MetGem Software for the Generation of **Molecular Networks** Based on the **t-SNE** Algorithm
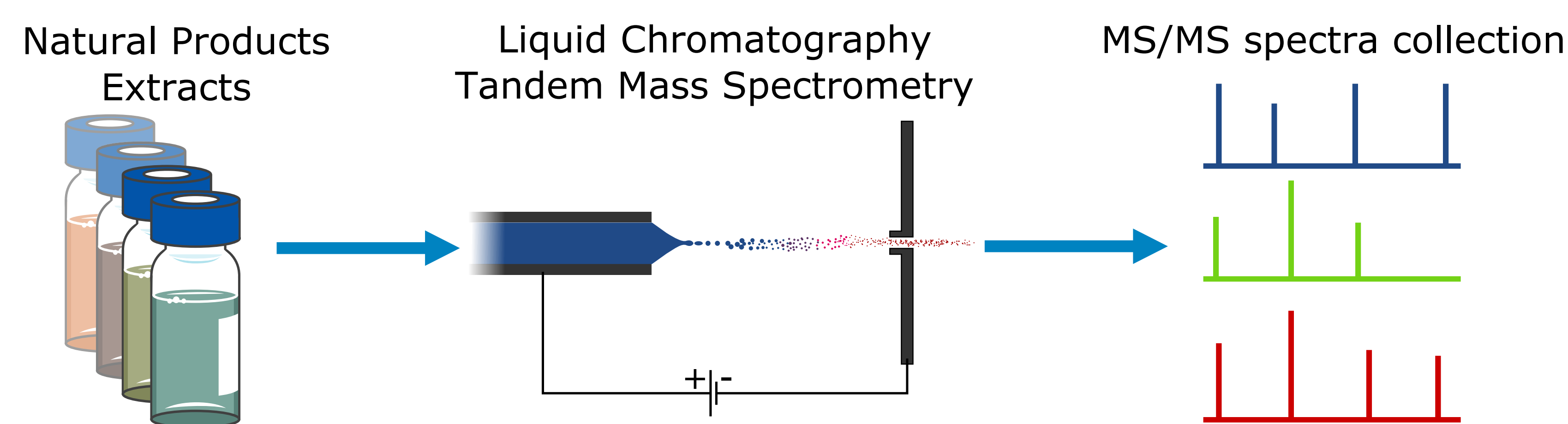
Nicolas Elie, Florent Olivon, Gwendal Grelier, Fanny Roussi, Marc Litaudon and David Touboul

Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France.

https://metgem.github.io

Download it from here

ICSN

CNRS

Molecular Networks

Natural Products Extracts

Liquid Chromatography Tandem Mass Spectrometry

MS/MS spectra collection

Spectra are compared by pairs
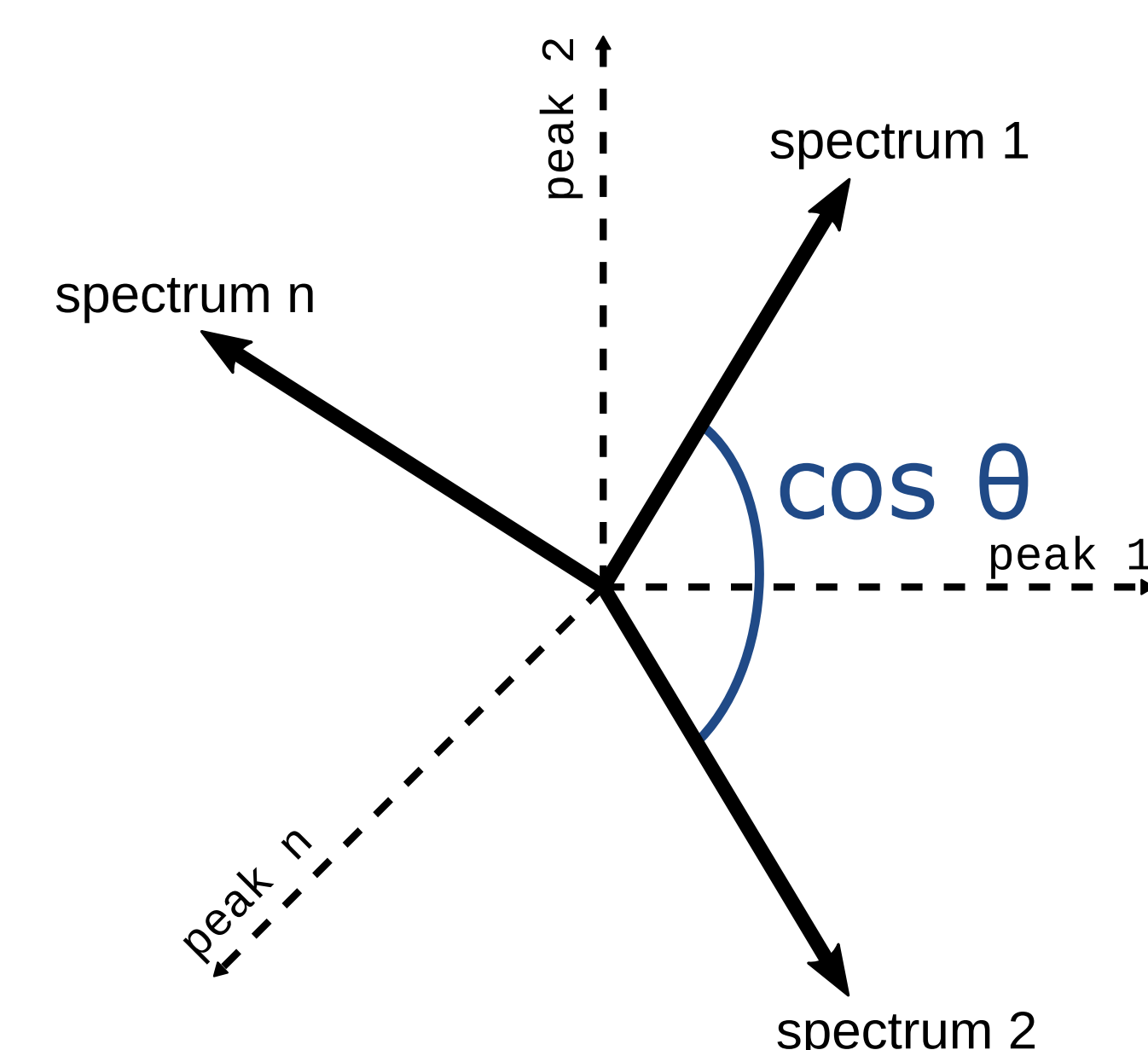
Nodes repulse each other like charged particles

Edges act as spring attracting nodes with stiffness proportional to cosine scores

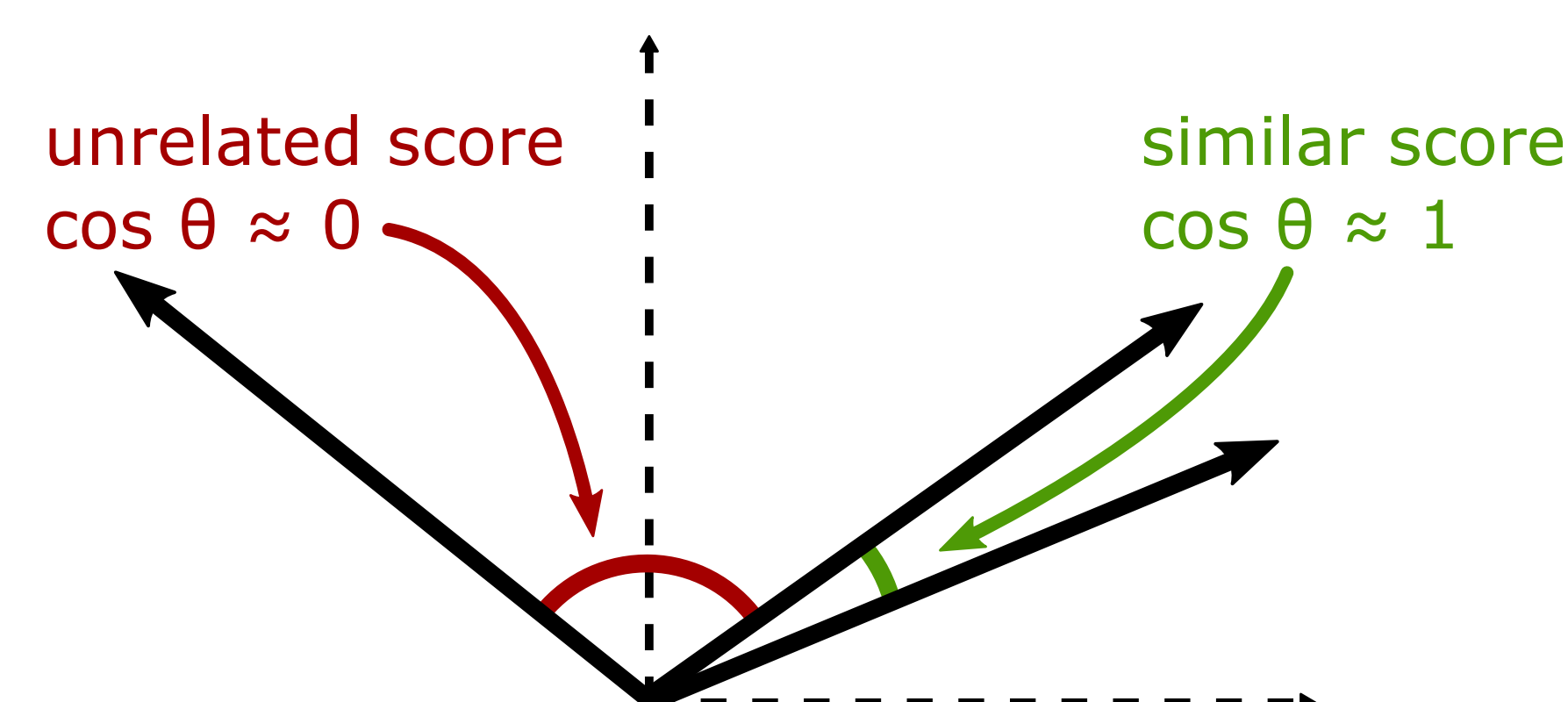Because of cosine threshold and topK algorithm, some links are cut, leading to information lost

GNPS

M. Wang *et al.*, *Nature Biotechnol.*, **2016**, *34*, 828-837

Cosine scores

peak 2

spectrum 1

spectrum n

cos θ

peak 1

peak n

spectrum 2

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2}\sqrt{\sum\limits_{i=1}^{n} B_i^2}}$$

unrelated score
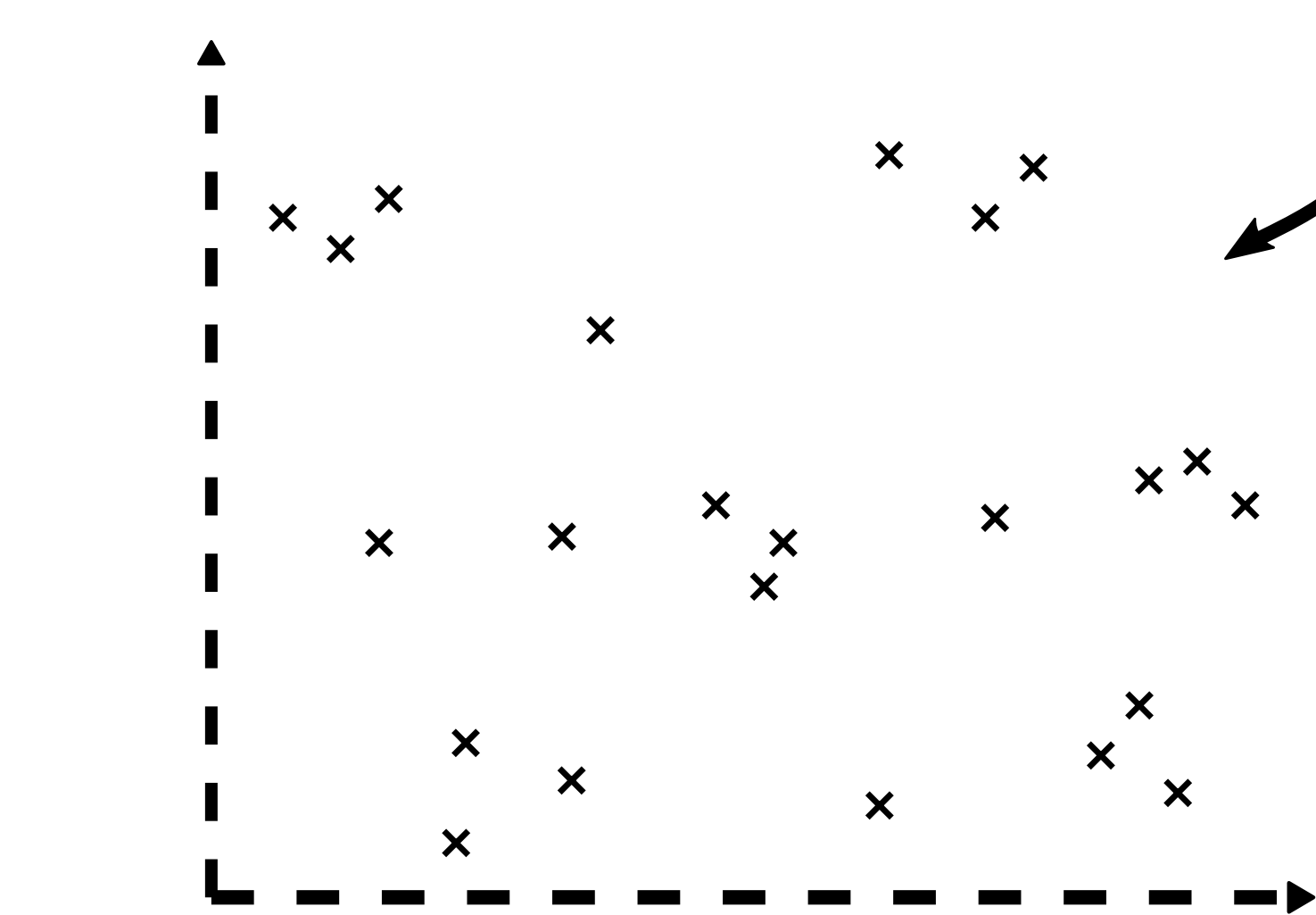cos θ ≈ 0

similar score
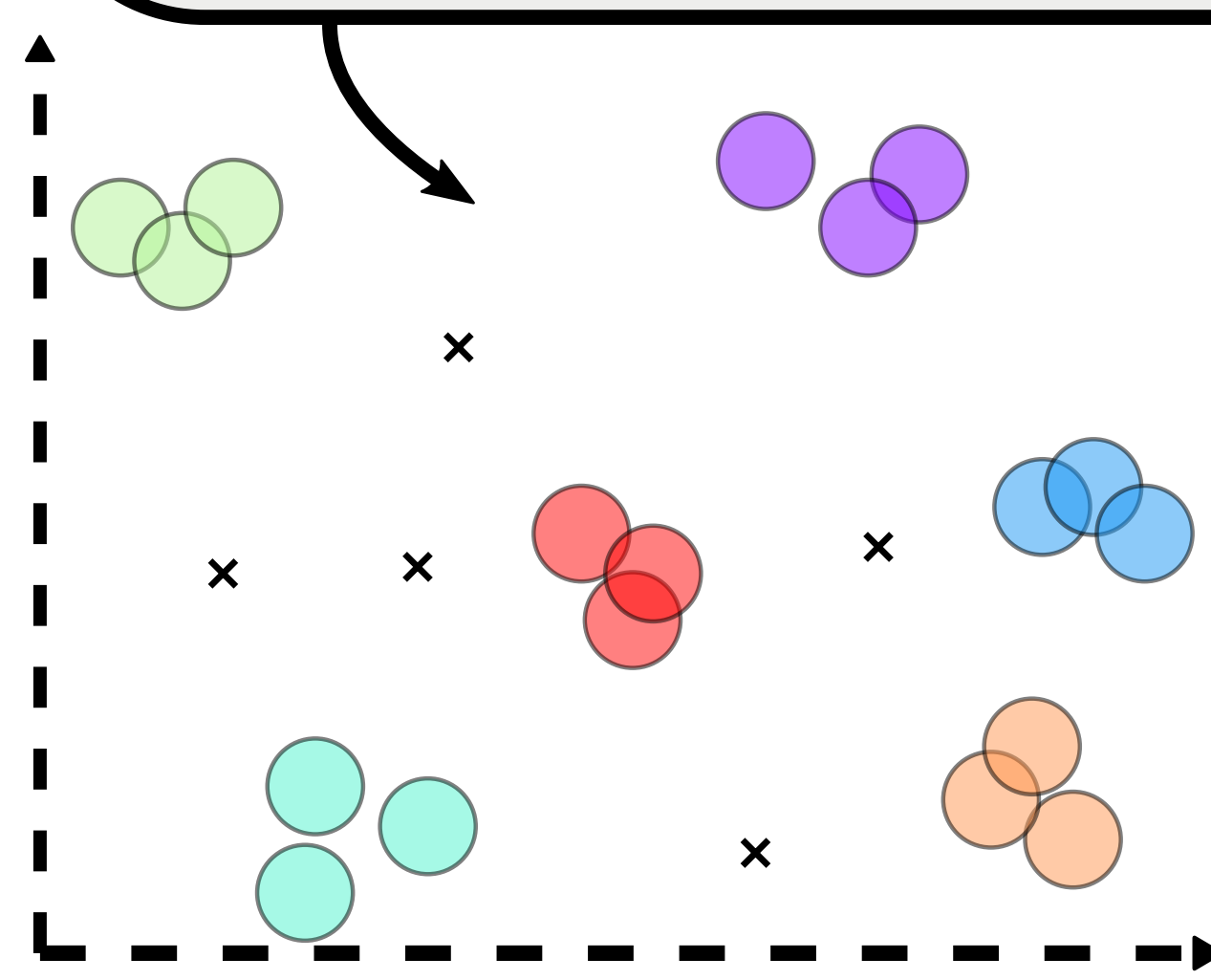cos θ ≈ 1

---

t-SNE

Non-linear projection of a *n*-dimensional space into a 2D space with priority to local distances over global distances.

How to represent these data in a space of lower dimension?

We are lucky, our data is 2D so we can plot it but real world data is usually way larger...
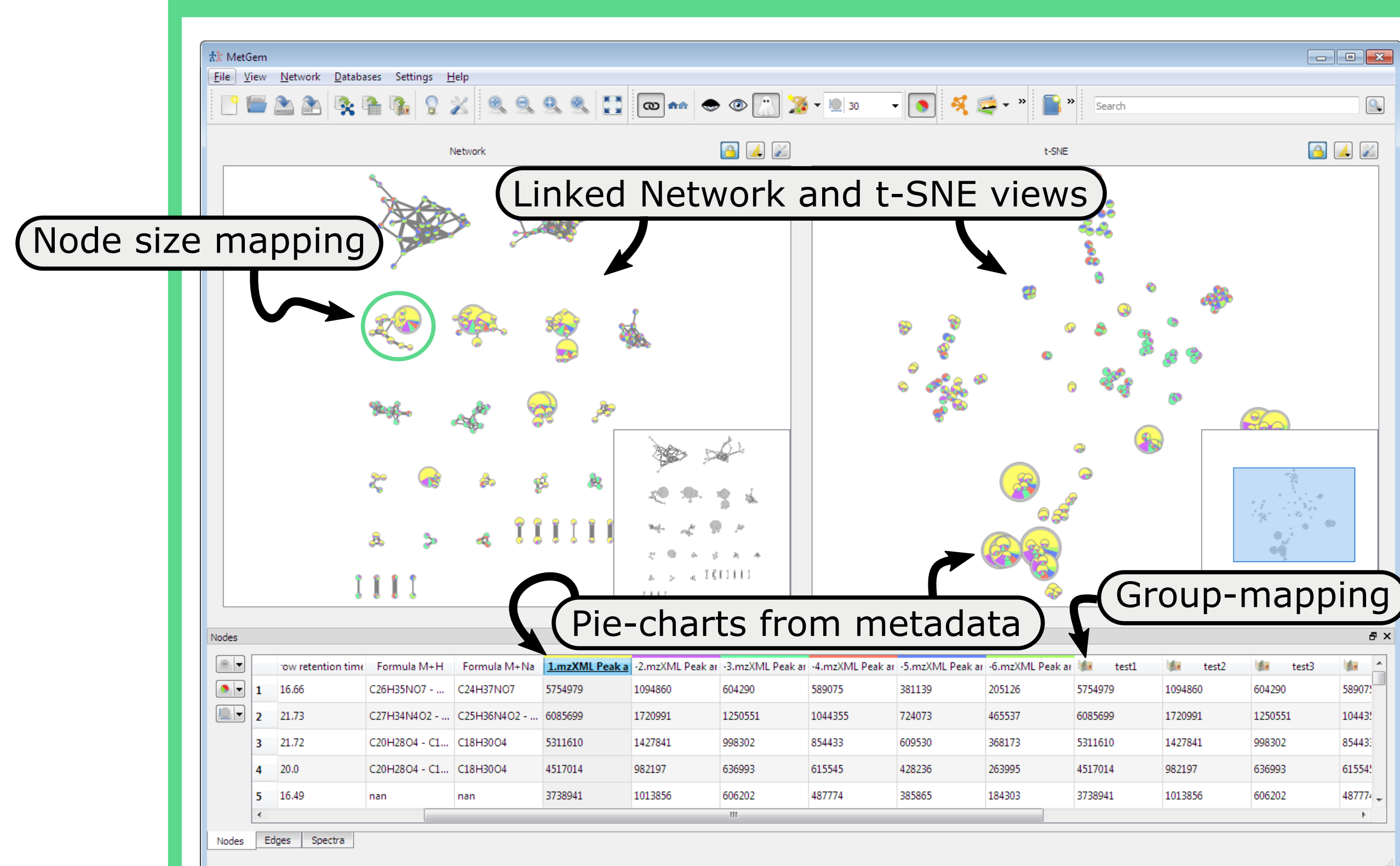
PCA

t-SNE

The common solution give bad results

t-SNE is doing better. It can even get improved if we exclude isolated points

Intercluster distances are not meaningfull!

L. Van der Maaten *et al.*, *J. of M. Learning Research*, **2008**, *9*, 2579-2605

StatQuest: t-SNE, Clearly Explained: youtu.be/NEaUSP4YerM

Key features

Node size mapping

Linked Network and t-SNE views

Pie-charts from metadata

Group-mapping

✓ Quickly and easily tune parameters

✓ Fast generation of molecular networks

✓ No internet connection required

✓ Read data from: MGF files exported by **MZmine 2**

MSP files (**NIST** file format)

✓ Use databases from different sources

MS-Dial   GNPS   ISDB

✓ Plugins can be easily created to add new databases

✓ One-click import of user databases

✓ Export both views to cytoscape

✓ Cross-platform   macOS

Free software: contributions, suggestions are welcome!

GPL v3 Free Software — Free as in Freedom

✉ nicolas.elie@cnrs.fr, david.touboul@cnrs.fr

🌐 https://github.com/metgem/metgem

Source code repository

MetGem is powered by: Qt OpenMP SQLite ARROW