Finding large independent sets in graphs with bounded degrees

A Data Management Plan created using DMPonline.be

Creator: Jan Goedgebeur https://orcid.org/0000-0001-8984-2463

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: G0AGX24N

ID: 204355

Start date: 01-01-2024

End date: 31-12-2027

Project abstract:

Sparse graphs are graphs whose number of edges is at most linear in their number n of vertices. Prime examples of sparse graphs include planar graphs, graphs excluding a fixed minor, and graphs with bounded degrees. It is well known that sparse graphs must contain large independent sets, of size linear in n. But how large an independent set are we guaranteed to find exactly? In this project, we set out to answer this question for an important family of sparse graphs, namely graphs having bounded maximum degree and no short cycles.

More precisely, given an upper bound on the maximum degree of a vertex, and a lower bound on the length of cycles, we are interested in determining the largest real r such that every graph satisfying the bounds has an independent set consisting of at least a fraction r of its vertices. This is a fundamental extremal problem that has been studied for the past 40 years. While several lower and upper bounds have been established for a wide range of the parameters, a precise answer is known only for a handful of cases. Based on recent developments in the area, including in a joint work of the two PIs, we believe that the time is ripe to make significant progress on the problems. Our approach to tackle these problems is to develop new mathematical ideas and proofs with the help of computers. The computational methods envisioned include classical exhaustive generation and a novel method based on deep reinforcement learning.

Last modified: 26-01-2024

Finding large independent sets in graphs with bounded degrees
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Finding large independent sets in graphs with bounded degrees
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Finding large independent sets in graphs with bounded degrees
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

In this research project we will collect and/or generate the following data types:
- Write source code for the computer programs which will be developed in this project.
- Collect lists of graphs (which will be generated by the algorithms which we will design in this project).
- Write manuscripts to disseminate the results obtained in this project.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Jan Goedgebeur and Gwenaël Joret will be responsible for the data preservation at KU Leuven and ULB, respectively.
2. The following storage capacity is in place and will be used (during and after the project):
   - All source code developed in this project will be released under an open source license and will be hosted on a public GitHub repository.
   - The new lists of graphs will be uploaded to our online searchable database of interesting graphs called the House of Graphs (https://houseofgraphs.org/).
   - All testdata will be backed up on the central backup server of the department (using procedures for automatic backups).
   - All manuscripts will be uploaded to the preprint server arXiv.org.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

N/A

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

N/A

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

N/A

Finding large independent sets in graphs with bounded degrees
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Papers	Manuscripts in which the new results will be presented	New data	Digital	Other	.tex and .pdf	<1GB	
Source code	Computer programs that we will develop	Both new data and reusing existing data	Digital	Software	C++, Java, and/or Python code	<100MB	
Graphs	New lists of graphs that we will generate	Both new data and reusing existing data	Digital	Graphs in graph6 or adjacency list format	.txt	<100GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

We will possibly reuse existing generators or existing lists of graphs. The sources are listed here.
1. http://users.cecs.anu.edu.au/~bdm/nauty/
2. https://houseofgraphs.org/

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

- All source code developed in this project will be released under an open source license and will be hosted on a public GitHub repository and will include a README file with clear instructions on how to use the software. The code will also contain comments so other researchers can also build upon the software, if wanted.
- The new lists of graphs will be uploaded to the online searchable database of interesting graphs called the House of Graphs (https://houseofgraphs.org).
- All manuscripts will be uploaded to the preprint server arXiv.org.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

3. Data storage & back-up during the research project

Where will the data be stored?

1. Manuscripts will be uploaded to arXiv.org (a usual habit in the field), KU Leuven's internal repository Lirias and KU Leuven's OneDrive system (which has automatic back-up capabilities).
2. Source code will be made available under an open source license on a public GitHub repository and also stored on KU Leuven's OneDrive system.
3. Graphs will be stored on the online graph database https://houseofgraphs.org/ and on KU Leuven's OneDrive system.

How will the data be backed up?

KU Leuven's OneDrive storage has automatic back-up capabilities. In case this would fail, there will also be a back-up available on arXiv, Lirias, GitHub or the House of Graphs.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

The size of the data that will be produced is small to moderately large and there is plenty of space available (e.g. every KU Leuven employee has 2 TB of available storage space).

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

KU Leuven's One Drive allows secure and personalized storage. Moreover, none of the data that will be produced are sensitive and we strive to make everything publicly available to allow other researchers to build on top of our work.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

No costs.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

The following data will be retained for at least five years:
1. papers
2. source code
3. graphs

Where will these data be archived (stored and curated for the long-term)?

- KU Leuven's OneDrive system
- Lirias
- arXiv
- GitHub
- House of Graphs

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No costs are expected. The storage is provided by KU Leuven and/or the department.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

All data will be made available as follows:
1. Manuscripts will be uploaded to arXiv.org;
2. The developed source code will be released under an open-source license and hosted on a public GitHub repository;
3. The new graphs will be uploaded to the House of Graphs.

If access is restricted, please specify who will be able to access the data and under what conditions.

N/A

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

1. Manuscripts will be uploaded to arXiv.org (a usual habit in the field), KU Leuven's internal repository Lirias and KU Leuven's OneDrive system (which has automatic back-up capabilities).
2. Source code will be made available under an open source license on a public GitHub repository and also stored on KU Leuven's OneDrive system.
3. Graphs will be stored on the online graph database https://houseofgraphs.org/ and on KU Leuven's OneDrive system.

When will the data be made available?

The data will be made available as soon as the preprint related to the research is made available on the preprint server arXiv.

Which data usage licenses are you going to provide? If none, please explain why.

All data will be made available as open access under a Creative Commons License.
The source code will be publicly available as open source software under the GNU GPLv3 license.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

What are the expected costs for data sharing? How will these costs be covered?

No costs.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Jan Goedgebeur

Who will manage data storage and backup during the research project?

Jan Goedgebeur

Who will manage data preservation and sharing?

Jan Goedgebeur

Who will update and implement this DMP?

Jan Goedgebeur