Molecular Networks

Abterra bio

A deep dive into the world of small molecules

First published July 29, 2016

- **IDENTIFY known compounds** using a large, natural product library.
- **EXPLORE 'analog' compounds** based on similarity to known molecules.
- **TRACK compounds** between samples and across time.

Motivation

Molecular Networking is a revolutionary tool for visualizing, identifying, and tracking natural products. By harnessing the high-throughput and high accuracy characteristics of mass spectrometry, we are using Molecular Networks to explore the world of small molecules. The network-based approach was recently touted as providing a "massively spectacular view" of chemical compounds [4].

The molecular networking paradigm

Tandem mass spectra (MS/MS) are generated from a complex sample using one of many available protocols targeting different types of natural products [3]. The Molecular Network framework is highly flexible, and capable of analyzing different classes of natural products, even within the same dataset. From the MS/MS data, Molecular Networks automatically derives a discrete collection of compounds along with meta-information about how the compounds are structurally similar. The network construction relies on two observations of mass spectra of compounds:

1. The MS/MS spectrum produced by a compound is influenced by its structure, and can serve as a fingerprint for that structure.
2. The MS/MS spectra of structurally similar compounds are similar.

While the first point is widely accepted, the second point has been demonstrated on several classes of molecules including peptides, primary and secondary metabolites, lipids, and glycans [3]. It is important to note that the **network can be constructed without the need for either compound to be identified, and without any structural knowledge of the compounds.**

Network topology

Figure 1 illustrates how a molecular network is constructed. First, multiple MS/MS spectra from the same compound are merged together to produce a 'consensus spectrum'. The consensus spectrum is a better version of each individual spectrum, retaining and amplifying the signal while reducing the noise. In addition, time is not wasted on re-analyzing spectra from the same compound. Depending on experimental conditions, this consensus-building step may reduce the number of spectra by 30%-70%. Window peak filtering and intensity normalization are also employed. After these pre-processing steps, each consensus spectrum becomes a node in the network.

Next, a similarity score is computed between all pairs of consensus spectra. We employ a generic score based on the cosine similarity of the two spectra. Similarity scores that are tailored to specific types of molecules, such as peptides, can be substituted.

The connectivity of the network is determined by applying a similarity score threshold. An edge is created between two consensus spectra if their similarity score exceeds the threshold.

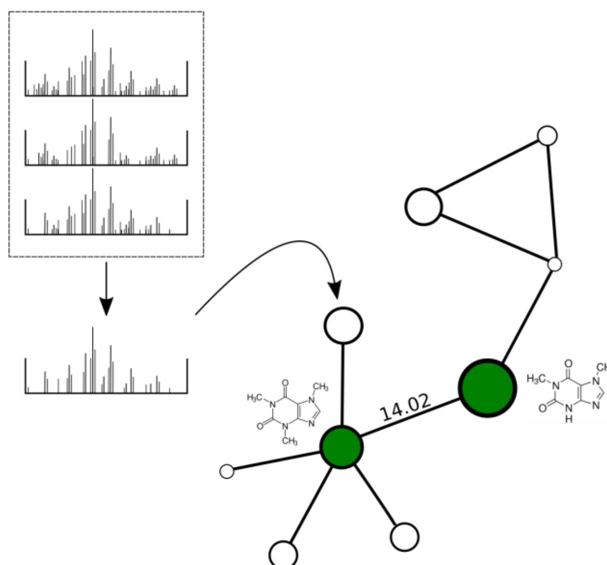


Figure 1: Example network constructed after spectral clustering and library identification.

Compound identification

Molecular Networks contains a large, built-in spectral library of compounds. From FDA-approved drugs, to phytochemicals, to NIH clinical compounds, our library contains over 10,000 library spectra from multiple instruments and multiple ion modes. Our library also includes publicly-contributed spectra from across the small molecule community [1]. After network construction, consensus spectra are matched to our library and annotated with the compound information. In Figure 1, two compounds (colored green) have been identified against the library. The astute observer will notice that one compound is caffeine, while the other compound is a derivative of caffeine, paraxanthine.

If we restrict our identifications to compounds that exactly match a compound in our spectral library, we would only be able to identify a small fraction of the compounds in an experiment. Manually identifying the remaining spectra is a monumental task, and discarding them is a waste. Molecular Networks can help!

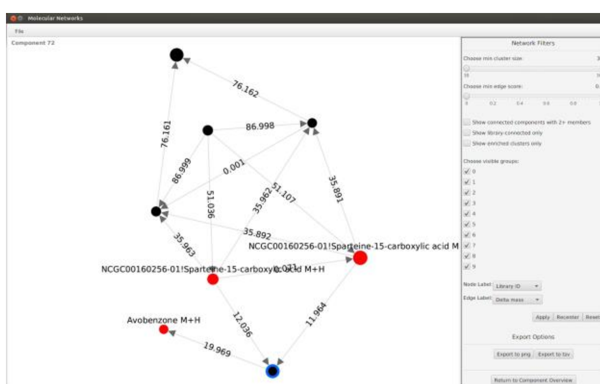
By viewing unknown compounds in a network together with known compounds, scientists can focus efforts on relevant molecules that are structurally similar to compounds with desirable chemical and pharmaceutical properties. In addition, by determining the mass change between a known compound and an unknown compound, scientists can quickly determine which compounds may simply be known compounds with uninteresting adducts, and which compounds are worthy of more attention.

Compound tracking

The ability to track compounds and their abundance across samples or time points is a powerful component of Molecular Networks. Each input spectrum can be assigned a user-defined label (e.g. 'Case' and 'Control, or 'T0', 'T1', and 'T2'). Using the interactive Molecular Networks viewer, compounds appearing exclusively in certain conditions or environments can be easily identified. In addition, the distribution of a compound across the labels can be monitored (see the real example in the next section). In fact, the viewer can be used to dynamically control the visualization, including which nodes and edges are shown and how they are labeled. For example, the 'Sub-Network List View' shown in Figure 2a allows filtering on a number of statistics, including network size, number of identified compounds, and distribution of spectra across labels, to identify the sub-networks that are the most interesting. The 'Sub-Network View' shown in Figure 2b shows the topology for a specific sub-network. The figure shows edge labels containing mass differences between compounds, while the node label contains the library identification. In addition, the user has chosen to highlight 'enriched' nodes (shown with a blue border). Enriched nodes represent compounds whose abundance diverges significantly from the background distribution of spectra across labels. The viewer is both flexible and intuitive, enabling researchers to view the compounds they wish to see using several filters at once. In addition, users can export component information into a tab-delimited file or save images of the network.

ComponentID	Num Clusters	Library IDs	Num Spectra	Groups										Enrichment Count	Go to component
				0	1	2	3	4	5	6	7	8	9		
394	99	7	9763	0.08	0.10	0.13	0.08	0.11	0.11	0.10	0.10	0.10	0.10	16	View
72	8	3	440	0.13	0.08	0.05	0.05	0.08	0.17	0.09	0.11	0.14	0.10	1	View
56	13	2	975	0.05	0.11	0.11	0.12	0.10	0.12	0.11	0.08	0.11	0.08	0	View
122	3	2	355	0.07	0.10	0.07	0.10	0.09	0.11	0.08	0.09	0.12	0.17	1	View
297	7	2	415	0.06	0.07	0.06	0.07	0.09	0.13	0.06	0.13	0.18	0.15	2	View
293	1	1	38	0.00	0.05	0.26	0.03	0.16	0.00	0.05	0.00	0.00	0.45	0	View
117	1	1	42	0.00	0.00	0.17	0.19	0.14	0.07	0.12	0.07	0.14	0.10	0	View
316	6	1	360	0.10	0.07	0.01	0.01	0.09	0.20	0.16	0.08	0.17	0.11	3	View
76	1	1	49	0.00	0.06	0.16	0.00	0.10	0.16	0.00	0.08	0.12	0.31	0	View
27	1	1	42	0.02	0.10	0.07	0.10	0.07	0.14	0.05	0.14	0.12	0.19	0	View
145	1	1	64	0.00	0.03	0.02	0.00	0.11	0.19	0.03	0.16	0.25	0.22	1	View
111	1	1	38	0.00	0.00	0.00	0.00	0.18	0.32	0.13	0.18	0.13	0.05	0	View
18	1	1	53	0.00	0.08	0.19	0.02	0.15	0.06	0.02	0.21	0.11	0.17	0	View
88	2	0	76	0.64	0.00	0.00	0.00	0.14	0.18	0.03	0.00	0.00	0.00	0	View
31	1	0	33	0.45	0.00	0.00	0.00	0.00	0.21	0.00	0.00	0.33	0.00	0	View
128	1	0	39	0.00	0.05	0.05	0.13	0.00	0.08	0.05	0.00	0.31	0.33	0	View
344	1	0	31	0.35	0.00	0.00	0.00	0.10	0.13	0.13	0.00	0.26	0.03	0	View
132	1	0	38	0.03	0.08	0.00	0.08	0.00	0.05	0.11	0.00	0.24	0.42	0	View
369	1	0	31	0.32	0.00	0.00	0.03	0.00	0.19	0.13	0.00	0.23	0.10	0	View
141	1	0	55	0.02	0.04	0.05	0.20	0.02	0.09	0.11	0.00	0.18	0.29	1	View

(a) Table view of sub-networks.



(b) A single sub-network of interest.

Figure 2: The Molecular Networks viewer showing the a) sub-network table and, b) a single sub-network with library IDs and highly divergent nodes.

A real world example

We demonstrate the Molecular Network approach using a publicly available dataset of human skin metabolites and beauty products (MSV000079559), previously described [2]. A single subject's face samples were analyzed across all available time points. The subject was instructed to abstain from using any beauty products for weeks 1-3, to use select products during weeks 4-6, and to resume their normal beauty regimen for weeks 7-9.

We constructed the network on 157,191 MS/MS spectra, resulting in 86,314 compounds (a 45% reduction). In order to quickly identify sub-networks of interest, a 'Sub-Network List View' is provided as a table, shown in Figure 2a. This table includes each sub-network as a separate row, with different statistics about the composition of that sub-network. Having chosen 'Component 72', we can view the topology of the sub-network. Figure 2b highlights three different library identifications, shown as red nodes. Each node can be clicked on to reveal relative spectral abundances of that molecule across the different time points.

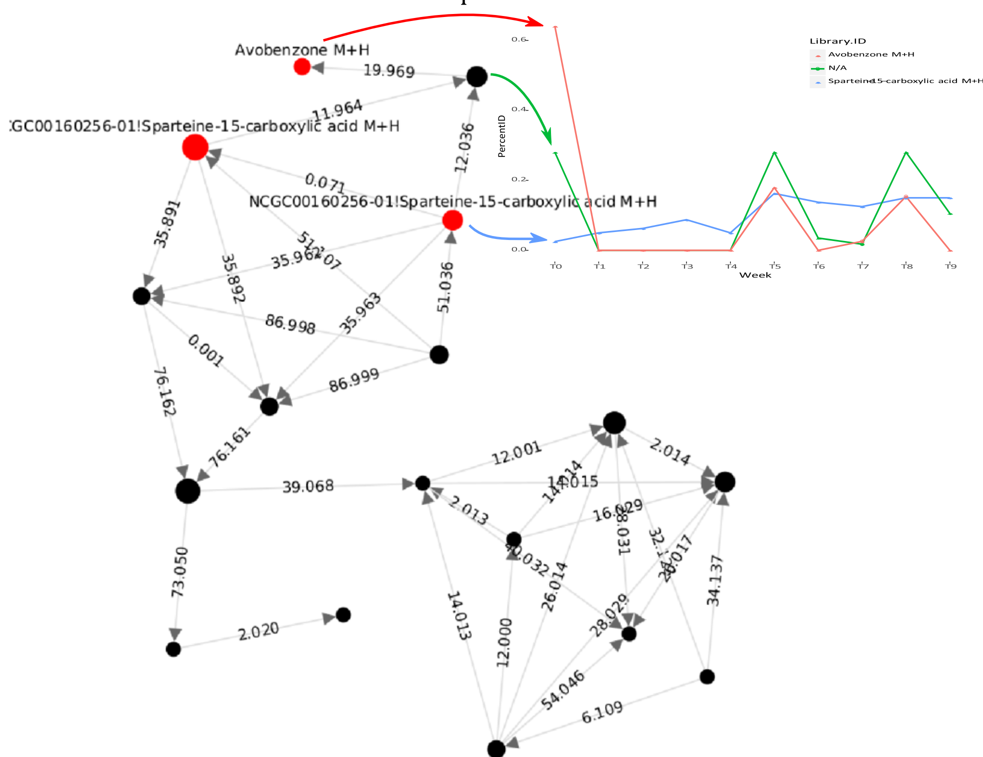


Figure 3: Avobenzone molecule and similar molecule tracking together over time.

In depth information of the sub-network can be exported to a CSV file for further analysis. For the component shown in Figure 2b, we plotted the spectral counts for each compound at each of the time points, and observed correlations among the connected nodes. Figure 3 shows a library identified node matching avobenzone, a UVA absorber in sunscreen, administered during weeks 4-6. Another molecule of 20 Da difference tracks closely with avobenzone.

References

[1] GNPS: Global Natural Products Social Molecular Networking. http://gnps.ucsd.edu.

[2] A. Bouslimani, C. Porto, C. M. Rath, M. Wang, Y. Guo, A. Gonzalez, D. Berg-Lyon, G. Ackermann, G. J. M. Christensen, T. Nakatsuji, et al. Molecular cartography of the human skin surface in 3d. *Proceedings of the National Academy of Sciences*, 112(17):E2120–E2129, 2015.

[3] A. Guthals, J. D. Watrous, P. C. Dorrestein, and N. Bandeira. The spectral networks paradigm in high throughput mass spectrometry. *Mol Biosyst*, 8(10):2535–2544, Oct 2012.

[4] M. F. Traxler and R. Kolter. A massively spectacular view of the chemical lives of microbes. *Proc. Natl. Acad. Sci. U.S.A.*, 109(26):10128–10129, Jun 2012.