HypergraphRepository: A Community-Driven and Interactive Hypernetwork Data Collection

Publisher:

Springer Nature Switzerland

Published version:

DOI:10.1007/978-3-031-59205-8_11

(Article begins on next page)

HypergraphRepository:
A Community-driven and Interactive Hypernetwork Data Collection

Alessia Antelmi[1][0000−0002−6366−0546], Daniele De Vinco*[2][0000−0003−0781−3744], and Carmine Spagnuolo[2][0000−0002−8267−9808]

[1] Università degli Studi di Torino, Italy
alessia.antelmi@unito.it
[2] Università degli Studi di Salerno
{ddevinco,cspagnuolo}@unisa.it

Abstract. Hypergraph research has been thriving over the past few years, with a growing interest in a plethora of domains. Despite this remarkable surge, the lack of a comprehensive platform for searching and downloading diverse and well-curated datasets poses a significant obstacle to the continued advancement of the field. This absence hinders the ability of researchers and practitioners to validate and benchmark their hypergraph algorithms and models effectively.

To bridge this gap, we present HypergraphRepository, a web-based data collection aiming to serve as a centralized hub for hypergraph datasets, fostering collaboration and knowledge exchange within the hypergraph research community. In this paper, we detail the platform's architecture, features, and the collaborative framework it offers. HypergraphRepository is a GitHub-based open-source project.

Keywords: Hypergraph repository · Community-driven · Interactive · Web-based platform · Network collection.

1 Introduction

Hypergraphs are the natural representation of a broad range of systems where high-order relationships exist among their interacting parts. In technical terms, a hypergraph extends the concept of a graph, allowing a (hyper)edge to connect an arbitrary number of nodes [11]. This extension proves invaluable when dealing with systems exhibiting highly non-linear interactions among their constituents [6]. In practice, hypergraphs become essential for modeling complex group interactions that cannot be adequately described through dyads (and, hence, via graphs). This modeling approach is particularly apt for capturing social systems where individuals engage in groups of varying sizes [23]. For instance, in the case of a co-authorship collaboration network, a hyperedge may represent an article and link together all authors (nodes) having collaborated on it [3]. Further, hypergraphs prove useful for embedding sociological concepts to explore the dynamics of opinion formation [27] and social influence diffusion [28] when groups are explicitly taken into account, as well as for modeling epidemic-spreading processes to expressly consider community structure [8], and group dynamics [5]. Similar situations involving high-order interactions are evident in biology, ecology, and neuroscience [6].

In contrast to graphs, the literature on hypergraphs is in its adolescence thanks to a recent rise of systematic studies demonstrating how the transformation of a hypergraph to a classical graph either leads to an unavoidable loss of information or creates a large number of extra nodes/edges that increases space and time requirements in downstream graph analytic tasks [3]. This recent attention toward hypergraphs is reflected in the absence of one or more established repositories that offer readily available benchmark hypergraphs. The lack of standardized benchmark datasets may pose a challenge for researchers aiming to assess and compare claims, hypotheses, and algorithms specifically designed for hypergraphs in different application scenarios. Addressing this gap could potentially contribute to the growth of hypergraph research and foster a more robust understanding of their applications and implications across diverse fields while favoring the open science principles [12]. To this end, our paper introduces *HypergraphRepository*, the first open-source, community-driven, and interactive hypergraph collection. Our project stands on two fundamental pillars. First, its primary objective is to establish a dedicated space crafted by and for the community, allowing users to contribute by uploading their own datasets. This collaborative approach ensures that the repository evolves organically, reflecting the diverse needs and interests of the hypergraph research community. Second, the repository's interactive design aims to facilitate the exploration and comparison of a broad spectrum of datasets. The contribution of our paper is two-fold and can be summarized as follows:

- We introduce *HypergraphRepository*, the first systematic hypergraph data collection, where researchers can quickly and interactively compare, explore, and analyze data in real-time via a web-based platform. HypergraphRepository is open-source and available at `hypergraphrepository.di.unisa.it`.
- We provide an open-source community-driven data repository where users can contribute by sharing their own data and insight through a system of pull requests handled via git. The hypergraph database is available at `github.com/HypergraphRepository/datasets`.

The remainder of the paper is organized as follows. Section 2 reviews existing network collections. Section 3 describes HypergraphRepository, detailing the dataset creation and uploading pipeline and the currently available analytics features. Section 4 concludes this work, giving an overview of the current work and potential directions for further development.

2 Related Work

This section reviews existing network repositories.

Graph repositories. The most famous graph repositories are probably SNAP [21] and Network Repository [26]. The `SNAP collection`, managed by the Stanford Graph Learning Research Group, encompasses more than 200 real-world network datasets from diverse domains and of various types. These datasets are provided in multiple formats, ensuring their compatibility with a broad spectrum of graph analysis tools. Launched in July 2009, the SNAP website primarily hosts datasets gathered to serve the specific research objectives of the research group. `NetworkRepository` is a web-based data repository for interactively exploring, visualizing, and comparing a large number of networks. This platform also integrates social and collaborative features, enabling users to engage in discussions, share observations, and exchange visualizations related to each network. Currently, the repository contains more than 5000 networks sourced from 19 diverse categories (e.g., social, biological). These collections cover various network types, such as bipartite and time-series networks, and span domains like social sciences, physics, and bioinformatics. Both repositories come with their own programmatic libraries or tools for graph analysis and possible support for machine learning frameworks.

The `KONECT Project` [19,18] is another example of a platform where users can download network datasets and visualize their statistics online. The project is run by Jérôme Kunegis, and the entire source code is available as free software. The repository also includes a network analysis toolbox for GNU Octave, a network extraction library, as well as code to generate all statistics and plots shown on the website. Currently, The KONECT project boasts a collection of 1,326 network datasets in 24 categories.

Two websites where it is possible to download large graph datasets are the website of the `Laboratory for Web Algorithmics` [10,9] and Amazon's project `GraphChallenge` [1]. The former offers a diverse array of graph categories accompanied by comprehensive statistics, including plots detailing degree distribution, the size of the giant component, average and median distance, as well as the harmonic diameter. The latter refers to the Graph Challenge data sets available to the community free of charge as part of the AWS Public Data Sets program.

The `Open Graph Benchmark` (OGB) [15] and the `Illinois Graph Benchmark` (IGB) [16] are two recent yet already mature projects tailored for advancing graph machine-learning tasks, aiming to facilitate the creation of novel models and the comparison with existing approaches. The OGB repository covers diverse scale graphs from various domains, such as biological networks, molecular graphs, academic networks, and knowledge graphs. This project fully automates dataset processing, providing graph objects that are fully compatible with Pytorch Geometric [13] and DGL [29], as well as standardized dataset splits and evaluators that allow for easy and reliable comparison of different models in a unified manner. The IGB project specifically focuses on training and evaluating graph neural network models. In particular, it provides highly labeled graphs, including both homogeneous and heterogeneous large-scale real-world citation graphs, with more than 40% of their nodes labeled. IGB is open-sourced and, like OGB, supports DGL and Pytorch Geometric frameworks.

Hypergraph repositories. As of our current knowledge, there is a lack of a comparable repository dedicated to hypergraphs. The limited online resources typically consist of personal web pages where authors provide downloadable datasets used in their own work. Although this practice contributes to the reproducibility of their research, it falls short of establishing a universal platform. Moreover, this approach does not guarantee the expansion of dataset availability, as it relies on individual efforts. The most notable example is Professor Austin R. Benson's personal website [7]. Among other datasets, on this webpage, users can access 17 temporal hypergraphs, 11 node-labeled hypergraphs, 10 edge-labeled hypergraphs, and 2 hypergraphs labeled with a core-fringe structure. A short list of resources can also be found in the following pages [25,22,24,17].

3 HypergraphRepository

HypergraphRepository stands as the first open-source, community-driven web platform designed to interactively explore, compare, and download hypergraph data. The ultimate objective of our platform is to assist users throughout the workflow of hypergraph-related tasks by serving as a centralized hub where they can access various hypergraph types (e.g., temporal, directed, weighted interactions) from diverse application domains (e.g., social, transportation, biological networks). This approach is designed to streamline the initial stages of hypergraph processing, providing users with readily available hypergraphs tailored for different tasks, such as community detection, hyperedge prediction, or node classification/ranking.

The dual nature of being both open-source and community-driven opens up a two-fold avenue for user contribution. First, users can actively participate in dataset creation, modification of existing datasets and take on roles as community reviewers. Second, users have the opportunity to contribute directly to the platform itself by engaging in discussions about potential updates to the repository. This collaborative process empowers users to request and add new features to the website, including the integration of novel interactive plots. The entire workflow, from dataset management to platform enhancements, operates seamlessly through the GitHub platform, fostering a transparent and collaborative environment. Specifically, users can utilize the pull request system on GitHub to propose modifications and improvements. This strategy not only promotes accessibility but also enhances the overall quality and integrity of the datasets. By implementing a stringent review mechanism as part of the submission process, the platform ensures that the available hypergraphs adhere to high standards driven by the community, thus enriching the data's exploitability for current research challenges.

Technically speaking, our platform is made up of two main software components: *(i)* a dataset manager and *(ii)* an interactive analytics manager. The dataset manager is responsible for handling all aspects related to datasets, spanning from their creation to their possible publication on the site. This includes various tasks such as dataset organization, maintenance, and ensuring seamless accessibility for users. The interactive analytics manager is dedicated to the computation of hypergraph statistics, dataset search, and comparison. This component provides users with robust capabilities for exploring and deriving insights from the datasets through interactive and analytical features. Together, these components synergize to offer a comprehensive and user-friendly environment for managing, analyzing, and sharing hypergraph data on our platform.

In the following, we detail our platform, its underlying hypergraph model (see Section 3.1), and its features, with a particular emphasis on the dataset creation and uploading pipeline (see Section 3.2). Additionally, we delve into the currently accessible analytics functionalities (see Section 3.3). More technical details about how to contribute can be found in the repositories of the platform[3], the hypergraph database[4], or on the website FAQ page[5].

3.1 Hypergraph representations in HypergraphRepository

Under the hood, HypergraphRepository exploits the SimpleHypergraphs.jl Julia library to represent and analyze the hypergraph datasets [2,4]. This library represents a hypergraph $H = (V, E)$ as an $n \times k$ matrix, where n is the number of vertices and k is the number of hyperedges. Vertices and hyperedges are uniquely identified by progressive integer ids, corresponding to rows $(1, \ldots, n)$ and columns $(1, \ldots, k)$, respectively. Each position (i, j) of the matrix denotes the weight of the vertex i within the hyperedge j. SimpleHypergraphs.jl also provides several constructors for defining meta-information type and enables the attachment of meta-data values of arbitrary type to both vertices and hyperedges. In this manner, the library naturally models a wide range of hypergraph types, such as heterogeneous, weighted, attributed, and temporal hypergraphs.

Currently, SimpleHypergraphs.jl offers two mechanisms to load and save a hypergraph from or to a stream. In our framework, we use the plain text storage type (named HGF format). In this format, the first line consists of two integers n and k, representing the number of vertices and the number of hyperedges of H, respectively. The following k rows describe the structure of H; specifically, each line represents a hyperedge as a list of all vertex-weight pairs within that hyperedge. While this format's simplicity facilitates seamless interoperability among hypergraph software libraries, its drawback lies in the necessity to define supplementary information (such as vertex metadata or hyperedge weights) in separate files. Still, it is worth noting that the framework-agnostic nature of the HGF format means it is not bound to the development of any specific library, including SimpleHypergraphs.jl, in our case. For instance, this flexibility allows users to upload directed hypergraph datasets, even though SimpleHypergraphs.jl currently lacks support for these structures.

[3] `github.com/HypergraphRepository/website`
[4] `github.com/HypergraphRepository/datasets`
[5] `hypergraphrepository.di.unisa.it/f-a-q`

3.2 A community-driven hypergraph collection

As previously discussed, our project's primary aim is to provide the research community with a comprehensive and expansive hypergraph collection that can evolve with the community's interests. This approach enables users to actively contribute by uploading their own datasets. Figure 1 summarizes the dataset management pipeline.

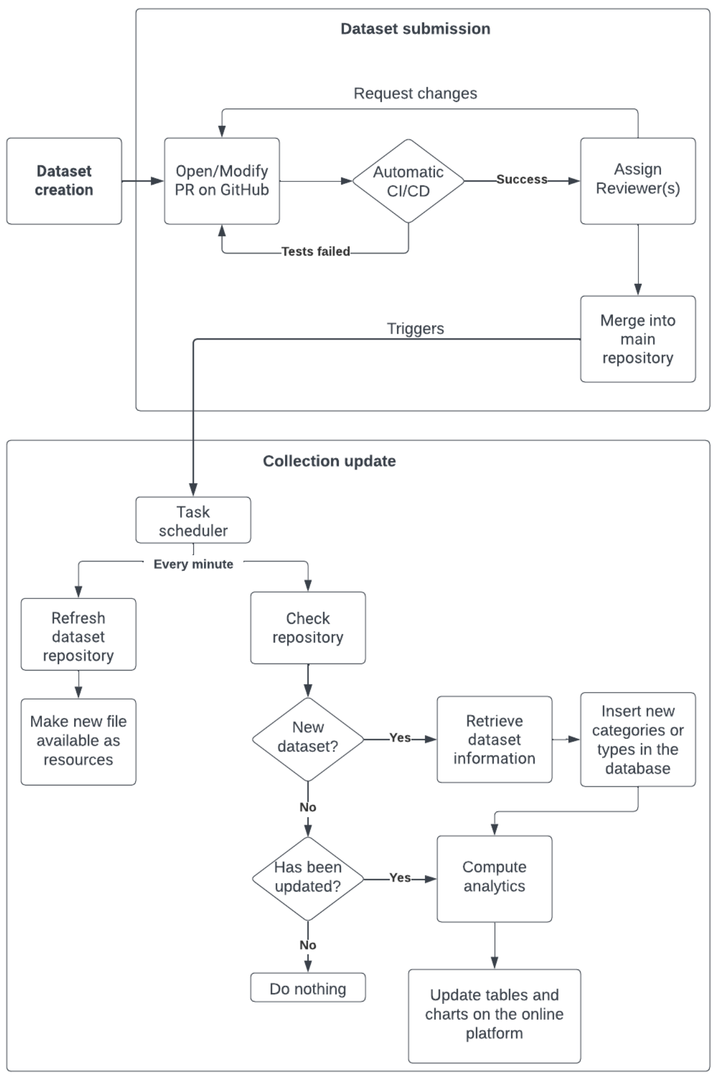

Fig. 1: Workflow of the dataset management pipeline.

Given that the dataset repository, storing all the hypergraphs accessible on the platform, is publicly accessible, any user can propose modifications. To facilitate this collaborative effort, we established a designated template to guide the process of uploading a new dataset into the repository. The overall pipeline comprises three main phases, described below.

Dataset creation phase. The first step of the pipeline is delegated to the end user who intends to upload a novel dataset. The same process applies if a user wants to modify existing hypergraph data. Upon forking the current dataset repository, the user has to create a new folder for each additional dataset they wish to contribute. This folder must include the hypergraph stored in the HGF format, a markdown-formatted file offering a concise dataset description along with any supplementary information the user deems pertinent, and a metadata file describing the hypergraph's type (e.g., homogeneous, temporal) and its application domain (e.g., social network, infrastructure network). These three files are essential for initiating a successful pull request (PR) on the main repository. Moreover, the user has the option to upload further files describing hypergraph features, such as node and vertex labels, timestamps, and hyperedge weights.

Dataset submission. Once the user is satisfied with the dataset and any supplementary data generated, they can submit all components to the central dataset repository. Specifically, the user must initiate a PR via the GitHub system, adhering to the provided template that outlines a checklist to be fulfilled before officially requesting a merge. Figure 2 illustrates the template.

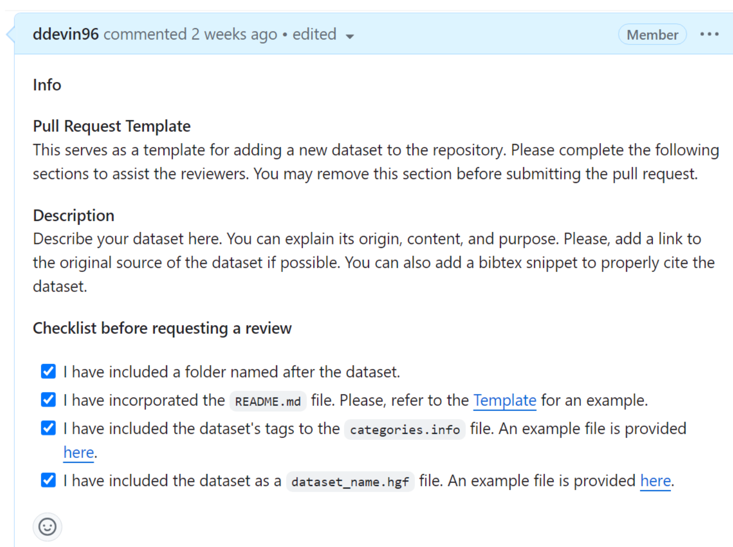

Fig. 2: PR template for adding or modifying a dataset.

Upon opening the PR, the first round of Continuous Integration/Continuous Development (CI/CD) starts. This automated process allows for validating the PR integrity before assigning one or more reviewers. This mechanism is handled through GitHub actions, which trigger specific Python scripts. The first action verifies the presence of all requested files and ensures that their format aligns with expectations (e.g., the hypergraph data adheres to the HGF format). If any checks fail, the action will halt, and the bot will promptly report the missing elements (as a comment on the PR), delineating necessary revisions. If all these conditions are satisfied, two additional actions come into play: assigning the appropriate label(s) to the PR (e.g., creation, change, documentation) and the subsequent updating of the PR after each new commit. Simultaneously, the final action, responsible for assigning a reviewer to the PR, is triggered. Each reviewer is chosen randomly from a list of volunteers, making this step a highly community-driven process open to collaboration from anyone interested. At this point, the assigned reviewer(s) can verify whether all uploaded files are compliant with the guidelines, and eventually merge the new/modified dataset into the main repository. Taking place on the GitHub platform, it is worth highlighting that the overall review process is public and unblinded. Figure 3 shows the GitHub actions pipeline.

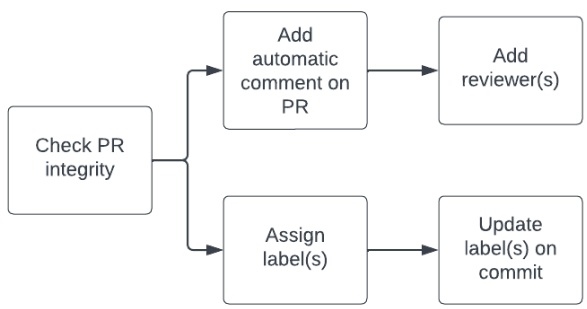

Fig. 3: CI/CD pipeline for validating PR integrity during dataset addition or modification, along with the assignment of at least one reviewer.

Collection update. The final step of the pipeline is entirely automated and is once again managed through CI/CD automation. This phase starts upon the successful merging of the PR into the main repository. Specifically, the server is configured to execute two primary scripts every minute.

– A Bash script triggers a pull operation on the dataset repository, capturing every change made. Then, it updates a folder associated with the public online collection, ensuring that all dataset files are promptly updated and made accessible for download.

- A Python script orchestrates all system calls, interacts with the GitHub APIs, and updates the database. In particular, this script makes an external call to a Julia script responsible for computing all metrics, statistics, and plots for the newly added datasets. Leveraging the SimpleHypergraphs.jl library, the Julia script operates behind the scenes, enabling the generation, manipulation, and analysis of hypergraphs at runtime. If an existing hypergraph has been modified, the script verifies which properties have changed and subsequently updates all fields associated with those changes. To accomplish this, we utilize the GitHub API to retrieve the commit timestamps and verify which files have been updated. The coordination between these two scripts ensures the seamless update of newly computed values on the platform's website.

3.3 An interactive hypergraph repository

This section overviews the features currently available on the platform. The design of these functionalities draws inspiration from established graph repositories to provide users with a user-friendly and intuitive experience as they interact with the web interface. At the time of writing, the web platform hosts two main web pages. Table 1 summarizes the statistics currently available on the website.

Table 1: Currently available statistics.

Symbols	**Description**
$\|V\|$	Number of nodes
$\|E\|$	Number of hyperedges
d_{max}	Maximum node degree
d_{avg}	Average node degree
d_{median}	Median node degree
e_{max}	Maximum hyperedge size
e_{avg}	Average hyperedge size
e_{median}	Median hyperedge size

- The homepage provides users with an overview of the datasets accessible on the platform, including key information such as the number of available hypergraphs, the variety of hypergraph types and categories, and a summary table detailing the notation conventions in use. Additionally, it lists the latest updated hypergraphs. As shown in Figure 4a, users can access details about dataset names, authors, categories, types, and the number of nodes and hyperedges. The table also includes timestamps indicating when each dataset was initially uploaded and when it was last modified. Finally, for a more in-depth exploration of a specific hypergraph, users can click on the eye icon to access its specific page.

Latest updated graphs

Search

Name	Author	Category	Type	|V|	|E|	Updated at	Created at
algebra	ddevin96	Collaboration network	undirected	423	1,268	Jan 5, 2024 15:43:50	Dec 20, 2023 10:50:03
amazon	ddevin96	Online reviews	undirected	5,000	1,176	Jan 5, 2024 15:43:50	Dec 20, 2023 11:36:09
dblp	ddevin96	Collaboration network	undirected	71,116	25,624	Jan 5, 2024 15:57:00	Dec 20, 2023 11:36:09
email-Enron	ddevin96	Email network	undirected	2,807	5,000	Jan 5, 2024 15:57:00	Dec 20, 2023 11:36:09
email-W3C	ddevin96	Email network	undirected	5,601	6,000	Jan 5, 2024 15:43:50	Dec 20, 2023 11:36:09

Per page 5 Next

(a) Partial view of the HypergraphRepository's homepage.

Group by

Search

Name	Type	|V|	|E|	d_{max}	e_{max}	d_{avg}	e_{avg}	d_{med}	e_{med}	
algebra	undirected	423	1,268	375	107	19.532	6.516	10	4	View Download (0.07 MB)
amazon	undirected	5,000	1,176	4	6	1.022	4.347	1	6	View Download (0.05 MB)
dblp	undirected	71,116	25,624	25	69	1.244	3.452	1	3	View Download (0.96 MB)
email-Enron	undirected	2,807	5,000	786	25	7.662	4.301	2	2	View Download (0.2 MB)
email-W3C	undirected	5,601	6,000	282	23	2.385	2.227	1	2	View Download (0.12 MB)

Showing 1 to 5 of 16 results Per page 5 1 2 3 4 >

(b) Partial view of the HypergraphRepository's dataset page.

Fig. 4: The HypergraphRepository's website.

– The dataset page showcases all hypergraphs within the collection, as shown in Figure 4b. Beyond the information available on the homepage, this table provides additional insights, such as the maximum, average, and median node degree, as well as the maximum, average, and median hyperedge size. Users can download a specific hypergraph of interest directly from this page, with the file size conveniently displayed for reference.

Interactive hypergraph search and comparison. An essential feature for any online data repository is the provision of two core functionalities: the ability to search for specific datasets and the capability to compare them. In the following, we elaborate on the implementation of these features within our platform.

Regardless of the page users visit, they can search for a specific dataset through the tables presented in Figures 4a and 4b. More precisely, users can search for a hypergraph based on its name, author, category, or type. Additionally, while navigating the dataset page, users can group datasets based on author, application domain (i.e., category), or type. Further, users can apply filters to hypergraph types and categories and define minimum thresholds for the number of nodes or hyperedges. As illustrated in Figure 5, users can customize the settings of the main table to identify the most relevant datasets for their specific use cases. All numeric values are sortable in the table view, enhancing the user's ability to navigate and explore the datasets efficiently.

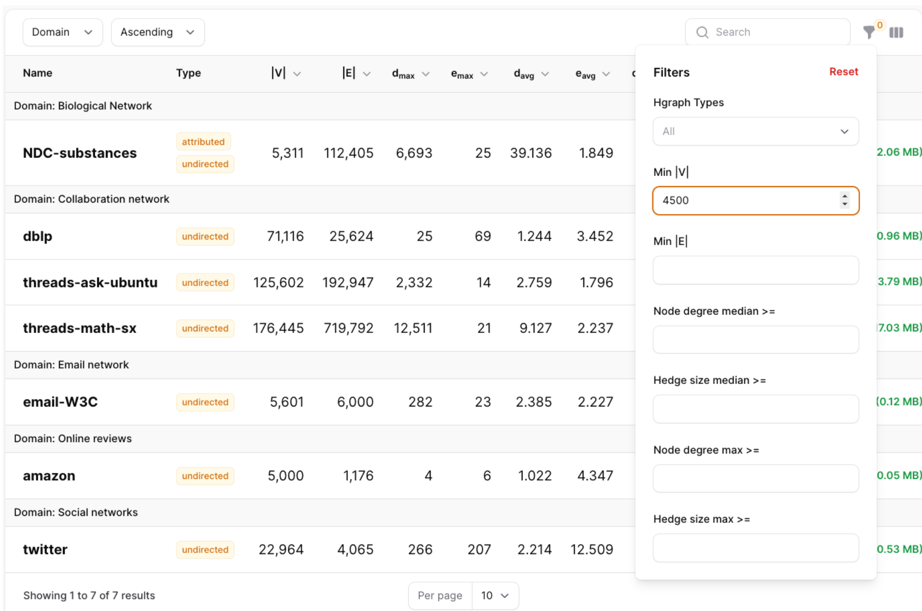

Fig. 5: Example of *filtering* and *group by* operations.

The tables accessible on both the home and dataset pages facilitate seamless dataset comparison for users. Leveraging the table headers introduced earlier (see Figure 4), users can effortlessly compare the datasets within the collection along the dimensions defined. Although trivial, this approach offers users a bird's-eye view of the primary characteristics of the datasets. A more in-depth comparison across datasets is the object of future work.

Interactive hypergraph analytics. Users can access a detailed view of each dataset by clicking on the "view" button, as illustrated in Figure 6. Currently, the user may access two different tabs reporting information about the dataset.

- The first tab reports all statistics computed for the given hypergraph, details about the dataset creator, and the dates of both the upload and the most

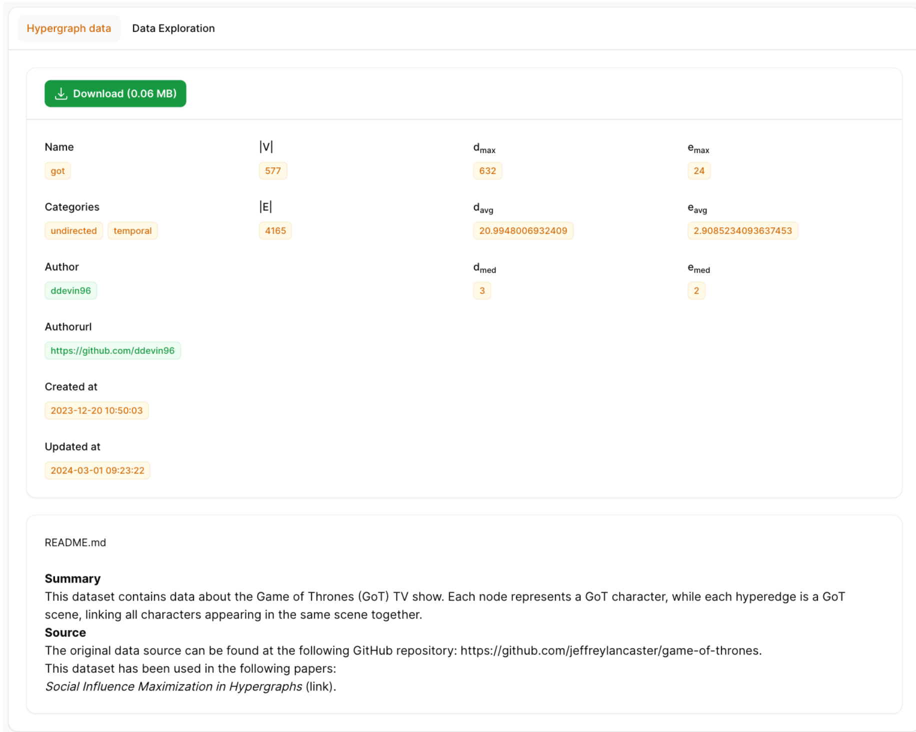

Fig. 6: A snapshot of the web page reporting the details of a hypergraph.

recent modification (retrieved through the GitHub API). This page also renders the README file with all the information provided at the time of the submission. Users are also provided with the option to download the dataset directly from this page.

– The second tab collects the interactive visualizations, providing users with the ability to visualize various hypergraph properties. This functionality is achieved by exploiting Chart.js[6], a widely acclaimed and highly customizable JavaScript library for data plotting. The incorporation of this library is designed to facilitate the development of community-driven custom plugins and the integration of features into our platform. Currently, our platform supports the visualization of node degree and hyperedge size distribution, visualized via a histogram and a scatter plot on a log-log scale, as shown in Figure 7.

[6] https://www.chartjs.org/

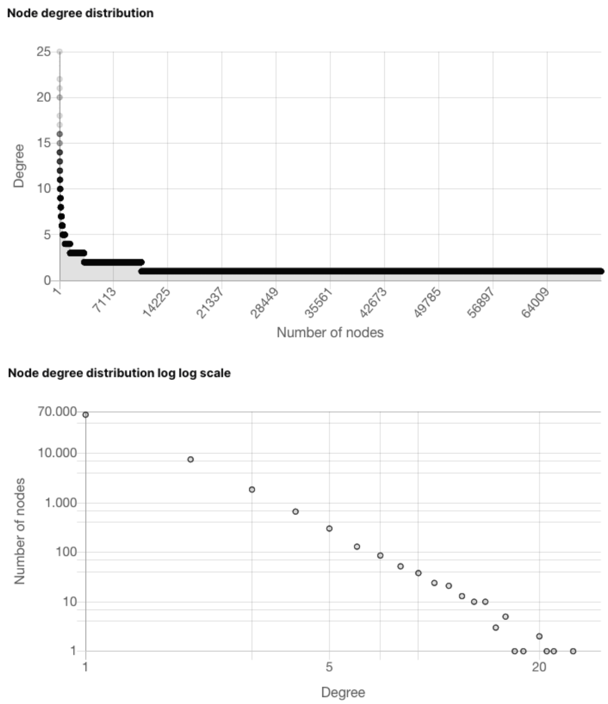

Fig. 7: Examples of node-related interactive charts.

4 Conclusion

This work presented HypergraphRepository, the first open-source, community-driven hypergraph repository. Implemented as a web-based platform, our project aims to serve as a centralized hub for hypergraph datasets, fostering collaboration, data quality assurance, and knowledge exchange within the hypergraph research community. HypergraphRepository is in its infancy, and our ongoing efforts focus on expanding the hypergraph collection in terms of both diversity in the nature and size of hypergraphs. In future updates, we also intend to:

- Enhance the set of statistics describing each hypergraph, such as adding descriptors for directed hypergraphs, information on the community composition, clustering coefficient, and high-order motifs [20];
- Offer users the option to select specific subsets of hypergraphs and directly compare them on the platform;
- Introduce interactive and scalable visualization features [14];
- Offer programmatic API access to the hypergraph database.

Acknowledgments. This work has been partially supported by the spoke "FutureHPC & BigData" of the ICSC – Centro Nazionale di Ricerca in High-Performance Computing, Big Data and Quantum Computing funded by European Union – NextGenerationEU.

References

1. Amazon, M.: GraphChallenge - Data sets. https://graphchallenge.mit.edu/data-sets (2017), accessed on 4th January 2024
2. Antelmi, A., Cordasco, G., Kamiński, B., Prałat, P., Scarano, V., Spagnuolo, C., Szufel, P.: Analyzing, Exploring, and Visualizing Complex Networks via Hypergraphs using SimpleHypergraphs.jl. Internet Mathematics **1**, 1–32 (2020). https://doi.org/10.24166/im.01.2020
3. Antelmi, A., Cordasco, G., Polato, M., Scarano, V., Spagnuolo, C., Yang, D.: A Survey on Hypergraph Representation Learning. ACM Computing Surveys **56**(1) (2023). https://doi.org/10.1145/3605776
4. Antelmi, A., Cordasco, G., Kamiński, B., Prałat, P., Scarano, V., Spagnuolo, C., Szufel, P.: SimpleHypergraphs.jl—Novel Software Framework for Modelling and Analysis of Hypergraphs. In: Avrachenkov, K., Prałat, P., Ye, N. (eds.) Algorithms and Models for the Web Graph. pp. 115–129. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-25070-6_9
5. de Arruda, G.F., Petri, G., Moreno, Y.: Social contagion models on hypergraphs. Physical Review Research **2**, 023032 (2020). https://doi.org/10.1103/PhysRevResearch.2.023032
6. Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.G., Petri, G.: Networks beyond pairwise interactions: Structure and dynamics. Physics Reports **874**, 1–92 (2020). https://doi.org/10.1016/j.physrep.2020.05.004
7. Benson, A.R.: Data! https://www.cs.cornell.edu/ arb/data/ (2021), accessed on 30th December 2023
8. Bodó, Á., Katona, G.Y., Simon, P.L.: SIS Epidemic Propagation on Hypergraphs. Bulletin of Mathematical Biology **78**, 713–735 (2016). https://doi.org/10.1007/s11538-016-0158-0
9. Boldi, P., Rosa, M., Santini, M., Vigna, S.: Layered Label Propagation: A Multiresolution Coordinate-Free Ordering for Compressing Social Networks. In: Proceedings of the 20th International Conference on World Wide Web. p. 587–596. WWW '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/1963405.1963488
10. Boldi, P., Vigna, S.: The Webgraph Framework I: Compression Techniques. In: Proceedings of the 13th International Conference on World Wide Web. p. 595–602. WWW '04, Association for Computing Machinery, New York, NY, USA (2004). https://doi.org/10.1145/988672.988752
11. Bretto, A.: Hypergraph Theory: An Introduction. Springer International Publishing (2013)
12. Commission, E.: Open science. https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en (2020), accessed on 29th December 2023
13. Fey, M., Lenssen, J.: Fast Graph Representation Learning with PyTorch Geometric. CoRR **abs/1903.02428** (2019), workshop paper at ICLR 2019
14. Fischer, M., Frings, A., Keim, D., Seebacher, D.: Towards a survey on static and dynamic hypergraph visualizations. In: 2021 IEEE Visualization Conference (VIS). pp. 81–85 (2021). https://doi.org/10.1109/VIS49827.2021.9623305
15. Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., Leskovec, J.: Open Graph Benchmark: Datasets for Machine Learning on Graphs. In: Advances in Neural Information Processing Systems. vol. 33, pp. 22118–22133. Curran Associates, Inc. (2020)

16. Khatua, A., Mailthody, V.S., Taleka, B., Ma, T., Song, X., Hwu, W.: IGB: Addressing The Gaps In Labeling, Features, Heterogeneity, and Size of Public Graph Datasets for Deep Learning Research. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. p. 4284–4295. KDD '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3580305.3599843
17. Kim, S., Lee, D., Kim, Y., Park, J., Hwang, T., Shin, K.: Datasets, tasks, and training methods for large-scale hypergraph learning. Data Mining and Knowledge Discovery **37**(6), 2216–2254 (2023). https://doi.org/10.1007/s10618-023-00952-6
18. Kunegis, J.: KONECT: The Koblenz Network Collection. In: Proceedings of the 22nd International Conference on World Wide Web. p. 1343–1350. WWW '13 Companion, Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2487788.2488173
19. Kunegis, J.: The KONECT Project. http://konect.cc/ (2017), accessed on 30th December 2023
20. Lee, G., Ko, J., Shin, K.: Hypergraph motifs: concepts, algorithms, and discoveries. Proc. VLDB Endow. **13**(12), 2256–2269 (2020). https://doi.org/10.14778/3407790.3407823
21. Leskovec, J., Krevl, A.: SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data (Jun 2014)
22. Li, P.: Datasets. https://sites.google.com/view/panli-purdue/datasets (2020), accessed on 30th December 2023
23. Lotito, Q.F., Musciotto, F., Montresor, A., Battiston, F.: Higher-order motif analysis in hypergraphs. Communications Physics **5**(1), 79 (2022). https://doi.org/10.1038/s42005-022-00858-7
24. Macgregor, P.: Datasets. https://pmacg.io/datasets.html (2021), accessed on 30th December 2023
25. Papachristou, M.: Datasets. https://papachristoumarios.github.io/datasets/ (2020), accessed on 30th December 2023
26. Rossi, R., Ahmed, N.: The Network Data Repository with Interactive Graph Analytics and Visualization. In: AAAI (2015), https://networkrepository.com
27. Sahasrabuddhe, R., Neuhäuser, L., Lambiotte, R.: Modelling non-linear consensus dynamics on hypergraphs. Journal of Physics: Complexity **2**(2), 025006 (feb 2021). https://doi.org/10.1088/2632-072X/abcea3
28. Wang, R., Li, Y., Lin, S., Xie, H., Xu, Y., Lui, J.C.S.: On Modeling Influence Maximization in Social Activity Networks under General Settings. ACM Transactions on Knowledge Discovery from Data **15**(6) (2021). https://doi.org/10.1145/3451218
29. Zheng, D., Wang, M., Gan, Q., Song, X., Zhang, Z., Karypis, G.: Scalable Graph Neural Networks with Deep Graph Library. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining. p. 1141–1142. WSDM '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3437963.3441663