Journal of the Association for Information Systems (2023) **24**(6), 1562-1593
doi: 10.17705/1jais.00810

RESEARCH ARTICLE

ISSN 1536-9323

Champions for Social Good: How Can We Discover Social Sentiment and Attitude-Driven Patterns in Prosocial Communication?

Raghava Rao Mukkamala,[1] Robert J. Kauffman,[2] Helle Zinner Henriksen[3]

[1]Department of Digitalization, Copenhagen Business School, Denmark, rrm.digi@cbs.dk
[2]Department of Digitalization, Copenhagen Business School, Denmark /
School of Information Systems, Singapore Management University, Singapore, rk.digi@cbs.dk
[3]Department of Digitalization, Copenhagen Business School, Denmark, hzh.digi@cbs.dk

Abstract

The UN High Commissioner on Refugees (UNHCR) is pursuing a social media strategy to inform people about displaced populations and refugee emergencies. It is actively engaging public figures to increase awareness through its prosocial communications and improve social informedness and support for policy changes in its services. We studied the Twitter communications of UNHCR social media champions and investigated their role as high-profile influencers. In this study, we offer a design science research and data analytics framework and propositions based on the *social informedness theory* we propose in this paper to assess communication about UNHCR's mission. Two variables—*refugee-emergency* and *champion type*—relate to the informedness of UNHCR champions' followers. Based on a Twitter sentiment and attitude corpus, we applied a *five-step design science analytics framework* involving machine learning and natural language processing to test how the emergency type and champion type impact social communication patterns. Positive and neutral sentiment dominated the tweets of the champions and their followers for most refugee-emergency types. High participation-intensity champions emphasized high-intensity emergencies with tweet patterns reflecting dominant positive or neutral sentiment and sharing/liking attitudes. However, we found that sports figures effects were limited in spreading UNHCR's message, despite their millions of followers. We demonstrate the power of data science for prosocial policy based on refugee crisis awareness and instantiate our methods and knowledge contributions in a research framework that derives knowledge, decisions, and actions from behavioral, design, and economics of information systems perspectives.

Keywords: Data Science for Social Good, Deep Learning, Influencers, Machine Learning, Natural Language Processing, Prosocial Behavior, Sentiment Analytics, Social Informedness Theory, Social Outreach, Twitter

Roger Chiang, Ahmed Abbasi, and Jennifer Xu were the accepting senior editors. This research article was submitted on April 15, 2021 and underwent three revisions. Raghava Rao Mukkamala is the corresponding author. This paper is part of the Special Issue on Data Science for Social Good.

> Speaking about how she came to adopt her eldest child, Maddox, Angelina Jolie explained, *"Cambodia was the country that made me aware of refugees. It made me engage in foreign affairs in a way I never had and join UNHCR. Above all, it made me a mom."* (Amy Mackelden, *Harper's Bazaar*, 2020)

1 Introduction

The Premier League soccer player, Mo Salah, a top player on the Liverpool team in the UK, is also known for a completely different kind of activity: *prosocial outreach*. The same is true for the American actress, Angelina Jolie, who adopted orphaned refugee children during times of political upheavals and humanitarian crises in Cambodia, Ethiopia, and Vietnam. Both have lent their names to philanthropic activities and services for social good and to the United Nations High Commissioner for Refugees (UNHCR) and have contributed a portion of their wealth to prosocial causes. This has positioned them as high-profile role models, ambassadors, and envoys, as well as social change influencers.

For more than half a century, the role of *champions for change* has been recognized as a driving force for supporting social group collective actions (Ginsberg & Abrahamson, 1991). [1] Champions of change are influencers that improve the informedness of others and shift attitudes and preferences in a beneficial prosocial direction. *Individual agency*—the capacity of a person to make their own free choices—is a central element (Lawless & Price, 1992) yet one that is subject to both promotion and pushback in the world of social media. As such, many prominent social media influencers have sought to become *champions for social good.*[2]

This research focuses on the social good that can arise from the impact of champions on prosocial organizations, the changed attitudes among their followers, and other prosocial outcomes. We ask:

1. Is *social informedness theory*, as proposed in this article for the first time, a useful lens for understanding how social sentiment can beneficially change the impacts of UNHCR and its stakeholders' work on displaced refugees?
2. Can abundant data be analyzed to understand how an organization's policy is aligned with changes in social sentiment that affect stakeholders' views of its mission and its relationship to leading societal issues?

The present study examines the role of champions in creating awareness of refugee services, focusing on UNHCR as a catalyst.[3] We pay particular attention to the profiles of those who collaborate with UNHCR and generate engagement on social media. A key issue is the impact of such engagement on the institution.

Using data science methods (Chang et al., 2014), we assessed patterns of social communication that can create positive awareness and may thus beneficially impact refugee emergency-related outreach. We studied regions experiencing political unrest and armed conflicts associated with UNHCR. Though we note Salah's and Jolie's UNHCR collaborations above, there are numerous examples of high-profile public figures who are working with UNHCR to promote social good for refugees.

We applied *machine learning* (ML) and *neural word-embedding analytics* for the *natural language processing* (NLP) of social sentiment data. We show how these methods support UNHCR's ambition to make people more aware of the difficulties experienced by refugees through information technology (IT). We also demonstrate the power of data science for prosocial policy based on refugee crisis awareness, instantiating our methods and knowledge contributions in a research framework that derives knowledge, decisions, and actions from behavioral, design, and economics of information systems (IS) perspectives.

2 Literature and Theoretical Background

Social media platforms have become key channels for organizations and individuals to share their ideas and promote public engagement (Xu et al., 2019). Below, we discuss the role that IT has played in social awareness and prosocial information sharing involving champions as influencers. We also introduce social informedness theory. We develop it in this article for application to the UNHCR context, though it has broader relevance and applicability. The basic idea is that increased social awareness by a group of people experiencing a refugee crisis or other social phenomenon is beneficial for an organization that seeks social support, charitable funding, and widespread agreement that its prosocial actions are justified to create social good.

2.1 Technology and Communication of Prosocial Information

Limited research has addressed how NGOs use messages posted on social media to foster civic engagement and address humanitarian crises (Perez-Cepeda & Arias-Bolzmann, 2021). In addition, academic studies of UNHCR and its work with refugees and the related social, economic, government, and military problems are sparse.

[1] The effectiveness of champions as social influencers has been described by epidemiological models that identify how ideas travel in social networks. The dynamics have been assessed with quantitative and qualitative methods. Recent technology developments have increased the *information intensity of society*, transforming organizational strategy and social informedness in the process (Clemons et al., 2017). Tracking sentiment-related communication among social network members also has been key to providing richness and depth in understanding the public's reactions to the prosocial activities of leading influencers (Tan & Tas, 2020).

[2] *Champions for Social Good* is also the name of a well-known podcast focusing on bringing "together thought leaders and practitioners in the social good movement to share learnings, educate, and inspire change across the philanthropic sector" (ListenNotes, 2021). This is indicative of how widespread and popular promoting social good and the role of champions have become in the philanthropic, crowdfunding, and social media spaces to support prosocial activities (Mayer, 2022).

[3] The UN is a nonprofit, *intergovernmental organization* (IGO), while the UNHCR is an agency that focuses on refugees globally and maintains relationships with more than 900 supporting *nongovernmental organizations* (NGOs).

There are two themes prevalent in the literature on digital technologies and refugees: *security and privacy* and *discrimination and sorting* (Jacobsen, 2015). For example, Tazzioli (2020) studied the experiences of asylum seekers with the card-based Cash Assistance Program in Greece, whose data were leveraged to enable effective governmental management of the people involved. This study examined the prosocial intentions related to asylum seekers' private-data extraction and identity datafication. In contrast, our study focuses on the awareness of people other than the affected refugees as a means of jump-starting fundraising and securing support (e.g., for UNHCR policy and services).

In a similar prosocial vein, Jacobsen and Sandvik (2018) assessed UNHCR's use of techno-bureaucratic practices for managing refugees—including biometric solutions, such as the use of biometric data collection and storage for identifying and registering refugees, as well as their children and relatives (Kingston, 2018). As a result of such efforts, UNHCR now recognizes that the maintenance of refugee ID information is a core information management and tech innovation activity that must be carried out even under extremely difficult social situations (UNHCR, 2020b).

The dark side of such technological innovation is that not everything has worked as well as the social planners at UNHCR had hoped. Jacobsen and Sandvik (2018) noted that biometric solutions created transparency for UNHCR's donors but did not improve refugee protection. Other researchers have pointed out that the collection of biometric data by international agencies—regardless of how well-meaning they are—can lead to conflicts. As Madianou (2019, p. 581) reported, their experience suggests "asymmetries between refugees and humanitarian agencies … ultimately entrenches inequalities in a global context."

Indeed, there have been reports of unrecognized biases in individual documentation and refugee status determination (Morand et al., 2012), failures in protecting personally identifying information (UNHCR, 2019), and the need for permission protocols for refugee-related data sharing involving NGOs, firms, and nations (Identification for Development, 2022). Our study shifts the focus to influencers' capacity for creating awareness of the problems faced by refugees and displaced citizens.

The literature has recognized the importance of accessing massive datasets from private and public sources, as well as data science methods that can provide new ways of uncovering and calibrating the social impacts of network influencers (Aral & Walker, 2012). The possibilities surrounding human communications involving IT and social media sentiment and big data analytics have become important sources of scientific achievement (Chen et al., 2012). For example, the creation of explanatory theories for the seemingly random and sometimes aggressive behavior observed in online social networks can now be explored using *computational social science* (CSS) research approaches (Carley & Gasser, 2013). The development of data science analytics tools and research perspectives emphasizing the information value chain for studying social media issues has also been key (Abbasi et al., 2016).[4]

2.2 Champions as Social Media Influencers

The Twitter platform—known today as X—has enabled many well-known influencers to reach millions of followers, offering a potent platform not only for product sponsorships but also for the promotion of causes related to their humanitarian work. For example, leading sports figures have been studied for their influence on the golf ball market and football-related sales to young adult consumers (Chung et al., 2013; Dix et al., 2010).

Although the value of sports figures in product ads has been well established, leveraging influencers to promote refugee causes has also become a central outreach strategy for UNHCR and similar organizations.[5] In 2020, Salah became UNHCR's Ambassador for Instant Network Schools, a Vodafone Foundation-UNHCR joint initiative[6] intended to provide young refugees, host communities, and teachers access to digital learning content and the internet.[7]

[4] Kauffman and Wood (2009) emphasized that a new paradigm in the philosophy of empirical social science had arisen, as innovative data analytics techniques permitted the study of topics that were uneconomical and technically infeasible to conduct earlier—though leading researchers had already expressed their scientific imagination for what would soon emerge (Oinas-Kukkonen et al., 2010).

[5] Traditionally, famous royalty have held positions as prosocial ambassadors—for example, Princess Diana on landmine removal, Prince William on African wildlife protection, and Prince Harry on landmines as scars of war (Barr 2019), and recently (though controversially), as chief impact officer for a Silicon Valley mental health start-up (Martin, 2021).

[6] Through his status as a sports icon, Salah also has a well-established social media profile. As of April 2022, he had 49.5 million Instagram followers and 16 million Twitter followers. In comparison, the Prince of Wales and the Duchess of Cornwall had more than 1 million followers as of February 2022. Thus, Twitter activities and news may provide a basis for attitude and sentiment analytics on salient impacts—for example, regarding Salah's and others' engagement with vulnerable social groups.

[7] The initiative promotes education in marginalized communities in Africa. In addition, the American actor and philanthropist, Ben Stiller, was appointed as a UNHCR Goodwill Ambassador in 2018. He was tasked with promoting awareness and solutions to vexing human rights problems

2.3 Social Awareness-Related Theory

Many people are now spending a significant portion of their lives on social media. Despite the prevalence of misinformation on social media platforms, people are increasingly using such channels to improve their awareness of current events and to inform themselves about products, brand reviews, services, targeted news, weather and natural disasters, refugee crises, political messaging, healthcare and wellness information, and many other topics. We refer to these kinds of sources as the basis for informedness, which serves as the foundation for the proposed theory in this article.

The digital traces of social media users are widely available for data science purposes, including new computational big data experimental methods that analyze "living analytics" in social, geographical, and cultural contexts, as well as issues that map to the lives of citizens, consumers, and corporations (Kauffman et al., 2017). Marketers have been particularly attentive to the marketing strategy impacts of social media, including brand posts and fan pages (de Vries et al., 2012), new modes of social influence, and predictive analytics for influencers' impacts (Appel et al., 2019). IS researchers have explored the implications of increased information abundance, using insights from earlier theory to understand its impacts in varied contexts (Clemons, 2008; Clemons et al., 2017). Prior research has focused on various issues,[8] including information provision for sustainable business network collaboration (Thimm & Rasmussen, 2010), consumer and firm informedness in train ticket pricing and hospitality (Li et al., 2014), household informedness in recycling (Lim-Wavde et al., 2017), household digital entertainment consumption (Hoang & Kauffman, 2018), and the strategic sustainability of rewards-based crowdfunding firms (Wessel et al., 2021).

In this study, we build a set of propositions that represent the UNHCR program, based on the communications of its champions and their followers to obtain project support (via positive sentiment and attitudes, etc.). Our findings indicate that *social informedness* is an intermediate benefit generated through social media to support the creation of final program outcomes intended to create social good. We proxy informedness through refugee issue-related awareness outreach using social sentiment and attitude polarity, which offer quantitative evidence of follower receptivity and the convergence of supportive social opinions. We show how this process can be disentangled via *deep learning* (DL) methods to offer new insights into how UNHCR can institute effective social policymaking. Our findings suggest that sports champions, leading actors, and other high-profile people—including former refugees—are well-positioned to serve as influencers for social change.

2.4 Data Science for Discovering Patterns in UNHCR Prosocial Communication

To analyze unstructured data, such as tweets and other textual corpora, we used DL methods for NLP and supervised aspect-level sentiment classification with unsupervised DL approaches for aspect extraction. While aspect-level sentiment classification is a fine-grained classification method (Schouten & Frasincar, 2015), most sentiment methods identify sentiment polarity based on whole sentences or the entirety of a textual document but ignore other important aspects—much like omitted variable bias in statistical hypothesis-testing approaches. Aspect-level sentiment methods focus on connection polarity for a potential aspect and the linguistic content of a sentence. We employ attention modeling-focused *long and short-term memory* (LSTM) recurrent neural networks (Wang et al., 2016) and bidirectional word assessment for semantic meaning.

The identification of Twitter corpus aspects (e.g., armed conflicts, political unrest, connections across issues that relate influencers to one another, sensitive topics discussed in Twitter messages exchanges, etc.) is important for classifying aspect-level sentiment (He et al., 2017). While topic modeling methods (e.g., latent Dirichlet allocation, LDA) are frequently used for aspect extraction (Blei et al., 2003), they are associated with certain drawbacks. For example, such methods do not produce highly coherent classification aspects due to their sparse data and are sometimes difficult to interpret in managerial contexts. Word-embedding models, in contrast, such as word and global vectors (Mikolov et al., 2013), identify words occurring in similar contexts. Thus, since word co-occurrences identified for word embedding using neural networks can produce more appropriate aspects, we use an unsupervised attention approach for aspect-oriented sentiment analysis (He et al., 2017).

3 Charitable Activities, Social Influencers, and UNHCR

3.1 Charitable Activities in IS and Marketing Research

IS researchers have addressed how technology can make a difference to people at the margins of society, focusing, in particular, on people living in rural environments (Jha et al., 2016) and healthcare for vulnerable groups (Venkatesh et al., 2016). However, refugees have received very limited attention in IS

experienced by Guatemalan, Honduran, El Salvadorian, and Syrian refugees.

[8] The constructs in these studies relate to information as an *acquirable commodity* (e.g., what drives willingness to collaborate, how train ticket and lodging price volatility impacts on bookings can be predicted, etc.) and how shifts in agency preferences and personal choices occur in the presence of data that yields greater informedness.

research. An exception is the work of Díaz and Doolin (2016, 2019), which illustrates how IS can strengthen the inclusion of refugees in society. They highlight how ICT provides a bridge to participation, social connectivity, and cultural understanding. Leong et al. (2015) examined the 2011 flooding in Thailand and how the ICT-based response by rural communities enabled effective local crisis management for people displaced from their homes.

In the present study, we focus on how social media can support charitable activities. Miranda et al. (2016) illustrated the emancipatory and hegemonic potential of social media and called for IS researchers to explore social media's effects on public discourse. Since our work is on the spread of tweets rather than the ideological stance communicated, our work does not emphasize whether the effects of the tweets are emancipatory or hegemonic. We next discuss charitable activity research in IS and related social science disciplines.

Charitable activities involve actions that people make to extend money, time, effort, service, or expertise to perceived worthy causes related to, for example, others' lack of resources, capabilities, good health, or challenging social situations. Beyond the recent surge in theoretical and empirical research on crowdfunding (Burtch et al., 2016), there is relatively little IS research on charitable activities. For example, in a retrospective, Sarker et al. (2020) only identified one paper published on this topic between 2013 and 2019 in the *Journal of the Association for Information Systems*—an evaluation of a fictitious charity website exploring the impact of campaigns' halo effects on donors' initial perceptions and subsequent willingness to contribute.

Other authors have examined donations in the crowdfunding context, offering useful findings that support our research (Kwak et al., 2019; Lee et al., 2018). Gleasure and Feller (2016) explored the extent to which giving in the crowdfunding context is a matter of passionate commitment vs. reasoned investment on the part of donors—in essence, *heart vs. head* decision-making. They found that, in contrast to traditional charities, crowdfunding donations are not driven by considerations such as guilt avoidance, reciprocal exchanges, image and reputation, vicarious enjoyment, or group-level benefits, as is the case with traditional charities. Rather, they concluded that the heart is more important than the head in making such decisions. Given the large number of emergencies worthy of attention, this research supports the importance of using recognizable social influencers to drive emotional responses to refugee emergencies for organizations like UNHCR.

Other technology research focused on charities has been presented at human-computing interaction conferences. For example, Schetgen et al. (2020) applied data science methods to investigate whether social media data can predict donation behavior. They emphasized the capabilities of various predictive variables, as well as the performance of multiple dimensionality-reduction and data-classification techniques. Also, Dokhanchi et al. (2019) addressed the relationship between the public sphere and social campaigns that include philanthropic activities carried out on digital platforms like Twitter.

Additionally, the marketing domain has devoted significant attention to the study of charitable action. Peng et al. (2022), for instance, studied individual information and donor propensities in making charitable monetary gifts prompted by information-bearing nudges to influence their behavior. The authors reported that positive sentiment and donation amount go hand-in-hand. A simulated charitable fundraising campaign experiment on social media by Xiao et al. (2021) further supports this view by emphasizing the role of information and how it is presented in philanthropic donations. The authors focused on abstract vs. concrete outcomes and loss-or-gain framing to emphasize desirable and undesirable outcomes. They found that perceived message credibility, perceived transparency, cognitive elaboration, and information presentation were more compelling in increasing donations; alternative framing, surprisingly, had no effect.[9]

Further, ample marketing research has examined the design of fundraising campaigns and the profiles of donors. Kim et al. (2021) emphasized how nonprofit organizations typically reach out for various monetary forms of charitable donations (e.g., memberships, donations, planned giving), finding that fundraising campaigns should address the informational needs of participants as the nature of their altruism changes over time. In addition, Kessler and Milkman (2018) suggested that fundraising campaigns that leverage donors' identities are often associated with increased donation rates. In this vein, Goswami and Urminsky (2016) conducted lab experiments and large-scale field studies to examine whether setting donation choice options (as defaults) had salient effects on increasing fundraising amounts. They reported that defaults interfered with other cues, such as the prosocial nature of a charitable campaign. Similar kinds of market behavior in rewards-based crowdfunding have also been studied in IS research. For example, Wessel et al. (2021) examined the use of minimum fundraising goals as defaults to protect donors from campaign failures caused by underfunding, when creators try to make do with insufficient funds, only to fail later.

[9] See Liu et al. (2018) for additional research findings on empathy, credibility, and individual donation behavior in charitable crowdfunding, and Chang and Lee (2008) for research on advertising framing for charitable donations.

3.2 UNHCR and Its Social Media Influencers

UNHCR operates with different categories of refugees and displaced persons. Regardless of the reason for their situations, we refer to them all as *refugees* here for simplicity. UNHCR's approach is to assist refugees with administrative, legal, health, and daily needs.[10] Their work is financed by voluntary contributions rather than state budgets. Over the years, UNHCR has been successful in its fundraising activities; however, the motivation for donations has varied from altruism to opportunism (Hantscher, 2019). Most funds have come from developed nations, especially the European Union, with about 10% of the total budget obtained from private donations (UNHCR, 2020a). In 2019, for example, the budget was US$8.6 billion, but there is clearly a need for much more funding. The UN (2020) reported that 1% of the world's population are refugees—almost 80 million people. This number doubled from 41.1 million people in 2010 to 79.5 million in 2019, and UNHCR has estimated an annual growth of 10 million refugees in recent years.[11] Providing support for refugees' access to information has also increased the financial needs of UNHCR and similar NGOs.

To meet its funding goals, UNHCR has chosen to connect with high-profile individuals to improve awareness of the scope of the challenges associated with its refugee assistance activities. It works with different types of representatives who seek to inform and influence the public through their advocacy and participative activities. The different roles of envoy, goodwill ambassador, and high-profile supporter, along with their focal activities (see Table 1).

UNHCR's selected influencers are a mix of sports stars, actors, directors, and former refugees. Current research by Perez-Cepeda and Arias-Bolzmann (2021) suggests that Twitter is a preferred channel for conversations related to refugee issues—including communication from refugees about their situations. We focus on how UNHCR's influencers communicate and create awareness about its work and about refugee situations around the world.

[10] The Statute of the UNHCR of 1950 was ratified in 1954 in the aftermath of World War II. The UN's perspective was that the need to support refugees would be temporally limited so that the UNHCR would have a short lifespan. However, since it was established, the UNHCR has had constant demand for its services. Moreover, today it employs 17,000+ people, of which 90% are in the field in 135 nations across the globe.

[11] Overall, 85% of refugees have been able to find refuge in developing countries, but this has challenged the ability of the host countries to live up to the minimum standards of treatment outlined in the UNHCR's statute. They include offering access to basic protection, health services, and education. Local constraints of the hosting nations also limit access to technologies such as laptops, tablets, mobile phones, and internet infrastructure.

4 Framework and Propositions on Social Informedness

We next develop a framework based on our proposed social informedness theory to address issues related to how champions create awareness about key refugee-related emergencies and other issues.

4.1 A Framework for UNHCR's Application of Social Informedness Theory

Understanding how communication creates a basis for changing the awareness of refugee issues of interest is important. The theoretical perspective we use in this work is presented based on a sequence of general relationships, which we assess in several different ways using data analytics methods from the ML and NLP domains and a dataset acquired from UNHCR.

Our framework posits that *observable tweets* posted about refugee emergencies (*issues*) and their champions and followers (*communicators with stakeholder characteristics*) are appropriate for empirical studies such as ours, which was developed in cooperation with UNHCR (see Figure 1). The type and intensity of refugee emergencies, along with the type and participation intensity of their champions and followers, are useful for assessing the benefits of UNHCR's strategy with respect to selected influencers.

The tweets we captured support the supervised and unsupervised ML and NLP methods used to measure the valence and strength of social sentiment. Semantic similarity measures further enabled the data mining of social sentiment and attitudes present in the sharing of prosocial communication via Twitter. These measures can proxy for the positivity of the tweets and interactions of champions and their followers and can be used as a rough gauge of the effectiveness of UNHCR's selected high-profile influencers (e.g., analogous to project backers in crowdfunding and influencers that affect stock prices in volatile markets).[12]

[12] An example supporting the generalizability of the empirical data analysis methods we have chosen to employ, occurred in 2021 (Su, 2021; Kim et al., 2023). The publicly traded firm, GameStop, was a target of millennial traders who participated in a social media subreddit (group) called "WallStreetBets" within Reddit. They engaged in speculative trading to support GameStop's price after it fell sharply, though institutional investors had "shorted" the stock—agreeing, in effect, to buy it in the future—never expecting higher GameStop stock prices. This episode in the financial markets is different based on the context, but similar in the relevant measures and data analytics methods for studying social sentiment as a driver of trading patterns over millions of trades.

Table 1. UNHCR's Supporters and Their Roles

Descriptors	Special envoy	Goodwill ambassadors	High-profile supporters
Role	Undertakes advocacy and represents UNHCR and High Commissioner at the diplomatic level	Internationally recognized people who bring UNHCR's message to every corner of the world through their influence: Ben Stiller, Yusra Mardini, Alphonso Davies, Kat Graham, Maya Ghazal, Nikki Samonas, Nujeen Mustafa, Emi Mahmoud	World-recognized profiles of leading people who are micro-influencers at local and regional levels
Profiles in our data sample	Angelina Jolie		Asmir Begovic, Douglas Booth, Dianna Agron, Mo Salah

Issue and stakeholder characteristics and tweets	leading to →	Metrics from ML and NLP analysis methods with bidirectional encoder words transforms (BERT) and Global Vectors of Words (GloVe) for tweets	which create →	Sentiment & attitude outcome measures	proxying for →	Champion and follower positivity that correlates with UNHCR's view of effectiveness in its social media strategy program

Note: We discuss BERT sentiment and attitude analysis, and GloVe global word-vector semantic similarity later.

Figure 1. A Framework for the Theory of Social Informedness-Based Relationships

Table 2. A Framework for Champions' and Their Followers' Social Informedness Impacts

Propositions		**Tweet valence and values**	**Sentiment and attitudes**
***Supervised methods* (*BERT*)**			
P1a	Refugee-emergency type + champion & follower tweets (A)	→ Tweet valence value (B)	→ Sentiment (D)
P1b	Refugee-emergency type + champion & follower tweets (A)	→	→ Attitude (E)
***Unsupervised methods* (*GloVe*)**			
P2a	Refugee-emergency intensity + champion & follower tweets (A)	→ Tweet semantic similarity (C)	→ Sentiment (D)
P2b	Refugee-emergency intensity + champion & follower tweets (A)	→	→ Attitude (E)
***Supervised methods* (*BERT*)**			
P3a	Refugee-emergency participation + champion & follower tweets (A)	→ Tweet valence value (B)	→ Sentiment (D)
P3b	Refugee-emergency participation + champion & follower tweets (A)	→	→ Attitude (E)
***Unsupervised methods* (*GloVe*)**			
P4a	Champion participation intensity + Champion & follower tweets (A)	→ Tweet semantic similarity (C)	→ Sentiment (D)
P4b	Champion participation intensity + Champion & follower tweets (A)	→	→ Attitude (E)

Note: P1-P4 assert how *tweets of champions and their followers* (A) can be analyzed with data science methods to determine tweet *valence values* (B) and *tweet semantic similarities* (C) for *sentiment* (D) and *attitude* (E) in the Twitter corpus, for refugee-emergency type and refugee-emergency intensity, respectively. We use arrows above (→) to represent "*can be analyzed with data science methods to determine …*" We use tweet valances and semantic similarity to gauge the extent of consistently positive social media communication with influencers supports greater informedness about refugee-emergencies, elevated armed conflicts, and political unrest.

This approach can be used as an evidentiary basis for assessing UNHCR's social media outreach strategy. We also offer a framework that represents details of the thinking behind our approach (see Table 2). It offers constructs for emergencies, stakeholders, data, measures, outcomes, and proxies to provide evidence of UNHCR champions' effectiveness and blends the analytics methods used. Our first analysis is related to different refugee emergencies, based on their locations in the world. We treated them as different *refugee-emergency types* based on their respective regions. These regions comprise 15 countries that have experienced problems due to various armed conflicts and political instabilities. We offer background information on UNHCR's refugee situations and emergencies, along with evidence for their intensity levels (see Appendix Table A1).

4.2 Combining Social Informedness Theory Propositions with Data Analytics Methods

The propositions explored here are related to observable aspects of social communication data that we were able to gather on UNHCR's influencers and their followers about UNHCR's core mission: assisting refugees from countries experiencing armed conflicts and political unrest resulting in displaced citizens and refugee emergencies.

4.2.1 Refugee-Emergency Type and Intensity

The groups for which we have data include refugee emergencies in nine African countries, three Middle Eastern countries, two Asian countries, and one South American country.[13] This enabled us to examine the tweet activities and their valence for UNHCR champions and their followers according to the refugee emergency for sentiment (positive, neutral, negative) and attitudes expressed (sharing, liking, unrelated, helplessness, disliking). While we expected to observe positive tweet patterns for most of the refugee emergencies, some were more likely to be picked up by champions and their followers in the available communications. We further expected to see similar patterns of social communication for the refugee-emergency types related to global word vectors for tweet similarity between champions' followers. Positive tweets are likely to be beneficial for UNHCR in terms of increasing followers' awareness of a refugee conflict and the general need for contributions. Our first proposition set and those that follow focus on *refugee-emergency types*.[14]

Proposition 1a (Refugee-emergency type and sentiment patterns proposition): *The posted tweets of champions and their followers revealing the type of refugee emergency display discernable patterns of sentiment and attitudes discoverable through ML-based measures of text valence.*

Proposition 1b (Refugee-emergency type and influencer-follower similarity proposition): *The posted tweets of champions and their followers revealing the type of refugee emergency display discernable patterns of tweet similarity discoverable through measures leveraging global word vectors.*

We further considered two analogously formulated propositions that center on the intensity of refugee situations and emergencies. We assigned the countries as having *high-intensity* and *low-intensity conflicts*, based on their duration and the number of refugees displaced. We coded for intensity levels by splitting the data by country based on the median levels observed across all countries. As emergencies reach a greater level of intensity, it becomes likely that positive-valence sentiment, as well as similar, consistent, and more positive tweets by champions and their followers, will yield more positive patterns of social sentiment and communication. This will make UNHCR's purposes more apparent to followers, who will be more likely to donate to the UNHCR. The following set of propositions assesses sentiment patterns and champions' social communication using global word vectors for tweet similarity.

Proposition 2a (Refugee-emergency intensity and sentiment patterns proposition): *The posted tweets of champions and their followers revealing the intensity of a refugee emergency display discernable sentiment and attitudes discoverable through ML-based measures of text valence.*

Proposition 2b (Refugee-emergency intensity level and influencer-follower similarity proposition): *The posted tweets of champions and their followers revealing the intensity of a refugee emergency display discernable patterns of tweet similarity discoverable through measures leveraging global word vectors.*

4.2.2 Champion Type and Participation Intensity

Our third set of propositions is related to the kinds of champion types represented in UNHCR's Twitter corpus, based on their career roles and past experiences. They include actors/directors, sports stars, and former refugees. We offer additional background on the champions' role types and their participation-intensity levels (see Appendix Table A2). The logic of our analysis is as described above, only this time, we are interested in learning about the extent of positivity and volume of tweets and retweets associated with the individual champions and their followers:

[13] The countries are Group 1—Burkina Faso, Mali, Niger, the Sahel Region, Ethiopia, Nigeria, Central African Republic, Burundi; South Sudan; and the Democratic Republic of the Congo; Group 2—Iraq, Syria, and Yemen; Group 3—Bangladesh and Myanmar; and Group 4—Venezuela. There are others the UNHCR has tracked (e.g., the Ukraine) but we could not access their data sources.

[14] We use the term *refugee emergency* without loss of generality, though the UNHCR has defined other *refugee situations.*

Proposition 3a (Champion type and followers' sentiment and attitude patterns proposition): *The posted tweets of different types of champions and their followers display discernable patterns of sentiment and attitudes discoverable through ML-based measures of text valence.*

Proposition 3b (Champion type and followers' tweet similarity proposition). *The posted tweets of different types of champions and their followers display discernable patterns of tweet similarity discoverable through measures leveraging global word vectors.*

The final set of propositions focuses on the participation intensity of the champions and their followers in social media communication regarding various refugee-emergency activities.

Proposition 4a (Champions' participation intensity, and sentiment and attitude patterns proposition): *The posted tweets of champions and their followers revealing different levels of participation intensity display discernable sentiment patterns and attitudes discoverable through ML-based measures of text valence.*

Proposition 4b (Champions' participation intensity and influencer-follower similarity proposition): *The posted tweets of champions and their followers revealing different levels of participation intensity display discernable patterns of tweet similarity discoverable through measures that leverage global word vectors.*

5 Data and Methodology

Our empirical research design involved user tweet and retweet data captured for this research project. We benefited from guidance from our organizational sponsor at the Liaison Office for UNHCR Northern Europe, located in UN City, Copenhagen, Denmark. In the first phase of the study, the conversations with the Liaison Office focused on the best possible approach to perform a study on "Data Science for Social Good." After presenting an initial analysis, we conducted detailed interviews to gain a deeper understanding of the role and function of ambassadors. During the interview, it became clear that the primary function of ambassadors is to create awareness of the refugee crisis rather than engage in direct fundraising. This enabled us to align our study with the newly proposed social informedness theory. Field notes from the interviews supported our presentation of the UNHCR refugee-emergency case and helped us define our research questions.[15]

5.1 Dataset Collection and Description

Our dataset focused on Twitter data attribute values for all tweet activities on UNHCR's refugee projects related to armed conflicts and political unrest. The dataset has tweet attributes and user values for UNHCR's refugee emergencies and contains 5.5 million data points for 2008-2021 (see Tables 3 and 4).[16]

We collected the Twitter profiles of UNHCR goodwill ambassadors, envoys, and high-profile supporters. We also obtained tweets from UNHCR's official Twitter account, UNHCR@Refugees. For several reasons, when collecting data for champions, we followed a different tactic by using four search queries for each champion (see Table 5 and Appendix Table B1).

First, some champions did not have an active personal Twitter profile.[17] Second, even if a champion had one, their tweet activities may have been mostly confined to their professional activities: for example, sports stars mostly tweet about sports. Third, UNHCR focuses on engaging champions through specific activities—such as visits to their camps to motivate the refugees—rather than limiting their engagement to social media communications. When the champions made field visits, UNHCR's outreach campaigns followed them on Twitter to spread awareness by tagging the related tweets and posting them on UNHCR@Refugees. Fourth, we also focused on measuring champions' engagement potential and presence on social media platforms, both within and outside the UNHCR sphere. This enabled us to assess the champions' effectiveness in spreading awareness about humanitarian activities.

Thus, from the champions' Twitter profiles, we identified tweets containing refugee- and UNHCR-related words. To measure their presence in UNHCR activities on Twitter, we collected tweets from those who mentioned both UNHCR and its champions together. Similarly, to measure the champions' personal or professional presence, we separately collected tweets that mentioned them without any reference to UNHCR-related activities.

[15] We also utilized the Twitter research API that returns data encoded using JavaScript object notation (JSON) in relational data tables (Twitter, 2020) (see Table 3). It yielded data attribute name-and-value pairs that may change states over time, to describe the objects of research interest—including tweets, retweets, users, and their followers.

[16] There are many tools to extract data from Twitter, including those requiring programming (e.g., Python), and browser interface, or purpose-built software. We wrote custom scripts in Python to collect data from the Twitter premium API.

[17] They included Angelina Jolie, Douglas Booth, and Mo Salah. The UNHCR views them as participating in refugee outreach from their overall activities though, and their many followers (loosely defined, rather than operationally defined by their Twitter posts) tweet related to what the champions have been doing with respect to refugee situations and emergencies.

Table 3. Data Attributes Collected from the Twitter JSON

Attribute	Description
Attributes of tweets	
Tweet_ID	Unique identifier for a tweet
Created_Date	Time when tweet was created
Tweet_Text	UTF-8 text of the status update
Tweet_URL	URL for the tweet
Retweet_Count	Number of times a tweet was retweeted
Favorite_Count	Number of likes/favorites
Reply_Count	Number of times tweet was replied to
Quote_Count	Approximate number of times tweet was quoted by Twitter users
Attributes of Twitter users	
Tweet_Username	Twitter username
Followers_Count	Number of followers Twitter user account currently has
Friends_Count	Number of users Twitter user account follows
Listed_Count	Number of public lists on which Twitter user is a member
User_Favorites_Count	Number of tweets Twitter user has liked over the account's lifetime
User_Status_Count	Number of tweets (including retweets) posted by Twitter user

Note: JSON dataset: 5.5 million tweets. Included retweet counts and users, as well as favorites, replies, and quote counts. Further, tweet post attributes were time-stamped, while user attributes only supported current queries.

Table 4. Dataset Description: Champions' Own Tweets and Their Presence on Twitter

Champions' names	Champion's own tweets with UNHCR	Champion's own tweets without UNHCR	Presence of champions with UNHCR	Presence of champions without UNHCR
UNHCR@Refugees	91,332	—	—	—
Alphonso Davies	15	152	3,981	196,755
Angelina Jolie	—	—	78,925	1,581,767
Asmir Begovic	15	4,640	253	169,836
Ben Stiller	152	3,400	8,602	555,391
Dianna Agron	6	3,000	191	183,174
Douglas Booth	0	7	1,057	68,439
Emi Mahmoud	29	90	3,119	3,326
Kat Graham	114	4,899	10,931	906,957
Maya Ghazal	33	58	1,377	701
Mo Salah	1	271	1,800	1,419,088
Nikki Samonas	55	10,045	178	160,062
Nujeen Mustafa	16	99	1,115	3,237
Yusra Mardini	36	26	14,332	6,026
Totals	91,804	26,687	125,854	5,254,752

Table 5. Twitter Profiles of UNHCR Ambassadors, Envoys and High-Profile Supporters

Champions' names	*Tweet _Count*	*Retweet _Count* [(a)]	*Followers _Count*	*Friends _Count*	*Listed _Count*
UNHCR@Refugees	86,696	41	2,639,435	1,956	13,958
Angelina Jolie[(b)]	—	—	—	—	—
Ben Stiller	8,232	9	5,621,026	1,469	26,533
Yusra Mardini	177	8	17,476	171	96
Alphonso Davies	356	12	318,197	40	445
Asmir Begovic	7,321	10	291,307	496	863
Douglas Booth[(c)]	8	22	325	52	0
Dianna Agron	3,815	11	1,775,687	477	12,767
Mo Salah[(d)]	1,527	13	15,293,381	148	5,986
Emi Mahmoud	173	10	5,962	225	36
Kat Graham	24,368	52	1,953,362	165	7,378

Maya Ghazal	546	4	1952	317	15
Nujeen Mustafa	460	9	4137	146	41
Nikki Samonas	54,022	20	256,395	826	90

Note: [a]*Retweet_Count* is an attribute on the tweet level; the other attributes are at the user level. Most of UNHCR's champions were not active in retweeting. [b]Angelina Jolie did not have an active Twitter profile, though she has mentioned Twitter often. We investigated how UNHCR tagged her in Twitter activities and found that her full name was used. Her data are from the search string "UNHCR Angelina Jolie," which produced the tweets where she is mentioned, along with UNHCR. [c]Douglas Booth had a Twitter profile, but it was inactive, while [d] Mo Salah did have an active Twitter account, but it did not show any affiliation to UNHCR. The followers of champions were relatively more active than the champions.

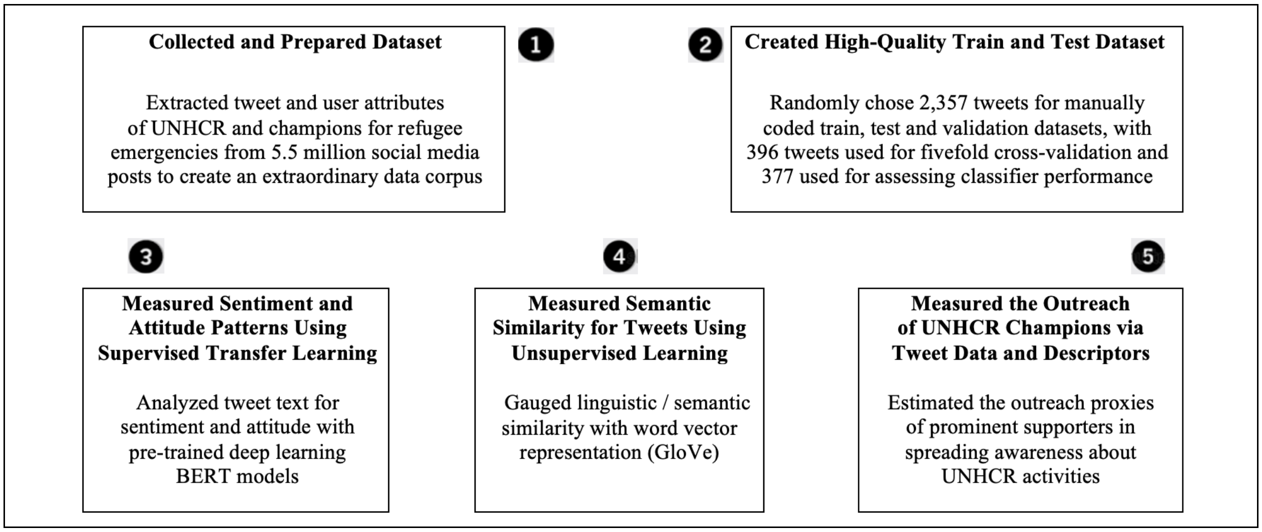

Figure 2. Five-Step Design Science Data Analytics Framework for Social Communication

Tweet attributes enabled us to gauge the intensity of posts related to refugee emergencies made by champions, followers, and friends from one month before to one month after the focal event. For this purpose, the features available for our analysis were the *Tweet_ID*, *Tweet_Username*, and the *Tweet_Text.*[18]

5.2 Methods Used in Our Five-Step Design Science Data Analytics Framework

Our research applied state-of-the-art data science methods based on supervised and unsupervised ML methods related to a textual corpus that required processing with NLP tools. These methods enabled us to extract words and word vectors from the tweet data that permitted our identification of sentiment valence (e.g., positive, neutral, or negative), as well as the linguistic similarities of tweets and the correlations between champions and their followers on the various refugee emergencies tracked by UNHCR. We next discuss our methods for data collection, modeling, and data analytics. We also offer a road map for the multistep methods sequence implemented in each of the five steps (see Figure 2).

For this work, we used pretrained text-classification models to create the basis for textual data analytics for Twitter user sentiment and attitudes (see Table 6). In Step 1 (dataset collection and preparation), we constructed the basic tweet dataset that we used in this research. Then, in Step 2 (high-quality dataset training and testing), we measured sentiment and attitude patterns and refined our data into a gold-standard, labeled, train-and-test dataset, following the manual content analysis approach guidelines of Morris (1994). We collected a random sample of 2,357 tweets from the corpus, labeled for the sentiment and attitude models by independent coders.

We further used the Cohen kappa (κ) statistic to guide our measurement of the reliability of the coding and the extent of intercoder agreement (Fleiss, 1981). The κ-values for coding sentiment and attitude models were 0.746 and 0.705, respectively, indicating reasonable agreement between coders in assessing the tweets based on our κ-value calculations (see Table 7).[19]

[18] We briefly discuss our text analysis methods to explain how we used supervised and unsupervised learning approaches to extract linguistic indicators of sentiment valence and global word vectors to establish semantic similarities in textual content. Data attribute values such as *Favorite_Count*, *Reply_Count*, and *Quote_Count* are constructs that proxy for *topic awareness* in related posts, and thus, tie in directly with our social informedness construct.

[19] This author suggests that intercoder reliability should be around 75% to enable a researcher to confirm that there was a reasonable, albeit still imperfect degree of agreement between the coders—a criterion that our dataset met.

Table 6. Text Classification Models: Sentiment and Attitude

Label	Explanation
Sentiment	
Negative	Direct criticism of the population of concern or UNHCR and its social actions.
Positive	Social posts that appreciate or compliment some person of concern, action, or initiative from an individual, or posts thanking UNHCR.
Neutral	Sharing UNHCR content or discussing current events, UNHCR-issued statistics, etc.; the posts typically do not express an opinion about any specific people.
Attitude	
Disliking	Expressing dislike of UNHCR's population of concern or disliking the work or an action of UNHCR or a UNHCR staff member.
Helpless	Expressing helplessness or asking for help.
Sharing	Just sharing some content from UNHCR.
Liking	Expressing that they like an action or initiative or sharing positive comments about UNHCR's population of concern or UNHCR's work.
Unrelated	Posting tweets unrelated to the population of concern or UNHCR.

Table 7. Intercoder Reliability and Cohen's κ for Sentiment Classification

Coder	Label	Coder 2				Cohen's κ
		Negative	*Positive*	*Neutral*	Total	
Sentiment						
Coder 1	*Negative*	701	28	59	788	p_o =0.834, p_c = 0.347 $\kappa = \frac{p_o - p_c}{1 - p_c}$ = 0.746
	Positive	32	455	121	608	
	Neutral	89	63	809	961	
	Total	822	546	989	2,357	0.746

Coder	Label	Coder 2						Cohen's κ
		Disliking	*Helpless*	*Sharing*	*Liking*	*Unrelated*	Total	
Attitude								
Coder 1	*Disliking*	249	82	16	11	29	387	p_o = 0.766, p_c = 0.207, $\kappa = \frac{p_o - p_c}{1 - p_c}$ = 0.705
	Helpless	51	503	19	25	19	617	
	Sharing	13	15	420	52	31	531	
	Liking	14	21	38	291	27	391	
	Unrelated	38	21	12	17	343	431	
	Total	365	642	505	396	449	2,357	0.705

Out of 2,357 tweets in the manually coded training dataset, we used 377 tweets for the test dataset and used the remaining 1,980 tweets to train the algorithms and test the performance of the models using fivefold cross-validation. For the validation, 20% of the 1,980 tweets were used; the validation data were rotated in each iteration. After training the models on the training datasets, we assessed our test dataset of 377 tweets to further evaluate the classifiers' performance, thereby avoiding overfitting the models and producing more generalizable results.

5.3 Modeling and Measuring Tweet Valence Patterns with Supervised Transfer Learning

Step 3 (measuring sentiment and attitude patterns using supervised transfer learning) supported the analysis of tweet text for models on the sentiment and attitudes of the relevant Twitter users. We used a supervised learning approach that involved mapping data inputs to outputs in a textual corpus based on ML analysis of training data. This was done so that the classification of interest—tweet valence in our case—could be identified. In applications of this method, the outcomes are subject to analysis procedure issues and shortcomings, including different kinds of data (continuous, discrete, and count values, etc.) and data interactions, nonlinear relationships and noisy variables, bias-variance with training data, and the dimensionality of the input data relative to the learning target (Hastie et al., 2009). ML tools for supervised learning circumvent these problems to some extent, though they persist when the setting and data are problematic.

We also applied a language representation approach called the *bidirectional encoder representations from transformers* (BERT) model (Devlin et al., 2018) from Google's AI Language Research Group. It was developed to offer NLP pretraining to improve Google's search engine performance for English language searches. In contrast to other context-free NLP models, BERT takes advantage of look-back and look-forward word assessments to improve its contextual understanding. This DL method was implemented via the layers of neural networks applied in this kind of data analysis (Goodfellow et al., 2016).[20]

We used contextual word embedding and a one-hidden-layer neural network classifier to perform supervised text classification using our own predefined training datasets in order to understand champions' and their followers' sentiment and attitudes. Our tweet classification for these models resulted in much better performance, with an accuracy of 80% and 68%, respectively. We also benchmarked the BERT classifier against five other ML text classification algorithms and used our train-and-test datasets to train them. We also included details and performance metrics for the classifiers (see Table 8). Overall, the BERT models performed best and resulted in higher accuracies at 10% - 15%.

5.4 Measurement of Semantic Similarity Using Unsupervised Learning

Based on the supervised learning approach that we used to understand tweet sentiment and attitudes, in Step 4 (measuring semantic similarity for tweets using unsupervised learning), we shifted to unsupervised learning (Chung et al., 2018). For this, we applied the *global vectors for word representation* (GloVe) model (Pennington et al., 2014), which supports the measurement of linguistic and semantic similarities related to word-vector representations. This connects champions to UNHCR's refugee emergencies and gauges the similarities in the sentiment and attitudes of their followers.

Word vectors are rows of numeric values in a matrix that capture a specific dimension of a word's meaning, based on linguistically similar words having similar numeric word-vector representations (Ganegedara, 2019).[21] This method extracts similarities and differences between word vectors by leveraging their semantic distances for the tweets of different users, a single user at different times, and so on, based on the textual corpus used. This made it ideal for us to support different kinds of comparisons: between champions, and between champions and their followers.

5.5 Measurement of the Outreach of UNHCR's Champions

For Step 5 (measuring the outreach of UNHCR champions via tweet data and some descriptors), we used different metrics from Twitter, which enabled us to assess how much content champions produced over time, both directly and through their followers. We further gauged whether metrics beyond champions' own posts (if any) reflected the extent of their participation in UNHCR social media outreach. In measuring outreach, we used three key metrics for the attributes of the tweets—*Retweet_Count*, *Favorite_Count*, and *Reply_Count*—which helped us determine the extent to which a champion influenced the tweet content of their followers. We further aggregated the values of the metrics to compare outreach levels across champions.

6 Results Related to UNHCR's Social Media Strategy

We next present the data analytics results obtained for the multiple sets of propositions to assess the social media activities of UNHCR's champions and their followers. The next two subsections blend the processes used in Steps 3 and 4 to obtain the overall findings regarding our proposed theory. Findings associated with Step 5 are presented later.

6.1 Results for the Refugee-Emergency Type Propositions (P1 and P2)

We first discuss the propositions on refugee-emergency type related to sentiment patterns (P1a) and influencer-follower tweet similarity patterns (P1b), as well as the refugee-emergency intensity level related to tweet sentiment patterns (P2a) and influencer-follower tweet similarities (P2b). The results indicate that the Twitter communication strategy adopted by UNHCR and the champions among the various emergency types is not uniform (see Table 9).

Some refugee issues associated with high emergency-intensity in certain regions—especially Syria and Iraq in the Middle East and the Rohingya emergency in Myanmar and Bangladesh—engendered more engagement than other emergencies. However, the same degree of engagement was observed for UNHCR's champions, whose coverage is limited, and may thus be due to UNHCR's priorities.

Next, we discuss sentiment and attitude patterns for a subset of emergencies. *Positive* and *neutral* sentiment and *sharing* and *liking* attitudes were the most frequently observed sentiment and attitudes for the refugee emergencies (see Table 10). Regarding the Iraq refugee emergency, however, some champions' tweets displayed *negative* sentiment and *helpless* or *disliking* attitudes (see Table 11).

[20] As an example, consider these two sentences: "The sick CFO is managing the company's acquisitions day to day" versus "The sick patient is managing a chronic illness day to day." BERT represents an improvement over noncontextual approaches (e.g., Word2Vec and GloVe), based on its forward and backward recognition of embedded words. The others focus on words or word vectors without context and do not distinguish the nuanced meanings of the word "managing."

[21] GloVe's innovation is to capture word co-occurrences in tweets or documents, as opposed to Word2vec's focus on local contextual information for individual words. Word vectors are high-dimensional and cannot be reduced to scalar metrics.

Table 8. Benchmarks: BERT Classifiers for Sentiment and Attitude Classification

Label	Logistic regression classifier			Naive Bayes classifier			BERT classifier		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Sentiment classification									
Negative	0.82	0.65	0.73	0.83	0.68	0.75	0.86	0.84	0.85
Positive	0.77	0.36	0.49	0.55	0.88	0.68	0.76	0.80	0.78
Neutral	0.57	0.85	0.68	0.80	0.10	0.18	0.79	0.75	0.77
Accuracy	—	—	0.66	—	—	0.64	—	—	0.80
Macro avg.	0.72	0.62	0.63	0.73	0.55	0.54	0.80	0.79	0.80
Weighted avg.	0.70	0.66	0.65	0.71	0.64	0.60	0.80	0.80	0.80
Attitude classification									
Sharing	0.49	0.81	0.61	0.46	0.75	0.57	0.62	0.65	0.64
Liking	0.60	0.61	0.60	0.86	0.36	0.51	0.73	0.66	0.69
Unrelated	0.62	0.54	0.58	1.00	0.21	0.35	0.67	0.56	0.61
Helpless	0.81	0.80	0.80	0.46	0.96	0.62	0.81	0.82	0.81
Disliking	0.57	0.75	0.65	0.70	0.19	0.29	0.53	0.65	0.58
Accuracy	—	—	0.63	—	—	0.53	—	—	0.68
Macro avg.	0.58	0.52	0.52	0.70	0.50	0.47	0.67	0.67	0.67
Weight avg.	0.63	0.63	0.61	0.69	0.53	0.48	0.68	0.68	0.68

Note: We benchmarked the BERT classifiers for sentiment and attitude, by applying five ML algorithms for both with the Scikit-learn and NLTK libraries of Python. They are the *multinomial naive Bayes*, *linear SVC*, *logistic regression*, *passive-aggressive*, and *SVM with stochastic gradient descent* (SGD) *classifiers*. To benchmark the BERT classifier for sentiment, we used the results from logistic regression and multinomial naive Bayes in the table. The performance measures of all other classifiers were in the same range with only 2%-3% variation. To compare the attitude classifiers, we used results from the logistic regression and multinomial naive Bayes classifiers only. Their performance had a 1% - 2% variation compared to the chosen classifier algorithms.

Table 9. Refugee-Emergency Type vs. Champion-Participation Intensity (High-H, Med-M)

Refugee emergency	UNHCR @Refugees	Angelina Jolie (H)	Ben Stiller (H)	Yusra Mardini (H)	Emi Mahmoud (M)	Maya Ghazal (M)	Kat Graham (M)
Syria (H)	11,389	25,347	1,325	6,858	28	389	52
Iraq (H)	2,355	5,814	557	87	522	266	193
Rohingya (H)	1,907	1,131	16	2	9	1	6
Yemen (H)	1,757	368	21	2	3	3	6
Ethiopia (M)	1,283	21	5	2	10	3	320
Venezuela (H)	1,220	597	10	1	1	1	0
South Sudan (H)	1,213	27	5	4	232	0	6
Nigeria (H)	766	95	10	5	6	1	1
Burundi (M)	765	5	0	0	0	0	0
Dem. Rep. Congo (H)	244	6	1	0	0	0	0
Central Afr. Rep. (M)	235	0	0	0	0	0	0
Sahel (H)	168	23	0	3	0	0	0

Note: Columns for the rest of the champions, Nujeen Mustafa (517 tweets on Syria), Mo Salah (177 retweets on Iraq), Alphonso Davies (100 tweets on Iraq), Douglas Booth (65 tweets on Iraq), Diana Agron (28 tweets on Syria), Asmir Begovic (4 tweets on Iraq), and Nikki Samonas (2 tweets on Iraq), are not included in the table.

Table 10. Emergency-Type Sentiment Patterns

Champions	*Positive*	*Neutral*	*Negative*	Total
Syria				
Angelina Jolie	2,409	22,418	520	25,347
UNHCR@Refugees	2,036	9,139	214	11,389
Yusra Mardini	1,249	5,588	21	6,858
Ben Stiller	334	843	148	1,325
All others	457	579	52	1,088
Iraq				
Angelina Jolie	555	5,192	67	5,814
UNHCR@Refugees	285	1,978	92	2,355
Ben Stiller	5	36	516	557
Emi Mahmoud	5	4	513	522
Maya Ghazal	1	10	255	266
Kat Graham	4	3	186	193
All others	13	69	361	443

***Rohingya* (*Myanmar and Bangladesh*)**				
UNHCR@Refugees	516	1,374	17	1,907
Angelina Jolie	194	917	20	1,131
All others	13	12	11	36
Ben Stiller	6	7	2	15

Table 11. High-Intensity Emergency Attitude Patterns

Champions	*Sharing*	*Liking*	*Unrelated*	*Helpless*	*Disliking*	Total
Syria						
Angelina Jolie	20,396	2,401	734	127	1,689	25,347
UNHCR@Refugees	5,599	3,106	1,075	691	918	11,389
Yusra Mardini	5,103	1,587	53	15	96	6,854
Ben Stiller	507	301	146	81	278	1,313
All others	460	511	11	50	56	1,088
Iraq						
Angelina Jolie	4,635	587	231	62	299	5,814
UNHCR@Refugees	1,012	501	226	447	169	2,355
Ben Stiller	4	6	1	513	33	557
Emi Mahmoud	0	4	0	507	11	522
Maya Ghazal	1	0	1	248	16	266
Kat Graham	0	0	0	179	14	193
All others	36	18	9	334	46	443
***Rohingya* (*Myanmar/Bangladesh*)**						
UNHCR@Refugees	1,158	560	37	22	130	1,907
Angelina Jolie	884	132	17	2	96	1,131
Ben Stiller	4	6	1	1	3	15
All others	9	12	2	3	10	36

The following is a sample tweet representing *negative* sentiment and a *helpless* attitude:

> *@UNRefugeeAgency @Refugees @RedHourBen UNHCR and humanitarian organizations please ... We are Iraqi refugees in Turkey from 2014 until now and so far we have not got a homeland ... Put yourselves in our place. Our situation is bad and our children, their future is unknown.*

This quote indicates that refugees from Iraq, because of their desperation, were trying to attract attention to their situation by leveraging communication from champions like Ben Stiller.

We next computed and plotted *semantic word-similarity* (see Figures 3 and 4). Regarding semantic word-similarity, word vectors for the refugee-emergency type (Syria, Iraq, etc.) and champion-influencer type, and words related to refugees included #Refugees, UNHCR, and World. A *semantic word* is represented by a 50-dimensional vector. We reduced the corresponding 50-dimensional vector to a two-dimensional vector using *principal component analysis* (PCA).

Words related to champions are grouped in the top-left quadrant, but various refugee-emergency types are grouped in the bottom left-to-middle quadrant of the figure (see Figure 3).

Words related to refugees are grouped on the right. Angelina Jolie is in the top-right quadrant, closer to UNHCR and refugees. This is attributable to her association with UNHCR and long-established role as a special envoy. Also, the word-vector representation of tweets in which Jolie was mentioned in the context of humanitarian activities is shown—along with those related to UNHCR but outside its sphere (see Figure 4). There is a clear pattern in the differences between the vectors representing prosocial communication on behalf of refugees on the left side of the figure in contrast to the word vectors representing Jolie's professional (movie, video, etc.) and personal (divorce, Brad, kids, etc.) presence on the right. The same pattern is observed for other champions.

Overall, the social media of UNHCR and its champions was limited to a few refugee-emergency types exhibiting high intensity—indicating the challenge of attracting the attention of champions. The sentiment patterns we observed were on the positive side. This suggests that champions were likely pursuing a specific strategy when attempting to increase social awareness regarding serious problems.

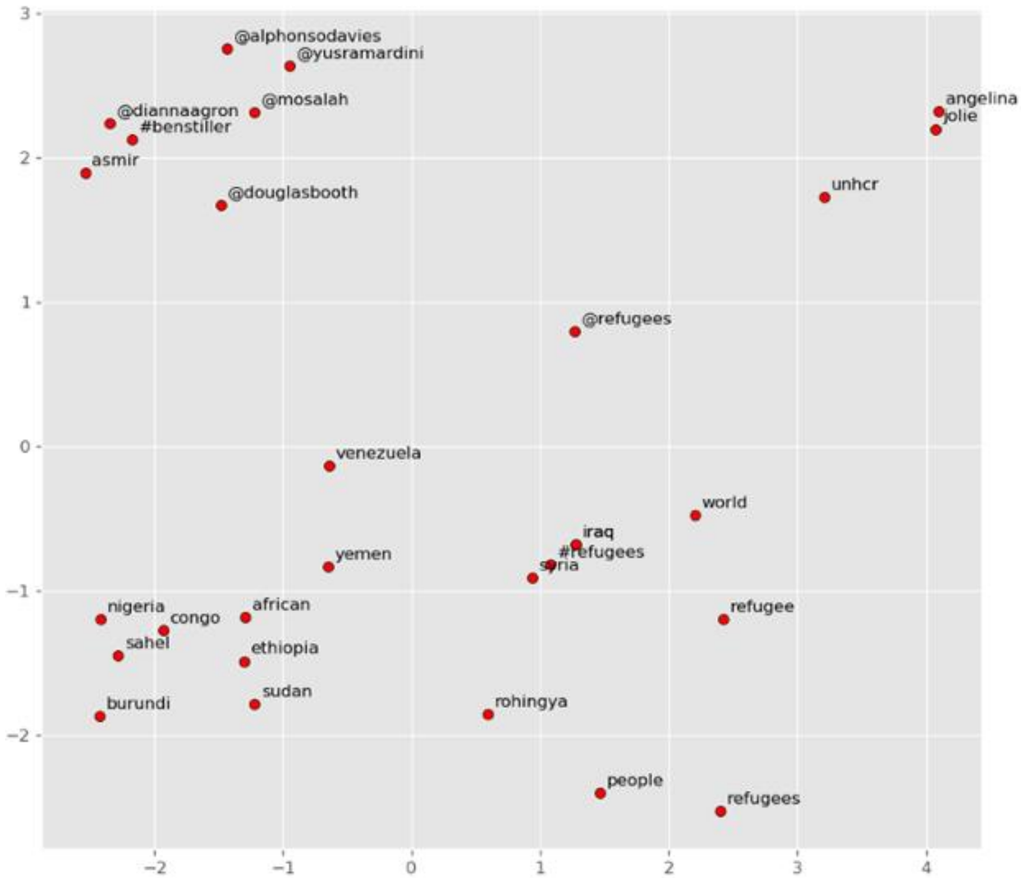

Figure 3. Semantic Word-Similarity among Refugee-Emergency Types and Champions

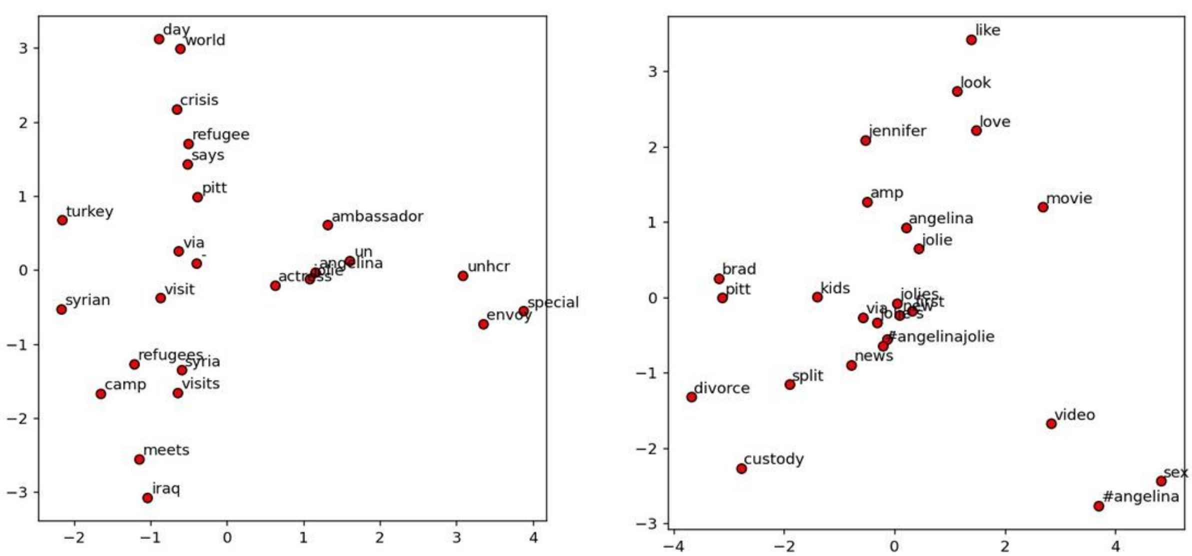

Figure 4. Word-Vectors of Angelina Jolie for Inside (Left)
and Outside UNHCR (Right), with *x*- and *y*-Axis Values from PCA

6.2 Results for the Propositions Involving Champions and Their Followers (P3, P4)

We next discuss the propositions involving champion types and their followers' tweet sentiment (P3a) and attitudes (P3b), as well as the champions' participation intensity and their sentiment (P4a) and attitudes (P4b). Most high participation-intensity champions focused on a few high-intensity refugee emergencies, in line with UNHCR's strategy. The dominant sentiment patterns were positive or neutral. Similarly, the dominant attitude patterns were *sharing* and *liking* (see Tables 12 and 13).

For some UNHCR champions, there were a substantial number of tweets revealing negative sentiment and *helpless* and *disliking* attitudes. Most were made by the refugees themselves, who mentioned the champions in their tweets, which, we believe, were aimed at bringing greater attention to their helpless plight and intolerable living conditions. In this way, the champions' activities served to spread awareness about UNHCR's activities. This kind of social communication was well-received by refugees since it exposed their predicament to a global audience.

6.3 Results for the Outreach Performance of UNHCR's Social Media Strategy

The outreach performance metrics of the champions indicate the spread of their messages during various types of emergencies. The UNHCR account had the highest outreach in terms of *the average for Retweet_Count* and the other metrics (see Table 14).

In addition, when we compared the outreach performance metrics for champions' presence inside and outside UNHCR, an interesting pattern emerged showing how the champions' followers perceived their participation in humanitarian activities. The champions' outreach metrics inside UNHCR were higher than their metrics outside it. This shows that the champions' participation in humanitarian activities was well-received by their followers, as a higher number of retweets, likes, and replies indicates deeper engagement and outreach.

Another observation is that, even though sports personalities (e.g., Alphonso Davies, Asmir Begovic, Mo Salah) with millions of followers were engaged as unpaid social media champions, their influence in spreading messages was limited—compared to the other kinds of champions, especially in high-intensity emergency contexts such as the Syria and Iraq emergencies and the Rohingya emergency in Myanmar and Bangladesh. The other champions, while a bit less popular, nevertheless appear to be relationally closer to people who read the tweets (see Table 15).

Though the activities of champions and their followers were limited to high-intensity refugee-emergencies, both the refugees and UNHCR appear to have perceived the role of champions as effective for presenting refugee issues to attract broader public attention.

7 Discussion and Implications

We next discuss the implications of our study by focusing on the problem of refugee crises. We demonstrate how the methods we applied inform UNHCR's social media strategy, how the effects of high-profile influencers work, the higher-level takeaways available, and what they mean for the effectiveness of UNHCR's social media strategy overall.

7.1 Data Science and the Impacts of UNHCR's Social Media Strategy

Online conversations have become a central source of information for individuals, impacting decisions involving product and service offerings (Jiang et al., 2021), as well as certain social issues. UNHCR faces difficulties in determining the outcome of its social media strategy due to the complexity of the services it offers and the unstable political contexts in which they are provided. Drawing comparisons across humanitarian tragedies in the presence of social misery creates challenges for champions for social good and their followers. Nevertheless, our application of data science methods to Twitter-based social media activities surrounding UNHCR's select group of high-profile influencers points to patterns that can be leveraged to strengthen UNHCR's reach and improve future strategies.

Three themes that emerged from our work show that data science offers valuable information that NGOs and IGOs can use to learn about the impacts of the tweet patterns they create. First, our findings indicate that Twitter data clearly identify the focus of UNHCR champions. While the majority of UNHCR's emergencies are located on the African continent, there has been almost no Twitter activity about African emergencies from UNHCR champions or their followers beyond the UNHCR@Refugees account. This is a gap that should be addressed so that UNHCR can enhance its social media impacts on African crises.

Table 12. Sentiment: High (H) and Med (M) Participation-Intensity Champions

Champions	*Positive*	*Neutral*	*Negative*	Total
UNHCR@Refugees	29,902	60,475	955	91,332
Angelina Jolie (H)	13,144	64,559	1,222	78,925
Yusra Mardini (H)	4,871	9,265	232	14,368
Kat Graham (M)	1,542	8,053	1,450	11,045
Ben Stiller (H)	2,967	3,945	1,842	8,754
Alphonso Davies (M)	1,904	1,784	308	3,996
Emi Mahmoud (M)	1,119	1,000	1,029	3,148
Mo Salah (H)	847	654	300	1,801
Maya Ghazal (M)	602	456	352	1,410
Nujeen Mustafa (M)	421	674	36	1,131
Douglas Booth (H)	406	587	64	1,057
Asmir Begovic (M)	119	135	14	268
Nikki Samonas (M)	112	109	12	233
Dianna Agron (H)	65	129	3	197

Table 13. Attitude-Patterns of High (H) and Medium (M) Participation-Intensity Champions

Name	*Sharing*	*Liking*	*Unrelated*	*Helpless*	*Disliking*	Total
UNHCR@Refugees	41,632	34,125	6,974	2,345	6,256	91,332
Angelina Jolie (H)	61,190	9,582	3,927	265	3,961	78,925
Yusra Mardini (H)	8,090	5,217	552	164	345	14,368
Kat Graham (M)	1,301	1,665	402	767	6,910	11,045
Ben Stiller (H)	1,624	3,040	1,124	1,224	1,742	8,754
Alphonso Davies (M)	1,018	2,119	380	142	337	3,996
Emi Mahmoud (M)	554	1115	288	668	523	3,148
Mo Salah (H)	251	877	280	246	147	1,801
Maya Ghazal (M)	195	663	103	321	128	1,410
Nujeen Mustafa (M)	545	448	56	17	65	1,131
Douglas Booth (H)	368	350	181	37	121	1,057
Asmir Begovic (M)	39	156	46	7	20	268
Nikki Samonas (M)	45	110	39	22	17	233
Dianna Agron (H)	29	13	4	0	0	46

Table 14. Comparison: Performance Metrics for Champions For or Without UNHCR

Champions	Champion outreach for UNHCR				Champion outreach outside UNHCR			
	Tweets	*Retweet _Count*	*Favorite _Count*	*Reply _Count*	*Tweets*	*Retweet _Count*	*Favorite _Count*	*Reply _Count*
UNHCR	91,332	53.49	46.47	2.77	—	—	—	—
Angelina Jolie	78,924	2.17	4.28	0.31	1,581,768	2.08	8.41	0.30
Yusra Mardini	14,331	4.23	7.81	0.21	6,025	2.53	7.22	0.23
Kat Graham	10,931	1.98	3.33	0.11	906,957	0.48	1.53	0.09
Ben Stiller	8,601	5.42	13.76	0.82	555,389	1.03	4.95	0.33
Alphonso Davies	3,980	11.33	83.17	1.03	196,754	3.4	31.85	0.87
Emi Mahmoud	3,119	8.88	17.66	0.88	3,326	1.85	5.64	0.24
Mo Salah	1,800	5.70	35.07	1.17	1,419,087	3.90	26.75	0.75
Maya Ghazal	1,377	7.41	21.31	1.05	701	1.68	5.13	0.33
Nujeen Mustafa	1,115	4.85	11.05	0.51	3,237	2.10	5.18	0.28
Douglas Booth	1,056	9.14	24.54	0.88	68,438	1.16	3.58	0.19
Asmir Begovic	252	5.28	23.17	0.67	169,835	1.51	4.04	0.28
Dianna Agron	190	6.71	11.63	0.26	183,173	1.50	5.75	0.22
Nikki Samonas	178	1.87	10.96	0.34	160,062	0.57	4.90	0.38

Note: All performance metrics were calculated as aggregated tweet measures when champions were mentioned. *Favorite_Count* was measured in terms of the number of *Likes* a tweet received from users. We noted a pattern indicating that *Retweet_Count*, *Favorite_Count*, and *Reply_Count* of champions for UNHCR were higher than the outreach outside UNHCR, except for the *Favorite_Count* of Angelina Jolie (4.28 < 8.41) and *Reply_Count* of Nikki Samonas (0.34 < 0.38).

Table 15. Performance Metrics for Champions Related to High-Intensity Emergencies

Champions	Tweets	Retweet _Count	Favorite _Count	Reply _Count
Syria				
UNHCR@Refugee	11,389	56.22	45.46	2.86
Angelina Jolie	25,347	1.60	1.91	0.15
Ben Stiller	1,313	13.79	40.55	1.58
Yusra Mardini	6,854	4.46	7.37	0.21
Iraq				
UNHCR@Refugees	2,355	40.72	38.96	2.41
Angelina Jolie	5,814	2.49	3.95	0.21
Ben Stiller	557	8.54	7.53	0.02
Emi Mahmoud	522	16.53	13.27	0.02
Maya Ghazal	266	6.64	5.86	0.03
Kat Graham	193	16.90	15.07	0.01
Rohingya (Myanmar and Bangladesh)				
UNHCR@Refugee	1,907	71.72	78.46	6.02
Angelina Jolie	1,131	7.54	16.47	1.06
Ben Stiller	15	10.67	21.47	0.73
Douglas Booth	11	5.00	6.91	0.64

Note: The metrics are averages across the tweets where champions with UNHCR were mentioned together.

Second, we found that the closer the relationship between the champion and a refugee emergency's *locus of misery*, the greater the champion's impact. This is illustrated by the Syrian Olympic swimmer, Yusra Mardini, who generated considerable attention for increasing her followers' informedness about Syria's underlying conflicts and the need for emergency donations. Moreover, we found that champions' participation in humanitarian activities was well-received by their followers, as indicated by the higher values of their outreach metrics compared to their presence outside UNHCR. Third, we found that champions with a large fan base outside UNHCR generated more activity within its context. This is illustrated by soccer players Mo Salah and Alphonso Davies, who each have many external followers. Champions epitomize the network effects that can arise in social media precisely because they have so many followers—beyond those associated with refugee emergencies.

7.2 Implications for Applications to UNHCR's Strategy Based on Social Informedness

Social media activities provide a communication channel that supports networked messages reaching a wide audience. Our application of data science methods assessed the effects of informative and awareness-creating champion-centric online communication. Thus, we believe it is appropriate to address the effectiveness of using champions, which has not been fully explored by UNHCR. Going forward, some groundwork will be needed to determine followers' awareness baselines for the refugee-emergency conversations in which they participate. This should be done in a minimally invasive way through pre-post research designs that gauge the *social information awareness effects* of follower participation in which social media is the conduit for sharing some form of information treatment. This is an attractive mechanism to conduct quasi-simulated experimental treatment-and-control tests of champion-influenced learning to support a staged approach to fundraising—whether for prosocial philanthropy, rewards-based entrepreneurial crowdfunding, or social networks of fintech start-up *initial public offering* (IPO) investors. This suggests the relatively high generalizability of our approach.

Further, to determine how charitable social media influencers and their followers will be, UNHCR could seek a reliable evaluation of how socially informed an influencer's followers are about the financial needs for a given refugee emergency and what they can do to help, based on their digital exposure. In this way, enhancing followers' social informedness may offer the side-benefit of creating greater *institutional informedness* about its key stakeholders' donation propensities.

Another takeaway is that the application of data science can produce insights into how information content travels in a networked environment—especially for retweets and replies. These are proxies for how messages are received by a broader network of followers. Though our analysis did not test the effects of content not posted on social networks or not centered around champion-influencer posting, we found that the soccer players we examined generally had high "favorite" counts, indicating that their followers enjoyed almost all their social media activities. Thus, an important application of our methods could be to assess what kinds of champions' or other followers' tweets have the capability to enhance the conversation so that more participants become committed to learning about a refugee emergency. This is especially true for emergencies that are less followed or those that are crowded out by competition with other issues demanding global attention.

Finally, since the term *champion* entered the research vocabulary, *strong vs. weak ties* and the dimensions of *homogeneity* and *heterogeneity* of role models relative to their followers have been discussed (Granovetter, 1973). Our analysis suggests that a mix of strong and weak ties plays a key role. Though this is not a new observation, it still is likely a useful one for UNHCR going forward. With soccer players, their followers' pattern seems to be to like whatever their stars post. Soccer is renowned for the intense emotional bonds that are formed between fans and players (Ronald & Damiano, 2019). Future studies could focus on such features.

8 Conclusion

The growth of social media analysis using data science methods has spawned the continuing development of tools and approaches to identify patterns and their meaning. This research highlights the scale of the problems associated with the massive numbers of refugees seeking support from their countries as well as organizations like UNHCR. This has created an urgent need for both compassion and donations. Our application of data science methods assesses the impact of UNHCR's social media strategy by identifying the sentiment, attitude, and outreach patterns generated by high-profile influencer champions.

Our research on prosocial communication followed a different path from that of prior literature on digital technologies and refugee problems. Earlier work has targeted the interface between technology and government services, exploring security and privacy issues in relation to resettlement, and addressing problems such as racism and assistance for refugees. Our goal was to conduct an empirical study rooted in our newly proposed social informedness theory focusing on the impact of the content of social media influencers' messages. For UNHCR, social influencers operate at the epicenter of a global network, creating awareness of UNHCR's mission and attracting donations.

8.1 Contributions of New Knowledge

We believe it is important to position the contributions of the paper in line with the theoretical and applied contexts of our research and the big data analytics and research methods we used.[22] For this, we will instantiate our knowledge creation and contributions using the "Big Data Agenda for IS Research Framework" (Abbasi et al., 2016, Figure 4). We believe this framework is appropriate for the present paper for two key reasons: first, it was published in this journal in 2016 and two of its authors are also guest editors of this special issue. Thus, it is clearly linked to our work in the current paper. Second, our paper is also aligned with the content of this framework, which includes design science, emphasizes big data methods and research inquiry, and also relates to our selected uses of social informedness theory with behavioral science and IS economics as bases for laying out our ideas (see Appendix Table C1 for an adapted representation of this framework that instantiates our research approaches and contributions to it).

An especially important contribution to academic knowledge is the new social informedness theory that we developed for this research. This theory touches on other developments related to different kinds of informedness that have emerged within the broader literature over time and entered the IS literature in the past decade. Our intention was to take a first step toward developing explanatory theory, which we used to construct four propositions on how evidence about social informedness can be seen in champion-centric, follower-participating communications. Our findings show that the types and intensity levels of emergencies play a key role in how social media followers are impacted by positive sentiment and attitudes. This leads to social informedness about refugee emergencies based on the champions they follow for reasons other than their UNHCR's activities. It also suggests that social informedness is a key ingredient that UNHCR can use to more effectively engage its supporters and government organizations in making donations to the refugee crisis response actions it promotes.

We also found that *not all champions are created equal* in terms of how they interact with their followers about the refugee emergencies they support. Actor/director champions appear to have broader global appeal, and the extent of their fame plays an important role in their level of influence. Actor/director champions are more likely than sports champions to engage their followers in a wider variety of refugee emergencies. Given that sports champions' popularity is more regional than that of actors/directors, their followers tend to be more narrowly focused on regional issues. Additionally, when a sports champion is also a former refugee, the scope of their impact is further narrowed to refugee emergencies in their home country.

[22] We warmly thank an anonymous reviewer for suggesting that we consider leveraging one of several "framework" articles from the IS journals in the past decade that emphasize research contributions related to conducting action design research (Sein et al. 2011), maximizing the impact of design science research (Gregor & Hevner, 2013) and big data analytics research for deriving knowledge, making decisions, and carrying out actions when behavioral science, design science, and economics of IS, and the different epistemological concerns and paradigmatic considerations are included.

In terms of the relevance of our work for practice, insights from Twitter data for prosocial empirical research point to the impacts of UNHCR's efforts to create social awareness of refugee emergencies. This study illustrates the relevance of social media discussions for developing impactful research on complex problems in economically and politically difficult social settings.[23]

We packaged our approach as a *five-step design science data analytics framework*, which has desirable features for its applicability in other IS and technology research settings (as suggested by its inclusion in Appendix Table C1 on Abbasi et al.'s big data research framework). The method is highly generalized for use in multiple contexts. One area of interest is the empirical investigation of the effectiveness of predictive models related to the stock or fixed-income securities investments associated with *environmental, social, and governmental* (ESG) portfolios aligned with the UN's 2030 Sustainability Goals. Tracking social sentiment is important in this context because there is currently no regulation that effectively defines whether a given firm is truly prosocial, operates a green enterprise, or controls its production of carbon externalities effectively, factors that impact the extent to which firms are making sustainability investments.

We recently conducted research on the impacts of social sentiment in meme-stock trading (e.g., GameStop, AMD, and Robinhood short selling in 2021, etc.) that demonstrates the efficacy of a similar approach (Kim et al., 2023). We are also using this approach to extract insights into how collective action based on Reddit posts and social communication influence less-informed investors' decisions to buy stock when hedge funds and professional investors predict sharp price declines. Our design science-based data analytics framework could also be applied to political election polling and predictions based on public advertising expenditures, social sentiment, and historically observed decision patterns for differentially informed members of a district or country electorate. The common thread is the *role of social information* in driving the observed outcomes.

Social informedness theory provides a useful frame of reference for outlining the problems addressed by UNHCR. Our research in this paper confirms the applicability of the theory in a different form compared to other work on firm, consumer, and household informedness.[24] Our applied contribution related to social informedness theory is to show how technology can make a difference in prosocial settings and support outcomes aligned with the UN's 2030 Sustainability Goals in the present research setting. Therefore, it is important to position this work on social informedness theory in a journal committed to scholarly submissions that contribute new theory to IS research and leverage data science and design methods for its practical application (Data Science for Social Good, 2021).

8.2 Limitations

Like all research, there are always limitations in what the authors are able to achieve—despite their best efforts. We offer several examples. In the UNHCR Twitter corpus we worked with, there was a lack of retweet counts and reply counts related to users. Instead, they were connected at the tweet-relation level. It would have been useful to calibrate the depth of communication and development of ties to assess other aspects of refugee-emergency issues. Further, though the data corpus was large overall, it was also sparse in some ways. For example, social media champions are more like central players and figures to rally around than actual instigators of specific communication threads that move conversations in different and more desirable directions.

A third consideration involves the use of social informedness theory to lay out hypotheses that permit the testing of relationships to determine the extent to which causal inferences can be asserted. We presented our findings in a more general way, akin to how leading field theories from psychology and sociology were originally shared. The emphasis was on participants' interactions with others in their environment before the details were worked out (Lewin, 1939; Martin, 2003). We focused on evaluating whether social informedness via social media regarding refugee emergencies can be better understood by considering the nature of the related social sentiment expressed by champions and their followers. Our experience suggests that it can—albeit in more nuanced ways than we originally imagined.

[23] An anonymous reviewer encouraged us to share information about how the authors' relationship with the UNHCR's staff beneficially supported our effort. It deepened our development of the new social informedness theory for the applied research setting, expanded the coverage of our data collection, increased the focus of our data science work, and produced meaningful—and unexpected findings with rich applications to the social media strategy involving champions in support of refugee-emergence social informedness. In the last of three stages of presentation meetings, we were asked to provide our full dataset of 5 million tweets and related observations, along with our codebase of Twitter API data queries and the related procedures to run the data analytics described in our five-step design science data analytics framework. We also offered to further assist the UNHCR with these things going forward.

[24] Other application domains have included transport tickets, hotel accommodations consumer purchase informedness, chronic disease healthcare, green collection of household waste, social trading of equities and fixed-income debt instruments in complex market settings, and building resident informedness for collective action regarding sustainable energy and water use.

Acknowledgments

This research is new and has not been submitted or previously handled by any other publication outlets. We benefited from the suggestions of the JAIS special issue guest editors and the anonymous reviewers. Raghava Mukkamala acknowledges the Global Communications Service team at the UNHCR office in Copenhagen for their input and close collaboration. Rob Kauffman wishes to thank the Copenhagen Business School, the Endowed Chair in Digitalization, and Singapore Management University for research support. Helle Henriksen further acknowledges the UNHCR office in Copenhagen for providing data and showing interest in this work. All errors and omissions are the authors' responsibility.

References

Abbasi, A., Sarker, S. & Chiang, R. H. L. (2016). Big data research: Toward an inclusive agenda. *Journal of the Association for Information Systems*, *17*(2), i-xxxii.

Appel, G., Grewal, L., Hadi, R. & Stephen, A. (2019). The future of social media in marketing. *Journal of the Academy of Marketing Science*, *48*, 79-85.

Aral, S. & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, *337*(6092), 337-341.

Barr, S. (2019). Price Harry follows in Princess Diana's footsteps by walking through minefield on royal tour. *Independent*, ttps://www. independent.co.uk/life-style/prince-harry-princess-diana-royal-tour-minefields-landmines-angola-africa-meghan-markle-a9122471.html

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(4-5), 933-1022.

Burtch, G., Ghose, A. & Wattal, S. (2016). Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Information Systems Research*, *27*(3), 478-496.

Burtch, G., Hong, Y. & Liu, D. (2018) The role of provision points in online crowdfunding. *Journal of Management Information Systems*, *35*(1), 117-144.

Carley, K. & Gasser, L. (2013). Computational organization theory. Chapter 7 in G. Weiss (ed.), *Multiagent systems* (2nd ed.), MIT Press.

Chang, C. & Lee, Y. (2009). Framing charity advertising: Influences of message framing, image valence, and temporal framing on a charitable appeal. *Journal of Applied Social Psychology*, *39*(12), 2910-2935.

Chang, M., Kauffman, R. J. & Kwon, Y. O. (2014). Understanding the paradigm shift in computational social science in the presence of big data. *Decision Support Systems*, *63*, 67-80.

Chen, H., Chiang, R. H. L. & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1188.

Chung, K. T. C, Derdenger, T. P. & Srinivasan, K. (2013). Economic value of celebrity endorsement: Tiger Woods' impact on sales of Nike golf balls. *Marketing Science*, *32*(2), 271-293.

Chung, Y. A., Lee, H. T. & Glass, J. (2018). Supervised and unsupervised transfer learning for question answering. *Proceedings of the 2018 Conference of the North American Association of Computational Linguistics: Human Language Technologies*.

Clemons, E. K. (2008). How information changes consumer behavior and consumer behavior determines corporate strategy. *Journal of Management Information Systems*, *25*(2), 13-40.

Clemons, E. K., Dewan, R., Kauffman, R. J. & Weber, T.W. (2017). Understanding the information-based transformation of strategy and society. *Journal of Management Information Systems*, *34*(2), 425-456.

Data Science for Social Good. (2021). *Data science to create social impact*. Retrieved on January 2, 2023 from https://www.datasciencefor socialgood.org/.

Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. (2018). *BERT: Pretraining of deep bidirectional transformers for language understanding.* Google.

de Vries, L., Gensler, S. & Leefland, P. S. H. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing*, *26*, 83-91.

Díaz, A. A. & Doolin, B. (2016). Information and communication technology and the social inclusion of refugees. *MIS Quarterly*, *40*(2), 405-416.

Díaz, A. A. & Doolin, B. (2019). Temporal enactment of resettled refugees' ICT-mediated information practices. *Information Systems Journal*, *29*(1), 145-174.

Dix, S., Phau, I. & Pougnet, S. (2010). Bend it like Beckham: The influence of sports celebrities on young adult consumers. *Young Consumers*, *11*(1), 36-46.

Dokhanchi, A., Manian, A., Amiri, M. & Hassanzadeh, A. (2019). Social campaigns on online platforms as a new form of public sphere in digital era: A critical review. *Journal of IT Management*, *11*(3), 81-95.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). Wiley.

Ganegedara, T. (2019). *Intuitive guide to understanding GloVe embeddings.* Towards Data Science. https://towardsdatascience.com/ light-on-math-ml-intuitive-guide-to-understanding-glove-embeddings-b13b4f19c010

Ginsberg, A. & Abrahamson, E. (1991) Champions of change and strategic shifts: The role of internal and external change advocates. *Journal of Management Studies*, *28*(2), 173-190.

Gleasure, R. & Feller, J. (2016). Does heart or head rule donor behaviors in charitable crowdfunding markets? *International Journal of Electronic Commerce*, *20*(4), 499-524.

Goodfellow, I., Benjio, Y. & Courville, A. *Deep learning*. MIT Press.

Goswami, I. & Urminsky, O. (2016). When should the ask be a nudge? The effect of default amounts on charitable donation. *Journal of Marketing*, *53*(5), 829-846.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360-1380.

Gregor, S. & Hevner, A. R. (2013) Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337-355.

Hantscher, S. (2019). *The UNHCR and disaster displacement in the 21st century*. Springer.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd ed.). Springer.

He, R., Lee, W. S., Ng, H. T. & Dahlmeier, D. (2017). An unsupervised neural attention model for aspect extraction. *Proceedings of the 2017 Conference of the Association of Computational Linguistics*.

Hoang, A. P. & Kauffman, R. J. (2018). Content sampling, household informedness, and the consumption of digital information goods. *Journal of Management Information Systems*, *35*(2), 575-609.

Identification for Development. (2022). *Practitioner's guide to biometric data*. The World Bank, https://id4d.worldbank.org/guide/biometric-data#:~:text=In%20addition%20to%20biographic%20data,/IEC%202382-37).

Jacobsen, K. L. (2015). Experimentation in humanitarian locations: UNHCR and biometric registration of Afghan refugees. *Security Dialogue*, *46*(2), 144-164.

Jacobsen, K. L. & Sandvik, K. B. (2018). UNHCR and the pursuit of international protection: Accountability through technology? *Third World Quarterly*, *39*(8), 1508-1524.

Jiang, C., Wang, J., Tang, Q. & Lyu, X. (2021). Investigating the effects of dimension-specific sentiments on product sales: The perspective of sentiment preferences. *Journal of the Association for Information Systems*, *22*(2), 6, 459-489.

Jha, S. K., Pinsonneault, A. & Dubé, L. (2016). The evolution of an ICT platform-enabled ecosystem for poverty alleviation: The case of eKutir. *MIS Quarterly, 40*(2), 431-445.

Kauffman, R. J., Kim, J., Kee, S. Y., Hoang, A. P. & Ren, J. (2017). Combining machine-based and econometrics methods for policy analytics insights. *Electronic Commerce Research and Applications, 25*, 115-140.

Kauffman, R. J. & Wood, C. A. (2009). Research strategies for e-business: A philosophy of science view in the age of the internet. In R.J. Kauffman and P.P. Tallon (Eds.), *Economics, information systems, and electronic commerce: Empirical research* (pp. 31-62). M. E. Sharpe.

Kessler, J. & Milkman, K. (2018). Identity in charitable giving. *Management Science, 64*(2), 845-859.

Kim, K., Lee, S.-Y. T. & Kauffman, R. J. (2023). Social informedness and investor sentiment in the GameStop short squeeze. *Electronic Markets*, *33*(1), Article 23.

Kim, S., Gupta, S. & Lee, C. (2021). Managing members, donors, and member-donors for effective nonprofit fundraising. *Journal of Marketing*, *85*(3), 220-239.

Kingston, L. F. (2018). Biometric identification, displacement, and protection gaps. In C.F. Maitland (ed.), *Digital lifeline: ICTs for refugees and displaced persons*. MIT Press.

Kwak, D. H., Ramamurthy, K. R. & Nazareth, D. L. (2019). Beautiful is good and good is reputable: Multiple-attribute charity website evaluation and initial perceptions of reputation under the halo effect. *Journal of the Association for Information Systems, 20*(11), 1611-1649.

Lawless, M. W. & Price, L. L. (1992). An agency perspective on new technology champions. *Organization Science, 3*(3), 342-355.

Leong, C. M. L., Pan, S. L., Racetam, P. & Kaewkitipong, L. (2015). ICT-enabled community empowerment in crisis response: Social media in Thailand flooding 2011. *Journal of the Association for Information Systems, 16*(3), 174-212.

Lewin, K. (1939). Field theory and experiment in social psychology. *American Journal of Sociology*, *44*(6), 868-896.

Li, T., Kauffman, R. J., van Heck, E., Vervest, P. & Dellaert, B. (2014). Consumer informedness and firm information strategy. *Information Systems Research*, *25*(2), 345-363.

Lim-Wavde, K., Kauffman, R. J. & Dawson, G. (2017). Household informedness and policy analytics for collection and recycling of household hazardous waste in California. *Resources, Conservation and Recycling, 120*, 88-107.

ListenNotes. (2021). *Champions for Social Good* [podcast]. https://www.listennotes.com/podcasts/champions-for-social-good-champions-for-n7ZV6fKWGND/

Liu, L., Suh, A. & Wagner, C. (2018). Empathy or perceived credibility? An empirical study on individual donation behavior in charitable crowdfunding. *Internet Research*, *28*(3), 623-651.

Mackelden, A. (2020). Angelina Jolie on honoring and learning from her adopted children's roots. *Harper's Bazaar*, https://www.harpersbazaar.com/celebrity/latest/a32926075/angelina-jolie-adopted-childrens-heritage-interview/

Madianou, M. (2019). The biometric assemblage: Surveillance, experimentation, profit, and the measuring of refugee bodies. *Television & New Media*, *20*(6), 581-599.

Martin, G. (2021). Chief impact officer for Silicon Valley mental health startup as his family battle continues through the press. *Forbes.com.* https://www.forbes.com/sites/guymartin/2021/03/23/prince-harry-takes-a-job-as-chief-impact-officer-for-silicon-valley-mental-health-startup-as-his-family-battle-continues-through-the-press/?sh=5d8bb70a29f3

Martin, J. L. (2003). What is field theory? *American Journal of Sociology*, *108*(1), 1-49.

Mayer, L. H. (2022). Regulating charitable crowdfunding. *Indiana Law Journal*, *97*(4), Article 6.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Proceedings of the 26th International Conference. on Neural Information Processing Systems*.

Miranda, S. M., Young, A., Yetgin, E. (2016). Are social media emancipatory or hegemonic? Societal effects of mass media digitization in the case of the SOPA discourse. *MIS Quarterly*, *40*(2), 303-329.

Morand, M., Mahoney, K., Bellour, S. & Rabkin, J. (2012). *The implementation of the UNHCR's policy on refugee protection and solutions in urban areas.* UNHCR.

Morris, R. (1994). Computerized content analysis in management research: A demonstration of advantages & limitations. *Journal of Management*, *20(4)*, 903-931.

Oinas-Kukkonen, H., Lyytinen, K. & Yoo, Y. (2010). Social networks and information systems. *Journal of the Association for Information Systems*, *11*(2), 61-68.

Peng, Y., Li, Y. & Wei, L. (2022). Positive sentiment and the donation amount: Social norms in crowdfunding donations during the COVID-19 pandemic. *Frontiers in Psychology*, *13*, 818510.

Pennington, J., Socher, R. & Manning, C.D. (2014). GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methodology in Natural Language Processing*, 1532-1543.

Perez-Cepeda, M. & Arias-Bolzmann, L.G. (2021). Refugee information consumption on Twitter. *Journal of Business Research*, *123*, 529-537.

Ronald, B. & Damiano, J. P. (2019). The role of commercialization of the European football business for the emotional bond between fans and clubs. *Advances in Management*, *12*(1), 8-29.

Sarker, S., Palvia, P., Oguz, A., Wu, C. & Xiao, X. (2020). A brief retrospective (2013-2019). *Journal of the Association for Information Systems*, *20*(12), 1901-1915.

Schetgen, L., Bogaert, M. & Van den Poel, D. (2020). Predicting donation behavior: Acquisition modeling in the nonprofit sector using Facebook data. *Decision Support Systems*, *141*, 113446.

Schouten, K. & Frasincar, D. (2015). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, *28*(3), 813-880.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., Lundgren, R. (2011). Action design research. *MIS Quarterly*, *35*(1), 37-56.

Su, E. (2021). *GameStop-related market volatility: What happened and what does it mean?* Congressional Research Service. https://crsreports.congress.gov/product/pdf/IN/IN11615

Tan, S. D. & Tas, O. (2020). Social media sentiment in international stock returns and trading activity. *Journal of Behavioral Finance*, June 16, 221-234.

Tazzioli, M. (2020). Extract, datafy and disrupt: Refugees' subjectivities between data abundance and data disregard. *Geopolitics*, *27*(1), 70-78.

Thimm, H. & Rasmussen, K. (2010). Informedness in collaborative networks through active information provisioning. In L. Camarinha-Matos, X. Boucher & H. Afsarmanesh (Eds.), *Collaborative networks for a sustainable world* (pp. 412-419). Springer.

Twitter. (2021). *Data dictionary: Standard v1.1 and introduction.* https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet

UN. (2020). *UN refugee chief laments nearly 80 million people forcibly displaced*. UN News. https://news.un.org/en/story/2020/06/ 1066492

UNHCR. (2019). *Evaluation of UNHCR's data use and information management approaches. Evaluation report*. UNHCR Evaluation Service.

UNHCR. (2020a). *Figures at a glance.* UNHCR.

UNHCR. (2020b). *Guidance on Registration and Identity Management*. UNHCR.

Xiao, A., Huang, Y., Bortree, D. M. & Waters, R. D. (2021). Designing social media fundraising messages: An experimental approach to understanding how message concreteness and framing influence donation intentions. *Nonprofit and Voluntary Sector Quarterly*, *51*(4), 832-856.

Venkatesh, V., Rai, A., Sykes, T. A. & Aljafari, R. (2016). Combating infant mortality in rural India: Evidence from a field study of ehealth Kiosk implementations. *MIS Quarterly*, *40*(2), 353-380.

Wang, Y., Huang, M., Zhao, L. & Zhu, X. (2016). Attention-based LSTM for aspect-level sentiment classification. *Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 606-615).

Wessel, M., Gleasure, R. & Kauffman, R. J. (2021). Sustainability of rewards-based crowdfunding: A quasi-experimental analysis of funding targets and backer satisfaction. *Journal of Management Information Systems*, *38*(3), 612-646.

Xu, J. J., Fedorowicz, J. & Williams, C. B. (2019). Effects of symbol sets and needs gratifications on audience engagement: Contextualizing police social media communication. *Journal of the Association for Information Systems*, *20*(5), 536-569.

Appendix A: UNHCR's Refugee Emergencies and Social Media Champions

Table A1. Refugee Emergencies and Evidence of Intensity

Country	Years	Basis of situation or emergency and refugees	Problem intensity
		Africa	
Burkina Faso, Mali, Niger/Sahel	2016-now	Characterized by sexual and gender-based violence, food insecurity, economic weakness, and limited educational opportunities for children, with 2+ million internally displaced peoples and 858,487 refugees and asylum seekers.	High
Ethiopia (Tigray)	2020-now	45,449 Tigray refugees who fled Ethiopia for Sudan, 96,000 refugees from Eritrea hosted in Ethiopia, and 100,000 people displaced in Tigray prior to 2020; resulted from armed violence, insecurity, food shortages, and a lack of basic infrastructure in the Tigray region.	Medium
Nigeria	2014-now	Armed conflict between the Nigerian military and Boko Haram insurgency resulting in sexual and gender-based violence, suicide bombings, forced recruitment, food instability, and malnutrition, resulting in 304,562 refugees, and 2.1 million people displaced.	High
Central African Republic	2017-now	Long one of the poorest countries in Africa, there remain armed conflicts and violence, election-induced political instability, and socioeconomic problems, with 632,108 refugees and 630,834 internally displaced people, despite improvements in 2016.	Medium
Burundi	2015-now	Approximately 337,000 refugees due to economic decline, food insecurity, disease, violence, and continuing and unresolved political problems.	Medium
South Sudan	2013-now	Internal instability and fears of violence and attacks, with 2.2 million refugees and asylum seekers in the neighboring countries, and nearly 2 million other people displaced internally.	High
Democratic Rep. Congo	2017-now	Ongoing armed conflicts leading to human rights violations, violence and killings, inhumane conditions, and political problems; there are about 918,000 refugees and asylum seekers, with 5 million people displaced within the country's borders.	High
		Middle East	
Iraq	2014-now	Over 3 million people displaced internally and 260,000 refugees outside Iraq, due to mass executions, systematic rape, violence, and lacking rule of law, with 11 million people in need of humanitarian services due to lacking infrastructure, accommodations, and food.	High
Syria	2011-now	Widely recognized as the worst refugee emergency in the world, armed conflict and political upheavals have resulted in 6.7 million internally displaced people, 6.6 million refugees, with 1 million still in Syria, and 13.4 million people in need of humanitarian assistance.	High
Yemen	2011-now	Ethnic conflict led to one of the largest humanitarian crises, with 66% of the people dependent on aid and 20.7 million people in need of humanitarian aid; 4 million people are displaced internally, and the country is also burdened with 137,000 African refugees and asylum seekers, damaged infrastructure, insufficient housing, and a threat of famine.	High
		South and Southeast Asia	
Bangladesh / Myanmar (Rohingya)	2017-now	Ethnic armed conflict involving the persecution and exodus of stateless, Muslim Rohingya minority people in Myanmar; 742,000 refugees have sought international protection and humanitarian assistance in Bangladesh, where capacity is highly limited.	High
		South and Central America	
El Salvador, Guatemala, Honduras	2016-now	Gang and drug-related violence that has not been effectively addressed, leading to security, social, economic, and political problems, which have affected 890,000 refugees and asylum seekers. The situation is not yet seen as an emergency.	Medium
Venezuela	2014-now	UNHCR refers to this as a situation rather than an emergency, though more than 5.9 million refugees and migrants have been displaced due to violence, lack of security, socioeconomic pressures, and political instability.	High

Note: Data source: UNHCR in 2020. *Situations* and *emergencies* identified by UNHCR have different degrees of intensity. Generally, situations are of lower intensity than refugee emergencies. Not all emergencies and situations had specific social media champions assigned to them by UNHCR, nor have we been able to track all of them with Twitter data.

Table A2. Social Media Champions: Types and Participation Intensities

Name	Background, brief bio	Participation intensity
UNHCR@ Refugees	**UNHCR's Twitter account:** Through this account, this IGO provides vital aid and protection to the forcibly displaced all around the world. Its posts involve institutional, not individual responses.	High
	Special Envoy	
Angelina Jolie	**Film actress, filmmaker, and humanitarian:** She has been working as a UNHCR Special Envoy since 2012. She was a Goodwill Ambassador for 10 years. As a Special Envoy, she visits refugee camps around the globe regularly and is highly visible on behalf of UNHCR's causes.	High
	Goodwill Ambassadors	
Ben Stiller	**American film actor and philanthropist:** He has supported UNHCR since 2016. He has traveled to meet refugees in Germany, Jordan, Guatemala, and Lebanon, supported many projects and campaigns, and used his voice in social media to advocate for refugees. He was appointed as a Goodwill Ambassador in July 2018. He traveled to Washington, DC, and testified in front of the Senate Foreign Relations Committee, imploring the US government to continue providing vital and much-needed support for refugees.	High
Yusra Mardini	**Syrian refugee and Olympic swimmer:** She was appointed the youngest-ever Goodwill Ambassador for UNHCR at age 19 in 2017. She advocates for refugees through sharing her own story and has become a powerful voice for the displaced across the world and an example of their resilience and determination to rebuild lives and positively contribute to host communities.	High
Alphonso Davies	**Ghana-born Canadian citizen and footballer in Germany:** In March 2021, he became the first footballer and Canadian appointed as a Global Goodwill Ambassador for UNHCR. He plays for FC Bayern Munich and Canada Men's National Football Team. Born in a refugee camp in Ghana to Liberian parents who fled the civil war in their home country, he knows firsthand what it means to be a refugee.	Medium
Maya Ghazal	**Syrian refugee and airline pilot:** She has been a Goodwill Ambassador since 2021, after years of active involvement in UNHCR initiatives. She is the first Syrian refugee to become a female air pilot. Through sharing her own inspiring story, Maya advocates for refugee inclusion and access to education and job opportunities and counters the negative stereotyping of refugees.	Medium
Nikki Simona	**Ghanaian film actress, TV presenter, and model:** She has supported UNHCR since December 2020. She is a household name, known for her work in movies and TV series, and a member of the LuQuLuQu campaign, which focuses on supporting forcibly displaced Africans. She has worked actively on fundraising and public engagement campaigns toward helping prevent the spread of coronavirus in refugee camps, by using both traditional and social media in her outreach.	Medium
Kat Graham	**American actress:** She has been a Goodwill Ambassador since December 2020. She supports UNHCR initiatives to bring attention to challenges faced by LGBTQI and female refugees around the world. This involves empowering them with new skills and providing the resources and knowledge they need to build brighter futures for themselves, their families, and their communities.	Medium
Emi Mahmoud	**Sudanese American slam poet:** She has been a Goodwill Ambassador since June 2018 and has used her talents to raise awareness around the refugee cause. She has also led poetry workshops in refugee camps and integrated her firsthand experiences in refugee camps into her poetry.	Medium
	High-Profile Supporters	
Asmir Begovic	**Italian soccer player and AC Milan goalkeeper:** He has supported UNHCR since early 2020, using his profile to raise awareness and fundraise for refugees. He was forced to flee his home in Bosnia and Herzegovina, first arriving in Germany before moving to Canada. He has enjoyed an illustrious career as a goalkeeper, playing for the biggest football clubs in the world. In 2014, he represented Bosnia and Herzegovina at the country's first-ever World Cup.	Medium
Douglas Booth	**British actor:** He has supported UNHCR since 2015, highlighting UNHCR's work in interviews and publicizing the global refugee crisis via his social media. He has supported key UNHCR fundraising events and campaigns, including World Refugee Day, the #iBelong Campaign to End Statelessness, and the Nansen Refugee Award. He is featured in film projects including *Words Matter* and *What They Took with Them*.	Medium

Dianna Agron	**American actor, singer, dancer, director:** She began supporting UNHCR in 2016 and visited resettled refugees in Austria. She traveled to Jordan to meet Syrian refugees in towns and camps. Through social media, she encourages her fans to support #WithRefugees campaign, calling for solidarity and responsibility for the refugee crisis. She continues to invoke empathy and support for refugees amid an environment of intolerance.	High
Mo Salah	**Egyptian soccer player and social media champion:** He is involved with UNHCR and an NGO-private partnership. In 2020, he became a high-profile UNHCR supporter. In a recent annual UNHCR report, Mo Salah stated his interest in helping refugee children gain access to education and also stressed the importance of the Instant Network Schools (INS) program—introduced in Egypt, where he was born and raised. INS is a partnership between UNHCR and the Vodafone Foundation, which promotes education in the most marginalized communities in Africa.	High
Nujeen Mustafa	**Syrian refugee:** She has been a UNHCR supporter since 2018. Born in Aleppo, Syria with cerebral palsy. At 16, she completed a 3,500-mile journey from Syria to Germany in a steel wheelchair. She is a powerful advocate for refugee youth and promotes active involvement and meaningful participation of people with disabilities in planning, development, and decision-making processes at all stages of the refugee response. She has spoken at UN events and for the media about the importance of keeping children's hopes alive.	Medium

Note: The common source of these brief bios is the www.unhcr.org website, which identifies its social media champions in this way.

Appendix B: The Twitter Search Queries Used for Data Collection

Table B1. Search Queries for Social Media Champions and UNHCR

Name	**Champion's own tweets for UNHCR**	**Champion's own tweets without UNHCR**	**Champion's presence with UNHCR**	**Champions presence outside UNHCR**
UNHCR	*from:Refugees -is:retweet lang:en*	—	—	—
Ben Stiller	*from:RedHourBen* (*UNHCR OR refugees OR refugee OR #refugees OR @refugees OR #unhcr*) *-is:retweet lang:en*	*from:RedHourBen* (*-UNHCR -refugee -refugees -unhcr -#refugees -@refugees -#unhcr -#WithRefugees*) *-is:retweet lang:en*	((*Ben Stiller*) *OR @RedHourBen*) (*UNHCR OR refugees OR refugee OR #refugees OR @refugees OR #unhcr OR #WithRefugees*) *-is:retweet lang:en*	((*Ben Stiller*) *OR @RedHourBen*) (*-UNHCR -refugee -refugees -unhcr -#refugees -@refugees -#unhcr -#WithRefugees*) *-is:retweet lang:en*

Note: In the above search queries, the *from:* operator indicates the Twitter username and the hyphen (-) is a negation operator. Prepending a hyphen to a keyword (e.g. *-refugee*) makes it so that it will not match the tweets containing that keyword. Similarly prepending a hyphen to an operator will negate its operation (e.g., *-is:retweet* will not match any retweets). Moreover, successive operators with a space between them indicate *AND* logic and hence they will fetch tweets if both operators are evaluated to be true. For example, *refugees -is:retweet lang:en* will match all the original tweets containing *refugees* as a search term where the tweets' language is English. Similarly, successive operators/terms with the *OR* operator will match the tweets if any single operator is evaluated to be true. For example, *refugee OR #refugees* will match tweets that either contain the word *refugee* as a search term or tweets containing the *refugee* hashtag (*#refugees*). Finally, for the rest of the champions, we used search queries like for Ben Stiller by replacing the respective Twitter handles of the champions (e.g., *RedHourBen*).

Appendix C: Knowledge, Decisions, and Actions in "Data Science for Social Good" Research

Contributions from a UNHCR Project Using a JAIS "Big Data Research Directions" Framework (Adapted from Abbasi et al., 2016, p. xxv)

The authors of this framework (with numbered dimensions in Table C1 below) wrote that it is intended to "*energize the conversation on big data in the broader IS community and provide a roadmap for advancing scholarship in the area*." From our point of view, it has been especially helpful since it embraces several key threads in our research. They include the new knowledge we produced was made possible by a blend of theoretical perspectives for ③ deriving knowledge, and ⑦ supporting decisions and actions.

The framework brings together ④ & ⑧ for behavioral science, ⑤ & ⑨ for design science perspectives, and ⑥ & ⑩ IS and economic theory in its view of big data research directions, mapping to the spectrum of research approaches that we have used. These approaches differ in for their epistemological concerns about methods, validity and scope, as well as in their paradigmatic considerations related to the kinds of explanations they produce, based on the data and evidence used, and the varied ways their theoretical perspectives are constructed and deliver contextual relevance. We further note that some of the entries below ④, ⑤, and ⑥ for deriving knowledge and ⑧, ⑨, and ⑩ for decisions and actions are our own (in italic font) and add to those originally proposed (in roman font) in the2016 JAIS article.

Table C1. Instantiation of Our Research Approach

The Behavioral, Design, and Economic Bases for Our UNHCR Research	
① Epistemological Concerns **② Paradigmatic Considerations**	
③ Deriving Knowledge: Topics	**⑦ Decisions and Actions: Topics**
④ *Behavioral Science*	**⑧ *Behavioral Science***
Computational social science	Nature of decision-making
Privacy and security concerns	Role of organizational culture
Ethical considerations	Effects on cognition and usability
Prosocial participation/communication	Trust and big data vs. intuition
Stakeholder characteristics	Adoption of big data technology
Champion and follower positivity	Big data outcomes
⑤ *Design Science*	**⑨ *Design Science***
Novel artifacts for prediction/description	Novel artifacts for decision support
Modeling formalisms, integration artifacts	Business process improvement
Sentiment classification and valence	Big data action design research
Supervised and unsupervised learning	*Social media influencers*
Social media influencers	*Tweet valence and semantic similarity*
Tweet valence and semantic similarity	*Five-step data analytics framework*
⑥ *Economics of IS*	**⑩ *Economics of IS***
Value of data volume and variety	Value and impact for 4Vs for decisions
Cost of veracity and velocity	Value of big data artifacts
Social media and economics of IS	*Information value and value chains*
Information asymmetries	*Process activities, action decisions*
Social informedness theory	*Quasi-experimental research design*

About the Authors

Raghava Rao Mukkamala is an associate professor and the director of the Centre for Business Data Analytics at the Department of Digitalization, Copenhagen Business School, Denmark. Raghava holds a PhD in theoretical computer science, and his current research focuses on data science, cyber security, blockchain, and distributed ledger technologies. Combining formal/mathematical modeling approaches with machine learning techniques, his current research program seeks to develop new algorithms for big data analytics. As part of his pro bono research collaboration with the United Nations High Commissioner for Refugees (UNHCR), his current research mainly focuses on identifying hate speech and bias against refugees using machine learning and deep learning approaches. He is also the program director for the Master's in Data Science Program at CBS and teaches several courses on machine learning. Before moving to research, Raghava worked for many years in programming and IT development in the Danish IT industry.

Robert J. Kauffman holds the Endowed Chair in Digitalization at Copenhagen Business School and is an emeritus professor of IS at Singapore Management University. His graduate degrees are from Cornell University (MA) and Carnegie Mellon University (MS, systems science and PhD, industrial administration). His research has focused on technology and strategy, the economics of IT, financial services and technology, managerial decision-making, sustainability economics, and e-commerce. He served as the associate dean (Faculty & Research) and chair of IS and Management at SMU's School of Computing and IS. He served as the W. P. Carey Chair in IS at Arizona State University and as a professor and the director of the MIS Research Center at the Carlson School of Management, University of Minnesota, where he chaired Information and Decision Sciences. His work has appeared in *Organization Science*, *Management Science,* the *Review of Economics and Statistics*, *Energy Policy*, *Information Systems Research*, *Journal of Management Information Systems*, *MIS Quarterly*, and the *Journal of the Association for Information Systems,* among others. He also served as a senior editor for *Journal of the Association for Information Systems* when it was founded.

Helle Zinner Henriksen is a professor (WSR) and the chair of the Department of Digitalization at Copenhagen Business School. She has a degree in law from the University of Copenhagen and a PhD in MIS from Copenhagen Business School. Her research focuses on the uptake and use of digital technologies in the public sector regarding digital services that are legally enforced for the benefit of citizens and businesses. Her work primarily focuses on the Danish context, where she applies qualitative research methods—more recently with a particular focus on critical perspectives to digitalization. Her work has appeared in IS journals including *European Journal of Information Systems, Information Systems Journal*, *Journal of Strategic Information Systems*, *Journal of Information Technology,* and *Information Systems Frontiers*, as well as e-government journals including *Government Information Quarterly*, *International Journal of Public Administration,* and *Information Polity.*