post them? Is that something that you have looked at?

The same sort of scale-free behavior that you see in all other cases of online behavior applies here too.

Cody Buntain: We have looked a little bit at this. Regarding the motivations for sharing, I have not very much insight into that, but I can say it is absolutely the case that there are accounts that share many more images than other accounts. For example, looking at something like 600 members of the US Congress over the past 10-ish years, the median number of images shared by these accounts is about 1,100 images. But there are certainly a small handful of politicians who have shared 100,000 images. These very few are very active, while other politicians have only shared a couple of images. The same sort of scale-free behavior that you see in all other cases of online behavior applies here too.

I had a conversation with Michael Bossetta who was looking at ads [8]. Bernie Sanders, e.g., posts the same ad, but translated into different languages. That is something that we miss because we are all looking only at the images right now. Politicians in particular have maybe a reason to share the same content many times.

GESIS: Different strategies are used to access populations. This last question is more about other modalities besides images, particularly video. Given that there is this kind of explosion of video-based media like TikTok, of course, and then Instagram Reels and YouTube Shorts, which I believe are also being heavily used by politicians and political campaigns. There are some politicians who are very active and try to take advantage of these platforms. Is that also something that you think about working with, and what are the kind of directions that you might go for?

Cody Buntain: This is definitely something that I am interested in working on. There are a number of barriers, though. Getting the actual videos itself is non-trivial. You can do this on some platforms more easily than others, but pulling down the videos from YouTube Shorts, Instagram Reels, or TikTok is not trivial. Let alone pulling videos from YouTube. YouTube has been around for a long time. There has been an opportunity to do work on YouTube videos, but mostly the work I have seen on that has been done on YouTube thumbnails. You get the thumbnail, and it is easier to deal with an image rather than a video.

Pulling down the videos from YouTube Shorts, Instagram Reels, or TikTok is not trivial. Let alone pulling videos from YouTube.

I am interested in understanding the propagation of video and who uses it on different platforms. There is a professor at Stanford, Jennifer Pan, who has looked specifically at Douyin, the Chinese mainland version of TikTok, to understand politicians in the Chinese Communist Party [9] – whether the national politicians share content that gets propagated to the local politicians or whether the local politicians create content and some of it filters up. Her answer is: it is both of these things. She and her colleagues at Stanford have done a good bit of analysis looking at how do you characterize video. I want to do that, but I have not gotten to do it yet. To handle video is just going to be even harder, both in storage and analysis and processing and the actual transmission of it, only a few elite institutions can actually deal with this. That is a very valid concern.

GESIS: That is an extremely good point. Not something that we have touched on yet in the conversations that we have had so far and definitely very important: the accessibility to social scientists for this type of modality.

Cody Buntain: One of the things we talked about in the conference today is trying to reach out from ICWSM to the global community to get their voices into the conference. But if a well-resourced institution in the West, like in the United States or Europe, has trouble dealing with the size of these datasets or just transferring them, how do we expect that people who are in research institutions that do not have a lot of resources in low- and middle-income countries could manage that? We know that they are consuming this content a lot, but we will not have a good sense of what that means for these contexts, which I am very concerned about.

GESIS: One of the questions that we always like to ask our experts is whether you – apart from your tutorial that you gave here – have some additional resources for people who want to get started in the field and do image and multimedia analysis. Could you

point to any tutorials or guidelines or specific papers that discuss the entry point in a concise manner.

Cody Buntain: I would like to point to three good papers. One is a paper called “Images that Matter” by Andreu Casas and Nora Webb Williams [10]. The authors talk about why images matter, they are looking at it from a social media perspective, the way people who are exposed to images of their friends demonstrating and being physically out in the street protesting, are much more likely to go join those protests themselves. More so than if their friends are posting textual messages around. Images really matter, and this is why. There is another paper called “Images as Data” by Jungseock Joo and Zachary Steinert-Threlkeld [11], which shows other ways that we deal with images. And then there is a third paper by Jennifer Pan and Han Zhang. They were looking at images of demonstrations in China, and they have a really great paper on that called “CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media” [12], which is a big dataset of protests and images of protests in China. Han Zhang went on to write another paper that is essentially guidance on how you do image clustering for these kinds of things [13]. It has some weaknesses in terms of the different kinds of models to use but has really great advice about different tasks and different considerations and how to deal with this. His paper is probably the best entry point right now for new researchers wanting to get involved in this.

GESIS: We talked about this area, which has some societal significance. But then there are also some drawbacks that we talked about, like providing access, processing, even though there are newer and better models coming up. If you could make a wish to the universe to get some type of research artifact – it could be a library, a package or maybe a handbook, a tutorial, or a guideline – what would it be that would make your research in this area much easier?

[The] thing that would have the most positive impact on my research would be infrastructure to help me understand who the actors are sharing similar kinds of images – to try and fuse this idea of similar image sharing with the social interactions that we see.

Cody Buntain: Right now, the thing that would have the most positive impact on my research would be infrastructure to help me understand who the actors sharing similar kinds of images are – to try and fuse this idea of similar image sharing with the social interactions that we see. Because right now, everything goes in the other direction.

If I look at a bunch of actors, I can say which kind of images they share. And I can do the same thing with text messages and more if I know a set of actors already. But going the opposite direction, looking at a set of images and finding people who are sharing these kinds of images on social media to help me understand what else they are sharing or how they are engaging with them.

There is no mechanism to do that right now or an infrastructure for that. Which means that if we want to understand a particular meme, that we know is being shared by an influence campaign, or if we want to use an image of damage after an earthquake to find other people who are sharing images of damages after an earthquake, there is no way to do that. And that is concerning when you have so many platforms that are visually oriented, lots of content being pushed to Instagram and Snapchat and all these places and we essentially have no way to do search to the user level from the image space. I think if we had that, that would help us understand a lot more about how people use imagery. But that is for my particular work.

A good handbook [...] and a good mechanism for sharing image datasets [...] would be great for the field.

A good handbook about how you engage with images, and a good mechanism for sharing image datasets, even though they are so large, would be great for the field. And it would be good for making the field more accessible to social and computational social science people who care a lot, who have real questions, but do not have the technical expertise or infrastructure to handle these problems.

GESIS: Thank you very much for this interview, Cody!

References

- 1 We talked with Cody Buntain on June 7, 2023, in the midst of Twitter's reorganization of the access tiers to its API. Amid other changes, the previously free access to Twitter data for academics was discontinued.
- 2 ImageNet (Mar 11, 2021). <https://www.image-net.org/index.php> [retrieved Nov 1, 2023]
- 3 Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv*. <https://doi.org/10.48550/arXiv.1409.1556>
- 4 He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2015). Deep Residual Learning for Image Recognition. *arXiv*. <https://doi.org/10.48550/arXiv.1512.03385>
- 5 González-Pizarro, F., & Zannettou, S. (2023). Understanding and Detecting Hateful Content Using Contrastive Learning. *Proceedings of the International AAAI Conference on Web and Social Media*, 17, 257–268. <https://doi.org/10.1609/icwsm.v17i1.22143>
- 6 Twitter/ X (s.d.). Civic Integrity. <https://about.twitter.com/en/our-priorities/civic-integrity#data> [retrieved Nov 1, 2023]
- 7 Twitter/ X (s.d.). Twitter Moderation Research Consortium. <https://transparency.twitter.com/en/reports/moderation-research.html> [retrieved Nov 1, 2023]

- 8 Bossetta, M. (2018). The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. *Journalism & Mass Communication Quarterly*, 95(2), 471–496. <https://doi.org/10.1177/1077699018763307>
- 9 Lu, Y., & Pan, J. (2021). The Pervasive Presence of Chinese Government Content on Douyin Trending Videos. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3794898>
- 10 Casas, A., & Williams, N. W. (2019). Images that Matter: Online Protests and the Mobilizing Role of Pictures. *Political Research Quarterly*, 72(2), 360–375. <https://doi.org/10.1177/1065912918786805>
- 11 Joo, J., & Steinert-Threlkeld, Z. C. (2022). Image as Data: Automated Content Analysis for Visual Presentations of Political Actors and Events. *Computational Communication Research*, 4(1). <https://doi.org/10.5117/CCR2022.1.001.JOO>
- 12 Zhang, H., & Pan, J. (2019). CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media. *Sociological Methodology*, 49(1), 1–57. <https://doi.org/10.1177/0081175019860244>
- 13 Zhang, H., & Peng, Y. (2022). Image Clustering: An Unsupervised Approach to Categorize Visual Data in Social Science Research. *Sociological Methods & Research*, 004912412210826 <https://doi.org/10.1177/00491241221082603>

Suggested citation

Buntain, C. (2024). *Expert Insights into Analyzing Images from Political Online Communication. An Interview with Cody Buntain* (GESIS Guides to Digital Behavioral Data, 5). Cologne: GESIS – Leibniz Institute for the Social Sciences.

Series editors

Maria Zens, Katrin Weller, Claudia Wagner

Publisher

GESIS Leibniz-Institute for the Social Sciences, Cologne

License

Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)