

The methods provided in the `parse_go` module allow for the creation of objects through which the Gene Ontology can be accessed programmatically. Currently, the module contains one object definition; `parse_obo`, which is capable of interpreting the Gene Ontology in the OBO flat-file format. Object instantiation accepts the following parameters:

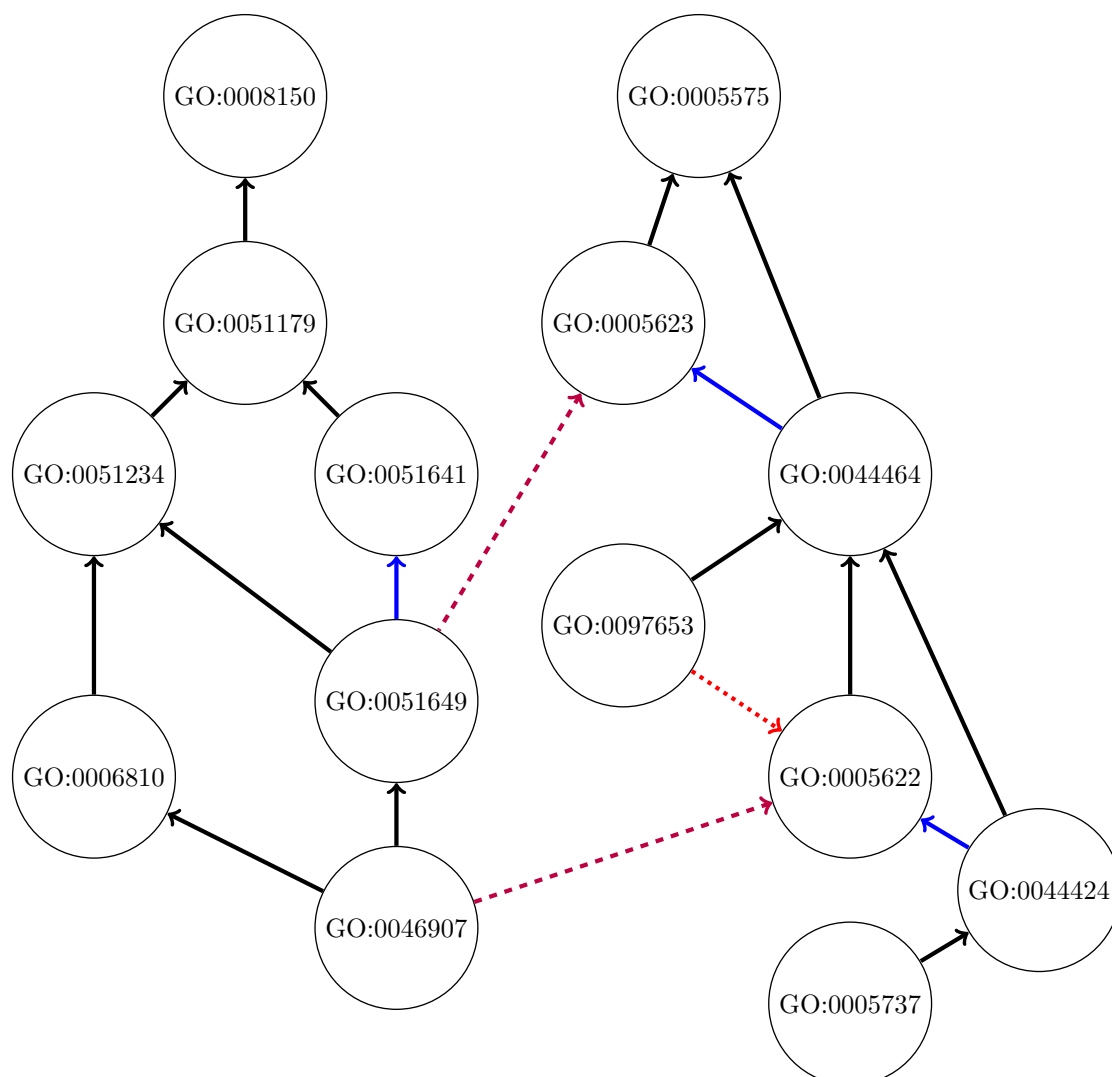
weights	A dictionary object which serves to map relations within the Gene Ontology to weights based off of how much the user wants to restrict movement along a given relation when reducing a list of GO terms. If a given relation is not included in the dictionary, then that relation is ignored entirely. The value of a weight is meaningless in of itself, and is used solely in order to score different paths, where a larger score represents a longer path. As an example, given the weight dictionary <code>{"is_a":1, "part_of":3}</code> , a path that consists of three consecutive “is_a” relations will be treated as if it is the same length as a path that consists of a single “part_of” relation. Currently, this dictionary defaults to <code>{"is_a": 1, "part_of": 2, "regulates": 5}</code> .
roots	The list of GO terms that serve to root the ontology in general. As of the time of writing, this list defaults to the GO terms “GO:0003674”, “GO:0005575”, and “GO:0008150”.
relations	The list of relations found in the Gene Ontology. As of the time of writing, this list defaults to “is_a”, “part_of”, “regulates”, “negatively_regulates”, “positively_regulates”, “has_part”, “occurs_in”, and “happens_during”.
disjoint	The list of relations that, when used to build a directed graph from the Gene Ontology, result in an acyclic graph. As of the time of writing, the sole relation in the Gene Ontology that meets this requirement is the “is_a” relation.
source_path	The filesystem path that points to the Gene Ontology OBO dump. If not provided, the script looks inside its own directory for a file entitled <code>go.obo</code> . If none is found, it attempts to download one from http://geneontology.org/ontology/go.obo .
save_detail	A boolean flag that, when set to <code>True</code> , creates an internal dictionary <code>obo_detail</code> that contains all information corresponding to a given GO term found in the original Gene Ontology OBO dump.
verbose	A boolean flag that, when set to <code>True</code> , will result in the object printing debug information during method calls.

Following object instantiation, the following methods can be called, with the caveat that each method’s output can vary wildly based off of what values were passed to each function parameter when the original `parse_obo` object was created.

`find_lsca(go_term1, go_term2):`

Central to the GO term set reduction functionality of the `parse_obo` object is the `find_lsca` method. When provided with two GO terms, the `find_lsca` method attempts to find the Lowest Single Common Ancestor¹ (abbreviated as LSCA) between the two GO terms. This method starts by locating the root term within `roots` that minimalizes the summed squared distance from both GO terms. If such a shared root is located, then we compute the list of all reachable terms for each of the two GO terms, and compute the intersection of these two term sets. At this point, the set intersection contains at least the shared root we found originally. For each GO term in the resulting intersection, we calculate the summed squared distance from that GO term to each of the two GO terms, as well as the distance between that GO term and the shared root term. Finally, we select the GO term from this intersection with the greatest distance from the shared root term, and that which minimizes the distance from the two original GO terms. This term is then labeled the LSCA between the two original GO terms. As a visual example, the selection process proceeds as follows:

¹Johannes Fischer, Daniel H. Huson, New common ancestor problems in trees and directed acyclic graphs, Information Processing Letters, Volume 110, Issues 89, 1 April 2010, Pages 331-335, ISSN 0020-0190, <http://dx.doi.org/10.1016/j.ipl.2010.02.014>. (<http://www.sciencedirect.com/science/article/pii/S0020019010000487>)



Provided in the above visual is a subsection of the Gene Ontology; in particular, it contains all nodes that can be reached from GO:0005737 (cytoplasm) or from GO:0046907 (intracellular transport). It contains four types of relations: the "**is_a**" relation, denoted by a black arrow, the "**part_of**" relation, denoted by a blue arrow, the "**occurs_in**" relation, denoted by a dashed, purple arrow, and finally the "**has_part**" relation, denoted by a dotted, red arrow. The process for finding the LSCA between GO:0005737 and GO:0046907 then occurs as follows:

1. Locate the shared root: in this case, GO:0005575.
2. Determine list of all reachable GO terms from either term:
 - (a) *GO:0005737*: GO:0044424, **GO:0005622**, **GO:0044464**, **GO:0005623**, **GO:0005575**.
 - (b) *GO:0046907*: GO:0006810, GO:0051649, **GO:0005622**, GO:0051234, GO:0051641, **GO:0044464**, GO:0051179, **GO:0005623**, GO:0008150, **GO:0005575**.
3. Determine list of shared terms: **GO:0005622**, **GO:0044464**, **GO:0005623**, **GO:0005575**.
4. Determine which term is furthest from the shared root: **GO:0005622** (intercellular).
5. As we only have one term, it is our LSCA. If there were more, the term with the smallest summed squared distance from both GO:0005737 and GO:0046907 would be selected as our LSCA.

```
reduce_list(go_terms, branch_count, avoid_roots, range_modifier, range_width,  
            path_lengths, dijkstra_lengths, verbose):
```

fetch_update(url, filepath):

url	The full URL from which to fetch an updated version of <code>go.obo</code> . If none is provided, it defaults to <code>http://geneontology.org/ontology/go.obo</code> .
filepath	The full or relative system path where the downloaded file should be saved. If none is provided, it defaults to <code>./go.obo</code> .

Method used to update the base `go.obo` file that the `parse_obo` relies on when building its internal graph representation of the Gene Ontology.

generate_detail():

Simple method that allows the user to create the `obo_detail` object if they set the `save_detail` flag to `False` during object instantiation without having to recreate the entire `parse_obo` object.

generate_circle_representation(size, verbose):

size	An integer value that defines the side length of a square that could contain the entirety of the generated representation. Defaults to 960, which can result in the placement of terms anywhere between the cartesian coordinate point $(-480, -480)$ and $(480, 480)$.
verbose	A boolean flag that, when set to <code>True</code> , will result in the method printing debug information during the function's runtime.

An as of yet unimplemented method intended to provide cartesian coordinate points for each term found in the Gene Ontology such that a roughly circular representation of the overall Gene Ontology can be constructed by associating each term with its matching (x, y) coordinate.