Reclaiming Stigmatized Narratives: The Networked Disclosure Landscape of #MeToo

RYAN J. GALLAGHER, Communication Media and Marginalization Lab, Network Science Institute, Global Resilience Institute, and College of Arts, Media and Design, Northeastern University
ELIZABETH STOWELL, Wellness Technology Lab, Khoury College of Computer Science and Bouvé College of Health Sciences, Northeastern University
ANDREA G. PARKER, Wellness Technology Lab, Khoury College of Computer Science and Bouvé College of Health Sciences, Northeastern University
BROOKE FOUCAULT WELLES, Communication Media and Marginalization Lab, Network Science Institute, Global Resilience Institute, and College of Arts, Media and Design, Northeastern University

The social stigma looming over disclosures of sexual violence discourages many women from publicly sharing their stories, limiting their ability to seek support and obscuring the epidemic of sexual violence against women. By inviting women to share their ordinarily silenced stories, the hashtag #MeToo surfaced a network of survivors to confront this stigma. Through a mixed-methods analysis of over 1.8 million tweets posted during the first two weeks after #MeToo gained widespread popularity in 2017, we map the landscape of disclosures that emerged and disentangle the effects of *network-level reciprocal disclosures*, or disclosures made in reaction to seeing others disclose. We detail how survivors disclosed a diversity of sexual violence experiences in solidarity with others, composing nearly half of all authored tweets and comprising a disproportionate number of interactions within the #MeToo network. Further, we show that the more disclosures an individual potentially saw prior to disclosing, the more likely they were to share details with their disclosure. We argue that such network-level reciprocal disclosures may have reduced stigma, creating a counterpublic space safe for disclosure which, subsequently, generated more disclosures. Our work illustrates how feminist hashtag activism, like #MeToo, can unify individual and collective narratives to dismantle the stigma surrounding disclosures of sexual violence. **Content warning:** *This article heavily discusses issues of sexual violence against women.*

1 INTRODUCTION

Sexual violence is a near ubiquitous experience among women, but, despite its prevalence, women's experiences of sexual violence are often relegated away from the public's view into private conversations. The stigma against disclosures of sexual violence manifests as a culture of silence where up to 40% of women never disclose to anyone [48] and, of those who do, most confide only in a close friend [38]. Women who do choose to disclose publicly risk reactions of disbelief or, worse, blame [56]. Blaming survivors of sexual violence discourages further disclosures, effectively chilling the collective narrative of those who have been sexually harassed and assaulted. The totality and consequences of social stigma against disclosures of sexual violence paint a bleak picture of a culture where sexual violence against women is not recognized as the pervasive public health issue which it is.

In October 2017, building on the foundations laid down by activist Tarana Burke over 10 years earlier [37], the hashtag #MeToo spread rapidly across Twitter and Facebook, broadcasting the ordinarily hidden and marginalized stories of sexual violence against women. In a shift away from the culture of silence manufactured by stigma against disclosures, #MeToo invited women to publicly and directly share their experiences of sexual violence to "give people a sense of the magnitude of the problem" [65]. In the first 48 hours, the hashtag #MeToo and phrase "me too" were used nearly 1 million times on Twitter and over 12 million times on Facebook, with Facebook reporting that 45% of its users had friends who disclosed "me too" [74]. By connecting the relentless

number of disclosures into a single network of solidarity and lived experiences, the hashtag surfaced the painfully commonplace nature of sexual violence against women in plain view.

The disclosure phenomena of #MeToo, in which many women simultaneously and publicly disclosed personal experiences with sexual violence, contrasts starkly with past observations on how individuals choose to disclose stigmatized experiences on social media. The anonymity of platforms like Reddit is often preferable for disclosing experiences of sexual violence because it allows survivors to safely seek support and advice without the repercussions of offline disclosure [3, 26]. To reconcile the gap between private, anonymous disclosures and public, direct disclosures online, Andalibi and Forte [2] introduced the concept of *network-level reciprocal disclosure*, which describes when individuals disclose to their social networks in reaction to seeing others in the network disclose their own stigmatized stories. They argue that disclosures of shared experiences within one's local network can reduce the perceived stigma around that experience, and encourage an individual to make their own self-disclosure. Importantly, network-level reciprocal disclosures move beyond the typical dyadic setting of disclosure and provide a framework for understanding how hashtags that address sexual violence against women, like #MeToo, #YesAllWomen, and #NotOkay, may arise through the reduction in stigma resulting from the propagation of disclosures across a social network.

Through a mixed-methods analysis of over 1.8 million tweets posted during the wave of #MeToo disclosures in October 2017, we disentangle the roles of network-level reciprocal disclosures and hashtag activism in the rise of #MeToo. We develop a predictive classifier of disclosures and couple it with with the follower-following networks of #MeToo users to chart potential exposures to disclosures. Our work shows that the more disclosures that an individual was potentially exposed to prior to disclosing, the more likely they were to share details when they disclosed. In turn, then, those who disclosed wove the communicative backbone of #MeToo and created a space conducive to more disclosures. Unifying our qualitative and quantitative analyses, we argue that the networked disclosure phenomenon that we observe is consistent with the theory that direct, public disclosures online encourage network-level reciprocal disclosures by reducing stigma.

Our work extends the theory of network-level reciprocal disclosures [2] to the context of emergent hashtag activism like #MeToo. In providing evidence that network-level reciprocal disclosures established the basis of #MeToo, we demonstrate that network-level reciprocal disclosures are a plausible mechanism for how hashtag campaigns addressing sexual violence against women gain traction. By connecting our analysis of network-level disclosures with theories of hashtag activism, our work lays the foundations for future research characterizing the micro mechanisms driving disclosures of stigmatized experiences on public social media platforms and the macro mechanisms governing the emergence of hashtag campaigns like #MeToo.

2 RELATED WORK

2.1 Disclosures of Sexual Violence and Stigmatization

While sexual violence impacts people of all genders and sexual orientations, sexual violence is particularly prevalent among sexual and gender minority groups. Because of these disparities, the #MeToo campaign, which builds off the movement founded by Tarana Burke [37], focuses specifically on sharing women's stories. In her lifetime in the United States, nearly 1 in every 3 women has experienced an act of sexual violence, 1 in 6 has experienced rape or attempted rape, 1 in 6 has experienced stalking, 1 in 3 has experienced intimate partner violence, and 1 in 2 has experienced sexual harassment in the workplace [13, 32, 83, 87]. Experience of sexual violence negatively impacts many facets of a person's life well after the incident, including mental and physical health [13, 24, 87, 90], financial stability [32, 59, 67, 83], and housing security [72]. In

addition to the direct impact on the survivor, sexual violence and the culture that enables it reinforce norms perpetuating gender inequities [16, 32, 43].

In the wake of an act of sexual violence, *self-disclosure*, the process of making information about oneself known to others [23], is a critical step for accessing formal assistance (e.g., medical or legal services) and informal support (e.g., social support) that can counteract numerous negative outcomes [48, 81]. Despite the potential benefits, the majority of experiences of sexual violence are never formally reported, and survivors who formally disclose often face disbelief and victim-blaming from authorities, resulting in retraumatization [18, 31, 56]. Of survivors who do disclose their experience, most disclose informally to a close friend [48, 81]. While informal disclosure is a powerful tool to solicit social support from one's friends and family, reactions to the disclosure have the potential to either redouble negative outcomes or support the individual's psychological well-being [48]. While negative reactions of a close friend are infrequent [48], such negative reactions to sexual violence disclosure are associated with poorer psychological well-being [57, 69].

The pervasiveness of sexual violence and infrequency of disclosure result from common rape myths that make up what has been termed "rape culture." These myths include beliefs about the victim's role in their own experience of sexual violence, perceptions of the prevalence of sexual violence, and explanations or excuses for acquaintances who commit these acts [10, 33, 49, 73]. The stigma resulting from these myths influences an individual's belief in own culpability or the likelihood that the survivor will label a past sexual assault experience as rape [16, 55, 81], expected and actual response to disclosure from family, friends, and service providers [40, 81, 89], and the cultural norms that impede wider discussions of the culture and prevalence of sexual violence [45]. These factors discourage disclosure, effectively isolating survivors and impeding their ability to receive needed support.

2.2 Affordances of Online Platforms for Disclosures

The repercussions of publicly disclosing offline often outweigh the potential for support, restricting the options for activating one's social network. In these cases, social media allows individuals to reach beyond their immediate networks and augment them with those who have shared experiences [75]. Reddit, in particular, has become a popular platform for forming communities to support those with stigmatized experiences, particularly around mental health [8, 26–28] and sexual abuse [4, 5, 68]. By accommodating the needs concerning particular stigmatized issues [82] and providing responsive emotional support [4], these communities are able to provide safe spaces for disclosing. This allows individuals to seek validation and support online and compartmentalize disclosures of stigmatized experiences away from their offline social ties, effectively partitioning their public and private lives. Notably, Reddit, in particular, helps make this partition by allowing users to make anonymous "throwaway" accounts [4, 26, 54]. Because they are not associated with any personally identifiable information, anonymous throwaway accounts result in "decreased feelings of vulnerability and increased self-disclosure when it comes to disclosure on mental health" [26], facilitating "intimate and open conversations" that replicate the offline "strangers on a train" phenomenon [79].

Using anonymity to draw a boundary between public and private mitigates the chances of receiving a negative response to a disclosure at the interpersonal level. However, it also neglects the potential support that may exist among one's personal, offline network [41, 92] which is hidden by others concealing their own stigmatized experiences. On platforms like Facebook and Twitter, where online and offline identities are more tightly interwoven [2, 75], individuals have to carefully balance between disclosing via dyadic private messages and via network-level status updates, public comments, and shared content [1, 62, 91]. A public disclosure risks context collapse [62], where colliding social spheres of friends, family, acquaintances, and colleagues within an individual's

network may have varying interpretations, reactions, or backlash towards a disclosure [12, 41, 62]. While some within the network may sympathize with, or even share, the disclosed experience, others may misinterpret or actively admonish it in the same way that individuals try to avoid offline.

Some users choose to navigate the potential of context collapse by hiding their disclosures in plain sight. By carefully choosing the types of images, videos, and links that they share, they are able to disclose publicly but *indirectly* to only the parts of their networks that understand that experience, without alerting the wider social network [5, 63]. Within the context of pregnancy loss, Andalibi, Morris, and Forte [5] formalized the motivations for publicly disclosing indirectly online into a framework of four decision factors: self-related (e.g., eliciting social support, memorializing the experience), audience-related (e.g., feeling out potential reaction), platform and affordance-related (e.g., using Reddit to avoid network overlap), and time-related (e.g., postponing a discussion of the loss). Moreover, Andalibi and Forte [2] demonstrated that the factors in this disclosure decision-making framework also apply to *direct*, public disclosures visibly broadcast to one's entire online network. That is, they found that women publicly made direct disclosures of pregnancy loss also for self-related, audience-related, platform and affordance-related, and time-related factors. The disclosure decision-making framework for online disclosures [2, 5] helps explain both *direct* and *indirect* disclosures made on public, identified platforms like Facebook and Twitter.

However, there are two additional factors which are unique to public, direct disclosures. Network-level and societal factors also influence the choice to publicly make a direct disclosure, and they go beyond the individual's single disclosure by connecting it to broader stigmatization [2]. Andalibi and Forte define network-level reciprocal disclosures as "disclosures to one's network that are motivated by observing others' disclosures" [2]. In the context of their study, some women felt more comfortable posting about their pregnancy loss when they saw another woman disclose her own pregnancy loss publicly. Relatedly, other women felt more comfortable disclosing in connection with Pregnancy and Infant Loss Awareness Month, a social media campaign designed to reduce stigma around pregnancy loss by illuminating the tragic frequency of the experience. In both cases, Andalibi and Forte argue that network-level reciprocal disclosures can reduce stigma around disclosure and create space for others to disclose [2].

The #MeToo campaign was characterized primarily by many direct, public disclosures, and so it is not well explained by most prior studies on online disclosures, which prioritize anonymity and, in particular, anonymity on Reddit. Instead, the concept of network-level reciprocal disclosure introduced by Andalibi and Forte [2] provides a viable framework for understanding the public disclosures of #MeToo on Twitter. Their framework and case study, however, do not immediately map onto #MeToo. A relatively emergent campaign[1] like #MeToo on a platform like Twitter, which emphasizes even more public sharing than Facebook, may not have been conducive to network-level reciprocal disclosures. By charting disclosures through #MeToo on Twitter, we are able to probe and establish boundaries on the theory of network-level reciprocal disclosures.

2.3 Networked Disclosures Through Hashtag Activism

Hashtags addressing sexual violence against women connect women's disclosures and amplify them to shift public discourse around their stigmatized experiences. Stigmatization depends on an implicit collective enforcement of the boundaries between public and private. Pregnancy loss, sexual assault, domestic violence, and other issues experienced by women have a long history of being excluded

[1] Although #MeToo had organizational infrastructure founded by activist Tarana Burke [37], that infrastructure is relatively small compared to that of Pregnancy and Infant Loss Awareness Month, a campaign which has a decades long history offline and which is internationally recognized by a number of countries [93].

from the public sphere by being deemed private matters [52, 80]. Social media renegotiates the boundaries between public and private by providing women a common platform for sharing their experiences. While their stigmatized experiences may be marginalized by the mainstream public, women can connect online and collectively interrogate and reimagine the narrative surrounding those experiences [70]. As a consequence, the unity that emerges from connecting with and supporting one another gives rise to *networked counterpublics*, online communication networks that both validate the experiences of women and amplify their narratives to a wider audience [34, 47, 84]. The dual functions of validation and amplification prioritize the feminist principle that the "personal is political," centering the everyday stories of women that have typically been marginalized through stigmatization and reframing them as public issues [45]. Together, this allows women to collaboratively weave their lived experiences into a collective narrative that exposes and challenges a broader culture of misogyny.

The #MeToo campaign has sustained an ongoing dialogue about sexual violence against women, but it is only the latest in a line of hashtags that have exposed and denounced such everyday violence. The hashtags #YesAllWomen and #NotOkay, which emerged after the Isla Vista shootings and Trump Access Hollywood video respectively, laid the discursive groundwork for #MeToo's emergence [45, 60, 77, 85]. Between these more visible hashtags, a number of networked counterpublics have coalesced around other hashtags addressing sexual violence against women, including #WhyIStayed [21], #IAmNotAfraidToSayIt [58], #IAmJada [94], #AskThicke [88], #safetytipsforladies [76], and #TheEmptyChair [45]. Collectively, these hashtag campaigns amplify a rejoinder to the mainstream public's inattention to sexual violence against women as an issue of public health [21, 78]. The theory of feminist networked counterpublics, however, only vaguely specifies the mechanisms that stitch these voices together into a cohesive network. In particular, there is a gap in understanding between how women choose to disclose individually and how their collective narrative emerges as a whole [21]. Consequently, the current networked counterpublics perspective on hashtags like #MeToo emphasizes broad social factors that impact sexual violence against women, but loses sight of interpersonal and intrapersonal reasons affecting women's choices to disclose.

A socioecological model, detailing the multi-level factors of society that impact women's health [50], establishes a framework for understanding how individual disclosure and network-level hashtag dynamics build off one another. Specifically, network-level reciprocal disclosure [2] is a promising theoretical mechanism to bridge the overlooked gap between how women disclose and hashtag campaigns emerge. Through #MeToo, we can not only study how the hashtag may have facilitated network-level reciprocal disclosures, but also how network-level reciprocal disclosures may encourage further hashtag activism. Clarifying these cyclical processes would help further explain the phenomenon of public direct disclosures of stigmatized experiences, and the propagation of hashtags addressing sexual violence against women. Taken as a whole, this multi-level view would more firmly situate hashtag campaigns like #MeToo as instruments for addressing sexual violence against women as a public health issue.

3 DATA AND METHODS

Our study was guided by three research questions, addressing the relationship between network-level reciprocal disclosures and hashtag activism through #MeToo:

RQ1. What patterns characterize the usage of #MeToo?
RQ2. How do the disclosure phenomena of #MeToo align with or diverge from the theory of network-level reciprocal disclosure?
RQ3. How did network-level reciprocal disclosures facilitate the emergence of #MeToo as a networked counterpublic?

We address these questions through a mixed-methods approach involving a qualitative content analysis, predictive modeling of disclosures, and a quantitative network analysis.

3.1 Data and Ethics

In February 2018 for a separate project, colleagues at the University of Michigan collected 7,025,786 tweets in English that contained the hashtag #MeToo from the social media analytics company Sysomos, which provides a Firehose-like historical archive covering the previous thirteen months. Using the tweet IDs that our colleagues shared with us, we retrieved 5,472,208 of those tweets by rehydrating them through the Twitter API in September 2018, reflecting a 22% data attrition rate. During our analysis, we uncovered a 24 hour period of missing data, which we filled by directly purchasing the missing tweets from Twitter. We focus our study on the first two weeks of the #MeToo campaign, which covers the wave of disclosures underlying the first rise and peak of the hashtag. Peaks in the hashtag after this period mainly coincide with scandals of prominent political and entertainment figures [6]. The final dataset of #MeToo tweets used for this study includes 1,536,068 tweets, including retweets. A subset of 503,614 of those tweets are *authored* tweets, tweets that are not retweets.

At the time of data collection, our colleagues also collected a one-step snowball sample of the follower-following network, the network of who follows whom on Twitter, using users in the original hashtag dataset as the seeds. They shared the edge list of this network with us for this study. In the time between the tweets being posted and the networks being collected, some users either deactivated or privatized their accounts. Collecting many follower-following networks was also computationally prohibitive, leaving us unable to collect complete follower-following information for 41.1% of those who disclosed (we detail how we identify who disclosed in Section 3.3).

Finally, to directly evaluate reciprocal interactions and social support between those who disclosed, it is critical to study replies to #MeToo tweets. However, most replies are not in the original sample because they do not contain the hashtag itself. We used the Python package Twint[2] to iteratively collect public replies threaded in response to the authored #MeToo tweets during the first two weeks, resulting in 295,008 additional tweets. We consider how the limitations of our data may affect our findings in more detail in the Discussion (see Section 5.4).

Although Twitter data is public and our work using such data is broadly approved by our institution's IRB, disclosures from survivors of sexual violence are particularly sensitive [3, 20]. To prevent identifiability, we do not report on specific users or directly quote any #MeToo tweets in this manuscript [7]. Instead, all examples presented in the paper were adapted by combining multiple, similarly-structured tweets, altering keywords, and removing or changing all identifiable details (e.g., location, name, age) while maintaining the overall sentiment and message of the tweets. Further, while we release the tweet IDs supporting this work to facilitate open science and replicability,[3] we do not indicate which tweets we have identified as disclosures in order to avoid creating a public list of survivors of sexual violence.

3.2 Qualitative Content Analysis Procedure

We randomly sampled 2,500 authored tweets from the two week sample window. If a tweet was retweet or it was a reply not using the hashtag #MeToo, then it was excluded from the sampling for qualitative coding. We chose this sample size to ensure that we had a sufficient number of tweets to later train a predictive model (see Section 3.3). Two authors undertook an iterative, open coding process of labeling emergent phenomena to inductively generate a set of codes. The two authors

[2]https://github.com/twintproject/twint

[3]Tweet IDs will be available through the Inter-university Consortium for Political and Social Research.

independently pilot coded the first 900 tweets in batches of 300 tweets and met after each batch to discuss and refine a codebook. Codes were developed inductively; however, the definitions of three codes, *Descriptive Disclosure*, *Hollywood Figures*, and *Political Figures* were guided by three codes *Personal Narrative*, *Mentions of Celebrities and Entertainment*, and *Mentions of Politics or Political Figures* which were derived by Pew Research Center in a report on #MeToo [6]. Coding was done on the level of individual tweets, and each tweet could be labeled with multiple binary thematic codes. For example, *Descriptive Disclosure* and *Disclosing Multiple Incidents* could both be applied to the same tweet.

After solidifying a codebook (see Appendix A), we trained two undergraduate research assistants to deductively code #MeToo tweets following the codebook. One author coded all 2,500 tweets and the research assistants coded two disjoint batches of 1,250 tweets each. Authors and research assistants met to discuss discrepancies in the codes. Intercoder agreement was calculated for each code, with a Cohen's kappa κ of 0.6 to 0.79 considered moderate agreement, and a κ greater than 0.8 considered strong agreement [25, 35, 64]. We report on a subset of codes that had above an intercoder agreement above 0.6 and that are relevant to the our findings on network-level reciprocal disclosure.

Finally, in order to utilize the full subsample of content analysis tweets to later train a predictive model, a second author who did not originally code all 2,500 tweets reconciled all disagreements on the label of *Disclosure*.

3.3 Automated Classification of #MeToo Disclosures

To scale our content analysis to the full set of 1.8 million #MeToo tweets, we automated the identification of disclosures using a predictive classifier trained on the data produced by the content analysis. Given that such classifications subsequently identify individuals who have experienced sexual violence, automating this process requires careful consideration. As we describe shortly, we primarily used language features that directly indicate disclosure [8, 29, 92]. We did not derive higher-level features using methods such as topic models or document embeddings and we did not use information beyond the tweet itself, that is, we did not engineer features based on past user language or activity. Avoiding these higher level features imitates our content analysis and avoids exploiting information that a user did not directly disclose in their #MeToo tweet [5, 20]. We note that focusing on the language of tweets meant that we were relying on the context of the first two weeks of #MeToo to make accurate predictions. For example, personal pronouns are predictive of a disclosure in our dataset, but they would not be in general. While this limits the generalizability of our classifier, it also limits its potential misuse while still giving us insights into network-level reciprocal disclosures [2]. Regardless, similar classifiers could be abused in other settings, which we reflect on further in the Discussion (see Section 5.3).

Using the binary label *Disclosure* from the content analysis subsample of 2,500 tweets, we trained a predictive model to make out-of-subsample classifications of disclosures among all 1.8 million #MeToo tweets posted during the first two weeks of the hashtag campaign. We note that the training set of 2,500 tweets was relatively balanced: 46% of tweets from the content analysis were identified as disclosures.

Using their knowledge from the content analysis, two of the authors manually and iteratively derived 50 model features (i.e. independent variables) that capture a tweet's content, structure, and language with respect to #MeToo and disclosure. Examples of content features include the number of links and whether a picture was embedded in the tweet. Structural features capture details such as whether the tweet only included "#MeToo" or if it was in reply to Alyssa Milano's original "me too" tweet. Lexical features group words into features thematically, where any word in a given group counts towards that feature. For example, "i," "me," "my," and "mine" all count towards the

Code	Cohen's κ	Percent of Total	n
Disclosures (General)	0.89	46.1%	1152
#MeToo	0.95	14.8%	372
Descriptive Disclosure	0.84	25.9%	645
Disclosing Multiple Incidents	0.71	6.4%	159
Prevalence and Pervasiveness of Sexual Violence	0.70	8.2%	206
Next Steps in Addressing Rape Culture	0.63	6.5%	163
Non-Women and Male Survivors	0.65	2.6%	65
Disbelieving / Maligning #MeToo	0.65	2.1%	53
Hollywood Figures	0.74	6.3%	157
Political Figures	0.72	4.8%	121

Table 1. Reported codes and their prevalences from the content analysis of 2,500 #MeToo tweets. Intercoder agreement is measured using Cohen's κ. See Appendix A for full code definitions.

feature of personal pronouns, while "skirt," "jeans," and "shirt" all count towards the feature of clothing phrases. The 42 lexical features encompass 212 phrases in total. See Appendix B.1 for a full description of the model features.

Using a logistic regression model with L2 regularization, these features achieved an average F1 score of 0.82 (precision = 0.86, recall = 0.80) and an average AUC score of 0.93 over 5-fold cross validation. Other tree-based and ensemble-based classifiers performed similarly in terms of predictive metrics while being less immediately interpretable, so we chose to use logistic regression as the final model. To determine whether our model had systematic patterns in the types of tweets it was and was not properly classifying as disclosures and, specifically, if it was marginalizing particular types of disclosures, we audited all misclassifications. We found that false negatives were primarily indirect disclosures which, as we discussed above, we would expect to misidentify given our chosen feature set (see Appendix B.2 for details).

4 RESULTS

4.1 Characterization of #MeToo Tweets

Through the content analysis, we distinguished tweets that were disclosures from those that were not. We further explored the characteristics of disclosure and of a non-disclosure tweets (Table 1). Of the 2,500 tweets, nearly half were disclosures of sexual violence (46.2%), tweets in which the user revealed information about or the existence of a personal experience of sexual violence [23]. Our criteria for what constituted an experience of sexual violence was broad, deferring to how the user defined their own experience and their choice to use the hashtag. Note, because of our focus on disclosures people made about their own experiences, we did not label disclosures a user made for someone else as disclosures (e.g., "This #MeToo is for my mother who is not alive to participate"). We labeled tweets as disclosures if they contained only "#MeToo" (14.8%) or if the tweet included discussion of the individual's personal experience of assault or harassment (25.8%). Tweets that contained any amount of detail beyond "#MeToo" were labeled as *Descriptive Disclosures*. Descriptive disclosures fell into a number of inductively created subcategories, the most common of which was *Disclosing Multiple Experiences*. In approximately 25% of descriptive disclosures, individuals opted to discuss multiple personal experiences, highlighting that sexual violence has been an ongoing reality or recurrent experience in their lives (e.g., "First assaulted at age 5 by an adult male, second time as a waitress as a teen. #MeToo") (6.4% of all #MeToo tweets).

Additionally, in only 2.2% of the disclosures did the author of the tweet explicitly identify as not a woman (e.g., "Not a woman but I'm going to add #MeToo"). This suggests that Alyssa Milano's original "me too" tweet, which explicitly called on women to speak about their experiences, set a strong expectation for who was intended to participate.

Of the 2,500 tweets, just over half were not labeled as disclosures (53.8%). These tweets included wider discussions of sexual violence and its impact on society. Connecting sexual violence within the political or entertainment worlds, numerous tweets included references to *Hollywood* (6.3%) and *Political Figures* (4.8%) connected to the #Metoo movement (e.g., "I hope that the #Weinstein scandal exposes more Hollywood scum. Very proud of the brave ppl saying #MeToo" and "These women spoke out against @realDonaldTrump and their voices matter. If you believe other women why not them? #MeToo"). The prevalences of *Descriptive Disclosures* (25.9% vs 14%), *Hollywood Figures* (6.3% vs 15%), and *Political Figures* (4.8% vs 7%) in our dataset differ from the findings of the Pew Research Center, which analyzed prevalence of these types of tweets in 2017 and 2018 during five separate weeks surrounding major events that coincided with later peaks in #MeToo usage (e.g., Time Magazine's Person of the Year, International Women's Day) [6]. During these times, #MeToo tweets focused less on descriptive disclosures and more on topics related to entertainment and political figures. Given that later peaks in #MeToo showed less disclosures and more discussion of current events, we see that individual disclosures were particularly abundant early in the hashtag campaign.

Additionally, users commented on the prevalence of sexual violence in general, often referencing the flood of #MeToo disclosures within their own feeds (e.g., "About 30 people I know have posted #MeToo in just the last 24h. What kind of world do we live in?") (8.2%). Finally, some non-disclosures went further and took the opportunity to open a dialogue on sexual violence and suggest actions that people can take to further challenge rape culture (e.g., "#MeToo was last week. Here are some ways to help combat sexual abuse today [hyperlink]") (6.5%). The many disclosures made through #MeToo enabled a discussion of sexual violence within culture more broadly, with tweets placing a spotlight on the prevalence of sexual violence and advocating for ways to change the culture that supports it.

Notably, few of the non-disclosure tweets were in opposition to the movement. Hashtags, in general, and hashtag activism, specifically, are susceptible to "hashtag hijacking," [47] in which users flood a hashtag with dissenting or opposing opinions to fragment the hashtag's connective power [36, 47]. For example, in the aftermath of the Isla Vista shootings, the hashtag #YesAllWomen faced backlash from men who also took to the hashtag #NotAllMen to dismiss women's experiences with sexual violence [77]. However, only about 2% of the manually-coded tweets containing #MeToo maligned the hashtag campaign or expressed disbelief of those who disclosed. The *absence* of these tweets likely played a role in establishing #MeToo as a space safe for disclosing.

4.2 Networked Disclosures

4.2.1 Timeline of Disclosures. Of the 1.8 million #MeToo tweets posted during the first two weeks of the hashtag campaign, we applied our disclosure classifier to the approximately 500,000 *authored* #MeToo tweets, i.e. those #MeToo tweets that are not retweets. Retweets of disclosures and replies to #MeToo tweets that did not contain the hashtag were not classified as disclosures. Fig. 1 shows the timeline of the number of descriptive, non-descriptive, and total disclosures during the first week of the two week sample window. Over the entire two week period, 201,395 individuals disclosed through the hashtag. Those users who disclosed composed 235,112 disclosure tweets,

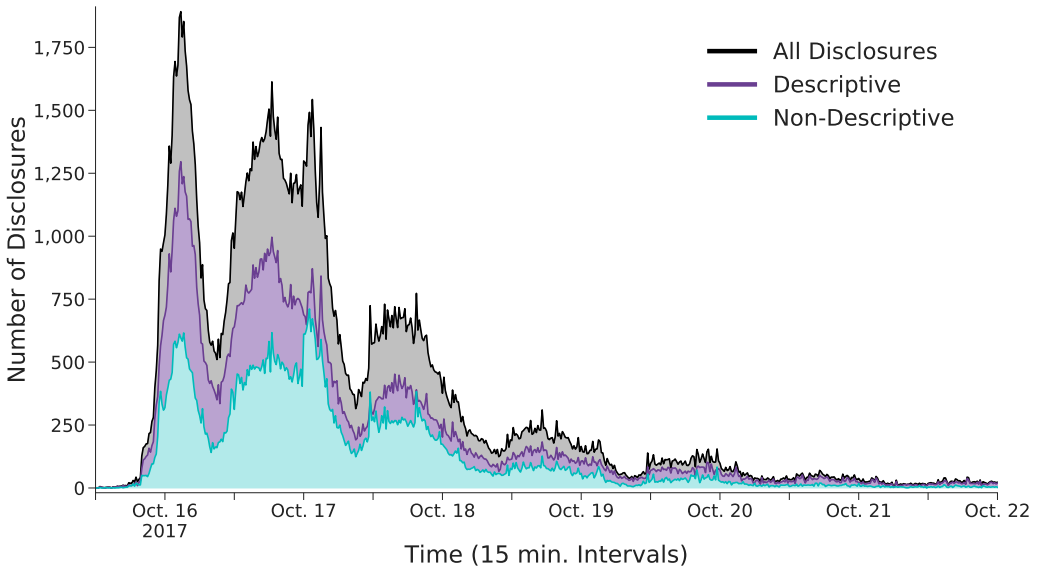

Fig. 1. Number of disclosure tweets during the first week of the #MeToo campaign. Inset shows disclosures as a fraction of all #MeToo tweets during the same time period. During the second week (not shown), #MeToo tweets continue to see a low-volume daily rhythm, while disclosures subside almost entirely.

which accounted for 51.7% of all authored #MeToo tweets [4] (15.1% of all #MeToo tweets including retweets). Nearly two thirds (63.7%) of disclosures were descriptive, containing details beyond only the "#MeToo" hashtag. Descriptive disclosures were consistently tweeted more than non-descriptive disclosures over nearly the entire sample window (Fig. 1). Proportionally, descriptive disclosures made up a larger fraction of all #MeToo tweets during the first week of the hashtag campaign, until tapering to a similar proportion of non-descriptive disclosures after the first week.

4.2.2 Descriptive Disclosures and Follower Disclosures. Although we do not have information on who could have disclosed during #MeToo but chose not to, we can measure variations in the levels of detail that individuals shared with their disclosures. We find that the probability of making a descriptive disclosure had a weak relationship with the number of followers a user had, beyond more than approximately 200 followers (Fig. 2A). A priori, we may expect that users with more followers may make less descriptive disclosures because they are more visible or they may make more descriptive disclosures because they are more sociable, but beyond the 200 follower threshold, the probability of making a descriptive disclosure is relatively constant. For those with less than 200 followers (41% of those who disclosed, Fig. 2C), there is a noticeable decline of 10 to 15 percentage points in the likelihood of making a descriptive disclosure. Even at these reduced rates though, nearly all of those who disclosed were more likely to make a descriptive disclosure than a non-descriptive disclosure, despite any level of public visibility.

Given the varying levels of publicity, we next consider how a descriptive or non-descriptive disclosure by an individual relates to the disclosures their followers later make. Using the follower

[4]Note, the number of disclosure tweets is more than the number of users who disclosed for two reasons. First, an individual can make multiple tweets disclosing experiences of sexual violence, which all count as disclosures. For all following calculations, we only use a user's first disclosure tweet as their "official" disclosure. Second, some of the excess tweets are misclassifications by our disclosure classifier.

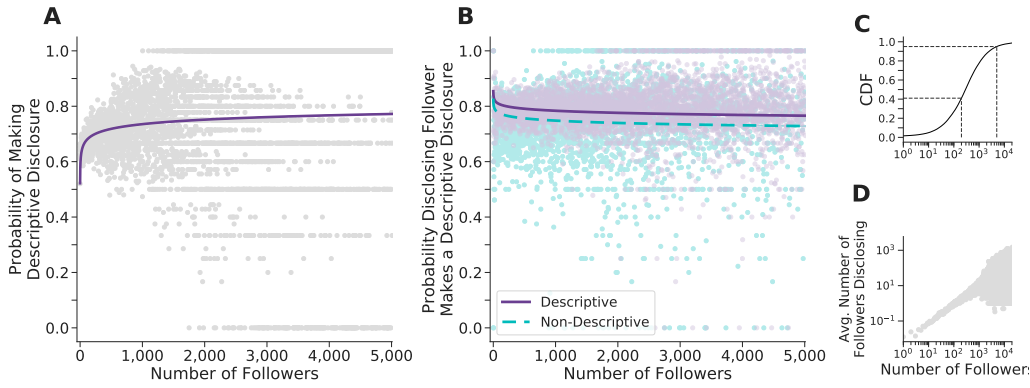

Fig. 2. **A)** Probability of making a descriptive disclosure. Raw probabilities per number of followers are shown in gray and fit with a logistic regression. **B)** Probability of a follower making a descriptive disclosure given that they make a disclosure, conditioned on whether the original user made a descriptive (solid purple) or non-descriptive (dashed teal) disclosure before the follower. **C)** Cumulative distribution function of number of followers of those who disclosed. Dashed lines indicate approximately 200 followers (elbows of curves in **A** and **B**) and 5,000 followers (95% of all those who disclosed). **D)** Average number of disclosures following a user's disclosure, given their number of followers. For **A** and **B**, logistic regressions were fit using an inverse hyperbolic sine transformation on the number of followers.

network, we measure the probability that a descriptive or non-descriptive disclosure is followed by a descriptive or non-descriptive disclosure from a follower who chooses to disclose (Fig. 2B). We fit one logistic regression with number of followers and type of disclosure as independent variables and the type of disclosure made by the follower as the dependent variable. Beyond the 200 follower threshold, there is no relationship between an individual's visibility (i.e. their number of followers) and the probability a disclosing follower later made a descriptive disclosure. Rather, while the disclosures of popular users were followed by more disclosures overall, likely because they generally had more followers (Fig. 2D), disclosures that were made to smaller followings (less than 200 followers) were more likely to be followed by disclosures that shared details. We find that a descriptive disclosure itself is slightly more likely to be followed by a descriptive disclosure from a follower who discloses ($p \ll 0.01$). As a whole, we see that individuals of various levels of popularity and visibility played a role in driving disclosures within #MeToo.

4.2.3 Exposures to Disclosures. From the perspective of an individual and their followers, we cannot establish that a single person's disclosure caused disclosures among their followers because those followers may have been exposed to multiple disclosures from others they were following. To better understand this dynamic, we shift to the perspective of individuals and who they were following, giving us a viewpoint on disclosures they may have been exposed to prior to disclosing. We estimate the number of disclosures an individual was exposed to by counting the number of disclosures posted or shared by those the individual was following, also known as their "friends." These are potential exposures because we do not know what any given individual may have seen on their Twitter timeline prior to disclosing; they may not have seen disclosures from or shared by their "friends" simply because they were not on Twitter at the time, or because the Twitter timeline

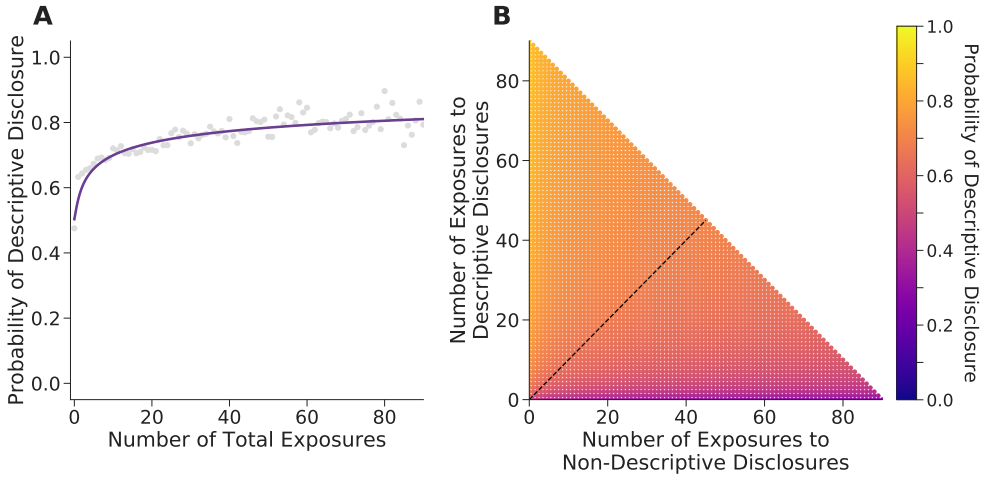

Fig. 3. **A)** Probability of making a descriptive disclosure versus the number of total potential exposures to disclosures prior to disclosing. **B)** Probability of making a descriptive disclosure given the number of potential exposures to descriptive and non-descriptive disclosures prior to disclosing. Dashed line denotes the line of equality. For **A** and **B**, logistic regressions were fit using an inverse hyperbolic sine transformation on the number of prior exposures. Bootstrapped 95% confidence intervals are plotted but are so tight they are not visible. Plots are bounded at 90 total potential exposures, which accounts for 95% of all those who disclosed.

algorithm prioritized some disclosures over others.[5] For brevity and consistency with prior work on information exposure on Twitter [39], we define "exposures" to disclosures as potential exposures to disclosures.

Up to 87.4% of users ($n = 102,546$) were exposed to a disclosure through their following network prior to disclosing with 19.5% exposed only from their friends and 63.7% exposed from both their friends and others via retweets. At most 4.1% of individuals saw disclosures only from people they were not following – disclosures of people who their friends retweeted but who they were not following themself – which highlights the locality of network-level reciprocal disclosures.

Considering disclosures made by friends and others outside the local ego network together, we observe that the more prior exposures that an individual potentially had to disclosures before disclosing themself, the higher the probability that their disclosure was descriptive (Fig. 3A). Through a logistic regression fit on the number of prior exposures, we find that every additional exposure increased the probability of a descriptive disclosure ($p \ll 0.01$). The effect of exposures is most stark among those who did not see any disclosures through their following network prior to disclosing: potentially seeing a single disclosure through one's local network increased the probability of making a descriptive disclosure by 15 percentage points. The more that one's friends made disclosures, the more likely that a person shared details about their stigmatized experience.

We further break down exposures by the number of exposures to descriptive and non-descriptive disclosures. Those who were exposed to only non-descriptive disclosures had a 52.9% chance of making a descriptive disclosure, while those seeing either only descriptive disclosures or both descriptive and non-descriptive disclosures had 71.7% and 71.1% chances respectively of making a

[5]Those who disclosed may have also been exposed to #MeToo through Twitter's trending hashtags, Facebook, or news media. This is certainly the case for those who were "not exposed" to any disclosures because, despite not having seen disclosures through their "friends," they still knew about the hashtag and used it for disclosing.

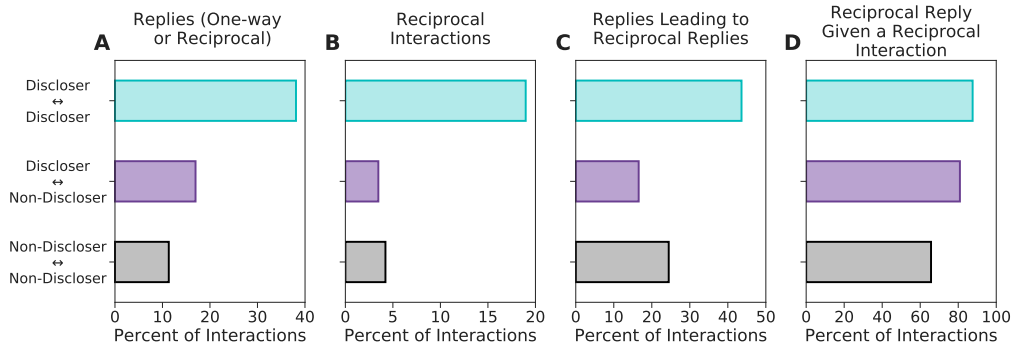

Fig. 4. Comparison of rates of interaction (retweets, replies, mentions) and reciprocity among different dyadic compositions of those who did and did not disclose. **A)** Percent of *all interactions* within each dyad type (each horizontal bar) that included a reply in at least one direction. **B)** Percent of *all interactions* within each dyad type that were reciprocal. **C)** Percent of *reply interactions* within each dyad type that were reciprocal replies (i.e. cases where someone replies to a reply they received). **D)** Percent of *all reciprocal interactions* within each dyad type that included reciprocal replies. Bars do not add up to 100% because each bar stands for interactions among a different dyadic composition.

descriptive disclosure. We decompose this phenomenon in Fig. 3B, which shows the probability of making a descriptive disclosure given a number of exposures to descriptive and non-descriptive disclosures each. The more that an individual saw others sharing details with their disclosures, the more likely that individual was to share details themself with their disclosure.

However, exposures to descriptive disclosures were not equally distributed throughout the #MeToo network. Although those who ultimately made either a descriptive or non-descriptive disclosure were equally likely to have been exposed to at least one non-descriptive disclosure prior to disclosing (62%), those who eventually made a descriptive disclosure were more likely to have seen a descriptive disclosure prior to disclosing (84.5%) than those who made a non-descriptive disclosure (78%). So even if exposures to descriptive disclosures did reduce stigma and encourage more descriptive disclosures, that stigma reduction may have already been concentrated in areas of the network where individuals felt more comfortable sharing details about their experiences. We discuss these potential inequities further in Section 5.2.2.

4.3 Reciprocal Interactions and Movement Building

4.3.1 Interactions Between those who Disclosed. Having examined how #MeToo as a hashtag facilitated network-level reciprocal disclosures, we now reverse the direction of inquiry and look at how network-level reciprocal disclosures helped promote #MeToo as a hashtag campaign. We first illuminate how those who disclosed through #MeToo made up a disproportionate fraction of the hashtag's communicative backbone by comparing dyadic interactions between those who disclosed to interactions between those who did not disclose and between one individual who disclosed and one who did not (Fig. 4).

Given the time needed to construct a reply, replies are generally considered a more effortful communicative act than retweets [51]. It is notable, then, that although only 25.1% of individuals who tweeted using #MeToo disclosed, they composed 51.7% of all authored #MeToo tweets overall and, in particular, 42.1% of all replies to #MeToo tweets. Among those who disclosed, 37.0% of all their dyadic interactions involved a reply, while, comparatively, only 10.7% of interactions between those who did not disclose involved a reply. Further, a reply from one person who disclosed to

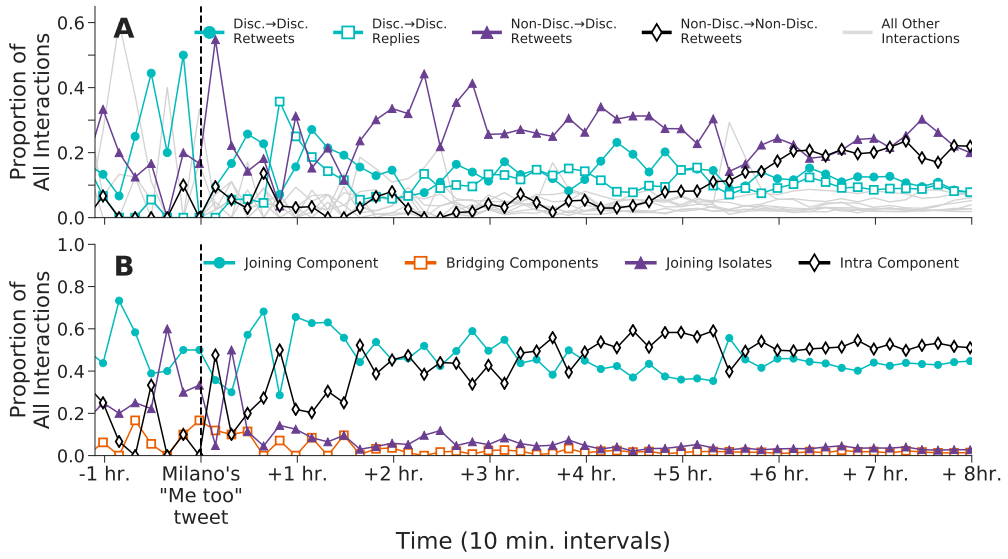

Fig. 5. Interactions (retweets, replies, mentions) within #MeToo during the first 9 hours of the hashtag. Dashed line indicates Alyssa Milano's "me too" tweet. **A)** Interactions between those who disclosed and those who did not. The interactions accounting for most activity are highlighted, while all others are shown in gray. **B)** Interactions that constructed the eventual giant weakly connected component of #MeToo. "Joining component" refers to an isolate node joining a network component of more than two nodes; "Bridging components" refers to two network components joining together; "Joining isolates" refers to a link being formed between two nodes that were previously isolated; "Intra component" refers to interactions between nodes that are already part of the largest connected component.

another led to a reciprocal reply 39.8% of the time, which is nearly twice the rate of replies among those who did not disclose (Fig. 4C).

Considering all dyadic interactions of retweets and replies within the #MeToo interaction network as a whole, the subnetwork between those who disclosed is twice as dense as the subnetwork among those who did not. Within these networks, dyads of those who disclosed are far more likely to have reciprocally interacted than any other dyads (Fig. 4B). All together, the exceptional rates of interactions between those who disclosed, as compared to those who did not, indicate the connective importance of network-level reciprocal disclosures within #MeToo (Fig. 4). The gulf between those who disclosed and those who did not in terms of their replies and reciprocity, which are particularly meaningful types of interaction [14, 51], suggests that those who disclosed through #MeToo played a fundamental role in establishing a social support structure underlying the broader hashtag campaign.

4.3.2 Emergence of #Metoo Network. The communicative structure created by those disclosing during #MeToo played an important role in fueling the hashtag during its early hours. In Fig. 5, we track the emergence of the #MeToo network's giant weakly connected component. Simultaneously, we analyze the relationship between interactions (retweets, replies, mentions) among those who disclosed and did not disclose (Fig. 5A) and how those interactions connected the emerging network (Fig. 5B). There are four ways an interaction could connect the #MeToo network:

(1) **Joining Component:** An interaction connected a single individual to a component of the #MeToo network. We consider a component to be any subnetwork with more than two individuals already connected.
(2) **Bridging Components:** An interaction connected two components of the #MeToo Network, each of which already connected more than two individuals.
(3) **Joining Isolates:** An interaction connected two individuals, neither of whom were connected to a component with other individuals.
(4) **Intra Component:** An interaction occurred between two individuals who were already both a part of the same component of the #MeToo network.

Parsing the emergence of the network in this way allows us to observe how interactions between those who disclosed and did not disclose coincide with how the #MeToo network was socially woven.

During the first 1.5 hours, retweets between those who disclosed rose from 20% to 30% to 50% of all interactions (Fig. 5A). An hour and a half into the hashtag campaign, Alyssa Milano made her "me too" tweet which prompted a number of tweets from those who did not disclose. While those who disclosed were still involved in over half of all interactions, they were retweeted primarily by those who did not disclose. By Fig. 5B, we see that this time coincides with many isolates joining the largest component of the #MeToo network.

Although retweets of those who disclosed subsided for the next 1–2 hours, replies between those who disclosed rose dramatically (Fig. 5A). Given that the number of intra-component interactions rose at the same time (Fig. 5B), the timings of these replies suggests they were in response to the disclosures retweeted in the immediately preceding hours of #MeToo and reinforcing the existing network. As these replies tapered proportionally (though those who disclosed maintained a higher rate of interaction than those who did not disclose) for the next 3 hours, those who did not disclose began amplifying those who did through significantly more retweets. About 7 hours into the #MeToo hashtag campaign, retweets of individuals who did not disclose began to climb in frequency, matching the rate of retweets of those who did disclose (Fig. 5A). Concurrently, balance shifted from individuals primarily joining the #MeToo network to individuals communicating within the existing network (Fig. 5B).

Through this interaction timeline, we see that early interactions between those who disclosed established the #MeToo network, laying the kindling for later #MeToo activity. Following these early interactions, allies and others who did not disclose helped give momentum to the hashtag's wildfire by retweeting those who did disclose. The shift from joining and centralizing marginalized narratives to amplifying them beyond those who are directly affected is a signature of hashtag activism [9]. In the case of #MeToo, early disclosures helped build the campaign within the first 9 hours. While those who did not disclose then played an important role in raising the visibility of those disclosures, those early disclosures helped the network-level reciprocal disclosure effects snowball, leading to many more disclosures in the following 72 hours. As we observe, when used strategically and together with allyship, network-level reciprocal disclosures can help construct a broader movement that generates disclosures in a self-sustaining way.

5 DISCUSSION

We summarize our findings and interpret them in terms of the theory of network-level reciprocal disclosures [2]. We argue that our work provides additional evidence in a new context that network-level reciprocal disclosures may reduce stigma locally, potentially generating more disclosures that helped fuel the #MeToo hashtag campaign. We then discuss future directions for our work, emphasizing the need to better understand inequities within and produced by campaigns like

#MeToo. Given that our findings point towards the need for more research on how public health inequities may be reproduced through hashtags like #MeToo, we finish by cautioning practitioners from designing interventions on disclosures on networked platforms.

5.1 Network-Level Reciprocal Disclosures and Hashtag Activism

By leveraging a mixed-methods approach, this paper details the number and types of disclosures made during the early weeks of the #MeToo hashtag campaign and contextualizes their spreading in terms of theories of network-level reciprocal disclosures [2] and feminist hashtag activism. Through a qualitative content analysis of 2,500 authored #MeToo tweets, we provide a comprehensive characterization of the stigmatized stories that emerged through the hashtag campaign. Using our insights from the content analysis, we then designed a reliable predictive model for identifying #MeToo tweets as disclosures, allowing us to scale our analysis up to over 500,000 authored #MeToo tweets. Together, the disclosures identified by this model and the underlying follower-following network of #MeToo users reveal that public disclosures by individuals were followed by additional disclosures by their followers, and that the more disclosures that an individual potentially saw prior to disclosing, the more likely they were to share details with their own eventual disclosure. Further, we found that those disclosing early in the hashtag campaign promoted and supported one another's stories through retweets and replies, weaving a networked counterpublic that allowed others to also join and share their stigmatized experiences.

5.1.1 What patterns characterize the usage of #MeToo? How do the disclosure phenomena of #MeToo align with or diverge from the theory of network-level reciprocal disclosure? We extend the work of Andalibi and Forte [2] on network-level reciprocal disclosures to the context of the #MeToo campaign, allowing us to map aspects of how network-level reciprocal disclosures emerged through the #MeToo movement. We found that individuals were more likely to make a descriptive disclosure if they potentially saw more disclosures, particularly if those disclosures also shared details. Among these descriptive disclosures, our close reading of #MeToo tweets revealed that a number of individuals explicitly mentioned disclosing after seeing the number of survivors in their personal networks (e.g., "My entire feed is full of #MeToo. Me too. I'm not an exception."). Others disclosed for the first time after seeing others disclose, referencing years of prior hesitation (e.g., "I could not speak out for YEARS due to social pressures. Things need to change. #MeToo") or shame (e.g., "I felt shame and guilt for years. I'm done with that. It was not my fault. #MeToo"). Together, our qualitative and quantitative findings suggest that individuals felt more comfortable sharing their stories the more that they saw the stories of others. These findings provide new evidence of the network-level effects that constitute the theory of network-level reciprocal disclosures [2], and the potential of such disclosures for reducing stigma.

Aspects of these findings are also consistent with additional motivations an individual may have had for disclosing through #MeToo, such as feeling a collective responsibility to name the issue of sexual violence against women [17, 22]. That is, as individuals saw more disclosures, they may not have felt a reduction in stigma but, nonetheless, felt obligated to share their own stories. A number of those disclosing actively addressed the public epidemic of sexual violence against women (e.g., "#MeToo - Amazed by everyone speaking out. We ALL must challenge #rapeculture. Here are ideas how. [hyperlink]"), placing more emphasis on the collective experience than their own feelings about sharing their story. Further, some survivors clearly felt pressure to share their stories as survivors, causing them to push back on the approach of #MeToo to addressing sexual violence (e.g., "Why is #MeToo asking me to relive the trauma and deal with people asking what happened? Focus on the abusers not the victims"). These reactions to #MeToo are related to the societal factors described by Andalibi and Forte [2] in their disclosure decision-making framework, which are

factors that are explicitly political reasons for publicly and directly disclosing. However, based on our content analysis, disclosures that simultaneously addressed the pervasiveness of sexual violence against women only made up 5.5% of all manually-coded disclosures (n=63), and at most 0.8% of all manually-coded #MeToo tweets (n=20) critiqued the necessity of sharing survivor stories. Further, those with the least followers were more likely to share details with their disclosures, which is less congruent with the hypothesis that disclosures were made for societal reasons because those individuals had the smallest platform to name the issue.

These findings suggest that the increased probability of making a descriptive disclosure upon seeing more descriptive disclosures is likely driven primarily by network-level reciprocal disclosures, rather than disclosures based on societal factors. This underscores how the "personal is political" in hashtag campaigns like #MeToo and the power of each woman's local network.

5.1.2 How did network-level reciprocal disclosures facilitate the emergence of #MeToo as a networked counterpublic? As discussed in Section 5.1.1, our findings show that although network-level and societal factors are intertwined in hashtag campaigns like #MeToo, hashtags addressing sexual violence against women can emerge organically when they solicit personal stigmatized experiences. So, we observe that while the hashtag #MeToo facilitated network-level reciprocal disclosures [2], conversely, network-level reciprocal disclosures also facilitated the propagation of the #MeToo campaign. Feminist networked counterpublics created by hashtags like #MeToo, #YesAllWomen, and #NotOkay are understood to unify stigmatized stories into a collective narrative to address the epidemic of sexual violence against women [21, 70]. Furthermore, the discursive work in these counterpublics prioritizes that the "personal is political" [11, 21, 45], but it has been less clear how individual stories coalesce around these hashtags. Operationalizing network-level reciprocal disclosure as a mechanism underlying feminist hashtag campaigns provides insight into this phenomenon.

Specifically, our analysis of the #MeToo network's emergence showed that early in the hashtag campaign those disclosing amplified the voices of others disclosing through retweets. Together then, those disclosing (e.g., "#MeToo - and to everyone else finding their voice today (and even those who are silent) #IBelieveYou") and allies (e.g., "To all the women, all the survivors. I hear you. I believe you. I support you. #MeToo") transformed the network into a space that survivors of sexual violence could join to disclose their personal experience. Coupled with our argument that #MeToo was primarily driven by network-level factors rather than societal factors, this suggests that the success of hashtags like #MeToo [21, 45] may be intimately related to their ability to reduce stigma locally in social networks and not just collectively, so that individuals feel comfortable sharing their stories. Centering network-level reciprocal disclosures [2] as the foremost mechanism underlying #MeToo emphasizes personal stories, and offers an explanation for how those stories are coupled to make a broader political message.

5.2 Avenues for Further Understanding Network-Level Reciprocal Disclosures

Our work connects network-level reciprocal disclosures with feminist networked counterpublics, laying the groundwork for several avenues of further research at the intersection and union of public health, human-computer interaction, and communication studies. We detail how our understanding of network-level reciprocal disclosures could be augmented by studying their relationship with social support, health outcomes, health inequities, and media effects.

5.2.1 Social Support and Network-Level Reciprocal Disclosures. The present work has focused on potential exposures to disclosures and how those may encourage network-level reciprocal disclosures [2]. These disclosures are tightly connected to the social support structure underlying the #MeToo network, as shown by the high rate of interactions between those who disclosed and

the importance of those interactions in driving #MeToo as a hashtag. However, while our work and other prior work has begun to provide insight into that social support structure [15, 44, 61], there are still dynamics between disclosures, replies, and perceived social support that remain to be unpacked. As also briefly noted by Andalibi and Forte [4], it is likely that, more than simply seeing a disclosure, the number of reactions to that disclosure and the supportiveness of those reactions also influences an individual's choice to make a network-level reciprocal disclosure. More firmly establishing the connection between network-level reciprocal disclosures and replies to those disclosures around hashtag campaigns like #MeToo would highlight how feminist networked counterpublics are able to not only conduct activist campaigns, but also empower and validate women who have had their experiences marginalized [45, 84].

5.2.2 Inequities of Network-Level Reciprocal Disclosures. With her first "me too" tweet, Alyssa Milano invited women to share their experiences so that they "might give people a sense of the magnitude of the problem." Our study has looked at how women chose to share those experiences, which are typically stigmatized through the prevalence of rape myths that discredit experiences of sexual violence. However, the level of acceptance of such rape myths varies between subpopulations, based on race, gender, age, and other cultural factors [49]. As a consequence, some groups of women, such as multiply marginalized Black women and trans women [81, 86], and men [81] may face higher levels of stigma against disclosing their experiences of sexual violence, making them less likely to engage in network-level reciprocal disclosures through hashtag campaigns like #MeToo. If this were the case, then the hashtag may have failed to create an environment in which everyone felt equally comfortable disclosing, and only a narrow portion of the population may have benefited from the impact of #MeToo. By definition though, our digital trace data does not include those who could have disclosed but chose not to do so. Further survey and interview work is needed to understand who engages with and sees value in hashtag campaigns addressing sexual violence, and who feels excluded and only marginally benefited by them.

Moreover, while it is not necessary for everyone who can disclose to participate in hashtag campaigns like #MeToo, it is important for everyone to have an equal *opportunity* to disclose and receive support if they choose to do so. At the individual level, a network-level reciprocal disclosure opens an individual up to receiving responses from their social network. In some cases, these responses may be supportive and commend the individual's choice to disclose [81]. In others, the responses may perpetuate rape myths and condemn the disclosure through harassment [40, 89]. In both cases, inequities potentially exist in who benefited from network-level reciprocal disclosures through #MeToo. So, in the way that offline support for sexual violence is inequitably distributed and accessed [49], hashtag campaigns like #MeToo may replicate and compound these inequities online. The cross-sectional nature of our study limits us from observing the outcomes of those network-level reciprocal disclosures [1] and calls for longitudinal research designs that track the impacts of network-level reciprocal disclosures made during feminist hashtag campaigns.

At the network level, there are further inequities among sexual violence survivors that could arise from and be exacerbated by hashtag campaigns like #MeToo as they gain momentum. Hashtag activism can centralize marginalized voices in the core of a networked counterpublic and amplify them to audiences outside of that core [9, 47], but not all marginalized narratives may be recognized equally. More specifically, while our study examines #MeToo disclosures in aggregate, it does not speak to what extent disclosures from various subpopulations were *heard*. For example, although there were likely a number of Black women, trans women, and men who disclosed their experiences with sexual violence despite the higher levels of stigma that they face, it is likely that those voices were not equally amplified through #MeToo [46, 53, 77]. Similar to how #MeToo may have failed to create a space that felt equally safe for disclosure among all those who could have disclosed,

the hashtag campaign may have also failed to include the voices of those who did ultimately choose to disclose, inequitably highlighting particular experiences with sexual violence. So, while feminist hashtag campaigns have the potential to benefit public health, they also emerge out of a crowdsourced, networked gatekeeping process [47, 71] that systematically amplifies some network-level reciprocal disclosures but not others. Better understanding these dynamics would allow us to evaluate the implications of such campaigns in the context of public health, and the role of network-level reciprocal disclosures in alleviating or reinforcing inequities of sexual violence along axes of race, gender identity, and sexual orientation.

5.2.3 Media Effects of and on Network-Level Reciprocal Disclosures. The networked gatekeeping of feminist networked counterpublics also interacts with the gatekeeping of traditional media. Not only is there the potential that stigma is reduced at the local level of individuals through network-level reciprocal disclosures, but it is also potentially reduced at a broader community level through the mainstream public's reactions to disclosures. However, because the mainstream media's framing of hashtag campaigns is made in negotiation with networked counterpublics [19, 47, 71, 95], any inequities in whose stories are amplified by the networked counterpublic may be reproduced through traditional media. Conversely, any inequities amplified by the traditional media may be embedded within and reproduced by the networked counterpublic. For example, some research has already found that while media outlets frame articles to be sympathetic towards women's #MeToo stories, they also still frequently portray accused men as powerful, even after the accusations [30]. Understanding the relationships between how stigma is reduced or compounded by interpersonal disclosures and media framing in these kinds of contexts would suggest strategies for reporting on feminist hashtag campaigns that equitably promote the voices of all those who have experienced sexual violence.

5.3 Perils of Designing Interventions on Networked Disclosures

As we have discussed, there are potentially many intricate layers of interactions underlying network-level reciprocal disclosures generated through hashtag campaigns. These interactions are complex because, fundamentally, network-level reciprocal disclosure is a networked phenomenon which exists beyond the context of any one woman's disclosure. The complexity of this phenomenon should give social media platforms and other applied practitioners pause when considering how they may intervene on those who choose to disclose publicly online about sexual violence, even if those interventions may be with good intent, such as connecting individuals who have shared experiences.

Our study begins to suggest inequitable distribution of descriptive disclosures through the #MeToo network by showing that those who may have felt more comfortable making a descriptive disclosure were already more likely to have seen descriptive disclosures than those who made a non-descriptive disclosure. As we argue in Section 5.2, this is only one type of inequity that may be produced by network-level reciprocal disclosures. These potential inequities align with prior theory from networked feminist counterpublics [46, 53, 77] that network-level reciprocal disclosures [2], activated through hashtag campaigns, may inadvertently reproduce existing inequities in whose stories of sexual violence are heard. Consequently, if platforms design algorithmic interventions for promoting (or discouraging) certain interactions around disclosure, they risk reproducing these inequities and, worse, exacerbating them. Because network-level reciprocal disclosures are a networked phenomenon, a community level process that depends on both intrapersonal and interpersonal actions, exacerbating these inequities through algorithmic interventions would be more than a harm to any individual woman, it would be a threat to women's public health more generally.

The perils discussed only consider direct disclosures so far. While indirect disclosures are less tightly related to network-level and societal factors for publicly disclosing [5], we did find evidence that they played a role during #MeToo. Some disclosures were made in response to and in solidarity with other women. However, through preliminary content analysis, we have found that there is considerable ambiguity between reciprocal disclosures made in solidarity ($\kappa = .47$) and general social support ($\kappa = .61$). Further, we did not classify any retweets of disclosures or even retweets of tweets containing only "#MeToo" as direct disclosures. The reasons for retweeting these disclosures are sufficiently ambiguous that they allow women to hide their intentions in plain sight [63], which may provide them the cover to disclose without social repercussions. However, automated classifiers of disclosures cannot easily account for this ambiguity and, if ever actually deployed, would almost certainly only be trained on public direct disclosures. In practice, this would risk making algorithmic decisions on disclosures from women who already felt comfortable enough disclosing, and not on disclosures from others who cannot disclose or are still testing out the waters in their social networks.

Given the potential public health consequences affecting women who have experienced sexual violence, social media platforms and other practitioners should be wary of making algorithmic interventions into network-level disclosures. Since feminist hashtag campaigns have the potential to be conducive for network-level reciprocal disclosures, social media companies should consider consulting with experienced feminist activist groups if they wish to make decisions that have the potential to impact women's health. Twitter, for example, has consulted with the group Women, Action, and the Media to improve the process of identifying abusive harassment of women on their platform [66], and with the Discourse of Online Misogyny (DOOM) project to better understand networked abuse [42]. However, to the best of our knowledge, none of these consultations have evolved into sustained discussions with activist and advocacy groups. We believe that feminist activists that have experience considering the multiplicity of experiences by women, including women of color, trans women, and LGBQ women, should be employed by social media companies to provide guidance on how they can effectively and ethically support women who disclose on their platforms.

5.4 Limitations

Our study provides the first computational approach to understanding network-level reciprocal disclosures [2], but it comes with its own limitations. First, a number of #MeToo tweets were deleted in between the beginning of the hashtag campaign and our data collection. If individuals did not feel supported, then they may have been more likely to delete their disclosures rather than leave them posted with no responses. Relatedly, we found a noticeable lack of hashtag hijacking, but this may be because Twitter itself was more likely to delete abusive #MeToo tweets. Further, if a user was either not supported or harassed, they would also be more likely to make their profile private, explaining a portion of the follower-following network missing data. In all these cases, our data and subsequent analysis may understate the particularly negative consequences of #MeToo that some users potentially experienced. This emphasizes the need for further work through surveys and interviews.

Our analysis of exposures to network-level reciprocal disclosures relies on potential, rather than actual, exposures to disclosures. Most users likely only saw a fraction of potential exposures because they were not logged onto Twitter during the entirety of #MeToo. Moreover, Twitter's algorithmic feed likely prioritized certain disclosures over others. On the other hand, by searching the #MeToo hashtag on Twitter, a number of users likely saw many more disclosures than we measured through the follower-following network. However, without server-side data, we cannot precisely account for overcounting and undercounting exposures to disclosures.

Additionally, our data is observational and we did not explicitly measure stigma. This limits our ability to unequivocally interpret the network-level reciprocal disclosure effects as the result of a reduction in stigma. Further, the fact that exposures to disclosures, and particularly descriptive disclosures, were not randomly distributed throughout the network also limits our ability to make strong causal claims. Together, these limitations call for additional research explicitly measuring stigma reduction in network-level reciprocal disclosures and the inequities that may arise through hashtag campaigns like #MeToo.

6 CONCLUSION

Our study implemented a mixed-methods research design scaling up a content analysis of 2,500 #MeToo tweets to a network analysis of over 1.8 million tweets sent during the first two weeks of the hashtag's widespread usage in 2017. Together, our qualitative and quantitative findings provide evidence that the public, direct disclosures that were made during the #MeToo campaign were prompted by network-level factors of seeing others disclose throughout their local social networks. We demonstrate that these network-level reciprocal disclosures [2], in turn, created a communication network in which others could share their stories, and we argue that this networked counterpublic generated further disclosures through #MeToo. Our findings corroborate the theory of network-level reciprocal disclosures and provide new perspectives on the interplay between individual disclosures and hashtag activism and their potential for reframing sexual violence against women as a public health crisis.

Although the networked #MeToo counterpublic appears to have flourished with a plurality of experiences, we stress that we must listen for whose voices were drowned out or silenced during the #MeToo campaign. As a vessel for renegotiating the stigma around disclosing experiences of sexual violence, #MeToo has the potential to leave the experiences of multiply marginalized groups, such as Black women, trans women, and LGBQ women, out of the shifting collective narrative. By connecting theoretical mechanisms of online public disclosures and feminist networked counterpublics, our work lays the foundations for mapping the extent of such inequities and how they are alleviated or compounded by hashtag activism campaigns like #MeToo, #YesAllWomen, and #NotOkay. Further, our study highlights the importance of applying qualitative close reading, predictive modeling, and social network analysis together in order to disentangle the multiscale phenomena of online networked disclosures. Understanding the intricacies of how disclosures reverberate through online social networks will help us imagine networked platforms that ethically and equitably promote the well-being of those who publicly disclose.

ACKNOWLEDGMENTS

We thank and strongly support the survivors of sexual violence who bravely shared their stories through #MeToo. We thank Jeffrey Lockhart for providing us with the #MeToo tweet IDs and follower-following networks that made this work possible. We also thank Sagar Kumar and Olivia Sterns for their assistance in the qualitative coding process. This research was partially supported by funding and equipment granted by Northeastern University's Global Resilience Institute and NVIDIA Corporation. We are grateful for their support.

REFERENCES

[1] Nazanin Andalibi. 2019. What Happens After Disclosing Stigmatized Experiences on Identified Social Media: Individuals, Dyadic and Social/Network Outcomes. In *Proceedings of the 2019 ACM CHI Conference on Human Factors in Computing Systems*. ACM.

[2] Nazanin Andalibi and Andrea Forte. 2018. Announcing Pregnancy Loss on Facebook: A Decision-Making Framework for Stigmatized Disclosures on Identified Social Network Sites. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 158.

[3] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3906–3918.

[4] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2018. Social Support, Reciprocity, and Anonymity in Responses to Sexual Abuse Disclosures on Social Media. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 5 (2018), 28.

[5] Nazanin Andalibi, Margaret E Morris, and Andrea Forte. 2018. Testing Waters, Sending Clues: Indirect Disclosures of Socially Stigmatized Experiences on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 19.

[6] Monica Anderson and Skye Toor. 2018. How Social Media Users Have Discussed Sexual Harassment Since #MeToo Went Viral. *Pew Research Center* (2018). https://www.pewresearch.org/fact-tank/2018/10/11/how-social-media-users-have-discussed-sexual-harassment-since-metoo-went-viral/

[7] John W Ayers, Theodore L Caputi, Camille Nebeker, and Mark Dredze. 2018. Don't Quote Me: Reverse Identification of Research Participants in Social Media Studies. *NPJ Digital Medicine* 1, 1 (2018), 30.

[8] Sairam Balani and Munmun De Choudhury. 2015. Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1373–1378.

[9] Pablo Barberá, Ning Wang, Richard Bonneau, John T Jost, Jonathan Nagler, Joshua Tucker, and Sandra González-Bailón. 2015. The Critical Periphery in the Growth of Social Protests. *PLoS ONE* 10, 11 (2015), e0143611.

[10] Susan A Basow and Alexandra Minieri. 2011. "You owe me": Effects of Date Cost, Who Pays, Participant Gender, and Rape myth Beliefs on Perceptions of Rape. *Journal of Interpersonal Violence* 26, 3 (2011), 479–497.

[11] W Lance Bennett and Alexandra Segerberg. 2012. The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics. *Information, Communication & Society* 15, 5 (2012), 739–768.

[12] Michael S Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. 2013. Quantifying the Invisible Audience in Social Networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 21–30.

[13] Michele Black, Kathleen Basile, Matthew Breiding, Sharon Smith, Mikel Walters, Melissa Merrick, Jieru Chen, and Mark Stevens. 2011. National Intimate Partner and Sexual Violence Survey: 2010 Summary Report. (2011).

[14] Catherine A Bliss, Isabel M Kloumann, Kameron Decker Harris, Christopher M Danforth, and Peter Sheridan Dodds. 2012. Twitter Reciprocal Reply Networks Exhibit Assortativity with Respect to Happiness. *Journal of Computational Science* 3, 5 (2012), 388–397.

[15] Katherine W Bogen, Kaitlyn K Bleiweiss, Nykia R Leach, and Lindsay M Orchowski. 2019. #MeToo: Disclosure and Response to Sexual Victimization on Twitter. *Journal of Interpersonal Violence* (2019).

[16] Gerd Bohner, Frank Siebler, and Jürgen Schmelcher. 2006. Social Norms and the Likelihood of Raping: Perceived Rape Myth Acceptance of Others Affects Men's Rape Proclivity. *Personality and Social Psychology Bulletin* 32, 3 (2006), 286–297.

[17] Christina A Byrne, Heidi S Resnick, Dean G Kilpatrick, Connie L Best, and Benjamin E Saunders. 1999. The Socioeconomic Impact of Interpersonal Violence on Women. *Journal of consulting and clinical psychology* 67, 3 (1999), 362.

[18] Rebecca Campbell and Sheela Raja. 1999. Secondary Victimization of Rape Victims: Insights from Mental Health Professionals who Treat Survivors of Violence. *Violence and Victims* 14, 3 (1999), 261–275.

[19] Andrew Chadwick. 2017. *The Hybrid Media System: Politics and Power*. Oxford University Press.

[20] Stevie Chancellor, Michael L Birnbaum, Eric D Caine, Vincent Silenzio, and Munmun De Choudhury. 2019. A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media. In *Proceedings of the 2nd ACM Conference on Fairness, Accountability, and Transparency*.

[21] Rosemary Clark. 2016. "Hope in a hashtag": The Discursive Activism of #WhyIStayed. *Feminist Media Studies* 16, 5 (2016), 788–804.

[22] Rosemary Clark-Parsons. 2019. "I SEE YOU, I BELIEVE YOU, I STAND WITH YOU": #MeToo and the Performance of Networked Feminist Visibility. *Feminist Media Studies* Forthcoming (2019).

[23] Paul C Cozby. 1973. Self-Disclosure: A Literature Review. *Psychological bulletin* 79, 2 (1973), 73.

[24] Kayleen A Culbertson, Peter W Vik, and Beverly J Kooiman. 2001. The Impact of Sexual Assault, Sexual Assault Perpetrator Type, and Location of Sexual Assault on Ratings of Perceived Safety. *Violence Against Women* 7, 8 (2001), 858–875.

[25] Sanchari Das, Javon Goard, and Dakota Murray. 2017. How Celebrities Feed Tweeples with Personal and Promotional Tweets: Celebrity Twitter Use and Audience Engagement. In *Proceedings of the 8th International Conference on Social Media & Society*. ACM, 30.

[26] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

[27] Munmun De Choudhury and Emre Kiciman. 2017. The Language of Social Support in Social Media and its Effect on Suicidal Ideation Risk. In *Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM)*.

[28] Munmun De Choudhury, Sanket S Sharma, Tomaz Logar, Wouter Eekhout, and René Clausen Nielsen. 2017. Gender and Cross-Cultural Differences in Social Media Disclosures of Mental Illness. In *Proceedings of the 2017 ACM conference on Computer Supported Cooperative Work and Social Computing*. ACM, 353–369.

[29] Sindhu Kiranmai Ernala, Asra F Rizvi, Michael L Birnbaum, John M Kane, and Munmun De Choudhury. 2017. Linguistic Markers Indicating Therapeutic Outcomes of Social Media Disclosures of Schizophrenia. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 43.

[30] Anjalie Field, Gayatri Bhat, and Yulia Tsvetkov. 2019. Contextual Affective Analysis: A Case Study of People Portrayals in Online #MeToo Stories. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 158–169.

[31] Henrietta H Filipas and Sarah E Ullman. 2001. Social Reactions to Sexual Assault Victims from Various Support Sources. *Violence and Victims* 16, 6 (2001), 673.

[32] Louise F Fitzgerald. 1993. Sexual Harassment: Violence Against Women in the Workplace. *American Psychologist* 48, 10 (1993), 1070.

[33] John D Foubert and Kenneth A Marriott. 1997. Effects of a Sexual Assault Peer Education Program on Men's Belief in Rape Myths. *Sex Roles* 36, 3-4 (1997), 259–268.

[34] Nancy Fraser. 1992. Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. In *Habermas and the Public Sphere*, Craig J Calhoun (Ed.). MIT Press, Chapter 5, 109–142.

[35] Deen G Freelon. 2010. ReCal: Intercoder Reliability Calculation as a Web Service. *International Journal of Internet Science* 5, 1 (2010), 20–33.

[36] Ryan J. Gallagher, Andrew J. Reagan, Christopher M. Danforth, and Peter Sheridan Dodds. 2018. Divergent Discourse Between Protests and Counter-Protests: BlackLivesMatter and AllLivesMatter. *PLOS ONE* 13, 4 (04 2018), 1–23.

[37] Sandra E Garcia. 2017. The Woman Who Created #MeToo Long Before Hashtags. *The New York Times* (2017). https://www.nytimes.com/2017/10/20/us/me-too-movement-tarana-burke.html

[38] Jacqueline M Golding, Judith M Siege, Susan B Sorenson, M Audrey Burnam, and Judith A Stein. 1989. Social Support Sources Following Sexual Assault. *Journal of Community Psychology* 17, 1 (1989), 92–107.

[39] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake News on Twitter During the 2016 US Presidential Election. *Science* 363, 6425 (2019), 374–378.

[40] Amy Grubb and Emily Turner. 2012. Attribution of Blame in Rape Cases: A Review of the Impact of Rape Myth Acceptance, Gender Role Conformity and Substance Use on Victim Blaming. *Aggression and Violent Behavior* 17, 5 (2012), 443–452.

[41] Oliver Lee Haimson. 2018. *The Social Complexities of Transgender Identity Disclosure on Social Media.* Ph.D. Dissertation. UC Irvine.

[42] Claire Hardaker and Mark McGlashan. 2017. Twitter Host CASS event: Twitter Rape Threats and the Discourse of Online Misogyny. *ESRC Centre for Corpus Approaches to Social Science* (2017). http://cass.lancs.ac.uk/twitter-host-cass-event-twitter-rape-threats-and-the-discourse-of-online-misogyny

[43] Lori L Heise. 1998. Violence Against Women: An Integrated, Ecological Framework. *Violence Against Women* 4, 3 (1998), 262–290.

[44] Alec R Hosterman, Naomi R Johnson, Ryan Stouffer, and Steven Herring. 2018. Twitter, Social Support Messages, and the #MeToo Movement. *The Journal of Social Media in Society* 7, 2 (2018), 69–91.

[45] Sarah J Jackson, Moya Bailey, and Brooke Foucault Welles. 2019. Women Tweet on Violence: From #YesAllWomen to #MeToo. *Ada: A Journal of Gender, New Media, and Technology* 15 (2019).

[46] Sarah J Jackson and Sonia Banaszczyk. 2016. Digital Standpoints: Debating Gendered Violence and Racial Exclusions in the Feminist Counterpublic. *Journal of Communication Inquiry* 40, 4 (2016), 391–407.

[47] Sarah J Jackson and Brooke Foucault Welles. 2015. Hijacking #myNYPD: Social Media Dissent and Networked Counterpublics. *Journal of Communication* 65, 6 (2015), 932–952.

[48] Angela J Jacques-Tiura, Rifky Tkatch, Antonia Abbey, and Rhiana Wegner. 2010. Disclosure of Sexual Assault: Characteristics and Implications for Posttraumatic Stress Symptoms Among African American and Caucasian Survivors. *Journal of Trauma & Dissociation* 11, 2 (2010), 174–192.

[49] Barbara E Johnson, Douglas L Kuck, and Patricia R Schander. 1997. Rape Myth Acceptance and Sociodemographic Characteristics: A Multidimensional Analysis. *Sex Roles* 36, 11-12 (1997), 693–707.

[50] Nancy Krieger. 2011. *Epidemiology and the People's Health: Theory and Context.* Oxford University Press.

[51] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a Social Network or a News Media?. In *Proceedings of the 19th International Conference on the World Wide Web (WWW)*. ACM, 591–600.

[52] Joan B Landes. 1988. *Women in the Public Sphere in the Age of the French Revolution.* Cornell University Press.

[53] Daniela Latina and Stevie Docherty. 2014. Trending Participation, Trending Exclusion? *Feminist Media Studies* 14, 6 (2014), 1103–1105.

[54] Alex Leavitt. 2015. This is a Throwaway Account: Temporary Technical Identities and Perceptions of Anonymity in a Massive Online Community. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*. ACM, 317–327.

[55] Kelly L LeMaire, Debra L Oswald, and Brenda L Russell. 2016. Labeling Sexual Victimization Experiences: The Role of Sexism, Rape Myth Acceptance, and Tolerance for Sexual Harassment. *Violence and Victims* 31, 2 (2016), 332–346.

[56] David Lisak, Lori Gardinier, Sarah C Nicksa, and Ashley M Cote. 2010. False Allegations of Sexual Assault: An Analysis of Ten Years of Reported Cases. *Violence Against Women* 16, 12 (2010), 1318–1334.

[57] Heather L Littleton. 2010. The Impact of Social Support and Negative Disclosure Reactions on Sexual Assault Victims: A Cross-Sectional and Longitudinal Investigation. *Journal of Trauma & Dissociation* 11, 2 (2010), 210–227.

[58] Tetyana Lokot. 2018. #IAmNotAfraidToSayIt: Stories of Sexual Violence as Everyday Political Speech on Facebook. *Information, Communication & Society* 21, 6 (2018), 802–817.

[59] Rebecca M Loya. 2015. Rape as an Economic Crime: The Impact of Sexual Violence on Survivors' Employment and Economic Well-Being. *Journal of Interpersonal Violence* 30, 16 (2015), 2793–2813.

[60] Megan K Maas, Heather L McCauley, Amy E Bonomi, and S Gisela Leija. 2018. "I was grabbed by my pussy and its #NotOkay": A Twitter Backlash Against Donald Trump's Degrading Commentary. *Violence Against Women* 24, 14 (2018), 1739–1750.

[61] Lydia Manikonda, Ghazaleh Beigi, Huan Liu, and Subbarao Kambhampati. 2018. Twitter for Sparking a Movement, Reddit for Sharing the Moment: #metoo Through the Lens of Social Media. *arXiv preprint:1803.08022* (2018).

[62] Alice E Marwick and danah boyd. 2011. I tweet honestly, I tweet passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media & Society* 13, 1 (2011), 114–133.

[63] Alice E Marwick and Danah Boyd. 2014. Networked Privacy: How Teenagers Negotiate Context in Social Media. *New Media & Society* 16, 7 (2014), 1051–1067.

[64] Mary L McHugh. 2012. Interrater Reliability: The Kappa Statistic. *Biochemia Medica* 22, 3 (2012), 276–282.

[65] Alyssa Milano. 2017. "If you've been sexually harassed or assaulted write 'me too' as a reply to this tweet.". *Tweet* (2017). https://twitter.com/alyssa_milano/status/919659438700670976

[66] Casey Newton. 2017. Twitter is Working with an Advocacy Group to Investigate the Harassment of Women. *The Verge* (2017). https://www.theverge.com/2014/11/6/7170447/twitter-is-working-with-an-advocacy-group-to-investigate-the

[67] Colleen E O'Connell and Karen Korabik. 2000. Sexual Harassment: The Relationship of Personal Vulnerability, Work Context, Perpetrator Status, and Type of Harassment to Outcomes. *Journal of Vocational Behavior* 56, 3 (2000), 299–329.

[68] Tully O'Neill. 2018. 'Today I Speak': Exploring How Victim-Survivors Use Reddit. *International Journal for Crime, Justice and Social Democracy* 7, 1 (2018), 44–59.

[69] Lindsay M Orchowski, Amy S Untied, and Christine A Gidycz. 2013. Social Reactions to Disclosure of Sexual Victimization and Adjustment Among Survivors of Sexual Assault. *Journal of interpersonal violence* 28, 10 (2013), 2005–2023.

[70] Zizi Papacharissi. 2016. Affective Publics and Structures of Storytelling: Sentiment, Events and Mediality. *Information, Communication & Society* 19, 3 (2016), 307–324.

[71] Zizi Papacharissi and Maria de Fatima Oliveira. 2012. Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt. *Journal of Communication* 62, 2 (2012), 266–282.

[72] Joanne Pavao, Jennifer Alvarez, Nikki Baumrind, Marta Induni, and Rachel Kimerling. 2007. Intimate Partner Violence and Housing Instability. *American Journal of Preventive Medicine* 32, 2 (2007), 143–146.

[73] Diana L Payne, Kimberly A Lonsway, and Louise F Fitzgerald. 1999. Rape Myth Acceptance: Exploration of its Structure and its Measurement using the Illinois Rape Myth Acceptance Scale. *Journal of Research in Personality* 33, 1 (1999), 27–68.

[74] Associated Press. 2017. More than 12M "Me Too" Facebook Posts, Comments, Reactions in 24 Hours. *CBS News* (2017). https://www.cbsnews.com/news/metoo-more-than-12-million-facebook-posts-comments-reactions-24-hours/

[75] Harrison Rainie and Barry Wellman. 2012. *Networked: The New Social Operating System.*

[76] Carrie Rentschler. 2015. #Safetytipsforladies: Feminist Twitter Takedowns of Victim Blaming. *Feminist Media Studies* 15, 2 (2015), 353–356.

[77] Michelle Rodino-Colocino. 2014. #YesAllWomen: Intersectional Mobilization Against Sexual Assault is Radical (Again). *Feminist Media Studies* 14, 6 (2014), 1113–1115.

[78] Michelle Rodino-Colocino. 2018. Me too, #MeToo: Countering Cruelty with Empathy. *Communication and Critical/Cultural Studies* 15, 1 (2018), 96–100.

[79] Zick Rubin. 1975. Disclosing Oneself to a Stranger: Reciprocity and its Limits. *Journal of Experimental Social Psychology* 11, 3 (1975), 233–260.

[80] Mary P. Ryan. 1992. Gender and Public Access: Women's Politics in Nineteenth Century America. In *Habermas and the Public Sphere*, Craig J Calhoun (Ed.). MIT Press, Chapter 11, 259–288.

[81] Chiara Sabina and Lavina Y Ho. 2014. Campus and College Victim Responses to Sexual Assault and Dating Violence: Disclosure, Service Utilization, and Service Provision. *Trauma, Violence, & Abuse* 15, 3 (2014), 201–226.
[82] Eva Sharma and Munmun De Choudhury. 2018. Mental Health Support and its Relationship to Linguistic Accommodation in Online Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 641.
[83] Sharon G Smith, Kathleen C Basile, Leah K Gilbert, Melissa T Merrick, Nimesh Patel, Margie Walling, and Anurag Jain. 2017. National Intimate Partner and Sexual Violence Survey (NISVS): 2010-2012 State Report. (2017).
[84] Catherine R Squires. 2002. Rethinking the Black Public Sphere: An Alternative Vocabulary for Multiple Public Spheres. *Communication Theory* 12, 4 (2002), 446–468.
[85] Shari J Stenberg. 2018. "Tweet Me Your First Assaults": Writing Shame and the Rhetorical Work of #NotOkay. *Rhetoric Society Quarterly* 48, 2 (2018), 119–138.
[86] Rebecca L Stotzer. 2009. Violence against transgender people: A review of United States data. *Aggression and Violent Behavior* 14, 3 (2009), 170–179.
[87] Patricia Tjaden and Nancy Thoennes. 1998. Prevalence, Incidence, and Consequences of Violence against Women: Findings from the National Violence Against Women Survey. Research in Brief. (1998).
[88] Emma Turley and Jenny Fisher. 2018. Tweeting Back While Shouting Back: Social Media and Feminist Activism. *Feminism & Psychology* 28, 1 (2018), 128–132.
[89] Rachel M Venema. 2016. Making Judgments: How Blame Mediates the Influence of Rape Myth Acceptance in Police Response to Sexual Assault. *Journal of Interpersonal Violence* (2016).
[90] Katrina A Vickerman and Gayla Margolin. 2009. Rape Treatment Outcome Research: Empirical Findings and State of the Literature. *Clinical Psychology Review* 29, 5 (2009), 431–448.
[91] Jessica Vitak. 2012. The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting & Electronic Media* 56, 4 (2012), 451–470.
[92] Yi-Chia Wang, Moira Burke, and Robert Kraut. 2016. Modeling Self-Disclosure in Social Networking Sites. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 74–85.
[93] Wikipedia. 2019. Pregnancy and Infant Loss Remembrance Day. *Wikipedia* (2019). https://en.wikipedia.org/wiki/Pregnancy_and_Infant_Loss_Remembrance_Day
[94] Sherri Williams. 2015. Digital Defense: Black Feminists Resist Violence with Hashtag Activism. *Feminist Media Studies* 15, 2 (2015), 341–344.
[95] Ying Xiong, Moonhee Cho, and Brandon Boatwright. 2019. Hashtag Activism and Message Frames Among Social Movement Organizations: Semantic Network Analysis and Thematic Analysis of Twitter During the #MeToo Movement. *Public Relations Review* 45, 1 (2019), 10–23.

A CONTENT ANALYSIS CODEBOOK

A.1 Conceptual Analysis Guidelines

Analysis will occur at the tweet level (the code is applied to the entire tweet). Apply the code to each tweet that it appears in, including if it only applies to part of the tweet. Multiple codes can be applied to one tweet.

- For tweets with pictures:
  - Interpret the tweet with the picture included.
  - Use text within picture as if it is part of the tweet.
- For tweets with links:
  - Do not follow the link.
  - Get context from the link itself (the URL) and the preview of the link (if available. See below for example).
- For tweets with quoted tweets:
  - Code for the original text only
  - Get context from the quoted tweet
- For tweets that are replies:
  - If context is needed, take into account only tweets one up or down in the thread.
- **DO NOT** infer gender of the author from the username.

A.2 Disclosures

A disclosure is a tweet that contains a discussion of harassment or assault that is explicitly about the user. This includes tweets that only contain "#MeToo".

Code	Definition	Examples
#MeToo	A tweet that only uses the hashtag #MeToo or only mentions other users and uses #MeToo ("@use1 @user2 #metoo") **DO NOT** code if *Personal Narrative* was coded	"#MeToo"
Personal Narrative or Details with Respect to Disclosure	Tweet contains discussion of harassment or assault that is explicitly about the user (PEW). Any amount of details are provided with the disclosure **DO NOT** code if • #MeToo or was coded • General comment about feminist • The author is not disclosing their own experience, but the experience of someone else	#MeToo Age 15. Never reported." "#MeToo - It happened. It was horrible." "#MeToo - by my first boyfriend." **DO NOT** Code: "#MeToo is so important. I just saw a teenage girl get harassed at the bus stop."
Disclosing Multiple Incidents	Disclosure of multiple incidents that happened to the author	"First assaulted at age 5 by an adult male. 2nd time waitressing as a teen. #MeToo" "#MeToo - too many times to count" **DO NOT** code: See examples from Personal Narrative

A.3 Pervasiveness of Sexual Violence and Inclusivity

Code	Definition	Examples
Prevalence and Pervasiveness of Sexual Violence	Discussion or reflection on: • The pervasiveness of sexual violence in the lives of women, either in general or specifically among one's friends and family • The frequency of the #metoo hashtag on social media feeds, either in general or specifically among one's friends and family	"About 30 people I know have posted #MeToo in just the last 24h. What kind of world do we live in?" "Pretty sure we are all catcalled at least once a week. It does not stop, even if running or walking #MeToo"
Next Steps in Addressing Rape Culture	Discussion of actions that individuals or groups can take to address rape culture.	"#MeToo was last week. Here are some ways to help combat sexual abuse today [hyperlink]" "We need to teach out kids about consent to end the cycle of #metoo"
Non-Women / Male Survivors	Discussion of survivors who are explicitly not women. The default of the #metoo discussion is to focus on survivors who are women. This code is applied to tweets that bring attention to men or non-women as survivors of sexual assault / harassment.	"No matter your gender, if you have experienced sexual assault I am here for you. #MeToo" "Even as a buff bodybuilding cop, I have a #MeToo from a Sargent. It can happen to anyone"

A.4 Pushback

Code	Definition	Examples
Disbelieving or Maligning #MeToo Tweets and Movement	Disbelief of people who are disclosing or of the importance of the #MeToo movement. Maligning of the #MeToo movement and those who support it.	"#MeToo. Another irrelevant brainwashing by the typical liberal mafia" "People dogpiling on the pseudo accusations just to feel relevant and part of 'social protest' SEE SOMETHING SAY SOMETHING #MeToo"

A.5 Entertainment and Politics

Code	Definition	Examples
Hollywood Figures	Discussion of public figures from Hollywood (e.g., actors, actresses,producers, etc.) with respect to the #metoo movement, sexual harassment, sexual assault. Also includes general discussions of Hollywood with respect to those topics [6].	"@TheEllenShow #MeToo" "Sicked by Quentin Tarantino's acceptance of Harvey Weinstein's abuse of women #MeToo #HeKnew"
Political Figures	Discussion of political or politics-adjacent figures (e.g., Al Franken, Donald Trump, Clarence Thomas) with respect to the #metoo movement, sexual harassment, sexual assault. Also includes references to political parties, the White House, or George Soros [6].	"This group of women spoke out against @realDonaldTrump for the sexual assault and rape he committed against then. Their stories matter. #MeToo" "@realDonaldTrump please look at how many women are posting #MeToo!"

B DISCLOSURE CLASSIFICATION MODEL

B.1 Feature Engineering

B.1.1 Lexical Features. Lexical features are grouped thematically. Any word or phrase under a particular feature grouping counts towards that feature. All lexical features are count variables.

Feature	Words and Phrases
Pronouns: Personal	i, me, my, mine
Pronouns: "Male"	he, him, his
Pronouns: "Female"	she, her, hers
Pronouns: Collective	we, our, ours, us
Pronouns: Second person	you, your, yours
Assault: Assault	assaulting, assaulted
Assault: Pain	pain, painful, hell
Assault: Harassment	harass, harassed, harassing, leave me alone, catcalled, catcalling, masturbate, masturbated, masturbating, stared, staring
Assault: Stalking	stalk, stalked, stalking, followed
Assault: Groping	grab, grabbing, grabbed, touch, touched, grope, groped, groping
Assault: Refusal of consent	said no
Assault: Yelling	yelled, yelling, shouted, shouting, catcalled, catcalling, called, calling
Feelings: Fear	afraid, fear, scary, scared, screamed
Feelings: Guilt	fault, guilt, guilty
Feelings: Shame	shame, embarrassed, ashamed, dehumanizing, bring myself
Details: Workplace	colleague, workplace, job, jobs, coworker, coworkers, boss
Details: Drugs	drink, drinking, drunk, drug, drugs, drugged
Details: Vehicle	car, truck
Details: Clothing	skirt, jeans, pants, shirt, shirts, shorts, bra
Details: Body parts	leg, legs, breast, breasts, boob, boobs, thigh, thighs, butt, ass, hair, vagina, pussy
Details: Control	powerless, helpless, trusted
Persons: Men	men
Persons: Women	women
Persons: Male figure	boy, boyfriend, bf, dad, uncle
Persons: Children	girls, boys, children
Persons: Female family	mom, moms, mother, mothers, sister, sisters, daughter, daughters
Persons: Transgender	non-binary, nb, genderqueer, trans, genderqueer
Experience: Experience	experience, experiences, happened, can't remember, detail, details
Experience: Story	story, stories
Experience: Number of times	multiple, instance, instances, times, can count, can remember
Experience: Reporting	not ready, did not report, didn't report, haven't told, hadn't told, have not told, had not told, came forward, spoke up
School: Year	freshman, freshmen, sophomore, junior, senior
School: Institution	college, university, campus, dorm, high school, middle school, elementary school
Movement: Trigger warning	trigger warning, tw:, content warning, cw:
Movement: Feminism	gender, problem, world, society, culture, community, ally, allies, consent, consensual
Movement: Right-wing	triggered, snowflake, liberal
Movement: Accusations	accusation, accusing, accused, accuser, false, fake, disbelieve, disbelief
Movement: Solidarity	powerful, strong, strength, brave, courage, courageous, love, believe you, i'm sorry, sharing, thank you, stand with
Figures: Political	clinton, hillary, congress, trump, @realdonaldtrump
Figures: Entertainment	weinstein, hollywood
Regex: Age	`r'\bI was (\d|\d\d)\b|\bage (\d|\d\d)\b|childhood\b|in \d\d\d\d'`
Regex: Name mention	`r'\b[A-Z]\w* [A-Z]\w*\b'`

B.1.2 Structural Features. Structural features are either binary or count variables, as indicated.

Feature	Description
Non-descriptive #MeToo	`r'^(#metoo)$\|^(((@\w+\| )+) #metoo)$|^(#metoo ((#\w+| )+))$`
In reply to Alyssa Milano	The tweet is in reply to or a quote retweet of Alyssa Milano's original "Me too" tweet
Number of commas	The total number of commas in the tweet
Number of colons	The total number of colons in the tweet

B.1.3 Content Features. Content Features are count variables.

Feature	Description
Number of photos	Total number of photos attached to the tweet
Number of hashtags	Total number of hashtags in the tweet
Number of disclosure hashtags	Total number of the following hashtags related to disclosure: #breakingsilence, #yesallwomen, #notokay, #iamworthit
Number of external links	Total number of links pointing outside of Twitter

B.2 Misclassifications Audit

Given our eye towards future work detailing the potential inequities that #MeToo may have replicated, we reviewed all misclassifications (n=331) to determine what disclosures our model may systematically misidentify. We did not observe trends in the false negatives or false positives that would suggest a problematic bias from our model. False negatives (tweets that should have been classified as disclosures which were not) occurred more frequently among indirect disclosures (e.g., "This #MeToo trend just hit my feed. Adding myself to it." or "Always unsure if you can look up and smile at a stranger w/o a response or checked out after helping someone #MeToo"). False positives (tweets that were classified as disclosures which should not have been) occasionally occurred in non-disclosure tweets using the first person, such as tweets providing emotional support (e.g., "To everyone posting #MeToo – I believe you, I support you, and I promise to speak out against sexual violence whenever I see it.")