🌻 The factors table

22 Sep 2025

Summary

This extension builds a **factors table** from the (already filtered) **links table**.

- **Input**: Links table (one row per coded causal claim), already filtered by whatever link-level filters are active (date range, source subset, link-type, polarity, etc.).
- **Output**: Factors table (one row per factor label) with optional **in/out splits**, optional **group breakdown columns**, optional **normalised views**, and optional **significance tests**.

The key idea is: **everything starts from links**, and the factors table is just an aggregation of link-derived "factor mentions".

What exactly is a "factor mention"?

Each link row typically contains at least:

- a **cause label** and an **effect label**
- a `source_id` (and source metadata available for grouping)

From each link row we derive **two factor-mention records**:

- one mention for the cause label (direction = `out` / "as cause")
- one mention for the effect label (direction = `in` / "as effect")

These mention records are the atomic units that the factors table aggregates.

How it works (links → factor mentions → factors table)

Step 0. Start from the already-filtered links table

All upstream filters apply first. This filter does **not** try to re-implement link filtering; it only consumes the current link set.

© Causal Map Ltd 2026 · causalmap.app · CC BY-NC 4.0

Step 1. Expand each link row into factor-mention records

For each link row r:

- emit mention
  $m_{out} = (factor = r.cause, direction = out, source_id = r.source_id, link_id = r.link_id)$
- emit mention
  $m_{in} = (factor = r.effect, direction = in, source_id = r.source_id, link_id = r.link_id)$

Notes:

- This is why "citation totals across factors" are not "number of links": each link yields (at least) **two mentions**.
- Self-loops can either yield 2 mentions with the same label, or be deduplicated; whichever rule is chosen must be stated consistently in the UI.

Step 2. Apply factor label transforms (rewrite layer)

Before aggregating, apply the current label-rewrite transforms to `m.factor`, such as:

- **Collapse factor labels** (bucket near-synonyms by search term)
- **Exclude bracket text** (e.g. `Education (primary)` → `Education`)

These are **temporary rewrites** for analysis/presentation; they do not change the underlying coding.

Step 3. Aggregate into the base factors table

Group mention records by factor label f. Compute one or more base columns (examples):

- **Citation Count (total)**: number of mention records for f
- **Citation Count (out / in)**: split by `direction`
- **Source Count (total)**: number of distinct sources that mention f at least once
- **Source Count (out / in)**: distinct-source counts split by `direction`

Derived role measures (examples):

- **Outcomeness**: a summary of "effect-like vs cause-like" usage (e.g. based on in/out counts)

The factors table is therefore an **interpretation layer**: it encodes choices about (a) rewrite rules, (b) evidence unit (citations vs sources), and (c) whether direction matters.

© Causal Map Ltd 2026 · causalmap.app · CC BY-NC 4.0

Groups: breakdown columns

Let G be a group variable defined on sources (e.g. `district`, `gender`, `section`). This filter can add **group breakdown columns** by aggregating mention records jointly by $(factor, group)$.

Step 4. Join source metadata to mentions (for grouping)

Join each mention record with its source attributes so each mention has $G(m)$.

Step 5. Add the group breakdown columns

For each factor f and each group level g, compute cells such as:

- **Citations mode**: You can't use 'macro parameter character #' in math mode
- **Sources mode**: You can't use 'macro parameter character #' in math mode

Optional: also split each cell by direction (`in` / `out`) if the UI supports it.

What group columns are for

They let you ask:

- “Which factors are disproportionately mentioned by group A vs B?”
- “Which outcomes differ by district / section?”

Totals and normalisations (what totals actually mean)

Because the factors table is built from factor mentions:

- **Citation totals across factors** are totals of **mentions**, not totals of **links**.
- **Source totals across factors** are totals of **source–factor incidences**, not “number of sources”.

Normalisation is a choice of baseline, not a cosmetic option:

- **Within-group totals** (sum over factors within a group column) measure that group’s overall mention volume under current filters.
- **Percent-of-baseline view** (within each group) shows *relative prominence*:

The little equation below is just intuition (you can ignore it if you want).

\[\text{share}(f,g) = \frac{\text{cell}(f,g)}{\sum_{f'} \text{cell}(f',g)} \]

This is useful when groups differ in overall verbosity or number of sources.

© Causal Map Ltd 2026 · causalmap.app · CC BY-NC 4.0

Optional inference: significance testing per factor (single grouping variable)

If exactly one group variable G is selected, this filter can compute a per-factor test asking whether mentions for factor f are distributed across group levels differently than expected given group baselines.

Intuition (chi-squared style):

> Even if group A has more mentions overall than group B, is factor f still *over-represented* in one group relative to that baseline?

For ordered groupings (e.g. age bands), an ordinal/trend framing can be more appropriate than treating levels as unordered categories.

Why this is useful

This helps you move from “what is mentioned most?” to “what differs by context/group?”.

- **Everything** (factor counts, roles, group comparisons, tests) is derived from the same link-derived mention records.
- Every result depends on the current link filters and label rewrite transforms.

© Causal Map Ltd 2026 · causalmap.app · CC BY-NC 4.0