Cohort Signatures in Human Spatiotemporal Behavior

Universal laws, demographic parameterization, and the detectability of synthetic identities

Kenshiki Labs - Technical Working Paper (draft for internal review)
June 2026
v0.2.1

Reading time: 26 minutes (~6,500 words)

Note on scope: this paper synthesizes two decades of large-scale trajectory research (call detail records, GPS panels, travel surveys). All numerical parameters are approximate and dataset-dependent; they shift with spatial resolution (tower vs. GPS), sampling regime (event-driven vs. network-driven), and observation window. The functional forms do not.

Abstract

Synthetic identity detection is shifting from attribute matching to life-process coherence testing. Static attributes - names, SSNs, addresses - can be fabricated or assembled; a coherent human *life process* is far harder to counterfeit, because large-scale empirical mobility research converges on a striking result: human movement exhibits a small set of recurrent statistical regularities whose **functional forms are invariant across every population studied**, while cohort membership - age, occupation, household structure, geography - determines the **parameters** of those regularities. Real people occupy a thin manifold in trajectory-statistic space; demographic cohorts occupy identifiable neighborhoods on it. A fabricated identity must therefore counterfeit a process, not a datum, and it can fail at four levels: violating the empirical invariants outright (archetypal failure), satisfying them with a parameter vector incoherent with claimed demographics (cohort failure), failing to cohere across independent channels - address history, carrier metadata, mobility exhaust - that no fraudster jointly controls (cross-channel failure), or failing joint consistency across a minted population (ring failure). This paper summarizes the invariants with their canonical sources, develops the cohort parameterizations under an explicit three-tier evidence labeling, characterizes what the carrier layer actually observes, maps synthetic archetypes to discriminating statistics, and - because the underlying signals are legally sensitive - specifies a deployment boundary under which raw mobility never reaches underwriting and the system's only output is a bounded, signed, replayable evidence packet.

1. Introduction

A lived life is an entropy source that is expensive to simulate. Every real adult emits a continuous stream of spatiotemporal exhaust c tower registrations, handoffs, dwell intervals, address changes - whose joint distribution is shaped by physiology (circadian biology), economics (labor schedules, housing markets), physics (travel speeds, road topology), and social structure (households, institutions). The remarkable empirical finding of the mobility-science literature is that this exhaust is *lawful*: a handful of scaling relations and distributional families describe essentially everyone, everywhere these phenomena have been measured,

from Boston to Abidjan to Seoul. Stated plainly: synthetic identities do not have lives, and the absence of a life is statistically legible.

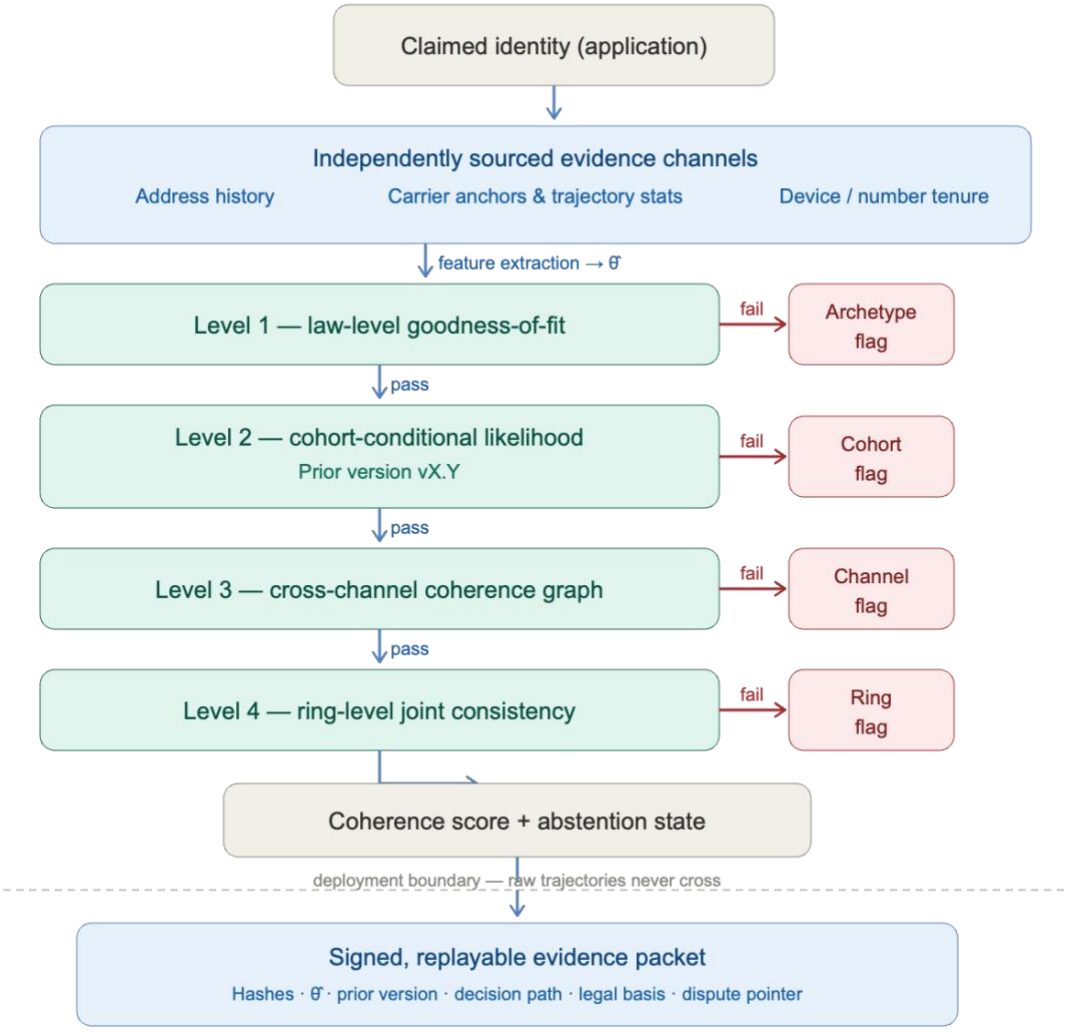
The equally important finding, less often stated explicitly, is that the laws are universal but the parameters are not. A 20-year-old student and a 70-year-old retiree both obey truncated power-law displacement statistics, both show Zipfian visitation, both are highly predictable. But the *scale* of their movement, the *phase* of their circadian activity, the *contrast* between their weekdays and weekends, their *exploration rate*, their *daily motif mixture*, and their *address-churn hazard* differ in characteristic, measurable ways. Formally: if a person's trajectory statistics are summarized by a parameter vector

θ = (spatial scale, circadian phase, weekday/weekend contrast, exploration rate, motif mixture, returner/explorer weight, address-churn hazard, seasonality amplitude, …)

then the universal laws fix the *family* of distributions, and cohort membership fixes a *prior* over θ. This is exactly the structure a detection system needs: the laws define what any plausible human trajectory must look like, and the cohort priors define what *this claimed* human's trajectory should look like.

Section 2 develops the universal layer. Section 3 - the heart of the paper - develops the cohort layer under an explicit three-tier evidence labeling. Section 4 describes the observational realities of carrier data. Section 5 maps synthetic-identity archetypes onto the invariants they violate and the statistics that discriminate them. Section 6 specifies the deployment boundary that keeps raw mobility out of underwriting, and Section 7 closes with the structural asymmetries that favor the defender. Appendix A specifies what a production system estimates, how, and when it must abstain.

Figure 1 - Identity-coherence pipeline and nested test structure. Raw trajectories never cross the deployment boundary (Section 6); the only artifact that does is the signed evidence packet.



2. The Universal Layer

A note on terminology before the inventory. The literature itself speaks of laws - Zipf's law of visitation, the universal visitation law - and where a result carries that name we keep it. In this paper's own voice, however, the operative term is *empirical invariant*: a regularity replicated across datasets, countries, and modalities whose functional form is stable but whose fitted parameters are sensitive to spatial resolution, sampling regime, time binning, and urban form. Nothing in the detection architecture rests on a constant of nature; everything rests on replicated form plus in-domain parameter estimation. The same humility is enforced on the cohort layer in Section 3 through explicit evidence tiers.

2.1 Displacement statistics and spatial scale

The founding result of the field is due to González, Hidalgo, and Barabási (2008), who analyzed the trajectories of roughly 100,000 anonymized mobile-phone users over six months. Jump lengths between consecutive observed locations follow a truncated power law:

$$P(\Delta r) \propto (\Delta r + \Delta r_0)^{(-\beta)} \cdot \exp(-\Delta r/\kappa)$$

with $\beta \approx 1.75 \pm 0.15$, a small-scale cutoff $\Delta r_0 \approx 1.5$ km, and an exponential truncation κ on the order of 10^2 km that varies with dataset and geography. Human travel is *not* a pure Lévy flight: the apparent population-level heavy tail is partly a superposition of individuals each moving at their own characteristic scale.

That characteristic scale is the **radius of gyration**,

$$r_g = \mathrm{sqrt}\left((1/n) \cdot \Sigma_i \, \|r_i - r_{cm}\|^2 \right),$$

the RMS distance of a person's observed positions from their trajectory's center of mass. The population distribution of r_g is itself heavy-tailed - most people live within a ~5–10 km envelope while a minority range over hundreds of kilometers - and, critically, when each individual's displacements are rescaled by their own r_g, the per-person distributions collapse onto a single universal curve. **People differ in scale, not in form.** This single sentence is the foundation of cohort-conditioned detection: cohorts shift r_g and κ; nothing human escapes the functional family.

A second temporal fact deserves more weight than it usually receives: $r_g(t)$ grows only *logarithmically* with observation time [1]. This is stranger than it sounds, because at short timescales the same trajectory is heavy-tailed - locally, a person resembles a Lévy flight, and a true Lévy flight never converges; its territory grows without bound forever. What makes a trajectory a *life*, statistically, is that recurrence beats diffusion: above the daily timescale the return forces win, and the spatial envelope collapses from power-law growth to logarithmic saturation. Nor is the saturation stasis - under the conserved-capacity result of Section 2.2, territories saturate *while still circulating*, gains balancing losses under a fixed cap. And the asymptote is piecewise: a residential move or a job change resets the exploration clock, launching a fresh transient that converges to a new plateau, with the renewal hazard itself lawful (the Rogers–Castro schedule of Section 3.1). A real life is thus a punctuated sequence of convergence curves with lawful resets (Figure 2) - *tail* of the curve rejects synthetics, which present either unbounded diffusion (the teleporter) or boundedness without flux (the static beacon); the *onset* of the curve admits genuine thin-file applicants, whose fresh, on-schedule exploration transient is exactly what a real newcomer looks like.

Figure 2 - Piecewise asymptotic territory: punctuated logarithmic convergence with lawful renewals, against the unbounded and frozen synthetic archetypes.

2.2 Exploration, preferential return, and Zipfian visitation

The generative mechanism behind these statistics is the **exploration and preferential return (EPR)** process[2]. Having already visited S distinct locations, a person visits a *new* location with probability

$$P_{new} = \rho \cdot S^{(-\gamma)}, \quad \rho \approx 0.6, \ \gamma \approx 0.21,$$

[1] Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. Nature Physics, 6, 818–823.

[2] Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. Nature Physics, 6, 818–823.

and otherwise returns to a previously visited location with probability proportional to its past visitation frequency. Three measurable invariants follow.

First, the number of distinct visited locations grows sublinearly, $S(t) \propto t^{\mu}$ with $\mu \approx 0.6$: real people discover new places at a steadily decelerating rate. Second, visitation frequencies obey Zipf's law, $f_k \propto k^{(-\zeta)}$ with $\zeta \approx 1.2$ when locations are ranked by frequency; for typical working adults the two top-ranked cells (home and workplace) account for a majority - commonly on the order of 60–70% - of hourly observations. Third, the rich-get-richer return dynamic concentrates time so strongly that everything downstream (entropy, predictability, motif structure) inherits its shape.

A complementary invariant from longitudinal GPS panels: the set of *actively familiar* locations is a **conserved quantity**. Alessandretti et al.[3] showed that people maintain roughly 25 familiar places at any given time, over years - new locations are adopted at the same rate old ones are dropped. Real exploration is *turnover under a capacity constraint*, not accumulation.

2.3 Entropy and the predictability ceiling

Song, Qu, Blumm, and Barabási (2010, *Science*) computed three entropies per user over hourly, tower-resolution trajectories: the random entropy $S_{rand} = \log_2 N$ (N = number of distinct visited locations), the temporal-uncorrelated entropy $S_{unc} = -\Sigma\ p_i\ \log_2\ p_i$, and the *actual* entropy S, estimated via Lempel–Ziv methods, which accounts for the full order of visits. The actual entropy distribution peaks near **0.8 bits** - meaning a person's next hourly location carries less uncertainty than a fair coin flip; the effective number of plausible next locations is $2^{0.8} \approx 1.7$. Through Fano's inequality this implies a maximum predictability Π_{max} **≈ 93%**, and the distribution of Π_{max} across the population is remarkably narrow.

The crucial nuance for this paper: **predictability is the most cohort-invariant observable in the literature.** Song et al. explicitly tested for demographic dependence and found that Π_{max} barely moves with age, gender, or radius of gyration - people whose territories differ by two orders of magnitude are *equally regular within them*. Cohorts move scale and phase; they do not move the regularity ceiling. A synthetic trajectory whose actual entropy sits far above the human band (too random) or far below it (too scripted) fails regardless of which cohort it claims.

Two qualifications keep this result honest. First, the 93% figure is task- and scale-specific: it bounds prediction of the *next hourly bin at tower resolution, self-transitions included* - and since people are usually where they just were, stationarity does much of the work. When the task is reframed as next-*place* prediction with self-transitions excluded, estimated ceilings drop substantially, to roughly 0.65–0.75 [4]. Second, and decisive for our purposes: the detection value lies not in any particular ceiling but in the existence of a *bounded regularity band* that real humans occupy under a fixed task definition. A trajectory outside the band fails under every one of these task formulations simultaneously. Production systems should therefore fix the

[3] Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., & Baronchelli, A. (2018). Evidence for a conserved quantity in human mobility. Nature Human Behaviour, 2, 485–491.
[4] Ikanovic, E. L., & Mollgaard, A. (2017). An alternative approach to the limits of predictability in human mobility. EPJ Data Science, 6, 10. Cuttone, A., Lehmann, S., & González, M. C. (2018). Understanding predictability and exploration in human mobility. EPJ Data Science, 7, 2.

task (binning, resolution, self-transition handling), estimate the band in-domain, and test against the band - never against a literature constant.

2.4 The returners/explorers dichotomy

Pappalardo et al.[5] defined the k-radius of gyration $r_g^{(k)}$, computed over only a person's k most-visited locations, and the ratio $s_k = r_g^{(k)}/r_g$. The population distribution of s_2 is strongly **bimodal**: "2-returners," whose total spatial spread is essentially captured by their top two locations, and "explorers," whose spread is distributed over many places. Both classes exist in every population studied; the *mixture weight* between them is cohort-dependent (returner share rises with age and job tenure; explorer share peaks in the student and early-career years). This is a clean example of the two-layer structure: the bimodality is universal; the weights are demographic.

2.5 Daily motifs

Schneider et al.[6] represented each person-day as a small, directed network - nodes are distinct visited locations, edges are trips - and found that although the combinatorial space of such networks is enormous, just **17 motifs account for roughly 90% of all observed person-days**, with a typical day touching about two to five distinct places. The motif census is stable across continents and across data modalities (travel surveys vs. phone records). The motif *mixture*, however, is one of the most cohort-sensitive observables we have: the simple home–work–home dumbbell dominates mid-career weekdays; chained motifs with fixed-time intermediate stops mark caregiving households; star motifs with many short-dwell leaves mark gig and delivery work (Section 3).

2.6 Temporal texture: burstiness and circadian return

Human activity in time is neither periodic nor Poissonian. Inter-event times are heavy-tailed [7]: long silences punctuated by bursts. The standard summary is the burstiness coefficient

$$B = (\sigma_\tau - m_\tau) / (\sigma_\tau + m_\tau),$$

where m_τ and σ_τ are the mean and standard deviation of inter-event times. Strictly periodic processes give $B \to -1$, Poisson processes give $B = 0$, and human activity sits at characteristically *positive* B. Naive synthetic schedulers almost always land at the wrong B - cron-like generators near −1, uniform-random generators near 0.

On top of the burstiness sits hard periodic structure. The probability that a person *returns* to a location peaks sharply at **24, 48, and 72 hours**after first visiting it [8], with a strong weekly line at 168 hours. The power spectrum of a real person's location time series therefore shows circadian and weekly spectral lines over a

[5] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. Nature Communications, 6, 8166.
[6] Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. Journal of the Royal Society Interface, 10, 20130246.
[7] Barabási, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. Nature, 435, 207–211.
[8] González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. Nature, 453, 779–782.

broadband bursty background - and the *phase* of the circadian line (when the day-anchor and night-anchor transitions occur) is set by cohort and occupation, while the *existence* of the lines is universal.

2.7 Dwell, pause, and motion physics

At sub-trip resolution, GPS studies [9] decompose movement into flights and pauses, both of which follow truncated power laws - the "Lévy-walk nature" of human mobility. Conditioned on place type, dwell times fall into recognizable families: overnight home dwells of roughly 8–14 hours, weekday workplace dwells of roughly 6–9 hours, retail and errand dwells of minutes to an hour, with heavy aggregate tails.

Motion itself is governed by modal physics: walking near 1.4 m/s, cycling near 4–6 m/s, urban driving, highway driving, rail, and air each occupy distinct speed bands, and transitions between high-speed modes require plausible interchange dwells (you do not go from pedestrian speeds to 250 m/s without an airport in between). Finally, the **travel-time budget**is one of the oldest invariants in transport science [10]: across cities, eras, and income levels, mean daily travel time hovers near an hour to an hour and a quarter. Faster modes buy *distance*, not time. A synthetic trajectory that spends four hours a day in motion, every day, has violated a constraint that has held from medieval villages to modern megacities.

2.8 Place-level and uniqueness invariants

Two further results complete the universal layer. Schläpfer et al.[11] established the **universal visitation law**: the density of visitors to any urban location, as a function of home distance r and visit frequency f, scales as

$$\rho(r, f) \propto (r \cdot f)^{-2}$$

- an inverse-square law in the product of distance and frequency, validated across cities on four continents. It provides a *place-level* check: in a realistic synthetic population, every venue's catchment must obey it, not just every person's trajectory.

And de Montjoye et al.[12] showed that in a dataset of 1.5 million users at hourly, tower-level resolution, **four randomly chosen spatiotemporal points uniquely identify 95% of individuals**. Trajectories are fingerprints. The privacy implication is well known; the detection implication is less appreciated: a cloned or donor-derived trajectory *collides* with its source under exactly this uniqueness property, and collisions are computable.

[9] Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., & Chong, S. (2011). On the Lévy-walk nature of human mobility. IEEE/ACM Transactions on Networking, 19(3), 630–643.
[10] Marchetti, C. (1994). Anthropological invariants in travel behavior. Technological Forecasting and Social Change, 47, 75–88.
[11] Schläpfer, M., Dong, L., O'Keeffe, K., et al. (2021). The universal visitation law of human mobility. Nature, 593, 522–527.
[12] de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. Scientific Reports, 3, 1376.

3. The Cohort Layer

The axes along which cohorts move θ are: **scale** (r_g, Δr_0, κ), **phase** (circadian offset of anchor transitions), **contrast** (weekday/weekend and seasonal asymmetry), **exploration** (effective ρ, capacity turnover rate), **motif mixture**, **returner/explorer weight**, **address-churn hazard**, and **seasonality amplitude**. What follows characterizes the major cohorts in those terms. Throughout, the cohort descriptions are *priors*, not stereotypes-as-rules: real populations are mixtures, and the correct statistical object is always a cohort-conditioned distribution, never a point.

One discipline governs everything below - the same discipline the detection product itself must be held to: claims are labeled by evidentiary status, and abstention is preferred to overstatement. Every cohort claim carries one of three tiers. *Tier 1* - empirically established: replicated and directly citable. *Tier 2* - operational hypothesis: grounded in a binding constraint (a statute, an institutional calendar, a physical mechanism) or in adjacent evidence, but whose trajectory-level signature still requires in-domain validation. *Tier 3* - candidate signal: a product-derived design target, stated explicitly so it can be estimated and falsified in production, and not to be represented as settled science. Untiered sentences are exposition; tier tags govern the claims they follow.

3.1 Life-course structure and address dynamics

Residential relocation - the slowest spatiotemporal channel, and the one most directly visible in credit-header data - has its own deep regularity. Age-specific migration rates follow the **Rogers–Castro model migration schedule**[13]: elevated rates in early childhood (children move when their parents move), a trough through the school years, a sharp **labor-force peak between roughly 22 and 30**, a long monotone decline through mid-life, and, in many populations, a secondary bump around retirement ages (62–67) with occasional late-life upturns associated with care transitions. This curve is one of demography's most replicated empirical regularities (Tier 1).

In the United States, annual mover rates have fallen for decades and now sit **under 10% of the population per year - roughly half the mid-twentieth-century rate**[14]. Most moves are short: the move-distance distribution is heavy-tailed but dominated by intra-county relocations, with long-distance moves disproportionately tied to education and employment events. Real address histories are also **calendar-textured**: moves cluster in summer, at lease boundaries (month-end/month-start), and - for student-age movers - in August and September (Tier 2: the lease-cycle and school-calendar mechanisms are binding, but the trajectory-level magnitudes want in-domain estimation).

The composite signature of a real address history is therefore *event coherence plus calendar texture*: discrete jumps whose timing co-varies with life events that leave traces in other channels (graduation, job change, marriage, first child, divorce, retirement) and whose calendar placement is non-uniform in characteristic ways (Tier 2). Synthetic address histories, by contrast, are typically generated uniformly over the calendar, at hazard rates flat in age, with no event correlates anywhere else in the record. The Rogers–Castro curve alone is a usable prior: a claimed 58-year-old with four address changes in three years is not impossible, but the likelihood penalty is real and quantifiable.

[13] Rogers, A., & Castro, L. J. (1981). Model Migration Schedules. IIASA Research Report RR-81-30.
[14] U.S. Census Bureau. Current Population Survey: Annual Geographic Mobility tables (series). Evidence level: Tier 1.

3.2 Students (≈17–24)

The student trajectory is anchored on campus and carries a **second seasonal frequency** on top of the universal weekly line: the academic calendar. Semester boundaries produce coherent territory shifts; summers commonly revert to a parental anchor, producing a *dual-anchor annual cycle* years before the retiree version of the same pattern appears (Tier 2: calendar-constraint-derived). Chronotype research and activity data agree that this cohort is **phase-delayed**: activity onset and offset shift two to three hours later than the working-adult baseline, with the delay amplified on weekends [15]. Exploration is at its lifetime maximum - effective ρ elevated, $S(t)$ growing near the top of the human band, explorer share at its peak in the returners/explorers mixture (Tier 2). Address churn is annual and sharply calendar-locked (August/September), and the modal mix skews pedestrian and transit on urban campuses, compressing displacement scale while keeping location diversity high.

3.3 Early career (≈22–35)

This cohort sits at the apex of the Rogers–Castro migration peak: long-distance moves tied to jobs are most probable here, and the address history legitimately shows its largest jumps. Within a metro, the defining transition is the **onset of commute bimodality**: tower-transition histograms develop the canonical 07:00–09:00 and 17:00–19:00 peaks as the workplace anchor stabilizes [16]. Weekend exploration remains high; the returner share begins rising with job tenure (Tier 2). Spatial scale r_g is set largely by metro structure and housing-cost geography rather than by preference.

3.4 Family formation (≈28–45)

Households with young children produce the most distinctive motif signature in the population: **trip-chaining**. The simple home–work dumbbell is replaced by chained motifs of four to five nodes with intermediate stops (school, daycare) pinned at fixed clock times, morning and afternoon, with low timing variance - regularity *increases* even as daily location count rises [17]. Late-evening entropy collapses; the weekend POI mixture rotates toward parks, athletic venues, and large-format retail. A canonical, well-documented life-event move sits in this window: the urban-core-to-suburban-ring relocation around the first child - a single coherent jump in the address channel that co-occurs with a motif change in the trajectory channel.[18]

This cohort also introduces a higher-order observable: **household coherence**. Two-adult households show correlated trajectories - a shared nighttime anchor with divergent daytime anchors, correlated weekend excursions, school-run alternation [19]. Real populations are organized into households; fabricated populations

---

[15] Roenneberg, T., et al. (2004). A marker for the end of adolescence. Current Biology, 14(24), R1038–R1039. Evidence level: Tier 1.

[16] Isaacman, S., et al. (2011). Identifying important places in people's lives from cellular network data. Pervasive Computing (LNCS 6696), 133–151. The bimodal commute signature is replicated across CDR studies. Evidence level: Tier 1.

[17] McGuckin, N., & Murakami, E. (1999). Examining trip-chaining behavior: Comparison of travel by men and women. Transportation Research Record, 1693, 79–85. Evidence level: Tier 1.

[18] Rossi, P. H. (1955). Why Families Move. Free Press. Evidence levels: Tier 1 for the residential move pattern; Tier 2 for its trajectory-channel signature.

[19] Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility. ACM KDD 2011. Toole, J. L., et al. (2015). Coupling human mobility and social ties. Journal of the Royal Society Interface, 12, 20141128.

almost never are. Household-graph structure is among the hardest things for a fraudster to counterfeit at scale, because it requires *jointly* consistent trajectories across identities (Tier 3: a design claim to be validated in production, not a finding).

3.5 Mid-career (≈45–60)

This is the routine maximum of the life course - not because the predictability *ceiling* rises (Section 2.3: Π_{max} barely moves with demographics) but because the *composition* of behavior shifts: the lowest new-location rates of adulthood, multi-year stability of r_g, the dumbbell motif at its modal peak, and the returner share at its lifetime high (Tier 2, consistent with the exploration decay of Section 2.2). Vacation seasonality becomes the dominant deviation from routine - annually recurring, corridor-consistent excursions of one to three weeks. A claimed member of this cohort presenting student-grade exploration statistics is exhibiting a parameter vector roughly two decades out of place.

3.6 Retirees (65+)

Retirement removes the commute and with it the population's strongest temporal landmark. The activity peak migrates to **mid-morning and early afternoon**; the weekday/weekend contrast - among the strongest features of working-age trajectories - **collapses toward symmetry**. Spatial scale contracts (smaller r_g, fewer distinct daily locations) while regularity stays high; recurring medical and service appointments introduce their own weekly and monthly periodicities (Tier 2; consistent with national time-use data on non-workers' day-of-week flattening). A well-defined subpopulation, the seasonal migrants ("snowbirds"), exhibits a clean **dual-anchor pattern with semiannual transitions**along plausible corridors to Sun Belt anchors [20]. The pattern is strongly stereotyped *and real* - which makes it a useful test case: fabricators who attempt it tend to render it either too clean (metronomic transition dates, zero year-to-year jitter) or incoherent with the claimed age and the address channel (Tier 3).

3.7 Occupational cross-cuts

Occupation cuts across age and often dominates it as a determinant of temporal structure.

Shift workers- roughly one in six U.S. workers on non-daytime schedules [21] - are the single most important trap in naive anomaly detection. Their circadian signature is **phase-shifted but fully regular**: nocturnal tower activity that recurs with high weekly fidelity. A detector scoring "nighttime activity" as anomalous will systematically flag a sixth of the legitimate workforce; the correct null is a *mixture over schedule phases*, under which the night-shift nurse is exactly as regular as the accountant, merely rotated.

Gig, delivery, and rideshare workers generate star and loop motifs with many short-dwell leaves, daily distances far above their residential cohort's median, and high tower diversity - yet retain a stable home anchor and a temporal envelope correlated with platform demand (meal peaks, weekend nights) (Tier 2 for the demand-correlated envelope; Tier 3 for the specific motif census). **Field-service trades** show depot-anchored, multi-site metro days at van speeds. **Long-haul truckers** are corridor-constrained with enormous r_g but high regularity: hours-of-service rules (the 11-hour driving window; 49 C.F.R. Part 395) *quantize*

[20] Smith, S. K., & House, M. (2006). Snowbirds, sunbirds, and stayers: Seasonal migration of elderly adults in Florida. Journals of Gerontology: Series B, 61(5), S232–S239. Evidence level: Tier 1.
[21] Bureau of Labor Statistics (2019). Job Flexibilities and Work Schedules — 2017–2018 Data from the American Time Use Survey. U.S. Department of Labor. Evidence level: Tier 1.

their dwell structure at a known network of stops - a regulatory constraint creating a verifiable statistical signature (the constraint is statute; the signature is Tier 2).

Remote and hybrid workers represent the largest structural break in the modern record. Post-2020, a substantial fraction of the white-collar workforce shows **attenuated or two-to-three-day commute bimodality**, midday neighborhood micro-trips, and home-anchor dwell fractions previously characteristic of retirees [22]. Any baseline trained on pre-2020 carrier data will misclassify this population; cohort priors are not static and must be versioned.

3.8 Geographic cross-cuts

Urban, suburban, and rural residents differ in both behavior and *measurement*. Behaviorally: urban trajectories show short displacements, high venue density, transit-speed segments, and large S; rural trajectories show longer displacements (larger effective Δr_0 and κ), vehicle dominance, and fewer distinct venues; suburban trajectories are car-modal with home-plus-retail-corridor anchoring (Tier 1 for the direction of these contrasts; Tier 2 for magnitudes, which are resolution-entangled). Observationally: tower density sets spatial resolution. A rural serving cell can cover orders of magnitude more area than an urban microcell, mechanically inflating apparent r_g and displacement lengths. **Cohort nulls must therefore be conditioned on serving-cell density**, or the model will confound geography of measurement with geography of behavior (Tier 1 - an engineering fact of cell geometry).

3.9 Socioeconomic and gender structure - and a fairness boundary

The literature documents real, replicated structure here, with important caveats. Visitation diversity and the income composition of visited places correlate with socioeconomic position - Moro et al.[23] showed that mobility patterns are tightly associated with *experienced income segregation* in large U.S. cities. But the relationship between SES and raw activity-space size is **city-dependent**: Xu et al.[24] found qualitatively different SES–mobility relationships in Singapore versus Boston, cautioning against any universal "wealth means range" rule. Several CDR studies [25] report gender gaps - smaller r_g, fewer unique locations, more trip-chaining among women in some cities - patterns driven by labor-market and caregiving structure, and varying with both (Tier 1 as cited, with the city-dependence caveat integral to the claim).

These regularities sharpen a boundary that a regulated deployment must draw explicitly. Cohort conditioning in this framework exists to **calibrate the null distribution of identity-coherence tests** - to ask whether a trajectory is consistent with the demographics the applicant *claims* - and must be architecturally firewalled from underwriting. Using cohort-correlated mobility features to alter credit terms walks directly into ECOA/Regulation B proxy territory; using them to ask "is this claimed person real?" does not, but the partition must be enforced and *documented in the evidence record*, not merely intended. The fairness argument cuts the other way too: a detector that ignores cohort structure imposes a disparate false-positive

[22] Barrero, J. M., Bloom, N., & Davis, S. J. (2021). Why working from home will stick. NBER Working Paper No. 28731. Evidence level: Tier 1 for the structural break; Tier 2 for specific trajectory-level signatures.
[23] Moro, E., Calacci, D., Dong, X., & Pentland, A. (2021). Mobility patterns are associated with experienced income segregation in large US cities. Nature Communications, 12, 4633.
[24] Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2018). Human mobility and socioeconomic status: Analysis of Singapore and Boston. Computers, Environment and Urban Systems, 72, 51–67.
[25] Gauvin, L., et al. (2020). Gender gaps in urban mobility. Humanities and Social Sciences Communications, 7, 11.

burden on legitimate outliers - shift workers, truckers, snowbirds, low-footprint elders - which is its own compliance failure mode. Cohort conditioning, done correctly, is not a fairness risk; it is the fairness control.

4. What the Carrier Layer Actually Observes

The invariants above are properties of behavior; carrier data observes behavior through a lens with its own statistics, and a detector that ignores the lens will mistake instrument for subject.

Sampling regime. Classical CDRs are *event-driven*: a location is recorded only when the device transacts (call, SMS, data session). The observation process is therefore itself bursty and entangled with behavior - heavy-tailed gaps in the record are partly communication behavior, not movement. Network-driven signaling data (periodic location/tracking-area updates, handovers) samples far more densely and more uniformly. Any fitted parameter - entropy especially - is conditional on the sampling regime, and cohort baselines must be estimated under the same regime they will be applied to.

Anchor inference. The standard constructions are robust: the home anchor is the modal serving cell during roughly 22:00–06:00; the work anchor is the modal weekday cell during roughly 09:00–17:00 excluding home [26]. The commute is then an *ordered handoff sequence* between anchors, and a real commute's sequence is consistent with road and rail topology and with modal speed bands - cells are traversed in a physically realizable order at realizable rates.

Authentic noise. Real carrier data contains characteristic artifacts, the most important being **ping-pong oscillation**: rapid alternation between adjacent sectors at a cell boundary driven by load balancing and signal fade, not motion. Its presence is a signature of authenticity. Naive synthetic traces are *too clean* - no oscillation, no missing intervals, no resolution heterogeneity - while crude random generators produce the opposite failure: handoff sequences between non-adjacent cells, which are physically impossible on a real network adjacency graph.

Device and account metadata. The trajectory rides on a device and a number, each with its own lawful history: handset replacement at plausible upgrade cadences (typically two to three years), porting events, roaming episodes coherent with travel in the trajectory channel, and number tenure. Real adults overwhelmingly carry numbers with years of history; synthetic identities skew toward young numbers and VoIP-originated blocks. Tenure alone is weak evidence; tenure *incoherent with the claimed life history* is strong evidence.

5. Synthetic Archetypes and the Invariants They Violate

Fabricated identities do not fail randomly; they fail in recognizable modes, because every generation shortcut leaves a statistical scar. Six archetypes cover the space observed and anticipated.

Table 1

Archetype	Generation shortcut	Violated invariants	Discriminating statistics

[26] Isaacman, S., et al. (2011). Identifying important places in people's lives from cellular network data. Pervasive Computing (LNCS 6696), 133–151.

Dormant / paper identity	No mobility exhaust at all; the identity exists only in bureau headers	Presence itself: a real claimed adult emits exhaust	Coverage likelihood vs. cohort base rates - treated as *evidence, not knockout* (legitimate low-footprint people exist)
Static beacon	A single fixed location pinged on schedule	$S(t)$ growth, motif census, travel-time budget	Distinct-location count, entropy below the human floor, motif degeneracy
Teleporter	i.i.d. location draws	Displacement law, return-probability peaks, motion physics	KS/AD tests on $P(\Delta r)$; absence of 24/168 h spectral lines; infeasible speeds; non-adjacent handoffs
Metronome	Scripted periodic schedule	Burstiness, heavy-tailed inter-event and pause times	$B \to -1$; quantized dwell times; actual entropy below the cohort floor; zero year-over-year jitter
Donor clone / replay	Copy of a real trajectory, possibly time-shifted or jittered	Uniqueness (4-point fingerprint); household and address coherence	Trajectory collision score against the corpus; cross-identity similarity (e.g., DTW); home anchor vs. claimed address mismatch
Stat-matched, cohort-incoherent	Learned generative model that reproduces the universal laws	None at the law level - fails at the cohort and cross-channel levels	Cohort-conditional likelihood ratio; cross-channel coherence graph

The first five archetypes fail level-one tests: goodness-of-fit against the universal functional forms (displacement and dwell distributions, spectral structure, burstiness, entropy band, motion feasibility on the network adjacency graph). These tests require no demographic information at all.

The sixth archetype is the frontier, because modern generative models *can* reproduce truncated power-law displacements, Zipfian visitation, and circadian lines. Against it, two further levels apply. **Level two** is the cohort-conditional likelihood ratio: given the claimed age, occupation, household status, and address geography, is the fitted parameter vector $\hat{\theta}$ plausible under the cohort prior? A student motif mixture on a claimed 68-year-old; rural-scale displacements on a claimed Manhattan address; a phase-regular nocturnal pattern alongside a claimed daytime employer; an address history with uniform calendar placement and a churn hazard flat in age - each is individually survivable and jointly damning, and the joint likelihood is a number, not a vibe. **Level three** is cross-channel coherence: the address timeline (bureau), the trajectory anchor timeline (carrier), number and device tenure (carrier metadata), and stated employment must tell *one* story - moves in the address channel should co-occur with anchor migrations in the trajectory channel, and both with plausible life events. No fraudster controls all of these channels jointly, which is precisely why the test has power.

One further level exists above the individual: the **ring level**. Synthetic identity fraud is industrial; identities are minted in batches. Generating one trajectory that passes levels one through three is expensive but conceivable. Generating *thousands* that are mutually consistent - no pairwise collisions under the uniqueness property, plausible household graphs, venue catchments that respect the $(r \cdot f)^{-2}$ visitation law, no shared

generation artifacts in dwell quantization or phase - multiplies the constraint set combinatorially. The defender's statistics compose across identities; the attacker's costs do too.

6. Deployment Boundary: No Raw Mobility in Underwriting

The science above answers whether spatiotemporal coherence *can* discriminate synthetic identities. A regulated reader's first question is different: *may* it be used, and under what containment? The signals involved are among the most legally sensitive a U.S. institution can touch. Device location held by carriers is customer proprietary network information under Section 222 of the Communications Act, and the FCC's 2024 forfeiture orders - approximately $200 million against the major carriers for unauthorized sharing of customer location data - made the enforcement posture explicit. The FTC's 2024 orders against location-data brokers (X-Mode/Outlogic; InMarket) established that sensitive location data cannot simply be bought and repurposed. State privacy statutes classify precise geolocation as sensitive personal information subject to heightened consent and use limits, and any signal that touches a consumer's eligibility for credit operates inside the FCRA's permissible-purpose regime and the ECOA/Regulation B fairness regime simultaneously.

The architecture therefore assumes the strictest reading and is designed so that the legal question never has to be relitigated per decision. Raw carrier trajectories are never exposed to lenders, never exported, and never retained beyond the computation window: feature extraction and all four test levels execute where the data already lawfully resides - carrier-side, bureau-side, or inside a contractual clean room - under a documented consent or authorization basis and a documented FCRA permissible purpose scoped to identity verification and fraud prevention. What crosses the boundary is a bounded assertion: a coherence score, an abstention state, and a signed, replayable evidence packet recording input hashes, the feature vector, the cohort-prior version, the test statistics and thresholds, the decision path, the legal basis, the retention clock, and a dispute pointer. The packet proves the computation without disclosing the trajectory.

Three containment rules complete the boundary. First, the output is scoped to identity verification and fraud prevention and is excluded by construction from pricing, terms, line assignment, and adverse-action reasoning about creditworthiness; a feature-exclusion ledger in the evidence record demonstrates that no mobility-derived or cohort-derived variable enters the underwriting model. Second, cohort priors exist to calibrate false-positive control - to keep the night-shift nurse, the long-haul trucker, and the snowbird from being penalized for being statistically unusual - and never function as decision factors about a person's merit. Third, retention, reinvestigation, and dispute workflows attach to the packet itself, so that a consumer challenge replays the exact computation rather than an analyst's recollection. None of this section is legal advice; deployment requires counsel-reviewed data-rights mapping for every channel in every jurisdiction. The claim made here is architectural: the system is buildable such that the most sensitive data never moves, and the only thing that moves is proof.

6.1 What this system does not do

It does not determine creditworthiness, price credit, set terms, or generate adverse-action reasons. It does not infer protected-class membership, and its cohort priors are firewalled from any underwriting feature space. It does not produce location histories, maps, or geofences for human review - no analyst sees a trajectory. It does not track devices in real time and holds no standing surveillance capability; it evaluates coherence over a bounded historical window for a single asserted identity under a documented purpose. And when evidence is insufficient - coverage below threshold, consent basis unresolved, prior version mismatched - it abstains rather than scores, and says so in the packet.

7. Closing

The empirical situation can be stated in one sentence: real human spatiotemporal behavior occupies a thin, well-charted manifold - truncated power-law displacements rescaled by personal radius of gyration, Zipfian visitation under a conserved location capacity, a 93% predictability ceiling that demographics barely move, seventeen daily motifs, positive burstiness beneath 24- and 168-hour spectral lines, an hour-a-day travel budget - and demographic cohorts occupy identifiable neighborhoods on that manifold, with parameters set by life course, occupation, household structure, and geography.

For detection, the structure dictates the architecture: nested tests, from universal goodness-of-fit, through cohort-conditional likelihood against the *claimed* demographic, to cross-channel coherence and ring-level joint consistency - with cohort priors versioned (the remote-work break is a warning), conditioned on measurement (serving-cell density), firewalled from underwriting, and contained behind the Section 6 boundary, across which only proof travels. Every statistic named in this paper is computable from the record, replayable from the same inputs, and attributable to a published empirical law. That is what makes spatiotemporal coherence not merely a fraud signal but *evidence* - the kind a regulated institution can stand behind when asked to show its work.

Appendix A - Production Validation Specification

This appendix specifies what a production deployment estimates, how, and when it refuses to answer. It is the contract that makes Section 3's tier system honest: every Tier 2 and Tier 3 claim corresponds to a quantity below that is estimated in-domain before it is relied upon.

A.1 Estimated quantities. Per identity, the system fits the parameter vector $\hat{\theta}$ comprising: displacement-law parameters ($\hat{\beta}$, $\hat{\kappa}$, $\hat{\Delta r}_0$) via maximum likelihood on the truncated power-law family with lower-cutoff selection per Clauset, Shalizi & Newman[27]; radius of gyration r_g and the returner ratio s_2; the visitation exponent $\hat{\zeta}$ from the rank–frequency distribution; the exploration exponent $\hat{\mu}$ from $S(t)$ growth; the entropy triple (S_{rand}, S_{unc}, S_{actual}) via Lempel–Ziv estimation with finite-sample correction, plus the implied Fano bound; burstiness B on inter-event times; circadian and weekly spectral line power and phase via the Lomb–Scargle periodogram, the appropriate estimator under the irregular, event-driven sampling of Section 4; the daily-motif mixture against the Schneider census; per-anchor-class dwell-distribution parameters; the handoff-adjacency feasibility rate against the serving network's cell graph; the uniqueness collision score against the reference corpus; and the address-event coherence score linking bureau-channel moves to trajectory-channel anchor migrations.

A.2 Normalization and stratification. All quantities are estimated conditional on sampling regime (event-driven vs. network-driven) and serving-cell density, per Sections 3.8 and 4; cross-regime comparison of raw statistics is prohibited. Minimum-coverage requirements - event count, observation span, and per-week coverage floors - are fixed before estimation and recorded in the packet.

A.3 Inference. The cohort-conditional test is a likelihood ratio of $\hat{\theta}$ under the claimed-demographic prior versus the population mixture, with the prior carrying an explicit version stamp recording its training window, source population, regime stratification, and handling of the post-2020 remote-work break. Confidence intervals come from block bootstrap over time, preserving temporal dependence. Thresholds are set against cohort-specific false-positive targets, and calibration is monitored per cohort so that error burden cannot silently concentrate on legitimate outlier populations (Section 3.9).

A.4 Abstention conditions. The system abstains - emitting an abstention state, not a score - when coverage falls below the fixed floors, when the consent or permissible-purpose basis for any contributing channel is unresolved, when the prior version does not match the deployment's regime stratification, when only a single evidence channel is available, or when the fitted model fails its own goodness-of-fit diagnostics. Abstention is a first-class output recorded in the packet, not an error.

A.5 Evidence packet fields. Input content hashes per channel; $\hat{\theta}$ with confidence intervals; cohort-prior version; per-level test statistics, thresholds, and outcomes; the decision path through Figure 1; abstention state and reason codes where applicable; legal basis (consent/authorization reference and permissible purpose); feature-exclusion ledger reference; retention clock; dispute pointer; and a signature over the foregoing.

[27] Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. SIAM Review, 51(4), 661–703.