# The Criminalization of Association: Surveillance, Infrastructural Identity, and the Manufacture of Conspiracy

Flyxion
Independent Researcher

October 2026

**Abstract**

Recent reporting on a federal investigation of activist organizations describes a surveillance apparatus that combines undercover infiltration, intercepted group communications, financial records, cloud infrastructure metadata, vehicle registrations, and commercially aggregated personal data. The essay concerns a change in the direction of inference: an investigation that begins with a population defined by political association and searches its accumulated traces for prosecutable relationships runs from association to crime, and so transfers the burden of proof from what was done to who was connected to whom. Graph theory is used as a diagnostic language and not as a theory of liability. The essay distinguishes the network of relationships that exists, the network that surveillance observes, and the network that prosecutors allege; shows that reachability, density, and structural centrality do not track criminal participation; and models evidentiary dependency as an AND/OR system, using Menger's theorem for its disjunctive case, to separate the number of evidentiary routes from the number of independent proofs. It proves a non-identifiability result for conspiracy under association-only observation, separates it from legal insufficiency, and characterizes the observation enrichments that would restore identifiability. It models investigative expansion as a feedback process whose growth can be driven by the connectivity of the community instead of by evidence of conduct, defines an amplification set that an independent reviewer could compute, distinguishes evidence merely discovered after designation from evidence generated by the surveillance itself, and states what provenance an investigative record must preserve for that review to be possible. A falsifiable protocol is proposed. The reported facts are treated as allegations, and the title states a hypothesis that the essay formulates and tests, not a finding it assumes; the historical comparison with COINTELPRO is treated as a question to be examined.

# 1 Introduction

A criminal investigation has traditionally been understood to begin with an offense and to proceed toward the persons responsible for it. A burglary occurs, a fraud is reported, an assault is alleged, and investigators work outward from the event, following physical evidence, testimony, and records until they arrive at suspects. The direction of inference runs from crime to participants. Modern surveillance technology makes possible a different direction. Authorities may begin with a population defined by political or social affiliation, accumulate the digital and institutional traces that population leaves, and search the resulting structure for relationships that can be described as prosecutable. The direction of inference then runs from association to crime, and the evidentiary burden is quietly transferred from the question of what was done to the question of who was connected to whom.

This essay concerns that transfer. It takes as its occasion a published investigation into federal surveillance of activist organizations in Minnesota [14], which reports, on the basis of discovery materials and interviews, that investigators assembled information from infiltrators, encrypted group communications, payment and cloud-hosting records, motor vehicle data, and commercial people-search databases. The reporting also makes comparisons to the historical program known as COINTELPRO and to the Central Intelligence Agency's domestic operation MH/CHAOS. Those comparisons, and the allegations of improper targeting on which they rest, are not accepted here as established. The allegations remain allegations unless and until they are confirmed by judicial findings or other independent verification. What the essay takes from the reporting is a pattern of capability: the combination of heterogeneous records into reconstructed networks of persons, devices, accounts, and organizations. The question is what that capability licenses.

The title states a hypothesis and not a finding. It names a possible process in which association is treated as a predicate for investigation and the investigation then produces records that appear to confirm the predicate, and the essay's purpose is to specify that process precisely enough that it can be tested and, if the evidence warrants, rejected.

The question is mathematical in an unusually literal sense. A reconstructed network of relationships is a graph, and the inferential steps that carry an investigator from a graph to an accusation can be stated precisely enough that their validity can be examined. Mikhail Lavrov's textbook *Start Doing Graph Theory* supplies the classical machinery used here: the modeling of relations by graphs, walks, paths and connected components, cut vertices and connectivity, and Menger's theorem on disjoint paths [1]. The application of that machinery to legal and political inference is not a theorem of graph theory, and nothing below should be read as claiming that it is. The argument is rather that graph theory is valuable here precisely because it is exact about what graphs can and cannot establish. Its theorems describe structure. They are silent on knowledge, agreement, and intent, which are the elements on which criminal liability for conspiracy depends.

The essay proceeds as follows. Section 2 states what the reporting alleges and what it

leaves open. Section 3 treats infrastructural identity and the aggregation of accurate records into inaccurate attributions of agency. Section 4 distinguishes three graphs and the maps between them. Sections 5 through 8 examine reachability, evidentiary weight, structural centrality, and independent corroboration. Section 9 proves a non-identifiability proposition and separates it from legal insufficiency. Section 10 models investigative expansion as a feedback process and defines the amplification set that an independent reviewer would examine. Section 11 asks what provenance an investigative record must preserve for that review to be possible. Section 12 relates the analysis to existing doctrine on conspiracy, association, and aggregation. Section 13 proposes an empirical protocol, and Section 14 concludes.

A note on notation. Edges of undirected graphs are written $uv$ or $(u, v)$ interchangeably, and arcs of directed graphs $(x, y)$. Persons are $P$, artifacts $D$, and the underlying, recorded, and alleged networks are $G^*$, $G_O$, and $G_C$, with $E^* = E(G^*)$. The observation process of Section 4 is $\Phi$, the inference process is $\Gamma$, and observation models over situations are written $\mathcal{O}$. Evidence is $Y$, the conspiracy hypothesis is $\mathcal{H}$, and the observed graph in Section 9 is $H$. Participation assignments are $\alpha$, relevance scores are rel, records are $r$, and derivations are $\Delta$. The numbers $\nu$ and $\tau$ always refer to the dependency system of Section 8.

## 2 What is alleged and what remains open

The reported investigation is notable less for any single technique than for the combination. Undercover infiltration is a long-established tool of American law enforcement, and nothing in the practice of investigating suspected crimes through informants is novel. The reporting nonetheless describes infiltration alongside the interception of group communications, the acquisition of financial and infrastructure records, and the use of commercially aggregated data, such that each source supplies a fragment of identity or association that is not apparent within any single dataset. The significance lies in the fusion.

One example in the reporting is particularly instructive. Investigators are described as having connected a cloud-hosted application to its operator by combining DNS information, SSH logs, connection records, account signup information, a payment-processor fingerprint, motor vehicle records, and a commercial investigative database. A payment account through which no transaction had ever been completed nonetheless contributed identifying information. The example shows that identification can proceed through infrastructure that a person never consciously used to disclose anything. An individual need not complete a purchase or publish a personal detail for the surrounding systems to generate traces that can be joined to a name.

There is also a question of scope. An inquiry that begins with allegations of assault, stalking, or obstruction may expand to encompass persons who communicated with the suspects, attended the same meetings, contributed to shared resources, or appear in an organizational database. Such persons may be legitimate sources of leads. They are not,

by virtue of proximity, participants in anything. The constitutional concern becomes acute where the investigative process cannot reliably distinguish participation in protected political activity from participation in a criminal agreement.

Two cautions about comparison are in order. The historical FBI program involved documented political disruption, infiltration, and abuses that exceeded any legitimate investigative purpose, and those findings were established through the congressional inquiries of the mid-1970s [3]. The contemporary reporting describes extensive surveillance and alleges improper targeting, but showing a comparable program of deliberate suppression would require evidence about authorization, institutional intent, investigative predicates, and operational conduct that discovery materials may or may not supply. Likewise, a comparison between the number of agents assigned to one investigation and the number of officers assigned to a historical domestic program does not establish relative scale, because the figures measure different organizational arrangements. The essay therefore treats the historical analogy as a hypothesis whose mechanism can be examined, and Section 13 describes how.

A different concern, perhaps more consequential than any single technique, is what may be called jurisdictional laundering. Authorities and procedures established for one purpose, such as counterterrorism, customs enforcement, or financial crime, can be repurposed for another. If each individual data request is procedurally authorized while the combined investigation exceeds the purpose that justified any of them, then oversight directed at isolated transactions can fail to detect the transformation of the whole. The unit of authorization and the unit of effect come apart. This theme returns in Sections 11 and 12.

## 3 Infrastructural identity and the aggregation problem

Let $P$ be a population of persons and $D$ a collection of artifacts: accounts, devices, addresses, vehicles, payment instruments, domains, and communication channels. Identification in the infrastructural setting is not the discovery of a single attribute. It is the accumulation of relations between persons and artifacts. Define the identity graph

$$G_I = (P \sqcup D,\, E_I),$$

a bipartite graph in which an edge joins a person to an artifact with which some recorded relation holds. The relations are of different kinds. A person may have registered an account, paid for it, accessed it from a particular address, or merely appeared in a record that mentions it. Each kind is a distinct relation, and the distinctions matter because they do not compose.

An investigative system tries to construct an identity resolution function $\rho : D \to P$ assigning to each artifact the person responsible for it. In real datasets this function is frequently not well defined. In terms of the identity graph, resolution is unambiguous for an artifact $d$ only if its neighborhood $N(d) \subseteq P$ has exactly one element. Devices are shared,

addresses contain several residents, accounts have several operators, identifiers are reused, and records contain errors. The more honest representation is a conditional distribution

$$q(p \mid d, Y),$$

where $Y$ is the available evidence, and in many cases that distribution is not concentrated.

A more fundamental difficulty remains even when it is. Suppose the relation of registration $R_{\text{reg}} \subseteq P \times D$ is established with high confidence for a given person and a given account. This does not establish that the person operated a particular application hosted through the account, authorized a particular transaction, or took part in a particular communication, since those are relations $R_{\text{op}}$, $R_{\text{auth}}$, and $R_{\text{comm}}$ that are not implied by $R_{\text{reg}}$. An investigation that treats the composition of identification with agency as automatic converts accurate records into an inaccurate attribution. This is the central hazard of infrastructural identity: *the aggregation of true statements about records can yield a false statement about conduct*, and the falsity is introduced not by any individual datum but by the inferential step that joins them.

The problem is aggravated by the fact that the aggregation is mostly invisible to the person it concerns. No one decides to be identifiable by the conjunction of a connection log, a signup field, a payment fingerprint, and a registration database. The identifying power arises from the combination, and the combination is performed by the investigator, so that the individual records can each appear innocuous and each be lawfully held while the assembled profile is something that no one consented to create.

## 4 Three graphs and the maps between them

An investigator observes a heterogeneous relational structure

$$\mathcal{G}_{\text{obs}} = (P \sqcup D,\, E,\, \text{typ},\, \sigma),$$

where $E$ is the set of observed relationships, typ assigns each edge a relationship type, and $\sigma$ records provenance, temporal scope, and evidentiary status. The types include communication, co-presence, financial exchange, shared infrastructure, organizational membership, and documented coordination. Lavrov's opening chapters make a point that is easy to overlook: a graph does not carry its own interpretation [1]. A map-coloring problem and a matching problem can both be modeled by graphs, but the edges mean different things, and the meaning is supplied by the decision about which relationship the question concerns. A communication edge is not an agreement edge. A financial transfer is not necessarily criminal financing. Two people attending the same meeting are not thereby engaged in a common enterprise.

An investigative system that suppresses these distinctions projects the heterogeneous

structure onto a simple graph

$$\pi : \mathcal{G}_{\text{obs}} \longrightarrow G_A = (P, E_A),$$

in which two persons are adjacent whenever some observed relationship connects them. The projection is information-losing. It preserves the existence of an association and discards the kind of association, the circumstances of its occurrence, and the reliability of its observation. It is also far from injective: structurally different bodies of evidence, such as a triangle of mutual correspondence and a triangle formed by three separate unrelated record matches, project to the same graph. A reader of $G_A$ cannot recover which is which. The danger arises when the resulting graph is read as though its edges represented a single substantive relation, so that a heap of ordinary activities acquires the look of a coherent association network. The graph is mathematically valid. Its interpretation may not be. The three structures are related as follows. $\mathcal{G}_{\text{obs}}$ is the heterogeneous structure that is observed, $G_A = \pi(\mathcal{G}_{\text{obs}})$ is its projection to a simple graph on persons, and $G_O$ is the projected network as it is recorded, in contrast with $G^*$, the projected network as it exists. The two coincide, $G_O = G_A$, when the recorded relationships are exactly the observed ones, and the recorded graph $G_t$ of Section 10 is $G_O$ at time $t$.

It is useful to separate three objects. Let $G^*$ be the network of social relationships that actually exists, $G_O$ the network recorded through surveillance, and $G_C$ the network of criminal relationships alleged by prosecutors. The observation process

$$G^* \xrightarrow{\Phi} G_O$$

is partial and may be biased. It omits relationships, introduces mistaken identity matches, and disproportionately records those who are easier to monitor, so that its output reflects the instrument as well as the population. The inferential process

$$G_O \xrightarrow{\Gamma} G_C$$

consists of investigative assumptions, evidence assessments, legal classifications, and prosecutorial decisions. Neither map is guaranteed to preserve the truth that matters. The requirement that follows is stated negatively:

$$(u, v) \in E(G_O) \quad \not\Rightarrow \quad (u, v) \in E(G_C).$$

Further, even an inferred criminal edge is not an adjudicated legal conclusion. Liability depends on the applicable offense, the conduct and mental state of the defendant, the admissibility and sufficiency of the evidence, and the governing standard of proof. The essay therefore resists replacing one unjustified simplification with another. Graph theory can expose structural and evidentiary weaknesses. It cannot substitute for legal reasoning.

## 5 Connectivity is not conspiracy

In an undirected graph, reachability is an equivalence relation: two vertices lie in the same connected component exactly when a walk joins them [1]. This gives an exact description of one mechanism by which investigative scope can grow. For a seed subject $s$ define the set of persons within $k$ steps,

$$B_k(s) = \{v \in P : \mathrm{dist}_{G_A}(s, v) \le k\}.$$

Then

$$B_0(s) \subseteq B_1(s) \subseteq B_2(s) \subseteq \cdots,$$

and in a finite graph the union is the connected component $C(s)$ containing $s$. The notation is used for other graphs and for sets of seeds with a superscript: for a graph $G$ and a vertex set $S$, $B_k^G(S) = \{v : \mathrm{dist}_G(s, v) \le k \text{ for some } s \in S\}$, and $C_G(s)$ is the component of $s$ in $G$, the superscript and subscript being omitted when $G = G_A$. An investigation that treats reachability as sufficient grounds for continued scrutiny can expand from one person to an entire connected community.

Consider an activist who maintains a public website. A journalist interviews the activist, a software developer hosts the site, a donor contributes to its operating costs, and several residents attend a public meeting advertised through it. All of them become connected through a small number of edges, though their activities, knowledge, intentions, and legal responsibilities differ greatly. Writing $u \rightsquigarrow_{G_A} v$ for the existence of a path in the association graph, and $u \sim_{\mathrm{crim}} v$ for an independently established criminal relationship between $u$ and $v$, the inferential fact is that

$$u \rightsquigarrow_{G_A} v \quad \not\Rightarrow \quad u \sim_{\mathrm{crim}} v,$$

and no theorem of graph theory licenses the implication.

The failure runs in both directions, and each direction can be exhibited concretely. Consider first a conspiracy that is sparse. Suppose an organizer instructs four couriers, none of whom knows of the others, and each delivers only to the organizer. The association graph is the star $K_{1,4}$. The organizer is a cut vertex and the couriers are pairwise nonadjacent, yet a prosecutor could allege an agreement between the organizer and each courier, and possibly a single agreement joining all five. A chain arrangement, in which each participant knows only the person before and the person after, has association graph the path $P_5$, with a diameter of four and no triangles. Neither graph is dense, and an investigator who required density as a sign of conspiracy would acquit both. Consider next a graph that is as dense as a graph can be. Five members of a neighborhood association who email one another about each item on a public agenda, and who all attend every meeting, form the complete graph $K_5$, in which every pair is adjacent and no vertex is a cut vertex. Every edge in it represents lawful communication, and the graph is indistinguishable, as a graph,

from five conspirators who all know one another. The same graph is thus realized by an entirely lawful situation and by a criminal one, and the sparse graphs above are realized by criminal situations that density-based reasoning would miss. Neither connectivity nor density is sufficient for criminal agreement, and neither is necessary for it. The topology of association is simply the wrong kind of quantity to settle the question.

## 6 Evidentiary weight and the likelihood ratio

The more insidious danger is not the mere collection of additional edges but the gradual conversion of uncertain associations into apparently corroborating evidence. Suppose an investigator assigns each observed edge a weight $w(e) \in [0, 1]$ representing confidence that the recorded relationship occurred. A tempting analytic step is to form a path score

$$W(p) = \prod_{e \in p} w(e)$$

and to read a high value as evidence of coordinated criminal activity. Even if the edge observations were independent and the weights perfectly calibrated, this product would measure the joint reliability of the observations along the path. It would say nothing about the probability of a criminal agreement. The weight of an edge is confidence that a relationship was correctly recorded. It is not confidence that the relationship was criminal.

For a conspiracy hypothesis $\mathcal{H}$ and observed evidence $Y$, the relevant evidentiary quantity is the likelihood ratio

$$\Lambda(Y) = \frac{\mathbb{P}(Y \mid \mathcal{H})}{\mathbb{P}(Y \mid \neg\mathcal{H})}.$$

An association becomes incriminating only to the extent that it is more expected under the criminal hypothesis than under the relevant lawful alternatives. This bears on political organizations because lawful collective action already generates dense networks of communication, shared resources, coordinated schedules, and repeated encounters, so that a dense or internally coherent graph is an outcome with substantial probability under $\neg\mathcal{H}$. That observation alone does not fix the value of $\Lambda(Y)$. The ratio approaches one only when the observed structure is about equally probable under the criminal and the lawful models, and whether that holds depends on how each model is specified. Density can be weak evidence under one pair of models and informative under another: a pattern that lawful organizations rarely produce, such as coordination confined to a narrow window around a specific offense, an unusual restriction of a channel to a small number of persons, or concealment practices that serve no stated lawful purpose, may be several times more probable under $\mathcal{H}$ than under $\neg\mathcal{H}$ even when ordinary density is not. A comparison of two patterns on the same dense graph shows the difference. In the first, a group channel is active continuously for months, with members joining and leaving, and this is about equally expected whether the group is a lawful organization or a criminal one, so that $\Lambda(Y)$ is close to one. In the second,

a channel is created hours before a specific event, restricted to persons present at the event, and abandoned afterward. That pattern is more expected under $\mathcal{H}$ than under a lawful organization with no event-specific purpose, but it is also expected under lawful event-day logistics such as arranging transport to a protest, so $\Lambda(Y)$ exceeds one by an amount that depends on which lawful alternative is modeled. Evidence of this kind is graded, and it is read from the specific pattern and the specific alternative, never from density alone. The claim defended here is therefore conditional. Structural features that lawful organizing produces readily carry little weight against it, and the weight of any feature must be assessed against an explicit lawful alternative rather than against the absence of any explanation. The investigative problem is one of competing explanations: the same structural evidence may be consistent with a lawful protest network, with an organization coordinating illegal conduct, or with a mixture in which different participants do different things, and graph structure alone does not adjudicate among them.

## 7 Cut vertices, intermediaries, and investigative leverage

A cut vertex is a vertex whose deletion increases the number of connected components, and vertex connectivity measures how many vertices must be deleted to disconnect a graph [1]. These notions explain why certain individuals become disproportionately important to investigators. Consider a volunteer who administers a shared channel connecting otherwise separate neighborhood groups. In the observed network the volunteer may be a cut vertex, and removing that person would separate several groups. Their structural position makes them an important intermediary. It does not establish that they direct, control, or even know the activities of everyone connected through them.

Three quantities should therefore be kept apart: the structural centrality $\mathrm{Cen}(v)$ of a vertex, its organizational authority $\mathrm{Auth}(v)$, and its legal responsibility $\mathrm{Resp}(v)$. These need not be correlated. A person may be structurally central because they perform clerical work, maintain technical infrastructure, or provide services to many unrelated participants, and such a person is attractive to investigators precisely because access to their accounts exposes several otherwise separate groups. The mathematical observation is that a vertex joining components can provide unusually broad access. The legal observation is that the utility of surveilling an intermediary does not establish the legitimacy of treating that intermediary as an organizer.

The point has an old doctrinal counterpart. A hub joined to many spokes, with no relations among the spokes themselves, is a star graph, and the star is exactly the configuration in which the hub is a cut vertex and the periphery is pairwise unconnected. The Supreme Court addressed this configuration in *Kotteakos v. United States*, where a central figure had arranged many separate fraudulent loan transactions for different participants who were not otherwise connected, and the Court reversed the convictions because the proof showed multiple conspiracies in place of the single conspiracy charged and the resulting prejudice could

not be dismissed as harmless [4]. The Court itself described the pattern as one of separate spokes meeting at a common center without a rim. In graph terms the star has no edges among the spokes, but the doctrinal question is not whether rim edges exist. Participants may share a single agreement without communicating directly with one another, and the question is whether the evidence establishes a common agreement. The absence of rim edges is the absence of one kind of evidence for such an agreement, and the structural fact that every spoke is reachable from every other through the hub is just reachability, which does not supply it. *Kotteakos* is a judicial recognition that reachability does not amount to unity of agreement.

## 8 Menger's theorem and the illusion of independent corroboration

Menger's theorem states that for two nonadjacent vertices the maximum number of internally vertex-disjoint paths between them equals the minimum size of a vertex cut separating them [2, 1]. The theorem suggests a way of examining claims that several investigative sources independently establish the same relationship. Suppose an individual is connected to an organization through five apparently different evidentiary paths: a payment processor, a website, a telephone record, an address database, and a vehicle registration. Five paths appear to be five confirmations. But if each depends in the end on the same account identifier, the same data broker, or the same mistaken identity match, then they share a point of failure.

To make this precise, the dependency structure must first be given a form that can express how evidence combines, because evidence combines in two ways. A conclusion may be supported by any one of several records, which is a disjunctive dependency, or it may require several records jointly, which is a conjunctive one. An identification of an account operator might require both an account registration and a separately established correspondence between an address and a person, neither of which supports the conclusion alone. An ordinary directed path cannot distinguish the two cases. The structure is therefore modeled as an *AND/OR dependency system*: a finite directed acyclic graph $G_D = (V_D, F)$ whose vertices are source records, derived records, identifiers, and inference steps, with an arc $(x, y)$ when $y$ is obtained from $x$, and with each vertex that has in-neighbors labeled either AND, meaning it is derivable only when all of its in-neighbors are available, or OR, meaning it is derivable when at least one is. The vertices without in-neighbors form the set $V_0$ of source records. Let $z \in V_D$ be a conclusion. A *derivation* of a vertex $y$ is a vertex set defined recursively from the sources. If $y$ is a source record, $\{y\}$ is its only derivation. If $y$ is an OR vertex with in-neighbors $x_1, \ldots, x_k$, a derivation of $y$ is $\{y\} \cup \Delta$ where $\Delta$ is a derivation of some $x_i$. If $y$ is an AND vertex with in-neighbors $x_1, \ldots, x_k$, a derivation of $y$ is $\{y\} \cup \Delta_1 \cup \cdots \cup \Delta_k$ where each $\Delta_i$ is a derivation of $x_i$. A *proof* of $z$ is a derivation of $z$ that is inclusion-minimal among derivations of $z$. Because $G_D$ is acyclic the recursion is

well founded, so every derivation is rooted in source records, and each records which parent supports each OR vertex it passes through. Two quantities describe the structure of proofs. The *independence number* $\nu(z)$ is the maximum cardinality of a family of proofs of $z$ no two of which share a vertex other than $z$. The *fragility number* $\tau(z)$ is the minimum size of a set $X \subseteq V_D \setminus \{z\}$ of vertices whose deletion leaves $z$ with no derivation, that is, the least number of records or origins whose failure destroys the conclusion. The deletable vertices include intermediate records as well as sources, since an intermediate record such as a broker entry can fail or be excluded. A deleted vertex is unavailable and is not recomputed from its surviving parents, because deletion models the failure or exclusion of that record as evidence, which is the sense in which deleting an intermediate vertex differs from deleting its sources.

Weak duality holds in general: $\nu(z) \leq \tau(z)$. If $X$ is a deleting set and $\Delta_1, \ldots, \Delta_n$ are proofs no two of which share a vertex other than $z$, then each $\Delta_i \setminus \{z\}$ must contain an element of $X$, for if $\Delta_i \cap X = \emptyset$ the derivation $\Delta_i$ survives deletion of $X$ and leaves $z$ derivable, and these elements are distinct by disjointness, so $|X| \geq n$. Equality can fail once conjunctive dependencies are present. Let $z$ be an OR vertex with in-neighbors $u_1, u_2, u_3$, each an AND vertex, where $u_1$ requires source records $r_1$ and $r_2$, $u_2$ requires $r_2$ and $r_3$, and $u_3$ requires $r_1$ and $r_3$. The three proofs of $z$ are $\{z, u_1, r_1, r_2\}$, $\{z, u_2, r_2, r_3\}$, and $\{z, u_3, r_1, r_3\}$, and any two of them share a source record, so $\nu(z) = 1$. Deleting any single vertex other than $z$ leaves one derivation intact, while deleting $\{r_1, r_2\}$ destroys all three, so $\tau(z) = 2$. The conclusion has no two independent proofs and yet is not destroyed by the failure of any one record. The two numbers answer different questions, how many separate proofs exist and how many failures the conclusion can survive, and an audit should report both.

Menger's theorem enters in the disjunctive case. If every derived vertex is an OR vertex, a proof is a directed path from a source record to $z$. Adjoin a new vertex $\hat{s}$ with an arc to every vertex of $V_0$ and treat $\hat{s}$ and $z$ as the terminals. The directed vertex version of Menger's theorem, valid when $\hat{s}$ and $z$ are not joined by an arc, states that the maximum number of directed $\hat{s}$–$z$ paths that are internally vertex-disjoint equals the minimum size of a vertex cut excluding both terminals, and therefore $\nu(z) = \tau(z)$ in the disjunctive case. The exclusion of the terminals is essential. If $\hat{s}$ were an admissible cut element the minimum would equal one whatever the structure, since removing $\hat{s}$ severs every route, and the quantity would carry no information. Because the cut may contain source records themselves, a single record on which every route depends, such as one data broker entry that feeds each downstream database, is counted as a common dependency, which is the intended behavior. The same value is a maximum flow when every vertex other than the terminals is given unit capacity, which expresses that a record either is or is not a single point of failure.

The count depends on two further modeling choices that should be stated. The first is granularity. If every source record is its own vertex, two records drawn from the same

upstream collection are counted as independent. If records are grouped by originating institution or collection, so that all entries from one commercial database share a vertex, the same evidence yields smaller values of $\nu$ and $\tau$. The first measures redundancy of records and the second redundancy of origin, and an auditor should report the second whenever correlated collection is plausible. The second is the representation of partial corroboration. A record confirmed by an independent registry should be represented structurally, as an OR vertex fed by two separate sources, and not by assigning the record a larger capacity according to an informal count of corroborations. A capacity would require an independently justified interpretation, and without one it would simply restate the analyst's confidence in numerical form.

A presentation asserts some number $m$ of routes, which need not be disjoint, and what matters is its relation to the structure beneath it. Define the *apparent redundancy* $m/\nu(z)$. A presentation with five routes and $\nu = 5$ has no apparent redundancy beyond what the structure supports, while five routes with $\nu = 1$ has a ratio of five and presents a single proof as five. If $\nu(z) = 1$ there are no two independent proofs, and the apparent multiplicity of supports is an artifact of presentation. The fragility number is a separate measure. If $\tau(z) = 1$ there is a single vertex, either a record or an origin, whose removal destroys every derivation, which implies $\nu(z) = 1$, whereas $\nu(z) = 1$ does not imply $\tau(z) = 1$, as the example above, with $\nu = 1$ and $\tau = 2$, shows. The distinction can be stated as

$$\text{number of evidentiary paths} \neq \text{number of independent proofs},$$

with the second quantity bounded above by the number of proofs and by $\tau(z)$.

A caution is necessary. The system $G_D$ records how derivations are drawn, and the numbers $\nu$ and $\tau$ characterize structural redundancy within it. They do not establish that the inference rules encoded in the AND and OR labels are valid, for instance that a shared network address identifies a person, and they do not establish statistical independence, since even disjoint derivations may rest on correlated sources or shared institutional assumptions, as when two databases are populated from the same upstream collection. Structural redundancy and evidentiary redundancy are therefore not automatically equivalent. The application is diagnostic. The analysis can expose shared dependencies that a presentation of evidence might obscure, and a separate probabilistic and legal assessment must determine whether the surviving evidence amounts to genuinely independent corroboration.

## 9 A non-identifiability result

The distinction between association and agreement can be given a precise form. Let $\Omega$ be a set of social situations. Each situation $\omega \in \Omega$ comprises a relational structure $R(\omega)$ on the population and a participation assignment $\alpha(\omega) : P \to \{0, 1\}$, where $\alpha(\omega)(v) = 1$ means that $v$ is party to a criminal agreement in $\omega$. An observation model is a map $\mathcal{O} : \Omega \to \mathcal{G}$ into

a space of graphs, the situation-level counterpart of the observation process $\Phi$ of Section 4. Say that the model is *association-only* if $\mathcal{O}(\omega)$ depends on $\omega$ only through the existence of recorded associations, and that it is *lawful-explainable* on an observed graph $H$ if at least one situation $\omega'$ with $\mathcal{O}(\omega') = H$ has $\alpha(\omega') \equiv 0$, so that the observed graph could have arisen with no criminal agreement present. This is a weak condition. It does not require that every situation producing $H$ have a lawful counterpart, only that the lawful explanation is available for $H$ itself.

**Proposition 1** (Non-identifiability of conspiracy from association topology)**.** *Let $H$ be an observed association graph on which the observation model is lawful-explainable, and suppose some situation $\omega$ with $\mathcal{O}(\omega) = H$ has $\alpha(\omega) \not\equiv 0$. Then no function $f : \mathcal{G} \to \{0,1\}^P$ satisfies $f(\mathcal{O}(\omega)) = \alpha(\omega)$ for every $\omega \in \mathcal{O}^{-1}(H)$.*

*Proof.* By hypothesis there is $\omega$ with $\mathcal{O}(\omega) = H$ and $\alpha(\omega) \not\equiv 0$. By lawful-explainability there is a situation $\omega'$ with $\mathcal{O}(\omega') = H$ and $\alpha(\omega') \equiv 0$. Then $\alpha(\omega) \neq \alpha(\omega')$ while $\mathcal{O}(\omega) = \mathcal{O}(\omega')$. Any function $f$ of the observation assigns the same value $f(H)$ to both situations, so it equals at most one of $\alpha(\omega)$ and $\alpha(\omega')$. Hence no such $f$ is correct on all of $\mathcal{O}^{-1}(H)$. □

The result is an identifiability statement under explicitly limited observations, and it is not a claim that conspiracy can never be inferred from communications or network evidence. If edges carry reliable information about agreement, intent, or conduct, the lawful-explainability hypothesis fails and the proposition does not apply. Nor does the proposition forbid probabilistic reasoning. A Bayesian observer with a prior over $\Omega$ can assign a posterior probability of participation given $H$, and that posterior will generally be neither zero nor one. What the proposition shows is that the posterior is not determined by the topology of $H$ but by the prior over situations and by the relative probabilities with which lawful and unlawful situations generate $H$, which is the likelihood-ratio point of Section 6 in another form. The proposition identifies the precise place at which further evidence becomes necessary: evidence that distinguishes situations producing the same graph.

Two notions that are easily conflated should be kept apart. *Observational non-identifiability* is a property of the observation map: a participation assignment is not identifiable from $\mathcal{O}$ when two situations with different assignments yield the same observation. It is a mathematical statement about what a given body of data can distinguish, and it can be established or refuted without reference to any legal standard. *Legal insufficiency* is a conclusion about whether the evidence meets the requirements of the offense charged, which depend on the elements of the offense, the admissibility of the evidence, and the governing standard of proof, and which is settled by substantive law and by the finder of fact. Neither implies the other. A graph could be observationally identifying of agreement, as when every edge is an authenticated record of an agreement, and still be legally insufficient because the agreement was to do something lawful. A graph could be observationally non-identifiable and the prosecution still legally sufficient, because other evidence in the case, such as testimony or documents expressing intent, distinguishes the situations that the graph alone cannot. The

proposition of this section establishes the first property for association-only observation. It supports an argument that association evidence alone should not be relied upon to carry the inference to agreement, and that argument is completed only by the legal standards and not by the mathematics.

The negative result invites a constructive question: what enrichment of the observation model would restore identifiability, and at what cost in intrusion? Observation systems can be compared by refinement. Write $\mathcal{O} \preceq \mathcal{O}'$ if $\mathcal{O} = g \circ \mathcal{O}'$ for some map $g$ that discards information, so that $\mathcal{O}'$ is at least as informative as $\mathcal{O}$. Strictly this is a preorder, since two distinct maps can determine each other, as when a binary observation is encoded in two different ways. Identifying maps with $\mathcal{O} \preceq \mathcal{O}'$ and $\mathcal{O}' \preceq \mathcal{O}$ yields a partial order on the resulting equivalence classes, and the order is not linear, because observation systems need not be nested. An observation containing evidence of conduct but lacking the content of communications is not comparable to one containing content but no conduct evidence. A convenient chain is obtained by cumulative construction. Let $\mathsf{Assoc}$ denote the association record, $\mathsf{Time}$ the timing of events, $\mathsf{Cont}$ the content of communications, and $\mathsf{Cond}$ evidence of conduct such as observed acts or documented transactions, and define

$$\mathcal{O}_1 = \mathsf{Assoc}, \qquad \mathcal{O}_2 = (\mathsf{Assoc}, \mathsf{Time}), \qquad \mathcal{O}_3 = (\mathsf{Assoc}, \mathsf{Time}, \mathsf{Cont}), \qquad \mathcal{O}_4 = (\mathsf{Assoc}, \mathsf{Time}, \mathsf{Cont}, \mathsf{Cond}),$$

so that $\mathcal{O}_k \preceq \mathcal{O}_{k+1}$ with $g_k$ the map that drops the last component. The chain is a choice of one path through the partial order. It is chosen as a natural cumulative order that roughly tracks one common progression of legal showings, though the correspondence is imperfect, since some conduct evidence, such as a photograph of a solitary act obtained by physical surveillance, requires no warrant while interception of content ordinarily does. Other paths, such as adding conduct before content, are equally legitimate.

Every observation map $\mathcal{O}$ induces an equivalence relation $\omega \equiv_{\mathcal{O}} \omega'$ if and only if $\mathcal{O}(\omega) = \mathcal{O}(\omega')$, and the participation assignment is identifiable from $\mathcal{O}$ exactly when $\alpha$ is constant on every equivalence class. Identifiability is monotone under refinement: if $\mathcal{O} \preceq \mathcal{O}'$ and $\alpha$ is identifiable from $\mathcal{O}$ through some $f$, then it is identifiable from $\mathcal{O}'$ through $f \circ g$. The converse fails, and the proposition above shows that the first observation $\mathcal{O}_1 = \mathsf{Assoc}$ already fails to identify $\alpha$. Timing can distinguish a communication pattern that precedes an offense from one that does not, and content can distinguish agreement from discussion, but neither guarantees identification, and in general only evidence of agreement and conduct comes close.

The hierarchy also tracks the legal showing required, with the caveat that the details depend on the statute and the technique employed and are not set out here. Non-content metadata is generally obtainable on a lower showing than the content of communications, which ordinarily requires a warrant or a court order issued on a heightened showing, and historical location records fall under the warrant requirement recognized in *Carpenter*. Evidence of conduct and intent obtained through a consenting informant or undercover agent

has been treated as outside the warrant requirement [9, 10], which is part of what makes infiltration both effective and consequential. The constructive point for oversight is that the enrichments that most nearly restore identifiability are not uniformly the most heavily regulated, and an inquiry that relies on the cheapest level to select subjects, and on the most informative level only after selection, has reversed the order in which the evidence is most informative and most constrained. The choice among enrichments is also a choice about what kind of investigation is being conducted. Moving from association to content, or to conduct evidence obtained through infiltration, moves the inquiry from a records-based one to one involving interception or undercover participation, with different authorization requirements and different institutional consequences, so that the comparison concerns the character of the investigation and not only the quantity of information gathered.

The comparison suggests a formal problem, stated here as a research direction and not as a completed theory. Fix an observation $\mathcal{O}$ and a family $\mathcal{E}$ of admissible enrichments. An enrichment $\varepsilon \in \mathcal{E}$ is *sufficient* if $(\mathcal{O}, \varepsilon)(\omega) = (\mathcal{O}, \varepsilon)(\omega')$ implies $\alpha(\omega) = \alpha(\omega')$, that is, if the combined observation separates every pair of situations with different participation assignments. The problem is

$$\min_{\varepsilon \in \mathcal{E} \text{ sufficient}} \operatorname{cost}(\varepsilon),$$

where cost measures the privacy cost, intrusiveness, or institutional burden of acquiring $\varepsilon$. The statement is incomplete in three respects. The admissible family, the cost function, and the population of situations over which sufficiency is required all need specification, and in practice sufficiency would be required only over the situations compatible with the facts of a given case. The cost is also unlikely to be additive across enrichments. Nor can an optimization criterion replace constitutional doctrine, which concerns what may be acquired and on what showing and not only what would be informative. The reformulation is nonetheless useful because it changes the question that oversight asks. The question is not simply how much surveillance is conducted but whether the information acquired actually resolves the uncertainty that is relevant to the alleged offense, and which of the available enrichments resolves it at the least cost.

## 10 Expansion as a feedback process

A further extension, which is not found in the textbook, treats the investigation as a dynamic graph process under idealized assumptions that are stated here so that the results can be applied with care. The underlying network $G^* = (P, E^*)$ is finite, undirected, and static over the period considered. Observation of a subject is complete in the sense that every edge of $G^*$ incident to a subject becomes recorded once that person is a subject, and observation of non-subjects is limited to what has been recorded through their ties to subjects. Let $E_t \subseteq E^*$ be the set of recorded edges at time $t$ and $I_t \subseteq P$ the set of subjects, with $I_0 = S$ a set of seeds, $I_\infty = \bigcup_t I_t$ the final set of subjects, and $G_t = (P, E_t)$ the recorded graph at

time $t$. Edges are written $uv$ for the unordered pair. The two coupled rules are

$$E_{t+1} = E_t \cup \{\, uv \in E^* : u \in I_t \text{ or } v \in I_t \,\},$$

which intensifies observation for subjects, and

$$I_{t+1} = I_t \cup \{\, v : \exists u \in I_t,\ uv \in E_{t+1},\ \mathrm{rel}(u, v) \geq \theta \,\},$$

which adds subjects by relevance, where rel is an investigative relevance score and $\theta$ a threshold. Relaxing the assumptions changes the results in predictable ways. If observation of edges incident to subjects is incomplete, growth is slower than the ball $B_t(S)$ of the lemma below, where in this section $B_t(S)$ abbreviates $B_t^{G^*}(S)$ and $C_{G^*}(s)$ denotes the component of $s$ in $G^*$, and if the underlying graph changes over time the lemma holds only for the graph in force at each stage. If ties are directed, as with one-way communications, reachability must be taken along arcs and the component statement becomes one about strongly or weakly connected sets according to which arcs create relevance.

The decisive issue is how rel is determined. Call the rule *proximity-sufficient* if $\mathrm{rel}(u, v) \geq \theta$ for every recorded association $uv$, so that being adjacent to a subject is enough to become one.

**Lemma 1.** *Let $G^*$ be finite, undirected, and static, and let observation be complete for edges incident to subjects. Under the proximity-sufficient rule, $I_t = B_t(S)$ computed in $G^*$, and $I_\infty = \bigcup_{s \in S} C_{G^*}(s)$.*

*Proof.* The recorded graph and the underlying graph agree where it matters: by the observation rule, the subgraph of $G_t$ formed by edges incident to $I_t$ equals the subgraph of $G^*$ formed by the same edges, so neighbors of subjects are neighbors in $G^*$. Induct on $t$. For $t = 0$, $I_0 = S = B_0(S)$. Suppose $I_t = B_t(S)$ in $G^*$. Every edge of $G^*$ at a vertex of $I_t$ is in $E_{t+1}$ by the observation rule, and every such edge satisfies the threshold by proximity-sufficiency, so $I_{t+1}$ adds exactly the neighbors in $G^*$ of vertices of $B_t(S)$, which is $B_{t+1}(S)$. Since $P$ is finite the sequence stabilizes at the union of the components containing the seeds. □

The contrasting case is best presented as a formal benchmark and not as a model of the law. Let $\mathrm{Pred}(v)$ denote a *predicate*: documented, offense-relevant information about $v$ obtained independently of the surveillance of $v$ and meeting whatever threshold the governing legal standard for the investigative step requires. Call the rule *predicate-conditioned* if $\mathrm{rel}(u, v) \geq \theta$ only when $\mathrm{Pred}(v)$ holds.

**Proposition 2.** *Under the predicate-conditioned rule, $I_\infty \subseteq S \cup \{v : \mathrm{Pred}(v)$ holds$\}$, whatever the topology of $G^*$.*

*Proof.* A vertex outside $S$ enters $I_{t+1}$ only if some adjacent subject $u$ has $\mathrm{rel}(u, v) \geq \theta$, which by definition requires $\mathrm{Pred}(v)$. Hence every vertex outside $S$ in $I_\infty$ satisfies Pred. □

This proposition follows from the definition of the rule, and it is stated for that reason as a benchmark, not as a discovery. Its content lies in what the benchmark leaves open. The predicate need not be evidence of completed criminal conduct. Investigators may lawfully open and pursue an inquiry on lesser information, including indications of planning or attempt, reliable tips, or specific and articulable facts, and the threshold varies with the investigative technique employed, so that a record request, a physical surveillance, and an interception order do not require the same showing. The benchmark is agnostic about the threshold. It requires only that whatever threshold the law sets be met by information that does not originate in the investigation's own prior surveillance of the person concerned, and that the information be about the person and not merely about their proximity to someone else. A realistic standard may sit well below proof of conduct and still differ sharply from the proximity-sufficient rule in the property that matters here: the extent of the investigation is bounded by the set of persons about whom a predicate exists, and the community's connectivity affects only how quickly investigators reach them.

Real investigations are neither pure case, and the model does not assert that any particular agency follows the first. What can be defined is the quantity on which the two rules differ, and the definition must separate two things that a cruder statement would conflate. Let the *unpredicated set* be

$$U = I_\infty \setminus \big(S \cup \{\, v : \mathrm{Pred}(v) \text{ held independently before } v \text{ was designated} \,\}\big),$$

the set of subjects who lacked a qualifying independent predicate when designated. Membership in $U$ does not show that proximity was the reason for designation. A subject might have been designated on a credible report that was later discredited, on a lead that did not meet the chosen definition of a predicate, or on some other consideration unrelated to proximity. Let the *amplification set* be

$$\mathrm{Amp} = \{\, v \in U : \text{the documented basis of designation of } v \text{ consists solely of ties to existing subjects} \,\},$$

so that $\mathrm{Amp} \subseteq U$. Under the predicate-conditioned rule $U$ and Amp are both empty. Under the proximity-sufficient rule, in the model, the only basis of designation is adjacency, so $\mathrm{Amp} = U = \big(\bigcup_{s \in S} C_{G^*}(s)\big) \setminus \big(S \cup \{v : \mathrm{Pred}(v)\}\big)$, which is a property of the community's connectivity and not of its conduct. The corresponding ratios are $|U|/|I_\infty \setminus S|$ and $|\mathrm{Amp}|/|I_\infty \setminus S|$, the second never exceeding the first. The first is an upper bound on how much of the investigation rests on proximity, and the second is the share for which proximity is the documented ground. Both are diagnostics and not verdicts, since a nonzero value can arise in a lawful investigation that is still developing its predicates, but a high value of the second ratio together with growth that tracks the ball $B_t(S)$ is the signature of the proximity-sufficient process. Computing Amp requires the documented basis of each designation and not only the timing of the evidence, which is why the record architecture of Section 11 must preserve the stated grounds.

The ratios are proposed here as metrics awaiting data and have not been applied to any investigation. A small illustration shows their behavior. Let $G^*$ be the path $s - p_1 - p_2 - p_3 - p_4$ with seed $S = \{s\}$, and suppose an independent predicate existed in advance only for $p_2$. Under the proximity-sufficient rule $I_\infty = \{s, p_1, p_2, p_3, p_4\}$, the non-seed subjects are $p_1, p_2, p_3, p_4$, and $U = \text{Amp} = \{p_1, p_3, p_4\}$, so both ratios equal three fourths. Suppose instead that the record shows $p_1$ was designated on a credible tip later discredited. Then $p_1 \in U$ but $p_1 \notin \text{Amp}$, so the first ratio remains three fourths while the second falls to two fourths. Under the predicate-conditioned rule $p_1$ is adjacent to a subject but lacks a predicate, so it is not admitted, and the investigation stops at $\{s\}$ until a predicate for some neighbor appears. The example illustrates the definitions and carries no empirical weight.

A distinction is needed for evidence acquired after designation, and it has a causal part as well as a temporal one. Evidence may be *discovered*, when its content existed before designation and investigators merely found it, as with a preexisting document located under a warrant obtained after designation. It may record a *subsequent* act, when the person independently performed an act after designation that investigators observed. Or it may be *generated*, when investigative activity produced the event or the record, as with a record of contacts created by monitoring the subject or an act induced by an undercover agent. None of the three was held at designation, so all leave the subject in $U$, and the audit of the designation is unaffected. Only generated evidence is manufactured by the process in the sense of the feedback mechanism, while discovered evidence and evidence of a subsequent independent act may fully corroborate a suspicion that was formed on other grounds. Timestamps order the events but cannot separate subsequent from generated evidence, so the classification requires a causal provenance field in the record, as Section 11 specifies. Evidence discovered through surveillance that the designation made possible remains discovered, and the question whether the designation was justified is separate from the question whether the discovery was productive.

The feedback character appears when the predicate variable is not independent of the process. If $\text{Pred}_t(v)$ is allowed to include records generated by the intensified observation of $v$ after $v$ has become a subject, then the relevance condition can be satisfied by material that the observation rule created. Individuals become subjects because they are associated with existing subjects, and their subsequent surveillance generates additional records that appear to corroborate the suspicion that selected them. In schematic form, the subject set leads to increased observation, observation leads to additional recorded edges and records, additional records lead to an enlarged association structure, and that structure leads to an enlarged subject set. A lawful investigation does expand when new evidence supports new suspicions. The problematic case is the one in which the expansion mechanism itself manufactures the apparent justification for further expansion, so that the process confirms the premise that set it in motion. Whether that case obtains for a given investigation is an empirical question that an independent reviewer can answer only if the record preserves the information needed to compute Amp, and Section 11 concerns what that information

must be.

## 11 Provenance-preserving projections and an auditable record

The projection $\pi$ of Section 4 discards the type, source, time, and reliability of each association. The diagnostics of the preceding sections, namely the dependency numbers $\nu$ and $\tau$, the designation times $T_I$, $T_E^{\mathrm{ind}}$, and $T_E^{\mathrm{all}}$ defined in Section 13, and the unpredicated and amplification sets, all require information that $\pi$ throws away. The question that follows the diagnosis is what an investigative record must retain for such diagnostics to be computable by someone other than the investigator who produced it.

Call a projection *provenance-preserving* if it is accompanied by a fiber map $\lambda$ assigning to each edge $uv$ of $G_A$ the set $\lambda(uv) = \pi^{-1}(uv)$ of observed relationships that gave rise to it, each retaining its type typ and its provenance record $\sigma$. A bare projection $\pi$ is the special case in which $\lambda$ is forgotten. An investigative decision is a claim, either an edge of the alleged graph $G_C$ or a designation of a person as a subject, and the record is *justification-complete* if every claim $c$ carries a basis $J(c)$, consisting of a finite set of vertices of the dependency system $G_D$, the stated grounds for the designation, the decision rule or standard applied at the time, and four timestamps for each element of the basis, together with a flag indicating whether the element was obtained before or after the person concerned became a subject. The four times are those at which the underlying event occurred, a record of it was created, investigators acquired the record, and it was first used to justify a decision. The timestamps order the events but do not establish causation. Evidence whose event and record precede designation but whose acquisition follows it was discovered, but evidence whose event and record both follow designation may have been generated by investigative activity or may record an act the person performed independently, and the timestamps cannot tell these apart. Each element of a basis therefore also carries an acquisition mechanism and a causal provenance field recording whether investigative activity produced the event or merely recorded it, with unknown as a permitted value. Acquisition time is the one used to audit a designation, and the causal provenance field is the one used to classify evidence as generated A record that is provenance-preserving and justification-complete supports the following computations: the type composition of each association, the numbers $\nu$ and $\tau$ on $G_D$ at both granularities of Section 8, the times $T_I(v)$, $T_E^{\mathrm{ind}}(v)$, and $T_E^{\mathrm{all}}(v)$ of Section 13, and the sets $U$ and Amp.

**Proposition 3.** *Let $L$ be a complete investigative record and let $\psi(L)$ be the part of it that is retained for review. A statistic $f(L)$ is computable from $\psi(L)$ only if $\psi(L) = \psi(L')$ implies $f(L) = f(L')$. If $\psi$ is the bare projection $\pi$, then none of $\nu$, $T_E^{\mathrm{ind}}$, and membership in $U$ or* Amp *is computable from $\psi(L)$ in general. Moreover, the classification of post-designation evidence as generated by investigative activity is not computable from a record that retains the projection and all four timestamps without a causal provenance field.*

*Proof.* The first statement is the definition of a function of $\psi(L)$. For the second, consider first the dependency numbers. Take two records that produce the same simple graph and support the same conclusion, namely that two persons $u$ and $v$ are associated, an edge of $G_A$. In $L$ the conclusion is supported by three independently obtained records, each sufficient, so that the dependency system has three pairwise disjoint proofs and $\nu = 3$ and $\tau = 3$. In $L'$ the same three records are all derived from a single entry of one data broker, so that every proof passes through that entry and $\nu = \tau = 1$. The projection retains the edge $uv$ and nothing of the supporting records, so it is the same for both. Hence $\nu$ takes different values on records with equal projections.

For the designation statistics, a different construction is needed, because a timestamp on an identity match does not by itself affect when offense-relevant evidence first existed. Let $v$ be a person designated as a subject at time $t_0$ on the documented basis of ties to existing subjects, and let $r$ be a record documenting an act of $v$ that is relevant to an element of the alleged offense, not generated by investigative activity, acquired at time $t_1$ and inducing no association edge, so that the projected association graph is unchanged by it. Offense-relevant evidence need not create an association edge, as with a financial record or a photograph of a solitary act. Let $L$ and $L'$ be identical except that $t_1 < t_0$ in $L$ and $t_1 > t_0$ in $L'$. The projection retains neither times nor bases and is the same for both. In $L$ the record $r$ is an independent predicate that held before designation, so $v \notin U$. In $L'$ no predicate preceded designation, so $v \in U$, and since the documented basis is proximity alone, $v \in \mathrm{Amp}$, while $T_E^{\mathrm{ind}}(v) = t_1 > t_0 = T_I(v)$. Hence $T_E^{\mathrm{ind}}$ and membership in $U$ and Amp differ on records with equal projections.

The argument for $T_E^{\mathrm{ind}}$ and for membership in $U$ and Amp uses acquisition time alone, so it is a lower bound on what the record must retain. The other fields are required for the final claim, in two steps. First, let $L$ and $L'$ agree in projection, in the designation time $t_0$, and in the acquisition time $t_1 > t_0$ of a record $r$ concerning $v$, but let the event documented by $r$ occur before $t_0$ in $L$ and after $t_0$ in $L'$. In $L$ the evidence was discovered, and in $L'$ it records an event that followed designation, so the event time is needed to separate the cases. Second, let $L''$ and $L'''$ agree in all four timestamps, with the event following designation, but let the event be an independent act of $v$ in $L''$ and an act induced by an undercover agent in $L'''$. All timestamps and the projection agree, yet the evidence is generated in one record and not in the other, so a causal provenance field is needed. □

The proposition is elementary, and the point it makes is architectural and not subtle. Whether a reviewer can detect amplification depends on what was kept, and the information a bare projection discards is exactly the information that distinguishes a well-supported investigation from a self-reinforcing one. Retaining the source and timestamp of every edge is also not enough. The statistics above depend on the original grounds for each designation and on the decision rule applied at the time, and a record that preserves only the edges and their times cannot reconstruct either. A defensible investigative data architecture would therefore satisfy six requirements that follow from the preceding definitions. Fibers of any

projection are retained. Each relationship carries its type and source. Each designation carries its stated grounds, the decision rule applied, and for each element of its basis the four timestamps, the acquisition mechanism, and the causal provenance field. The dependency relation among records is recorded, including which dependencies are conjunctive and which disjunctive, so that $\nu$ and $\tau$ can be computed and common origins identified. Identity edges are typed by the relation they assert, registration, operation, authorization, or communication, so that the composition of Section 3 cannot occur silently. Finally, the record is retained in a form that cannot be revised after the fact without a trace, so that the timestamps on which the diagnostics depend are those of the time. None of these requirements restricts what investigators may collect. They concern what is preserved of the reasoning, and they make it possible for a court, an inspector, or a defense expert to apply the diagnostics of this essay without relying on the account of the party under examination.

## 12 Association, conspiracy, and aggregation in doctrine

The preceding analysis connects to doctrine in several places, and it is worth stating the connections carefully, because the mathematical argument is not a substitute for legal analysis and the doctrines have their own contours.

The first connection concerns the constitutional status of association itself. In *NAACP v. Alabama ex rel. Patterson*, the Supreme Court recognized that compelled disclosure of an organization's membership could burden the freedom of association, because the exposure of affiliations can deter participation [6]. The point is relevant to graph construction because a membership or contact graph is exactly a structure that records affiliation. The harm the Court identified did not depend on a criminal charge. It arose from the exposure of a network. A more recent documented instance is the litigation over the New York Police Department's mapping and surveillance of Muslim communities, *Raza v. City of New York* [11], which was resolved by a 2017 settlement that bars investigations in which religion or ethnicity is a substantial factor and requires articulable factual information before investigating protected activity. The settlement is not precedent and is mentioned only as a record of the reforms the parties agreed to. A police official had reportedly acknowledged in related proceedings that the program had not produced a terrorism lead. The example is offered as a case of association-based collection whose evidentiary yield was disputed in court, and not as evidence about the investigation discussed in this essay.

The second concerns the relationship between membership and criminal liability. The principle that association alone does not establish culpability is also stated outside the Smith Act setting. In *NAACP v. Claiborne Hardware Co.*, the Court rejected liability founded on association with a group some of whose members committed unlawful acts and required proof that the individual held a specific intent to further the unlawful aims [8], and it is a principle stated throughout the federal case law of conspiracy that mere association or presence, without more, does not establish knowing participation in an agreement. In *Scales*

*v. United States*, the Court upheld a conviction under the membership clause of the Smith Act on a construction under which the clause reaches only active, knowing membership in an organization advocating forcible overthrow of the government, held with a specific intent to bring about that overthrow, and not mere passive or nominal membership [7]. The doctrinal structure there is the same as the structure of the non-identifiability proposition: the fact of belonging does not establish the mental element, and the mental element must be proved by something that distinguishes the culpable member from the member who shares only the connection.

The third concerns the law of conspiracy itself. The offense requires an agreement and, ordinarily, knowing participation in it, and a doctrine of co-conspirator liability for substantive offenses committed in furtherance of the agreement, associated with *Pinkerton v. United States*, makes the question of who belongs to the agreement unusually consequential [5]. Where liability for the acts of others depends on membership in an agreement, an inference to membership from graph proximity carries high stakes, and the discipline of *Kotteakos* on distinguishing one agreement from many is correspondingly important.

The fourth concerns aggregation. The Court has recognized in related contexts that the compilation of records can reveal more than the records separately. In *Carpenter v. United States*, the Court held that the government's acquisition of an extended period of historical cell-site location records is a search requiring a warrant, reasoning in part from the comprehensive picture of a person's movements that accumulated location data provides, and concurring opinions in *United States v. Jones*, which turned on the attachment of a GPS device to a vehicle, had earlier articulated concern about long-term monitoring producing what is sometimes called a mosaic [13, 12]. These decisions concern particular categories of data and should not be stretched to hold more than they do. They are consistent with, though they do not establish, the structural observation that the legal significance of an investigative act can differ from the sum of the significances of the records it combines. The practical implication is that the object of oversight should be the combined process, including what population was selected, which relationship types were used, how the population expanded, what information was aggregated, and which inferences were drawn, in addition to the individual request.

The essay does not dispute that the government may investigate crimes committed in the course of political activity. It may, subject to applicable law. The question is whether political association itself becomes a substitute for evidence of criminal conduct.

## 13 An empirical protocol for testing manufactured conspiracy

The framework yields a research design that could be applied to the investigation described in the reporting, using discovery materials where they are available, and it converts the essay from a conceptual argument into a potentially falsifiable account.

Investigative records would be organized into a temporal, provenance-aware graph in which every recorded association retains its source, its acquisition date, its relationship type, and the investigative relevance stated for it. For each subject $v$ define $T_I(v)$ as the time of initial investigative designation, and let $t_{\mathrm{acq}}(r)$ be the acquisition time of an item $r$ of offense-relevant evidence about $v$. Two times are defined, with the convention that the infimum of an empty set is $+\infty$:

$$T_E^{\mathrm{ind}}(v) = \inf\{\, t_{\mathrm{acq}}(r) : r \text{ is offense-relevant evidence about } v \text{ not generated by investigative activity} \,\},$$

$$T_E^{\mathrm{all}}(v) = \inf\{\, t_{\mathrm{acq}}(r) : r \text{ is any acquired offense-relevant evidence about } v \,\}.$$

Classification as generated uses the causal provenance field of Section 11, and evidence that is discovered, or that records an independent later act, counts as not generated. Because the second set contains the first, $T_E^{\mathrm{all}}(v) \le T_E^{\mathrm{ind}}(v)$, and both are defined even when the independent evidence does not exist, since the empty infimum is infinite. The difference between them is a diagnostic of investigative dependence: a finite $T_E^{\mathrm{all}}(v)$ with $T_E^{\mathrm{ind}}(v) = +\infty$ means that all the offense-relevant evidence acquired about $v$ was generated by investigative activity, and a long interval between the two means that independent evidence arrived late relative to evidence the investigation itself produced. The diagnostic is only as reliable as the provenance classification. Evidence that was merely discovered after designation, or that records a later independent act, can satisfy $T_I(v) < T_E^{\mathrm{ind}}(v)$ without having been manufactured by the process, which is why the classification and not the ordering carries the weight. The ordering $T_I(v) < T_E^{\mathrm{ind}}(v)$ does not by itself show misconduct, since investigators may lawfully develop suspicion before obtaining decisive evidence. A systematic pattern in which designation precedes substantive evidence, and in which the stated ground for designation is chiefly network proximity to existing subjects, would nonetheless warrant closer examination of investigative predicates.

Four further measurements are suggested. First, whether apparent corroboration survives the removal of common data sources, which is the computation of $\nu$ and $\tau$ of Section 8 on the actual dependency system at both granularities. Second, whether the categories of association were preserved through the analysis or collapsed by the projection $\pi$ of Section 4, which can be assessed by comparing the heterogeneous records with the relationship summaries presented to decision makers. Third, whether exculpatory information, such as the identification of an alternative operator of an account, changed the investigative graph or left it intact. Fourth, the observed pattern of expansion could be compared against a counterfactual model in which subject selection requires independently documented conduct as in the predicate-conditioned rule of Section 10, and the unpredicated set, the amplification set, and their ratios could be computed directly from a justification-complete record of the kind described in Section 11. A close match between actual growth and the proximity-sufficient model, a high amplification ratio, and a poor match to the predicate-conditioned model would together count in favor of the feedback hypothesis. The opposite finding would

count against it, and mixed findings would be inconclusive and should be reported as such. A high value of $|U|$ with a low value of $|\mathrm{Amp}|$, for example, would indicate designation without a documented predicate but not designation on documented proximity, and it would call for examination of the stated grounds and not support the feedback hypothesis. The ratios are metrics proposed in this essay and have not been applied to any investigation, so that nothing in the present argument depends on their values. The dependency measurements address a separate question from the expansion comparison, and any correlation between the two is a further hypothesis and not a consequence of either model. The hypothesis could fail, and that is what distinguishes it from a rhetorical characterization. The test also depends on what is available. Where the investigative record preserves only projected associations, the protocol cannot be completed, and that limitation is itself a finding about the auditability of the record.

## 14 Conclusion: the topology of suspicion

A graph is a statement about what can be connected, and graph theory is exact about what follows from such a statement. What follows is structure: which persons are reachable from which, which are indispensable intermediaries, how many routes lead to a conclusion and how many of them are truly separate. What does not follow is any statement about what particular people knew or agreed to do. The reported surveillance apparatus is capable of reconstructing extensive networks from digital traces and institutional records, and the preceding sections have tried to show, one diagnostic at a time, that connectivity is not coordination, density is not agreement, centrality is not authority, and a multiplicity of paths is not a multiplicity of independent sources. A network can accurately represent observed association and remain insufficient to identify criminal agreement, and an investigation whose subject selection depends on network proximity can amplify that insufficiency into a growing population of subjects, a process whose extent a reviewer can measure only if the record preserves the provenance of each designation.

The hypothesis named in the title is therefore a claim about process, and it can be tested where the record permits. Two dimensions are tested separately. Selection amplification, measured by the unpredicated set, the amplification set, and the growth of the subject set, is predicted to be large under a proximity-sufficient rule and empty under a predicate-conditioned one. Evidentiary dependency, measured by $\nu$, $\tau$, and the dependency system, tests independently whether apparent corroboration rests on a few common sources. Neither model implies the other, since an investigation can expand by proximity and still acquire genuinely independent evidence, and a predicate-conditioned investigation can rest on several records derived from one unreliable source. Finding both would strengthen the account of manufactured conspiracy, and finding only one would identify a more limited mechanism. Neither dimension can be checked from a bare association graph, and an investigative record that cannot support the check has already failed the condition of auditability

that the argument requires. A surveillance graph describes what an institution can connect. A criminal conspiracy requires proof of what particular people knowingly agreed to do. Additional evidence can close the distance between those two propositions, and communications establishing an agreement or records of conduct may do so, but only evidence that separates the lawful and the criminal possibilities can. No accumulation of observations that remain invariant between lawful association and criminal agreement can identify the distinction between them, since more instances of their shared structural features add edges and leave the distinguishing element missing.

## References


[1] M. Lavrov, *Start Doing Graph Theory*, 2026.

[2] K. Menger, Zur allgemeinen Kurventheorie, *Fundamenta Mathematicae* 10 (1927), 96–115.

[3] U.S. Senate, Select Committee to Study Governmental Operations with Respect to Intelligence Activities, *Final Report*, 94th Congress, 1976.

[4] *Kotteakos v. United States*, 328 U.S. 750 (1946).

[5] *Pinkerton v. United States*, 328 U.S. 640 (1946).

[6] *NAACP v. Alabama ex rel. Patterson*, 357 U.S. 449 (1958).

[7] *Scales v. United States*, 367 U.S. 203 (1961).

[8] *NAACP v. Claiborne Hardware Co.*, 458 U.S. 886 (1982).

[9] *Hoffa v. United States*, 385 U.S. 293 (1966).

[10] *United States v. White*, 401 U.S. 745 (1971).

[11] *Raza v. City of New York*, 998 F. Supp. 2d 70 (E.D.N.Y. 2013); see also the settlement agreement and stipulation of 2017.

[12] *United States v. Jones*, 565 U.S. 400 (2012).

[13] *Carpenter v. United States*, 585 U.S. 296 (2018).

[14] A. Winston, ICE Is Running One of the Largest Anti-Left Surveillance Operations in US History, *WIRED*, October 28, 2026.