Week 1

Introduction

This lecture provides an overview of online social media and its impact on society. The lecturer PK discusses key concepts like privacy, security, social networks, social media data, and case studies.

What is Privacy and Security?

- Privacy is the state of not being watched or tracked. It involves keeping information secret and sharing it selectively.
- Security refers to protecting information - confidentiality, integrity and availability. It involves securely transferring information from one place to another without tampering.

Social Media Concepts

- Six degrees of separation - The idea that anyone on Earth can be connected to anyone else through about 6 people on average. On social media like Facebook, this number reduces to around 3.5 because of increased connectivity.
- Strength of weak ties - Weak social ties or acquaintances (people you barely know) expose you to more new opportunities and information than strong ties like close friends.
- Social media includes more than just social networks like games and video calls which allow interactions.
- Key properties of social media data are velocity, variety, volume, veracity (truthfulness) and value. The pace and amount of data generated has vastly increased.

Growth and Impact

- People now spend more time checking social media than general news and information. Social interactions have become a big priority.
- Social media allows predicting real-world events to some extent by analyzing data, though why and how people post can't be determined. Some examples are predicting protests, stock prices, health outbreaks.
- It has been used in crisis response, mass movements like Arab Spring, recruitment by malicious groups, spreading fake news etc.

Key Takeaways

- Social media usage has exploded globally. Our online activities generate huge amounts of data about us.
- This data can be used to manipulate behavior and decisions on a mass scale, threatening privacy.
- Studying social media can help build solutions for privacy, security, misinformation, crisis response etc. Students are encouraged to explore projects and ideas beyond the coursework.

Summary of Video on Online Social Media

What is Social Media?

Social media refers to various online platforms and applications that allow users to create and share content or participate in social networking. Some popular social media platforms include Facebook, Twitter, YouTube, Instagram, Snapchat, WhatsApp etc.

The instructor explains that social media allows generation of different kinds of content like text, images, videos etc. He also talks about the massive amount of data generated on social media platforms - for example, Facebook gets 3.5 million posts per minute and YouTube gets 400 hours of video uploaded per minute.

Different Categories of Social Networks

The instructor discusses different categories of social networks:

Traditional Social Networks

These include platforms like Facebook, Twitter, LinkedIn etc. that allow users to connect with friends/followers and share diverse information.

- Facebook: Allows users to share text, photos, videos; like, comment and share posts; connected as friends. Stores data as graphs.
- Twitter: Microblogging platform to share short text updates. Allows users to tweet, retweet, like, reply and mention others.
- LinkedIn: Professional networking platform focused on career connections.

Location Based Networks

These networks focus on check-ins at locations and providing location-specific information e.g. Foursquare.

Ephemeral Networks

These social networks emphasize short-term sharing of content that expires quickly e.g. Snapchat, Whisper.

Anonymous Networks

These platforms allow anonymous posting of content e.g. Whisper.

Building Blocks of Social Networks

The instructor discusses the core building blocks of some popular social platforms:

- `Facebook:` Posts, friends, pages, groups, likes, comments, shares
- `Twitter:` Tweets, followers, likes, retweets, replies, mentions, hashtags
- `Foursquare:` Check-ins, tips
- `YouTube:` Video uploads, likes, comments, subscribing to channels
- `Pinterest:` Images, boards
- `Tinder:` Swiping left/right to connect with others nearby

Online Social Media - An Introduction

What is Online Social Media?

- Online social media refers to internet-based platforms that allow people to interact, share content, collaborate, and build communities.
- Some key examples are Facebook, Twitter, LinkedIn, Google+, WhatsApp, Instagram etc.

Scale and Impact of Social Media

- Social media usage has grown exponentially in recent years.
- As per 2021 statistics, there are 4.2 billion social media users globally. This is over half the world's population.
- On average, a person spends 2 hours and 25 minutes on social media per day.
- In one minute on the internet in 2021 - Facebook users shared 500,000 comments and 293,000 statuses; Instagram users shared 347,222 stories; Twitter users tweeted 511,200 tweets.
- This demonstrates the speed, volume and variety of content shared on social media.

Positive and Negative Effects of Social Media

- Social media has been used for many productive causes like reuniting lost children with parents, spreading awareness about social issues, coordinating disaster relief etc.
- However, there are downsides like cyberbullying, spread of misinformation, compromised accounts leading to security issues etc.

- It is important to be aware of both the positive and negative effects of social media.

Understanding Social Media Identifiers

- Social media platforms collect a lot of data about users based on how they interact on the platform.
- They use this for targeted advertising, behavior analysis etc.
- So it's important to understand what data social media collects about you through your profile and activity.
- Aspects like privacy and security need more awareness.

Week 2

Online Social Media APIs and Tools for Data Collection

Introduction

This video discusses using online social media APIs and tools to collect data for analysis. It covers the popular social media platforms like Facebook, Twitter, YouTube etc. and the data formats used by them.

What is Social Media

- Social media refers to various social networking sites like Facebook, Twitter, Instagram etc. where users interact and share content online.
- There are different types of social media content like text posts, images, videos etc.
- Some examples of classical social media services are Facebook, Twitter, YouTube. There are also ephemeral social networks like Snapchat and anonymous social networks like Whisper.

Key Characteristics of Social Media

Social media is characterized by:

- Volume - Large amount of data is generated every minute. For example, 400 hours of video are uploaded to YouTube every minute.
- Velocity - Data is generated and spread very fast on social media.
- Variety - Many different types of data like text, images, video etc. are shared.
- Veracity - Issues with fake content on social media.
- Value - Data can provide useful insights.

Role of Social Media

- Social media has played important roles in real world events like Arab Spring protests, disaster response etc.

- It has also caused issues like spread of misinformation, job losses etc.

APIs for Accessing Social Media Data

- API or Application Programming Interface allows collecting data programmatically from social media sites.
- Popular social media platforms like Facebook, Twitter provide APIs for data access.
- APIs return data in JSON format which contains information like user IDs, names etc.

Programming Languages and Databases

- Python is a popular language used to write programs to collect social media data using APIs.
- MySQL database can be used to store the collected data and query it.
- MongoDB is another database used to store social media data.

Visualization Tools

- Tools like phpMyAdmin and Robomongo help visualize the collected data.

Conclusion

- Social media provides APIs to collect data which can then be stored and analyzed using programming languages like Python and databases like MySQL, MongoDB.
- Visualization tools help explore the collected data.

Week 3

Using Social Media APIs to Detect Misinformation

Introduction

The video discusses how we can use APIs and data from social media platforms like Twitter and Facebook to detect misinformation and build models for determining post trustworthiness.

Recap of Concepts from Previous Weeks

- Learned about APIs and how to access data from Twitter and Facebook
- Looked at data storage with MongoDB and visualization with phpMyAdmin
- Discussed issues around trust and credibility on social media
- Saw examples of misinformation spreading on platforms
- Learned how to analyze tweets (who, what, when, where, why, how)

Features Available in Tweets

- User features - details about the user posting the tweet
- Tweet features - details about the tweet itself
- Can categorize features to decide if a tweet is real or fake

Using Web of Trust (WOT)

- WOT is a service that rates domains for trustworthiness
- Provides a reputation score that can indicate if a post contains misinformation
- Useful feature for analyzing Facebook posts

Facebook Inspector Plugin

- Browser extension that analyzes Facebook posts
- Shows annotations about possible misinformation
- Flags posts as potentially false or malicious

Applying Twitter Approaches to Facebook

- Similar architecture can be used across platforms
- Extract features, build model to classify posts, provide API for predictions
- Customized for the different types of data available

Conclusion

Using social media APIs and data, we can gain insights into misinformation and build models that help determine post credibility and trustworthiness.

Privacy and Social Media in India - Study Findings

Overview of the Study

This video discusses findings from a large-scale privacy study conducted in India. The study collected survey data from over 10,000 people across India to understand citizens' privacy preferences and attitudes towards online social networks.

Some key details about the study:

- Collected 10,427 survey responses with 83 questions from all states except one
- One of the largest global studies on privacy perceptions
- Survey data collected both online and in-person at malls, hotels, railway stations etc.

Key Privacy Attitude Segments

The study categorized Indian citizens into three privacy attitude segments based on their willingness to share personal information:

Fundamentalists

- 25% of respondents
- Very concerned about privacy, will not provide personal details easily

Pragmatists

- Make situational decisions based on context
- Willing to share more for higher value

Unconcerned

- 27% of respondents
- Willing to share personal info easily

Select Study Findings

Here are some key findings from the study questions related to social media:

- 42% believe their privacy settings protect their data even though that may not be true
- 19% express concern but still share personal info
- 27% accepted friendship requests from opposite gender, only 10% for nice profile pic

The full survey report is available for free and provides more detailed insights from the study data.

Week 4

Week 5

Police Use of Social Media in India

Social media services like Facebook and Twitter are being widely adopted by police organizations in India. Police departments are using these platforms to communicate with citizens and share information. However, there are some key issues around privacy, data collection, and fake accounts that need more attention.

Widespread Adoption by Police

In the last few years, many state and city police departments in India have created accounts on Facebook and Twitter. Some popular handles include Bangalore City Police, Delhi Traffic Police, and Hyderabad City Police. Even small towns and districts have a presence.

Police organizations use these platforms to post about traffic updates, request help from citizens, showcase their work and officers, and have conversations with the public. Many handles have hundreds of thousands of followers, showing their popularity.

Data Collection from Social Media

Researchers have shown that data from location-based services like Foursquare can be used to infer a person's home location. Similarly, data from police handles on Facebook and Twitter can be collected and analyzed.

Metrics like number of posts, comments, likes, follower growth, types of content etc. can give insights into how active and engaging these handles are. As social media and data analytics skills grow, more advanced analysis will be possible.

Issues Around Fake Accounts

There is a problem of fake accounts impersonating legitimate police handles. It is difficult for citizens to verify which handles are genuine. Even profiles with no name changes can be fake. Getting the blue 'Verified' check is one way for police organizations to establish authenticity.

Another issue is that some handles may be operated not by police but random users. So the source of information is unclear. Tech savvy youth need to pay more attention to identifying such fake accounts.

Balancing Privacy and Benefits

While social media adoption by police has many benefits, privacy issues around data collection and surveillance need more discussion. Citizens may not even be aware of what data can be inferred from social media. Tech policymakers need to find the right balance here.

Overview of Using Social Media for Policing

The video discusses how police organizations can use social media like Facebook and Twitter to gather information for crime prevention and get feedback from citizens.

Gathering Actionable Information

- Police can't be present everywhere in society at all times. So social media provides a way to gather a lot of public information about what's happening in society. This can be useful for crime prevention.
- Posts with specific details like time, location, events etc. can provide "actionable information" that police can act on. For example, a post saying there is a pothole at a certain intersection that caused a car accident provides actionable info.
- Posts can also provide information indirectly, like someone posting that their friend in a certain area is facing issues.

Analyzing Citizen Sentiment

- Looking at the types of posts citizens make and the comments they leave can provide insights into public perception of the police.
- Analysis found more posts praising or thanking the police versus complaining. But comments expressing dissatisfaction outnumbered thanks.
- Positive stories of police success had significantly more likes than other types of posts.

Differences in Communication Style

- Communication from police is mostly formal, while citizens often use informal language.
- This shows social media can help police connect strongly with society. But both sides have a responsibility in how they communicate.

Understanding Citizen Needs

- Analysis of posts can provide insights into citizen concerns, desires, and expectations of police. This can help police address issues better.
- Citizens also have a responsibility to provide useful information, have discussions with police, and participate in keeping the city safe.

Potential Uses

- Police organizations can use analysis of social media posts to be more responsive and improve engagement with citizens.
- They can provide examples of ideal responses to certain types of posts. This can increase police productivity in responding.

Analysis of Police-Citizen Interactions on Social Media

Introduction

The video discusses analysis of police-citizen interactions on social media platforms like Facebook. It aims to understand the nature of content and engagement between police departments and citizens.

Research Questions

The research tries to answer questions around:

- Contextual characteristics: What topics are discussed in social media threads between police and citizens?
- Engagement characteristics: How are citizens and police engaged in social media discussion threads?
- Emotional exchange: What is the nature of emotions and affective expression exhibited in social media?
- Cognitive and social orientation: What are the linguistic features that characterize cognitive and social response processes in police Facebook pages?

Key Insights

- Police posts focus on topics like rules, safety, violations while citizens discuss diverse issues like media articles, tips, complaints etc.
- Engagement is higher when police comment on citizen posts versus when citizens comment on police posts.
- Negative affect is higher in citizen threads versus police-citizen threads indicating citizens are expressing opinions strongly.
- Concern levels are lower when police engage with citizens in threads. Police engagement provides some reassurance.

Conclusion

Analysis of social media data created by police-citizen interactions can help improve community policing and sensing. It also helps understand behavioral attributes like engagement, emotions, social support. Technologies can be built to aid such interactions and predictive analysis. This can help police organizations make better decisions and provide citizens a safer life.

Week 6

Introduction

The professor provides an overview of various types of cybercrime that happen on online social media platforms. He covers phishing, fake customer service accounts, fake comments on popular posts, fake live streaming videos, fake online discounts, fake online surveys, clickbaiting, hashtag hijacking, account hijacking, impostor accounts, work from home scams, and more.

Phishing

Phishing involves sending emails that impersonate legitimate companies and trick people into clicking malicious links that steal login credentials. Phishing has evolved on social media through fake notifications and alerts that lure users to fake login pages.

Fake Customer Service Accounts

Scammers create fake customer service accounts on social media that impersonate real companies. They engage with customers having issues to obtain personal information.

Fake Comments on Popular Posts

Scammers post fake comments on trending and viral posts pretending to be real users. This is done to spread malicious links through the comments.

Fake Live Streaming Videos

Fake live streams of popular events like sports games are posted to lure viewers to malicious websites. The links don't contain real videos.

Fake Online Discounts

Scammers create fake social media posts offering discounts and deals impersonating real businesses. This is done to promote products and services.

Fake Online Surveys

Fake online surveys are shared offering rewards for participation. These harvest personal information which may be misused.

Clickbaiting

Clickbait posts entice users to click links unrelated to the content which often route to malicious websites.

Hashtag Hijacking

Hashtags unrelated to the content are used to increase reach and visibility. For example, using trending hashtags to promote products and services.

Account Hijacking

Compromised accounts are used to spread fake information and scams. This is done by stealing login credentials.

Impostor Accounts

Fake social media accounts impersonating real people are created using publicly available information. These are used to spread scams and harvest data.

Work From Home Scams

Fake opportunities to earn money from home are promoted through posts and ads. The links provided route to malicious websites.

Week 7

Online Social Media Link Farming

Background

Link farming refers to the practice of artificially creating inbound links to manipulate search engine rankings. Link farmers use social media platforms like Twitter to promote their own content.

- They create fake accounts that follow each other and retweet promoted content.
- This artificially inflates engagement metrics, increasing search rankings.

Researchers analyzed data from ~200k suspected link farming accounts on Twitter. They examined account features like follower/following ratio, reciprocal in-links etc.

Key Findings

- Top link farmers have very high in-degree and out-degree compared to spammers and random users.
- Top link farmers have ~1 follower to following ratio, similar to genuine popular accounts.
- Link farmers discuss internet marketing, social media, money etc. indicating commercial intent.
- Genuine popular accounts like celebrities also participate in link farming.
- Link farming effectively increases social capital and influence by inflating engagement metrics.

Takeaways

- Link farmers have distinct account patterns like high degree, reciprocity etc.
- Genuine and fake accounts both engage in link farming on social networks.
- Link farming can artificially inflate social standing and should be detected.

Using Technology to Help Users Make Better Decisions

Introduction

The video discusses how technology can be used to help users make more informed decisions when posting content online. It focuses on an MIT study that tested different "nudges" to get users to pause and reconsider their social media posts.

The Problem: Users Don't Read Privacy Policies

- Research shows most people don't actually read privacy policies before agreeing to them.
- One study estimated it would cost $781 billion annually in lost time if every US citizen read the privacy policy of each website they visited monthly.
- This highlights the need for tools to help users make informed decisions about their data.

MIT Study: Testing Different "Nudges"

- Researchers created three different "nudges" as Chrome browser extensions:
    - `Photo Nudge:` Shows profile pictures of who will see your post.
    - `Timer Nudge:` Gives you 10 seconds before posting.
    - `Sentiment Nudge:` Analyzes emotional sentiment of your post.

- They tested these with 21 participants using surveys and interviews.
- They measured privacy setting changes, deleted/edited posts, posting frequency and content sensitivity.

Study Results: Nudges Changed User Behavior

- One user narrowed their audience from "friends of friends" to just friends after seeing the wider reach.
- Another user deleted some posts because the profile photos made them reconsider sharing so widely.
- The timer nudge surprised users and made them reconsider spur of the moment posts.
- Sentiment analysis was flawed at deciphering emotions accurately.
- Many deleted or edited posts to avoid negative backlash.
- Posting frequency decreased from 13 to 7 posts per day on average.

Key Takeaways

- The nudges helped users make more informed sharing decisions, especially regarding negative posts.
- More research is needed on what works best in different contexts.
- Creating technologies that encourage mindful sharing is an important area of research.

Social Engineering through Phishing Attacks

What is a Semantic Attack?

A semantic attack is when a system is targeted by manipulating how it understands meaning and concepts. These attacks exploit the difference between how a system models what users are doing versus what users think the system is doing. This gap is called the semantic gap. Larger semantic gaps make issues harder to correct.

Phishing is a Type of Semantic Attack

Phishing involves sending emails that appear legitimate to trick users into revealing personal information. Phishing emails often include urgent calls to action, links to fake websites, and threatening messages. Users frequently fall victim due to the sense of urgency.

Case Study of Social Phishing Attack

Researchers conducted a study in 2005 at an Indian university where they collected public social media data of students aged 18-24. They crafted phishing emails appearing to come from friends which contained links to a fake university website. When users clicked the link and entered their credentials, their information was collected.

The study found:

- 72% of participants in the social phishing experiment entered their credentials compared to 16% in the control group. This shows the effectiveness of social phishing.

- Women were more frequently phished than men. Emails from the opposite gender were more successful.
- Younger students like freshmen were more susceptible compared to seniors.
- Science majors showed the biggest difference between social and control conditions.

Takeaways

- Most users don't realize how much public information can be used against them in phishing attacks.
- Phishing is very effective, especially when emails appear to come from the opposite gender.
- Education efforts, browser solutions, rapid takedowns, and digital signatures can help combat phishing.
- Social media sites should limit personal information sharing.

Week 8

Social Engineering through Phishing Attacks

What is a Semantic Attack?

A semantic attack is when a system is targeted by manipulating how it understands meaning and concepts. These attacks exploit the difference between how a system models what users are doing versus what users think the system is doing. This gap is called the semantic gap. Larger semantic gaps make issues harder to correct.

Phishing is a Type of Semantic Attack

Phishing involves sending emails that appear legitimate to trick users into revealing personal information. Phishing emails often include urgent calls to action, links to fake websites, and threatening messages. Users frequently fall victim due to the sense of urgency.

Case Study of Social Phishing Attack

Researchers conducted a study in 2005 at an Indian university where they collected public social media data of students aged 18-24. They crafted phishing emails appearing to come from friends which contained links to a fake university website. When users clicked the link and entered their credentials, their information was collected.

The study found:

- 72% of participants in the social phishing experiment entered their credentials compared to 16% in the control group. This shows the effectiveness of social phishing.

- Women were more frequently phished than men. Emails from the opposite gender were more successful.
- Younger students like freshmen were more susceptible compared to seniors.
- Science majors showed the biggest difference between social and control conditions.

Takeaways

- Most users don't realize how much public information can be used against them in phishing attacks.
- Phishing is very effective, especially when emails appear to come from the opposite gender.
- Education efforts, browser solutions, rapid takedowns, and digital signatures can help combat phishing.
- Social media sites should limit personal information sharing.

Week 9

Online Social Media Privacy Issues Around Location Sharing

Introduction

The video discusses privacy issues related to location sharing in online social media platforms like Foursquare, Yelp, Facebook etc.

Key Points

- Location-based services allow users to check-in and share their location. This raises privacy concerns as others can see where a user is.
- Foursquare encourages check-ins by providing rewards like badges, mayorships and tips. This gamifies the platform.
- Check-ins are visible to friends but tips, mayorships and user lists are public. So home location of a user can potentially be derived from public data.
- Research shows that for 75% of Foursquare users, their home city can be determined within 75 hours of analyzing check-in data.
- Many don't consider regional language notifications from service providers during travel as a privacy breach. But it does reveal location.
- Location data combined with tips and tweets can reveal if a user is away from home. This data was earlier exploited by the website 'Please Rob Me' to highlight privacy issues.

- Researchers have used Foursquare data for urban mobility analysis and understanding usage of location-based platforms.

Video Summary: Analyzing check-in data to infer home locations

Introduction

The video discusses a research paper that analyzed check-in data from Foursquare to infer users' home locations. Foursquare is a location-based social network where users can check in at venues they visit.

Data Collection

- The dataset contains check-in data for 13 million Foursquare users, comprising 10 million tips and 15 million check-ins globally
- Multiple tools like reverse geocoding were used to map check-in locations and extract geographic coordinates

Data Cleaning

- Non-geographic and ambiguous check-in locations were removed
- Only check-ins with valid geographic information were retained for analysis

Analysis

- The distribution of tips, check-ins and mayorships per user follows a power law distribution - a small percentage of users contribute a large percentage of content
- Check-ins, tips and mayorships are concentrated in a few cities like New York, Jakarta and Sao Paulo
- The time interval between consecutive tips or check-ins by a user shows a steep distribution for short intervals
- Almost 50% of users posted consecutive tips within 1 hour, indicating high frequency of content generation
- The distribution of displacement (distance) between consecutive check-ins shows most users have small displacements around 0-150 kms
- The distribution of time intervals between check-ins at a location shows a clear 24 hour periodicity, indicating daily patterns

Inferring Home Locations

- Users were classified into 3 classes based on activity patterns:
    - Class 0: Single activity - Only check-ins/tips/mayorships
    - Class 1: Multiple activities but one predominant location
    - Class 2: Multiple activities but no single predominant location
- For Class 1 users, home city can be inferred with 67% accuracy using mayorships

- Comparison between inferred and declared home cities shows 78% are within 50 kms of actual location

Conclusion

- The analysis shows check-in data can be used to infer home locations of a large fraction of users with good accuracy
- However, the resolution is limited to city-level, and home country is harder to infer accurately

Week 10

Inferring Home Location in Social Networks

Introduction

The video discusses a research paper that estimates users' home locations using data from social media networks like Foursquare, Google Plus, and Twitter. The goal is to analyze the geographic information leaked from these networks and how accurately a user's city and residence can be inferred.

Data Collection

- Data was collected from Foursquare, Google Plus, and Twitter between 2011-2012
- Foursquare: 15 million check-ins, 11 million tips, 10 million likes
- Google Plus: 27 million profile pages crawled, 7 million with at least one location, 5000 address details, 7 million with education info, 6 million with employment info
- Twitter: 120 million geo-tagged tweets from 20 million users collected via API

Data Analysis

- Users grouped into 3 classes based on location data availability:
    - Class 0: Only one unambiguous home city option
    - Class 1: One main location plus others
    - Class 2: Multiple ambiguous locations
- Models built for each network using features like check-ins, tips, friends etc.
- Precision measured as % of users in Class 0 & 1 correctly identified
- Results: Precision of 67%, 72% and 82% for home city for Foursquare, Google Plus and Twitter respectively
- Twitter had highest precision due to tweets being explicitly geo-tagged

Estimating Residence Location

- Similar models built for estimating user's residence within their home city
- Twitter model identified residence within 20km for 73% of users

- Foursquare model had 52% precision for <5km radius, 77% for <20km
- Google Plus had only 5% residence precision due to lack of check-in type features

Understanding Username Change Behavior on Twitter

A recent paper titled "On the dynamics of username change behavior on Twitter" analyzes how and why Twitter users change their usernames. The paper examines data from 7.7 million Twitter users over a 2-month period.

Key Findings

- 73.21% of the 8.7 million Twitter users changed profile features and assigned new values, while around 10% changed their usernames at least once.
- 20% of users change usernames frequently and repetitively over short intervals, while 80% of users rarely change at all, following a Pareto distribution.
- One user changed their username 113 times over 14 months through repetitive, organic posts of half-complete tweets, tweets with the same text, and frequent short-term posts. This suggests an inorganic, automated user rather than a human user.
- Within one day of a previous username change, around 20% of usernames started changing again.
- Usernames changing behavior follows the Pareto principle, with 10% of usernames changing after one hour of a previous username change.
- 65% of users chose a new username unrelated to the old one, while 35% reused the old name.
- Usernames with 11 or fewer characters tend to add letters in new usernames, while most old usernames with over 11 characters opt to remove letters from their new usernames, indicating a desire for space savings.

Reasons for Username Changes

The paper also examined some reasons users gave for changing usernames, including:

- Gaining space in posts
- Matching trending events like sports
- Obscuring identity
- Adapting to real-life events
- Fighting boredom
- Avoiding abusive impersonation
- Religious reasons

Relationship Between Popularity and Username Changes

The study found a weak but positive correlation between username change frequency and number of followers, suggesting username changes increase slightly with popularity. However, the weak correlation does not imply causation.

Overall, this paper provides interesting analysis into the dynamics of username changes on Twitter. Expanding the dataset could potentially yield further insights.

Week 11

Online Social Media Analysis Course Summary

Introduction

This is a summary of an online social media analysis course. The instructor provides an overview of the topics covered throughout the 11-week course.

Week 1: Introduction to Online Social Media

- Social media generates massive amounts of data at high velocity
- Main concerns with social media: variety, volume, velocity, veracity, value
- Social media plays role in crisis management and spreading misinformation

Week 2: Collecting Data and Building Blocks

- Learned how to collect data from APIs like Facebook and Twitter
- Looked at basics of machine learning like feature extraction and modeling
- Covered programming languages like Python for collecting and analyzing data

Week 3: Trust and Credibility

- Analyzed features to determine tweet credibility
- Collected Facebook data and factored credibility of links
- Looked at concepts like network analysis and data visualization

Week 4: Privacy

- Discussed how users can be identified from public social media data
- Talked about reading privacy policies and economics of privacy
- Introduced nudges to make users more privacy aware

Week 5: Law Enforcement

- Analyzed interactions between citizens and police on social media
- Discussed how police can use posts to aid decisions and operations
- Covered linguistic analysis of posts and response patterns

Week 6: Cybercrime

- Learned about different cybercrime techniques like phishing and scams
- Discussed hashtag hijacking and how it spreads unrelated content
- Analyzed link farming and how spammers manipulate search rankings

Week 7: More on Privacy and Identity

- Talked about costs of reading privacy policies
- Introduced semantic attacks using phishing case study
- Covered identity resolution across social media accounts

Week 8: Anonymous Social Networks

- Analyzed anonymity features of networks like Whisper
- Saw how anonymity changes user behavior and content sharing
- Compared community structure to non-anonymous networks

Week 9-11: Research Papers

- Read and analyzed real research papers on social media
- Learned how researchers frame problems and present analysis
- Covered a variety of topics like privacy, identity, and anonymity

Conclusion

The course provided an introduction to social media analysis concepts and hands-on data analysis skills.