CHAPTER 2: DEEP WEB

Compared to the Surface Web and the Dark Web, the Deep Web is the most topographically complex of the three cyber domains, both in terms of the technology used to create the Deep Web and the way people use it. Although easily accessible with the web browsing software that is used to access the Surface Web, finding the valuable information that is locked within the Deep Web is more of an art than the science of Surface Web mining. A good way to conceptualise the investigative approach to the Deep Web is that information within this layer of cyberspace that is relevant to an investigation is not hidden; it's just slightly out of sight.

Successfully mining the Deep Web can provide the most immediate gains in terms of an increased volume of raw data. This data can prove valuable for the OSINT investigator involved in any work that deals with people or groups as the focus of the investigation. To borrow from a law enforcement term, OSINT investigation within the Deep Web layer of the Internet typically has a focus on a 'person of interest,' a generic term for an individual who may become a suspect or is somehow closely linked to an investigation. The remainder of this chapter carries forward this concept, with the objective of equipping the investigation with enhanced tools to mine the Deep Web.

The Deep Web and social media technology

If there is one web technology that has driven the growth of the Deep Web it is social media technology. Platforms such as the two giants, Facebook and Twitter, turned practices such as 'tweeting'[38], 'trolling'[39], 'spamming'[40], and 'flaming'[41] into global pastimes. The author cannot overstate the importance of social media to the growth of the Deep Web; Facebook alone reached more than one billion users by about 2013[42]. With the current world population assessed at about seven billion, Facebook user rates mean that incredibly more than one-sixth of the entire world population has a Facebook account. This unprecedented representation of the global human population has been facilitated largely by the flexibility of the technology that social media is constructed from. With the number of Internet users predicted to reach more than five billion by 2020[43], the importance of social media to OSINT professional practice will only grow.

The difference between social media technology and social media platforms

A social media platform is the term used for a specific social media service provider, e.g. Facebook, LinkedIn, Bebo and so on. Social media technology is a generic term used to encapsulate all the different platforms that are currently contributing to the huge growth of the 'social web', i.e. the web content on the Internet being generated solely on social media platforms.

Social media technology is built to grow on a huge scale, with Facebook and similar providers building vast server farms to hold the millions of new profiles created by users,

[38] When a user posts a message on the Twitter platform.

[39] Deliberately posting provocative messages to lure new users to a forum into posting

[40] Flooding a user or forum with a deluge of messages.

[41] Deluge of aggressive messages.

[42] *http://thenextweb.com/facebook/2014/01/29/facebook-passes-1-23-billion-monthly-active-users-945-million-mobile-users-757-million-daily-users/*.

[43] *www.networkworld.com/article/2238913/wireless/10-fool-proof-predictions-for-the-internet-in-2020.html*.

who require no programming skills to design and deploy a website. Given the fact that social media is virtually unregulated, it is obvious that the Deep Web can prove to be a treasure trove for the skilled investigator.

Who watches the watchers?

Fact: did you know that if you collected together all the employees of Facebook, Twitter and LinkedIn, past and present, they would comfortably fit into Madison Square Garden? This fact is surprising given that these three services collectively hold accounts and personal data of over one-fifth of the entire world population. Given the disparity in the number of service users compared to the number of service providers, it becomes obvious that there is no way that service providers like Facebook can track the content of each post and profile on their platform. This explains the sometimes illegal and extreme content of many social media platforms.

A standard rule is that all content posted to any social media platform is uniquely attributable to an individual account, and almost all social media accounts track back to one identifiable individual. Taking this point into account, and the fact that social media is virtually unregulated, if used effectively then information derived from social media (SOCMINT, to borrow an acronym from Sir David Omand[44] *et al*) can provide a direct feed of primary source information to the skilled investigator.

Although investigation within the social media web space may be a relatively new practice, attempting to understand the collective behaviour patterns of users and ways to interpret them is a well-understood process for some professions. Marketing professionals in particular tend to have an innate understanding about the effects of social

[44] Omand, D, Bartlett, J, Miller, C. (2012). #INTELLIGENCE. Demos. Available from *www.demos.co.uk/files/_Intelligence_-_web.pdf?1335197327*

media and its influence on behaviour. Of the many excellent works within this field the book *33 Million People in the Room*[45] states three core principles concerning social media that are particularly relevant for the Internet investigator:

1. Different types of (social) networks exist for different audiences and different purposes, and each ***is a microcosm unto itself:***
   a. *Relevance to the investigator*: most persons of interest will have accounts on more than one social media platform to reflect different aspects of their personal and professional lives, e.g. Facebook for friends and family, LinkedIn for professional advancement.
2. There is no one size fits all solution to social networking and each social network is created with ***different users and uses*** in mind:
   a. *Relevance to the investigator*: different strata of society use different social media platforms for different purposes, e.g. Bebo is targeted at the teen and young adult market, whereas LinkedIn is intended for the established professional market. The point of this is that the investigator can expect different persons of interest to have different social media accounts depending upon factors such as age, gender, social status and profession.
3. A ***fundamental understanding*** of the differences between networks is key to making the best use of the tools:
   a. *Relevance to the investigator*: think of each social media platform as a unique self-contained world

[45] Powell, J. (2009). *33 Million People in the Room: How to Create, Influence, and Run a Successful Business with Social Networking*. Financial Times Press; 1st edition (February 10, 2009)

within itself, with a distinct set of rules and customs that any new interloper is required to learn to be effective within that environment. This point has specific repercussions for aspects of an investigation such as security; some social media platforms such as LinkedIn allow account holders to see who has viewed their account whereas others such as Facebook do not. Knowing the suitable technical nuances between social media platforms distinguishes the seasoned OSINT professional from the more novice Internet investigator.

To highlight some of these points in more depth, shown below are several of the main social media platforms:

Platform name	**Platform description**	**Intended audience**	**Data validation rules?**	**Number of users (2014)/yearly growth rate (where available)**	**Target language/ regional focus**	**Interest to the investigator**
Facebook	Text- and image-focused platform but with some more advanced mapping and instant message functionality coming online	Possibly the most generic of all social media platforms in that it is intended to capture the social life of users from the very young to the very old	Some verification of users via phone and external email, but aside from that very little	1,310,000,000 /22% 2012–2013[46]	Predominantly European and North American but rapid penetration into developing regions such as India, Africa and Russia as user base grows	The first port of call for almost all investigations as this shows the personal life and connections of a person of interest

[46] www.statisticbrain.com/facebook-statistics/.

LinkedIn	Similar to Facebook but with a cleaner, more streamlined look (white space is king on this platform). Users have the ability to list résumés on the site and create and join specialist interest groups	Business-minded Professionals	Some verification with regard to types of qualification, e.g. medical degrees from certain schools and certain requirements to create business-specific pages and members groups	313,000,000[47]	European but with growing user groups in the Americas and the developing world	Very useful for investigations involving persons of interest who have to maintain an outwardly respectable persona (typically fraudsters). If profiles for the same individual can be found on both LinkedIn and Facebook then the investigator can often build a very comprehensive picture of that person of interest
Tencent QQ	More of an amalgamation of social media tools such as instant messaging, email-style chat and game playing that has grown out of this platform's original	Technologically minded Chinese youths	Previous requirements for a People's Republic of China identity card number; however, this is no longer needed to gain an account on	200,000,000[48]	Initially deployed in the People's Republic of China, but increasingly popular in the wider Pacific Rim region. An English	Very useful for geostrategic researchers looking to gain access to difficult to reach populations in the Pacific Rim area

[47] www.statista.com/statistics/274050/quarterly-numbers-of-linkedin-members/.

[48] www.techinasia.com/qq/.

	roots as an instant messaging tool		this platform		language version may see the platform gain wider use.	
Twitter	A microblogging site allowing messages or 'tweets' of 140 characters or fewer. Trends are created with the now eponymous 'hash tag' (#)	Technologically minded youth and those used to communicating in shortened 'text speak'-style prose. Breaking through to celebrity and mainstream users	None	271,000,000 active users per month[49] (highly volatile user group, hence uncertainties about true number of users)	Initially English but quickly spreading across many language spheres	This platform is of particular interest to geopolitical analysts due to its role in the so-called 'Twitter Revolutions' and the popularity of this platform with extremists of all persuasions
Deviant Art	A digital art gallery for aspiring artists to post their work and connect with other likeminded artists and fans	Digital art community	None needed; anonymity preferred	1,000,000[50]	Anglophone	Even though 99 percent of the work on this platform is legal, there have been instances of more suggestive works being placed onto the site with the intention of signalling to certain viewers the

[49] https://about.twitter.com/company.
[50] www.quantcast.com/deviantart.com.

						presence of more hardcore images on request
Myspace	Strong focus on music with a social network built around fan groups (co-owned by musician Justin Timberlake)	Teen and young adult market	None	36,000,000[51]	Anglophone	Often used by paedophiles to groom and connect with victims
Tumblr	Somewhat like Twitter but longer-form blogs including images	Teen and young adult market with a strong bent to the creative market	None	300,000,000[52]	Multilingual	Popular with protest groups and other semi-legal groups

The preceding list is by no means an exhaustive one of all the current social media platforms; indeed there are several hundred general platforms and thousands of smaller niche-interest platforms on the web. The reason for examining these platforms in particular is that they all reveal a specific aspect of social media technology in general.

The examples of Facebook and QQ underline the principle of cyber geography and how the user group, even for the largest social media platforms, is very specific to both the geographical and linguistic spheres. Platforms such as

[51] *http://expandedramblings.com/index.php/myspace-stats-then-now/#.U_J3CpRdU00*.
[52] *http://allthingsd.com/20130521/how-many-users-does-tumblr-really-have/*.

LinkedIn, when compared to Twitter, show how the different functionality of a platform can affect the types of data that can be available across different social media sites. DeviantArt shows how the scalability and anonymity of social media technology can be used to hide sinister content. MySpace shows how designers of different platforms aim their products at different demographics broken down by factors such as age range, gender and social status. When comparing all the platforms together the different formats of data that social media technology collectively produces (e.g. images for Tumblr, text streams for QQ and so on) becomes obvious.

Like looks for like, and like likes what like finds

This curious phrase encapsulates the way that the Internet facilitates the development of connections between individuals with extremely niche interests and viewpoints. Niche-interest groups can form on mainstream social media platforms as well as bespoke social media sites dedicated to an interest. One such niche-interest site was titled Cannibal Café and services users who fantasised about human cannibalism. It was on this site that Armin Meiwes successfully advertised for a victim to be willingly killed and eaten by Meiwes[53]. The Meiwes case, aka The Internet Cannibal, has now become possibly the most notorious example of the Internet's ability to connect individuals with interests even at the most extreme fringes of normal human behaviour.

By examining even just this one aspect of OSINT, the core themes outlined at the start of this book concerning cyberspace (multilayered, cyber geographies, mixed medium, tangibility) become obvious. The question is: how to make sense of this space and start to effectively investigate it?

[53] *http://news.bbc.co.uk/1/hi/world/europe/3230774.stm*

The core principles of networks and how they affect investigations on Deep Web social media platforms

One of the main drivers behind the incredible success of social media is that it exploits people's natural tendency to form networks based upon shared interests and values.

Within the context of social media technology, networks are created by the connection functionality inbuilt within all social media. The exact form that the connection functionality takes varies between social media platform, e.g. 'Friends' in Facebook, 'Circles' in Google Hangout. However, the symbolism of a connection between users is the same, i.e. that two connected users have a closer relationship than the other unconnected users on the same platform.

Networks form around almost any form of human interest, from the benign to the most extreme forms of criminal and terrorist activities. What social media technology allows is the creation of networks that are free from the traditional constraints on network size such as geography and the number of people another person can canvass for similar interests within a certain period of time. As such, when conducting an investigation within a social media web environment, a key shift in thinking for an investigator accustomed to investigations within the physical world is that within the Deep Web the best approach is to focus on investigating networks as opposed to just focusing purely on the web presence of individuals.

As an example of the necessity to shift away from targeting individuals to targeting networks within the scope of an Internet investigation, consider the following scenario. Imagine you are a geopolitical analyst tasked with assessing the organisational stability of a country's governmental system. One approach you could adopt would be to look for signs of discord and interpersonal tension within the Twitter feeds of the ruling party. The government you are looking at is typical of the Western model of government in that any public or media relations are closely controlled by a centralised media team; as such, the Twitter feeds of the ruling parties' elite (President, First Minister and so on) present a homogenised 'on message' flow of not very insightful content. However, the ministers on the periphery of the party will typically provide a far more representative view of how things really are within the party, as they are free from the tight controls on central party members or lack the staff resources to have someone else manage their social media feeds. This theoretical example of investigating the core of a network by examining the data generated by those individuals on the periphery of that network is a technique used by investigative professionals since well before the dawn of the Internet. However, given that the concept of network building is integral to social media technology, the idea is clearly worth restating.

Look for the weak link: principle in action

Shown in this panel is a self-portrait of Alexander Sotkin, a Russian soldier who posted multiple images of himself on the social media platform Instagram circa July 2014. The images, with accompanying messages, apparently showed Sotkin servicing Russian military mobile surface to air missile platforms, from various locations in Ukraine. The images posted by Sotkin ran completely contrary to official Russian Government claims that there was no large-scale deployment of Russian troops within Ukraine at the time Sotkin was posting the images.

Given the hostilities that were occurring within Ukraine at the time and that the Malaysian Airlines flight MH-17 had been shot down by a surface to air missile system, fired from disputed Ukraine territory on 17 July 2014, Sotkin's photo posts were politically incendiary when they were released.

This example is just one of many of how one individual's actions, facilitated by social media, can compromise the operational security of an entire organisation engaged in a clandestine activity

The two essential principles when seeking to investigate any network are target the weak link to gain access to the network and then laterally move through the network to gain access to the target information or individual:

1. *Target the weak link*: within any network of people (especially one engaged in nefarious activity) there will always be one individual who cannot help posting revealing information about themselves and the activities of the rest of the network, and in so doing compromises the operational security of that group. For the

investigation the challenge is not so much locating this individual,[54] but ascertaining how many degrees of separation the indiscreet individual is from the core person of interest. Once the closeness or distance between individuals within a network has been ascertained, the investigator is then in a position to draw inferences about the relevancy of the data gained from a third-party profile to the core person of interest within an investigation. Think of this stage as finding the initial thread of an investigation.

2. *Move laterally through the network*: the goal of this stage is to move through the target network to reach the final objective of the investigation. This could be as specific as connecting to the profile of a person of interest or as general as moving about a network to find all the connections to certain key persons of interest. Many social media platforms such as LinkedIn present barriers to unrestricted lateral movement, meaning that users can only form connections with other users if they are within a couple of degrees of separation. These technological barriers necessitate a 'softly softly' approach to movement within a network as the investigator connects to users' profiles that allow further penetration into the target network. In addition to the technological challenges, there are a couple of methodological issues that the investigator needs to be mindful of when moving through a network, namely which connection is a pathway deeper into the network

[54] I have successfully investigated networks involved in drug running, contract killing, sectarian violence, terrorism, espionage and prostitution rings, and in every case there has always been at least one member of that network who has been so indiscreet about themselves and their co-conspirators that I have been able to worm into the network and fully illuminate it. The 'weak link' is always a fixture of any network.

(good) and which connection will lead the investigator out of the network and into a dead end (bad). This process of assessing who is connected to who and for what reason requires a level of judgement from the investigator, as all people active on a social media platform are ultimately all connected to one another if the network is expanded wide enough[55].

Theory into practice

Just as the meta search engines discussed in *Chapter 2* allowed the investigator to throw a wider net over the Surface Web than merely going directly to single-source search engines as a first port of call, so there are a number of useful meta search engines that allow an investigator to query multiple social media platforms:

Tool name[56]	**Freeware?**	**Social media platform(s) tool focuses on**	**Functional description**
Social Searcher[57]	Yes	Facebook, Twitter and Google+	Provides a clear side-by-side display for search terms from Facebook, Twitter and Google+

[55] Frigyes Karinthy developed a theory that every person on the planet is connected by six or fewer degrees of separation. Although the theory has been challenged and refined over the years, the basic principle is that everyone can be connected if you expand their personal network enough. The risk this poses to any investigation is that false links can be created between people who are only distantly related.

[56] Due to the volatile nature of the technology that is used to build social media sites, many of the preceding social media meta search engines are often offline as the makers of the tools struggle to keep up with the rapidly changing technology used by social media companies. The solution for the investigator is not to rely on one tool but to develop skills in a number of tools.

[57] *www.social-searcher.com/*.

TweetDeck[58]	Yes	Twitter only	Allows the investigator to view the content of multiple accounts simultaneously. Additionally, this tool now runs within the browser, therefore not requiring an install of specific software that may be barred within some public sector organisations
Social Mention[59]	Yes	Does not specify but would appear to pull from a broad range of social media tools	Very useful for keyword monitoring of trending topics. Also has the useful function of allowing an export of search results to CSV/Excel format
IFTTT[60] (If This Then That)	Yes	Facebook, LinkedIn and Email[61]	A very powerful tool that lets users specify actions based on conditions that occur in data: If [condition occurs] then [take this action]. Can be used for a number of useful purposes such as searching Facebook for the occurrence of certain images and so on
One Million Tweet Map[62]	Yes	Twitter only	World spot map view of Twitter posting. Has the ability to allow users to filter Tweets based on postcode and keywords

Aside from increased coverage of the Deep Web that the preceding meta engines provide, the tools capture a unique feature of the way people use social media technology in general. As *33 Million People in the Room* states, different social media platforms exist for different purposes and this by implication leads to the conclusion that a single person of interest will have accounts on multiple social media

[58] https://tweetdeck.twitter.com/.

[59] www.socialmention.com/.

[60] https://ifttt.com/wtf.

[61] Many social media platforms will offer updates via email. As such, IFTTT can be used to filter these incoming emails.

[62] http://onemilliontweetmap.com/.

platforms, all fulfilling different social needs (Facebook for family and friends, LinkedIn for professional contacts and so on). By using social media meta search engines the investigator is given a far more strategic view of an individual's web presence, and crucially, this allows the investigator to cross-reference different profiles on different social media platforms to build up a more coherent picture of a person of interest's digital profile.

> ***"So social media is the Deep Web?"***
>
> Well, no, not entirely. Although social media technology creates a huge amount of Deep Web content, much of the Deep Web is made up of content generated from non-social media platforms. The important point is that Deep Web content is not searchable by mainstream search engines, irrespective of whether this content is generated by social media platforms or some other form of web publishing mechanism.

Platform-specific search tools

Separate from the tools that aggregate content from multiple social media platforms are the tools that can focus solely on just one. The advantage of these tools over the meta tools is that although both classes examine the same content, one-platform specialist tools often retrieve much more metadata that can be analysed in creative ways. Data derived via specialist tools can be used to generate insight such as the time zone that the user of an account is based within, the average distribution of posts over a given time period and even the psychological profile of a given user. Used appropriately and dependent upon the context of the investigation, these facts can provide great insight into the habits and behaviour of individual users that can prove useful to the OSINT researcher.

What can be gained from metadata analysis?

As part of a research project looking at social media use by extremist groups active in Sub-Saharan Africa, the Author gathered a large number of statistics derived from the metadata attached to more than 150 extremist websites. Two pieces of data that were collected for every Twitter account within the study's dataset was the number of followers of the site and the number of sites that account was following. Analysis of the data showed an interesting relationship to the follower/following ratio, which for the standard Twitter user is on average about 1 to 1, i.e. for every one person who follows an account, that account will either reciprocate the follow or find someone else to follow. The ratio for the extremist Twitter accounts was typically 0 to 1, i.e. they followed no one even when they themselves often had thousands of followers. This observation, coupled with the fact that the extremist Twitter accounts almost never retweeted someone else's posts, indicated that extremists use the Twitter service as a broadcast medium for propaganda rather than a place for conversation and debate that the service was intended to provide. The full report can be read at this link[63]. However, for the purposes of this book the key point to understand is that this kind of insight can only be gained by taking an analytical step back from the raw data and by comparing data across multiple accounts. By doing this, patterns and deeper nuanced analysis typically emerges.

The following are specialist tools that can be applied to various social media platforms:

Tool name	Freeware?	Functional description	Social media platform(s) tool focuses on	Situations where this tool is useful for the investigator
NodeXL[64]	No charge for tool, but user must have a copy of Microsoft Excel (2007 or 2010) and Windows (XP, Vista or 7) for	Focuses on mapping the social networks between users across various social media platforms	Can retrieve network data from Twitter, YouTube, Flickr, email (.pst files), Facebook, Exchange, Wikis and WWW hyperlinks (easily the most	Useful for mapping special-interest groups that span across social media platforms, e.g. protest groups and so on

[63] Bertram, S and Ellison, K. (2014). Sub Saharan Africa Terrorist Groups' Use Of Internet. Journal of Terrorism Research Volume 5 Issue 1. Available from http://ojs.st-andrews.ac.uk/index.php/jtr/article/view/825/704.

[64] http://nodexl.codeplex.com/.

	the add-on to work		comprehensive tool featured within this chapter)	
Lococitato[65]	Users charged for software but very modest fee (<£100) for download	Focused purely on mapping Twitter relationships, including follower relationships and retweets	Twitter only (makers do produce an equivalent piece of software that maps Facebook networks; however, this is only available to law enforcement)	Mapping the flow of memes between Twitter users as well as the day-to-day work of looking at who is following who
Maltego Tungsten[66]	Users charged for software	A very powerful tool for the advanced user allows automated analysis of network infrastructure as well as many features around social media platforms. Comes complete with own easy to use scripting language	Twitter, YouTube and Facebook are all included along with a huge number of functions that can be applied to Surface Web sites	Particularly useful for an investigation involving a large crossover between Surface Web and Deep Web platforms
TweetPsych[67]	Free web-based tool	Draws up a psychological profile of a Twitter user based on their tweets[68]	Twitter only	Useful for a strategic view on the sentiment in a user's tweets e.g. positive, negative and so on
Snap Bird[69]	Free web-based tool	Retrieves all a user's tweets, even ones posted to a separate user profile	Twitter only	Very useful for providing a complete picture of a user over an extended period of time. A

[65] www.lococitato.com/.
[66] www.paterva.com/web6/.
[67] http://tweetpsych.com/.
[68] Many professional psychologists would challenge this tool's ability to construct an accurate profile on an individual based on such limited data. I leave the final decision over its validity to your judgement.
[69] http://snapbird.org/

				specific tool for use on profiles that are identified as belonging to persons of interest
TweetStats[70]	Free web-based tool	Provides a statistical analysis of a user's Twitter activity based on time of day, date, posting frequency and so on	Twitter only	Useful when considering attribution of a Twitter account to a specific time zone or user demographic, e.g. if all posts are Monday to Friday 9 am to 5 pm US Pacific time, it is a fair assessment to make that the user is based in Pacific America and is posting from work

These tools give a small flavour of the types that can be applied to various social media platforms. Although there is a focus on Twitter in the preceding list, new tools that can be used on other platforms are published almost daily.

The important point to take away from these tools is that social media platform-specific tools – when used correctly – can give a far more granular picture of a specific social media account than the 'top-down' strategic view that more generic tools provide. Finding and effectively applying these tools is largely up to the investigator, as any published list is out of date almost as soon as it is published due to the rapid pace of social media technology in general.

[70] www.tweetstats.com/

The importance of identity online

One assumption that many novice investigators make is that everyone who is online wants to be anonymous – this could not be further from the truth. There are many thousand of examples of individuals being highly overt about associating online content with their identity; indeed tangible financial rewards can be created on social media platforms such as YouTube by creating a following around a well-defined web identity. Of course users are not required to stick to their given physical-world identity within cyberspace, but even when users fabricate online personas there is often a strong incentive to maintain this persona over years or even decades.

The reason behind this behaviour maps back to our most basic human desire to be accepted and respected within a social group, and this trait affects us just as much in the digital world as it does in the real world. The work of cognitive anthropologists Jean Lave and Etienne Wenger (Lave & Wenger 1991)[71] is particularly useful for understanding clustering around social media resources, specifically with regard to their concept of a 'Community of Practice' (CoP). In Lave and Wenger's original theory[72] CoPs were defined as groups of individuals who share a trade or some distinct profession. This common interest unifies the group and motivates the group to acquire more knowledge on the subject that the group is themed on. Although Lave and

[71] Lave, Jean; Wenger, Etienne (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press. ISBN 0-521-42374-0.; first published in 1990 as Institute for Research on Learning report 90-0013.

[72] The concepts surrounding Communities of Practice have evolved significantly since Lave and Etienne's original work. The original concept is used here for its applicability to open source research as opposed to providing an overview of the current thinking about CoPs.

Wenger originally developed the concept based on an observation of tradesmen active in the physical world, the theory is just as applicable in explaining group behaviour on the Internet.

Leetspeak

Certain CoPs go so far as to develop their own versions of languages. The most well-known example of this is Leetspeak, which uses a combination of letters and numbers to create a version of the English language that can only be written. Originally the sole preserve of a hacker, Leetspeak has been adopted by others on the Internet, particularly to create very 'sticky' usernames. Try investigating the background of th3j35t3r (The Jester), a notorious cyber vigilante who has pursued a multiyear campaign against online Jihadists.

Within CoPs online there almost always develops a hierarchy of user, whether the topic of interest for the forum is benign or malicious in nature. Once a hierarchy has been established within a forum, those at the top tend to want to impose that respected and powerful status on other similar-interest groups spread across the Internet. The easiest mechanism for doing this, which avoids a lengthy climb from obscurity, is to reuse a distinctive username across multiple social media platforms.

By joining the dots between forums the investigator has the potential to create a comprehensive profile on an online identity. There are tools that facilitate this process and although they were created to check the availability of a username across multiple social media platforms, they work just as well for tracing the reoccurrence of a username across the Deep Web. Two recommended tools are:

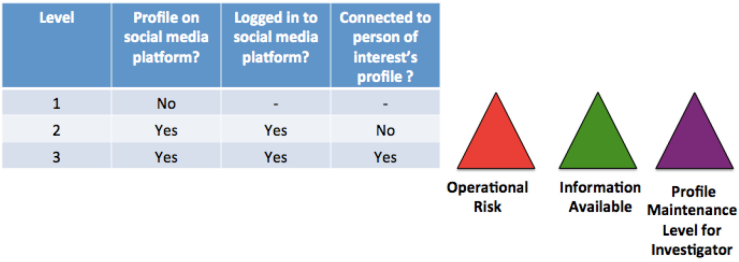
1. *http://knowem.com/*
2. *http://namechk.com/*

Final theoretical points

Levels of connection and volume of data

You might have noticed that while attempting some of the techniques within this chapter, the quantity of data that a person of interest's profile reveals depends largely upon whether the investigator's profile is connected to that person or not. This observation reveals the basic principles that apply across almost all social media platforms: the more connected an investigator's account is to a profile, the more data will be revealed to the investigator. Clearly the obvious approach to social-media-based research is to connect to as many profiles as possible during an investigation; however, this approach is not without its risks.

Figure 10 shows the three levels of connection that two profiles can have on a social media platform (the table on the left) mapped to the operational risk, information gained from the level of connection and level of effort on the part of the investigator to maintain the profile used to connect to the person of interest (the triangles on the right):



Level	Profile on social media platform?	Logged in to social media platform?	Connected to person of interest's profile ?
1	No	-	-
2	Yes	Yes	No
3	Yes	Yes	Yes

Figure 10: Three levels of connection available within a social media platform mapped to the risk level and the information that can potentially be gained (width of triangle represents level of risk (red), volume of data (green) and level of effort to maintain the profile (purple))

On first review of *Figure 10* you may wonder what the operational risk triangle represents and why it is increasing across the connectivity levels. The width of the triangle literally represents the amount of risk that a connectivity level presents to the investigator within the context of an investigation. The risk level increases due to the fact that when an investigator conducts activities on social media platforms (creating and connecting profiles and so on), the investigator in effect moves closer to the person of interest's profile, going from distant observation at Level 1 to the digital equivalent of a direct conversation at Level 3. Functionally this requires the investigator to create a fake profile or 'sock puppet' that requires a credible backstory to make the profile appear authentic (Annex A gives some tips on best practice for building credible fake profiles). The closer an investigator gets to a person of interest, the more this backstory (or 'legend' to use classic intelligence tradecraft speak) comes under greater scrutiny by the person of interest. This level of contact requires great focus and concentration on the part of the investigator and any slip in the sock puppet's legend or behaviour can lead to operational compromise (this concept is covered in more depth in *Chapter 5*).

Taking the risk factor into account, examining *Figure 10* in detail shows that at Level 1 (no profile, not logged in or connected) the least amount of data is available from a profile. Although revealing the least amount of data, an investigation operating at Level 1 is the least risky of the three levels as it does not require a profile to be created on a social media platform that could alert the person of interest to the investigator's activities. This is the typical scenario when an investigator finds a social media profile of a person of interest via a Surface Web search engine and clicks on the link.

An investigator operating at Level 2 (profile and logged in but not connected) has a good mix of access to data versus the operational risk of exposure. As such, Level 2 tends to be the most common configuration for most investigations as this level hits the sweet spot that balances good access to data and a managed level of operational risk without undue levels of effort to maintain a relationship.

A Level 3 operation (profile, logged in and connected) gains the greatest amount of data, but incurs the greatest amount of effort from the investigator for profile maintenance as well as exposing the investigator to the highest level of operational risk. Typically this level of operation is carried out by specially trained personnel who have a specific remit to connect to identified profiles of persons of interest within the context of a much wider investigation.

Much can be gained from being a bold investigator in cyberspace, and I have participated in and am aware of many cases of high-level investigations involving the routine creation and use of a fake profile to connect to persons of interest. Linked with any online investigation involving social media is the possibility of compromise to the investigating agency. Although this possibility can never be 100 percent guarded against, a solid risk assessment by the investigator, taking into account the levels of connectivity outlined in *Figure 10*, is a measured and structured way to pursue an online investigation.

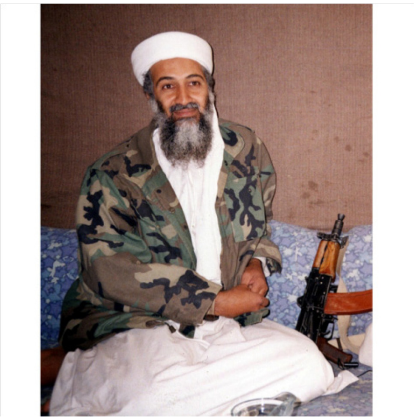
Conclusion

This chapter differed from the previous chapter in that the key to investigating the Deep Web is seeing beyond the raw information to visualise the social networks that created the web content in the first instance. Think of data within the Deep Web not as isolated islands, but as complex interconnected chains of information.

The skills required for effectively exploiting the Deep Web are heavily based in social science techniques such as social network analysis, content analysis and theories such as Lave and Wenger's concept of CoPs. The inquisitive investigator will find much material to develop their critical thinking with regard to Deep Web investigations in social science literature that examines group social phenomenon such as ganging, flash mobbing, closed communities and other forms of collective human behaviour.

Message within the medium

Social media analysis is not simply about assessing the message communicated within a written text. Often large visual cues are present within the images that individuals and groups place onto the Internet. Take, for example, the image of Osama Bin Laden: what do you notice about the weapon he is holding and what does it mean to you?



The weapon in the picture is an AKS-74U, a 'stubby' variant of the more common AK-47 assault rifle. Bin Laden is carrying this weapon due to its status as a trophy item within the Jihadist community. The AKS-74U holds this status due to the fact that specialist Russian soldiers such as tank and helicopter gunship crews carried it. Possessing a weapon like this implies that the owner has been skilful enough in battle to acquire the enemy's weapons, adding to the power and status of the individual carrying it. This one example serves to show how contextual data can alter the message that an image is trying to convey[73].

[73] As much as the author would like to claim the credit for this clever analysis, I cannot. Instead, the reader should consult Chivers, C.J. (2010). *The Gun. The Story of the AK-47*. Page 383 for the original analysis that contributed to this piece.