Network Analysis for Criminal Investigators:
Is it possible to identify criminal networks using mobile phone data?

2401049
Joshua.A.J.McManus
The University of Buckingham
BA(Hons) Security, Intelligence, and Cyber

Table of Contents

SECTION 1:

Introduction.

Introduction

In the study of modern criminal organisations, the primary obstacle for law enforcement remains the 'information problem'. That is, the inherent difficulty of acquiring high-quality data to map a network actively invested in its own concealment[1]. While modern technology has allowed for the expanse of these criminal enterprises onto the global domain[2], mobile phone data has emerged as a tool for resolving this problem, offering a pervasive digital footprint of both association intent. However, the framing of the question, "Is it possible to identify a criminal network?", is analytically reductive IF it implies a binary outcome. Therefore, it is the understanding of this essay that 'identification is not JUST an act of passive knowledge but a process of active construction, one whose validity is bounded as much by the assumptions embedded in its analytical methods, as it is by criminal countermeasures.

This essay argues that while mobile data provides the essential raw material for network identification, the transition from fragmented digital traces to true intelligence[3] rests on a sequential triad:

1) The capacity to structure unstructured data,
2) The application of mathematical diagnostics to detect hidden actors, and
3) The ability to overcome sophisticated countermeasures.

---

[1] Sparrow, M.K. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. Social Networks, [online] 13(3), pp.251–274. doi:https://doi.org/10.1016/0378-8733(91)90008-h.

[2] García, D.P. and Felbab-Brown, V. (2026). How technology is transforming crime and terrorism. [online] Brookings. Available at: https://www.brookings.edu/articles/how-technology-is-transforming-crime-and-terrorism/ [Accessed 12 May 2026].

[3] This essay defines Information as Data with Context, and Intelligence as "Informational Actionable".

SECTION 2:

Main Body.

The Information Problem

As stated in the introduction, the foundational challenge in network analysis is the two fold 'information problem'. In its technical form: the vast majority of phone-derived data is unstructured (meaning the investigator must analyse raw audio non-standard text, and disconnected metadata) and is in such an immense volume it becomes nigh-on impossible to differentiate the "signal from the noise", a SIGINT (Signals Intelligence) phrase pioneered from transatlantic radio telephony[4]. Before a network can be truly identified, it must be actively structured into nodes (individuals) and ties (relationships). Data such as Call Detail Records (CDRs) provide the primary skeleton, allowing for preliminary identification based on topology. Metrics such as degree centrality can flag a communicative hub, while betweenness centrality identifies structural brokers who control information flow between otherwise disconnected subgroups, a classic role for command-and-control nodes[5].

However, to treat this topological skeleton as an 'identified network' is to commit what this essay can only describe as a cartographical fallacy. Metadata proves only that a contact occurred; it is silent on the nature, purpose, or power dynamics present in that contact. A node with high betweenness centrality could be a pivotal crime lord, or equally, a legitimate logistics coordinator unaware of the criminal purpose being served. Resolving these ambiguities requires bridging the gap. While human investigators provide the interpretative hypotheses and crucial training labels, in the modern age it is thought that machine learning can scale to process the volume of non-volatile memory and logs necessary to classify nodes into operational 'roles'[6]. This human-computer synthesis is a viable solution to the scaling problem, yet it introduces a profound epistemological dependency. The model's output is entirely contingent on its training data and labelled examples it is fed. The AI does not discover roles; its pattern-matches against a pre-defined database of criminal behaviour supplied by the human analyst[7]. The resulting network therefore reflects investigator hypothesis as much as suspect reality.

---

[4] Arnold, H.D. and Espenschied, L. (1923). Transatlantic Radio Telephony. Bell System Technical Journal, [online] 2(4). Available at: https://archive.org/details/bstj2-4-116/mode/2up [Accessed 12 Oct. 2025].

[5] Calderoni, F. (2023). Predicting Organized Crime Leaders. Crime Prevention Studies, [online] 28, pp.89–110. doi:http://hdl.handle.net/10807/68084.

[6] Ferrara, E., De Meo, P., Catanese, S. and Fiumara, G. (2014). Detecting criminal organizations in mobile phone networks. Expert Systems with Applications, 41(13), pp.5733–5750. doi:https://doi.org/10.1016/j.eswa.2014.03.024.

[7] Europol (2025). AI Bias in Law enforcement: a Practical Guide. https://www.europol.europa.eu/publications-events/publications/ai-bias-in-law-enforcement#downloads.

Evaluating the Network

To move from an uncharted heap of undifferentiated contacts to a functional map of a criminal hierarchy, the ties between nodes must be assigned a numerical weight derived through content analysis. This is the process of adding qualitative depth to quantitative structure. Criminal networks rarely operate in plain language; they deliberately employ linguistic countermeasures such as 'leet' speak[8], code-switching, or slang to create a lexical haze designed against automated monitoring. The Levenshtein distance metric is essential here[9]. By algorithmically calculating the minimum single-character edits required to transform an obfuscated string into a known dictionary word, investigators can normalise this language back into a searchable database. However, this method can introduce analytical risk if the setting of the edit-distance threshold is too low, causing investigators to fail to capture creative misspellings, while setting it too high generates false-positive matches to innocent language, potentially manufacturing connections that do not exist in reality.

The true power of content identification lies in statistical semantics, which attempts to bypass surface noise. Term Frequency-Inverse Document Frequency (TF-IDF) provides a mathematically rigorous method to weigh terms by their informational rarity. Common words like "car" or "meet" appear frequently across many transcripts and are thus analytically thin. In contrast, TF-IDF highlights words that are locally frequent in a specific set of intercepts but globally rare across a wider linguistic databases[10].The recurring use of an unusual code for a specific ideas and locations would cause a spike on a TF-IDF analysis, instantly increasing the weight of any tie carrying that term as possible criminal operations rather than regular online social exchanges. N-grams further enhance this by modelling word sequences to provide context to fragmented or truncated text messages, while sentiment classification can assign a register, algorithmically distinguishing a command from a request[11]. The critical point, however, is that these tools are not transparent windows into meaning. TF-IDF identifies statistical rarity, not criminal intent, and sentiment analysis models trained on product reviews may misattribute the flat, affectless register of a disciplined conspirator as neutral rather than operational. The weighted network is therefore a second-order construct,

[8] IONOS (2021). Leetspeak: A Guide for Beginners. [online] IONOS Digital Guide. Available at: https://www.ionos.co.uk/digitalguide/online-marketing/social-media/what-is-leetspeak/ [Accessed 12 May 2026].

[9] Mukherjee, S. (2025). Levenshtein Distance: A Comprehensive Guide. [online] Digitalocean.com. Available at: https://www.digitalocean.com/community/tutorials/levenshtein-distance-python [Accessed 13 May 2026].

[10] Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, [online] 24(5), pp.513–523. doi:https://doi.org/10.1016/0306-4573(88)90021-0.

[11] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, [online] 2(1-2), pp.1–135. Available at: https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf [Accessed 13 May 2026].

merely a mathematical abstraction of a linguistic reality already designed to be twice removed from the analyst.

The Spatial Dimensions

A network analysis graph (regardless of how richly weighted by content) remains a purely social mapping until fused with geolocation data, which promises a physical, verifiable dimension. The correlation of social distance with physical proximity is a powerful indicator. If two nodes sharing a highly weighted, secretive tie are also shown via cell tower trilateration or GPS to be co-located at a specific time (see CDRs from earlier), the inference of a physical criminal meeting is significantly strengthened. Geofencing, the creation of a virtual perimeter around a known high-risk zone[12], transforms this from purely mapping to proactive and real-time intelligence.

Yet, it is precisely at this point of apparent concrete proof that geolocation reveals itself as a site of fragility. The production of true location from a mobile device is an error-laden process, and to treat it as ground truth carries significant risk. Multipath error, common in urban environments where radio signals reflect off buildings before reaching a handset[13], can introduce displacements of hundreds of metres[14]. A data point that appears to place a suspect inside a specific building may, in reality, be a phantom signal generated by a distant reflection. Furthermore, the internal logic of the device creates structured gaps in the data timeline. This can, however, be combatted by employing an intelligent form of Global Navigation Satellite System (GNSS) mixed with 3D Mapping, also known as Shadow Mapping. Shadow Mapping works by "comparing the measured signal availability with that predicted over a grid of candidate positions using 3D mapping" [15].

Without Cellular Assisted GPS, which uses cellular network data to accelerate a position fix, a cold-start GPS receiver can require a significant signal acquisition window. This is not a simple technical glitch but a predictable, exploitable temporal blind spot. The spatial 'position' produced by the device is therefore not a true point, but a probability cloud bounded by an error ellipse[16], and the network node exists within (what this essay is calling)

---

[12] Jones, R. (2024). Geofencing software to direct specialist resources to monitor preventative orders. [online] College of Policing. Available at: https://www.college.police.uk/support-forces/practices/geofencing-software-direct-specialist-resources-monitor-preventative-orders [Accessed 13 May 2026].

[13] European Space Agency (2011). Multipath. [online] Esa.int. Available at: https://gssc.esa.int/navipedia/index.php/Multipath [Accessed 13 May 2026].

[14] Groves, P., Jiang, Z., Rudi, M. and Strode, P. (2013). A Portfolio Approach to NLOS and Multipath Mitigation in Dense Urban Areas. [online] Available at: https://discovery.ucl.ac.uk/id/eprint/1394968/1/ION_GNSS13_B6_2_Groves_et_al_1_0%20%28NLOS%29.pdf [Accessed 13 May 2026].

[15] Groves, P., Adjrad, M., Gao, H. and Ellul, C. (2016). Intelligent GNSS Positioning using 3D Mapping and Context Detection for Better Accuracy in Dense Urban Environments. [online] Available at: https://discovery.ucl.ac.uk/id/eprint/1524033/2/Groves_INC2016%20Groves%20et%20al%20EDITED.pdf [Accessed 13 May 2026].

[16] Richter, P. and Toledano-Ayala, M. (2017). Ubiquitous and Seamless Localization: Fusing GNSS Pseudoranges and WLAN Signal Strengths. Mobile Information Systems, [online] pp.1–16. doi:https://doi.org/10.1155/2017/8260746.

a ‘zone of likelihood’. A skilled legal defence does not need to disprove this data as it all it needs to do is establish the reasonableness of the error radius as doubt.

Mathematical Diagnostics

Beyond direct communication analysis, mathematical diagnostics offer the possibility of identifying actors who are structurally invisible within the mapped network itself, allowing for analysis of the shape of the unobserved structure. Benford's Law, the "counter-intuitive"[17] statistical principle that in many naturally occurring numerical datasets, the leading digit '1' appears at a higher than probable rate, with larger digits being proportionally less frequent, provides a tool for a system-level diagnostic of an entire network's 'naturalness'[18]. The application, while indirect, can be seen through extracting the degree centrality score for every node in a large network, and then plotting the distribution of the first digits of those scores. This allows investigators to statistically test for conformity to, or deviation from, the expected Benfordian curve.

The working assumption is that an organic social network's degree distribution, when spanning multiple orders of magnitude, will closely mirror Benford's Law. Therefore, a statistically significant deviation (determined via a Chi-Square (χ^2) goodness-of-fit test at a conventional significance threshold, such as $\alpha = 0.05$) constitutes an anomaly indicating that the network structure may be manufactured or 'unnatural'. This distortion can happen for several reasons. For example, a single handler managing hundreds of unregistered 'burner' phones would create a massive cluster of nodes with a degree centrality of 1 or 2. This could artificially inflate the frequency of low leading digits, causing the network to fail the Benford test.

However, this application demands extreme analytical caution. While research has demonstrated that Benford's Law CAN apply to the socio-behavioural metrics of digital communication, its usage in a criminal intelligence context presses against the limits of this natural conformity[19]. The network's degree distribution is itself a product of prior analytical choices; such as the threshold at which a contact constitutes a tie, and the temporal boundary of the dataset. A deviation from Benford's Law may therefore reflect those choices as readily as it reflects criminal manipulation. As such, a Benford failure cannot serve as primary evidence. Within the context of the information problem, this debate does not invalidate the method, but it precisely defines its role. Benford's Law is a tool for generating investigatory leads and not forensic proof. It flags nodes for the reallocation of other surveillance resources, where more legally defensible evidence can be gathered.

---

[17] Jamain, A. (2001). Benford's Law. [online] Available at: https://www.ma.imperial.ac.uk/~nadams/classificationgroup/Benfords-Law.pdf [Accessed 13 May 2026].

[18] Nigrini, M.J. (2012). Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection. John Wiley & Sons, Inc. eBooks. John Wiley & Sons. doi:https://doi.org/10.1002/9781119203094.

[19] Golbeck, J. (2015). Benford's Law Applies to Online Social Networks. PLOS ONE, [online] 10(8), p.e0135169. doi:https://doi.org/10.1371/journal.pone.0135169.

Countermeasures and Legality

The ultimate boundary of the possibility of identification is not the sophistication of the toolkit used, but the adaptive counter-strategies of the criminal network and the evidentiary standards of the law. Every technique designed to fuse data and construct the network is met with an equally deliberate evasion designed to deconstruct it. Faraday bags are used to create portable signal blackouts, encrypted communication applications are used to limit back-end access, Virtual Private Networks to obscure IP geolocation, and the ubiquitous unregistered 'burner phone'. These are techniques utilised globally by criminal networks and terrorist organisations alike. They are active, rational strategies with the sole aim of fragmenting the bigger picture before it can even be assembled[20]. When this fragmentation is successfully enforced by a disciplined network, the state's view is reduced to disconnected, meaningless shards. Identification becomes impossible not because the individuals are hidden, but because their relational structure cannot be mathematically derived.

This operational limit is then decisively compounded by the juridical limit, which imposes a categorically different standard of truth. This showcases the semantic divide between intelligence and evidence. Intelligence is, as previously stated, 'Information Actionable', while evidence can be defined as "information given to the court and the jury to help them decide if a crime has been committed or not… [and]…tends to prove the truth or probability of truth about a fact put before the court and jury"[21], the standards of which are, largely, governed by the Criminal Justice Act (2003)[22].

The transition from an intelligence product (such as Network Model that is operationally true and tactically useful) to an admissible piece of evidence is frequently the most underestimated phase of the identification process. The common law hearsay rule exists to exclude out-of-court statements whose reliability cannot be tested through cross-examination[23]. A text message extracted from a physical seizure may be incriminating on its face, but before it can become a legal fact, a digital forensics expert must establish (through an unbroken and meticulously documented chain of custody) that the message was typed on that specific device, at that specific time, by an uncoerced individual whose identity can be proven. Furthermore, it must be proven that the file containing the message has not been altered since seizure, whether by human tampering, a system process, or a remote-access

[20] Hutchings, A. and Holt, T.J. (2014). A Crime Script Analysis of the Online Stolen Data Market: Table 1. British Journal of Criminology, [online] 55(3), pp.596–614. doi:https://doi.org/10.1093/bjc/azu106.

[21] HM Government (2026). Evidence in criminal investigations (accessible). [online] GOV.UK. Available at: https://www.gov.uk/government/publications/evidence-in-criminal-investigations/evidence-in-criminal-investigations-accessible [Accessed 13 May 2026].

[22] HM Government (2003). Criminal Justice Act 2003. [online] Legislation.gov.uk. Available at: https://www.legislation.gov.uk/ukpga/2003/44/contents [Accessed 13 May 2026].

[23] Mason, S. (2012). Electronic Evidence. LexisNexis Butterworths.

trojan. A single undocumented handover of the device between officers, or the theoretical possibility of malware, can sever the evidential chain at its weakest link, rendering a key node juridically invisible. The network that is ‘identified’ in an intelligence briefing is therefore a fundamentally different, and far more fragile, entity than the network that can be proven to the criminal standard of beyond reasonable doubt in a courtroom.

SECTION 3:

Synthesis.

Conclusion

In the study of modern criminal organisations, the primary obstacle for law enforcement remains the 'information problem'. The evolution of technology brought with it both the advancements in the ways that a criminal organisation may attempt to hide itself, and the advancements in the ways that law enforcement and criminal investigators can find weak links in the armour of these organisations. On whether or not it is possible to identify criminal networks through mobile phone data, it must be understood that this 'possibility' is a deeply contingent construct, rather than just a binary fact. This process to do so includes numerous sequential variables, such as the capacity to structure unstructured data, the application of statistical diagnostics, and the ability to overcome sophisticated countermeasures, all while gathering enough valid evidence to make the identification of such criminal organisations admissible in court.

However, at each stage of this construction, the proposed 'network' that emerges is shaped as much by the assumptions and limitations of the analytical tools, the thresholds set, the training data chosen, the error radii accepted, and by the empirical traces left by the suspects themselves. While the deployment of bleeding-edge analytical indicators (such as the Chi-Squared (χ^2) goodness-of-fit test alongside Benford's Law on chatlog data) offers a sophisticated diagnostic framework, it must be repeated that its usage in a criminal intelligence context constantly presses against the limits of natural socio-behavioural conformity. Network analysis, even when propped up with analytical backing, never truly paints a complete or final picture. Criminal networks are perpetually unstable, and the analysis of which is done by running probabilistic models whose success is measured by its temporary advantage in an ongoing, battle between obfuscation and illumination. So while it is possible (and often probable) that a criminal network can not only be investigated but also identified by mobile phone data, the extent to which this is useful is subject to numerous external and unpredictable variables

THANK YOU.

SECTION 4:

Accreditation.

Sources Used.

1. Arnold, H.D. and Espenschied, L. (1923). Transatlantic Radio Telephony. *Bell System Technical Journal*, [online] 2(4). Available at: https://archive.org/details/bstj2-4-116/mode/2up [Accessed 12 Oct. 2025].
2. Calderoni, F. (2023). Predicting Organized Crime Leaders. *Crime Prevention Studies*, [online] 28, pp.89–110. doi:http://hdl.handle.net/10807/68084.
3. European Space Agency (2011). *Multipath*. [online] Esa.int. Available at: https://gssc.esa.int/navipedia/index.php/Multipath [Accessed 13 May 2026].
4. Europol (2025). *AI Bias in Law enforcement: a Practical Guide*. *https://www.europol.europa.eu/publications-events/publications/ai-bias-in-law-enforcement#downloads*.
5. Ferrara, E., De Meo, P., Catanese, S. and Fiumara, G. (2014). Detecting criminal organizations in mobile phone networks. *Expert Systems with Applications*, 41(13), pp.5733–5750. doi:https://doi.org/10.1016/j.eswa.2014.03.024.
6. García, D.P. and Felbab-Brown, V. (2026). *How technology is transforming crime and terrorism*. [online] Brookings. Available at: https://www.brookings.edu/articles/how-technology-is-transforming-crime-and-terrorism/ [Accessed 12 May 2026].
7. Golbeck, J. (2015). Benford's Law Applies to Online Social Networks. *PLOS ONE*, [online] 10(8), p.e0135169. doi:https://doi.org/10.1371/journal.pone.0135169.
8. Groves, P., Adjrad, M., Gao, H. and Ellul, C. (2016). *Intelligent GNSS Positioning using 3D Mapping and Context Detection for Better Accuracy in Dense Urban Environments*. [online] Available at: https://discovery.ucl.ac.uk/id/eprint/1524033/2/Groves_INC2016%20Groves%20et%20al%20EDITED.pdf [Accessed 13 May 2026].
9. Groves, P., Jiang, Z., Rudi, M. and Strode, P. (2013). *A Portfolio Approach to NLOS and Multipath Mitigation in Dense Urban Areas*. [online] Available at: https://discovery.ucl.ac.uk/id/eprint/1394968/1/ION_GNSS13_B6_2_Groves_et_al_1_0%20%28NLOS%29.pdf [Accessed 13 May 2026].
10. HM Government (2003). *Criminal Justice Act 2003*. [online] Legislation.gov.uk. Available at: https://www.legislation.gov.uk/ukpga/2003/44/contents [Accessed 13 May 2026].
11. HM Government (2026). *Evidence in criminal investigations (accessible)*. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/evidence-in-criminal-

investigations/evidence-in-criminal-investigations-accessible [Accessed 13 May 2026].

12. Hutchings, A. and Holt, T.J. (2014). A Crime Script Analysis of the Online Stolen Data Market: Table 1. *British Journal of Criminology*, [online] 55(3), pp.596–614. doi:https://doi.org/10.1093/bjc/azu106.
13. IONOS (2021). *Leetspeak: A Guide for Beginners*. [online] IONOS Digital Guide. Available at: https://www.ionos.co.uk/digitalguide/online-marketing/social-media/what-is-leetspeak/ [Accessed 12 May 2026].
14. Jamain, A. (2001). *Benford's Law*. [online] Available at: https://www.ma.imperial.ac.uk/~nadams/classificationgroup/Benfords-Law.pdf [Accessed 13 May 2026].
15. Jones, R. (2024). *Geofencing software to direct specialist resources to monitor preventative orders*. [online] College of Policing. Available at: https://www.college.police.uk/support-forces/practices/geofencing-software-direct-specialist-resources-monitor-preventative-orders [Accessed 13 May 2026].
16. Mason, S. (2012). *Electronic Evidence*. LexisNexis Butterworths.
17. Mukherjee, S. (2025). *Levenshtein Distance: A Comprehensive Guide*. [online] Digitalocean.com. Available at: https://www.digitalocean.com/community/tutorials/levenshtein-distance-python [Accessed 13 May 2026].
18. Nigrini, M.J. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. *John Wiley & Sons, Inc. eBooks*. John Wiley & Sons. doi:https://doi.org/10.1002/9781119203094.
19. Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, [online] 2(1-2), pp.1–135. Available at: https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf [Accessed 13 May 2026].
20. Richter, P. and Toledano-Ayala, M. (2017). Ubiquitous and Seamless Localization: Fusing GNSS Pseudoranges and WLAN Signal Strengths. *Mobile Information Systems*, [online] pp.1–16. doi:https://doi.org/10.1155/2017/8260746.
21. Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, [online] 24(5), pp.513–523. doi:https://doi.org/10.1016/0306-4573(88)90021-0.
22. Sparrow, M.K. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks*, [online] 13(3), pp.251–274. doi:https://doi.org/10.1016/0378-8733(91)90008-h.

Network Analysis for Criminal Investigators:

Is it possible to identify criminal networks using mobile phone data?

2401049

Joshua.A.J.McManus

The University of Buckingham

BA(Hons) Security, Intelligence, and Cyber