

The *GOSim* package

Holger Fröhlich

February 7, 2014

1 Introduction

The Gene Ontology (GO) has become one of the most widespread systems for systematically annotating gene products within the bioinformatics community and is developed by the Gene Ontology Consortium (The Gene Ontology Consortium, 2004). It is specifically intended for describing gene products with a controlled and structured vocabulary. GO terms are part of a Directed Acyclic Graph (DAG), covering three orthogonal taxonomies or "aspects": *molecular function*, *biological process* and *cellular component*. Two different kinds of relationship between GO terms exist: the "is-a" relationship and the "part-of" relationship. Providing a standard vocabulary across any biological resources, the GO enables researchers to use this information for automated data analysis.

The *GOSim* package (Fröhlich et al., 2007) provides the researcher with various information theoretic similarity concepts for GO terms (Resnik, 1995, 1999; Lin, 1998; Jiang and Conrath, 1998; Lord et al., 2003; Couto et al., 2003, 2005). Moreover, since version 1.1.5 *GOSim* contains several new similarity concepts, which are based on so-called diffusion kernel techniques (Lerman and Shakhnovich, 2007). Additionally *GOSim* implements different methods for computing functional similarities between gene products based on the similarities between the associated GO terms (Speer et al., 2005; Fröhlich et al., 2006; Schlicker et al., 2006; Lerman and Shakhnovich, 2007; del Pozo et al., 2008). This can, for instances, be used for clustering genes according to their biological function (Speer et al., 2005; Fröhlich et al., 2006) and thus may help to get a better understanding of the biological aspects covered by a set of genes.

Since version 1.1 *GOSim* additionally offers the possibility of a GO enrichment analysis using the topGO package (Alexa et al., 2006). Hence, *GOSim* acts now as an umbrella for different analysis methods employing the GO structure.

2 Usage of *GOSim*

To elucidate the usage of *GOSim* we show an example workflow and explain the employed similarity concepts. We create a character vector of Entrez gene IDs, which we assume to be from human:

```
> library(GOSim)
> genes=c("207", "208", "596", "901", "780", "3169", "9518", "2852", "26353", "8614", "7494")
```

Next we investigate the GO annotation within the current ontology (which is *biological process* by default):

```
> getGOInfo(genes)
```

	207	208	596	901	780
go_id	Character,101	Character,29	Character,111	Character,3	Character,17
Term	Character,101	Character,29	Character,111	Character,3	Character,17
Definition	Character,101	Character,29	Character,111	Character,3	Character,17
IC	Numeric,101	Numeric,29	Numeric,111	Numeric,3	Numeric,17
	3169	9518	2852	26353	8614
go_id	Character,22	Character,3	Character,39	Character,3	Character,13
Term	Character,22	Character,3	Character,39	Character,3	Character,13
Definition	Character,22	Character,3	Character,39	Character,3	Character,13
IC	Numeric,22	Numeric,3	Numeric,39	Numeric,3	Numeric,13
	7494				
go_id	Character,11				
Term	Character,11				
Definition	Character,11				
IC	Numeric,11				

2.1 Term Similarities

Let us examine the similarity of the GO terms for genes "8614" and "2852" in greater detail:

```
> getTermSim(c("GO:0007166", "GO:0007267", "GO:0007584", "GO:0007165", "GO:0007186"), m=1)
```

	GO:0007166	GO:0007267	GO:0007584	GO:0007165	GO:0007186
GO:0007166	0.2628131	0.1806383	0.1266641	0.1945233	0.2628131
GO:0007267	0.1806383	0.3551639	0.0000000	0.1806383	0.1806383
GO:0007584	0.1266641	0.0000000	0.5128961	0.1266641	0.1266641
GO:0007165	0.1945233	0.1806383	0.1266641	0.1945233	0.1945233
GO:0007186	0.2628131	0.1806383	0.1266641	0.1945233	0.4016432

This calculates Resnik's pairwise similarity between GO terms (Resnik, 1995, 1999):

$$sim(t, t') = IC_{ms}(t, t') := \max_{\hat{t} \in Pa(t, t')} IC(\hat{t}) \quad (1)$$

Here $Pa(t, t')$ denotes the set of all common ancestors of GO terms t and t' , while $IC(t)$ denotes the information content of term t . It is defined as (e.g. Lord et al. (2003))

$$IC(\hat{t}) = -\log P(\hat{t}) \quad (2)$$

i.e. as the negative logarithm of the probability of observing \hat{t} . The information content of each GO term is already precomputed for each ontology based on the empirical observation, how many times a specific GO term or any of its direct or indirect offsprings appear in the annotation of the GO with gene products. GOSim provides a normalized version of Resnik's similarity measure, which divides the information content of the minimum subsumer by the maximum information content of all GO terms, hence obtaining a number between 0 and 1.

```
> data("ICsBPHumanall")
> IC[c("GO:0007166", "GO:0007267", "GO:0007584", "GO:0007165", "GO:0007186")]

GO:0007166 GO:0007267 GO:0007584 GO:0007165 GO:0007186
3.006413 4.062846 5.867200 2.225221 4.594539
```

This loads the information contents of all GO terms within "biological process". Likewise, the data files ICsMFHumanall and ICsCCHumanall contain the information contents of all GO terms within "molecular function" and "cellular component" for human. Since GOSim version 1.1.4.0 the information content of GO terms relies on the mapping of primary gene IDs (mainly Entrez) to GO terms provided by the libraries org.Dm.eg.db (fly), org.Hs.eg.db (human), org.Mm.eg.db (mouse), etc. Additionally, it is possible to pass a user provided mapping via the function `setEvidenceLevel`. Please refer to the manual pages for details. If only GO terms having certain evidence codes should be considered, one must explicitly calculate the corresponding information contents in the function `calcICs`. Again, more information on this function can be found in the manual pages.

To continue our example from above, let us also calculate Jiang and Conrath's pairwise similarity between GO terms, which is the default, for comparison reasons (Jiang and Conrath, 1998):

```
> getTermSim(c("GO:0007166", "GO:0007267", "GO:0007584", "GO:0007165", "GO:0007186"), v=

GO:0007166 GO:0007267 GO:0007584 GO:0007165 GO:0007186
GO:0007166 0.9505312 0.5105747 0.2498911 0.7587689 0.7519293
GO:0007267 0.5105747 0.9828000 0.0000000 0.5740054 0.4169139
GO:0007584 0.2498911 0.0000000 0.9971692 0.2740140 0.2119568
GO:0007165 0.7587689 0.5740054 0.2740140 0.8919565 0.5820734
GO:0007186 0.7519293 0.4169139 0.2119568 0.5820734 0.9898931
```

Jiang and Conrath's similarity measure is defined as

$$sim(t, t') = 1 - \min(1, IC(t) - 2IC_{ms}(t, t') + IC(t')) \quad (3)$$

i.e. the similarity between t and t' is 0, if their normalized distance is at least 1.

Likewise, we can also compute Lin's pairwise similarity between GO terms (Lin, 1998):

```
> getTermSim(c("GO:0007166", "GO:0007267", "GO:0007584", "GO:0007165", "GO:0007186"), m
```

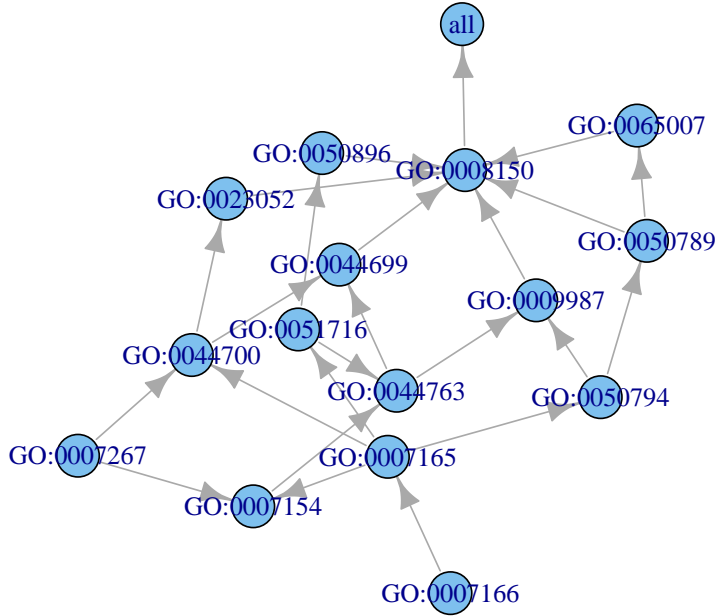
	GO:0007166	GO:0007267	GO:0007584	GO:0007165	GO:0007186
GO:0007166	1.0000000	0.5846115	0.3265762	0.8506792	0.7910622
GO:0007267	0.5846115	1.0000000	0.0000000	0.6572401	0.4773693
GO:0007584	0.3265762	0.0000000	1.0000000	0.3581018	0.2770009
GO:0007165	0.8506792	0.6572401	0.3581018	1.0000000	0.6525805
GO:0007186	0.7910622	0.4773693	0.2770009	0.6525805	1.0000000

It is defined as:

$$sim(t, t') = \frac{2IC_{ms}(t, t')}{IC(t) + IC(t')} \quad (4)$$

Resnik's, Jiang-Conraths's and Lin's term similarities all refer to $IC_{ms}(t, t')$, the information content of the minimum subsumer of t and t' , i.e. of the lowest common ancestor in the hierarchy. For illustration let us plot the GO graph with leaves GO:0007166 and GO:0007267 and let us compute their minimum subsumer (see Fig. ??):

```
> library(igraph)
> G = getGOGraph(c("GO:0007166", "GO:0007267"))
> G2 = igraph.from.graphNEL(G)
> plot(G2, vertex.label=V(G2)$name)
```



```
> getMinimumSubsumer("GO:0007166", "GO:0007267")
```

```
[1] "GO:0023052"
```

In contrast to the above defined similarity measures Couto et al. (Couto et al., 2005) introduced a concept, which is not based on the minimum subsumer, but on the set of all disjunctive common ancestors. Roughly speaking, the idea is not to consider the common ancestor having the highest information content only, but also others, if they are somehow "separate" from each other, i.e. there exists a path to t or to t' not passing any other of the disjunctive common ancestors.

```
> getDisjCommAnc("GO:0007166", "GO:0007267")
```

```
[1] "GO:0007154" "GO:0009987" "GO:0023052" "GO:0044699" "GO:0044700"
```

```
[6] "GO:0044763"
```

In this case the set of disjunctive common ancestors consists of the minimum subsumer, GO:0007154, and its parent, GO:0009987, because from both there exists a path to GO:0007166 not passing any other disjunctive common ancestor(see Fig. ??).

Based on the notion of disjunctive common ancestors Resnik's similarity concept can be extended by defining:

$$sim(t, t') = IC_{share}(t, t') = \frac{1}{|DisjCommAnc|} \sum_{t \in DisjCommAnc} IC(t) \quad (5)$$

Likewise, Jiang-Conraths's and Lin's measures can be extended as well by replacing $IC_{ms}(t, t')$ by $IC_{share}(t, t')$.

```
> getTermSim(c("GO:0007166", "GO:0007267", "GO:0007584", "GO:0007165", "GO:0007186"), m
```

	GO:0007166	GO:0007267	GO:0007584	GO:0007165	GO:0007186
GO:0007166	3.006413	1.507568	1.448955	1.545332	2.225221
GO:0007267	1.507568	4.062846	0.000000	1.507568	1.507568
GO:0007584	1.448955	0.000000	5.867200	1.448955	1.448955
GO:0007165	1.545332	1.507568	1.448955	2.225221	1.545332
GO:0007186	2.225221	1.507568	1.448955	1.545332	4.594539

Finally, it should be mentioned that also the depth and density enriched term similarity by Couto et al. (Couto et al., 2003) has been integrated into *GOSim*:

```
> setEnrichmentFactors(alpha=0.5, beta=0.3)
> getTermSim(c("GO:0007166", "GO:0007267", "GO:0007584", "GO:0007165", "GO:0007186"), m
```

	GO:0007166	GO:0007267	GO:0007584	GO:0007165	GO:0007186
GO:0007166	9.038517	0.00000	0.00000	4.456869	3.732446
GO:0007267	0.000000	16.50672	0.00000	0.000000	0.000000
GO:0007584	0.000000	0.00000	34.42404	0.000000	0.000000
GO:0007165	4.456869	0.00000	0.00000	4.951609	0.000000
GO:0007186	3.732446	0.00000	0.00000	0.000000	21.109787

Since version 1.1.5 *GOSim* contains several new similarity concepts, which are based on so-called diffusion kernel techniques (Lerman and Shakhnovich, 2007) rather than on the information theoretic ideas presented before. For using these similarity measures it is necessary to pre-compute a diffusion kernel on the Gene Ontology graph via `calc.diffusion.kernel`. This will take some time and result in a kernel/similarity matrix that is stored in a file called e.g. 'diffKernelpowerBPhumanall.rda' (meaning matrix power diffusion kernel for ontology BP in human using all evidence codes) in the current working directory. Once the kernel is created, it has to be loaded into the environment first `load.diffusion.kernel`. Afterwards GO term similarities can be computed via function `getTermSim`. Please check the manual pages for details.

Since version 1.2 *GOSim* also contains Schlicker et al.'s GO term similarity measure (Schlicker et al., 2006), which is an adaption of Lin's similarity measure. Moreover, the graph information content similarity by Pesquita et al. has been implemented (Pesquita et al., 2007).

```
> getTermSim(c("GO:0007166", "GO:0007267", "GO:0007584", "GO:0007165", "GO:0007186"), m)
```

	GO:0007166	GO:0007267	GO:0007584	GO:0007165	GO:0007186
GO:0007166	0.9505312	0.5105747	0.2498911	0.7587689	0.7519293
GO:0007267	0.5105747	0.9828000	0.0000000	0.5740054	0.4169139
GO:0007584	0.2498911	0.0000000	0.9971692	0.2740140	0.2119568
GO:0007165	0.7587689	0.5740054	0.2740140	0.8919565	0.5820734
GO:0007186	0.7519293	0.4169139	0.2119568	0.5820734	0.9898931

2.2 Functional Gene Similarities

The special strength of *GOSim* lies in the possibility not only to calculate similarities for individual GO terms, but also for genes based on their complete GO annotation. Since *GOSim* version 1.1.5 for this purpose the following ideas have been implemented:

1. Maximum (Couto et al., 2003) and average pairwise GO term similarity
2. Average of best matching GO term similarities (Schlicker et al., 2006).
3. Computation of a so-called *optimal assignment* of terms from one gene to those of another one (Fröhlich et al., 2006).
4. Similarity derived from Hausdorff distances between sets (del Pozo et al., 2008).
5. Embedding of each gene into a feature space: (Speer et al., 2005; Fröhlich et al., 2006) proposed to define feature vectors by a gene's maximum GO term similarity to certain prototype genes. More simple (but probably also less accurate), (Mistry and Pavlidis, 2008) recently proposed to represent each gene by a feature vector describing the presence/absence of all GO terms. The absence of each GO term is additionally weighted by its information content. Within a feature space gene functional similarities naturally arise as dot products between feature vectors. These dot products can be understood as so-called *kernel functions* (Schölkopf and Smola, 2002), as used in e.g. Support Vector Machines (Cortes and Vapnik, 1995). Depending on the choice of later normalization (see below) one can arrive at the cosine similarity (Eq. 6), at the Tanimoto coefficient (Eq. 7) or at a measure similar to Lin's one (Eq. 8, Eq. 4).

2.2.1 Normalization of Similarities

Often, people want to normalize similarities, e.g. on the interval $[0, 1]$, for better interpretation. To do so, we can perform the transformation

$$sim_{gene}(g, g') \leftarrow \frac{sim_{gene}(g, g')}{\sqrt{sim_{gene}(g, g)sim_{gene}(g', g')}} \quad (6)$$

Provided $sim_{gene} \geq 0$, the consequence will be a similarity of 1 for g with itself and between 0 and 1 for g with any other gene. In case of a feature space embedding this transformation is equivalent to computing the cosine similarity between two feature vectors.

Another possibility is to use Lin's normalization (see Eq. 4):

$$sim_{gene}(g, g') \leftarrow \frac{2sim_{gene}(g, g')}{sim_{gene}(g, g) + sim_{gene}(g', g')} \quad (7)$$

Furthermore, one can use a normalization in the spirit of the Tanimoto coefficient:

$$sim_{gene}(g, g') \leftarrow \frac{sim_{gene}(g, g')}{sim_{gene}(g, g) + sim_{gene}(g', g') - sim_{gene}(g, g')} \quad (8)$$

In case of a feature space embedding the transformation corresponds exactly to the Tanimoto coefficient between two feature vectors.

We now give a more detailed overview over the different similarity concepts mentioned above.

2.2.2 Maximum and Average Pairwise GO Term Similarity

The idea of the maximum pairwise GO term similarity is straight forward. Given two genes g and g' annotated with GO terms t_1, \dots, t_n and t'_1, \dots, t'_m we define the functional similarity between g and g' as

$$sim_{gene}(g, g') = \max_{\substack{i = 1, \dots, n \\ j = 1, \dots, m}} sim(t_i, t'_j) \quad (9)$$

where sim is some similarity measure to compare GO terms t_i and t'_j . This idea is, for instance, realized in FuSSiMeg (Couto et al., 2003). Instead of computing the maximum pairwise GO term similarity one may also take the average here.

2.2.3 Average of Best Matching GO Term Similarities

The idea of this approach (Schlicker et al., 2006) is to assign each GO term t_i occurring in gene g to its best matching partner $t'_{\pi i}$ in gene g' . Hence multiple GO terms from gene g can be assigned to one GO term from gene g' . A similarity score is computed by taking the average similarity of assigned GO terms. Since, however, genes can have an unequal number of GO terms the result depends on whether GO terms of gene g are assigned to those of gene g' or vice versa. Hence, in Schlicker et al. (2006) it was proposed to either take the maximum or the average of both similarity scores. Both strategies are implemented in *GOSim*.

2.2.4 Optimal Assignment Gene Similarities

To elucidate the idea of the optimal assignment (Fröhlich et al., 2006), consider the GO terms associated with gene "8614" on one hand and gene "2852" on the other hand:

```
> getGOInfo(c("8614", "2852"))
```

	8614	2852
go_id	Character,13	Character,39
Term	Character,13	Character,39
Definition	Character,13	Character,39
IC	Numeric,13	Numeric,39

Given a similarity concept *sim* to compare individual GO terms, the idea is now to assign each term of the gene having fewer annotation to exactly one term of the other gene such that the overall similarity is maximized. More formally the optimal assignment problem can be stated as follows: Let π be some permutation of either an n -subset of natural numbers $\{1, \dots, m\}$ or an m -subset of natural numbers $\{1, \dots, n\}$ (this will be clear from context). Then we are looking for the quantity

$$sim_{gene}(g, g') = \begin{cases} \max_{\pi} \sum_{i=1}^n sim(t_i, t'_{\pi(i)}) & \text{if } m > n \\ \max_{\pi} \sum_{j=1}^m sim(t_{\pi(j)}, t'_j) & \text{otherwise} \end{cases} \quad (10)$$

The computation of (10) corresponds to the solution of the classical maximum weighted bipartite matching (optimal assignment) problem in graph theory and can be carried out in $O(\max(n, m)^3)$ time (Mehlhorn and Näher, 1999). To prevent that larger lists of terms automatically achieve a higher similarity we may further sim_{gene} divide 10 by $\max(m, n)$.

In our example, using Lin's GO term similarity measure the following assignments yielding a corresponding similarity matrix are found:

```
> getGeneSim(c("8614", "2852"), similarity="OA", similarityTerm="Lin", avg=FALSE, verb
```

```
filtering out genes not mapping to the currently set GO category ... ==> list of 2
term similarity matrix:
```

	GO:0006874	GO:0006979	GO:0007166	GO:0007267	GO:0007566	GO:0007584
GO:0006874	1.00000000	0.00000000	0.2274750	0.08408554	0.00000000	0.00000000
GO:0006979	0.00000000	1.00000000	0.3275458	0.00000000	0.00000000	0.2475126
GO:0007166	0.22747501	0.3275458	1.00000000	0.58461150	0.00000000	0.3265762
GO:0007267	0.08408554	0.00000000	0.5846115	1.00000000	0.00000000	0.00000000
GO:0007566	0.00000000	0.00000000	0.00000000	0.00000000	1.00000000	0.00000000
GO:0007584	0.00000000	0.2475126	0.3265762	0.00000000	0.00000000	1.00000000
GO:0030968	0.16376373	0.4414432	0.4733334	0.39514856	0.00000000	0.4434258
GO:0033280	0.00000000	0.2092914	0.2631646	0.00000000	0.00000000	0.8458709

G0:0040015	0.13039013	0.0000000	0.1739665	0.00000000	0.1860972	0.0000000
G0:0043434	0.00000000	0.2605260	0.3496181	0.00000000	0.0000000	0.4877125
G0:0046697	0.00000000	0.0000000	0.0000000	0.00000000	0.7783503	0.0000000
G0:0071456	0.06225648	0.4052713	0.3727105	0.07111270	0.0000000	0.4071629
G0:2001256	0.12046150	0.0000000	0.1581194	0.00000000	0.0000000	0.0000000
G0:0030968		G0:0033280	G0:0040015	G0:0043434	G0:0046697	G0:0071456
G0:0006874	0.1637637	0.0000000	0.1303901	0.0000000	0.0000000	0.06225648
G0:0006979	0.4414432	0.2092914	0.0000000	0.2605260	0.0000000	0.40527132
G0:0007166	0.4733334	0.2631646	0.1739665	0.3496181	0.0000000	0.37271054
G0:0007267	0.3951486	0.0000000	0.0000000	0.0000000	0.0000000	0.07111270
G0:0007566	0.0000000	0.0000000	0.1860972	0.0000000	0.7783503	0.00000000
G0:0007584	0.4434258	0.8458709	0.0000000	0.4877125	0.0000000	0.40716293
G0:0030968	1.0000000	0.4483505	0.1368560	0.5528910	0.0000000	0.54474045
G0:0033280	0.4483505	1.0000000	0.0000000	0.4859237	0.0000000	0.35097269
G0:0040015	0.1368560	0.0000000	1.0000000	0.0000000	0.1781009	0.00000000
G0:0043434	0.5528910	0.4859237	0.0000000	1.0000000	0.0000000	0.42580898
G0:0046697	0.0000000	0.0000000	0.1781009	0.0000000	1.0000000	0.00000000
G0:0071456	0.5447404	0.3509727	0.0000000	0.4258090	0.0000000	1.00000000
G0:2001256	0.1268545	0.0000000	0.1074404	0.0000000	0.0000000	0.00000000

G0:2001256

G0:0006874	0.1204615
G0:0006979	0.0000000
G0:0007166	0.1581194
G0:0007267	0.0000000
G0:0007566	0.0000000
G0:0007584	0.0000000
G0:0030968	0.1268545
G0:0033280	0.0000000
G0:0040015	0.1074404
G0:0043434	0.0000000
G0:0046697	0.0000000
G0:0071456	0.0000000
G0:2001256	1.0000000

optimal assignment:

	G0:0006874	G0:0006979	G0:0007166	G0:0007267	G0:0007566	G0:0007584
G0:0006874	1	0	0	0	0	0
G0:0006979	0	1	0	0	0	0
G0:0007166	0	0	1	0	0	0
G0:0007267	0	0	0	1	0	0
G0:0007566	0	0	0	0	1	0
G0:0007584	0	0	0	0	0	1
G0:0030968	0	0	0	0	0	0

G0:0033280	0	0	0	0	0	0
G0:0040015	0	0	0	0	0	0
G0:0043434	0	0	0	0	0	0
G0:0046697	0	0	0	0	0	0
G0:0071456	0	0	0	0	0	0
G0:2001256	0	0	0	0	0	0
G0:0030968	G0:0033280	G0:0040015	G0:0043434	G0:0046697	G0:0071456	
G0:0006874	0	0	0	0	0	0
G0:0006979	0	0	0	0	0	0
G0:0007166	0	0	0	0	0	0
G0:0007267	0	0	0	0	0	0
G0:0007566	0	0	0	0	0	0
G0:0007584	0	0	0	0	0	0
G0:0030968	1	0	0	0	0	0
G0:0033280	0	1	0	0	0	0
G0:0040015	0	0	1	0	0	0
G0:0043434	0	0	0	1	0	0
G0:0046697	0	0	0	0	1	0
G0:0071456	0	0	0	0	0	1
G0:2001256	0	0	0	0	0	0
G0:2001256						
G0:0006874	0					
G0:0006979	0					
G0:0007166	0					
G0:0007267	0					
G0:0007566	0					
G0:0007584	0					
G0:0030968	0					
G0:0033280	0					
G0:0040015	0					
G0:0043434	0					
G0:0046697	0					
G0:0071456	0					
G0:2001256	1					

=====

term similarity matrix:

G0:0001934	G0:0002695	G0:0007186	G0:0007204	G0:0008284	G0:0008285	
G0:0001934	1.00000000	0.22007801	0.26720333	0.18199644	0.52148479	0.25414486
G0:0002695	0.22007801	1.00000000	0.22310634	0.15560831	0.22053153	0.47416475
G0:0007186	0.26720333	0.22310634	1.00000000	0.18460669	0.26787217	0.25819191
G0:0007204	0.18199644	0.15560831	0.18460669	1.00000000	0.18238716	0.17670495
G0:0008284	0.52148479	0.22053153	0.26787217	0.18238716	1.00000000	0.83460599

G0:0008285	0.25414486	0.47416475	0.25819191	0.17670495	0.83460599	1.00000000
G0:0045909	0.32808696	0.14359053	0.16544703	0.41038426	0.32866354	0.15961704
G0:0010579	0.60758037	0.14396173	0.59899810	0.11735835	0.31932412	0.15779813
G0:0010628	0.90656841	0.20142197	0.24723804	0.19163405	0.48876477	0.23444188
G0:0010629	0.88341800	0.46988842	0.24081684	0.18738474	0.23738239	0.54501949
G0:0014068	0.39446408	0.17371799	0.35976742	0.14240737	0.39520019	0.19427372
G0:0030263	0.53768169	0.04850166	0.05499861	0.04983805	0.05452169	0.05328585
G0:0030264	0.05301288	0.04736763	0.05354497	0.04864143	0.05309282	0.05192021
G0:0030335	0.47047325	0.20218546	0.24127903	0.16663498	0.47152076	0.23058080
G0:0030518	0.22184371	0.19057139	0.40096172	0.15671929	0.22230455	0.21559634
G0:0030819	0.78464087	0.18031913	0.21077711	0.14800225	0.41256254	0.20256677
G0:0032024	0.39756345	0.17490633	0.33674516	0.43624498	0.39831119	0.19576113
G0:0032962	0.62390673	0.14740834	0.16715416	0.12024546	0.32791800	0.16194864
G0:0043065	0.51862086	0.21951705	0.26637688	0.18151326	0.51989403	0.25339710
G0:0043280	0.39856694	0.18926695	0.22310634	0.15560831	0.22053153	0.21392831
G0:0045742	0.37722034	0.16706246	0.46456731	0.13678042	0.37789345	0.18598749
G0:0045745	0.36650265	0.16288774	0.68956821	0.13325805	0.36713802	0.18082797
G0:0045944	0.84089759	0.21718211	0.26294644	0.17950313	0.51328965	0.25029088
G0:0050728	0.18169213	0.37518069	0.24528662	0.14833294	0.18204697	0.42158261
G0:0050769	0.43429381	0.18880885	0.22247007	0.15521828	0.43518626	0.21334324
G0:0051053	0.67953459	0.40137701	0.21182945	0.14865546	0.20950699	0.45113618
G0:0051281	0.36463635	0.33666581	0.18637737	0.82201990	0.36526526	0.37356134
G0:0070374	0.42187574	0.18414548	0.38510107	0.15125169	0.42271783	0.20740823
G0:0070474	0.30304961	0.13387850	0.15268474	0.12552444	0.30354148	0.14770597
G0:0071333	0.15791354	0.13765840	0.31553102	0.68149839	0.15820761	0.15391441
G0:0071356	0.06965070	0.06022107	0.33604865	0.06229511	0.06978875	0.06777665
G0:0071375	0.07934080	0.06733108	0.38350699	0.06993437	0.07951998	0.07691810
G0:0071392	0.05871209	0.05186616	0.28268505	0.05339731	0.05881015	0.05737480
G0:0090004	0.31837899	0.16055023	0.18425690	0.13128821	0.31892193	0.17795176
G0:0090200	0.37850501	0.16756089	0.34503835	0.13720133	0.37918272	0.18660546
G0:2000353	0.33377457	0.14995556	0.17043710	0.12238158	0.33430145	0.16502841
G0:2001238	0.36826583	0.16357655	0.45341130	0.13383884	0.36890733	0.18167726
	G0:0045909	G0:0010579	G0:0010628	G0:0010629	G0:0014068	G0:0030263
G0:0001934	0.3280870	0.60758037	0.9065684	0.8834180	0.39446408	0.53768169
G0:0002695	0.1435905	0.14396173	0.2014220	0.4698884	0.17371799	0.04850166
G0:0007186	0.1654470	0.59899810	0.2472380	0.2408168	0.35976742	0.05499861
G0:0007204	0.4103843	0.11735835	0.1916341	0.1873847	0.14240737	0.04983805
G0:0008284	0.3286635	0.31932412	0.4887648	0.2373824	0.39520019	0.05452169
G0:0008285	0.1596170	0.15779813	0.2344419	0.5450195	0.19427372	0.05328585
G0:0045909	1.0000000	0.22610477	0.3421795	0.1674750	0.26724622	0.00000000
G0:0010579	0.2261048	1.00000000	0.5427966	0.5344116	0.26673083	0.51599856
G0:0010628	0.3421795	0.54279659	1.0000000	0.9394387	0.36427001	0.55793145

G0:0010629	0.1674750	0.53441156	0.9394387	1.0000000	0.17807925	0.54907611
G0:0014068	0.2672462	0.26673083	0.3642700	0.1780793	1.00000000	0.04534089
G0:0030263	0.0000000	0.51599856	0.5579314	0.5490761	0.04534089	1.00000000
G0:0030264	0.0000000	0.03821905	0.0000000	0.0000000	0.04434833	0.76892582
G0:0030335	0.3050573	0.29944313	0.4383232	0.2134480	0.36519261	0.05104167
G0:0030518	0.1444517	0.25798039	0.2031207	0.1987664	0.31164093	0.04876162
G0:0030819	0.2761455	0.80570540	0.7085433	0.6943227	0.32880043	0.61856523
G0:0032024	0.3715106	0.26814434	0.3672641	0.1795146	0.47284146	0.04558657
G0:0032962	0.2309581	0.51185417	0.5579314	0.5490761	0.27270051	0.39229369
G0:0043065	0.3273733	0.31824798	0.4859168	0.2360344	0.39355320	0.46346781
G0:0043280	0.1435905	0.49791837	0.4187852	0.4098813	0.17371799	0.41372555
G0:0045742	0.2581913	0.31489797	0.3476513	0.1701040	0.55698685	0.04395192
G0:0045745	0.2524689	0.47177884	0.3373554	0.1651560	0.54406382	0.04306930
G0:0045944	0.3243970	0.95475829	0.9232019	0.8691041	0.38975697	0.71394167
G0:0050728	0.1379022	0.16129923	0.1904048	0.4447040	0.39145296	0.00000000
G0:0050769	0.4216132	0.28436546	0.4029113	0.1965692	0.34301195	0.04841018
G0:0051053	0.1381562	0.60661177	0.7122031	0.6978366	0.16680380	0.62064129
G0:0051281	0.3474814	0.25275048	0.3355652	0.3916027	0.29804305	0.04291406
G0:0070374	0.2812671	0.27898835	0.3908254	0.1907941	0.60937914	0.04747334
G0:0070474	0.5371171	0.21392453	0.3150341	0.1544103	0.25039531	0.00000000
G0:0071333	0.3687019	0.21088121	0.1651188	0.1619543	0.25109319	0.04527597
G0:0071356	0.0000000	0.21985244	0.0000000	0.0000000	0.26391601	0.04724700
G0:0071375	0.0000000	0.23921956	0.0000000	0.0000000	0.29232594	0.05151490
G0:0071392	0.0000000	0.19568509	0.0000000	0.0000000	0.22984118	0.04194583
G0:0090004	0.2492503	0.22145124	0.3316331	0.1624036	0.26076939	0.26105619
G0:0090200	0.2588723	0.25933705	0.3488871	0.1706975	0.45444183	0.37580949
G0:2000353	0.2345299	0.23752704	0.3060727	0.1500891	0.27710074	0.34357588
G0:2001238	0.2534153	0.30973233	0.3390474	0.1659695	0.54619932	0.36863425
G0:0030264	G0:0030335	G0:0030518	G0:0030819	G0:0032024	G0:0032962	
G0:0001934	0.05301288	0.47047325	0.22184371	0.78464087	0.39756345	0.62390673
G0:0002695	0.04736763	0.20218546	0.19057139	0.18031913	0.17490633	0.14740834
G0:0007186	0.05354497	0.24127903	0.40096172	0.21077711	0.33674516	0.16715416
G0:0007204	0.04864143	0.16663498	0.15671929	0.14800225	0.43624498	0.12024546
G0:0008284	0.05309282	0.47152076	0.22230455	0.41256254	0.39831119	0.32791800
G0:0008285	0.05192021	0.23058080	0.21559634	0.20256677	0.19576113	0.16194864
G0:0045909	0.00000000	0.30505729	0.14445173	0.27614547	0.37151057	0.23095808
G0:0010579	0.03821905	0.29944313	0.25798039	0.80570540	0.26814434	0.51185417
G0:0010628	0.00000000	0.43832321	0.20312066	0.70854332	0.36726410	0.55793145
G0:0010629	0.00000000	0.21344797	0.19876643	0.69432266	0.17951462	0.54907611
G0:0014068	0.04434833	0.36519261	0.31164093	0.32880043	0.47284146	0.27270051
G0:0030263	0.76892582	0.05104167	0.04876162	0.61856523	0.04558657	0.39229369
G0:0030264	1.00000000	0.04978728	0.04761553	0.04564481	0.04458335	0.03895394

G0:0030335	0.04978728	1.00000000	0.20367476	0.37996912	0.56180407	0.30698758
G0:0030518	0.04761553	0.20367476	1.00000000	0.18150277	0.29138827	0.14819840
G0:0030819	0.04564481	0.37996912	0.18150277	1.00000000	0.33095101	0.61359704
G0:0032024	0.04458335	0.56180407	0.29138827	0.33095101	1.00000000	0.27417819
G0:0032962	0.03895394	0.30698758	0.14819840	0.61359704	0.27417819	1.00000000
G0:0043065	0.45136246	0.46917809	0.22127373	0.41076798	0.39663822	0.32678326
G0:0043280	0.40405208	0.20218546	0.19057139	0.32656259	0.17490633	0.26696031
G0:0045742	0.04301861	0.35036498	0.29962867	0.31673193	0.45594997	0.26434662
G0:0045745	0.04217273	0.34110028	0.29209690	0.30914130	0.44530823	0.25903819
G0:0045944	0.05250039	0.46379271	0.21890146	1.22952914	0.39278254	0.53234541
G0:0050728	0.00000000	0.16763172	0.21114524	0.15028506	0.14596298	0.12383164
G0:0050769	0.04728037	0.39907660	0.19010697	0.35601605	0.34535312	0.29116065
G0:0051053	0.04579454	0.19288020	0.18228255	0.75349131	0.16789912	0.46277665
G0:0051281	0.04202387	0.51848390	0.16311437	0.30781242	0.60979728	0.25810450
G0:0070374	0.04638634	0.38856648	0.33047271	0.34762783	0.49912439	0.28552603
G0:0070474	0.00000000	0.28329493	0.13462683	0.25819131	0.25180640	0.21826401
G0:0071333	0.04428622	0.14621792	0.27339872	0.13167215	0.38944335	0.10923842
G0:0071356	0.04617023	0.06418704	0.28867021	0.05746362	0.05579138	0.04724700
G0:0071375	0.05023743	0.07232767	0.32300623	0.06390262	0.06184133	0.05151490
G0:0071392	0.04109495	0.05478137	0.24839124	0.04980768	0.04854645	0.04194583
G0:0090004	0.53455768	0.51304937	0.16148789	0.26923567	0.57617755	0.22610477
G0:0090200	0.36781069	0.35147299	0.30052807	0.31763715	0.45721814	0.26497687
G0:2000353	0.33687814	0.31257520	0.15077324	0.28552603	0.27862662	0.24224950
G0:2001238	0.36093485	0.34262700	0.29333945	0.31039482	0.44706654	0.25991773
	G0:0043065	G0:0043280	G0:0045742	G0:0045745	G0:0045944	G0:0050728
G0:0001934	0.51862086	0.39856694	0.37722034	0.36650265	0.84089759	0.1816921
G0:0002695	0.21951705	0.18926695	0.16706246	0.16288774	0.21718211	0.3751807
G0:0007186	0.26637688	0.22310634	0.46456731	0.68956821	0.26294644	0.2452866
G0:0007204	0.18151326	0.15560831	0.13678042	0.13325805	0.17950313	0.1483329
G0:0008284	0.51989403	0.22053153	0.37789345	0.36713802	0.51328965	0.1820470
G0:0008285	0.25339710	0.21392831	0.18598749	0.18082797	0.25029088	0.4215826
G0:0045909	0.32737331	0.14359053	0.25819131	0.25246887	0.32439701	0.1379022
G0:0010579	0.31824798	0.49791837	0.31489797	0.47177884	0.95475829	0.1612992
G0:0010628	0.48591680	0.41878522	0.34765126	0.33735540	0.92320189	0.1904048
G0:0010629	0.23603440	0.40988134	0.17010395	0.16515597	0.86910408	0.4447040
G0:0014068	0.39355320	0.17371799	0.55698685	0.54406382	0.38975697	0.3914530
G0:0030263	0.46346781	0.41372555	0.04395192	0.04306930	0.71394167	0.0000000
G0:0030264	0.45136246	0.40405208	0.04301861	0.04217273	0.05250039	0.0000000
G0:0030335	0.46917809	0.20218546	0.35036498	0.34110028	0.46379271	0.1676317
G0:0030518	0.22127373	0.19057139	0.29962867	0.29209690	0.21890146	0.2111452
G0:0030819	0.41076798	0.32656259	0.31673193	0.30914130	1.22952914	0.1502851
G0:0032024	0.39663822	0.17490633	0.45594997	0.44530823	0.39278254	0.1459630

G0:0032962	0.32678326	0.26696031	0.26434662	0.25903819	0.53234541	0.1238316
G0:0043065	1.00000000	0.70063704	0.37638727	0.36571620	0.51051478	0.1812531
G0:0043280	0.70063704	1.00000000	0.16706246	0.16288774	0.39332239	0.1574053
G0:0045742	0.37638727	0.16706246	1.00000000	0.52582647	0.37291353	0.3770573
G0:0045745	0.36571620	0.16288774	0.52582647	1.00000000	0.36243577	0.3680040
G0:0045944	0.51051478	0.39332239	0.37291353	0.36243577	1.00000000	0.1794246
G0:0050728	0.18125313	0.15740531	0.37705728	0.36800401	0.17942456	1.0000000
G0:0050769	0.43318996	0.18880885	0.32989845	0.32167182	0.42859502	0.1570415
G0:0051053	0.20859120	0.32795640	0.16065814	0.15679366	0.89993969	0.3596719
G0:0051281	0.36385788	0.16215778	0.28809266	0.28179904	0.36061057	0.3235381
G0:0070374	0.42083403	0.18414548	0.58659183	0.57227618	0.41649616	0.4139151
G0:0070474	0.30244062	0.13387850	0.24242927	0.23737736	0.29989865	0.1289204
G0:0071333	0.15754964	0.13765840	0.24219195	0.23658152	0.15603302	0.1929161
G0:0071356	0.06947992	0.06022107	0.25410019	0.24793151	0.06876871	0.2031519
G0:0071375	0.07911928	0.06733108	0.28033111	0.27284186	0.07819834	0.2259687
G0:0071392	0.05859069	0.05186616	0.22236050	0.21762227	0.05808412	0.1760366
G0:0090004	0.31770691	0.16055023	0.25214097	0.24668075	0.31490302	0.1344452
G0:0090200	0.60911953	0.53480750	0.43881764	0.42895194	0.37416899	0.3781370
G0:2000353	0.64427772	0.47861622	0.26847935	0.26300535	0.33039825	0.1258941
G0:2001238	0.59267736	0.52209060	0.52782094	0.51620176	0.36415995	0.3694990
	G0:0050769	G0:0051053	G0:0051281	G0:0070374	G0:0070474	G0:0071333
G0:0001934	0.43429381	0.67953459	0.36463635	0.42187574	0.3030496	0.15791354
G0:0002695	0.18880885	0.40137701	0.33666581	0.18414548	0.1338785	0.13765840
G0:0007186	0.22247007	0.21182945	0.18637737	0.38510107	0.1526847	0.31553102
G0:0007204	0.15521828	0.14865546	0.82201990	0.15125169	0.1255244	0.68149839
G0:0008284	0.43518626	0.20950699	0.36526526	0.42271783	0.3035415	0.15820761
G0:0008285	0.21334324	0.45113618	0.37356134	0.20740823	0.1477060	0.15391441
G0:0045909	0.42161318	0.13815619	0.34748145	0.28126713	0.5371171	0.36870189
G0:0010579	0.28436546	0.60661177	0.25275048	0.27898835	0.2139245	0.21088121
G0:0010628	0.40291129	0.71220309	0.33556516	0.39082535	0.3150341	0.16511880
G0:0010629	0.19656919	0.69783664	0.39160266	0.19079413	0.1544103	0.16195433
G0:0014068	0.34301195	0.16680380	0.29804305	0.60937914	0.2503953	0.25109319
G0:0030263	0.04841018	0.62064129	0.04291406	0.04747334	0.0000000	0.04527597
G0:0030264	0.04728037	0.04579454	0.04202387	0.04638634	0.0000000	0.04428622
G0:0030335	0.39907660	0.19288020	0.51848390	0.38856648	0.2832949	0.14621792
G0:0030518	0.19010697	0.18228255	0.16311437	0.33047271	0.1346268	0.27339872
G0:0030819	0.35601605	0.75349131	0.30781242	0.34762783	0.2581913	0.13167215
G0:0032024	0.34535312	0.16789912	0.60979728	0.49912439	0.2518064	0.38944335
G0:0032962	0.29116065	0.46277665	0.25810450	0.28552603	0.2182640	0.10923842
G0:0043065	0.43318996	0.20859120	0.36385788	0.42083403	0.3024406	0.15754964
G0:0043280	0.18880885	0.32795640	0.16215778	0.18414548	0.1338785	0.13765840
G0:0045742	0.32989845	0.16065814	0.28809266	0.58659183	0.2424293	0.24219195

G0:0045745	0.32167182	0.15679366	0.28179904	0.57227618	0.2373774	0.23658152
G0:0045944	0.42859502	0.89993969	0.36061057	0.41649616	0.2998986	0.15603302
G0:0050728	0.15704152	0.35967194	0.32353808	0.41391508	0.1289204	0.19291606
G0:0050769	1.00000000	0.18066935	0.32023328	0.36355289	0.3931526	0.13735308
G0:0051053	0.18066935	1.00000000	0.32412456	0.17639483	0.1291423	0.13218892
G0:0051281	0.32023328	0.32412456	1.00000000	0.31343039	0.2364896	0.57233839
G0:0070374	0.36355289	0.17639483	0.31343039	1.00000000	0.2626632	0.26493336
G0:0070474	0.39315259	0.12914229	0.23648960	0.26266324	1.0000000	0.11357774
G0:0071333	0.13735308	0.13218892	0.57233839	0.26493336	0.1135777	1.00000000
G0:0071356	0.06008011	0.05770114	0.05184028	0.27924902	0.0000000	0.51560532
G0:0071375	0.06715492	0.06419648	0.05702387	0.31125615	0.0000000	0.57100148
G0:0071392	0.05176156	0.04998602	0.04552712	0.24138385	0.0000000	0.44913575
G0:0090004	0.27999240	0.15462663	0.53976174	0.27410197	0.2345299	0.11827606
G0:0090200	0.33088061	0.16111904	0.28884138	0.47866667	0.2430296	0.24286041
G0:2000353	0.29618227	0.14477534	0.26204290	0.29035356	0.2214512	0.11099850
G0:2001238	0.32302924	0.15743179	0.28284025	0.57463937	0.2382139	0.23750866
	G0:0071356	G0:0071375	G0:0071392	G0:0090004	G0:0090200	G0:2000353
G0:0001934	0.06965070	0.07934080	0.05871209	0.31837899	0.3785050	0.33377457
G0:0002695	0.06022107	0.06733108	0.05186616	0.16055023	0.1675609	0.14995556
G0:0007186	0.33604865	0.38350699	0.28268505	0.18425690	0.3450383	0.17043710
G0:0007204	0.06229511	0.06993437	0.05339731	0.13128821	0.1372013	0.12238158
G0:0008284	0.06978875	0.07951998	0.05881015	0.31892193	0.3791827	0.33430145
G0:0008285	0.06777665	0.07691810	0.05737480	0.17795176	0.1866055	0.16502841
G0:0045909	0.00000000	0.00000000	0.00000000	0.24925026	0.2588723	0.23452986
G0:0010579	0.21985244	0.23921956	0.19568509	0.22145124	0.2593371	0.23752704
G0:0010628	0.00000000	0.00000000	0.00000000	0.33163309	0.3488871	0.30607267
G0:0010629	0.00000000	0.00000000	0.00000000	0.16240362	0.1706975	0.15008907
G0:0014068	0.26391601	0.29232594	0.22984118	0.26076939	0.4544418	0.27710074
G0:0030263	0.04724700	0.05151490	0.04194583	0.26105619	0.3758095	0.34357588
G0:0030264	0.04617023	0.05023743	0.04109495	0.53455768	0.3678107	0.33687814
G0:0030335	0.06418704	0.07232767	0.05478137	0.51304937	0.3514730	0.31257520
G0:0030518	0.28867021	0.32300623	0.24839124	0.16148789	0.3005281	0.15077324
G0:0030819	0.05746362	0.06390262	0.04980768	0.26923567	0.3176371	0.28552603
G0:0032024	0.05579138	0.06184133	0.04854645	0.57617755	0.4572181	0.27862662
G0:0032962	0.04724700	0.05151490	0.04194583	0.22610477	0.2649769	0.24224950
G0:0043065	0.06947992	0.07911928	0.05859069	0.31770691	0.6091195	0.64427772
G0:0043280	0.06022107	0.06733108	0.05186616	0.16055023	0.5348075	0.47861622
G0:0045742	0.25410019	0.28033111	0.22236050	0.25214097	0.4388176	0.26847935
G0:0045745	0.24793151	0.27284186	0.21762227	0.24668075	0.4289519	0.26300535
G0:0045944	0.06876871	0.07819834	0.05808412	0.31490302	0.3741690	0.33039825
G0:0050728	0.20315185	0.22596869	0.17603662	0.13444522	0.3781370	0.12589407
G0:0050769	0.06008011	0.06715492	0.05176156	0.27999240	0.3308806	0.29618227

G0:0051053	0.05770114	0.06419648	0.04998602	0.15462663	0.1611190	0.14477534
G0:0051281	0.05184028	0.05702387	0.04552712	0.53976174	0.2888414	0.26204290
G0:0070374	0.27924902	0.31125615	0.24138385	0.27410197	0.4786667	0.29035356
G0:0070474	0.00000000	0.00000000	0.00000000	0.23452986	0.2430296	0.22145124
G0:0071333	0.51560532	0.57100148	0.44913575	0.11827606	0.2428604	0.11099850
G0:0071356	1.00000000	0.60516659	0.47000721	0.05134081	0.2548361	0.04804196
G0:0071375	0.60516659	1.00000000	0.69984532	0.05642010	0.2812271	0.05246140
G0:0071392	0.47000721	0.69984532	1.00000000	0.04514144	0.2229238	0.04257122
G0:0090004	0.05134081	0.05642010	0.04514144	1.00000000	0.5860990	0.22952691
G0:0090200	0.25483610	0.28122706	0.22292384	0.58609898	1.0000000	0.43406582
G0:2000353	0.04804196	0.05246140	0.04257122	0.22952691	0.4340658	1.00000000
G0:2001238	0.24894993	0.27407573	0.21840651	0.24758423	0.8919121	0.42565095
G0:2001238						
G0:0001934	0.3682658					
G0:0002695	0.1635765					
G0:0007186	0.4534113					
G0:0007204	0.1338388					
G0:0008284	0.3689073					
G0:0008285	0.1816773					
G0:0045909	0.2534153					
G0:0010579	0.3097323					
G0:0010628	0.3390474					
G0:0010629	0.1659695					
G0:0014068	0.5461993					
G0:0030263	0.3686343					
G0:0030264	0.3609348					
G0:0030335	0.3426270					
G0:0030518	0.2933394					
G0:0030819	0.3103948					
G0:0032024	0.4470665					
G0:0032962	0.2599177					
G0:0043065	0.5926774					
G0:0043280	0.5220906					
G0:0045742	0.5278209					
G0:0045745	0.5162018					
G0:0045944	0.3641599					
G0:0050728	0.3694990					
G0:0050769	0.3230292					
G0:0051053	0.1574318					
G0:0051281	0.2828402					
G0:0070374	0.5746394					
G0:0070474	0.2382139					

G0:0071333 0.2375087
 G0:0071356 0.2489499
 G0:0071375 0.2740757
 G0:0071392 0.2184065
 G0:0090004 0.2475842
 G0:0090200 0.8919121
 G0:2000353 0.4256510
 G0:2001238 1.0000000

optimal assignment:

	G0:0001934	G0:0002695	G0:0007186	G0:0007204	G0:0008284	G0:0008285
G0:0001934	1	0	0	0	0	0
G0:0002695	0	1	0	0	0	0
G0:0007186	0	0	1	0	0	0
G0:0007204	0	0	0	1	0	0
G0:0008284	0	0	0	0	1	0
G0:0008285	0	0	0	0	0	1
G0:0045909	0	0	0	0	0	0
G0:0010579	0	0	0	0	0	0
G0:0010628	0	0	0	0	0	0
G0:0010629	0	0	0	0	0	0
G0:0014068	0	0	0	0	0	0
G0:0030263	0	0	0	0	0	0
G0:0030264	0	0	0	0	0	0
G0:0030335	0	0	0	0	0	0
G0:0030518	0	0	0	0	0	0
G0:0030819	0	0	0	0	0	0
G0:0032024	0	0	0	0	0	0
G0:0032962	0	0	0	0	0	0
G0:0043065	0	0	0	0	0	0
G0:0043280	0	0	0	0	0	0
G0:0045742	0	0	0	0	0	0
G0:0045745	0	0	0	0	0	0
G0:0045944	0	0	0	0	0	0
G0:0050728	0	0	0	0	0	0
G0:0050769	0	0	0	0	0	0
G0:0051053	0	0	0	0	0	0
G0:0051281	0	0	0	0	0	0
G0:0070374	0	0	0	0	0	0
G0:0070474	0	0	0	0	0	0
G0:0071333	0	0	0	0	0	0
G0:0071356	0	0	0	0	0	0
G0:0071375	0	0	0	0	0	0

G0:0071392	0	0	0	0	0	0
G0:0090004	0	0	0	0	0	0
G0:0090200	0	0	0	0	0	0
G0:2000353	0	0	0	0	0	0
G0:2001238	0	0	0	0	0	0
G0:0045909 G0:0010579 G0:0010628 G0:0010629 G0:0014068 G0:0030263						
G0:0001934	0	0	0	0	0	0
G0:0002695	0	0	0	0	0	0
G0:0007186	0	0	0	0	0	0
G0:0007204	0	0	0	0	0	0
G0:0008284	0	0	0	0	0	0
G0:0008285	0	0	0	0	0	0
G0:0045909	1	0	0	0	0	0
G0:0010579	0	1	0	0	0	0
G0:0010628	0	0	1	0	0	0
G0:0010629	0	0	0	1	0	0
G0:0014068	0	0	0	0	1	0
G0:0030263	0	0	0	0	0	1
G0:0030264	0	0	0	0	0	0
G0:0030335	0	0	0	0	0	0
G0:0030518	0	0	0	0	0	0
G0:0030819	0	0	0	0	0	0
G0:0032024	0	0	0	0	0	0
G0:0032962	0	0	0	0	0	0
G0:0043065	0	0	0	0	0	0
G0:0043280	0	0	0	0	0	0
G0:0045742	0	0	0	0	0	0
G0:0045745	0	0	0	0	0	0
G0:0045944	0	0	0	0	0	0
G0:0050728	0	0	0	0	0	0
G0:0050769	0	0	0	0	0	0
G0:0051053	0	0	0	0	0	0
G0:0051281	0	0	0	0	0	0
G0:0070374	0	0	0	0	0	0
G0:0070474	0	0	0	0	0	0
G0:0071333	0	0	0	0	0	0
G0:0071356	0	0	0	0	0	0
G0:0071375	0	0	0	0	0	0
G0:0071392	0	0	0	0	0	0
G0:0090004	0	0	0	0	0	0
G0:0090200	0	0	0	0	0	0
G0:2000353	0	0	0	0	0	0

G0:2001238	0	0	0	0	0	0
	G0:0030264	G0:0030335	G0:0030518	G0:0030819	G0:0032024	G0:0032962
G0:0001934	0	0	0	0	0	0
G0:0002695	0	0	0	0	0	0
G0:0007186	0	0	0	0	0	0
G0:0007204	0	0	0	0	0	0
G0:0008284	0	0	0	0	0	0
G0:0008285	0	0	0	0	0	0
G0:0045909	0	0	0	0	0	0
G0:0010579	0	0	0	0	0	0
G0:0010628	0	0	0	0	0	0
G0:0010629	0	0	0	0	0	0
G0:0014068	0	0	0	0	0	0
G0:0030263	0	0	0	0	0	0
G0:0030264	1	0	0	0	0	0
G0:0030335	0	1	0	0	0	0
G0:0030518	0	0	1	0	0	0
G0:0030819	0	0	0	0	0	0
G0:0032024	0	0	0	0	1	0
G0:0032962	0	0	0	0	0	1
G0:0043065	0	0	0	0	0	0
G0:0043280	0	0	0	0	0	0
G0:0045742	0	0	0	0	0	0
G0:0045745	0	0	0	0	0	0
G0:0045944	0	0	0	1	0	0
G0:0050728	0	0	0	0	0	0
G0:0050769	0	0	0	0	0	0
G0:0051053	0	0	0	0	0	0
G0:0051281	0	0	0	0	0	0
G0:0070374	0	0	0	0	0	0
G0:0070474	0	0	0	0	0	0
G0:0071333	0	0	0	0	0	0
G0:0071356	0	0	0	0	0	0
G0:0071375	0	0	0	0	0	0
G0:0071392	0	0	0	0	0	0
G0:0090004	0	0	0	0	0	0
G0:0090200	0	0	0	0	0	0
G0:2000353	0	0	0	0	0	0
G0:2001238	0	0	0	0	0	0
	G0:0043065	G0:0043280	G0:0045742	G0:0045745	G0:0045944	G0:0050728
G0:0001934	0	0	0	0	0	0
G0:0002695	0	0	0	0	0	0

G0:0007186	0	0	0	0	0	0
G0:0007204	0	0	0	0	0	0
G0:0008284	0	0	0	0	0	0
G0:0008285	0	0	0	0	0	0
G0:0045909	0	0	0	0	0	0
G0:0010579	0	0	0	0	0	0
G0:0010628	0	0	0	0	0	0
G0:0010629	0	0	0	0	0	0
G0:0014068	0	0	0	0	0	0
G0:0030263	0	0	0	0	0	0
G0:0030264	0	0	0	0	0	0
G0:0030335	0	0	0	0	0	0
G0:0030518	0	0	0	0	0	0
G0:0030819	0	0	0	0	1	0
G0:0032024	0	0	0	0	0	0
G0:0032962	0	0	0	0	0	0
G0:0043065	1	0	0	0	0	0
G0:0043280	0	1	0	0	0	0
G0:0045742	0	0	1	0	0	0
G0:0045745	0	0	0	1	0	0
G0:0045944	0	0	0	0	0	0
G0:0050728	0	0	0	0	0	1
G0:0050769	0	0	0	0	0	0
G0:0051053	0	0	0	0	0	0
G0:0051281	0	0	0	0	0	0
G0:0070374	0	0	0	0	0	0
G0:0070474	0	0	0	0	0	0
G0:0071333	0	0	0	0	0	0
G0:0071356	0	0	0	0	0	0
G0:0071375	0	0	0	0	0	0
G0:0071392	0	0	0	0	0	0
G0:0090004	0	0	0	0	0	0
G0:0090200	0	0	0	0	0	0
G0:2000353	0	0	0	0	0	0
G0:2001238	0	0	0	0	0	0
G0:0050769 G0:0051053 G0:0051281 G0:0070374 G0:0070474 G0:0071333						
G0:0001934	0	0	0	0	0	0
G0:0002695	0	0	0	0	0	0
G0:0007186	0	0	0	0	0	0
G0:0007204	0	0	0	0	0	0
G0:0008284	0	0	0	0	0	0
G0:0008285	0	0	0	0	0	0

G0:0045909	0	0	0	0	0	0
G0:0010579	0	0	0	0	0	0
G0:0010628	0	0	0	0	0	0
G0:0010629	0	0	0	0	0	0
G0:0014068	0	0	0	0	0	0
G0:0030263	0	0	0	0	0	0
G0:0030264	0	0	0	0	0	0
G0:0030335	0	0	0	0	0	0
G0:0030518	0	0	0	0	0	0
G0:0030819	0	0	0	0	0	0
G0:0032024	0	0	0	0	0	0
G0:0032962	0	0	0	0	0	0
G0:0043065	0	0	0	0	0	0
G0:0043280	0	0	0	0	0	0
G0:0045742	0	0	0	0	0	0
G0:0045745	0	0	0	0	0	0
G0:0045944	0	0	0	0	0	0
G0:0050728	0	0	0	0	0	0
G0:0050769	1	0	0	0	0	0
G0:0051053	0	1	0	0	0	0
G0:0051281	0	0	1	0	0	0
G0:0070374	0	0	0	1	0	0
G0:0070474	0	0	0	0	1	0
G0:0071333	0	0	0	0	0	1
G0:0071356	0	0	0	0	0	0
G0:0071375	0	0	0	0	0	0
G0:0071392	0	0	0	0	0	0
G0:0090004	0	0	0	0	0	0
G0:0090200	0	0	0	0	0	0
G0:2000353	0	0	0	0	0	0
G0:2001238	0	0	0	0	0	0
G0:0071356 G0:0071375 G0:0071392 G0:0090004 G0:0090200 G0:2000353						
G0:0001934	0	0	0	0	0	0
G0:0002695	0	0	0	0	0	0
G0:0007186	0	0	0	0	0	0
G0:0007204	0	0	0	0	0	0
G0:0008284	0	0	0	0	0	0
G0:0008285	0	0	0	0	0	0
G0:0045909	0	0	0	0	0	0
G0:0010579	0	0	0	0	0	0
G0:0010628	0	0	0	0	0	0
G0:0010629	0	0	0	0	0	0

G0:0014068	0	0	0	0	0	0
G0:0030263	0	0	0	0	0	0
G0:0030264	0	0	0	0	0	0
G0:0030335	0	0	0	0	0	0
G0:0030518	0	0	0	0	0	0
G0:0030819	0	0	0	0	0	0
G0:0032024	0	0	0	0	0	0
G0:0032962	0	0	0	0	0	0
G0:0043065	0	0	0	0	0	0
G0:0043280	0	0	0	0	0	0
G0:0045742	0	0	0	0	0	0
G0:0045745	0	0	0	0	0	0
G0:0045944	0	0	0	0	0	0
G0:0050728	0	0	0	0	0	0
G0:0050769	0	0	0	0	0	0
G0:0051053	0	0	0	0	0	0
G0:0051281	0	0	0	0	0	0
G0:0070374	0	0	0	0	0	0
G0:0070474	0	0	0	0	0	0
G0:0071333	0	0	0	0	0	0
G0:0071356	1	0	0	0	0	0
G0:0071375	0	1	0	0	0	0
G0:0071392	0	0	1	0	0	0
G0:0090004	0	0	0	1	0	0
G0:0090200	0	0	0	0	1	0
G0:2000353	0	0	0	0	0	1
G0:2001238	0	0	0	0	0	0
G0:2001238						
G0:0001934	0					
G0:0002695	0					
G0:0007186	0					
G0:0007204	0					
G0:0008284	0					
G0:0008285	0					
G0:0045909	0					
G0:0010579	0					
G0:0010628	0					
G0:0010629	0					
G0:0014068	0					
G0:0030263	0					
G0:0030264	0					
G0:0030335	0					

G0:0030518	0
G0:0030819	0
G0:0032024	0
G0:0032962	0
G0:0043065	0
G0:0043280	0
G0:0045742	0
G0:0045745	0
G0:0045944	0
G0:0050728	0
G0:0050769	0
G0:0051053	0
G0:0051281	0
G0:0070374	0
G0:0070474	0
G0:0071333	0
G0:0071356	0
G0:0071375	0
G0:0071392	0
G0:0090004	0
G0:0090200	0
G0:2000353	0
G0:2001238	1

=====

term similarity matrix:

	G0:0001934	G0:0002695	G0:0007186	G0:0007204	G0:0008284	G0:0008285
G0:0006874	0.18956951	0.16111130	0.1924032	0.96327998	0.1899935	0.18383542
G0:0006979	0.00000000	0.00000000	0.2776981	0.00000000	0.00000000	0.00000000
G0:0007166	0.32192373	0.26000870	0.7910622	0.21665702	0.3228951	0.30893321
G0:0007267	0.09322283	0.07707063	0.4773693	0.08050070	0.0934703	0.08989595
G0:0007566	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
G0:0007584	0.00000000	0.00000000	0.2770009	0.00000000	0.00000000	0.00000000
G0:0030968	0.22401252	0.19216964	0.4049366	0.15808124	0.2244824	0.21764415
G0:0033280	0.00000000	0.00000000	0.2299946	0.00000000	0.00000000	0.00000000
G0:0040015	0.15268665	0.32216121	0.1543554	0.12676207	0.1529372	0.35578731
G0:0043434	0.00000000	0.00000000	0.2934024	0.00000000	0.00000000	0.00000000
G0:0046697	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
G0:0071456	0.06712798	0.05832589	0.3237220	0.06026934	0.0672562	0.06538553
G0:2001256	0.14034178	0.12539701	0.1417504	0.11735835	0.1405534	0.13744912
	G0:0045909	G0:0010579	G0:0010628	G0:0010629	G0:0014068	G0:0030263
G0:0006874	0.4229081	0.1204615	0.2000490	0.1954228	0.1470025	0.05118915
G0:0006979	0.00000000	0.1747083	0.00000000	0.00000000	0.2128199	0.00000000

G0:0007166	0.1881854	0.4372131	0.3017171	0.2922087	0.4127581	0.06154269
G0:0007267	0.0000000	0.2790705	0.0000000	0.0000000	0.3490917	0.05702883
G0:0007566	0.1968291	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000
G0:0007584	0.0000000	0.1744320	0.0000000	0.0000000	0.2124102	0.00000000
G0:0030968	0.1455048	0.2596200	0.2052091	0.2007658	0.3140368	0.04907904
G0:0033280	0.0000000	0.1545423	0.0000000	0.0000000	0.1836310	0.00000000
G0:0040015	0.3546344	0.1074404	0.1587928	0.3721151	0.1259208	0.00000000
G0:0043434	0.0000000	0.1807964	0.0000000	0.0000000	0.2219232	0.00000000
G0:0046697	0.1879061	0.0000000	0.0000000	0.0000000	0.0000000	0.00000000
G0:0071456	0.0000000	0.2145087	0.0000000	0.0000000	0.2562529	0.04607250
G0:2001256	0.1126994	0.1011779	0.1454839	0.1432364	0.1174040	0.00000000
	G0:0030264	G0:0030335	G0:0030518	G0:0030819	G0:0032024	G0:0032962
G0:0006874	0.04992759	0.17296137	0.1623026	0.15297184	0.45042420	0.12350529
G0:0006979	0.00000000	0.00000000	0.2347282	0.00000000	0.00000000	0.00000000
G0:0007166	0.05972824	0.28502725	0.4679114	0.24341527	0.38679808	0.18704318
G0:0007267	0.05546740	0.08368833	0.3910747	0.07261140	0.69207836	0.05702883
G0:0007566	0.00000000	0.00000000	0.0000000	0.00000000	0.00000000	0.00000000
G0:0007584	0.00000000	0.00000000	0.2342299	0.00000000	0.00000000	0.00000000
G0:0030968	0.04791817	0.20550140	0.3449739	0.18295194	0.29364399	0.14916313
G0:0033280	0.00000000	0.00000000	0.1997148	0.00000000	0.00000000	0.00000000
G0:0040015	0.00000000	0.14263294	0.1359240	0.12987743	0.12663682	0.10963679
G0:0043434	0.00000000	0.00000000	0.2458512	0.00000000	0.00000000	0.00000000
G0:0046697	0.00000000	0.00000000	0.0000000	0.00000000	0.00000000	0.00000000
G0:0071456	0.04504801	0.06203847	0.2795270	0.05573554	0.05416098	0.04607250
G0:2001256	0.00000000	0.45733181	0.1260533	0.12083618	0.45377292	0.10312334
	G0:0043065	G0:0043280	G0:0045742	G0:0045745	G0:0045944	G0:0050728
G0:0006874	0.18904533	0.16111130	0.1410142	0.1372734	0.18686592	0.1533251
G0:0006979	0.00000000	0.00000000	0.2042329	0.1988652	0.00000000	0.4135955
G0:0007166	0.32072488	0.26000870	0.5295439	0.5122653	0.31576488	0.2833794
G0:0007267	0.09291715	0.07707063	0.3329883	0.3230203	0.09164957	0.00000000
G0:0007566	0.00000000	0.00000000	0.0000000	0.0000000	0.00000000	0.00000000
G0:0007584	0.00000000	0.00000000	0.2038555	0.1985074	0.00000000	0.6009321
G0:0030968	0.22343135	0.19216964	0.3018427	0.2942007	0.22101285	0.3967368
G0:0033280	0.00000000	0.00000000	0.1772024	0.1731474	0.00000000	0.5165396
G0:0040015	0.15237651	0.13516123	0.1218802	0.1193189	0.15108208	0.3101200
G0:0043434	0.00000000	0.00000000	0.2126020	0.2067916	0.00000000	0.2317917
G0:0046697	0.00000000	0.00000000	0.0000000	0.0000000	0.00000000	0.00000000
G0:0071456	0.06696934	0.05832589	0.2469888	0.2411566	0.06630835	0.3672759
G0:2001256	0.14007972	0.12539701	0.1138838	0.1116445	0.13898504	0.1210370
	G0:0050769	G0:0051053	G0:0051281	G0:0070374	G0:0070474	G0:0071333
G0:0006874	0.16069324	0.15366975	0.78669393	0.1564457	0.1290810	0.70344634
G0:0006979	0.00000000	0.00000000	0.00000000	0.2263471	0.00000000	0.21241474

G0:0007166	0.25914495	0.24481983	0.21144708	0.4464537	0.1718473	0.36189434
G0:0007267	0.07683990	0.07299105	0.06385840	0.3748591	0.0000000	0.06923274
G0:0007566	0.28241049	0.00000000	0.00000000	0.0000000	0.1844556	0.00000000
G0:0007584	0.00000000	0.00000000	0.00000000	0.2258837	0.0000000	0.39781989
G0:0030968	0.19169741	0.18374426	0.16428383	0.3331681	0.1355411	0.53917361
G0:0033280	0.00000000	0.00000000	0.00000000	0.1936150	0.0000000	0.40847560
G0:0040015	0.39691205	0.31065885	0.28332785	0.1321480	0.3342800	0.11459004
G0:0043434	0.00000000	0.00000000	0.00000000	0.2366725	0.0000000	0.49348504
G0:0046697	0.29190372	0.00000000	0.00000000	0.0000000	0.1765969	0.00000000
G0:0071456	0.05819365	0.05595895	0.05042972	0.2706841	0.0000000	0.45733281
G0:2001256	0.12516602	0.12123256	0.63522824	0.1227992	0.1066283	0.10685043
	G0:0071356	G0:0071375	G0:0071392	G0:0090004	G0:0090200	G0:2000353
G0:0006874	0.06442044	0.07262417	0.05495129	0.13518396	0.1414616	0.12575988
G0:0006979	0.22489114	0.25319268	0.19213013	0.00000000	0.2048747	0.00000000
G0:0007166	0.38914493	0.45423752	0.31933754	0.20872197	0.3934867	0.19116349
G0:0007267	0.07395014	0.08496810	0.06173779	0.06310219	0.3341852	0.05819108
G0:0007566	0.00000000	0.00000000	0.00000000	0.00000000	0.0000000	0.00000000
G0:0007584	0.42113872	0.47401578	0.35989601	0.00000000	0.2044950	0.00000000
G0:0030968	0.56953487	0.63789352	0.48951240	0.16263407	0.3027555	0.15177190
G0:0033280	0.42901743	0.47442112	0.82129547	0.00000000	0.1776853	0.00000000
G0:0040015	0.00000000	0.00000000	0.00000000	0.11787561	0.1221846	0.11125041
G0:0043434	0.52378374	0.97040777	0.62086240	0.00000000	0.2132975	0.00000000
G0:0046697	0.00000000	0.00000000	0.00000000	0.00000000	0.0000000	0.00000000
G0:0071456	0.48115018	0.53408293	0.41795952	0.04995693	0.2476841	0.04682810
G0:2001256	0.00000000	0.00000000	0.00000000	0.38299885	0.1141495	0.10454967
	G0:2001238					
G0:0006874	0.1378898					
G0:0006979	0.1997499					
G0:0007166	0.5150975					
G0:0007267	0.3246580					
G0:0007566	0.0000000					
G0:0007584	0.1993888					
G0:0030968	0.2954612					
G0:0033280	0.1738176					
G0:0040015	0.1197429					
G0:0043434	0.2077483					
G0:0046697	0.0000000					
G0:0071456	0.2421200					
G0:2001256	0.1120156					
optimal assignment:						
	G0:0006874	G0:0006979	G0:0007166	G0:0007267	G0:0007566	G0:0007584
G0:0001934	0	0	0	1	0	0

G0:0002695	0	0	0	0	0	0
G0:0007186	0	0	1	0	0	0
G0:0007204	0	0	0	0	0	0
G0:0008284	0	0	0	0	0	0
G0:0008285	0	0	0	0	0	0
G0:0045909	0	0	0	0	0	0
G0:0010579	0	0	0	0	0	0
G0:0010628	0	0	0	0	0	0
G0:0010629	0	0	0	0	0	0
G0:0014068	0	0	0	0	0	0
G0:0030263	0	0	0	0	0	0
G0:0030264	0	0	0	0	0	0
G0:0030335	0	0	0	0	0	0
G0:0030518	0	0	0	0	0	0
G0:0030819	0	0	0	0	0	0
G0:0032024	0	0	0	0	0	0
G0:0032962	0	0	0	0	0	0
G0:0043065	0	0	0	0	0	0
G0:0043280	0	0	0	0	0	0
G0:0045742	0	0	0	0	0	0
G0:0045745	0	0	0	0	0	0
G0:0045944	0	0	0	0	0	0
G0:0050728	1	0	0	0	0	0
G0:0050769	0	1	0	0	0	0
G0:0051053	0	0	0	0	1	0
G0:0051281	0	0	0	0	0	1
G0:0070374	0	0	0	0	0	0
G0:0070474	0	0	0	0	0	0
G0:0071333	0	0	0	0	0	0
G0:0071356	0	0	0	0	0	0
G0:0071375	0	0	0	0	0	0
G0:0071392	0	0	0	0	0	0
G0:0090004	0	0	0	0	0	0
G0:0090200	0	0	0	0	0	0
G0:2000353	0	0	0	0	0	0
G0:2001238	0	0	0	0	0	0
G0:0030968						
G0:0033280						
G0:0040015						
G0:0043434						
G0:0046697						
G0:0071456						
G0:0001934	0	0	0	0	0	0
G0:0002695	0	0	0	0	0	0
G0:0007186	0	0	0	0	0	0
G0:0007204	0	0	0	0	0	0
G0:0008284	1	0	0	0	0	0

G0:0008285	0	0	0	0	0	0
G0:0045909	0	0	0	0	0	0
G0:0010579	0	0	0	0	0	0
G0:0010628	0	0	0	1	0	0
G0:0010629	0	0	0	0	0	0
G0:0014068	0	0	0	0	0	0
G0:0030263	0	0	0	0	0	0
G0:0030264	0	0	0	0	0	0
G0:0030335	0	0	0	0	0	1
G0:0030518	0	0	0	0	0	0
G0:0030819	0	0	0	0	0	0
G0:0032024	0	0	0	0	0	0
G0:0032962	0	0	0	0	0	0
G0:0043065	0	0	0	0	0	0
G0:0043280	0	0	1	0	0	0
G0:0045742	0	0	0	0	0	0
G0:0045745	0	0	0	0	0	0
G0:0045944	0	0	0	0	0	0
G0:0050728	0	0	0	0	0	0
G0:0050769	0	0	0	0	0	0
G0:0051053	0	0	0	0	0	0
G0:0051281	0	0	0	0	0	0
G0:0070374	0	0	0	0	0	0
G0:0070474	0	0	0	0	0	0
G0:0071333	0	0	0	0	0	0
G0:0071356	0	1	0	0	0	0
G0:0071375	0	0	0	0	1	0
G0:0071392	0	0	0	0	0	0
G0:0090004	0	0	0	0	0	0
G0:0090200	0	0	0	0	0	0
G0:2000353	0	0	0	0	0	0
G0:2001238	0	0	0	0	0	0
G0:2001256						
G0:0001934	0					
G0:0002695	0					
G0:0007186	0					
G0:0007204	0					
G0:0008284	0					
G0:0008285	0					
G0:0045909	0					
G0:0010579	0					
G0:0010628	0					

```

GO:0010629      0
GO:0014068      0
GO:0030263      0
GO:0030264      0
GO:0030335      0
GO:0030518      0
GO:0030819      1
GO:0032024      0
GO:0032962      0
GO:0043065      0
GO:0043280      0
GO:0045742      0
GO:0045745      0
GO:0045944      0
GO:0050728      0
GO:0050769      0
GO:0051053      0
GO:0051281      0
GO:0070374      0
GO:0070474      0
GO:0071333      0
GO:0071356      0
GO:0071375      0
GO:0071392      0
GO:0090004      0
GO:0090200      0
GO:2000353      0
GO:2001238      0
=====
      8614      2852
8614 1.000000 0.344252
2852 0.344252 1.000000

```

Note the difference to a gene similarity that is just based on the maximum GO term similarity and to a gene similarity that is based on the average of best matching GO terms:

```

> getGeneSim(c("8614","2852"),similarity="max",similarityTerm="Lin",verbose=FALSE)

filtering out genes not mapping to the currently set GO category ... ==> list of 2
      8614      2852
8614 1.0000000 0.8751547
2852 0.8751547 1.0000000

```

```
> getGeneSim(c("8614", "2852"), similarity="funSimMax", similarityTerm="Lin", verbose=1)
filtering out genes not mapping to the currently set GO category ... ==> list of 2
      8614      2852
8614 1.0000000 0.6139785
2852 0.6139785 1.0000000
```

2.2.5 Gene Similarities In the Spirit of Hausdorff Metrics

Hausdorff metrics are a general concept for measuring distances between compact subsets of a metric space. Let X and Y be the two sets of GO terms associated to genes g and g' , and let $d(t, t')$ denote the distance between GO terms t and t' . Then the Hausdorff distance X and Y is defined as

$$d_{Hausdorff}(X, Y) = \max\{\sup_{t \in X} \inf_{t' \in Y} d(t, t'), \sup_{t' \in Y} \inf_{t \in X} d(t, t')\} \quad (11)$$

Using Hausdorff metrics for measuring gene functional distances was proposed in del Pozo et al. (2008). We translate the idea to define a similarity measure between g and g' (see the difference to previous GOSim versions):

$$sim_{gene}(g, g') = \exp(-d_{Hausdorff}(g, g')) \quad (12)$$

```
> getGeneSim(c("8614", "2852"), similarity="hausdorff", similarityTerm="Lin", verbose=1)
filtering out genes not mapping to the currently set GO category ... ==> list of 2
      8614 2852
8614    1    1
2852    1    1
```

2.2.6 Feature Space Embedding of Gene Products

The Simple Approach Mistry and Pavlidis (2008) proposed to represent each gene by a feature vector describing the presence/absence of all GO terms. The absence of each GO term is additionally weighted by its information content. In the feature space similarities arise as dot products. Hence, the similarity between two GO terms t and t' is implicitly defined as the product of their information content values, hence ignoring the exact DAG structure of the Gene Ontology as employed by the GO term similarity measures explained in the beginning of this document.

```
> getGeneSim(c("8614", "2852"), similarity="dot", method="Tanimoto", verbose=FALSE)
filtering out genes not mapping to the currently set GO category ... ==> list of 2
      8614 2852
8614    1 NaN
2852 NaN    1
```

This will calculate the Tanimoto coefficient between feature vectors as a similarity measure. It is possible to retrieve the feature vectors via:

```
> features = getGeneFeatures(c("8614","2852"))
```

filtering out genes not mapping to the currently set GO category ... ==> list of 2

Embeddings via GO Term Similarities to Prototype Genes This approach is due to Speer et al. (2005); Fröhlich et al. (2006). The idea is to define a feature vector for each gene by its pairwise GO term similarity to certain prototype genes, i.e. the prototype genes form a (nonorthogonal) basis, and each gene is defined relative to this basis. The prototype genes can either be defined a priori or one can use one of the heuristics implemented in the function `selectPrototypes`. The default behavior is to select the 250 best annotated genes, i.e. which have been annotated with GO terms most often, but here we just use 5 for computational reasons:

```
> proto = selectPrototypes(n=5,verbose=FALSE)
```

We now calculate for each gene g feature vectors $\phi(g)$ by using their similarity to all prototypes p_1, \dots, p_n :

$$\phi(g) = (sim'(g, p_1), \dots, sim'(g, p_n))^T \quad (13)$$

Here sim' by default is the maximum pairwise GO term similarity. Alternatively, one can use other similarity measures for sim' as well. These similarity measures can by itself again be combined with arbitrary GO term similarity concepts. The default is the Jiang-Conrath term similarity.

Because the feature vectors are very high-dimensional we usually perform a principal component analysis (PCA) to project the data into a lower dimensional subspace:

```
> PHI = getGeneFeaturesPrototypes(genes,prototypes=proto,verbose=FALSE)
```

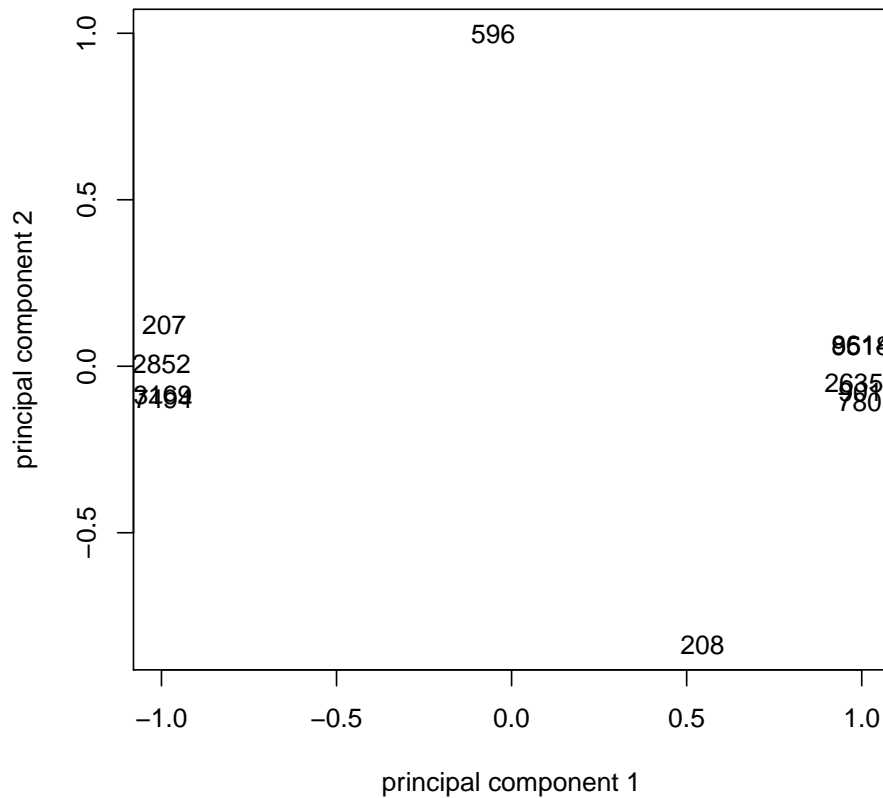
filtering out genes not mapping to the currently set GO category ... ==> list of 1

filtering out genes not mapping to the currently set GO category ... ==> list of 5

This uses the above define prototypes to calculate feature vectors and performs a PCA afterwards. The number of principal components is chosen such that at least 95% of the total variance in feature space can be explained (this is a relatively conservative criterion).

We can now plot our genes in the space spanned by the first 2 principal components to get an impression of the relative "position" of the genes to each other in the feature space (see Fig. ??). The feature vectors are normalized to Euclidian norm 1 by default:

```
> x=seq(min(PHI$features[,1]),max(PHI$features[,1]),length.out=100)
> y=seq(min(PHI$features[,2]),max(PHI$features[,2]),length.out=100)
> plot(x,y,xlab="principal component 1",ylab="principal component 2",type="n")
> text(PHI$features[,1],PHI$features[,2],labels=genes)
```



Finally, we can directly calculate the similarities of the genes to each other, this time using the Resnik's GO term similarity concept. These similarities may then be used to cluster genes with respect to their function:

```
> sim = getGeneSimPrototypes(genes,prototypes=proto,similarityTerm="Resnik",verbose=TRUE)
> h=hclust(as.dist(1-sim$similarity),"ward")
> plot(h,xlab="")
```

This produces a hierarchical clustering of all genes using Ward's method (see Fig. 1).

2.2.7 Combination of Similarities from Different Ontologies

It should be mentioned that up to now all similarity computations were performed within the ontology "biological process". One could imagine to combine functional similarities between gene products with regard to different taxonomies. An obvious way for doing so would be to consider the sum of the respective similarities:

$$sim_{total}(g, g') = sim_{Ontology1}(g, g') + sim_{Ontology2}(g, g') \quad (14)$$

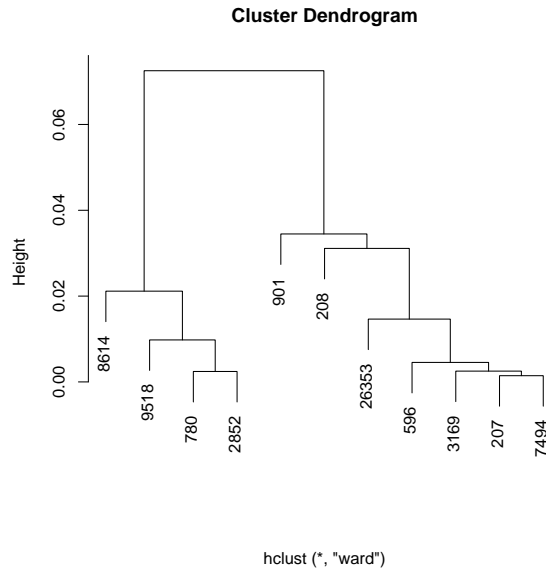


Figure 1: Possible functional clustering of the genes using Ward's method.

Of course, one could also use a weighted averaging scheme here, if desired.

2.3 Cluster Evaluations

GOSim has the possibility to evaluate a given clustering of genes or terms by means of their GO similarities. Supposed, based on other experiments (e.g. microarray), we have decided to put genes "8614", "9518", "780", "2852" in one group, genes "3169", "207", "7494", "596" in a second and the rest in a third group. Then we can ask ourselves, how similar these groups are with respect to their GO annotations:

```
> ev = evaluateClustering(c(2,3,2,3,1,2,1,1,3,1,2), sim$similarity)
> plot(ev$clustersil,main="")
```

A good indication of the clustering quality can be obtained by looking at the cluster silhouettes (Rousseeuw, 1987) (see Fig. 2). This shows that clusters 1 and 2 are relatively homogenous with respect to the functional similarity of the genes contained in it, while the genes in cluster 3 are more dissimilar.

2.4 GO Enrichment Analysis

Since version 1.1 *GOSim* also offers the possibility of a GO enrichment analysis. Suppose, we may now want to get a clearer picture of the genes involved in cluster 1. For this purpose we use the topGO tool (Alexa et al., 2006).

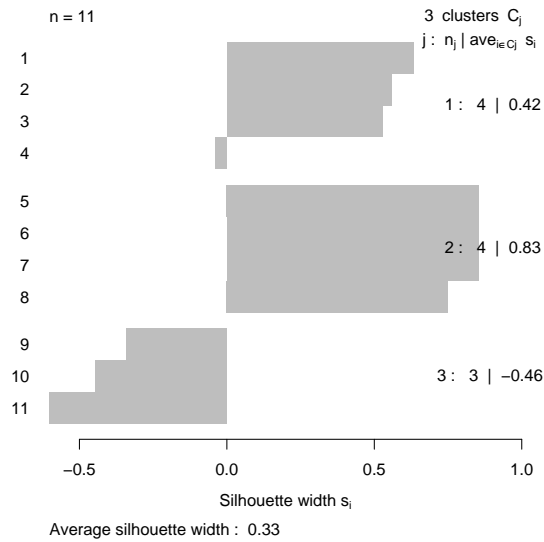


Figure 2: Silhouette plot of a possible given grouping of genes.

```
> library(org.Hs.eg.db)
> library(topGO)

groupGOTerms:          GOBPterm, GOMFterm, GOCCTerm environments built.

> allgenes = union(c("8614", "9518", "780", "2852"), sample(keys(org.Hs.egGO), 1000))
> GOenrichment(c("8614", "9518", "780", "2852"), allgenes) # print out what clusters

Building most specific GOs .....      ( 1220 GO terms found. )

Build GO DAG topology .....          ( 3494 GO terms and 7596 relations. )

Annotating nodes .....                ( 324 genes annotated to the GO terms. )

-- Elim Algorithm --

the algorithm is scoring 624 nontrivial nodes
parameters:
  test statistic: fisher
  cutOff: 0.01

Level 16:          2 nodes to be scored          (0 eliminated genes)
```

Level 15:	4 nodes to be scored	(0 eliminated genes)
Level 14:	4 nodes to be scored	(0 eliminated genes)
Level 13:	5 nodes to be scored	(0 eliminated genes)
Level 12:	12 nodes to be scored	(0 eliminated genes)
Level 11:	28 nodes to be scored	(0 eliminated genes)
Level 10:	42 nodes to be scored	(5 eliminated genes)
Level 9:	57 nodes to be scored	(10 eliminated genes)
Level 8:	71 nodes to be scored	(10 eliminated genes)
Level 7:	91 nodes to be scored	(10 eliminated genes)
Level 6:	95 nodes to be scored	(35 eliminated genes)
Level 5:	102 nodes to be scored	(54 eliminated genes)
Level 4:	66 nodes to be scored	(74 eliminated genes)
Level 3:	28 nodes to be scored	(87 eliminated genes)
Level 2:	16 nodes to be scored	(87 eliminated genes)
Level 1:	1 nodes to be scored	(87 eliminated genes)

\$GOTerms

	go_id	Term
16505	GO:0006874	cellular calcium ion homeostasis
16996	GO:0007167	enzyme linked receptor protein signaling pathway
17259	GO:0007267	cell-cell signaling
17671	GO:0007566	embryo implantation
18469	GO:0008285	negative regulation of cell proliferation
24848	GO:0014070	response to organic cyclic compound
33138	GO:0022411	cellular component disassembly
50942	GO:0043434	response to peptide hormone stimulus
64631	GO:0048609	multicellular organismal reproductive process
71264	GO:0051924	regulation of calcium ion transport
78481	GO:0071310	cellular response to organic substance

82851 GO:0097305

response to alcohol

16505

16996

Any series of molecular signals initiated by the binding of an extra

17259

17671

18469

24848

33138

50942 Any process that results in a change in state or activity of a cell or an orga

64631

71264

78481

82851

\$p.values

GO:0048609	GO:0007267	GO:0071310	GO:0014070	GO:0008285	GO:0022411
0.0085413037	0.0039531761	0.0043128952	0.0085413037	0.0072578620	0.0060737836
GO:0007566	GO:0051924	GO:0097305	GO:0007167	GO:0043434	GO:0006874
0.0003425728	0.0011324466	0.0023583590	0.0015538816	0.0049904262	0.0049904262

\$genes

\$genes\$`GO:0048609`

[1]	"122042"	"140801"	"1738"	"2560"	"4086"	"64067"	"644890"	"6667"
[9]	"7042"	"780"	"8614"	"881"	"9825"			

\$genes\$`GO:0007267`

[1]	"1081"	"10947"	"1131"	"1142"	"1201"	"1230"	"1813"	"2560"
[9]	"2703"	"273"	"2852"	"292"	"29993"	"348980"	"356"	"389692"
[17]	"4082"	"5208"	"5376"	"54207"	"54899"	"56659"	"5715"	"6310"
[25]	"7042"	"7474"	"796"	"84896"	"8614"	"8839"	"9402"	"9518"
[33]	"952"	"9758"						

\$genes\$`GO:0071310`

[1]	"1053"	"1081"	"1230"	"1439"	"16"	"1646"	"203100"	"23318"
[9]	"27316"	"2852"	"3122"	"3454"	"4086"	"466"	"4864"	"51094"
[17]	"5167"	"527"	"534"	"5770"	"64397"	"6667"	"6773"	"7041"
[25]	"7042"	"7132"	"7474"	"8021"	"842"	"8614"	"9125"	"9365"
[33]	"9459"	"9518"	"983"					

\$genes\$`GO:0014070`

[1]	"1813"	"23318"	"2560"	"2852"	"4086"	"4864"	"7042"	"7132"	"7832"
-----	--------	---------	--------	--------	--------	--------	--------	--------	--------

```

[10] "842"    "8614"   "952"    "983"

$genes$`G0:0008285`
[1] "11285"  "122042" "1813"   "2852"   "4086"   "5376"   "6801"   "7041"
[9] "7042"   "7474"   "780"    "7832"

$genes$`G0:0022411`
[1] "1299"   "1509"   "196528" "2852"   "51119"  "6144"   "6171"   "6227"
[9] "780"    "8021"   "983"

$genes$`G0:0007566`
[1] "780"    "796"    "8614"

$genes$`G0:0051924`
[1] "1230"   "1813"   "2852"   "796"    "8614"

$genes$`G0:0097305`
[1] "1813"   "2852"   "7132"   "842"    "8614"   "952"    "983"

$genes$`G0:0007167`
[1] "203100" "2852"   "356"    "4086"   "466"    "5167"   "527"    "534"
[9] "5770"   "6310"   "64397"  "6667"   "7041"   "7042"   "7474"   "780"
[17] "79923"  "819"    "84159"  "842"    "9365"   "9459"   "9518"   "983"
[25] "9846"

$genes$`G0:0043434`
[1] "2852"   "5167"   "527"    "534"    "5770"   "64397"  "7832"   "8614"   "9365"
[10] "983"

$genes$`G0:0006874`
[1] "1201"   "1230"   "1813"   "2587"   "2852"   "6915"   "778"    "796"    "8614"   "952"

```

References

- Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, 22(13):1600 – 1607.
- Cortes, C. and Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20:273 – 297.
- Couto, F., Silva, M., and Coutinho, P. (2003). Implementation of a Functional Semantic

- Similarity Measure between Gene-Products. Technical Report DI/FCUL TR 03–29, Department of Informatics, University of Lisbon.
- Couto, F., Silva, M., and Coutinho, P. (2005). Semantic Similarity over the Gene Ontology: Family Correlation and Selecting Disjunctive Ancestors. In *Conference in Information and Knowledge Management*.
- del Pozo, A., Pazos, F., and Valencia, A. (2008). Defining functional distances over gene ontology. *BMC Bioinformatics*, 9:50.
- Fröhlich, H., Speer, N., Poustka, A., and Beissbarth, T. (2007). GOSim - An R-Package for Computation of Information Theoretic GO Similarities Between Terms and Gene Products. *BMC Bioinformatics*, 8:166.
- Fröhlich, H., Speer, N., and Zell, A. (2006). Kernel based functional gene grouping. In *Proc. Int. Joint Conf. Neural Networks*, pages 6886 – 6891.
- Jiang, J. and Conrath, D. (1998). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of the International Conference on Research in Computational Linguistics*, Taiwan.
- Lerman, G. and Shakhnovich, B. E. (2007). Defining functional distance using manifold embeddings of gene ontology annotations. *Proc Natl Acad Sci U S A*, 104(27):11334–11339.
- Lin, D. (1998). An information-theoretic definition of similarity. In Kaufmann, M., editor, *Proceedings of the 15th International Conference on Machine Learning*, volume 1, pages 296–304, San Francisco, CA.
- Lord, P., Stevens, R., Brass, A., and Goble, C. (2003). Semantic similarity measures as tools for exploring the gene ontology. In *Proceedings of the Pacific Symposium on Biocomputing*, pages 601–612.
- Mehlhorn, K. and Näher, S. (1999). *The LEDA Platform of Combinatorial and Geometric Computing*. Cambridge University Press.
- Mistry, M. and Pavlidis, P. (2008). Gene ontology term overlap as a measure of gene functional similarity. *BMC Bioinformatics*, 9:327.
- Pesquita, C., Faria, D., Bastos, H., Falcao, A., and Couto, F. (2007). Evaluating go-based semantic similarity measures. In *Proc. 10th Annual Bio-Ontologies Meeting 2007*, volume 2007, pages 37 – 40.
- Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, volume 1, pages 448–453, Montreal.

- Resnik, P. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11:95–130.
- Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comp. and Applied Mathematics*, 20:53–65.
- Schlicker, A., Domingues, F. S., Rahnenführer, J., and Lengauer, T. (2006). A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*, 7:302.
- Schölkopf, B. and Smola, A. J. (2002). *Learning with Kernels*. MIT Press, Cambridge, MA.
- Speer, N., Fröhlich, H., Spieth, C., and Zell, A. (2005). Functional grouping of genes using spectral clustering and gene ontology. In *Proc. Int. Joint Conf. Neural Networks*, pages 298 – 303.
- The Gene Ontology Consortium (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32:D258–D261.