Journal of Visualization and Interaction
JoVI review snapshot created at Jan 27, 2026 08:55.

Issue #17 (open): [REVIEW] GLaDOS: Graph Layout Algorithm Benchmark Datasets for Open Science â€“ GD Benchmark Sets

@mjwybrow on Apr 28, 2025 08:36: [opened]

@mjwybrow on Apr 28, 2025 08:36:

Conflicts of interest

- ☒ I declare that I have no known conflicts of interest with the authors.

Reviewed version

f41e831

Review

The article describes GlaDOS, a website providing a collection of benchmark graph datasets. GLaDOS provides the graphs in several common formats or links to their source (where this is actively curated elsewhere). Additionally, the site contains various tags and statistics, a summary of each data set and a description of its origin and uses of those datasets within the literature.

Strengths

- The presented collection of datasets were collected through a clearly time-consuming process that even involved recreating graphs through author instructions or contacting lab members to secure missing data.
- The presented collection also includes lots of valuable information on the origin and uses of each dataset.
- It makes a valuable contribution in making these available in perpetuity both through the GLaDOS website and providing the underlying code and instructions behind the site.

Weaknesses

- The entries for datasets within the article and on the GLaDOS website are hand-authored and inconsistent in terms of the information provided and the formatting. This brings into question the reliability and accuracy of this data.
- The high-level classification of datasets is also unclear and some surprising choices, raising questions about the value of the "taxonomy" contribution.
- Ongoing plans for upkeep and maintenance of the GLaDOS are not described.
- The included datasets are based on papers mostly from the Graph Drawing conference during a set period and with several citations (plus associated papers). While this is reasonable selection criteria it means the collection is not exhaustive.
- The article has a number of minor issues, described below, that require more explanation or

clarification.

Review content

- Article says "Our organization by features and statistics supports rapid identification of relevant graphs". While the article itself lets people filter by feature, the website (the actual collection) doesn't provide any filtering or search, just a long table-style list with a "Features" column. To satisfy its intended purpose, the website needs some search/filtering capability.
- When mentioning the work of W. Hu et al. having a different focus, the article should explain why it is not relevant.
- The article says they looked "into internal storages of research groups". Please clarify what this means. Does it mean the authors contacted someone else from their previous group, if the author was no longer there and didn't have the data themselves, and they looked in the internal storage?
- The figure (data collection process) at the beginning of Section 3 should be labelled and have a caption.
- "the graph structure is very difficult to piece together" This could be explained. Is it because it is in an uncommon format, or that the source data is something like Twitter posts.
- The section on file formats (3.1) is unclear. It says "We chose to convert and store several of the datasets in a uniform JSON representation". Why only "some"? It says "we have also converted and made available all graphs in three additional commonly-used formats: GraphML, GEXF, and GML". Why have "multiple accessible formats"? Do these all have exactly the same information, or do some lack the extra information ("timestamps, labels, or belonging to a clusters, and edges having weights"). The article and website should make clear if there is a master format for everything and then convenience formats and what information they each contain. Otherwise people could end up comparing graphs that are actually different (for info like clusters) when thinking they are the same.
- "the original data we downloaded when the files were small enough to be uploaded to GitHub". What does this mean? Why does it matter?
- For Figure 3, it says the top 20 most used (should be "most-used") data sets but everything after the first 14 only have 1 graph. It would make more sense just to show those 14 with a comment that the others are all single graphs.
- Figure 5 is not needed to show a single number, especially when it has a different scale to other figures.
- Taxonomy/Classification
    - "This work proposes an overarching taxonomy of datasets", but I feel it is a stretch to call it this. There are a number of inconsistencies and questions about the categorisation, as described below. The classification should be made more thorough and precise, or this claim reduced.
    - Some of the Datasets are classes of graph rather than a single data set. For example "Social Networks" is from 4 data sources. Does this mean all the papers that use this use all 4 or just 1. If the later, this makes the presentation of the most-used data sets questionable with Social Networks being number 4.
    - Similar to this, if the purpose of the archive is for comparison why split entries that refer to the same graphs but the authors mislabelled.
    - The difference between "Uniform Benchmark" and "Established Network Repository" is not clear, especially as it related to subset collections.
    - Relatedly, Figure 4 shows 6 collections that are a "subset of another collection". Where are these on the GLaDOS website? This isn't clear. Similarly the website has a section "Other" with 7 datasets, stating "A few more datasets that did not fit in the preivous categories." (note typo on "previous"). Where are these in Figure 4? All together the categorisation appears arbitrary and inconsistent, and can hardly be called a taxonomy.
    - The article says "additional information about the nature of these tags can be found in Section 3 and Section 3.1" except those sections really don't contain info on the tags other than that tagging was done by two people. This should be properly explained.

    - The article says "Uniform Benchmark datasets" "aims to provide a general overview of the performance of a graph layout algorithm by testing on a large amount of graphs varying in size and density, rather than focusing on a specific type of graph". It is not clear what "specific type of graph" means. For instance, it could be argued that all graphs in the Storylines dataset are a specific type of graph.
    - Why categorise the North DAGs and AT&T graphs separately, and then in the North NAGs description, just have See AT&T". Given the focus on reproducibility, It would make more sense to have this as an alias for the AT&T dataset and combine them in the listing.

- The KnownCR set lists the "Known Crossing Number" tag, but is this value included in the data (graph downloads) for each graph in the dataset?

- For the "Established Network Repositories" the article says "we do not include here any storage of the data (which would be redundant) or report statistics on them" but provided information (size information, min/max nodes, node distribution and summary charts) would be useful since that allows people to select the appropriate data set. If the issue with generating stats is that they could becoe outdated if the collection is added to, you could say the information was based on a date and then potentially update these as part of the maintenance of GLaDOS.

- The article says "One more of such collections is Konect. At the time of writing, though, the website for Konect has been down for a while. Both the data and the website are still accessible through web archiveâ€”thus we do not consider this a lost collection." Why is the Konect dataset not included in the collection then?

- For a dataset like "Car Features" where the features are not known, the dataset is unavailable and the author won't disclose the origin of the data, is there value including it at all, except maybe in a list of papers that fit this category (I imagine there might be many).

- Custom-made Datasets

    - The final part of the article talks about Custom-made datasets. It says "we also found several instances of custom-made datasets". From the chart it looks like close to 80 instances. It would be good to say this and be precise.

    - It also talks about sub-categories of Replicable vs. Reproducible vs non-replicable. Where does this information appear in the GLaDOS website? This is not clear.

- In Section 4.2 "Random Generation"

    - The article says "The list of features to take into account to claim that a synthetic graph is comparable to another one would be long, and perhaps out of the scope of this publication. These are just a few examples of what could be relevant:" What follows looks like a list produced by GenAI with no explanation of where this came from (at least I got similar result crafting a quick prompt from the text above). I question what this list contributes to the paper. If it is to be included, it should be shortened and properly justified.
    - References should be given for "ErdÅ‘s-RÃ©nyi", "BarabÃ¡si-Albert" and also the sentence beginning "Conversely, the BA model produces scale-free networks with... "

- In Section 5, the difference between gray and blue dots in Figure 8 is almost impossible to distinguish. Please use more visually distinct colours or a different mark.

- The future maintainability of the work is unclear. That is, there is no commitment or plan in the article of the future maintainability of the archive as new graph data sets are published. It would be good to see a plan for this, whether it is the authors or as an open project. If the later there might be a need to document the processes for maintenance to allow others to update the GLaDOS website.

Issues with GLaDOS content included in article vs. website

- The descriptions on the website are clearly hand-authored and are inconsistent in the presented information. For example most don't list the number of graphs, sometimes the Size information is formatted differently. It would be useful for all of them to list the number of graphs in the dataset (some are listed in paper, but not on website). No degree statistics are listed for Storylines (is this because all nodes are degree 1 or 2? No degree statistics are listed for Co-Phylogenetic Trees dataset. The Militarized Interstate Disputes (MID) dataset has no descriptive stats at all. All these entries

should be consistent, so the user of this set can have a reasonable expectation of the available info, even if this is blank or N/A for some entries.
- Sometimes under Usage examples each paper lists the authors and years (GraphViz in article) other times just [link] (Graphviz on website, Matrix Market in both article and website). Relatedly, links to papers are formatted inconsistently throughout, for example in the Co-Phylogenetic Trees dataset it says "Collected by the authors of â€œhttps://almob.biomedcentral.com/articles/10.1186/s13015-014-0031-3â€." It would help to have consistent formatting like with the Author name, and year, like a paper reference, with a clickable hyperlink. This applies to both the website and the embedded expandable version in the paper.
- The article makes it clear that all datasets are sourced from the review of literatures, but the GLaDOS website doesn't give any details of where they were sourced, other than the "Benchmark datasets" category saying "These are collections of graphs that have been frequently used in graph drawing papers" which is then not listed under the other categories. There should be a sentence or two on the website that explains all datasets came from graph drawing literature.
- The Enron entry has a typo "webiste" in both the article and the GLaDOS website.
- The Transportation Networks entry has a typo "feautre" in the article only.

Minor typographical issues

- "although all the links to it in previous papers are now broken" -> "with all the links to it in previous papers now broken"
- "by still documenting what we could find about them" -> "by documenting what we could still find about them"
- "with an accent on encouraging efforts towards replicability" -> "with an focus on encouraging replicability"
- "Following this process, we tried to track down" -> "We used the following process to track down"
- "The chart below shows" -> "Figure 1 shows"
- "The following one, instead, shows" -> "Figure 2 shows"
- "last 7 years" Would be better to define the period.
- "redistirbution" -> "redistribution"
- "This phenomenon, as long as the information" -> "This phenomenon, as well as the information"
- "particularly useful when researchers are particularly useful for this task" -> "particularly useful for this task"
- "A dataset that is not anymore accessible renders" -> "A dataset that is not accessible anymore renders"
- "gaining more and more attention." -> "gaining more attention."
- "but has found no universal solution yet" -> "but has not resulted in a universal solution"

Openness/Transparency

A long-term-archive provides the code and data for the website, as well as code used for the reconstruction of "lost" data sets. As such, this work is open and people can scrutinise and build on it.

Submission categories

- ☐ Registered Report
- ☐ Replication Study
- ☐ Empirical Research - Quantitative
- ☐ Empirical Research - Qualitative
- ☐ Systems or design research
- ☐ Commentary
- ☒ Systematic Literature Review

Suggested outcome

Minor revisions: this paper requires some smaller changes, after which I am confident I would be able to endorse it.

Requested changes

- Ensure consistency of the metadata for datasets within the article and on the GLaDOS website, and the describe the process for accuracy.
- Address or explain the inconsistencies/questions about the classification of datasets, and potentially downplay or remove the contribution of a taxonomy.
- Add search/filtering to the GLaDOS website, similar to the article.
- Describe plans for upkeep and maintenance of the GLaDOS archive in the future.
- Explain the properties of different graph download formats and whether they all contain equivalent information.
- Address other minor issues that require explanation or clarification.

ORCID

https://orcid.org/0000-0001-5536-7780

@mjwybrow on Dec 14, 2025 11:22:

Endorse:* I am willing to endorse this paper, with at most minor copyediting.

Thank you. I have reviewed the changes, My only suggestion is that the text "We chose to convert and store several of the datasets in a uniform JSON representation" be replaced with "We chose to convert and store datasets in a uniform JSON representation", i.e., make it clear this is the main format used for all data sets rather than just some.