



ARTICLE



<https://doi.org/10.1057/s41599-020-0513-5>

OPEN

Phrase-level pairwise topic modeling to uncover helpful peer responses to online suicidal crises

Meng Jiang ^{1✉}, Brooke A. Ammerman², Qingkai Zeng¹, Ross Jacobucci² & Alex Brodersen²

Suicide is a serious public health problem; however, suicides are preventable with timely, evidence-based interventions. Social media platforms have been serving users who are experiencing real-time suicidal crises with hopes of receiving peer support. To better understand the helpfulness of peer support occurring online, this study characterizes the content of both a user's post and corresponding peer comments occurring on a social media platform and present an empirical example for comparison. It introduces an approach that uses pairwise topic models to transform large corpora of discussion into associated topics of user and peer posts. The key advantages of this approach include: (i) modeling both the generative process of each type of the corpora (i.e., user posts and peer comments) and the associations between them, and (ii) using phrases, which are more informative and less ambiguous than words, in addition to words, to represent social media posts and topics. The study evaluated the method using data from Reddit *r/SuicideWatch*. It examined how the topics of user and peer posts were associated and how this information influenced the peer perceived helpfulness of the peer support. Then, this study applied structural topic modeling to data collected from individuals with a history of suicidal crisis as a means to validate findings. The observations suggest that effective modeling of the association between the two lines of topics can uncover helpful peer responses to online suicidal crises, notably providing the *suggestion* of professional help. The proposed technology can be applied to "paired" corpora in many applications, such as technical support forums, question-answering sites, and online medical services.

¹ Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA. ² Department of Psychology, University of Notre Dame, Notre Dame, IN, USA. ✉email: mjiang2@nd.edu

Introduction

As suicide rates continue to rise across the United States (for Disease Control, 2018), an increasing number of individuals are also experiencing non-lethal suicidal thoughts and behaviors (i.e., suicidal ideation, suicide attempts). To support individuals through suicidal crises, and ultimately prevent deaths by suicide, national crisis resources have become increasingly important (i.e., National Suicide Prevention Lifeline); however, due to concerns of call volumes exceeding available resources (Service, 2018), resulting in delayed assistance, and fears of unwanted consequences in using such services (e.g., forced hospitalization (Drum et al., 2009)), many individuals in crisis are instead turning to informal crisis resources, such as peer support.

A leading avenue for peer support of stigmatized topics, such as suicidality (Berry et al., 2017; Birnbaum et al., 2017), is social media platforms. Reddit is one such, widely used platform with more than 330 million active users. Like many other platforms, Reddit depends almost solely on crowd-sourced interactions, which occur on topic-based conversation “threads”, with minimal involvement from moderators (Reddit, 2018). Within Reddit, there is a growing thread called *r/SuicideWatch* (n.d., 2019), where users who are experiencing real-time suicidal crises post messages with hopes of receiving peer support. Theories of suicidal behavioral may aid in understanding the potential impact of online peer support via platforms like *r/SuicideWatch*. Despite the existence of several suicide theories (i.e., Interpersonal Psychological Theory (Cowgell, 1977); Three-Step Theory (Simon et al., 2001)), the Fluid Vulnerability Theory was one of the first to articulate the temporal dynamics of suicidality (Rudd, 2006), a process that is particularly relevant for understanding the process, and importance of timely, online peer support (American Foundation for Suicide Prevention, 2019). More specifically, this theory postulates two phases, or dimensions, of suicidality. First, is baseline risk that is one’s chronic or stable level of suicide risk, which may be influenced by factors such as genetics or demographics. Next, is acute risk, which represents factors that fluctuate in response to one’s environment (i.e., interpersonal stress, job loss) or internal experiences (i.e., hopelessness, depressed mood). The interaction between these can then lead to person-specific experience of “suicide mode”, which is dynamic time-dependent process (Rudd, 2006) where the potential to engage in suicidal behavior is high. It is during this acute experience of suicide mode that an individual experiences elevated distress, and thus may seek external support (i.e., online peer support). Indeed, crisis resources, formal and informal, are centered around this idea of a dynamic suicide mode that can be intervened upon (Woodward and Wyllie, 2016).

For many individuals, online peer support is an acceptable and often effective avenue for coping with mental health symptomatology (Ali et al., 2015; Horgan and Sweeney, 2010; Naslund et al., 2019); this may be particularly salient for those struggling with sensitive mental health conditions (Gowen et al., 2012; Haker et al., 2005; Kummervold et al., 2002), like suicidality. More specifically, individuals with severe mental illness may be more likely to seek online peer support (Ciarrochi et al., 2003; Gowen et al., 2012; Hope et al., 2005). Further underscoring their importance, those interactions may be used as a catalyst to seek formal care for mental health difficulties (Lawlor and Kirakowski, 2014). A specific platform for online peer support among those experiencing acute suicidal crises, is *r/SuicideWatch*. Despite the fact that *r/SuicideWatch* may serve as a primary suicide crisis intervention for some, there has been limited research aimed to understand both the content and quality of support received by individuals via *r/SuicideWatch*. Research that has been conducted in this area has worked to improve automatic suicide risk

determinations of users posting as a way to triage resources (Althoff et al., 2016; Coppersmith et al., 2018; De Choudhury et al., 2016; DeMasi et al., 2019; Milne et al., 2016, 2019; Morris, 2015; Shing et al., 2018) or have provided descriptive information on the elements of a user’s posts that may elicit peer responses (De Choudhury and De, 2014; Huang and Bashir, 2016), rather than evaluating the quality (i.e., helpfulness) of the peer support actually received. This particular avenue of research is important given evidence that appropriate support can mitigate crisis states (Turner et al., 1983), and may even ultimately prevent the transition from suicidal crisis to suicide behavior (Woodward and Wyllie, 2016). As such, the current study had two primary aims. First, we aimed to characterize the content of both a user’s post and corresponding peer comments as a way to better understand the peer support interactions occurring on *r/SuicideWatch*. Furthermore, we examined how the topics of user and peer posts were associated and how this information influenced how helpful peers perceived the given peer support, which was quantified by the number of “upvotes,” or likes, a specific peer post receives (i.e., “Study 1”). This metric of peer perceived helpfulness is particularly important on Reddit, where the visibility of a peer comment is based on upvotes: peer comments with the greatest number of upvotes appear at the top of the thread. Second, we aimed to validate the findings related to peer perceived helpfulness by presenting empirical data from individuals with a personal history of suicidal crisis (i.e., “Study 2”). Such findings may highlight ways that peer support on *r/SuicideWatch* can be leveraged, or improved, to best support individuals through suicidal crises.

As prior research on the use of *r/SuicideWatch* for the resolution of suicidal crises may be limited by available methodologies, our key methodological contributions are as follows: (i) We introduce PairwiseLDA, a new topic model that allows for analyzing the topics of users’ posts, the topics of peers’ comments, and *their paired relation*. The pairwise topic model discovers a set of multiple topics of posts from users experiencing suicidal crises and another set of multiple topics of the associated peer comments. (ii) Documents were represented as bags of words and *phrases*. PairwiseLDA discovers representative words and phrases, giving a more comprehensive description for each topic. (iii) When topics have been captured, comparisons between the volume (defined by the number of comments) and peer perceived helpfulness (defined by the number of “upvotes”) of peer comment topics by user post topics. Findings also offer advances in both the practical and theoretical understanding of online peer support for suicidal crises. Indeed, the present findings contribute to the theoretical understanding of the proximal risk factors for suicidal crises and have the potential to identify aspects of peer support that may serve to reduce suicidal crisis intensity, having implications for educational efforts aimed at improving informal crisis resources. Furthermore, beyond the topic of suicide, our technology can be applied to “paired” corpora in many applications such as tech support forums, question-answering sites, and online medical services, if data are available for readers.

Study 1: Results

Data collection. Data was scraped using the Reddit API from *r/SuicideWatch* channel across a period of 18 months, ranging from December 1, 2013 to May 31, 2015. All procedures were approved by the authors’ institutional review board for research ethics (Protocol #19-06-5390). We extracted original user’s posts (i.e., individuals experiencing a suicidal crisis), subsequent peer comments (i.e., received peer support responses), and the number of upvotes for peer comments (i.e., peer perceived helpfulness of

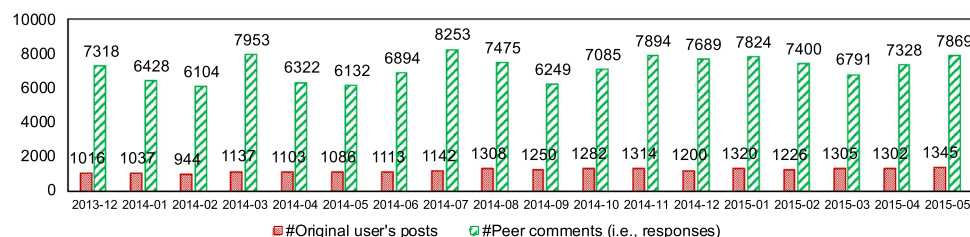


Fig. 1 The number of original user's posts and peer responses from the Reddit *r/SuicideWatch* forum across 18 months.

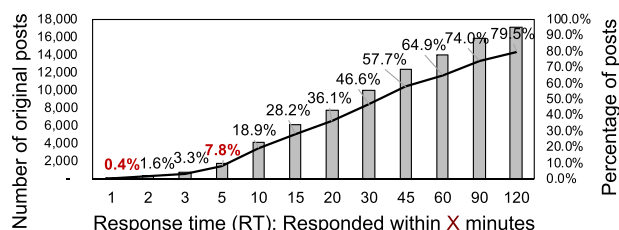


Fig. 2 Very few original posts were timely responded: fewer than 1% posts received at least one response within 1 min. Fewer than 10% were responded within 5 min.

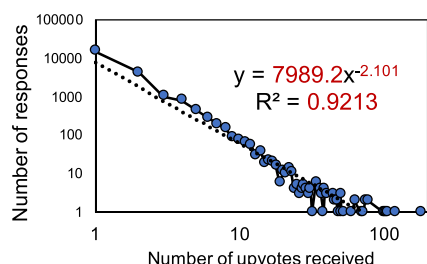


Fig. 3 Very few peer responses had more than a couple of upvotes: the long-tail distribution fits power law. Only 6.1% had at least two upvotes.

responses). Of note, upvotes are given by any user of the *r/SuicideWatch* forum to comments that are perceived to be particularly helpful or useful in response to the original user's post, and have a direct impact on the visibility of that peer's comment. There were a total of 21,430 original posts and 129,008 peer comments. Each original post was found to have received 6.02 peer comments, on average. The number of original user's posts and the number of peer comments by month are detailed in Fig. 1. The number of original posts increased across the time span, as the forum had more and more users.

We calculated the response time (RT) for peer comments, defined as the length of time taken for the first peer comment in response to the original post. Figure 2 presents the number of posts and percentage of posts vs. RT. The curve remained at a very low level before 5 min and then rose steadily thereafter. Only 0.4% posts received the first response within 1 min, and fewer than 8% posts were responded within 5 min.

Figure 3 presents the distribution on the number of upvotes of responses in logarithm-logarithm scale. The high R^2 value shows the distribution follows *power law* (Adamic, 2000) ($y = ax^{-k}$) with the law's exponent of $k = 2.101$. Most responses had no or very few upvotes. Only 18.0% responses had at least one upvote and only 6.1% responses had at least two upvotes.

Topics of original user's posts. Our method identified five topics among original users' posts. Each post document has a topic distribution and is assigned to the topic of the highest probability. Then for each topic, the *volume* (i.e., number and percentage of

posts) was determined. Table 1 presents the descriptive information for each topic, representative words defining each topic, and phrases found within each topic. Table 2 gives an example post with the greatest percentage of words from each topic. Based on this information, we provided a label for each identified topic. The largest topic (i.e., of the highest volume) was labeled *psychological pain*. As can be noted by the representative words (such as curse words) and phrases (i.e., "I do not want to live", "I want to die"), these posts represent significant psychological pain closely associated with expressions of suicidal crises, which were often not event specific but instead broader emotional states. The second largest topic was labeled *relationship stress*, which included representative words of "relationship", "friend", and "love", in addition to phrases of "talk to me" and "talk to someone". This topic predominantly focused on recent disruptions in friendly or romantic relationships. The third largest topic was labeled *psychiatric disorder* and included representative words of "depression", "anxiety" and "panic", and phrases of "intrusive thoughts", "panic attack", and "suicide attempt." In contrast to broader psychological pain represented by the first topic, this topic appeared to be focused on specific psychological problems or diagnoses that were contributing to the user's current state. The fourth topic was labeled *academic difficulty*. Posts representative of this topic used words such as "college", "class" and "grade", and phrases such as "go to college" and "first year" to discuss current difficulties experienced while enrolled in or the prospective of enrolling in high school or college (CBS News, 2018). The last topic, which was the smallest proportion of user's posts, was labeled *financial stress* (U.S. News, 2015). The representative words (i.e., "money", "afford", "debt") and phrases (i.e., "enough money", "lost my job") highlight acute and chronic financial difficulties an user was discussing as part of his/her current crisis state.

Topics of peer responses. Our method identified ten topics among peer comments. Table 3 presents the descriptive information of each peer comment topic, key words defining each topic, and phrases found within each topic. Table 4 gives an example comment with the greatest percentage of words from each topic. The largest topic was labeled *asking questions*. Peer comments queried for additional information presented in original user's posts (representative phrases: "do you think", "do you mind"), which often seemed to inform suggested problem-solving strategies. The second largest topic was labeled *communication support*, which was best characterized by peers suggesting or offering time to talk (representative phrases: "you need someone to talk", "talk to me"). The third topic was labeled *academic encouragement*. Peer comments representative of this topic provided specific accounts or encouragement related to high school and/or college experiences (representative phrases: "high school", "community college"). The next three topics (i.e., *interest development*, *life meaning*, *distraction*, and *entertainment*) represented <10% of peer comments each, with the remaining four topics (i.e., *professional help suggestion*, *relationship/loss support*, *thanks and*

Table 1 Five topics in original user's Posts, their volume, representative words, and phrases.

Topic	Volume	Representative words	Representative phrases
P.PP. Psychological pain	22,075 (41.2%)	Hate, f**k, f**king, h**l, worthless, s**t, b***s**t, blame, d**n, pain, etc.	I do not want to live, i do not want to be, bad person, i just want to die, reason to live, care about me, etc.
P.RS. Relationship stress	13,682 (25.5%)	Relationship, friend, love, girlfriend, boyfriend, close, talk, talked, etc.	Best friend, good friend, close friend, talk to me, talk to someone, talk to him, talk to her, etc.
P.PD. Psychiatric disorder	9606 (17.9%)	Depression, depressed, depressants, anxiety, anxious, panic, disorder, etc.	Suicidal thoughts, intrusive thoughts, suicide attempt, panic attack, social anxiety, severe anxiety, etc.
P.AD. Academic difficulty	5990 (11.2%)	School, college, university, class, high school, semester, grade, course, etc.	Middle school, high school, community college, public school, grad school, go to college, first year, etc.
P.FS. Financial stress	2268 (4.2%)	Worked, jobs, boss, money, afford, paid, paying, debt, bills, insurance, etc.	Child support, enough money, health insurance, new job, student loans, lost my job, place to live, etc.

Table 2 Representative excerpts of the five topics in original user's Posts.

Topic	Representative excerpt
P.PP. Psychological pain	Nothing is as beautiful as it should be. I want to see and be beautiful but I'm surrounded by ugly things. All my life I've been surrounded by ugly things and cruel people...
P.RS. Relationship stress	...the girl I liked started ignoring me completely, that itself isn't much of a problem but then one of my closest friends suddenly became my enemy and she brought a lot of people against me...
P.PD. Psychiatric disorder	...I get panic attacks quite frequently and am getting really paranoid and scared about getting them... My question for this is how can I make panic attacks less frequent or stop them?...
P.AD. Academic difficulty	I'm a 16-year-old guy in high school. When i signed up for freshman year my parents made sure I took the hardest classes... I'm at the point where I cannot stay in this damn room any longer...
P.FS. Financial stress	Been homeless for a few weeks now, quit everything to take care of my mother and she kicked me out, burned through what little money I could get to with only my ID haven't showered...

Table 3 Ten topics in peer Responses, their volume, representative words, and phrases.

Topic	Volume	Representative words	Representative phrases
R.AQ. Asking questions	16,798 (31.3%)	Like, feeling, thoughts, etc.	Do you think, do you mind, sounds like, etc.
R.CS. Communication support	11,239 (21.0%)	Hear, listen, care, know, etc.	You need someone to talk, talk to me, etc.
R.AE. Academic encouragement	6062 (11.3%)	School, degree, class, etc.	High school, community college, etc.
R.ID. Interest development	5259 (9.8%)	Hobby, beautiful, etc.	Meeting new people, great things, etc.
R.LM. Life meaning	4472 (8.3%)	Believe, matter, life, etc.	Only thing, only reason, live your life, etc.
R.DE. Distraction and entertainment	3967 (7.4%)	Game, play, food, music, etc.	Video games, tv shows, spare time, etc.
R.PH. Professional help suggestion	2355 (4.4%)	Hospital, psychologist, etc.	Professional help, suicide hotline, etc.
R.RS. Relationship/loss support	1667 (3.1%)	Boyfriend, deserve, etc.	Loved ones, sorry for your loss, etc.
R.TA. Thanks and appreciation	917 (1.7%)	Thanks, appreciation, etc.	Thank you so much, good night, etc.
R.TM. Treatment and medication	886 (1.7%)	Therapy, doctor, health, etc.	Mental illness, erectile dysfunction, etc.

appreciation, treatment, and medication) each representing <5% of peer comments.

Topic associations on volume between original user's posts and peer responses. Figure 4 presents a heatmap of the distribution of volume of specific peer comment topics to specific user's post topics. The cell values are percentages; for example, the cell "32.38" means that among all the peer comments to the user post topic of *psychological pain*, 32.38% were assigned to the topic *asking questions*. For each peer response topic, the green fonts highlight the highest volume among the original post topics. We summarize Fig. 4 in the following five points: (1) Peer comment topics of *interest development* and *life meaning* had a higher representation in response to the user post topic of *psychological pain* than other original user's post topics. (2) Peer comment topics of *communication support* and *relationship/loss support* had a higher representation in response to the user post topic of *relationship stress* than those of other user post topics. (3) Peer comment topics of *asking questions* and *treatment and medication* had a higher representation in response to user post topics of *psychiatric disorder* than other user post topics. (4) Peer comment

topics of *academic encouragement* had a higher representation in response to the user post topic of *academic difficulty* than other user post topics. (5) Peer comment topics of *distraction and entertainment* and *professional help suggestion* had a higher representation in response to the topic of *financial stress* than other user post topics.

Topic associations with peer perceived helpfulness between original posts and peer responses. We examined two potential factors in generating a response perceived as helpful by peers (i.e., received a greater number of upvotes). The first factor is *comment length*. The average number of words per original user's post was 272.2, with a median of 183. The average number of words per peer comment was 75.2, with a median of 44. This is because users need more words to describe their situations in original posts than to respond with peer support. Writing a peer comment depends on many factors such as the respondent's background, personality, and available time. On the other side, the average number of upvotes per comment was 0.53. The correlation between comment length and upvote was *NOT significant*,

Table 4 Representative excerpts of the ten topics in peer Responses.

Topic	Representative excerpt
R.AQ. Asking questions	Why, for example, do you feel that you're acutely aware of your mortality? What kinds of things in the past have you tried that have helped you cope before?
R.CS. Communication support	I want to be able to help you. Is there anything I can do to help? If you'd like to just vent or if you just want to be able to talk to someone, I'm willing to listen.
R.AE. Academic encouragement	...it may be hell but once you're done you will have a extremely successful life... I have a friend... he quit school after 2 months. He is already starting to regret it...
R.ID. Interest development	...you need to find a hobby could give you something to live for. What kinds of things do you enjoy doing outside of work, like hobbies or volunteer activities?
R.LM. Life meaning	Instead of reasons why not to, perhaps think of reasons to live, reasons to care. Because as long as you have 1 reason to live, you have no reason to die...
R.DE. Distraction and entertainment	What is your favorite game? I am a huge nerd and I try to incorporate gaming into that but I am pretty terrible at most games.
R.PH. Professional help suggestion	Have you tried getting any professional help? Talking to a therapist or may be considered consulting a psychiatrist?
R.RS. Relationship/loss support	I am really sorry for your loss... You don't have to leave her. If you're best friends, just sit down and tell her what you told us...
R.TA. Thanks and appreciation	Thank you. I appreciate you taking the time, your kind words and advice.
R.TM. Treatment and medication	Are you currently being treated for depression (meds, therapy, etc)? Hey, if you are up to it, I recommend medication for anxiety. It worked for me...

	R.AQ	R.CS	R.AE	R.ID	R.LM	R.DE	R.PH	R.RS	R.TA	R.TM
PPP	32.38	18.99	10.44	12.75	9.89	6.76	3.09	2.89	1.34	1.47
P.RS	28.96	26.39	9.86	7.13	8.19	6.90	4.93	3.99	2.47	1.18
P.PD	36.18	19.83	9.71	6.86	6.27	7.63	6.06	2.71	1.62	3.13
P.AD	28.08	17.76	20.53	10.45	6.03	8.18	4.22	2.59	1.15	1.00
P.FS	23.46	20.55	10.85	8.11	9.04	13.54	7.14	3.04	2.56	1.72

Fig. 4 The heatmap table presents the percentage of a topic of peer comments (i.e., each column) for each topic of original user's posts (i.e., each row). Darker background colors represent peer response topics occurring at a higher percentage to a specific user's post topic (i.e., **a** communication support R.CS and relationship/loss support R.RS to relationship stress P.RS, **b** treatment and medication R.TM to psychiatric disorder P.PD, and **c** academic encouragement R.AE to academic difficulty P.AD), and green fonts represent the peer response topic with the highest volume by user's post topic.

	R.AQ	R.CS	R.AE	R.ID	R.LM	R.DE	R.PH	R.RS	R.TA	R.TM	Average
PPP	0.42	0.46	0.37	0.42	0.53	0.45	0.29	0.44	0.36	0.20	0.43
P.RS	0.56	0.58	0.46	0.49	0.71	0.42	0.79	0.41	0.49	0.43	0.56
P.PD	0.41	0.45	0.40	0.36	0.48	0.38	0.51	0.82	0.42	0.26	0.43
P.AD	0.45	0.47	0.52	0.41	0.47	0.40	1.17	0.44	0.23	0.17	0.49
P.FS	0.66	0.64	0.41	0.62	0.95	1.27	0.79	2.43	0.90	0.67	0.80
Average	0.46	0.51	0.43	0.43	0.58	0.49	0.62	0.57	0.44	0.28	

Fig. 5 The heatmap table presents the average number of upvotes that a topic of comments receive when responding to a topic of original posts. Response topics such as relationship/loss support, professional help suggestion, and treatment and medication show significant patterns on response helpfulness.

$r = 0.036$. Therefore, writing a long comment was not more likely to receive a greater number of upvotes.

The second factor is *topic associations*. Figure 5 presents a heatmap of the distribution of upvotes per peer comment topics by original user's post topics. Given a pair of a specific original post topic and a specific peer comment topic, the cell value is the average number of upvotes for the peer comment topic. For each row, the highest values are colored in green and the lowest is in red. The figure has an additional row/column to give the average upvote of each user post/peer comment topic. We make the following observations:

- The peer comment topic of *treatment and medication* received the least number of upvotes (0.28) compared to other peer comment topics in response to three of the five original user's post topics.
- The original user's post topic of *financial stress* received responses with the highest number of upvotes (0.80 vs. the overall average of 0.53), though it received a variety of responses as shown in Fig. 4.
- The original user's post topic of *relationship stress* received responses with the second highest number of upvotes. Interestingly, the response topic with the lowest number of

- upvotes (i.e., 0.41) was *relationship/loss support*, despite being one of the most frequently used topics. On the other hand, the comment topic of *professional help suggestion* had the highest average number of upvotes in response to *relationship tress* (i.e., 0.79).
- The original user’s post topic of *psychiatric disorder* mainly (i.e., >36%) received peer comments of *asking questions*, however, the average number of upvotes was 0.41, below the overall average 0.53. The highest number was 0.82 by the peer response topic of *relationship/loss support* (2.7% of the peer comments to *psychiatric disorder*).
 - The original user’s post topic of *academic difficulty* received peer comment topics of *academic encouragement* (average number of upvotes was 0.52); however the peer response topic with the most upvotes was *professional help suggestion* with an average of 1.17 upvotes. Of note, this was only 4.2% of the peer comments in response to *academic difficulty*.
 - The peer comment topic of *life meaning* received the highest number of upvotes, on average, (0.53) in response to user topic of *psychological pain*, however, made up only 9.9% of the peer comments.
 - According to Table 3, the three *most frequent* peer comment topics were *asking questions*, *communication support*, and *academic encouragement*. Figure 5 shows, however, that the three *most helpful* peer comment topics were *professional help suggestion*, *life meaning*, and *relationship/loss support*. The three most frequent peer comment topics make-up 63.6% of all peer comments, whereas the three most helpful peer comment topics make-up only 15.8% of the peer comments.

Study 2: Results

Data collection. Data was collected from an independent online sample of individuals from the community via Amazon Mechanical Turk (mTurk). As part of mTurk, potential participants are allowed to browse available tasks (i.e., focus groups, marketing surveys, research studies), and select those that they are interested in completing. Participants were recruited as part of a larger study advertised as, “A Research Study on Emotion, Health, and Suicidality;” participants were not selected based on history of suicidality. For the current analysis, one question regarding lifetime history of suicidal ideation (which was used as an indicator of personal history of a suicidal crisis) and one open-ended question were utilized. Participants were asked, “If you had a friend that was going through a difficult time, what advice would you give?”, in addition to, “How would your advice differ if they had thoughts of suicide?” The current study utilized responses from only the latter half of the question (pertaining to suicide), to increase the equivalency of responses to peers responding on the *r/SuicideWatch* forum. All procedures were approved by the authors’ institutional review board for research ethics (Protocol #18-12-5050).

The final sample consisted of 491 participants, of which 42.16% who reported a lifetime history of suicidal crisis (i.e., suicidal ideation; *n* = 207). For analysis of participant responses, stop-words and punctuation were removed, and only words that were used greater than five times were included.

Topics associated with suicidal ideation. We used the structural topic model (Roberts et al., 2013), using history of suicidal crisis as a covariate to explain topic prevalence. Analyses are based on the assumption that an individual’s response is indicative of advice they view as helpful. That is, if an individual is offering this advice to a friend struggling with thoughts of suicide, it is likely they believe their response would be helpful or supportive. Further, through the structural topic model, results will demonstrate

Table 5 Twelve topics from the follow up study, along with the coefficient values and p-value. The association is considered to be significant when p-value is smaller than 0.05.

	Representative words	Coefficient	p-value
1	Seek, love, needed, recommend, suicide	−0.02	0.03 (<0.05)
2	Call, hotline, thing, friend, crisis	−0.01	0.42
3	Professional, live, contact, feeling, suicidal	−0.03	0.001 (<0.05)
4	Thoughts, family, stay, differ, medical	−0.001	0.83
5	Encourage, things, care, suggest, therapy	0.02	0.01 (<0.05)
6	Suicide, find, hospital, change, situation	0.03	0.001 (<0.05)
7	Problem, temporary, permanent, solution, suicide	0.004	0.78
8	Life, people, god, offer, trust	−0.03	0.02 (<0.05)
9	Therapist, immediately, advise, good, leave	0.02	0.02 (<0.05)
10	Advice, give, time, dont, person	0.025	0.024 (<0.05)
11	Talk, feel, make, bad, times	0.011	0.35
12	Worth living, loved, listen, matter	−0.02	0.016 (<0.05)

if topics differ based on individual history of suicidal crisis, highlight any similarities, or discrepancies, with peer perceived helpfulness (i.e., upvotes) on the Reddit *r/SuicideWatch* forum.

Based on the correlated topic model, we tested 2–15 topics, using hold out likelihood (on random subset of 100 observations), exclusivity, and semantic coherence to select a final number of topics. The results of this analysis identified 12 topics, eight of which were significantly associated with suicidal crisis history (see Table 5). We note that in interpreting these coefficients, positive values indicate they were more likely to be used among those with a history of suicidal crisis. As it is difficult to discern topic content solely from the most representative words, we depict the most representative responses for each of the significant topics, displayed in Table 6.

Eight topics were significantly associated with suicidal crisis history; we focus our topic interpretations here. Four topics were identified as less likely to be used among those with a history of suicidal crisis: topics 1, 3, 8, and 12. These topics convey the notion of a need to seek professional help (topics 1 and 3) and offer suggestions of reasons for living, including seeing one’s own worth or having faith in a higher power (topic 8), in addition to the pain they may cause others by dying (topic 12). Similarly, four topics were identified as more likely to be used among those with a history of suicidal crisis. While some of these topics also centered around professional help, they appear to suggest or encourage professional help, and perhaps, more specifically, therapy (topic 5). Further, topics provided a potential context or understanding for why someone might feel suicidal (topic 6) and offer more direct or immediate support for the individual (topic 9). A final topic emerged that suggested individuals with a history of suicidal crisis would not differ in the advice they would give to a friend having a hard time more generally vs. thinking about suicide (topic 10).

Discussion

The current study introduced a new methodological framework that allowed us to conduct a nuanced investigation of peer support interactions in times of suicidal crisis occurring on Reddit *r/SuicideWatch*, in addition to responses from participants in an

Table 6 Most representative responses from each of the significant topics.

Representative excerpt	
1	"I would recommend that they seek professional help"
3	"I would tell them to contact a professional to help them or, if necessary, I would contact someone on their behalf".
5	"I would encourage them to seek professional help...therapy"
6	"It would depend entirely on the situation If the thoughts of suicide were related to a terminal illness, for example, it doesn't seem appropriate to interfere In other plausible scenarios, I couldn't conceive of another situation where I would be approached by another person for this sort of serious advice, and if I were I would demure and avoid interacting to the extent possible by the situation"
8	"I think I would try to help them see how worthy they are of life People make small differences in other people's lives, sometimes they don't even realize that they've made an impact Helping someone see that someone else's life is better just because of some encounter that they have shared with that person"
9	"I'd tell them they have to call me every hour or that I will call 911 immediately, and that I love them, and it would really suck if they killed themselves, and to call their therapist immediately"
10	"The advice wouldn't differ at all I would give the same advice"
12	"I would tell them they are loved and too many people will miss them if they kill themselves I would also tell them that there in nothing in life that is worth not living"

independent sample. Utilizing Reddit *r/SuicideWatch* data, we aimed to (1) characterize both the user's posts (individuals experiencing a suicidal crisis) and peer comments in response to users' post; (2) examine the association between topics discussed in users' posts and peer comments; and (3) investigate how this may inform the peer perceived helpfulness (as indicated by upvotes) of peer comments. Empirical findings from an independent sample of individuals with a personal history of a suicidal crisis were also examined to further inform the helpfulness of peer comment topics provided on the Reddit *r/SuicideWatch* forum. Methodologically, the findings from this study support the use of the newly developed method *PairwiseLDA* and highlight its use to advance our understanding of online peer support during suicidal crises. Substantially, results provide further evidence of topics, or triggers and acute states, that may be directly contributing to suicidal crises. The current study also highlights the ways in which Reddit *r/SuicideWatch* is supporting individuals through suicidal crises, while providing potential opportunities for improvement to better resolve suicidal crises among individuals electing to use informal crisis resources.

Findings demonstrated that five different topics were discussed by original users (i.e., individuals experiencing suicidal crises). These topics support the Fluid Vulnerability Theory (Rudd, 2006) of suicide. More specifically, the current findings lend insight to the acute factors that may be interacting with individuals' baseline risk, leading to a heightened experience of suicidal crisis (i.e., suicide mode). While the findings may be specific to those individuals who opt to seek online peer support, they are consistent with acute factors proposed in theory. Indeed, there is a direct relation between each of the five topics and factors theorized to contribute to suicide mode, including triggers (i.e., relationship stress, academic difficulties, and financial stress), cognition (i.e., psychological pain ["blame"]) and emotions (i.e., psychiatric disorders ["depression", "anxiety"]; psychological pain ["hate?"]). Beyond theoretical support, these findings also highlight factors that may be particularly relevant for individuals who seek online peer support, but have received relatively limited attention in the literature. For example, academic difficulties was a predominant topic in original user's posts with 11% of users discussing this triggers. Echoing this, findings are consistent with the empirical literature demonstrating broad risk factor categories that contribute to suicide outcomes (i.e., suicide attempts (Franklin et al., 2017)). Similarly, these findings may help elucidate the specific experiences within the broad risk categories that are important. For example, social factors have been identified as one of the risk categories with the greatest weighted hazard ratios in predicting suicide attempts (Franklin et al., 2017), and the current findings

may help elucidate relationship stress as a particularly relevant social factor.

Peer comments, on the other hand, displayed a variety of topics discussed, from content specific topics (i.e., academic encouragement) to broader support (i.e., emotional support). On average, peers were more varied in their topics discussed compared to original user's, with four of the topics each consisting of <5% of the overall peer comments. There are a few potential explanations for this. First, it is possible that this reflects a discomfort and uncertainty among peers regarding the most appropriate response to a suicidal crisis (underscoring asking questions as being the most common peer comment topics). This is also indirectly supported by a long-held finding that talking about suicide increases psychological and physiological distress among the non-suicidal individual (Cowgell, 1977). However, it is unclear if peers providing support on *r/SuicideWatch* have a personal history of suicidal crises. As such, it is also possible that peers have reflected on their own experiences of times in (suicidal or non-suicidal) distress, identifying that personalized support during crises that have been the most meaningful, resulting in attempts to provide idiographic responses on the *r/SuicideWatch* thread. This notion is supported by the fact that peer comments focused on academic encouragement were one of the most common responses to user's posts discussing academic difficulties. Findings from our independent empirical example, however, are not congruent with the peer post topics on *r/SuicideWatch*. For example, in the empirical example, individuals with a history of suicidal crisis presented topics that largely reflect broader support topics (i.e., encouragement, support, listen) vs. those that are personalized. While it is possible this is due to the prompt given in the empirical advice (requesting general advice), it may also suggest a potential discrepancy between support provided online and desired support by the individual in a suicidal crisis. Given the significance of this, it is worthy of further investigation.

In examining the associations between post and comment topics and peer perceived helpfulness, there was a disconnect between the topics peers most frequently discussed and those rated as most helpful. For example, the three most frequent peer comment topics were asking questions, communication or emotional support, and academic encouragement (63.6% of all peer comments), however, the peer comment topics perceived as most helpful by peers were professional help suggestions, life meaning, and relationship/loss support (15.8% of all peer comments). Part of the reason for this disagreement may be that peer perceived helpfulness appeared to be dependent on both the user's post and peer's response topic, as noted above. For example, peer responses of relationship/loss support were rated as most helpful

by peers in responses to user posts expressing financial stress or psychological difficulties. On the other hand, peer responses focused on the meaning of life were rated as most helpful in response to user posts of psychological pain. Such discrepancies highlight that, despite the prevalence of peer responses being similar across original post topics, peers are attuned to the topic of the original post and attempt to tailor their response accordingly.

Our empirical example highlights areas of both consistency and potential discrepancy between *peer* perceived helpfulness and perceptions of individuals with a personal history of a suicidal crisis. One of the most notable consistencies across samples was the promotion of professional help suggestions. Indeed, this was one of the peer comment topics with highest number of upvotes and was more likely to be used by those with a history of suicidal crisis in the independent sample. This underscores that suggesting professional help, as opposed to being more directive regarding treatment (i.e., *needing* to get professional help), may be a supportive response for those experiencing a suicidal crisis. It will be beneficial for future educational and prevention efforts to target the use of such support in an empathetic manner. There were also discrepancies in topics perceived as helpful by peers (i.e., relationship/loss support) that were not used by those with a history of suicidal crisis in the independent sample. Furthermore, results also demonstrate there was a topic perceived as helpful by peers that was *less* likely to be used by those with a history of suicidal crisis: reasons for living/meaning in life. This suggests that while peers, or those without a personal history of suicidal crises, may think it helpful to remind others of reasons to keep living, it may not be perceived that way to those experiencing a suicidal crisis. If replicated, these findings have important implications for peer education programs for suicide prevention.

A final point of discussion with regard to this studies' findings is the timeliness of peer responses. Only 7.8% of users received a peer response within 5 min of their post and only 64.90% received a response within one hour. This aspect of online peer support is particularly important to consider not only due to timeliness being a cited concern for the use of more formal suicidal crisis resourced, but because consistent with the acute risk phase of the Fluid Vulnerability Theory, 25–40% of individuals who attempt suicide make the decision to act on their suicidal thoughts within 5 min and 70% make the decision within an hour (Simon et al., 2001; Williams et al., 1980), leaving a limited window for intervention. Finally, despite the perceived helpfulness of professional help suggestions by both peers and individuals with a history of suicidal crisis, this topic was rarely used in peer comments. While noted above, this suggests the need for educational efforts for peers providing online support. The fact that social media platforms are being used for its intended purpose of (only) peer support and not necessarily for the purpose of providing referral resources may be an important element of this support avenue; however, it is also appears important to improve peers' understanding that many individuals experiencing suicidal crises may have a treatable mental illness (Nock et al., 2010), and that warmly suggesting or encouraging professional help may actually be perceived as helpful by those in a suicidal crisis.

The current study's findings should be viewed in the context of its limitations. Foremost, we relied on a combination of upvotes by peers and responses from an independent sample among those with a suicidal crisis as metrics of helpfulness. Given that neither metric directly assesses the original user's perception of peer support received online, there is a need for further research in this area. Similarly, we do not have outcome data for user's posting about their suicidal crisis to determine if the peer support that was received aided in reducing the intensity of their suicidal crisis. It will also be important for future research to compare findings

from the current study, which utilized one specific social media platform, to other avenues for online peer support for suicidal crises. Demographic information to characterize the original users and peers utilizing *r/SuicideWatch* is not available. While prior research examining online peer support for mental illness more broadly has successfully included individuals across developmental lifespans (Naslund et al., 2019), it is unclear how the current findings will generalize among all ages, or, if Reddit users and those from our independent sample were similar. Finally, the current study is also limited by the inability to ensure the validity of peer responses provided online. That is, it is possible that some of the peer comments included in the current analysis were not "real" responses as has been suggested in recent media reports (Bergstrom, 2011), which may have had an impact on current findings.

Results provide valuable insights into the process of receiving online peer support during suicidal crises. Topics discussed by users experiencing a suicidal crisis are consistent with theoretical models of suicide (i.e., Fluid Vulnerability Theory), which serves to inform current suicide risk assessment procedures. Moreover, current findings should be utilized to provide educational efforts to peers to promote the use of helpful responses, such as the suggestion of professional help rather than reasons for living, as way to improve online support for suicidal crises. Such educational efforts may help peers provide a greater proportion of comments that may be perceived as helpful, thus increasing the overall quality of support received during a crucial time of distress. Similarly, it may be beneficial for efforts to not only increase the quality of peer support, but also the quantity of individuals available to provide support as a way to improve the timeliness of support received.

Methods

Automated phrase mining. Representing text with *quality phrases* instead of words or *n*-grams can improve computational models for applications such as information retrieval, taxonomy construction, and sentiment analysis (El-Kishky et al., 2014; Li et al., 2016). We use the state-of-the-art automated phrase mining tool AutoPhrase (Shang et al., 2018) to find phrases from the text corpus—original user's post to peer response, which is domain-independent and with no additional human effort on the target text data. Figure 6 illustrates the workflow of the tool. The first step is to use frequent *n*-gram mining (based on frequent sequential pattern mining (Pei et al., 2001; Yan et al., 2003)) to find phrase candidates from the Reddit corpus. The *support* (i.e., frequency in the corpus) of each candidate is not smaller than a given minimum support threshold. Short phrase candidates such as "great at" and "to talk" may have a large support but not good quality. Therefore, the follow step is to classify the phrase candidates into two categories, quality phrases and non-quality phrases. AutoPhrase adopts a random forest-based classifier and has innovative designs on labeling and feature selection:

- *Distant positive-only labeling and robust negative label sampling:* Knowledge bases (KBs) such as Wikipedia and Freebase have been widely used for *distant supervision* (Mintz et al., 2009). Positive labels can be assigned to the phrase candidates that can be linked to the KBs, such as "community college", "social anxiety", and "health insurance". The strategy of *random negative sampling* has been applied to generate negative labels for classification, which assumes that the majority of samples are likely to be negative. However, the noise in the negative pool can be harmful to the classifier's performance. A robust negative sampling method was developed in AutoPhrase to reduce the noise based on ensemble subsampling (Chen et al., 2013).

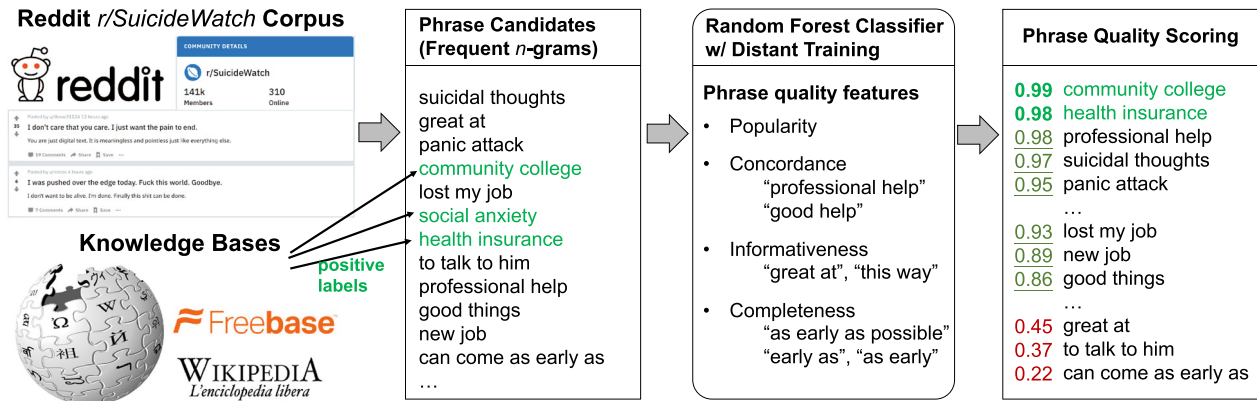


Fig. 6 Illustration of AutoPhrase (Shang et al., 2018). It generates phrase candidates with frequent n -gram mining and builds a random forest classifier with *distant supervision* (Mintz et al., 2009) from general knowledge bases (e.g., Freebase and Wikipedia). The classifier has (1) positive labels from knowledge bases and negative labels through robust sampling and (2) a rich set of features of phrase quality. Applied to the corpus, it assigns good phrase candidates with a high score.

- **Statistical features of phrase quality:** If the phrase quality is defined as the probability of a word sequence being a complete semantic unit, then a quality measurement should include the following criteria: (1) *popularity*: quality phrases should have sufficient frequency in the given text corpus; (2) *concordance*: the collocation of tokens should have significantly higher probability than expected due to chance; (3) *informativeness*: it should have tokens that indicate specific topics, events, or concepts; (4) *completeness*: a phrase is deemed complete when it can be interpreted as a complete semantic unit in the context. Statistical features such as probabilities and Part-of-Speech tag counts are used to generate the set of features.

We applied AutoPhrase to our dataset. It assigns scores between 0 and 1 to every phrase candidate. A higher score indicates a better phrase. We determine the set of quality phrases with a threshold score.

Parameter settings: We chose the minimum support threshold as 2. The threshold score of phrase quality was 0.75.

AutoPhrase has been recognized as an available and useful tool for phrase mining in the NLP community (Kshirsagar et al., 2017; Nguyen et al., 2019). The tool was published in 2018 and has been cited 78 times by December 17, 2019 on Google Scholar.

Pairwise latent Dirichlet allocation (LDA). LDA is a generative probabilistic model of text data. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words (Blei et al., 2003). In our study, we have two text corpora, original user's post corpus, and peer response corpus. The topics of the two corpora are different and associated. So on one hand, we develop two LDA models separately to discover the topics on each corpus; on the other hand, we use the post-comment pairs in the dataset to supervise the models with the association information. Figure 7 presents the proposed *pairwise* LDA model. Specifically, the models have different number of topics, $K^{(P)}$ and $K^{(R)}$ for original user's post and peer response, respectively. The observable post-response pairs can be generated by pairs of topic distributions of post documents and comment documents. Moreover, our model generates documents with both words (i.e., single-words) and multi-word phrases. We use the term "phrases", as words can be considered as single-word phrases and one of the unique points of the model is to learn topics at the phrase level.

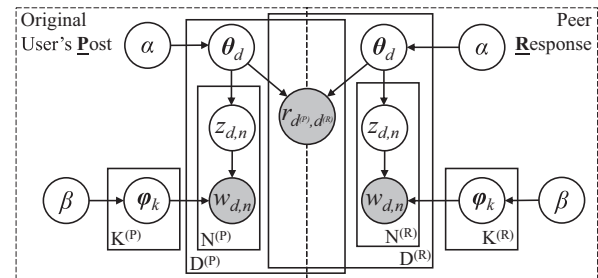


Fig. 7 Plate notation of the proposed PairwiseLDA to discover topics in post-response text corpora.

Formally, we define the following terms:

- A *phrase* is the basic unit of data, defined as an item from a vocabulary indexed by $\{1, \dots, V^{(\cdot)}\}$, where $(\cdot) \in \{(P), (R)\}$ for original user's Post and peer Response.
- A *document* is a sequence of phrases. We denote the n th word in the d th document by observable $w_{d,n} \in \{1, \dots, V^{(\cdot)}\}$.
- A *corpus* is a collection of $D^{(\cdot)}$ documents. We have a post corpus and a response corpus.
- A *post-response document pair* $(d^{(P)}, d^{(R)})$ contains the $d^{(P)}$ th document of the post corpus and the $d^{(R)}$ th document of the response corpus. We denote it by $r_{d^{(P)}, d^{(R)}} = 1$ if the response is associated with the post; otherwise, $r_{d^{(P)}, d^{(R)}} = 0$.

Based on the notations, Fig. 7 presents the plate design of the proposed model. PairwiseLDA assumes the following generative process for each document in each corpus:

1. Draw topic proportion for each document $\theta_d \sim \text{Dir}(\alpha)$, where $\text{Dir}(\cdot)$ denotes Dirichlet distribution;
2. Draw a distribution over *phrases* for each topic $\phi_k \sim \text{Dir}(\beta)$;
3. For each phrase position d, n :
 - (a) Choose a topic $z_{d,n} \sim \text{Multi}(\theta_d)$, where $\text{Multi}(\cdot)$ denotes multinomial distribution;
 - (b) Choose a phrase $w_{d,n}$ from $p(w_{d,n} | \phi_{z_{d,n}})$.
4. For a pair of documents $d^{(P)}, d^{(R)}$, determine the association $r_{d^{(P)}, d^{(R)}} = \sigma(\theta_{d^{(P)}} \Sigma \theta_{d^{(R)}})$, where $\Sigma \in \mathbb{R}^{K^{(P)} \times K^{(R)}}$ is a feature transformation matrix from the post topic space to the comment topic space. $\sigma(\cdot)$ is a sigmoid function.

We make inference with the proposed probabilistic graphical model as follows. The inference problem is to compute the output parameters ϕ_k , θ_d , $z_{d,m}$ and Σ in the *pairwise* supervised topic model (for each corpus) given input observed data $w_{d,n}$ and $r_{d^{(P)}, d^{(R)}}$. However, exact computation is intractable in topic models (Blei et al., 2003). Thus, we are interested in approximated inference of the model parameters. Collapsed Gibbs sampling is a straightforward, easy to implement, and unbiased approach that converges rapidly to a known ground-truth (Mcauliffe and Blei, 2008; Porteous et al., 2008). It is typically preferred over other possible approaches in large scale applications of topic models (Rehurek and Sojka, 2010).

Topic number selection. We use a heuristic approach to search for a proper number of topics (Zhao et al., 2015). Based on the following expectations/assumptions, we determine the search range for each corpus as $K^{(P)} \in \{1, \dots, 10\}$ and $K^{(R)} \in \{1, \dots, 20\}$. First, the number of the topics should not be too large; otherwise, the topics would be too sparse and the topic associations would be of too high complexity, i.e., $O(K^{(P)}K^{(R)})$. Second, the number of response topics $K^{(R)}$ is expected to be larger than the number of post topics $K^{(P)}$. There could be a small number of types of needs when people had suicide thoughts but the replies could vary to a large extent, particularly given the larger number of responses than posts. So we chose 10 and 20 as the maximum number of topics to search.

We use two metrics to evaluate the results of topic modeling with respect to different topic numbers. The first metric is well accepted and widely used, called *topic coherence* (Röder et al., 2015; Stevens et al., 2012). It is the average coherence of all topics. For a single topic, the coherence is defined as the sum of pairwise distributional similarity scores over a set of topic *phrases*:

$$\text{coherence}(V) = \sum_{(v_i, v_j) \in V} \text{score}(v_i, v_j, \epsilon), \quad (1)$$

where V is a set of phrases describing the topic and ϵ indicates a smoothing factor, which guarantees that *score* returns real numbers. Here, we use $\epsilon = 1$ as the original authors used. The similarity *score* defines the score to be based on document co-occurrence, known as the UMass metric (Stevens et al., 2012):

$$\text{score}(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)}, \quad (2)$$

where $D(x, y)$ counts the number of documents containing phrases x and y and $D(x)$ counts the number of documents containing x . Significantly, this metric computes these counts over the original corpus used to train the topic models. It attempts to confirm that the models learned data known to be in the corpus. Note that the topic coherence score is *negative* and a *higher* coherence score means a better result of topic modeling.

The second metric is new in this study. Given the topic distribution of a document, we assign the document to the topic of the highest probability. The “volume” of a single topic is the number of documents that they were assigned to. As we are expecting volume balance among topics but also aware of the variety of topic popularity, we measure the degree of volume imbalance, called Maximum Relative Volume Gap (MRVG). We sort the volume (in percentage) from high to low: $p_1 \geq p_2 \geq \dots \geq p_K$, where K is the number of topics. The metric is defined as

$$\text{MRVG} = \arg\max_{i=1}^{K-1} \frac{p_{i+1} - p_i}{1/K} = K \cdot \arg\max_{i=1}^{K-1} (p_{i+1} - p_i). \quad (3)$$

A close-to-1 MRVG indicates a more balanced volume distribution.

Here, we list the set of expectations to a *good* result of topic modeling:

1. $K^{(R)} > K^{(P)}$: The number of topics of responses $K^{(R)}$ should be larger than the number of topics of original posts $K^{(P)}$.
2. Higher *coherence* score means a better result.
3. Close-to-1 MRVG indicates a better result.

We set $K^{(R)} = 2K^{(P)}$ and apply the proposed pairwise LDA model with $K^{(P)} \in \{3, 4, 5, 6, 7, 8, 9, 10\}$. Figures 8, 9, 10, and 11 present the coherence score and MRVG on each topic distribution. First, when $K^{(P)} = 3, 4, 5$, the coherence score is high. It significantly drops from the point of $K^{(P)} = 6$. Similar

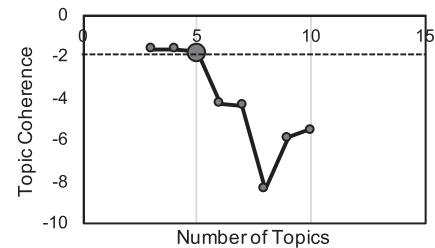


Fig. 8 This is the coherence score when we choose $K^{(P)} = 5$ for *original* user's post text.

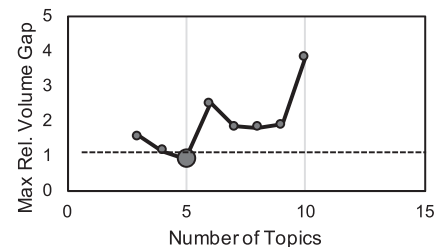


Fig. 9 This is the MRVG when we choose $K^{(P)} = 5$ for *original* user's post text.

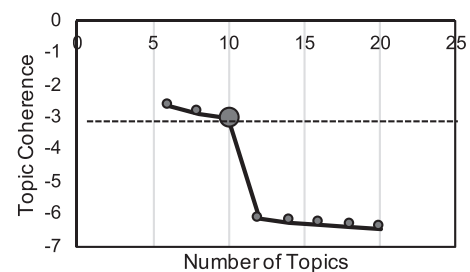


Fig. 10 This is the coherence score when we choose $K^{(R)} = 10$ for *peer* response text.

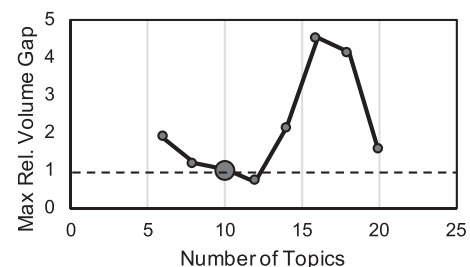


Fig. 11 This is the MRVG when we choose $K^{(R)} = 10$ for *peer* response text.

Table 7 Representative words and phrases generated by traditional LDA model (Blei et al., 2003).

Topic ID	Representative words	Representative phrases
P.1	Know, school, help, family, friends, care, friend	Long time, go back to school, i do not want to live
P.2	Want, things, time, going, anymore, work, thoughts	Last year, first time, few weeks, suicidal thoughts
P.3	Feel, people, told, stop, actually, home, person	Only one, only time, only person, grocery store
P.4	Live, years, left, talk, right, hate, love, world, said	Best friend, friend of mine, whole life, most people
P.5	Life, like, think, trying, need, living, year, money	Only thing, other people, end my life, entire life
R.1	Like, good, right, school, friends, understand, talking	Talk to someone, make you happy, hey man, best thing
R.2	Better, thoughts, care, little, mother, hear, thought, fact	Same time, best friend, let me know, many others
R.3	Life, help, talk, trying, happy, sounds, true, felt, future	Other things, you want to talk, best way, talk to me
R.4	Family, means, thank, wanted, months, tell, smile	So many people, good thing, talk to people
R.5	People, think, suicide, love, makes, lost, therapist	Long time, do you have, let me tell, pm me if you want
R.6	Person, sorry, live, point, kind, different, telling	High school, good luck, next thing, family member
R.7	Feeling, years, reason, helps, great, look, place, plain	Suicidal thoughts, most people, few years, take care
R.8	Know, things, living, thing, world, probably, hate	Other people, own life, you need someone to talk
R.9	Want, need, time, sure, hard, matter, course, best	Same way, best friends, same thing, tough time
R.10	Feel, depression, work, thanks, hope, change, nice	Few months, do you think, even if it is, good ones

Table 8 Representative words and phrases generated by cross-lingual LDA model (Yang et al., 2019).

Topic ID	Representative words	Representative phrases
P.1	Brother, father, class, boyfriend, play, sister, grades	Middle school, good grades, few friends, freshman year
P.2	Sick, single, problem, self, commit, trust, date	Makes me feel, single day, mental illness, bad person
P.3	Feelings, suffer, nobody, enjoy, human, alive, feeling	Rest of my life, first place, I do not want to die
P.4	Mental, issues, therapist, therapy, realize, handle	Panic attacks, talk to me, few times, suicide hotline
P.5	Graduate, afford, jobs, different, debt, pay, degree	Student loans, last week, new job, have been
R.1	Meet, stress, focus, alive, similar, small, drink	New friends, even if it is, good things, similar situation
R.2	Difficult, brain, sense, figure, emotional, express	Mental health, entire life, not everyone, great deal
R.3	Doctor, therapy, mother, medication, health, treat	Family member, mental health, best wishes, good idea
R.4	Social, girl, women, relationships, girls, date, grow	Social anxiety, talk to them, social skills, few friends
R.5	Thought, easy, games, happiness, video, pretty, free	Do you know, good person, free time, makes you happy
R.6	Reach, issues, feeling, second, improve, trust	Mental illness, talk to someone, suicide hotline, talk to you
R.7	Home, months, whole, hospital, face, words, wait	Few months, whole life, few days, kind words
R.8	Control, internet, fault, deserve, young, strong, lead	Right direction, other side, deep breath, keep in mind
R.9	Deserve, night, open, wife, option, girlfriend, true	Last night, do you feel, other options, let me know
R.10	Fail, death, anxiety, imagine, self, normal, suffer	Panic attacks, loved ones, not something, is not something

observations can be found on the response text. When $K^{(R)} = 6, 8, 10$, the coherence score is high. It drops to the bottom from the point of $K^{(R)} = 12$. Second, the MRVG scores are very close to 1 when $K^{(P)} = 5$ and $K^{(P)} = 10$, respectively. Therefore, we choose $K^{(P)} = 5$ and $K^{(P)} = 10$ for the topic numbers in this study.

- Assessing associations between post topics and comment topics:* As we have assigned every document to a proper topic in both post and comment data, we can project associations between post and comment documents (in the original data) into associations between post and comment topics. Therefore, we can calculate the number of document pairs associated with every topic pair, as presented in Fig. 4. As the document associations come along with the number of upvotes, we can calculate the average number of upvotes between every topic pair, as presented in Fig. 5.
- Topic comparison analysis 1: pairwise LDA vs. traditional LDA:* We run traditional LDA model (Blei et al., 2003) separately on the post and comment corpora in Reddit data. Table 7 presents the representative words and phrases (discovered by AutoPhrase). We met difficulty in finding proper topic names for any of the topics of original posts and/or peer responses. Our pairwise LDA model was able to generate meaningful topics in Tables 1 and 3.

- Topic comparison analysis 2: pairwise LDA vs. cross-lingual LDA:* A recent work on topic model proposed to connect cross-lingual topics (Yang et al., 2019). We implemented this model and applied an analogy: The analogy would be that original posts are in one “language,” and peer responses are in the other language. Then this model can be considered as a competitive baseline with our proposed pairwise topic model. The assumption of the cross-lingual topic model is that the topics across the two languages follow the same distributions and thus can be aligned. This is true for the cross-lingual setting because usually we can find proper word pairs of the same semantics between languages. However, as we have observed, the topics in the original posts and the topics in the peer responses are significantly different. They cannot be simply aligned with a one-to-the-other transformation regularization. Table 8 presents the representative words and phrases. Though they make more sense than the traditional LDA, we still cannot easily find a proper name for each topic. Our pairwise topic model assumes that if the topic distributions of any original post and peer response were predictive of the pairing of the two documents (paired or not), then the topics would be effectively, jointly learned.

Conclusions

In this study, we characterized the content of both a user's post and corresponding peer comments occurring on a social media platform and present an empirical example for comparison. We introduced an approach that uses pairwise topic models to transform large corpora of discussion into associated topics of user and peer posts. Our approach modeled both the generative process of each type of the corpora (i.e., user posts and peer comments) and the associations between them. It used phrases, which are more informative and less ambiguous than words, in addition to words, to represent social media posts and topics. We evaluated the method using data from Reddit *r/SuicideWatch*. We examined how the topics of user and peer posts were associated, and how this information influenced the peer perceived helpfulness of the peer support. Then, we applied structural topic modeling to data collected from individuals with a history of suicidal crisis as a means to validate findings. Our observations suggested that effective modeling of the association between the two lines of topics can uncover helpful peer responses to online suicidal crises, notably providing the suggestion of professional help.

Data availability

The datasets analyzed during the current study are available in the Dataverse repository: <https://doi.org/10.7910/DVN/BDZ4LI>. Besides the datasets, code is available in the repository. These datasets were derived from a public domain resource: https://archive.org/download/2015_reddit_comments_corpus/reddit_data/.

Received: 28 August 2019; Accepted: 2 June 2020;

Published online: 15 July 2020

References

- U.S. News (2015) The dark link between financial stress and depression. <https://health.usnews.com/health-news/health-wellness/articles/2015/02/25/the-dark-link-between-financial-stress-and-depression>
- CBS News (2018) College students so stressed. <https://www.cbsnews.com/news/1-in-5-college-students-so-stressed-they-consider-suicide/>
- American Foundation for Suicide Prevention (2019) The five biggest myths about crisis text line. <https://afsp.org/five-biggest-myths-crisis-text-line/>
- Adamic LA, Huberman BA (2000) Power-law distribution of the world wide web. *Science* 287:2115–2115
- Ali K, Farrer L, Gulliver A, Griffiths KM (2015) Online peer-to-peer support for young people with mental health problems: a systematic review. *JMIR Ment Health* 2:e19
- Althoff T, Clark K, Leskovec J (2016) Large-scale analysis of counseling conversations: an application of natural language processing to mental health. *Trans Assoc Comput Linguist* 4:463–476
- Benton A, Coppersmith G, Dredze M (2017) Ethical research protocols for social media health research. In: Proceedings of the First ACL workshop on ethics in natural language processing. Association for Computational Linguistics. pp. 94–102, <https://doi.org/10.18653/v1/w17-1612>
- Bergstrom K (2011) Don't feed the troll?: shutting down debate about community expectations on reddit.com. *First Monday* 16, <https://doi.org/10.5210/fm.v1i6.83498>
- Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S (2017) #Why-WeTweethMH: understanding why people use twitter to discuss mental health problems. *J Med Internet Res* 19:e107
- Birnbaum ML, Rizvi AF, Correll CU, Kane JM, Confino J (2017) Role of social media and the internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. *Early Interv Psychiatry* 11:290–295
- Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Buntain C, Golbeck J (2014) Identifying social roles in reddit using network structure. In: Proceedings of the 23rd international conference on world wide web. ACM Press. pp. 615–620, <https://doi.org/10.1145/2567948.2579231>
- Chen L, Dou WW, Qiao Z (2013) Ensemble subsampling for imbalanced multi-variate two-sample tests. *J Am Stat Assoc* 108:1308–1323
- Ciarrochi J, Wilson CJ, Deane FP, Rickwood D (2003) Do difficulties with emotions inhibit help-seeking in adolescence? The role of age and emotional competence in predicting help-seeking intentions. *Counsel Psychol Q* 16:103–120
- Coppersmith G, Leary R, Crutchley P, Fine A (2018) Natural language processing of social media as screening for suicide risk. *Biomed Inform Insight* 10:1178222618792860
- Cowgell VG (1977) Interpersonal effects of a suicidal communication. *J Consul Clin Psychol* 45:592
- De Choudhury M, De S (2014) Mental health discourse on reddit: self-disclosure, social support, and anonymity. In: Proceedings of the 8th international AAAI conference on web and social media. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8075>
- De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M (2016) Discovering shifts to suicidal ideation from mental health content in social media. In: Proceedings of the 2016 Chi conference on human factors in computing systems. ACM, pp. 2098–2110, <https://doi.org/10.1145/2858036.2858207>
- DeMasi O, Hearst MA, Recht B (2019) Towards augmenting crisis counselor training by improving message retrieval. In: Proceedings of the 6th workshop on computational linguistics and clinical psychology. Association for Computational Linguistics, pp. 1–11, <https://doi.org/10.18653/v1/w19-3001>
- Centers for Disease Control (2018) Suicide rates rising across the U.S. <https://www.cdc.gov/media/releases/2018/p0607-suicide-prevention.html>
- Drum DJ, Brownson C, Burton Denmark A, Smith SE (2009) New data on the nature of suicidal crises in college students: shifting the paradigm. *Prof Psychol Res Prac* 40:213
- El-Kishky A, Song Y, Wang C, Voss CR, Han J (2014) Scalable topical phrase mining from text corpora. *Proc VLDB Endow* 8:305–316
- Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, Musacchio KM, Jaroszewski AC, Chang BP, Nock MK (2017) Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull* 143:187
- Gilbert E (2013) Widespread underprovision on reddit. In: Proceedings of the 2013 conference on computer supported cooperative work. ACM Press, pp. 803–808, <https://doi.org/10.1145/2441776.2441866>
- Gowen K, Deschaine M, Gruttadara D, Markey D (2012) Young adults with mental health conditions and social networking websites: seeking tools to build community. *Psychiatr Rehabil J* 35:245
- Haker H, Lauber C, Rössler W (2005) Internet forums: a self-help approach for individuals with schizophrenia? *Acta Psychiatrica Scand* 112:474–477
- Hope A, Dring C, Dring J (2005) College lifestyle and attitudinal national (CLAN) survey. *Public Health Well*
- Horgan A, Sweeney J (2010) Young students' use of the internet for mental health information and support. *J Psychiatr Ment Health Nursing* 17:117–123
- Huang H-Y, Bashir M (2016) Online community and suicide prevention: investigating the linguistic cues and reply bias. In: Proceedings of the conference on human factors in computing systems
- Klonsky ED, May AM (2014) Differentiating suicide attempters from suicide ideators: a critical frontier for suicidology research. *Suicide Life-Threat Behav* 44:1–5
- Klonsky ED, May AM (2015) The three-step theory (3st): a new theory of suicide rooted in the "ideation-to-action" framework. *Int J Cogn Ther* 8:114–129
- Kshirsagar R, Morris R, Bowman S (2017) Detecting and explaining crisis. In: Proceedings of the 4th workshop on computational linguistics and clinical psychology-from linguistic signal to clinical reality. Association for Computational Linguistics, Preprint at 1705.09585, <https://doi.org/10.18653/v1/w17-3108>
- Kummervold PE, Gammon D, Bergvik S, Johnsen J-AK, Hasvold T, Rosenvinge JH (2002) Social support in a wired world: use of online mental health forums in norway. *Nord J Psychiatry* 56:59–65
- Laranjo L et al. (2018) Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 25:1248–1258
- Lawlor A, Kirakowski J (2014) Online support groups for mental health: a space for challenging self-stigma or a means of social avoidance? *Comput Hum Behav* 32:152–161
- Li B, Wang B, Zhou R, Yang X, Liu C (2016) CITPM: A cluster-based iterative topical phrase mining framework. In: Database systems for advanced applications. Springer International Publishing, pp. 197–213, https://doi.org/10.1007/978-3-319-32025-0_13
- Linthicum KP, Schafer KM, Ribeiro JD (2019) Machine learning in suicide science: applications and ethics. *Behav Sci Law* 37:214–222, <https://doi.org/10.1002/bsl.2392>
- Mcauliffe JD, Blei DM (2007) Supervised topic models. In: Advances in neural information processing systems. Proceedings of the 20th International Conference on Neural Information Processing Systems. Curran Associates

- Inc., Vancouver, British Columbia, Canada, pp. 121–128, <https://doi.org/10.5555/2981562.2981578>
- Milne DN, McCabe KL, Calvo RA (2019) Improving moderator responsiveness in online peer support through automated triage. *J Med Internet Res* 21:e11410
- Milne DN, Pink G, Hachey B, Calvo RA (2016) CLPsych 2016 shared task: triaging content in online peer-support forums. In: *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology*. Association for Computational Linguistics, pp. 118–127, <https://doi.org/10.18653/v1/w16-0312>
- Mintz M, Bills S, Snow R, Jurafsky D (2009) Distant supervision for relation extraction without labeled data. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, vol. 2. Association for Computational Linguistics, pp. 1003–1011, <https://doi.org/10.3115/1690219.1690287>
- Morris RRR (2015) Crowdsourcing mental health and emotional well-being. Ph.D. thesis, Massachusetts Institute of Technology
- Naslund JA, Aschbrenner KA, McHugo GJ, Ünützer J, Marsch LA, Bartels SJ (2019) Exploring opportunities to support mental health care using social media: a survey of social media users with mental illness. *Early Interv Psychia* 13:405–413 (n.d.), R. (2019) *r/suicidewatch*. <https://www.reddit.com/r/SuicideWatch>
- Nguyen D, Liakata M, DeDeo S, Eisenstein J, Mimno D, Tromble R, Winters J (2019) How we do things with words: analyzing text as social and cultural data. Preprint at arXiv:1907.01468
- Nock MK, Hwang I, Sampson NA, Kessler RC (2010) Mental disorders, comorbidity and suicidal behavior: results from the national comorbidity survey replication. *Mol Psychiatry* 15:868
- Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M-C (2001) PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth. In: *Proceedings 17th international conference on data engineering*. IEEE, Comput. Soc, pp. 215–224, <https://doi.org/10.1109/icde.2001.914830>
- Porteous I, Newman D, Ihler A, Asuncion A, Smyth P, Welling M (2008) Fast collapsed gibbs sampling for latent Dirichlet allocation. In: *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM Press, pp. 569–577, <https://doi.org/10.1145/1401890.1401960>
- Reddit (2018) Reddit year in review. <https://redditblog.com/2018/12/04/reddit-year-in-review-2018/>
- Rehurek R, Sojka P (2010) Software framework for topic modelling with large corpora. In: *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks*. ELRA, Valletta, Malta, pp. 45–50, <http://is.muni.cz/publication/884893/en>
- Roberts ME et al. (2013) The structural topic model and applied social science. In: *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*. Harrahs and Harveys, Lake Tahoe, vol. 4, pp. 1–20
- Röder M, Both A, Hinneburg A (2015) Exploring the space of topic coherence measures. In: *Proceedings of the 8th ACM international conference on web search and data mining-WSDM'15*. ACM Press, pp. 399–408, <https://doi.org/10.1145/2684822.2685324>
- Rudd MD (2006) Fluid vulnerability theory: a cognitive approach to understanding the process of acute and chronic suicide risk. In: *Cognition and suicide: theory, research, and therapy*. American Psychological Association, pp. 355–368, <https://doi.org/10.1037/11377-016>
- Scherr S, Reinemann C (2016) First do no harm: cross-sectional and longitudinal evidence for the impact of individual suicidality on the use of online health forums and support groups. *Comput Hum Behav* 61:80–88
- Serafini G, Muzio C, Piccinini G, Flouri E, Ferrigno G, Pompili M, Girardi P, Amore M (2015) Life adversities and suicidal behavior in young individuals: a systematic review. *Eur Child Adolesc Psy* 24:1423–1446
- Service PB (2018) As calls to the suicide prevention lifeline surge, under-resourced centers struggle to keep up. <https://www.pbs.org/newshour/nation/suicide-prevention-lifeline-centers-calls>
- Shang J, Liu J, Jiang M, Ren X, Voss CR, Han J (2018) Automated phrase mining from massive text corpora. *IEEE Trans Knowl Data Eng* 30:1825–1837
- Shen JH, Rudzicz F (2017) Detecting anxiety through reddit. In: *Proceedings of the 4th workshop on computational linguistics and clinical psychology? From linguistic signal to clinical reality*. Association for Computational Linguistics, pp. 58–65, <https://doi.org/10.18653/v1/w17-3107>
- Shing H-C, Nair S, Ziriky A, Friedenberg M, DauméIII H, Resnik P (2018) Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: *Proceedings of the 5th workshop on computational linguistics and clinical psychology: from keyboard to clinic*. Association for Computational Linguistics, pp. 25–36, <https://doi.org/10.18653/v1/w18-0603>
- Simon TR, Swann AC, Powell KE, Potter LB, Kresnow M-j, O'Carroll PW (2001) Characteristics of impulsive suicide attempts and attempters. *Suicide Life-Threat Behav* 32:49–59
- Stevens K, Kegelmeyer P, Andrzejewski D, Buttlar D (2012) Exploring topic coherence over many models and many topics. In: *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*. Association for Computational Linguistics, Jeju Island, Korea, pp. 952–961, <https://www.aclweb.org/anthology/D12-10871>
- Turner RJ, Frankel BG, Levin DM (1983) Social support: conceptualization, measurement, and implications for mental health. *Res Community Ment Health* 3:67–111, <https://psycnet.apa.org/record/1984-20538-001>
- Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner Jr. TE (2010) The interpersonal theory of suicide. *Psychol Rev* 117:575
- Williams CL, Davidson JA, Montgomery I (1980) Impulsive suicidal behavior. *J Clin Psychol* 36:90–94
- Woodward A, Wyllie C (2016) Helplines, tele-web support services, and suicide prevention. *The international handbook of suicide prevention*, pp. 490–504, <https://doi.org/10.1002/9781118903223.ch28>
- Yan X, Han J, Afshar R (2003) CloSpan: mining: closed sequential patterns in large datasets. In: *Proceedings of the 2003 SIAM international conference on data mining*. Society for Industrial and Applied Mathematics, pp. 166–177, <https://doi.org/10.1137/1.9781611972733.15>
- Yang W, Boyd-Graber J, Resnik P (2019) A multilingual topic model for learning weighted topic links across corpora with low comparability. In: *Proceedings of the 2019 conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, pp. 1243–1248, <https://doi.org/10.18653/v1/d19-1120>
- Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, Zou W (2015) A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinform* 16:S8

Acknowledgements

Our research is supported by National Science Foundation IIS-1849816.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.J.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020