Qualitative analysis of adolescent's online risks on instagram

Sociology, Communication

Introduction

According to the most recent research from Pew Research Center " Fully 95% of teens have access to a smartphone, and 45% say they are online ' almost constantly'". As it is depicted in the Figure 01 from the Pew Research survey, about half of U. S. teens ages 13 to 17 say they use Facebook, considerably less than the shares who use YouTube, Instagram or Snapchat. This statistics about adolescent usage of internet and social media raise the concern for the parents and society of the risks and consequences that technology may bring to the developments of these youth. As the survey found, even the teen themselves do not have a clear consensus about the effect that social media has on their lives.

One of the important topics in adolescent interaction with technology that is receiving increasing attention from researchers is sexting. Sexting has been defined in different ways among researchers. Some researcher consider sexting definition as '' sexually explicit content communicated via text messages, smart phones, or visual and web 2. 0 activities such as social networking sites". Other researchers consider sexting as " the interpersonal exchange of self-produced sexualized texts and above all images (photos, videos) via cell phone or the internet". There needs to be consensus among researchers with a broader definition of sexting.

Some research indicates that adolescent sexting may have a correlation with other risky behaviors such as sexual risk-taking and substance abuse and alcohol consumption. Sexting may have legal, social, and health

consequences. One of the worst scenario is when an adolescent’s sexually explicit photo spreads on the internet or among peers without consent which could be used for cyberbullying or harassment. From a legal stand point, adolescent sexting may be considered under child pornography.

Socially, research has shown that sexting can related to emotional and psychological conditions, such as anxiety disorders to the extent of depression and contemplating suicide. Out of 132 articles in the adolescent online safety research area, 83% of teen-focused studies asked teens to quantify their own online risk experiences and behaviors via self-reported surveys. Small number of studies involving teens used other techniques such as interviews (9%), focus groups (9%), or analysis of publicly available data (2%). Only 6% of the studies used multi-method approaches. Previous findings suggest that most of the studies relied on self-reporting measures of sexting behavior, but the extent in which these measures reflect actual behavior is unknown. Although usually in these studies participants were assured that their answers would remain anonymous, it might be possible that teenagers do not want to disclose this information and provide socially desirable answers.

At this time, it does not appear that any studies have been conducted utilizing real-world data of adolescent communications on Instagram to analyze the actual sexting behaviors of adolescents. Thus our objective is to qualitatively analyze real-word Instagram data, which is used by more than 70% of adolescent in the U. S. Preliminary research questionsIn the present study, we direct our attention toward a better understanding of the content

that adolescents exchange on Instagram. The primary goals of this research are to:

1. develop a better understanding of how adolescents engage with others online by analyzing the information that they are sharing on Instagram.
2. have a better understanding of adolescent sexting and risky behaviors utilizing real-world Instagram data.
3. To qualitatively label Instagram conversations so they can be used as a ground truth for training machine learning algorithms.

The associated research questions that we are going to examine include:

RQ1 – How can we qualitatively code a large dataset?

RQ2 – What types of information do adolescents share on Instagram? What are the different categories?

RQ3 – How can we identify and categorize sexting? What other risky behaviors do teens conduct on Instagram?

RQ4 – How can we determine if the conversations are between strangers or friends? Are there any grooming (predatory) conversations in the dataset?

RQ5 – Can metadata such as temporal information helps in detecting sexting?

RQ6 – Does the teen initiate the sexting, or do they just receive sextual content?

How does the teen reply when someone sends a sexual content to them?

RQ7 – How does the sexting experiences differentiate between genders?

RQ8 – How can we analyze the public, semi-public, and private social media interactions of adolescents?

What trends are different in private conversations vs public posts and comments? The research contributions of this work are dataset and empirical. This research will help the research community to build a qualitatively labeled dataset and continue working on the problem with applying machine learning algorithms. The quality in which a dataset is labeled is very important, but there has not been much research in this area and usually machine learning engineers tend to find a labeled dataset and use it for training their machine learning algorithms.

In recent years, there has been increased attention from researchers in the field of adolescent online safety, although our research will contribute to a better understanding of the nature of risky behaviors such as sexting among adolescents in a real-world dataset. Literature reviewWe will conduct a structured literature review, which will consist of literature search, relevancy coding, and article content coding. The research will be done mostly using the peer-reviewed journals and conferences which include the words " adolescent" or " teen" in combination of one or some of the following key words: " Sexting", " Online Safety", " Online risk", " Qualitative analysis", " Social Media", " Instagram", " Online predator", " Machine learning".

The organization of the topics that will be in our literature review will consist of two main parts. First, papers that used a qualitative analysis method in social media content analysis. Second, research papers that have used machine learning methods in their social media content analysis. The first group of papers will be prioritize based on their relativity to the subject, so research papers that conducted qualitative analysis in sexting or risky behaviors in social media (Instagram) content will be considered first. MethodsWe plan to conduct a qualitative content analysis to address the research questions, and to gain insight into how and why adolescents engage in risky behaviors using a real-world data. It has been shown that qualitative techniques are suitable to iteratively identify patterns in unstructured data, and help researchers to answer “ how”, “ what”, and “ why” for a complex, unexplored, and socially-based phenomena that cannot be quantified [5]. Data CollectionThe Instagram data used in this study are provided from a third-party company who developed a mobile parental monitoring app named RAKKOON [15].

RAKKOON uses teen’s social media activities and machine learning algorithms to flag “ risky” text and images and then notifies parents. The dataset contains Instagram data about 140 teens including approximately 1. 6 million images, 1. 3 million posts, 217 thousand private messages, and 4. 8 thousand conversations. The dataset also contains descriptive metadata such as time and date, number of followers, captions and comments for each post. Approximately 400 randomly selected conversations will be used to identifying a classification scheme to develop a codebook of the content of

all the conversations. Classifications will cover the factors to answer the research questions. If sexting occurred in the conversation, then the type of sexting will be coded. This structured codebook will be created using a joint inductive and deductive approaches. The codebook will be refined based on themes in the data. Therefore, the coding frame will have both concepts-driven (defined in advance) and data-driven (derived from data during coding). Timeline for each weekThe following table demonstrates a research plan for conducting the mentioned research by the end of the semester.