Substructure Index-based Approximate Graph Alignment

RRID:SCR_003434
Type: Tool

Proper Citation

Substructure Index-based Approximate Graph Alignment (RRID:SCR_003434)

Resource Information

URL: http://portal.ncibi.org/gateway/saga.html

Proper Citation: Substructure Index-based Approximate Graph Alignment (RRID:SCR_003434)

Description: SAGA (Substructure Index-based Approximate Graph Alignment) is a tool for querying a biological graph database to retrieve matches between subgraphs of molecular interactions and biological networks. SAGA implements an efficient approximate subgraph matching algorithm that can be used for a variety of biological graph matching problems such as the pathway matching SAGA uses to compare pathways in KEGG and Reactome. You can also use SAGA to find matches in literature databases that have been parsed into semantic graphs. In this use of SAGA, portions of PubMed have been parsed into graphs that have nodes representing gene names. A link is drawn between two genes if they are discussed in the same sentence (indicating there is potential association between the two genes). SAGA lets you match graphs between different databases even though the content is distinct and the databases organize pathways in different ways. This cross-database matching is achieved by SAGA's flexible approximate subgraph matching model that computes graph similarity, and allows for node gaps, node mismatches, and graph structural differences. Comparing pathways from different databases can be a useful precursor to pathway data integration. SAGA is very efficient for querying relatively small graphs, but becomes prohibitory expensive for querying large graphs. Large graph data sets are common in many emerging database applications, and most notably in large-scale scientific applications. To fully exploit the wealth of information encoded in graphs, effective and efficient graph matching tools are critical. Due to the noisy and incomplete nature of real graph datasets, approximate, rather than exact, graph matching is required. Furthermore, many modern applications need to query large graphs, each of which has hundreds to thousands of nodes and edges. TALE is an approximate subgraph matching tool for matching graph queries with a large number of nodes and edges. TALE employs a novel indexing technique that achieves a high pruning power and scales linearly with the database size.

Abbreviations: SAGA

Synonyms: SAGA (Substructure Index-based Approximate Graph Alignment), SAGA - Substructure Index-based Approximate Graph Alignment, SAGA: A Fast and Flexible Graph Matching Tool

Resource Type: software resource, software application

Defining Citation: PMID:17110368

Keywords: gene, algorithm, alignment, biological, graph, interaction, literature, molecular, pathway, query, reactome, structural, subgraph, substructure, tool, graph similarity

Funding: Microsoft ;
NIDA 1-U54-DA021519-01A1

Availability: Access free via the web; Binaries are also freely available. If you use SAGA for any work that leads to a publication you must cite paper.

Resource Name: Substructure Index-based Approximate Graph Alignment

Resource ID: SCR_003434

Alternate IDs: nif-0000-33157

Record Creation Time: 20220129T080219+0000

Record Last Update: 20250421T053355+0000

Ratings and Alerts

No rating or validation information has been found for Substructure Index-based Approximate Graph Alignment.

No alerts have been found for Substructure Index-based Approximate Graph Alignment.

Data and Source Information

Source: SciCrunch Registry

Usage and Citation Metrics

We found 23 mentions in open access literature.

Listed below are recent publications. The full list is available at NIF.

Bianchini M, et al. (2025) Modeling Climate-Driven Vegetation Changes Under Contrasting Temperate and Arid Conditions in the Mediterranean Basin. Ecology and evolution, 15(1), e70753.

Khanal S, et al. (2024) Environmental correlates of the forest carbon distribution in the Central Himalayas. Ecology and evolution, 14(6), e11517.

Burch J, et al. (2024) Wright was right: leveraging old data and new methods to illustrate the critical role of epistasis in genetics and evolution. Evolution; international journal of organic evolution, 78(4), 624.

Fang H, et al. (2024) Evaluating the impact of soil erosion on soil quality in an agricultural land, northeastern China. Scientific reports, 14(1), 15629.

Fleischauer J, et al. (2023) TGF? Inhibitor A83-01 Enhances Murine HSPC Expansion for Gene Therapy. Cells, 12(15).

Fan PY, et al. (2023) The spatial configuration of local climate zones reveals effects on wayfinding in human walking. PloS one, 18(9), e0289780.

Opgenoorth L, et al. (2021) The GenTree Platform: growth traits and tree-level environmental data in 12 European forest tree species. GigaScience, 10(3).

Schwarzer A, et al. (2021) Predicting genotoxicity of viral vectors for stem cell gene therapy using gene expression-based machine learning. Molecular therapy : the journal of the American Society of Gene Therapy, 29(12), 3383.

Ball JW, et al. (2020) Fine-scale species distribution modelling and genotyping by sequencing to examine hybridisation between two narrow endemic plant species. Scientific reports, 10(1), 1562.

Ruckman SN, et al. (2020) The March of the Beetles: Epistatic Components Dominate Divergence in Dispersal Tendency in Tribolium castaneum. The Journal of heredity, 111(5), 498.

Wang Y, et al. (2019) Optimizing the Predictive Ability of Machine Learning Methods for Landslide Susceptibility Mapping Using SMOTE for Lishui City in Zhejiang Province, China. International journal of environmental research and public health, 16(3).

Yates CJ, et al. (2019) High species diversity and turnover in granite inselberg floras highlight the need for a conservation strategy protecting many outcrops. Ecology and evolution, 9(13), 7660.

Pulla S, et al. (2016) Controls of Soil Spatial Variability in a Dry Tropical Forest. PloS one,

11(4), e0153212.

Zayasu Y, et al. (2016) Unexpectedly complex gradation of coral population structure in the Nansei Islands, Japan. Ecology and evolution, 6(15), 5491.

Homburger H, et al. (2015) Patterns of livestock activity on heterogeneous subalpine pastures reveal distinct responses to spatial autocorrelation, environment and management. Movement ecology, 3, 35.

Hufford KM, et al. (2014) Genetic variation among mainland and island populations of a native perennial grass used in restoration. AoB PLANTS, 6.

Leon JX, et al. (2014) Incorporating DEM uncertainty in coastal inundation mapping. PloS one, 9(9), e108727.

Nmor JC, et al. (2013) Topographic models for predicting malaria vector breeding habitats: potential tools for vector control managers. Parasites & vectors, 6, 14.

López-Collado J, et al. (2013) Geographic distribution of habitat, development, and population growth rates of the Asian citrus psyllid, Diaphorina citri, in Mexico. Journal of insect science (Online), 13, 114.

Jensen HR, et al. (2012) The Red Queen and the seed bank: pathogen resistance of ex situ and in situ conserved barley. Evolutionary applications, 5(4), 353.