

Analytical workflow

The workflow treats communities as **time-varying, overlapping groups of authors**, rather than as one fixed partition of the author population.

1. **Construct temporal collaboration networks.** Convert papers into yearly or multi-year sliding-window co-authorship-network snapshots. Nodes are disambiguated DBLP authors; weighted edges represent repeated collaboration. Sliding windows reduce instability caused by sparse annual data.
2. **Represent research foci.** Derive paper-topic distributions from titles, abstracts, and keywords. Aggregate these distributions for authors and later for communities. Topics characterise communities but should not automatically define them unless the analyst explicitly chooses a topic-aware community model.
3. **Detect dynamic, overlapping communities.** Fit a temporal overlapping-community model to the collaboration snapshots. The model should produce weighted author–community affiliations for each period, allowing an author to belong to several communities and change affiliations over time. Temporal continuity constraints should link communities across adjacent periods without preventing genuine splits or mergers.
4. **Derive community histories and leadership measures.** Form community instances for each time period and link them into longer-lived community lineages. Compute:
 - persistent, high-affiliation members as candidate cores;
 - collaboration centrality, continuity, and optionally citation impact as complementary leadership indicators;
 - member overlap between communities;
 - topical profiles and topical change over time;
 - continuity relations indicating emergence, merger, split, persistence, and dissolution.
5. **Visual inspection and iterative refinement.** Provide linked views: temporal alluvial/ community-evolution views, co-author network views, author-by-community affiliation matrices, topic-profile timelines, and views of candidate core members. The analyst interprets whether detected communities are coherent and substantively meaningful, assesses whether parameter choices create implausible fragmentation or over-merging, and refines the temporal window, affiliation threshold, temporal-smoothness strength, or model family as needed.
6. **Synthesize findings.** Once the results are stable and interpretable, formulate evidence-based conclusions about the number and nature of communities, their

leaders and stable cores, research foci, overlaps, mobility, and detected structural events. Explanations for merges or splits should be framed as interpretations supported by collaboration and topic evidence—not as automatically established causal claims.

ATWL representation

atwl

```
workflow ieee-vis-community-evolution
  template: define-unit -> characterise -> build-model ->
    loop(build-model (dynamic overlapping communities) ->
      characterise -> define-unit -> characterise ->
      contextualise -> visualise -> abstract -> assess) ->
    generate-knowledge
  description: "Human-in-the-loop workflow for identifying, characterising
    and interpreting evolving overlapping co-authorship communities in the
    IEEE VIS literature from 1990 to 2024."

# =====
# INPUT DATA AND INITIAL ANALYTICAL SPECIFICATIONS
# =====

artifact D_papers : entities
  origin: given
  internal structure: elementary
  embedment: time
  features:
    - id: paper_id
      value structure: atomic
      value type: reference
      description: "Unique paper identifier"
    - id: publication_year
      value structure: atomic
      value type: temporal
      description: "Publication year from 1990 to 2024"
    - id: authors
      value structure: list
      value type: reference
      description: "DBLP-disambiguated author identifiers, retained with
        their original author-name strings"
    - id: title_abstract_keywords
```

```
    value structure: list
    value type: text
    description: "Textual metadata used to characterise research foci"
  - id: citations
    value structure: atomic
    value type: numeric
    description: "Citation count associated with the paper record"
  description: "3596 IEEE VIS, InfoVis, VAST, and SciVis paper records."
```

artifact D_periods : entities

```
  origin: given
  internal structure: elementary
  embedment: time
  features:
    - id: temporal_extent
      value structure: atomic
      value type: temporal
      description: "Yearly or analyst-defined sliding temporal window"
  description: "Temporal reference periods used to organise network snapshots and community evolution."
```

artifact S_temporal_network : specification

```
  origin: given
  representation form: "configuration schema"
  description: "Initial choices for temporal window width and overlap, co-authorship edge weighting, treatment of single-author papers, minimum activity threshold, and inclusion of authors with weak ties."
```

artifact S_topic_model : specification

```
  origin: given
  representation form: "parameter settings and method choice"
  description: "Text preprocessing, topic-model family, topic granularity and use of titles, abstracts, and keywords."
```

artifact S_community_model_0 : specification

```
  origin: given
  representation form: "configuration schema"
  description: "Initial dynamic overlapping-community model settings, including affiliation threshold, temporal smoothness, resolution, minimum community size, and optional topic-aware regularisation."
```

artifact K_domain_context : knowledge(D_papers)

```

origin: given
representation form: "annotations and interpretive criteria"
description: "Analyst knowledge about visualization research areas,
  conference history, known terminology changes, and criteria for a
  substantively meaningful research community."

# =====
# STEP 1: CREATE TEMPORAL CO-AUTHORSHIP NETWORK SNAPSHOTS
# =====

transform T_construct_collaboration_snapshots :
  intent: define-unit
  manner: "time-partitioning and co-authorship-network construction"
  input: D_papers, D_periods, S_temporal_network
  output: D_collaboration_snapshots
  actor: machine
  description: "Aggregate papers within each temporal period into weight
    co-authorship-network snapshots. Nodes are unique DBLP authors and
    edge weights represent the strength or frequency of collaboration."

artifact D_collaboration_snapshots : entities
  internal structure: formation
  embedment: time
  features:
    - id: edge_weight
      value structure: atomic
      value type: numeric
      description: "Weighted collaboration strength between two authors"
    - id: active_author_count
      value structure: atomic
      value type: numeric
      description: "Number of authors active in the snapshot"
    - id: paper_count
      value structure: atomic
      value type: numeric
      description: "Number of papers contributing to the snapshot"
  description: "Temporal sequence of weighted co-authorship-network
    snapshots for the IEEE VIS community."

transform T_characterise_collaboration_structure :
  intent: characterise
  manner: "compute network roles and temporal collaboration descriptors"

```

```

input: D_collaboration_snapshots
output: F_author_structural_roles
actor: machine
description: "Compute author-level structural descriptors such as weight, degree, centrality, brokerage, collaboration persistence, and change in collaboration neighbourhood."

artifact F_author_structural_roles : feature(D_collaboration_snapshots)
  value structure: relational configuration
  representation form: "time-indexed author-role attributes"
  description: "Network-derived descriptors used to identify candidate leaders, persistent members, brokers, and structurally changing authors."

# =====
# STEP 2: DERIVE RESEARCH-TOPIC REPRESENTATIONS
# =====

transform T_fit_topic_model :
  intent: build-model
  manner: "fit-topic-model"
  input: D_papers, S_topic_model, K_domain_context
  output: M_topic_model
  actor: hybrid
  description: "Build a topic model over titles, abstracts, and keywords. The analyst reviews topic granularity and vocabulary while the machine estimates topic distributions."

artifact M_topic_model : model(D_papers)
  model type: "topic model"
  representation form: "document-topic and topic-term distributions"
  description: "Computational representation of recurring research foci in the IEEE VIS paper corpus."

transform T_characterise_paper_topics :
  intent: characterise
  manner: "apply topic model"
  input: D_papers, M_topic_model
  output: F_paper_topics
  actor: machine
  description: "Assign each paper a distribution over research topics."

artifact F_paper_topics : feature(D_papers)

```

```

value structure: vector
value type: numeric
representation form: "topic-probability vector per paper"
description: "Topic mixture describing the research foci of each paper"

transform T_abstract_topic_semantics :
  intent: abstract
  manner: "name-patterns"
  input: M_topic_model, F_paper_topics, K_domain_context
  output: P_topic_semantics
  actor: human
  description: "Inspect topic terms and representative papers, then assign
    meaningful research-focus labels to coherent topics."

artifact P_topic_semantics : pattern(M_topic_model, F_paper_topics)
  representation form: "named topic descriptions"
  description: "Human-interpretable descriptions of major visualization
    research topics and problem areas."

# =====
# STEP 3: INITIALISE THE COMMUNITY-MODEL SPECIFICATION
# =====

assign:
  S_community_model := S_community_model_0

# =====
# STEP 4: INTERACTIVE COMMUNITY-DISCOVERY AND VALIDATION LOOP
# =====

loop L_community_refinement:
  purpose: "Identify temporally persistent but overlapping collaboration
    communities and refine modelling choices until the results are both
    structurally credible and interpretable by the analyst."
  until: "Community trajectories, cores, overlaps, topical profiles, and
    structural events are sufficiently stable, coherent, and useful for
    answering the analyst's questions."
  body:

    # =====
    # Step 4a: Fit a dynamic overlapping-community model
    # =====

```

```

transform T_fit_dynamic_communities :
  intent: build-model
  manner: "fit dynamic overlapping-community model"
  input: D_collaboration_snapshots, F_author_structural_roles,
         S_community_model
  output: M_dynamic_communities
  actor: hybrid
  description: "Fit a temporal community model that permits weighted
    multiple memberships, author mobility, and persistence of community
    identity across adjacent periods. Topic information may be used
    if specified as a regularisation or validation choice."

artifact M_dynamic_communities : model(D_collaboration_snapshots,
    F_author_structural_roles)
  model type: "dynamic overlapping-community model"
  representation form: "time-indexed weighted author-community
    affiliation parameters with temporal continuity relations"
  description: "Formal model of evolving collaboration communities in
    which authors may belong to several communities simultaneously."

# =====
# Step 4b: Derive affiliations and period-specific community instances
# =====

transform T_characterise_affiliations :
  intent: characterise
  manner: "infer weighted community affiliations"
  input: D_collaboration_snapshots, M_dynamic_communities
  output: F_author_community_affiliations
  actor: machine
  description: "Infer each author's weighted affiliation with one or
    more communities in each temporal period."

artifact F_author_community_affiliations : feature(D_collaboration_s
  value structure: relational configuration
  representation form: "time-indexed weighted author-community
    bipartite affiliation relations"
  description: "Affiliation strengths that represent overlapping
    membership, entry, exit, and movement between communities."

transform T_define_community_periods :

```

```

intent: define-unit
manner: "group-by-weighted community affiliation"
input: D_collaboration_snapshots, F_author_community_affiliations,
      S_community_model
output: D_community_periods
actor: machine
description: "Create a community instance for each identified
             community in each temporal period, retaining weighted membership
             rather than forcing authors into one exclusive group."

artifact D_community_periods : entities
  internal structure: group/cluster
  embedment: time
  features:
    - id: member_authors
      value structure: list
      value type: reference
      description: "DBLP author identifiers belonging to the community
                   instance, with associated affiliation strengths"
    - id: active_size
      value structure: atomic
      value type: numeric
      description: "Number of active members in the period"
  description: "Period-specific overlapping co-authorship communities"

# =====
# Step 4c: Derive lineages, core members, leadership, and topical focus
# =====

transform T_characterise_community_evolution :
  intent: characterise
  manner: "compute temporal continuity, overlap, and community roles"
  input: D_community_periods, F_author_community_affiliations,
        F_author_structural_roles
  output: F_community_continuity, F_membership_overlap,
         F_core_and_leadership
  actor: machine
  description: "Compute similarity and continuity between community
               instances, shared memberships between concurrent communities, and
               persistent high-affiliation members with structural leadership roles"

artifact F_community_continuity : feature(D_community_periods)

```

```

value structure: relational configuration
representation form: "directed temporal similarity graph"
description: "Relations linking community instances across periods
    one-to-many and many-to-one relations provide evidence of splits
    and mergers."

artifact F_membership_overlap : feature(D_community_periods)
    value structure: matrix
    value type: numeric
    representation form: "pairwise weighted membership-overlap matrix
        per temporal period"
    description: "Degree to which communities share authors, including
        authors with substantial simultaneous affiliations."

artifact F_core_and_leadership : feature(D_community_periods)
    value structure: list
    value type: reference
    representation form: "ranked author lists with persistence,
        affiliation, centrality, brokerage, and optional citation measures"
    description: "Candidate stable core members and leaders, derived from
        sustained membership and collaboration-network roles; citation
        measures are contextual rather than defining community leadership"

transform T_define_community_lineages :
    intent: define-unit
    manner: "link temporally continuous community instances"
    input: D_community_periods, F_community_continuity
    output: D_community_lineages
    actor: machine
    description: "Link related period-specific community instances into
        longer-lived community histories while retaining branching relations
        for mergers and splits."

artifact D_community_lineages : entities
    internal structure: sequence
    embedment: set
    features:
        - id: lifespan
            value structure: atomic
            value type: temporal
            description: "Temporal extent from appearance to dissolution"
        - id: stability_score

```

```

        value structure: atomic
        value type: numeric
        description: "Persistence of the community and its core across
            successive periods"
description: "Evolving community trajectories composed of linked
    community instances over time."

transform T_characterise_community_foci :
    intent: characterise
    manner: "aggregate topic distributions by community and period"
    input: D_papers, F_paper_topics, D_community_periods,
        F_author_community_affiliations
    output: F_community_topic_profiles
    actor: machine
    description: "Aggregate papers and topic distributions associated
        community members to estimate each community's research foci and
        their change through time."

artifact F_community_topic_profiles : feature(D_community_periods)
    value structure: vector
    value type: numeric
    representation form: "time-indexed topic-distribution profiles"
    description: "Research-focus profiles for each community instance,
        enabling comparison of topical specialisation and topical evolution"

# =====
# Step 4d: Arrange and visualise the evolving community structure
# =====

transform T_contextualise_community_evolution :
    intent: contextualise
    manner: "timeline-based"
    input: D_community_periods, D_periods, F_community_continuity
    output: A_community_timeline
    actor: machine
    description: "Place community instances along the temporal reference
        periods and organise lineage connections to support tracing of
        persistence, emergence, merging, splitting, and dissolution."

artifact A_community_timeline : arrangement(D_community_periods)
    context: D_periods
    principle: "temporal placement by publication period; lineage link

```

```

    derived from continuity relations"
description: "Timeline arrangement of community instances and their
    temporal continuity relations."

transform T_visualise_communities :
    intent: visualise
    manner: "coordinated temporal network and affiliation views"
    input: A_community_timeline, D_collaboration_snapshots,
        D_community_lineages, F_author_community_affiliations,
        F_membership_overlap, F_core_and_leadership,
        F_community_topic_profiles, P_topic_semantics
    output: V_community_evolution
    actor: machine
    description: "Create linked interactive views for inspecting
        community evolution, collaboration structure, overlapping
        memberships, candidate cores and leaders, and topic changes."

artifact V_community_evolution : visualisation(A_community_timeline,
    D_collaboration_snapshots, F_author_community_affiliations,
    F_membership_overlap, F_core_and_leadership,
    F_community_topic_profiles)
    layout: "coordinated alluvial timeline, temporal node-link views,
        author-by-community affiliation matrix, and topic-profile timeline"
    form: "flows, nodes and weighted edges, coloured matrix cells,
        ranked lists, and stacked topic distributions"
    encoding: "timeline position from A_community_timeline; community
        colour by lineage; flow width by membership continuity; author
        affiliation intensity by membership weight; node prominence by
        structural role; topic colour by P_topic_semantics"
    description: "Interactive visual environment for examining community
        identities, cores, leadership, overlap, mobility, and topical changes"

# =====
# Step 4e: Human interpretation and quality assessment
# =====

transform T_abstract_community_patterns :
    intent: abstract
    manner: "perception-and-interpretation"
    input: V_community_evolution, D_community_lineages,
        F_core_and_leadership, F_membership_overlap,
        F_community_topic_profiles, P_topic_semantics,

```

```

        K_domain_context
output: P_community_dynamics
actor: human
description: "Identify and name salient community structures:
    stable cores, bridging researchers, persistent overlap, topical
    shifts, emerging communities, mergers, splits, and dissolutions.

artifact P_community_dynamics : pattern(D_community_lineages,
    F_community_continuity, F_community_topic_profiles)
representation form: "annotated community histories and event type
description: "Interpreted patterns of community persistence, mover
    overlap, topical development, merger, split, emergence, and
    dissolution."

transform T_assess_community_solution :
    intent: assess
    manner: "evaluate dynamic-community quality and interpretability"
    input: V_community_evolution, M_dynamic_communities,
        D_community_lineages, F_membership_overlap,
        F_core_and_leadership, P_community_dynamics,
        K_domain_context, S_community_model
    output: K_community_quality, S_community_model_r
    actor: hybrid
    description: "Evaluate structural fit, temporal stability, robustness
        to reasonable parameter changes, coherence of core members,
        plausibility of overlap and event detection, topic interpretability
        and usefulness for the research questions."

artifact K_community_quality : knowledge(M_dynamic_communities,
    D_community_lineages, V_community_evolution)
representation form: "quality judgment and diagnostic notes"
description: "Assessment of whether the current community solution
    over-fragmented, over-merged, temporally unstable, insufficiently
    overlapping, or otherwise inadequate for interpretation."

artifact S_community_model_r : specification
    representation form: "revised configuration settings"
    description: "Refined temporal window, resolution, affiliation
        threshold, temporal smoothness, community-size rule, or model
        family selected in response to assessment."

if K_community_quality indicates further refinement needed:

```

```

    then:
        assign:
            S_community_model := S_community_model_r
    else:
        exit loop L_community_refinement

end loop L_community_refinement

# =====
# STEP 5: FORMULATE ANALYTICAL FINDINGS
# =====

transform T_synthesise_findings :
    intent: generate-knowledge
    manner: "formulate-statements"
    input: D_community_lineages, F_core_and_leadership,
           F_membership_overlap, F_community_topic_profiles,
           P_community_dynamics, K_community_quality,
           P_topic_semantics
    output: K_community_findings
    actor: human
    description: "Formulate evidence-based conclusions about the major
        collaboration communities, their stable cores and leaders, research
        foci, overlap and mobility, and the timing and interpretation of
        structural changes."

artifact K_community_findings : knowledge(D_community_lineages,
    P_community_dynamics, F_community_topic_profiles)
    representation form: "analytical statements, annotated timelines, and
        qualified explanations"
    description: "Findings addressing the number of communities; their
        stability, leaders, and cores; their main and evolving research foci
        membership overlap and mobility; and documented merger, split, and
        dissolution events with supporting collaboration and topical evidence"

```