**Proximity measures**

Script **Network Mapping**:

- **Edges** tab, choose « proximity measure »
- And / Or tab **« Network Analysis and layout »** when « **Add information from a 3rd variable to tag clusters or produce a heatmap** » is activated (https://docs.cortext.net/analysis-mapping-heterogeneous-networks/mapping/#tagging-heatmap-specificity-measure)

| proximity measures | type of network | normalisation | special properties |
|---|---|---|---|
| raw | interaction network (*e.g.* social network) | no | - |
| $X^2$ | homogeneous & heterogeneous | yes | normalization tend to create links toward higher degree nodes |
| MI | homogeneous & heterogeneous | yes | Inspired from information theory |
| Cramer | homogeneous & heterogeneous | yes | - |
| cosine | homogeneous network (eg. semantic) | yes | Classical measure (originating from scientometrics) |
| distributional | homogeneous network (eg. semantic) | yes | very robust measure (coming from computational linguistics) |
| cosine_het | affiliation network (eg. users sharing the same hashtags ) | yes | two fields are required but the final network is homogeneous |
| dot_product_het | affiliation network (eg. users sharing the same hashtags ) | no | two fields are required but the final network is homogeneous |

## Heterogenous networks

| | | | |
|---|---|---|---|
| cosine_het | affiliation network (eg. users sharing the same hashtags ) | yes | two fields are required but the final network is homogeneous |
| dot_product_het | affiliation network (eg. users sharing the same hashtags ) | no | two fields are required but the final network is homogeneous |

- *https://docs.cortext.net/metrics-definitions/#heterogeneous-dot-product*
- *https://docs.cortext.net/metrics-definitions/#heterogeneous-cosine* (see below)

## For homogenous and heterogenous networks

**Raw** corresponds to the raw value (count, frequency). Raw co-occurrence measure. For example, we will count 1 for the pair {carrots, leeks} each time carrots and leeks appear together in a recipe. Relevant measure for the construction of collaboration networks (no particular correction of the information is necessary; respect of the data). Is based on the assumption that a link corresponds to an effective interaction.

- Consider it for social networks, collaboration networks (collaborations between individuals or between organizations)

- To go further: *https://docs.cortext.net/metrics-definitions/#raw*

**Chi²**: measures the intensity of the link between two terms by assessing the deviation from the expected value. The Chi² is therefore a measure of specificity. The expected value between two terms is equal to the sum of all the co-occurrences of the first term (with, therefore, all the other terms) multiplied by the sum of all the co-occurrences of the second term (with, therefore, all the other terms), over the sum of all the co-occurrences between them (sum of the rows plus the sum of the columns of the co-occurrence matrix, that is to say, in fact, the total number of observed co-occurrences). When it is positive, the difference between the actual value of the co-occurrences of two terms and the expected value indicates an over-representation of the link between these two terms and therefore a **specificity**.

- The "**Add information from a 3rd variable to tag clusters or produce a heatmap**" tab is generally used to identify the specificity of a value in a context (for example with the "**Network Analysis and layout**" tab)
- reference : *https://docs.cortext.net/metrics-definitions/#chi2*

## Homogenous networks

**Distributional** : stemming from recent linguistics, it allows to show relevant relations, although rare. Distributional proximity is based on the direct measurement of Mutual Information presented above. For a given word, the Mutual Information is the amount of information brought by the presence of this word in the context of the appearance of another word. The Distributional measure, for two terms (i and j), compares the n-dimensional vectors of Mutual Information of these two terms, in other words the similarity of the contexts of appearance of these terms. This allows the detection of synonyms, i.e. terms that do not necessarily co-occur but have identical contexts of occurrence. It has therefore a property of interchangeability: carrot and porreau being vegetables, they have common characteristics, possibilities of association with other close ingredients as well as similar cooking methods.

- Useful for semantic analysis: very efficient for extracting the underlying structure of texts, by presenting words that play similar functions in the texts
- reference: *https://docs.cortext.net/metrics-definitions/#distributional*

**Cosine**: the cosine similarity measure was introduced by Salton (Gerard Salton & McGill, 1983) with the idea that the similarity between two documents can be measured by comparing the two vectors (G Salton, Wong, & Yang, 1975) formed by the list of terms and the frequencies of these terms for these documents. This measure is commonly used in text mining. It is also popular in scientometrics, especially in the analysis of co-citations (Eck & Waltman, 2009; Hamers et al., 1989). Applied to a graph, the cosine similarity measure aims at comparing the co-occurrence patterns of two nodes (word, citation, author...) in a set of documents. For two terms i and j , their similarity is based on the two rows of the terms of the initial co-occurrence matrix, i.e. on the set of co-occurrences of these two terms. The co-occurrence profiles of the two terms are treated as n-dimensional vectors (where n is the total number of terms with which they co-occur) whose angle is measured. The smaller the angle,

the more similar the profiles are (the closer the two lists of terms with which they co-occur, as well as the associated frequencies, are).

- The advantage of this measure is that it does not favor frequent terms (important nodes) over rare terms. Thus, two terms can have a high proximity even if they co-occur little together or little with other terms.
- Reference: *https://docs.cortext.net/metrics-definitions/#cosine*