ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

Bottom-to-Top OSINT Investigation: A Structured Approach for Relational Attribution in Cyber Threat Analysis

Gagan Jain B. S.[1]

CEO & Founder, CyberSafe Bangalore Master's in Information Security And Networking - Durham Canada

CEH, CND, CompTIA Pentest+, EC-Council CEI

Co-Author: Gagan Jain B. S.[1]

Publication Date: 2026/05/04

Abstract: Open Source Intelligence (OSINT) plays a critical role in modern cyber incident response, enabling investigators to extract actionable insights from publicly available data. However, traditional approaches often rely on top-down logic—starting from a hypothesis and attempting to confirm it through data. This research introduces a novel Bottom-to-Top OSINT methodology focused specifically on relational attribution: the process of linking individuals, threat actors, and infrastructure based on verified open-source attributes.

Our approach begins with a singular data point—such as an email or phone number—and builds upward through a structured process of contextual enrichment, relational mapping, and validation. The methodology is implemented through a custom-built tool named Sycek, which extracts and indexes breach-derived attributes and displays entity relationships via visual graphing and confidence scoring.

➢ ***The Paper Contributes:***

- **A structured, bottom-up methodology for relational attribution.**
- **Demonstrated examples using synthetic and breach-derived data.**
- **An iterative hybrid loop that integrates bottom-up discovery with top-down hypothesis validation.**

How to Cite: Gagan Jain B. S. (2026) Bottom-to-Top OSINT Investigation: A Structured Approach for Relational Attribution in Cyber Threat Analysis. *International Journalof Innovative Science and Research Technology*, 11(4), 3020-3033. https://doi.org/10.38124/ijisrt/26apr1405

I. INTRODUCTION

Open Source Intelligence (OSINT) refers to the practice of collecting, analysing, and leveraging publicly accessible information for investigative or intelligence purposes. In cybersecurity, OSINT is a foundational capability—used to detect leaked credentials, profile threat actors, map digital infrastructure, and identify risks before escalation.

Traditionally, OSINT investigations follow a Top-Down approach: the analyst begins with a clear hypothesis (e.g., "this may be linked to group X") and collects data to support or refute that claim. Conversely, Bottom-Up approaches are more exploratory—starting from isolated data points (e.g., a phone number or alias) and building toward broader intelligence through linkage and enrichment.

However, current methodologies fall short in one critical dimension: structured relational attribution. While individual tools and techniques may identify links between entities, there is no widely adopted framework for systematically constructing relational maps from raw OSINT data in a validated, auditable way.

This research introduces the Bottom-to-Top Relational Attribution (BTRA) model as a novel contribution. It combines structured bottom-up data gathering with iterative top-down analysis to produce robust identity profiles, all operationalised via the custom-built Sycek platform.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

II. LITERATURE REVIEW

➢ *Contemporary OSINT Practices Typically Adhere to One of three Models:*

- Top-Down (Deductive): Begins with a hypothesis or target profile and drills down into data to find supporting evidence.
- Bottom-Up (Inductive): Starts with a low-level clue (e.g., a handle or breached record) and builds a broader picture through pivoting.
- Hybrid Approaches: Combine both top-down reasoning and bottom-up exploration, often in iterative cycles.

While these methods are functionally effective for data collection or surface-level enrichment, they lack consistency when it comes to relational attribution—i.e., reliably linking people, entities, or infrastructure with real-world contextual evidence.

➢ *Key Limitations Include:*

- Lack of structure in relational discovery: Most methods rely on analyst intuition rather than repeatable logic.
- Risk of misattribution: Without multi-source validation and scoring, inferred connections can be incorrect.
- Difficulty in identifying weak or hidden links: Pattern recognition across noisy or partially structured datasets is often left to manual interpretation.

To date, no formally defined, bottom-up framework for structured relational attribution has been published in academic or applied OSINT literature. Our work fills this gap by introducing the BTRA model, supported by a fully operational tooling layer through Sycek.

➢ *Proposed Methodology – "Bottom-to-Top Relational Attribution" (BTRA):*

Define and formalise your unique approach step-by-step:

- *Step 1: Initial Incident or Attribution Identification*

✓ Start from the smallest traceable indicators (phone numbers, emails, usernames, etc.).

- *Step 2: Contextual Expansion (Horizontal Pivoting)*

✓ Identify additional open-source data related to these small identifiers.

- *Step 3: Relational Linking (Vertical Pivoting)*

✓ Clearly define categories of relational attribution:
✓ Direct associations (clear evidence of direct connection).
✓ Indirect associations (shared intermediary connections).
✓ Thematic associations (similar interests, thematic overlaps).

- *Step 4: Verification and Filtering*

✓ Apply multiple-source validation to avoid misinterpretation or false attribution.
✓ Document clearly which attributions are confirmed, probable, or unconfirmed.

- *Step 5: Building the Relational Profile*

✓ Systematically organise the validated data into structured profiles.
✓ Clearly visualise relationships using network graphs and diagrams to establish clarity and facilitate analysis.

- *Step 6: Iterative Loop with Hybrid Methodology*

✓ Iterate through this hybrid loop until a clear, validated, and actionable profile emerges.

➢ *Step 1: Initial Incident or Attribution Identification*

In a Bottom-to-Top Relational Attribution (BTRA) OSINT investigation, the initial incident or attribution identification phase involves clearly isolating a singular foundational data point obtained from open-source or incident-related data. This carefully selected data point serves as the investigative entry point, laying the groundwork for subsequent attributional discovery phases.

Investigators typically select foundational attributes from the following structured categories:

- *Identity-Based Attributes*

✓ Personal identifiers: Names, usernames, aliases.
✓ Contact identifiers: Email addresses.
✓ Familial identifiers: Names of immediate or extended family members.
✓ Official identifiers: Government-issued documents (passports, national IDs, tax IDs).

- *Phone-Based Attributes*

✓ Contact numbers: Mobile numbers, landlines, VoIP numbers.

- *Social-Based Attributes*

✓ Platform handles: Social media usernames or unique IDs.
✓ Linked attributes: Emails or phone numbers associated with social profiles.
✓ Visual identifiers: Profile pictures, photographs, digital avatars.

- *Geo-Based Attributes*

✓ Explicit locations: Street addresses, city names, area codes.
✓ Coordinate-level data: GPS coordinates, geotags, IP-based location information.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

- *Organisational or Syndicate-Based Attributes*

✓ Collective identities: Affiliations with organised crime groups, syndicates, or extremist entities.

➢ *Illustrative Example:*

To demonstrate clearly, consider the initial data point selected is an email address from a commonly utilised email service provider:

- *Chosen Data Point:*

john.doe123@gmail.com

This singular Gmail address provides a clear and relatable starting attribute, from which the investigator will systematically expand horizontally (contextual enrichment) and vertically (relational linking) in subsequent investigation phases.

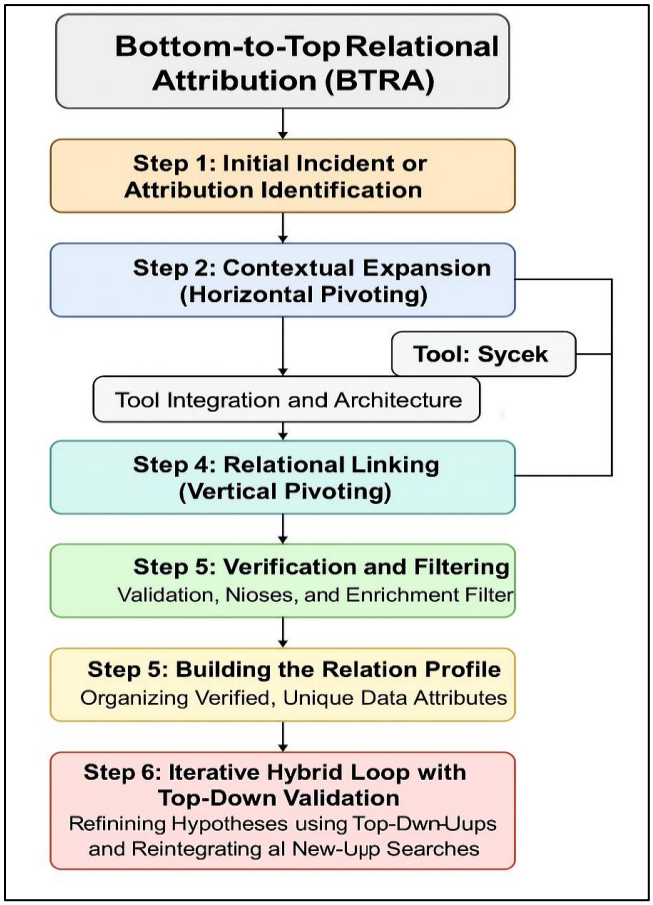


Fig 1 Chosen Data Point

➢ *Step 2: Contextual Expansion (Horizontal Pivoting)*

Upon clearly establishing the initial foundational attribute (e.g., an email address), the next critical phase, known as Contextual Expansion or Horizontal Pivoting, systematically expands the investigation to gather extensive contextual information from diverse open-source datasets.

The core strength of the Bottom-to-Top Relational Attribution (BTRA) method here lies in its deliberate and structured data-aggregation process, prioritising data validation and evidentiary reliability.

➢ *Formalised Investigative Process:*

The investigator initiates the Contextual Expansion process by following a structured, iterative approach:

- *Phase 1: Reverse Attribute Discovery*

Starting with the given email (e.g., john.doe123@gmail.com), the investigator performs comprehensive reverse searches to uncover associated digital footprints. This is done across multiple open-source intelligence domains:

✓ General search engines (Google, Bing, Yandex) for publicly indexed references.
✓ Social media platforms (Twitter, Facebook, LinkedIn, Instagram, Reddit, Telegram) to identify user accounts.
✓ Deep and Dark web intelligence portals (specialised OSINT databases, forums, marketplaces, and onion-based sites) for hidden references.

- *Phase 2: Structured Attribute Checks*

Following the initial reverse discovery, the investigator conducts structured attribute checks to build a robust data profile:

✓ *Username Checks:*

Utilise specialised tools (e.g., Sherlock, WhatsMyName, or custom username-checkers) to locate potential usernames linked directly to the identified email.

✓ *Email Alias and Similarity Checks:*

Identify variants and aliases of the original email, using techniques such as wildcard searches and email permutation tools (e.g., Epieos, Hunter.io, and custom scripts) to detect related email variations.

✓ *Recent Online Reviews and Activity:*

Extract publicly available reviews, comments, or interactions linked explicitly to the email or username from platforms like Google Reviews, Yelp, or forum postings.

✓ *Geolocation Data Gathering (1st Link of Attribute):*

Locate explicitly tagged geographical data (addresses, recent city-based references, GPS coordinates) clearly linked to the user's known activities, digital posts, or public mentions.

✓ *Visual Attribute Discovery (Photo/Avatar):*

Capture publicly indexed profile images or avatars to further validate account consistency across various platforms through reverse image searches (Google Images, Yandex, TinEye).

✓ *Recent Activity Timestamps (2nd Link of Attribute):*

Obtain activity data timestamps (last-seen dates and times, recent posts or interactions) via intelligence databases (Epieos, platform-specific "last active" metadata) to anchor findings temporally.

- *Phase 3: Breached Database Cross-validation*

The investigator explicitly leverages breached datasets (via reliable sources like HaveIBeenPwned, Dehashed, or

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

internal collections) to cross-validate attributes identified during the earlier phases. Unlike traditional top-down approaches, the use of breached databases offers verified and authenticated data points (real-world usage):

- ✓ Validated emails and phone numbers clearly verified in real-world user transactions.
- ✓ Confirmed usernames and aliases explicitly linked through breached records.
- ✓ Authentic secondary contact details (secondary phones, emails, or backup authentication methods), providing additional verification.
- ✓ Proven addresses and transactional details, as breached datasets typically represent confirmed, real-world interactions (purchases, sign-ups, deliveries).

➢ *Evidential Strength of BTRA Contextual Expansion:*

The methodological advantage of this structured Bottom-to-Top expansion lies explicitly in its evidentiary reliability. Unlike top-down approaches—where attributes (usernames, phone numbers, etc.) might initially be hypothesised or inferred—this method rigorously grounds each discovered attribute in confirmed, real-world transactional data, substantially enhancing investigative accuracy and reducing potential misattribution.

Through this method, each gathered attribute undergoes structured validation, providing greater evidential confidence, explicitly linking the user reliably to subsequent attributes.

➢ *Illustrative Example (Continued):*

Starting from the initial data point (john.doe123@gmail.com):

- Reverse search results yield usernames (john.doe123, john.d1990), publicly listed on Twitter and LinkedIn.
- Alias and similarity checks via Epieos detect additional linked emails (johndoe1990@gmail.com, john.doe_work@gmail.com).
- Reviews from Google highlight recent interactions at businesses in London, providing reliable geographical context.
- Public profile images obtained from Twitter align visually with LinkedIn profiles, confirming identity consistency.
- Breached data search cross-verifies usernames, secondary emails (john.secondary@gmail.com), and real-world transactional data (e.g., an online purchase), significantly enhancing evidentiary credibility.

➢ *Step 3: Relational Linking (Vertical Pivoting)*
(*With Custom OSINT Engine: Sycek Integration*)

After horizontal pivoting enriches the context around an initial data point, the Bottom-to-Top Relational Attribution (BTRA) methodology advances to Relational Linking, or Vertical Pivoting. This stage involves synthesising the expanded dataset into concrete, evidence-backed connections that help identify real-world relationships—between entities, infrastructure, or behavioural patterns.

To systematise and scale this step, the author developed a purpose-built OSINT engine named Sycek—an intelligence system optimised to process breached datasets, resolve identifiers, and extract relational insights with evidentiary precision.

- *Sycek: A Custom-Built OSINT Engine for Relational Attribution*

Sycek is an OSINT orchestration platform designed to ingest, structure, index, and query large-scale datasets—particularly breach dumps, leaked user records, and intelligence-enriched archives—on a global or country-specific basis. The tool transforms traditionally messy and fragmented breach datasets into a coherent relational intelligence layer accessible via API or scripted queries.

➢ *Key Capabilities and Workflow of Sycek*

- *Curated, Country-Agnostic Data Processing Pipeline*

- ✓ Sycek supports integration of both global breach datasets (e.g., collections from HIBP, Cit0Day, Combo Lists) and country-specific sources (e.g., electoral rolls, telecom data leaks, ecommerce breaches).
- ✓ A parsing pipeline is applied to extract and retain high-value, attribution-grade attributes, including:
- ✓ email, username, name, phone, secondary_phone, address, secondary_address, city, IP address, GPS coordinates, device models, passport/ID numbers, and more.

- *Structured Text Conversion and Attribute Normalisation*

- ✓ All ingested data is preprocessed into structured formats:
- ✓ Cleaned for encoding issues, standardised into field-value pairs.
- ✓ Normalised (e.g., phone formatting, casing, location formatting).
- ✓ Redundant attributes and low-confidence fields are filtered out.
- ✓ Final structured outputs are text-optimised and schema-aligned, suitable for indexing and search.

- *ElasticSearch-Driven Search Engine Architecture*

- ✓ Preprocessed data is indexed into Elasticsearch for scalable and high-speed search.
- ✓ A custom multi-field, fuzziness-enabled query structure enables both exact and near matches:

```
"query": {
"multi_match": {
"query": "<input_value>",
"fields": ["email", "phone", "username", "name", "passport",
"address", "city", "raw"],
"type": "best_fields",
"fuzziness": "AUTO"
}
}
```

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

✓ *This Supports:*

- Misspelt inputs (e.g., john.doe@gmai.com still resolves john.doe@gmail.com).
- Variant detection (e.g., john_doe, john.doe1, john-doe.official).
- Multi-field correlation across identities.

- *Attribute Retrieval and Evidence Clustering*

Once queried, Sycek returns enriched result sets:

✓ Each result includes linked attributes such as secondary phone numbers, multiple usernames, historical addresses, device fingerprints, IP blocks, and geocoordinates.
✓ These attributes are clustered into entity-centric nodes.

➢ *For Instance, Querying john.doe@gmail.com might return:*

- Multiple usernames: john_doe, jd_1990, doe.johnny
- Verified phones: +447XXXXXX12, +447XXXXXX99
- Historical and current addresses from breach records
- Reused device models (e.g., Pixel 6, iPhone SE2) and IP subnet data
- Repeated employer name from resumes or form submissions in leaked data

➢ *Relational Inference Logic*

From this structured dataset, Sycek performs a relational linkage analysis:

- Direct Links: Same IP range or GPS coordinate across multiple attributes.
- Indirect Links: Different usernames sharing the same mobile number or secondary email.
- Contextual Bridges: Common device ID used across otherwise distinct aliases.
- Behavioural Similarity: Same username pattern, timezone behaviour, or signup habits.

➢ *Evidence Graph Output and Cross-Person Attribution*

Sycek optionally exports a relational evidence graph, mapping:

- All attributes connected to the target identity.
- Cross-person links (e.g., mutual addresses, shared phone usage).
- Confidence scores (based on multi-source correlation).
- Temporal behaviour chains (e.g., same IP used on same date by two usernames).

➢ *This Enables the OSINT Investigator to:*

- Confirm identity convergence through real-world, verifiable data.
- Detect impersonation, alias use, or shared infrastructure.
- Pivot upward—from raw data to structured attribution with actionable connections.

➢ *Integration into the BTRA Workflow*

By integrating Sycek at the vertical pivoting stage of the BTRA model, the investigator gains:

- Speed and scale in resolving complex identities.
- Confidence scores based on repeated real-world use of attributes (e.g., addresses used in multiple transactional leaks).
- Precision linking between multiple partial or disguised identities.

Moreover, once Sycek refines and narrows down the evidence-supported entity map, the investigator is positioned to transition into a Top-Down exploratory loop—using the established relational core to guide deeper investigations (e.g., social, ideological, or geopolitical profiling).

➢ *Illustrative Flow Example (Extended)*

Input: john.doe@gmail.com

- *Sycek Performs:*

✓ *Fuzzy match resolution → finds john_doe1@gmail.com, johnny.d@xyz.com*
✓ *Retrieves:*

- *IP blocks from Eastern Europe, linked GPS to London*
- *Username jd_phantom used with identical device metadata*
- *Purchase address from a 2019 telecom breach*
- *Secondary phone numbers verified in food delivery service leaks*

✓ *Detects same GPS cluster used by j.doe@pm.me and johnnydoe_xx@outlook.com*
✓ *Constructs a visual graph:*

- *3 aliases*
- *2 shared IPs*
- *1 reused phone*
- *Common city with geotemporal overlap*

✓ *Provides confidence tagging (e.g., High Certainty Match) for 2 aliases*

- *Conclusion: Sycek's Role in BTRA Vertical Pivoting*

Sycek enhances the vertical pivoting phase of the Bottom-to-Top OSINT model by:

✓ Turning fragmented breach data into structured, queryable intelligence.
✓ Offering cross-entity resolution through high-value field mapping.
✓ Enabling pattern-based relational mapping and entity profiling at speed.
✓ Feeding evidence-rich nodes into follow-up investigative stages.

This tool-driven approach ensures that relational links aren't inferred loosely but are instead built on hard, breach-

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

proven data—bringing forensic-grade reliability into OSINT workflows.

These visual outputs not only provide transparency to the underlying data structures but also reinforce the evidentiary strength and traceability of the Bottom-to-Top OSINT model, particularly when driven by purpose-built systems like Sycek.

To reinforce the vertical pivoting process enabled by Sycek, the following figures illustrate various capabilities of the tool in real-time OSINT investigations:

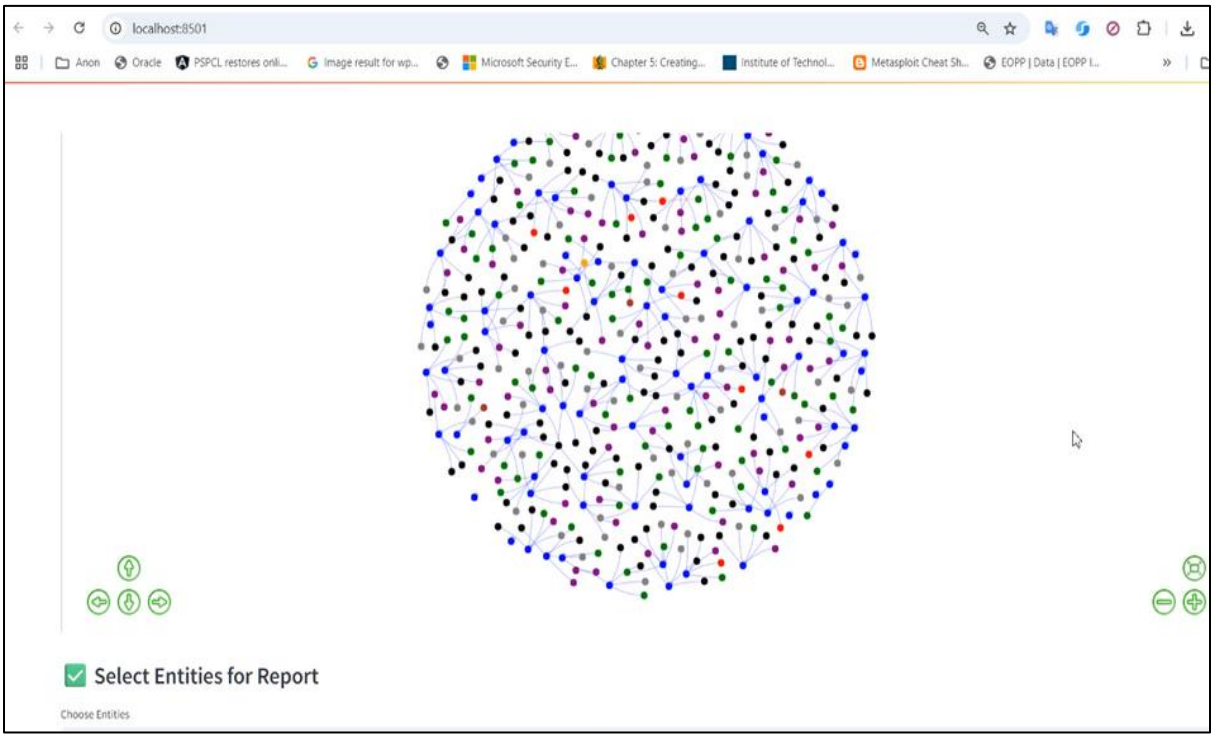


Fig 1 High-Density Entity Graph View

A zoomed-out graph view depicting the dense relational structure derived from multiple breach datasets. Each node represents an entity (email, device ID, phone, etc.), and edges signify inferred or confirmed relationships.

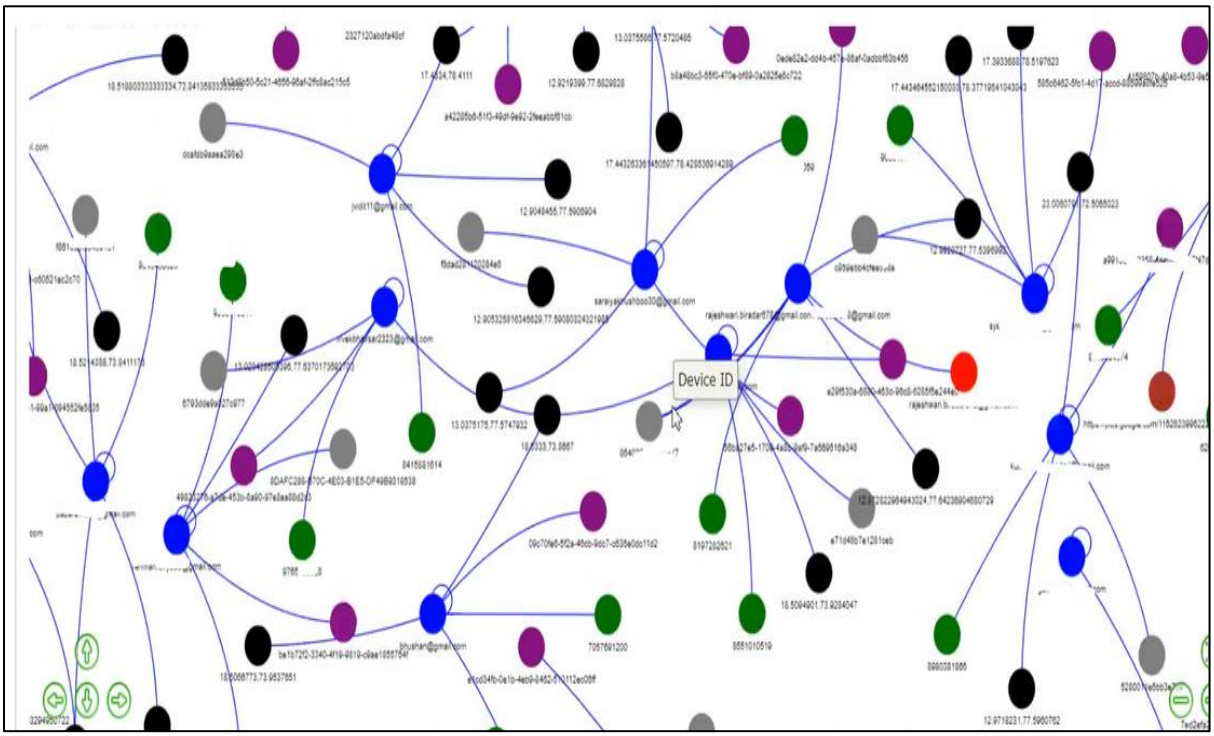


Fig 2 Mid-Level Zoom – Device-Level Linkage

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

The graph shows how device IDs (central blue nodes) connect with multiple users and attributes. This is critical in identifying shared infrastructure, impersonation, or secondary accounts tied to the same hardware or access points.

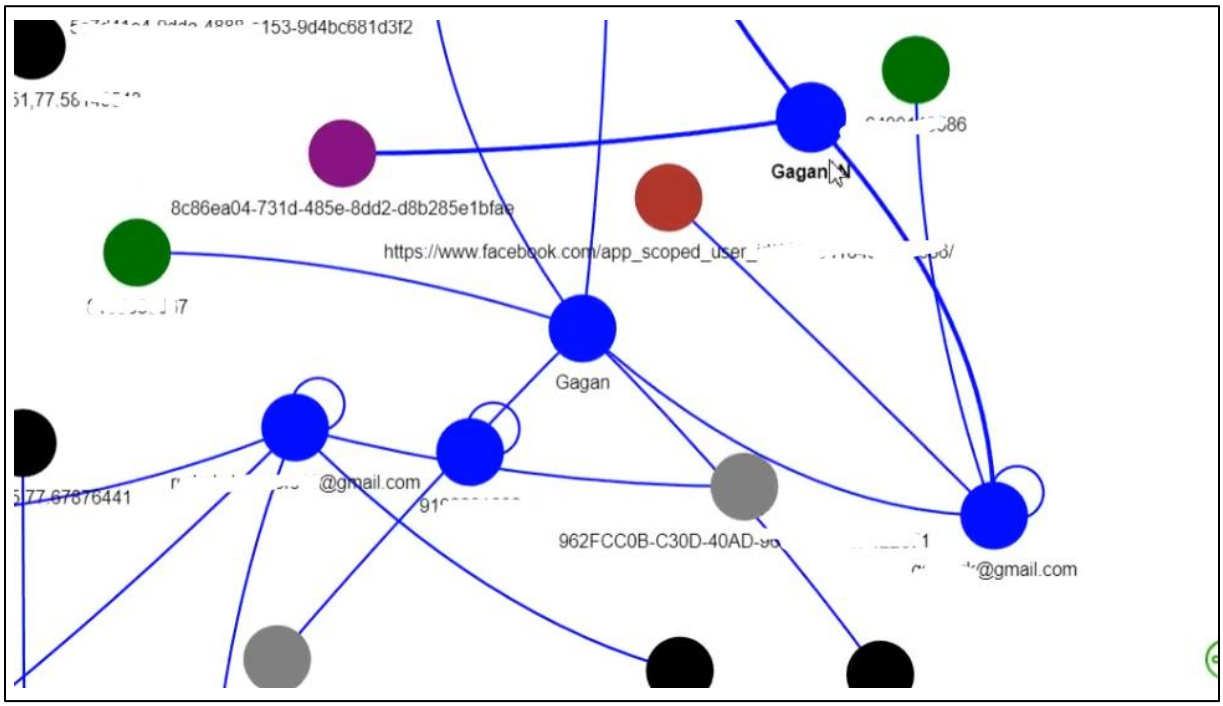


Fig 3 Alias and Social Relationship Mapping

Demonstrates cross-platform identity correlation. In this example, the node "Gagan" is tied to multiple Gmail IDs, a Facebook user ID, and correlated device or location attributes.

Fig 4 Attribute Preview for Entity Reporting

Real-time preview of selected entity attributes before report generation. This includes email, phone, location metadata (GPS, city IDs), device model, and identifiers—integrated seamlessly into a report generation pipeline.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

➢ *Step 4: Verification and Filtering*

The Verification and Filtering phase is critical within the Bottom-to-Top Relational Attribution (BTRA) model. While the earlier stages focus on gathering and mapping potentially connected attributes, this phase is dedicated to establishing credibility, consistency, and trustworthiness of the discovered links and entities—filtering out false positives, weak signals, and speculative associations.

- *Objective:*

To distinguish verified, high-confidence attribution data from loosely correlated or noise-prone artifacts—ensuring that all relationships used in subsequent conclusions are evidence-backed and logically defensible.

➢ *Key Processes in Verification and Filtering:*

- *Multi-Source Cross Validation*

✓ Purpose: Confirm that an attribute (e.g., a phone number, address, username) appears in more than one independent dataset or platform.
✓ Execution: For example, if a phone number appears in both a breached ecommerce dataset and a financial services dataset, it significantly increases the likelihood of authenticity.
✓ Tool Integration: In Sycek, cross-validation is automated by correlating attributes across independent indexed datasets with metadata tagging (source, year, type of leak).

- *Temporal and Contextual Consistency*

✓ Purpose: Ensure that associated attributes make sense over time and in context.
✓ Execution:

▪ Confirm that addresses, devices, or IPs aren't anachronistic (e.g., using a device before its release date).
▪ Check that location metadata aligns with known travel patterns or employment history.
▪ Compare last-seen timestamps for logical consistency.

- *De-Duplication and Noise Reduction*

✓ Purpose: Remove redundant or repeated entries that could bias frequency-based inference.
✓ Execution: Cluster entities with near-identical fields (e.g., same email used with slight username variations) and consolidate them into singular profiles unless divergent data is found.

- *Confidence Scoring*

✓ Each link or entity in the graph is assigned a confidence level based on:

▪ Source Reliability: Was the data from a verified breach or an unverifiable forum post?
▪ Volume of Occurrence: Is the attribute seen multiple times across datasets?
▪ Relational Strength: Is the link direct (same IP, device ID) or inferred (pattern-based aliasing)?

✓ *Example Scoring Framework:*

▪ High Confidence (≥90%): Multi-source, breach-verified, time-aligned, high-precision match.
▪ Moderate Confidence (60–89%): Single-source, partial validation, logically sound.
▪ Low Confidence (<60%): Speculative or uncorroborated link.

- *Analyst-in-the-Loop Judgement*

✓ The tool provides automation and suggestions, but final filtration allows a human investigator to:

▪ Remove entities that are statistically valid but contextually irrelevant.
▪ Manually correct false positives based on real-world logic.
▪ Add notes or flag ambiguous data for future review.

This hybrid human-in-the-loop model balances speed with accuracy, and helps avoid blind trust in automation, especially in sensitive investigations.

➢ *Illustrative Example (Verification Scenario)*

From the Sycek graph generated earlier, we identified:

- john.doe123@gmail.com linked to:

✓ jd_phantom on Telegram
✓ john_doe1990@pm.me
✓ A device ID reused across both profiles
✓ A GPS location tied to both accounts within 3 days

- *Verification Passes:*

✓ All emails were found in two independent breach sources.
✓ Device ID matched across both accounts.
✓ Location overlap consistent with known behavioural pattern.
✓ Timestamp consistency supports shared ownership theory.

- *Confidence Assigned:*

✓ Alias jd_phantom → High Confidence (92%)
✓ Secondary email john_doe1990@pm.me → Moderate Confidence (78%)

These confidence scores guide the decision to include or exclude entities from the final relational attribution graph and reporting.

- *Importance in OSINT Investigations*

Unlike speculative or surface-level analysis often found in top-down models, BTRA's Verification and Filtering step ensures:

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

- ✓ Investigations are defensible and evidence-led.
- ✓ Findings are traceable, timestamped, and backed by data provenance.
- ✓ Human biases and automation artifacts are kept in check.

Visual Demonstration – Verification & Filtering via Sycek

To operationalise verification and relational confidence scoring, Sycek enables visual exploration and filter-based refinement of connections using graph interfaces and dynamic matching logic.

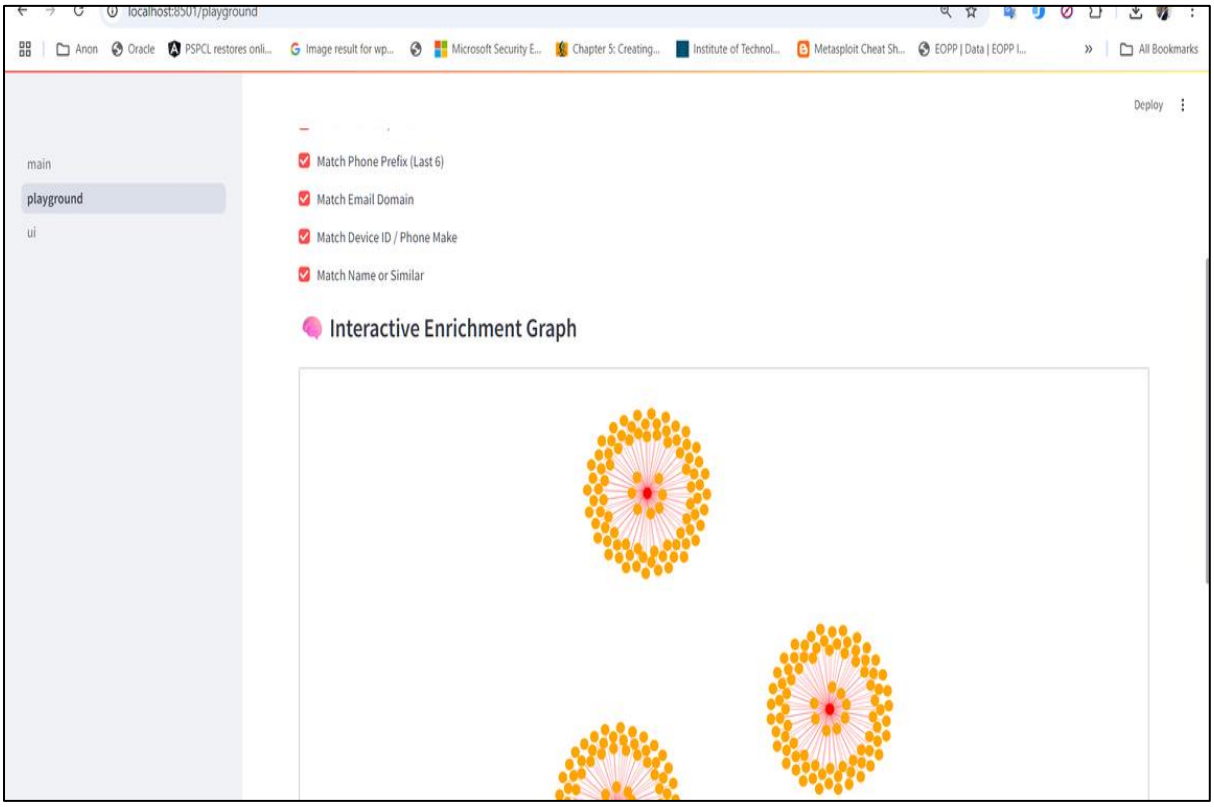


Fig 5 Interactive Enrichment Graph – Filtered Clusters

Each enriched cluster represents a unique entity with radial relationships based on shared patterns (e.g., phone prefixes, device types, or similar email usernames). These clusters are generated using selected filter criteria and can be interactively examined.

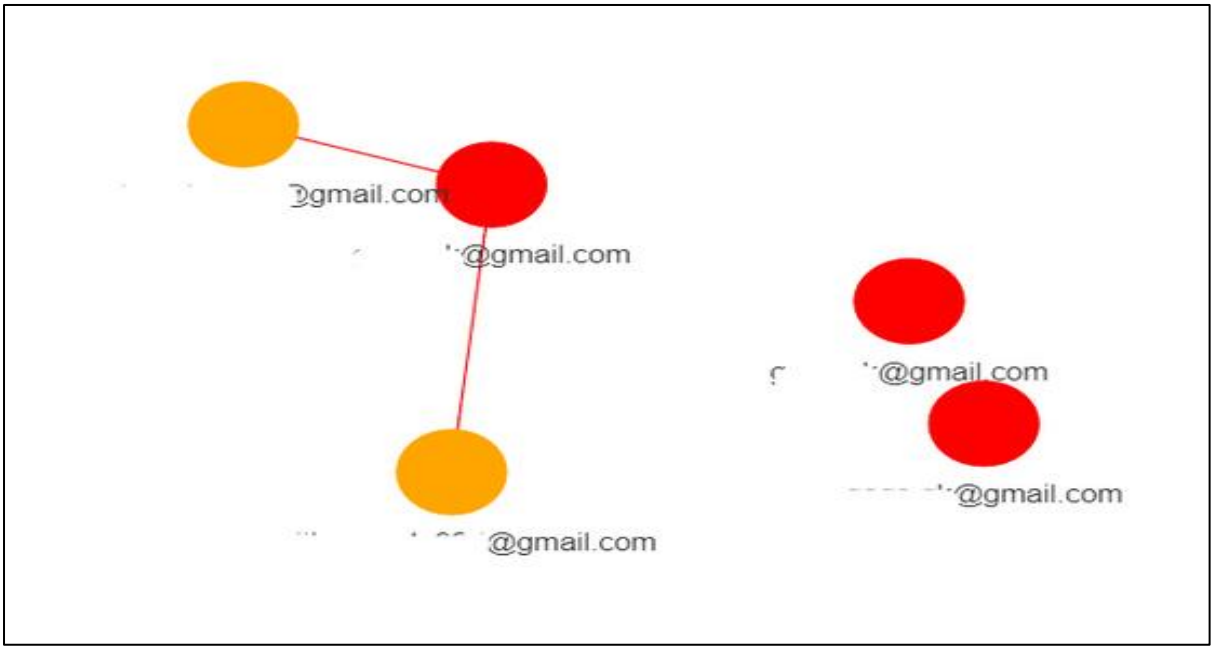


Fig 6 Filtered Micrograph of Similar Email Entities

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

Precision-focused mini-graph displaying 1:1 relationships between email addresses that share verified secondary signals (e.g., device IDs or domain usage). Ideal for identifying alias clusters.

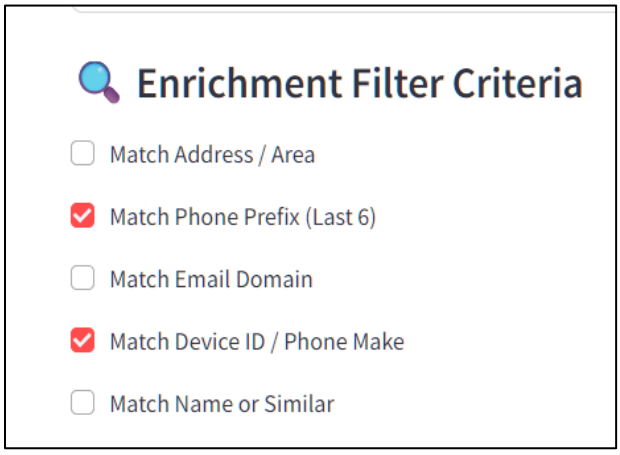


Fig 7 Enrichment Filter Criteria Selection Panel

Analysts are given full control to apply or remove relational filters—such as phone prefix, email domain, or device identifiers. This ensures noise reduction while maintaining investigative precision.

➢ *Step 5: Building the Relational Profile*

After verified data points have been filtered and scored for confidence, the Bottom-to-Top Relational Attribution (BTRA) model culminates in the construction of a comprehensive relational profile. This profile acts as the synthesized outcome of the OSINT investigation—organising verified connections, contextual details, and behavioural patterns into a structured, report-ready intelligence product.

- *Objective:*

To consolidate all enriched, verified intelligence into a singular, evidence-based profile that reflects both direct and indirect linkages tied to the original attribution point.

- *Core Components of the Relational Profile:*

The relational profile is built using a node-centric approach, where each entity (individual, alias, identifier, etc.) is represented as a node, and each verified relationship becomes a link or edge. The final profile includes:

✓ *Identity Summary*

- Primary Attribution Point: The initiating identifier (e.g., john.doe@gmail.com)
- Verified Identifiers: All validated aliases, usernames, secondary emails, phone numbers, and social handles.
- Scoring: Confidence ratings per attribute based on source count, type, and cross-confirmation.

✓ *Location & Device Metadata*

- Known Locations: First-seen and last-seen city/region data (with geocoordinates where available).
- Device IDs: Device models, IP ranges, and hardware fingerprints linked to the entity.
- Behavioural Timeline: Key activity timestamps drawn from breach data, reviews, posts, or social activity.

- *Relational Network Mapping*

✓ Connected Entities: Individuals who share attributes such as:

- Device usage
- IP/subnet overlap
- Location co-presence
- Shared employers/schools

✓ Organisational or Group Associations: Detection of possible syndicate, criminal, ideological, or community group ties.

✓ *Visual Representation*

Use network graphs generated via Sycek to:

- Visualise direct vs. indirect relationships.
- Identify cluster centres or influencer nodes.
- Segment clusters by attribute type (e.g., phone-based, geo-based, alias-based).

- *Machine-Generated Report Summary*

Sycek allows for on-demand report generation with summaries enriched via LLM (Large Language Model) prompting:

✓ Narrative Summary: Synthesised story based on evidence trail.
✓ Confidence Notes: Alerts on low-confidence areas or gaps.
✓ Exportable Output: PDF or JSON format, enabling audit, escalation, or case filing.

- *Visual Example:*

The investigator can preview linked attributes and generate an LLM-enriched narrative report directly from the UI.

➢ *Illustrative Output (Sample Profile Structure)*

Primary Email: john.doe@gmail.com
Aliases:
- john_doe
- j.doe1990@proton.me
- jd_phantom (Telegram)
Verified Phones:
- +447900XXXXXX (breach confirmed)
- +447911YYYYYY (matched via device reuse)
Known Locations:
- London (Geo Confirmed)
- Delhi (IP Confirmed, Medium Confidence)
Devices:
- Samsung Galaxy A52 (x2)
- Shared IP block 212.48.11.0/24
Relationships:

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

- Alias overlap with 2 other known identities
- Co-location pattern with user "rajesh.kumar@..." on 3 occasions
Confidence Score: 92%
Risk Level: Moderate to High (Pending further social linkage analysis)
Generated Report: Attached LLM Summary

- *Value of Relational Profiles in OSINT*

This structured profile achieves what traditional methods cannot:

✓ Proves connections using hard data (not just inference).
✓ Draws linkages across identity layers, even when obfuscated.
✓ Enables both strategic analysis and tactical response, such as social takedowns, attribution handovers, or legal escalation.

Note: *At the time of writing, Sycek has been fully developed and functionally validated across multiple synthetic and breach-verified datasets, with visual graph and attribute-based matching confirmed through internal testing. While real-world deployment in ongoing investigations is forthcoming, the tool has been engineered with readiness for operational use in law enforcement, cybersecurity investigations, and private intelligence environments.*

➢ *Step 6: Iterative Hybrid Loop with Top-Down Validation*

The Bottom-to-Top Relational Attribution (BTRA) model does not end with profile generation. The final—and perhaps most powerful—step involves iterating back into a Top-Down mode using the structured relational profile as a springboard to test, validate, or enhance the investigation further.

- *Objective:*

To reframe the verified profile into investigative hypotheses, align them with known threat actor patterns or organisational profiles, and continue exploration where new leads emerge.

➢ *Core Process Flow:*

- *Transition from Evidence to Hypothesis*

✓ Take the relational graph generated from Sycek and identify:

▪ Clusters of behaviour
▪ Potential group-level affiliations
▪ Unusual combinations of geography, devices, or aliases

✓ These are framed as investigative questions:

▪ "Is this cluster part of a coordinated fraud ring?"
▪ "Do these profiles match the TTPs of known threat actor X?"

- *Align with Known Intelligence Repositories*

✓ Cross-check profile traits with:

▪ Threat actor databases (MITRE ATT&CK, Recorded Future, MISP)
▪ Known scam syndicates
▪ Government watchlists
▪ Historic case files

✓ Sycek's structured output (including device, IP, alias patterns) can be exported for ingestion into SIEMs, threat intelligence platforms, or graph correlation tools.

- *Analyst-Led Exploration*

✓ Investigators may now go top-down:

▪ Search known group TTPs or published IOCs that match Sycek output.
▪ Use structured reports to identify related targets.
▪ Validate hypotheses or disprove false positives using grounded data.

- *Feedback Loop for Re-enrichment*

✓ If new attributes or entities are discovered from this top-down pass, they are:

▪ Re-ingested into Sycek
▪ Processed through the same Bottom-to-Top logic
▪ Verified, scored, and linked back to the main profile

This results in a live, evolving investigation graph—an iterative hybrid cycle.

➢ *Illustrative (Synthetic) Flow:*

- Bottom-Up Input: john.doe@gmail.com
- Sycek builds profile → Links 2 aliases, 3 phones, 1 IP block
- Analyst observes shared device + IP in Singapore + UK
- Hypothesis: Could this be an actor in a mule network or drop-shipping fraud?
- Cross-reference with vendor fraud reports → Similar alias cluster spotted
- Analyst adds that new alias into Sycek → more connections emerge

This dynamic loop enables deep insight generation—a departure from rigid, one-directional OSINT workflows.

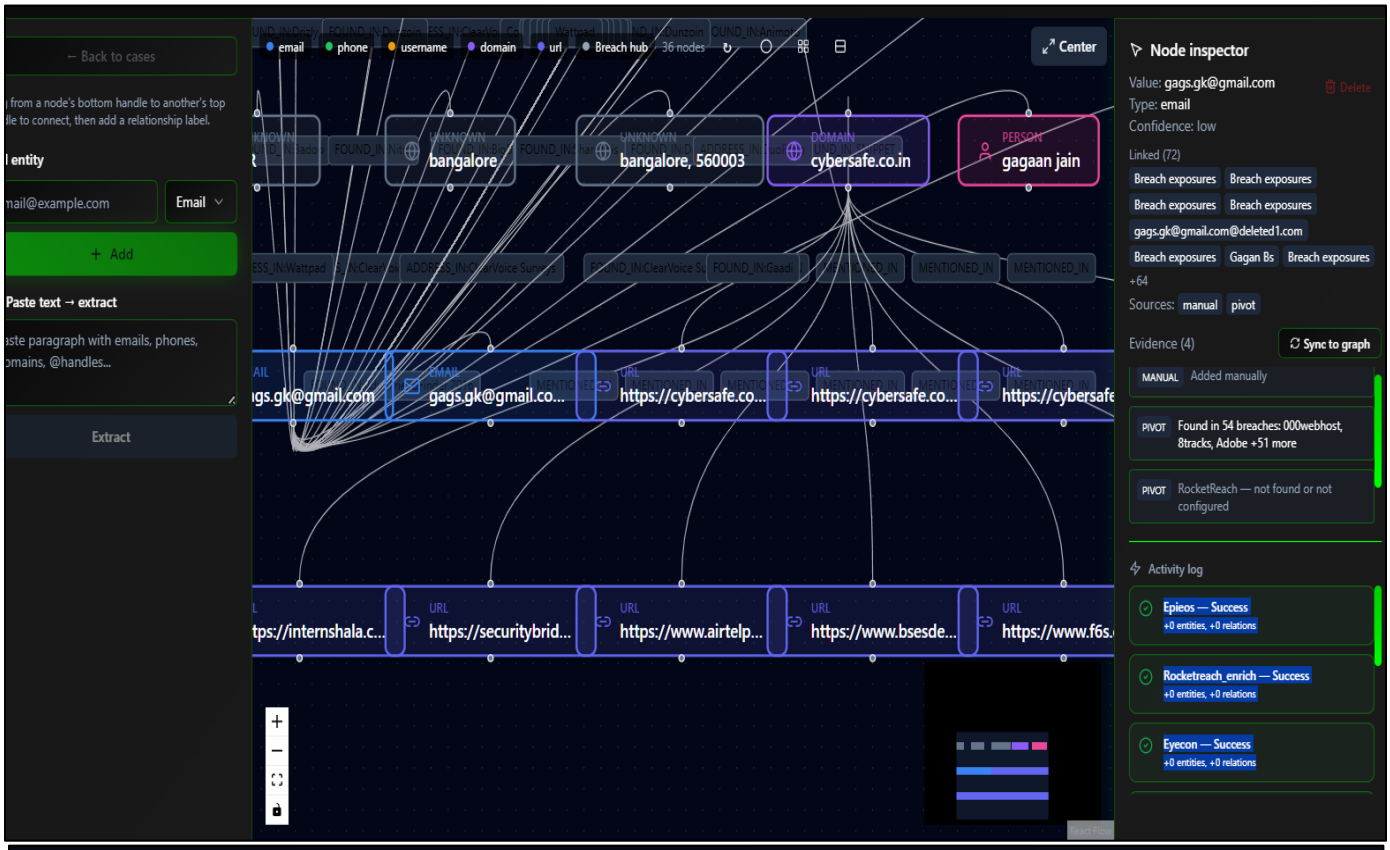
➢ *Why This Hybrid Loop Matters*

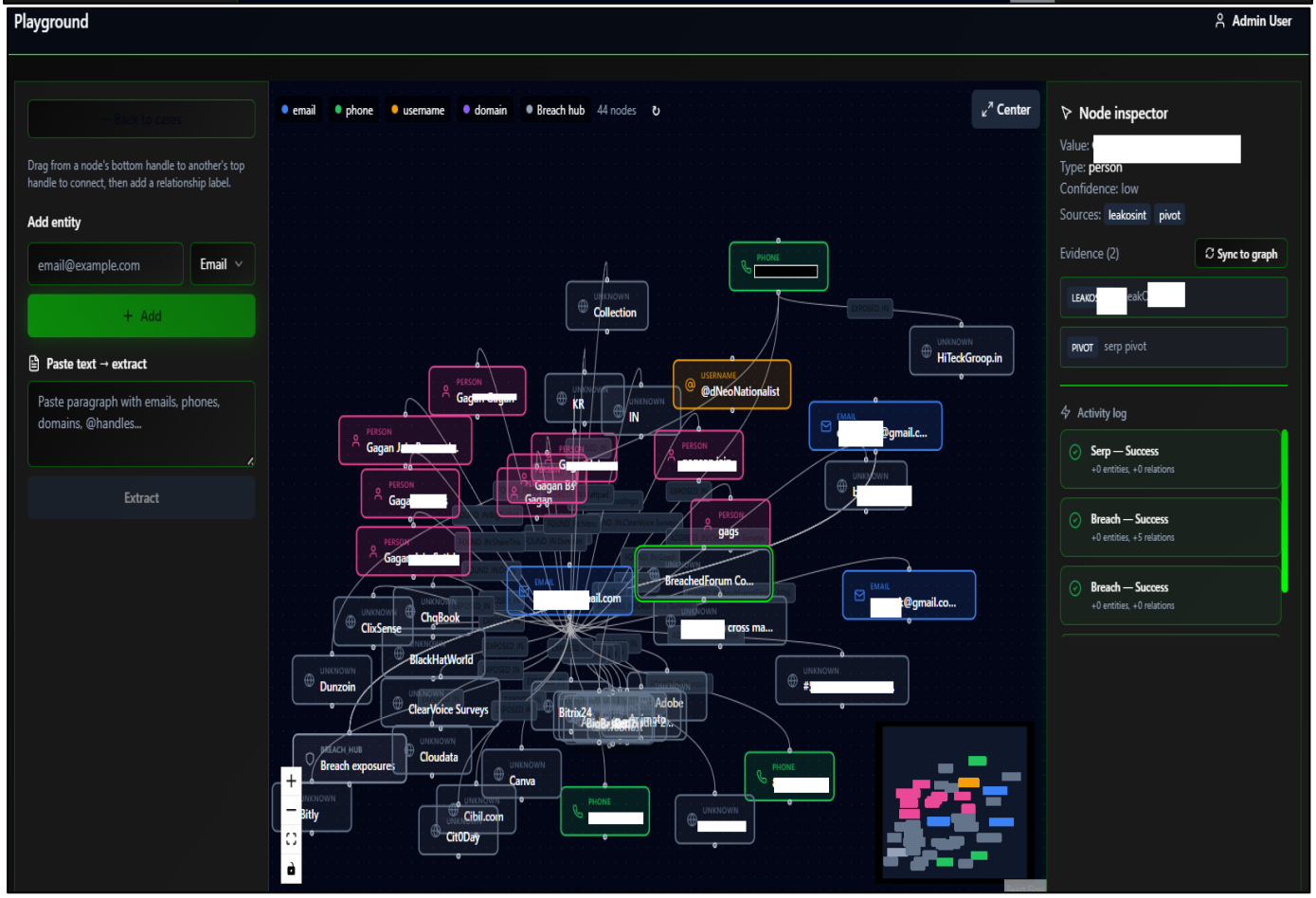
- Pure Top-Down investigations risk bias and assumption.
- Pure Bottom-Up models risk being aimless or inefficient.
- The BTRA hybrid model allows:

✓ Rapid entry via minimal input
✓ Evidence-driven validation
✓ Hypothesis-driven extrapolation
✓ Repeatable, documented logic

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

III. SYCEK TOOL SHOWCASE





ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

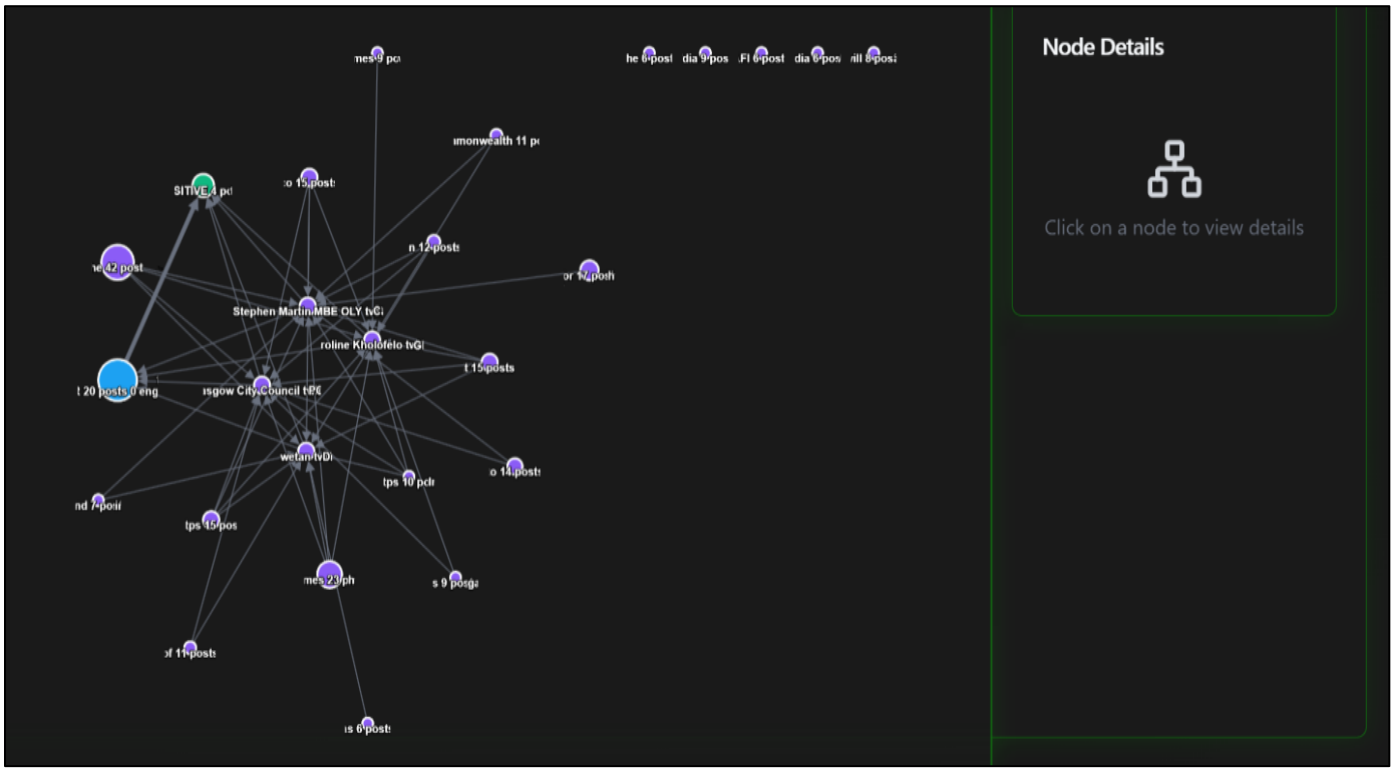


Fig 8 Sycek Tool Showcase

IV. CONCLUSION

This research presents a structured, evidence-driven methodology for OSINT investigations through the Bottom-to-Top Relational Attribution (BTRA) model. Departing from conventional top-down approaches that rely heavily on hypotheses and predefined intelligence frameworks, BTRA begins with minimal data points—such as an email or phone number—and incrementally builds validated, interconnected profiles through contextual expansion, relational linking, and rigorous verification.

At the core of this methodology lies Sycek, a custom-built OSINT engine designed to ingest, index, and visualise high-value attributes from global breach datasets. By leveraging fuzzy search, pattern recognition, and graph-based analysis, Sycek allows investigators to map digital footprints, detect alias clusters, and uncover behavioural links that may otherwise remain hidden.

Through six clearly defined steps—from initial data capture to hybrid iteration—this model introduces a repeatable, scalable workflow that prioritises data integrity, evidentiary value, and analytical transparency. The integration of a feedback loop between bottom-up discovery and top-down validation further enhances the model's adaptability to both tactical investigations and strategic threat actor profiling.

While this paper focuses on foundational capabilities and synthetic use cases, future real-world deployments of Sycek are expected to expand its applicability in law enforcement, corporate forensics, and national security operations. The BTRA methodology stands as a new standard for structured relational intelligence, bridging the gap between fragmented data and actionable insight in modern cyber investigations.

REFERENCES

[1]. Acharya, B., Saha, D., & Mohan, S. (2023). *The social structures of OSINT: Investigating crowdsourced cyber investigations*. In Proceedings of the 2023 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW). https://doi.org/10.1145/3579480

[2]. Haycox, D. (2023). *Anatomy of a phish: Breaking down email attacks like an expert*. Infosec Institute. https://resources.infosecinstitute.com

[3]. Mavroeidis, V., & Bromander, S. (2017). *Cyber threat intelligence model: An evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence*. In 2017 European Intelligence and Security Informatics Conference (EISIC). IEEE. https://doi.org/10.1109/EISIC.2017.20

[4]. MITRE Corporation. (n.d.). *MITRE ATT&CK® framework*. https://attack.mitre.org

[5]. Mohamad, A. S., Al-Qershi, O., & Nizam, M. (2022). *Evolving techniques in cyber threat hunting: A systematic review*. IEEE Access, 10, 74521–74540. https://doi.org/10.1109/ACCESS.2022.3190110

[6]. Recorded Future. (2021). *Threat intelligence handbook: A practical guide for security teams*. https://www.recordedfuture.com

[7]. SEON. (2023). *Open source intelligence (OSINT) techniques for fraud prevention*. https://seon.io/resources/osint-techniques

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1405

[8]. Skopik, F., Settanni, G., & Fiedler, R. (2016). *A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing*. Computers & Security, 60, 154–176. https://doi.org/10.1016/j.cose.2016.03.003

[9]. Tegl, J. (2020). *OSINT for business: How open source intelligence can drive strategy*. LinkedIn Articles. https://www.linkedin.com/pulse/osint-business-johan-tegl

[10]. Europol. (2022). *Internet organised crime threat assessment (IOCTA).* https://www.europol.europa.eu