# Building bridges between computer-assisted graph theory and operations research

*A Data Management Plan created using DMPonline.be*

**Creator:** Jorik Jooken

**Affiliation:** KU Leuven (KUL)

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

**Template:** FWO DMP (Flemish Standard DMP)

**Principal Investigator:** Jorik Jooken

**Grant number / URL:** 1222524N

**ID:** 204049

**Start date:** 01-10-2023

**End date:** 30-09-2026

**Project abstract:**

Many problems in graph theory can be solved by algorithmically exploring a large search space. The key ingredient for most state-of-the-art computer-assisted graph theory (CAGT) approaches consists of carefully designed algorithms that are capable of generating all graphs of a given class while avoiding redundant isomorphic copies. The strategic goal of this research proposal is to extend and complement CAGT techniques with a rich toolbox of operations research (OR) algorithms. To demonstrate the potential of this integrated CAGT-OR approach, a number of important open problems in graph theory for which such an approach seems promising will be tackled. These problems can be subdivided into three distinct research lines.

The first research line concerns problems on cycles that contain all fixed-size vertex subsets of a given graph. The aim is to answer questions about the largest such subsets and to investigate questions in which an additional constraint is imposed on cycles so that certain fixed-size vertex subsets are forbidden.

The second research line revolves around the average order of connected induced subgraphs of simple graphs and three important subclasses, namely cubic graphs, trees and graphs with minimum vertex degree three.

The third and final research line aims to improve the currently best lower and upper bounds on the order of cages (the smallest regular graphs with a given girth) and edge-girth-regular graphs (a closely related extension).

**Last modified:** 12-01-2024

# Building bridges between computer-assisted graph theory and operations research
# DPIA

---

**DPIA**

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

- Not applicable

# Building bridges between computer-assisted graph theory and operations research
# GDPR

---

### GDPR

**Have you registered personal data processing activities for this project?**

- Not applicable

# Building bridges between computer-assisted graph theory and operations research
# 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)**

The following datatypes will be generated (G) and reused (R):
1. G: Manuscripts (including LaTeX code and figures);
2. G and R: Source code of algorithms that we will develop;
3. G and R: Graphs that are the result of conducting experiments with the algorithms.

**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)**

Jorik Jooken is responsible for preserving the data. All data will be stored in multiple locations to avoid data loss. The different data types will be stored as follows:
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 (a usual habit in the field) 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.

**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)**

NA

**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)**

NA

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

None.

# Building bridges between computer-assisted graph theory and operations research
# 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:*<br>• Generate new data<br>• Reuse existing data | *Please choose from the following options:*<br>• Digital<br>• Physical | *Please choose from the following options:*<br>• Observational<br>• Experimental<br>• Compiled/aggregated data<br>• Simulation data<br>• Software<br>• Other<br>• NA | *Please choose from the following options:*<br>• .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, …<br>• NA | *Please choose from the following options:*<br>• <100MB<br>• <1GB<br>• <100GB<br>• <1TB<br>• <5TB<br>• <10TB<br>• <50TB<br>• >50TB<br>• NA | |
| Papers | Manuscripts in which the research will be described | Generate new data | Digital | Other | .tex and .pdf | <1GB | NA |
| Software | The source code of algorithms that will be developed | Generate new data and reuse existing data | Digital | Software | .c, .cpp, .py | <1GB | NA |
| Graphs | Mathematical objects known as "graphs" that are typically the output of these algorithms | Generate new data and reuse existing data | Digital | Experimental | .g6 | <100GB | NA |
| | | | | | | | |

**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 probably reuse existing software that was developed earlier (by our group or other groups). We will use:
1) nauty (http://users.cecs.anu.edu.au/~bdm/nauty/)
2) various algorithms that were previously developed in our group (https://kulak.kuleuven.be/algo/software)
We will also reuse existing graphs that can be found on the graph database 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).**

- The source code of the algorithms will be made publicly available on a GitHub repository, where we will add a README file that contains clear instructions outlining the content of the repository (see https://github.com/JorikJooken/ for some examples).
- The graphs will be uploaded to the graph database House of Graphs (https://houseofgraphs.org), which allows to annotate the graphs by textual comments and indication of interesting invariants.
- The papers will be uploaded to the preprint server arXiv.

**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 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?**

There will be no costs for me for data storage and backup.

**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)?**

1) KU Leuven's OneDrive system
2) Lirias
3) arXiv
4) GitHub
5) House of Graphs

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

There will be no costs for me for data storage and backup.

**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

1) The source code of the algorithms will be made publicly available on a GitHub repository, where we will add a README file that contains clear instructions outlining the content of the repository (see https://github.com/JorikJooken/ for some examples).
2) The graphs will be uploaded to the graph database House of Graphs (https://houseofgraphs.org), which allows to annotate the graphs by textual comments and indication of interesting invariants.
3) The papers will be uploaded to the preprint server arXiv.

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

NA

**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 (https://github.com/JorikJooken/) 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.**

I typically apply a CC-BY license (papers) or a GPLv3 license (source code).

**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?**

There will be no costs for me.

**6. Responsibilities**

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

Jorik Jooken

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

Jorik Jooken

**Who will manage data preservation and sharing?**

Jorik Jooken

**Who will update and implement this DMP?**

Jorik Jooken