



Published in final edited form as:

Am Psychol. 2020 April ; 75(3): 365–379. doi:10.1037/amp0000477.

Assortativity of Suicide-Related Posting on Social Media

Ian Cero, Tracy K. Witte

Auburn University

Abstract

Since the turn of the century, interdisciplinary research on networks—their formation, structure, and influence—has advanced so rapidly, it is now a science unto itself, offering new and powerful quantitative tools for studying human behavior, whose potential psychologists are just beginning to glimpse. Among these tools is a formula for quantifying *assortativity*, the propensity of similar people to be socially connected with one another more often than their dissimilar counterparts. With this formula, this investigation establishes a foundation for examining assortative patterns in suicidal behavior and highlights how they can be exploited for improved prevention. Specifically, the established clustering of suicide fatalities in time and space implies such fatalities have assortative features. This suggests other forms of suicide-related behavior may as well. Thus, the assortativity of suicide-related verbalizations (SRVs) was examined by machine coding 64 million posts from 17 million users of a large social media platform—Twitter—over 2 distinct 28-day periods. Users were defined as socially linked in the network if they mutually replied to each other at least once. Results show SRVs were significantly more assortative than chance, through 6 degrees of separation. This implies that if a person posts SRVs, their friends’ friends’ friends’ friends’ friends’ friends are more likely than chance to do the same even though they have never met. SRVs also remained significantly assortative through 2 degrees, even when mood was controlled. Discussion illustrates how these assortative patterns can be exploited to improve the true-positive rate of suicide risk screenings.

Keywords

suicide; network; assortativity; ideation; social media

Social clustering is an important feature of suicidal behavior (Gould, Wallenstein, & Davidson, 1989), causing significant concern in the communities in which it is perceived (Haw, Hawton, Niedzwiedz, & Platt, 2013). Research has identified both different types of clusters and different mechanisms by which clustering might occur (Joiner, 1999). For example, atypical elevations in the rate of suicide have been described in narrow time frames (e.g., after mass media exposure; Gould, 1990), narrow geographic spaces (e.g., underground trains; Farmer, O’Donnell, & Tranah, 1991), as well as narrow temporal-geographic spaces (e.g., several suicides during a short period in a single school; Brent et al., 1989). Nonfatal suicide attempts have also been shown to cluster (Gould, Petrie, Kleinman,

& Wallenstein, 1994), and randomized experiments have even shown a clustering pattern emerges in suicidal ideation (Joiner, 2003).

Proposed mechanisms for suicide clustering are numerous. The most popular accounts involve some kind of *contagion* process, by which exposure to suicidal behavior transmits novel risk for suicide to those who are exposed (e.g., related to social learning, projective identification, priming, imitation, or some other mechanism; Haw et al., 2013). However, a more parsimonious account of suicide clustering involves simple *homophilic preference* (or more simply, *homophily*), a social phenomenon in which people prefer to associate with others who are similar to themselves (i.e., “birds of a feather, flock together”; Newman, 2003). In this way, already suicidal people may be socially clustered, but no meaningful risk is produced by their interaction with one another; if they die at similar times or places, it is because they already had similar risk profiles and backgrounds. Finally, an even more basic account implies suicide clustering could simply be an artifact of shared history. For example, economic downturn could simultaneously increase the suicide risk in a group of workers laid off from the same large factory. Even if they never actually knew one another, they may end their lives in an apparently patterned way. Along these lines, many examples of suicide clustering that appear to result from contagion could also result from other, simpler processes (Joiner, 1999).

In actuality, all of these possibilities are specific hypotheses about the origin of a more general phenomenon known as *assortative mixing*, or just simply *assortativity* (Aral, Muchnik, & Sundararajan, 2009). *Assortativity* is the observation that any two people with similar attributes are more likely to be linked in some way than any two people with dissimilar attributes. This pattern could be produced by one person influencing friends toward similarity (contagion); it could be the result of already-similar people seeking each other out (homophily); it could be the result of shared history and context in a group of people; or it could be any combination thereof.

Regardless of its origin, assortativity is nearly ubiquitous in human social networks. In fact, it is so common that it is argued to be one of the statistical signatures that distinguishes human social networks from their nonsocial counterparts (i.e., technological or microbiological networks; Newman & Park, 2003). For instance, assortative mixing has been documented across numerous and diverse types of human behavior, ranging from smoking to obesity (Christakis & Fowler, 2007, 2008). Even measures of happiness have been shown to be assortative in both online and offline networks (Bliss, Kloumann, Harris, Danforth, & Dodds, 2012; Fowler & Christakis, 2009). Perhaps even more surprising than its breadth, assortativity also exhibits substantial reach in human networks. In each of these examples people were observed to be more similar than would be expected by chance, even when there were three degrees of separation between them (Christakis & Fowler, 2013). That is, the friends of your friends’ friends are more like you (e.g., in terms of happiness) than would be predicted by chance alone, even though you will likely never meet them.

However, because each of these causal mechanisms described earlier can produce exactly the same assortative pattern in an observed network, it is mathematically impossible to discern which of them gave rise to that pattern in any particular observational context

(Shalizi & Thomas, 2011). A prohibitively difficult—and in the case of suicide, likely unethical—experimental manipulation of the large-scale network structure itself would be required. Thus, although researchers have now spent decades arguing over what gives rise to the clustering of suicidal behavior (for recent reviews, see Cheng, Li, Silenzio, & Caine, 2014; Haw et al., 2013), the unforgiving mathematical landscape will likely prevent meaningful resolution for years to come. This investigation thus represents an intentional departure from the origin of assortativity. Instead, the focus is now on deciphering the assortative pattern itself (e.g., is nonfatal suicidal behavior also assortative; through how many degrees of separation) and how it can be exploited for suicide prevention. After all, if the pattern by which birds of a feather flock together is discovered, then only a few birds are needed to infer the location of the whole flock—even when their impulse to come together remains enigmatic.

Quantifying Assortativity

The assortativity coefficient of a network is closely related to the Pearson correlation coefficient and is bounded by the interval -1.0 to 1.0 . A value of 1.0 represents complete similarity across links (e.g., a marriage network of only gay couples); a value of -1.0 represents complete dissimilarity across links (e.g., a marriage network of only straight couples); and a value of 0 represents random pairing of people on either side of a link (e.g., a marriage network with a high proportion of bisexual members). Consider Figure 1, in which an example network of seven nodes (e.g., people), in which each node is given a score on some attribute of interest (e.g., sex, scored 0 or 1, light or dark gray). Imagine also that this network includes six total links among those seven members and numeric ID numbers are used to label each link for clarity. If that network were represented by a table, in which each link is a row and the attributes of the nodes at either end of that link are columns, there would be two columns for our attribute of interest (e.g., the sexes of the people on either side of a link). Note that any node with multiple links will appear in this table multiple times, once for each link to which it is adjacent. When the data are arranged in this way, the network assortativity of a given node attribute is simply Pearson correlation between those two attribute columns.¹ Thus, assortativity is the correlation of different people's attributes across the links that connect them.

Working With Assortativity

Although most research on the assortativity of suicidal behavior has focused on identifying the origins of that pattern (Haw et al., 2013), more basic (and unanswered) questions about the fundamental pattern of suicidal assortativity continue to have growing importance for prevention. For example, mobile apps are especially scalable (Sanderson, 2009) compared with traditional mental health treatment, and they have now been shown effective for the mitigation of suicidal and nonsuicidal self-injury in multiple randomized control trials (Franklin et al., 2016). Harnessing multiple aspects of social media for large-scale risk

¹Note that assortativity coefficients are more general than Pearson coefficients and can be computed on unordered nominal variables as well (e.g., race). There, the relation to the Pearson correlation will not hold. In such cases, as well as with binary variables, the interval bounding the assortativity coefficient may contract to be narrower than -1.0 to 1.0 but will never be wider (Newman, 2002).

detection and prevention efforts has been the focus of increased attention in recent years (Franklin, 2016; Luxton, June, & Fairall, 2012; Robinson et al., 2016) for many of the same reasons.

Delivery of these new techniques can potentially be achieved through existing digital mechanisms (e.g., targeted advertising on social media sites), but to whom should they be delivered? Effective models of assortativity offer an important basis for answering this question. If assortative patterns of suicide risk can be characterized mathematically, then the identification of one at-risk individual will also constitute the probabilistic identification of many others as well. If birds of a feather flock together, finding a small number of birds reveals the location of the whole flock. Thus, establishing effective models of assortativity for multiple kinds of suicidal behavior has important prevention potential, especially for emerging population-scale interventions and screening approaches.

Unresolved Questions in Suicide Assortativity

Even though several features of suicide-related clustering have been debated in the literature for decades, to the authors' awareness, no study has ever directly estimated the network assortativity of any form of suicidal behavior. Unfortunately, although assortative clustering of suicide fatalities have been documented in several studies (Haw et al., 2013), little is known about the degree to which assortativity is involved in other forms of suicidal behavior, including attempts, ideation, or other gestures (e.g., online verbalizations). This is likely because fatalities are more discrete than other forms of nonlethal suicidal behavior (e.g., ideation), and they are more likely to produce contact with the medical and legal systems in which they occur. Thus, they are also more likely to be included in population-sized data sets capable of uncovering clustering (e.g., data sets with contextual variables like time and location of event). In contrast, suicidal ideation and attempts generally require disclosure on the part of people who experience them, which has historically implied traditional survey methods for collecting data on such behavior. Unfortunately, these are hampered by a range of shortcomings, including nondisclosure because of stigma and retrospective biases (Batterham, Calear, & Christensen, 2013; Klimes-Dougan, Safer, Ronsaville, Tinsley, & Harris, 2007). Collecting a data set that is sufficiently large to study assortativity in this relatively uncommon behavior, along with the necessary precision about the contexts in which the behavior occurred to assess whether such behavior is assortative, is thus rarely achieved.

Moreover, when research on assortativity in nonfatal forms of suicidal behavior has been conducted, the designs have so far been unable to assess how far into the network clustering extends. For example, in one study, college roommates who chose to live together were found to have more strongly correlated suicidal ideation scores than roommates who were randomly assigned (Joiner, 2003). However, this study did not incorporate information on participants' broader social networks, instead collecting only individual dyads. Thus, it still unknown whether suicidal behavior is assortative beyond just one degree of separation, as has been shown for positive emotional states, which are assortative even up to three degrees (Bliss et al., 2012; Fowler & Christakis, 2009). The ability to assess for both subtler behavior and to assess its reach is one of the key advantages of data collected from online

networks. For example, many suicide attempts have been disclosed online that likely would not have been registered in a medical database (Wood, Shiffman, Leary, & Coppersmith, 2016).

When assortativity is expected, it has also been difficult to assess what the actual dimension of similarity is: Are people clustered along suicide-specific characteristics or merely along mood states that are correlated with those characteristics (e.g., depression)? For example, in a study of over 3,000 adolescents from 46 high schools, Randall, Nickel, and Colman (2015) used logistic regression to show that exposure to a friend's suicide attempt increased a participant's risk of suicide attempt over the next year. They attempted to assess for "assortative relating" by controlling for 50 potential confounds (e.g., depression, anxiety), finding that exposure to suicide was still an important predictor of subsequent attempts after inclusion of control variables. But the technique employed—logistic regression—does not actually evaluate assortativity in any of the confounds they attempted to control. Specifically, they would need to model the similarity between the attempter and the participant on these confounds to demonstrate or refute assortativity. Instead, only the overall level of each confound observed in the participant was controlled, not the similarity (i.e., assortativity) between the attempter and the participant on that confound. Thus, it remains an open question whether any form of suicidal behavior—fatal or nonfatal—is actually assortative at all. It may in fact be that another dimension of behavior, such as low mood, is the primary driver of these relationships, and suicidal behavior is ancillary. For effective testing of the assortativity in suicidal behavior, more rigorous, assortativity-specific control procedures must be implemented.

This deficit in our understanding of assortativity in nonfatal suicidal behavior should be especially concerning to researchers, as histories of nonfatal suicidal behavior are widely regarded as the most potent risk factors for subsequent death by suicide (Ribeiro et al., 2016). Indeed, the reader will quickly recognize how improbable it is that any well-established suicide clusters emerged *sui generis*, with members sharing no risk factors prior to their deaths. Clearly more plausible is the expectation that when a cluster of suicide fatalities occurred, some clustering of nonfatal suicidal behavior (e.g., ideation, attempts) preceded it. Thus, for an effective prediction and prevention scheme, quantitative modeling of cluster-like phenomena in nonfatal suicidal behavior is a high priority.

Suicide and Twitter

Social media platforms of all kinds have offered an unprecedented volume, velocity, and variety of data to social scientists (Kern et al., 2016; Kosinski, Matz, Gosling, Popov, & Stillwell, 2015; Russell, 2013). Among these, the most consistently studied is likely Twitter (Twitter, 2016), a "microblogging" platform in which users broadcast 140-character posts directly to one another or to the entire user base of the platform simultaneously. Twitter's scientific popularity is likely owed to how easily accessible it makes its data, the fact that most data collection schemes can be implemented for free, and the ease of data management (Russell, 2013). For example, Twitter limits all posts to 140-character units, which are more easily stored and analyzed than a corpus of long-form blogposts from different servers, each with their own formatting and embedded media (e.g., video, images).

Like other social media platforms, Twitter has also been a boon to suicide researchers, who can now observe the behavior of individual people unobtrusively, collecting time-sensitive information that might not otherwise be shared because of stigma. For example, Wood et al. (2016) were able to analyze 125 users who publicly announced they had attempted suicide, finding that there were distinct signals in their preattempt posts that could have been used to predict their attempts. This is a relatively large sample, given the base rate of attempts (Drapeau & McIntosh, 2016), and it is possible that many of those participants would never have otherwise been studied, especially if their attempts were not severe enough to produce contact with the medical system. Elsewhere, strong correlations have now been documented between the use of suicide-related keywords on Twitter and state-specific age-adjusted suicide rates (Jashinsky et al., 2014); users posting suicide-related content are now understood to exhibit distinct linguistic profiles from those who do not (O’Dea et al., 2015); and unique temporal posting patterns have been detected for service members who died by suicide relative to controls who died from other causes (Bryan et al., 2018). Taken together, these results provide compelling support for the value of the verbal content people post on social media platforms, especially on Twitter. In a range of designs, their verbal behavior has provided unique insight into suicidal behavior of various forms.

Unfortunately, although the scraping and analysis of data retrieved from social network platforms has exploded in suicide research, our understanding of the relationship between social networks and suicide has not. To our awareness, no one has yet studied the actual social network contained in their suicide-related social network data. Given that the structure of the social landscape plays a significant role even in self-destructive behavior of nonsocial bacteria (Berngruber, Lion, & Gandon, 2013), this is a critical deficit in our understanding of suicidal behavior, and it is addressed directly in this investigation.

The Present Investigation

To the authors’ awareness, the network assortativity coefficient has never been directly estimated in any form of suicidal behavior, lethal or otherwise. However, despite this and a few other methodological shortcomings, existing research does imply two important questions about the assortativity of suicidal behavior. First, is any form of suicidal behavior assortative, beyond one degree of social separation? Second, does that assortativity persist, even after another plausible source of assortativity (i.e., mood) has been controlled? That is, holding the distribution of mood in the network constant, will the observed assortativity in suicide-related behavior still be higher than chance? Following previous research on assortative social behavior in humans, it was predicted that suicide-related behavior online would be significantly more assortative than chance, up to three degrees of separation but not further (Christakis & Fowler, 2013). It was predicted the same pattern would hold even when accounting for the distribution of mood in the network. To address these two questions, the present investigation utilized novel bootstrapping methods for the evaluation of assortative mixing in social networks (Christakis & Fowler, 2013; Newman, 2002) using data from a large online social media network, Twitter (Twitter, 2016).

Method

Methodological Overview

This investigation involves several steps, a visual depiction of which is given in Figure 2. As shown by the large arrows in the top left side of the figure, the study began by collecting a random sample of real-time posting activity from Twitter—specifically, two nonconsecutive 28-day periods. The second period was a replication sample for the first. Concurrently, Mechanical Turk (MTurk; Amazon, 2016) raters produced *suicide-relatedness* ratings for each of the 10,222 most common words in contemporary English. When both processes were complete, a machine scored each Twitter post according to whether it used words that were determined by MTurk raters to be highly related to suicide and whether it used words from a preexisting list of “sad” words, which were later used to infer users’ general moods (see second hypothesis). Note that these sad words had been previously rated and validated by a large group of MTurk raters for their relatedness to happiness and sadness, similar to the procedure used to determine suicide relatedness in the current study (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011).

After each post was scored by machine, a new group of MTurk raters then evaluated a random subsample of machine-scored posts, to ensure the validity of machine scoring. Twitter users were then given total scores for both suicide-related verbalizations (SRVs) and for sadness, which were simply the total numbers of posts that used words for the suicide-related and sad word lists, respectively. After user SRV and sadness scores had been totaled for each Twitter user, the social network itself was estimated. Specifically, two users were defined as linked if they engaged in at least one reciprocal reply with one another (i.e., Alice replies to at least one of Bob’s posts, and Bob replies to at least one of Alice’s posts).

Analysis of the resulting network involved three phases. First, the observed assortativity (r) of SRVs in the network was estimated using standard formulas (Newman, 2002). Second, bootstrapping was used to estimate what the SRV assortativity distribution would look like if SRV posting were random in the network—the *marginal null distribution*—and whether the observed r was significantly different from that distribution was assessed (Christakis & Fowler, 2013). Third, bootstrapping was then used again to estimate what the SRV assortativity distribution would look like if SRV scores were conditional on sadness scores in the network—the *mood conditional null distribution*—and whether the observed r was significantly different from that distribution was reassessed. Thus, the observed r at a given degree of separation remains constant but is compared with two different null distribution benchmarks: (a) the *marginal* benchmark distribution, and (b) the *mood conditional* benchmark distribution. This process was conducted at different degrees of separation to assess whether SRV assortativity wanes with increasing social distance in the network. Specifically, this investigation evaluated one through six degrees of separation, as six degrees is expected to be the approximate average social distance between any two people in the world (Dodds, Muhamad, & Watts, 2003; Travers & Milgram, 1967).

Participants

The participants under investigation in this study were all active, English-speaking users of Twitter who posted between July 9, 2016, and August 6, 2016, or between September 12, 2016, and October 10, 2016. This collection procedure yielded a final dataset of 64,499,981 posts, produced by 17,438,868 unique users. There were 456,358 user interactions that involved at least one reciprocal reply between users. This is the total number of links in the network per our definition. The mean number of posts across all users was 3.70 ($SD = 11.82$), with a median of one post per user. Thus, this distribution had a strong positive skew, with 50% of users posting approximately one time, but with the top 5% posting more than 12 times throughout the entire study period. Although traditional demographic information about the users in this sample (e.g., gender, age, race) is not available through Twitter, previous research has indicated that, in general, Twitter users are more likely to be urban and male than the U.S. general population (Mislove, Lehmann, Ahn, Onnela, & Rosenquist, 2011).

Raters

In addition to the Twitter users (participants), this investigation also recruited two distinct groups of MTurk raters (Amazon, 2016). The first ($n = 2,429$) was used to create a measure of whether posts included suicide-related words, used in the scoring procedure described later in this section. The second ($n = 162$) was used as a validity check to evaluate whether the scoring procedure described below indeed tracked intuitive differences in perceived risk of suicidal behavior in the users who produced SRV-positive versus SRV-negative posts.

Measures

Sadness and happiness.—Previous research has established a measurement system—termed a *hedonometer* by its creators—for the assessment of mood in online networks, including Twitter (Dodds et al., 2011). This measure includes mood rating the 10,222 of the most frequently used words in English, including digital slang (e.g., *lol*, *:p*) rated on a scale of 1 (*sad*) to 9 (*happy*). In this study, we operationally define a word as any set of symbols in a Twitter post that are separated by a space on each side or by punctuation and is contained in the hedonometer word list. Additionally, a post is regarded as *sad* (scored 0, 1), if it includes any word with a mood score below 4 (n words = 1,043; 10% of the entire corpus). A user's overall sadness score is then simply the number of posts that were marked as sad during a given study period.²

SRV scores.—The present investigation extends work of Dodds et al. (2011) on mood measurement to the assessment of suicide-related verbalizations in online social media. This process consisted of a measure creation phase and a validity check phase. First, the same 10,222 words utilized for the hedonometer were rerated by a new group of MTurk raters ($n = 2,429$) for their relatedness to suicide. Each rater responded to approximately 100 randomly selected words. Ratings were made on a 1 (*not at all related*) to 9 (*highly related*) scale and the mean rating for a given word was retained. The average reported age of raters was 36.43

²Note that words with the same root are treated as separate words (e.g., *killer* is unique from *killed*), meaning they may have different mood ratings as well as SRV scores.

years ($SD = 12.22$); 56% reported being female, 82% reported being White, and 7% reported being Hispanic or Latino. Raters were paid \$0.50 for their ratings.

Any word with an average suicide-relatedness score at or below 5 was removed. Scores above 5 denote that the word was judged as more related to suicide than not.³ When scoring SRVs for each post (0 or 1), posts that included any happy word (i.e., any word with a hedonometer mood score above 6) were also removed. This latter standard was included to help exclude disingenuous or otherwise colloquial suicide-related posts (e.g., *I'm so happy I could kill myself*).⁴ Thus, a post was treated as SRV-positive (scored 1) by our machine-scored algorithm if it (a) contained at least one word with a suicide-relatedness score greater than 5, and (b) did not contain any happy words (with hedonometer scores above 6). It was regarded as SRV-negative (scored 0) otherwise. Examples of hypothetical posts, how they would be scored, and how those scores are aggregated to create user total scores are depicted in Table 1. All scoring was conducted by machine.

SRV descriptive statistics.: Exactly 219 words were determined to be highly suicide related (see Figure 3). In both study periods combined, the SRV post scoring algorithm flagged 6,453,415 (1%) posts as including at least one word with high suicide relatedness and 24,072,770 (37%) posts as including at least one sad word. Among posts with at least one highly suicide-related word, 5,638,419 (87%) also included a happy word and were ruled out, resulting in a total of 814,996 (1%) of posts marked SRV-positive.

SRV validity.: After each post was scored as SRV-positive or SRV-negative by machine, a new group of SRV-blind MTurk raters ($n = 162$) was recruited to evaluate a random subsample of these posts ($n = 1,500$; 50% SRV-positive), which contained various combinations of individual words, in order to demonstrate that our SRV coding scheme was identifying people who truly appeared to be at risk for suicide. Specifically, raters were asked “Is the author of the following post at risk for suicide?” Ratings were made on a 0 (*extremely unlikely*) to 5 (*extremely unlikely*) scale. Raters were paid \$0.50 for their ratings. Each rater rated approximated 65 posts. The average reported age of raters was 35.96 years ($SD = 12.39$); 66% reported being female, 80% reported being White, and 3% reported being Hispanic or Latino.

If scoring a post SRV-positive tracks something like suicidal ideation or suicide risk more broadly, then the users who authored SRV-positive posts should be perceived by external observers of their posts to be at greater risk for suicide than users who authored SRV-negative posts. Indeed, results showed SRV-positive posts were perceived by MTurk raters as having been produced by users with significantly higher risk for suicide ($b = 0.59$, $SE = .04$, $p < .001$, Cohen's $d = .82$). Thus, SRV-positive posts are unlikely to be dominated simply by

³This is admittedly a somewhat arbitrary cutoff and future research will be required to develop a more comprehensive way to rate suicide-related features of text. However, we note that follow-up analyses revealed setting the cutoff lower (at 4) included roughly 5% of the whole corpus (far too large to be specific to suicide), but setting it higher excluded numerous words that were clearly related to suicide (leading to a set that was too small to be meaningful). As such, the cutoff of 5 was retained in this study.

⁴Note that no additional criteria of this kind were applied to the scoring of sad posts. This was to maintain comparability with previous research using the hedonometer, in which post mood scores were assigned based only on which words are present in a post (regardless of which words are absent).

casual or sarcastic comments, with raters having been explicitly told to rate such comments as having lower risk.

Network Assembly

Following Bliss et al. (2012) work on the assortativity of mood on Twitter, two users were designated as *linked* in the network when they have each replied to one another at least once. For example, Alice and Bob are linked if (a) Alice has replied to at least one of Bob's posts, and (b) Bob has replied to at least one of Alice's posts (i.e., reciprocal replies) in the period under consideration. As opposed to other link-coding approaches (e.g., liking another user's posts, retweeting another user's posts, following another user), this reciprocal-reply requirement is an especially conservative criterion for regarding people as socially connected because it requires they have each intentionally interacted with each other at least once (Bliss et al., 2012).

Network Analysis

Observed assortativity.—A network is a collection of nodes (people) and the links (interactions) among them (Newman, 2010). Here, the nodes also have attributes of interest: their total numbers of SRV-positive posts and sad posts. The propensity of linked nodes to have similar numbers of SRV-positive posts is estimated with the assortativity coefficient, r , detailed in the introduction (Newman, 2002).

Degrees of separation.—Assortativity quantifies the similarity of nodes across links in a network. To assess how that quantity changes at increasing levels of social distance, the assortativity coefficient was estimated for each network at one through six degrees of separation. That is, the assortativity of direct friends (links of Degree 1), friends of friends (links of Degree 2), friends of friends of friends (links of Degree 3), and so on up to Degree 6, were all estimated separately.

Hypothesis tests.—It is important to assess whether that observed level of assortativity could plausibly be the result of chance. The observed assortativity coefficient was thus compared with two benchmarks. The first benchmark was the *null marginal assortativity distribution* (or simply, *the marginal null*), which represents the range of values an observed network's assortativity coefficient could take on if (a) the network structure remained the same, but (b) SRV scores were assigned to the nodes at random.⁵ This is achieved by a bootstrapping technique described by Christakis and Fowler (2013).

The second benchmark was termed the *null mood conditional assortativity distribution* (or simply, the conditional null), which represents the range of values an observed network's assortativity coefficient could take on if (a) the network structure remained the same, (b) the user sadness scores in the network remained the same, and (c) the SRV scores were shuffled within sadness score groups. Phrased statistically, the conditional null represents what SRV

⁵Note that the network remains constant during this bootstrapping process, based on recommendations from previous literature on this topic (Christakis & Fowler, 2013).

assortativity scores would look like if SRVs were random, once the sadness scores of users were controlled.

Results

As expected from previous research (Bliss et al., 2012; Christakis & Fowler, 2013), observed SRV assortativity values generally declined in magnitude as social distance (degrees of separation) increased.⁶ Specifically, users' SRV scores were most similar to the SRV scores of people with whom they interact directly ($r_1 = .08-.10$, $p_{\text{marg}} < .001$ for both periods, $p_{\text{cond}} < .001-.004$). Then, they become slightly less like the SRV scores of people who are two degrees away ($r_2 = .06$ for both periods, $p_{\text{marg}} < .001$ for both periods, $p_{\text{cond}} < .001$ for both periods). Lastly, they are least similar to the scores of those who are six degrees away ($r_6 = .01-.02$, $p_{\text{marg}} < .001$ for both periods, $p_{\text{cond}} = .002-.772$).

Results of the assortativity hypothesis tests are depicted in Figure 4. To walk the reader through the figure, recall that we are concerned with a few priorities when testing assortativity values: (a) What is the observed value of assortativity, (b) does that value change at higher degrees of separation, (c) is that observed value greater than would be expected by random chance, and (d) is it still greater than chance after mood is controlled? To represent these priorities, we first depict observed assortativity values as white circles. The farther right in the figure that these circles are located, the greater the assortativity that was observed. To depict different degrees of separation, we give each observed assortativity value its own horizontal continuum to move along—12 in total, for six degrees in each of two study periods. Looking only at the white circles, notice that the observed assortativities decrease (move leftward) as the degrees of separation increase (move upward).

To assess whether these observed values are significantly greater than what would be expected by random chance, we plot the bootstrapped marginal null distribution (dark hills on left side of the plot) for each observed assortativity value. Comparing the observed points with their respective marginal nulls, we see that the observed points are all far greater (farther right) than would be expected by random chance. That is, all of these points occupy a low-probability region of their null marginal distributions (dark gray hills)—either the end of or completely off the distribution's tail (top 2.5%). To see whether this effect was still significant when mood was controlled, we plotted the mood conditional null distributions in light gray. Unlike the marginal null, some of the observed assortativity circles are not significantly greater than what we would expect to see, given the known mood pattern in the data. To aid the reader, assortativity scores that are significant even when mood is controlled are marked with asterisks.

Observed Assortativity Versus Marginal Null (Random SRVs)

Recall that the first hypothesis in this investigation was that SRVs would be significantly assortative, up through three degrees of separation. Results show this prediction was met and

⁶A few small wrinkles are observed in this pattern. For example, the transition from three to four to five degrees of separation in Period A is not a perfectly decreasing trend (i.e., four degrees is slightly less assortative than five). However, statistical noise is likely the best explanation for this small deviation from the broader expected pattern, given that unexpected trend was not observed in both periods.

exceeded, with SRVs demonstrating significant assortativity all the way through six degrees of separation. Specifically, all observed assortativity values were as high or higher than the top 2.5% of bootstrapped values from the null marginal distribution regardless of study period or degree of separation. Phrased practically, a randomly selected person from this network is more similar to the friends of their friends' friends' friends' friends' friends than would be expected by chance. The practical implications of this finding for prevention is further demonstrated via simulation later in the Discussion.

Observed Assortativity Versus Mood-Conditional Null

The second hypothesis in this investigation was that SRVs would remain significantly assortative, even after mood was controlled in the network. In both study periods, observed assortativity values were significantly greater than would be expected at both one and two degrees of separation, even when sadness scores were controlled. In other words, the assortativity of SRV at one and two degrees of separation is too high to be a plausible artifact of mood assortativity. Thus, results indicate the observed assortativity of SRV on Twitter is not statistically attributable to the already well-documented assortativity in mood (Bliss et al., 2012).

However, at social distances greater than two, the observed assortativity at each increasing degree of separation was typically significant in only one of the two study periods, with only Degree 5 being significant in both periods. Absent a compelling justification for why five degrees of separation are in some way unique from three, four, and six degrees, a scientifically conservative interpretation of these results is that the observed assortativity coefficients for Degrees 3 through 6 cannot be reliably distinguished from what would be expected given the already established assortativity of mood. That is, following the often-neglected insight that negative scientific evidence is more powerful than positive (Meehl, 1978), it is argued these mixed results for degrees of three and above should be treated as “functionally null.”

Discussion

This study investigated the connection between a social network (i.e., the Twitter network) and the suicide-related behavior enacted by individuals who occupy it (i.e., verbal behavior involving suicide-related words, or SRVs). Although several studies have previously attempted to evaluate the similarities of people at risk for suicide or the clustering of suicidal behavior, to our awareness, this is the first study to directly estimate the assortativity of any kind of suicide-related behavior. It is also among the first of its kind to evaluate a suicide-related network of this scale, having analyzed over 64 million posts from more than 17 million unique users in two 28-day periods. Consistent with the expectations of previous research on the assortativity of mood (Bliss et al., 2012; Fowler & Christakis, 2009), SRVs were observed to be significantly more assortative than would be expected by chance alone—remaining significant up to at least six degrees of separation between individuals in the network. When mood was controlled, the assortativity of SRVs was still significantly higher than would be expected by chance, up to at least two degrees of separation.

Taken together, our results suggest that verbal behavior related to suicide (i.e., SRVs) occurs in a socially assortatively patterned way on Twitter, with individuals who engage in such verbal behavior having a significantly higher likelihood of interacting with one another than chance. Although the magnitude of this correlation-style estimate was modest (i.e., $r_1 = .08-.10$), its reach was quite substantial—observed at up to six degrees of separation.⁷ Moreover, in the neighborhood encompassed by two degrees of separation, the assortativity of SRVs was not attributable to the already known assortativity of mood.

Implications

There are two primary implications arising from the findings presented earlier: one pragmatic and one etiological. Pragmatically, the results imply that for at least one form of suicide-related behavior, birds of a feather flock together in a statistically systematic way. Assuming that more serious forms of suicidal behavior (e.g., suicide attempts) follow a similar pattern—a priority question for future research—only a small number of known birds are needed to locate much of the flock. It is emphasized again that this is true regardless of the causal origin of that assortative flocking pattern.

The most obvious way to exploit such a pattern is to use snowball sampling when screening for suicide risk, which we now illustrate. First, consider Figure 5, which depicts a random subset of the first period's network. In this subsample, there are 107 total people, 29 of which produced at least one SRV-positive post during the study period (dark nodes). Among those 29 SRV-positive users, 17 knew at least one other SRV-positive user (highlighted with stars). Thus, the probability an SRV-positive user knew at least one other is $17/29 = .59$ and the probability an SRV-negative user knew at least one SRV-positive user is $42/78 = .54$. This difference may appear modest, but it is practically quite powerful.

For example, imagine the network in Figure 5 represents a small high school, in which one student has recently taken his life. Concerned about the suicide risk of the remaining students, an administrator asks the district's lone school psychologist to locate any other students that are at risk so they can be referred for psychotherapy. Unfortunately, the school psychologist knows she only has the time and resources to perform 25 total screens, guaranteeing 82 students will remain unscreened. She has a list of every student in the school, but how does she choose who to screen?

The most likely approach is simple random screening (without replacement), in which the school psychologist chooses 25 people from her list in advance, then screens them one after another. Such a scenario will follow a hypergeometric distribution (Casella & Berger, 2001), in which the mean number of true positives she will locate after 25 screens is 6.78 (i.e., a sensitivity of $6.78/29 = .23$). However, if she is aware that suicide risk has nonzero assortativity, she can exploit that pattern with snowball sampling: (a) Ask a teacher who the

⁷For readers looking for examples of common assortativity strengths, see Bliss et al. (2012) but also Newman (2002, 2003). The assortativity observed here is similar to the similarity between math and biology coauthorships. In a loose sense, academic readers can interpret this coefficient as indicating two people on Twitter are about as similar on SRVs as researchers are with their coauthors. That is, by no means are coauthors expected to be identical, but by knowing a little information about only one author on an article, it is possible to guess (with a likely surprising accuracy) much of the information about the others. For example, they are likely interested in similar research topics and likely share many demographic characteristics.

decendent's closest friends were and screen them; (b) ask any positives from that group to list their closest friends and screen those friends; (c) ask any positives from that new group who their closest friends are, screen them, and so on; and (d) if ever there are no more positives in a friend group, screen students at random until a positive is found and begin the snowball procedure again until resources run out.

If this strategy is simulated 1,000 times using the data in Figure 5, the average number of true positives using 25 screens is 8.30 (sensitivity = .29), about 22% higher than random sampling's average. Beyond the averages, comparing their overall distributions shows that exploiting even the modest assortativity of $r_1 = .12$ from this network will cause the snowball-style screening to produce a higher rate of true positives than random sampling 68% of the time, even though the same total number of people are screened in each case. Readers will also note these results are from a small network, with an even smaller number of available screens. But as the size of the network and available screens grows (e.g., online), the superiority of the assortativity-informed snowball approach accelerates nonlinearly. For example, if there were 1,000 screens applied to the whole first-period network ($n = 168,098$, 21% SRV-positive), the random sampling approach would find on average 207 true positives, but the snowball approach would find an average of 358 (across 10 simulated runs), a 73% improvement. The advantage of this assortativity-exploiting sampling approach thus grows nonlinearly with the size of the network, making it ideal for improving the accuracy and resource efficiency of suicide prevention in large scale contexts like the military, Veterans Affairs, school districts, and online social media platforms.

Aside from the obvious benefits of enhancing the rate of true positives located during mass screening, capitalizing on assortativity is likely to have other benefits for suicide research as well. For example, researchers struggling to acquire participants from a low-base-rate group like suicide attempters could employ assortativity-informed recruitment strategy like the snowball approach.⁸ Second, and more speculatively, it is possible that assortativity-informed interventions could produce positive ripple effects throughout the social network after the treatment of only a few at-risk nodes. Such a possibility would improve both issues related to limited resources and difficulties reaching at-risk people who would not otherwise present for treatment (i.e., because they would indirectly receive a dose of treatment through their friends who were treated). It is emphasized, however, that this treatment approach first requires confirmation about the causes of suicide-related assortativity in social networks, which the current study was not designed to disentangle. Although the current results provide an initial hint of an assortativity-informed treatment approach, additional research will be required prior to piloting interventions of this kind.

A final comment here: although our school example was employed only for its simplicity, the reader may still wonder, what do the current results say about the assortativity of suicide-related behavior in offline networks? In the interest of transparency, it is also noted the current results say little about behavior in offline networks because no offline behavior was

⁸Researchers concerned about this admittedly nonrandom sampling approach are reminded that (a) their existing samples were already highly likely to be nonrandom, (b) this kind of nonrandom sampling bias is better than the status quo because it is quantifiable and thus able to be accounted in the final model, and (c) knowledge of the social relationships between participants in a study actually opens a whole new dimension of investigation (e.g., the effects of the network on the study topic).

studied here. However, as explained in the introduction, assortativity is one of the defining statistical signatures of a vast majority of human social networks studied so far (Newman, 2002, 2003). Thus, there is good reason to expect that if it has been observed online, it will be observed offline as well. For example, previous research on mood has shown it to be assortative in both online and offline contexts (Bliss et al., 2012; Fowler & Christakis, 2009). In addition, the observed clustering of suicides in both time and space imply that at least some kind of assortative pattern must apply to at least some kinds of suicidal behavior (Haw et al., 2013).

Although further research is needed, there is good reason to expect some forms of suicidal behavior (e.g., ideation, attempt risk) in offline networks to be as assortative, as was observed here, if not more so. Indeed, it is even plausible that the long-form interaction available offline will facilitate even higher assortativity than was observed here on Twitter, where interactions are limited only to 140-character bursts available for all the public to see. Instead, the privacy, information density, and potential for a much higher number of interactions (albeit with fewer others) in offline social relationships likely allows for much more assortativity in suicidal behavior to emerge over time, and the authors look forward to future research investigating this possibility.

The second implication of these results is related to our more basic understanding of the etiology of suicide. Specifically, the mood-controlled results suggest that the observed clustering of suicide-related behavior in the social environment is not merely a side effect of preexisting clusters of people with similar moods, who also happen to be at greater risk for suicidal behavior because their moods are low. Rather, results imply that a suicide-specific mechanism is likely to be involved in the assortative clustering of at least suicide-related verbal behavior.

Thus, although it is not currently possible to disentangle the causal mechanism of assortative patterns in nonexperimental network settings (Shalizi & Thomas, 2011), the current results have still narrowed the search. That is, whether the cause of this assortative pattern is social contagion, homophily, or shared history, it is likely to be a suicide-specific class of one of those mechanisms. For example, people may exhibit similar SRV scores because one high-SRV member of the network unintentionally influences his neighbors to engage in such suicide-related verbal behavior as well (social contagion); however, the present results imply it is unlikely that the similarity of SRV scores among socially connected people is merely the result of one having influenced the others to become sad, which had the secondary consequence of also increasing their propensity to utilize highly suicide-related words.

Interpretive Caveats and Future Directions

This study breaks ground in a new area of suicide research, utilizing a model capable of reconciling large group-level patterns with the individual-level behavior that gives rise to them; however, as an initial study in this area, several caveats are important for the interpretation of results. Most importantly, this investigation employed a relatively coarse-grained measure of individual behavior, SRVs. This was intentional, as the present study was designed to demonstrate the significance of potential large-scale assortativity in suicidal

behavior rather than to screen and refer people for intervention. However, the limitations of the present study do still imply a few important agenda items for future research to resolve.

First, it will be important to confirm that more refined methods for measuring suicide-related verbal behavior online remain assortative, like the SRV measure here. For example, an intuitive next step would be to assign each post a score (e.g., cosine similarity) based on its grammatical similarity to a known set of suicide notes (Pestian et al., 2012). Using such scores, it would be possible to ask whether people with a propensity to write content with high similarity to a suicide note associate more often with one another than chance. More directly, is suicide-note-like posting also assortative?

A related question is what kinds of posts should be included in any kind of SRV calculation? For example, this investigation counted “retweets”—cases in which one user reposted another user’s original post—in the sadness and SRV calculations. Thus, if one user produced an SRV-positive post and a second user retweeted that post, the second user’s SRV score would increase. This was done to treat all posts as equal, as little empirical means were available to prioritize one kind of post over another. It is worth noting this decision could be either desirable or undesirable for the measurement of SRVs. On one side, being at elevated risk for suicide may increase the likelihood a user retweets a suicide-related post, making it important to include such reposts in their SRV score. On the other side, it may also be that users predominantly retweet other users’ suicide-related posts out of concern for those users, making such reposting unreflective of the reposters actual suicide risk. It is conceded that this is an open question in need of empirical resolution. Depending on its resolution, the assortativity coefficients in this study could increase or decrease; however, it is the authors’ general expectation that they would increase, as resolving this question would reduce another form of random error variance, resulting in more precise measurement.

Second, it is critical to verify whether the kinds of suicide-related verbal behavior that are assortative are also predictive of more serious suicidal behavior offline. Specifically, there is already good reason to believe that online verbal behavior is predictive of offline suicide risk (Bryan et al., 2018; Wood et al., 2016), but it is not clear whether the features of verbal behavior that are risk-predictive are themselves assortative. For example, imagine that the ratio of verbs to nouns a user posted online predicted offline suicide attempts in some way but that verb–noun ratios were not assortative among users—there was no “flocking” pattern. In such a case, the practical advantages accrued from network models would likely be null: Assortativity-driven snowball screening would be no better than random screening, and would probably be worse.

Importantly, our validity check shows that external raters perceive the users who produced SRV-positive posts to be at greater risk for suicide attempts. Coupled with the fact that this study observed reliable assortativity in those SRV scores across two independent periods, this implies that at least some features of suicide risk are likely to be assortative. However, a study that can directly test the assortativity of a strong index of suicide risk—or of suicidal behavior directly—is an important next step in this line of research, without which strong recommendations cannot be made.

Lastly, this study was only able to evaluate one online network, Twitter. There are important features of Twitter that make it different from offline networks (e.g., the ability to directly approach and interact with celebrities) and from other online networks (e.g., nearly everything is public, posts are all limited to 140 characters). Additionally, the demographic characteristics of at least U.S. Twitter users are different from the general U.S. population, a difference that may hold for other online and offline networks as well. For these reasons, a variety of network-related phenomena could likely manifest differently in other networks. Establishing quantitative models of their potential differences is thus a priority before augmenting prevention schemes in those networks with the approaches discussed here.

Conclusion

Assortativity is a hallmark of human social networks (Newman, 2002), observed in numerous and disparate forms of behavior (Christakis & Fowler, 2013). This investigation demonstrates that suicide-related verbal behavior is also among them and suggests that other more serious forms of suicidal behavior might be as well. Notably, the coefficients in this investigation were of only modest magnitude, but they demonstrated substantial social reach—lasting up to at least six degrees of separation. If exploited, such a pattern has the capacity to substantially improve screening approaches in both small and large networks alike, increasing the number of people at risk for a suicide attempt found in advance, using only existing resources.

Biographies



Ian Cero



Tracy K. Witte

References

- Amazon. (2016). Amazon Mechanical Turk. Retrieved from <https://www.mturk.com/mturk/help?helpPage=overview>
- Aral S, Muchnik L, & Sundararajan A (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 21544–21549. 10.1073/pnas.0908800106 [PubMed: 20007780]
- Batterham PJ, Calear AL, & Christensen H (2013). Correlates of suicide stigma and suicide literacy in the community. *Suicide and Life-Threatening Behavior*, 43, 406–417. 10.1111/sltb.12026 [PubMed: 23556504]

- Berngruber TW, Lion S, & Gandon S (2013). Evolution of suicide as a defence strategy against pathogens in a spatially structured environment. *Ecology Letters*, 16, 446–453. 10.1111/ele.12064 [PubMed: 23331662]
- Bliss CA, Kloumann IM, Harris KD, Danforth CM, & Dodds PS (2012). Twitter reciprocal reply networks exhibit assortativity with respect to happiness. *Journal of Computational Science*, 3, 388–397. 10.1016/j.jocs.2012.05.001
- Brent DA, Kerr MM, Goldstein C, Bozigar J, Wartella M, & Allan MJ (1989). An outbreak of suicide and suicidal behavior in a high school. *Journal of the American Academy of Child & Adolescent Psychiatry*, 28, 918–924. 10.1097/00004583-198911000-00017 [PubMed: 2808263]
- Bryan CJ, Butner JE, Sinclair S, Bryan ABO, Hesse CM, & Rose AE (2018). Predictors of emerging suicide death among military personnel on social media networks. *Suicide and Life-Threatening Behavior*, 48, 413–430. 10.1111/sltb.12370 [PubMed: 28752655]
- Casella G, & Berger RL (2001). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury Press.
- Cheng Q, Li H, Silenzio V, & Caine ED (2014). Suicide contagion: A systematic review of definitions and research utility. *PLoS ONE*, 9, e108724. 10.1371/journal.pone.0108724 [PubMed: 25259604]
- Christakis NA, & Fowler JH (2007). The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine*, 357, 370–379. 10.1056/NEJMsa066082 [PubMed: 17652652]
- Christakis NA, & Fowler JH (2008). The collective dynamics of smoking in a large social network. *The New England Journal of Medicine*, 358, 2249–2258. 10.1056/NEJMsa0706154 [PubMed: 18499567]
- Christakis NA, & Fowler JH (2013). Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine*, 32, 556–577. 10.1002/sim.5408 [PubMed: 22711416]
- Dodds PS, Harris KD, Kloumann IM, Bliss CA, & Danforth CM (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, 6, e26752. 10.1371/journal.pone.0026752 [PubMed: 22163266]
- Dodds PS, Muhamad R, & Watts DJ (2003). An experimental study of search in global social networks. *Science*, 301, 827–829. 10.1126/science.1081058 [PubMed: 12907800]
- Drapeau CW, & McIntosh JL (2016). U.S.A. suicide 2015: Official final data Washington, DC: American Association of Suicidology Retrieved from <http://www.suicidology.org/Portals/14/docs/Resources/FactSheets/2015/2015datapgsv1.pdf?ver=2017-01-02-220151-870>
- Farmer R, O'Donnell I, & Tranah T (1991). Suicide on the London Underground System. *International Journal of Epidemiology*, 20, 707–711. 10.1093/ije/20.3.707 [PubMed: 1955255]
- Fowler JH, & Christakis NA (2009). Dynamic spread of happiness in a large social network: Longitudinal analysis of the Framingham Heart Study social network. *BMJ*, 338, 23–27. 10.1136/bmj.a2338
- Franklin JC (2016). How technology can help us move toward large-scale suicide risk detection and prevention [Webinar]. Retrieved from <http://www.suicidology.org/store/BKctl/ViewDetails/SKU/AASTECHRISKWE>
- Franklin JC, Fox KR, Franklin CR, Kleiman EM, Ribeiro JD, Jaroszewski AC, ... Nock MK (2016). A brief mobile app reduces nonsuicidal and suicidal self-injury: Evidence from three randomized controlled trials. *Journal of Consulting and Clinical Psychology*, 84, 544–557. 10.1037/ccp0000093 [PubMed: 27018530]
- Gould MS (1990). Suicide clusters and media exposure In Blumenthal SJ & Kupfer DJ (Eds.), *Suicide over the life cycle: Risk factors, assessment, and treatment of suicidal patients* (pp. 517–532). Arlington, VA: American Psychiatric Association.
- Gould MS, Petrie K, Kleinman MH, & Wallenstein S (1994). Clustering of attempted suicide: New Zealand national data. *International Journal of Epidemiology*, 23, 1185–1189. 10.1093/ije/23.6.1185 [PubMed: 7721521]
- Gould MS, Wallenstein S, & Davidson L (1989). Suicide clusters: A critical review. *Suicide and Life-Threatening Behavior*, 19, 17–29. 10.1111/j.1943-278X.1989.tb00363.x [PubMed: 2652386]
- Haw C, Hawton K, Niedzwiedz C, & Platt S (2013). Suicide clusters: A review of risk factors and mechanisms. *Suicide and Life-Threatening Behavior*, 43, 97–108. 10.1111/j.1943-278X.2012.00130.x [PubMed: 23356785]

- Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, & Argyle T (2014). Tracking suicide risk factors through Twitter in the U.S. *Crisis*, 35, 51–59. 10.1027/0227-5910/a000234 [PubMed: 24121153]
- Joiner TE Jr. (1999). The clustering and contagion of suicide. *Current Directions in Psychological Science*, 8, 89–92. 10.1111/1467-8721.00021
- Joiner TE Jr. (2003). Contagion of suicidal symptoms as a function of assortative relating and shared relationship stress in college roommates. *Journal of Adolescence*, 26, 495–504. 10.1016/S0140-1971(02)00133-1 [PubMed: 12887937]
- Kern ML, Park G, Eichstaedt JC, Andrew H, Sap M, Smith LK, & Ungar LH (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods*, 21, 507–525. 10.1037/met0000091 [PubMed: 27505683]
- Klimes-Dougan B, Safer MA, Ronsaville D, Tinsley R, & Harris SJ (2007). The value of forgetting suicidal thoughts and behavior. *Suicide and Life-Threatening Behavior*, 37, 431–438. 10.1521/suli.2007.37.4.431 [PubMed: 17896883]
- Kosinski M, Matz SC, Gosling SD, Popov V, & Stillwell D (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70, 543–556. 10.1037/a0039210 [PubMed: 26348336]
- Luxton DD, June JD, & Fairall JM (2012). Social media and suicide: A public health perspective. *American Journal of Public Health*, 102(Suppl. 2), S195–S200. 10.2105/AJPH.2011.300608 [PubMed: 22401525]
- Meehl PE (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806–834. 10.1037/0022-006X.46.4.806
- Mislove A, Lehmann S, Ahn YY, Onnela JP, & Rosenquist JN (2011, 7). Understanding the demographics of Twitter users. In Fifth international AAAI conference on weblogs and social media Retrieved from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2816/3234>
- Newman ME (2002). Assortative mixing in networks. *Physical Review Letters*, 89, 208701 10.1103/PhysRevLett.89.208701 [PubMed: 12443515]
- Newman ME (2003). Mixing patterns in networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 67, 026126 10.1103/PhysRevE.67.026126
- Newman M (2010). *Networks: An introduction*. Oxford, UK: Oxford University Press 10.1093/acprof:oso/9780199206650.001.0001
- Newman ME, & Park J (2003). Why social networks are different from other types of networks. *Physical Review E*, 68, 036122.
- O’Dea B, Wan S, Batterham PJ, Calear AL, Paris C, & Christensen H (2015). Detecting suicidality on Twitter. *Internet Interventions: The Application of Information Technology in Mental and Behavioural Health*, 2, 183–188. 10.1016/j.invent.2015.03.005
- Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzuner O, Wiebe J, ... Brew C (2012). Sentiment analysis of suicide notes: A shared task. *Biomedical Informatics Insights*, 5(Suppl. 1), 3–16. 10.4137/BII.S9042 [PubMed: 22419877]
- Randall JR, Nickel NC, & Colman I (2015). Contagion from peer suicidal behavior in a representative sample of American adolescents. *Journal of Affective Disorders*, 186, 219–225. 10.1016/j.jad.2015.07.001 [PubMed: 26253902]
- Ribeiro JD, Franklin JC, Fox KR, Bentley KH, Kleiman EM, Chang BP, & Nock MK (2016). Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: A meta-analysis of longitudinal studies. *Psychological Medicine*, 46, 225–236. 10.1017/S0033291715001804 [PubMed: 26370729]
- Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, & Herrman H (2016). Social media and suicide prevention: A systematic review. *Early Intervention in Psychiatry*, 10, 103–121. 10.1111/eip.12229 [PubMed: 25702826]
- Russell MA (2013). *Mining the social web: Data mining Facebook, Twitter, LinkedIn, Google+, GitHub, and more* (2nd ed.). Sebastopol, CA: O’Reilly Media.

- Sanderson D (2009). Programming google app engine: Build and run scalable web apps on Google's infrastructure. Retrieved from https://books.google.com/books?hl=en&lr=&id=6cL_kCZ4NJ4C&oi=fnd&pg=PR7&dq=mobile+app+scalable&ots=sKhhMVMZfh&sig=_3qSzEYQK-Q2cyD_erRIPv-jP2o
- Shalizi CR, & Thomas AC (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40, 211–239. 10.1177/0049124111404820 [PubMed: 22523436]
- Travers J, & Milgram S (1967). The small world problem. *Psychology Today*, 1, 61–67.
- Twitter. (2016). Company: About. Retrieved from <https://about.twitter.com/company>
- Wood A, Shiffman J, Leary R, & Coppersmith G (2016). Language signals preceding suicide attempts CHI 2016 Computing and Mental Health Workshop, San Jose, CA Retrieved from http://chi2016mentalhealth.media.mit.edu/wp-content/uploads/sites/46/2016/04/language_signals_preceding_suicide_attempts.pdf

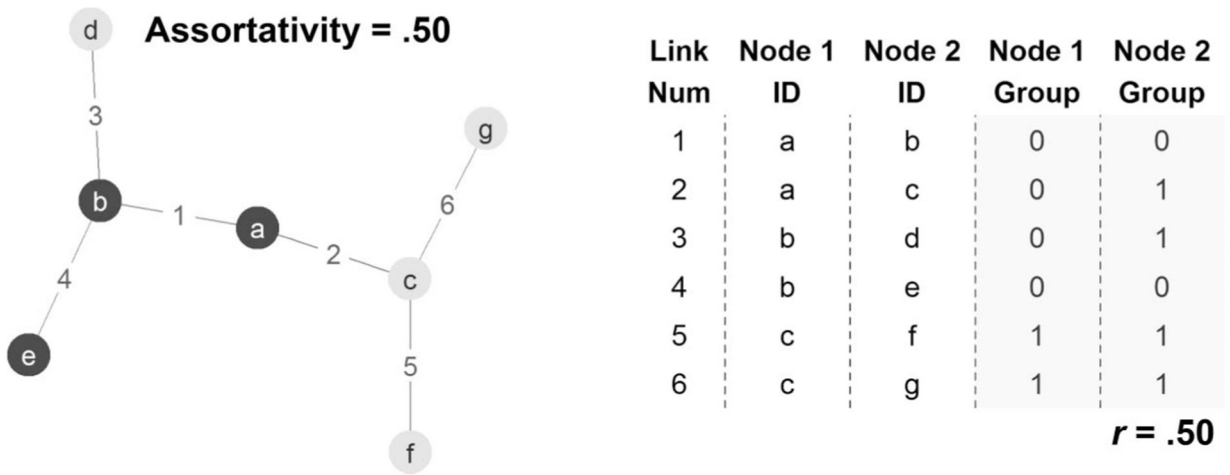


Figure 1.

Illustration of assortativity as correlation of node attributes across links. The left panel depicts an example seven-node/six-link network, in which each node belongs to one of two groups (0 = dark gray, 1 = light gray). Nodes are labeled with arbitrary letter IDs, and links are labeled with arbitrary number IDs. Group membership is assortative in this network ($r = .50$), with 60% of links holding like-type nodes on either end. The right panel shows the same network, now represented as an “edge list,” in which each row is a link and each column is an attribute of that link (e.g., its link ID number) or of the nodes residing on either of its ends. Note that when a network is represented in this way, the Pearson correlation between the two columns of a given node attribute—group membership in this case—is identical to the assortativity of that attribute in the network (with some specific exceptions).

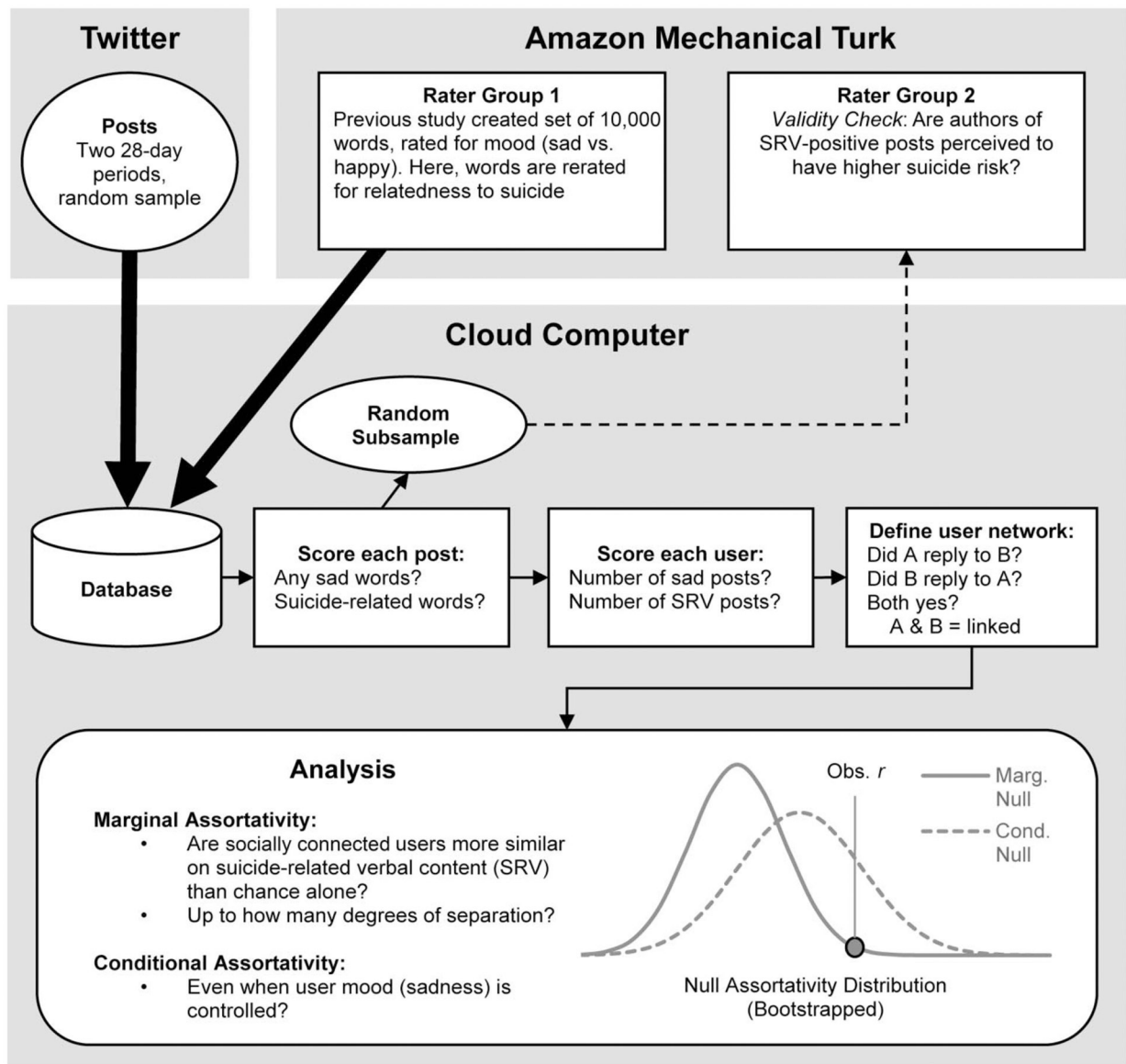


Figure 2.

Diagram of study process. Investigation begins with the collection of two discontinuous 28-day periods of Twitter posts (a random sample of all posts). Simultaneously, Mechanical Turk (MTurk) Group 1 rerated an existing set of 10,000 English words for relatedness to suicide (MTurk Group 1). Word ratings were used to define sad posts and posts with suicide-related verbal content (SRVs). Posts were then scored by machine, but a random subsample were also reevaluated by humans (MTurk Group 2, blind to machine rating). Users were assigned total scores base on their post scores and linked to other users with whom they mutually interacted to establish a network. Data were analyzed as described in the Method section. Obs. r = observed SRV assortativity value for the observed network. Marg. = Marginal; Cond = Conditional.

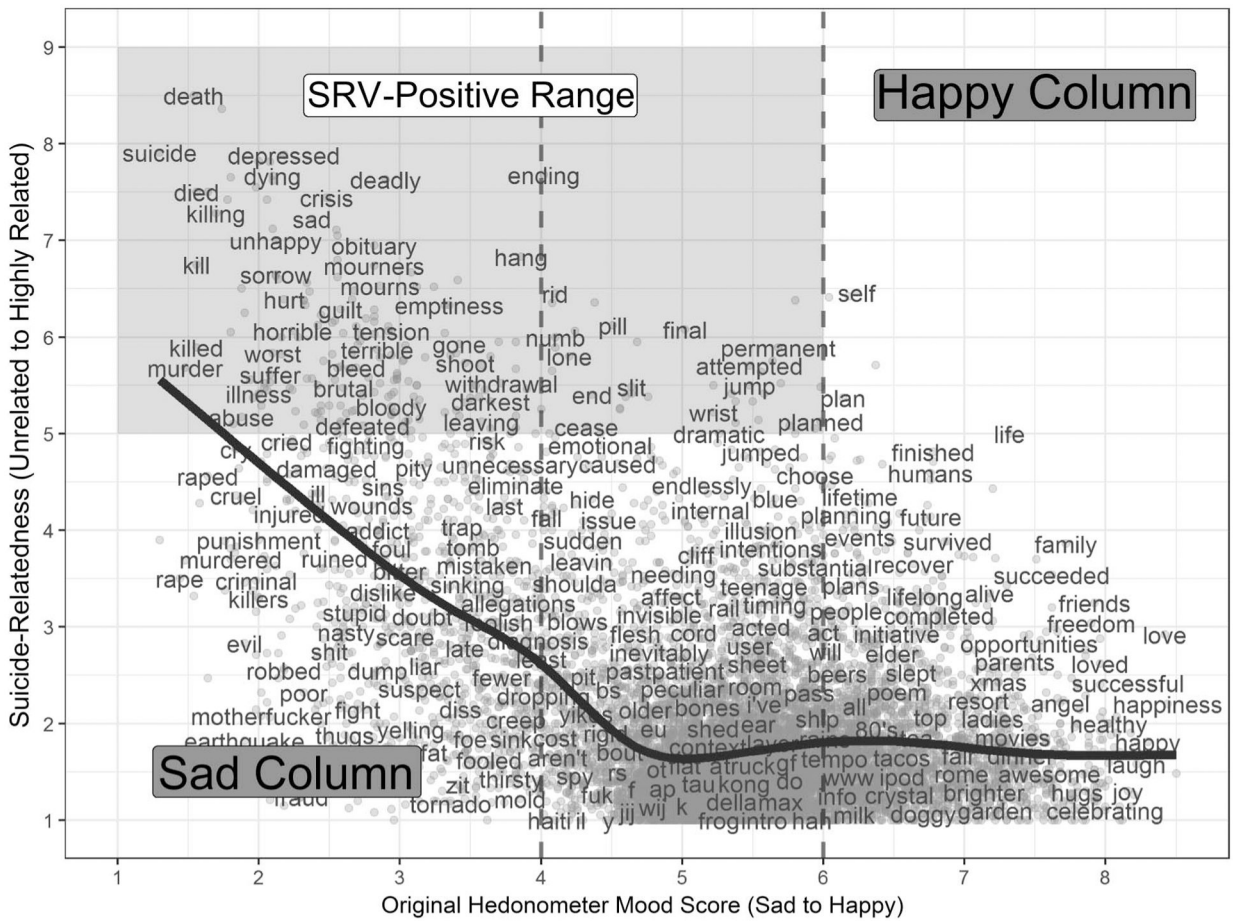


Figure 3. Mood and suicide relatedness of the 10,222 most common words in contemporary English. Original mood scores are plotted along the horizontal axis and the suicide-relatedness ratings from the current study are plotted along the vertical axis. Each word is represented by a gray dot, a random subsample of which also have text displayed for the reader. The curved solid line represents the smoothed relationship between mood scores (sad to happy) and suicide-relatedness scores among words. The vertical dashed lines demarcate sad and happy words (mood scores below 4 or above 6, respectively). The shaded box represents the set of words leading to a positive SRV score for a post. Application of these scores is further described in the Method section and illustrated for hypothetical posts in Table 1. SRV = Suicide-related Verbalization.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

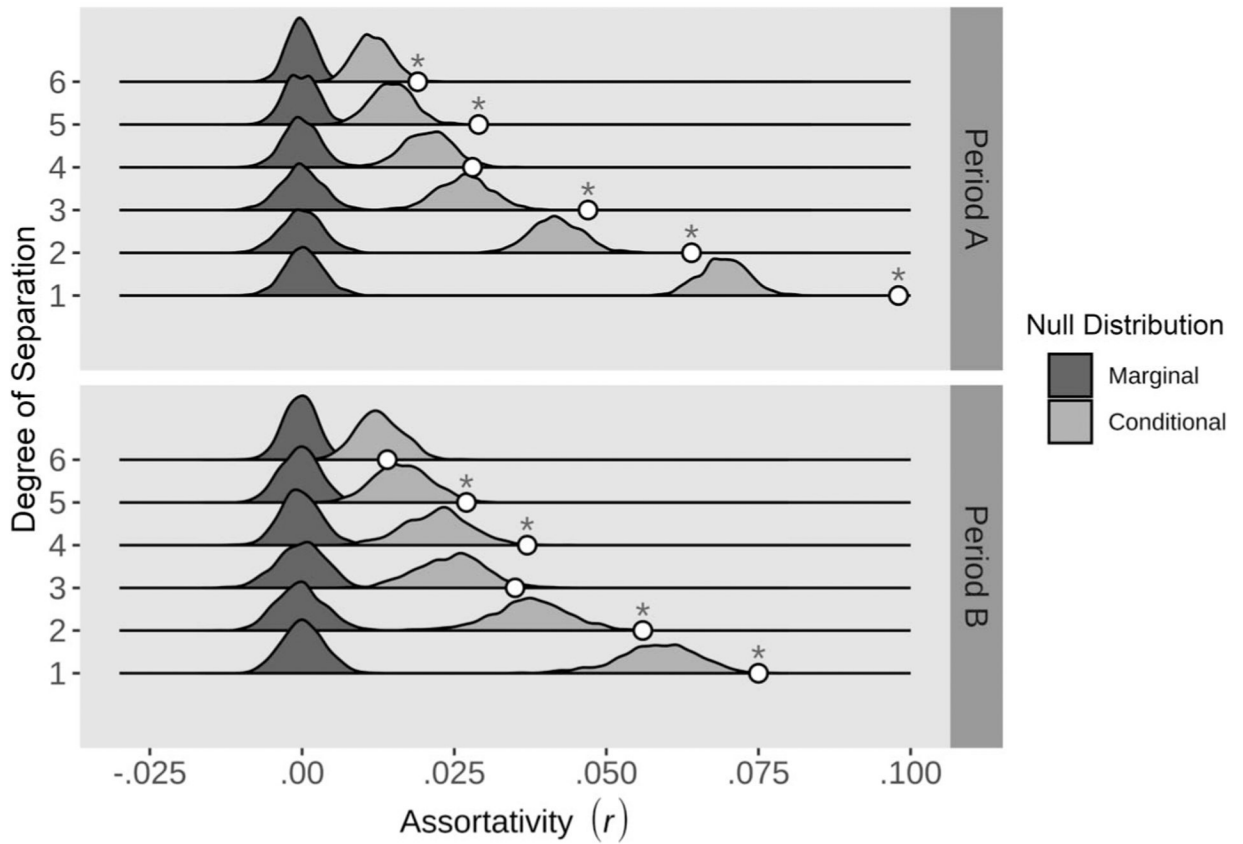


Figure 4.

Observed assortativity (r) up to six degrees of separation and across two periods. Each white circle represents an assortativity estimate (horizontal axis) for the network at a given degree of separation (vertical axis). The bootstrapped null distributions for both marginal suicide-related verbalization (SRV) assortativity (dark) and conditional SRV assortativity (light) are also provided as benchmarks for evaluating the statistical significance of those sample assortativity estimates. Consistent with traditional hypothesis testing, a sample assortativity estimate is significant when it occupies a low density (low probability) region of its null distribution. In this case, all observed assortativity estimates are significantly different than the marginal null (completely random SRVs). Asterisks (*) denote which observed assortativity values are also significantly different from the conditional null (i.e., they are not simply an artifact of mood at a given degree of separation). Note that because marginal null distributions are completely random, they will always center around zero assortativity; however, because the conditional null distributions account for the assortativity in sadness (which is known to decline as a function of social distance), these distributions will center at difference values. Specifically, they will be centered on the most plausible SRV assortativity value one would expect, knowing the sadness scores of everyone in the network.

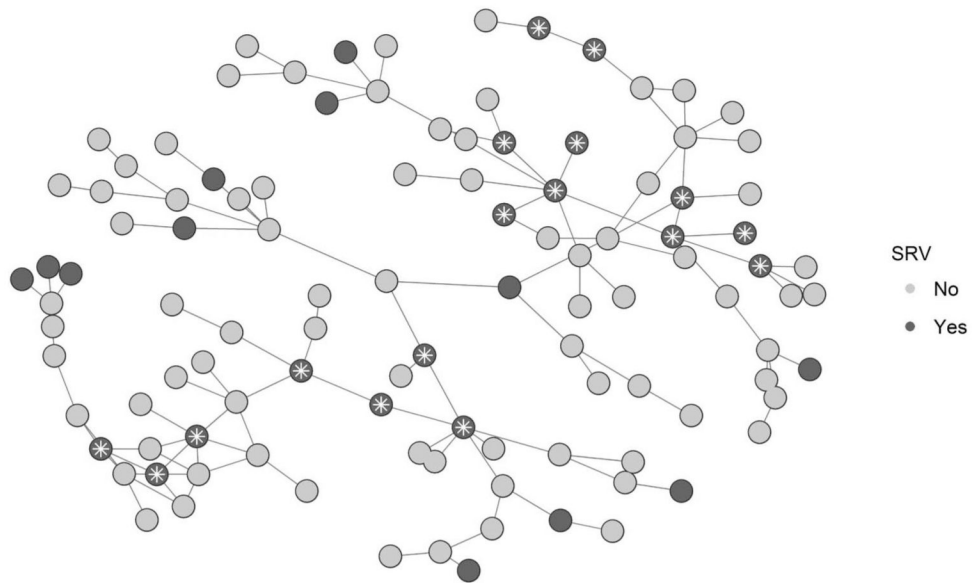


Figure 5. Visual demonstration of actual suicide-related verbalization (SRV) assortativity in a subsample of nodes from the first study period. Light circles represent nodes with no SRV-positive posts during the first study period; dark circles represent nodes with one or more SRV-positive posts during that period. Stars mark the SRV-positive nodes that are directly linked to at least one other SRV-positive node in the subsample. In this subsample, 59% of SRV-positive nodes are directly connected to another SRV-positive node; likewise, 54% of SRV-negative nodes are directly connected to at least one SRV-positive node.

Table 1

Example Posts With Application of Scoring Rules From Methods

User	Text	Sad word	Happy word	Suicide-related word	Sad post	SRV post	User sadness	User SRV score
A	"I'm so sad! Gonna kill myself	sad, kill		kill	1	1		
A	"I'm the worst lol!"	worst	lol,;	worst	1	0		
A	"My final day on Earth ..."			final	0	1	3	2
A	"Just got in a fight"	fight			1	0		
B	"It's a sad day"	sad			1	0		
B	"I love my life"		love, life		0	0	1	0

Note. "User" represents the ID of a hypothetical Twitter user, whose posts are given in the "Text" column. The "Sad word," "Happy word," and "Suicide-related word" columns identify which words (if any) from the user's post fall into the sad, happy, or suicide-related word categories described in the Method section and depicted in Figure 3. "Sad post" and "SRV post" represent dummy-coded variables representing whether the post would be classified as sad (scored 1 if the post included any sad word; 0 otherwise) or SRV-positive (scored 1 if the post included a suicide-related word but no happy words; 0 otherwise). "User sadness" and "User SRV score" columns represent the total number of sad and SRV-positive posts by each hypothetical user; these are the scores being analyzed in this investigation (i.e., Are SRV scores more assertative than chance? Do they remain so after controlling for sadness scores?). SRV = suicide-related verbalization.