Exploring Timelines of Confirmed Suicide Incidents through Social Media

University of Colorado Boulder

Xiaolei Huang, Linzi Xing, Jed R. Brubaker, Michael J. Paul

Departments of Information Science & Computer Science, University of Colorado Boulder

Abstract

Suicide is one of leading causes of death worldwide, yet little data is available about the lives of suicide victims because most people do not seek treatment[3]. Research has shown that people express suicidal ideation in social media[2], which can potentially be tapped to improve our understanding of the thoughts and behaviors of people prior to suicide. In this work, we introduce a novel dataset of Chinese social media accounts of 130 people who committed suicide between 2011&2016. We describe the demographic and geographic composition of the users, then conduct a longitudinal text analysis of their post histories, showing observable changes in content leading up to the time of death. With encouraging exploratory findings, we discuss directions for future research.

Background & Motivation

- **800,000 death[1]**: Over 800,000 people die due to suicide per year, which is underestimated.
- **20 attempts[1]**: For each adult who died of suicide, there may have been more than 20 others attempting suicide.
- **2nd leading cause of death[1,2]**: Suicide is the 2nd leading cause of death in 15-29-year-olds.
- **7% of treatment coverage[3]**: Chinese official statistics can't provide fine-grained information of suicide. Suicide is under-treated in China: only 7% of people who committed suicide had received health treatment.
- **300 million users[4,5]**: In 2016, Weibo has over 300 million monthly active users, 82% of whom are 15 to 29 years old. Weibo might be a supplementary source to learn about suicidal issues in China.

Dataset Overview

- 130 confirmed suicide accounts in Sina Weibo: 30 males, 100 females;
- 37610 followers and fans; 104,229 microblogs;
- 2 million interactions: 1,363,813 comments; 536,687 reposts; 462,410 likes

Figure 1. A widely-publicized example of suicidal microblog

2011	12
2012	39
2013	18
2014	22
2015	17
2016	22

Age/Gender	Male	Female	Overall
13-20	9	29	38
21-30	14	49	63
31-40	6	8	14
Unknown	1	14	15
Overall	30	100	130

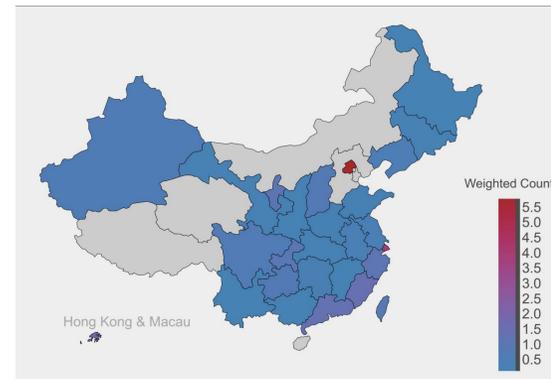**Table 1.** The number of users by Year (left), Demographics (Right)

Figure 2. Number of users (weighted count) in each location of death, normalized by the population of each location. Gray indicates no users in our dataset from that province. Hong Kong and Macau were counted as one province (bottom left).

Top 2 rankings: **Beijing** & **Shanghai**.

Daily Life Topic 26	睡觉 (sleep); 吃 (eat); 回家 (back home); 一天 (a day); 下雨 (raining); 工作 (work); 天气(weather); 洗澡 (shower); 起床 (get up); 学校(school)
Entertainment Topic 6	演员 (actor); 相声 (crosstalk); 曲艺 (Beijing Opera); 演出 (perform); 嘉宾 (guest on show); 单曲 (single song); 电影票 (movie ticket); 电影院 (movie theater)
News/opinions Topic 16	国家 (nation); 社会 (society); 人民 (the people); 政府(government); 老百姓 (civilian); 教育 (education); 文化 (culture); 法律 (law); 制度 (social system)
Family Topic 21	老婆 (wife); 老公 (husband); 爸爸 (Dad); 妈妈 (Mom); 弟弟 (Brother); 姐姐 (Sister); 家里 (home); 闺蜜 (bestie)
Sentiment (-) Topic 3	难过 (frustrated); 悲伤 (sad); 害怕 (fear); 孤独 (lonely); 寂寞 (lonely); 伤心 (sad); 哭 (cry)
Sentiment (+) Topic 18	哈哈 (laugh); 偷笑 (laugh); 好 (good); 可爱 (cute); 赞 (awesome); 给力 (awesome); 酷 (cool); 鼓掌 (applause)

Table 2. Examples of Topics from Each Category

Experiments & Approaches

- **Time Series Analysis**: We calculated statistics to summarize the activity levels and content of the users' social media feeds in time intervals with respect to the date of death--**10 months**, **10 weeks**, **10 days** prior to death.
- **Bootstrap resampling**: We performed **bootstrap resampling** to measure uncertainty in our estimates in each time period. We created 100 estimates of the statistics, where each estimate was calculated by randomly sampling with replacement from the 130 users.
- **Measurements**: **Three measurements** were utilized to learn suicide users.

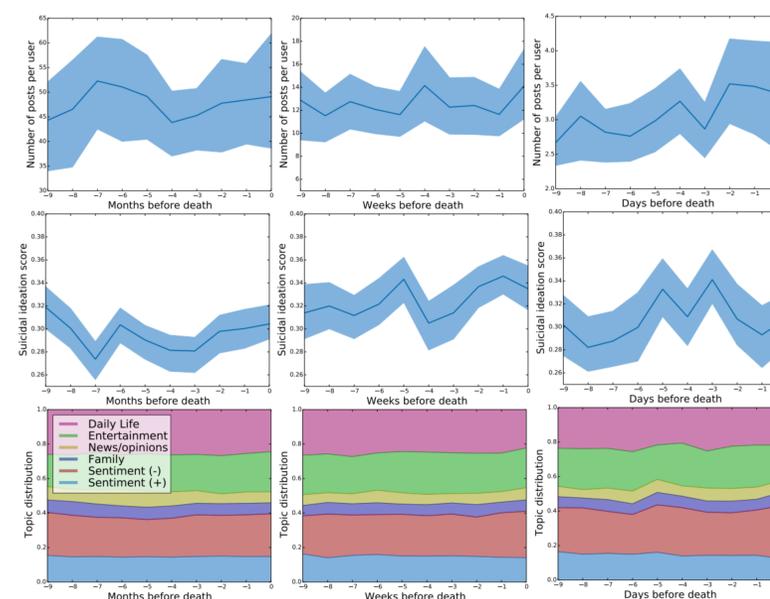| Measurements of Text (see Figure 3)		
Posting frequency (Count-based)	Linguistic indicators (Lexicon-based)	Topic composition (Topic Model-based)

Figure 3. The post frequency (top row), the linguistic suicidal ideation score(middle row) & the distribution of topic categories(bottom row) for 10 time intervals at three granularities: monthly(left), weekly(center), and daily(right).

Analysis & Discussion

- **Posting frequency**: An increase in posting frequency in the last week before death, with a fairly steady increase through the final days.
- **Linguistic indicators**: It shows an increase in levels of language suggestive of suicidal ideation in the weeks leading up to death. There is high variability in the scores in the daily plot, so it is less clear what the trend is in the final days, but the score reaches high points in the 3–5 days before death.
- **Topic composition**: The topic distribution is stable over time, suggesting that users largely share similar content leading up to their suicide. But, we observe an increase in negative sentiment topics close to the time of death, with a decrease in the entertainment and daily life topics.
- **Comparing Suicide Contexts**: **Broken relationship** VS. **Other Causes of Suicide**. (see Figure 3 and 4)
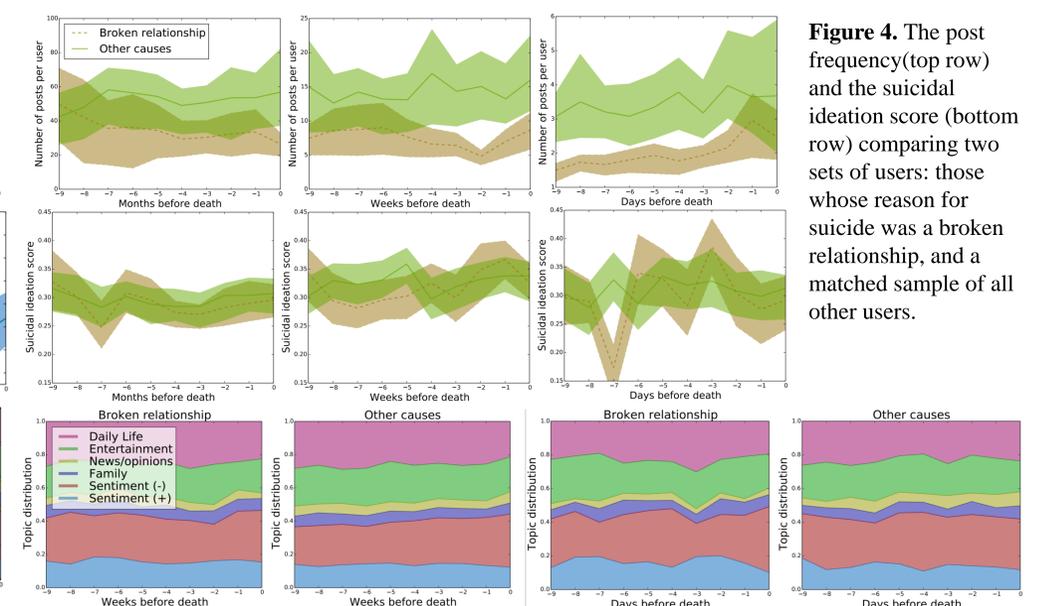

Figure 4. The post frequency(top row) and the suicidal ideation score (bottom row) comparing two sets of users: those whose reason for suicide was a broken relationship, and a matched sample of all other users.

Figure 5. The topic distribution by week (left) and day (right) for broken relationship users and other causes.

PRIVACY AND ETHICS

References

1. World-Health-Organization, "Preventing suicide: A global imperative," 2014. [Online]. Available: http://www.int/mental health/suicide-prevention/world report 2014/en/
2. Jared Jashinsky, Scott H Burton, Carl L Hanson, Josh West, Christophe Giraud-Carrier, Michael D Barnes, and Trenton Argyle. Tracking suicide risk factors through Twitter in the US. Crisis, 2014
3. S. Law and P. Liu, "Suicide in China: unique demographic patterns and relationship to depressive disorder," Current Psychiatry Reports, vol. 10, no. 1, pp. 80–86, 2008.
4. Weibo, "China statistical report on weibo user development," Jan 2017. [Online]. Available: http://data.weibo.com/report/reportDetail?id=346
5. J. Lin, "他在微博那头自杀，最后你却拉黑了他," Dec 2014. [Online]. Available: http://news.sina.com.cn/zl/zatan/2014-12-02/18032759.shtml