Statement of Research Interests

My research focuses on human behavior on online platforms—the measurement, understanding, design, and assessment of implications. We use mobile devices any time to access the internet, read the news, watch videos, search for nearby restaurants, chat with friends, and leave posts on social networking sites. Online platforms serve all of these actions; billions of people use them. On average, an individual spends 143 minutes on social media [1] and is online for 6 hours and 42 minutes every day [2].

These tremendous interactions on online platforms are recorded as digital footprints. With appropriate consideration of privacy, these footprints enable us to understand individual or collective human behavior: what people like or hate, how people feel about various topics, and how people behave and engage. Thus, it has become crucial to understand human behavior on these online platforms.

My fundamental research goal is to develop new computational methods and tools for understanding, predicting, and changing human behavior on online platforms. One of the challenges posed by online data is the diversity and complexity of the datasets. I explore various types of large-scale data, investigate and compare existing tools to overcome its limitations and use them in the right way, and develop new measurements, machine learning models, and linguistic methods to understand human behaviors online and, furthermore, solve real-world problems. In this context, I summarize my past research and plans in the following paragraphs.

Novel computational methods for measuring and characterizing online human behavior

Among various formats of online data, my research focuses on textual content to facilitate the nuanced analyses of human behavior on online platforms. Not only is textual content amply available and rich with nuanced detail, but users' online posts and conversations are a key indicator for understanding online human behavior. Content can provide insight into the emotions, opinions, and actions of individuals, groups, or the public in the past, current, and future. My recent research addresses a range of problems relating to the analysis of structured and unstructured textual information used in, or arising from, content analysis for understanding online human behavior. A selection of my research in these areas is described below.

Methods for characterizing linguistic behaviors. In online human behavior research, one of the critical tasks is a 'comparative text analysis'—how do two (or more) groups differ in terms of essential topics perceived, tones and emotions, semantics (frames), and writing styles? For example, 'Word Cloud' is a simple and intuitive automated method for comparative text analysis. These automated methods can be used to compare languages of liberals vs. conservatives, people who are more vs. less extroverted, and consumer reviews on AT&T vs. Verizon, etc. They are particularly handy for characterizing the semantic content of textual data, which is so large that manual human examination is cost-prohibitive.

Studies that have considered comparative text analysis have mostly focused on issue salience using topic analysis or tone differences via sentiment analysis. My research has a different emphasis: leveraging semantic changes or variations of words through the comparison of texts from two groups. In our recent work [3], we compare the interaction and linguistic patterns of Trump and Clinton supporters using 2.5M posts and 39.8M comments on Reddit. Applying the word embedding-based method for quantifying semantic change [4], we found that Trump and Clinton supporters change the meaning of words significantly, depending on whom they talk to, while still using similar sets of words. For example, when interacting with Clinton supporters, Trump supporters use the term 'women' in the context of homosexuality (e.g., lesbians, homosexuality) and sexual harassment (e.g., accusers, harassed). However, among Trump supporters, context words point to fantasizing about women (e.g., fantasized, fantasizing, fantasizes). Since it was possible to quantify the semantic differences given to words between the two groups, we were then able to find words for which the semantic

meanings changed dramatically between them. These words were mostly related to political issues, such as 'war,' 'women,' 'job,' 'tax,' and 'healthcare.' Our approach identified meaningful differences in the linguistic behaviors of users that may be invisible when merely focusing on the volume of words in use.

Inspired by this analysis, I led the development of a technique, 'SemAxis,' to measure semantic changes in words across communities, which was published at ACL 2018 (a top-tier NLP conference) [5]. Using transfer learning of word embeddings, SemAxis offers a framework to examine and interpret words on diverse semantic axes (732 systematically created semantic axes that capture common antonyms, such as respectful vs. disrespectful). For example, how would Trump supporters perceive 'gun' compared to Sanders supporters? One may expect that 'gun' would be more positively perceived for Trump supporters than for Sanders supporters, but, can we go beyond the simple sentiment and characterize the semantic differences in a more nuanced way? Our approach can examine how 'gun' is perceived in two communities in relationship to 'Arousal' and 'Safety' axes. Using posts and comments on the two subreddits, /r/The_Donald and /r/SandersForPresident, we found that Trump supporters are generally more positive about gun-related issues, but Sanders supporters associate 'gun' more with arousal and danger. Our study can facilitate further investigations on context-dependent text analysis techniques and applications.

Methods for measuring public opinion. Fundamental questions about online human behaviors are often challenged by the difficult measurement tasks of aggregating large but noisy and non-representative data and inferring the demographic attributes of online users. Since online users do not tend to be a representative sample of the population, using raw online data may incur sampling bias in the analyses. To avoid such bias, one should accurately know who those online users are; this is challenging as online user attributes, including demographics, are not readily available. In my research program, I have been tackling challenges in measurement problems and seeking new methods to resolve them.

With a growing interest in using online data to study and quantify phenomena in the offline real world, a lot of effort has been made to use social media to measure public opinion, or for now-casting indices such as levels of flu activity or unemployment. In this context, social media users, who publicly report their health status or job losses, may act as 'social sensors.' For example, in our study [6], which was published at *Scientific Report*, we showed high correlations between Google and Twitter search volumes linked to Middle East Respiratory Syndrome (MERS)-related keywords and the number of confirmed MERS cases with a four-day delay. The result provides evidence that digital surveillance using a search engine and Twitter data is useful for monitoring the outbreak of an emerging infectious disease.

Yet, to create reliable and auditable social sensing systems, it is critical to understand who those people are that are talking online and whether they are a representative sample of the population. Are there particular demographic groups who should be given more or less attention? To this end, we analyzed a 10% sample of all public tweets in 2014, which were in their billions. We investigated how different sampling strategies (first-person [real user, not agencies], gender, age, geography, number of followers, etc.) affect the performance of the now-casting of two common offline indices: flu activity and unemployment [7]. We found that now-casting performs best when the fraction of 'normal' people, who have a complete profile, but with not too many followers, and who generally babble about their personal lives, is at its highest. This work provides not only methodological contributions for user attribute inference (e.g., first-person classifier for Twitter users) but also useful guidelines for designing a social sensing system.

To avoid sampling bias in online and social media data, inferring demographic attributes of online users is a key component in the social sensing system. Researchers have used, for example, a provided name to infer gender or race. However, guessing their age and race typically involves training a classifier, and this can be a daunting task for a multilingual corpus. Instead, state-of-the-art face analysis technology can infer the demographic attributes of a user, their age, gender, and race all at once, from a profile image. Considering the potential algorithmic biases toward minor groups (e.g., dark-skinned women) [8], we examined the feasibility of face

analysis technologies for demographic inference [9]. We also compared multiple face analysis technologies with three benchmark datasets for their performance on demographic inferences [10]. This work has been utilized as a guide to using this approach in multiple studies [11, 12, 13].

Human behavior and interaction on online platforms

My research is heavily driven by questions bearing on understanding human behavior in both academia and industry. As a leading researcher in online social media analyses, the rise of new online platforms leads me to answer questions, such as how people interact with new platforms and services or how those platforms affect human behavior and collective dynamics. However, content popularity or user engagement is also an intriguing topic. In the era of attention scarcity, but ironically under information overload, brands, businesses, and organizations strive to understand what makes users engage with them because it holds the key to their success. In several other projects, I investigated factors relating to online human behavior in various domains.

Effect of interactions on human behavior. The core functionality of social media platforms is 'social connection.' On such platforms, people connect, talk, and share information with others. People even change their behaviors online, depending on with whom they are connected. Thus, it is essential to understand the effect of user interaction on online human behavior.

Using information on the following links and tweets of 80 popular media sources and their 14 million audience members in late 2009, we built a network of news media using co-following behaviors on Twitter, revealing the landscape in modern media systems [14]. As the first study of news exposure on Twitter, we observed that the 'indirect media exposure' through social links (a.k.a., retweets) also broadened the opportunity for users to receive updates from politically diverse media outlets. It was truly amazing to observe people who share stories on both political sides. Thus, we further examined motivations on why people retweet political news [15]. We determined that, on Twitter, some of those who tend to be diversity-averse in their consumption of political news continue to promote stories they disagree with; they do so because these stories are relevant to their online friends. While this work promised a positive political engagement by increasing the diversity of news consumption, later, in another study [16], we observed 'partisan sharing' on both Facebook and Twitter; people selectively share political news based on their political leanings. This study was conducted based on Facebook data (44,999 news articles and 12,495 Facebook user profiles) and Twitter data (1,008 political news articles posted by 71 users who were recruited via our VotingTime application).

In the political regime, online platforms are increasingly considered as places for selective exposure, self-affirmation, and polarization. However, in other domains, social ties might be key to breaking the echo chamber. Using Charlie Hebdo as a case study, we analyzed social media responses to the terrorist attack to understand the social factors that contributed to individuals' online behavior [17]. At an individual level, we found social context played a significant role, with non-Arabs living in Arab countries using #JeSuisAhmed ("I am Ahmed") five times more often when they are embedded in a mixed Arab/non-Arab (mention) network. The results demonstrate the importance of understanding human behavior in the appropriate context, including social pressures.

Effect of content on human behavior. The content itself is one of the most substantial factors that affects how individuals respond (e.g., with a click, like, share, or reply). Through several projects, I investigated the content features that are associated with human behavior. In particular, I conducted a series of work looking deeply into news sharing behaviors to unveil one crucial question for news organizations, "what determines the popularity of news?" Since I aim to help editors, journalists, and social media managers at news organizations, I focus more on content features than user features. As a first step, we examined the sentiment of the headlines and how they are related to news popularity measured by the number of clicks on news URLs [18]. We found that the majority of news articles have negative headlines. We demonstrated that extremely negative and positive

headlines tend to attract more popularity, while neutral headlines tend to be less attractive.

Social media news posts are generally composed of two parts: a few sentences summarizing the news story and a URL to the news article. It is the social media managers' job to write the summary. We wondered whether the writing style of social media posts affects the popularity of news articles. To test that, we constructed paired datasets of social media posts (tweets) and their corresponding news headlines of eight news media. We conducted a propensity score matching analysis to control for the effect of the news topics. The results showed that 'clickbait'-style tweets result in increased user engagement on Twitter [19].

Writing an attractive social media post for one platform may not be too difficult. However, most news organizations appear on three or four social media platforms (Facebook, Instagram, Twitter, and YouTube) to distribute their news stories. Thus, for social media managers, it is vital to understand the similarities and differences of user engagement across multiple platforms. We considered how stylistic features [20], topics [21], and textual features [22] of news content are associated with popularity across multiple platforms using 3.1M news posts of 53 news organizations on Facebook, Twitter, Instagram, and YouTube. We then provided prediction models to determine whether news content becomes popular or not on each of the platforms. As one would expect, our results showed that factors associated with user engagement vary across platforms and news organizations. These findings highlight the need for a comprehensive analysis to cover multiple platforms and content providers, much like the work we have done here. Through a series of work using cross-platform datasets, we have provided a framework to examine the semantic and syntactic features of news posts in users' engagement.

In addition to the news industry, I have also looked into user engagement on a crowd-funding platform and examined what prompts users to fund Kickstarter projects [23], featured by *Fast Company*. We found frequent investors tend to fund efforts that are well-managed and match their interests. We also provided a statistical model to recommend potential investors from Twitter. The result highlights that modeling users' interests is also important for understanding their engagement.

Lastly, beyond textual features, I analyzed the visual features of content, in particular, online advertising images, on user engagement for brands on social media. With face detection-based demographic inferences, we also examined how demographics of models in advertisements play a role in brand engagement [12]. Using 14K posts and 850K comments relating to 69 US brands on Instagram and Facebook, we provided evidence for the resonance between the demographics depicted in a particular post and that of engaging users (a.k.a, in-group preference). This work provides a computational methodology for studying user engagement via visual features in online advertising, which can lead to more in-depth content analysis, such as sector-specific break downs.

Effect of platforms on human behavior. Platforms often influence online human behavior. For example, 'trending topics' on social media influence our perceptions of globally significant issues. However, would they genuinely trend for all? To answer this question, we conducted the first large-scale study presenting details on how various demographic groups use different hashtags on Twitter [9]. By analyzing tweets and profile images of 350K Twitter users from New York, we showed that such trending topics are mostly driven by a dominant, majority demographic group as the influence of the data for minorities is too small to register. We found some overlap for popular hashtags, but many group-specific hashtags also exist. Our work showed that a population-level analysis of hashtags and trends on Twitter is likely to miss the complexities induced by demographic-specific behavior.

Another example of such features is 'most read news,' which is common in news portals or on online news sites. These lists often give perceptions on which news topics are more salient. However, who are these topics important to? Using the daily top-30 news items for each gender and age group in Daum News, the second most popular news portal in South Korea, for the whole of 2015, we conducted multi-level analyses of gender and age differences in news consumption at four different levels by actual news items, sections, topics, and subtopics [24]. We found that there is a striking difference between the popular news items across groups,

and that the differences are mainly driven by subtopic preferences, not by section or topic preferences. The results implied that significant news issues for each demographic group could be very different from each other. However, on most news platforms, the 'most read news' is compiled from the news consumption of the majority group online and, thus, fails to promote a fair representation of information. This study highlights the fact that these platforms need to go beyond analyzing population-level behavior, as majority groups will always dominate this.

Platforms & societal challenges While it is tempting to think of online platforms as safe and diverse places with lots of 'good' opportunities, unfortunately, this may not always be true. Online spaces can be full of hate, bias and prejudice, stereotypes, and fakery; thus, it is crucial to understand the risks and limitations of online platforms and their negative influences on individuals and society.

One of such efforts made by us is Tanbih. Tanbih, an Arabic term meaning 'alert,' is a news aggregator, aimed at combating misinformation and promoting news literacy [25]. We recently presented our demo at EMNLP 2019 [26], a top NLP conference. As a news aggregator, it shows various meta-information regarding news articles (e.g., whether it is propagandistic or not, a frame expressed, etc.) and provides media profiles, including factuality, centrality, hyper-partisanship, inclinations for popular topics, and social media readership. Tanbih has been featured in reports by more than 50 public and university media outlets, including Wired, MIT News, Fast Company, and Forbes.

Another area I find essential to monitor is online advertising. Minority groups are continuously under- or misrepresented in advertisements. Since big brands, such as Starbucks or Jeep, have millions of followers on social media, the portrayal of different demographic groups in ads can play a crucial role in shaping the formation of the identities of those groups. To this end, we characterized the gender and racial diversity among ads of top US brands on Instagram and Facebook using 350K online advertising images [13]. Our work provided a framework for building a real-time system that monitors diversity and stereotyping in advertisements at a low cost.

Future directions

My research goal is to understand, predict, and change human behavior on online platforms by developing theoretical models, methods, and tools using large-scale real-world data. My interests span across a broad area of computational social science, including communication and journalism, politics, and health; I believe that the rich interdisciplinary environment of the school will enable me to successfully pursue my research objectives.

Developing computational methods for computational social science. As textual data is a primary source of online data, I apply and develop NLP methods in my work. I will further leverage and pursue rich linguistic analysis methods to answer questions in both academia and industry. In particular, I plan to develop unsupervised learning models for comparative text analysis by extracting important topics, contextual differences (frames), opinions, values, and beliefs. My short-term goal includes the development of context-dependent text analysis techniques for non-English languages, particularly resource-scarce languages, by extending SemAxis [5] to foster research in other non-English cultures.

Images and videos are presently abundant on social media. These visual data can be crucial cues to understanding the psycho-sociological status of individuals. However, due to the difficulty in processing large-scale visual data, the relationship between visual data and human behavior on social media is understudied. With advances in deep learning, researchers have a new opportunity to study news photos at scale. By using existing tools such as Google Cloud Vision API, it is possible to automatically recognize the objects, faces, and sentiments apparent in a given photo. Alternatively, since pre-trained image neural networks such as Residual-Net are readily available, we can easily apply computer vision approaches in our analyses. For example, we analyzed

two million news photos published in 2016 using Google Vision API to understand portrayals of different presidential candidates by various news outlets in their news photos [27]. In another study, we analyzed 9K videos published by Al Jazeera to predict user engagement based on video features [28]. Continuing working in this direction, I aim to use visual data to understand individuals' interests and behaviors, and to develop tools that make it more comfortable to analyze visual data for other research communities.

Media analysis and its impact on society. Understanding human behavior on online platforms is the core of my research. I would like to continue collaborations with scholars in other disciplines and practitioners in industries to solve real-world problems. One such effort is our on-going project 'Discursive power of contemporary media system' funded by the Volkswagen Foundation. Under the current media system, anyone can contribute to communication spaces to introduce, amplify, and maintain topics, frames, and speakers, thus shaping public discourses and controversies; this concept is called as discursive power [29]. With colleagues at the University of Konstanz and the University of Bamberg, I will reveal the power relationships in contemporary media systems and provide theoretical, methodological, and practical implications in the fields of political science and computer science. We are currently building software that collates news articles, political talk, and social media interactions in real-time, which other scholars will be able to access in the future.

Nudging human behavior. Large-scale observational online data and computational power enables us to work on 'understanding' and 'predicting,' but not 'changing' online human behavior. I aim to pursue working in the area of 'nudging' human behavior. I am currently working on a fascinating project, Precision Public Health Campaigns, to determine how we could (i) use data analysis and machine learning to create the right message for the right person; and (ii) evaluate the impact on offline behavior beyond simple impressions and click-based metrics. Digital channels facilitate the tailoring and distribution of messages to very specific target audiences and decide on an effective approach and content. It enables us to do A/B testing and evaluates which message works best for which sub-population in creating physical participation, and not just a digital 'Like.' We address the opportunities and challenges of using targeted advertising tools for personalized health messaging for public health campaigns [30]. We show that tailored messages by gender, culture, and psychological attributes are more effective for user engagement. I further aim to investigate its effectiveness on real-world activities—how online advertisement increases offline activities, such as benefiting from vaccinations or cancer screenings.

Promoting media literacy and critical thinking. Lastly, misinformation and bias on online platforms is one of the biggest threats to our society. It can lead to several implications such as personalized attacks, mass mobilizations, and grave health consequences, among others. Continuing working on the Tanbih project, I plan to work towards educating and promoting media literacy, and critical thinking.

Concluding remarks. Data science has opened a new door to developing methods that can leverage online data to understand individuals and society, provide decision-makers with insights into the impacts of online data, and build systems to support brands, businesses, and organizations. I have remained fascinated by the opportunities within data science. Fortunately, I have not only worked with researchers of computer science but also with communication scholars, political scientists, and journalists. This diversity of academic training has brought unique perspectives to the table. While collaborations are always difficult, they are also the most effective and engaging way to perform research; I plan to continue conducting research in this way in the future.

References

[1] Famemass. How much time do people spend on social media in 2019?, 2019. https://famemass.com/time-spent-on-social-media.

[2] Saima Salim at Digital Information World. More than six hours of our day is spent online, 2019. https://www.digitalinformationworld.com/2019/02/internet-users-spend-more-than-a-quarter-of-their-lives-online.html.

[3] Jisun An, Haewoon Kwak, Oliver Posegga, and Andreas Jungherr. Political discussions in homogeneous and cross-cutting communication spaces. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 68–79, 2019.

[4] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501. Association for Computational Linguistics, 2016.

[5] Jisun An, Haewoon Kwak, and Yong-Yeol Ahn. SemAxis: A lightweight framework to characterize domain-specific word semantics beyond sentiment. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, July 2018.

[6] Soo-Yong Shin, Dong-Woo Seo, Jisun An, Haewoon Kwak, Sung-Han Kim, Jin Gwack, and Min-Woo Jo. High correlation of middle east respiratory syndrome spread with google search and twitter trends in korea. *Scientific reports*, 6:32920, 2016.

[7] Jisun An and Ingmar Weber. Whom should we sense in "social sensing" - analyzing which users work best for social media now-casting. *EPJ Data Science*, 4(1):22, 2015.

[8] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91, 2018.

[9] Jisun An and Ingmar Weber. # greysanatomy vs.# yankees: Demographics and hashtag use on twitter. In *Tenth International AAAI Conference on Web and Social Media*, 2016.

[10] Soon-Gyo Jung, Jisun An, Haewoon Kwak, Joni Salminen, and Bernard Jim Jansen. Assessing the accuracy of four popular face recognition tools for inferring gender, age, and race. In *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[11] Julio Reis, Fabricio Benevenuto, Raquel de Melo, P Vaz 'd Prates, Haewoon Kwak, and Jisun An. Breaking the news: First impressions matter on online news. In *ICWSM'15: Proceedings of The International Conference on Weblogs and Social Media*, 2015.

[12] Jisun An and Ingmar Weber. Diversity in online advertising: A case study of 69 brands on social media. In *International Conference on Social Informatics*, pages 38–53. Springer, 2018.

[13] Jisun An and Haewoon Kwak. Gender and racial diversity in commercial brands' advertising images on social media. In *International Conference on Social Informatics*, pages 79–94. Springer, 2019.

[14] Jisun An, Meeyoung Cha, Krishna Gummadi, and Jon Crowcroft. Media landscape in twitter: A world of new conventions and political diversity. In *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[15] Jisun An, Daniele Quercia, Meeyoung Cha, Krishna Gummadi, and Jon Crowcroft. Sharing political news: the balancing act of intimacy and socialization in selective exposure. *EPJ Data Science*, 3(1):12, 2014.

[16] Jisun An, Daniele Quercia, and Jon Crowcroft. Partisan sharing: Facebook evidence and societal consequences. In *Proceedings of the second ACM conference on Online social networks*, pages 13–24. ACM, 2014.

[17] Jisun An, Haewoon Kwak, Yelena Mejova, Sonia Alonso Saenz De Oger, and Braulio Gomez Fortes. Are you charlie or ahmed? cultural pluralism in charlie hebdo response on twitter. In *Tenth International AAAI Conference on Web and Social Media*, 2016.

[18] Julio Cesar Soares Dos Rieis, Fabrício Benevenuto de Souza, Pedro Olmo S Vaz de Melo, Raquel Oliveira Prates, Haewoon Kwak, and Jisun An. Breaking the news: First impressions matter on online news. In *Ninth International AAAI conference on web and social media*, 2015.

[19] Kunwoo Park, Haewoon Kwak, Jisun An, and Sanjay Chawla. Sharing of news articles on social media: How should news media post a tweet for better engagement? In *Under submission*, 2020.

[20] Kholoud Khalil Aldous, Jisun An, and Bernard J. Jansen. Stylistic features usage: Similarities and differences using multiple social networks. In *International Conference on Social Informatics*, pages 309–318. Springer, 2019.

[21] Kholoud Khalil Aldous, Jisun An, and Bernard J Jansen. View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 47–57, 2019.

[22] Kholoud Khalil Aldous, Jisun An, and Bernard J. Jansen. Predicting audience engagement across social media platforms in the news domain. In *International Conference on Social Informatics*, pages 173–187. Springer, 2019.

[23] Jisun An, Daniele Quercia, and Jon Crowcroft. Recommending investors for crowdfunding projects. In *Proceedings of the 23rd international conference on World wide web*, pages 261–270. ACM, 2014.

[24] Jisun An and Haewoon Kwak. Multidimensional analysis of the news consumption of different demographic groups on a nationwide scale. In *International Conference on Social Informatics*, pages 124–142. Springer, 2017.

[25] QCRI. Tanbih, 2019. `https://tanbih.org`.

[26] Yifan Zhang, Giovanni Da San Martino, Alberto Barrón-Cedeño, Salvatore Romeo, Jisun An, Haewoon Kwak, Todor Staykovski, Israa Jaradat, Georgi Karadzhov, Ramy Baly, Kareem Darwish, James Glass, and Preslav Nakov. Tanbih: Get to know what you are reading. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 223–228, Hong Kong, China, November 2019. Association for Computational Linguistics.

[27] Haewoon Kwak and Jisun An. Revealing the hidden patterns of news photos: Analysis of millions of news photos through gdelt and deep learning-based vision apis. In *Tenth International AAAI Conference on Web and Social Media*, 2016.

[28] Jisun An and Haewoon Kwak. Viewer retention of online video content: Analyzing effect of entities and shot changes. In *Work In Progress*, 2019.

[29] Andreas Jungherr, Oliver Posegga, and Jisun An. Discursive power in contemporary media systems: A comparative framework. *The International Journal of Press/Politics*, 24(4):404–425, 2019.

[30] Jisun An, Haewoon Kwak, Luis Fernandos-Luque, and Ingmar Weber. Precision public health campaigns:. In *Under submission*, 2020.