



Mental health discourse and social media: Which mechanisms of cultural power drive discourse on Twitter

Alina Pavlova^{a,*}, Pauwke Berkers^{b,**}

^a Department of Media and Communication, Erasmus University Rotterdam, Rotterdam, the Netherlands

^b Erasmus School of History, Culture and Communication, Erasmus University Rotterdam, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:

Mental health
Discourse analysis
Social media
Cultural power
Stigma
Topic modeling
Cognitive-emotional currents
Emotional energy

ABSTRACT

The global burden of mental health disorders has increased steadily during the past decade. Today, mental illness is the leading cause of total years lived with disability. At the same time, global mental health policies and budgets fall short of addressing the societal burden as mental health discourse languishes in the shadows due to stigma. As social media have become an increasingly popular source of information, they create opportunities as well as threats for mental health discourse. On the one hand, social media can help to bring awareness to stigmatized topics as they give marginalized members of society the possibility to share experiences and voice their discontent. On the other hand, mental health discourse on social media may lead to stigmatization. To date, little is known about social media mental health discourse and what drives it. This study addresses these research gaps by (1) mapping the mental health discourse on Twitter and (2) analyzing mechanisms of cultural power through which some mental health topics take prevalence over the others. Drawing on Twitter data, this research employs innovative methods of topic modeling, sentiment analysis, and panel data regression analyses. Theoretically, it combines, in a multidisciplinary fashion, concepts such as emotional energy and cognitive focus from sociology and bandwagon behavior from economics. Our findings show that low-cost attention mechanisms are ineffective in fostering online mental health discourse, whereas emotional energy and discursive variability have a positive influence by engaging audiences, creating online solidarity, and speaking to worldviews of audiences from different walks of life. Social media mental health discourse is also shown to be quite diverse and more stigma-neutral than such discourse in traditional media.

1. Introduction

One in four people in the developed world is affected by mental illness (WHO, 2016; Whiteford et al., 2015). Approximately 25 cents in every dollar are lost from the global economy due to mental illness as productivity forgone (Chisholm et al., 2016). However, mental illness has not been sufficiently addressed as a societal problem (IHME, 2010). Even though many mental health concerns are preventable and responsive to treatment, mental health related issues can be characterized by delays in help-seeking, low treatment adherence, lack of policy, and prejudiced provision of care, mostly due to being heavily stigmatized (Chisholm et al., 2016; Thornicroft et al., 2013; Whitley and Wang, 2017).

Creating positive awareness about stigmatized topics, such as that of

mental health, is not considered to be easy (Wahl, 1997). The rising popularity of social media as a source of information creates opportunities as well as threats for mental health discourse. On the one hand, the internet can provide a safe place characterized by anonymity and solidarity where marginalized individuals can voice their opinions, express frustrations, and share knowledge, creating positive awareness and cultural change (Leung and Lee, 2014; Bail, 2016). On the other hand, online environments are also said to have a 'dark side', exemplified by heightened stigma from hate speech and cyberbullying (Campbell, 2005; Beckman et al., 2019).

Consequently, it is important to understand online mental health discourse today and to analyze the mechanisms through which it develops. Specifically, this research aims to understand how the retrievability of discursive frames within online mental health discourse

* Corresponding author.

** Corresponding author. Erasmus School of History, Culture and Communication Department of Arts & Culture Studies, Burg. Oudlaan 50, 3062 PA, Rotterdam, the Netherlands.

E-mail addresses: apav715@aucklanduni.ac.nz (A. Pavlova), berkers@eshcc.eur.nl (P. Berkers).

<https://doi.org/10.1016/j.socscimed.2020.113250>

Received in revised form 9 July 2020; Accepted 22 July 2020

Available online 6 August 2020

0277-9536/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

changes (i.e. when discursive frames emerge, fall behind or thrive) by providing audiences with either (1) meaningful and emotionally charged online conversations, (2) low-cost engagement such as liking and retweeting posts, or (3) cognitively focused communication. Theoretically, we draw upon the concepts of cultural power at a macro-level and apply the theories of emotional energy, bandwagon behavior, and cognitive focus at a micro-level. Accordingly, this research, first, provides an *empirical contribution* outlining current online mental health discourse on English-language tweets in the years 2007–2017. Second, it makes a *theoretical contribution* by suggesting – and empirically testing – online mechanisms of attention generation viewed through the lens of cultural power and cultural change, building upon and contributing to previous sociological works on the roles of cultural carrying capacity and emotional currents (Bail, 2016; Bail et al., 2017). Third, the study makes a *methodological contribution* by applying innovative mixed methods of topic modeling, sentiment, and panel data regression analyses using big data.

1.1. The cultural power of online discourse

1.1.1. Cultural power: retrievability and resonance

Cultural power refers to the extent to which culture can influence one's actions and beliefs. Because people tend to know more culture than they use, such power does not rely on the depth of one's understanding or internalization (Swidler, 1986), but is affected by the interaction between an individual's internal cognitive make-up and the external environment (Lizardo, 2016; DiMaggio, 1997). This is because the external environment provides cultural cues. Consequently, the cues which are more publicly available facilitate action by social actors seeking to reconstruct and enact the culture they know, while at the same time discouraging competing cultural references (Lizardo and Strand, 2010; Swidler, 1986).

Discursive frames, or topics through which the interpretation of a phenomenon occurs, can function as external cultural cues as they stipulate a central organizing idea that specifies the meaning for this given phenomenon (Gamson and Modigliani, 1994). Therefore, it is imperative to know which discursive frames are used to portray certain topics. The more retrievable a discursive frame is, the more likely that it leads to engagement, triggers action, and limits the availability of competing frames. Hence, *retrievability* can be considered the first component of a discursive frame's cultural power.

Secondly, for a discursive frame to have cultural power, it should create *resonance*. Schudson (1989) defined resonance as the relevance (of the frame) to the audience or the fit with the audience's worldview. While widely applied in the academic literature on framing and discourse, this concept has been criticized as often tautological; frames that work are said to be resonant, and frames which are resonant work. McDonnell et al. (2017a), therefore, theorized that resonance is an emergent process that develops through interactions. Their notion of resonance then, as opposed to its more passive form, would come from active individual-frame engagement and the establishment of a tacit connection.

Yet, retrievability and resonance do not operate independently. We argue that resonance influences retrievability. When a discursive frame is retrievable it does not automatically translate into cultural power, unless this frame is being engaged with. Conversely, when frame/actor interactions happen, whether the actor planned it or not, the frame gets charged with a personal meaning or relevance which reinforces further interactions, creating a positive feedback loop and contributing to its retrievability. As such, the more a discursive frame is engaged with, the more retrievable it becomes (McDonnell et al., 2017a).

We posit that resonance as an interactional process can be observed in three processes: emotional energy, bandwagon behavior, and cognitive focus.

1.1.1.1. Emotional Energy. Emotional energy arises from deep engagement with something (Csikszentmihalyi, 1996), or in interaction by intense involvement and commitment, often accompanied by strong emotions and feelings of solidarity, confidence, conviction, and collective effervescence (Collins, 2014; DiMaggio et al., 2017). We argue that emotional energy is crucial for providing a discursive frame with momentum. This form of resonance might lead to more continuous discourse and action and therefore enhance retrievability. It has been empirically shown that emotional energy is a more likely driving force for the public domain discourse than reason (Bail et al., 2017).

To measure emotional energy on Twitter, we assume that each form of communication represents a distinct speech genre that possesses own norms and conventions (DiMaggio et al., 2017) by which emotional energy can be analyzed. When texts alone can change one's mental state, and emotional and cognitive contagion can occur when the shifts of mental states happen collectively with or without co-presence, online environments can enable discursive cascades about social problems (see Bail et al., 2017).

Although our data does not account for the emotions felt by people who engage in Twitter discourse, recent studies show that default mode network (DMN) – a neural network which is also responsible for the emotional evaluation and social categorizing – activates and intercorrelates between subjects also through reading the same texts (Simony et al., 2016), which warrants the use of the psychometric properties of texts as valid indicators of the text's emotionality. While texts can invoke strong emotions, felt similarly by two or more people, we would also like to strengthen the argument by assessing whether these emotions elicit interactions by users engaging in Twitter conversations, making sure that the emotional energy is functional (Collins, 2014).

Focusing on user engagement, we also make sure that the interaction ritual pre-requisite of exclusive co-presence and synchronicity is met (Collins, 2014; Maloney, 2013). Synchronicity is satisfied because Twitter allows for interactions to happen in real-time. Exclusive co-presence is possible because these interactions create a clear boundary to outsiders. Additionally, identifying the presence of a conversation can help to ensure emotional presence. Garas et al. (2012) empirically verified that online interactions are not significantly different on the emotional level from the interactions offline. Hence, in this article, we define emotional energy as a combination *engagement* operationalized via Tweet replies, *affectivity* operationalized via emotional intensity, *confidence* and *solidarity*, both operationalized via linguistic markers of these concepts (see 'Operationalization of variables' section) as those are the aspect of the emotional energy that we can measure.

1.1.1.2. Bandwagon Behavior. Bandwagon behavior, as opposed to emotional energy, is the low-cost form of engagement. Designers of online social networks often rely on instruments such as Favorites and Retweets to attract users to generate online content (Levina and Arriaga, 2014). While favoriting and/or retweeting can be a reflection of an actual commitment to mental health issues, we argue that actual replies show a greater engagement and are more likely to encourage interaction. Favoriting and retweeting are much easier and only require one click of a mouse. Today favorites and retweets have become an online currency of attention. When people see a tweet with a lot of likes or retweets, they might jump on the bandwagon, extrinsically motivated to achieve social distinction or satisfy a need to fit in (Van Herpen et al., 2009). Research has shown that online social media use can be predicted by both received and observed gains in reputation (here – 'favorites'), confirming the theory of bandwagon (Meshi et al., 2013).

We posit that bandwagon behavior has the potential to reduce retrievability going forward as jumping on the bandwagon with low-cost mechanisms could cut the discourse short due to lack of (interest in) interaction and the consequent shifting of attention to more novel

topics. Thus, topics that produce shallow attention will eventually stagnate in analogy with any other fashion or fad (Stigler and Becker, 1977). Bandwagon behavior could likewise result in “slacktivism” – a low-cost form of participation characterized by short-lived attention (Wicks, 2014) which does not have sufficient momentum to manifest in cultural change (Bail et al., 2017).

1.1.1.3. Cognitive Focus. Cognition is a form of awareness responsible for perception, reasoning, remembering, and focusing attention, among others (APA, 2020). Attention is a limited cognitive resource. Especially in the realms of online social networks characterized by abundance and fast flows of information, a message should first catch one’s attention to get engaged with, a process we refer to as *cognitive focus*. We argue that more focused discourse will give the audiences more clarity and create more room for engagement since it is less ambiguous and easier to understand. This argument is based on three different ideas. First, the theory of Interaction Ritual Chains posits that to engage one’s attention there should be a mutual focus of attention (Collins, 2014), which calls for a smaller number of topics so that multiple people reading the same tweet could focus on a single subject. Second, cultural carrying capacity theory suggests that to retain clarity the number of topics shall not be overly diverse to avoid people’s intolerance for complexity (Bail, 2016). Third, clarity of a discursive frame can come from the resolution – defined as a call for action – which gives a sense of direction and enhances cognitive focus even when discursive frames compete (Schudson, 1989).

1.1.2. Hypotheses

To answer the question “to what extent can resonance mechanisms of emotional energy, bandwagon behavior or cognitive focus predict future retrievability of mental health discursive frames?”, we formulated the following hypotheses:

H1. *Emotional energy in a previous period has a significant positive effect on the retrievability of a discursive frame in the current period.*

H2. *Bandwagon behavior in a previous period has a significant negative effect on the retrievability of a discursive frame in the current period.*

H3. *Cognitive focus in a previous period has a significant positive effect on the retrievability of a discursive frame in the current period.*

H4. *Bandwagon behavior in a previous period has a significant positive effect on the retrievability of a discursive frame in the current period.*

1.1.3. The role of stigma

In addition to testing these hypotheses on general mental health discourse, we also aim to assess whether there is a difference in the mechanisms of cultural power for neutral topics versus the topics which stigmatize mental health. For instance, we are not sure if the emotional energy hypothesis will still hold. Stigma affects self-esteem and self-efficacy, discouraging disclosure, and confidence in engagement in mental health discourse (Livingston and Boyd, 2010; Corrigan et al., 2006). Speaking about mental health in a stigmatizing way (i.e. making fun of a disliked politician by calling them ‘mentally challenged’) can on contrary be easy (i.e. prejudice is systemic, often implicit and is based on conformity), intensifying emotional affectivity by the use of pejorative terms (Dovidio et al., 2008). Exploring these questions will help our understanding of the effect of stigma on cultural power mechanisms.

2. Data and methods

2.1. Data sources

2.1.1. Twitter data

In this analysis, we used time-series data derived from Twitter, one of the biggest social network/microblogging services that allow network

participants to engage in open conversations with each other. Compared to other social media platforms (e.g. Facebook), Twitter allows researchers to download publicly available conversations and, hence, analyze the discourse (Bail, 2012). Using search terms “Mental Health” and #mentalhealth, we downloaded ten years of publicly available English language tweets (2007–2017) through the Jefferson-Henrique Python script – widely used in social media research. Since this script uses publicly available data and does not require signing up for a Twitter account or accepting the Terms of Service, using the script does not breach any legal obligations. The tool can be deemed ethical as it cannot derive any private information or breach copyright in any other way. To avoid computational power hiccups due to high volumes of data, we collected a cross-sectional sample for the time-series in question by gathering one day (24h) of tweets per month in the middle of the month/week and avoiding public holidays. In total, our data comprised of 695,414 tweets by 339,493 unique users. Besides texts, the data also included the date stamp, username, number of replies and retweets and favorites, hashtags (#), and mentions (@). Additionally, a 1-year sample (2015) of users’ meta-data of the number of followers per tweet was collected.

2.1.2. Textual sentiment

We analyzed sentiment characteristics of tweets which we automatically derived using LIWC – *Linguistic Inquiry and Word Count* software. LIWC is a dictionary-based, automated text analysis software package that assigns sentiment characteristics to texts (Pennebaker et al., 2015). Specifically, for operationalization of the emotional energy, we used LIWC to assess the emotionality of the tweets (representation of positive and negative emotions) and emotional linguistic markers of confidence and solidarity; for the operationalization of cognitive focus, we used LIWC to assess the resolution component (see ‘Operationalization of Variables’). Additionally, to describe the discourse we looked at what negative emotions precisely were more prominent (i.e. anxiety, anger, sadness). We also used LIWC to create custom dictionaries to quantify stigma-related vocabularies (see ‘Topic Modeling’).

2.1.3. Topic modeling

To map mental health discourse, Latent Dirichlet Allocation (LDA) technique was used to discover main themes related to mental health, using Mallet software (McCallum, 2002). LDA (Blei et al., 2003) explores latent structures in texts by clustering words that ‘occur in documents together more frequently than one would expect by chance’ (DiMaggio et al., 2013: 578). Prior to analysis, the textual data of the tweets were cleaned (i.e. stopwords, corpus specific and rare words were removed, stemming and lowercasing implemented) to ensure better topical coherence. For our purpose, we ran the LDA analysis where we set the number of topics to 30, based on 20 top words which is the methodologically optimal solution (Rehurek and Sojka, 2010; Röder et al., 2015 – see Supplemental Info).

In addition, discursive frames were grouped according to two criteria. Firstly, we categorized topics by the level of mental health stigmatization by assigning a topic to one of two categories: stigma-related or stigma-neutral. We decided whether a topic is stigma-related or stigma neutral if either of two, or both, conditions were met. First, we made a quantitative assessment by looking if a topic exhibited above mean levels of stigma-related vocabularies in a given tweet (see Supplemental Info for custom dictionaries). Second, we did a qualitative thematic assessment, i.e. if a topic related to violence, criminalization or danger, framed mental health as madness or looked at mental illness with skepticism (Rose et al., 2007; Thornicroft et al., 2013).

Secondly, we grouped discursive frames by their trend patterns by analyzing them graphically. We compared the discursive frames which expressed a ‘pick and through pattern’ (i.e. an indication of short-lived attention or “slacktivism”), with discursive frames which displayed

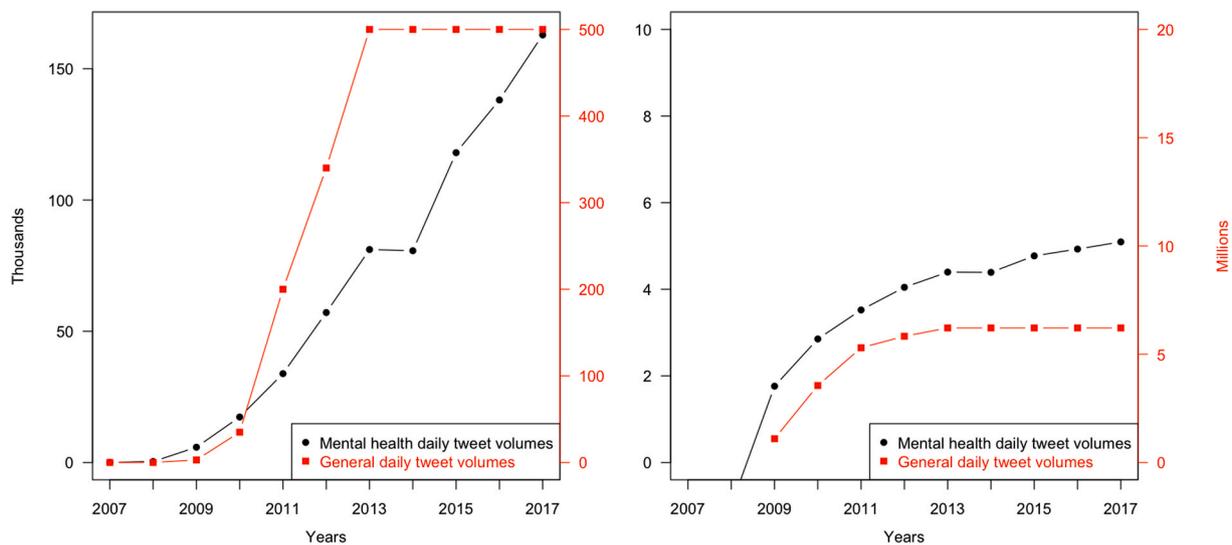


Fig. 1. Daily tweet volumes (left: real data, right: log). Source (General daily tweets volume): *Internet Live Stats (2018)*

more usual trend pattern either representing a flat pattern, growth or decline retrievability trends.

2.2. Statistical analyses

Two statistical techniques were used to analyze the dynamics of discourse. When looking at the mental health discourse in general, we assessed trend patterns of discourse characteristics by using Mann-Kendall trend test (McLeod, 2011), which indicates whether or not the trend observations are significant. Mann-Kendall trend test has been performed in R by using *Kendall* package. We also looked at the differences in discourse characteristics (i.e. sentiment, mechanisms of online attention generation such as ‘Favorites’ and ‘Replies’) between stigma-related and stigma-neutral frames, and short-lived versus stable attention retrievability pattern by comparing means with independent

samples *t*-test by using R *t.test* function.

2.2.1. Regression analysis

We used regression analysis to examine the predictive power of either emotional energy, bandwagon behavior of cognitive focus on the retrievability of the discourse. We used the mean-aggregated data per topic per quarter, resulting in a sample of 1,320 observations (30 topics/ 44 periods).

2.3. Operationalization of variables

2.3.1. Dependent variable

Retrievability is the dependent variable, being the main pre-requisite of cultural power. Frame retrievability was measured through the discursive prevalence as widely prevalent discursive frames are more

Table 1
Descriptive statistics all tweets (N = 695,414).

Variable	Mean	Min	Max	StDev	Skew	Kurtosis	Mean [Twitter] ^a
Tweet Characteristics							
Replies	0.20	0	817	2.6	191	50,256	
Retweets	1.64	0	52,980	126	277	86,850	
Favourites	2.62	0	116,239	257	293	98,364	
User social capital	10,028	0	11,562,825	184,594	59	3,671	
Hashtags	0.44	0	101	1.72	3.39	32.95	
Mentions	0.27	0	19	0.74	3.13	18.30	
Word count	13.91	0	62	2.78	1.57	4.93	
Sentiment Characteristics							
Sentiment	1.17	-66.67	72.22	5.96	0.16	2.80	3.34
positive	3.04	0	72.22	4.40	2	5	5.48
negative	1.88	0	68.42	3.62	2	9	2.14
anxiety	0.40	0	50.00	1.58	5	29	0.24
anger	0.40	0	60.00	1.63	5	44	0.75
sad	0.39	0	37.50	1.54	5	30	0.43
Affectivity	4.58	0.00	72.22	4.93	1.25	2.65	5.88
Confidence	65.01	1.00	99.00	25.75	1	0	63.02
Solidarity	0.62	0	47.06	1.95	4	18	0.47
Discourse Characteristics							
Intra-topic focus	6.25	0	9.97	2.17	1.30	0.89	
Resolution	5.29	0	41.67	4.57	1	0	
Stigma	0.30	0	18.00	0.75	3	15	

^a (Pennebaker et al., 2015).

Table 2
Mann-kendall trend tests (N = 44).

	Slope (10 years trend)	Slope (5years trend)
Mental health discourse	0.925***	0.737***
Stigma vocabularies	0.071	0.05
Emotional Energy	0.896***	0.747***
Replies	0.875***	0.726***
Replies per tweet	0.734***	0.354**
Sentiment	0.556***	0.179
Positive emotions	0.256*	0.01
Negative emotions	0.48***	0.021
Intensity	0.882***	0.021
Confidence	0.592***	0.653***
Solidarity	0.78***	0.442*
Bandwagon Behaviour	0.801***	0.863***
Favourites	0.961***	0.905***
Favourites per tweet	0.897***	0.926***
Duplication rate	0.228*	-0.37*
Cognitive Focus	-0.408***	-0.579***
Inter-topic focus	-0.474***	-0.632***
Intra-topic focus	0.314**	-0.026
Resolution	0.433***	0.158
Controls		
Echo chamber	0.09	0.189
Hashtags (#)	0.878***	0.516**
Hashtags per tweet	0.687***	-0.06
Mentions (@)	0.831***	0.305
Mentions per tweet	0.333**	-0.63***
Other		
Followers	0.191	0.52**
Tweets per user	0.368***	0.073

*p < 0.05; **p < 0.01; ***p < 0.001 (two-sided tests).
Note: A slope of >0 indicates an increasing trend, while a slope of <0 indicating a decreasing trend. The larger the value is, the more degree of slope for the trend line.

likely to ensure that a given discursive frame is both more easily retrievable than others (diffusion) and can be sustained for longer periods (duration) (McDonnell et al., 2017b). Retrievability was operationalized by the number of documents in which a specific topic from the topic modeling analysis occurred (diffusion). Furthermore, our Fixed Effects model considered the time-effect by lagging the predictor variables (duration).

A document was attributed to a specific topic and counted if topic representation exceeded a cut-off value of 20% as a proportion of the document, which means that one tweet could represent up to five topics. We chose 20% as a cut-off rate as we argue that co-occurrence of no less than two words must apply for a tweet to be assigned to a topic. With an average length of a tweet being 34 characters and an average length of an English word being 4.5, a 20% threshold would ensure suitable co-occurrence (Twitter, 2017; MacKenzie and Tanaka-Ishii, 2007).

2.3.2. Independent variables

Emotional energy was operationalized as a combination of engagement, affectivity, confidence, and solidarity as those are the aspects of emotional energy that we could measure. First, we used the number of replies a tweet attracts as a proxy of engagement. Unlike favorites or retweets, replies satisfy the interaction ritual chains criteria of synchronicity and exclusive co-presence as they allow for interactions to happen in real-time and keep the users ‘inside’ the conversation by sending notifications. Additionally, Twitter users can only use the replies where the owner of the Tweet has not restricted it - the restriction does not apply to favorites or retweets. Second, affectivity was operationalized via emotional intensity which was calculated as the square root of the sum of squares of positive and negative emotions proportions (see Lee and Nerghe, 2017), where positive and negative emotions were derived from LIWC dictionaries. Third and fourth, confidence and solidarity represented two other emotional tools that help to build up

emotional energy and continue the interaction ritual chain. In this research, we assessed confidence by LIWC composite ‘Clout’. ‘Clout’ identifies confidence cues such as low count of auxiliary verbs and questions, using more social words, fewer first-person singular pronouns (I), fewer negations, and swear words (Xu and Zhang, 2018). Higher clout suggests high expertise and confidence, while lower clout means tentative, anxious style which is less engaging (Pennebaker et al., 2015). Solidarity is expressed by increased use of plural form first-person pronouns (i.e. we, us, ours), indicating shared identity, empathy, and predicting cognitive convergence (Scheibman, 2004).

The emotional energy variable is a composite variable that averaged normalized ($\frac{(x-MIN)*10}{RANGE}$) values of engagement, affectivity, confidence, and solidarity.

Bandwagon Behavior has been defined by the use of low-cost attention generating mechanisms and was operationalized by the average of the normalized values of *favorites* and *tweet duplication rate* variables. Due to the high correlation between retweets and favorites (0.96, p < 0.05), we decided to only focus on favorites in the general analysis and as a covariate, and leave out retweets (both were operationalized as bandwagon) as they also contribute to the overall discourse and appear in searches.

Cognitive focus was operationalized as a normalized average of intra-topic (within-topic) focus, inter-topic (between-topics) focus, and resolution. *Intra-topic focus* was calculated as one minus the median value of a proportional split of 30 topics within a single tweet (intra-tweet). We measured *inter-topic focus* as one minus the median value of a proportional split of 30 topics aggregated per period. A higher (inter/) intra-topic focus would then entail that the discourse (per given period) is more concentrated, which meant that it possessed a lower number of discursive frames of high prominence. *Resolution* was operationalized per the notion of illocutionary force in speech-act theory. Particular linguistic compositions of speech can display urgency and call for action, requiring focused attention (Austin, 1962). The speech will usually imply action by making use of imperative verbs (usually present tense second-person verbs “Come here”, “Read this”), via modal auxiliary verbs (“Could you ... ?”) or the use of performative verbs which convey warnings (“I advise you to ...”) (Allen and Core, 1997). Since LIWC allows us to detect verbs and present focus, we composed a proxy to reflect resolution in a text by averaging the proportions of verbs, present focus, and modal auxiliary verbs using a custom-made dictionary.

2.3.3. Control variables

Echo-chamber became a control variable because increased attention to the topic, even in the presence of emotional energy, could be restricted to a certain closed network (Kretschmer et al., 1999). Echo chamber here was operationalized via inverse *Krackhardt’s E/I Ratio*:

$$E - I \text{ Index} = \frac{(EL - IL)}{(EL + IL)} \times (-1)$$

where *EL* represents the number of edges that are external to a given topic per period (quarter) and *IL* is the number of edges internal to – or between – vertexes within that topic. This measure varies on a scale from –1 to 1 with –1 representing a perfectly open community and +1 representing an absolute echo chamber (Krackhardt and Stern, 1988). Using usernames related to tweets, Internal Edges were calculated as a total number of users minus a unique number of users per topic per period, while the External Edges were calculated by a difference of a total number and a unique number of users of all topics per quarter minus Internal Edges of the topic.

Additionally, we controlled for the influence of hashtags and mentions which could affect the retrievability of discursive frames. Hashtags (#) are used to “index keywords or topics on Twitter. This function [...] allows people to easily follow topics they are interested in [...] or categorize those tweets and help them show more easily in Twitter search” (Twitter, 2018a). Mentions (@) are used to “tag other users in

the discourse for either acknowledgment or engaging these users in conversation” (Twitter, 2018b).

Although we predict emotional energy to positively influence retrievability, there might be a tipping point after which retrievability will subside. According to the theory of cognitive-emotional currents, fast-rising emotions may reach a saturation effect if the content of the discourse is not strong enough to induce the cognitive process (Bail et al., 2017). We accounted for the saturation effect by lagging multiple periods.

In the Supplemental Info, the reader can find an extended operationalization table, Tweet examples of the variables, custom-made dictionaries, and step-by-step calculation of intra- and inter-topic focus.

2.4. Model

For the panel data, the most appropriate regression model is the econometric panel linear model (R plm package Croissant and Millo, 2008) for the following reasons. First, this panel data are time-series data where the goal of the analysis is to scrutinize the relationships between covariates and the dependent variable, taking into consideration two dimensions – topics and time. Second, the econometric panel model allows accounting for individual-specific effects implied by the data (Baltagi, 2008). Third, the econometric model solves

non-stationarity by differencing dependent and independent variables (Schmidheiny and Basel, 2011). Four, at large sample sizes, the economic panel model proximates the continuous exponential model.

Based on the Hausman test ($p < 0.05$), a diagnostic for the assumption of whether individual-specific effects are more or less likely to be correlated with explanatory variables, the Fixed Effects (FE) model was chosen. Whereas the Random Effects model assumes that (1) ‘the individual-specific effect is a random variable that is uncorrelated with the explanatory variables of all past, current and future time periods of the same individual’ and (2) that ‘individual-specific effect is of constant variance’, the FE model allows for ‘individual-specific effect to be correlated with the explanatory variables’ (Schmidheiny and Basel, 2011). As such, it controls for omitted variables by controlling the variables against each other (unobserved effect).

The FE model can be written as follows:

$$\ddot{y}_{it} = \beta_{EE} \ddot{x}_{(EE)it} + \beta_{BW} \ddot{x}_{(BW)it} + \beta_{CF} \ddot{x}_{(CF)it} + \beta_{CV} \ddot{x}_{(CV)it} + \ddot{u}_{it}$$

,where \ddot{y}_{it} is a count of documents representing retrievability, from which time averages have been subtracted ($\ddot{y}_{it} = y_{it} - \bar{y}_{it}$, where $\bar{y}_{it} = 1/T \sum_t y_{it}$) per subject i per time t .

Table 3
LDA-derived MH discourse topic model (N = 695,414).

TOPIC	WEIGHT	Top words (10)
PROBLEMATIZATION^b	5.37%	issues people problems care important good life talk feel illness
FEELINGS	3.03%	day good issues bad love people shit break feel lol care head mind
COMMUNITY, AWARENESS, EVENTS	2.97%	event conference support awareness aid training join Community group meeting
EDUCATION, RESEARCH, SCHOOLING	2.93%	care services students support research school education training Community social childrens youth improve
ANTI-STIGMA AWARENESS	2.51%	Stigma talk awareness people lets change campaign conversation support open
EVERYDAY LIFE	2.42%	day time good important work days school lives today bad risk taking put feel back year
DEPRESSION, ANXIETY, BPD, PTSD (#)	2.08%	#depression #anxiety #mentalillness disorder #bipolar #ptsd illness #bpd # Stigma #suicide
ACCESSIBILITY OF MH CARE	2.07%	care services system treatment State coverage access veterans substance abuse addiction insurance program
HEALTHY LIFESTYLE, PHYSICAL EXERCISE	1.93%	physical improve important exercise body mind benefits stress healthy life positive care boost
SOCIAL MEDIA AND SHARING	1.89%	blog #mhsm stories read post share news social story chat
SUBSTANCE ABUSE	1.76%	abuse substance addiction study risk drug women treatment teens poor
FUNDING CUTS, CRISIS	1.73%	services funding support cuts minister crisis government budget call reform money
WELLBEING AND MINDFULNESS, SELFCARE, LOVE	1.69%	#mindfulness #wellness #wellbeing # recovery #therapy #love #happiness # self care #inspiration #motivation
STRESS AND PHYSIC AL SYMPTOMS^b	1.53%	depression study brain risk research anxiety linked stress sleep eating
YOUNG PEOPLE, LEARNING DISABILITY	1.46%	people young children support adults experience parents struggling learning disabilities
GUNS AND VIOLENCE^a	1.37%	gun control violence laws people checks background shootings school reform bill mass ban
NURSING JOBS, WORKING AT A HOSPITAL^b	1.34%	nurse worker services registered job Community therapist care manager hospital
ENGLAND, NHS SYSTEM, CRISIS	1.24%	care services crisis nhs patients beds England provide staff system
CHARITY AND HELPING	1.12%	support charity online follow interested send resources love people helping
MH AWARENESS MONTH, CAMPAIGNING	1.11%	awareness week #mhaw month raise facts support campaign everybodys video
DEATH AND MURDER^a	0.99%	death hospital stabbed experts police murder facility patient people worker
MEN, VETERANS, PTSD, SUICIDE, MILITARY	0.89%	suicide military crisis veterans #suicide #ptsd ptsd # veterans men care
DONALD TRUMP^a	0.89%	donald trump president doctor exam physical journalists america expert cognitive evaluation
WORK AND MH	0.80%	World place work employees conditions benefits wear apps wristbands employers candidate
DONATING TO MH PROGRAMS	0.75%	#bulletstalk tweet awareness donate canada programs raise money initiatives support
PSYCHIATRIC PROFESSION, PSYCHIATRIC NURSING	0.71%	nurse Psychiatric job practitioner position counselor psychiatry therapist counseling professional
WORLD MH DAY	0.60%	day world #wmhd depression global october happy #depression link theme
AUTISM, ADHD, CHILDREN	0.55%	children #parenting #autism #asd #adhd sciences autism crisis adhd syndrome
POLICE TRAINING, PRISONS	0.52%	police raises million training aid officers spending people boosts budget reform cells
PARENTING	0.46%	american parenting baby monitor exmobabybychkov visit receive medication assistance coddling
	Σ 48.7%	

^a contextually stigmatizing topics.

^b topics featuring high stigma-related vocabularies.

Table 4
Mann-kendall trend test (n = 44/20/12).

	Slope (10 years trend)	Slope (5 years trend)	Slope (3 years trend)
Stigma-related discourse	-0.532***	0.0116	0.697**

*p < 0.05; **p < 0.01; ***p < 0.001 (two-sided tests).

Note: A slope of >0 indicates an increasing trend, while a slope of <0 indicating a decreasing trend. The larger the value is, the more degree of slope for the trend line. We use proportions of topics per period for observations, as it takes into account the discursive competition and assesses performance of the topics relative to each other. The retrievability metric of document count would not be helpful in assessing trends as the numbers of documents for mental health discourse have been increasing for most of the topics during the last 10 years.

\ddot{x}_{it} is an independent variable calculated in a similar manner (where *EE* stands for Emotional Energy operationalization, *BW* - Bandwagon Effect, *CF* – Cognitive Focus operationalization and *CV* - a control variable(s)),

β coefficients which measure the effect of the explanatory variables (*x*'s) on the dependent variable (*y*),

\ddot{u}_{it} is an idiosyncratic error term which is, likewise, time-adjusted.

The logic behind the FE model, therefore, entails the subtraction of time averages from both sides of the initial model (equals first differencing for 2-period model), which yields a “within model” solution for heterogeneity bias by canceling out the time-related individual-specific effect, the intercept and time-invariant regressors. This also means that the coefficients derived by the FE model can be used to describe the independent variables’ effects autonomously from other covariates (Schmidheiny and Basel, 2011). The model’s assumptions have been checked with no discrepancies identified (see Supplemental Info).

3. Results

3.1. Descriptive statistics

Twitter data showed a growing interest in mental health discourse, not only in absolute numbers but also relative to general Twitter discourse (Fig. 1). Mental health discourse represented 0.003% of all Twitter discourse, which was rather small. However, the average five-

year growth rate of 15% relative to total discourse indicated a growing interest in the subject.

Table 1 shows that while only 20% of the tweets elicited conversation in terms of replies, an average tweet was “liked” and retweeted 1.6 and 2.6 times, respectively. On average, 44% of tweets used hashtags, and 28% of tweets mentioned other users. People who tweeted about mental health had high social capital (median = 1026 followers). However, looking at the absolute numbers, 55% of the discourse was initiated by users with less than 1000 followers (usually private individuals), 30% by accounts with 1000 to 5000 followers (usually small enterprises, consultants), and the remaining 15% by media outlets, businesses, NGOs and NPOs, celebrities or other influencers (doctors, consultants, researchers) (see Supplemental Info). About a quarter of the tweets collected were either retweets (3%) or duplicate tweets (22%) which copied the texts word-by-word, posted by the same or different users (possibly bots).

With regards to discourse characteristics, the average tweet focus (intra-topic focus) on a 0–10 scale was 6.25 (0 being very dispersed, 10 being very thematically focused). The proportion of directedness of an average tweet (resolution) was 5.29 out of 100. Texts were on average of moderate confidence (65.01/100), which was slightly higher than the general Twitter level of confidence (63.02/100). The tweet sentiment was tilted slightly towards positive, however, it is lower than Twitter sentiment in general. The level of perceived anxiety, as a negative emotion, in mental health discourse was more pronounced in Twitter mental health discourse than in general; however, the levels of anger and sadness were lower. Solidarity was also more pronounced in mental

Table 5
Descriptive Statistics [Stigma] with *t*-test (N = 1,320).

Variable	Stigma-related discourse		Stigma-neutral discourse		All mental health discourse		Sig.
	Mean	StDev	Mean	StDev	Mean	StDev	
Twitter Characteristics							
Retrievability	1061	1153	741	892	826	978	***
Replies per tweet	0.11	0.12	0.12	0.13	0.13	0.34	
Favourites per tweet	1.12	6.80	0.86	2.75	0.93	4.22	
Duplication rate	1.06	0.31	1.07	0.41	1.07	0.39	
Tweets per user	1.09	0.33	1.09	0.44	1.09	0.41	
User social capital	5315	3535	12,060	13,049	10,261	11,707	***
Hashtags per tweet	0.17	0.20	0.35	0.62	0.30	0.55	***
Mentions per tweet	0.16	0.27	0.25	0.47	0.23	0.43	***
Echo Chamber	0.16	0.43	0.23	0.36	0.21	0.38	**
Sentiment Characteristics							
positive	2.86	1.62	2.24	1.70	2.41	170	***
negative	1.93	1.47	1.35	1.34	1.51	1.39	***
Sentiment	0.93	2.10	0.89	2.15	0.90	2.14	
Affectivity	4.43	2.06	3.40	2.03	3.67	2.09	***
Confidence	57.40	17.80	57.14	19.65	57.21	19.17	
Personal	0.38	0.30	0.40	0.44	0.40	0.42	
Discourse Characteristics							
Intra-topic focus	7.43	0.82	7.37	0.93	7.39	0.90	**
Resolution	4.57	2.79	3.97	2.60	4.13	2.67	***
Resonance Mechanisms							
Emotional Energy	3.48	1.32	3.20	1.37	3.28	1.36	**
Bandwagon	4.09	1.57	3.61	1.77	3.74	1.73	***
Cognitive Focus	3.87	0.86	3.78	0.95	3.81	0.93	*

*p < 0.05; **p < 0.01; ***p < 0.001 (two-sided tests).

health related tweets.

From the Mann-Kendall test of trend significance (Table 2), mental health discourse showed a trend in affectivity rising over time, mainly due to an increase in negative – and to a lesser extent positive – sentiment. We noted higher levels of solidarity, exhibiting an upward trend. The level of confidence was also steadily growing. The negative inter-topic focus trend indicated an increase in mental health diversity of topics. The intra-topic focus decrease, however, signposted an increase in tweets discursive focus making the tweets less fuzzy. The call to action, or resolution, has become significantly more prominent as well.

Concerning tweet characteristics, the use of favorites and replies per tweet grew. The use of hashtags saw a significant increase during the last 10 years. The use of mentions, however, significantly declined in the last 5 years. Replies per tweet also grew, pointing out an increase in engagement. This was also confirmed by the decrease in duplication rate, while the tweets per user ratio remained stable. Neither stigma vocabularies nor echo chamber showed a significant trend. In fact, stigma-related vocabularies were rare. Only 0.3% of the discourse featured words and phrases associated with pejorative portrayals of mental health, as compared to almost 28% in the press (Whitley and Wang, 2017).

With regards to resonance mechanisms, indicators of bandwagon behavior have been growing at the highest pace, followed by emotional energy. The cognitive focus of the discourse has seen a significant decline, indicating that mental health discourse has become fuzzier.

3.2. Topic modeling results

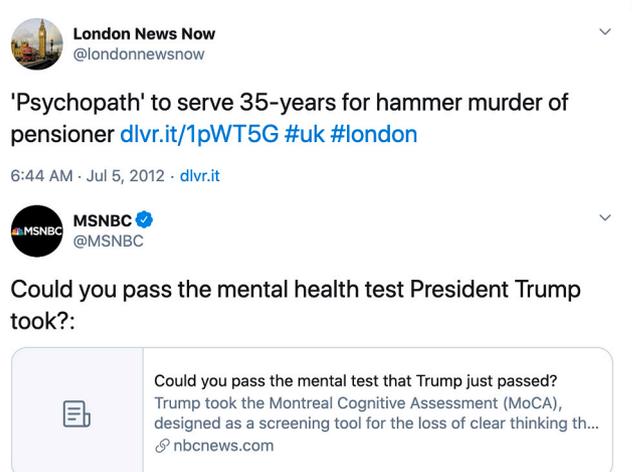
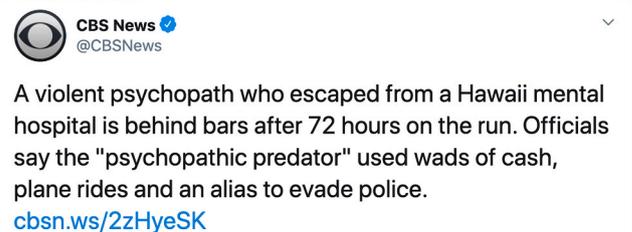
Thirty topics emerged via topic modeling analysis (Table 3). The topics have been sorted by prominence and represent 48.7% of the mental health discourse on Twitter. The remaining 51.3% of discourse consists of more fragmented topics (represent <0.5% of discourse); hence, they could not be meaningfully identified by an LDA model. The topics were labeled according to their keywords and qualitative assessment looking at the tweets they represented. For instance, the topic called 'Feelings' was labeled as such because the tweets which represent this topic are feeling-related, where the author of the tweet speaks about how he or she feels about one's self, one's mind or other people. The topic called 'Problematicization' speaks about mental health as an 'issue', 'problem' or 'illness', and so on. The topics below were used to research the mechanisms behind how discursive frames rise and demise.

Regarding stigma-related discourse, the contextual theme of 'Stigma' emerged from topic modeling (Table 3 a). The 'Stigma' theme is comparable to previous literature on mental health discourse on media (Thornicroft et al., 2013; Whitley and Wang, 2017) and revolved around risk, danger and violence (topics of 'Guns and Shooting' and 'Death and Murder'), and the connection of mental illness to madness and insanity as portrayed by its contextual use within 'Donald Trump' topic. The contextual theme of 'Stigma', however, only represented 3.2% of the discourse which is, again, much lower in comparison to traditional media discourse. Additionally, certain topics exhibited above-average levels of stigma-related vocabularies (Table 3 b). Both contextual 'Stigma' topics and topics high in stigma-related vocabularies (Table 3 topics in **bold**) formed the analytical category of stigma-related discourse, as opposed to stigma-neutral discourse.

Looking at the trends of thematic prominence (Table 4), we note that stigma-related discourse has significantly stagnated during the last 10 years, but recovered and showed a significant increase in the last 3 years relative to stigma-neutral discourse.

We discovered that stigma-related discourse was significantly more retrievable by assessing the differences in the discourse characteristics between stigma-related and stigma-neutral discourse (Table 5). Unlike stigma-neutral tweets, stigma-related tweets were characterized by sensationalism and fluidity (i.e. not a continuous discourse, but stories of a similar theme replacing each other overtime). Oftentimes these

types of tweets came from media outlets (see examples).



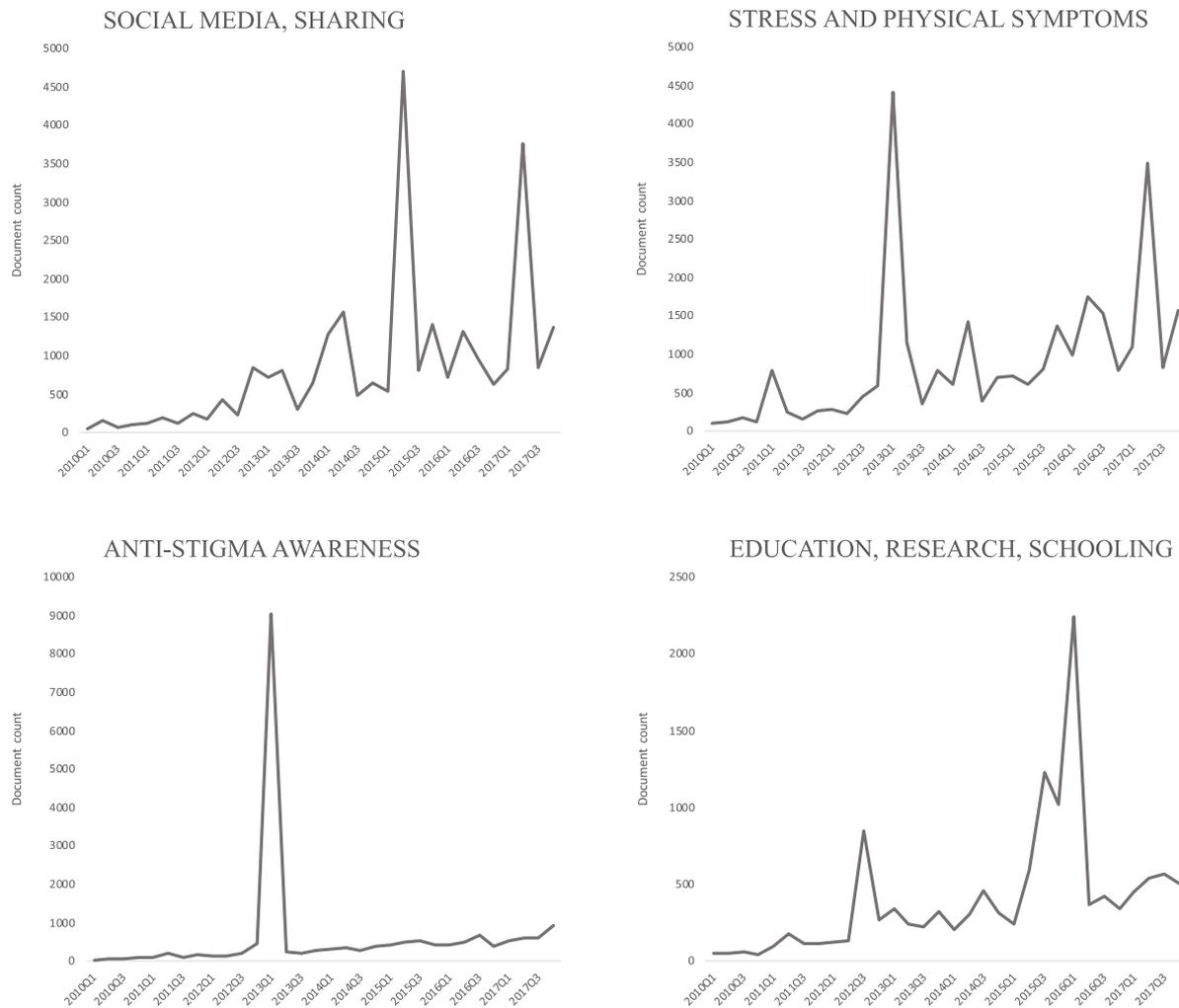


Fig. 2. Topics with short-lived retrievability pattern.

Stigma-related discourse also displayed significantly higher levels of emotional intensity and levels of negativity. Paradoxically, stigma-related discourse also showed significantly higher levels of positive emotions, making stigma-related discourse sentiment higher than that of

stigma-neutral. Upon closer investigation, several examples emerged which indicated that positivity of stigma-related mental health discourse could stem from sarcasm (see the anonymized examples).

Murderer of New Addington teenager says escape from mental health unit was 'best three months of his life'
[#Croydon croydonadvertiser.co.uk/Lerone-Boye-mu... ..](#)

5:52 AM ·  2014 · Twitter Web Client

Them - "I really think Trump is doing a good job!"
 Me - "Please pass the ham."

If you can't suck it up for 2 hours, stay home from Thanksgiving and Google "mental health providers in my area"

 Salon  @Salon · 

How to survive Thanksgiving if you have to spend it with diehard Trump supporters ift.tt/2mhqhye

11:13 AM ·  · Twitter for Android

Replying to @USATODAY

How callous after the [#TexasChurchMassacre](#). Typical [#Trump](#) it doesn't affect me mentality! Is this what he means by making America military strong? Intentionally Arming mental health patients makes his constituents feel warm and fuzzy as they walk about the life?

9:04 AM ·  · Twitter for iPhone

The Link Between Genius and Insanity -
naturibeauty.com/link-between-g... [#psychology](#) [#art](#) [#mentalhealth](#)



The Link Between Genius and Insanity - Psychology of Beauty
 Many of us are familiar with Edvard Munch's famous painting "The Scream", but few know the details of the artist's life and the inspiration for this evocative ...
naturibeauty.com

1:42 AM ·  2016 · 

With respect to bandwagon behavior, stigma-related discourse was

associated with significantly higher levels of bandwagon behavior than the discourse which was stigma-neutral. The mean of cognitive focus for stigma-related discourse was also significantly higher, due to both - higher resolution and intra-topic focus.

Lastly, tweets belonging to stigma-related discourse used significantly fewer hashtags and mentions and were initiated by users with significantly lower social capital. Stigma-related discourse also tended towards higher polarisation.

Looking at the retrievability trends over time, the following four topics were identified as having relatively short-lived attention spikes: 'Social Media and Sharing', 'Anti-stigma awareness', 'Education, Research, Schooling', and 'Stress and Physical symptoms' (see Fig. 2).

The topics with short-lived attention (Table 6) showed significantly lower emotional energy and bandwagon behavior than topics with continuous attention patterns, but higher cognitive focus. The short-lived attention topics also displayed less negative vocabulary and more solidarity and exhibited significantly higher use of hashtags and mentions. Discourse characterized by the hype, moreover, was initiated by users with significantly lower social capital.

3.3. Regression analysis

The generalized Fixed Effect panel regression Model 1 (Table 7) explained 42.1% of the variance in discursive retrievability, which was statistically significant [$F_{(6,1254)} = 151.79, p < 0.001$]. Regression results indicated that we can reject the H_0 for $H1$ since topics with higher emotional energy were persistently driving the discourse ($p < 0.001$), mostly by engagement ($p < 0.001$) and to a lesser extent by high confidence and solidarity ($p < 0.01$). We can accept H_0 for $H2$ since bandwagon behavior had no significant effect on the retrievability. However, we can observe that duplicates within the discourse had a significant negative effect on discursive retrievability, with levels of significance increasing with time ($p_{lag2,3} < 0.05, p_{lag4} < 0.001$). In a shorter time-frame, the cognitive focus had a significant positive effect on the retrievability of the discourse ($p_{lag1} < 0.05$). Rejecting H_0 for $H3$, however, proved difficult, as inter-focus and intra-focus affected the discourse in different directions: we observed a significant positive effect from the intra-focus ($p < 0.001$), a significant negative effect from the inter-focus ($p < 0.001$), and no effect from the resolution. In other words, the within-tweet discursive competition affected the discourse negatively, while higher diversity of mental health related topics on Twitter in general drove more attention to the topics within the discourse.

Looking at controls, using hashtags to thematically tag the discourse had a significant positive effect on the continuation of this discourse ($p < 0.01$). It is also worth noting that, *ceteris paribus*, explained variance of Model 2 has increased with time, meaning that mechanisms which affect the discourse work long-term.

Looking at the potential differences in stigma-related, stigma-neutral, and short-lived attention discursive mechanisms (Table 8), the attention-driving mechanisms for these subsets were very similar to the mechanisms which drive mental health discourse in general and in between themselves. Emotional energy had a significant positive effect. Bandwagon behavior effect was not significant. The notable differences included a high level of emotional arousal and the use of mentions, which seemed to affect stigma-related discourse negatively, while engagement and intra-topic focus lost significance.

For the discourse characterized by short-lived attention, we can accept H_0 for $H4$ since the bandwagon behavior had no significant effect on retrievability. We also observed that emotional intensity ($p < 0.001$) and the duplication rate ($p < 0.05$) could be considered the main drivers. Nevertheless, a given model only explains 18.6% of the variance in what drives the short-lived attention up, meaning that topics characterized by short-lived attention cannot be assessed similarly to the rest of the frames and are probably driven by other unobserved exogenous factors.

Table 6
Descriptive Statistics [Retrievability Patterns] with *t*-test (N = 1,320).

Variable	Short-lived attention pattern		Stable attention pattern		All mental health discourse		Sig.
	Mean	StDev	Mean	StDev	Mean	StDev	
Twitter Characteristics							
Retrievability	503	953	876	973	826	978	***
Replies per tweet	0.12	0.16	0.12	0.12	0.13	0.34	
Favourites per tweet	1.02	3.00	0.91	4.39	0.93	4.22	
Duplication rate	1.11	0.61	1.06	0.34	1.07	0.39	
Tweets per user	1.04	0.43	1.06	0.41	1.09	0.41	
User social capital	7,706	1,560	10,655	12,516	10,261	11,707	***
Hashtags per tweet	0.88	0.82	0.21	0.43	0.30	0.55	***
Mentions per tweet	0.68	0.60	0.16	0.34	0.23	0.43	***
Echo Chamber	0.22	0.31	0.21	0.39	0.21	0.38	
Sentiment Characteristics							
positive	2.42	1.86	2.40	1.68	2.41	1.70	
negative	1.25	1.32	1.55	1.40	1.51	1.39	**
Sentiment	1.17	2.28	0.86	2.11	0.90	2.14	
Affectivity	3.47	2.13	3.70	2.08	3.67	2.09	
Confidence	55.23	21.55	57.52	18.77	57.21	19.17	
Personal	0.54	0.49	0.37	0.38	0.40	0.42	***
Discourse Characteristics							
Intra-topic focus	7.14	1.12	7.42	0.86	7.39	0.90	***
Resolution	5.10	2.95	3.98	2.59	4.13	2.67	***
Resonance Mechanisms							
Emotional Energy	3.04	1.38	3.32	1.35	3.28	1.36	*
Bandwagon	2.60	1.83	3.91	1.65	3.74	1.73	***
Cognitive Focus	3.85	1.06	3.80	0.90	3.81	0.93	**

p* < 0. 05; *p* < 0.01; ****p* < 0.001 (two-sided tests).

4. Discussion and conclusion

In this study, we have empirically demonstrated that mental health discourse is growing, both in absolute terms and relative to total Twitter discourse. From sentiment analysis, we derived that mental health discourse online was timid, of low emotional intensity, and high anxiety. Although over the years the discourse has become characterized by more emotionally charged tweets, especially with regards to positive emotions, it was still heavily tilted towards negative sentiment, further problematizing mental health, whereas positive sentiment often masked sarcasm. Not accounting for sarcasm, online social media mental health discourse exhibited less pronounced stigmatizing tendencies. On the contrary, online mental health discourse was more oriented towards creating awareness about mental health, talking about feelings, and encouraging conversation, possibly due to online environments being more inclusive and built on bottom-up user-generated content. Still, these findings should be taken with caution. Even though the stigma-related discourse on Twitter was low relative to stigma-neutral discourse, it has shown a significant upward trend in the last three years.

With regard to cultural power mechanisms, bandwagon behavior could be considered negligible, only negatively affecting the themes characterized by short-lived attention. Tweet duplicates, moreover, affected the discourse negatively. Either filtered out through Twitter algorithms (Gillespie, 2014) or as an ecological density problem (Van Venrooij, 2015) duplicate tweets diluted the discourse and caused the loss of interest in the discursive frames from one period to another.

Emotional energy, however, could be seen as a driving force behind increased discourse. Emotional energy increased the discourse by engagement ('replies') and tweets that displayed higher levels of solidarity and confidence. The emotional affectivity of the discourse showed to have contributed to increased levels of attention for topics for which the attention was short-lived and affected stigma-related discourse negatively. The theory of cognitive-emotional currents can assist in understanding this observation, suggesting that there might be an emotional saturation point and that emotional neutrality makes the discourse more sustainable than increasingly intensifying emotions (Bail et al., 2017). In the case of stigma-related discourse, however, the cognitive-emotional currents worked to maintain the discourse, as sensationalism takes care of feeding a variety of emotional

ever-changing dramatic news stories. Hence, constantly renewing, the neither emotional nor cognitive content of a discursive frame is ever able to reach a peak and the discourse then rolls over resulting in increased attention and sustainability.

Moving to the cognitive focus, the evidence of its effect on discursive retrievability was conflicting. On a macro level, with the higher number of topics discussed within the general mental health discourse (inter-topic) came better retrievability of a single topic within this discourse. Hence, when mental health discourse could cater to the worldviews of larger audiences, it helped the retrievability of the topics within this discourse (Benford and Snow, 2003; Collins, 2001). On a micro level, on the contrary, tweets that are more subject-focused can be considered more cognitively engaging, driving the retrievability up in accord with the theory of cultural carrying capacity as theorized by Bail (2016). Finally, the use of hashtags affected the discursive retrievability positively, possibly by helping the audiences to (1) better classify a tweet or assign it to a certain cognitive category, making it more (consciously or subconsciously) retrievable for a consequent action, and (2) make certain discursive frames more searchable.

4.1. Academic and societal contribution

In addition to our empirical contributions discussed above, we made several academic contributions. First, this study bridged the macro and micro-levels in analyzing how the online mechanisms of attention generation affect cultural power, also between subsets of the discourse (i.e. stigmatizing vs stigma-neutral, and topics with short-lived attention). In short, the results point to the link between retrievability and resonance as also discussed by McDonnell et al. (2017a, 2017b) where resonance creates a feedback loop that increases retrievability. However, and second, we discovered that there are indeed various types of resonance that are qualitatively different. Some forms of engagement are more habitual and implicit, while others are more intense and explicit (Tavory and Timmermans, 2014). In this research, we have demonstrated that the former – here the bandwagon behavior – is ineffective for creating an action, while the latter - emotional energy - positively contributes to the discourse.

Third, we also learned that intense emotions – attributed to more explicit resonance – are not necessarily a pre-requisite to intense

Table 7

Model summaries of Fixed Effect Panel Regression assessing the mechanisms of retrievability of mental health discourse (N = 1,290).

	Model 1				Model 2			
	Lag1 B (SE)	Lag2 B (SE)	Lag3 B (SE)	Lag4 B (SE)	Lag1 B (SE)	Lag2 B (SE)	Lag3 B (SE)	Lag4 B (SE)
Emotional Energy	390.094 (71.01)***	361.391 (65.98)***	315.8744 (65.46)***	378.387 (63.62)***				
replies					1.83 (0.47)***	1.805 (0.54)**	2.623 (0.33)***	2.966 (0.38)***
affectivity					-0.243(7.72)	-12.64(9.26)	-4.027(10.58)	1.505(6.76)
confidence					8.059(2.64)**	8.503(2.49)**	6.122(1.51)***	6.265(1.58)***
solidarity					219.711(76.33)**	226.792(62.38)***	86.398(83.41)	202.212(62.73)**
Bandwagon Behaviour	8.789 (51.27)	-17.175 (48.12)	-1.032 (41.75)	-41.589 (43.73)				
favourites					-0.005 (0)	-0.004 (0)	-0.001(0)	0.031 (0.03)
duplication rate					-125.195(82.56)	-207.98(81.01)*	-157.213(63.09)*	-123.34(36.94)**
Cognitive Focus	88.982 (37.40)*	-18.359 (37.84)	-38.183 (37.37)	-31.273 (38.94)				
Inter-topic focus					-82.724 (10.07)***	-123.334 (12.31)***	-102.329 (9.36)***	-84.415 (5.57)***
Intra-topic focus					141.644(34.05)***	154.381(29.07)***	102.05(25.66)***	92.87(24.75)***
Resolution					-4.415(8.47)	-12.417(8.14)	-4.974(6.04)	-7.93(5.25)
Controls								
Echo chamber	11.650 (70.37)	59.210 (73.17)	15.745 (73.82)	61.493 (71.01)	33.661 (44.19)	75.309 (51.38)	26.87 (48.59)	50.504 (42.68)
Hashtags	0.564 (0.19)**	0.620 (0.21)**	0.587 (0.24)*	0.397 (0.28)	0.414 (0.15)**	0.429 (0.15)**	0.472 (0.14)**	0.218 (0.16)
Mentions	0.095 (0.47)	-0.045 (0.50)	0.231 (0.57)	0.392 (0.66)	-0.066 (0.26)	-0.234 (0.24)	-0.279 (0.2)	-0.136 (0.24)
N	1,290	1,260	1,230	1,200	1,290	1,260	1,230	1,200
df	6/1,254	6/1,224	6/1,194	6/1,164	12/1,248	12/1,218	12/1,188	12/1,158
R²	42.07%	39.80%	38.98%	38.22%	59.00%	58.68%	63.24%	64.24%
F-test	151.79***	134.89***	127.13***	120.02***	149.666***	144.168***	170.33***	173.32***

^p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001 (two-sided tests).

engagement. Possibly this is because of the sensitive subject matter (mental health), but perhaps heightened emotions could also be seen as something more habitual/implicit and not create emotional energy in itself. It is engagement and the more complex emotions (i.e. portrayals of confidence, solidarity) which seem to drive the discourse.

Fourth, the content has shown to be highly relevant as emotions can reach saturation and interest can only be maintained by focusing on the subject. We propose that cognitive focus is also a form of explicit engagement, where an actor needs to interact with the topic cognitively and explicitly. Cognitive focus can be aided by external factors such as (a) maintaining of the novelty via cognitive-emotional currents or (b) making the message highly understandable by focusing on fewer topics and ensuring optimal cultural carrying capacity (Bail, 2016; Bail et al., 2017). However, cognitive focus as a form of explicit engagement only works on a micro-level (intra-topic focus), as on the macro-level (inter-topic focus) discursive diversity is more helpful to drive the public's interest in the subject.

Our findings could be of use to mental health advocacy organizations and policymakers. Their outreach efforts could be improved by actively engaging in stigma-reducing mental health discourse as it (1) creates awareness even in the absence of a direct contribution and (2) dilutes stigma-inducing topics by creating attention-resource competition.

Moreover, the problem of stigma can be addressed by policy work concerning sensationalism and the use of sarcasm in the stigma-associated discourse, particularly targeting traditional media channels on social media. Lastly, policymakers should be cautious not to idealize stigma-neutral mental health discourse as there can be danger of misinformation by uninformed parties providing inaccurate information and perpetuating myths.

4.2. Limitations and avenues for future research

This research is not without limitations. First, we acknowledge that this research can only tell us about the mechanisms which drive mental health discourse indirectly from the nature of the tweets, rather than claiming what subjective feelings or thoughts triggered by these tweets will point towards the discursive contagion. Second, by relying on available data of tweet characteristics (i.e. texts, 'replies', sentiment), we had to forgo some of the micro-elements of this research. The opportunity for future research lays in the analysis of conversation threads and interaction rituals between users (i.e. how fast some tweets elicit the replies, what is the speed of interaction, etc.). Namely, there is a need to assess to what extent Twitter interactions can show synchronicity. Thirdly, we used a long timeframe. Although our models have a

Table 8

Model summaries of Fixed Effect Panel Regression assessing the mechanisms of retrievability of mental health discourse [thematic] (N = 1,290).

	Stigma-related B (SE)	Stigma-neutral B (SE)	Short-lived attention B (SE)	Stigma-related B (SE)	Stigma-neutral B(SE)	Short-lived attention B (SE)
Emotional Energy	288.544(128.53)*	321.532(72.54)***	150.078(74.08)*			
replies				0.577(0.53)	1.355(0.40)**	-0.072(0.36)
affectivity				-77.695(38.11)*	-0.010(9.89)	30.933(6.68)***
confidence				7.357(5.65)	7.816(3.05)*	-4.789(2.60)*
solidarity				843.598(289.00)**	204.486(98.84)*	450.579(209.62)*
Bandwagon Behaviour	-111.376(130.34)	-76.845(45.36)	-111.347(72.64)			
favourites				-0.003(0)	-0.028(0.02)	-0.003(0.01)
duplication rate				-469.753(290.18)	-149.528(99.72)	86.667(38.02)*
Cognitive Focus	-134.73(86.75)	-72.907(33.57)*	-0.194(27.64)			
Inter-topic focus				-152.621(27.03)***	-107.371(13.62)***	5.659(39.93)
Intra-topic focus				89.774(74.01)	148.197(40.77)***	-41.345(45.56)
Resolution				-4.756(24.91)	-12.191(10.85)	0.703(15.41)
Controls						
Echo chamber	-237.515(148.72)	105.251(62.17)*	-99.399(63.59)	-85.208(169.89)	76.415(46.68)	-84.456(98.175)
Hashtags	3.484(0.95)***	0.601(0.19)**	0.96(0.65)	2.62(0.33)***	0.424(0.17)*	1.012(0.49)*
Mentions	-0.967(0.51)	-0.401(0.39)	-0.282(1.11)	-1.041(0.16)***	-0.231(0.32)	-0.557(1.09)
N	352	968	176	352	968	176
df	6/332	6/920	6/164	12/324	12/912	12/156
R²	55.14%	32.81%	17.08%	61.71%	41.49%	18.60%
F-test	69.040***	76.75***	5.73***	44.452***	55.13***	3.01***

†p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001 (ten-sided tests).

significant predictive power, increasing with longer lags which shows that mechanisms of online attention work long-term, comparison of how these mechanisms work over a shorter timescale (i.e. on a day-to-day basis) would be desirable.

Notwithstanding our methodological contribution of trying to analyze the discourse quantitatively and longitudinally (i.e. by applying mixed methods of topic modeling, sentiment and panel data regression analyses), this research could benefit from cross-checking between different topic modeling algorithms (i.e. Biterm Topic Model specifically designed for short texts, Dynamic Topic Models) and improving the search term optimization to further fine-tune the research. If we were to extend our search terms, we would probably find more stigma because someone might speak about schizophrenia without tagging their tweets with the words “mental health” or #mentalhealth. By applying topic modeling, this research has given us a near-exhaustive list of terms that can be used to further our research about stigma. Current research, however, was instrumental in deriving these search terms. Additionally, future research could also relate our findings to a different empirical reality to assess whether the mechanisms driving the retrievability differ across contexts (i.e. subject, media type, cross-culturally).

Credit author statement

Alina Pavlova: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Visualization, Pauwke Berkers: Validation, Writing - review & editing, Supervision.

Ethics approval/Statement EA not required

No ethical approval was necessary to conduct this study as it is completely based on publicly available Twitter's data. With respect to the data privacy, as Twitter users agree to forgo some confidentiality and anonymity by signing Twitter terms and conditions, no further user consent is required for their data to be used in this research.

Acknowledgements

This research has been undertaken as a part of a Research Masters degree in Cultural Sociology at Erasmus University Rotterdam under the supervision of Dr. Alex van Venrooij (University of Amsterdam, Department of Sociology) who we thank for stimulating discussions and his support in this research project. We would also like to thank Dr. (Jay) JS Lee, Frank Weij, and Rens Wilderom for their help with the Twitter data collection and topic modeling scripts. We want to thank Dr. Callan L. Attwell for his help with the editing of this manuscript.

Preliminary results of the article have been presented at the “Dutch Sociology Day” conference held in Rotterdam (14th June 2018). The authors would like to thank the conference participants for feedback on the earlier version of this article.

References

- Allen, J., Core, M.G., 1997. November). Coding dialogs with the DAMSL annotation scheme. In: AAAI Fall Symposium on Communicative Action in Humans and Machines, vol. 56.

- A, PA dictionary of psychology, 2020. Cognition. American Psychological Association, Washington, DC.
- Austin, J.L., 1962. How to Do Things with Words. University Press, Oxford.
- Bail, C.A., 2012. The fringe effect: civil society organizations and the evolution of media discourse about Islam since the September 11th attacks. *Am. Socio. Rev.* 77 (6), 855–879.
- Bail, C.A., 2016. Cultural carrying capacity: organ donation advocacy, discursive framing, and social media engagement. *Soc. Sci. Med.* 165, 280–288.
- Bail, C.A., Brown, T.W., Mann, M., 2017. Channelling hearts and minds: advocacy organizations, cognitive-emotional currents, and public conversation. *Am. Socio. Rev.* 82 (6), 1188–1213.
- Baltagi, B., 2008. *Econometric Analysis of Panel Data*. John Wiley & Sons.
- Beckman, L., Hellström, L., von Kobyletzki, L., 2019. Cyber bullying among children with neurodevelopmental disorders: a systematic review. *Scand. J. Psychol.*
- Benford, R., Snow, D., 2003. Framing processes and social movements: an overview and assessment. *Annu. Rev. Sociol.* 26, 611–639.
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3 (Jan), 993–1022.
- Campbell, M.A., 2005. Cyber bullying: an old problem in a new guise? *J. Psychol. Couns. Sch.* 15 (1), 68–76.
- Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., Saxena, S., 2016. Scaling-up treatment of depression and anxiety: a global return on investment analysis. *The Lancet Psychiatry* 3 (5), 415–424.
- Collins, R., 2001. Social movements and the focus of emotional attention. In: *Passionate Politics: Emotions and Social Movements*. University of Chicago Press, Chicago, pp. 27–46.
- Collins, R., 2014. *Interaction Ritual Chains*. Princeton university press.
- Corrigan, P.W., Watson, A.C., Barr, L., 2006. The self-stigma of mental illness: implications for self-esteem and self-efficacy. *J. Soc. Clin. Psychol.* 25 (8), 875–884.
- Croissant, Y., Millo, G., 2008. Panel data econometrics in R: the plm package. *J. Stat. Software* 27 (2), 1–43. <https://doi.org/10.18637/jss.v027.i02>.
- Csikszentmihalyi, M., 1996. *Flow and the Psychology of Discovery and Invention*. Harper Collins, New York.
- DiMaggio, P., 1997. Culture and cognition. *Annu. Rev. Sociol.* 23 (1), 263–287.
- DiMaggio, P., Nag, M., Blei, D., 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of U.S. government arts funding. *Poetics* 41, 570–606.
- DiMaggio, P., Bernier, C., Heckscher, C., Mimno, D., 2017. *I. CSSO Working Paper #16 Interaction Ritual Threads: Does IRC Theory Apply Online?*. Retrieved from: <https://csso.princeton.edu/file/251/download?token=1q12ntrd>.
- Dovidio, J.F., Glick, P., Rudman, L.A. (Eds.), 2008. *On the Nature of Prejudice: Fifty Years after Allport*. John Wiley & Sons.
- Gamson, W.A., Modigliani, A., 1994. The changing culture of affirmative action. *Equal Employ. Oppor.: Lab. Mark. Discrim. Publ. Pol.* 373–394.
- Garas, A., Garcia, D., Skowron, M., et al., 2012. Emotional persistence in online chatting communities. *Sci. Rep.* 2, 402. <https://doi.org/10.1038/srep00402>.
- Gillespie, T., 2014. The Relevance of Algorithms. *Media technologies: Essays on communication, materiality, and society*, p. 167.
- IHME, 2010. *Financing Global Health 2010: Development Assistance and Country Spending in Economic Uncertainty*. Institute for Health Metrics and Evaluation, Seattle, WA, 2010.
- Internet Live Stats, 2018. *Twitter Usage Statistics*. Retrieved in May, 2018 from: <http://www.internetlivestats.com/twitter-statistics/>.
- Krackhardt, D., Stern, R.N., 1988. Informal networks and organizational crises: an experimental simulation. *Soc. Psychol. Q.* 123–140.
- Kretschmer, M., Klimis, G.M., Choi, C.J., 1999. Increasing returns and social contagion in cultural industries. *Br. J. Manag.* 10 (s1), 61–72.
- Lee, J.S., Nerghes, A., 2017, July. Labels and sentiment in social media: on the role of perceived agency in online discussions of the refugee crisis. In: *Proceedings of the 8th International Conference on Social Media & Society*. ACM, p. 14.
- Leung, D.K., Lee, F.L., 2014. Cultivating an active online counterpublic: examining usage and political impact of Internet alternative media. *Int. J. Press/Pol.* 19 (3), 340–359.
- Levina, N., Arriaga, M., 2014. Distinction and status production on user-generated content platforms: using Bourdieu's theory of cultural production to understand social dynamics in online fields. *Inf. Syst. Res.* 25 (3), 468–488.
- Lizardo, O., Strand, M., 2010. Skills, toolkits, contexts and institutions: clarifying the relationship between different approaches to cognition in cultural sociology. *Poetics* 38 (2), 205–228.
- Lizardo, O., 2016. Cultural symbols and cultural power. *Qual. Sociol.* 39 (2), 199–204.
- Livingston, J.D., Boyd, J.E., 2010. Correlates and consequences of internalized stigma for people living with mental illness: a systematic review and meta-analysis. *Soc. Sci. Med.* 71 (12), 2150–2161.
- Maloney, P., 2013. Online networks and emotional energy: how pro-anorexic websites use interaction ritual chains to (re) form identity. *Inf. Commun. Soc.* 16 (1), 105–124.
- McCallum, A.K., 2002. MALLETT: A Machine Learning for Language Toolkit. Available at: <http://mallet.cs.umass.edu/>.
- McDonnell, T.E., Bail, C.A., Tavorly, I., 2017a. A theory of resonance. *Socio. Theor.* 35 (1), 1–14.
- McDonnell, T.E., Jonason, A., Christoffersen, K., 2017b. Seeing red and wearing pink: trajectories of cultural power in the AIDS and breast cancer ribbons. *Poetics* 60, 1–15.
- McLeod, A.I., 2011. Kendall: Kendall Rank Correlation and Mann-Kendall Trend Test. R Package Version 2.2. Retrieved from: <https://CRAN.R-project.org/package=Kendall>.
- Meshi, D., Morawetz, C., Heekeren, H.R., 2013. Nucleus accumbens response to gains in reputation for the self relative to gains for others predicts social media use. *Front. Hum. Neurosci.* 7, 439.
- MacKenzie, I.S., Tanaka-Ishii, K., 2007. Text Entry Using a Small Number of Buttons. *Text entry systems: Mobility, accessibility, universality*, pp. 105–121.
- Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K., 2015. *The Development and Psychometric Properties of LIWC2015*. UT Faculty/Researcher Works. Retrieved from: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf.
- Rehurek, R., Sojka, P., 2010. Software framework for topic modeling with large corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Retrieved from: <http://is.muni.cz/publication/884893/en>.
- Röder, M., Both, A., Hinneburg, A., 2015, February. Exploring the space of topic coherence measures. In: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, pp. 399–408.
- Rose, D., Thornicroft, G., Pinfold, V., Kassam, A., 2007. 250 labels used to stigmatize people with mental illness. *BMC Health Serv. Res.* 7 (1), 97.
- Scheibman, J., 2004. Inclusive and exclusive patterning of the English first person plural: evidence from conversation. *Lang. Cult. Mind* 377–396.
- Schmidheiny, K., Basel, U., 2011. Panel data: fixed and random effects. *Short Guides to Microeconometrics*. Retrieved from: <https://www.schmidheiny.name/teachin/gpanel2up.pdf>, 2-7.
- Schudson, M., 1989. How culture works. *Theor. Soc.* 18 (2), 153–180.
- Simony, E., Honey, C.J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., Hasson, U., 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* 7, 12141.
- Stigler, G.J., Becker, G.S., 1977. De gustibus non est disputandum. *Am. Econ. Rev.* 67 (2), 76–90.
- Swidler, A., 1986. Culture in action: symbols and strategies. *Am. Socio. Rev.* 273–286.
- Tavorly, I., Timmermans, S., 2014. *Abductive Analysis: Theorizing Qualitative Research*. University of Chicago Press.
- Thornicroft, A., Goulden, R., Shefer, G., Rhydderch, D., Rose, D., Williams, P., et al., 2013. Newspaper coverage of mental illness in England 2008–2011. *Br. J. Psychiatr.* 202 (s55), s64–s69.
- Twitter, 2017. *Number of Characters*. Retrieved from: https://blog.twitter.com/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html.
- Twitter, 2018a. *How to Use Hashtags*. Retrieved from: [Twitter Help Centre on 10 July from: https://help.twitter.com/en/using-twitter/how-to-use-hashtags](https://help.twitter.com/en/using-twitter/how-to-use-hashtags).
- Twitter, 2018b. *About Replies and Mentions*. Retrieved from: [Twitter Help Centre on 10 July from: https://help.twitter.com/en/using-twitter/mentions-and-replies](https://help.twitter.com/en/using-twitter/mentions-and-replies).
- Van Herpen, E., Pieters, R., Zeelenberg, M., 2009. When demand accelerates demand: trailing the bandwagon. *J. Consum. Psychol.* 19 (3), 302–312.
- Van Venrooij, A., 2015. A community ecology of genres: explaining the emergence of new genres in the UK field of electronic/dance music, 1985–1999. *Poetics* 52, 104–123.
- Wahl, O.F., 1997. *Media Madness: Public Images of Mental Illness*. Rutgers University Press.
- Whiteford, H.A., Ferrari, A.J., Degenhardt, L., Feigin, V., Vos, T., 2015. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PloS One* 10 (2), e0116820.
- Whitley, R., Wang, J., 2017. Good news? A longitudinal analysis of newspaper portrayals of mental illness in Canada 2005 to 2015. *Can. J. Psychiatr.* 62 (4), 278–285.
- Wicks, P., 2014. The ALS Ice Bucket Challenge—Can a splash of water reinvigorate a field? *Amyotrop. Lateral Scler. Frontotemporal Degeneration* 15 (7–8), 479–480.
- WHO, 2016. *Data and Statistics*. Retrieved from: [World Health Organization Europe. http://www.euro.who.int/en/health-topics/noncommunicable-diseases/mentalhealth/](http://www.euro.who.int/en/health-topics/noncommunicable-diseases/mentalhealth/). <http://www.euro.who.int/en/health-topics/noncommunicable-diseases/mentalhealth/data-and-statistics>.
- Xu, W.W., Zhang, C., 2018. Sentiment, richness, authority, and relevance model of information sharing during social Crises—the case of # MH370 tweets. *Comput. Hum. Behav.* 89, 199–206.