🌻 Tribes. The most relevantly different subgroups in your data (by causal story)

The **Tribes** filter answers a very specific analysis question:

> **What are the most relevantly different subgroups in the data in terms of the causal stories they tell?**

In other words: if your sources contain *different narratives about how the system works*, Tribes tries to find those subgroups for you.

This aligns with the broader idea that causal mapping helps you make sense of **many causal claims from many sources** (not to build a single "true" causal model).

When is Tribes useful?

Use it when you want to:

- surface **distinct narrative patterns** without predefining groups
- compare "how the system works" across sources (not just what's frequent overall)
- identify disagreement or tension (e.g. different explanations of the same outcomes)

It's less useful when:

- you have very few sources (any clustering will be unstable)
- the sources mostly tell the same story (there may be no strong subgroups to find)

What Tribes clusters (conceptually)

Tribes clusters **sources**, not links.

Each source is summarised by the pattern of causal links it contains (cause→effect) and the sources are thus grouped into tribes with similar patterns.

(To be more precise, we create separate buckets for cause→effect bundles, plus sentiment buckets, then the **k-means** procedure groups sources that have similar patterns.)

The output is a new field in the links table: `custom_tribeId` (plus similarity diagnostics), so you can analyse the same map “through the lens of tribes”.

Controls (how to think about them)

Number of clusters (k)

This is “how many subgroups do you want to see?”.

Practical workflow:

1. start with k = 2–4
2. inspect the result (do the tribes look meaningfully different?)
3. adjust k up/down (too low merges stories; too high splits hairs)

Similarity cutoff + Drop unmatched

Each source has a similarity to its assigned tribe. These two controls work together. If drop unmatched is off, the similarity cut-off is meaningless.

- **Similarity cutoff**: minimum similarity for a source to count as a “good fit”
- **Drop unmatched**:
- ON: remove links from weakly-assigned sources (cleaner tribes, less coverage)
- OFF: keep them (more coverage, more mixing)

Min cluster %

Prevents the “1 big cluster + many tiny clusters” pattern.

Clusters below the threshold are discarded and their sources are reassigned to the nearest surviving cluster (subject to similarity + drop rules).

View Tribe Report: Sources vs Citations

The **View Tribe Report** button generates chi-square tables for *all* categorical fields by tribe.
You can choose what “counts” mean:

- **Sources** (default): each source contributes at most 1 to a cell (more robust if some participants produce many links)
- **Citations**: each link contributes 1 (more sensitive to “talkative” sources)

Using the Statistics (Pivot) tab to explain *why* tribes differ

Alternatively you can use the Pivot tab in the normal way to create cross-tabulations between your tribes and other fields.

Once you have `custom_tribeId`, the next question is usually:

> "Do these tribes line up with anything we already know about our sources (gender, region, program arm, interview round…)?"

A practical workflow:

1. run Tribes (so links have `custom_tribeId`)
2. open **Statistics / Pivot**
3. use the dataset that reflects your **current filtered links**
4. compare tribes against known characteristics (e.g. `custom_gender`, `custom_region`)

This helps you distinguish:

- tribes that are "purely story-based" (not explained by known demographics)
- tribes that correlate with known subgroups (e.g. region-specific causal mechanisms)

Toy example

Imagine 30 sources. Tribes with k=3 might reveal:

- Tribe A: "training → productivity → income"
- Tribe B: "prices → debt → stress"
- Tribe C: "weather → crops → migration"

Next steps:

- use Statistics/Pivot to see whether (say) Tribe C is concentrated in drought-prone regions
- use Custom Links Label (next post) to label edges by tribe composition and spot where the narratives diverge in the map

Related

- chapter intro