



Neutral molecular networks: Polarity-independent tool for mass spectrometry data

Nicola Dalla Valle^{a,b,c}, Pietro Franceschi^{b,*} , Mar Garcia-Aloy^b , Peter Robatscher^c, Sabine Oettl^c , Michael Oberhuber^c

^a University of Trento, Trento, TN, 38100, Italy

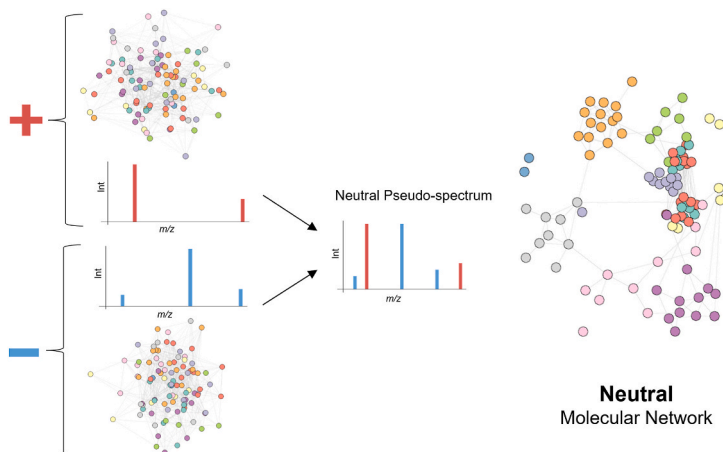
^b Research Innovation Centre, Fondazione Edmund Mach, San Michele a/A, TN, 38098, Italy

^c Laimburg Research Centre, Laimburg 6 – Pfatten (Vadena), Auer (Ora), BZ, 39040, Italy

HIGHLIGHTS

- NMN unify positive and negative MS/MS spectra.
- Unique polarity-independent workflow for untargeted data.
- NMN improves clustering reliability and annotation.
- NMN favoured the discovery of novel toxin derivatives.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Untargeted metabolomics
Molecular network
Alternaria
Mass spectrometry
Electrospray ionization

ABSTRACT

Untargeted mass spectrometry-based metabolomics generates large-scale fragmentation data, typically analyzed separately in positive and negative ionization modes. The fragmentation patterns of the same molecule usually capture distinct but complementary structural information across polarities. Evaluating them simultaneously, rather than in separate molecular networks, can enhance the overall informativeness and speed the analysis. In this paper, we introduce the Neutral Molecular Network (NMN) concept, a novel strategy that unifies complementary fragmentation patterns from both polarities into a single "neutral pseudo-spectra". NMNs outperformed polarity-specific networks on several large-scale publicly available MS/MS libraries in terms of chemical reliability of spectral matches and clustering capability. The improved performances were further confirmed through the analysis of a biological untargeted dataset of *Alternaria* fungal extracts, where this approach facilitated the identification of previously uncharacterized toxin derivatives. NMN offers a polarity-independent

* Corresponding author.

E-mail address: pietro.franceschi@fmach.it (P. Franceschi).

<https://doi.org/10.1016/j.aca.2026.345708>

Received 4 March 2026; Received in revised form 21 May 2026; Accepted 23 May 2026

Available online 23 May 2026

0003-2670/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

framework for the structural interpretation of untargeted MS/MS data. This method can improve the efficiency of metabolite annotation pipelines and can be applicable to diverse metabolomics workflows.

1. Introduction

Untargeted mass spectrometry-based metabolomics aims to provide the chemical profile of biological systems, offering quantitative insights on the chemical bases of biological complexity by profiling hundreds or even thousands of small molecules [1–3]. However, translating the vast and complex datasets generated by mass spectrometry into biological knowledge remains a challenge [4]. Data-dependent acquisition (DDA) offers a valuable approach to enhance compound characterisation by fragmenting detected ions to generate MS/MS spectra. These spectra play a central role in compound identification, which continues to be the most demanding step in untargeted metabolomics [5,6]. To navigate this complexity, Molecular Network (MN) has emerged as a powerful computational strategy that organizes spectral data into maps where nodes represent molecules and edges connect structurally similar compounds. This framework enables researchers to explore chemical relationships across complex datasets and facilitates the annotation of unknowns through their proximity to known molecules [7,8]. In practice, molecular networks are constructed by calculating similarities between MS/MS spectra. One of the most used similarity score calculations is the one implemented by the Global Natural Products Social Molecular Networking (GNPS), which computes spectral similarity based on both shared m/z values and neutral losses [9,10], enhancing annotation of structurally similar molecules [11] that are expected to have similar functions.

In a comprehensive untargeted metabolomic analysis, the entire workflow is performed separately for positive and negative ionization modes [12], creating two disconnected networks from the same biological samples [13]. Because molecular fragmentation depends on polarity [14], identical compounds may occupy different positions and cluster memberships in the two networks, introducing redundancy and polarity-related bias [15,16]. While this separation complicates MN interpretation and often requires extensive manual inspection [17,18], the resulting fragmentation patterns are frequently complementary, each concerning distinct structural regions of the molecule. Assessing both ionization modes together provides a more complete and chemically informative view; at worst, the information overlaps, reinforcing structural interpretation. Although complementary in principle, positive and negative ionization data are rarely integrated effectively. Previous integrative strategies, such as multi-layer networks combining molecular and biochemical data [19,20], can partly mitigate single-polarity biases by expanding contextual coverage, but they do not directly integrate information across ionization modes and remain complex and not suited to untargeted analyses [21].

Here, we present the Neutral Molecular Network (NMN) concept, a molecular network constructed from neutral pseudo-spectra that merges the fragmentation spectra from both positive and negative ionization modes into a single representation. By integrating complementary fragmentation profiles, NMNs reduce the need for subjective polarity selection while increasing chemical informativeness. This holistic representation of molecular structure promises to contain ambiguities in compound annotation and reveal chemical relationships previously obscured by polarity-dependent biases. After providing an overview of neutral pseudo-spectra, we evaluated NMNs against polarity-specific networks using publicly available spectral libraries and subsequently validated the approach on a biological case study of *Alternaria alternata* extracts, where this application enabled the identification of previously unreported toxin derivatives.

2. Materials and methods

2.1. Publicly available libraries data

Fragmentation spectra from GNPS, HCD Natural Product Library (HCD-NPL), and Vaniya/Fiehn Natural Products Library (VF-NPL) were downloaded from the MassBank of North America (<https://mona.fiehnlab.ucdavis.edu/downloads>, 20 April 2025). Only compounds with both $[M+H]^+$ and $[M-H]^-$ spectra and chemical classes with ≥ 10 members were retained. Spectra with >30 ions were cleaned according to Dalla Valle et al. [22]; spectra with ≤ 30 ions were considered pre-cleaned. Data were organised by instrument type (Orbitrap or ToF), resulting in a total of 5 datasets (Table S1). Between multiple spectra of the same compound within a dataset, the lowest entropy spectrum was used [23]. Results for the largest dataset considered (VF-NPL-Qex) are shown in the main text, while results for the remaining datasets are presented in the Supplementary Information.

2.2. *Alternaria* sample preparation and data acquisition

Thirty *Alternaria alternata* strains (Table S2) were analyzed with LC–MS using a reversed-phase water/acetonitrile gradient with 0.1% formic acid. Data-dependent acquisition was performed for all samples. QC samples were injected every 10 biological samples, with six additional concentrations of pooled QC in triplicate (Table S3). Thirty *Alternaria alternata* strains were grown on potato dextrose agar (PDA) plates for 4 days at 25 °C (8 h light/16 h dark). Three biological replicates per strain were cultured in 50 mL Czapek-Dox Broth +0.5% yeast extract for 14 days, respectively. Culture supernatants were filtered (22–25 μ m) and stored at –80 °C. For each analysis, 2.5 μ L of the *Alternaria* extract was injected into the LC–MS system (Orbitrap Exploris 240, Thermo Fisher Scientific, Bremen, Germany). Chromatographic separation was performed on an ACQUITY UPLC HSS T3 column (2.1 $\times$ 150 mm, 1.8 μ m particle size; Waters, Milford, MA, USA). Chromatographic separation was carried out at a flow rate of 0.4 mL/min. The mobile phase consisted of water with 0.1% formic acid (phase A) and acetonitrile with 0.1% formic acid (phase B). The gradient started at 5% B and was held for the first minute, followed by a linear increase to 100% B over the interval 1–10 min. The composition was maintained at 100% B until 13.5 min, then returned to 5% B between 13.5 and 14.5 min, and finally held at this condition until 19 min. The instrument was operated in DDA mode, scanning an m/z range of 90–1350 Da. Stepped normalized collision energies of 20, 40, and 60 were applied. The MS1 resolution was set to 60,000, and the MS2 resolution to 30,000. The time window between consecutive MS1 acquisitions was 0.7 s. To avoid redundant fragmentation, a dynamic exclusion of 5 s was applied for ions already fragmented twice. A prefilter threshold of 50,000 was used. For each strain, three biological replicates were analyzed. The sample injection order was randomized to minimize analytical bias. A blank and two quality control (QCP5) samples were injected after every 10 biological samples.

Feature extraction and representative MS/MS selection were performed with XCMS [24], and feature linearity across QC dilution series was assessed using a linear model.

2.3. *Alternaria* compound identification

Compounds were first matched by exact mass against a predefined candidate list, initially derived from the *Alternaria* metabolite catalogue of Lou et al. (2013) [25] and subsequently expanded using additional literature. Identification levels were assigned as: Level 2 – match with

MS/MS spectra from publicly available libraries (spectral similarity >0.7 with matching precursors m/z) or putatively annotated via experimental MS/MS comparison with reference spectra [26,27]; Level 3 – supported by in silico fragmentation (MetFrag, Compound Discoverer 3.3) when no reference spectra were available. A total of 157 features were identified with the following confidence partitioning: 2: 101; 3: 56. Putative identifications are listed in Table S4. Chemical subclasses were assigned to identified compounds using the ClassyFire taxonomy according to their InChIKey, previously extracted via MS/MS annotation and manually refined. For features lacking identification, chemical classes were predicted using the SIRIUS software [28], where for each feature the chemical subclass with the highest probability considering ClassyFire lineage was considered.

2.4. Neutral pseudo-spectra creation

Positive and negative MS/MS spectra corresponding to the same molecule were merged using the Spectra R package [29]. Before merging, the intensity values of each spectrum were normalized to the maximum peak intensity (set to 100) and m/z were adjusted according to polarity (difference of 1.00728 Da) to align both spectra to the neutral mass. Duplicate ions were averaged within a 0.0005 m/z tolerance. Precursor m/z and retention time values were averaged across polarities.

For MS/MS libraries, neutral spectra were generated by merging pre-selected $[M + H]^+$ and $[M - H]^-$ ions of the same compound. The resulting library sizes are reported in Table S1. For *Alternaria* samples, complementary $[M + H]^+$ and $[M - H]^-$ were merged following the ion identity molecular network (IIN) approach [30]. After peak picking and feature extraction (xcms package), only features with at least one MS/MS spectra and a retention time between 0.5 and 10 min were selected. Chromatographic groups were created according to retention time tolerance (5 s), correlation across samples (calculated on $\log_2$ -transformed area values) and peak-shape correlation, both using a threshold of 0.7 (MsFeatures package). Within each chromatographic group adduct species were identified by considering the most common adducts (positive ESI: $[M+H]^+$, $[M+Na]^+$, $[M + NH_4]^+$, $[2M + H]^+$, $[2M + Na]^+$, $[2M + K]^+$, $[M+K]^+$; negative ESI: $[M - H]^-$, $[M+Cl]^-$, $[M+Br]^-$, $[M + HCOO]^-$, $[M + CH_3COO]^-$, $[M-2H + Na]^-$, $[2M - H]^-$, $[2M + Cl]^-$, $[2M + HCOO]^-$), neutral losses (H_2O , 2^*H_2O , 3^*H_2O , CO , $COOH$, CO_2), and +1, +2 and +3 isotopes. For groups containing only one feature, or groups where none of the described adduct species was explicitly identified, all features were treated as potential $[M+H]^+/[M - H]^-$. Only identified $[M+H]^+/[M - H]^-$ species were considered for neutral mass calculation, applying the appropriate theoretical mass offset (e.g., subtracting or adding 1.00728 Da). Cross-polarity pairs were then matched considering the neutral mass (± 5 ppm) and retention time (± 5 s). Only pairs with a correlation (calculated on $\log_2$ -transformed area values) across samples of both polarities above 0.7 were considered as species generated from the same chemical entity. In case of duplicated matches, the one with the strongest correlation was selected.

Importantly, single-polarity MS/MS denoising is a crucial step in creating neutral pseudo-spectra [22]. Without this process, merging spectra from both polarities would result in an overcrowding of noisy peaks, increasing the likelihood of random ion matches and potentially compromising the reliability of MS/MS comparisons.

2.5. Code and computational efficiency

The script for the creation of the 544 neutral pseudo-spectra for the *Alternaria* biological dataset took 18 h, using a standard performance PC. This includes chromatographic group formation, adducts recognition and neutral pseudo-spectra creation. The analysis started from 3093 to 4969 features from negative and positive XCMS data processing, respectively, in line with the expected size of a relatively complex untargeted dataset. Almost the entirety of the time was necessary for

chromatographic group creation. Full code for neutral pseudo-spectra and related MN creation is available at: https://github.com/NicolaDalla/Neutral_Molecular_Network.

2.6. MNs creation and topology evaluation

Only the subset of compounds/features successfully merged into the NMN was used to reconstruct different modes of MNs, maintaining an identical node set across all comparisons. Modularity [31] and assortativity [32] metrics were used for MNs structure comparison.

Fragmentation similarity matrices were computed using MsCoreUtils [29] R package (tolerance = 0.01), considering absolute m/z and neutral loss matches (GNPS similarity score). MNs were generated and analyzed with iGraph [33]. Network topology was evaluated with 25 resampling iterations on 80% stratified subsets. Edge thresholds ranged 0–1 (step 0.025). Local modularity (LQ) was calculated as described by Clauset et al. [34]. Node membership was defined by compound chemical classes. For the *Alternaria* dataset, only chemical subclass assignments that remained consistent across both positive and negative ionization modes were considered valid for subsequent modularity analysis.

2.7. Chemical structural comparison

Molecular similarity was assessed using Tanimoto and Overlay scores via fmcSR [35]. SDF inputs were generated from SMILES using the smiles2sdf function from ChemmineR [36].

2.8. Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Google Gemini to improve grammar and spelling. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

3. Results and discussion

To evaluate the advantages of a complementary dual-polarity architecture, we benchmarked the NMN workflow against six diverse datasets. First, five public spectral libraries were sourced from the MassBank of North America (MoNA), including the GNPS library, the HCD Natural Product Library (HCD-NPL), and the Vaniya/Fiehn Natural Products Library (VF-NPL), which were further separated by instrument type (ToF vs. Orbitrap) to ensure the concept's robustness across different platforms. For these benchmark datasets, defined chemical classes and subclasses were used as the ground truth for measuring network performance. To validate the concept in real-world complexity, we generated an untargeted LC-HRMS dataset from *Alternaria alternata* fungal extracts, a species known for its diverse and complex secondary metabolome [25–27].

Since neutral-pseudo spectra have the aim to merge complementary chemical information from positive and negative ionization modes, it is essential to align ion species from both polarities. In this study, we focused on $[M+H]^+$ and $[M - H]^-$ pairs. This selection was made for several reasons: these ions are almost ubiquitous in untargeted metabolomics and they are generally transversal across diverse chemical classes. Moreover, alternative adducts may exhibit variable fragmentation behaviors which reflect adduct-dependency rather than the chemical structure [30]. To accurately isolate these specific pairs within untargeted datasets, we started from the IIN approach [30]. First, features with an MS/MS spectra were merged into chromatographic groups based on retention time similarity, chromatographic and intensity correlations. Within each group, various adducts, isotopes and in-source fragments were recognized. From these annotated groups, only the $[M+H]^+$ and $[M - H]^-$ species were considered for inter-polarity pairing to generate neutral pseudo-spectra creation. Limitations related to

inter-polarity pairing are discussed in section 3.4.

3.1. Neutral pseudo-spectra

The core of the NMN architecture lies in the synthesis of neutral pseudo-spectra, an approach that merges fragmentation information from both positive and negative ionization modes into a single representation, capturing complementary structural insights as demonstrated in Fig. 1. The plot shows the neutral pseudo-spectrum of folic acid: the negative-mode spectrum reveals small, specific fragment ions corresponding to individual subunits, while the positive-mode spectrum retains larger, less fragmented ions reflecting the overall molecular structure. Ions representing the dihydropteridine moiety appear in both polarities, suggesting it is a core structural motif consistently shown across ionization modes. Because the spectra are intensity-normalized before merging, the polarity that produces a richer fragmentation pattern naturally contributes more ions and carries greater relative weight in the final pseudo-spectrum. This combined approach enhances chemical informativity, offering a balanced fragmentation profile that can improve structural annotation and compound identification.

To assess whether neutral pseudo-spectra yield more chemically informative matches, we randomly selected 50 compounds and matched their MS/MS spectra against their original public spectra library with the self-spectrum excluded, across positive, negative and neutral modes. In library searches, strong library matches are expected to be between similar compounds, so only strong spectral similarity scores (score ≥ 0.8) were considered. For each query compound, we evaluated the chemical similarity between the library match and the query using Tanimoto and Overlay scores, which quantify overall structural overlap and substructure alignment, respectively. Results for the VF-NPL-Qex library are shown in Fig. 2. Neutral pseudo-spectra provide significantly more reliable matches, with higher Tanimoto and Overlay scores, reflecting greater overall chemical similarity. Similar trends were observed across most of the considered libraries (Fig. S1), where neutral pseudo-spectra either outperformed both single polarities or were comparable to the best-performing one. As can be expected, considering all spectral matches regardless of their spectral similarity score, the chemical similarity increases with stronger spectral matches. Notably, this effect is particularly pronounced for neutral pseudo-spectra (Fig. S2), highlighting the enhanced confidence of pseudo-spectra matches with a high

similarity score.

3.2. Public libraries molecular networks

While neutral pseudo-spectra yields more informative matches, demonstrating superior network quality requires quantitative metrics. We employed assortativity and modularity metrics [31,32] to evaluate how well the networks group related compounds. Assortativity measures the tendency of nodes to connect within the same category, while modularity quantifies the strength of separation between distinct communities. For both metrics, higher values indicate superior clustering performance. Using chemical class and subclass metadata as node labels, we computed both metrics across all five public datasets by incrementally increasing the spectral similarity threshold from 0 to 1, retaining only edges exceeding the cutoff.

Fig. 3 shows results for the VF-NPL-Qex dataset at subclass-level. In all cases, modularity and assortativity of the networks increase as the similarity threshold increases. This is in agreement with the expected presence of stronger connections between structurally related compounds. Interestingly, the increase was larger for the networks built from neutral pseudo-spectra. In particular, NMNs turned out to be more structurally informative in the range of commonly used thresholds for MNs analysis (i.e., 0.6–0.8). This indicates that neutral pseudo-spectra preferentially capture chemically meaningful edges, especially when considering relatively “strong” connections. Going towards higher similarity thresholds, assortativity and modularity of the NMN showed a clear decrease. A likely explanation is that since neutral pseudo-spectra integrate more fragment ions from both polarities, similarity values close to 1 are rare in NMN. This leads to fewer edges for the higher part of similarity thresholds (Fig. S3), which at its extreme, can affect assortativity and modularity, although the remaining connections are expected to be more structurally meaningful. These trends were consistent across all datasets and for both class- and subclass-based analyses (Figs. S4 and S5).

The visual impact of an improvement in the network structure in terms of modularity and assortativity from single polarity to NMN is illustrated in Fig. 3B, which shows MNs from the VF-NPL-Qex dataset at a similarity threshold of 0.7. In the figure, the chemical subclasses flavonoid glycosides and carbohydrates and their conjugates are highlighted. The NMN exhibits a clearer and more interpretable organization

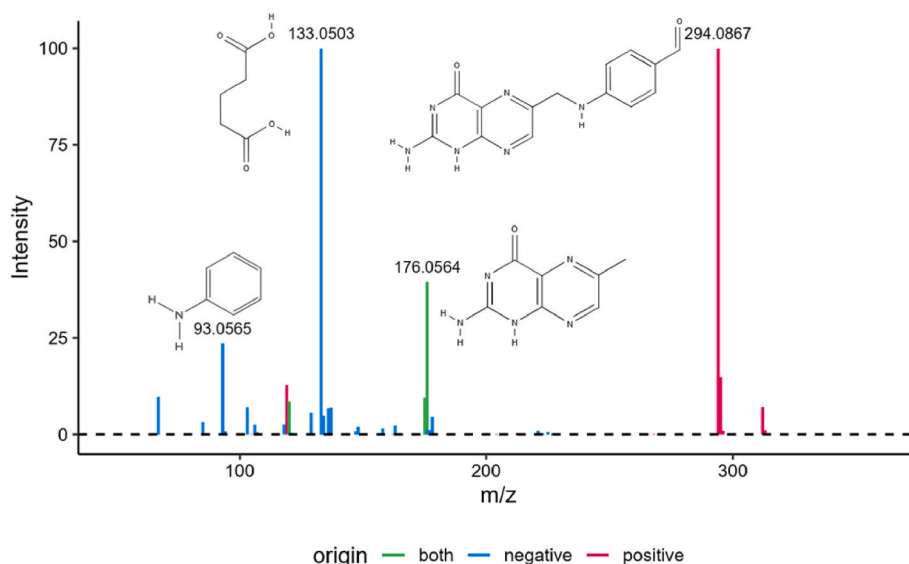


Fig. 1. Neutral pseudo-spectrum of folic acid from the VF-NPL-Qex dataset (positive spectrum ID: VF-NPL-QEHF027649, negative spectrum ID: VF-NPL-QEHF027654). Fragments from negative ionization mode are shown in blue, those from positive mode in red and fragments shared in both modes in green. The molecular substructures corresponding to the most intense fragments are annotated through MetFrag website. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

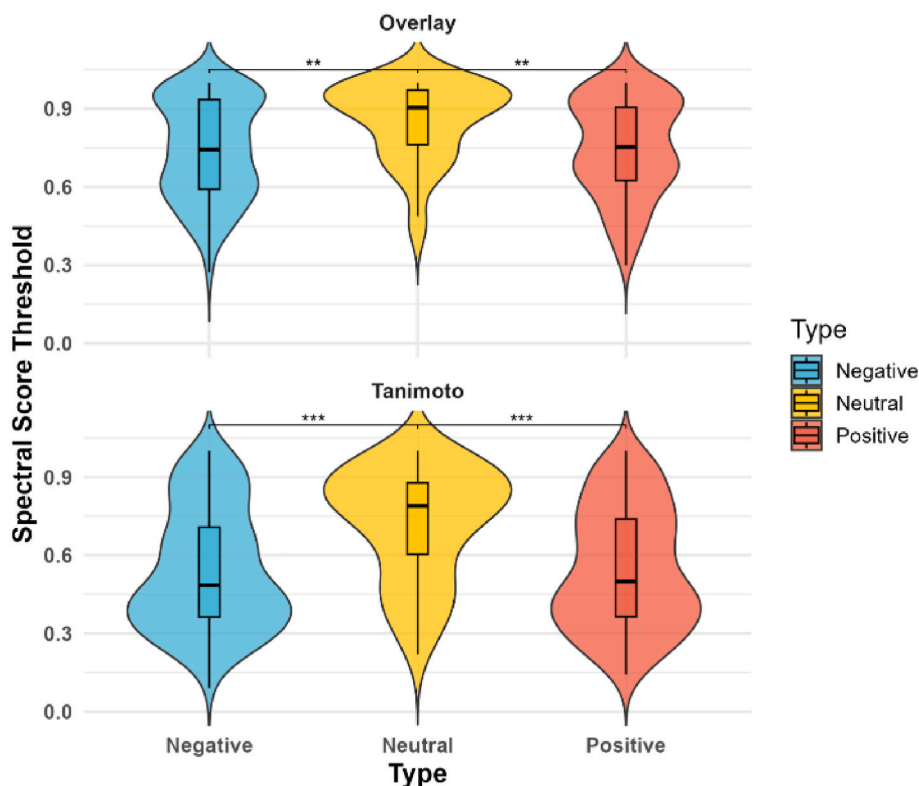


Fig. 2. Tanimoto and Overlay chemical similarity scores of 50 randomly selected compounds for strong spectral matches (compounds considered with a spectral similarity match equal or higher than 0.8). Mann-Whitney p-value: ** <0.01, *** <0.001.

than the single-polarity MNs, showing better subclass separation and fewer, but more meaningful, connections (Fig. 3B). Among the single polarity MNs, the negative-mode network shows the poorest organization, with extensive overlap and crowding of subclasses, leading to reduced interpretability. This figure provides a visual example of the higher assortativity and modularity described above, in agreement with Fig. 3A.

To compare network structural quality independently of similarity thresholds, we constructed MNs for positive, negative, and neutral spectra at matched assortativity values. NMNs achieved comparable assortativity requiring a lower similarity threshold (Table S5, Fig. S6; Supplementary Note 1), maintaining the informativity of the best-performing single-polarity MN while preserving overall network structure. This indicates that the lower edge count in NMNs, generally observed across datasets (Fig. S3), results from stricter filtering rather than a loss of structural information, retaining only the most chemically meaningful connections.

Observed assortativity and modularity trends were further supported by subclass-prediction analyses, which investigate the network's ability to predict chemical subclass annotations for unknown compounds. In these analyses, each node was treated as 'unknown' and its subclass was predicted from the annotations of its nearest neighbors, ranked by spectral similarity. Across all datasets, neutral pseudo-spectra matched or exceeded the prediction performance of single-polarity networks, particularly when only the strongest spectral matches were considered (see Supplementary Note 2 and Fig. S7).

Considered network metrics (i.e., assortativity and modularity), as previously mentioned, are edge-based metrics, so they do not take into consideration the community size. To completely understand the general architecture of the molecular networks, singletons and communities size of different MNs were considered for all datasets. Results are presented in Fig. S8, and they clearly indicate that the NMN approach effectively addresses the "hairball" problem by decreasing the node density of the largest community. In addition, as far as the singletons are

concerned, an increase in the absolute number of singletons is generally observed in NMN. This is not unexpected due to the higher specificity required for cross-polarity connection. Interestingly, however, this increase does not result in a disproportionate sparse network since the proportion of singletons relative to the total number of communities remains comparable to that of single-polarity MNs because the NMN simultaneously generates a higher overall number of distinct communities. This suggests a topological balance where the NMN achieves a more refined separation of nodes, avoiding the overcrowding of the biggest communities without over-isolating individual nodes.

3.3. *Alternaria* extracts molecular networks

To evaluate the performance of NMN in a real-world bioanalytical scenario, this workflow was applied to an untargeted metabolomics dataset of 30 *Alternaria* fungal strains (Fig. 4A, Fig. S9), in which we were able to identify 544 feature pairs used for neutral pseudo-spectra creation. Given the absence of a comprehensive ground truth for the *Alternaria* metabolome, we constructed a validation framework based on chemical subclass consistency, where subclass definitions were derived through a two-layer identification strategy (spectral library matching and in-silico structural annotation) supplemented by SIRIUS chemical subclass predictions for previously unannotated features. Beyond global modularity metrics, we also assessed local modularity for the most prevalent chemical subclasses in order to quantify the clustering density of nodes relative to their first neighbors, where higher values indicate superior group cohesiveness and better-defined structural clustering. To assess the intrinsic variability of the network structure, we constructed 25 independent networks by randomly selecting different MS/MS spectra among the ones available for each feature. This estimate of the variability is expressed as standard deviation in the plots presented in Fig. 4B. The plots clearly demonstrate that across the commonly utilized similarity thresholds (0.6–0.8), the NMN achieves global modularity performance that is equal to, or significantly superior to, the best-

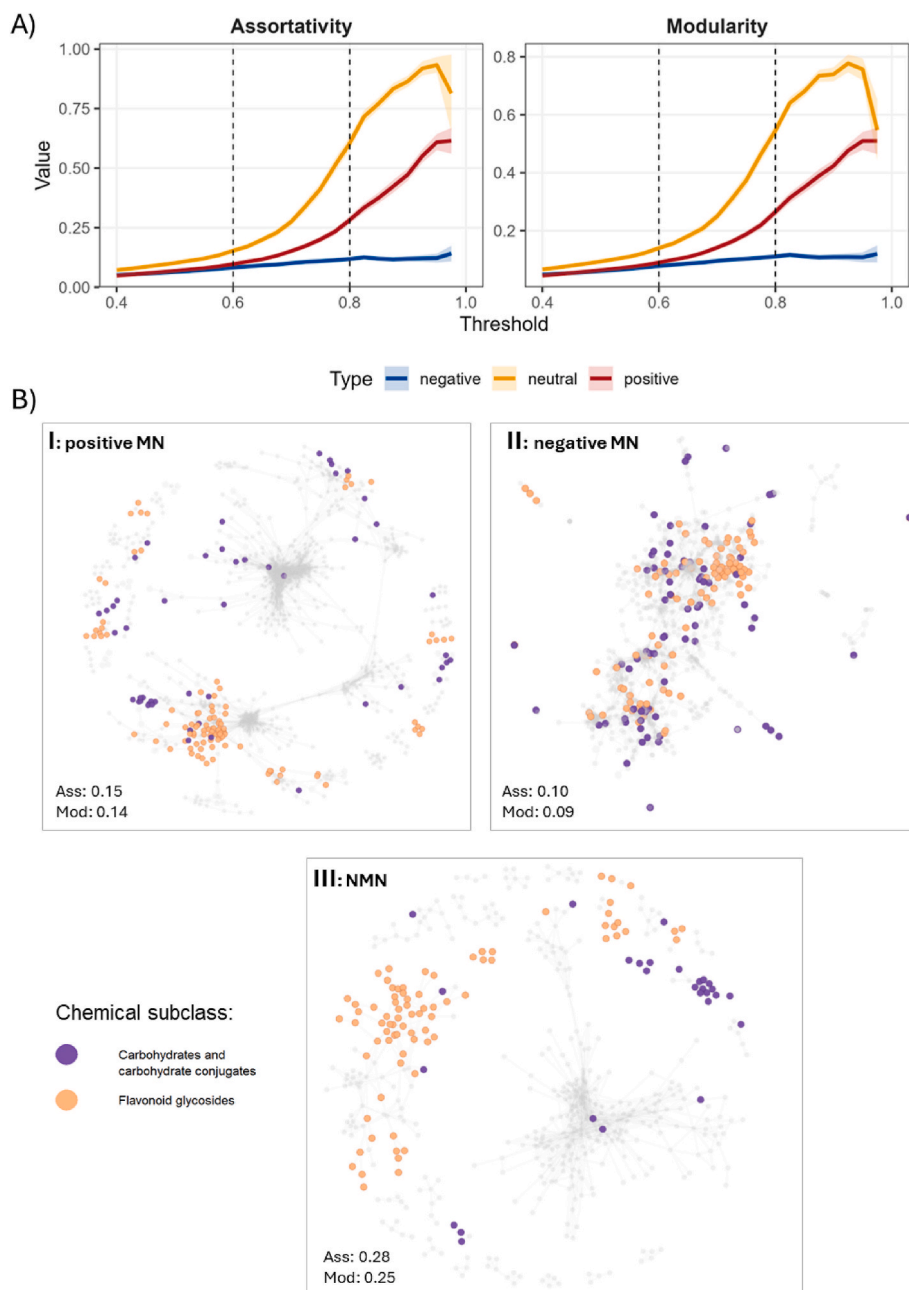


Fig. 3. Results from analysis of VF-NPL-Qex dataset across the three MNs. A) Modularity and assortativity values based on all chemical subclasses. Wider lines represent the average values calculated from 25 stratified resampling iterations, with the corresponding standard deviation shown as colored low-opacity areas. Results from the NMN are shown in yellow, while blue and red indicate negative and positive ionization modes, respectively. Vertical dashed lines are relative to the commonly used MN similarity score interval threshold (i.e., 0.6–0.8). B) MNs for similarity threshold of 0.7 built using: I) positive MS/MS spectra, II) negative MS/MS spectra and III) neutral pseudo-spectra. “Flavonoid glycosides” and “Carbohydrates and their conjugates” subclasses are highlighted. For better visual interpretation, only subgroups with more than 3 nodes are shown. Global assortativity and modularity values for the entire networks are present in the bottom-left corner. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

performing single-polarity network (Fig. 4B I), in line with what is observed in public libraries analysis. Regarding local modularity metric, results for the three most prevalent chemical subclasses are shown in Fig. 4B II. The NMN provided significantly enhanced grouping for amino acids and peptides compared to both positive and negative ion modes. For other chemical subclasses such as Sesquiterpenoids and Terpene lactones, the results were comparable to the best-performing single polarity, which in this case was the positive ESI. This is consistent with the ionization chemistry of these metabolites. For terpenoids structure, the neutral pseudo-spectra are generally primarily influenced by the positive ionization mode, where fragmentation is more informative,

representing the majority of m/z signals. For instance, for Tricycloalternarene 2b (Fig. S10), over 70% of the fragment ions in its neutral pseudo-spectra originate from the positive mode. Consequently, the positive polarity acts as the primary driver of similarity scores within the NMN for these specific compounds.

3.4. Limitations and methodological considerations

It is important to point out that our current feature-pairing protocol, by focusing on neutral pseudo-spectra created by merging $[M+H]^+$ / $[M-H]^-$ species, represents a trade-off between metabolic coverage and

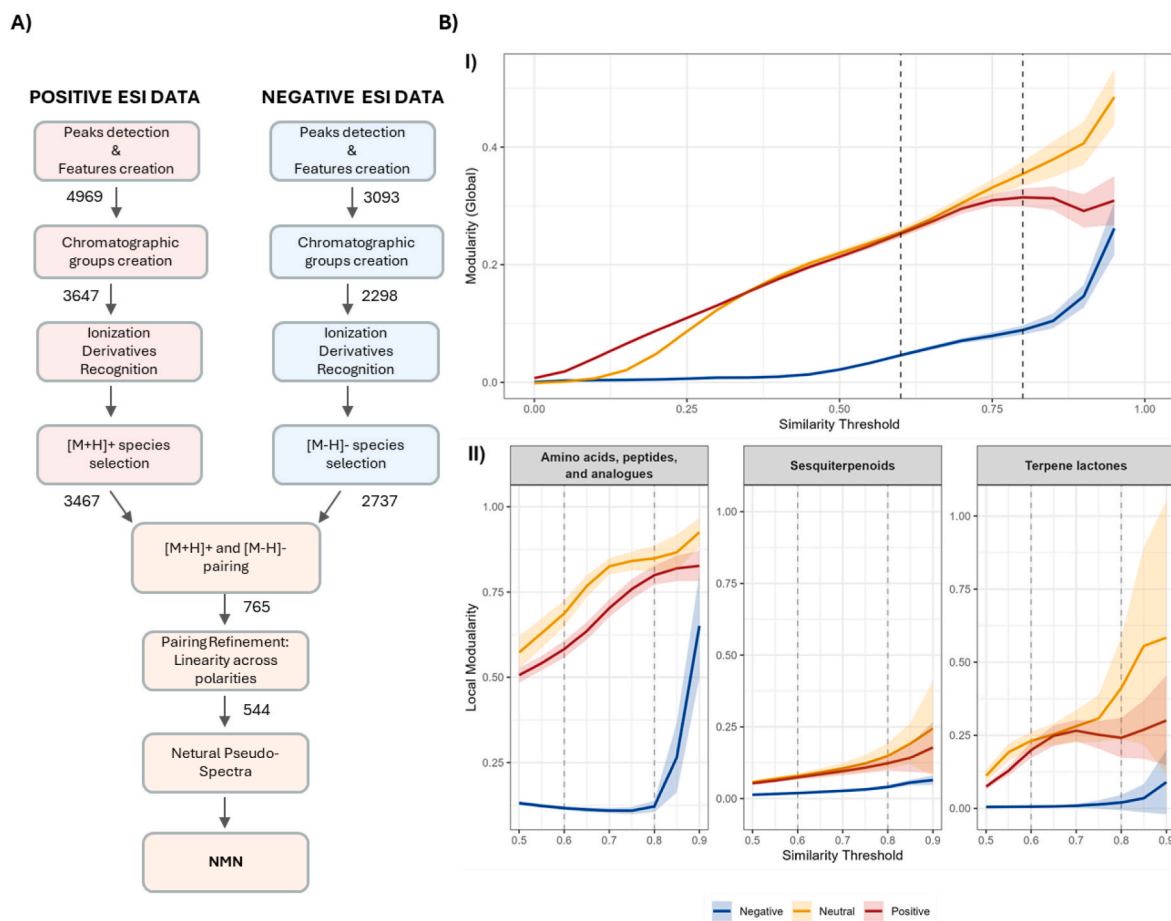


Fig. 4. Neutral Molecular Networking analysis for the *Alternaria* untargeted dataset. A) Workflow used for the analysis of an untargeted dataset. The number of retained features/pairs after each step is shown. B) Comparative evaluation of network grouping performance based on chemical subclass labelling: I) Results for modularity calculated across the entire network as a function of the similarity threshold; II) Local modularity scores through different similarity thresholds for the three most prevalent chemical subclasses. Variability, visualized as shaded areas, is expressed as standard deviation and has been calculated with 25 MNs repetition randomly selecting different MS/MS spectra of the same feature.

network structural accuracy. Features corresponding to other common adducts, in-source fragments, or to molecules that ionize in only one polarity are excluded from the NMN construction. To a certain extent, their elimination can reduce redundant signals in the final network, indeed, feature selection following the pairing process for neutral pseudo-spectra creation have a positive influence in modularity and assortativity metrics of single polarity MNs (Supplementary Note 3, Fig. S11), but a certain level of information loss is inherent to this choice.

The primary limitation of the proposed workflow is related to chemical entities that ionize exclusively as alternative adducts or that are detected only in a single polarity. The latter will be inevitably excluded in this framework, while for the former the pairing protocol can be tailored to merge specific alternative adducts if a study requires a targeted focus on chemical classes, which are detectable under specific adducts (e.g., $[M+Na]^+$; $[M+NH_4]^+$). In this study, we specifically focused on $[M+H]^+$ and $[M-H]^-$ pairs because they are the most frequently present and transversal across the diverse chemical classes commonly found in complex untargeted datasets but the NMN idea can be applied to different scenarios where the two ionization modes can be linked. As a general comment we would like to remark that many of the considerations which are valid for NMN, regarding adducts and ionization, are also valid in classical MNs. Alternative adducts often fail to connect with parent ions due to divergent fragmentation behaviours [30], and chemical classes can be biased by the specific ionization efficiencies of the considered polarity [37]. The pairing protocol, with the discussed related limitation, can lead to a loss of annotation coverage

compared to the single polarities. For instance, in the *Alternaria* dataset, the pairing protocol excluded 21 of the 157 features initially annotated across the separate polarities.

Another aspect that it is important to emphasise is that neutral pseudo-spectra are not developed to maximize the exact spectral matching scores against reference libraries. A neutral pseudo-spectrum indeed merges fragments from both ionization modes, so the resulting similarity score against a reference typically reflects a compromise of the positive and negative performances. Within the single spectrum, the results are largely driven by whichever polarity generates the highest number of informative fragment ions. Consequently, in direct library matching scenarios, the optimal single-polarity spectrum will frequently yield an absolute superior similarity score compared to the merged neutral spectrum (Supplementary Note 4, Fig. S12). This limitation could potentially be addressed by modifying the standard cosine similarity metric implemented within the GNPS ecosystem, which may not be perfectly calibrated for neutral pseudo-spectra. This improvement goes beyond the scope of our research and may represent an interesting direction for future developments.

The fundamental aim of NMN is to capture the widest possible chemical informativity by integrating complementary structural fragments. For this reason, its application is particularly suitable in the MN context, as we have shown in the case of the *Alternaria* dataset. Considering the identified features, the neutral pseudo-spectra, despite showing a lower number of connections, provide a more robust and reliable foundation for MN analysis in clustering compounds that are

chemically similar (i.e., belonging to the same chemical subclass, Fig. S13), confirming the general results of modularity investigation. In addition, the global distribution of $[M+H]^+$ and $[M-H]^-$ features across the m/z and RT range (as shown in Fig. S14) confirms that the neutral-pseudo spectra maintain a coverage across the majority of the chromatographic run, likely capturing pairs across a wide range of different chemical polarities. As a further check of the reliability of the retained features, we evaluated their signal linearity across a dilution series, an established indicator of genuine metabolite signals [38]. Most NMN features showed strong linearity (92% with $R^2 > 0.7$, Table S7), confirming that they likely represent real metabolites rather than noise signals. Moreover, the overall NMN workflow demonstrates a limited false discovery rate for both inter-polarity feature pairing protocol and network connectivity (Supplementary Note 5).

For all the mentioned reasons, NMN can be used as a valuable tool for a faster and more robust initial annotation and interpretation stage, where NMN provides the primary, structurally coherent MN which can then be supplemented with the exploration of remaining single-polarity features to maximize compound discovery.

3.5. Identification of unknown toxin derivatives

The NMN approach facilitated a comprehensive understanding of the *Alternaria* metabolome. Annotation propagation using the NMN enabled the sequential (tentative) identification of previously unknown derivatives. The anchor of the annotation propagation was a known *Alternaria* toxin (LT = 3,6-dihydroxy-8-oxo-9-eremophilene-12-oic acid), for which 3 isomers were detected (Fig. S15).

Detailed neutral pseudo-spectral analysis of an unknown compound (blue node in Fig. 5A; C₁₇H₂₄O₆; M: 324.1571 Da) alongside the known LT-toxin enabled its identification as an O-acetylated LT-toxin derivative (7'-O-Ac-LT). This discovery was facilitated by the synergistic contributions of both ionization polarities. The positive ionization mode (red ions in Fig. 5B) provided essential structural context, revealing shared core scaffold fragments that linked the unknown node to the LT-toxin family. However, on its own, the positive mode lacked the diagnostic resolution to understand the specific modification. Conversely, beyond providing an additional match (NL 62 Da), the negative ionization mode revealed the essential functional fingerprint, contributing critical evidence of the acetoxy moiety ($-O-CO-CH_3$) through a diagnostic fragment at m/z 60.0211 (CH₃COOH). While these negative fragments were modification-specific, they lacked the scaffold data required to connect the molecule back to the LT-toxin family in a negative MN (involved

subregion comparison in Fig. S16). By merging information from both polarities, the NMN created a unified pseudo-spectrum where the positive-mode scaffold and negative-mode acetylation were evaluated together.

Further structural elucidation focused on a node connected to the 7'-O-Ac-LT, with a neutral mass of 306.1466 Da (purple in Fig. 5), representing a mass difference relative to a water moiety. This feature represents an independent compound. According to the neutral pseudo-spectral mirror plot (Fig. S17), it was identified as the dehydrated form of 7'-O-Ac-LT (anhydro-7'-O-Ac-LT). These identifications are consistent with the sequential increase in hydrophobicity and retention time: LT-toxin (4.07 - 4.17 min) < 7'-O-Ac-LT (5.60-5.84 min) < anhydro-7'-O-Ac-LT (6.76 min).

From a biosynthetic perspective, acetylation represents a critical post-synthetic modification in fungal secondary metabolism that is often essential for biological functionality [39]. Secondary metabolites modifications, which modulate molecules' polarity, can function as a mechanism for self-protection, stabilize the toxin against oxidative stress, or influence its cellular transport [40,41]. The discovery of these derivatives could represent a step forward in understanding the toxicity of the *Alternaria* genus. These modifications could play a role in the self-protection and transportation of the LT-toxin, a compound previously demonstrated to be harmful for apple plants [26]. However, the specific effects of the acetylated and anhydro forms are not yet confirmed. Therefore, future studies are required to test their phytotoxic activity compared to the parent toxin.

This annotation propagation cascade highlights the advantage of the NMN approach. By overcoming the limitations of single-polarity analysis, where critical functional group data can be separated from scaffold information, the NMN strategy effectively navigates the complex *Alternaria* metabolome.

4. Conclusion

In this study, we introduced the Neutral Molecular Network (NMN) concept, a new strategy that merges fragmentation spectra from both positive and negative ionization modes into unified neutral pseudo-spectra. By integrating complementary spectral information, NMNs increase chemical informativeness, yielding higher-quality spectral matches and improved structural relevance compared with single-polarity networks. The resulting networks exhibit greater interpretability, forming more coherent clusters of structurally related compounds. In the *Alternaria* case study, NMNs facilitated the discovery of

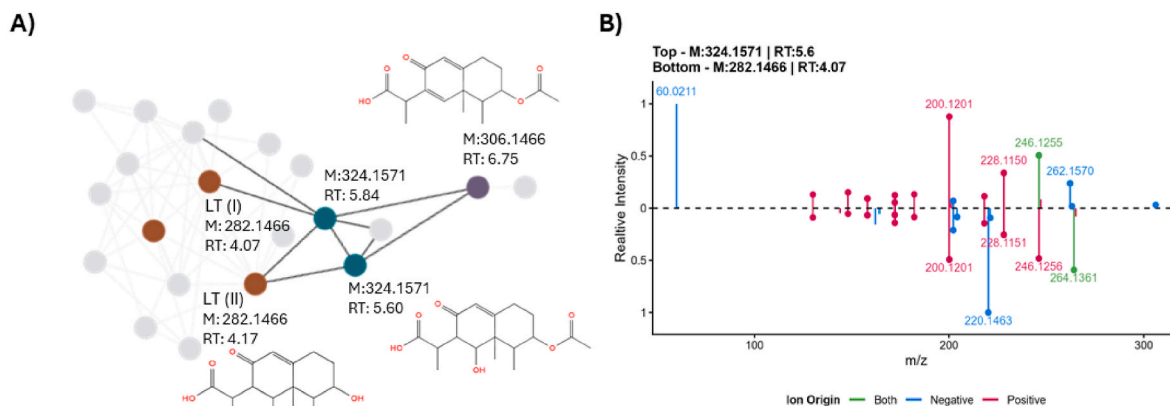


Fig. 5. Annotation propagation using NMN on the untargeted *Alternaria* dataset. A) NMN sub-region highlighting an uncharacterized compound and its isomer (blue node, M: 324.1571; RT: 5.60 & 5.84) and their connections (dark edges) to two LT-toxin isomers (brown nodes) and to another node with mass of 306.1466 (purple node). Known chemical structures are shown for the LT-toxin and a putative structure is provided for the acetylated and anhydro-acetylated derivatives. B) The mirror pseudo-spectrum illustrates the relationship between the first unknown compound (top) and LT-toxin I (bottom). Peaks are colored by their ionization source: red for positive mode, blue for negative mode, and green for fragments detected in both polarities. Matched ions, for either m/z or neutral loss match, are highlighted with top dots. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

two previously uncharacterized derivatives, which may play a biological role in *Alternaria* phytotoxicity.

Overall, NMNs offer a more interpretable and structurally meaningful representation of molecular relationships, serving as a faster data inspection tool that can be complemented by traditional single-polarity analyses. Although the approach inevitably excludes features that do not ionize in both modes, the retained signals represent high-confidence metabolites, reducing redundancy and spectral noise. By streamlining annotation and improving network quality, this approach can accelerate untargeted metabolomics workflows and help researchers transition more efficiently from raw spectra to biological insight and chemical discovery.

CRediT authorship contribution statement

Nicola Dalla Valle: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Pietro Franceschi:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Mar Garcia-Aloy:** Conceptualization, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. **Peter Robatscher:** Funding acquisition, Methodology, Supervision, Writing – review & editing. **Sabine Oettl:** Methodology, Writing – review & editing. **Michael Oberhuber:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The PhD-Project of N.D.V., entitled “Development of molecular networking approaches for the analysis of untargeted metabolomics datasets (MolNet)”, was funded by the Fondazione Edmund Mach and Laimburg Research Centre. Laimburg Research Centre is funded by the Autonomous Province of Bozen-Bolzano.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aca.2026.345708>.

Data availability

Data will be made available on request.

References

- [1] M. Sindelar, G.J. Patti, Chemical discovery in the era of metabolomics, *J. Am. Chem. Soc.* 142 (20) (2020) 9097–9105, <https://doi.org/10.1021/jacs.9b13198>, American Chemical Society.
- [2] I. Gertsman, B.A. Barshop, Promises and pitfalls of untargeted metabolomics, *J. Inher. Metab. Dis.* 41 (3) (2018) 355–366, <https://doi.org/10.1007/s10545-017-0130-7>, Springer Netherlands.
- [3] E. Stancliffe, M. Schwaiger-Haber, M. Sindelar, M.J. Murphy, M. Soerensen, G. J. Patti, An untargeted metabolomics workflow that scales to thousands of samples for population-based studies, *Anal. Chem.* 94 (50) (2022) 17370–17378, <https://doi.org/10.1021/acs.analchem.2c01270>.
- [4] J. Guo, H. Yu, S. Xing, T. Huan, Addressing big data challenges in mass spectrometry-based metabolomics, *Chem. Commun.* 58 (72) (2022) 9979–9990, <https://doi.org/10.1039/d2cc03598g>.
- [5] R. Chaleckis, I. Meister, P. Zhang, C.E. Wheelock, Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics, *Curr. Opin. Biotechnol.* 55 (2019) 44–50, <https://doi.org/10.1016/j.copbio.2018.07.010>.
- [6] E. Defossez, J. Bourquin, S. von Reuss, S. Rasmann, G. Glauser, Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics, *Mass Spectrom. Rev.* 42 (1) (2023) 131–143, <https://doi.org/10.1002/mas.21715>.
- [7] L.F. Nothias, D. Petras, R. Schmid, K. Dührkop, J. Rainer, A. Sarvepalli, I. Protosyuk, M. Ernst, H. Tsugawa, M. Fleischauer, F. Aicheler, A.A. Aksenov, O. Alka, P. M. Allard, A. Barsch, X. Cachet, A.M. Caraballo-Rodríguez, R.R. da Silva, T. Dang, P.C. Dorrestein, Feature-based molecular networking in the GNPS analysis environment, *Nat. Methods* 17 (9) (2020) 905–908, <https://doi.org/10.1038/s41592-020-0933-6>.
- [8] J.S. Yu, L.F. Nothias, M. Wang, D.H. Kim, P.C. Dorrestein, K. bin Kang, H.H. Yoo, Tandem mass spectrometry molecular networking as a powerful and efficient tool for drug metabolism studies, *Anal. Chem.* 94 (2) (2022) 1456–1464, <https://doi.org/10.1021/acs.analchem.1c04925>.
- [9] A.T. Aron, E.C. Gentry, K.L. McPhail, L.F. Nothias, M. Nothias-Esposito, A. Bouslimani, D. Petras, J.M. Gauglitz, N. Sikora, F. Vargas, J.J.J. van der Hooft, M. Ernst, K. bin Kang, C.M. Aceves, A.M. Caraballo-Rodríguez, I. Koester, K. C. Weldon, S. Bertrand, C. Roullier, P.C. Dorrestein, Reproducible molecular networking of untargeted mass spectrometry data using GNPS, *Nat. Protoc.* 15 (6) (2020) 1954–1991, <https://doi.org/10.1038/s41596-020-0317-5>.
- [10] M. Wang, J.J. Carver, V.v. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.v. Melnik, M.J. Meehan, W.T. Liu, M. Crüsemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, N. Bandeira, Sharing and community curation of mass spectrometry data with global natural products social molecular networking, *Nat. Biotechnol.* 34 (8) (2016) 828–837, <https://doi.org/10.1038/nbt.3597>.
- [11] S. Bourcier, Y. Hoppilliard, Use of diagnostic neutral losses for structural information on unknown aromatic metabolites: an experimental and theoretical study, *Rapid Commun. Mass Spectrom.* 23 (1) (2009) 93–103, <https://doi.org/10.1002/rcm.3852>.
- [12] A. Nordström, E. Want, T. Northen, J. Lehtiö, G. Siuzdak, Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics, *Anal. Chem.* 80 (2) (2008) 421–429, <https://doi.org/10.1021/ac701982e>.
- [13] M. Holcápek, K. Volná, D. Vaněrková, Effects of functional groups on the fragmentation of dyes in electrospray and atmospheric pressure chemical ionization mass spectra, *Dyes Pigments* 75 (1) (2007) 156–165, <https://doi.org/10.1016/j.dyepig.2006.05.040>.
- [14] S. Banerjee, S. Mazumdar, Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte, *Inter. J. Analytical Chem.* (2012) 1–40, <https://doi.org/10.1155/2012/282574>, 2012.
- [15] A. Aisporna, H.P. Benton, A. Chen, R.J.E. Derks, J.M. Galano, M. Giera, G. Siuzdak, Neutral loss mass spectral data enhances molecular similarity analysis in METLIN, *J. Am. Soc. Mass Spectrom.* 33 (3) (2022) 530–534, <https://doi.org/10.1021/jasms.1c00343>.
- [16] W. Bittremieux, R. Schmid, F. Huber, J.J. van der Hooft, M. Wang, P.C. Dorrestein, Comparison of cosine, modified cosine, and neutral loss based spectrum alignment for discovery of structurally related molecules, <https://doi.org/10.1101/2022.06.01.494370>, 2022.
- [17] P. Liigand, K. Kaupmees, K. Haav, J. Liigand, I. Leito, M. Girod, R. Antoine, A. Krueve, Think negative: finding the best electrospray ionization/MS mode for your analyte, *Anal. Chem.* 89 (11) (2017) 5665–5668, <https://doi.org/10.1021/acs.analchem.7b00096>.
- [18] S.Y. Choi, J. Park, J. Kim, J. Lee, H. Yang, Investigation of chemical profiles of different parts of *Morus alba* using a combination of molecular networking methods with mass spectral data from two ionization modes of LC/MS, *Plants* 10 (8) (2021), <https://doi.org/10.3390/plants10081711>.
- [19] A. Aleta, Y. Moreno, Multilayer networks in a nutshell, *Annu. Rev. Condens. Matter Phys.* 10 (2019) 45–62, <https://doi.org/10.1146/annurev-conmatphys.10.2019.45>.
- [20] M. Zitnik, J. Leskovec, Predicting multicellular function through multi-layer tissue networks, *Bioinformatics* 33 (14) (2017) i190–i198, <https://doi.org/10.1093/bioinformatics/btx252>.
- [21] A. Amara, C. Frainay, F. Jourdan, T. Naake, S. Neumann, E.M. Novoa-del-Toro, R. M. Salek, L. Salzer, S. Scharfenberg, M. Witting, Networks and graphs discovery in metabolomics data analysis and interpretation, *Front. Mol. Biosci.* 9 (2022), <https://doi.org/10.3389/fmolb.2022.841373>.
- [22] N. Dalla Valle, M. Garcia-Aloy, P. Robatscher, P. Franceschi, M. Oberhuber, Improving spectral similarity and molecular network reliability through noise signal filtering in MS/MS spectra, *Anal. Chem.* 97 (29) (2025) 15873–15882, <https://doi.org/10.1021/acs.analchem.5c02109>.
- [23] Y. Li, T. Kind, J. Folz, A. Vaniya, S.S. Mehta, O. Fiehn, Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification, *Nat. Methods* 18 (12) (2021) 1524–1531, <https://doi.org/10.1038/s41592-021-01331-z>.
- [24] P. Louail, C. Brunius, M. Garcia-Aloy, W. Kumler, N. Storz, J. Stanstrup, H. Treutler, P. Vangeenderhuysen, M. Witting, S. Neumann, J. Rainer, *Xcms* in peak form: now anchoring a complete metabolomics data preprocessing and analysis software ecosystem, *Anal. Chem.* (2025), <https://doi.org/10.1021/acs.analchem.5c04338>.
- [25] J. Lou, L. Fu, Y. Peng, L. Zhou, Metabolites from *alternaria* fungi and their bioactivities, *Molecules* 18 (2013) 5891–5935, <https://doi.org/10.3390/molecules18055891>.
- [26] S. Oettl, M. Bosello, K. Marschall, A. Gallmetzer, Y. Reyes-Domínguez, C. Kreutz, M. Tollinger, P. Robatscher, M. Oberhuber, (3 ζ ,4 ζ ,5 ζ ,6 ζ ,7 ζ ,11 ζ)-3,6-Dihydroxy-8-oxo-9-eremophilene-12-oic acid, a new phytotoxin of *Alternaria alternata* ssp. *teuissima* isolates associated with fruit spots on apple (*Malus × domestica* Borkh.), *J. Agric. Food Chem.* 69 (48) (2021) 14445–14458, <https://doi.org/10.1021/acs.jafc.1c02847>.

- [27] R. Spina, A. Ropars, S. Bouazzi, S. Dadi, P. Lemiere, F. Dupire, A. Khiralla, S. Yagi, J.P. Fripiat, D. Laurain-Mattar, Screening of anti-inflammatory activity and metabolomics analysis of endophytic fungal extracts; identification and characterization of perylenequinones and terpenoids from the interesting active *Alternaria* Endophyte, *Molecules* 28 (18) (2023), <https://doi.org/10.3390/molecules28186531>.
- [28] K. Dührkop, M. Fleischauer, M. Ludwig, et al., Sirius 4: a rapid tool for identifying metabolite structure from tandem mass spectrometry data, *Nat. Methods* 16 (2019) 299–302, <https://doi.org/10.1038/s41592-019-0344-8>.
- [29] J. Rainer, A. Vicini, L. Salzer, J. Stanstrup, J.M. Badia, S. Neumann, M.A. Stravs, V. V. Hernandez, L. Gatto, S. Gibb, et al., A modular and expandable ecosystem for metabolomics data annotation in R, *Metabolites* 12 (2) (2022) 8577–8582, <https://doi.org/10.3390/metabo12020173M>.
- [30] E. J. Newman, Modularity and community structure in networks, *Proc. Natl. Acad. Sci. USA* 2006, 103.
- [31] R. Schmid, D. Petras, L.F. Nothias, M. Wang, A.T. Aron, A. Jagels, H. Tsugawa, J. Rainer, M. Garcia-Aloy, K. Dührkop, A. Korf, T. Pluskal, Z. Kamenik, A. K. Jarmusch, A.M. Caraballo-Rodríguez, K.C. Weldon, M. Nothias-Esposito, A. A. Aksenov, A. Bauermeister, P.C. Dorrestein, Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment, *Nat. Commun.* 12 (1) (2021), <https://doi.org/10.1038/s41467-021-23953-9>.
- [32] M.E.J. Newman, Assortative mixing in networks, *Phys. Rev. Lett.* 89 (20) (2002) 208701, <https://doi.org/10.1103/PhysRevLett.89.208701>.
- [33] M.E.J. Newman, Modularity and community structure in networks, *Proc. Natl. Acad. Sci. U. S. A* 103 (23) (2006) 8577–8582, <https://doi.org/10.1073/pnas.0601602103>.
- [34] G. Csárdi, T. Nepusz, The igraph software package for complex network research, *InterJournal, Complex Systems* 1695 (2006). <https://igraph.org>.
- [35] A. Clauset, Finding local community structure in networks, *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* 72 (2) (2005), <https://doi.org/10.1103/PhysRevE.72.026132>.
- [36] Y. Wang, T. Backman, K. Horan, T. Girke, *fmcsR*: mismatch tolerant maximum common substructure searching, *Bioconductor* (2025), <https://doi.org/10.18129/B9.bioc.fmcsR>.
- [37] Y. Cao, A. Charisi, L.-C. Cheng, T. Jiang, T. Girke, ChemmineR: a compound mining framework for R, *Bioinformatics* 24 (15) (2008) 1733–1734, <https://doi.org/10.1093/bioinformatics/btn307>.
- [38] A. Nikolopoulos, D. van Herwerden, V. Turkina, A. Krueve, M. Baerenfaenger, S. Samanipour, Ionization efficiency prediction of electrospray ionization mass spectrometry analytes based on molecular fingerprints and cumulative neutral losses, *J. Cheminf.* 17 (1) (2025) 183, <https://doi.org/10.1186/s13321-025-01129-7>.
- [39] D. Broadhurst, R. Goodacre, S.N. Reinke, J. Kuligowski, I.D. Wilson, M.R. Lewis, W. B. Dunn, Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies, *Metabolomics* 14 (6) (2018), <https://doi.org/10.1007/s11306-018-1367-3>.
- [40] N.P. Keller, Translating biosynthetic gene clusters into fungal armor and weaponry, *Nat. Chem. Biol.* 11 (9) (2015) 671–677, <https://doi.org/10.1038/nchembio.1897>.
- [41] J. Collemare, M.F. Seidl, Chromatin-dependent regulation of secondary metabolite biosynthesis in fungi: is the picture complete? *FEMS Microbiol. Rev.* 43 (6) (2019) 591–607, <https://doi.org/10.1093/femsre/fuz018>.
- [42] J.J. Coleman, E. Mylonakis, Efflux in Fungi: La Pièce de Résistance, *PLoS Pathog.* 5 (6) (2009) e1000486, <https://doi.org/10.1371/journal.ppat.1000486>.