

Analytical Workflow (Natural-Language Description)

Goal: From geolocated microblogs, a map, zone population statistics, and weather data, determine (a) where and when the outbreak began, and (b) which transmission mechanism (person-to-person, airborne, waterborne, or other) best explains its spread — to guide resource deployment. Because the raw signal (social-media text) is noisy and the causal question requires domain interpretation, the workflow is human-in-the-loop throughout, with two explicit refinement loops.

1. Extract an illness signal from microblogs, refined interactively. Symptom/severity/death-related keywords (a *specification*, initially analyst-supplied) are matched against message text to produce a symptom-indicator feature, which is then used to filter the full microblog set down to a set of *illness reports*. Because keyword-based filtering is imprecise, a loop maps the reports geographically, lets the analyst visually inspect their spatial/temporal density and sample content, and assesses whether the set looks like a genuine illness signal or noise. If not, the keyword lexicon is revised (a new specification) and extraction/filtering are re-run, until the filtered set is judged reliable.

2. Establish the spatial-temporal footprint and locate the origin. Illness reports are aggregated into daily counts per zone and normalised by zone population to obtain per-capita incidence rates (correcting for the fact that zones differ greatly in population). The incidence rates are visualised as an animated choropleth map over the outbreak period, and the analyst interprets this display to identify the earliest-affected zone(s), approximate onset time, and the overall shape of the spread (e.g., a single radiating focus vs. several near-simultaneous foci) — the candidate outbreak origin.

3. Test transmission-mechanism hypotheses, refined iteratively. Three mechanism-specific signatures are computed from the illness reports and contextual data: (i) alignment between the daily spread direction from the origin and prevailing wind direction/speed (airborne signature), (ii) proximity of cases to water bodies/water infrastructure on the map (waterborne signature), and (iii) whether the zone-to-zone spread follows population density and road connectivity in a gradual, contact-like pattern versus appearing simultaneously in disconnected zones (person-to-person vs. point-source signature). These are displayed together on coordinated map overlays (wind vectors, water layer, zone-connectivity graph) so the analyst can visually compare the observed spread against each candidate signature, informed also by injected epidemiological domain knowledge about what each mechanism typically looks like spatio-temporally. The analyst assesses whether the evidence is conclusive; if not, the analysis scope (time window,

spatial resolution, map layers considered) is adjusted and the mechanism-specific features are recomputed, iterating until one mechanism (or ranked set) is clearly best supported.

4. Synthesise findings. The identified origin location/time and the transmission-mechanism conclusion are combined into an explicit knowledge artifact summarising where the outbreak started and how it is most plausibly spreading, to inform prioritisation of emergency-management resources (e.g., targeted containment appropriate to airborne, waterborne, or person-to-person spread).

ATWL Representation

atwl

```
workflow vastopolis-outbreak-origin-and-transmission
  template: characterise -> define-unit ->
    loop(contextualise -> visualise -> assess -> define-unit) ->
    characterise -> characterise ->
    contextualise -> visualise -> abstract ->
    loop(characterise -> visualise -> assess -> generate-knowledge)
    generate-knowledge
  description: "Human-in-the-loop workflow to identify the spatial-temporal
    origin of a flu-like outbreak in Vastopolis and to determine the most
    plausible transmission mechanism (person-to-person, airborne,
    waterborne, or other), combining social-media evidence with map,
    population, and weather context to support resource deployment
    decisions"

# =====
# INPUT
# =====

artifact D_microblogs : entities
  origin: given
  internal structure: elementary
  embedment: {set, space, time}
  features:
    - id: text_content
      value structure: atomic
      value type: text
      description: "Free text content of the message"
    - id: location
```

```
    value structure: atomic
    value type: spatial
    description: "GPS coordinates of posting"
  - id: timestamp
    value structure: atomic
    value type: temporal
    description: "Posting time"
  description: "Geolocated microblog messages posted from mobile devices
    over the outbreak period"
```

artifact D_zones : entities

```
  origin: given
  internal structure: region
  embedment: space
  features:
    - id: population
      value structure: atomic
      value type: numeric
      description: "Resident population of the zone"
    - id: zone_name
      value structure: atomic
      value type: categorical
      description: "Zone identifier/name"
  description: "13 city zones with population statistics, used as spatial
    units for incidence normalisation"
```

artifact D_weather : entities

```
  origin: given
  internal structure: elementary
  embedment: time
  features:
    - id: wind_speed
      value structure: atomic
      value type: numeric
      description: "Daily wind speed"
    - id: wind_direction
      value structure: atomic
      value type: numeric
      description: "Daily wind direction"
  description: "Daily weather observations over the outbreak period"
```

artifact D_map_context : entities

```
origin: given
internal structure: elementary
embedment: space
features:
  - id: layer_type
    value structure: atomic
    value type: categorical
    description: "Map layer category: highway, hospital, landmark,
      water_body"
  - id: geometry
    value structure: atomic
    value type: spatial
    description: "Geometry of the map feature"
description: "Static geographic layers of the metropolitan area (road
  network, hospitals, landmarks, water bodies), used as spatial refere
  context for mapping and mechanism analysis"
```

```
artifact S_keyword_spec : specification
```

```
origin: given
representation form: "keyword lexicon"
description: "Initial lexicon of flu-like symptom, severity, and
  death-related terms used to identify illness-related messages"
```

```
artifact K_domain_transmission_knowledge : knowledge(D_weather, D_map_co
```

```
origin: given
representation form: "domain rules"
description: "Epidemiological domain knowledge describing expected
  spatial-temporal signatures of airborne, waterborne, and
  person-to-person transmission, used to interpret observed spread
  patterns"
```

```
# =====
# STEP 1: EXTRACT AND FILTER ILLNESS-RELATED REPORTS
# =====
```

```
transform T_extract_symptom_signal :
```

```
intent: characterise
manner: "keyword/text feature extraction"
input: D_microblogs, S_keyword_spec
output: F_symptom_signal
actor: machine
description: "Match message text against the keyword lexicon to derive
```

```

    indicators of symptom mention, severity, and death-related content"

artifact F_symptom_signal : feature(D_microblogs)
  value structure: vector
  value type: categorical
  representation form: "boolean indicators for symptom mention, severity
    keywords, and death-related keywords"
  description: "Text-derived illness-indicator features per message"

transform T_filter_illness_reports :
  intent: define-unit
  manner: "filter"
  input: D_microblogs, F_symptom_signal
  output: D_illness_reports
  actor: machine
  description: "Select messages whose content indicates flu-like illness"

artifact D_illness_reports : entities
  internal structure: elementary
  embedment: {set, space, time}
  description: "Geolocated, timestamped messages indicative of flu-like
    illness"

# =====
# STEP 2: FILTER REFINEMENT LOOP
# =====

loop L_filter_refinement:
  purpose: "Refine the keyword lexicon and filtering criteria so the
    illness-report set is a reliable proxy for actual illness incidence"
  until: "Visual inspection confirms filtered reports are plausibly
    illness-related and their spatial-temporal density looks consistent
    with a genuine outbreak signal rather than noise"
  body:

    transform T_contextualise_reports_map :
      intent: contextualise
      manner: "map-based"
      input: D_illness_reports, D_map_context
      output: A_reports_map
      actor: machine
      description: "Position illness reports on the metropolitan map"

```

```

artifact A_reports_map : arrangement(D_illness_reports)
  context: D_map_context
  principle: "geographic coordinates"
  description: "Spatial placement of illness reports relative to
    roads, hospitals, landmarks, and water bodies"

transform T_visualise_reports :
  intent: visualise
  manner: "coloured points on map with time slider"
  input: A_reports_map, D_illness_reports
  output: V_reports_check
  actor: machine
  description: "Display raw spatial-temporal distribution of filtered
    reports for inspection of content and density"

artifact V_reports_check : visualisation(A_reports_map, D_illness_reports)
  layout: "map with time slider"
  form: "coloured points"
  encoding: "position from A_reports_map; colour/opacity from
    timestamp recency"
  description: "Interactive view for judging plausibility of the
    filtered illness signal"

transform T_assess_filter :
  intent: assess
  manner: "judge relevance and noise level of filtered reports"
  input: V_reports_check, F_symptom_signal
  output: K_filter_quality
  actor: human
  description: "Assess whether filtered reports plausibly reflect
    genuine illness incidence or contain excessive noise/false
    positives"

artifact K_filter_quality : knowledge(D_illness_reports, F_symptom_signal)
  representation form: "quality judgment"
  description: "Judgment of filtering adequacy and need for lexicon
    refinement"

if K_filter_quality indicates refinement needed:
  then:
    transform T_refine_keywords :

```

```
intent: generate-knowledge
manner: "formulate methodological decision"
input: K_filter_quality, F_symptom_signal
output: S_keyword_spec_r
actor: human
description: "Revise keyword lexicon/filtering rules based on
    observed false positives/negatives"

artifact S_keyword_spec_r : specification
    representation form: "keyword lexicon"
    description: "Revised keyword lexicon for illness-indicative
        content"

transform T_extract_symptom_signal_r :
    intent: characterise
    manner: "keyword/text feature extraction"
    input: D_microblogs, S_keyword_spec_r
    output: F_symptom_signal_r
    actor: machine
    description: "Re-extract illness-indicator features using the
        revised lexicon"

artifact F_symptom_signal_r : feature(D_microblogs)
    value structure: vector
    value type: categorical
    description: "Updated illness-indicator features per message"

transform T_filter_illness_reports_r :
    intent: define-unit
    manner: "filter"
    input: D_microblogs, F_symptom_signal_r
    output: D_illness_reports_r
    actor: machine
    description: "Re-select illness-related messages using updated
        features"

artifact D_illness_reports_r : entities
    internal structure: elementary
    embedment: {set, space, time}
    description: "Updated set of illness-related messages"

assign:
```

```

        S_keyword_spec := S_keyword_spec_r
        F_symptom_signal := F_symptom_signal_r
        D_illness_reports := D_illness_reports_r
    else:
        exit loop L_filter_refinement

end loop L_filter_refinement

# =====
# STEP 3: SPATIAL-TEMPORAL FOOTPRINT AND ORIGIN IDENTIFICATION
# =====

transform T_aggregate_zone_counts :
    intent: characterise
    manner: "aggregate"
    input: D_illness_reports, D_zones
    output: F_case_counts
    actor: machine
    description: "Spatially join illness reports to zones and aggregate
        daily case counts per zone"

artifact F_case_counts : feature(D_zones)
    value structure: vector
    value type: numeric
    representation form: "daily case-count time series per zone"
    description: "Daily number of illness reports per zone"

transform T_normalise_incidence :
    intent: characterise
    manner: "compute-statistics (per-capita normalisation)"
    input: F_case_counts, D_zones
    output: F_incidence_rate
    actor: machine
    description: "Normalise daily case counts by zone population to obtain
        comparable per-capita incidence rates"

artifact F_incidence_rate : feature(D_zones)
    value structure: vector
    value type: numeric
    representation form: "daily per-capita incidence rate time series per
        zone"
    description: "Population-normalised daily illness incidence per zone"

```

```

transform T_visualise_incidence_spread :
  intent: visualise
  manner: "animated choropleth map with time slider"
  input: D_zones, F_incidence_rate, D_map_context
  output: V_incidence_map
  actor: machine
  description: "Show spatial-temporal progression of normalised illness
    incidence across zones and days"

artifact V_incidence_map : visualisation(D_zones, F_incidence_rate, D_ma
  layout: "map with time slider"
  form: "choropleth coloured zones"
  encoding: "zone geometry from D_map_context; colour intensity from
    F_incidence_rate per day"
  description: "Reveals onset location and diffusion pattern of the
    outbreak"

transform T_abstract_origin_pattern :
  intent: abstract
  manner: "perception-and-interpretation"
  input: V_incidence_map, V_reports_check
  output: P_outbreak_origin
  actor: human
  description: "Identify the zone(s) and approximate time where the
    outbreak appears to have started, and characterise the overall
    spatial diffusion pattern"

artifact P_outbreak_origin : pattern(D_zones, D_illness_reports)
  representation form: "spatial-temporal description of earliest cluster
    and diffusion direction"
  description: "Candidate outbreak origin (zone/time) and qualitative
    description of how the outbreak has spread so far"

# =====
# STEP 4: TRANSMISSION-MECHANISM ANALYSIS LOOP
# =====

artifact S_mechanism_spec : specification
  origin: given
  representation form: "analysis scope settings"
  description: "Initial scope for transmission-mechanism analysis: full

```

time period, all zones, primary map layers (water bodies, road network)"

loop L_transmission_analysis:

purpose: "Iteratively test and refine mechanism-specific evidence until the spread pattern is clearly best explained by one (or a ranked set of) transmission mechanism(s)"

until: "Assessment yields a clear, defensible conclusion about the dominant transmission pathway"

body:

transform T_characterise_wind_alignment :

intent: characterise

manner: "define-relations (directional correlation)"

input: P_outbreak_origin, D_illness_reports, D_weather, S_mechanism

output: F_wind_alignment

actor: machine

description: "Compute daily alignment between the direction of case spread from the origin and prevailing wind direction/speed (airborne-transmission signature)"

artifact F_wind_alignment : feature(D_zones)

value structure: vector

value type: numeric

representation form: "daily wind-spread alignment score per zone"

description: "Evidence for airborne transmission"

transform T_characterise_water_proximity :

intent: characterise

manner: "compute-statistics (spatial distance)"

input: D_illness_reports, D_map_context, S_mechanism_spec

output: F_water_proximity

actor: machine

description: "Compute distance of illness-report locations to nearest water body/water infrastructure (waterborne-transmission signature)"

artifact F_water_proximity : feature(D_illness_reports)

value structure: atomic

value type: numeric

description: "Evidence for waterborne transmission"

```

transform T_characterise_diffusion_shape :
  intent: characterise
  manner: "define-relations (spatial diffusion pattern)"
  input: D_illness_reports, D_zones, D_map_context, S_mechanism_spec
  output: F_diffusion_shape
  actor: machine
  description: "Characterise zone-to-zone spread over time relative
    population density and road connectivity, distinguishing gradual
    contact-like diffusion (person-to-person) from near-simultaneous
    onset in disconnected zones (point-source)"

artifact F_diffusion_shape : feature(D_zones)
  value structure: relational configuration
  representation form: "time-stamped zone-adjacency graph weighted by
    case-onset lag"
  description: "Evidence distinguishing person-to-person from
    point-source spread"

transform T_visualise_mechanism_overlays :
  intent: visualise
  manner: "coordinated overlay views"
  input: P_outbreak_origin, F_wind_alignment, F_water_proximity,
    F_diffusion_shape, D_map_context
  output: V_transmission_overlays
  actor: machine
  description: "Overlay wind vectors, water-proximity gradient, and
    zone-connectivity graph on the map for visual comparison against
    candidate transmission signatures"

artifact V_transmission_overlays : visualisation(P_outbreak_origin,
  F_wind_alignment, F_water_proximity, F_diffusion_shape, D_map_context)
  layout: "map with wind-vector, water-layer, and zone-adjacency
    overlays, small multiples per day"
  form: "vector arrows, coloured zones, network edges"
  encoding: "wind arrows from F_wind_alignment; water-proximity color
    gradient from F_water_proximity; connectivity edges from
    F_diffusion_shape"
  description: "Visual comparison of observed spread against airborne
    waterborne, and person-to-person signatures"

transform T_assess_transmission_hypotheses :
  intent: assess

```

```

manner: "compare alternative results produced under different
specifications"
input: V_transmission_overlays, F_wind_alignment, F_water_proximity,
      F_diffusion_shape, K_domain_transmission_knowledge
output: K_transmission_evidence
actor: human
description: "Judge which candidate transmission mechanism(s) best
explain the observed spread, informed by domain knowledge of
typical mechanism signatures; determine if evidence is conclusive"

artifact K_transmission_evidence : knowledge(F_wind_alignment,
      F_water_proximity, F_diffusion_shape)
representation form: "comparative quality judgment"
description: "Evaluation of which transmission mechanism(s) are
best supported by current evidence, and whether further analysis
is needed"

if K_transmission_evidence indicates refinement needed:
  then:
    transform T_adjust_analysis_scope :
      intent: generate-knowledge
      manner: "formulate methodological decision"
      input: K_transmission_evidence
      output: S_mechanism_spec_r
      actor: human
      description: "Decide how to refine the mechanism analysis, e.g.
        narrow time window to early outbreak days, focus on specific
        zones, or incorporate additional map layers"

    artifact S_mechanism_spec_r : specification
      representation form: "analysis scope settings"
      description: "Revised scope for transmission-mechanism analysis"

    assign:
      S_mechanism_spec := S_mechanism_spec_r
  else:
    exit loop L_transmission_analysis

end loop L_transmission_analysis

# =====
# STEP 5: SYNTHESISE FINDINGS

```

```
# =====

transform T_synthesise_findings :
  intent: generate-knowledge
  manner: "formulate-statements"
  input: P_outbreak_origin, K_transmission_evidence, K_filter_quality
  output: K_outbreak_findings
  actor: human
  description: "Synthesise conclusions about the outbreak's spatial-temporal
    origin and its most plausible transmission mechanism(s), to inform
    prioritisation of emergency-management resources"

artifact K_outbreak_findings : knowledge(D_illness_reports, D_zones,
  D_weather, D_map_context)
  representation form: "statements and recommendations"
  description: "Conclusions on outbreak origin location/time and dominant
    transmission mechanism, with implications for resource deployment"
```