2026

BSCCS 405

Privacy and Security in Social Media - II

Dr. Babasaheb Ambedkar Open University

Course Writer, Editor & Reviewer

Prof. (Dr.) Nilesh Modi
Professor and Director,
School of Computer Science,
Dr. Babasaheb Ambedkar Open University,
Ahmedabad, Gujarat

ISBN:

Printed and published by: Dr. Babasaheb Ambedkar Open University, Ahmedabad. While all efforts have been made by editors to check accuracy of the content, the representation of facts, principles, descriptions and methods are that of the respective module writers. Views expressed in the publication are that of the authors, and do not necessarily reflect the views of Dr. Babasaheb Ambedkar Open University. All products and services mentioned are owned by their respective copyright's holders, and mere presentation in the publication does not mean endorsement by Dr. Babasaheb Ambedkar Open University. Every effort has been made to acknowledge and attribute all sources of information used in preparation of this learning material. Readers are requested to kindly notify missing attribution, if any.

Source acknowledgement: The units in this paper have been adapted and rewritten from the NPTEL course "Privacy and Security in Online Social Media" by Prof. Ponnurangam Kumaraguru, Department of Computer Science and Engineering, Indian Institute of Technology, Madras, and from the associated published research literature, for the purpose of self-instructional learning material.

BSCCS 405

Privacy and Security in Social Media - II

Block-1: FUNDAMENTALS OF ONLINE SOCIAL MEDIA

UNIT-1
Introduction to the Course

UNIT-2
Policing and Online Social Media

UNIT-3
eCrime on Online Social Media

Block-2: PRIVACY, POLICING AND ECRIME ON SOCIAL MEDIA

UNIT-4
Link Farming in Online Social Media

UNIT-5
Privacy in Location-Based Social Networks

UNIT-6
Inferring Home Location in Social Networks

UNIT- 1
INTRODUCTION TO THE COURSE

1.1 OBJECTIVES

After successful completion of this unit, you will be able to-

- Understand the aim, scope and learning outcomes of the course Privacy and Security in Social Media – II.
- Identify the major themes covered across the units of this course, including policing and eCrime on social media.
- Appreciate the methodology used in this course for collecting, analysing and visualising social media data.
- Understand the significance of identity resolution across multiple online social media platforms.

1.2 INTRODUCTION

Online social media platforms such as Facebook, Twitter (X), Instagram and LinkedIn have become deeply embedded in everyday personal, professional and civic life. As billions of users share text, images, location and behavioural data on these platforms every day, questions of privacy, security, trust and safety have moved from being purely technical concerns to matters of significant societal importance. This course, Privacy and Security in Social Media – II, builds upon the foundational concepts introduced in Privacy and Security in Social Media – I and examines, in greater depth, how social media is used, misused, secured and governed in practice.

The course draws upon research and practical case studies from the domain of computational social science, examining online social media not merely as communication platforms but as rich sources of behavioural data that can be collected, analysed and visualised to answer meaningful questions about human behaviour, institutional behaviour and criminal activity. Two themes anchor this course: how public institutions, particularly the police, use social media to engage with citizens and manage crises; and how criminals and fraudsters exploit the same platforms to commit electronic crime (eCrime).

The material in this course is grounded in the lived experience of teaching and researching this subject over several years, across different institutions and, in one instance, in a different country altogether. Workshops on privacy and security in online social media conducted at the Universidade Federal de Minas Gerais (UFMG) in Brazil, for example, eventually grew into a full academic conference on online social networks, and a full-credit version of this course was taught over two consecutive summers in Brazil. This history reflects a broader point that recurs throughout the course: the problems of privacy and security on social media are not confined to any one country or platform, but are global, evolving concerns that benefit from being studied comparatively.

This unit orients the learner to the objectives, structure and pedagogical approach of the course before the subsequent units examine policing on social media (Unit 2) and eCrime on social media (Unit 3) in detail.

1.3 SCOPE AND FOCUS OF THE COURSE

By the end of this course, learners will be able to appreciate a range of privacy and security concerns on online social media, including spam, phishing, fraud and identity theft, and will be able to relate these concerns to real-world incidents. The primary focus of the course is on the different dimensions of privacy and security as they manifest specifically on social media platforms, as distinct from privacy and security concerns in computing more broadly.

The course is organised around the idea that social media generates continuous, large-scale behavioural data, and that meaningful insight can be drawn from this data through systematic collection, analysis and visualisation. Learners are introduced to real platforms and real incidents – rather than only abstract concepts – so that the theoretical material is consistently grounded in observable, current phenomena on social media.

Two broad application areas receive particular attention in this course:

- Policing and citizen engagement: how police organisations across the world, and particularly in India, have adopted social media platforms to communicate with citizens, respond to emergencies, and build public trust – along with the challenges this creates, such as impersonation of official accounts.
- eCrime on social media: how criminals exploit the trust, virality and openness of social media to carry out phishing, fraud, impersonation, account compromise and reputation manipulation, and how such abuse can be recognised and studied.

Beyond these two anchor themes, the broader field of study that this course sits within also covers trust and credibility on social media – that is, how far the content that is posted and shared can be believed – and the general privacy implications of the large volumes of personal information, including photographs and location data, that users routinely share. While these broader topics are treated in greater depth elsewhere in the course sequence, they form essential

background for understanding why policing and eCrime, the two themes of this unit's companion units, matter as much as they do: both are, at their core, applied problems of trust and credibility playing out on social media.

1.4 METHODOLOGY: DATA COLLECTION, ANALYSIS AND VISUALISATION

A distinguishing feature of this course is its emphasis on empirical methodology. Rather than treating privacy and security as purely theoretical subjects, the course exposes learners to the practical process of collecting data from online social networks such as Facebook and Twitter using their respective Application Programming Interfaces (APIs), storing this data in databases such as MySQL and MongoDB, and analysing it to answer specific research questions.

For instance, a learner may wish to investigate whether the followers of a particular Twitter account are genuine or fake, or whether the content posted by a police organisation's social media handle is generating meaningful citizen engagement. Answering such questions requires the ability to collect relevant data, process and analyse it, and finally visualise the results in a form that supports clear interpretation and decision-making.

Learners are introduced to tools and techniques for social network analysis and text analytics, including natural language processing toolkits, which allow textual content posted on social media to be analysed systematically – for example, to understand the underlying needs, wants or emotional tone expressed by citizens interacting with an organisation's social media page. Visualisation tools are used to represent patterns in the collected data, including geo-location data, in a form that is easy to interpret.

The practical, hands-on component of the course is organised progressively. Learners first set up their working environment (a Linux-based operating system and the Python programming language), and then move on to collecting data through platform-specific APIs, such as the Twitter API. Once collected, data is stored using database systems – MySQL for structured, tabular data and MongoDB for the more flexible, semi-structured data typically returned by social media APIs – before being analysed and visualised.

Specific tools are introduced for specific purposes: NLTK (the Natural Language Toolkit) for analysing textual content; graph and network visualisation tools such as Gephi and ORA for studying the structure of social connections; and charting libraries such as Plotly and Highcharts for producing interactive visualisations of trends over time, including geo-located data drawn from posts that carry latitude and longitude information. Together, these tools allow a learner to move from a raw research question – such as whether a particular account's followers are genuine – through data collection and analysis to a clear, visual answer.

1.5 IDENTITY RESOLUTION ACROSS SOCIAL MEDIA PLATFORMS

A recurring technical challenge addressed in this course is identity resolution: the problem of determining whether accounts on different social media platforms – for example, a Facebook profile, a Twitter handle and a YouTube channel – belong to the same real-world individual or organisation. This is a non-trivial problem because usernames, display names and profile details often differ across platforms, and because deliberate impersonation further complicates matters.

Consider a simple, illustrative case: an individual's Facebook handle, Twitter handle and YouTube channel name might all be entirely different strings of text, with no obvious textual similarity connecting them. Confirming that these three accounts belong to the same person is, in general, a hard problem – it typically requires combining multiple weak signals, such as overlapping profile photographs, similar writing style, shared social connections, consistent posting times, or cross-links that the account owner may have placed between their own profiles.

Identity resolution has several practical applications. It can help advertising and marketing agencies present more relevant content to users by connecting their activity across platforms. It can help distinguish a genuine, official organisational account from an unverified or fake account – a problem that recurs throughout this course, particularly in the context of police organisations (Unit 2) and impersonation-based eCrime (Unit 3). Techniques for identity resolution, therefore, provide an important conceptual bridge between the study of legitimate social media use and the study of its misuse.

It is worth noting that identity resolution is a double-edged capability. The same techniques that help a citizen confirm that a police department's Twitter handle is genuine can, in principle, also be used to link together a private individual's otherwise separate online identities without their consent. This tension – between identity resolution as a tool for establishing legitimate trust and identity resolution as a potential privacy risk – is a theme that this course returns to repeatedly, and it is one of the reasons identity resolution is introduced early, in this unit, before the more applied discussions of policing and eCrime that follow.

1.6 COURSE STRUCTURE AND LEARNING SUPPORT

This course is organised into units that combine conceptual explanation with illustrative, real-world examples drawn from the study of social media platforms. Each unit begins with clearly stated learning objectives and concludes with a set of Check Your Progress questions that allow learners to test their understanding before proceeding to the next unit.

Beyond the written material, learners are encouraged to treat the course as a participatory experience rather than a purely passive one. In its original delivery, this course placed considerable emphasis on an online discussion forum, with learners expected to post at least one question, answer or comment related to the week's topic. This expectation was not incidental: research on learning behaviour has repeatedly shown that learners who remain active outside formal class time, through discussion and peer interaction, tend to perform better on the topics being studied. Weekly homework assignments were designed around the same principle – built directly from the content covered in that week's material, often requiring learners to apply a concept (such as collecting a small sample of social media data) rather than simply recall a definition.

Learner support in this course also extended to structured opportunities for direct interaction – including scheduled office hours conducted online, in an informal "Ask Me Anything" style, and, where learners were based in the same city as the instructing institution, optional in-person laboratory sessions where teaching assistants were available to help learners work through technical exercises and strengthen their conceptual understanding. A team of teaching assistants, each pursuing doctoral research in related areas – including social reputation on social networks, the use of phone numbers and over-the-top (OTT) technologies on social media, malicious content on Facebook, and the use of social media by police organisations – supported the creation of homework questions, the running of laboratory sessions, and the overall delivery of the course.

Learners are encouraged to engage actively with the subject matter in the same spirit: to observe how organisations and institutions around them use social media, to critically examine posts and accounts they encounter for signs of illegitimacy or manipulation, and to relate the concepts discussed in this course to their own use of social media platforms. This applied, observational approach is central to developing a practical understanding of privacy and security in social media, and it is the approach carried forward into Unit 2, on policing, and Unit 3, on eCrime.

1.7 CHECK YOUR PROGRESS

- Explain the primary focus and learning outcomes of the course Privacy and Security in Social Media – II.
- Describe the role of data collection, analysis and visualisation as a methodology for studying privacy and security concerns on social media.
- What is meant by 'identity resolution' across social media platforms? Why is it a difficult problem?
- List the two broad application areas – policing and eCrime – that anchor this course, and briefly describe why each is significant.
- Why is it important to study privacy and security specifically in the context of social media, rather than computing in general?

UNIT- 2
POLICING AND ONLINE SOCIAL MEDIA

2.1 OBJECTIVES

After successful completion of this unit, you will be able to-

- Understand how online social media has become a tool for crisis response and citizen engagement by police organisations.
- Trace the global and Indian adoption of social media platforms by police departments.
- Identify the problem of fake and unverified police accounts on social media and its implications.
- Understand how social media data generated by police organisations can be collected and analysed to draw meaningful insights.

2.2 INTRODUCTION

Online social media has moved well beyond its origins as a platform for personal communication. Over the last decade, public institutions – and police organisations in particular – have increasingly adopted platforms such as Facebook and Twitter to interact directly with citizens, disseminate information, respond to emergencies and build public trust. This unit examines how policing has evolved on social media, both globally and in the Indian context, and highlights the technical and social challenges that accompany this evolution, particularly the problem of fake or unverified accounts that impersonate legitimate police organisations.

This theme follows naturally from the study of privacy on social media. Earlier discussion of privacy has shown how information that individuals post online – photographs, check-ins, and other publicly visible data – can be pieced together to reveal a great deal about a person, including, in some documented studies, their home location or even partial predictions of sensitive identifiers. If private individuals leave such a rich, exploitable trail of data through their ordinary use of social media, it follows that public institutions, which post far more frequently and at much greater scale, generate an even richer trail of publicly observable data. This unit turns that observation into a constructive line of enquiry: rather than asking what can be learned about a person from their social media activity, it asks what can be learned about an institution – specifically, the police – from its social media activity, and how that institution's presence on social media can, in turn, be misused.

The discussion in this unit therefore builds on the broader theme of privacy on social media by illustrating a specific, socially significant application: the use of publicly available social media data to understand institutional behaviour, measure citizen engagement, and identify problems such as impersonation that have direct implications for public trust and safety.

2.3 SOCIAL MEDIA AS A TOOL FOR CRISIS RESPONSE AND CITIZEN ENGAGEMENT

One of the earliest and most striking demonstrations of the power of social media in a crisis occurred in January 2009, when US Airways Flight 1549 made an emergency landing on the Hudson River in New York. Photographs and eyewitness accounts of the incident circulated on Twitter before conventional news media or first responders had fully reported on the event; one of the earliest posts came from a bystander on the riverbank, sharing a photograph and a message offering to help. Until that point, Twitter and similar platforms had been used largely for casual, everyday communication – status updates, light-hearted hashtags, and conversations between friends. This incident marked one of the first occasions on which social media was used, in real time, to communicate information about an unfolding crisis, rather than merely to socialise – establishing a template that police and emergency services would later adopt deliberately.

Police organisations have since used social media not only to respond to crises but also to run deliberate public engagement campaigns. A notable example is the hashtag campaign #myNYPD, launched by the New York Police Department, which invited members of the public to share photographs of themselves with NYPD officers, with the promise that selected photographs might be featured on the department's own Facebook page. While the campaign generated considerable public participation, it also illustrates a recurring risk of such initiatives: hashtag campaigns can attract responses very different from those intended, including critical, satirical or embarrassing content submitted precisely because the hashtag was open to anyone. The same #myNYPD format was subsequently adopted – and, in some cases, used critically – in other cities under variants such as #myLAPD, spreading both the format and its risks well beyond the department that started it.

These two examples together illustrate the two faces of police engagement with social media that recur throughout this unit: social media as a genuinely valuable channel for real-time information and citizen engagement, and social media as an open, public space whose openness cannot always be controlled once an organisation invites participation. Both

faces are necessary to keep in mind when studying how police organisations, particularly in India, have gone on to build their own social media presence.

2.4 GLOBAL ADOPTION OF SOCIAL MEDIA BY POLICE ORGANISATIONS

Police departments in a number of major cities, including New York, Boston and Baltimore, as well as several forces in the United Kingdom, have established an active presence on Facebook and Twitter. Publicly visible metrics such as the number of likes and followers, the year a page was established, and whether the organisation allows the public to post directly on its page, provide useful indicators of how actively a given police organisation is using social media to engage with citizens.

One particularly informative metric is whether an organisation's Facebook page permits citizens to post directly on the page's own timeline, rather than only allowing comments on posts the organisation has made. Facebook allows page administrators to control this setting, and the choice an organisation makes is itself revealing: departments that allow open citizen posting are, in effect, opting into a more participatory and less controlled relationship with the public, whereas departments that restrict posting to their own content retain tighter control over their page's tone and content, at some cost to open engagement.

Comparative analysis of such metrics across cities and countries allows researchers to study patterns in institutional social media adoption – for example, whether departments that allow open posting by citizens generate higher engagement, or whether adoption has accelerated over particular periods, such as after 2010 in many jurisdictions. Such analysis is a useful illustration of how publicly available social media data can be used for institutional and policy-relevant research, and it sets up a natural comparison with the Indian context discussed next.

2.5 POLICING THROUGH SOCIAL MEDIA IN INDIA

In India, adoption of social media by police organisations has grown rapidly, particularly from around 2010–2011 onward, and has accelerated markedly in the years since. City and state police departments – including the Bangalore City Police, Bangalore Traffic Police, Delhi Traffic Police, Hyderabad City Police, UP Police, Mumbai Police, Kolkata Police, Pune Police, Guwahati Police and others – maintain active Facebook pages and Twitter handles that are used to share traffic advisories, public safety announcements, appeals for information, and updates on ongoing operations. Almost all of these official pages allow citizens to post directly, reflecting a broadly participatory approach across Indian police social media presence.

The Bangalore City Police Twitter handle offers a clear illustration of this growth. At one point of observation the account had grown to roughly 335,000 followers; within about a year, the same account had grown further, to nearly 688,000 followers, while following only around 60 accounts itself and having posted roughly 10,000 tweets – indicating both a rapidly growing citizen audience and a sustained, high level of posting and interaction activity. Typical posts include operational updates, such as announcements of traffic signal synchronisation across several city corridors to improve traffic flow, as well as posts recognising the efforts of individual police personnel, and occasional posts offering cash rewards for information leading to the resolution of a case. Citizens frequently engage with such posts through likes, comments and shares, providing police organisations with a direct channel of communication that complements traditional media.

This growing adoption has also created new research opportunities: by systematically collecting posts, likes, comments and follower data from official police accounts – as, for example, from a purpose-built directory that captures state and city police Facebook pages and Twitter handles across India – researchers can study how effectively different organisations are using social media, what kinds of content generate the highest citizen engagement, and how such engagement changes over time.

2.6 THE PROBLEM OF FAKE AND UNVERIFIED POLICE ACCOUNTS

A significant challenge that accompanies the growth of official police accounts on social media is the proliferation of fake and unverified accounts that closely mimic legitimate ones. Because anyone can create an account using a police organisation's name, logo or profile picture, citizens often find it difficult to distinguish an authentic account from an impersonating one – a difficulty that extends even to researchers who study these accounts closely.

Platform-provided verification badges (such as Twitter's blue tick) offer one signal of authenticity, confirming that the platform itself has taken steps to verify that the account genuinely belongs to the named organisation or individual. However, verification of this kind is not granted automatically, and historically only a small fraction of all accounts on a platform – numbering in the low millions out of hundreds of millions of users – have carried a verification badge. This

means many legitimate police handles remain unverified, so a verification badge alone cannot be relied upon as the only cue. Analysts and citizens must also consider signals such as the age of the account, its posting history and consistency of tone, whether its profile picture has ever been changed (an unchanged, generic "egg" profile picture on an old account is one common warning sign), links to the organisation's official government website, and consistency with other known official accounts of the same organisation.

Examples of this problem are well documented. Fake or parody handles resembling "Delhi Traffic Police" with subtly altered spellings, unofficial pages mimicking Kolkata Police using an identical profile picture but explicitly labelled as fake or satirical, and accounts using names such as "Mumbai Cops" without clear confirmation of their legitimacy, have all been observed. Some states have chosen to consolidate their social media presence into a single official page per organisation, while others maintain multiple pages at the district level for different activities – a choice that itself affects how easily citizens can distinguish genuine from fake accounts, since more official pages generally means more surface area for impersonation.

Systematically cataloguing official handles – by cross-referencing them against verified government websites (such as an official ".gov.in" domain) and platform verification status, and recording the confirmed source for each entry – is one practical approach to building a reliable, continuously updated directory of legitimate police social media accounts. Such a directory, covering dozens of state and city police pages, helps both citizens and researchers avoid the confusion created by impersonating pages, and only a minority of the entries in any such directory are typically found to carry platform verification, underscoring why cross-referencing against multiple signals remains necessary.

2.7 COLLECTING AND ANALYSING POLICE SOCIAL MEDIA DATA

The study of policing on social media in this course follows the same empirical methodology introduced in Unit 1: collecting data from official accounts using platform APIs, storing this data systematically, and analysing it to answer specific questions. For police social media pages, relevant data includes the volume and timing of posts, the number of likes and comments received, and the growth of followers over time. A purpose-built visual dashboard, for instance, can let an analyst start from a map of India, select a state, drill down to a specific city and organisation, and view the collected Facebook or Twitter data for that account – including how posting activity and citizen engagement have varied over specific days, weeks or months.

Such analysis can reveal, for instance, which days or months saw the highest posting activity – in one observed case, a single police page recorded 59 posts on a single day in late December – which kinds of posts (traffic advisories, appeals for information, recognition of personnel) receive the greatest citizen engagement, and how an organisation's online activity compares with that of similar organisations elsewhere. This kind of analysis has practical value: it can function as an early-warning indicator of emerging public concerns, support predictive analysis of citizen sentiment, and ultimately help build technologies that make it easier both for police to make better-informed decisions and for citizens to interact with police more effectively, contributing to safer communities.

More broadly, this line of research is not limited to Indian police pages; comparable studies have been conducted on police social media use in cities across the world, reflecting a shared, global interest in understanding how public institutions can use social media responsibly and effectively to serve citizens. The techniques introduced here – systematic data collection, engagement analysis and visualisation – recur, in a very different context, in Unit 3, which examines how the same platforms are exploited for eCrime.

2.8 CHECK YOUR PROGRESS

- Describe the Hudson River (US Airways Flight 1549) incident of 2009 and explain its significance in the history of social media use during crises.
- What risks are associated with public hashtag campaigns run by police organisations? Illustrate with an example.
- What does it reveal about a police organisation's approach to citizen engagement if it allows the public to post directly on its Facebook page?
- Discuss the growth of social media adoption by Indian police organisations, with reference to at least two examples.
- What is the problem of fake and unverified police accounts on social media? Why can platform verification alone not fully solve this problem?
- Explain how data collected from official police social media accounts can be analysed, and describe two practical uses of such analysis.

UNIT- 3
ECRIME ON ONLINE SOCIAL MEDIA

3.1 OBJECTIVES

After successful completion of this unit, you will be able to-

- Understand the concept of eCrime and its specific manifestations on online social media.
- Identify different forms of phishing that occur on social media platforms.
- Recognise common fraudulent practices, including fake customer service accounts, fake comments, fake live streams, fake discounts and fake surveys.
- Understand how social reputation can be manipulated, and how clickbaiting, hashtag hijacking, account compromise and impersonation occur on social media.

3.2 INTRODUCTION

Electronic crime, or eCrime, refers broadly to criminal activity that is carried out using electronic means, including the internet and connected digital platforms. While eCrime can occur across many parts of the internet, this unit focuses specifically on the forms of eCrime that occur on online social media platforms such as Facebook and Twitter, where the openness, virality and inherent trust that users place in their social connections can be exploited by criminals for financial gain, data theft or reputational harm.

This unit follows directly from the discussion of policing in Unit 2. There, the focus was on how a legitimate institution builds and uses a social media presence, and how that presence can be undermined by impersonation. Here, the focus widens to the broader landscape of eCrime aimed not only at institutions but at ordinary users, brands and public figures alike. The same underlying dynamic connects both units: social media rewards content and accounts that appear familiar, trustworthy and popular, and precisely this dynamic can be exploited – whether by someone impersonating a police department, or by a fraudster impersonating a bank, a celebrity or a friend.

This unit surveys the major categories of eCrime observed on social media – phishing, various forms of fake accounts and fraudulent activity, social reputation manipulation, clickbaiting, hashtag hijacking, account compromise and impersonation – and explains, for each, the underlying mechanism, the way it manifests on social media specifically, and why it is effective. Recognising these patterns is the first step toward building the technical and behavioural defences discussed elsewhere in this course.

3.3 PHISHING ON SOCIAL MEDIA

Phishing is the practice of deceiving a person into revealing sensitive information, such as login credentials, by presenting a fraudulent request that appears to come from a legitimate and trusted source. In its traditional form, phishing is carried out through email: a message purporting to be from a bank or service provider asks the recipient to click a link and “verify” or “update” their account, leading them to a fake website that may or may not closely resemble the genuine site. On this fake site, when the user enters their username and password to “log in”, they are, in effect, handing their credentials directly to the attacker. Such emails, and the sophisticated attacks built around them, have circulated for many years and continue to evolve.

Variants of phishing include spear phishing, in which a specific, narrowly defined group of people is targeted with a message tailored to appear especially credible to that group – for example, an email appearing to come from a course instructor and addressed only to students enrolled in that course, referencing course-specific details to increase its credibility – and whaling, in which senior executives of an organisation are specifically targeted because of the high value of the credentials or information they hold. Many other variants exist, but all share the same basic structure of a trusted-looking request leading to a fraudulent destination.

On social media, the same underlying mechanism reappears in platform-specific forms: a link posted on a Twitter timeline promising a cash reward, or a message claiming there has been a problem with a Facebook account and asking the user to click a link to “verify” or “update” their password. A link may appear on Facebook, on Twitter, or arrive by email directing the user to click an image or link for “more information” about a topic of interest, only to lead to a fraudulent site.

Because a compromised social media account reveals a great deal about a person – their social connections, interests, and behavioural patterns – social media credentials have become an increasingly attractive target for phishing, in some cases even more valuable to attackers than email credentials or financial account details. Common real-world examples include messages claiming to be from “Facebook technical support” asking a user to verify a supposedly compromised account, or messages announcing a new Facebook “login system” and asking the user to click a link to migrate or merge their existing account into it – both designed to capture login credentials under the guise of routine account maintenance.

3.4 FAKE ACCOUNTS AND FRAUDULENT PRACTICES ON SOCIAL MEDIA

Beyond phishing, several distinct categories of fraudulent activity involving fake accounts have been observed on social media platforms. Each exploits a slightly different aspect of how users behave and what they trust:

- Fake customer service accounts: when a customer publicly tags a legitimate organisation (such as a bank) in a complaint – for example, posting that a bank's website has repeatedly failed to work – fraudsters monitor such posts and reply using an account resembling the organisation's real account, directing the customer to a fake “secure sign-on” link to “resolve” their issue and thereby capturing login credentials from someone who believes they are dealing with genuine support.
- Fake comments on popular posts: when a post from a prominent figure or on a widely followed topic – a major sporting event, for instance – becomes very popular, scammers create accounts that post comments containing malicious links, exploiting the visibility of the original post to reach a large number of potential victims.
- Fake live-streaming links: during major events such as sporting tournaments, scammers post content claiming to offer a live video stream; users seeking to watch the event on their laptop or phone are instead directed to fraudulent or malware-laden websites, with no genuine stream ever existing.
- Fake online discounts: fraudulent pages imitate well-known brands – a streaming service, a retailer, or any recognisable business – and advertise implausibly large discounts to lure users into providing payment details or personal information on a page with no real connection to the brand.
- Fake online surveys and contests: users are invited to complete a survey – often framed as a personality test – or enter a contest promising a cash prize, in exchange for a small amount of personal information; the actual purpose is typically to harvest personal data or spread malicious content.
- Fake tips on location-based networks: on networks such as Foursquare, where users can “check in” to a venue and leave a short “tip”, scammers post tips containing promotional or phishing links entirely unrelated to the venue, exploiting the visibility of the location-based feed.

3.5 MANIPULATION OF SOCIAL REPUTATION

Social reputation – measured through visible metrics such as the number of followers, likes, endorsements and reviews – has become an important currency on social media, shaping how individuals, organisations and products are perceived. This is no longer limited to celebrities or public figures: even among ordinary friends, the number of followers or likes a person has has become a casual measure of social standing. Facebook likes, Amazon and Flipkart reviews, YouTube likes, and LinkedIn endorsements for particular skills have all become, in their own way, informal measures of influence or credibility. Because these metrics are visible and comparative, they create a strong incentive to manipulate them artificially.

Online reviews are a well-documented example: sellers or manufacturers may arrange for large numbers of favourable but fabricated reviews to be posted about a product, artificially inflating its perceived quality. Such practices have drawn legal action in some cases – for instance, a major e-commerce platform has pursued legal proceedings against a large number of individuals found to be posting fake reviews at scale. Similar manipulation can occur with follower counts and endorsements, where accounts purchase or exchange artificial followers to appear more influential than they genuinely are.

This is a genuinely difficult problem to study, because manipulated reputation signals are, by design, made to look identical to genuine ones. Because reputation signals on social media are increasingly used to inform real decisions – purchasing choices, trust in an information source, even which accounts are worth engaging with – manipulation of these signals is a significant form of eCrime, connecting directly to the broader challenge of distinguishing genuine from fake accounts discussed throughout this course.

3.6 CLICKBAITING AND HASHTAG HIJACKING

Clickbaiting refers to the practice of presenting a headline, image or description that is deliberately sensational or misleading in order to induce users to click through, only to be directed to content that does not match what was promised – or to a fraudulent website designed to capture information or install malicious software. A typical example is a post or link that presents an exaggerated or emotionally charged claim connected to a news story or event a user is already following, luring them to click without offering any of the substance the headline implied.

Hashtag hijacking is a related but distinct problem: a hashtag that is trending because of its association with a particular topic or event is deliberately co-opted by an unrelated party – often for commercial promotion – without regard for the

original context of the hashtag. Because a hashtag exists precisely to group together posts on a shared topic, using a popular hashtag automatically places a post in front of everyone already following that topic, which is what makes hijacking attractive to marketers, and occasionally to malicious actors as well.

A widely cited example involved a well-known beverage brand using a hashtag associated with a royal birth announcement to promote its own product, drawing criticism for being tangential to, and appearing to commercially exploit, the occasion. Another example involved a food brand's use of a hashtag that was, in fact, being used by survivors to share accounts of domestic violence; the brand's unrelated promotional post caused considerable public backlash and required a public apology, with the brand acknowledging that it had not checked the context in which the hashtag was actually being used before posting.

These examples illustrate that hashtag hijacking is not only a matter of poor marketing judgement but can cause genuine reputational harm – both to the party whose hashtag has been hijacked and to the party doing the hijacking – when the hijacked hashtag carries serious or sensitive connotations that the hijacking party has failed to recognise before posting.

3.7 ACCOUNT COMPROMISE AND IMPERSONATION

Account compromise occurs when an unauthorised party gains access to a legitimate, established social media account – typically through leaked or stolen credentials, often obtained via exactly the kind of phishing described earlier in this unit – and uses it to post content the genuine owner did not authorise. A widely cited example is the 2013 compromise of a major news organisation's verified Twitter account, which was used to post a false report of an attack on a government building in which a senior political figure was said to have been injured. The false report caused a brief but measurable disruption in financial markets before it was identified as fabricated and retracted.

This example illustrates how the trust associated with a verified, well-established account can be exploited to give false information much greater initial credibility, and much wider circulation, than it would otherwise have – a reminder that verification and reputation, discussed earlier in this unit, are valuable precisely because they can be exploited when compromised.

Impersonation involves the creation of an account that mimics a real individual or organisation without authorisation, using publicly available details such as photographs, city of residence and biographical information gathered from online sources. Impersonation affects both individuals – for example, complaints have been filed by private citizens whose photographs and personal details were used, without consent, to create unauthorised accounts in their name – and organisations, including police organisations, as discussed in Unit 2.

Because the information needed to create a convincing impersonating account (photographs, workplace, city, social connections, even writing style) is often readily available from a person's own public social media activity, impersonation is a persistent and difficult problem to fully prevent. It is best addressed through a combination of platform reporting mechanisms that allow victims to flag impersonating accounts, public awareness of how easily such accounts can be created, and, where available, verification systems of the kind discussed in Unit 2 – underscoring, once again, how closely the problems of policing and eCrime on social media are connected.

3.8 CHECK YOUR PROGRESS

- Define eCrime and explain why online social media has become a significant context for such crime.
- Distinguish between phishing, spear phishing and whaling, giving an example of each in the context of social media.
- Describe any three types of fake accounts or fraudulent practices discussed in this unit, with an example of each.
- How is social reputation manipulated on social media, and why does this matter?
- Differentiate between clickbaiting and hashtag hijacking, using examples.
- What is meant by ‘account compromise’? Illustrate with the example discussed in this unit and explain its consequences.
- How does impersonation occur on social media, and what information do perpetrators typically rely upon?

UNIT- 4
LINK FARMING IN ONLINE SOCIAL MEDIA

Unit Structure

4.1 OBJECTIVES

After successful completion of this unit, you will be able to—

- Define link farming and explain why it occurs in online social networks.
- Explain the concept of link reciprocity and the role of spam followers.
- Interpret the relationship between a user's in-degree and the probability of reciprocating a follow request.
- Identify the structural signatures of top link farmers.
- Explain why legitimate and verified accounts participate in link farming, and why this makes detection difficult.

4.2 INTRODUCTION

On platforms such as Twitter (now X), the number of followers a user holds is not merely a vanity statistic. It determines how many accounts receive a message directly, it feeds the ranking algorithms that decide whose content appears in search results, and it serves as a visible proxy for credibility. Follower links therefore carry real economic and reputational value.

Wherever a valuable quantity can be accumulated, there is an incentive to accumulate it artificially. On the World Wide Web this produced *link farms*—clusters of websites linking to one another purely to inflate search rankings. Social media has produced an equivalent. Users, and spammers in particular, acquire large numbers of follower links through aggressive and indiscriminate following. This is called **link farming**.

This unit examines who supplies those links, what the accounts that accumulate them look like, and the uncomfortable finding that the main beneficiaries on Twitter were not anonymous spam accounts but popular, active and often verified users. The discussion draws on an empirical study of the Twitter follow graph by Ghosh and colleagues, published at the World Wide Web Conference in 2012, which analysed the links acquired by over 40,000 suspended spammer accounts. The figures describe Twitter as it was at that time; the behavioural principles remain directly relevant today.

4.3 UNDERSTANDING LINK FARMING

Link farming is the practice of acquiring a large number of incoming follower links through indiscriminate reciprocal following, in order to inflate one's apparent audience, influence and ranking rather than to build genuine social ties.

The follow relationship is *directed*: A following B does not oblige B to follow A. The network is therefore a directed graph, and three terms are essential:

- **In-degree** — the number of accounts that follow a user (follower count).
- **Out-degree** — the number of accounts a user follows (following count).
- **Reciprocity** — the probability that when A follows B, B follows A back.

Reciprocity is what makes link farming possible at scale. A spammer cannot compel anyone to follow it, but it can follow very many accounts and rely on a fraction following back. Every reciprocated follow converts cheap outgoing effort into valuable incoming social capital.

The acquired links are valuable through three channels. They enlarge the **direct audience** receiving the spammer's posts. They raise the spammer's score in **link-based ranking systems**, which are conceptually related to PageRank and assign a node a score based on the links pointing to it and the importance of the accounts supplying them. And they increase **perceived credibility**, since ordinary users treat follower count as a shortcut for trustworthiness.

4.4 SPAM FOLLOWERS AND RECIPROCITY

The first question is where the spammers' followers come from. The answer is that link supply is sharply concentrated. The **top 100,000 spam followers** accounted for approximately **60 per cent of all links acquired by the spammers**. Because of this central role, these accounts are called the **top link farmers**.

When these accounts are ranked by how many links they supply, and the fraction of reciprocated incoming links is plotted against that rank, the highest-ranked accounts show a reciprocation fraction close to 1.0—they follow back almost everyone who follows them. The fraction declines with rank but remains substantial across the top ranks.

The implication is direct. The accounts responsible for the bulk of spammers' follower links are exactly those that reciprocate almost unconditionally. A spammer needs no sophisticated targeting: it need only follow accounts known to follow back. Its in-degree rises, its ranking score rises with it, and its content is surfaced more often.

4.5 RESPONSIVENESS AND IN-DEGREE

Which targeted users are most likely to respond? Intuition suggests that users with few followers would be flattered into reciprocating while heavily followed users would ignore an unknown account. The data show the opposite.

Plotting the probability of response against in-degree reveals that **responsiveness increases with the number of followers**. Users with low follower counts largely do not reciprocate links from spammers. Users with high in-degree tend to follow back far more often.

Several explanations are consistent with this pattern. Accounts pursuing audience growth often adopt an explicit follow-back policy as a routine cost of doing business. Accounts managed at scale frequently use automated tools that follow back without human judgement. An account receiving hundreds of new followers daily cannot assess each one, and the marginal cost of following back seems zero. Finally, following back is widely read as a social courtesy. Whatever the mix of reasons, the users best placed to confer ranking value on a spammer are also the users most likely to do so.

4.6 WHO ARE THE TOP LINK FARMERS?

Status of account	Number	Approximate share
Suspended by the platform	18,826	About 19%
Not found (deleted or deactivated)	4,768	About 5%
Still active	—	About 76%
Of the active accounts: verified	235	—

Table 4.1: Status of the top 100,000 link farmers at the time of analysis

Roughly one in five had been suspended, indicating the platform had classified them as malicious. But about three-quarters were still active accounts in good standing. **The top link farmers were, for the most part, not spammers.** Further, 235 were **verified accounts**, carrying the blue tick that confirmed the account genuinely belonged to the person or organisation it claimed to represent.

To test whether these were genuinely legitimate, researchers drew a random sample of 100 of the 235 verified accounts and had human volunteers inspect each one, using more than one assessor per account. **86 were confirmed real.** Their content clustered consistently around business, internet marketing, entrepreneurship, money and social media.

Profile biographies of the highest-ranked link farmers reinforce this: an internet affiliate marketer, an artist and founder, a technology enthusiast who states openly that they will follow back, a social media manager. These are self-promotional accounts, and at least one advertises its follow-back policy explicitly. The accounts with the highest ranking scores overall included those of Barack Obama and his campaign staff, Britney Spears, the politics desk of NPR, the office of the UK Prime Minister and JetBlue Airways.

> **The Central Finding**
>
> Link farming on Twitter was not confined to malicious accounts. It was substantially carried out by legitimate, popular and highly active users—bloggers, marketers, experts, politicians, brands and celebrities—building their own social capital. Their indiscriminate follow-back behaviour, adopted for lawful reasons, is precisely what spammers exploited.

4.7 STRUCTURAL CHARACTERISTICS OF LINK FARMERS

The study compared the top 100,000 link farmers against two reference groups: known spammer accounts, and a large random sample of ordinary users. Comparing three groups rather than two reveals whether link farmers behave more like spammers or more like ordinary users.

On **in-degree**, the cumulative distribution for top link farmers lies far to the right of both comparison groups. At a follower count of about one thousand, only around 10 per cent of link farmers have been accounted for, meaning roughly 90 per cent have more followers than that. For spammers and random users, the same threshold already covers most of the group. On **out-degree**, the picture is similar: link farmers follow far more accounts than either comparison group. They are hubs in both directions.

The most useful result concerns the **ratio** between the two. A celebrity account has a ratio far above 1—followed by millions, following few. An aggressive spammer has a ratio well below 1—following many, followed by few. An ordinary user sits close to 1.

Most top link farmers have a ratio close to 1, clustering around 10^0 on a logarithmic axis, while spammers and random users show visibly different ratios. This is what makes link farming hard to police. A ratio near 1 is the signature of an ordinary legitimate user. An account with 50,000 followers and 48,000 followings does not look suspicious. Any detection rule based on this ratio alone would either miss the link farmers or sweep up unacceptable numbers of ordinary users.

Content analysis shows a matching pattern. Link farmer biographies are dominated by commercial vocabulary—market, online, internet, social—while random users describe themselves personally, with words such as life, music and live. The word *love* appears in

both, a reminder that no single word is diagnostic. Link farmers also post far more links to external websites than ordinary users do.

4.8 IMPLICATIONS FOR SOCIAL CAPITAL AND DETECTION

Social capital is the value a user derives from their network position—the reach of their voice, the credibility others assign them, the influence they exercise. On social media this is measured almost entirely through visible quantities: follower counts, engagement and content propagation. Because indiscriminate reciprocal following raises these cheaply and lawfully, link farming is an effective strategy. Its practitioners are best understood not as criminals but as **social capitalists**, pursuing status within the platform's own incentive structure.

Detection is therefore difficult on every front. Link farmers are structurally indistinguishable from ordinary users on the in-to-out ratio. Most are legitimate accounts, making account-level enforcement inappropriate. Many are verified, meaning the platform has itself certified them. And the behaviour violates no rule.

The response must therefore operate on the ranking system rather than the account. If the value of an acquired link is what motivates link farming, reducing that value reduces the incentive. Links can be weighted by the discrimination of the account supplying them, so that a link from an account that follows back everyone is worth very little, and links acquired through automated reciprocation can be discounted. The strategy then loses its point without anyone being penalised.

4.9 LET US SUM UP

- Link farming is the acquisition of follower links through indiscriminate reciprocal following, to inflate audience, ranking and perceived influence.
- Link supply is concentrated: the top 100,000 spam followers supplied about 60 per cent of all links acquired by spammers, and reciprocate at very high rates.
- Responsiveness increases with in-degree. Users with few followers ignore spammers; heavily followed users follow back.
- About 76 per cent of top link farmers were still active and 235 were verified. Of 100 inspected, 86 were genuine accounts oriented towards business and marketing.
- Link farmers show very high in-degree and out-degree but a ratio close to 1—the signature of an ordinary user—which defeats simple detection rules.
- The motive is social capital. Since the behaviour is lawful and practised by legitimate users, mitigation belongs in ranking design, not account suspension.

4.10 CHECK YOUR PROGRESS

Descriptive Type Questions

1. Define link farming and explain, with reference to at least three distinct channels of benefit, why acquiring follower links is valuable to a spammer.
2. Distinguish between in-degree, out-degree and reciprocity, and explain how reciprocity makes large-scale link farming feasible.
3. Why is the finding that responsiveness increases with follower count counter-intuitive? Give at least three explanations that account for it.
4. Explain why the in-degree to out-degree ratio makes link farmers difficult to detect, contrasting the expected ratios for a celebrity, an ordinary user and a spammer.
5. “The top link farmers were not spammers.” Discuss with reference to the account-status data and the manual verification exercise, and explain why this complicates enforcement.
6. What is meant by describing link farmers as “social capitalists”, and why does this suggest ranking-system design as a better response than account suspension?

Multiple Choice Questions (MCQs)

1. Link farming in an online social network refers to:

   a) Sharing agricultural information through social media groups

   b) Acquiring large numbers of follower links to inflate audience and ranking

   c) Encrypting links between a user’s device and the platform server

   d) Removing inactive followers to improve engagement

   Answer: b

2. The in-degree of a user in a directed follow graph is:

   a) The number of accounts the user follows

   b) The number of accounts that follow the user

   c) The number of posts the user has published

   d) The number of reciprocated links the user has

   Answer: b

3. Approximately what share of all links acquired by spammers came from the top 100,000 spam followers?

a) 20 per cent

b) 40 per cent

c) 60 per cent

d) 90 per cent

Answer: c

4. According to the study, the probability that a targeted user follows a spammer back:

a) Decreases as the user’s in-degree increases

b) Increases as the user’s in-degree increases

c) Is unrelated to the user’s in-degree

d) Is the same for all users at about 50 per cent

Answer: b

5. Most top link farmers were found to have an in-degree to out-degree ratio:

a) Close to 1, resembling an ordinary user

b) Far above 1, resembling a celebrity account

c) Far below 1, resembling an aggressive spammer

d) Exactly 0, since they had no followers

Answer: a

6. Of 100 randomly selected verified accounts inspected by volunteers, how many were confirmed to be real?

a) 46

b) 62

c) 86

d) 100

Answer: c

7. The most appropriate response to link farming by legitimate users is to:

a) Suspend every account with more than 10,000 followers

b) Reduce the ranking value of links acquired through indiscriminate reciprocation

c) Prohibit users from following anyone they have not met in person

d) Remove follower counts from all profiles permanently

Answer: b

Fill in the Blanks

1. The probability that a user follows back an account that follows them is known as _____.

Answer: reciprocity

2. The number of accounts that follow a given user is called the user's _____.

Answer: in-degree

3. The top _____ spam followers accounted for about 60 per cent of all links acquired by spammers.

Answer: 100,000

4. Of the top 100,000 link farmers, _____ were verified accounts carrying a blue tick.

Answer: 235

5. Because most top link farmers had an in-degree to out-degree ratio close to _____, they were hard to distinguish from ordinary users.

Answer: 1 (one)

6. The value a user derives from their network position, measured through reach, credibility and influence, is called _____.

Answer: social capital

4.11 REFERENCES

1. Ghosh, S., Viswanath, B., Kooti, F., Sharma, N. K., Korlam, G., Benevenuto, F., Ganguly, N., & Gummadi, K. P. (2012). Understanding and combating link farming in the Twitter social network. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12)* (pp. 61–70). Lyon, France: ACM. https://doi.org/10.1145/2187836.2187846

2. Kumaraguru, P. (n.d.). *Privacy and Security in Online Social Media* [NPTEL online course, Week 7.1: Link Farming in Online Social Media]. Department of Computer Science and Engineering, Indian Institute of Technology Madras.

3. Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the Web* (Technical Report 1999-66). Stanford InfoLab.

4. Gyöngyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). Combating web spam with TrustRank. In *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB '04)* (pp. 576–587). Toronto, Canada: VLDB Endowment.

5. Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting spammers on Twitter. In *Proceedings of the 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS)*. Redmond, WA.

6. Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)* (pp. 10–17). Washington, DC.

Note on sources: The empirical findings in this unit—the concentration of link supply, the reciprocation and responsiveness patterns, the account-status figures, the verification exercise and the degree distributions—are drawn from Reference 1 above and from the NPTEL lecture at Reference 2, which discusses that study. The figures describe the Twitter network at the time of the original analysis and should be read as illustrative of the underlying behavioural mechanisms rather than as current platform statistics.

UNIT- 5
PRIVACY IN LOCATION-BASED SOCIAL NETWORKS

Unit Structure

5.1 OBJECTIVES

After successful completion of this unit, you will be able to—

- Define location-based services and location-based social networks, and name the major platforms in this category.
- Explain the privacy concerns that arise when location information is shared.
- Interpret survey evidence on how Indian users perceive location privacy.
- Define the core Foursquare concepts of check-in, venue, mayorship, tip and done, and distinguish private from public attributes.
- Describe how a majority-voting model infers a user's home city from public attributes, and interpret the reported accuracy.
- Outline the standard structure of a research paper in this field.

5.2 INTRODUCTION

Almost every smartphone in use today carries a GPS receiver, and this single fact has reshaped social media. Where a status update once conveyed only what a person was thinking, it now routinely conveys where they were standing while thinking it. Geographic information has been woven into ordinary sharing so thoroughly that most users no longer consciously notice they are disclosing it.

This unit examines what that disclosure makes possible. The central argument is uncomfortable but important: a user may decline to state where they live, and their home city can still be determined with high accuracy from information shared for unrelated reasons. Privacy is not controlled only by the fields a user leaves blank, but also by what can be *inferred* from the fields they fill.

The study examined here is *We Know Where You Live: Privacy Characterization of Foursquare Behavior* by Pontes and colleagues, presented at the ACM Conference on Ubiquitous Computing in 2012, based on a crawl of about 13.5 million users.

5.3 LOCATION-BASED SERVICES AND SOCIAL NETWORKS

A **location-based service (LBS)** uses the geographic position of a device to deliver relevant information—navigation, nearby-restaurant search, local weather. A **location-based social network (LBSN)** is a social network in which geographic information is attached to shared

content. Its distinguishing feature is not that the platform knows where the user is, but that it *publishes* that knowledge into a social context.

Platform	Nature of location functionality
Foursquare	Check-ins at venues with rewards for frequent visits
Yelp	Location-anchored reviews of restaurants and local businesses
Gowalla	An early check-in based LBSN
Facebook	Check-ins attached to posts; location tagging of photographs
Twitter	Optional geo-tagging of posts with exact coordinates

Table 5.1: Major location-based services and social networks

A traveller at a Delhi airport terminal who posts an update announcing their departure, with the terminal tagged, has acted voluntarily and entirely ordinarily. They have also published the fact that they are leaving, and by implication that their home is now unoccupied.

5.4 PRIVACY CONCERNS IN LOCATION SHARING

Once a location is published, others know where the user is, and if the post is public that knowledge reaches anyone at all. Beyond this immediate exposure lies a more serious harm. A single disclosure reveals a point; a *sequence* reveals a pattern, and patterns are far more revealing. From an accumulated location history one can derive:

- **Movement patterns and routine** — where a person goes regularly, when, and by which route.
- **Residence and workplace** — the two locations that dominate most histories, identifiable without either being stated.
- **Interests and habits** — the categories of venue frequented, supporting inferences about health, beliefs and affiliations.
- **Absence from home** — the inverse inference, which converts a travel announcement into an opportunity for burglary.

Each platform adds concerns from its own design—which fields are visible by default, how easily data can be retrieved in bulk. The gap between what a user believes is visible and what actually is visible is where most location privacy harm originates.

5.5 USER PERCEPTIONS OF LOCATION PRIVACY IN INDIA

Technical exposure and perceived exposure differ. A privacy survey conducted in India, with 10,427 respondents, included a section on online social media.

Reported privacy setting for location on Facebook	Share of respondents
Everyone	39.0%
Friends	33.0%
Friends of friends	12.5%
Network	5.7%
I have customised it	3.1%

Table 5.2: Reported privacy settings for location information

The largest group—39 per cent—had location visible to everyone, and only 3.1 per cent had customised the setting. Respondents answered “to the best of your knowledge”, so these are *reported* settings, not audited ones; where users are mistaken, the error usually runs towards believing themselves more protected than they are.

A second question tested recognition of *indirect* disclosure. A subscriber holding a Delhi mobile connection travels with the handset switched off: callers hear the unavailability announcement in Marathi while the subscriber is in Mumbai, in Malayalam while in Kerala. The subscriber has shared nothing, yet the language discloses their linguistic region to every caller. Roughly 54 per cent considered this privacy invasive and about 22 per cent did not, with a substantial group neutral—indicating that many users have simply not considered such indirect channels.

> **Relevance to the Indian Context**
>
> This survey was conducted entirely in India. Since then, India has enacted the Digital Personal Data Protection Act, 2023, which establishes consent requirements for processing personal data including location data. As you read Section 5.7, consider the question this raises: if a home city can be *derived* from data shared for another purpose, what does meaningful consent require?

5.6 FOURSQUARE: TERMINOLOGY AND PUBLIC ATTRIBUTES

Term	Meaning
Venue	A page representing a real physical location. City and address are mandatory when it is created.
Check-in	Registering presence at a venue via GPS. Shared with the user's friends only.
Badge	A reward for checking in at particular venues or reaching a set number of check-ins.
Mayorship	Awarded to the most frequent visitor over the preceding 60 days; retained in the user's history.
Tip	A comment left at a venue about the user's experience.
Done / To-do	Markers on another user's tip signalling agreement or intention to visit.

Table 5.3: Core Foursquare terminology

Gamification—the use of points, badges and titles in non-game activities to increase engagement—was central to Foursquare's design, and the rewards were not only symbolic. Businesses commonly gave the mayor of their venue discounts or reserved parking, creating a real incentive to check in repeatedly at venues near daily life. Here gamification and privacy intersect: the mechanism that makes the platform engaging is the same mechanism that concentrates public activity around where the user actually lives. A mayorship is, by construction, evidence of habitual presence.

The critical design feature is an asymmetry in visibility. **Check-ins are shared only with friends, but the lists of mayorships, tips and dones are public.** Users treat check-ins as sensitive, while a mayorship feels like an achievement and a tip like a review. Yet each is anchored to a venue carrying a mandatory address, and the full history with timestamps is publicly readable. The user has published a timestamped list of places visited without perceiving it.

This is an **information leakage surface**: the set of publicly exposed attributes from which unintended inferences can be drawn. It arises not from a breach but from the system operating exactly as designed.

5.7 CASE STUDY: INFERRING THE HOME CITY OF USERS

On Foursquare the home city field is optional, free text and unvalidated, so a user who does not wish to disclose it can leave it blank. The research question follows: *despite being*

private data the user may choose not to reveal, can the home city still be inferred from publicly available mayorships, tips and dones?

Element of the dataset	Count
Users	13,570,060
Distinct venues	15,898,484
Mayorships	15,149,981
Tips	10,618,411
Dones	9,989,325

Table 5.4: Composition of the Foursquare dataset (crawled August–October 2011)

Since Foursquare had about 10 million users in June 2011 and 15 million by December, this crawl approximates the entire platform rather than a sample. Free-text fields were standardised using a geo-coding service: around 98 per cent of users supplied a valid home city and only 0.2 per cent left it empty. Ambiguous entries were discarded rather than guessed at. About 30 per cent of all users, some 4.2 million, had at least one mayorship, tip or done.

The method rests on one assumption: **users tend to have mayorships, tips and dones at venues where they live.** For each user, the locations of all venues attached to their public attributes are collected and the most frequent taken as the inferred home city—a **majority voting** scheme, chosen deliberately for its simplicity so the result would show the *minimum* capability an adversary would have. Users fall into three classes: **Class 0** have a single activity, **Class 1** have multiple activities with one predominant location, and **Class 2** have no predominant location and cannot be handled. Between 87 and 91 per cent of eligible users were in Classes 0 and 1.

Model	Class 0	Class 1	Overall
Mayorship	51.61%	67.41%	60.12%
Tip	51.52%	67.29%	59.11%
Done	50.09%	61.74%	55.89%
All three attributes	51.46%	64.86%	59.85%

Table 5.5: Accuracy of home city inference by model and class

Mayorships are the strongest single attribute, with tips only marginally weaker. Combining attributes slightly reduces accuracy because tips and dones add noise, but substantially increases *coverage*: the Mayorship model applies to about 1.81 million users, the All model to about 2.82 million, making All the best model in absolute terms. Coarser geography is easier—home state reached about 75 per cent and home country over 90 per cent.

Examining the errors shows many are not badly wrong: about **46 per cent of error distances were under 50 kilometres**, a plausible gap between neighbouring cities. Combining correct inferences with these near-misses gives the headline result:

Principal Finding

The home city of approximately **78 per cent** of analysed users can be inferred to within **50 kilometres**, using only publicly available mayorships, tips and dones and a simple majority-voting method.

A Commonly Repeated Misstatement

This finding is sometimes quoted as “75 per cent of users within 58 hours”. That version is incorrect and originates in a transcription error. The unit is distance, not time: **78 per cent within 50 kilometres**, as the paper’s abstract states. A result whose units make no sense is a signal that something has been garbled—always check quantitative claims against the primary source.

The significance lies not in the percentage but in the principle. A user who withholds their home city is not thereby protected: the information can be reconstructed from data published for different reasons, by a method simple enough for anyone with API access to implement.

5.8 ANATOMY OF A RESEARCH PAPER

Papers in this field follow a stable structure, and knowing it lets you navigate one efficiently and write one yourself.

Section	Function
Abstract	Compressed account of problem, method, data and finding
Introduction	Opens broadly, narrows to the problem, closes with the question and contributions
Related work	Engages with prior literature and states how this work differs
Background and dataset	Defines terminology; describes collection period and data validity
Characterisation	Describes the shape of the data before inference
Methodology	States the assumption, defines models, specifies evaluation
Results and conclusions	Presents findings with interpretation; acknowledges limitations

Table 5.6: Standard structure of a research paper in this field

The characterisation section exists to support **reproducibility**: by describing precisely how data was collected, the authors enable another researcher to verify the findings.

5.9 LET US SUM UP

- Location-based social networks attach geographic information to social sharing. Examples include Foursquare, Yelp, Gowalla, Facebook and Twitter.
- Location disclosure exposes not only present position but, cumulatively, routine, residence, workplace, interests and absence from home.
- In a large Indian survey, 39 per cent had location visible to everyone and only 3.1 per cent had customised the setting; just over half considered regional-language call announcements privacy invasive.
- On Foursquare, check-ins are visible only to friends, but mayorships, tips and dones are public and anchored to venues with mandatory addresses—a public information leakage surface.
- Gamification concentrates public activity around where users live and work, which is what makes the inference possible.
- Using majority voting on 13.57 million users, about 78 per cent of home cities were inferred within 50 kilometres; home state about 75 per cent and home country over 90 per cent.

- Withholding a field does not protect the information in it, if that information can be reconstructed from other fields the user chose to share.

5.10 CHECK YOUR PROGRESS

Descriptive Type Questions

1. Distinguish between a location-based service and a location-based social network. Name four platforms and describe the location functionality of each.
2. Explain why a *sequence* of location disclosures is more privacy-sensitive than a single disclosure, identifying at least four categories of derivable information.
3. Present the survey findings on reported Facebook location settings. What does the phrase “to the best of your knowledge” imply about how these figures should be read?
4. Describe the regional-language call announcement scenario and explain why it is an example of indirect disclosure.
5. Define check-in, venue, mayorship, tip and done. Explain how gamification created an incentive structure that concentrated public activity near users’ homes.
6. What is an information leakage surface? Using the visibility asymmetry between check-ins and mayorships, explain how one arises from ordinary design decisions.
7. Explain the majority-voting method and the three-class grouping. Why did the researchers choose a simple method rather than machine learning?
8. The All model was less accurate than the Mayorship model yet was described as better overall. Explain this with reference to accuracy and coverage.
9. State the principal finding of the study and explain why home country could be inferred far more accurately than home city.

Multiple Choice Questions (MCQs)

1. Which best describes a location-based social network?

   a) Any application that uses GPS to provide driving directions

   b) A social network in which geographic information is attached to shared content

   c) A private network accessible only within one geographic region

   d) A network of physical servers distributed across cities

   Answer: b

2. On Foursquare, a mayorship is awarded to the user who:

a) Created the venue page in the system

b) Posted the first tip at the venue

c) Was the most frequent visitor in the last 60 days

d) Has the most friends who visited the venue

Answer: c

3. Which Foursquare attribute was shared only with friends rather than publicly?

a) Mayorships

b) Tips

c) Dones

d) Check-ins

Answer: d

4. What share of survey respondents reported their Facebook location was visible to everyone?

a) 3.1 per cent

b) 12.5 per cent

c) 33.0 per cent

d) 39.0 per cent

Answer: d

5. Approximately how many Foursquare users were in the study’s dataset?

a) 1.3 million

b) 13.5 million

c) 135 million

d) 1.35 billion

Answer: b

6. Class 2 consisted of users who:

a) Had exactly one mayorship, tip or done

b) Had multiple activities with one predominant location

c) Had multiple activities with no single predominant location

d) Had left their home city field blank

Answer: c

7. Which single attribute gave the highest home city accuracy?

a) Done

b) Tip

c) Mayorship

d) Check-in

Answer: c

8. The principal finding was that about 78 per cent of home cities could be inferred within:

a) 5 kilometres

b) 20 kilometres

c) 50 kilometres

d) 500 kilometres

Answer: c

9. Home country was inferred more accurately than home city because:

a) Country names are shorter and easier to process

b) Users rarely leave their country, so activity is concentrated within it

c) Foursquare required users to declare their country

d) Venues do not record city-level information

Answer: b

10. Applying game design elements such as badges and titles to non-game activities is called:

a) Geo-fencing

b) Gamification

c) Crowdsourcing

d) Geo-coding

Answer: b

Fill in the Blanks

1. A social network in which geographic information is attached to shared content is called a _____.

Answer: location-based social network (LBSN)

2. On Foursquare, a page representing a real physical location is called a _____.

Answer: venue

3. A mayorship is awarded to the most frequent visitor over the preceding _____ days.

Answer: 60

4. The inference method that takes the most frequently occurring location among a user's public attributes is called _____.

Answer: majority voting

5. The set of publicly exposed attributes from which unintended inferences can be drawn is the information _____ surface.

Answer: leakage

6. About _____ per cent of analysed users had their home city inferred within 50 kilometres.

Answer: 78

7. India's _____ Act, 2023 establishes consent requirements for processing personal data, including location data.

Answer: Digital Personal Data Protection (DPDP)

5.11 REFERENCES

1. Pontes, T., Vasconcelos, M., Almeida, J., Kumaraguru, P., & Almeida, V. (2012). We know where you live: Privacy characterization of Foursquare behavior. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)* (pp. 898–905). Pittsburgh, PA: ACM. https://doi.org/10.1145/2370216.2370419

2. Kumaraguru, P. (n.d.). *Privacy and Security in Online Social Media* [NPTEL online course, Week 9.1: Privacy in Location Based Social Networks, Part 1]. Department of Computer Science and Engineering, Indian Institute of Technology Madras.

3. Vasconcelos, M., Ricci, S., Almeida, J., Benevenuto, F., & Almeida, V. (2012). Tips, dones and to-dos: Uncovering user profiles in Foursquare. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining (WSDM '12)*. Seattle, WA: ACM.

4. Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011). Tweets from Justin Bieber's heart: The dynamics of the location field in user profiles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Vancouver, Canada: ACM.

5. Vicente, C. R., Freni, D., Bettini, C., & Jensen, C. S. (2011). Location-related privacy in geo-social networks. *IEEE Internet Computing, 15*(3), 20–27.

6. Zheng, Y. (2011). Location-based social networks: Users. In Y. Zheng & X. Zhou (Eds.), *Computing with Spatial Trajectories* (pp. 243–276). New York: Springer.

7. Ministry of Electronics and Information Technology, Government of India. (2023). *The Digital Personal Data Protection Act, 2023* (No. 22 of 2023). New Delhi: Government of India.

Note on sources: The dataset composition, validity statistics, inference models and accuracy figures in Section 5.7 are drawn from Reference 1 above. The survey findings in Section 5.5 are drawn from the NPTEL lecture at Reference 2. All figures describe the platforms as they existed at the time of the original studies and should be read as illustrative of the underlying inference mechanisms rather than as current platform statistics.

UNIT- 6
INFERRING HOME LOCATION IN SOCIAL NETWORKS

Unit Structure

6.1 OBJECTIVES

After successful completion of this unit, you will be able to—

- Explain why extending a location inference study across multiple social networks strengthens its conclusions.
- Describe the datasets collected from Foursquare, Google+ and Twitter, and identify what each contributes to the inference.
- Distinguish between valid unambiguous, valid ambiguous, non-geographic and empty location information.
- Describe the inference models used and interpret the accuracy results.
- Explain the reasoning behind the refined friend model.
- Distinguish home city inference from home residence inference and interpret the error distance distributions.
- Assess the implications of cross-platform inference for users and platform designers.

6.2 INTRODUCTION

The previous unit examined a study that inferred the home city of Foursquare users from their public mayorships, tips and dones. It left one question open: was the result a peculiarity of Foursquare's gamified design, or does it reflect something general about how people share location information? This unit examines the study that answers it. Pontes and colleagues extended the analysis across **Foursquare, Google+ and Twitter**, three platforms differing substantially in purpose, design and the location information they expose.

The headline result is stated at the outset: home city was inferred with accuracies of about **67 per cent for Foursquare, 72 per cent for Google+ and 82 per cent for Twitter**. As you work through the method, keep asking why Twitter should be the most vulnerable. The answer is instructive about the relationship between data precision and privacy risk.

6.3 FROM SINGLE-PLATFORM TO CROSS-PLATFORM INFERENCE

A finding established on one platform is open to the objection that it is an artefact of that platform's design. Foursquare invited exactly this: its mayorship mechanism rewards repeated visits to the same venue, so it would be unsurprising if activity clustered around users' homes. Extending the study to platforms that work differently answers the objection. Google+ has no check-in mechanism, drawing its signal from profile fields; Twitter has no

venue concept, drawing its signal from coordinates attached to posts. If home location can be inferred on all three, the finding concerns location disclosure generally.

Platform	Location-relevant attributes	Nature of signal
Foursquare	Mayorships, tips, likes (dones), friends	Venue-anchored; each points to a place with an address
Google+	Places lived, address, education, employment, friends	Profile-declared; cities and institutions stated by the user
Twitter	User location field, geo-tagged posts	Coordinate-based; exact latitude and longitude

Table 6.1: Location-relevant attributes across the three platforms

These signals differ in kind. A mayorship is *behavioural* evidence of what the user did; an education entry is *declarative*; a geo-tagged tweet is *positional*. Their precision differs greatly, and precision translates directly into inference accuracy.

6.4 THE THREE DATASETS

Platform	Dataset composition
Foursquare	13 million users; 16 million venues; 15 million mayorships; 11 million tips; 10 million likes
Google+	27 million profiles crawled; 7 million declaring a place lived; 7 million declaring education; 6 million declaring employment; 5,000 giving a full address
Twitter	120 million tweets from 20 million users (April–June 2012); 0.5 per cent geo-tagged, giving about 700,000 tweets from 300,000 users

Table 6.2: Composition of the three datasets

The Foursquare data is the collection used in Unit 5. Google+ was tied to the Google account system, so a Gmail account holder effectively held a profile; its fields are completed to be found by former classmates or to signal professional standing. Almost nobody gives a street address, but seven million declaring where they have lived is a large geographic signal.

The Twitter figures show an important asymmetry. Only half of one per cent of posts carried geographic tags, so most users do not enable the feature—but those tags are of entirely different quality, carrying the exact latitude and longitude of the device rather than a city name requiring interpretation. This trade-off, **few users but extremely precise data**, shapes the Twitter results throughout.

6.5 QUALITY OF GEOGRAPHIC INFORMATION

Before inference can be attempted, raw location strings must be assessed, since users type whatever they wish and much of it is unusable.

Category	Meaning	Example
Valid unambiguous	Resolves to exactly one real place	“New Delhi”; a coordinate pair
Valid but ambiguous	A real place, but which one is unclear	“near Taj Mahal”; a city name shared by several places
Non-geographic	Not a place at all	“in your heart”; a number or email address
Empty	Field left blank	—

Table 6.3: Classification of location field contents

For the Foursquare home city field, about 95 per cent of entries were valid and unambiguous, 2.6 per cent ambiguous, 1.8 per cent non-geographic and 0.2 per cent empty: users are, on the whole, honest about where they live. Two further figures are instructive. **Twitter geo-tags were essentially 100 per cent valid and unambiguous**, which follows from what a geo-tag is—a coordinate pair cannot be ambiguous or facetious. **Google+ education was about 53 per cent valid and unambiguous**, since an entry naming a university resolves cleanly to one institution at a known address.

Granularity is consistent across platforms: about 80 per cent of Foursquare users and venues, 79.63 per cent of Google+ users and 62.54 per cent of Twitter users give location at **city** level. City is the granularity users naturally reach for, which is why city-level inference is the natural target.

The distributions of attributes per user are all **heavy-tailed**: most users have very few mayorships, tips or geo-tagged posts while a small minority have many. This is the pattern you have met as the **power law** and **Pareto principle**, recurring so consistently that its absence would be grounds for suspecting the dataset. But sparse data is not weak data: if 94 per cent of Google+ users declare only one place lived, no majority vote is needed—the single entry *is* the answer, informative because nothing contradicts it.

6.6 THE INFERENCE METHODOLOGY AND RESULTS

The methodology follows Unit 5. **Class 0** users have one location vote, **Class 1** users have multiple votes with one predominant location, and **Class 2** users have none standing out

and cannot be handled. Accuracy is the fraction of correct inferences among Class 0 and Class 1 users.

Separate single-attribute models isolated each attribute's contribution: Mayorship, Tip, Like and Friends for Foursquare; Friends, Education and Employment for Google+; and geo-tagged location for Twitter, the only attribute available.

The **friends model** differs in kind. It infers a home city not from anything the user disclosed but from the declared locations of their connections, resting on homophily—the tendency of people to be connected to others living near them. This is the most privacy-significant model of the set, because a user's location can be estimated even if they have disclosed nothing. Their privacy depends on their friends' behaviour.

Platform	Best-performing model	Home city accuracy
Foursquare	Mayorship (Class 1)	67%
Google+	Friends, without refinement	72%
Twitter	Geo-tagged location	82%

Table 6.4: Summary of home city inference accuracy

Using mayorships alone, about 51.61 per cent of Class 0 and about 67 per cent of Class 1 Foursquare users were correctly identified. The gap is instructive: one mayorship gives roughly even odds, while several concentrated in one place give a genuine signal. For Google+ the friends model performed best, confirming homophily as a strong predictor. Twitter was the most accurate, and the reason is now apparent: **the data is a coordinate**—there is no interpretation step and no ambiguity to resolve.

> **The Central Lesson on Precision**
>
> Twitter was the most vulnerable platform despite only 0.5 per cent of its posts carrying location data. Precision beats volume: a user who geo-tags ten posts discloses more usable location information than one who writes a hundred tips at named venues. Ask not only *how much* location data you have shared, but *how precise* each item is.

6.7 THE REFINED FRIEND MODEL

The friends model was improved by filtering. Not every friend is equally informative: users with fewer than a minimum number of friends, $k_{\min}$, provide too little evidence, while users with more than a maximum, $k_{\max}$, are unlikely to have strong relationships with all their connections. The refinement excludes both.

Filtering improved accuracy unevenly. **For Google+ the gain was about 21 per cent**, while for Foursquare it was smaller, because friendship means different things on the two platforms. A Google+ connection could be a classmate in another city or a contact met online, with geography playing little part. Foursquare connections are already anchored in shared physical places, leaving less noise to remove.

Refinement buys accuracy at the cost of coverage, since excluded users receive no inference at all. Whether the trade is worthwhile depends on purpose: an adversary targeting one individual cares only about accuracy, one profiling a population cares about both.

6.8 ERROR DISTANCES AND HOME RESIDENCE INFERENCE

Reporting accuracy alone treats every error as equally bad, which for location inference is misleading—inferring that a user in Ahmedabad lives in Gandhinagar differs entirely from inferring Kolkata. The distance between inferred and declared city was therefore computed for every incorrect inference.

Platform	Error distances under 50 km	Accuracy within 50 km
Foursquare	46%	78.5%
Google+	27%	64%
Twitter	27%	87%

Table 6.5: Error distances and accuracy within a 50 kilometre radius

Fifty kilometres is roughly the gap between neighbouring cities and often falls within one metropolitan region, so an error of that size may not be meaningful—the user may live in one city and work in the adjacent one. The analysis was then repeated at a far finer granularity: the user’s **home residence** rather than home city. This is substantially more demanding, and the results diverge sharply.

Platform	Exact residence	Within 5 km	Within 20 km
Twitter	35%	—	73.67%
Foursquare	—	52.73%	77%
Google+	5.23%	—	—

Table 6.6: Home residence inference accuracy by platform

Each row is explained by the underlying data. For **Twitter**, 35 per cent of residences were inferred *exactly* and 73.67 per cent within 20 kilometres, since geo-tagged tweets carry exact coordinates and people tweet a great deal from home. For **Foursquare**, 52.73 per cent fall within 5 kilometres—venue data is precise, but the venue is a cafe or station near the home rather than the home. For **Google+**, only 5.23 per cent of exact residences could be inferred, which is expected since education and employment locate where a person studies or works, not where they sleep.

> **Matching the Attribute to the Question**
>
> Google+ performed well at home *city* inference (72 per cent) and poorly at home *residence* inference (5.23 per cent), using the same data. The informativeness of an attribute is not fixed; it depends on the question asked. An education record locates you within a city reliably, and says nothing about which street you live on.

6.9 IMPLICATIONS FOR USERS AND PLATFORMS

The cross-platform design settles the question the earlier study left open. Home location can be inferred with high accuracy on three dissimilar networks, using behavioural, declarative and positional evidence. The vulnerability is not an artefact of any one platform's design; it is a general property of how people share location.

For **users**, four conclusions follow. Precision matters more than volume, so disabling geo-tagging is among the most effective measures available. Profile fields such as education and employment function as geographic identifiers. Privacy depends partly on others, since the friends model locates a user through their connections. And withholding a field does not protect its contents, because aggregation is the mechanism of harm.

For **platform design and regulation**, defaults carry great weight, since whether geo-tagging is opt-in or opt-out determines the exposure of the majority who never change one. Public APIs enable bulk inference, making rate limits a privacy control rather than merely an operational one, and granularity can be reduced at the point of sharing. Finally, inferred data raises a live question for consent: under India's Digital Personal Data Protection Act, 2023, where a home location is *derived* from attributes shared for a different purpose, whether the original consent extends to it remains unsettled.

6.10 LET US SUM UP

- The study extended location inference to Foursquare, Google+ and Twitter, establishing that the vulnerability is general rather than platform-specific.

- Home city was inferred with accuracies of about 67 per cent for Foursquare, 72 per cent for Google+ and 82 per cent for Twitter.
- The datasets comprised roughly 13 million Foursquare users, 27 million Google+ profiles and 120 million tweets from 20 million users, of which 0.5 per cent were geo-tagged.
- Twitter geo-tags were essentially 100 per cent valid and unambiguous; Google+ education about 53 per cent. All attribute distributions were heavy-tailed.
- The refined friend model improved accuracy by about 21 per cent for Google+ but less for Foursquare, whose friendships are already geographically anchored.
- Allowing a 50 kilometre radius, accuracy reached 78.5 per cent for Foursquare, 64 per cent for Google+ and 87 per cent for Twitter.
- For home residence, Twitter inferred 35 per cent exactly and 73.67 per cent within 20 kilometres; Foursquare 52.73 per cent within 5 kilometres; Google+ only 5.23 per cent exactly, since education and employment are the wrong evidence.
- Precision of disclosure matters more than volume, and a user's privacy depends partly on the disclosure behaviour of their connections.

6.11 CHECK YOUR PROGRESS

Descriptive Type Questions

1. Explain why extending the study from Foursquare alone to three platforms strengthens its conclusions. What specific objection does the extension answer?
2. Describe the location-relevant attributes contributed by each platform, classifying each as behavioural, declarative or positional evidence.
3. Distinguish between valid unambiguous, valid ambiguous, non-geographic and empty location information, giving an example of each.
4. Only 0.5 per cent of collected tweets carried geographic tags, yet Twitter produced the highest accuracy of the three platforms. Explain this apparent paradox.
5. What is a heavy-tailed distribution? Explain why such distributions recur in social media data and why their presence supports confidence in a dataset.
6. Explain the friends model and the assumption of homophily. Why is it the most privacy-significant of the models examined?

7. Describe the refined friend model, including the roles of k_{min} and k_{max}. Why did refinement help Google+ much more than Foursquare, and what trade-off does it involve?

8. Why is it insufficient to report accuracy without also reporting error distances? Illustrate with reference to the 50 kilometre threshold.

9. Google+ achieved about 72 per cent for home city but only 5.23 per cent for exact residence. Explain this difference and the general principle it illustrates.

10. Discuss the implications of these findings for individual users and for platform design, including the question of whether consent to share an attribute extends to attributes derived from it under India's DPDP Act, 2023.

Multiple Choice Questions (MCQs)

1. The three social networks examined in this study were:

a) Facebook, Instagram and LinkedIn

b) Foursquare, Google+ and Twitter

c) Yelp, Gowalla and Brightkite

d) WhatsApp, Telegram and Signal

Answer: b

2. The home city accuracies for Foursquare, Google+ and Twitter respectively were approximately:

a) 82, 72 and 67 per cent

b) 67, 72 and 82 per cent

c) 51, 64 and 78 per cent

d) 46, 27 and 27 per cent

Answer: b

3. What proportion of collected tweets carried geographic tags?

a) 0.5 per cent

b) 5 per cent

c) 35 per cent

d) 50 per cent

Answer: a

4. An entry reading “near Govindpuri Metro Station” would be classified as:

a) Valid unambiguous geographic information

b) Valid but ambiguous geographic information

c) Non-geographic information

d) An empty entry

Answer: b

5. Geo-tagged tweets were valid and unambiguous in approximately what proportion of cases?

a) 53 per cent

b) 72 per cent

c) 95 per cent

d) 100 per cent

Answer: d

6. A distribution in which most users have very few of an attribute while a few have very many is described as:

a) Uniform

b) Normal

c) Heavy-tailed

d) Bimodal

Answer: c

7. Class 2 users were excluded from the accuracy calculation because:

a) They had deleted their accounts

b) They had no predominant location, so no inference could be made

c) They had opted out of the study

d) Their data fell outside the study period

Answer: b

8. The friends model infers a user’s home city from:

a) The user’s own geo-tagged posts

b) The declared locations of the user’s connections

c) The IP address from which the user logs in

d) The language in which the user posts

Answer: b

9. Refinement of the friends model improved accuracy for Google+ by approximately:

a) 5 per cent

b) 12 per cent

c) 21 per cent

d) 45 per cent

Answer: c

10. Allowing a 50 kilometre radius, accuracy for Twitter was approximately:

a) 64 per cent

b) 73 per cent

c) 78.5 per cent

d) 87 per cent

Answer: d

11. For what proportion of Twitter users was the exact home residence inferred?

a) 5.23 per cent

b) 35 per cent

c) 52.73 per cent

d) 73.67 per cent

Answer: b

12. Google+ inferred exact home residence for only 5.23 per cent of users because:

a) The Google+ dataset was too small

b) Education and employment locate where a person studies or works, not where they live

c) Google+ did not permit data collection

d) Google+ users deliberately entered false information

Answer: b

Fill in the Blanks

1. The assumption that people tend to be connected to others who live near them is called _____.

Answer: homophily

2. A distribution in which most users have small values and a few have very large values is described as _____.

Answer: heavy-tailed

3. In the refined friend model, users with fewer than _____ friends are excluded from the inference process.

Answer: k_{min}

4. Approximately _____ per cent of Twitter users had their exact home residence inferred at a distance of zero.

Answer: 35

5. For Foursquare, about 52.73 per cent of residence inferences fell within _____ kilometres.

Answer: 5 (five)

6. The requirement that another researcher collecting comparable data should obtain comparable results is called _____.

Answer: reproducibility

7. Twitter geo-tags were essentially _____ per cent valid and unambiguous, because a coordinate pair cannot be ambiguous.

Answer: 100

6.12 REFERENCES

1. Pontes, T., Magno, G., Vasconcelos, M., Gupta, A., Almeida, J., Kumaraguru, P., & Almeida, V. (2012). Beware of what you share: Inferring home location in social networks. In *Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW)* (pp. 571–578). Brussels, Belgium: IEEE.

2. Pontes, T., Vasconcelos, M., Almeida, J., Kumaraguru, P., & Almeida, V. (2012). We know where you live: Privacy characterization of Foursquare behavior. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp ’12)* (pp. 898–905). Pittsburgh, PA: ACM. https://doi.org/10.1145/2370216.2370419

3. Kumaraguru, P. (n.d.). *Privacy and Security in Online Social Media* [NPTEL online course, Week 10.1: Beware of What You Share — Inferring Home Location in Social Networks]. Department of Computer Science and Engineering, Indian Institute of Technology Madras.

4. Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: Improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th International Conference on World Wide Web (WWW ’10)* (pp. 61–70). Raleigh, NC: ACM.

5. Mahmud, J., Nichols, J., & Drews, C. (2012). Where is this tweet from? Inferring home locations of Twitter users. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM ’12)*. Dublin, Ireland.

6. Mislove, A., Viswanath, B., Gummadi, K. P., & Druschel, P. (2010). You are who you know: Inferring user profiles in online social networks. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM '10)* (pp. 251–260). New York, NY: ACM.

7. Vicente, C. R., Freni, D., Bettini, C., & Jensen, C. S. (2011). Location-related privacy in geo-social networks. *IEEE Internet Computing, 15*(3), 20–27.

8. Ministry of Electronics and Information Technology, Government of India. (2023). *The Digital Personal Data Protection Act, 2023* (No. 22 of 2023). New Delhi: Government of India.

Note on sources: The dataset descriptions, geographic information quality figures, inference models, accuracy results, refinement gains and residence inference distances in this unit are drawn from Reference 1 above and from the NPTEL lecture at Reference 3, which discusses that study. The figures describe the three platforms as they existed at the time of the original analysis. Google+ has since been discontinued for consumer accounts and Twitter now operates as X; the figures should be read as illustrative of the underlying inference mechanisms rather than as current platform statistics.