

Recommended workflow: Dynamic, overlapping co-authorship communities with topical characterization

This workflow treats the VIS literature as an evolving collaboration network, while preserving the fact that researchers may belong to several communities. It combines ideas from **NodeTrix**, **Reducing Snapshots to Points**, **Human-Steered Topic Modelling**, **SOMFlow**, and **StratomeX**.

1. Audit and prepare the publication data

First, validate the paper–author records:

- Use DBLP author identifiers as the primary identity key; retain raw names for display and manual checking.
- Identify duplicate papers, incomplete author lists, missing abstracts/keywords, and unusually large author teams.
- Inspect yearly publication counts by venue, because the four venues have different histories and may create apparent community changes caused by coverage rather than genuine intellectual change.
- Normalize titles, abstracts, and keywords: remove generic terms such as *visualization*, *visual analytics*, and venue names; retain domain terms and methodological phrases.
- If “citations” means citation links between papers, construct a paper-citation network. If it means only citation counts, use those as impact indicators rather than network links.

Create two linked data structures:

1. A **paper–author bipartite network**, preserving full author teams.
2. A **weighted co-authorship network**, where authors are nodes and a collaboration edge is created for each co-authored paper.

To avoid large author teams dominating the network, fractionally weight each paper’s collaboration contribution. For a paper with n authors, distribute a total contribution of 1 across its author pairs rather than assigning every pair a full weight.

2. Define temporal views at more than one scale

Construct overlapping temporal snapshots, for example:

- annual snapshots for detailed inspection;
- rolling 3- or 5-year windows for more stable community detection;
- broader eras, such as 1990–1999, 2000–2009, 2010–2019, and 2020–2024, for high-level reporting.

The analyst should interactively vary the window width. Short windows reveal rapid membership changes but may fragment communities; longer windows reveal durable structures but can conceal splits or temporary collaborations.

For each snapshot, calculate:

- weighted co-authorship strength;
 - publication count;
 - weighted degree and within-community degree;
 - betweenness or brokerage measures;
 - number of recurring collaborators;
 - venue composition;
 - citation impact, where available.
-

3. Discover candidate overlapping communities

Use an **overlapping community model**, not a hard partition. Suitable approaches include link-community detection, mixed-membership stochastic block models, or overlapping modularity methods. The important design choice is that each author receives a membership weight for one or more communities in each time window.

Run the detection at several resolutions and parameter settings. Do not treat one algorithmic result as definitive. Instead, compare candidate solutions according to:

- number and size distribution of communities;
- internal collaboration density;
- stability under bootstrap resampling of papers or edges;
- persistence of communities across adjacent time windows;
- interpretability of their topical profiles;
- plausibility according to visualization-domain knowledge.

The analyst should use a **community-resolution control** to merge overly fragmented clusters or examine subcommunities within large ones. This follows the human-guided refinement logic of **SOMFlow: Guided Exploratory Cluster Analysis**.

Report community number in two forms:

- the number of **active communities per time window**;
- the number of distinct **community lineages** across the full 1990–2024 period.

This distinction prevents temporary splits or renamed continuations from being counted as entirely new communities.

4. Identify cores, leaders, and peripheral members

For every community and time window, classify members by both structural role and temporal persistence.

A practical classification is:

- **Stable core:** authors with sustained, high membership in the community over multiple adjacent windows and repeated collaborations with other core members.
- **Leaders:** stable-core authors with high within-community collaboration strength, high publication contribution, or strong brokerage between the community and others.
- **Peripheral contributors:** authors with limited, short-lived, or low-strength membership.
- **Boundary spanners:** authors with meaningful membership in multiple communities or high cross-community collaboration.

Do not define leaders solely by publication count or citation count. Present a composite profile including:

- years active in the community;
- within-community weighted degree;
- number of repeated co-authors;
- cross-community brokerage;
- paper and citation impact;
- topical contribution.

The analyst should be able to inspect the underlying paper list for any proposed leader, especially when a high centrality score is driven by one large team paper or a short-lived collaboration.

5. Track community evolution and lineage events

Match communities between consecutive temporal snapshots using a combination of:

- weighted author-membership overlap;
- overlap of stable-core members;
- topical similarity;
- collaboration-structure similarity.

Use core overlap more heavily than peripheral overlap when deciding whether a community persists. A community whose peripheral membership changes substantially but whose core remains stable should normally be treated as a continuation.

Classify temporal events using explicit rules, subject to analyst review:

Event	Operational indication
Continuation	One community in period t strongly matches one in $t+1$
Growth / contraction	Continuation with substantial size change
Merge	Multiple prior communities substantially contribute to one later community
Split	One prior community substantially contributes to multiple later communities
Birth	A later community has no sufficient predecessor
Dissolution	A community has no sufficient successor
Temporary disappearance	No successor for a short gap, followed by a strong later match
Reorientation	Stable core persists but topical profile changes substantially

Visualize the result as a **community lineage diagram**: communities appear as bands over time, band width encodes active membership or publication volume, and links encode continuation, splitting, and merging. Membership overlap should be visible separately from lineage; otherwise, a few mobile authors can falsely imply a merger.

Use the dynamic-network perspective of **Reducing Snapshots to Points** as a complementary overview. Represent each temporal snapshot by a vector of network and community features—such as community-size distribution, core persistence, overlap level, modularity, and topic mixture—and project snapshots into a “network-state map.” This can reveal periods with unusually rapid restructuring, such as the emergence of new venues, interdisciplinary expansion, or changes in collaboration scale.

6. Characterize each community’s research foci

For every community–time-window pair, build a weighted document corpus from papers authored by its members.

When an author belongs to multiple communities, fractionally allocate the paper to each community according to that author’s membership weight. Retain a second view of full paper membership so the analyst can inspect interdisciplinary papers without forcing them into one topic.

Apply an initial topic model, preferably NMF or another interpretable topic method, to titles, abstracts, and keywords. Then use the interactive refinement strategy from **Human-Steered Topic Modelling**:

- merge redundant topics;
- split topics that mix distinct research areas;
- remove generic terms;
- seed missing domain concepts using key terms or selected papers;
- compare topic assignments against paper-level evidence.

For each community, show:

- its top topics and representative keywords;
- representative papers for each topic;
- topic prevalence over time;
- emerging, declining, and persistent topics;
- topic diversity and topical concentration;
- topics shared with other communities.

If citation links are available, add citation-flow summaries: which communities cite one another, which external fields they draw from, and whether a topical change coincides with new citation connections.

7. Provide coordinated visual analysis views

A research analyst should work across linked views rather than rely on a single network picture.

View	Main purpose
Community lineage view	Identify persistence, births, splits, merges, and dissolutions
Temporal membership matrix / author ribbons	Inspect individual affiliation histories and multi-community participation
Overlap view	Compare communities through shared members; show pairwise overlap and its temporal change
NodeTrix time-slice view	Inspect selected communities as matrices while retaining inter-community links, following NodeTrix
Core–periphery view	Compare leaders, stable cores, peripheral members, and brokers
Topic timeline / heatmap	Examine each community’s topical profile and topic drift
Paper and citation inspector	Validate patterns through titles, abstracts, keywords, authorship teams, and citations
Snapshot-state projection	Identify stable eras, recurrent structures, and anomalous periods, following Reducing Snapshots to Points
Provenance panel	Record community parameters, topic edits, event interpretations, and analyst annotations, following SOMFlow

For overlap analysis, the analyst should be able to select two or more communities and see:

- shared authors;
- authors with strong membership in both;
- shared papers;

- common topics;
- cross-community collaboration edges;
- whether overlap is stable, increasing, or limited to a particular period.

This follows the set-comparison reasoning used in **StratomeX**, adapted from patient-set overlap to author-community overlap.

8. Investigate major evolutionary events with evidence

When the lineage view identifies a split, merge, reorientation, or dissolution, open an event-focused comparison workspace:

1. Compare the predecessor and successor communities' stable cores.
2. Identify members who stayed, left, joined, or became boundary spanners.
3. Compare their collaboration matrices before and after the event.
4. Compare topic distributions, representative papers, and key terms.
5. Inspect venue shifts and, if available, citation flows.
6. Review a timeline of highly central papers and new collaborations around the event.
7. Record an analyst interpretation of the likely cause.

Potential interpretations should be framed as evidence-backed hypotheses, not automatic causal claims. For example:

- a merge may coincide with increased collaboration between previously separate methodological and application-oriented groups;
 - a split may reflect divergence into visual analytics versus domain-specific visualization work;
 - a topical reorientation may coincide with new data types, algorithms, venues, or influential papers;
 - a dissolution may be an artifact of sparse publication periods rather than the end of an intellectual community.
-

9. Iterate, validate, and document uncertainty

The analyst should repeat detection and tracking under reasonable alternatives:

- different rolling-window widths;
- different edge-weighting schemes;
- different community resolutions;
- inclusion versus exclusion of weak, one-off ties;

- alternative topic-model sizes and preprocessing choices.

Flag conclusions that are sensitive to these choices. In particular, distinguish:

- **robust communities**, visible across parameter settings;
- **tentative communities**, present only at a narrow resolution or in sparse periods;
- **uncertain lineage events**, where several predecessors or successors are similarly plausible.

Maintain an analysis history containing parameters, manually adjusted community boundaries, topic edits, event labels, and supporting paper-level evidence. This makes the resulting intellectual-history account reproducible and open to later revision.

Expected analytical outcomes

The workflow should yield:

1. A defensible estimate of the number of active communities and long-term community lineages.
2. For each community, a profile of its stable core, leaders, peripheral members, and boundary spanners.
3. A temporally resolved account of membership overlap and researcher mobility.
4. Topic profiles showing each community's major research foci and topical evolution.
5. A reviewed event history of community births, merges, splits, reorientations, and dissolutions.
6. Evidence-backed interpretations linking structural changes to papers, collaborations, topics, venues, and citation patterns.