College of Computing and Digital Media
Dissertations

College of Computing and Digital Media

Spring 5-12-2021

Quantitative intersectional data (QUINTA): a #metoo case study

Alicia E. Boyd
DePaul University, ABOYD20@depaul.edu

Follow this and additional works at: https://via.library.depaul.edu/cdm_etd

Part of the Computer Sciences Commons, and the Data Science Commons

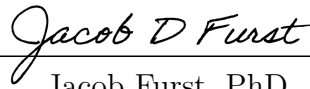QUANTITATIVE INTERSECTIONAL DATA (QUINTA): A #METOO CASE
STUDY

BY

ALICIA ELIZABETH BOYD

A DISSERTATION SUBMITTED TO THE SCHOOL OF COMPUTING,

COLLEGE OF COMPUTING AND DIGITAL MEDIA OF

DEPAUL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

COMPUTER SCIENCE

DEPAUL UNIVERSITY

CHICAGO, ILLINOIS MAY 2021

This dissertation entitled:
Quantitative Intersectional Data (QUINTA): A #metoo case study
written by Alicia E. Boyd has been approved
on Wednesday, May 12, 2021

Jacob D Furst

Jacob Furst, PhD

Brian Keegan

Brian Keegan, PhD

Sheena Erete

Sheena Erete, PhD

John Rogers

John Rogers, PhD

The final copy of this dissertation has been examined by the signatories, and we find
that both the content and the form meet acceptable presentation standards of
scholarly work in the above mentioned discipline.

Abstract

This research began as an investigation of the #metoo movement, with the initial impetus to illuminate the voices located on the margins, those who often go unheard or are never recognized. This work aimed to understand the intersectional aspects of how these hashtag variations of the hashtag #metoo (i.e. #metoomosque, #churchtoo, #metoodiable, #metooqueer, #metoochina, etc) reveal the inequities of the #metoo movement on Twitter. The proliferation of these hashtag variations has often been ignored by scholars, and therefore absorbed into the larger #metoo movement conversation on Twitter. Therefore, the term 'hashtag derivative' was created to describe the variation on the theme of its original hashtag, strongly reflecting its composition.

Moreover, a critical theory such as Intersectionality is well-equipped to explore how overlapping identities encounter structure social reality relationship to power. Amid a pandemic and racial unrest, the true capabilities of Intersectionality to describe inequities and injustices beyond the singular social position of race and gender are not widely understood. Data science, is not absolved of its role in inequities and injustices merely by dint of being a quantitative field that claims to "objectivity". Social scientists have illuminated the racism, sexism, ableism, transphobia, homophobia, prejudice, bigotry, and bias embedded in data science's technology, tools, and algorithms. This has, direct and indirectly, grave consequences on an entire community as a whole as well as marginalized communities.

The application of Intersectionality into a quantitative field can provide researchers a formal structure to be more conscientious about how to critique, develop, and design their data science processes, while also reckoning with their own positioning in relationship to the data. In this way, Intersectionality is inclusive in terms of data equity yet adds an additional layer of accountability to the researcher. This research leads to the three critical contributions of this work: (1) creating a more concise terminology to describe the phenomenon of hashtag variation, known as hashtag derivatives, (2) defining the historical context of Intersectionality and building a formal case for this to be properly contextualized in the Computer Science field (in particular Data Science), and (3) developing the Quantitative Intersectional Data (QUINTA) Framework which data scientists and scholars can use to be more equitable, inclusive and accountable for their role in the data science process.

Dedication

First and foremost, I dedicate this to the Lord. With God everything is possible. To my family–past and present, thank you for paving a way for me. I am your wildest dreams. Special dedication to my grandmothers, Evail and Frances.

Acknowledgments

"For I know the plans I have for you, declares the Lord, plans for welfare and not for evil, to give you a future and a hope." -Jeremiah 29:11, *English Standard Version*

Thank you and I love you, Auntie Gwen.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The emergence of this work began with the investigation of online social movements, particularly the #metoo movement. The #metoo movement created awareness of sexual assault and violence. The hashtag variations on the theme of the original hashtag, #metoo, further illuminated social media's impacts on communities outside of Hollywood. These hashtag derivatives highlighted the voices located on the margins who often go unheard and ever recognized. Furthermore, these proliferations were absorbed into the larger conversation of sexual assault and violence; however, as scholars and researchers have argued, #metoo is a reflection of women of color's experience inside the feminist movements in the United States [228]. I would further contend the exclusion of other marginalized communities not limited by gender and race but by social identities such as but not limited to, religion, sexual orientation, able-bodiedness, *etc.*

Intersectionality is a critical framework that is well-equipped to explore how overlapping identities structure social reality in relationship to power. Using this critical framework was equipped to discuss the inequities of the #metoo movement on Twitter though the investigation eventually led to incorporating Intersectional-

ity into data science. Leading to the question, what is the role of Intersectionality within data science? The crux of this dissertation begins at the confluence of two concepts–Intersectionality and data science– both epistemologically different yet used synergistically to analyze the #metoo movement. The application of Intersectionality into data science can help researchers be more critically conscientious about how to critique, develop and design the data science process while also calling to their own positioning in relationship to the data. In this way, Intersectionality is inclusive in terms of data equity yet adds a layer of accountability to the researcher.

Therefore, this dissertation posits that the intersectional framework can be used in two ways: (1) to allow researchers to account for the complexities behind the data we seek to analyze and interpret, and (2) to allow researchers to be reflexive in their positioning in relationship to the data. Adopting an intersectional approach in data science research allows researchers to acknowledge this integral relationship between social positionings and power structures. Else-Quest and Hyde elaborate on this by stating the "existence of multiple social categories intersecting and are constructed by and within power relations and empowering individuals and groups to transcend the constraints imposed upon them by those categories and linked inequities." [92, p. 319] Researchers have advocated for the intersectional approach to be used in social science fields such as psychology, because it broadens the research questions to examine heterogeneity within social categories and explore how power and inequality are constructed. This heterogeneity exists in our data where social categories and power and inequality among these social constructions exist. However, data does not represent marginalized communities [113, 252]; demographic data, and sampling can fall short of reflecting a population [92, 282].

One of the dissertation's critical contributions is the development of Quantitative Intersectional Data (QUINTA). QUINTA is the result of coupling Intersection-

ality and data science linked with the concept of reflexivity. Collins and Bilge [70] and later Collins [69] further discuss the importance of being self-reflexive when using Intersectionality as a form of critical inquiry and praxis. The application of reflexivity in data science allows researchers to call attention to their own practices in the context of Intersectionality. I challenge further that reflexivity can be used as a coupling mechanism for intersectionality and data science to co-exist simultaneously. I adapted Cole's work [61] for QUINTA in order to interrogate the data science process.

As researchers and data scientists, it is imperative that we select frameworks, methods, and tools that account for (and question) the complexity of the data and allow us to be conscientious during each step. Although collected data is often so complex and vast that this conscientiousness is not an easy task, a critical theory that guides our decisions, designs, processes, and interpretations allows us as researchers to be reflexive[1] in the analysis. Researchers [27, 28, 129, 136, 181] call attention to how technologies have embedded hegemonic ideologies that mirror the white, patriarchal, capitalistic society and perpetuate oppression's dominating structures. Knowing this, and that numerous researchers and commentators have been actively calling attention to these matters, there is a need for a critical theory such as Intersectionality to guide us through the research process.

This work will not be a savior to the ills of the world, nor entirely bridge the long-standing gap between Intersectionality and data science. However, this dissertation offers a starting point for this conversation and examines how to navigate the tenuous relationship between this framework and its two parent disciplines.

---

[1]This term will be explained in Chapter 2.

1.0.1   Positionality Statement

For this dissertation, I include a positionality statement using the methodology of Standpoint Theory [141]. The purpose of this positionality statement is to discern the difference between my voice as a cis-gendered, upper-middle-class, able-bodied Black woman and the diverse perspectives of Black people across the globe. Even though Black people's life experiences are not a monolith, sadly, our collective humanity is not seen as valuable compared to white people. Within the United States of America alone, there is evidence of Black massacres, for instance, East Saint Louis in 1917 [253], Chicago in 1919 [298], Tulsa in 1921 [180], New York in 1863 [12], and Charleston Church in 2015 [90] name a few. The historical treatment of and for Black people has been embedded and imprinted into our technology [22, 129]. Scholars have articulated how the data represent and impact people [227]. However, the processes we as data scientists utilize have origins of white-eurocentric male ethos that center and reinforce the status quo [27, 28, 136, 257].

Lastly throughout this dissertation, I will capitalize the letter 'b' in the word "Black" to acknowledge the cultural identity of Black people [63]. Furthermore as Erete and colleagues have eloquently used the inclusive terminology in their work stating:

> "Black to refer to those from the African diaspora and is inclusive of people born of African descent across the diaspora, including Africa, the Caribbean, North and South America, Canada, Europe, and Asia. Racial discrimination, as evidenced in the U.S., is applicable to and impacts everyone in the African diaspora … Those who are from the African diaspora experience racial discrimination based on the color of their skin in the U.S., not their place of origin. At the same time, we acknowledge that

4

we do not speak for all Black people since Black people are not monolithic and represent diverse perspectives and life experiences." [94, p.6]

Therefore, I take a similar position in this dissertation.

1.1   Problem Statement

Using the #metoo movement as a case study, this research seeks to better center and amplify the experiences of the most marginalized survivors in order to address their erasure and its intersections with racism, gender, and sexual orientation by looking at the co-occurrence of both hashtags and hashtag derivatives in tweets. This brings us to the overarching question of this work: **How do co-occurrence of both hashtags and their derivatives articulate and highlight intersectional differences for participants in the broader #metoo Twitter conversation?**

This dissertation will explore two topics, each representing a unique gap in the literature:

1. the usage and proliferation of co-occurrence of both hashtags and their derivatives in the #metoo movement,

2. the application of intersectional network analysis to understand and analyze these interrelated positions.

1.2   Research Questions

The following five research questions, partitioned into two separate studies (Chapters 3 and 4), chart the course of exploration and inquiry of this dissertation.

1. **RQ1a:** Which categories of hashtags are frequently used during #metoo conversation?

2. **RQ1b:** What temporal hashtag patterns occurred in the first week of #metoo?

3. **RQ1c:** How do Twitter bios relate to categories and patterns of hashtags?

4. **RQ2a:** What are the characteristics of the participants using co-occurrence hashtag derivatives on Twitter?

5. **RQ2b:** How does intersectionality modify the traditional network analysis?

1.3   Dissertation Outline

Through these research questions, this dissertation situates the theoretical framework of Intersectionality within the data science discipline using the #metoo viral phenomenon as a case study for analysis.

Chapter 2 contours the landscape of relevant work at the convergence of Intersectionality, Twitter hashtags, #metoo, and data science, recounting previous research and identifying missing perspectives in current literature.

Chapter 3 addresses the literature's first missing gap, the co-occurrence and derivative hashtags that appeared in the #metoo movement.

Chapter 4 addresses the literature's second missing gap, examining how to extend the social network analysis into a more intersectional approach.

Chapter 5 introduces and discusses what Quantitative Intersectional Data (QUINTA) means within the data science world. I suggest recommendations on navigating the data science process and introduce a table to help guide research scientists.

Chapter 6 concludes by summarizing contributions, personal reflections through-
out the research journey, limitations of this work, and future research directions that
I intend to pursue.

Chapter 2

Related Work

The core of this chapter is to establish an understanding of Intersectionality as the theoretical framework for this work by reviewing its rich lineage. The application of Intersectionality here is motivated by the study of #metoo, and that movement was popularized by a hashtag, so we must first start by exploring the history and current state of research on social media hashtags and #metoo. After discussing Intersectionality, we will tie it into these topics to create the foundation for the new research being presented.

2.1   Twitter Hashtags

Twitter provides users a way to connect with other users via tweets; a tweet, as Kwak *et al.* [184] noted, can be a simple statement generated by a user or a reply to another tweet. Twitter users can tweet using 280-character messages (formerly 140) which contain text, hyperlinks and emojis to interact with one another or convey a message. Twitter is particularly known for the immediacy and brevity with which information can be shared in a single tweet or string of tweets. There has been a

great depth of research into the diffusion of information on Twitter [270, 272, 301]. Researchers have found that people use Twitter to communicate and broadcast information about news [158, 232], elections [283], natural hazards and human-induced disasters [168, 226], television shows [49, 205], and social media campaigns [169]. One significant mechanism for the diffusion of information on Twitter and other social media platforms is via hashtags.

Hashtags, initially introduced on Twitter, can facilitate subsequent search and filtering, classification, and clustering of messages. As Tsur and Rappoport [281] define them, a hashtag is a case-insensitive string of non-whitespace (i.e., no punctuation or spaces) characters preceded by the hash (#) character [159, 236, 266, 281], which come in various lexical forms such as a single word (#patience), a combination of words (#metoo), or a combination of terms and abbreviations (#metook12). Acronyms and abbreviations are often used due to tweet length constraints [100, 261]. Hashtags can be used as topical markers for conversations, to denote events [71], and to express ideas and feelings or signal community membership [303]. When rendered in a finalized tweet, hashtags create links reminiscent of a topically-focused variation of the classic social media feed page, by organizing tweets together on a single infinite scroll page featuring recent tweets that contain the hashtag. This allows people to connect to those who would not otherwise appear in their personal network. In this manner, hashtagging has become an excellent way to research and categorize posts and microblogs on social media.

Hashtags have been used for many more reasons than those already mentioned [261], including: advertising, indicating a specific object described in the post or the situation, describing feelings and thoughts—for instance, #yolo and #tbt[1]—or mentioning personal words (or groups of characters/symbols) to be understood only

---

[1] "You Only Live Once" and "Throw Back Thursday"

by the user or its intended audience. Researchers have also documented the use of "personal hashtags" to define hashtags that have latent meanings for specific users [261]. One example would be #leon to represent a user's cat—named, of course, Leon. In addition, hashtags are employed to join a public discussion [159], categorize messages, or build communities around specific topics of interest [291, 303].

2.1.1 Hashtags and Memes

Hashtags are conversations starters and spread and diffuse information on Twitter. Huang *et al.* [159] explored how hashtags were used to start conversations and spur hashtag adoption. They noted a phenomenon they termed "micro-meme" to describe how users create variants of popular hashtags to spawn multiple conversations, desynchronized from, but nevertheless related to the original. They focused their dataset on a period of time when hashtags became widely adopted on Twitter and compared the tagging behavior between Twitter and another platform —del.icio.us[2]—to understand the adoption of hashtags. The authors termed the behavior of Twitter users using their hashtags in a conversational aspect, as "asynchronous massively-multi-person" behavior. They noted that a user's motivation to participate in this micro-meme phenomenon, "is to [have] their tweets displayed in the filtered stream of messages with that tag attach." [159, p. 176]

Because of the strong phenomenological associations between hashtags and memes, researchers have often used the two terms interchangeably [45, 281], which may not have been a critical error at the time, but poses a nuisance to the field today. Likewise, other researchers [124, 290] have used different terminology to describe the memefication of hashtags. Gonzalez-Bailon, who also studied the Occupy movement, describes Occupy hashtags as "labels self-assigned by users to identify streams of

---

[2]http://del.icio.us

information that are relevant to particular issues" as variations [124, p. 217]. Memes are the spread of thoughts and ideas which go unchanged throughout social media. When focusing our attention on Twitter, hashtags carry similar properties as memes. Hashtags promote ideas where users can use the hashtag in their tweets. Even though memes and hashtags are very similar, there is a uniqueness in their implementation. Kidd [173] stresses that, "a hashtag in and of itself is not a meme," providing an example in which a company creates a hashtag and controls how it is employed as part of a marketing or communications promotion, but nevertheless will fail to become a meme if users do not adopt the hashtag. Kidd identifies the transition from hashtag to meme (memefication) as when "users apply the hashtag to circumstances that differ from the original context." [173, p. 162] Additionally, "further memeification happens when those actors [or users] shift the hashtag in some way to make it unique to them, even as it still indicates a connection to the original" [173, p. 162-163].

2.1.2   Hashtag: Co-Occurrence and Proliferation Research

The body of literature on hashtag co-occurrence has been quite broad and largely focuses on the same topics as research on singular hashtag occurrences. Researchers have investigated the co-occurrence of hashtags in the realms of topic modeling sentiment [284, 290, 291], networks [96, 290], prediction [296], popularity [230], virality [290], and social movements [115]. The main subject areas within this body of literature in relationship to hashtag co-occurrence is based on classification algorithms to understand semantic and temporal meaning.

Among the earliest research on tagging co-occurrence as a generic topic was done by Begelman *et al.* [20]. They proposed an algorithm based on counting the number of tag co-occurrences within the same page, labeling any two given tags

as semantically related if their co-occurrences were greater than a cutoff threshold. Although this research was not on the Twitter platform, it still lays the foundation for tag and clustering analysis on Twitter and a number of other platforms such as Flickr[3], del.icio.us and technorati[4]. Wartena *et al.* [293] proposed another approach to calculate the similarity between tag co-occurrence distribution and the user profile, whereby tags with high similarity would be recommended to the user. Belem *et al.* [21] proposed new heuristic methods based on textual features. All these approaches are based on two assumptions: tags are assigned to resources beforehand and most resources have two or more tags. This latter assumption does not hold on Twitter, however, as most tweets only contain one hashtag or none at all.

Defining the co-occurrence incidence in a tweet has been broadly established in the literature, but there are two competing definitions due to the fact that each researcher has their own definition based on the context of their own subfields. For instance, Antenucci *et al.* [6] and Poschko [235] both defined co-occurrence as when multiple hashtags occur in the same tweet. Whereas Li *et al.* [190] defined co-occurrence as a keyword and a hashtag appearing in the same tweet.

Researchers have studied co-occurrence phenomenon in relationship to various social movements on Twitter such as Brazil protest [246], Occupy Wall Street [290], and JeNeSuisPasCharlie [115]. Each of these studies constructed a co-occurrence network to study how to hashtags depicted different meanings and functions for their respectively online social movements. These co-occurrence networks revealed the personalization of messages narrating the protests, mainly focusing on the mobilization and the narration of the events in each place where they occurred [246]. Within these co-occurrent networks, researchers found that strategic combinations of viral

---

[3]https://www.flickr.com
[4]https://technorati.com

hashtags by users would mobilize public figures and other influential actors toward the movement via usage patterns of hashtags which aided in visibility [115, 290]. It is encouraging to see the discussion of co-occurrence of hashtags appearing in events and social movements such as Brazilian protest, Occupy Wall Street and Charlie Hebdo Shooting, respectively along with more consistent terminology. However the focus is more on prediction rather than understanding the proliferation these hashtags signalling social intersections within these larger movements.

We begin to see more discussion in the literature about the proliferation of co-occurrence hashtags in their classification as memes in the hashtag activism spaces, like the Occupy Movement. The conversation about these hashtag co-occurrences focuses on how connections are created from community to community on Twitter, as opposed to the post-to-post and user-to-user connections that most research is concerned with. One way in which community-to-community connections are created is by rhyzomaticity in hashtags. Kidd [173] discusses in his book, *"Social media freaks: Digital identity in the network society"*, the Occupy movement but uses the "memeification" terminology to describe hashtags (#OccupyNYC and #Occupyfarm). He describes that hashtags can take on a "rhizomatic" structure in social movement conversations. The idea of such a structure originates from biology, where rhizomes are underground horizontal offshoots of plants' stems within their roots, which form a decentralized network that allows nearby plants above ground to survive even when they've been cut off from the original plant. Kidd's work suggests that on Twitter, hashtags are the building blocks of rhizomatic conversations, where the hashtag may create several local conversations which can function independently of the original or global conversation. He cites an example of these being the Occupy Wall Street Movement being naturally rhizomatic, given its decentralized, leaderless structure and use of hashtags to organize local chapters while also uniting the

broader movement. Gonzalez-Bailon [123] also found evidence of this same pattern of hashtag proliferation and "variation of hashtags" in her work on Occupy location movements (#OccupyNYC and #OccupyLondon). One of the major distinctions between Kidd [173] and Gonzalez-Bailon [123] however, is that they do not specifically conceive of their hashtag variations as "derivatives". Nor do they clarify whether these hashtag variations are present in the tweets on of other hashtags with the original #occupywallstreet hashtag.

2.1.3 Hashtag Social Networks

Community building and formation on Twitter has been studied [133, 134], and one way to build these communities has been through the usage of hashtags. Hashtags can be used to start communities and groups through health-focused social support [75], fostering awareness [121], call to action [109], adoption of metrics [303] and advocacy. As stated earlier Jackson *et al.* [164] found certain hashtags fostered community support and advocating. While there have been discussions of hashtags being created for particular purposes to carve out spaces for particular groups, like hashtag feminism[5] [197].

Researchers can use social networks as a perspective[6] for exploring hashtags and community structures [278]. Social network analysis (SNA) is a field of data analytic involving the usage of concepts of networks and graph theory in order to understand social structures [294]. SNA techniques can also be applied to networks outside of the societal realm. Hashtags play a strategic role in mobilizing Twitter users, which created these networked connections within social movements [195, 263,

---

[5]Hashtags used to create spaces for women who can communicate about issues they encounter.
[6]This social network perspective is not the same as social networking. Social networking is the use of Internet-based social media sites to stay connected with friends, family, colleagues, customers, or clients.

306]. Wang *et al.* argued that though hashtags can function as mobilizers, they can stimulate messages within the network. In their work, they created a hashtag co-occurrence network to identify the flow of information during the viral moment in Occupy Wall Street movement. They found that "different types of hashtags often help attract attention from clustered Twitter Users within their network." [290, p. 861] Furthermore, the researchers found people in the network were more likely to use specific hashtags to gain public attention toward a particular collective cause. While their study used a hashtag co-occurrence network, their definition of hashtag co-occurrence is based on the incidence of any pairing of hashtags together in a tweet, not a particular hashtag variation. Also, the construction of their hashtag typology did include categorical elements based on the Occupy protests. However, I note that these categories (*e.g.,* class, location, identity) and hashtags presented a missed opportunity to use Intersectionality.

Researchers also create these social network models to predict virality and examine the structure of communities with the networks. There are two components missing in the social network modeling literature (1) the use of intersectionality as a guiding framework, and (2) the development and implementation of intersectional methods to create these network models. Social Network Analysis and Intersectionality were employed in Geerlings' work [112] by examining the correspondence of Rosey E. Pool, Dutch anthologist, during the period 1966-1971. We should note that Geerlings makes the unusual move of using collective identity theory in tandem with Intersectionality, as opposed to independently. Nor does she apply intersectional methodologically in the creation of network analysis. However, Geerlings does use intersectionality (again, and collective identity) to guide the interpretation of the social network analysis.

2.2   #metoo Movement

In this section, the origins of 'me too' will be discussed, followed by the impacts of #metoo on various groups of people. This section will review previous work completed surrounding the usage of the #metoo movement.

2.2.1   History: Analog to Digital

In 2006, Tarana Burke started to use the phrase 'me too' to discuss the sexual violence she encountered in her work with young women survivors of sexual assault [186]. Burke first got the idea for the phrase when she encountered a 13 year-old girl who disclosed that her mother's boyfriend had sexually assaulted her. The young girl's revelation left Burke speechless and horrified, and she referred the young girl to a counselor. In response to this incident, Burke reflected:

> *"And as much as I love children, as much as I cared about that child, I could not find the courage that she had found. I could not muster the energy to tell her that I understood, that I connected, that I could feel her pain. . . I couldn't even bring myself to whisper. . . me too."* (original emphasis) [47]

Burke started the organization "Just Be Inc.", based on this 'me too' idea as its crux, and used it to promote healing via "empowerment through empathy" for the young survivors and fostering community and connection among them [47].

In October 2017, Hollywood producer Harvey Weinstein's disturbing history of sexual misconduct resurfaced to be reckoned with by a society growing more and more aware of sexual violence, and finally ready to apply the mass public scrutiny that he had avoided in his previous close calls. On Sunday, October 15, 2017, actress

Alyssa Milano tweeted (see Figure 2.1):

Figure 2.1: Alyssa Milano's 'me too' tweet.

In its first day on Twitter, there were more than 70 thousand replies to Milano's first tweet [310]. The 'me too' phrase transformed into a global hashtag literally overnight. Since then, #metoo has resonated with those who have experienced such assaults and harassment; millions of users shared their narratives on various social media platforms, including Twitter, where the actress encouraged others to do so. The hashtag was reportedly used in at least 85 countries [98]. By her own account, Milano's impetus for writing this tweet came from her own experience of sexual assault, and encouragement from an unknown friend, as well as the prior week's uproar in Hollywood over the new Weinstein allegations. As the hashtag ignited virally and caught popular attention, journalists and others began to credit Milano with the creation of this 'me too' conversation.

As these stories were shared, journalist Britini Danielle was one of the first several people to point out that activist Tarana Burke "began the crusade 10 years ago

particularly for women of color [146]." Abby Ohlheiser, a writer from *The Washington Post* describes Burke's reaction as #metoo exploded:

> *"Burke saw the hope and empathy that she knew those words could inspire the women who chose to tweet. But her experience told her that the viral hashtag could, despite its best intentions, undo much of what she's worked to build, at a moment when there was so much more to be done."* [224]

In her own account, Burke's work aimed to unify those who have been victimized by sexual violence. Starting as a grassroots movement, she began 'me too' to aid sexual assault survivors in underprivileged communities "where [rape crisis centers and sexual assault workers] were not going." [146] In the days after Milano's tweet went viral, Burke later stated that:

> "It ['me too'] wasn't built to be a viral campaign or a hashtag that is here today and forgotten tomorrow. It was a catchphrase to be used from survivor to survivor to let folks know that they were not alone and that a movement for radical healing was happening and possible." [146].

As stated earlier, the motto of Burke's 'me too' campaign was "empowerment through empathy." Recalling what she did as she began to tweet, Burke thought, "I had to ring the alarm. One, before my work is erased, and two because if I can support people, I have to do that." Burke tweeted:

> "It made my heart swell to see women using this idea–one that we call 'empowerment through empathy'," she tweeted, "to not only show the world how widespread and pervasive sexual violence is, but also to let other survivors know they are not alone. #metoo ¨ [224]

Followed by:

> "It's beyond a hashtag. It's the start of a larger conversation and a move-
> ment for radical community healing. Join us. #metoo" [224]

The allure of the #metoo was, and still is, that it invokes confession as conver-
sation and shifts the cultural dynamics across both analog and digital spaces. Zahara
Hill of *Ebony* described the hashtag as a "bold and poignant demonstration of solidar-
ity among those who've experienced sexual assault and/or harassment." [146] Since
#metoo hashtag went viral, several people have been convicted and publicly accused.
In February 2020, Harvey Weinstein has been convicted of and sentenced for sexual
assault in New York and is awaiting another trial in California. Robert E. Kelly,
better known as R. Kelly, has been convicted and sentenced as well. Bill Cosby was
found guilty in April 2018 and sentenced in September 2018. Kevin Spacey, Louis CK,
Matt Lauer, Tavis Smiley, Al Franken, Roy Moore are a few of the people publically
accused since viral hashtag. An expanded understanding of survivors' psychology has
led more states to extend their statutes of limitations in sexual abuse cases, allowing
survivors more time to report. People have begun to have more conversations about
sexual assault and violence and their impacts on communities. Progress has taken
place; and yet, echoing Burke's sentiment from earlier, there remains much work to
be done within this space.

2.2.2 Co-option

While the appearance of the #metoo movement unified those who have ex-
perienced sexual assault and violence, there is a juxtaposition in the intersectional
positioning of its participants. Specifically, it is that of a white woman populariz-
ing 'me too' digitally when the analog 'me too' movement was in truth created by

a Black woman a decade prior. Discussion of co-option began when people discovered that the 'me too' phrase had a foundation rooted in Burke's work. This new co-option narrative questioned the solidarity of white women and their commitment to the inclusion of Black women in this viral movement. There is a long history of such tensions [66, 72, 189, 280] between Black and white women, both in offline spaces, and digitally as on Twitter. The 2017 women's march, for instance, was staged at the nation's capital in response to Trump's election tackled questions of diversity, inclusion on where to include racial minorities, and transgender women, LGBTQ+ people so they could discuss the issues that impact their communities [132].

In the weeks before #metoo went viral, actress Rose McGowan's allegations against Weinstein made their way to social media. Twitter banned McGowan for violating Terms of Service when her tweet included a personal phone number of a screen-capped email. She took to Instagram and posted, "TWITTER HAS SUS-PENDED ME. THERE ARE POWERFUL FORCES AT WORK. BE MY VOICE #ROSEARMY #whywomendontreport" (original emphasis) [250]. When Twitter banned her, white feminists showed solidarity by leaving the social media platform. Black women, however, were hesitant to show support for McGowan. Activist Ashley C. Ford explained her reservations:

> "Where was the boycott for ESPN sports journalist Jemele Hill when her
> employer suspended her from her job citing a vague social media policy?
> Where was the boycott when actress-comedian Leslie Jones was harassed
> by trolls to the point of deleting her account for months?" [105]

Black women's hesitancy stems from the lack of outrage by white women when Black women are "put in vulnerable positions by rich white men" [146]. Furthermore, Black women are not exclusively placed in these vulnerable positions by white and

Black men, but also by the same white women who claim the mantle of "feminists" as well. For instance, in 2013, writer and activist Mikki Kendall created the hashtag #SolidarityIsForWhiteWomen in response to the inaction of prominent white feminist bloggers who failed to acknowledge the racist and sexist behavior of a white male contributor who identified as a "male feminist" while viciously attacking women of color [83].

Scholar Jessie Daniels discusses white feminism on the internet, where she cites examples of how women of color face challenges [83]. These challenges reflect real-life experiences mirrored on social media platforms, such as Twitter, where women of color are harassed and bullied. Daniels quotes Mikki Kendall, who argues, that "Twitter changes everything, which forces people to hear women of color because Twitter provides them with their own microphones." [83, p. 56] Yet as Daniels points out, while these microphones provide amplification of their voices, there is still a white supremacy environment present. Diehl [87] discussed the reproduction of colonial ideologies on Instagram and Twitter when she investigated violence against indigenous women and found they were missing from #metoo conversation. Diehl [87] points out the contrast between social media being a tool for activists, yet marginalized groups are left out of the conversation and have the highest rates of sexual violence. She concludes, "Overall social media lies heavily on white women when talking about sexual violence compared to other groups of women." [87, p. ii]

It should be noted that this labeling of Milano's actions as a co-option does not imply malice on her part. When Milano realized[7] [116] Burke had created the 'me too' movement, she gave her credit. It is more than likely that her intentions and experiences with sexual assault are valid, and that she genuinely intended to call attention to the environment in the entertainment industry. Nevertheless, there were conse-

---

[7]Milano realized 3 days later and retroactively credited Burke [116].

21

quences of the co-option which lead to the characterization of the movement being a movement for upper class white white-nationally and globally a movement described for western white women [122]. These characterizations leads people to believe hashtag campaign is only for a select people which do not include them. Although those subjected to sexual assault and violence are not confined to this particular group of people, women of color and men, transgender, genderqueer/gender non-conforming individuals, undocumented immigrants and children can, unfortunately, experience this. An example of this can be seen four months after #metoo went viral, user NicoleCCarlton tweeted [114, p.51-52]:

"I do not feel like I have a place in the #metoo movement. Black women have been forgotten again. When I first heard about the movement people made it seem like @Alyssa_Milano created it. I wish people would credit you more. I think it would help Black women feel more a part of it."

Burke responded:

"I see you sis. The work *I* do sees you and acknowledges you. Don't get caught up in what the media says this movement is–it is about ALL survivors finding resources to heal and working to interrupt sexual violence."

NicoleCCarlton replied:

"I truly am thankful for the work you have put into the movement. I had to go out of my way to actually learn you found it, but I am happy I didn't depend on just the media to each me. Thanks for still pushing to help Brown/Black girls like me the media seems to have forgotten."

The tweet exchange between Burke and Carlton highlights several themes: (1) misconceptions of #metoo movement on social media, (2) the erasure of Black

women within the movement, (3) implicitly acknowledges a shared commonality with the user affirmation and confirmation that Burke "sees" the user, and (4) that Burke calls attention to her work.

After the tweet exchange between the two women, Burke later wrote a series of tweets:

"I don't say much online about what is going on with the #MeTooMVMT because honestly there is so much wrongheaded chatter all of the time I don't have time to chase it all down. But I'm particularly challenged by this framing in the media #MeToo

While it's true that I have been widely recognized as the 'founder' of the movement-there is virtually no mention of my leadership. Like I just discovered something 12 years ago and in 2017 it suddenly gained value.#MeTooMVTMT #MeToo

Founder is acknowledgment. But watch carefully who are called "leaders" of the movement. It's as if 25+ years of on the ground movement building work is not enough or maybe spending most of that time being invested in the lives Black and brown Girls isn't enough. #MeTooMVMT

I'm not trying to ignite debate I meant it when I said I want to be in service and not conflict, but I have to say it is very posaible [sic] to be acknowledged and erased. #watchforthehook #MeTooMVMT

And please don't make this about @Alyssa_Milano. She has been an ally and friend from the moment she found out. She both acknowledges me as founder and has often differed to [sic] my vision and leadership. #MeTooMVMT"

Finally, Burke tweeted (Figure 2.2):

23

Figure 2.2: Burke's tweet

These exchanges—first between Burke and Carlton, and then Burke's reflection on the media's framing of the movement she had created over a decade prior—occurred a whole four months after #metoo went viral, and these trends were—and are still —happening. The movement itself has been co-opted several times over, diluting its original effectiveness. Hollywood celebrities splintered off into their #timesup movement [198], and before that, they had already made #metoo primarily focused on calling out male celebrities, instead of a movement of solidarity and empathy for survivors and the inclusion of the marginalized.

While Burke feels like her work was never "hijacked" by those in Hollywood and is intact, there is still the inclusion and diversity of the #metoo movement, which is even called into question. The #metoo movement largely focuses on the salacious scandals that distract and detract the focus from the survivors' healing and empathy. Burke states, "The women of color, trans women, queer people–our stories get pushed aside, and our pain is never prioritized. We don't talk about indigenous women. Their stories go untold." [98]

This diffusion of focus has continued: as recently as April 6, 2020, Milano used the #metoo hashtag once again, in the face of new allegations against former

Vice President and presumptive 2020 Democratic presidential nominee, Joe Biden, for whom she justified her continued support by saying (Figure 2.3):

Figure 2.3: Milano's #metoo political tweet

Critics of Milano tweeted their frustrations at the actress for using the #metoo hashtag in this perceived hypocrisy. One Twitter user tweeted (Figure 2.4):

Figure 2.4: A Twitter user's comment

Tweets such as this one by Twitter user Joementia illustrate the complexity in tracing the history and intersectionality of 'me too's' co-option. Joementia is a white man calling out a white woman for stealing a Black woman's work in order to use its moral authority to support a different white man, a politician[8]. These

---

[8]The eventual outcome was that the accusation against Biden fell apart under media scrutiny,

people are all ostensibly supporters of #metoo, and yet their politics and identities vary dramatically, especially in the context of Burke's original intent for the 'me too' movement.

2.2.3   Criticism and Quantitative Research

Researchers and scholars have made an array of critiques on the #metoo movement. These criticisms range from co-option [15, 38, 88, 116, 122, 198, 206, 228, 274], westernized focused [225, 286, 307], exclusion of marginalized communities [15, 38, 88, 117, 117, 157, 200, 206, 228, 255, 274], mainstream media highlighting salacious stories [30,59], systems of law, power, prison, politics [37,56,77,110,157,200,228,248,251], hyper-visualization and amplification of white women experiences [15,88,116,117,122, 198, 206, 228, 274]. The common thread among these criticisms was the impression the #metoo movement was only for cis-gendered, heterosexual, able-bodied, affluent white women particularly in Hollywood which excluded the stories from marginalized communities. Clark-Parsons [59] describes this as the "ideal victim" where those fit into this description will most likely be seen, heard and their "allegations of sexual violence taken seriously." [59, p. 12] However as researchers have noted, there was more than just Hollywood actresses using the #metoo hashtag. Furthermore activist have argued that #metoo "needed to better center and amplify the experiences of those most marginalized victims in order to address sexual violence and its intersections with racism, classism, heterosexism and transphobia." [59, p.12]

In the realm of data analytic research on the #metoo movement has examined a myriad of areas such as content analysis of tweets [144,202,215,259,302], public discourse [59,196,213,271], tweet classification [155], temporal dynamics [213,271], story

---

but this is irrelevant to the point being made here about the intersectional positioning of the actors represented in these tweets.

sharing [144], politics [59, 144, 247], social network analysis [163, 278, 302], tweet prediction [213, 275], engagement [192, 202, 203], storytelling [163], and reporting sexual crimes [188]. However within the body of literature there is summative amalgamation of communities under a single topic sexual violence which gives the impression that everyone is inherently impacted by sexual violence in the same manner. This proposes a dangerous impact on marginalized communities. In this section, there is an exploration in details of what the researchers did in their work to due to initial work in this space, in order to identify the gaps in their research in relationship to harms to social identities.

2.2.3.1 #metoo Co-occurrence Patterns

Manikonda *et al.* [202] performed a comparative analysis across Twitter and Reddit platforms to investigate the subtopics while using the hashtag #metoo. From their work, we can ascertain two things: (1) that the co-occurrence pattern is observed on Twitter, not on the Reddit platform, and (2) the authors did not capture any hashtag variation of the #metoo in their corpus nor co-occurrence analysis. There was an vague explanation how the authors sampled their Twitter data other than stating they collected their data from the API using Python.

In discussing this work, Suk, *et al.* used the term "hashjacking" to describe a strategic way how Twitter users 1) can "invade other people's discussions" or 2) make a tweet go viral. They list several hashtags, such as #churchtoo, #metoomilitary, #mosquemetoo as the change of discourse for #metoo conversation and became sub-movements for activists. The authors ended up categorizing these hashtags as activism discourse and placing them along with other hashtags that do not share the characteristics of morphing the #metoo. For instance, they added what I call hashtag derivatives to other hashtags - #himthough, #howiwillchange, #nomore, and

#stoprape. In this work, I focus on hashtag derivatives. I argue that these types of hashtags are more than just invading other's discussions its about calling attention to who is missing from the conversation by morphing the hashtag to highlight an intersectional position outside of the ideal victim framework which framed by Milano's initial tweet.

2.2.3.2 Temporal Analysis of #metoo Movement

Several researchers have investigated the virality of the #metoo in the first few days and weeks of the movement. Lindgren [192] examines the viral hashtag movement's momentum using topic modeling and sentiment analysis to analyze over four million tweets collected from the Twitter API from October 15-18, 2017. The researcher's aim was threefold: (1) to assesses to what degree the #metoo campaign maintained its focus on Twitter, (2) to analyze the tone of the hashtag campaign on Twitter, and (3) to explore the clicktivism and disengagement of the #metoo campaign. The author found that #metoo quickly started to lose its momentum after the initial and explosive impact, with noise, antagonism, and sloganization increasingly weighing down and diluting the campaign. While the author's findings are quite vital in regards to the main #metoo hashtag activity in the first few days of virality, it is unclear why the loss of topical focus of #metoo coincided with the divergence of hashtag variations.

Hassan, *et al.* [144] collected a similarly large sample of 441,925 #metoo tweets (from 323,813 unique users) collected from Twitter between October 15-31, 2017, to better understand how online social movements disseminate information, educate, and develop community. They performed a quantitative analysis by identifying the gender and age of Twitter users with Face++[9] determine the sentiment and topic

---

[9]https://www.faceplusplus.com

28

modeling. They found that several people and organizations were using the hashtag to promote products and services instead of discuss sexual assault. One of the most popular other results included women sharing stories of their experiences. While they included the gender and geolocation in their work, once again, there is still a collapse of intersections for gender and not include non-binary trans identities.

2.2.3.3   #metoo Movement Participation

Xiong, *et al.* [302] looked at social movement organizations' (SMO) participation and use of hashtags in the #metoo movement. Their work sought to understand how social movement organizations used digital platforms and functioned during hashtag activism. They collected tweets from each social movement organization's account from October 15, 2017 to January 7, 2018, corresponding to when Milano's tweet went viral and Golden Globe awards, respectively. A total of 408 tweets were collected. Most of the accounts focused on issues of women's health, well-being, and safety. The organizations focused on feminism-centered and activism-centered networks, and thus the researchers created an ego-centered network[10], which showed a total of 58 words, and the top five most-frequently-mentioned reflect feminism, sexual harassment, and survivors, movements, and stories, and then performed a semantic analysis on that network by collecting tweets from a third-party source where they looked for SMO Twitter handles that included #metoo in their tweets. In their work, the researchers mentioned the various iterations of the #metoo (*e.g.,* #metoocongress, #hertoo, #ustoo) being defined as directly related to the movement however they included #timesup in the same classification as the iterations. This work provides an additional impetus to further explore these various iterations #metoo, the authors

---

[10]An ego-centered network focuses on a distinct individual in a network which is referred to an ego network [294].

summed up iterations under a single theme being movement related. However there are tensions with these different themes alluding to different place, and person and groups that should not be summed under a single monolithic theme summation.

Suk, *et al.* [271] created retweet networks with Twitter users as nodes and the tweets between them as edges. Networks based on the opinion leaders and hidden influencers to other understand activism and networked acknowledgment discourse. While their work revealed various people participating in the public acknowledgment of #metoo by sharing their stories and experiences, the networks focused more on the main events occurring in the media and influential people such as the actresses in Hollywood. Even though the authors did mention several ordinary users and grassroots organizations present in the network, they did not go into depth about their function.

2.2.3.4 Identity

Hosterman, *et al.* [155] analyzed 2,782 tweets during the first six months of #metoo on Twitter, and manually identified Twitter users by their gender (man or woman) using their profiles. They used the twitter metrics of favorites and retweets to understand engagement in the #metoo movement. They found that most tweets were informational support messages tweeted by both individuals and organizations. They found men shared more informational support compared to women, who shared more emotional support. A gap in this research is limiting the gender to a binary category, and to only considering global metrics for analyzing engagement.

Improving on the demographic front, Mueller, *et al.* [215] examined intersectional identities of gender and race or ethnicity in a sample of 660,237 tweets from 256,650 unique users during the first year of #metoo. Their work used topic modeling to examine the broader content and storytelling of tweets. The researchers found

notable differences in representation and storytelling across Black and white women. Black women's tweets involved discussions of emotional support and inequities in the justice system for sexual assault, while white women's tweets surrounded conversations about sexual assault and violence by public figures and more general political discussions. While the researchers included race and gender in their demographics, they unfortunately left out other key intersectional identities of non-binary gender, and trans.

Trott [278] examined the intersectional concerns that complicate the scale, reach, and seeming permanence of the millions of testimonials suspended online that constitute the #metoo movement. Trott's work stems from Onwuachi-Willig [228] essay discussing marginalized women's voices, and those that are more vulnerable were not as heavily circulated online when #metoo's virality occurred. The researcher investigated 81,408 tweets in the first three days of the viral hashtag to look at intersectional narratives. She looked for terms that signified color, queerness, sexuality, and gender identity. She then conducted a qualitative discourse analysis on the dataset from the keyword search to understand how identity and relationality was discussed within the #metoo network. In her work she combined social network analysis and discourse analysis to look at the network feminism and brings attention to the voices that were erased from the dominant narratives of #metoo. Trott [278] found two dominant modes within the dataset: (1) a call to action for "all the women" to discuss their experiences and (2) sexual assault does not discriminate and for people to stand together regardless of individual positionings. One point that Trott mentions in her work about using intersectionality is that her work was "not written with the intention of retelling intersectional narratives [to] marginalised women but to draw attention to the exclusivity of popular and networked feminism and to elevate the voices of the multiply marginalized survivors of sexual violence who were erased from

31

the dominant narratives of #metoo." [278, p. 1]

2.2.3.5   Disclosing Sexual Violence

Modrek and Chakalov [213] examined the frequency of disclosures of sexual abuse or assault expressed by singular identities (gender and, separately, race/ethnicity that were inferred using a commercial marketing software) in 11,935 tweets during the first week of #metoo. The researchers found that 11% of the tweets disclosed an incidence (of abuse or assault), and these tweets were overwhelmingly authored by white women. A gap in this work is the limiting of demographic identities (*e.g.,* race and gender) captured by the researchers in their investigation of the #metoo research. Building on this theme, some research has applied qualitative analysis of the demographics of #metoo movement contributors [215, 231]. PettyJohn, *et al.* [231] looked at the hashtag #HowIWillChange in tweets during the first week of #metoo hashtag went viral. The authors found that the hashtag tweets were associated with allyship by discussing strategies to dismantle rape culture, including open hostility.

Lowenstein-Barkai [196] sought to address the literature gap to examine whether men and women elicit different responses based on gender stereotypes. They performed a quantitative content analysis of 2,635 responses to 734 self-disclosures of male and female survivors of sexual victimization published on Facebook and Twitter, during the first 3 weeks of the #metoo and #WhyIDidntReport protests in Israel (October 2017 and October 2018, respectively). They found that social networks were supportive environments for survivors of both sexes. However, there were differences in the type of support each gender received, for instance women received more emotional support compared to men who received retributive support[11] [196]. While this

---

[11]Authors defined retributive support as "calls for punishing the perpetrator and for restoration of justice." [196, p. 9]

literature looks at the #metoo movement from a non-western perspective, there is still a gap in how users are classified in according to gender. Researchers only define gender being two categories.

Suk, *et al.* [271] sought to predict the relationship between network acknowledgment and #metoo activism discourses. They defined network acknowledgment as "when online communities sustain a discourse that allows public testimony about trauma, provides a space for open discussion about claims, highlights common experiences, and affirms faith in the stores of survivors." [271, p. 5] They examined the temporal dynamics of these discourses within the #metoo movement on Twitter by analyzing a 1% sample of the global Twitter stream associated with a set of search strings – "metoo," "timesup," "sexual assault*," "sexually assault*," "sexual harass*," "sexually harass*," "sexual molest*," "sexually molest*," "sexual misconduct," "feminism," "feminist*"– from over five months, from October 2017 to February 2018. This yielded a total number of 395,037 tweets, which were reduced to 296,387 English-only tweets. Suk, *et al.* found a variation in the volume of monthly tweets, with small spikes on particular days, consisting of sharing of personal stories and expressing of solidarity; the spikes gradually waned over time. However, a different temporal pattern was observed in the activism discourse, which is marked by punctuated moments of growth throughout the entire period.

2.3   #metoo movement: Theories and frameworks

When examining the #metoo movement, the guiding frameworks and theories used to study this social movement have stemmed from legal [37, 77, 110, 228, 286], post-feminist [2], hashtag feminist [59, 210, 302], and generic feminist underpinnings [3, 60, 114, 186, 234, 234]. Researchers have use theories and frameworks with psychology-

based [122] and communication-based theories such as affective publics [271], mental health theories such as social support theory [155, 196]. Researchers, in addition, have used frameworks and theories such as guilt-redemption cycle and consubstantiality [30], intersectional-feminism [198], and counterpublics [278] in conjunction with intersectionality to highlight power structures at the core of exclusion dynamics in the #metoo hashtag. Whether these theories are either using intersectionality tangentially, partially, there is a sparse amount of research where we see intersectionality being utilized exclusively.

For instance, Leung and Williams [187] implement a comparative analysis looking at the Robert Kelly[12] scandal and responses from Black women compared to the backlash after the Harvey Weinstein scandal. They used intersectionality to understand the #metoo movement analyzing these two scandals. They argued that intersectionality of the #metoo first emerged in the summer of 2017 when articles were published alleging that singer R. Kelly sexually, physically, and mentally abused a group of African-American females and outlined his history of predatory behavior against young and underage women of color. This comparative study highlighted the injustices and the differences between treatments of white and Black victims. Within this work, we see the intersectionality framework utilized as a critique in a comparative analysis.

Mueller, *et al.* [215] used the intersectionality framework in order to understand intersectional identities of gender, race and ethnicity to understand the demographic representation in the movement. Their reasoning for utilizing the framework stems from intersectionality being a critically important component in historical movements. Mueller and colleagues' utilization of intersectionality warrants grave caution because intersectionality is not amenable to the a la carte approach they took. As will be dis-

---

[12]also known as R. Kelly

cussed later in the Intersectionality portion, because it aspires to be a comprehensive framework, omitting any component of it compromises the rest of the analysis. For these reasons, there is still a gap in the literature where researchers are not relying on this framework exclusively in the space of quantitative methods and begs the need for appropriate and proper application of the Intersectionality to be used in data science space.

Van Rjiswijk [286] uses historian Michael Rothberg's "implicated subject" to analyze the #metoo in conceptualizing responsibility for sexual violence in the context of Australian colonialism. The 'implicated subject' framework draws from the disciplines of law and politics to "think responsibility [sic] for sexual violence laterally–beyond the limited figures of victim, perpetrator, and bystander–and also temporally, connecting the contemporary sexual harms with the legacies of colonialism and slavery." [286, p. 32] Van Rjiswijk continues,

"Approaching #metoo as implicated subjects means re-thinking both the nature of the harm and also the institutional settings that are at the focus of #metoo. With the use of implicated subjects, the legal scholar encourages the #metoo conversation should extend beyond institutions such as schools, and corporations and the focus should include prisons, policing, healthcare and 'family welfare services to dismantle the 'colonial patriarchy." [286, p. 32]

She urges lawyers and legal theorists to bring our these "implicated histories in law to reframe #metoo in colonial contexts" to understand how indigenous people are impacted by sexual assault and violence in the court of law [286, p. 33]. While I agree with the legal scholar that there is a need for a more in-depth understanding of a conceptual context, intersectionality does that.

As Hsu [157] points out, "while intersectional theories and activisms are hardly new, as with conversations about sexual violence, advocates must continually struggle against their erasure." [157, p. 271] In their work, the author proposes trans-ing of the #metoo moment that addresses the entanglement of gender violence and carceral politics [157]. This is important because trans identity is often left out of intersectional analyses despite being very much a definitional aspect of intersectionality. However, what is interesting is that Hsu accomplishes this outside of the intersectional framework. What we can do in this work is to emphasize building trans identity into our own intersectional framework, instead of neglecting it.

Few researchers utilized social network analysis to analyze the #metoo movement using Twitter data to assess the prominent hashtags' discourse either from retweets, replies, mentions. Some work by Trott [278] constructed a social network using the first 50,000 tweets from the first day of the #metoo movement on October 15, 2017; not all the tweets included hashtags. The social network visualizations were created using Gephi to illustrate publics and counterpublics within the #metoo network. There were 40,904 users in this network and 50,018 tweets (mentions or replies), represented as nodes and edges, respectively. Trott indicated a large cluster in the center of the network surrounded by a peripheral rim, the disconnected outer rim of participants who had no ties or connection with any other participants. Trott notes those nodes located on the periphery received no response or engagement to their participation with the hashtag and further raised questions on who is heard and overlooked within the #metoo conversation. When focusing on the network's center, there were users Trott mentioned, which stood out by out-degree: Women's March, Alyssa Milano, Lady Gaga, Amy Siskind, and Dana Loesch. Trott included users Tarana Burke, BlackLivesMatter, and Rose McGowan to indicate their network position. Burke was positioned within the rim, which alluded to being "distant and

loosely connected to the beginning of the #metoo movement." [278, p. 9] Trott said, "This distance indicates a weakness of connectivity between her and the celebrity feminists who are positioned front and centre." [278, p. 9] Trott suggests throughout the movement, particularly when Burke received coverage by mainstream media outlets from *Time Magazine* being a person of the year, perhaps looking at her position to see if it is closer to the center than the periphery.

Two themes emerged from Trott's work: exclusion and erasure of marginalized communities. The first theme stemmed from Milano's tweet, a call for "all the women," which set the agenda to focus on women who experienced sexual harassment and violence, thus excluding all other experiences of men, transmen, nonbinary folks, which share high rates of sexual violence. This corresponds with Boyd and McEwan [38] critique of Milano's initial tweet. The second theme was involved confounding the sexual assault does not discriminate regardless of one's position and should unite because "we are all victims." This illusive homogeneous solidarity only creates further erasure and displacement. Something Crenshaw argues, "the displacement of the 'other' as the presumed victim of domestic violence works primarily as a political appeal to rally white elites." [79, p. 1260] Trott contends cis-women to be added to this list.

The literature's common theme on the #metoo movement is that Intersectionality is being used as a "buzzword" [84] to superficially call attention to the social identities present in—or omitted from—the corpus of data. There is ample discussion of Intersectionality on these terms, but scant highlighting of who is actually being left out of these conversations, and scant criticism of the systems and structures of power involved. These questions of exclusion and power, however, are critical imperatives to the competent study and application of Intersectionality, without which there is no hope of ever effectively calling attention to the inequities it seeks to describe. This is

the heart of the literature gap I have sought to illuminate, and which this dissertation seeks to close by furthering the usage and application of the intersectional framework especially in the discussion of #metoo movement research.

2.4 Intersectionality

"Intersectionality is a way of understanding and analyzing the complexity in the world, in people, and in human experiences. The events and conditions of social and political life and the self can seldom be understood as shaped by one factor. They are generally shaped by many factors in diverse and mutually influencing ways. When it comes to social inequality, people's lives and the organization of power in a given society are better understood as being shaped not by a single axis of social division, be it race or gender or class, but by many axes that work together and influence each other. Intersectionality as an analytic tool gives people better access to the complexity of the world and of themselves." [70, p. 193]

Though term 'Intersectionality' is fairly recent, the concept is not new. It draws from a long genealogy which stretches back to its intellectual foundations in Reconstruction- and Abolitionist-era justice work from women of color. Pioneers like Sojourner Truth, Harriet Tubman, Ida B. Wells-Barnett, Mary Church Terrell[13], the Combahee River Collective[14], Deborah King[15] and many others, all of whom used their own lives

---

[13]In 1940, Terrel argued that Black women had a "double-handicap" of race and sex in her book *A Colored Woman in a White World*.

[14]In the 1980s, they were the first public declaration to mention the framing of identity through an intersectional lens in their statement, arguing "[that] major systems of oppression are interlocking," [64, p. 264] and the compositions of interlocking oppressions impacts the conditions of our everyday lives [64].

[15]King coined the term multiple jeopardy, which better explained the simultaneous oppression instead of taking an additive approach but a multiplicative relationship between racism, sexism, and classism.

to illustrate the overlapping injustices they experienced as Black women [34, 64, 73, 152, 279]. Throughout the decades leading up to the coinage of the term, other Black women contributed to this body of knowledge which argues against binary thinking such as gender-only and race-only experiences [40]. In the late 1980s, a formal formulation of Intersectionality emerged. The term Intersectionality, coined and elaborated upon by legal theorist Kimberlè Crenshaw, how the categories of "race and gender as mutually exclusive categories of experience and analysis." [78, p. 140] Crenshaw's innovation was to synthesize the body of Black feminist theory into a single term. This was motivated by her legal experience with Black women who were discriminated against as "Black women, not only as women and not only as Blacks." [73, p. 4]

Intersectionality complicates analyses by centering the ambiguities, conflicts, and complexities arising from the experiences of people living within overlapping marginalized social categories [57, 64, 65, 70]. It is through this complication that Intersectionality explores how these social identities are confronted by power structures and systems that impact the quality of life and lived experiences of those with such identities. Therefore, Intersectionality is a comprised of two tenets: (1) seeking to understand how multiple social identities are not independent and unidimensional, but rather multiple, interdependent, and mutually constitutive, and (2) emphasizing how these multiple interlocking identities confront structures and systems of power [35, 70].

Understanding the implementation of Intersectionality in the #metoo movement requires a discussion of its genealogy and tumultuous travels across academic disciplines[16], where misunderstandings born of this amazing breadth have led to its mutilation and misunderstood or misguided criticisms. Learning from the ills of other

---

[16]Social Sciences, for example: women gender studies, sociology, education, psychology, legal, library sciences, communication.

disciplines where Intersectionality has traveled, data science now has a earnest opportunity to appropriately apply the rich intellectual framework in a way that will be recognizable to the original founding communities of Black feminism.

2.4.1 Misconceptions and Misuse of Intersectionality

There have been numerous critics [114, 185, 219, 297] who have misarticulated and misread Intersectionality. Through its travels across the academic landscape away from its Black feminist genealogy [17, 73, 189], there are both internal and external criticisms of Intersectionality, with notably significant overlaps. The internal criticisms predominantly center around the of the "flattening" that the discipline has experienced, under the concept of being a "traveling theory," per Sirma Bilge:

> "like other 'traveling theories' that move across disciplines and geographies, Intersectionality falls prey to widespread misrepresentation, tokenization, displacement, and disarticulation. Because the concept of Intersectionality emerged as a tool to counter multiple oppressions, there are multiple narratives about its orgins [sic], as well as tensions over the legibility of its stakes." [26, p. 410]

This flattening of Intersectionality has come from misinterpreting its first tenet, which identifies social identities (or positionings) such as gender, race, sexuality, religion, class, ablebodiedness, *etc.* Alexander-Floyd [1] highlights this as the "universalizing tendency."

However, among the more valid external criticisms, perhaps the best summary of these is given by Carbado [50], who identified six main complaints (or criticisms).

1. The first criticism is the notion that intersectionality is "only or largely about Black women, or only about race and gender." Carbado [50] notes here that

Kimberlè Crenshaw, was responding to two very specific failures in the law, and thus indeed her initial examples focused on the context of Black women who were being failed by the law. Indeed, Carbado [50] highlights that only focusing on the "double-jeopardy" question obscures the deeper context and is an example of the exact power dynamics that Intersectionality criticizes. To paraphrase him, we agree with the conclusion that, "Intersectionality does not have to be solely about Black women—which is in fact antithetical to it—but it does have to respect its genealogy of Black feminism." [50, p. 813]

2. The second criticism is that Intersectionality is an "identitarian framework" - in other words, that it seeks to turn its analysis of identity into what is often mocked as "oppression Olympics." This only makes sense if one sees identity as static, monistic, and uncomplex. Whereas Intersectionality fundamentally calls for the opposite: dynamism, overlapping multiplicity, and complexity.

3. The third criticism is that Intersectionality is a "static theory that does not capture the dynamic nor contingent processes of identity formation." Per Carbado [50]: "Nor is the theory an effort to identify, in the abstract, an exhaustive list of intersectional social categories and to add them up to determine—once and for all—the different intersectional configurations those categories can form." [50, p. 813]

4. The fourth criticism is that Intersectionality is "overly invested in subjects". This line of argument is confused, however: Intersectionality is a framework for studying subjective truths, and this inherently requires an investment in subjects.

5. The fifth criticism is the sort of meta-criticism that involves the traveling the

Intersectional framework has taken we can no longer gain additional knowledge [50]. But to the extent that Intersectionality has failed or stumbled, it has primarily been failed in its misapplications and co-optations. Intersectionality is a more complex framework than its critics give it credit for, and this complexity is incompatible with the relatively rudimentary multiple regression models that have attempted to describe it. There certainly remains far more left for it to teach data science, for one. And it's difficult to characterize the array of fields Intersectionality has traveled to as "exhaustive", given its only recent widespread dissemination.

6. The final criticism Carbado [50] identifies is the idea that Intersectional framework can not function independently "or at least applied in conjunction with [fill in the blank]." [50, p. 815] Scholars have attempted to do this, and it could even be alleged that this dissertation fits that description. However, this dissertation differs in that it does not purport to be the final word on quantitative applications of Intersectionality in data science, nor to replace anything that was not already an inconsistent or misconceived application of Intersectionality in the literature.

The common thread among the external criticisms [138, 139, 219] is that they invariably arise from misconceptions that have already been readily identified by its practitioners' internal criticisms. We can conclude that Intersectionality is decidedly not a theory which has failed to respond to its external critics or generate its own internal criticisms, but rather one where the sheer spread of its misconceptions has hamstrung most attempts to resolve or address them. Furthermore, as Intersectionality has traveled across disciplines [26], its misuse and misconceptions perpetuate their own self-reinforcing cycle [1, 73].

If we allow the Intersectionality framework to remain intact there are several opportunities to shift from the dominant perspectives. As Cooper puts it, "while it [Intersectionality] brings into focus marginalized people...[and] while the relations of power Intersectionality exposes [structural inequalities], Intersectionality does not tether black women to a certain epistemological standpoint." [73, p. 5] Collins and Bilge [70] therefore describe Intersectionality as encompassing a dual focus on inquiry and praxis which will be used in this work.

Furthermore, Cooper describes Crenshaws's revisiting of Intersectionality due to the ideas about identity and cultural battles over identity politics. Crenshaw explicitly states that Intersectionality is not "some new totalizing theory of identity rather" [79, p. 1244] a way to understand how these social identities impacted in relationship to power in the social constructed world [79]. Cooper summarizes Crenshaw's two intersectional works stating the "essays catalyzed a tectonic shift in the name of feminist theorizing by suggesting that Black women's experiences demanded new paradigms in feminist theorizing, creating an analytic framework that exposed through use of a powerful metaphor exactly what it meant for systems of power to be interactive, and explicitly trying the political aims of an inclusive democracy to theory and account of power." [73, p. 2]

Intersectionality has been used as a tool by researchers to solve analytical problems where there are multiple identities in play. For instance, Collins and Bilge cite examples of how colleges and universities use Intersectionality as a "useful analytical tool for thinking about and developing strategies to achieve campus equity." [70, p. 2] Other examples include using Intersectionality as a better framework to grapple with workers' rights in complex cases of discrimination [70]. Intersectionality has also been used as an analytic tool to examine the structural inequities of power relations in Brazil during FIFA World 2014 Cup. Collins and Bilge further elaborate,

43

"Race, class, gender, sexuality, dis/ability, ethnicity, nation, religion, and age are categories of analysis, terms that reference important social divisions. But they are also categories that gain meaning from power relations of racism, sexism, heterosexism, and class exploitation. One way of describing the organization of power identifies four distinctive yet interconnected domains of power: interpersonal, disciplinary, cultural, and structural. These four dimensions of the organization of power provide opportunities of using Intersectionality as an analytical tool to better understand the 2014 FIFA World Cup." [70, p. 7]

2.5   Quantitative Methods in Intersectionality

Intersectional frameworks are increasingly influential within behavioral sciences like psychology, public health, and epidemiology as critical, qualitative, mixed-methods, and quantitative methods [207]. Researchers have advocated for a mixed methods intersectional approach, because it first establishes similarities and differences among the populations qualitatively, following up quantitatively with measures appropriate to analyze the data [36]. Bowleg & Bauer hold that the mixed-methods approach in intersectional research 'capitalize(s)' on the advantages of quantitative and qualitative methods [36], and furthermore, that it encourages quantitative intersectional researchers to learn qualitative approaches to help understand the data.

However, applying Intersectionality quantitatively on its own allows researchers to capture stories and lived experiences numerically outside of the dominant context. Covarrubias argues a benefit of using quantitative Intersectionality "is the ability to uncover often concealed anomalies that may require further investigation." [76, p. 102] Researchers have found quantitative applications of Intersectionality in context with

social-historical context have revealed patterns for various social identities [35,76] and aided in the understanding how marginalized communities were impacted by power structures (*i.e.,* police violence, sexual assault) [2]. It is imperative for researchers not to neglect the socio-historical context and structural inequalities confronted by real people based on their social positionings [34, 92, 99, 299]. Lacking a social-historical context shifts to an over-emphasization and hyper-focus on group differences between social identities [92, 299].

Another pitfall Bauer cautions against is that "[the] quantitative applications of Intersectionality can be obfuscated by the predominance of mathematical-like language in Intersectionality theory, though its use there is conceptual rather than strictly mathematical." [16, p. 12] Bauer is alluding to Bowleg's paper titled *"When black + lesbian + woman ≠ black lesbian woman.* In this paper, Bowleg calls attention to not only the additive identity fallacy where people's experiences are not separate, independent nor summative, but also that the manner in which Crenshaw described gender, race and sexuality as interacting multiplicatively, does not literally translate in a mathematical sense using a multiplicative-scale statistical interaction model. Bauer further adds, "If Intersectionality is to be implemented in quantitative research, then terminology will need to be disaggregated in order to allow for clear communication and to... prevent the conflation of identical-or similar sounding concepts." [16, p. 12] This presents the opportunity for intersectional quantitative researchers to investigate these mathematical tools to see whether and how they are applicable and to explore or develop alternatives that consider non-binary results.

Bowleg impeccably captures the tenuous juxtaposition between Black lesbian poet Audre Lord's famous quote, "The master's tools will never dismantle the master's house" [194, p. 111] and incompatibility with statistical methods, which I hold to be extendable further to mathematical models and algorithms than in the presently

prevalent cases of misuse. Bowleg writes:

> "That is, the statistical methods, even those that test interactions, were not designed with the study of intersectionality in mind. Rather, statisticians rooted in positivistic [sic] paradigms developed statistical assumptions of linearity, unidimensionality of measures, uncorrelated error components and the like that do not reflect the real world complexities of intersections of race, sex/gender and sexual orientation. In short, we need new analytical tools and strategies to assist us in understanding the complexities of intersectionality." [34, p. 320]

Calling attention to the limitations and incongruities of statistical methods and the shortsightedness of utilizing solely quantitative methods, the current employment of the intersectional framework appears mostly and despondently inefficacious, especially in a quantitative field like data science. If we are to return to good knowledge production within intersectional framework, holding onto its genealogy and intersectional scholarship, we need to rely on its concept of reflexivity. Collins and Bilge [69, 70] discuss the importance of being self-reflexive when using Intersectionality as a form of critical inquiry and praxis. This reflexivity allows researchers to call attention to their own practices in the context of Intersectionality [26] while navigating the research process. To apply this reflexivity, researchers need to be aware of their own positioning in relationship to the data in how their prejudices and bias are imparted onto the processes (*e.g.,* quantitative methodologies) of their investigations. This reflexivity is embodied within the core constructs and guiding premises of Intersectionality, and is demonstrated throughout this dissertation in its dedication to presenting its methods and framework critically.

2.5.1   Core constructs and guiding premises

Intersectionality's core constructs consist of relationality, power, social inequality, social context, complexity, and social justice, which inform and remain in conversation with one another [69, 70]. Varying intersectional projects place varying degrees of emphasis on each of the core constructs. Collins reminds us not to use the core constructs as a checklist to see which boxes a particular project checks off; instead, she insists it's more useful to investigate how and in what ways these concepts reappear either singularly or in combination with an intersectional inquiry. These core constructs were introduced by Collins and Bilge [70, p. 194-204] and further elaborated by Collins [69, p. 45-50]:

Relationality. Relationality rejects the either/or binary thinking (*e.g.,* Black or white), instead embracing the "both/and" frame. The focus of relationality shifts to analyzing what distinguishes entities: the differences between race and gender, to examining their interconnections, mutual engagement and relationships. Systems of power (i.e. race, gender, sexual orientation to name a few) are composed and upheld through relational processes gaining meaning through these relationships' nature [67]. "The analytic importance of relationality and intersectional scholarship demonstrates how various social positions necessarily acquire meaning and power (or lack thereof) related to other social positions."

Power. Intersecting power relations produce social divisions of race, gender, class, sexuality, ability, age, country of origin, and citizenship status that are unlikely to be adequately understood in isolation from one another. Non-intersectional scholarship assumes that race, class, and gender are unconnected variables or features of social organization that can be studied as singular phenomena, gender or race as discrete aspects of individual identity, or patriarchy or racism

mono-categorical systems of power. Intersectionality posits that systems of power co-produce one another in ways that reproduce unequal material outcomes in the distinctive social experiences that categorize people's experiences within social hierarchy.

Social Inequality. Using Intersectionality as an analytic tool encourages us to move beyond seeing social inequality in race-only or class only lenses. Instead, Intersectionality encourages understandings of social inequality based on interactions among various categories.

Social Context. Social context is important for understanding how interpretive communities organize knowledge production. This premise applies to internal dynamics for a given interpretive community and how communities of inquiry are hierarchically arranged and valued.

Complexity. These core themes of social inequality, power, relationality, and social context are intertwined, introducing complexity into intersectional analysis. Intersectionality itself is a way of understanding and analyzing the complexity in the world. Using Intersectionality as an analytic tool is difficult, precisely because Intersectionality itself is complex.

Social Justice. Intersectionality is not a simple substitute for social justice. The construct of social justice raises questions about the ethics of intersectional scholarship and practice. Social justice within the context of Intersectionality allows for the challenge of norms that place "social justice, freedom, equality, and similar ethical issues as secondary concerns within acceptable scholarship." [69, p. 47]

Furthermore, Collins introduces four guiding premises for these six core con-

structs of Intersectionality [69, p. 49]:

1. Race, class, gender, and similar systems of power are interdependent and mutually construct one another.

2. Intersecting power relations produce complex, interdependent social inequalities of race,class, gender, sexuality, nationality, ethnicity, ability and age.

3. The social location of individuals and groups within intersecting power relations shapes their experiences within and perspectives on the social world.

4. Solving social problems within a given local, regional, national, or global context requires intersectional analysis.

Therefore, nucleus of Intersectionality is, as described by Collins is this idea of reflexive accountability, using these core constructs within this framework paradigmatic thinking [69]. These core constructs can be used by researchers and scientists to guide their research process.

2.5.2 Intersectionality and #metoo Research

Understanding the implications of whiteness and white supremacy as a domination and power system, how it became institutionalized, perpetuated and upheld, allows us to be conscious and recognize ways whiteness is embedded into our technology where people are exploited and oppressed. Whiteness is a social construct used for social, cultural, political, and economic domination (even without overt or explicit force) that also obscures its power by being passive, invisible, and taken-for-granted by the people privileged with it [102, 145]. Acknowledging the existence of, participation in, and benefits from these constructs is an important first step to-

wards recognizing and reforming these invisible structures through which social power flows [153, 209, 223].

Applying Intersectionality to the #metoo movement makes sense because there is a highly diverse sociohistorical context (*i.e.,* sexual assault and harassment) being collapsed under a single sensationalized hashtag. These popular narratives presume a homogeneous and universal womanhood and manhood and obscure the ways in which sexual assault is racialized [200]. Unfortunately, sexual assault, violence, and harassment are ubiquitous across far more social identity dimensions than race, and there are systems and structures of power in place which prevent the needed healing. As Mack and McCann [200] poignantly point out, "while #metoo galvanize [sic] survivors to speak up, many victims of sexual assault cannot speak up for fear of further violence from individuals or do not want to speak up because they fear violent state sanctioned responses from their already marginalized communities." [200, p. 331]

Intersectionality has been used to investigate the #metoo movement, either quantitatively (See section #metoo Research: Criticism and Quantitative Research) or in conversation with other theories. However, there is a gap in the scholarship of utilizing this frame reflexively, in its totality, and with its genealogy intact.

With the Intersectionality framework challenging systems and structures of power to make visible the inequities for marginalized groups, there is an opportunity to interrogate these mathematical algorithms in data science.

Suk, *et al.* [271] found that various users participated in networked acknowledgement around #metoo, exposing the prevalence of sexual assault and sexual abuse, and building a sense of shared experience and identity.

2.6 Data Science and Power

"Digital Black studies" refers to a rich tradition of exploring the overlaps of cultural, media, and ethnic studies [80–82, 104]. These researchers are engaged in questions of how digital systems represent identity [62, 218], how communities convene attention and conversations [103, 208], and how platform affordances[17] shape social movements [165, 183]. More recent work has characterized how the gender, race, and sexuality biases embedded within invisible technologies continue to directly and indirectly harm marginalized communities [22, 42, 130, 221]. The gap in the literature is how we, as researchers, do better with this is in a quantitative space and working with quantitative data. As Hampton [137] eloquently argues there needs to be distinction between social science and STEM discussions on race and technology, its a grave error to homogeneously collect all these works - largely produced by Black women - under the single heading of 'race and technology.' While it is quite helpful and necessary for the social sciences to illuminate these issues, there is an absence of quantitative methods critiquing these processes and algorithms mathematically. Therefore, this dissertation presents the opportunity to critique dual juxtapositions: (1) between the social science establishment which eschews the need for rigorously intersectional quantitative data science methods, and (2) between the data science establishment perspective which all too often discounts Black womens' contributions to data science under the heading of 'race and technology.'

Recently, there has been a much-needed reckoning within the computer and information sciences about how the methods, tools, and practices of data science can be unethically applied to very efficiently achieve malign or benignly negligent or harmful goals. Although the seeds of human-centered computing go back decades [11,

---

[17]Refer as reciprocal interactions between a technology application, its users, and its social context.

166, 175, 285], there has been an explosion of interest around initiatives branded as "fairness, accountability, transparency, and ethics" (FATE) [23, 108, 160, 217, 260], "data science for social good" (DSSG) [308], and "human-centered machine learning" (HCML) [101, 119, 239] or "human-centered data science" (HCDS) [8, 177]. In the following sections, there were will be a brief review around the different initiatives.

2.6.1 FATE and Critical Digital/Data Studies

These HCML and HCDS efforts highlight a tendency to center the technical needs (*e.g.* better data and methods) over human needs of safety and justice and are reluctant to explore how they reinforce social structures that accommodate injustice and reward exploitation. Efforts like fairness, accountability, transparency, and ethics (FATE) are good-faith attempts to grapple with how technical systems reproduce and reinforce the problems of the social systems within which they are embedded [108], but these nevertheless tend to privilege incremental technical innovation over developing policies for regulating for new capabilities, much less attending to how technologies reinforce systems of social power [23, 160, 217, 260]. Fields like critical digital/data studies have examined the historical, economic, and political pressures that shape the very material properties of computing and data technologies [39, 161, 174], but these critiques are often ignored within engineering cultures with their unclear language and absence of technical advice that engineers can implement [13].

The intersectional framework application into the data science field allows the complete immersion not only from (nor limited to) techniques, methods and tools, but also holds researchers and scientists accountable with this reflexive component of Intersectionality [69, 70]. Reflexivity is the key difference between previous work in HCML, HCDS and data science communities. Once again, in order to apply this

reflexivity–and to larger extent Intersectionality–into data science, researchers need to be aware of their own positioning in relationship to the data in how their prejudices and biases are imparted onto the processes of data science. Furthermore, not to be tempted to mutated and contort the Intersectionality framework away from its rich Black feminist genealogy.

2.6.2 Human-centered Machine Learning

Machine learning and data science processes like clustering, classification, recommendation, and forecasting are often conceived of as impersonal, abstract, and passive operations. Standardized benchmark data sets exist, there is consensus on defining metrics for evaluating performance, and researchers focus on building algorithms that accommodate the unique properties of data or task that ideally generalize to other kinds of data and tasks [151]. The increasing complexity and brittleness of many machine learning deployments has led to calls for greater interpretability [89, 170] and explainability [222, 289]. Other themes within human-centered machine learning (HCML) have focused on greater interactivity and usability by lowering barriers for building models with automatic, interactive, human-in-the-loop, and machine teaching models [101, 239, 288]. These formulations of HCML continue to center improving the performance or adoption of ML systems but have not reflected on how human-AI collaborations can be designed for human enrichment rather than extraction.

2.6.3 Human-centered Data Science

Human-centered data science (HCDS) is billed as one of these bridges for scaling up qualitative methods like virtual and trace ethnography, integrating quantitative methods into qualitative workflows, and sensitizing data science cultures to humane

values of privacy, ethics, and sustainability [8]. Subsequent articulations of HCDS have emphasized leveraging powerful computational techniques while accounting for the nuances and situated nature of digital trace data [177]. But these formulations continue to center methodological concerns around the collection, analysis, and interpretation of data over the implications for humans subjects to intrusive and coercive models of data science.

Before conclusion of this chapter, and there is an intermittent review on the data science process, any discussion of "data science" as a domain must begin with some definitional hand-wringing. I scope the definition in a bottom-up strategy by attending to data scientists' practices rather than a top-down mapping of disciplinary identities. There are no shortage of frameworks and flow-charts for describing data science processes of mining data into knowledge [24, 269]. While data debugging and cleaning is estimated to take up most of a data scientist's time [58, 111], these practices occur downstream of other decisions around the design of the data and databases and operational missions and research strategies for logging behavior as digital traces [97, 233]. Once the data has been designed and captured, the subsequent engineering practices by filtering out irrelevant data, handling missing data, transforming the data into the standards of other libraries, and engineering unstructured data into compliant features [216, 254]. Subsequent steps of modeling the data modeling, validating, and deploying data models are characterized by increasingly formalized roles and workflows [211, 309], which are rife with opportunities for biases—unconscious or not—to creep into an analysis [106, 156, 282]. The six standard data science practices[18] are summarized below to structure the case studies and further the discussion.

---

[18]These practices are not intended to be definitive or exhaustive, just recognizable to the median data analyst or scientist.

1. Design Pipeline - Determining how to structure and implement the process.

2. Data collection - Querying and sampling raw data for analysis.

3. Cleaning - Cleaning and transforming into "clean" data.

4. Explore - Optimizing algorithms to find patterns.

5. Model - Measuring and interpreting model performance.

6. Interpret - Sharing models as products and papers.

Chapter 3

Research Study 1

For the two case studies, the Intersectionality framework will be simultaneously applied in various ways throughout the data science process and be used as a guiding framework. Therefore, this chapter includes the research data collection, the methodology used, and discussions on its rationale.

There have been several scholarship discussions about the co-occurrence of hashtags in tweets for various reasons outside of, and including, social movements. We have also seen the appearance of the hashtag variation phenomenon that has been described in other online social movements such as Occupy Wall Street and #metoo. Researchers have collapsed or summed these incidents under the main hashtag. Therefore we have neither precise nor consistent terminology describing these hashtag variations. Researchers have also reported that #metoo variation resulted from subsequent hashtags created by individuals during these viral movements to call attention to various related matters. But once again, this results in a collapse of the larger movement. Introducing the Intersectionality framework with these different phenomena, we are encouraged to pause and observe the present social identities as they are confronted with power systems. For our #metoo case study, the social iden-

tities are represented by the variation of hashtags of #metoo and the systems and structures of power are replicated on this digital platform [129, 130, 218].

For instance, there are variations of #metoo that brought attention to particular communities that were impacted by sexual assault and/or violence in unique ways, which simultaneously occurred with the #metoo hashtag, Figure 3.1, for example, illustrates a tweet from Stop Sexual Assault in Schools (SSAIS), a non-profit organization calling attention to the sexual violence in kindergarten and throughout high school education. This organization is not part of the Hollywood community and yet participated in the #metoo conversation using the hashtag, as well as a hashtag derivative of their own to call attention to the sexual misconduct in kindergarten and throughout high school. The #metook12 represents an age intersection challenging the dominant narrative that sexual violence only affects women, when in truth it unfortunately affects anyone regardless of age. This again echoes Burke's sentiment when she reflected on her encounter with the teenager.

Figure 3.1: Example of #metoo hashtag variation

3.1 Data Collection

Sampling is not unique to Intersectionality research [91, 92]. When selecting a sampling method, researchers stress that the type of sampling methods utilized should be inclusive of marginalized populations and capture intersectional positions instead of using other, methods, such as random sampling [36, 91, 92]. Therefore, snowball sampling was appropriate for this work because this technique is known to find hidden populations that would not ordinarily be included using more conventional sampling techniques [277]. As a non-probabilistic technique for iteratively expanding a sample based on information within the sample, the snowball technique has mainly been used in the social sciences, information science, human computer interaction, and health care fields [32, 44, 256, 262].

This work implemented snowball sampling in a novel way while acknowledging its inherent seed-selection bias[1] [44, 301]. To address this bias, seeds were chosen by selecting the Twitter accounts of the two women who played the most critical roles in the movement [25]: Tarana Burke and Alyssa Milano. I picked these two women for their unique roles in the #metoo movement as (respectively): (1) the creator of the 'me too' movement, and (2) the celebrity widely credited for the 'me too' phrase and hashtag going viral on Twitter, respectively. The tailoring of these seeds also aids in the intersectional immersion of this work. The overall topic #metoo was about sexual assault and violence; however, Burke and Milano were located in different intersections, Black woman and white, affluent, Hollywood woman.

The point of using Intersectionality in this #metoo work is not to solve the problems and tensions of racism, sexism, and any other interlocking systems of oppression under this one movement. Instead, this work will provide a foundation

---

[1]Another possible bias may be introduced by spam tweets and hashtags. Identifying spam in Twitter is beyond this work; thus, we cannot estimate the bias introduced by spammers.

for intersectional work when investigating communities that possess multiple social identities. This work's contribution approaches its analysis from a vantage point on the margin to tell something about singular, non-intersecting axes of identity. Furthermore, finding these hidden populations aids in the argument that the #metoo movement should not be collapsed under a single topic (*e.g.,* sexual assault and violence) due to the variance in how communities receive support. Refer to the Appendix for Snowball Sampling.

In order to collect data using snowball sampling for the data collection, there were two inclusion criteria for a Twitter account to be collected: (1) having been "@"-mentioned by a user in the previous generation, and (2) having used the hashtag #metoo (case-insensitive) in a tweet. This sampling process proceeded in generations, until eight generations of accounts were collected. The collection process was stopped at eight generations due to diminishing numbers of includable accounts.

Although the #metoo hashtag went viral on October 15, 2017, the eight generations of Twitter accounts using snowball sampling occur between the dates of January 1, 2017 and June 20, 2018. While the original date that the #metoo hashtag went viral was in October 2017, data collection started in 2017 to see if there were any previous mentions of #metoo from any of the users, which could potentially indicate a bias in the two seed points, to which there were not[2]. As mentioned previously, the snowball sampling date was terminated due to the reduced number of accounts.

The features collected were tweet metadata consisting of follower counts, following counts, hashtags, author id, time, date, retweets, URLs, favorites, the text of the tweet, and permalinks. There were over 8000 Twitter accounts mentioned over all eight generations in the initial corpus, before applying any exclusion criteria. The

---

[2]The full set of users' tweets were collected, rather than those tweets relevant only to the focal event.

data were downloaded in CSV (comma-separated values) format to perform other types of analysis in Excel, Python and R.

Twitter accounts were discarded from the corpus when they had zero tweets, had been deactivated, were retweet-only accounts, or were private accounts, reducing the total number to 6,944. Of this amount, there were only 3,965 Twitter accounts that satisfied the two inclusion criteria mentioned earlier, totaling 51,602 tweets. For more details on how the Twitter Accounts were classified, refer to the Appendix for User Classification. This dataset will be used in various ways to answer the subsequent research questions in for these studies.

3.1.1 Ethical Considerations

These studies did not directly involve human subjects, but rather the content they produced on social media. Data collection procedures followed the ethical merits of Internet-based research using publicly available data only and keeping users' anonymity intact [31]. These studies were considered exempt from Institutional Review Board (IRB) approval and no formal informed consent process was required. In addition, all identifying information of both disclosures and respondents (except for disclosers' gender, race/ethnicity, orientation in profile) was removed from the database to maintain ethical standards of online data and to protect users' anonymity.

3.2 Methodology

In this work, I provide more consistent definitions of the terms "hashtag co-occurrence" and "hashtag derivatives." "Hashtag co-occurrence" will be defined as the co-appearance of any two or more hashtags in one tweet. This expands upon the definitions given by other researchers [164, 235, 290], who tend to limit it to co-

appearances of only two hashtags in one tweet. This decision not to limit to two stresses the need for rigorous formal definitions of terminologies: the limit inherently stifles the critical source of information represented by multiple co-occurrences. Secondly, "hashtag derivative" is defined as a hashtag whose composition varies on, but strongly reflects, the original hashtag. The #metoo, for instance, is the original hashtag, and an example of a hashtag derivative would be #metook12 as well as #churchtoo. These derivatives follow formulas such as #metoo+suffix or #prefix+too. With these definitions established, it is important to note a key distinction: a derivative may co-occur with the parent hashtag, but a co-occurrent hashtag is not necessarily a derivative.

All programming was done in Python[3].

3.2.1   RQ1a: Hashtag Categories

Question. Which categories of hashtags are frequently used during the #metoo conversation?

Of the initial data set constituting 51,602 tweets (3,965 unique users), there were 27,370 tweets from 1,805 individual users using hashtags. Given the global reach of #metoo, there were tweets in various languages such as Dutch, Korean, Mandarin, Japanese, and Spanish in the collected corpus where some of the tweets contained non-ASCII characters. Therefore, 14 non-English and non-ASCII tweets were discarded, which left 27,356 tweets with hashtags and 1,804 unique users[4]. Non-alphanumeric characters were not removed, as this was part of the conversation (*i.e.,* 4life, 2upset).

Previous work exploring which features best capture social influence on Twitter involved exploring the number of follows and PageRank [184,295], investigating social

---

[3]https://www.python.org/
[4]As a consequence of removing the tweets, a user was removed from the set

influence metrics (*e.g.*, in-degree, retweets, mention and lists) [51, 305], and gathering user attributes that allude to influence (*i.e.,* followers, friends and tweets, date of joining, URLs) [10, 85]. The main difference between this and previous work is that in this work, the social features were tailored to those that have more granularity to the #metoo movement. For instance, there are variations of #metoo which brought attention to particular communities impacted by sexual assault and/or violence in unique ways, which simultaneously occurred with the #metoo hashtag, as is shown in the tweet in Figure 3.1.

Within Occupy hashtags, these variations were a way to reconstruct the network of communication among protesters by location [123, 125]. In the viral #metoo movement, hashtag variations can be a sign of different types of community formation processes and communication among groups of people. Therefore, creating features that capture these variations of #metoo hashtags in the viral movement might highlight intersectional positions to provide more insight into the social influence within the movement. These new features accounted for the frequency of #metoo derivatives in a tweet and the number of hashtag co-occurrences in tweet. For more details on how the features were created visit the Appendix on Feature Engineering.

For the categorization of hashtag variations, the coding process was conducted in two stages: first, two coders classified the hashtags independently based on tweet context, and then they discussed the classification. Categorization is used to understand the various intersections present in the dataset. Therefore, hashtags were classified based on the different intersectional identity domains such as (but not limited to) race, ethnicity, age, religion, class, nationality, gender, sexual orientation, ability, *etc.*

3.2.2 RQ1b: Time Series

Question. What temporal hashtag patterns occurred in the first week of #metoo?

For this study, the aim was to cover the period immediately following the initial tweet on October 15, 2017. The focus is on the Twitter users who participated for the first week of the #metoo hashtag going viral. Like Modrek and colleagues [213], the selection of this week was due to it having the greatest activity in the viral movement and moved uncommonly rapidly from hashtag creation to virality compared to other social movements[5]. The objective of this choice was to gauge how the participants were enacted with the hashtags and view the presence of hashtag derivatives.

Of the initial dataset constituting 51,602 tweets (3,965 unique users), the data was subsetted to the first week of the viral hashtag #metoo between October 15-21, 2017 so that 1,997 tweets containing the hashtag #metoo and/or #metoo hashtag derivatives which were published by 645 unique users, were used. The dataset was created using Python.

3.2.3 RQ1c: Twitter Bio Participants

Question. How do Twitter bios relate to categories and patterns of hashtags?

In previous work, scholars [155, 213, 231] have analyzed demographic groups contributions to the #metoo movement, but did not have the Intersectionality framework to guide them in their work. As Mueller and colleagues [215] noted, the use of Intersectionality to examine identity often falls short of a rigorous application of it as a framework. Therefore, in this study, I extend the prior work to be more intentionally inclusive of the Intersectionality framework, also to include and examine

---

[5]Modrek *et al.* [213] showed that social movements like Black Lives Matter did not take off until months after the hashtag was created. the #metoo hashtag went viral instantly.

intersectional identities of race or ethnicity, gender orientation orientation, pronoun usage, sexual orientation, and transgender, as an important lens [278] for understanding the extent of sexual assault and harassment and not reflect a universal approach built on systems of exclusion that inevitably focus attention on a single group of affluent white women.

The acquisition of social media posts in themselves, including tweets, can garner little contextualizing information [215]. Therefore, aligning this data with the collection of demographic information from users' profiles can be essential for more controlled analyses [215].

Demographics—in our case, gender, orientation, race, and ethnicity—have been used in a variety of studies to contextualize online conversations, especially in #metoo [196, 215]. Numerous approaches exist for demographic inference on social media, including text-based [212, 244, 245] and image-based approaches [201, 220]. It is widely acknowledged that both of these approaches have inherent biases and limitations. For example, known biases include, but are not limited to, a tendency for image-based approaches to incorrectly classify darker skin tones, especially for women [46], and for text-based approaches to display racial disparities [29]. Methods that are reliant on previous postings or social networks are potentially subject to insufficient historical data or prohibitively expensive calls to an API for a user's historical data [300]. Additionally, all of these approaches limit the inference of gender, race, and ethnicity to a binary classification (*e.g.,* users are either male or female and users are limited to a singular racial/ethnic category).

Once again, in the #metoo movement there was a collapsing of voices summed under one topic which appeared to be universal, therefore leading to this simplification that overlooks the complexity of identities in the movement by summing it under gender and race/ethnicity. Therefore, the study's objective is to collect bio-information

from the selected data and then use the features created from RQ1a as part of this study.

For this question, the data collection consisted of collecting Twitter bio data from the 645 unique users from Research Question 1b. Of the 645 unique users, the focus will only be on people and not media and organizations. Therefore this amount is 339 users. Using a similar methodology as Chakraborty, *et al.* [52] to collect these Twitter users' demographic information, we can only use publicly available information about a user, such as their name, profile description, and the tweets they posted. In Twitter's user bio information, a user can post a myriad of things in their profile description. The user can choose to post information about their race, ethnicity, likes, dislikes, *etc.* Therefore, two independent annotators manually annotated the accounts for the presence of race/ethnicity, gender orientation, pronoun usage, and sexual orientation on whether the participants disclose that information in their profile. I opted to use manual annotation (1) due to the relatively small number of accounts, (2) researchers' observed challenges and biases using facial recognition software [46,52] and (3) to gain a better idea of the participants in the data. Therefore, I did not base any decisions on the image of the user. We scraped the textual information from the profile. However, I went to the person's bio if there were ASCII information in their Twitter profile information. Limitation the bio data was collected post-viral #metoo. Furthermore, I am using "presense" of these different attributes to infer instead of explicitly saying which users identified as what to keep in accordance of ethical considerations, as mentioned in Ethical Consideration.

The decision to look for these key features (race/ethnicity, gender orientation, sexual orientation, pronoun usage) is due to (1) the ubiquitous nature of sexual assault and violence (2) understanding how Intersectionality plays a role in various hashtag usage.

The collected data in Research Question 1b involves three (3) variables to understand hashtags usage in a user's tweet. These variables were:

1. Total number of hashtags used.

2. Number of hashtag derivatives used.

3. Number of #metoo hashtags used.

With the information collected from the Twitter profiles and hashtag variables, a canonical correlation was performed. Scholars [135] have said canonical correlation can be seen as extension of multiple regression. In multiple regression, there is a single dependent variable and several independent variables. Although the difference is in canonical correlation, there are several dependent variables correlate simultaneously instead of a single dependent variable [135].

In this research, canonical correlation will be used to determine whether there is a relationship between bioinformation (dependent variables) and hashtag usage (independent variables), each defined within the intersectional framework. Particularly, I am interested in how many dimensions (canonical variables) are necessary to understand the association between the two sets of variables. The null hypothesis is: No relationship between the hashtag patterns and Twitter profile demographic data. The alternative hypothesis: There is a relationship between hashtag patterns and Twitter profile demographic data. At a confidence level of 95% (p-value < .05), we can reject the null hypothesis.

3.3 Analysis and Results

3.3.1 RQ1a: Hashtag frequency

Data Analysis. Create a descriptive report hashtags including number of unique hashtags, co-occurrence of hashtags, their derivatives, and total frequency of each hashtag; categorize hashtags based on keyword analysis and qualitative assessment.

I analyzed the hashtags for 27,356 tweets. There were 19,781 tweets (72.31%) with one hashtag and 27.64% (7,575) with two or more hashtags. In Table 3.1, the distribution of hashtags per tweet is given. The most frequently used co-occurrence hashtags were #metoo #timesup, followed by #metoo #metoo, see Table 3.2.

Number of Hashtags	Counts	Percentage
1	19781	72.31
2	4897	17.90
3	1720	6.29
4	604	2.21
5	186	.68
7	37	.25
8	20	.13
9	14	.07
10	13	.05
11	5	.02
12	2	.01
13	2	.01
17	2	.01
14	1	.003
15	1	.003
16	1	.003
23	1	.003

Table 3.1: Distributions of Hashtag Co-occurrence in Tweets

Out of the total number of 2,500 unique hashtags, 1,470 hashtags were used only once, while 1,030 were used twice or more. The #metoo was the top used

Hashtag Name	Counts
#metoo #timesup	1,033
#metoo #metoo	171
#metook12 #changetheratio #diversity	103
#metook12 #metoo #timesup	89
#metoo #metook12	83
#metoo #ptsd #mst #usarmy	76
#metook12 #metoo	74
#metoo #aidtoo	65
#metoo #tictocnews	62
#metoo #healmetoo	57

Table 3.2: Top Ten Frequently Used Co-occurrence Hashtags

hashtag in this dataset (n=26,105), followed by #timesup (n=1,952). The top most frequently used hashtags can be found in Table 3.3.

Hashtag Name	Counts	Percentage	Cumulative Percent
#metoo	26,105	66.30	66.30
#timesup	1,952	4.96	71.25
#metook12	1,190	3.02	74.27
#metoomilitary	266	.068	74.95
#changetheratio	218	.554	75.50
#diversity	185	.470	75.97
#sexualharassment	175	.444	76.41
#metoowhatnext	153	.389	76.81
#metooonpbs	148	.376	77.18
#ptsd	135	.343	77.52

Table 3.3: Top Ten Frequently Used Hashtags

There were 169 #metoo hashtag derivatives; 50 were composed of the suffix of "too" (*e.g.,* '-too') while the remaining 119 had the prefix of "metoo" (*e.g.,* 'metoo-'). The most frequently used hashtag derivative was #metook12 (1,190) followed by the second most used #metoomilitary (266). The top most frequently used hashtag derivatives can be found in Table 3.4.

For a qualitative assessment, we see from the individual hashtags multiple topics present in this dataset: race, politics, organizations, ability, nationality, gender, ethnicity, age, and events/movements.

Hashtag Name	Counts
#metook12	1,190
#metoomilitary	266
#metoowhatnext	153
#metooonpbs	148
#aidtoo	127
#metoophd	125
#metoochat	97
#metoomvmt	85
#metoocongress	66
#kidstoo	59

Table 3.4: Top Ten Frequently Used #metoo Hashtags Derivatives

3.3.2 RQ1b: Temporality

Data Analysis. Aggregate the number of tweets by day and examine the volume and proportion of tweets with hashtags (co-occurrences, derivatives) over time; visualization of the trends and possible fragmentations within the data.

For this question, I used the original snowball sampling data collected from Twitter. This dataset consists of 1,997 tweets containing the #metoo hashtag and or #metoo hashtag derivatives published by 645 unique users between October 15, 2017 and October 21, 2017 for the first week of the #metoo movement.

Figures 3.2 and 3.3 shows the time series plot of the first week of the #metoo movement viral event. A total of 2,544 hashtags were used over the course of this 7-day period. There were 245 unique hashtags used. Out of the total amount of hashtags used, 2009 (78.6%) were #metoo hashtags, 23 were #metoo hashtag derivatives (1.06%), and 516 (20.2%) were neither. Figure 3.5 shows the breakdown of hashtags per day in the first week when the #metoo went viral.

There were 12 unique hashtag derivatives of the 23 total used during the first week. The list of hashtag derivatives along with the counts can be seen in Table 3.6. The number of co-occurrent hashtags peak after day #metoo goes viral and declines

69

steadily throughout the week with a slight bump the second to the last day of the first week.

In regards to the number of co-occurrences in the data, the total number of co-occurrent hashtags were 368. Out of this amount, there were 222 unique co-occurrences of hashtags. In Table 3.7 is a list of dates and the frequency of co-occurrent hashtags. There is a large amount of co-occurrent hashtags on the second day (10-16-2017) after the #metoo went viral. Afterwards, there is steady trend of decline throughout the week. In Table 3.8 is a list of number of the top ten frequently used co-occurrences in the first week. The only co-occurrent hashtag derivative present is the #metoo #healmetoo.
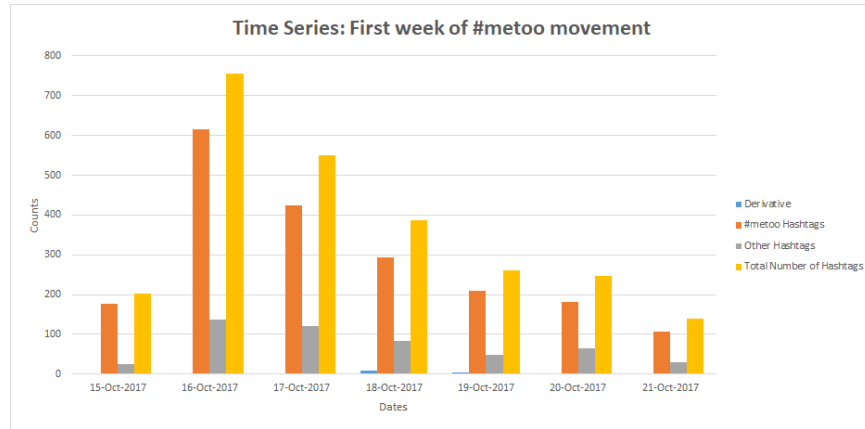
Figure 3.2: Time Series (Regular Scale)

Date	Accounts	Derivative	#metoo Hashtags	Other Hashtags	Total Hashtags
10-15-2017	95	0	177	26	203
10-16-2017	**315**	3	**615**	**138**	**756**
10-17-2017	229	3	425	122	550
10-18-2017	170	**10**	293	83	386
10-19-2017	121	4	210	48	262
10-20-2017	122	1	181	66	248
10-21-2017	72	2	108	29	139

Table 3.5: Numbers for First Week #metoo
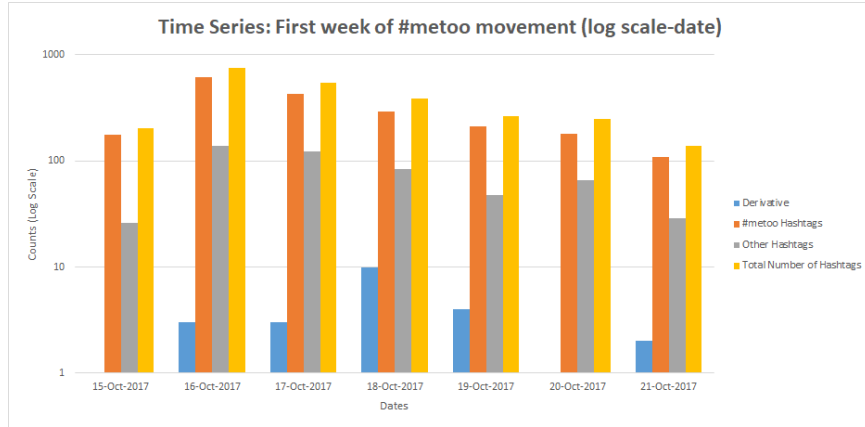
Figure 3.3: Time Series (Log-Scale)

Figure 3.4: Times Series for First Week of #metoo Movement.

List of Hashtag Derivatives	Counts
#metoomovement	5
#wetoogether	5
#ustoo	2
#mentoo	2
#metoomvmnt	1
#himtoo	1
#aintiawomantoo	1
#metooimsorryiwillnowwhat	1
#ptsdtoo	1
#womenabusetoo	1
#youtoo	1
#wetoo	1
#toomanytimes	1

Table 3.6: List of Hashtag Derivatives for First Week

Date	Number of Co-Occurrent Hashtags
10-15-2017	23
10-16-2017	**107**
10-17-2017	81
10-18-2017	65
10-19-2017	35
10-20-2017	38
10-21-2017	19

Table 3.7: Frequency of Co-occurrent Hashtags for First Week

Hashtag Name	Counts
#metoo #himthough	13
#metoo #waitingforjustice	12
#metoo #befiercebook	10
#metoo #wwv17	8
#metoo #howiwillchange	8
#metoo #changetheratio #diversity	8
#metoo #whatilearnedtoday	7
#metoo #silentepidemic #voiceforvoiceless #erinslaw #psa	6
#metoo #ibelieveyou	6
#metoo #healmetoo	5

Table 3.8: Top Ten Frequently Used Co-occurrence Hashtags During the First Week of #metoo

3.3.3 RQ1c: Participant Demographics

Data Analysis. Scrape the Twitter bios for race, gender and sexual orientation; quantitatively and qualitatively analyze the scraped timelines of the selected people to understand their activity during the movement.

Out of the 645 unique users, there were only 339 unique counts that were individual people. Therefore for this question I only focused on these people, and not media or organizational Twitter accounts. The presence of demographic information on the Twitter's users can be found in Table 3.9. It appears most Twitter users in this sample did not allude to their identity and positioning on their Twitter profile.

Biographic Info	Yes	No
Race-Ethnicity	118 (35%)	**221 (65%)**
Gender Orientation	1 (1%)	**338 (99%)**
Sexual Orientation	14 (4%)	**325 (96%)**
Pronoun-usage	44 (13%)	**295 (87%)**
Trans-gender	3 (1%)	**336 (99%)**

Table 3.9: Breakdown of Users ($n = 339$) Demographic

Below are the cross-tabulation results for Presence of (Race-Ethnicity (RE), Gender Orientation (GO), Sexual Orientation (SO), Pronouns usage (PNU) , Trans-

gender (TG)) Bio-graphic data in twitter profiles with hashtag categories (see Table 3.10). Twitter users who did not self-identify their positionings had overwhelming greater numbers in derivative, #metoo hashtag and overall hashtag usages. Those that self-identified their race and ethnicity had the next greater usage.

RE	GO	PNU	SO	TG	Derivative	metoo hashtag	Total Hash num
no	no	no	no	no	4	587	822
no	no	no	no	yes	0	1	1
no	no	no	yes	no	0	10	10
no	no	yes	no	no	0	77	80
yes	no	no	no	no	3	219	263
yes	no	no	no	yes	0	2	4
yes	no	no	yes	no	0	12	14
yes	no	yes	no	no	1	53	63
yes	no	yes	yes	no	0	3	3
yes	yes	yes	yes	no	0	1	1

Table 3.10: Cross-Tabulation for Biographic Data and Hashtag Categories ($n = 339$)

3.3.4   RQ1c: Canonical Correlation Analysis

Canonical correlation analysis is a machine learning algorithm for exploring the relationships between two multivariate sets of variables [135]. In this study, the variate sets are bio-information and hashtag characteristics. Generally speaking, canonical correlation is to used to describe the relationship between a dependent and independent set of variables, but with this study methods, we do not necessarily think of either set as independent or dependent - although we do not rule out such an approach, either.

To perform the canonical correlation analysis, I broke up the data into two different variate sets. The first variate is comprised of hashtag variables and the second variate is comprised of biographic variables, represented as X and Y, respectively.

Looking at the correlations among the X and Y variables independently, we

can observe that the X variables are highly correlated to one another, whereas there is low to moderate correlations amongst the Y variables. This can be seen both numerically and visually in Figure 3.5. Thus, we use 3 dimensions of hashtags and 5 dimensions of bio data variables in the analysis. Looking at the correlation among the X variables (hashtag variables) there are ranges form low to high correlations. For the Y variables, there are low correlation among the variables and some negative correlations present. When looking at the cross correlations between X and Y variates, these are moderately negative correlations.
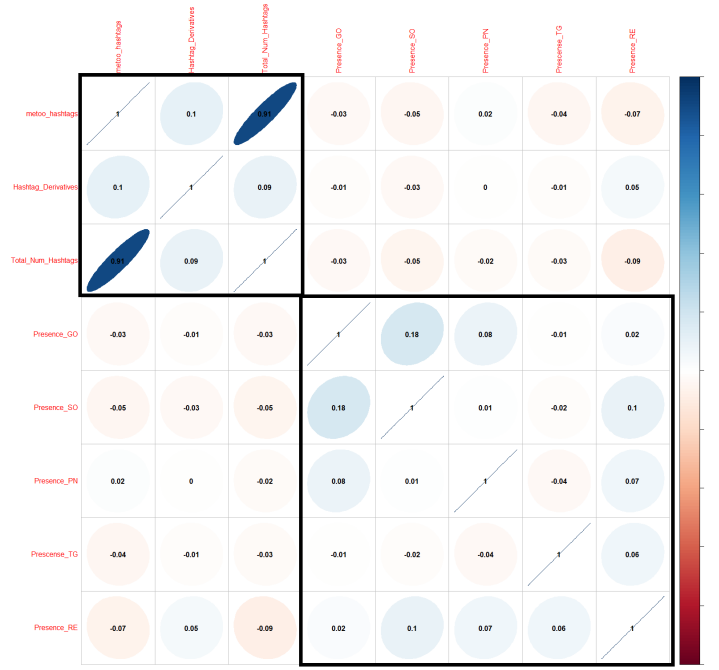
Figure 3.5: Numeric and Visual Correlation between X and Y variates

The first three canonical variate pairs selected have correlations of 0.1160, 0.0915 and 0.0597 respectively (see Table 3.11).

The eigenvalues are functions of the squared canonical correlations. For instance, the largest eigenvalue is equal to the "largest squared correlation/(1-largest squared correlation)." The size of an eigenvalue captures the proportion of the vari-

CV1	CV2	CV3
0.1160	0.0915	0.0597

Table 3.11: Canonical Correlations amongst 3 variates

ance of a particular pair of canonical variates that can be explained by the corresponding canonical correlation. Hence, the first canonical function explains 53.15% of the variation of the first pair of canonical variables, while the second pair of the three canonical functions together explain 86.04% of the total variation in the first two pairs of canonical variates. The 3 resulting canonical correlations and their corresponding eigenvalues are given in Table 3.12.

Dimension	Eigenvalues	Cum. per.	Canon cor.	Sq. cor.
1	0.0136	53.15	0.1160	0.0135
2	0.0084	86.04	0.0915	0.0084
3	0.0036	100.00	0.0600	0.0036

Table 3.12: Eigenvalues and Canonical Correlations

The raw canonical coefficients shown in Table 3.13. The raw canonical coefficients are interpreted in a manner analogous to interpreting regression coefficients (*i.e.,* for the variable hashtag derivative, a one unit increase in hashtag derivatives leads to a 2.047 decrease in the first canonical variate of set 1 when all the other variables are held constant. Another example, having #metoo hashtag in a tweet leads to a 0.3212 decrease in the dimension 1 for the Hashtag Characteristics with the other predictors held constant.

Next, we compute the correlations between the variables and the canonical variates (also the loadings of the variables on the canonical dimensions). Usually, the number of canonical dimensions are the same as the count of variables in the smaller set. The number of canonical dimensions that are significant in explaining the relationship between the two sets of variables may, however, be smaller than the

xcoef	[,1]	[,2]	[,3]
Hashtag Derivative	-2.0470	0.9584	6.2144
metoo hashtags	-0.3212	-0.4588	-0.0624
Total number of hashtags	0.3033	0.1775	0.0736

ycoef	[,1]	[,2]	[,3]
Race-Ethnicity	-1.6058	0.8508	1.0666
Gender Orientation	-0.3002	3.9796	-1.9703
Sexual Orientation	-0.9776	1.0387	-3.8271
Pronouns-usage	-1.5841	-2.1185	-0.9634
Transgender	0.4298	4.5915	-3.7221

Table 3.13: Raw Canonical coefficients for the Hashtag Variables and Bio Data

number of variables in the smaller data set. In our case, there are 3 dimensions. This can be seen in Table 3.14.

corr.X.xscores	[,1]	[,2]	[,3]
Hashtag Derivative	-0.2658	0.05515	0.9625
metoo hashtags	0.4349	-0.8573	0.2755
Total number of hashtags	0.7498	-0.5705	0.3350

corr.Y.xscores	[,1]	[,2]	[,3]
Race-Ethnicity	-0.0950	0.0377	0.0231
Gender Orientation	-0.0141	0.0264	-0.0179
Sexual Orientation	-0.0323	0.0262	-0.0439
Pronoun-usage	-0.0685	-0.0616	-0.0177
Trans-gender	0.0018	0.0435	-0.0172

corr.X.yscores	[,1]	[,2]	[,3]
Hashtag Derivative	-0.0308	0.0050	0.0575
metoo hashtags	0.0504	-0.0784	0.0165
Total number of hashtags	0.0870	-0.0522	0.0200

corr.Y.yscores	[,1]	[,2]	[,3]
Race-Ethnicity	-0.8193	0.4124	0.3868
Gender Orientation	-0.1220	0.2883	-0.2990
Sexual Orientation	-0.2786	0.2865	-0.7357
Pronoun-usage	-0.5902	-0.6736	-0.2959
Trans-gender	0.0153	0.4760	-0.2891

Table 3.14: Canonical Loadings

Next is to determine the statistical significance of the dimensions, see Table 3.15. The stat result yields value of the statistic for the statistical statistic in ques-

tion. The approximate value is the value of the corresponding F-approximation for the statistic. DF1 stands for the numerator degrees of freedom for the F-approximation. The DF2 denominator degrees of freedom for the F-approximation. The last column is the p-value. The results are using four significance tests used to compute where the canonical correlations are significantly results, the results may vary slightly depending on the test statistic implemented. Therefore, the significant tests were used to yield p-values of approximately.

Significance tests were used to compute which canonical correlations are statistically significant, although they may vary slightly depending on the test statistic implemented. Therefore, the four significant tests used were Wilks, Hotellings, Pillais and Roy's[6] which yielded p-values of approximately $p > 0.05$. It is determined that the significance of none of the dimension, which is not statistically significant in this case since $p > 0.05$. Therefore, it is a failure to reject the null hypothesis. Since the null hypothesis cannot be rejected , we will cease from going further to analyze this data.

3.4   Discussion

Research Study 1's goal was to understand the usage and proliferation of co-occurrence of both hashtags and their derivatives in the #metoo movement. To move toward this goal, the formalization of the definition hashtag derivative had to be established. Therefore, a hashtag derivative was defined as a hashtag whose composition varied on but strongly reflected the original hashtag. Besides, the definition of hashtag co-occurrence had to be defined, which was the presence of two hashtags present in a single tweet. Upon formalizing these two definitions, the next steps were

---

[6]On principle, Roy's Largest Root takes only rho[1] into account, hence one p-value is calculated only.

Dimensions	Wilks	F-Approx	DF1	DF2	p-value
1 to 3	0.9748	0.5661	15	914	0.9016
2 to 3	0.9881	0.4986	8	664	0.8576
3 to 3	0.9964	0.3875	3	333	0.0752
Dimensions	HLT	F-Approx	DF1	DF2	p-value
1 to 3	0.0257	0.5639	15	989	0.9032
2 to 3	0.0120	0.4983	8	995	0.8580
3 to 3	0.0036	0.3983	3	1001	0.7543
Dimensions	PLT	F-Approx	DF1	DF2	p-value
1 to 3	0.0254	0.5685	15	999	0.9000
2 to 3	0.0120	0.5019	8	1005	0.8554
3 to 3	0.0036	0.4013	3	1011	0.7521
Dimensions	RLR	F-Approx	DF1	DF2	p-value
1 to 1	0.0135	0.9083	5	333	0.4757

Table 3.15: Test of Canonical Dimensions: Significance test: Wilks, Hotellings, Pillais and Roy's

to explore the categorization of these hashtags and temporal patterns and investigate the relationship between Twitter bios and hashtags.

3.4.1 RQ1a: Hashtag Frequency

To answer the question, "Which categories of hashtags are frequently used in the sampled dataset?", first had to investigate which hashtags frequently occurred in the data set. In doing this, we tallied the unique number of hashtags, the co-occurrence of hashtags, and hashtag derivatives. The most frequently used hashtag was #metoo. This finding is unsurprising. The data collection criteria were based on the #metoo and the 'me too' phrase, which Milano used in her initial tweet to inspire the hashtag, with her celebrity powering the hashtag's viral popularity and discussion. The #metoo hashtag understandably dominates in frequency compared to other hashtags in the dataset. The second most frequently used hashtag was #timesup. The #timesup was Hollywood's spin-off from #metoo to focus more on

Hollywood's specific fight against sexual assault and violence.

Similarly, other researchers [302] observed the #timesup hashtag was the most prevalent in their corpus next after #metoo, even though they explored social movement organization mobilization during the #metoo. After the #timesup hashtag, we begin to see hashtag derivatives appear, such as #metook12[7] and #metoomilitary, which call attention to sexual assault and violence occurring in schools from kindergarten to senior high school [162] and military, respectively. #metoomilitary was a hashtag that illuminated the pervasiveness of sexual assault and violence in the military. Early January 2018, Service Women's Action Network (SWAN) used the hashtag #metoomilitary to identify themselves as they protested outside of the Pentagon. Arnold [9] points out that even though the #metoo hashtag made a broader impact for awareness, the hashtag "lacks a focus on these service members." They also examine the long history and policies of sexual assault and violence in the United States military and marked the beginning of the issue when women were granted military status in 1942. #changetheratio is a social movement hashtag founded by lawyer and writer Rachel Sklar to address the differing ratio between men and women in the workplace in 2010 [154]. This hashtag resurfaced again with a discussion of #metoo [147]. Other hashtags such as #diversity and #sexualharassment were most frequently used in the #metoo corpus, a similar finding as Xiong and colleagues [302]. #metoowhatnext is a hashtag derivative that addresses the legal shortcomings and creates a strategic plan for Strengthening Workplace Sexual Harassment Protections and Accountability and details recommendations for reform [238]. #metooonpbs was a special on PBS that discussed whether or not the balance of power could be shifted, exposing the cultural biases related to patriarchy, equal pay, corporate culture hosted

---

[7]Stop Sexual Assault in Schools (SSAIS) created the #metook12 to bring awareness to the sexual assault and violence occurring in school from kindergarten to senior high school [162].

by Zainab Salbi, an Iraqi women's rights activist. The last hashtag #ptsd is an abbreviation of post-traumatic stress disorder, where people may experience sexual assault and violence. The large count totals involve the main discussion category of sexual assault violence but fine graining to the category of class with #timesup and category of age with #metook12. After looking at the frequency of hashtags, we discovered remaining hashtags categories encompass more broadly shift to incorporate other categories of social positions such as #diversity, yet do not explicitly define diverse for whom.

Looking at the dynamics of hashtag co-occurrences, most occurrences were a single hashtag in a tweet. Other researchers have observed that many people used the #metoo hashtag to share their stories and positionings in relation to the hashtag [213, 271]. This observation can be further extended when people used multiple hashtags in a single tweet to emphasize a particular topic or subject within the larger #metoo conversation. In Table 3.2, we find a repeat of frequently used hashtags reappear again as co-occurrent hashtags along with the #metoo. Once again, similar topics and categories reappear, focusing on age (*e.g.,* #metook12), occupation (#usarmy), and social location (#timesup). Once again, several of the co-occurrence combinations surrounded the sexual assault and violence in the military, along with calling out the pervasiveness in primary, secondary, and senior high schools.

When focusing on the various hashtag co-occurrences and their frequencies, I tracked appearance order because of how it reflects users' conscious and subconscious priorities (*i.e.,* #metoo #metook12 and #metook12 #metoo); unsurprisingly, the most frequent co-occurrence was the #metoo #timesup combination, due to the influence of Milano and her Hollywood-oriented priorities. When looking at the most frequently used hashtag co-occurrences in Table 3.2, we see the presence of the co-

occurrences are related to organizations[8]. Notably, some interesting we observe some of the frequently used hashtags reappear again in co-occurrent hashtags. Some additional hashtags appear alongside these co-occurrences, such as another abbreviation, #mst, which stands for military sexual trauma, which co-appear along with #metoo, #ptsd and #usarmy [131]. The #metoo #aidtoo hashtag combination called out in the wake of allegations of harassment and abuse at nonprofits where it has elicited a momentum of change in some charities [149]. #aidtoo hashtag highlights the sexual assault and violence occurring in the nonprofit/nongovernmental organization sector. Gillespie, *et al.* [118] discussed more "how women have experienced sexual violence in the context of NPOs/NGOs and draw on an intersectional feminist theory lens to highlight the context that enables violence to persist, and which requires more than implementing bureaucratic accountability reforms." [118, p. 1] The #aidtoo hashtag was created by an organization called Devex, citing the aid industry having problems with sexual assault and misconduct [74].

One surprising phenomenon captured here is organizations like Bloomberg using #metoo with their #tictocnews to spread news globally on Twitter. Bloomberg has a Twitter newsfeed called Bloomberg TicToc, which streams global news live. This was renamed to QuickTake in December 2019 mainly to avoid confusion with Tik Tok the Application [268].

Looking at the top hashtag derivatives by frequency of use (Table 3.4), the first five hashtag derivatives as seen earlier (*e.g.,* #metook12, #metoomilitary, #metoowhatnext, #metooonpbs, and #aidtoo). The most frequently used was #metook12. Now we have introduced five additional hashtag derivatives. #metoophd, created by former university professor Karen Kelsky, provides a virtual space for survivors in academia to share their stories. [172] Some hashtag derivatives were encouraging

---

[8]In this first dataset (RQ1a), there are social movement organizations in this corpus.

users to speak out, such as #metoochat. We observe the #metoomvmt hashtag in this list, representing Tarana Burke's work and referencing her work [48]. Congresswoman Jackie Speier from California encouraged women working in Capitol Hill to share their stories [127], hence the #metoocongress. Moreover, lastly, the #kidtoo focuses on kids and sexual abuse. The most frequent derivatives fall into categorical themes involved calling out sexual assault and harassment in spaces involving schools, the military, and charity and direct calls to protect children.

3.4.2   RQ1b: Temporality

I sought to investigate the temporal hashtag patterns in the first week of when #metoo went viral for this snowball sampled dataset; there are 645 unique users among them. The highest usage of hashtags was on the second day after the hashtag went viral. The virality's impact corresponds with the reports of not receiving the full effect until the following day after Milano's initial tweet [265]. The highest usage of hashtag derivatives did not occur until the 3rd day. It is presumed that there was not a heavy usage until much later, given the divergence of the ubiquity of sexual assault. Therefore the temporal patterns for overall hashtag usage appear to be decreasing throughout the week. Unsurprisingly, there are no hashtag derivatives because of the newness of the hashtag.

It should be noted that the #timesup is not present in this corpus during the first week. Within the first week of the #metoo going viral, there were 23 instances of hashtag derivatives, and of those, there were 12 unique hashtag derivatives that appeared in that same week. What is striking about these is that they all have a coherent theme, adhering to the core #metoo idea initially but over time deviating from it to bring different counter- and co-narratives to bear. This

was a similar observation made by [302]. These unique hashtag derivatives were: #metoomovement, #wetoogether, #ustoo, #mentoo, #himtoo, #aintiawomantoo, #metoomvment, #metooimsorryiwillnowwhat, #ptsdtoo, #womenabusetoo, #wetoo and #youtoo. These unique derivatives are calling attention to other positionings besides the dominant positioning of the #metoo movement as an affluent, white movement. Out of these unique hashtags, the ones we see on the second day calls for solidarity (*e.g.*, #wetoo, #youtoo), which then transitions overtime to highlight other positions, including mental health and wellness (*i.e.*, #psdtoo), as well as in #mentoo and #himtoo, to call attention to the fact that men [9] can experience sexual assault. As the week progresses on, it was observed hashtags calling attention to the #metoomovement and #metoomvmnt. These two hashtags are calling attention to Burke's work for the analog 'me too', which is different from #metoo movement started on Twitter, which Burke and other scholars have called attention to the difference between the 'me too' saying and '#metoo' [48, 116, 171, 186]. Lastly, towards the end of the week, some Twitter users highlighted the counter-narrative hashtag #womenabusetoo to express that there are groups of people who can be perpetrators. We also saw people calling attention to groups left out the #metoo conversation who experience sexual abuse and trauma in the hashtags such as #aintiawomantoo. This hashtag is dubbing off the famous Sojourner Truth's speech in 1851 titled, *Ain't I a woman* where she spoke out for the rights of African Americans and women during and after the Civil War. Boyd and McEwan [38] offer an intersectional critique through the examination on viral version of #metoo perpetuated by Alyssa Milano reified the social construction of inequalities and interlocking systems of oppression for Black and other women of color. Scholars [15,38,88,116,122,198,206,228,274] have

---

[9]Example of this actor Terry Crews who publicly announced that he experienced sexual assault on the Expendables movie set.

discussed the co-option of 'me too' phase from marginalized spaces into a dominant-white space has altered the meaning and context of the initial movement.

Table 3.8, the most frequently top used co-occurrent hashtags were associated with the #metoo in the first week of the #metoo hashtag going viral, compared to the most frequently top used co-occurrent hashtags in RQ1a. These co-occurrent hashtags are involved in reflecting how pervasive sexual assault and violence are in society and solidarity, which coincides with other scholars [59,210,302] who acknowledge their research of the unity in the movement. The only used co-occurrence hashtag with a derivative was #healmetoo which co-occurred with #metoo. Other co-occurrent hashtags with #metoo involve self-promoting attention to Fox and Friend's Gretchen Carlson promoting her book *Be Fierce*. The #wwv17 stands for 'week without violence in 2017.' This hashtag is related to the YWCA organization, where 'Week Without Violence is part of a global movement with YWCAs across the country and around the world to end violence against women and girls [287].' The longest chain of co-occurrence of hashtags was five hashtags lobbying every state to pass Erin's Law (e.g., #erinslaw), which requires all public schools to implement a prevention-oriented child sexual abuse program [4].

3.4.3   RQ1c: Demographics

For this research question I investigated if there was a relationship between the characteristics in Twitter Bios and patterns of hashtags. I accounted for the presence five demographics (or social categories) for Twitter Users' bios, which were race-ethnicity, gender orientation, sexual orientation, pronoun usage and trans-gender expression. Most people did not self-identify their position on their profile, refer to Table 3.9. This finding is in alignment with a larger canonical conversation where the

Internet continues to be racialized and therefore users are perceived to be white which implies a laced veil of anti-Blackness sentiment unless otherwise users have chosen to self-disclose their racial identity [148, 218, 263]. This is no different on Twitter and can be further extended to other social categories as defaults unless explicitly expressed: white, male, able-bodied, cis-gender, heterosexual. Thus Twitter supports, maintains and perpetuates "power structures of white supremacist capitalist cisheteropatriarachy." [137] There has been several investigations garnering whether Twitter user's are raced and gendered, yet there has not been work linking characteristics of Twitter users' profiles to hashtag usage.

Sharma [263] explored racialized identities in digital spaces and how these assemblages constitute racialized identities online via hashtags. He argues these 'racialized hashtags' can be a way to trace and even identify propagating messages in the network which can reveal different raced communities. If we can extend these thoughts to this work, by challenging this, hashtag derivatives can identify different communities that go beyond race and elude to other social identities, which hint at intersectional positionings, such as orientation, occupation, religion, to name a few. Furthermore this question sought to investigate whether there is a link between the various hashtag usage to possibly identify Twitter users' social positioning.

Therefore to explore this possible relationship between Twitter Bio characteristics and hashtag characteristics, the utilization of a machine learning of canonical correlation analysis was implemented. The objective was to discern if there was relationship between the biographic data and hashtag categories. There was a failure to reject the null hypothesis as an unintended consequence of the data's sparsity. The dataset had a lot of Twitter users that did not self-identify, resulting in missing data for those users. Similarly for the hashtag categories, there were several people who did not use hashtags in their tweets. Once again, given the sparsity of zeros in

the data led yielded to inability for the null hypothesis being rejected. Though the results yield the null hypothesis being rejected, this still offers the opportunity for future exploration. Hence, I would challenge that revisit the study and be able to collect more users who self-identify along with hashtag usage.

We also discuss how, over the demographics (social categories), in Table 3.9 retrieved from the Twitter profiles, most people did not self-identify their positionality on their Twitter profile. This echos from the larger conversation where the Internet continues to raced and are perceived to be white (littered with anti-Black sentiment) unless otherwise they self-disclose [148, 218]. This is no different on Twitter and can be further extended to other social categories as defaults unless explicitly expressed: white, male, able-bodied, cis-gender, heterosexual. Once again this further extends, Hampton's [137] work on oppression, since there is a given precedence of a preserved and targeted user why would not the technologies be used to in support, maintain and perpetuate the "power structures of white supremacist capitalist cisheteropatriarchy."

Moving to the Cross Tabulation Table 3.10 where there is a breakdown among profile data in relationship to the hashtag derivatives, #metoo hashtags and the total number of hashtags. Those that made the largest contribution to hashtag categories are those who did not cite specific biographical data. When people cited some type of race or ethnicity and no other biographical data was the next largest group that contributed to the large amounts of hashtags, followed those that used pronouns in their bios.

One goal of the Canonical Correlation Analysis (CCA) was to see if there was a relationship between the biographic data and hashtag categories. Given the data's sparsity, which was an unintended consequence, the null hypothesis could not be rejected. The dataset had a lot of Twitter users that did not self-identify, resulting in zero for that missing data. Similarly for the hashtag categories, there were several

86

people who did not use hashtags in their tweets. Once again, given the sparsity of zeros in the data led yielded to inability for the null hypothesis being rejected.

3.5   Summary

In summary, the first research question (RQ1a) adds to the body of #metoo scholarship by intentionally looking for hashtag derivatives in the larger #metoo conversation; we include more breadth and depth to the conversation. With this exploration's breadth and depth, we see that multiple conversations are being discussed about sexual assault and violence in different locations and spaces. As Cole and Atuk [60] point out, even though the #metoo was being used in a (re)tweet fashion to spread the word, the numbers do not include the proliferation of conversations about other hashtag derivatives. They have noted that these conversations are taking place as sub-threats on hashtag derivatives. This creates a further collapse and foundational misconception that the #metoo movement was monolithic, one monomaniacally focused on the sole topic of sexual assault and violence. This monolithic mythologization has led too many observers to the incorrect conclusion that all sexual violence experiences have been impacted in the same manner. I am not the first researcher to point this truth out in their work: that people were excluded from the conversation and/or felt the conversation was meant for a particular type of woman [54, 228, 276, 278]. In short, just because white women cry out "for all" does not mean that everyone else has been received in the same serious manner as they. Continuing with the Intersectionality framework, we observe that social identities are included in the overall themes in the corpus; however, the most frequent usage of these derivatives comes from spaces automatically perceived as white. It is noteworthy, even though the hashtag #diversity appeared in the most frequently used

and co-occurred hashtags (refer to Tables 3.2 and 3.3) does not count and should not be included as an acknowledgment of 'everyone.' Milano's rally call of the 'all women' caused harm towards marginalized communities [38]. This echo's Crenshaw's words saying, "Not only are women of colour in fact overlooked, but their exclusion is reinforced when white women speak for and as women. The authoritative universal voice—usually white male subjectivity masquerading as non-racial, non-gendered objectivity—is merely transferred to those who, but for gender, share many of the same cultural, economic and social characteristics." [78, p. 33] We can further extend Crenshaw's words to include those further invisible and often forgotten survivors of sexual assault violence such as queer, trans, non-able-bodied, elderly communities.

Research question (RQ1b) reveals the prevalence of hashtag derivatives were relatively minimal the first week of the #metoo going viral compared to the main #metoo hashtag. As time progressed throughout the week, we see a variety of hashtags derivatives appear calling attention to solidarity, the original 'me too' movement, discourse of who can experience abuse and how is lift out of the conversation. For future exploration is to explore the weeks after the #metoo hashtag had gone viral and whether these hashtag derivatives are in response to event-related news. Also to further explore the exploration of how names of celebrities became hashtags regardless if the celebrity had an Twitter account or not, appeared to get further amplification. As Modrek and Chakalov [213] point out in their analysis of the first week of #metoo going viral, there was an overrepresentation of white women on Twitter when discussing #metoo. Given this over-representation, it is not an unfair assessment that most of these discussions about #metoo mostly pertain to white spaces and are not inclusive of marginalized communities.

Research question 1c reveals that not too many people from this collected data sample self-disclosed their social identity. It should be noted that the time of my data

collection came well after the the period for which it was collected. Researcher [249] found that Twitter users' profiles are quite fluid and possibly change as result of different events on and off social media, making my bio information not necessarily an accurate representation of bio information at the time. However, it is difficult to say which direction the inaccuracy goes. There are potentially valid reasons for any and all particular classes to remove or add bio information, and in the absence of a solid model of this aspect, we are left with the data we have.

Chapter 4

Research Study 2

In this second research study, I seek to ask two questions for inquiry. The first involves looking at the Twitter users within the selected dataset and investigating their involvement in the #metoo movement. We will also revisit the implementation of canonical correlation analysis on this dataset due to its lesser sparsity than Research Question 1c's dataset. The second question compares the two pathways, each representing different techniques (1) traditional network analysis and (2) intersectional analysis.

The omnipresence (or ubiquity) of networks is everywhere; they occur naturally (*i.e.,* neurons, genomes), socially (*i.e.,* Facebook, Tik Tok, Instagram, Twitter), or artificially (*i.e.,* globalization, transportation). The intricacy of these networks can vary from simple to complex. Networks are studied in various fields such as (but not limited to) the social sciences, biology, mathematics, and data science. By studying networks, we can gain additional insight into how entities relate and operate in relationships through their links (or connections).

Kadushin [167] defines a network as a set of relationships between a collection of objects. These objects, or "nodes," could be people, things, or places, and the links

between these nodes are known as "edges." The network can be considered as directed or undirected, depending on the researcher's central analysis. Network analysis is a study that helps researchers and scientists to understand (or unpack) the change or process of the structural phenomenon of relationships in networks [41]. Within these networks, there could be clusters of nodes that represent communities. Girvan and Newman [120] defined a community as a group of nodes that possesses more connections amongst its nodes than other nodes in the remainder of the network [294].

Community detection is a process of uncovering groups of clusters in a network. The implementation of detecting (or finding) communities in a network uses community detection algorithms. Researchers can deploy community detection algorithms to understand better the groups of nodes and links between and within a community. Given the complexity of a network and the need for analysis, Yang, *et al.* [304] point out that the primary focus of choosing particular community algorithms is based on the algorithm's speed and whether it can be optimized for performance. While it is essential to have an algorithm built for speed and optimized for performance, especially if the network is extensive, this presents an impasse where inclusion and appreciation for complexity are left up to the researcher.

Furthermore, with the Intersectionality framework's utilization, questions emerge inquiring about the complexity-oriented toward dominant nodes in the network. They impede the nodes that represent more nuanced and salient stories. Another place of intersectional exploration stems from how researchers select algorithms rather than speed and optimization, instead adopt an investigative inquiry on how these algorithms are constructing these networks of nodes and creating these links among them. Mainly these algorithms fall under two types of hierarchical clustering: agglomerative and divisive. The term 'hierarchical' from an intersectional perspective is not exactly one to one in a quantitative field.

91

As Bauer reminds us, "the very language used in intersectionality can create confusion for quantitative researchers." [16, p. 12] She continues to expand, "If Intersectionality is to be implemented in quantitative research, then terminology will need to be disaggregated in order to allow for clear communication and to prevent the conflation of identical- or similar-sounding concepts." [16, p. 12] Hence, the general idea, for something to be 'hierarchical' is to hold precedence over another, from an intersectional vantage point, cautions to investigate further explore what the algorithms are doing in this 'hierarchical clustering' of nodes and creating linkages. So aligning with Bauer, the terminology must be clearly defined. Yet, I challenge extending further and urging us to review the mathematical equations on how these community detection algorithms are composed and then ask if this algorithm supports an intersectional framework. Previous researchers and scholars have discussed how algorithms embed mathematical concepts rooted in eugenics and white supremacy [257]. Thus, the researcher must investigate these algorithms mathematically to see how these data are calculated. Simultaneously, the most complex models can still be fundamentally flawed, although if the methodology for pre-processing negates to be intersectional, then the algorithms results are invalid. Therefore, the study's innovation uses the intersectional framework to guide the method, construct a network, and compare it to the non-intersectional network analysis approach.

4.1   Data Collection

For this study, a random sample was collected of approximately 438 users from the original snowball data (3,965), so roughly just over 10% of the corpus. The two seed users, Burke and Milano, were included as well given their integral role in the dataset. The collected Twitter Bio data information was collected in the same

manner as RQ1c (refer to RQ1c-Methodology section) for each of the 438 users. User metadata was collected: number of hashtags, user follower counts, user following counts, retweets, and favorites. Refer to the Appendix on Feature Engineering.

4.2 Methodology

4.2.1 RQ2a: Participant Characteristics

Question. What are the characteristics of the participants using co-occurrence hashtag derivatives on Twitter?

From the collected data, the following features were created: total number of tweets, #metoo activity, #metoo derivatives, #metoo tweets, and other hashtags. The other features collected from the metadata - the number of hashtags, followers, following, retweets, favorites, total number of tweets - were used to analyze the characteristics of these participants based on whether they used co-occurrent hashtag derivatives.

I performed a canonical correlation analysis on this dataset as well, again in a similar strategy as employed in RQ1c, refer to Twitter Bio Participants. The analysis was implemented in RStudio[1] using cancor package[2]. The null hypothesis is: No relationship between co-occurrent derivative use and Twitter profile demographic data. The alternative hypothesis is: There is a relationship between co-occurrent derivative use and Twitter profile demographic data. Above our confidence level of 95% (p-value < 0.05), we will reject the null hypothesis.

---

[1]https://www.rstudio.com
[2]https://www.rdocumentation.org/packages/candisc/versions/0.8-5/topics/cancor

4.2.2 RQ2b: Network Analysis

Question. How does Intersectionality modify the traditional network analysis?

For this question, the network analysis will explicitly focus on the hashtag layer of the collected user data[3]. The purpose of this question is to see how traditional network analysis and intersectional network analysis differ as methodological approaches. Unpacking the similarities and differences help us to understand the relationship between #metoo and the other hashtags featured in this data.

In order to create the network graph, the node and edge lists must first be constructed. This was done in a similar manner to how Tüker, *et al.* [284] created their own co-occurrence hashtag network. From the 438 Twitter users, there were 4,966 total tweets. Of those, 1,441 had hashtag co-occurrences. Tweets with less than two hashtags were removed from the dataset as an first step. Now, we have a corpus with tweets consisting only of co-occurrence hashtags. The illustration shows how the node and edges lists were created from the tweets, in Figure 4.1.

4.2.2.1 Node List Creation

In order to create the node list, the co-occurrence of two or more hashtags in a single tweet represents a link between those hashtags used. Multiple co-occurrences were split into multiple nodes: for instance, a tweet containing three hashtags - #metoo, #hashtag1 and #hashtag2 - would have three nodes instantiated: #metoo, #hashtag1 and #hashtag2. Each node contained a list of features consisting of the name of the hashtag, frequency count of the unique hashtag, and type of hashtag[4]. After this, I was left with a co-occurrence network nodeset consisting of 7,257 tags,

---

[3]We are not focusing on the actual participants due to ethical considerations.

[4]Hashtags were classified either as a 'derivative', a 'person', or 'other.' The #metoo hashtag was classified as 'main.'

out of which 804 were unique hashtags. Therefore, there were a total of 804 hashtags in this node list.
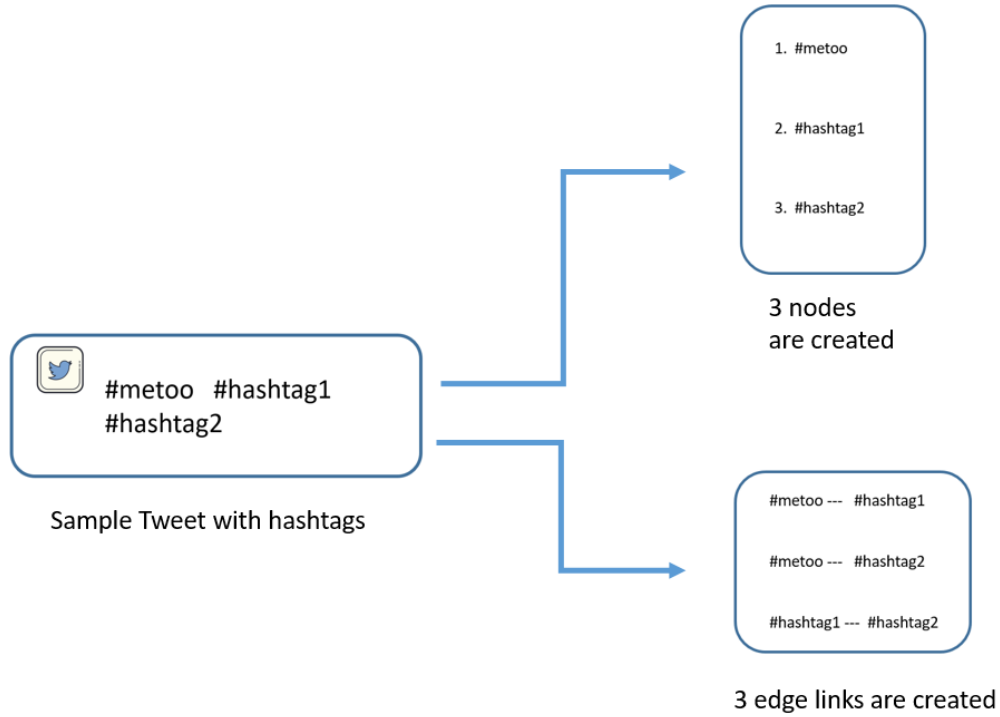
Figure 4.1: Illustration of Node and Edge List Construction from Sample Tweet

4.2.2.2 Edge List Creation

The links were established for the edge list construction by the presence of co-occurrence hashtags in a tweet. Returning to the illustrated example, if a tweet contained three hashtags, #metoo, #hashtag1, and #hashtag2, then there would be three links established. For instance, (1) #metoo and #hashtag1 represent one link, (2) #metoo and #hashtag2 represent the second link and (3) #hashtag1 and #hashtag2 represent the third link[5]. There were a total of 2334 links, 843 unique, in the edge list. The edge list's values were weighted by the count of co-occurrence

---

[5]A link that originates from and terminates at the same node is called self-loop.

repetitions for each pair of nodes.

The node and edge lists were created using Python and exported into .csv files. Upon creating the node and edge list, the network was an undirected[6], simple[7] graph G(N, E), where N represents the hashtags and E represents the links to the co-occurrence weight between the set of tags. This initial undirected-weighted simple network graph can be seen in Figure 4.2 where it has 804 nodes, 843 edges, and a diameter of 72.
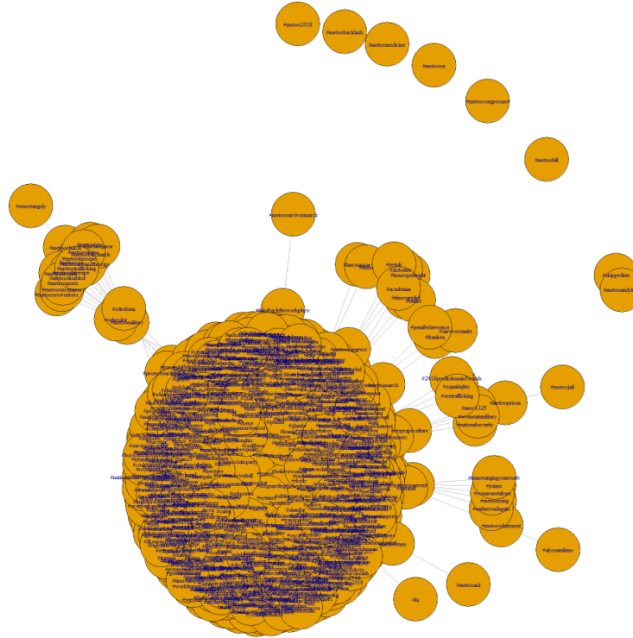
Figure 4.2: Original Network

In Figure 4.3 is an illustration of the two pathways for the network analysis[8]. Next, we will discuss the methodology for each pathway under each dedicated pathway.

---

[6]This is undirected instead of directed because the current subject of this study is not the causality between individual hashtags.

[7]There were some singleton nodes removed because there were tweets that consisted of two repetitions of the same hashtag.

[8]I will discuss later that this is not the only possible methodology; there are other ways to implement network analysis. This is only applicable for this particular problem and exploration. Further discussion on this implementation will be in the Discussion Chapter.

Similarly to Howison, *et al.* [156], we will provide a context and chain of reasoning for each step within its respective pathway. There were three stages for each pathway after creating the network graph. The three stages are: (1) node removal (either small or large in degree), (2) extracting the largest component[9] in the network and (3) using community detection algorithms. Since a community is formed when nodes exhibit some structural or dynamic affinity, we can expect that these nodes share a similar meaning to other clustered nodes [204].

4.2.2.3   Community Detection Algorithms

In the third stage, I ran the community detection algorithms on the network graph. There are several community detection algorithms; in this work, I applied Edge-Betweenness[10] and WalkTrap onto the data.

Per Yang, *et al.* [304], these two community detection algorithms were selected based on the number of nodes in the network and their efficiency in detecting communities for small networks with nodes less than 1000. These algorithms will be based on hierarchical clustering: agglomerative (top-down) and divisive (bottom-up). Agglomerative clustering start with each node in their own "community," and these communities are then iteratively linked together until there is one community containing all nodes [150]. This method starts with an empty graph with no edges, but edges are added iteratively one-by-one to the graph starting with the stronger to weaker edges. We stop dividing a community if the undivided community's total modularity score is greater than the divided community's total modularity score. The divisive clustering is the opposite of agglomerative. This method starts with all the nodes in a community then separated them into smaller and smaller communities

---

[9]Most networks have a single giant connected component that includes most nodes. Most studies of networks actually focus on the giant component.

[10]This algorithm is also known as the Girvin and Newman's algorithm.

until each individual is in their own community. The edges are iteratively removed from the highest weight.

As discussed earlier, guided by the Intersectionality framework, the term 'hierarchical' brings forth different connotations in the data science space. However, it is essential to investigate the definition and mathematical concepts about how the algorithms are implemented.

Edge-Betweenness is a divisive approach to clustering. At the same time, WalkTrap is an agglomerative approach to clustering communities. Based on the description of how each algorithm forms communities, there are aspects where Edge-Betweenness presents a more "intersectional", unlike WalkTrap is more towards "non-intersectional approach" technique. Although, looking at the mathematical construction of these clustering algorithms provides some fascinating insight.

Edge-Betweenness algorithm proceeds as [150, p. 5]:

1. Calculate betweenness[11] for every edge in the graph;

2. Remove edge with the highest betweenness score;

3. Recalculate betweennesss for all remaining edges;

4. Repeat from Step 2

"The final partition of the nodes is selected by calculating Q at every split, and selecting the final number of communities to correspond to the maximum value of Q, where is:

$$Q = \frac{1}{2L} \sum_{i,j} [x_{i,j} - \frac{d_i d_j}{2L}] \delta(i,j) \tag{4.1}$$

---

[11]Betweenness is calculated when you take every pair of the network and count how many times a node can interrupt the shortest paths between the two nodes of the pair [294].

L is equal to the number of edges where the jth edge is represented as l_j and $j = 1, ..., L$.

In the WalkTrap algorithm, there is driving assumption that as one traverses (or "walks") throughout the graph one should get trapped within the communities and therefore the community is found. A transition matrix $\mathbf{P}_{NxN} = p_{i,j}$ is formed where $p_{i,j} = \frac{x_{i,j}}{d_i}$ is the transition probability from n_i to n_j at any step. For a random walk of length m starting at n_i, the probability of ending at n_j is $P^m_{i,j}$. From here a distance measure, $D(i,j)$, between n_i and n_j is calculated:

$$D(i,j) = \sqrt{\sum_{n=1}^{N} \frac{(P^m_{n_i} - P^m_{n_j})^2}{d_n}} \tag{4.2}$$

This is then generalized to a distance measure between communities:

$$D(C_k, C'_k) = \sqrt{\sum_{n=1}^{N} \frac{(P^m_{C_k n} - P^m_{C'_k n})^2}{d_n}}, \tag{4.3}$$

where $P^m_{C_k j} = \frac{1}{N_k} \sum_{i \epsilon C_k} P^m_{i,j}$ is probability of going from C_k to the n_j, where $n_j \notin C_k$, in m steps.

WalkTrap algorithm proceeds as [150, p.6-7]:

1. Begin with N communities (every node is its own community) and calculate the distances $D(\text{i,j})$ for each pair;

2. Use Ward's criterion [12] to merge two communities (minimize the average squared distance between each node and its community):

$$\sigma_k = \frac{1}{N} \sum_k \sum_{i \epsilon C_k} D^2(C_k i) \tag{4.4}$$

---

[12]The Ward's criterion depends on the selection of cluster pairs in order to merge each step which is hinged on the objective function's optimal value [150].

3. Update the distances between adjacent communities.

4. Until all nodes are clustered in the same community, steps 2 and 3 are repeated.

Finally, the best choice of K is selected from the sequence of communities with increasing k based on the maximum modularity.

The Edge-Betweenness algorithm is less involved when it comes to finding community clusters, in comparison to the WalkTrap, which involves matrices, averages probabilities, and Ward's criterion. Some caution is required when using these algorithms due to the impact of varying distances between nodes. In addition, a drawback to be conscious of is the historical misuse of statistical methods in quantifying racial dynamics. Therefore, I will proceed with caution given the uncertain impact on our results.

Non-Intersectional Network Pathway

| Network graph → Eliminate the low degree node(s) → Extract the Largest Component → Community Detection Algorithm(s) → Analysis and investigation of results |

| Network graph → Eliminate the high degree node(s) → Extract the Largest Component → Community Detection Algorithm(s) → Analysis and investigation of results |

Intersectional Network Pathway

Figure 4.3: Illustration of Network Analysis Methodology

4.2.3 Non-Intersectional Network Pathway

In the non-intersectional network analysis, generally researchers encourage removing the edges and nodes with lower weights to give priority to more prevalent nodes within network. The lower weight edges and lower degree nodes represent the

infrequent links and hashtags, respectively, in the co-occurrence corpus. Therefore, from an initial 804 nodes and 843 edges ranging from weights of 0 to 142, we chose to reduce the number of edges by deleting all edges of the weight less than the mean, and likewise deleted the nodes that had a degree less than one[13]. The edge removal left only 146 edges, and then the node removal left 136 nodes with the same 146 edges. This leaves us with a graph that constitutes the largest and only component. These two intermediate graphs for non-intersectional network analysis can be seen in Figures 4.4a and 4.4b.

(a) Intermediate Graph 1 (804N, 146E)   (b) Intermediate Graph 2 (136N, 146E)
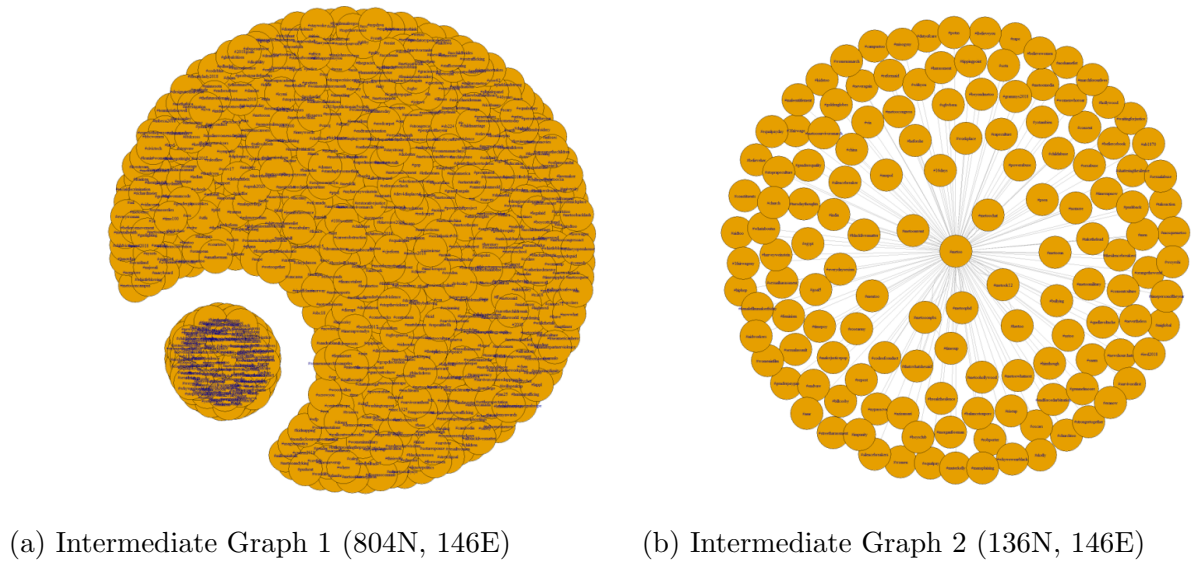
Figure 4.4: Non-Intersectional Network Analysis: Intermediate Graphs

Next, we extract the largest component from the graph, which happens to be the intermediate graph 2 Figure 4.4b). I then use this component and perform community detection algorithms on to the network to see how the nodes are clustered together. Using the Edge-Betweenness and WalkTrap community detection algorithms, we located one and 3 communities, respectively, which can be seen in Figures 4.5a and 4.5b.

---

[13]We did not remove aggressively because that would have made the network extremely sparse.

(a) Edge-Betweenness Graph   (b) WalkTrap Graph
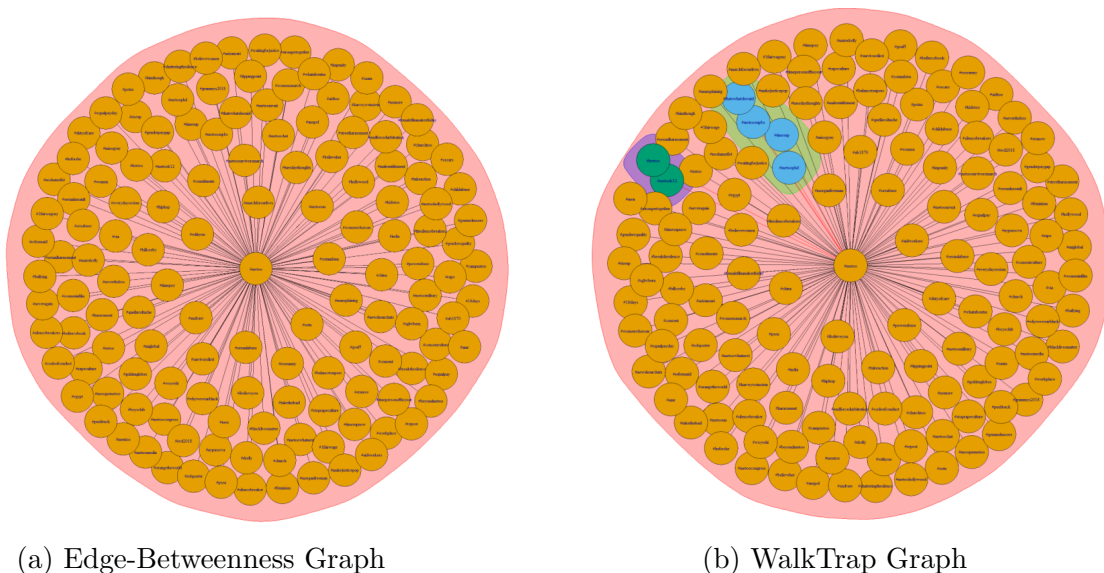
Figure 4.5: Non-Intersectional Network Analysis: Community Detection Algorithms

4.2.4 Intersectional Network Pathway

In the intersectional network analysis, we are led to travel to the margins of the data instead of focusing on the most monolithic node - that is, #metoo. Therefore, we removed the #metoo node from the network, to give more visibility to the other nodes. By removing the largest node, #metoo, this reduced the node strength range from 0 to 18. Also as a consequence this drastically reduced the number of edges from 83 to 65. The rationale behind intersectional network methodology is the ability to visualize the nodes that are not readily seen therefore revealing other nodes present in the network which would not have been viewed previously. This can be seen in Figure 4.6a. What cannot be readily seen in the figure is that there are numerous singleton nodes as a result of removing the #metoo node. In this pathway, since we are focused on the nodes located in margins, we do not remove any edges from the network nor do we filter any nodes. Finally, we extracted the largest component[14]

---

[14]As a reminder, a component is a group of nodes that are connected to each other, but not to the rest of the nodes.

which consisted of 66 nodes and 65 edges. This component was used in the remainder of the analysis as shown in Figure 4.6b.

(a) Intermediate Graph 1 (803N, 83E)

(b) Intermediate Graph 2 (66N, 65E)
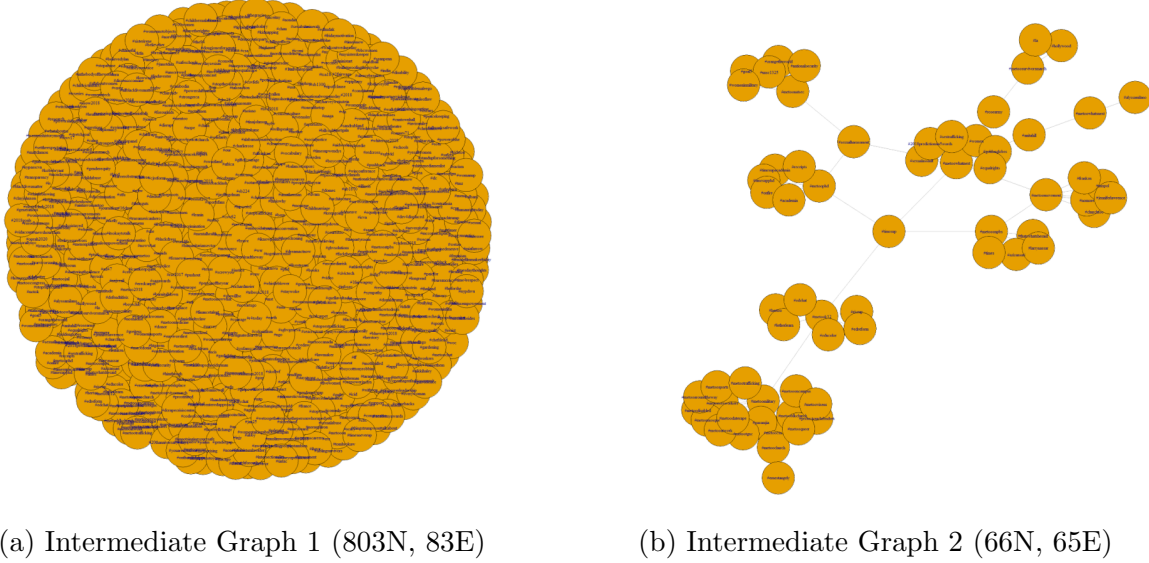
Figure 4.6: Intersectional Network Analysis: Intermediate Graphs

After extracting the largest component, we then perform Edge-Betweenness and WalkTrap community detection algorithms. Using the Edge-Betweenness and Walk-Trap community detection algorithms, we located 8 communities each, which can be seen in figures 4.7a and 4.7b.

R studio was used for network visualizations and analysis. Given the relatively dense structure of the networks, the Fruchterman Reingold layout algorithm[15] was used. As Wang and Liu [292] point out, using particular layout algorithms allows for a more readable visualization where the nodes are pushed apart, preventing crowding and adjusting for the remaining overlap.

---

[15]The Fruchterman-Reingold algorithm is a type of layout algorithm which is force-directed. The concept behind the force-directed algorithm is to visually represent the force between the nodes in the network [107].

(a) Edge-Betweenness Graph   (b) WalkTrap Graph
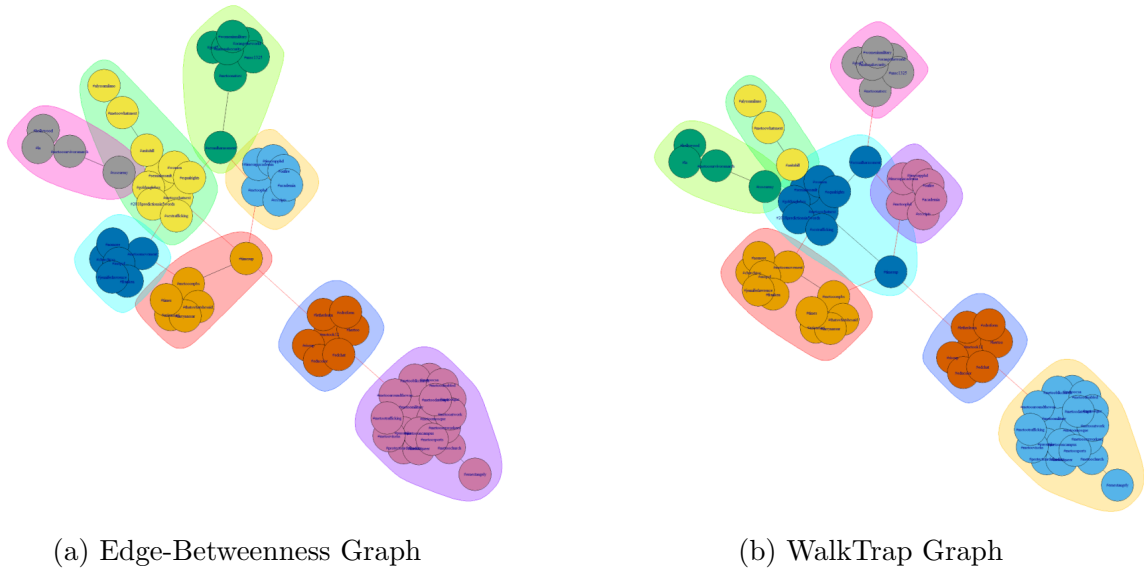
Figure 4.7: Intersectional Network Analysis: Community Detection Algorithm

4.3 Results

4.3.1 RQ2a: Participants Revisited

Data Analysis. Create a descriptive report using co-occurrence hashtag derivatives, then implemented the canonical correlation analysis on the selected Twitter users.

4.3.1.1 Participant Demographics

There were 438 participants in the dataset. The overall participant demographics can be found in Table 4.1. Compared to other demographic categories, gender orientation was the most-disclosed by participants in their Twitter Bios.

There were 50 of the 438 participants (11%) who used co-occurrence of hashtag derivatives in their tweets. The participant summaries can be found in Table 4.2 broken down by feature categories. Those who used co-occurring hashtag derivatives surprisingly led in several feature categories: hashtag derivatives, other hashtags,

Presence of...	No	Yes
Race-Ethnicity	**301 (69%)**	137 (31%)
Gender Orientation	4 (1%)	**434 (99%)**
Sexual Orientation	**427 (97%)**	11 (3%)
Pronoun-usage	**392 (89%)**	46 (11%)
Trans-gender	**436 (99%)**	2 (1%)

Table 4.1: Participant ($n = 438$) Demographics for Research Question 2a

favorites, and retweets.

Features	Non-Participant	Participant	Total
User count	388	50	438
Total Tweets	**1,481,192**	301,004	1,782,196
MeToo Tweets	**3,448**	2,000	5,448
MeToo Activity	**40,469**	9,065	49,534
Total Number of Hashtags	**3,933**	3,458	7,391
Hashtag Derivatives	33	**727**	760
MeToo Hashtags	**2,946**	1,746	4,692
Other Hashtags	954	**986**	1940
Favorites	18,355,283	**82,030,269**	100,385,552
Retweets	5,214,510	**22,189,724**	27,404,234
Followers	**198,474,418**	6,081,263	204,555,681
Following	**798,295**	161,355	959,650

Table 4.2: Feature Comparison of Participants used Co-occurrence Hashtags

In Table 4.3 and Table 4.4, we compare the demographics between the co-occurrent derivative participant and non-participant groups, respectively. In both groups, the overwhelming presence of gender orientation (or disclosure) was present in participants in Twitter Bios compared to the other Presence of other demographic information.

4.3.1.2 Canonical Correlation Analysis

Moving on to the canonical correlation analysis, I separated the data into into two different variate sets. The first variate is comprised of hashtag variables and the second variate is comprised of biographic variables, represented as X and Y,

Presence of...	No	Yes
Race-Ethnicity	**34 (68%)**	16 (32%)
Gender Orientation	1 (2%)	**49 (98%)**
Sexual Orientation	**50 (100%)**	0 (0%)
Pronoun-usage	**40 (80%)**	10 (20%)
Trans-gender	**50 (100%)**	0 (0%)

Table 4.3: Participant ($n = 50$) Demographics for those who used co-occurrence of hashtags

Presence of...	No	Yes
Race-Ethnicity	**267 (69%)**	121 (31%)
Gender Orientation	3 (1%)	**385 (99%)**
Sexual Orientation	**377 (97%)**	11 (3%)
Pronoun-usage	**352 (91%)**	36 (9%)
Trans-gender	**386 (99%)**	2 (1%)

Table 4.4: Participant ($n = 388$) Demographics for non-usage co-occurrence of hashtags

respectively. The results are presented in the numeric and visual correlation chart in Figure 4.8. We can observe that the X variates are highly correlated to one another whereas there is low to moderate correlations amongst the Y variates, and the X and Y variates are slightly cross-correlated.

I used (3) three dimensions of hashtags and (5) five dimensions of bio data variables in the analysis, thus resulting in three canonical correlations among the three pairs of canonical variate.The first three canonical variate pairs selected have correlations of 0.1830, 0.0840 and 0.0515 respectively (see Table 4.5).

CV1	CV2	CV3
0.1830	0.0840	0.0515

Table 4.5: Canonical Correlations amongst 3 variates

The eigenvalues are functions of the squared canonical correlations. For instance, the largest eigenvalue is equal to the "largest squared correlation/(1-largest

106
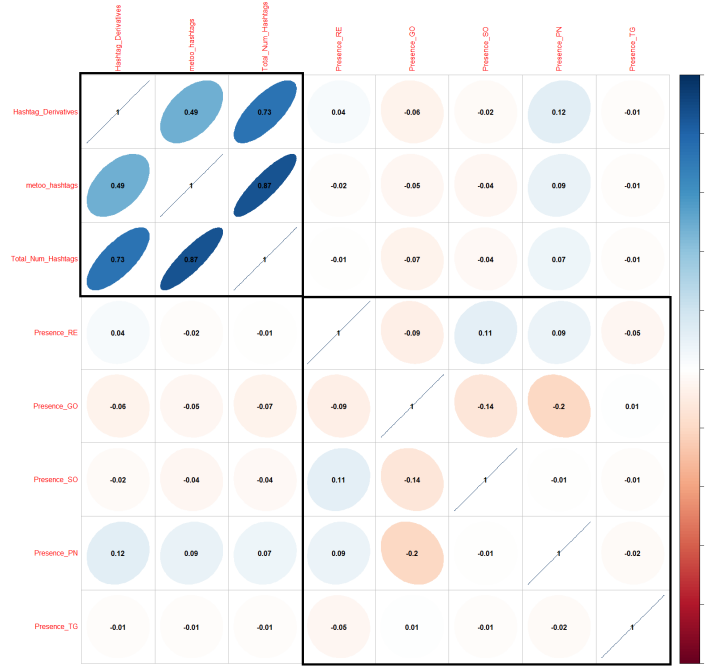
Figure 4.8: Numeric and Visual Correlation between X and Y variates

squared correlation)." The size of an eigenvalue captures the proportion of the variance of a particular pair of canonical variates that can be explained by the corresponding canonical correlation. Hence, the first canonical function explains 78.01% of the variation of the first pair of canonical variables, while the second pair of the three canonical functions together explain 94.02% of the total variation in the first two pairs of canonical variates. The 3 resulting canonical correlations and their corresponding eigenvalues are given in Table 4.6.

Dimension	Eigenvalues	Cum. per.	Canon cor.	Sq. cor.
1	0.0346	78.01	0.1830	0.0335
2	0.0071	94.02	0.0840	0.0071
3	0.0027	100.00	0.0515	0.0026

Table 4.6: Eigenvalues and Canonical Correlations

The raw canonical coefficients shown in Table 4.7 are interpreted in a manner

analogous to interpreting regression coefficients [16]

xcoef	[,1]	[,2]	[,3]
Hashtag Derivative	-0.1222	-0.0478	0.0560
metoo hashtags	-0.0778	-0.0331	-0.0724
Total number of hashtags	0.0458	0.0431	0.0177
ycoef	**[,1]**	**[,2]**	**[,3]**
Race-Ethnicity	-0.3409	-0.6540	1.8043
Gender Orientation	-0.9283	-9.6164	-4.2145
Sexual Orientation	0.8608	-2.8747	1.1984
Pronouns	-3.1887	-0.3233	-0.6632
Transgender	0.1938	-2.4034	1.3983

Table 4.7: Raw Canonical coefficients for the Hashtag Variables and Bio Data

Next, I computed the correlations between the variables and the canonical variates. Usually, the number of canonical dimensions is the same as the count of variables in the smaller set. The number of canonical dimensions that are significant in explaining the relationship between the two sets of variables may, however, be smaller than the number of variables in the smaller data set. In our case, there are 3 dimensions. This can be seen in Table 4.8.

Next is to determine the statistical significance of the dimensions. The results can be seen in Table 4.9, for each of four test statistics - Wilks' Lambda, Hotelling-Lawley Trace, Pillai-Bartlett Trace, and Roy's Largest Root. F-Approx is the approximate value of the corresponding F-statistic. DF1 stands for the F-statistic's numerator, and DF2 the denominator, both in degrees of freedom.

Reviewing the output from the four significance tests, it is determined that the Roy's Largest Root is significant in this case at the at $p < 0.05$ level. This is largely due to how the Roy's Largest Root is computed. This particular statistical test

---

[16]i.e., for the variable "Hashtag derivative", a one-unit increase in hashtag derivatives leads to a 0.1222 decrease in the first canonical variate of set 1, when all other variables are held constant. In another example, having #metoo hashtag in a tweet leads to a .0778 decrease in dimension 1 for the Hashtag Characteristics, when other predictors are held constant.

corr.X.xscores	[,1]	[,2]	[,3]
Hashtag Derivative	-0.0689	0.5250	0.5000
metoo hashtags	-0.4999	0.7250	-0.4746
Total number of hashtags	-0.3729	0.9276	-0.0228

corr.Y.xscores	[,1]	[,2]	[,3]
Race-Ethnicity	-0.0405	-0.0230	0.0450
Gender Orientation	0.0192	-0.0678	-0.0237
Sexual Orientation	0.0250	-0.0298	0.0172
Pronoun-usage	-0.1786	0.0056	-0.0027
Transgender	0.0075	-0.0123	0.0029

corr.X.yscores	[,1]	[,2]	[,3]
Hashtag Derivative	-0.1260	0.04411	0.0257
metoo hashtags	-0.0913	0.0609	-0.0244
Total number of hashtags	-0.0682	0.0779	-0.0012

corr.Y.yscores	[,1]	[,2]	[,3]
Race-Ethnicity	-0.2213	-0.2736	0.8753
Gender Orientation	0.1049	-0.8075	-0.4601
Sexual Orientation	0.1367	-0.3550	0.3361
Pronoun-usage	-0.9759	0.0664	-0.0534
Trans-gender	0.0411	-0.1469	0.0556

Table 4.8: Canonical Loadings

focuses on a single dimension while computing the largest eigenvalue of the resulting matrix. Since the resulting matrix is relatively large, the Roy's root takes on a large value and hence why we rejected the null hypothesis, meaning there is a relationship between the hashtag characters and Twitter Bio data.

Now, there may be a strong cause of alarm at rejecting the null hypothesis, because the first three statistical tests –Wilks' Lambda, Hotelling-Lawley Trace, Pillai-Bartlett Trace– mirror one another. However, if we observe how these three statistical tests are calculated, a common thread among them is that they each take into account the entire matrix, instead of a single value within the matrix like Roy's Root. Furthermore, if were to look at the first dimension of each the first each statistical test, which p-values ranges from 0.2142 to 0.2189, we can see that are close to a 80% confidence interval. This is dramatically more favorable, if we are to re-

109

call the statistical tests from RQ1c (Table 3.15) where the first dimension first three
statistical test confidence interval was 10% and Roy's Root was 60%.

Dimensions	Wilks	F-Approx	DF1	DF2	p-value
1 to 3	0.9572	1.2657	15	1187.422	0.2165
2 to 3	0.9903	0.5257	8	862	0.8377
3 to 3	0.9973	0.3821	3	432	0.7659
Dimensions	**HLT**	**F-Approx**	**DF1**	**DF2**	**p-value**
1 to 3	0.0444	1.2657	15	1186	0.2142
2 to 3	0.0098	0.5255	8	862	0.8379
3 to 3	0.0027	0.3827	3	432	0.7655
Dimensions	**PLT**	**F-Approx**	**DF1**	**DF2**	**p-value**
1 to 3	0.0431	1.2620	15	1296	0.2189
2 to 3	0.0097	0.5282	8	1302	0.8359
3 to 3	0.0026	0.3850	3	1308	0.7638
Dimensions	**RLR**	**F-Approx**	**DF1**	**DF2**	**p-value**
1 to 1	0.0335	2.993	5	432	**0.0114**

Table 4.9: Test of Canonical Dimensions: Significance tests

In Figure 4.9, we focus on the first two dimensions for illustration purposes. These
plots allow us to interpret the variables in two dimensional spaces in relationship to the
Twitter accounts. The first and second canonical variates which are represented using
x and y, respectively. In the first figure on the left, we see the hashtag characteristics
(red) and Twitter-Bio Data (blue). The Twitter Bio variables are tightly clustered
at the center of the diagram, while the Hashtag characteristic variables are further
away from the center. This indicates a strong difference between the Hashtag and
Twitter Bio variates. Moving to the diagram on the right, this plot shows all the
Twitter accounts in relationship to one another. We observe that the distribution is
dramatically different from our left graph, and the data is centered in the bottom-
right corner. This plot is not centered because there are several outlier users present
who land far away from from the clustered mass of Twitter accounts centered at
the origin. These were outliers because they had the most extreme usage for these

hashtag characteristics. As a point of reference, Burke is #416 and Milano is #19. It is observed that both women are not apart of the large mass of Twitter accounts though Milano is much closer than Burke. Burke had more incidents of hashtag derivatives and #metoo hashtags.

Now if were were to imagine these two plots superimposed onto one another, we would see that the clustered mass of Twitter accounts mostly had similar Twitter Bio Characteristics, and thus would lay above the blue cluster. The outlier accounts, by contrast are more closely linked by Hashtag Characteristics such as the usage of #metoo derivatives and hashtags, and would lay near the red cluster.

The canonical correlation will be discussed further in the Analysis section.
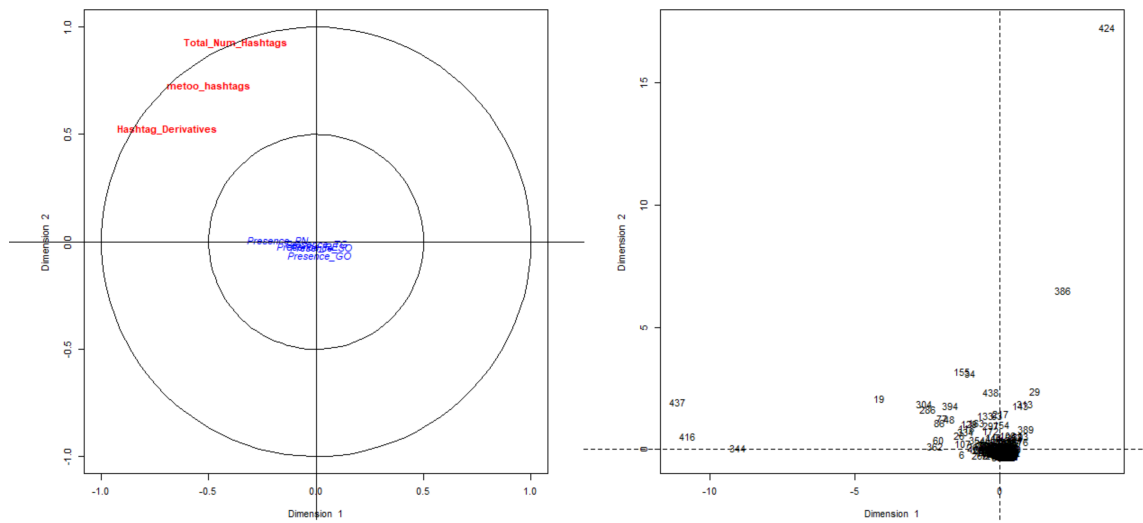

Figure 4.9: Visualization of the variables

4.3.2 RQ2b: Network Analysis

Data Analysis. Compare and contrast two types of community detection algorithms for the hashtag derivative using traditional network analysis and intersectional

approach.

In Table 4.10 we can observe the various characteristics of the non-intersectional and intersectional networks we created. The non-intersectional network has significantly more nodes and edges, while the intersectional network has larger diameter and modularities[17] for each of the community detection algorithms.

In networks, people, and things tend to associate with those whom they perceive as being similar to themselves. When nodes belong to one of two different groups, we can use metrics to capture the interplay between the network structure and the underlying social network [229]. Similarity can be anything such as age group, gender, origin, hobbies, *etc.* This concept is known as homophily. Heterophily, by contrast, assesses the connectedness between nodes of different groups. Krachard and Stern [179] proposed a metric to assess the relative prevalence of between and within group connectedness, in a measure called *E-I Index*. This index is both normalized and polar, so values closer to positive one (+1) indicate heterophily, while values closer to negative one (-1) indicate homophily.

Across all the networks, the *E-I Index* indicates there is strong heterophilous connectedness of the between and within group nodes.

Characteristics	Non-Intersectional Network	Intersectional Network
(Node, Edges)	(138, 146)	(65, 66)
Diameter	2	8
Modularity: EB	0	0.7042
Modularity: WT	0.0528	0.6935
***E-I Index*: EB**	1	1
***E-I Index*: WT**	.9858	1

Table 4.10: Final Network Characteristics

---

[17]In community detection, modularity is a metric used to measure the network structure. Within a network, this type of metric is used to determine the strength of node. Networks with dense connections between nodes tend to have high modularity whereas, less dense connections among nodes have low modularity. [294, 304]

Looking more closely at the community detection algorithms performed on each of the networks, we can observe from the Non-Intersectional Network yielded a smaller number of communities. The community detection algorithms performed on the Intersectional Networks yields larger communities in size but different numbers of hashtags within each community. This can be seen in Table 4.11.

Non-Intersectional: Edge-Betweenness (Divisive)

Communities	Most Important Hashtag	M	D	P	O	Hashtags
1	#metoo	1	21	10	106	138

Non-Intersectional : WalkTrap (Agglomerative)

Communities	Most Important Hashtag	M	D	P	O	Hashtags
1	#metoo	1	17	10	104	132
2	#timesup, #metooonpbs	0	2	0	2	4
3	#metook12, #hertoo	0	2	0	0	2

Intersectional: Edge-Betweenness (Divisive)

Communities	Most Important Hashtag	M	D	P	O	Hashtags
1	#metooonpbs	0	1	2	3	6
2	#metoophd	0	1	0	5	6
3	#metoonatsec	0	1	0	6	7
4	#metoowhatnext	0	2	2	6	10
5	#metoomovement	0	2	2	2	6
6	#metook12	0	2	0	5	7
7	#metoomilitary	0	16	0	3	19
8	#metoosurvivorsmarch	0	1	1	2	4

Intersectional: WalkTrap (Agglomerative)

Communities	Most Important Hashtag	M	D	P	O	Hashtags
1	#metoomovement	0	3	4	4	11
2	#metoomilitary	0	16	0	3	19
3	#metoosurvivorsmarch	0	1	1	2	4
4	#metoowhatsnext	0	1	2	0	3
5	#metoowhatnext	0	1	0	8	9
6	#metook12	0	2	0	5	7
7	#metoophd	0	1	0	5	6
8	#metoonatsec	0	1	0	5	6

Table 4.11: Community Detection Algorithms from each Pathway

If we further look how the community detection algorithms clustered the various hashtags together whether divisively and agglomerativitly, we see more interesting

clustering results when the Intersectional Pathway is taken. Looking further at the clustered hashtag communities, under the Intersectional Pathway, we see similarities along the two types of algorithms (Edge-Betweenness and WalkTrap) among the number of communities yielded and most important hashtags[18] in each community.

Looking for more granularly at clustered hashtag communities, we observed slight differences in the yielding of number of hashtags along with the hashtag was determined the most important in the community. Both algorithms produced the same number of communities and nearly the same important hashtags ranging from plus minus 1 to 5, with the exception of #metooonpbs with six hashtags and #metoowhatsnext with three hashtags, for Edge-Betweenness and WalkTrap respectively. The largest hashtag communities are both #metoomilitary for both community detection algorithms. Looking closer at the types of hashtags within each community both yield the same number of hashtags and same types of hashtags 16 derivatives and 3 other miscellaneous hashtags. However, both algorithms yield different results in regards to smallest communities. The Edge-Betweenness algorithm yields a community that contains four hashtags with the #metoosurvivorsmarch being most important hashtag. Whereas the WalkTrap algorithm yields a community that contains three hashtags with the #metoowhatsnext being its most important hashtag.

4.4 Discussion

4.4.1 RQ2a: Participants Revisited

Canonical Correlation Analysis (CCA) reveals complexities in data in a way that popular methods like linear regression do not [135]. In the second research study,

---

[18]The most important hashtag is determined by using the betweenness centrality measure. This particular measure based on centrality in a graph based on a shortest path.

I decided to perform CCA due to the similarity among features. This dataset was less sparse compared to Research Question 1c, which contributed to a significant result from the Roy's test. My presumption was that if there was an opportunity for it more people would have participated in using more hashtag characteristics associated with derivatives and likewise self-identified the significance tests would have been more fruitful.

One interesting revelation from the participants' demographics was that more people self-disclosed their gender orientation (e.g., male, female, non-binary). This was particularly notable because the same keyword-search method was used for all demographic categories, and yet yielded such a stark contrast. It may be that the wave of discussions about sexual assault and violence, being highly gender-charged topic, had a specific effect on social awareness of gender which perhaps made participants more aware of and alert to be open about their positioning. Another nuance in the results is the ally-ship taken place with the usage of hashtag derivatives to call attention to various communities that may not have an amplified voice on Twitter.

Highlighting the issues of the #metoo movement targets overlapping forms of stigma, prejudice, and oppression in addition to presenting themes on strategies that vulnerable populations have to navigate the confluence of their intersections and the convergence of oppression.

4.4.2   RQ2b: Network Analysis

Network analysis enables the mapping of links and the clustering of hashtags, along with their derivatives, to see how various hashtags clustered together to form communities. With our second research question, we were investigating network analysis by modifying the more "traditional" approach which is non-intersectional[19]

---

[19]*i.e.,* the typical "race as a correlative variable" model.

compared to a pathway guided by an Intersectional framework. Therefore, a first step for the Intersectional network pathway's methodology is to eliminate the highest degree node, in this case #metoo, in order to see nodes on the fringes. Moreover, the purpose for removing the #metoo is that it represents a consuming, monolithic conversation which overshadows others around it. Other data scientists have pointed this out in their work [278], and this correlates to a similar commentary in the social sciences [86, 228, 274] that sexual violence in Hollywood, obscures its ubiquity elsewhere. In contrast to Intersectionality, the Non-Intersectional network pathway does not have any well-established ideography for eliminating this overshadowing, due to its base assumption that all data is inherently equally objective.

We start to get a slightly different emergence of hashtags when we use different community detection algorithms. However we still have the same exposure the similar hashtags from Research Study 1, for instance, #metook12, #timesup, #metooonpbs, with the exception of #hertoo[20], which is first time explicitly mentioning this hashtag. We are safe to assume when we take an Intersectional approach when it comes to node removal we are able to obtain a more granular picture and see other types of hashtags (in this particular research study hashtag derivatives) become more visible.

Edge-Betweenness algorithm is a divisive which looks which detects communities for smaller networks [304]. The communities are clustered together based on the "assumption that edges connecting nodes of separate communities will be given high importance scores." [150, p. 5] This importance is referred to as centrality which is a measurement which determines the importance of a node within its network.

The WalkTrap algorithm uses short "random walks" to find communities in the networks. The random walk concept is the idea that walks stay inside the same

---

[20]I had to investigate qualitatively to explore the relationship between this hashtag and its relationship with #metook12. It was discovered that #hertoo represented...

community being explored which can increase the modularity for the identified community [304].

Utilizing these different community detection algorithms, we are able to see a drastic difference stemming from the whether we are able to remove a largest to smaller degree nodes. Based on the node removal pathway, we do see differences based on which community detection algorithm. However we look within each pathway the difference between each type of community detection algorithm yielded, not so much.

Referring to the literature on what does it mean to have community detection algorithms to yield similar results and whether or not this is a good or bad thing stems from a few factors (1) was is your initial research question and (2) is this the application of intersectionality.

Future direction is starting with a random sampled data and preform an intersectional network analysis to see if there are other connections of hashtags that appear which may(not) be hashtag derivatives to locate other trends or communities.

Returning to the mathematical implications of these clustering algorithms onto the dataset, its appears to be minimal. However it is interesting how the hashtags are clustered together to form particular themes, though some might be nonsensical to call attention. A future exploration is to try other community detection algorithms to see if the similar or different yielding of results.

4.5   Summary

In summary, the participants who were using co-occurrence hashtags had disclosed their gender orientation whether it be by gender binary assignment (*e.g.,* woman or man) or non-binary. This finding is not unusual given the nature of the hashtag conversation. Those participants were more likely to use hashtag derivatives,

use other hashtags within their tweets, as well as retweet and favorites other tweets.

When looking at the machine learning algorithm, there are more users who are similar in Twitter profile characteristics, hence the majority of Twitter users being clustered as the center. However, there are Twitter users who are outliers who have extremely usage of hashtag patterns. This alludes to more involvement in the hashtag #metoo conversation online.

When looking at the network analysis, we observed a differentiation among non-Intersectional and Intersectional framework approaches. Depending on the pathway, this is reflected differently in the community detection output. We discovered more communities and emerging themes from some of the clusters. Also we are reminded by scholars [16, 34, 92] to be attentive with the language and terminology used since Intersectional and quantitative science are not a one to one match. Bear in mind, it is imperative to explore the inter workings of these algorithms to be mindful of the possible sorted history they were formulated from.

Non-Intersectional Network Analysis: Edge-Betweenness Community Detection Algorithms
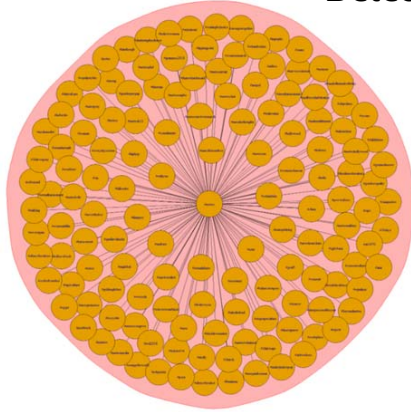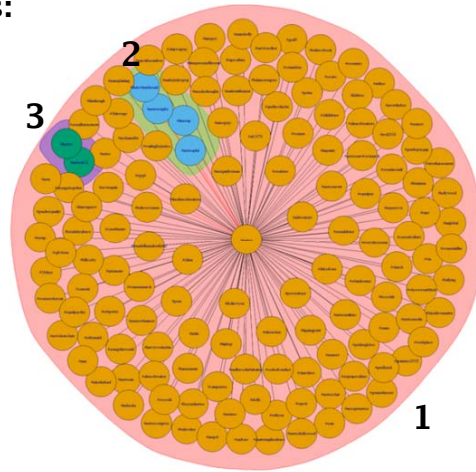

Cluster 1:

"#metoo" "#timesup" "#metoophd" "#metooonpbs" "#metoomvmt" "#aidtoo" "#believeher" "#ustoo" "#metoocongress" "#sexualharassment" "#metoowhatnext" "#befiercebook" "#noshamefist" "#churchtoo" "#metook12" "#hertoo" "#sexualassault" "#strongertogether" "#oscars" "#metooun" "#metoomilitary" "#women" "#metoochat" "#rapeculture" "#kidstoo" "#waitingforjustice" "#harveyweinstein" "#nypasscva" "#quellavoltache" "#goal5" "#rosearmy" "#goldenglobes" "#constituents" "#orangetheworld" "#takeaction" "#saam" "#hollywood" "#miglobal" "#silencebreaker" "#silencebreakers" "#nomore" "#timepoy" "#robporter" "#blacklivesmatter" "#thatswhatshesaid" "#breakthesilence" "#16days" "#1fairwage" "#neveragain" "#maleentitlement" "#azizansari" "#campustoo" "#billcosby" "#metoomedia" "#timesupnow" "#genderequality" "#consent" "#metoosurvivorsmarch" "#balancetonporc" "#reformaid" "#rape" "#himthough" "#mosquemetoo" "#shatteringthesilence" "#sexualabuse" "#withyou" "#hiphop" "#india" "#equalpay" "#dutyofcare" "#feminism" "#psea" "#womenwhoroar" "#grammys2018" "#womensmarch" "#workplace" "#riseup" "#streetharassment" "#1fairwageny" "#auspol" "#muterkelly" "#morganfreeman" "#beyondmetoo" "#repost" "#impunity" "#newshourchats" "#endforcedarbitration" "#nevertheless" "#thesilencebreakers" "#men" "#ibelieveyou" "#stoprapeculture" "#whataboutus" "#timepersonoftheyear" "#childabuse" "#sotu" "#genderpaygap" "#misogyny" "#femalefilmmakerfriday" "#potus" "#yotambien" "#harassment" "#rkelly" "#gemmelmoore" "#powerabuse" "#consentculture" "#codeofconduct" "#bullying" "#marchforourlives" "#survivorsfirst" "#eranow" "#endvaw" "#takethelead" "#whywewearblack" "#via" "#mansplaining" "#ab1870" "#china" "#believewomen" "#mentoo" "#boysclub" "#makejusticepop" "#everydaysexism" "#equalpayday" "#tippingpoint" "#egypt" "#womeninfilm" "#pushback" "#metoohollywood" "#sgbvbern" "#tuesdaythoughts" "#sexabuse" "#aidworkers" "#amr" "#heforshe" "#church" "#iwd2018" "#woyeshi"

Figure 4.10: Non-Intersectional Network Analysis: Edge-Betweenness Community Detection Algorithms

Non-Intersectional Network Analysis: WalkTrap Community Detection Algorithms

Cluster 1:

"#metoo" "#metoomvmt" "#aidtoo" "#believeher" "#ustoo" "#metoocongress" "#sexualharassment" "#metoowhatnext" "#befiercebook" "#noshamefist" "#churchtoo" "#sexualassault" "#strongertogether" "#oscars" "#metooun" "#metoomilitary" "#women" "#metoochat" "#rapeculture" "#kidstoo" "#waitingforjustice" "#harveyweinstein" "#nypasscva" "#quellavoltache" "#goal5" "#rosearmy" "#goldenglobes" "#constituents" "#orangetheworld" "#takeaction" "#saam" "#hollywood" "#miglobal" "#silencebreaker" "#silencebreakers" "#nomore" "#timepoy" "#robporter" "#blacklivesmatter""#breakthesilence" "#16days" "#1fairwage""#neveragain" "#maleentitlement" "#azizansari" "#campustoo" "#billcosby" "#metoomedia" "#timesupnow" "#genderequality" "#consent" "#metoosurvivorsmarch" "#balancetonporc" "#reformaid" "#rape" "#himthough" "#mosquemetoo" "#shatteringthesilence" "#sexualabuse" "#withyou" "#hiphop" "#india" "#equalpay"   "#dutyofcare" "#feminism" "#psea" "#womenwhoroar" "#grammys2018" "#womensmarch" "#workplace" "#riseup" "#streetharassment" "#1fairwageny" "#auspol" "#muterkelly" "#morganfreeman" "#beyondmetoo" "#repost" "#impunity" "#newshourchats" "#endforcedarbitration" "#nevertheless" "#thesilencebreakers" "#men" "#ibelieveyou" "#stoprapeculture" "#whataboutus" "#timepersonoftheyear" "#childabuse" "#sotu" "#genderpaygap" "#misogyny" "#femalefilmmakerfriday" "#potus" "#yotambien" "#harassment" "#rkelly" "#gemmelmoore" "#powerabuse" "#consentculture" "#codeofconduct" "#bullying" "#marchforourlives" "#survivorsfirst" "#eranow" "#endvaw" "#takethelead" "#whywewearblack" "#via" "#mansplaining" "#ab1870" "#china" "#believewomen" "#mentoo" "#boysclub" "#makejusticepop" "#everydaysexism" "#equalpayday" "#tippingpoint" "#egypt" "#womeninfilm" "#pushback" "#metoohollywood" "#sgbvbern" "#tuesdaythoughts" "#sexabuse" "#aidworkers" "#amr" "#heforshe" "#church" "#iwd2018" "#woyeshi"
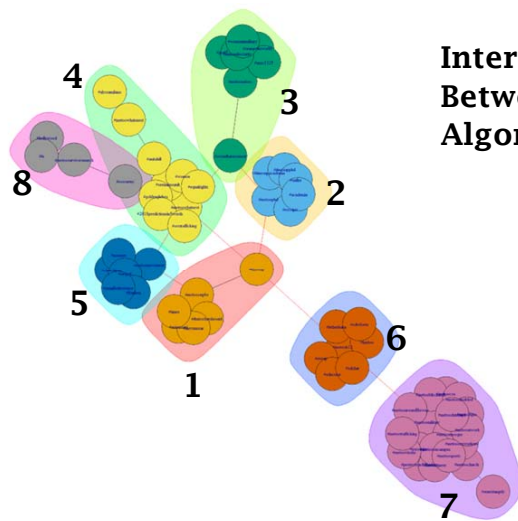

Cluster 2:

"#timesup" "#metoophd" "#metooonpbs" "#thatswhatshesaid"

Cluster 3:

"#metook12" "#hertoo"

Figure 4.11: Non-Intersectional Network Analysis: WalkTrap Community Detection Algorithms

Intersectional Network Analysis: Edge-Betweenness Community Detection Algorithms

Cluster 5:

"#churchtoo" "#metoomovement" "#nomore" "#auspol" "#franken" "#jenniferlawrence"

Cluster 1:

"#timesup" "#metooonpbs" "#thatswhatshesaid" "#azizansari" "#larrynassar" "#times"

Cluster 6:

"#metook12" "#hertoo" "#riseup" "#educolor" "#letherlearn" "#edchat" "#edreform"

Cluster 2:

"#metoophd" "#timesupacademia" "#receipts" "#timesupphd" "#onfire" "#academia"

Cluster 7:

"#metoomilitary" "#metoocsa" "#passmjia" "#metoochurch" "#metooblkchurch" "#metoosexworkers" "#protectourdefenders" "#metoodisabled" "#metoomosque" "#metooaroundtheway" "#metootgnc" "#metooatwork" "#metooqueer" "#metoooncampus" "#metoodaterape" "#metoovixens" "#metoosports" "#metootrafficking" "#ernestangely"

Cluster 3:

"#sexualharassment" "#goal5" "#orangetheworld" "#metoonatsec" "#unsc1325" "#nationalsecurity" "#womeninmilitary"

Cluster 4:

"#metoowhatnext" "#sexualassault" "#women" "#goldenglobes" "#anitahill" "#metoowhatsnext" "#equalrights" "#2018predictionsin5words" "#alyssamilano" "#sextrafficking"

Cluster 8:

"#rosearmy" "#hollywood" "#metoosurvivorsmarch" "#la"

Figure 4.12: Intersectional Network Analysis: Edge-Betweenness Community Detection Algorithms
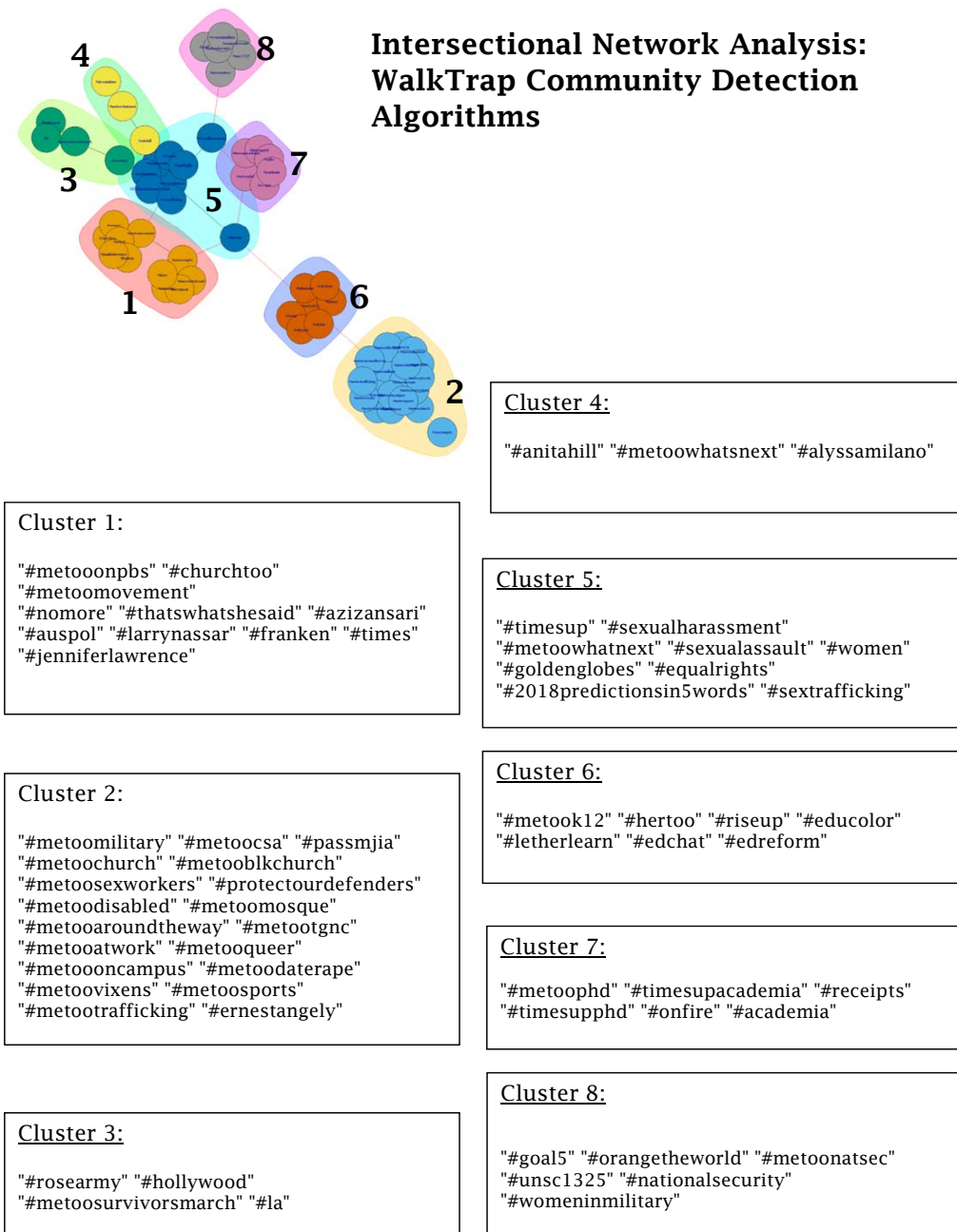
Intersectional Network Analysis: WalkTrap Community Detection Algorithms

Cluster 4:

"#anitahill" "#metoowhatsnext" "#alyssamilano"

Cluster 1:

"#metooonpbs" "#churchtoo"
"#metoomovement"
"#nomore" "#thatswhatshesaid" "#azizansari"
"#auspol" "#larrynassar" "#franken" "#times"
"#jenniferlawrence"

Cluster 5:

"#timesup" "#sexualharassment"
"#metoowhatnext" "#sexualassault" "#women"
"#goldenglobes" "#equalrights"
"#2018predictionsin5words" "#sextrafficking"

Cluster 6:

"#metook12" "#hertoo" "#riseup" "#educolor"
"#letherlearn" "#edchat" "#edreform"

Cluster 2:

"#metoomilitary" "#metoocsa" "#passmjia"
"#metoochurch" "#metooblkchurch"
"#metoosexworkers" "#protectourdefenders"
"#metoodisabled" "#metoomosque"
"#metooaroundtheway" "#metootgnc"
"#metooatwork" "#metooqueer"
"#metoooncampus" "#metoodaterape"
"#metoovixens" "#metoosports"
"#metootrafficking" "#ernestangely"

Cluster 7:

"#metoophd" "#timesupacademia" "#receipts"
"#timesupphd" "#onfire" "#academia"

Cluster 8:

"#goal5" "#orangetheworld" "#metoonatsec"
"#unsc1325" "#nationalsecurity"
"#womeninmilitary"

Cluster 3:

"#rosearmy" "#hollywood"
"#metoosurvivorsmarch" "#la"

Figure 4.13: Intersectional Network Analysis: WalkTrap Community Detection Algorithms

Chapter 5

Discussion

"The master's tools will never dismantle the master's house."

— Audre Lorde [194]

In Chapter 2, I introduced how data science and machine learning researchers have begun to grapple with the implications of the technologies they develop and practices they implemented. A solution to this is adopting a "human-centered" framing in order to appeal to the humane potential of technology. However, questions of "who is regarded as human?" and "are *all* humans valued with humanity?" emerged. If we are to recall, there are white supremacist ideologies embedded into the scientific methods [257], mathematics (*i.e.*, statistics) [257] and technology [28,142]. These ideologies use tactics and exploit people in order lessen their humanity [19,65,143]; hence, the dehumanization of people, and in turn fortifies, strengths and maintains white supremacy while those on the margins are deemed inferior. Therefore, I argue there is some flawed foundational concepts with these "human-centered" framings because not all people consider others as human and furthermore seen as possessing humanity. Unfortunately there are a plethora of examples in our society, reminding communities of there lack of humanity whether it be explicitly and implicitly (not limited to)

Black women (*e.g.,* Breonna Taylor[1], Sandra Bland[2], Rekia Boyd[3]), Trans-gender or gender non-conforming community (*e.g.,* Tony McDade[4], Selena Reyes-Hernandez[5], Brayla Stone[6], Nina Pop[7], Alexandria Winchester[8]), Indigenous community (*i.e.,* I lift up these hashtags that call attention to the injustices and murders that are committed on Native-owned land to the native people by non-native people: #MMIW[9] and #MMIWG2S[10]) to name a few.

There are subliminal messages being internalized to affirm who is and is not be valued as human. Cathy O'Neil [227] discussed in her book, *Weapons of Math Destruction*, how we live in the age of algorithms. These various algorithms make decisions on who gets hired, the amount of loans dispersed and insurance rates. O'Neil argued these mathematical models are not being challenged nor regulated and are seen as objective when making decisions on social issues (*i.e., jail/prison sentences, loans, insurance rates, admissions into schools, job interviews, to name a few*). As Birhane [27] points out there is a vial created by mathematization and formalization which purports as objective and situates as "value-free, neutral, and amoral." In the book *Superior: The Return of Race Science*, author Angela Saini illuminates the tenuous and infamous history of race science and how it reappears in the 21st century. Saini [257] quotes British History professor Gavin Schaffer discusses how the "absence of introspection was rooted in the ability to point fingers at other people for being

---

[1]In Louisville, Kentucky, Taylor was fatally shot in her apartment on March 13, 2020.

[2]In Waller County, Texas jail, Bland was found hung in her cell on July 13, 2015. She was 28.

[3]In Chicago, IL, Boyd was fatally shot by an off-duty Chicago police detective, on March 21, 2012. She was 22.

[4]In Tallhassee, FL, McDade, a Black transgender man, was killed in on May 27, 2020.

[5]In Chicago, IL, Reyes-Hernandez, a 37-year old transgender woman, was killed on May 31, 2020

[6]In Little Rock, Arkansas, Stone, a 17-year old Black transgender girl, was found killed on June 25, 2020.

[7]In Sikeston, Missouri, Pop, a Black transgender woman, was killed on May 3, 2020..

[8]Winchester a 24-year-old Latina trans woman, was killed on Dec. 26, 2020 in the Bronx, New York.

[9]This hashtag stands for Missing and Murdered Indigenous Woman.

[10]This hashtag stands for Missing and Murdered Indigenous Women, Girls and Two-Spirit People.

responsible for the perversion of science." [257, p. 52] Hence there is glossing over a narrative and facts, that these racial theories are the foundations have permeated into others fields not to limited politics, policy, mathematics, biology, social sciences and natural sciences. We can expand further this idea "absence of intropection" to "human-centered" data science and machine learning where researchers need to heed the underlying dehumanization of communities excluded by white supremacy. This can start with being intentional with analysis of the multiplicity of social identities are impacted by social power invested in and benefiting from these constructs.

I also encourage machine learning and data science researchers need to consider how their tools, processes, interpretations–which are far from being neutral—re-create and re-inforce existing structures of power and domination. Now this presents "slight" conundrum. The Discussion chapter with a quote from a Black queer woman poet Audre Lorde, "The master's tools will never dismantle the master's house." That is, the methods, analytical tools, and strategies that are currently available to us unfortunately do not reflect the real-world intersectional complexities of (but not limited to) race, gender, class, sexual orientation [34]. In addition, these algorithms and technologies can embody white, patriarchal, capitalistic, sexist ideologies to promote and dominate ways of thinking and analyzing, to inherently maintain oppressive structures. Gray [129] offers a temporary approach to combat these ideologies. Therefore, we can begin to implement in data science, Gray says, "though we can not use these tools to destroy this culture (in our case algorithms and machine learning), we can instead use them for temporary or partial gains in counteracting the Establishment." [129, p.178-179] This temporary approach offers future researchers and data scientists the opportunity to utilize new frameworks that challenge the way we approach problems. In the meantime, while we wait for the new toolboxes to handle the complexities of these multiple intersections [34], we can start to turn to-

ward methodological frameworks, such as Intersectionality, that are not grounded in whiteness and power, but instead rooted in the vast complexity, further accounts for this introspection that is missing.

Also in Chapter 2, introduced an intersectional framework as a lens, instead of a "human-centered" data science and machine learning, to hold up a mirror to these emerging traditions' commitments to navigate technical efficiencies calling attention to possible dehumanization of communities whom to do fit to the white supremacist ideology. This is a call for being more intentional and methodical in our approach by selecting a machine learning and statistics to understand the data and proactively engage with the mathematics of these tools. This canonical work should be understood as a conversation starter to begin to critique the tradition approaches to choosing tools and analysis and encourage the use of Intersectionality which seeks to address the study design and execution phases the sorts of statistical and systematic biases that typically are only ever discussed—if at all —as "confounding factors" to analyses that are already completed.

By using Intersectionality as a guiding framework, we have to be intentionally mindful of the tools we utilize to analyze (and interpret) data and explore the inter-workings of the algorithm to see how its imposing mathematical hierarchies onto our findings. In order to attend to the complexity and nuances in the data especially when using Intersectionality, a lean towards more multivariate tools to engage in critical research which allows for more complex hypotheses [53, 92].

One might ask, how to incorporate the intersectional lens into the data science process? Now, data science process and Intersectionality are quite epistemologically different in how we think about these two concepts. Data science is very ridge and linear process as it purports the ideology it is objective and agnostic, especially neutral, of any bias. Birhane uses philosophy to argue that data science—which can include

126

the wider computer science field—and its practices takes on this Western, straight, white, cis male ideology as universal-standardized default to how researchers approach scientific inquiry [27]. Therefore, there are misconceptions of the world being static; hence creating constraints that everyone and everything must fit into these confined compartments of thoughts. Furthermore any ambiguity, uncertainty and complexity is not tolerated. Whereas Intersectionality is fluid and dynamic; there is no masquerade. Intersectionality makes the invisible visible provides context to exist and to be acknowledged. This framework starts on the margins where conscious raising how the multiplicity exist in the world.

I challenge the connection between data science and Intersectionality is this concept of reflexivity. To recall, reflexivity allows researchers to call attention to their own practices in the context in relationship to the data science process. Having a framework that consciously is aware and accounts for the inclusively present in society (and the world) from a non-white-cis-hetro-patriarchal-ablebodiedness. Furthermore, this concept goes beyond simple race or gender equality but inclusive of other social identities and positioning.

One may ask, how do we began to visualize Intersectionality and the Data Science Process cohesively together? Therefore, I introduce a visualization how Intersectionality and data science synergistically co-exist together.

In the Figure 5.1, the data science process, as described earlier in Chapter 2: Data Science and Power, is this linear process represented in six steps. It is inside of Intersectionality which is fluid which flows. Both concepts are joined by the reflexivity which is represented in this spiral coil that goes through the data science Process. The implementation of this diagram will be discussed in the next section.
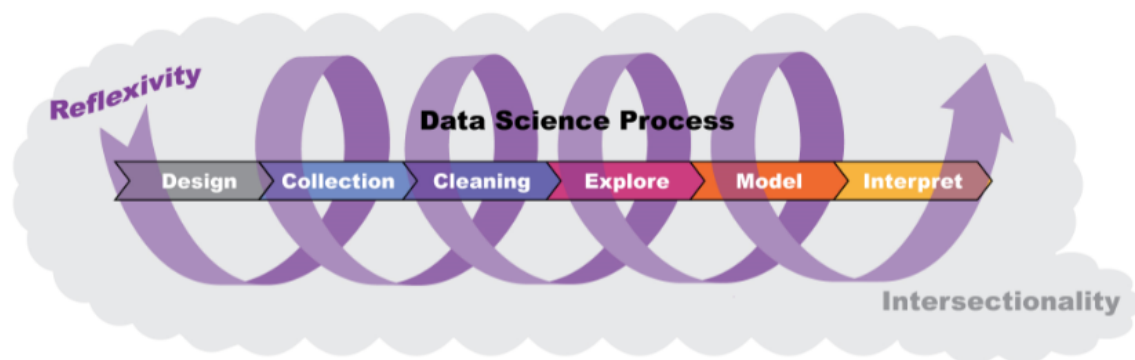
Figure 5.1: Intersectionality meets Data Science

5.1 Quantitative Intersectional Data Science

Chapter 2 discussed how Intersectionality was used quantitatively, mainly in the social sciences. As Intersectionality travels to data science, how do researchers properly implement this critical framework into their work without stripping and bastardizing this historically rooted grounded theory? There are two things to consider.

First, honor its origins rooted in multiply-marginalized context (*i.e.*, Black women). Now, this will be difficult and challenging especially for those who are accustomed to being centered and or approaching tasks from a dominant lens. The beautiful and unique thing about Intersectionality is the ability to step away from knowledge from dominant perspective to actively and intentionally approach a problem from a multiplicative lens. It is from this position, the researcher can begin to see things from a different perspective.

Now one may say, "Can I use Intersectionality to center white, disabled queer people?" No, you cannot because Black, Indigenous, and other disabled queer people not only exist but also warrant attention. As so many researchers have shouted from the words on the page, as Chan and Erby points out "a white gay man may discount race as a factor...assuming discrimination is solely related to affectional orientation,

exemplifying white privilege." [53, p. 1259] Carbado further elaborates on whiteness and Intersectionality stating, "The fact that whiteness is intersectionally unmarked across each of the preceding social positions, as well as others, shores up whiteness as the default and normative racial category through which gender, sexuality, class, and so on are expressed." [50, p. 823] Hence any marginalized social identity is alloyed by whiteness as a privilege.

To effectively apply an Intersectional lens in practice, especially in quantitative research, secondly researchers must acquire both historical and current understanding of privilege and power not only in the data but must go further beyond to see how the privilege and power are in the context of the mathematical equations, techniques and numbers reporting. There are several examples identified in the inequality in the numbers we present.

For quantitative techniques and measures (including data science, machine learning, and artificial intelligence), the linear, binary approach to reduce (or eliminate), the complexity of data to narrowly summarize into neatly binarized variables eliminates the nuanced layers of complexity in the data [34, 53, 91, 92].

As Intersectionality travels to data science, once again we can learn from the social sciences about their successes and missteps. In 2009, Cole [61] proposed a guide consisting of three questions for psychologists who wanted to implement Intersectionality in their research process. These questions were not independent of one another instead built from the previous. Although these questions interrogated social categories usage in psychology research, these questions still probed who was included, what roles did inequality exist in, and challenged looking at similarities rather than differences. I argue we can transfer and adapt these questions to the data science process. Furthermore, we can use the concept of reflexivity introduced by Collins and Bilge [70] as a vehicle to engage in these questions proposed by Cole.

129

Now, it should worth bearing in mind; this is not *THE* guide but a guide to start interrogating the data science process and hold the research accountable for their role in the process. For future research, I encourage researchers and scholars to adapt this table better to fit the proposed research process better, hence further the conversation. Cole argues, "to translate the theoretical insights of Intersectionality into psychological research does not require the adoption of new set of methods; rather, it requires a reconceptualization of the meaning and consequences of the social categories." [61, p. 176] I agree and further extend this argument to the data science process. Reconceptualization is what it means to step away from a prescribed and perfunctory approach to a problem and consider how machine learning algorithms and mathematics impact society.

Therefore, I introduce a similar Table 5.1 as Cole did in 2009 and adapt it to the data science space. I call this Table 5.2 the **Qu**antitative **In**tersectional Da**ta** (QUINTA). This table introduces the implications for the three questions for each state of the data science process. I will use these questions as I review the #metoo case study in the next section.

Research Stage	Who is included with this category?	What role does inequality play?	What are the similarities?
Generation of hypotheses	Is attuned to diversity within categories	Literature review attends to social and historical contexts of inequality	May be exploratory rather than hypothesis testing to discover similarities
Sampling	Focuses on neglected groups (*i.e.* who are we centering and where are those not centered?)	Category memberships mark groups with unequal access to power and resources	Includes diverse groups connected by common relationships to social and institutional power
Operationalization	Develops measures from the perspective of the group being studied	If comparative, differences are conceptualized as stemming from structural inequality (upstream) rather than as primarily individual-level differences	Views social categories in terms of individual and institutional practices rather than primarily as characteristics of individuals
Analysis	Attends to diversity within a group and may be conducted separately for each group studied	Tests for both similarities and differences	Interest is not limited to differences
Interpretation of findings	No group's findings are interpreted to represent a universal or normative experience	Differences are interpreted in light of groups' structural positions	Sensitivity to nuanced variations across groups is maintained even when similarities are identified

Table 5.1: Cole (2009): Implication of the Three Questions for Each Stage of the Research Process

DS Process	Who is included with the data?	What role does inequity play in machine learning and statistics?	What is your positionality with the data?
Design	Pipeline attuned for diversity and inclusion; Literature review attends to the social and historical contexts of inequality of the social issue(s)	Literature review attends to the social and historical contexts of inequality of the proposed techniques	Why are you doing this research? What is the personal benefit for you? How are you profiting from this research? Are you causing harm and erasure? What story are you trying to tell? What is the purpose of your involvement in the research?
Collection	Focuses on neglected groups (*i.e.* who are we centering and who are we marginalizing?)	Techniques used exasperate unequal visibility and further enhance structural inequality	Who, why, and how are you silencing/amplifying those included or excluded from the collection?
Cleaning	Are normalizing and cleaning techniques reinforcing a dominant reframing or are they promoting inclusivity?	Techniques used exaggerate unequal visibility and further enhance structural inequality	Who, why, and how are you silencing/amplifying those included or excluded from the cleaning?
Exploring	Develop and investigate measures to apply data	Differences are conceptualized as stemming from structural inequalities produced from the techniques	Who, why, and how are you silencing/amplifying those included or excluded from the exploring?
Model	Attends to diversity and inclusion within the data	Test for both similarities and differences (*i.e.* implicit bias)	Who, why, and how are you silencing/amplifying those included or excluded from the modeling?
Interpret	Findings are interpreted to represent a universal or normative experience	Discuss how similarities and differences are interpreted and how the structural inequalities are enhanced by the algorithms and statistics	Who, why, and how are you silencing/amplifying those included or excluded from the interpretation?; Sensitivity to nuanced variations in the data

Table 5.2: Boyd (2021) Implication of the 3 Questions for Stage of Data Science Process

5.2 Intersectional Case Study: #metoo

In this dissertation, I wanted to build upon the Black feminist thinkers who laid the intellectual ground work for, "Intersectionality" and extend upon others work into this data science and the wider discipline of computer science. I incorporated the Intersectional framework into the data science process and used #metoo movement as a case study. Many #metoo researchers [202, 203, 225, 286, 302, 307] have begun their point of origin from a single-axis, white women. Utilizing Intersectionality, the point of origin for 'metoo' begins with Burke, a Black woman. She entered the digital conversation from the margins, and her work focused on those in the margins. In essence, this is where the genealogy, a new lineage, for the intersectional framework begins in this research.

The viral #metoo hashtag movement brought global interaction and involvement into the discussion of sexual violence and harassment. Admittedly, this conversation is boundless when it comes to its survivors[11]; there is no exclusion of race, gender, sexual orientation, able-bodiedness, class, etc. from the ranks of the vulnerable. The movement is/was very intersectional. Unfortunately, little of the surrounding research reflects intersectional analysis using machine learning [192, 202, 203, 271, 278, 302]. Most, in fact, focus on white feminist perspectives, which sometimes masquerade as more universalist in intention than in fact, and occasionally focus on the perspectives of survivors who are Black female, LGBTQ+, disabled, male, or any other category. The erasure of the movement's original voices led to several misconceptions about it, such as unfounded allegations of male-bashing, and the mistaken notion that it was founded by and for white women [59, 146, 193]. Unfortunately, given the pervasiveness of sexual assault and violence, the need to be thoughtful and take strides

---

[11]Tarana Burke uses the term "survivor" to refer to people who have experienced sexual assault and sexual harassment.

of the tools we utilize to encompass everyone. Boyd and McEwan [38] argues that there was a sudden increase of visibility which, paradoxically, led to the erasure of the very Black female and LGBTQ+ voices who had originally established the 'me too' movement, thus causing harm to their community. Therefore, this prompts a shifting in design of pipelines that can be more equitable and inclusive and the work towards and show the dynamics of the movement. Also investigating what the mathematical terminology means while using Intersectionality as a guiding framework. Bauer [16] and Bowleg [34] have called out how intersectional terminology is not one-to-one to in a quantitative space. This suggests the need for additional investigation. Using the Table 5.2 and Figure 5.1, I revisit the data science process while being guided by the intersectional framework.

5.2.1 Design

Given the multiple critiques in particular co-option [15, 38, 88, 116, 122, 198, 206, 228, 274] and exclusion of marginalized communities [15, 38, 88, 117, 117, 157, 200, 206, 228, 255, 274] from the #metoo movement, designing a pipeline that captured the hidden users in the movement. In order to do so, the dataset collected needed to be inclusive by first selecting a strategy that encapsulates the communities formed around the movement's two focal women (Burke and Milano). This strategy avoids saturation and thus effective erasure due to the presence of louder or better-platformed voices. In order to find out how the participants in this collected data set participated in this viral conversation, then, create features that would understand their activity in the conversation. There were Twitter users who began to alter the hashtag (#metoo) use to call attention to different communities, this phenomenon was acknowledged by several researchers [60, 164], yet hashtag variations were collapsed under the larger

conversation of hashtag #metoo. After the creation of the selection and features, I used a machine learning algorithm to understand the relationship between the feature usage and participants followed by a validation technique, seeking to understand how networks can be implemented differently to account for the usage of Intersectionality.

Considering the potential of erasure of voices, as a data scientist, I wanted to design a pipeline for a sampling strategy that is sensitive to connections, but resists the saturation that comes from a pure reliance on engagement. Plus, I wanted to avoid the introduction of biases in the results, which might otherwise have to be removed, explained, or simply tolerated; and when they cannot be avoided, we want to understand them.

Once again, the motivation of this work was to encompass the totality of those impacted by the #metoo movement and avoid defaulting to centering on the positioning of whiteness in general, and Milano's (hetero, cis, white upper class) positioning in particular. This type of centering reinforces the false perception that the #metoo movement is for white women, which both further stigmatizes, and disenchants and disenfranchises, the people it ignores [59, 146, 193]. Therefore, I would want to include Burke, not solely for her creation of the original "me too" movement, but for her positioning as a Black woman as well.

Another complexity in this dataset is that we would want to see who is participating in this movement solely from these two women's involvement. What most research does and what is easy enough to execute would be to indiscriminately apply whatever method might be most convenient to us as researchers - easiest to use, assembles a large data set (presumably to become "objective" with the power of large numbers, despite the fact that sampling theory tells us that large data sets are more biased), or simply whatever our lab might use most often. In this case, I chose to emphasize unsupervised algorithms in order to leverage their relative objectivity com-

pared to manual classification. I then also chose a collection method that stood a reasonable chance of collecting the users I wished to focus on, which will be discussed in the next section.

5.2.2 Collection

The aspects of retrieval outlined earlier were querying, sampling, and joining data from multiple sources. In this case study, the first and last are relatively simple: I adapted a Twitter API mass-retrieval script and as such had no multiple sources of data to join. With regards to sampling methods, researchers [36, 91] emphasize the use of those that are suited for finding "marginalized populations who are not readily accessible using more traditional sampling methods such as random sampling" and garden-hose samplings [36, p. 338]. Therefore in this work, I decided to use snowball sampling.

As researchers [81, 128, 129, 181] have thoroughly discussed that different technologies algorithms, and tech institutions, which includes social media like Twitter, are generally created by people who skew white and male. It is well-documented that this skew has created biases in tech's outcomes that harm marginalized, intersectional groups. Therefore, these power relations of whiteness and patriarchy might create skews on Twitter when discussing sexual assault and violence, at the expense of those impacted the most by them. These social constructions are put into place to co-maintain and co-produce power and inequalities [137].

With this in mind, I would want to use a sampling strategy that best captures the actual intersectional population involved in #metoo, despite these structural biases. Snowball sampling inclusion criteria accomplishes this by building the sample around two critical focal points. This also avoids the hordes of people randomly

136

spouting talking points at each other - and may even receive high engagement as a result - who might otherwise be collected by a garden-hose technique. Instead, a group of people was collected who are known to actually be conversing with each other, and from there we might analyze their hashtags to probe which Intersectional subcommunities they participate in.

5.2.3 Cleaning

Before using the machine learning and network analysis, several pre-processing steps were performed. First, I removed all tweets containing non-ASCII characters, along with non-English language tweets in order to focus on the English language tweets. By removing the non-English language and non-ASCII tweets, I acknowledge that they are centering a specific language and the cultures that speak it. According to Pew Researchers, a majority of #metoo tweets in high-volume periods have been in English (71%), but Pew included other languages from Afrikaans, Somali, and Spanish, to name a few [5]. Nevertheless, this data-scrubbing for my convenience is a literal erasure that must be noted. Further analyses should take care to observe whether these groups are underrepresented in our data, and from an intersectional data science perspective, researchers should refrain from making conclusions that center the groups likely to be strongly represented among the excluded tweets. Since the primary focus is the American #metoo movement and context, there is a negligible risk of this happening.

5.2.4 Explore

All the features were numeric; however the distribution of features was highly positively skewed. The normalization of features allows for their equal distribution,

which is useful for classification. However, in this case study the distribution was highly positively skewed, which alerts us that there were large number of users who participated for a short amount of time in comparison to those users who were quite vested into the movement conversation. Hearkening back to Boyd and McEwan's [38] earlier-stated research motives, the latter group is whom we are more concerned about. Therefore, there was no scaling of variables in this data set. Next, I created and selected features for the dataset followed by normalizing the data. I did not normalize the data because the purpose is not be representative.

The approach for the creation and selection of the features was based on the participants involvement in the movement. The impetus is to find the users that were more invested into the conversation, compared to those who are passing by. The approach to create features instead of relying on the traditional metrics like follower and following counts - which are rooted in popularity and algorithms skewed towards whiteness and maleness - allows us to dilute the effects of power and hierarchy dynamics. Instead, the created features allow us to classify people's involvement as a basis for their influence. As Cha, *et al.* [51] demonstrated, having a million followers does not always mean much in the Twitter world.

5.2.5 Modeling

In Research Studies 1c and 2a, the selection of CCA as the unsupervised algorithm stems from the goal of not imparting any bias from our position as researchers. Though this does not negate the inherit bias from the algorithms. Also, the algorithm was selected other understand the relationship between two sets of variables to garner if there is a relationship among them.

In Research Study 2b, the implementation of network analysis stemmed from

ability to visualize the various hashtag communities in a network. There were two strategies executed one that is more intersectional, which entails removing the largest node that takes up the majority of the conversation. The second strategy involved non-intersectional pathway that captures the general the monolithic hashtag #metoo which obscures some of the nuances the smaller communities. Furthermore, in each pathway, there was the use of community detection algorithms that are hierarchical in nature based on how the algorithms cluster similar objects together. The word 'hierarchical' implies there is a rank (or order) to how objects are grouped together, therefore additional need to investigate the mathematics of each community detection algorithm to what was happening behind the scenes.

5.2.6   Interpret

These validation tools, much like statistical tools, were not designed with Intersectionality in mind [34] and thus do not necessarily reflect the real world complexity. As Bowleg [34] points out, until there are new analytical tools created that are designed to account for the complexity of intersectional and validation techniques, we are left figuring out how to best explain the complexity through the collapsed lens of unsupervised algorithms. In the meantime, the emerging themes from these research studies were (1) more than just the "ideal victim" being impacted by sexual assault and violence, (2) observed the usage of hashtag derivatives that call attention other intersectional communities.

Utilizing Intersectionality in the deployment phase of the data science practice, as researchers we must be mindful of both the critical inquiry and praxis stands to gain its meaning within specific social context. The clusters represent granular involvement in the context of the collected data while investigating the #metoo

movement. Returning to the QUINTA diagram which is a framework that integrates Intersectionality and data science (Figure 5.2), there is a summary of the steps taken for the #metoo case study.
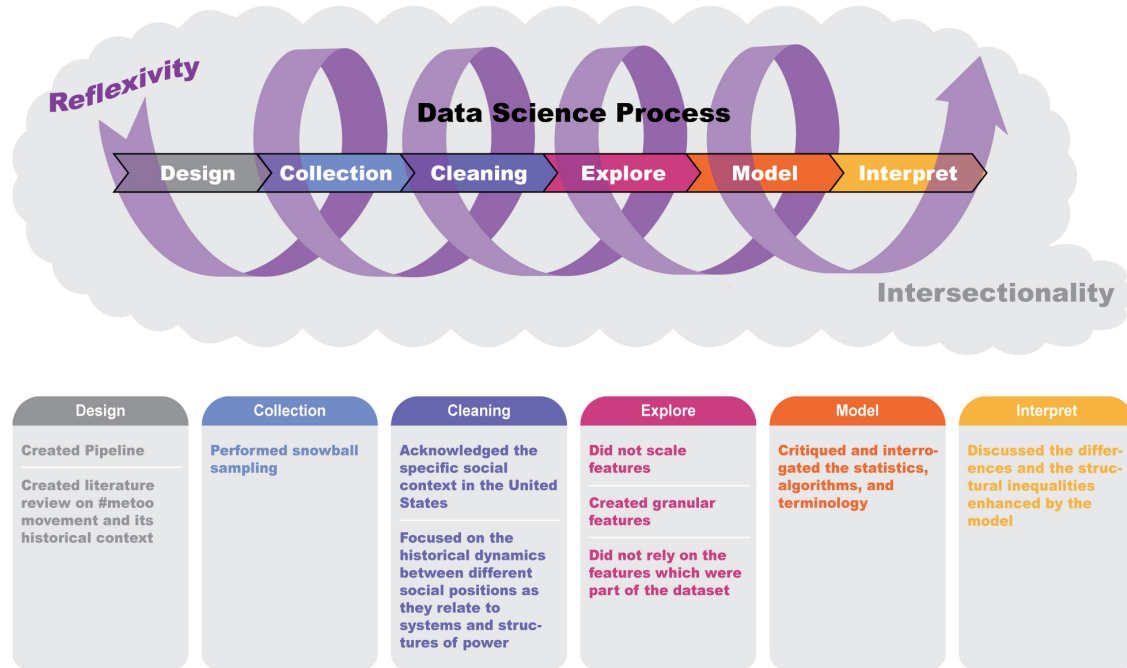

Figure 5.2: Intersectionality meets Data Science: #metoo case study

5.3   Summary, Challenges and Limitations

The #metoo case study illuminates how traditional data science processes of retrieving, engineering, modeling, and validating data from social media streams privilege particular kinds of defaults. The choice of the case study involving a massively viral hashtag, significant disruptions to high society, and white feminist appropriation

of a social movement led by a woman of color all invited significant reflexivity about how social power manifests within and through technology, class, sex, and race. These social facts shaped subsequent research design and methods decisions to consider and employ alternative data science processes. These alternative processes included privileging alternative forms of sampling, attending to rather than normalizing the data, engineering more appropriate quantitative features, embedding quantitative methods within qualitative workflows for interpretation, and attending to how power remains invisible and taken-for-granted within the technical infrastructures and processes of data science.

In adopting an intersectional framing, we are being critical and intentional about our processes while utilizing these methods and algorithms. I pose that bringing Intersectionality into the data science field shifts the perspectives and centers marginalized groups whom are located on the margins which decenters the defaults. Furthermore, reflexivity allows researchers to call attention to their own practices in the context of Intersectionality. In addition, this work contributes to an emerging strand of race-conscious machine learning, computing, and data science scholarship by illustrating how Intersectionality's commitments to relationality, power, inequality, context, complexity, and justice can be brought into conversation with quantitative epistemologies, theories, and methods of data science.

As reminder, Intersectionality does not argue for centering race and gender as the only or most important social identities for analysis. Nor is Intersectionality a matter of concern only for Black women [50]. Nor is it a monolithic framework with definitions and implications to which everyone unanimously subscribes [57, 68]. Rather, Intersectionality attends to how social categories like male, white, or heterosexual are privileged as default, invisible, and individualized while other categories are abnormal, salient, and collectivized. These privileged categories, as Birhane [27, 28]

points out are embedded into machine learning systems which sort, categorize, predict, and classify the world and how its operated from this context.

Traveling theories have a tendency for their core critiques and commitments to become diluted as they move from their home fields. In the case of Intersectionality, it is important to center its attention to constructs like power, inequality, and justice and not simply to multi-category classifications [178]. As Collins says, "while an oppressed group's experiences may put them in a position to see things differently, their lack of control over the apparatuses of society that sustain ideological hegemony makes the articulation of their self-defined standpoint difficult." [69, p. 185] "Human-centered" data science and machine learning researchers must consider how their tools, processes, and interpretations—far from being neutral—re-create and re-inforce existing structures of power and domination. Such inequities can start with categorically-aware analyses of how results vary by race, gender, sexuality, class, and geography but must also point out the types of social power invested in and benefiting from these constructs.

There were some limitations first while trying to implement this approach in my research have stemmed from only considering race-ethnicity, gender, and sexuality were only considered in RQ1c and RQ2a, because they are the only intersectional categories that are easily inferred from the available data. Location, ability, nationality, and other categories are less easily inferred. Leading into another limitation is the collection of Twitter Bio graphic information was post-viral incident. Therefore, the Twitter profiles represented during the time of the viral event might not strongly representative of the user at the time. Furthermore, these Twitter profiles are not static but fluid as the frequency.

5.4   Preemptive Rebuttals

Engaging technical cultures and quantitative epistemologies with critical ideas is likely to meet resistance along several dimensions. I outline three lines of critique to introducing Intersectionality into the data science and offer preliminary rebuttals to these critiques.

5.4.1   Fidelity and Traveling Theory

Intersectionality is a well-known example of a "traveling theory" that has left its origins in critical, ethnic, and legal studies and has been embraced in fields with different theoretical, methodological, epistemic, and political commitments like sociology, psychology, and public health. Traveling theories run the risk of being diluted and mistranslated, and this is especially a concern when translating a framework like Intersectionality grounded in critical, historical, and qualitative ways of knowing into data science and machine learning grounded in positivist, technical, and quantitative ways of knowing [176,214,258]. Psychology [91,92,273] and public health [17,18,34,36] both serve as models for how Intersectionality has been brought into dialogue with quantitative methods while respecting its core constructs. Researchers and scholars in the computing and HCI field have engaged in the utilization of Intersectionality in their work [43, 142, 199, 237, 240–243]. I believe it is essential to attempt to translate intersectional perspectives and frameworks for data science and machine learning audiences because of the increasing amount of social power that is being invested in these fields to manage products and platforms and shape "evidence-based" policies affecting billions of people.

5.4.2   Realism and Reformism

The increasing pervasiveness and capabilities of computing means these technologies and the assumptions of their designers are now in the hands of more people with multiply-marginalized identities [93,182]. A similar dynamic is happening within data science as the barriers to access are falling: the growing availability of "raw" data (APIs, repositories, *etc.*), increasing usability of powerful data analysis tools (software libraries, "auto-ML", *etc.*), and greater access to support resources (tutorials, Q&A communities, *etc.*) puts the capabilities of inference, prediction, and classification in the hands of many more people without the same professional norms around responsibility and ethics shared by, say, doctoral graduates of American computing programs. As much as a radical might want to abolish unjust systems of surveillance or extraction and the technologies that enable them, the proverbial genie is out of the bottle in terms of the "dual use" capabilities of data science and machine learning methods for classification, clustering, and forecasting. Admitting intersectional perspectives into the high temples of computing only to secure incremental reform and regulations is deeply unsatisfying given the unambiguous and existential threats humanity faces. But incrementalism of asking data science and machine learning researchers to incorporate intersectional values can secure greater adoption and harm reduction before socio-technical-ecological path dependencies foreclose on the possibilities for change.

5.4.3   Defensiveness and Whiteness

Members of dominant groups commonly become defensive when marginalized people critique the systems that marginalize them. Even if the objectors can get past naïve arguments like colorblindness [33], the discomfort associated with acknowledging one's privilege and complicity manifests as defensiveness—"But *I'm* not

racist!," "Data is objective!," "Algorithms are objective!," *etc*—that derails arguments about larger, structural processes [209]. The need for the dominant group to center their newfound discomfort *and* expect support from groups describing their on-going marginalization distracts from efforts to build solidarity around dismantling marginalizing systems [7]. There is a related tendency too for performative allyship, tokenism, and appropriation of marginalized perspectives while evading the reflexivity that Intersectionality demands. Intersectionality is not a matter of prioritizing, say, Black women's needs over anyone else's but to create space—especially in elite circles like academic data science and machine learning—for multiply-marginalized people to share their experiences dealing with oppressive structures and systems. Acknowledging the taken-for-granted assumptions, biases, and mechanisms designed into data science tools and processes as well as "who," "why," and "how" questions of how social power works invisibly through them creates opportunities for dialogue to imagine and create alternative configurations.

Chapter 6

Conclusion

This dissertation's impetus started with the investigation of the #metoo movement to illuminate the voices located on the margins who often go unheard or ever recognized. The Intersectional framework's initial usage was solely a guide to help reveal the inequities of the #metoo movement on Twitter. Unbeknownst to me, I was unaware originally of the capabilities and power that encompassed Intersectionality. Amid this pandemic and racial unrest, there are still inequities and injustices currently going beyond a singular social position of race and gender. Being a research data scientist during this pivotal moment in world history, I was heavily influenced and asked myself what my role and responsibility are? How can I use my voice and position to be intentional and elevate others who are more marginalized?

During this dissertation process, I had to allow myself to be open and vulnerable, in addition, I wanted to build upon the Black feminist thinkers. They laid the intellectual groundwork into data science. The openness and vulnerability are represented traveling to other academic disciplines to learn how the Intersectional framework was implemented (*i.e.,* what worked and didn't). While this work is far from perfect it denotes a conversation starter, where I encourage other fields and disci-

plines wherever there is data to think about how their processes impact marginalized people and think about how their processes might not be applicable for everyone and built-in exclusivity. Also, holding accountable the researcher doing the work, no matter what your position is in relationship to the data. Being an intentional researcher, there was a need to understand the power dynamics and the social inequalities present within and outside of the movement to better inform my decisions of the tools and techniques used.

This dissertation critically explored how to bring Intersectionality into dialogue with data science. The coupling of these two epistemologies is linked with the concept of reflexivity. Now, this is not meant to be a one-way conversation of forcing the epistemologies, theories, and methods of data science and machine learning into intersectional frameworks. Recognizing the challenges of translating critical and qualitative frameworks like Intersectionality into quantitative cultures like data science requires synthesizing new frameworks and perspectives. A critical effort in this line for future work is to take the Table 5.2 and apply it to other datasets (*i.e.,* COVID-19) and fields (*i.e.,* medicine and power industry). Wherever there is data, and those data points represent people, inequalities and systems and structures of power are bound to exist. I believe QUINTA can guide data scientists and researchers' practices, tool building, and product deployment while also respecting Intersectionality's commitments to power, inclusivity, context, and justice.

QUINTA is not exhaustive and comprehensive, as Collins [69] reminds us, these core constructs and guiding premises should not be used as a checklist instead as an iterative-reflexive process of how knowledge is produced in data science practice impacting those not traditionally centered. Using Intersectionality helps us as data scientists and researchers to recognize how our positions impact the data science process. Until algorithms and statistical tools are created to account for intersec-

tional positions, we must remain conscientious of Lorde's [194] "master's tools will never dismantle the master's house" and evermore vigilant not reinforce the white supremacy, whiteness, prejudices, and bias in our work.

Hence, this dissertation's purpose is a reflexive piece exploring and navigating how to translate what has thus been far a qualitative methodology into a quantitative one that transfers our approach to data science. While this canonical dissertation is imperfect, its still does not invalidate the opportunity to build and explore the foundation which it establishes. I encourage future researchers and scholars to hold on to the ancestral intersectional tenets and see how this work can be applied in other spaces.

Intersectionality is not meant to be complicated. It becomes complex when researchers try to constrain and contort it in ways to make it fit into dominant ridge ways of thinking. Furthermore, Intersectionality helps us to not stay comfortable and complacent about the world around us. Chan and Erby so aptly explains about Intersectionality "reinforces a commitment to highlight multiply marginalized groups as sources of knowledge and values, requiring critiques that interrupt systems of oppression to prevent the reproduction of injustices and subordination." [53, p. 1251] Therefore, I sought to critically examine the data processing methods when looking at quantitative data from a holistic and reflexive perspective using the #metoo movement was as a case study.

Returning to the conversation of universalism brought up by Alexander-Floyd [1], this dissertation can itself be - at least superficially - classified within the 'universalist' line of thought, it is crucial that we address this potential criticism.. Alexander-Floyd [1] specifically objects to the 're-subjugation of Black women' through the universalization of Intersectionality. However, from a data science perspective, universalization does not necessarily nor inherently re-subjugate Black feminism; rather,

resubjugation must be committed by the research itself.

For starters, this dissertation makes a clear and significant effort to examine the genealogy of Intersectionality and apply it to the problem in question. The further expungement of the origins of the #metoo movement allowed for the more grander narrative that 'all women' was impacted by sexual assault and violence. Albeit true that sexual assault is unfortunately ubiquitous in this way and can impact anyone no matter of their social positioning whether it be single or multiple. Therefore the purpose was to recenter the work of a Black feminist - Tarana Burke - and to apply the Intersectional work of other Black feminists faithfully in the field of data science, thereby avoiding the "mentioned, not used" failure which Alexander-Floyd criticizes. Salem [258] notes that the core mechanism of co-optation is the denial of genealogy. Burke's erasure from #metoo is a prime example, and thus correcting it must be the starting point for intersectionalizing data science methods in this #metoo project.

Second, even as Alexander-Floyd acknowledges, universalism itself is not inherently bad. If anything, in universalizing Intersectionality, we can stay true to its progenitors' original intent by surfacing overlooked identities, including those that may be buried within "Black women" - for instance, Black trans women, or Black-Indian-American women like Kamala Harris - as well as other identities that are similarly buried but not necessarily exclusively Black nor female.

Alexander-Floyd's main criticism of Hancock's re-subjugation of Black feminism hinges on the fact that Hancock actually ignored existing Black feminist work that had already covered the subject matter she purported to be pioneering. Indeed, while no academic can claim to have perfectly researched their background, I have chosen to root this dissertation deeply in Intersectionality's Black feminist genealogy, which has had the direct benefit of avoiding pitfalls like Hancock's and others'. Most researchers of Intersectionality effectively rely on word-of-mouth definitions of it and

mistake it for prescribing a linearly multiplicative relationship between identity and oppression, which contributes to the "flattening" phenomenon that is widely detested by stalwarts of Intersectionality. In other words, we are building the same universalist quantitative approach Hancock aimed for, but not presuming we are correcting the theory of Intersectionality or attempting to turn it into something it's not.

Finally, Alexander-Floyd [1] criticizes the universalizing tendency for privileging dominant modes of knowledge production. As Salem noted, this can be traced to the simultaneous rise of neoliberalism, which co-opted Intersectionality. What separates this dissertation is not only the faithful approach to Intersectionality's genealogy, but also the use of Intersectionality as an ideographic imperative to challenge the dominant - and non-intersectional - modes of knowledge production within data science.The dissertation's objective is to bridge the gap between the data science academy's traditional resistance to and misapplication of "flattened" Intersectionality, and an actual implementation of a quantitative approach built out of intersectional methods. Alexander-Floyd pointed out that many attempts to systematize Intersectionality erase the people it attempts to center. The goal of my quantitative intersectional approach is thus to surface the empirical and experiential aspects within the data not with a singular "intersectional" approach, but a framework for intersectional analyses of quantitative data.

The #metoo movement has challenged gender norms and the rules they play in sexual assault and harassment. There are still people that remain invisible in the movement. Unfortunately, the use of these hashtags does not accurately describe the nuances of the politics of police surrounding sexual assault in United States. These popular narratives presume a homogeneous and universal womanhood and manhood and kind of scare the ways in which sexual assault is racialized [200]. But further can be gendered, and so on. Just as we look at sexual violence we should also ask the

questions whose voices are questioned, mistrusted, deleted whose innocence credibility is assumed which structures or hierarchies do such maneuvers sanction?

While this research takes on new steps and depths by introducing a new framework into data science and looking at a hashtag movement, we should not forget how all of this started—it began with Tarana Burke and her desire to communicate empathy which prompted her to found the 'me too.' The 'me too' movement is based on empowerment through empathy, a transformative empathy that promotes listening and healing. Secondly, the movement has become an agent for exposing the systems of oppression and privilege of sexual harassment and assault or cause and effect.

6.1   Future Research Directions

Deploying Intersectionality as research paradigm has received considerable attention due to its revolutionary tenets advancing critical scholarship and aiding alternative view to approach complexity of social identities impacted by systems and power structures. These systems and powers can be in the form of algorithms where we can make decisions and ascribe results and findings which can color our views of a particular group of people whom harmed, marginalized, and vulnerable.

There are three key areas for future directions involving the #metoo research. The first is looking more closely at media and organization accounts. There were several media and organizations who participated in this movement and generated their own hashtag derivatives. Researchers have explored their impact on the #metoo movement [30, 302]. Furthermore, this is something that Burke mentioned in her tweets saying that there were influence, especially from the media organizations that were responsible for changing the narrative of the #metoo.

A second area of future research is looking at these hashtag derivatives from a

perspective of sub-communities. Instead of Cole and Atuk [60, 302] mentioning these communities in passing, to be intentional about what these derivatives mean for each of this communities. For instance, just because of topic is about sexual assault we can explore what about sexual assault is in conflict of social positions and power the community is up against. Looking at the network visualizations there were several hashtag derivatives that clustered together around other hashtags. Questions surface in regards what do these hashtags mean and what are there associations around and away from the context of the original hashtag?

Lastly, I want to further explore the semantic parsing of hashtag derivative candidates. For the purpose of this dissertation, hashtags like #himthough were not considered #metoo derivatives. However, on semantic grounds, this candidate could potentially be included: it mimics the formula of combining a pronoun ("me" or "him") with an adverb ("too" or "though"). The #himthough was excluded because semantics can be difficult to identify manually, especially in large data sets. Semantic parsing could help highlight these candidates for inclusion.

6.2   Summary of Contributions

Intersectionality is not theoretical vehicle arguing for centering race and gender as the only or most important social identities for analysis. Nor is Intersectionality is a matter of concern only for Black women [50]. Nor is it a monolithic framework with definitions and implications to which everyone unanimously subscribes [57, 68]. Rather, Intersectionality attends to how social categories like male, white, or heterosexual are privileged as default, invisible, and individualized while other categories are abnormal, salient, and collectivized. Therefore, a main contribution of this dissertation is the introduction of the Intersectional framework into the data science field.

Intersectionality will assist what data scientists routinely grapple with the intersectional complexities of datasets and how to be held accountable.

Next is the implementation of snowball sampling, I identified two critical seeds whose causal roles in the movement help mitigate the inherent bias of the technique. Among the sampling method presented, the first was co-occurrence of both hashtags and their derivatives. This shows that the terminology to date is inconsistent. The creation of terminology that specifies and alludes to different communities on Twitter outside the context of original hashtag.

Lastly, the proposed method was the intersectional network analysis. While multidimensional analysis and correlational terms are nothing new to statistics, network analyses are rarely built from an intersectional perspective. This work contributed to further empirical analysis by providing these new tools, upon which future work can be built. It should be noted, however, that this work is exploratory. This dissertation's findings have substantial implications for an Intersectional framework as grounded theory and guiding analysis for future work to explore those in the margins, who are often ignored by or obscured from researchers.

The contribution of this work is to explore the #metoo hashtag movement from an intersectional position instead of using traditional theories and methodologies to gather understandings. Starting from an intersectional framework allows researchers new ways to collect data, explore new features, and interpret findings that one may not arrive at using broader contexts, and approaching problems from this context challenges us how to see and analyze our work. To reiterate, the purpose of this work will not–nor is it looking to–be a savior, solve racism, sexism, classism, or any other'-ism' and its relative. Instead, this work is a foundation, a set of tools, and a guide to traversing the data from a new perspective. The hope for this dissertation is to allow for new approaches to view data from a different perspective, especially

153

those who have multiple intersecting positions and expound on their experiences.

"If we don't center the voices of marginalized people, we're doing the wrong work."

— Tarana Burke

Appendices

Appendix A Snowball Sampling

Snowball sampling is a type of sampling method which begins with identifying a limited number of seeds, typically 1-2. Figure 1 gives an visual representation of how the method is implemented. From each seed, all connections are retrieved. Each connection is used as a seed for the next generation of the process until no new connections are produced or until time and resources have been exhausted [140].
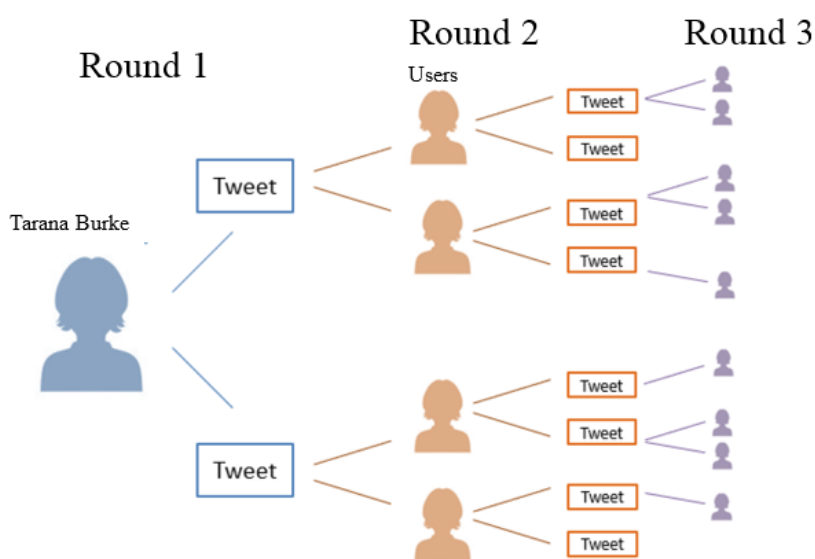

Figure 1: Snowball Sampling Description.

Depending on the type of research, as Handcock and Giles [139] point out, defining the term "snowball sampling" may lead to some confusion due to inconsistent terminology and descriptions, as the term has been used in multiple fields. The first recorded and informal usage of snowball sampling began in 1940s where the Bureau of Applied Social Research were interested in personal influence via media where opinion leaders and followers were identified described by Barton [14]. Paul Lazarsfeld led this research, which remains relevant to today's study of social media [139].

There are two main classes of methods in the literature, both of which have

shared the name "snowball sampling" [139, 267]. In 1958, Coleman described snowball sampling as "a method of interviewing a person's immediate social environment by using socio-metric questions in the interview for sampling purposes." [126, p. 347] Later in 1961, Goodman [126] described the sampling method as way to statistical estimate the mutual connections with individuals in a population. Thompson [277] further elaborates on Goodman's method "as individuals in the sample are asked to identify a fixed number of other individuals, who in turn are asked to identify other individuals, for a fixed number of stages, for the purpose of estimating the number of mutual relationships or social circles in the population." [277, p.182] In other words, Coleman's was a method to be used for acquiring a population sample to calculate traditional statistics on hard-to-reach populations, while Goodman's method was intended to examine network dynamics within the population sampled. In this research, I implemented Goodman's conception of snowball sampling. Happily for my purposes, Twitter users eagerly identify themselves and their relationships using their names, hashtags, and mentions.

Traditionally, snowball sampling has been used as a form of recruitment in the sociology and medical fields [32, 44, 256, 262]. Researchers have limited this method to use recruiting underrepresented participants from otherwise hard-to-find populations for studies, but this has additionally aided in discovering characteristics about hidden populations that were unknown at the beginning of the study [44, 256]. Faugier, *et al.* [95] have used snowball sampling to find hidden populations of AIDS, HIV, and drug addicts that would be beneficial for nursing research. Brown [44] discussed how snowball sampling was used to recruit participants from social networks to gain access to non-heterosexual women. Sadler, *et al.* states this technique comes in great use "when representation from diverse communities is needed, and the research team can't include a representative of all the communities sought." [256, p. 370]

In online social networks such as Twitter, Wu, *et al.* [301] used snowball sampling as a method to distinguish between different types of users in Twitter lists. The researchers sought to look at the flow of information among different categories of users instead of ranking individual users in terms of various influence measures. Elite users were classified as: media, celebrities, organizations and bloggers. Furthermore, Wu, *et al.* [301] sought to determine whether the prominence of elite users affected how these different categories listened to each other. In a related study, Shmargad [264] implemented snowball sampling to look at celebrity and non-celebrity politicians who retweet in a network. He found these two groups had significant impacts on the 2016 United States congressional races; however, celebrity politicians had better opportunities compared to candidates who sought endorsements from politicized celebrities. Chen [55] used snowball sampling in order to collect participants for his exploratory study to investigate whether the usage of Twitter meets the needs to feel connected to other users on the social media platform. The author claims the sampling method was a benefit to reaching users given the type of exploratory research conducted.

Importantly, there are caveats to this technique. Snowball sampling is non-probabilistic, meaning there is no random sampling. Due to its non-probabilistic nature, this technique can be viewed as biased [44, 301] based on the seed selection. Biernacki and Waldorf [256] argue that researchers must know when to terminate the method. Hanneman and Riddle [140] further point out there is no guarantee to find all connections in a population when implementing snowball sampling.

Another use of snowball sampling has been for utility-based modeling. Li, *et al.* [191] created a novel utility-based model to explain the evolution of social networks. They combined expectation-maximization and snowball sampling techniques to explore how social networks evolve on social media. Researchers found that snowball sampling was:

"(1) effective in recovering true parameter values, because a larger seed set size implies more accurate estimates for a given network size, and a larger network needs a larger seed set size, (2) The unobserved meeting states are a crucial factor in analysing network evolution. The meeting probability affects the efficiency of the analytic approach, and considering the meeting states generally yields relatively better estimates and predictions, (3) The choice of sampling method is related to the designed utility function. If a utility function contains the information of long-distance friends, more waves of snowball sampling is necessary and (4) The developed algorithm is scalable so that it can be widely adopted no matter the problem size is large or small." [191, p. 9]

To address the termination of snowball sampling collection, I continued collecting successive generations until there was a decrease in participation. Previous researchers who utilized snowball sampling in their research did not indicate any serious analysis for their cut-off criteria; they only wanted to gather 'enough' participants for their study. It is helpful to note that these researchers were typically in the medical field and the humanities, and thus not universally familiar with applying vigorous statistical or algorithmic rigor to their numerical methods. Therefore, the justification to terminate collection was that the decline of users participating in the original conversation (as defined later in the inclusion criteria) implies the algorithm is reaching a saturation point; also, the ratio of mentions to inclusions goes dramatically up after the first several generations, so the sampling takes longer to sift through non-included users in later generations and reflects a growing lack of interest among mentionees to participate in the conversation.

Appendix B   User Classification

In order to classify users in the data, there were two approaches: subjective labels and assignment via original seed will be explained below.

B.1   Classifying Twitter Accounts: Twitter and Bio-Wikipedia user classification

As previously described, users were classified according to four categories: celebrity, media, non-celebrity and organizations. It was originally planned to scrape Twitter bios to determine which of the four categories a user would fall into. However, it was observed that not all Twitter accounts gave accurate descriptions of a person's activities in order to be classified into the predetermined category. For instance, Alyssa Milano's Twitter bio does not mention that she is/has been an actress. Figure 2 shows the comparison between Twitter and Wikipedia Bios. Therefore, scraping Twitter bio as a sole means of user classification did not suffice. I decided to use Wikipedia to determine the classification for each user by using the Twitter accounts' Display Names as a way to search for user's corresponding Wikipedia page. Then, scraping the first line of each Wikipedia page provided a brief synopsis of the user's history. A representative list of keywords were created for each of the four categories. The following keywords were hand-selected based on (1) representativeness of the desired categories and (2) lack of overlap between categories. Table 1 shows the keyword classification.

There are some limitations to this text analysis classification implementation. First, not all Twitter accounts had bios, and not all Twitter accounts had corresponding Wikipedia pages associated with them. Therefore, I had to classify approximately 800 accounts after automated classification was performed manually. For future work

Figure 2: Actress Alyssa Milano's Twitter and Wikipedia Bios.

Category	Keywords
Celebrity	actor, actress, politician, singer, athlete, TV host
Media	news, media, podcast, radio, television, show, series, documentary, magazine
Organizations	company, organization, corporation, for-profit, non-profit
Non-celebrity	activist, author, writer, journalist

Table 1: Keywords for classifying users.

after completion of PhD, I will explore new techniques to improve on this user classification process. Second, several Twitter users had multiple professions. In this particular study, I decided to only focus on each user's most dominant role to classify their group.

Appendix C   Feature Engineering

The creation of features were based on the involvement of the movement on a micro level instead of macro level. Features such as favorites and retweets capture global dynamics. To determine the popularity on a global level, researchers have been using follower and following counts. However, for this project, feature creation will be based on local dynamics and focus on distinguishing between users who are invested in the movement. Using an intersectional analysis to guide feature engineering refocuses the methodology from dominant but information-poor features to more granular and information-rich ones. Below are some examples of features that were created:

Num_hashtags. This feature was created to count the total number of case-insensitive hashtags a user used. The justification for including this feature is to count how many times a user used hashtags in his/her tweets. This feature allows the researcher to determine the usage frequency of hashtag usage.

Num_metoohash. This feature was created to count the total number of case-insensitive #metoo hashtags a user used. The justification for including this feature is to count how many times a user used #metoo hashtag in his/her tweets. This feature allows for the researcher to determine the usage frequency of #metoo hashtag usage.

Derv_num. This feature was created to count the total number of case-insensitive derivatives of #metoo a user used. The hashtag #metoo, for instance, is the original hashtag, and an example of a hashtag derivative would be #metook12 as well as #churchtoo. The hinging aspect is that these derivatives follow formulas such as #metoo+suffix or #prefix+too. With these definitions established, it is important to note a key distinction: a derivative may co-occur with

162

the parent hashtag, but a co-occurrent hashtag is not necessarily a derivative. The justification for including this feature is to count the number of #metoo derivatives that users called attention to in their tweets. The inclusion of this feature could signify calling attention to additional communities, topics, and/or possible divergence of topics.

Unique_hashtags. This feature was created to count the unique hashtags used by a user. The justification for including this feature is to capture the user's breadth of hashtag use.

All_tweets. This feature was created to count the total number of tweets over the entire collection period. This measures the user's baseline level of activity.

Total_tweets. This feature was created to count the total number of tweets only from the start of the viral hashtag to the end of the collection period. The purpose of this feature is to determine how active the user generally was during the #metoo conversation.

metoo_tweet. This feature was created to count the total number of tweets which had the #metoo hashtag present[1]. The purpose of this feature was to determine how active and consistent the user was during the #metoo conversation.

Indegree. This feature was created to count the total number of times a user is mentioned in tweets. The justification for including this feature is to determine the user's relative popularity in the #metoo context.

metoo_activity. This feature was created to determine how long a user was active from the start of the viral hashtag to the end of the collection period. The feature is in the units of days.

---

[1]This did not include derivative hashtags of #metoo.

Date of account creation. This feature is defined as the number of days, where subtracted the collection date (which was 6-20-2018) from when the Twitter account was first created. For example: Milano's twitter account was created on March 26, 2009, therefore her account is 3,373 days old. NOTE: This feature was created afterwards; therefore it is not included in the machine learning algorithm.

Appendix D   Abstract and Keyword Form

Appendix

Abstract and Keyword Form

Author's name: _____

Thesis dissertation title:

_____

_____

Keywords Please enter **up to 6 keywords or phrases** to enhance the indexing and retrieval of your dissertation or thesis. Separate the keywords by commas.

_____

_____

Subject Categories. Please select one or two subject categories that best fit your Dissertation/Thesis. Subject Categories will help the indexing of your dissertation or thesis and the eventual retrieval of it by interested parties. View the list of available categories here: http://via.library.depaul.edu/assets/taxonomy.pdf

1. _____   2. _____

Thesis Abstract:

Committee Members: Chair: _____

Committee member 1.: _____

Committee member 2.: _____

Committee member 3. : _____

Committee member 4. : _____

Bibliography

[1] Nikol G Alexander-Floyd. Disappearing acts: Reclaiming intersectionality in the social sciences in a post—black feminist era. *Feminist Formations*, pages 1–25, 2012.

[2] Nikol G Alexander-Floyd. *Re-imagining Black Women: A Critique of Post-feminist and Post-racial Melodrama in Culture and Politics*. NYU Press, 2021.

[3] MC Andersen. *Getting to the root of #metoo—Through the fourth wave of feminism*. PhD thesis, (Master's thesis). University of Copenhagen, Copenhagen, Denmark. doi:10.13140/RG. 2.2. 20534.14403, 2018.

[4] Gwendolyn D Anderson. Child sexual abuse prevention policy: An analysis of erin's law. *Social work in public health*, 29(3):196–206, 2014.

[5] Monica Anderson and Skye Toor. How social media users have discussed sexual harassment since #metoo went viral. *Pew Research Center*, Aug 2020.

[6] Dolan Antenucci, Gregory Handy, Akshay Modi, and Miller Tinkerhess. Classification of tweets via clustering of hashtags. *EECS*, 545:1–11, 2011.

[7] Barbara Applebaum. Comforting discomfort as complicity: White fragility and the pursuit of invulnerability. *Hypatia*, 32(4):862–875, 2017.

[8] Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, and Joseph Bayer. Developing a research agenda for human-centered data science. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, CSCW '16 Companion, page 529–535, New York, NY, USA, 2016. Association for Computing Machinery.

[9] Chelsea Sandra Lee Arnold. Examining united states military sexual misconduct policy processes. *International Journal of Sociology and Social Policy*, 2019.

[10] Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the*

fourth ACM international conference on Web search and data mining, pages 65–74, 2011.

[11] Liam Bannon. Reimagining hci: Toward a more human-centered perspective. *Interactions*, 18(4):50–57, July 2011.

[12] David Barnes. *The Draft Riots in New York, July, 1863*. Applewood Books, 2009.

[13] Solon Barocas and Danah Boyd. Engaging the ethics of data science in practice. *Commun. ACM*, 60(11):23–25, October 2017.

[14] Allen H Barton. Paul lazarsfeld as institutional inventor. *International Journal of Public Opinion Research*, 13(3):245–269, 2001.

[15] Judy E Battaglia, Paige P Edley, and Victoria Ann Newsom. Intersectional feminisms and sexual violence in the era of me too, trump, and kavanaugh. *Women & Language*, 42(1):133–143, 2019.

[16] Greta R Bauer. Incorporating intersectionality theory into population health research methodology: Challenges and the potential to advance health equity. *Social science & medicine*, 110:10–17, 2014.

[17] Greta R Bauer and Ayden I Scheim. Advancing quantitative intersectionality research methods: Intracategorical and intercategorical approaches to shared and differential constructs. *Social Science & Medicine*, 226:260–262, 2019.

[18] Greta R Bauer and Ayden I Scheim. Methods for analytic intercategorical intersectionality in quantitative research: Discrimination as a mediator of health inequalities. *Social Science & Medicine*, 226:236–245, 2019.

[19] Frances M Beal. Double jeopardy: To be black and female. *Meridians*, 8(2):166–176, 2008.

[20] Grigory Begelman, Philipp Keller, Frank Smadja, et al. Automated tag clustering: Improving search and exploration in the tag space. In *collaborative web tagging workshop at WWW2006, Edinburgh, Scotland*, pages 15–33, 2006.

[21] Fabiano Belém, Eder Martins, Tatiana Pontes, Jussara Almeida, and Marcos Gonçalves. Associative tag recommendation exploiting multiple textual features. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1033–1042, 2011.

[22] Ruha Benjamin. *Race After Technology: Abolitionist Tools for the New Jim Code*. John Wiley & Sons, 2019.

[23] Sebastian Benthall and Bruce D. Haynes. Racial categories in machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 289–298, New York, NY, USA, 2019. Association for Computing Machinery.

[24] Przemyslaw Biecek. Model development process. *arXiv preprint arXiv:1907.04461*, 2019.

[25] Patrick Biernacki and Dan Waldorf. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, 10(2):141–163, 1981.

[26] Sirma Bilge. Intersectionality undone: Saving intersectionality from feminist intersectionality studies. *Du Bois Review: Social Science Research on Race*, 10(2):405–424, 2013.

[27] Abeba Birhane. Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2):100205, 2021.

[28] Abeba Birhane and Olivia Guest. Towards decolonising computational sciences. *arXiv preprint arXiv:2009.14258*, 2020.

[29] Su Lin Blodgett and Brendan O'Connor. Racial disparity in natural language processing: A case study of social media african-american english. *arXiv preprint arXiv:1707.00061*, 2017.

[30] Emma Frances Bloomfield. Rhetorical constellations and the inventional/intersectional possibilities of #metoo. *Journal of Communication Inquiry*, 43(4):394–414, 2019.

[31] Katherine W Bogen, Kaitlyn Bleiweiss, and Lindsay M Orchowski. Sexual violence is #notokay: Social reactions to disclosures of sexual victimization on twitter. *Psychology of Violence*, 9(1):127, 2019.

[32] Billie Bonevski, Madeleine Randell, Chris Paul, Kathy Chapman, Laura Twyman, Jamie Bryant, Irena Brozek, and Clare Hughes. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC medical research methodology*, 14(1):42, 2014.

[33] Eduardo Bonilla-Silva. *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States*. Rowman & Littlefield Publishers, 2006.

[34] Lisa Bowleg. When black+ lesbian+ woman≠ black lesbian woman: The methodological challenges of qualitative and quantitative intersectionality research. *Sex roles*, 59(5-6):312–325, 2008.

[35] Lisa Bowleg. The problem with the phrase women and minorities: intersectionality—an important theoretical framework for public health. *American journal of public health*, 102(7):1267–1273, 2012.

[36] Lisa Bowleg and Greta Bauer. Quantifying intersectionality. *Psychology of Women Quarterly*, 40(3):337–341, 2016.

[37] Jamillah Bowman Williams. Maximizing #metoo: Intersectionality and the movement. 2020.

[38] Alicia Boyd and Bree McEwan. An intersectional analysis of #metoo: Effects of viral audience on established support communities. *Paper accepted to Social Media Society Conference.*, 2020.

[39] Danah Boyd and Kate Crawford. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5):662–679, 2012.

[40] Avtar Brah and Ann Phoenix. Ain't i a woman? revisiting intersectionality. *Journal of international women's studies*, 5(3):75–86, 2004.

[41] Ulrik Brandes. *Network analysis: methodological foundations*, volume 3418. Springer Science & Business Media, 2005.

[42] Meredith Broussard. *Artificial unintelligence: How computers misunderstand the world.* MIT Press, 2018.

[43] Quincy Brown, Neha Kumar, Jakita Thomas, Alexandra To, and Yolanda Rankin. Discovering intersectionality part i: researcher interrupted. *Interactions*, 28(3):73–77, 2021.

[44] Kath Browne. Snowball sampling: using social networks to research non-heterosexual women. *International journal of social research methodology*, 8(1):47–60, 2005.

[45] Axel Bruns and Stefan Stieglitz. Quantitative approaches to comparing communication patterns on twitter. *Journal of technology in human services*, 30(3-4):160–185, 2012.

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

[47] Tarana Burke. The inception. `http://justbeinc.wixsite.com/justbeinc/the-me-too-movement-cmml`, 2013.

[48] Tarana Burke. Get to know us: Tarana burke, founder. *me too. Movement*, Jul 2020.

[49] Christopher Buschow, Beate Schneider, and Simon Ueberheide. Tweeting television: Exploring communication activities on twitter while watching tv. 2014.

[50] Devon W Carbado. Colorblind intersectionality. *Signs: Journal of Women in Culture and Society*, 38(4):811–845, 2013.

[51] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P Gummadi. Measuring user influence in twitter: The million follower fallacy. In *fourth international AAAI conference on weblogs and social media*, 2010.

[52] Abhijnan Chakraborty, Johnnatan Messias, Fabricio Benevenuto, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. Who makes trends? understanding demographic biases in crowdsourced recommendations. *arXiv preprint arXiv:1704.00139*, 2017.

[53] Christian D Chan and Adrienne N Erby. A critical analysis and applied intersectionality framework with intercultural queer couples. *Journal of homosexuality*, 65(9):1249–1274, 2018.

[54] Devika Chawla. My #metoos before the #metoo. *Women & Language*, 42:165–168, 2019.

[55] Gina Masullo Chen. Tweet this: A uses and gratifications perspective on how active twitter use gratifies a need to connect with others. *Computers in Human Behavior*, 27(2):755–762, 2011.

[56] Eva Cherniavsky. Keyword 1: #metoo. *differences*, 30(1):15–23, 2019.

[57] Sumi Cho, Kimberlé Williams Crenshaw, and Leslie McCall. Toward a field of intersectionality studies: Theory, applications, and praxis. *Signs: Journal of Women in Culture and Society*, 38(4):785–810, 2013.

[58] Xu Chu, Ihab F. Ilyas, Sanjay Krishnan, and Jiannan Wang. Data cleaning: Overview and emerging challenges. In *Proceedings of the 2016 International Conference on Management of Data*, SIGMOD '16, page 2201–2206, New York, NY, USA, 2016. Association for Computing Machinery.

[59] Rosemary Clark-Parsons. "i see you, i believe you, i stand with you": #metoo and the performance of networked feminist visibility. *Feminist Media Studies*, pages 1–19, 2019.

[60] Alyson Cole and Sumru Atuk. What's in a hashtag? feminist terms for tweeting in alliance. *philoSOPHIA*, 9(1):26–52, 2019.

[61] Elizabeth R Cole. Intersectionality and research in psychology. *American psychologist*, 64(3):170, 2009.

[62] Beth Coleman. Race as technology. *Camera Obscura: Feminism, Culture, and Media Studies*, 24(1 (70)):177–207, 2009.

[63] Nancy Coleman. Why we're capitalizing black. *The New York Times*, Jul 2020.

[64] Combahee River Collective. The combahee river collective statement. *Home girls: A Black feminist anthology*, pages 264–74, 1983.

[65] Patricia Hill Collins. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Routledge, 1991.

[66] Patricia Hill Collins. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Routledge, 2002.

[67] Patricia Hill Collins. Social inequality, power, and politics: Intersectionality and american pragmatism in dialogue. *The Journal of Speculative Philosophy*, 26(2):442–457, 2012.

[68] Patricia Hill Collins. Intersectionality's definitional dilemmas. *Annual Review of Sociology*, 41(1):1–20, 2015.

[69] Patricia Hill Collins. *Intersectionality as critical social theory*. Duke University Press, 2019.

[70] Patricia Hill Collins and Sirma Bilge. *Intersectionality*. John Wiley & Sons, 2016.

[71] Michael D Conover, Clayton Davis, Emilio Ferrara, Karissa McKelvey, Filippo Menczer, and Alessandro Flammini. The geospatial characteristics of a social movement communication network. *PloS one*, 8(3):e55957, 2013.

[72] Anna Julia Cooper. *A Voice from the South*. Oxford University Press, 1988.

[73] Brittney Cooper. Intersectionality. In *The Oxford handbook of feminist theory*. 2016.

[74] Natacha Cornaz. An analysis of the#aidtoo movement on twitter: What impacts can a hashtag achieve on sexual exploitation and abuse in the aid sector? 2019.

[75] Neil S Coulson. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. *Cyberpsychology & Behavior*, 8(6):580–584, 2005.

[76] Alejandro Covarrubias. Quantitative intersectionality: A critical race analysis of the chicana/o educational pipeline. *Journal of Latinos and Education*, 10(2):86–105, 2011.

[77] Bridget J Crawford, Linda S Greene, Lolita Buckner Inniss, Mehrsa Baradaran, Noa Ben-Asher, I Bennett Capers, Osamudia R James, and Keisha Lindsay. Talking about black lives matter and #metoo. *Wis. JL Gender, & Soc'y*, 34:109, 2019.

[78] Kimberle Crenshaw. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and anti-racist politics. *u. Chi. Legal f.*, page 139, 1989.

[79] Kimberle Crenshaw. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stan. L. Rev.*, 43:1241, 1990.

[80] Jessie Daniels. *Cyber racism: White supremacy online and the new attack on civil rights*. Rowman & Littlefield Publishers, 2009.

[81] Jessie Daniels. Race and racism in internet studies: A review and critique. *New Media & Society*, 15(5):695–719, 2013.

[82] Jessie Daniels. "my brain database doesn't see skin color": Color-blind racism in the technology industry and in theorizing the web. *American Behavioral Scientist*, 59(11):1377–1393, 2015.

[83] Jessie Daniels. The trouble with white feminism: Whiteness, digital feminism and the intersectional internet. *Digital Feminism and the Intersectional Internet (February 16, 2015)*, 2015.

[84] Kathy Davis. Intersectionality as buzzword: A sociology of science perspective on what makes a feminist theory successful. *Feminist theory*, 9(1):67–85, 2008.

[85] Munmun De Choudhury, Nicholas Diakopoulos, and Mor Naaman. Unfolding the event landscape on twitter: classification and exploration of user categories. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 241–244. ACM, 2012.

[86] Sarah Amira de la Garza. No more magic mirrors: Performing intercultural communication in a (post?) fairy tale world. *Journal of Intercultural Communication Research*, 49(5):494–497, 2020.

[87] Sophie Diehl. *Is Social Media Only for White Women?: From #METOO to #MMIW*. PhD thesis, Northern Arizona University, 2019.

[88] Cristy Dougherty and Bernadette Marie Calafell. Before and beyond #metoo and #timesup: Rape as a colonial and racist project. *Women & Language*, 42:181–185, 2019.

[89] Mengnan Du, Ninghao Liu, and Xia Hu. Techniques for interpretable machine learning. *Commun. ACM*, 63(1):68–77, December 2019.

[90] Debbie Elliott. 5 years after charleston church massacre, what have we learned? *NPR*, Jun 2020.

[91] Nicole M. Else-Quest and Janet Shibley Hyde. Intersectionality in quantitative psychological research: I. theoretical and epistemological issues. *Psychology of Women Quarterly*, 40(2):155–170, 2016.

[92] Nicole M. Else-Quest and Janet Shibley Hyde. Intersectionality in quantitative psychological research: Ii. methods and techniques. *Psychology of Women Quarterly*, 40(3):319–336, 2016.

[93] Sheena Erete, Aarti Israni, and Tawanna Dillahunt. An intersectional approach to designing in the margins. *Interactions*, 25(3):66–69, April 2018.

[94] Sheena Erete, Yolanda A Rankin, and Jakita O Thomas. I can't breathe: Reflections from black women in cscw and hci. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–23, 2021.

[95] Jean Faugier and Mary Sargeant. Sampling hard to reach populations. *Journal of advanced nursing*, 26(4):790–797, 1997.

[96] Ma Feicheng and Li Yating. Utilising social network analysis to study the characteristics and functions of the co-occurrence network of online tags. *Online information review*, 2014.

[97] Melanie Feinberg. A design perspective on data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 2952–2963, New York, NY, USA, 2017. Association for Computing Machinery.

[98] Edward Felsenthal. Silence breakers time person of the year 2017: How we chose. *Time*, Dec 2017.

[99] Sarah Fenstermaker and Abigail J Stewart. *Gender, Considered: Feminist Reflections Across the US Social Sciences*. Springer, 2020.

[100] Paolo Ferragina, Francesco Piccinno, and Roberto Santoro. On analyzing hashtags in twitter. In *Ninth International AAAI Conference on Web and Social Media*. Citeseer, 2015.

[101] Rebecca Fiebrink and Marco Gillies. Introduction to the special issue on human-centered machine learning. 8(2), June 2018.

[102] Crystal Marie Fleming. *How to Be Less Stupid About Race: On Racism, White Supremacy, and the Racial Divide.* Beacon Press, 2018.

[103] Sarah Florini. Tweets, tweeps, and signifyin' communication and cultural performance on "black twitter". *Television & New Media*, 15(3):223–237, 2014.

[104] Center for Critical Race and Digital Studies. Critical race & digital studies syllabus. 2020.

[105] Ashley C. Ford. Why i'm not boycotting twitter today. 2017.

[106] Deen Freelon. On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media*, 58(1):59–75, 2014.

[107] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[108] Simson Garfinkel, Jeanna Matthews, Stuart S. Shapiro, and Jonathan M. Smith. Toward algorithmic transparency and accountability. *Communications of the ACM*, 60(9):5, August 2017.

[109] Alicia Garza. A herstory of the #blacklivesmatter movement. *A Herstory of the BlackLivesMatter Movement by Alicia Garza – The Feminist Wire*, Oct 2014.

[110] Alison Gash and Ryan Harding. #metoo? legal discourse and everyday responses to sexual violence. *Laws*, 7(2):21, 2018.

[111] Sneha Gathani, Peter Lim, and Leilani Battle. Debugging database queries: A survey of tools, techniques, and users. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–16, New York, NY, USA, 2020. Association for Computing Machinery.

[112] Lonneke Geerlings. A visual analysis of rosey e. pool's correspondence archives. biographical data, intersectionality, and social network analysis. In *BD*, pages 61–67, 2015.

[113] Jen Jack Gieseking. Size matters to lesbians, too: queer feminist interventions into the scale of big data. *The Professional Geographer*, 70(1):150–156, 2018.

[114] Carly Gieseler. *The voices of #MeToo: From grassroots activism to a viral roar.* Rowman & Littlefield, 2019.

[115] Fabio Giglietto and Yenn Lee. A hashtag worth a thousand words: Discursive strategies around #jenesuispascharlie after the 2015 charlie hebdo shooting. *Social Media+ Society*, 3(1):1–15, 2017.

[116] Meghan Gilbert-Hickey. #metoo, moving forward: How reckoning with an imperfect movement can help us examine violent inequality, past and present, in order to dismantle it in the future. *South Central Review*, 36(2):1–16, 2019.

[117] Rosalind Gill and Shani Orgad. The shifting terrain of sex and power: From the 'sexualization of culture' to# metoo. *Sexualities*, page 1363460718794647, 2018.

[118] Elizabeth M Gillespie, Roseanne M Mirabella, and Angela M Eikenberry. #metoo/ #aidtoo and creating an intersectional feminist npo/ngo sector. In *Nonprofit Policy Forum*, volume 10. De Gruyter, 2019.

[119] Marco Gillies, Rebecca Fiebrink, Atau Tanaka, Jérémie Garcia, Frédéric Bevilacqua, Alexis Heloir, Fabrizio Nunnari, Wendy Mackay, Saleema Amershi, Bongshin Lee, Nicolas d'Alessandro, Joëlle Tilmanne, Todd Kulesza, and Baptiste Caramiaux. Human-centred machine learning. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, page 3558–3565, New York, NY, USA, 2016. Association for Computing Machinery.

[120] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[121] Jennifer Golbeck, Summer Ash, and Nicole Cabrera. Hashtags as online communities with social support: A study of anti-sexism-in-science hashtag movements. *First Monday*, 2017.

[122] Jennifer M Gómez and Robyn L Gobin. Black women and girls & #metoo: Rape, cultural betrayal, & healing. *Sex Roles*, 82(1-2):1–12, 2020.

[123] Sandra González-Bailón. Online social networks and bottom-up politics. 2013.

[124] Sandra González-Bailón, Javier Borge-Holthoefer, and Yamir Moreno. Broadcasters and hidden influentials in online protest diffusion. *American behavioral scientist*, 57(7):943–965, 2013.

[125] Sandra González-Bailón and Ning Wang. Networked discontent: The anatomy of protest campaigns in social media. *Social networks*, 44:95–104, 2016.

[126] Leo A Goodman. Comment: On respondent-driven sampling and snowball sampling in hard-to-reach populations and snowball sampling not in hard-to-reach populations. *Sociological Methodology*, 41(1):347–353, 2011.

176

[127] Lucia Graves. #metoocongress: Women in politics take action in wake of weinstein, Nov 2017.

[128] Kishonna L Gray. Intersecting oppressions and online communities: Examining the experiences of women of color in xbox live. *Information, Communication & Society*, 15(3):411–428, 2012.

[129] Kishonna L Gray. Race, gender, and virtual inequality: Exploring the liberatory potential of black cyberfeminist theory. *Produsing theory*, 2, 2015.

[130] Kishonna L. Gray. *Intersecctional Tech: Black Users in Digital Gaming*. LSU Press, 2020.

[131] Phyllis E. Greenberger. The #metoo movement is opening a dialogue on ptsd and sexual trauma, Sep 2020.

[132] Evan Greer. The women's march left trans women behind, January 2017.

[133] Anatoliy Gruzd and Caroline Haythornthwaite. Enabling community through social media. *Journal of medical Internet research*, 15(10):e248, 2013.

[134] Anatoliy Gruzd, Barry Wellman, and Yuri Takhteyev. Imagining twitter as an imagined community. *American Behavioral Scientist*, 55(10):1294–1318, 2011.

[135] JF Hair, B Black, B Babin, and RE Anderson. *Multivariate Data Analysis 7th Pearson Prentice Hall*, volume 7. Upper Saddle River, NJ, 2010.

[136] Lelia Marie Hampton. Black feminist musings on algorithmic oppression. *arXiv preprint arXiv:2101.09869*, 2021.

[137] Lelia Marie Hampton. Black feminist musings on algorithmic oppression. *ACM Conference on Fairness, Accountability, and Transparency*, 2021.

[138] Ange-Marie Hancock. Intersectionality as a normative and empirical paradigm. *Politics & Gender*, 3(2):248, 2007.

[139] Mark S Handcock and Krista J Gile. Comment: On the concept of snowball sampling. *Sociological Methodology*, 41(1):367–371, 2011.

[140] Robert A Hanneman and Mark Riddle. *Introduction to social network methods*. University of California Riverside, 2005.

[141] Sandra G Harding. *The feminist standpoint theory reader: Intellectual and political controversies*. Psychology Press, 2004.

[142] Christina N Harrington and Anne Marie Piper. Informing design through socio-cultural values: Co-creation with low-income african-american older adults. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 294–298, 2018.

[143] Saidiya V Hartman et al. *Scenes of subjection: Terror, slavery, and self-making in nineteenth-century America*. Oxford University Press on Demand, 1997.

[144] Naeemul Hassan, Manash Kumar Mandal, Mansurul Bhuiyan, Aparna Moitra, and Syed Ishtiaque Ahmed. Nonparticipation of bangladeshi women in# metoo movement. In *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*, pages 1–5, 2019.

[145] F Henry and C Tator. *The colour of democracy: Racism in Canadian society (3rd edn.)*. Thomson Nelson: Toronto, 2006.

[146] Zahara Hill. A black woman created the "me too" campaign against sexual assault 10 years ago. 2017.

[147] Shanu SP Hinduja. Empowering women in the digital age. *Our World: 100 thought leaders shape our common future*, Jan 2019.

[148] Janell Hobson. Digital whiteness, primitive blackness: Racializing the "digital divide" in film and new media. *Feminist Media Studies*, 8(2):111–126, 2008.

[149] Kate Hodal. Aid sector forced into greater transparency by #metoo movement, Oct 2019.

[150] Michaela Hoffman, Douglas Steinley, Kathleen M Gates, Mitchell J Prinstein, and Michael J Brusco. Detecting clusters/communities in social networks. *Multivariate behavioral research*, 53(1):57–73, 2018.

[151] Jake M Hofman, Amit Sharma, and Duncan J Watts. Prediction and explanation in social systems. *Science*, 355(6324):486–488, 2017.

[152] bell hooks. *Black women and feminism*. Routledge, 1981.

[153] bell hooks. *Talking back: Thinking feminist, thinking black*. South End Press, 1989.

[154] Brittany Hoskins. How does a digital movement keep movin' for five years?, Feb 2015.

[155] Alec R Hosterman, Naomi R Johnson, Ryan Stouffer, and Steven Herring. Twitter, social support messages, and the #metoo movement. *The Journal of Social Media in Society*, 7(2):69–91, 2018.

[156] James Howison, Andrea Wiggins, and Kevin Crowston. Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12(12):2, 2011.

[157] V Jo Hsu. (trans) forming #metoo: Toward a networked response to gender violence. *Women's Studies in Communication*, 42(3):269–286, 2019.

[158] Mengdie Hu, Shixia Liu, Furu Wei, Yingcai Wu, John Stasko, and Kwan-Liu Ma. Breaking news on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2751–2754. ACM, 2012.

[159] Jeff Huang, Katherine M Thornton, and Efthimis N Efthimiadis. Conversational tagging in twitter. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, pages 173–178. ACM, 2010.

[160] Ben Hutchinson and Margaret Mitchell. 50 years of test (un)fairness: Lessons for machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 49–58, New York, NY, USA, 2019. Association for Computing Machinery.

[161] Andrew Iliadis and Federica Russo. Critical data studies: An introduction. *Big Data & Society*, 3(2):2053951716674238, 2016.

[162] Stop Sexual Assault in Schools (SSAIS). #metook12 campaign: Stop sexual assault in schools.

[163] Sarah Jackson, Moya Bailey, and Brooke Foucault Welles. Women tweet on violence: From #yesallwomen to #metoo. *Ada: A Journal of Gender, New Media, and Technology*, (15), 2019.

[164] Sarah Jackson, Moya Bailey, and Brooke Welles. *#hashtagactivism: Networks of Race and Gender Justice*. MIT Press, 2020.

[165] Sarah Jackson and Brooke Foucault Welles. Hijacking #mynypd: Social media dissent and networked counterpublics. *Journal of Communication*, 65(6):932–952, 2015.

[166] Alejandro Jaimes, Daniel Gatica-Perez, Nicu Sebe, and Thomas S Huang. Human-centered computing–toward a human revolution. *Computer*, 40(5):30–34, 2007.

[167] Charles Kadushin. *Understanding social networks: Theories, concepts, and findings*. Oup Usa, 2012.

[168] Nattiya Kanhabua and Wolfgang Nejdl. Understanding the diversity of tweets in the time of outbreaks. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1335–1342. ACM, 2013.

[169] Habib Karbasian, Hemant Purohit, Rajat Handa, Aqdas Malik, and Aditya Johri. Real-time inference of user types to assist with more inclusive and diverse social media activism campaigns. In *Proceedings of the 1st AAAI/ACM Conference on AI, Ethics, and Society*, 2018.

[170] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

[171] Caitlin L Kelly. #metoo or "me too"?: Defining our terms. *ABO: Interactive Journal for Women in the Arts, 1640-1830*, 10(2):4, 2020.

[172] Karen Kelsky. A crowdsourced survey of sexual harassment in the academy. *The Professor Is In*, Jun 2020.

[173] Dustin Kidd. *Social media freaks: Digital identity in the network society*. Routledge, 2018.

[174] Rob Kitchin. Big data, new epistemologies and paradigm shifts. *Big data & society*, 1(1):2053951714528481, 2014.

[175] Rob Kling and Susan Leigh Star. Human centered systems in the perspective of organizational and social informatics. *SIGCAS Computers and Society*, 28(1):22–29, March 1998.

[176] Gudrun-Axeli Knapp. Race, class, gender: Reclaiming baggage in fast travelling theories. *European Journal of Women's Studies*, 12(3):249–265, 2005.

[177] Marina Kogan, Aaron Halfaker, Shion Guha, Cecilia Aragon, Michael Muller, and Stuart Geiger. Mapping out human-centered data science: Methods, approaches, and best practices. In *Companion of the 2020 ACM International Conference on Supporting Group Work*, GROUP '20, page 151–156. ACM, 2020.

[178] Beth E Kolko. Erasing@ race: Going white in the (inter) face. *Race in cyberspace*, pages 213–32, 2000.

[179] David Krackhardt and Robert N Stern. Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, pages 123–140, 1988.

[180] Randy Krehbiel. *Tulsa, 1921: Reporting a massacre*. University of Oklahoma Press, 2019.

[181] Tricia M Kress. In the shadow of whiteness:(re) exploring connections between history, enacted culture, and identity in a digital divide initiative. *Cultural Studies of Science Education*, 4(1):41–49, 2009.

[182] Neha Kumar and Naveena Karusala. Intersectional computing. *Interactions*, 26(2):50–54, February 2019.

[183] Rachel Kuo. Racial justice activist hashtags: Counterpublics and discourse circulation. *New Media & Society*, 20(2):495–514, 2018.

[184] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. AcM, 2010.

[185] Peter Kwan. Complicity and complexity: Cosynthesis and praxis. *DePaul L. Rev.*, 49:673, 1999.

[186] Heather Lang. #metoo: A case study in re-embodying information. *Computers and Composition*, 53:9–20, 2019.

[187] Rebecca Leung and Robert Williams. #metoo and intersectionality: An examination of the #metoo movement through the r. kelly scandal. *Journal of Communication Inquiry*, 43(4):349–371, 2019.

[188] Roee Levy and Martin Mattsson. The effects of social movements: Evidence from #metoo. *Available at SSRN 3496903*, 2020.

[189] Gail Lewis. Unsafe travel: experiencing intersectionality and feminist displacements. *Signs: Journal of Women in Culture and Society*, 38(4):869–892, 2013.

[190] Yang Li, Ting Liu, Jing Jiang, and Liang Zhang. Hashtag recommendation with topical attention-based lstm. Coling, 2016.

[191] Yongli Li, Peng Luo, and Paolo Pin. Utility-based model for characterizing the evolution of social networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, (99):1–12, 2017.

[192] Simon Lindgren. Movement mobilization in the age of hashtag activism: Examining the challenge of noise, hate, and disengagement in the#metoo campaign. *Policy & Internet*, 11(4):418–438, 2019.

[193] P.R. Lockhart. Women of color in low-wage jobs are being overlooked in the #metoo moment. 2017.

[194] Audre Lorde. *Sister outsider: Essays and speeches*. Trumansburg, NY: Crossings, 1984.

[195] Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, et al. The arab spring— the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International journal of communication*, 5:1375–1405, 2011.

[196] Hila Lowenstein-Barkai. #me (n) too? online social support toward male and female survivors of sexual victimization. *Journal of interpersonal violence*, page 0886260520905095, 2020.

[197] Susana Loza. Hashtag feminism, #solidarityisforwhitewomen, and the other #femfuture. 2013.

[198] Ritty Lukose. Decolonizing feminism in the #metoo era. *The Cambridge Journal of Anthropology*, 36(2):34–52, 2018.

[199] Stephanie Lunn, Leila Zahedi, Monique Ross, and Matthew Ohland. Exploration of intersectionality and computer science demographics: Understanding the historical context of shifts in participation. *ACM Transactions on Computing Education (TOCE)*, 21(2):1–30, 2021.

[200] Ashley Noel Mack and Bryan J McCann. Critiquing state and gendered violence in the age of #metoo. *Quarterly Journal of Speech*, 104(3):329–344, 2018.

[201] Erno Makinen and Roope Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE transactions on pattern analysis and machine intelligence*, 30(3):541–547, 2008.

[202] Lydia Manikonda, Ghazaleh Beigi, Subbarao Kambhampati, and Huan Liu. #metoo through the lens of social media. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pages 104–110. Springer, 2018.

[203] Lydia Manikonda, Ghazaleh Beigi, Huan Liu, and Subbarao Kambhampati. Twitter for sparking a movement, reddit for sharing the moment:# metoo through the lens of social media. *arXiv preprint arXiv:1803.08022*, 2018.

[204] Juan Martinez-Romo, Lourdes Araujo, Javier Borge-Holthoefer, Alex Arenas, José A Capitán, and José A Cuesta. Disentangling categorical relationships through a graph of co-occurrences. *Physical Review E*, 84(4):046108, 2011.

[205] Alice Marwick and Danah Boyd. To see and be seen: Celebrity practice on twitter. *Convergence*, 17(2):139–158, 2011.

[206] R Maule. Not just a movement for famous white cisgendered women. *Me Too and intersectionality. Gender and Women's Studies*, 2(3):1–13, 2020.

[207] Leslie McCall. The complexity of intersectionality. *Signs: Journal of women in culture and society*, 30(3):1771–1800, 2005.

[208] Charlton McIlwain. Racial formation, inequality and the political economy of web traffic. *Information, Communication & Society*, 20(7):1073–1089, 2017.

[209] Peggy McIntosh. White privilege: Unpacking the invisible knapsack. 1988.

[210] Kaitlynn Mendes, Jessica Ringrose, and Jessalynn Keller. #metoo and the promise and pitfalls of challenging rape culture through digital feminist activism. *European Journal of Women's Studies*, 25(2):236–246, 2018.

[211] Justin Middleton, Emerson Murphy-Hill, and Kathryn T. Stolee. Data analysts and their software practices: A profile of the sabermetrics community and beyond. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW), May 2020.

[212] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and James Rosenquist. Understanding the demographics of twitter users. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 2011.

[213] Sepideh Modrek and Bozhidar Chakalov. The #metoo movement in the united states: text analysis of early twitter conversations. *Journal of medical internet research*, 21(9):e13837, 2019.

[214] Bonnie Moradi and Patrick R Grzanka. Using intersectionality responsibly: Toward critical epistemology, structural analysis, and social justice activism. *Journal of counseling psychology*, 64(5):500, 2017.

[215] Aaron Mueller, Zach Wood-Doughty, Silvio Amir, Mark Dredze, and Alicia L Nobles. Demographic representation and collective storytelling in the me too twitter hashtag activism movement. *arXiv preprint arXiv:2010.06472*, 2020.

[216] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–15, New York, NY, USA, 2019. Association for Computing Machinery.

[217] Deirdre K. Mulligan, Joshua A. Kroll, Nitin Kohli, and Richmond Y. Wong. This thing called fairness: Disciplinary confusion realizing a value in technology. *Proc. ACM Human Computer Interact.*, 3(CSCW), November 2019.

[218] Lisa Nakamura. *Cybertypes: Race, ethnicity, and identity on the Internet*. Routledge, 2013.

[219] Jennifer C Nash. Re-thinking intersectionality. *Feminist review*, 89(1):1–15, 2008.

[220] Choon-Boon Ng, Yong-Haur Tay, and Bok-Min Goi. A review of facial gender recognition. *Pattern Analysis and Applications*, 18(4):739–755, 2015.

[221] Safiya Umoja Noble. *Algorithms of oppression: How search engines reinforce racism.* NYU Press, 2018.

[222] Ingrid Nunes and Dietmar Jannach. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3-5):393–444, 2017.

[223] Ihudiya Finda Ogbonnaya-Ogburu, Angela D.R. Smith, Alexandra To, and Kentaro Toyama. Critical race theory for hci. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–16, New York, NY, USA, 2020. Association for Computing Machinery.

[224] Abby Ohlheiser. The woman behind 'me too' knew the power of the phrase when she created it — 10 years ago. *The Washington Post*, 2017.

[225] Anna Oleszczuk et al. #hashtag: How selected texts of popular culture engaged with sexual assault in the context of the me too movement in 2019. *New Horizons in English Studies*, 5(1):208–217, 2020.

[226] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 994–1009. ACM, 2015.

[227] Cathy O'Neil. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown, 2016.

[228] Angela Onwuachi-Willig. What about#ustoo: the invisibility of race in the#metoo movement. *Yale LJF*, 128:105, 2018.

[229] Juyong Park and Albert-László Barabási. Distribution of node characteristics in complex networks. *Proceedings of the National Academy of Sciences*, 104(46):17916–17920, 2007.

[230] Nargis Pervin, Tuan Quang Phan, Anindya Datta, Hideaki Takeda, and Fujio Toriumi. Hashtag popularity on twitter: Analyzing co-occurrence of multiple hashtags. In *International Conference on Social Computing and Social Media*, pages 169–182. Springer, 2015.

[231] Morgan E PettyJohn, Finneran K Muzzey, Megan K Maas, and Heather L McCauley. #howiwillchange: Engaging men and boys in the#metoo movement. *Psychology of Men & Masculinities*, 20(4):612, 2019.

[232] Swit Phuvipadawat and Tsuyoshi Murata. Breaking news detection and tracking in twitter. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pages 120–123. IEEE, 2010.

[233] Kathleen H. Pine and Max Liboiron. The politics of measurement and action. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 3147–3156, New York, NY, USA, 2015. Association for Computing Machinery.

[234] Camille Point. #mosquemetoo: Islamic feminism in the twitter sphere. 2019.

[235] Jan Pöschko. Exploring twitter hashtags. *arXiv preprint arXiv:1111.6553*, 2011.

[236] Hemant Purohit and Jennifer Lisa Chan. Classifying user types on social media to inform who-what-where coordination during crisis response. In *Proceedings of the International ISCRAM Conference*, volume 2017, pages 656–665, 2017.

[237] Beth Quinn, Colleen M Lewis, Gretchen Achenbach, Cynthia Lee, Kyla Mc-Mullen, and Vidushi Ojha. Terms to know and videos to help: Gender-identity, sex, sexual orientation, pronouns, race, intersectionality, privilege, & bias. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 1234–1235, 2021.

[238] Maya Raghu and Joanna Suriani. #metoowhatnext: Strengthening workplace sexual harassment protections and accountability. *National Women's Law Center*, 2017.

[239] Gonzalo Ramos, Jina Suh, Soroush Ghorashi, Christopher Meek, Richard Banks, Saleema Amershi, Rebecca Fiebrink, Alison Smith-Renner, and Gagan Bansal. Emerging perspectives in human-centered machine learning. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–8, New York, NY, USA, 2019. Association for Computing Machinery.

[240] Yolanda A Rankin and Jakita O Thomas. Straighten up and fly right: Rethinking intersectionality in hci research. *Interactions*, 26(6):64–68, 2019.

[241] Yolanda A Rankin and Jakita O Thomas. The intersectional experiences of black women in computing. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pages 199–205, 2020.

[242] Yolanda A Rankin, Jakita O Thomas, and Sheena Erete. Real talk: saturated sites of violence in cs education. *ACM Inroads*, 12(2):30–37, 2021.

[243] Yolanda A Rankin, Jakita O Thomas, and Nicole M Joseph. Intersectionality in hci: Lost in translation. *Interactions*, 27(5):68–71, 2020.

[244] Delip Rao and David Yarowsky. Detecting latent user properties in social media. In *Proc. of the NIPS MLSN Workshop*, pages 1–7. Citeseer, 2010.

[245] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, pages 37–44. ACM, 2010.

[246] Raquel Recuero, Gabriela Zago, Marco T Bastos, and Ricardo Araújo. Hashtags functions in the protests across brazil. *SAGE Open*, 5(2):2158244015586000, 2015.

[247] Eugenia Ha Rim Rho and Melissa Mazmanian. Hashtag burnout? a control experiment investigating how political hashtags shape reactions to news content. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–25, 2019.

[248] Michelle Rodino-Colocino. Me too, #metoo: Countering cruelty with empathy. *Communication and Critical/Cultural Studies*, 15(1):96–100, 2018.

[249] Nick Rogers and Jason J Jones. Using twitter bios to measure changes in social identity: Are americans defining themselves more politically over time. 2019.

[250] Aja Romano. Twitter's suspension of rose mcgowan epitomizes the site's most infuriating problem. *Vox*, Oct 2017.

[251] Catherine Rottenberg. #metoo and the prospects of political change. *Soundings*, 71(71):40–49, 2019.

[252] Bonnie Ruberg and Spencer Ruelos. Data for queer lives: How lgbtq gender and sexuality identities challenge norms of demographics. *Big Data & Society*, 7(1):2053951720933286, 2020.

[253] Elliott M Rudwick. *Race Riot at East St. Louis, July 2, 1917*, volume 82. University of Illinois Press, 1964.

[254] Adam Rule, Aurélien Tabard, and James D. Hollan. Exploration and explanation in computational notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery.

[255] Tess Ryan. For indigenous women, #metoo is a fight against racism and oppression, Oct 2019.

[256] Georgia Robins Sadler, Hau-Chen Lee, Rod Seung-Hwan Lim, and Judith Fullerton. Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy. *Nursing & health sciences*, 12(3):369–374, 2010.

[257] Angela Saini. *Superior: the return of race science*. Beacon Press, 2019.

[258] Sara Salem. Intersectionality and its discontents: Intersectionality as traveling theory. *European Journal of Women's Studies*, 25(4):403–418, 2018.

[259] Kimberly T Schneider and Nathan J Carpenter. Sharing #metoo on twitter: incidents, coping responses, and social reactions. *Equality, Diversity and Inclusion: An International Journal*, 2019.

[260] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 59–68, New York, NY, USA, 2019. Association for Computing Machinery.

[261] Sungwon Seo, Jong-Kook Kim, Sung-Il Kim, Jeewoo Kim, and Joongheon Kim. Semantic hashtag relation classification using co-occurrence word information. *Wireless personal communications*, 107(3):1355–1365, 2019.

[262] Abdolreza Shaghaghi, Raj S Bhopal, and Aziz Sheikh. Approaches to recruiting 'hard-to-reach' populations into research: a review of the literature. *Health Promotion Perspectives*, 1(2):86, 2011.

[263] Sanjay Sharma. Black twitter? racial hashtags, networks and contagion. *New Formations*, 78(78):46–64, 2013.

[264] Yotam Shmargad. Twitter influencers in the 2016 us congressional races. *Journal of Political Marketing*, pages 1–18, 2018.

[265] Rozina Sini. How 'metoo' is exposing the scale of sexual abuse, Oct 2017.

[266] Laura M Smith, Linhong Zhu, Kristina Lerman, and Zornitsa Kozareva. The role of social media in the discussion of controversial topics. In *2013 International Conference on Social Computing*, pages 236–243. IEEE, 2013.

[267] Tom AB Snijders. Estimation on the basis of snowball samples: how to weight? *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 36(1):59–70, 1992.

[268] Todd Spangler. Bloomberg renames tictoc news service quicktake, partly to avoid confusion with chinese tiktok app (exclusive), Dec 2019.

187

[269] Victoria Stodden. The data science life cycle: A disciplined approach to advancing data science as a science. *Commununications of the ACM*, 63(7):58–66, June 2020.

[270] Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *2010 IEEE Second International Conference on Social Computing*, pages 177–184. IEEE, 2010.

[271] Jiyoun Suk, Aman Abhishek, Yini Zhang, So Yun Ahn, Teresa Correa, Christine Garlough, and Dhavan V Shah. #metoo, networked acknowledgment, and connective action: How "empowerment through empathy" launched a social movement. *Social Science Computer Review*, page 0894439319864882, 2019.

[272] Minjung Sung and Jang-Sun Hwang. Who drives a crisis? the diffusion of an issue through social networks. *Computers in Human Behavior*, 36:246–257, 2014.

[273] Moin Syed. Disciplinarity and methodology in intersectionality theory and research. *American Psychologist*, 65:61–62, 2010.

[274] Ashwini Tambe. Reckoning with the silences of #metoo. *Feminist Studies*, 44(1):197–203, 2018.

[275] Ana Tarano and Dana Murphy. Tracking #metoo on twitter to predict engagement in the movement.

[276] Toniesha L. Taylor. Dear nice white ladies: A womanist response to intersectional feminism and sexual violence. *Women & Language*, 42:187–190, 2019.

[277] S.K. Thompson. *Sampling*. CourseSmart. Wiley, 2012.

[278] Verity Trott. Networked feminism: counterpublics and the intersectional issues of# metoo. *Feminist Media Studies*, pages 1–18, 2020.

[279] Sojourner Truth. Ain't i a woman? In *Discurso pronunciado en Women's Rights Convention, Ohio*, volume 29, 1851.

[280] Sojourner Truth. And ain'tla 'woman? *Feminism: The Essential Historical Writings. New York: Random House*, pages 93–95, 1851.

[281] Oren Tsur and Ari Rappoport. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 643–652. ACM, 2012.

[282] Zeynep Tufekci. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *arXiv preprint arXiv:1403.7400*, 2014.

[283] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. *Icwsm*, 10(1):178–185, 2010.

[284] İlker Türker and Eyüb Ekmel Sulak. A multilayer network analysis of hashtags in twitter via co-occurrence and semantic links. *International Journal of Modern Physics B*, 32(04):1850029, 2018.

[285] Nancy A Van House. Science and technology studies and information studies. *Annual Review of Information Science and Technology*, 38:3–86, 2004.

[286] Honni van Rjiswijk. #metoo under colonialism: Conceptualising responsibility for sexual violence in australia. *Journal of Perpetrator Research*, 3(1), 2020.

[287] YWCA Week Without Violence. Ywca week without violence. `https://ywcaweekwithoutviolence.org/`, October 2020.

[288] Dakuo Wang, Justin D. Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019.

[289] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. Designing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–15, New York, NY, USA, 2019. Association for Computing Machinery.

[290] Rong Wang, Wenlin Liu, and Shuyang Gao. Hashtags and information virality in networked social movement. *Online Information Review*, 2016.

[291] Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1031–1040, 2011.

[292] Yuan Wang, Jie Liu, Yalou Huang, and Xia Feng. Using hashtag graph-based topic model to connect semantically-related words without co-occurrence in microblogs. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1919–1933, 2016.

[293] Christian Wartena, Rogier Brussee, and Martin Wibbels. Using tag co-occurrence for recommendation. In *2009 Ninth International Conference on Intelligent Systems Design and Applications*, pages 273–278. IEEE, 2009.

[294] Stanley Wasserman, Katherine Faust, et al. *Social network analysis: Methods and applications*, volume 8. Cambridge University Press, 1994.

[295] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 261–270, 2010.

[296] Jason Weston, Sumit Chopra, and Keith Adams. # tagspace: Semantic embeddings from hashtags. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1822–1827, 2014.

[297] Robyn Wiegman. *Object lessons*. Duke University Press, 2012.

[298] Lee E Williams. *Anatomy of four race riots: racial conflict in Knoxville, Elaine (Arkansas), Tulsa, and Chicago, 1919-1921*. Univ. Press of Mississippi, 2008.

[299] Stacey L Williams, Sarah A Job, Emerson Todd, and Kelsey Braun. A critical deconstructed quantitative analysis: Sexual and gender minority stress through an intersectional lens. *Journal of Social Issues*, 2020.

[300] Zach Wood-Doughty, Nicholas Andrews, Rebecca Marvin, and Mark Dredze. Predicting twitter user demographics from names alone. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 105–111, 2018.

[301] Shaomei Wu, Jake M Hofman, Winter A Mason, and Duncan J Watts. Who says what to whom on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, pages 705–714. ACM, 2011.

[302] Ying Xiong, Moonhee Cho, and Brandon Boatwright. 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):10–23, 2019.

[303] Lei Yang, Tao Sun, Ming Zhang, and Qiaozhu Mei. We know what@ you# tag: does the dual role affect hashtag adoption? In *Proceedings of the 21st international conference on World Wide Web*, pages 261–270. ACM, 2012.

[304] Zhao Yang, René Algesheimer, and Claudio J Tessone. A comparative analysis of community detection algorithms on artificial networks. *Scientific reports*, 6:30750, 2016.

[305] Shaozhi Ye and Felix Wu. Measuring message propagation and social influence on twitter. com. *International Journal of Communication Networks and Distributed Systems*, 11(1):59–76, 2013.

[306] Md Hassan Zamir. Diffusion of protest information in twitter during shahbag movement of bangladesh. *Proceedings of the American Society for Information Science and Technology*, 51(1):1–4, 2014.

[307] Dubravka Zarkov and Kathy Davis. Ambiguities and dilemmas around #metoo: #forhow long and #whereto?, 2018.

[308] Ellen Zegura, Carl DiSalvo, and Amanda Meng. Care and the practice of data science for social good. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '18. ACM, 2018.

[309] Amy X. Zhang, Michael Muller, and Dakuo Wang. How do data science workers collaborate? roles, workflows, and tools. *Proceedings of ACM Human-Computer Interaction*, 4(CSCW), May 2020.

[310] Claire Zillman. A new poll on sexual harassment suggests why 'me too' went so insanely viral. 2017.