**VAST 2025 MC1 Data Description**

Important notes:

* All of the data is synthetic.  Resemblance to real people or situations is incidental.
* Use only data from this mini challenge to answer the questions.  Data may be inconsistent between challenges.  No external sources are required to answer the questions posed.

**Graph Description**

* File name: MC1\_graph.json
* Directed Multigraph
* # nodes: 17,412
* # edges: 37, 857
* 18 connected components
* Possible node types are: {Person, Song, RecordLabel, Album, MusicalGroup}
* Possible edge types are: {MemberOf, PerformerOf, ComposerOf, ProducerOf, LyricistOf, InStyleOf, InterpolatesFrom, CoverOf, LyricalReferenceTo, DirectlySamples, RecordedBy, DistributedBy}
* The graph format is a JSON file generated by Python’s [network.node\_link\_data()](https://networkx.org/documentation/stable/reference/readwrite/generated/networkx.readwrite.json_graph.node_link_data.html) function. It can likewise be loaded to a networkx object using the corresponding [node\_link\_graph()](https://networkx.org/documentation/stable/reference/readwrite/generated/networkx.readwrite.json_graph.node_link_graph.html#networkx.readwrite.json_graph.node_link_graph) function. The root-level JSON object consists of graph-level properties specifying that it is directed and a multigraph, a “nodes” key which holds the list of nodes, and a “links” key which holds the list of edges.

The data for this challenge comes from two different sources

* an online, crowdsourced repository of musical influence, where contributors have manually notated instances in which songs or albums have sampled, covered, or otherwise drawn inspiration from previous work.
* aggregate-level descriptions of song popularity, as labeled by our journalist, Silas Reed. These notations (*notable* and *notoriety\_date*) are directly based on both the number of sales and streams that the song/album has achieved as well as the whether or not the work landed on a top record chart.

**Graph Schema**

**Nodes**

Person nodes: These can be anyone in the music industry, including singers, producers, instrumentalists, composers, etc.

* **Node Type** **(string)** – the type of node
* **name** **(string)** – the name of the person
* **stage\_name (string)** – if provided, the stage name of the musician

MusicalGroup nodes: Bands, quartets, small choirs, or other officially organized entities formed by musicians to make music.

* **Node Type** **(string)** – the type of node
* **name** **(string)** – the name of the person

RecordLabel nodes: These are organizations—professional, commercial, or otherwise institutional—involved in the recording, production, or distribution of the music.

* **Node Type (string)** – the type of node
* **Name (string)** – the name of the organization

Song nodes:

* **Node Type (string)** – the type of node
* **single (boolean)** – if provided, specifies whether or not the song was released as a standalone, or as part of a larger album
* **genre (string)** – the song’s genre
* **notable (boolean)** – whether or not the song has appeared on a top record chart
* **release\_date (string)** – the year in which the song was released
* **notoriety\_date (string)** – if provided, the year in which the song first appeared on a top record chart
* **written\_date (string)** – if provided, the year in which the song was written

Album nodes:

* **Node Type (string)** – the type of node
* **genre (string)** – the album’s genre
* **notable (boolean)** – whether or not the album has appeared on a top record chart
* **release\_date (string)** – the year in which the album was released
* **notoriety\_date (string)** – if provided, the year in which the album first appeared on a top record chart
* **written\_date (string)** – if provided, the year in which the album was written

**Edges**

For all of the edges in this graph, the only property is **Edge Type**. These types are listed below.

* **PerformerOf** – Indicates a that the source node (Person or MusicalGroup) performed the destination node (Song or Album)
* **ComposerOf** – Indicates that the source node (Person) composed the destination node (Song or Album)
* **ProducerOf** – Indicates that the source node (Person or RecordLabel) participated in the production of the destination node’s work (Song, Album, Person, or MusicalGroup)
* **LyricistOf** – Indicates that the source node (Person) wrote lyrics for the destination node (Song or Album)
* **RecordedBy** – Indicates that the destination node (RecordLabel) aided in the recording process for the source node (Song or Album)
* **DistributedBy –** Indicates that the destination node (RecordLabel) aided in the distribution process for the source node (Song or Album)
* **InStyleOf –** Indicates that the source node (Song or Album) was performed at least partly in the style of the destination node (Song, Album, Person, or MusicalGroup)
* **InterpolatesFrom –** Indicates that the source node (Song or Album) [interpolated](https://en.wikipedia.org/wiki/Interpolation_(popular_music)) a melody from the destination node (Song or Album).
* **CoverOf –** Indicates that the source node (Song or Album) is a [cover](https://en.wikipedia.org/wiki/Cover_version) of the destination node (Song or Album)
* **LyricalReferenceTo –** Indicates that the source node (Song or Album) makes a lyrical reference to the destination node (Song or Album)
* **DirectlySamples –** Indicates that the source node (Song or Album) consists of (an) audio recording(s) that directly reuse a portion of the audio recording of the destination node (Song or Album) via [sampling](https://en.wikipedia.org/wiki/Sampling_(music))
* **MemberOf** – Indicates that the source node (Person) is (or was) a member of the destination node (MusicalGroup)