Semantic Similarity in the Gene Ontology

Catia Pesquita

Abstract

Gene Ontology-based semantic similarity (SS) allows the comparison of GO terms or entities annotated with GO terms, by leveraging on the ontology structure and properties and on annotation corpora. In the last decade the number and diversity of SS measures based on GO has grown considerably, and their application ranges from functional coherence evaluation, protein interaction prediction, and disease gene prioritization.

Understanding how SS measures work, what issues can affect their performance and how they compare to each other in different evaluation settings is crucial to gain a comprehensive view of this area and choose the most appropriate approaches for a given application.

In this chapter, we provide a guide to understanding and selecting SS measures for biomedical researchers. We present a straightforward categorization of SS measures and describe the main strategies they employ. We discuss the intrinsic and external issues that affect their performance, and how these can be addressed. We summarize comparative assessment studies, highlighting the top measures in different settings, and compare different implementation strategies and their use. Finally, we discuss some of the extant challenges and opportunities, namely the increased semantic complexity of GO and the need for fast and efficient computation, pointing the way towards the future generation of SS measures.

Key words Gene ontology, Semantic similarity, Functional similarity, Protein similarity

1 Introduction

The graph structure of the Gene Ontology (GO) allows the comparison of GO terms and GO-annotated gene products by semantic similarity. Assessing similarity is crucial to expanding knowledge, because it allows us to categorize objects into kinds. Similar objects tend to behave similarly, which supports inference, a crucial task to support many applications including identifying protein–protein interactions [1], suggesting candidate genes involved in diseases [2] and evaluating the functional coherence of gene sets [3, 4].

Semantic similarity (SS) assesses the likeness in meaning of two concepts. It has been a subject of interest to Artificial Intelligence, Cognitive Science, and Psychology for the last few decades, and an important tool for Natural Language Processing. It has been used

in this context to perform word sense disambiguation, determining discourse structure, text summarization and annotation, information extraction and retrieval, automatic indexing, lexical selection, and automatic correction of word errors in text [5].

Sometimes, research literature uses SS, relatedness, and distance as interchangeable terms, but they are in fact not identical. Semantic relatedness makes use of various relations between two concepts (i.e., hyponymic, hypernymic, meronymic, antonymic, and any kind of functional relations including has-part, is-made-of, and is-an-attribute-of). SS is more limited since it usually only makes use of hierarchical relations, such as hyponymy/hyperonymy (i.e., is-a), and synonymy. Most authors support that semantic distance is the opposite of similarity, but it is sometimes also used as the opposite of semantic relatedness.

The basis for much of the earlier research in SS is the WordNet, a large lexical database of the English language, freely available online. However, the last decade has witnessed an explosion in the number of applications of SS to biomedical ontologies, and specifically in the GO [6]. The GO structure provides meaningful links between GO terms, based on the various relationships it establishes. This structure allows us to capture the similarity between GO terms. In general, the closer two terms are in the GO graph, the more similar their meaning is. Moreover, we can also determine the similarity between two GO-annotated gene products by expanding on this notion to compare sets of GO terms. This provides a measure of the functional similarity between two proteins, which has numerous applications in biomedical research.

The remainder of this chapter provides an overview of SS between GO terms and gene products annotated with GO terms, the different kinds of approaches used in this research area, the issues that affect their performance and evaluation and challenges and future directions.

2   SS Measures

A SS measure can be defined as a function that, given two ontology terms or two sets of terms annotating two entities, returns a numerical value reflecting the closeness in meaning between them [7]. For a theoretical framework for SS measures please refer to [8], where the core elements shared by most SS measures are identified and a foundation for the comparison, selection, and development of novel measures is laid out.

In the context of GO, SS measures can be applied to compute the similarity between two GO terms, *term similarity*, or to compute the similarity between two gene products each annotated with a set of GO terms, *gene product similarity*.

In recent years there have been several categorizations of SS measures [7, 9], and we advise readers to refer to both surveys for a more detailed classification and survey of SS measures and their applications.

2.1 Term Similarity

When considering SS between concepts organized in a taxonomy, as is the case of GO, there are two basic approaches: internal methods based on ontology structure and external methods based on external corpora.

The simplest structural methods calculate distance between two nodes as the number of edges in the path between them [10]. If there are multiple paths, the shortest path or an average of all possible paths can be used. For instance, in Fig. 1, the distance between *heme binding* and *anion binding* is 5. This measure depends only on the structure of the graph and it assumes that all semantic links have the same weight. Accordingly, SS is defined as the inverse score of the semantic distance. This edge-counting approach is intuitive and simple but disregards the depth of the nodes, since it considers
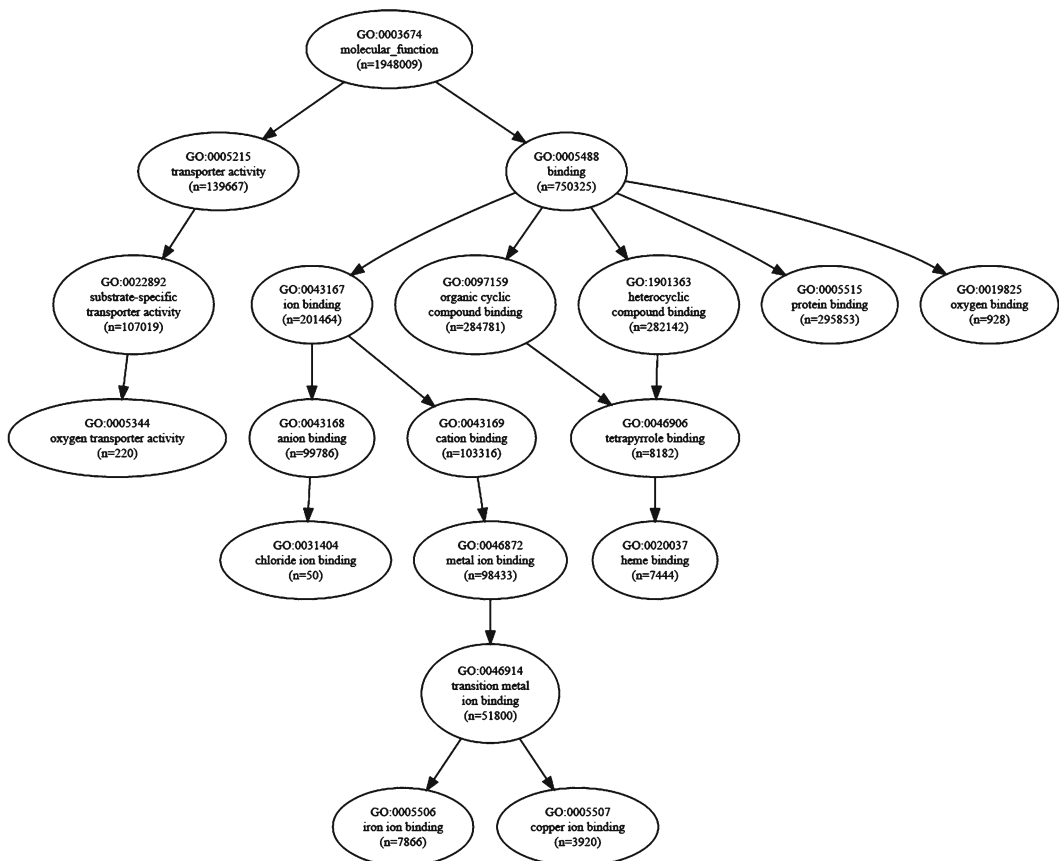
Fig. 1 Subgraph of GO covering the annotations of hemoglobin subunit alpha and hemocyanin II proteins. The number of gene products annotated to each term in GOA (January, 2016) are indicated by *n*

paths of equal length to equate to the same degree of similarity, regardless if they occur near the root or deeper in the ontology. For instance, in Fig. 1, the classes *transport* and *binding* are at a distance of two edges, the same distance that separates *iron ion binding* and *copper ion binding*.

To overcome this limitation of equal distance edges, some approaches give edges different weights to reflect some degree of hierarchical depth. It is intuitive that the deeper the level in the taxonomy, the smaller the conceptual distance, so weights are reduced according to depth. Other factors can be used to determine weights for edges such as node density and type of link.

However these methods have two important limitations, they rely heavily on the assumption that nodes and edges in an ontology are uniformly distributed and that nodes at the same level correspond to the same semantic distance, which are untrue in the case of GO. For instance, in Fig. 1, although *oxygen binding* and *ion binding* are both at a depth of 2, the former is a more specific concept and is actually a leaf node. More recent approaches attempt at mitigating some of these issues using for instance the depth of the lowest common ancestor (LCA) [11], distance to nearest leaf node [12], and depth of distinct GO subgraphs [1]. Related approaches, also based on the structure of the ontology, combine distance metrics with node structural properties, such as number of subclasses and distance to the lowest common ancestor between the terms [13].

External methods typically make use of information-theoretic principles. This type of approach has been demonstrated to be less sensitive or not at all to the issue of link density variability [14], i.e., that the ontology graph may be unbalanced and edges linking nodes may not be evenly distributed, so that the same depth or distance indicate a different level of specificity or similarity. Information content (IC)-based measures are based on the intuition that the similarity between two concepts can be given by the extent to which they share information.

The IC of a concept c is a measure of how likely the concept is to occur, which can be quantified as the negative log likelihood, $-log\ p(c)$ where $p(c)$ is the probability of occurrence of c in a specific corpus, usually estimated by the annotation frequency in the Gene Ontology Annotation database. A normalized version of IC was introduced in [15], whereby IC values are expressed in a range of uniformly scaled values, making them easier to interpret. Taking Fig. 1 again as an example, the frequency of annotation of *binding* is 750,325/1,948,009, making its IC 1.38 and its normalized IC 0.066.

When the concept of IC is applied to the common ancestors two terms have, it can be used to quantify the information they share and thus measure their SS. There are two main approaches for doing this: the most informative common ancestor (MICA technique), in which only the common ancestor with the highest

IC is considered [14]; and the disjoint common ancestors (DCA technique), in which all disjoint common ancestors (the common ancestors that do not subsume any other common ancestor) are considered. There are several methods to compute the DCA [16–18], which allow IC-based measures to take into account multiple common ancestors.

Several measures have been used to measure the information shared by two GO terms. The simplest of these measures, Resnik's, takes the IC of the MICA as the similarity between two terms, and was among the first to be applied to GO [19]. The MICA of *chloride ion binding* and *iron ion binding* is *ion binding*, making the Resnik similarity between these terms to be 0.066. Other measures combine the IC of terms with the IC of the MICA and weight them according to the MICA's IC [20].

More recently, hybrid measures that combine both edge and IC-based strategies have been proposed [21]. Corpus-independent IC measures have also been proposed, based on number of descendants [22], depth and descendants [23] and on the notion of entropy [24].

2.2 Gene Product Similarity

Since gene products can be annotated with several GO terms within each of the three GO categories, gene product SS measures need to compare sets of terms rather than single terms. Several approaches have been proposed for this, most following one of two strategies: pairwise or groupwise.

Pairwise approaches take the individual similarities between all terms annotating two gene products and combine them into a global measure of functional similarity. Any term similarity measure can be applied with this strategy, where each gene product is represented by its set of direct annotations. Typical combination strategies include the average, maximum, or sum, and these can be applied to every pairwise combination of terms from the two sets or only the best-matching pair for each term.

Groupwise approaches calculate gene product similarity directly by one of three approaches: set, graph, or vector. Set approaches consider only direct annotations and are calculated using set similarity techniques. Set-based measures are limited in that they do not take into account the shared ancestry between GO terms. Graph approaches represent gene products as the subgraphs of GO corresponding to all their annotations. Functional similarity is then calculated either using graph-matching techniques or by less computationally intensive approaches such as set similarity. This approach takes into account all annotations (direct and inherited) providing a more comprehensive model of the annotations. Vector approaches represent gene products in vector space, with each term corresponding to a dimension, and functional similarity is calculated using vector similarity measures. Groupwise approaches can also make use of the IC of terms, by using it to weigh set similarity

computations, such as simGIC [15], which compares two sets of terms based on a IC-weighted Jaccard similarity; as scalar values in vectors, such as IntelliGO [25], which combines IC and the evidence content of annotations; or to compute the IC of shared subgraphs, such as the SS measure proposed in [14].

3   Issues and Challenges in SS

Guzzi et al. [9] have identified several issues affecting SS measures, which they categorize into external issues, which are usually related to annotation corpora, and internal issues, inherent to the design of the measures. They do however recognize that both kinds of issues can be entangled, for instance when measures make errone-ous assumptions about the corpora.

The most relevant external issues are the shallow annotation problem, the annotation length bias, and the use of Evidence Codes. The shallow annotation problem stems from the fact that many proteins are only annotated to very general GO terms, thus for instance two proteins can share 100 % of their terms and still be very dissimilar. SS measures need to account for this issue, which can be especially relevant in the electronic annotations. Nevertheless, the quality and specificity of these annotations has been increasing over the years [26].

The annotation length bias refers to the positive correlation between SS scores and the number of annotations that some mea-sures produce. This is due to the fact that annotations are not uni-formly distributed among the proteins within an annotation corpus (and also vary among different organisms corpora), with some pro-teins being very well annotated while others have a single annota-tion. Both of these issues stem from incomplete annotations, which have been shown to have a significant impact in the performance of information-theoretic measures [27]. Finally, SS approaches need to be aware of the impact that using electronic annotations (evi-dence code IEA) can have.[1] Although in general the use of IEA annotations has a positive or null effect on the measures perfor-mance, in some cases and particularly when employing the maxi-mum combination approach over pairwise similarities it can have a detrimental effect and decrease the measure's ability to capture similarity as conveyed by evaluation metrics [9, 17].

There are three levels at which internal issues can occur: term specificity, term similarity, and gene product similarity. At the term specificity level, both typically used approaches (term depth and IC) have their advantages and drawbacks. IC-based measures can be affected by the corpus bias effect [29] whereby rarely used but generic terms possess a high IC but are not biologically specific.

---

[1] Please *see* Chap. 3 [28] for more information on evidence codes.

This issue is particularly relevant when using specific corpora that may be incomplete. Term depth measures on the other hand, while being independent of annotation corpora, are unable to handle the fact that terms at the same depth rarely have the same biological specificity, given the fact that GO's regions have varying node and edge density.

At the term similarity level, distance-based measures suffer from the same issues as term depth term specificity. Moreover, since most measures rely on the concept of common ancestors to measure similarity between two terms, SS measures need to define the set of common ancestors over which similarity is computed. While the most informative common ancestor (or lowest common ancestor in the case of edge-based measures) is commonly used and usually provides good performance, it has been argued that measures taking into account all ancestors or a selection of them can more adequately portray the whole gamut of function.

At the gene product similarity level, and in particular for pairwise measures, special care needs to be taken when choosing a combination approach. The maximum approach is unsuitable to assess their global similarity, since it focuses on the single most similar aspect. The average approach, on the other hand, by making an all-against-all comparison of the terms of two gene products, produces counterintuitive results for gene products with multiple distinct functional aspects. For instance, two gene products both annotated with the same two unrelated terms, *t1* and *t2*, will be 50 % similar under the average approach, because similarity will be calculated between both the matching (*t1–t1,t2–t2*) and the opposite (*t1–t2,t2–t1*) terms of the two gene products. The best-match approach would rely on comparing just (*t1–t1,t2–t2*), since these are the best-matching term pairs in the annotations set. The best-match average approach generally provides a better performance by considering all terms but only the most significant matches.

4 Evaluating and Comparing SS Measures

Evaluating the reliability of SS measures or determining the best measure for each application scenario is still an open question since there is no gold standard. Furthermore, each of the existing measures formalizes the notion of function similarity in slightly different ways and for that reason it is not possible to define what the best SS measure would be, since it becomes a subjective decision. Ultimately, SS measures attempt to capture functional similarity based on GO annotations, so one possible solution is to compare SS measures to other measures or proxies of functional similarity. These include sequence similarity, family similarity, protein–protein interactions, functional modules and complexes, and expression profile similarity. Table 1 details the best performing measures

Table 1
Best performing SS measures according to different protein similarity measures or proxies. Sequence, Pfam, and ECC similarity correspond to correlation evaluated using CESSM

Similarity proxy or measure	Best performing SS measures
Sequence similarity	SSDD [13], SimGIC [15], HRSS [21]
Pfam similarity	SORA [23], SSDD, SimGIC
ECC similarity	SSDD, HRSS, SORA
Expression similarity	TCSS [1], SimGIC, SimIC, Best-Match-Avg (Resnik [15])
Protein–protein interaction	TCSS, SimIC, Max(Resnik)

Results compiled from refs. 9, 13, 21, 23, 30

for each aspect according to a recent survey of literature. Although more classic measures of SS such as Resnik still provide top results in some settings, it is the newer generation of measures that provides the best results. And if until recently [9] GOA-based IC measures were regarded as the best performing measures for most settings, the new wave of more complex structural-based measures, such as SSDD [13], SORA [23] and TCSS [1] are now on the lead, though closely followed by SimGIC. SSDD is based on the concept of semantic "totipotency" whereby terms are assigned values according to their distance to the root and the number of descendants for each of the levels in that path, and then similarity corresponds to the smallest sum of "totipotencies" along a path between two terms. SORA uses an IC based on structural information that considers depth and number of descendants, and then applies set similarity to gene products. TCSS divides the GO graph into subgraphs and considers gene products more similar if they belong to the same subgraph. We postulate that the recent success of structural and hybrid measures, is not only due to their ability to more accurately capture the complexity of the GO graph, but also due to the evolution of GO itself, which has grown considerably since the "classic" measures were proposed. Linear correlation to sequence similarity is one of the most used measures, and in general a positive correlation between sequence and SS has been found, particularly on binned data. Nonlinear regression analysis found that the normal cumulative distribution fits data for many different SS measures, confirming the positive yet, nonlinear agreement between sequence and SS [15]. Linear correlation has also been used to compare SS to Pfam-based and Enzyme Commission Class similarity.

One of the most relevant efforts in this area is the Collaborative Evaluation of Semantic Similarity Measures (CESSM) tool [30], which was created in 2009 to answer this need. It enables the

comparison of new GO-based SS measures against previously published ones considering their relation to sequence, Pfam, and Enzyme Commission Class (ECC) similarity. Since its inception, CESSM has been adopted by the community and used to evaluate several novel SS measures.

The predictive power of SS measures in identifying protein–protein interactions is also commonly employed in SS evaluation [9]. In general SS measures are good predictors of PPI, but the most effective are groupwise or maximum combination approach measures. This is unsurprising given that proteins can interact when sharing a single functional aspect.

5   Tools

There are two main kinds of available tools to compute SS measures in GO: webservers, which typically provide easy to use solutions with fewer parametrizations possible; and software packages, which are more customizable, though more complex to use.

Many of the recently proposed SS measures provide specific webservers, but some online tools provide a wider array of measures, such as ProteInOn [31], FunSimMat [32], or GOssToWeb [33]. These tools rely on their own GO and GOA versions, and though they can output similarity scores with an input of just GO terms or Uniprot accession numbers, these scores are based on the tool's ontology and annotation versions.

If a user needs more control over the parametrization of the input data, then the best option is to employ a software package. Options include R packages (e.g., GoSemSim [34]) or standalone programs (GOssTo [33]), which give the user more freedom in terms of ontology and annotation versions as well as in programmatic access or the computation of SS for larger datasets. A Java library has been recently developed for ontology-based SS calculations [35], which includes over 50 different SS measures and accepts input ontologies in a number of formats, including OWL, OBO, and RDF. This library is well suited for large input datasets, being able to run over 100 million comparisons in under 1 h. In the case of webtools, we advise readers to check their update frequency to ensure that recent versions of GO and the annotations are in use.

6   Challenges and Future Directions

The last decade has witnessed a growing interest in GO-based SS, with dozens of new measures being proposed and applied in different settings. Although measures have become increasingly sophisticated, there remain several challenges and opportunities.

GO-based SS measures are inherently dependent on GO's development and its use in annotations. Measures should evolve with GO, striving to provide ever more accurate metrics for gene product functional similarity. In recent years there have been several developments of GO which SS measures are still not exploring. For instance, the different kinds of regulatory and occurrence relationships, the categorization of evidence codes, logical definitions and internal and external cross-products, can all in principle be explored by SS approaches.

The need to provide more semantically sound measures of SS for biomedical ontologies has been argued [36], and though GO is commonly viewed as a DAG for a controlled vocabulary it is actually well axiomatized in OWL [37]. The presence of these axioms should be considered by SS measures, and the exploration of disjointness in SS has been recently proposed in ChEBI [38].

In general, the computational complexity of SS measures has not been addressed. Current GO-based SS applications happen in an offline context where computational speed is not a relevant factor. However, for applications such as similarity-based search, which so far are based on precomputed similarities [32], performance should be taken into consideration. In addition, the growth in size of biomedical datasets spurred by genomic scale studies in the last few years, also places further computational constraints on SS measures. The challenge of handling very large datasets is increasingly recognized, and recent implementations of SS measures allow for parallel computation [35], but the development of SS measures is not taking this issue into consideration *a priori*.

The next generation of SS measures should take into account these two aspects, on one hand, the possibility for increased complexity in SS measures to provide more accurate similarity scores, and on the other the need for efficient SS computation, and strive to achieve a balance between increased accuracy and efficiency.

7 Exercises

Consider the subgraph of GO represented in Fig. 1 and the number of annotations for each GO term it shows.

1. Calculate the IC of the term "heme binding" considering that the total universe of annotations corresponds to the number of annotations to the root term.

2. Transform the IC value calculated in 1 to a uniform scale [0,1]. Consider that the maximum IC is given to a term with a single annotated gene product, and an IC of zero corresponds to the IC of the root term, "molecular function."

3. Calculate the SS between the terms "chloride ion binding" and "iron ion binding," and "oxygen transporter activity," and

"tetrapyrrole binding," following the minimum edge distance measure.

4. Calculate Resnik's SS between the same terms as in c.

5. Calculate the similarity between the protein *hemoglobin subunit alpha* annotated with [ion iron binding, copper ion binding, protein binding, heme binding, oxygen binding, oxygen transporter activity], and the protein *hemocyanin II* annotated with [chloride ion binding, copper ion binding, oxygen transporter activity]:

   (a) Using the average of all pairwise Resnik's similarities

   (b) Using the maximum of all pairwise Resnik's similarities

   (c) Using the simGIC measure, which corresponds to the ratio between sum of the IC of the shared terms between the two proteins and the sum of the IC of the union of all terms between the two proteins.

   (d) Compare the obtained results with your perception of the actual functional similarity between the two proteins.

References

1. Jain S, Bader GD (2010) An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology. BMC Bioinformatics 11(1):562

2. Li X, Wang Q, Zheng Y, Lv S, Ning S, Sun J, Li Y (2011) Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer. Nucleic Acids Res 39(22):e153

3. Richards AJ, Muller B, Shotwell M, Cowart LA, Rohrer B, Lu X (2010) Assessing the functional coherence of gene sets with metrics based on the Gene Ontology graph. Bioinformatics 26(12):i79–i87

4. Bastos HP, Clarke LA, Couto FM (2013) Annotation extension through protein family annotation coherence metrics. Front Genet 4:201

5. Budanitsky A, Hirst G (2001) Semantic distance in WordNet: an experimental, application-oriented evaluation of five measures. In Workshop on WordNet and other lexical resources, vol 2, pp 2–2

6. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Sherlock G et al (2000) Gene ontology: tool for the unification of biology. Nat Genet 25(1):25–29

7. Pesquita C, Faria D, Falcao AO, Lord P, Couto FM (2009) Semantic similarity in biomedical ontologies. PLoS Comput Biol 5(7):e1000443

8. Harispe S, Sánchez D, Ranwez S, Janaqi S, Montmain J (2014) A framework for unifying ontology-based semantic similarity measures: a study in the biomedical domain. J Biomed Inform 48:38–53

9. Guzzi PH, Mina M, Guerra C, Cannataro M (2012) Semantic similarity analysis of protein data: assessment with biological features and issues. Brief Bioinform 13(5):569–585

10. Rada R, Mili H, Bicknell E, Blettner M (1989) Development and application of a metric on semantic nets. IEEE Trans Syst Man Cybernet 19(1):17–30

11. Yu H, Gao L, Tu K, Guo Z (2005) Broadly predicting specific gene functions with expression similarity and taxonomy similarity. Gene 352:75–81

12. Cheng J, Cline M, Martin J, Finkelstein D, Awad T, Kulp D, Siani-Rose MA (2004) A knowledge-based clustering algorithm driven by gene ontology. J Biopharm Stat 14(3):687–700

13. Xu Y, Guo M, Shi W, Liu X, Wang C (2013) A novel insight into Gene Ontology semantic similarity. Genomics 101(6):368–375

14. Resnik P (1999) Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language. J Artif Intell Res (JAIR) 11:95–130

15. Pesquita C, Faria D, Bastos H, Ferreira AE, Falcão AO, Couto FM (2008) Metrics for GO based protein semantic similarity: a systematic evaluation. BMC Bioinformatics 9(Suppl 5):S4

16. Couto FM, Silva MJ, Coutinho PM (2005) Semantic similarity over the gene ontology: Family correlation and selecting disjunctive ancestors. Proceedings of the ACM conference in information and knowledge management

17. Couto FM, Silva MJ (2011) Disjunctive shared information between ontology concepts: application to Gene Ontology. J Biomed Semantics 2:5

18. Zhang SB, Lai JH (2015) Semantic similarity measurement between gene ontology terms based on exclusively inherited shared information. Gene 558(1):108–117

19. Lord P, Stevens R, Brass A, Goble C (2003) Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. Bioinformatics 19:1275–1283

20. Schlicker A, Domingues FS, Rahnenführer J, Lengauer T (2006) A new measure for functional similarity of gene products based on gene ontology. BMC Bioinformatics 7:302

21. Wu X, Pang E, Lin K, Pei ZM (2013) Improving the measurement of semantic similarity between gene ontology terms and gene products: insights from an edge-and IC-based hybrid method. PLoS One 8(5):e66745

22. Seco N, Veale T, Hayes J (2004) An intrinsic information content metric for semantic similarity in wordnet. ECAI, pp 1089–1090

23. Teng Z, Guo M, Liu X, Dai Q, Wang C, Xuan P (2013) Measuring gene functional similarity based on group-wise comparison of GO terms. Bioinformatics:btt160

24. Warren A, Setubal J (2012) Using entropy estimates for DAG-based ontologies. In Proceedings of the 15th bio-ontologies special interest group meeting of ISMB 2012

25. Benabderrahmane S, Smail-Tabbone M, Poch O, Napoli A, Devignes MD (2010) IntelliGO: a new vector-based semantic similarity measure including annotation origin. BMC Bioinformatics 11(1):588

26. Škunca N, Altenhoff A, Dessimoz C (2012) Quality of computationally inferred gene ontology annotations. PLoS Comput Biol 8(5):e1002533

27. Jiang Y, Clark WT, Friedberg I, Radivojac P (2014) The impact of incomplete knowledge on the evaluation of protein function prediction: a structured-output learning perspective. Bioinformatics 30(17):i609–i616

28. Gaudet P, Škunca N, Hu JC, Dessimoz C (2016) Primer on the gene ontology. In: Dessimoz C, Škunca N (eds) The gene ontology handbook. Methods in molecular biology, vol 1446. Humana Press. Chapter 3

29. Mistry M, Pavlidis P (2008) Gene ontology term overlap as a measure of gene functional similarity. BMC Bioinformatics 9:327

30. Pesquita C, Pessoa D, Faria D, Couto F (2009) CESSM: collaborative evaluation of semantic similarity measures. In: JB2009: challenges in bioinformatics, vol 157, p 190

31. Faria D, Pesquita C, Couto FM, Falcão A (2007) Proteinon: a web tool for protein semantic similarity. Department of Informatics, University of Lisbon

32. Schlicker A, Albrecht M (2008) FunSimMat: a comprehensive functional similarity database. Nucleic Acids Res 36(Suppl 1):D434–D439

33. Caniza H, Romero AE, Heron S, Yang H, Devoto A, Frasca M et al (2014) GOssTo: a stand-alone application and a web tool for calculating semantic similarities on the Gene Ontology. Bioinformatics 30(15):2235–2236

34. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 26(7):976–978

35. Harispe S, Ranwez S, Janaqi S, Montmain J (2014) The semantic measures library and toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies. Bioinformatics 30(5):740–742

36. Couto FM, Pinto HS (2013) The next generation of similarity measures that fully explore the semantics in biomedical ontologies. J Bioinforma Comput Biol 11(05): 1371001

37. Mungall CJ, Dietze H, Osumi-Sutherland D (2014) Use of OWL within the Gene Ontology. Proceedings of the 11th international workshop on OWL: experiences and directions. Riva del Garda, Italy, 2014

38. Ferreira JD, Hastings J, Couto FM (2013) Exploiting disjointness axioms to improve semantic similarity measures. Bioinformatics 29(21):2781–2787