

Goal

Build an **interactive temporal, overlapping community-analysis workflow** rather than relying on one clustering result. Communities should be inferred from collaboration patterns, then interpreted and validated using topics, venues, citations, and expert review.

1. Prepare and validate the data

1.1 Normalize paper and author records

Use the DBLP author identifier as the primary author key; retain raw names for display.

For each paper, retain:

- Paper ID
- Year
- Conference/track: VIS, InfoVis, VAST, SciVis
- DBLP author IDs and display names
- Title, abstract, keywords
- Citation count or citation links, if available

1.2 Perform quality checks

Before community detection, provide an analyst-facing audit view for:

- Papers with missing or unusually long author lists
- Authors with implausibly high publication counts
- Duplicate papers across conference datasets
- Missing abstracts/keywords by year
- Very recent papers with immature citation counts
- Changes in conference naming or scope over time

Do not use raw author strings to define identity; use DBLP IDs. Flag possible remaining errors for manual inspection.

2. Model collaboration as a temporal network

2.1 Create an author–paper bipartite network

Represent:

- Authors as one node type
- Papers as another node type
- An edge when an author wrote a paper

This preserves the original collaboration structure and avoids prematurely converting every multi-author paper into a dense clique.

2.2 Derive an author collaboration network

For each time period, project the bipartite graph to an author network:

- Node: author
- Edge: coauthorship relationship
- Edge weight: strength of collaboration

A useful fractional weighting is:

$$w_{ij} = \sum_{p: i,j \in A_p} \frac{1}{|A_p|-1}$$

where A_p is the set of authors on paper p . This prevents a paper with many authors from contributing disproportionately many strong ties.

Optionally incorporate recency:

$$w_{ij,t} = \sum_{p: i,j \in A_p, y_p \leq t} \frac{\exp[-\lambda(t-y_p)]}{\sum_{p: i,j \in A_p, y_p \leq t} \exp[-\lambda(t-y_p)]}$$

This emphasizes current collaboration while retaining historical ties.

2.3 Use multiple time views

Create several network representations:

View	Purpose
Full-period network, 1990–2024	Identify broad historical communities

View	Purpose
Non-overlapping periods, e.g. 1990–1994, 1995–1999	Compare eras clearly
Rolling windows, e.g. 5-year windows	Track gradual evolution
Year-by-year network	Inspect sudden changes, but may be sparse
Conference-specific networks	Distinguish VIS, InfoVis, VAST, and SciVis structures

A five-year rolling window is a reasonable initial choice, but it should be adjustable interactively.

3. Identify candidate communities

No single algorithm should be treated as ground truth. Run several methods and compare results.

3.1 Establish a non-overlapping baseline

Apply a standard weighted community detection method such as:

- Leiden
- Louvain
- Degree-corrected stochastic block model

This gives an easy-to-explain baseline partition and helps identify the largest collaboration regions.

3.2 Use overlapping community detection

Because researchers may participate in multiple communities, use methods that support multi-membership, such as:

- Link communities / edge clustering
- BigCLAM
- Mixed-membership stochastic block models
- Overlapping stochastic block models
- OSLOM

For each author i , retain a membership score rather than only a single label:

$$m_{ik} \in [0,1] \quad \forall i \in [0,1] \quad k \in [0,1]$$

where m_{ik} represents the strength of author i 's membership in community k .

This distinguishes:

- **Core members:** high and persistent membership in one community
- **Boundary spanners:** meaningful memberships in multiple communities
- **Transient contributors:** brief or weak membership
- **New entrants:** recently appearing members

3.3 Make algorithm parameters inspectable

The analyst should be able to adjust:

- Time-window length
- Minimum edge weight
- Resolution parameter
- Minimum community size
- Whether repeated collaboration is required
- Inclusion/exclusion of single-paper authors

Show how results change under these choices. A community that appears only under one narrow parameter setting should be treated cautiously.

4. Track communities through time

Community labels generated independently in each time window are not inherently comparable. Align them across adjacent windows.

4.1 Match communities between periods

For communities $C_{k,t}$ and $C_{l,t+1}$, calculate similarity using weighted membership overlap, for example weighted Jaccard similarity:

$$J(C_{k,t}, C_{l,t+1}) = \frac{\sum_i \min(m_{ik,t}, m_{il,t+1})}{\sum_i \max(m_{ik,t}, m_{il,t+1})}$$

Match highly similar communities across consecutive windows.

4.2 Identify temporal events

Classify changes using membership flow and overlap:

Event	Evidence
Continuation	One community strongly matches one later community
Growth / contraction	Same lineage, but substantial membership-size change
Merge	Multiple prior communities substantially contribute to one later community
Split	One prior community substantially contributes to multiple later communities
Birth	A new community has no strong predecessor
Dissolution	A community has no strong successor
Reorganization	Membership persists but is redistributed across communities

Do not assign an event solely from an overlap threshold. Present candidate events to the analyst with supporting evidence.

4.3 Display a community lineage graph

Use an alluvial/Sankey-style diagram or temporal lineage graph:

- Each block represents a community in one time window.
- Block size represents active members, papers, or weighted collaboration volume.
- Links represent author-membership flow.
- Link width reflects overlap strength.
- Color reflects community lineage or primary research focus.

Allow the analyst to click a transition and inspect the authors and papers responsible for it.

5. Define stability, cores, and leaders

5.1 Measure community stability

Use several complementary measures:

Measure	Interpretation
Membership retention	Fraction of members retained from one period to the next
Core retention	Fraction of core members retained
Lineage duration	Number of consecutive windows in which a community persists
Internal collaboration persistence	Stability of internal coauthor ties
Topic persistence	Similarity of research foci over time
Structural cohesion	Internal edge density or conductance relative to outside ties

A stable community need not retain every member; the key question is whether its central members and collaboration structure persist.

5.2 Define the stable core

For each community, define a core using a combination of:

- High membership strength in that community
- Presence in multiple consecutive time windows
- Strong internal collaboration ties
- Repeated publication within the community
- Low dependence on a single one-off collaboration

For example, classify an author as a core member if they:

1. Have membership above a chosen threshold in the community,
2. Appear in at least rr consecutive or nearby windows, and
3. Have substantial weighted collaboration with other community members.

The analyst should be able to inspect and alter these thresholds.

5.3 Identify leaders using multiple roles

“Leader” should not mean only the most cited person. Report several roles:

Role	Possible indicators
Collaboration hub	Weighted degree, internal degree
Structural broker	Betweenness, cross-community ties
Persistent core leader	Long-term core membership and repeated coauthorship
Productive contributor	Papers within the community
Influential contributor	Citation impact, field-normalized by publication year
Topic leader	Strong association with a major topic or emerging topic

Show a profile for each leader, including coauthors, papers, topics, temporal role, and affiliations with multiple communities.

Citation counts should be normalized by publication age, since papers from 2023–2024 have had little time to accumulate citations.

6. Characterize each community’s research foci

6.1 Build a paper text representation

Use title, abstract, and keywords. Recommended preprocessing:

- Standardize keyword variants
- Preserve domain terms and acronyms
- Remove generic publication language
- Track terms whose meaning changes over time
- Retain conference/track labels as metadata, not as text features

Represent papers using both:

- Keyword and phrase features for interpretable labels
- Modern text embeddings for semantic similarity

6.2 Discover paper-level topics

Use a combination of:

- Embedding-based topic modeling, such as BERTopic
- Hierarchical clustering of paper embeddings
- Keyword co-occurrence analysis
- Manual expert labeling of top topic clusters

Potential high-level VIS topics may include interaction, graph visualization, scientific visualization, volume visualization, visual analytics, geospatial analysis, uncertainty, machine learning for visualization, evaluation, and storytelling. These labels should emerge from the data and be reviewed by domain experts rather than imposed in advance.

6.3 Aggregate topics to community–time profiles

Assign each paper’s topic contribution fractionally to its authors’ communities. A paper coauthored by people in multiple communities should contribute to each relevant community rather than being forced into one.

For every community and time window, report:

- Top topics
- Representative keywords and phrases
- Most representative papers
- Emerging topics
- Declining topics
- Topic diversity
- Topic similarity to other communities

6.4 Visualize topic evolution

Useful views include:

Visualization	Use
Community-topic heatmap	Compare research foci across communities and periods
Topic streamgraph	Show rise and decline of topics within a community

Visualization	Use
Temporal keyword map	Reveal emerging terminology
Paper embedding map	Inspect semantic clusters and outliers
Representative-paper list	Validate automatic topic labels
Topic similarity network	Identify intellectually related communities

Require analyst review of topic labels. Automatically generated labels should be treated as suggestions.

7. Analyze overlaps and boundary-spanning researchers

7.1 Measure community overlap

For each pair of communities, compute:

- Shared active members
- Shared core members
- Fractional membership overlap
- Cross-community coauthorship volume
- Topic similarity
- Shared papers or recurring author teams

Display this as a matrix and as an inter-community network.

7.2 Identify bridges

For each author, calculate:

- Number of active community memberships
- Membership entropy
- Cross-community coauthor tie strength
- Betweenness centrality
- Topics associated with each membership

High-overlap authors may be:

- Interdisciplinary contributors
- Collaborators linking InfoVis and VAST, for example

- Researchers transitioning between communities
- Members of a temporary cross-community project team

The analyst should inspect individual publication trajectories before assigning an interpretation.

8. Explain merges, splits, and dissolutions

For each candidate event, create an evidence panel containing:

1. Membership evidence

- Which authors moved, remained, or left?
- Were core members retained?

2. Collaboration evidence

- Did cross-group coauthorship increase before a merge?
- Did internal ties weaken before a split?

3. Topic evidence

- Did communities converge on similar topics?
- Did a new topic create a separate subgroup?

4. Publication evidence

- Which papers, author teams, or research programs drove the change?

5. Conference evidence

- Did the event coincide with shifts in VIS, InfoVis, VAST, or SciVis participation?

6. External contextual evidence

- Major methodological shifts, new datasets, emerging application domains, institutional changes, or field-wide trends.

The final “why” should be an expert interpretation supported by the evidence, not an automated causal claim.

9. Build an analyst-facing visual workspace

A practical dashboard should include linked views:

View	Main purpose
Temporal community map	Explore collaboration structure in a selected period

View	Main purpose
Community lineage/Sankey view	Track communities, merges, splits, births, and dissolution
Community profile panel	Inspect members, leaders, papers, topics, and metrics
Author trajectory view	Follow one researcher's memberships and collaborators over time
Topic heatmap and streamgraph	Compare and track research foci
Inter-community overlap matrix	Identify shared members and bridges
Paper embedding map	Inspect content-level structure
Parameter/sensitivity panel	Compare results across windows, thresholds, and algorithms
Event evidence panel	Examine candidate merge, split, or dissolution events

All views should support filtering by year, conference, topic, community size, author role, and citation threshold.

10. Validate and assess robustness

10.1 Compare alternative specifications

Repeat analyses across:

- Different time-window widths
- Different edge-weighting schemes
- Multiple community detection algorithms
- Different resolution parameters
- Inclusion/exclusion of weak or one-off ties

A robust community should recur across reasonable choices.

10.2 Use quantitative diagnostics

Assess:

- Community cohesion versus null models
- Temporal membership stability
- Agreement between algorithms
- Topic coherence
- Sensitivity of leaders and cores to parameter changes
- Whether apparent communities are simply conference-specific or truly collaborative groups

10.3 Conduct expert validation

Ask visualization researchers to review:

- Community labels
- Key leaders
- Representative papers
- Major claimed mergers/splits
- Interpretation of emerging topics

Record annotations and disagreements. These are valuable outputs, not merely corrections.

11. Recommended final deliverables

1. **Historical community atlas** for 1990–2024
2. **Community lineage graph** showing temporal evolution
3. **Community profiles** with cores, leaders, collaboration patterns, and research foci
4. **Author mobility and bridge analysis**
5. **Topic evolution report** for each major community
6. **Candidate merge/split/dissolution event catalog** with evidence
7. **Sensitivity and validation report** documenting which findings are robust
8. **Interactive visual analytics system** for expert exploration

The central principle is to treat communities as **dynamic, overlapping, and interpretable research formations**, not as fixed clusters produced once by an algorithm.