



HAL
open science

Retrieving Adversarial Cliques in Cognitive Communities: A New Conceptual Framework for Scientific Knowledge Graphs

Renaud Fabre, Otmane Azeroual, Patrice Bellot, Joachim Schöpfel, Daniel Egret

► **To cite this version:**

Renaud Fabre, Otmane Azeroual, Patrice Bellot, Joachim Schöpfel, Daniel Egret. Retrieving Adversarial Cliques in Cognitive Communities: A New Conceptual Framework for Scientific Knowledge Graphs. *Future internet*, 2022, 14 (9), pp.262. 10.3390/fi14090262 . hal-03772163

HAL Id: hal-03772163

<https://hal.univ-lille.fr/hal-03772163>

Submitted on 28 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Article

Retrieving Adversarial Cliques in Cognitive Communities: A New Conceptual Framework for Scientific Knowledge Graphs

Renaud Fabre ¹, Otmane Azeroual ^{2,*}, Patrice Bellot ³, Joachim Schöpfel ⁴ and Daniel Egret ⁵¹ Dionysian Economics Laboratory (LED), University of Paris 8, 93200 Saint-Denis, France² German Centre for Higher Education Research and Science Studies (DZHW), 10117 Berlin, Germany³ CNRS, LIS, Aix Marseille University (AMU), 13007 Marseille, France⁴ GERiiCO-Labor, Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication, University of Lille, 59000 Lille, France⁵ Observatoire de Paris, Paris Sciences & Lettres University (PSL), 75006 Paris, France

* Correspondence: azeroual@dzhw.eu; Tel.: +49-30-206417738

Abstract: The variety and diversity of published content are currently expanding in all fields of scholarly communication. Yet, scientific knowledge graphs (SKG) provide only poor images of the varied directions of alternative scientific choices, and in particular scientific controversies, which are not currently identified and interpreted. We propose to use the rich variety of knowledge present in search histories to represent cliques modeling the main interpretable practices of information retrieval issued from the same “cognitive community”, identified by their use of keywords and by the search experience of the users sharing the same research question. Modeling typical cliques belonging to the same cognitive community is achieved through a new conceptual framework, based on user profiles, namely a bipartite geometric scientific knowledge graph, SKG GRAPHYP. Further studies of interpretation will test differences of documentary profiles and their meaning in various possible contexts which studies on “disagreements in scientific literature” have outlined. This final adjusted version of GRAPHYP optimizes the modeling of “Manifold Subnetworks of Cliques in Cognitive Communities” (MSCCC), captured from previous user experience in the same search domain. Cliques are built from graph grids of three parameters outlining the manifold of search experiences: mass of users; intensity of uses of items; and attention, identified as a ratio of “feature augmentation” by literature on information retrieval, its mean value allows calculation of an observed “steady” value of the user/item ratio or, conversely, a documentary behavior “deviating” from this mean value. An illustration of our approach is supplied in a positive first test, which stimulates further work on modeling subnetworks of users in search experience, that could help identify the varied alternative documentary sources of information retrieval, and in particular the scientific controversies and scholarly disputes.

Keywords: community detection; cliques; graph completion; graph subnetwork; model interpretability; meta learning; search history; entity alignment; multiplex



Citation: Fabre, R.; Azeroual, O.; Bellot, P.; Schöpfel, J.; Egret, D. Retrieving Adversarial Cliques in Cognitive Communities: A New Conceptual Framework for Scientific Knowledge Graphs. *Future Internet* **2022**, *14*, 262. <https://doi.org/10.3390/fi14090262>

Academic Editors: Diego Reforgiato Recupero and Alessandro Sebastian Podda

Received: 3 August 2022

Accepted: 2 September 2022

Published: 7 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction: Representing the Manifold of Cognitive Communities with Their Adversarial Cliques

The variety and diversity of published content are currently expanding in all areas of science, with the simultaneous growth of interdisciplinarity. Powerful new tools and new technical infrastructures such as scientific knowledge graphs (SKG) have been developed [1], to help users navigate the flood of scientific information. However, the search experience requires more precision because retrieval systems do not benefit from a rich panel of content descriptors, in the context of model-based information retrieval allowing personalized answers to queries. At the same time, captured queries are rich in a knowledge manifold that could be exploited to the benefit of a more personalized and efficient search. To achieve this result in the future we explore in this article three conditions: the

first condition is to design a “cognitive community” to represent on knowledge graphs all the cliques of users of the same keyword; the second condition is to model, inside each community, a classifier of the interacting cliques, specifying each possible type of documentary need of the users of available items; the third condition is to optimize the efficient information use, by allowing all users of a keyword to access the mapping of all registered cliques to help them, if necessary, to refine or modify their choice.

To fulfill that program, in the following, we will call a group of researchers sharing similar search practices and neighbor documentation routes a “cognitive community”. Within the perimeter of these cognitive communities that share a common research question, our objective is to identify cliques sharing similar expectations, and accessing common articles, or, conversely, to identify divergent cliques in communities, through the analysis of differences in their uses and consultations. For that, we propose here to exploit as basic data the “logs” of databases and documentary services so that it becomes possible to identify the query strategies used by researchers during their consultation of documentary resources. In cases where knowledge gives rise to controversies and disputes, it seems particularly useful to distinguish cliques in the “cognitive communities” taking one side or the other of the controversy (imagine here, for example, the controversies over the use of a medical treatment or vaccination strategies giving rise to debate). Knowledge graphs help represent the alternative routes taken by the search experience of identified cliques. *Note that the identification of the features of the user experience does not in any case constitute a scientific evaluation:* the knowledge graph that we propose represents the “adversarial” documentary pathways without prejudging their use and their scientific interpretation in any way. Future studies will test differences and systemic classification of “documentary profiles” of the users, their meaning in various possible contexts, as proposed by a massive recent study (4M papers) of “disagreements in scientific literature” [2]. This study enumerates the obstacles to a direct characterization of scientific disagreements and exposes the helpful results of various operational approaches to search profiles and to the extraction of documentary features.

In this article, we propose to measure the practices of identifiable cliques in a cognitive community, by retaining three parameters: mass of users, intensity of items and attention of users to items. (Ratio of mass to intensity creates an «attention» parameter that could be measured in search experience: from mean behaviour of the users in the uses of items (readings, downloadings, clicks, etc.); it can be computed as we did, from observed values remaining or deviating from the mean. In our modeling we distinguish two typical situations: one where the ratio user/items tends to approach the mean of observed results, and the other where the values of the «attention» parameter, tend to deviate from the mean. «Attention» has been recently outlined as a parameter of «feature augmentation» in information retrieval literature (see [3]).) Such a combination of documentary features makes it possible to build triplets structuring a community, and cliques inside a community, with the help of a specifically designed hypergraph that we describe below. Document consultation logs have been used for a long time to model user profiles. We adopt a similar approach with our methodology for characterizing documentary practices. However, to the best of our knowledge, we bring a new conceptual solution for profiling cliques inside communities of knowledge, stemming from the modelization of all query results obtained from a common research question.

In all three sequences of this approach, we have to specify an interaction between “community” and “cliques” that has to be semantically contextualized; finding a community structure in networks needs to split the network in elementary groups, then merge subgraphs by a modeling of detected cliques [4]. We define in this article a “cognitive community” as the community of users of the same keyword, and “cliques” as subnetworks inside this network (see Section 2). Indeed, cliques are themselves communities of practice (here, search experience) [5]; however, they find their semantic root in a first ranked community, here referred to as the “cognitive community”, that, in a nutshell, gathers all the users of the same keyword. All the cognitive communities live today with the same hopes and documentary challenges.

Recently deployed tools, in principle, constitute solutions for the knowledge representation problem that we are exploring in this study; these are scientific knowledge graphs (SKGs) which already offer many innovative services in support of science, and which also address the challenges of evaluating the impact of research [1].

SKGs create interactions of nodes that explore information spaces representing research results. Unfortunately, SKGs suffer from the flaw of being barely designed to compare the diverse pathways to information, which often lead to contradictory results, a situation frequently encountered in scientific approaches, and more specifically in the context of scientific controversies. This insufficient data capture severely limits the expressive power of the SKGs in representing the research workflow in its entirety and versatility as, indeed, the results of a documentary query may involve quite different methods, assumptions, or theoretical approaches and conditions.

This article explores emerging SKGs technologies leveraging the variety of possible answers to a research question; capturing this rich diversity opens new tools to identify the scientific approaches of specific cliques inside cognitive communities sharing a common profile in their search for information. These cliques can be delineated as entities sharing similar pathways to answers, revealed by data captured in their search history on a given research question. Data capture is very efficient because meaningful structures partitioned into cliques inside communities reveal strongly connected groups of online documents [6]. Moreover, by identification of cliques inside cognitive communities, we exceed the current range of retrieval perimeters that propose current methods of citation analysis (see Sections 2 and 3 below).

The expression of the diversity of cognitive communities, possibly engaged in contradictory arguments of a scholarly dispute, has been traditional in academia, since medieval times, because it strongly contributes to the advancement of knowledge through the confrontation of contradictory opinions, and the systematic refutation of opposing statements. However, while this immemorial practice makes it possible to follow and keep track of the knowledge pathways leading to scientific results, it seems paradoxical in the digital age that the expression of scholarly disputes is barely achievable with current SKG architectures. Indeed, data structures that would identify the expressive power of adversarial scientific communities (i.e., groups of researchers who appear clustered through their common usage of different sets of documents, each set being representative of one side of a scientific dispute) are not commonly available today. (We will show in Section 2 how recent literature has taken up the new concept of “generative adversarial networks” in the context of machine learning.) The search for a solution to the problem of representing and delimiting alternative scientific paths and outcomes is addressed in many recent publications (see, e.g., [7,8]) with the aim to alleviate obstacles to recall and integrity in the assessment of research impact, but not to capture the representation of the controversy itself, as we do. Our research goal is to make visible in the knowledge graphs not only the scientific result itself, but also the complete path towards new knowledge, with its various cliques, and its place among other alternative paths.

Insufficient performance towards data capture of SKGs has well-known consequences, as it prevents research impact assessment from delineating the real-world behaviors of the challenging “invisible colleges” [9] in scholarly publication. Nevertheless, the need exists, and recording user experience in a diversity of knowledge expressions is a well-known claim: the Open Knowledge Research Graph project [10] recommends the use of “techniques that acquire scholarly knowledge in machine actionable form as knowledge is generated along the research lifecycle”. This goal is considered a hot priority: “Organizing scholarly knowledge is one of the most pressing tasks for solving current and upcoming societal challenges” [10].

1.1. Research Question

Our ambition is to propose solutions to represent all the scientific choices available to the researcher in her search for information, choices determined by the documentary

practices commonly used, and established by collecting the query paths previously used by researchers. This research goal is the subject of many recent developments that we will attempt to briefly discuss below.

On one side, the endemic growth in the variety and diversity of published content is a feature of contemporary knowledge construction practices, as can be seen from the emergence of new metrics of “atypicality” and “disruption” in the assessment of scientific publication; these new indices testify to the positive weight of “disconnection” and “discord” [11] on the construction and advancement of fruitful orientations in scientific ideas and projects. Moreover, the interactions between categories of scientific vocabularies, measured by metrics of the “cognitive extent of science”, with a method based on the lexical diversity of titles of fixed quotas of research articles [12], denote greater interdisciplinarity with larger networks of collaborations, in most disciplines.

On the other side, a recent comprehensive analysis of trends in SKG outcomes [1] observes that SKG efficiency is primarily significant in global encyclopedic databases of scientific results, with the ability to enrich “classic metadata” about publications. However, knowledge graph “incompleteness” causes serious obstacles to the expressive power of graphs as a whole [13] (we refer to the definition of graphs given in the article quoted above, where graphs «accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities»), and even more in the applications of scientific knowledge graphs [14] and namely in full research activities, where SKGs perform poorly (Open Research Knowledge Graph (<https://www.orkg.org/orkg/> 2 August 2022), MAG...), “due to the heterogeneity, inhomogeneity, and evolution of scholarly communication” [8], while it is clear that, “these KGs are ambiguous due to a lack of standard terminology used across the literature and poses domain-specific challenges for KG completion task”. Trade-offs are sought between the reduction of parsimony and the admission of complexity in the representation of scientific results, which are “incomplete by design” as science explores always a supposedly missing path [15]. Moreover, existing solutions delivered by the SKGs “are still relatively static” [16].

Considering these challenges to the contributions of SKGs to the assessment of research impact, our study addresses the issue of understanding insufficient data capture in knowledge representations, in search interfaces, and in matching structured objects in graphs applied to communities of cognitive information, and to their subnetworked cliques. As a scientific baseline, we refer, as the application of C.E. Shannon’s theory remarks, to the idea that, “a computational scheme of a cognitive process, may itself be deemed as a form of cognition” [17]. In that direction, we note also that a “cognitive information space” has been mathematically described and is shaped with differentiable manifolds, representing as proposed here for cliques, a “manifold atlas topology” which serves as a descriptive “organism” of modeled cognitive information [17]. For this, our article will mainly follow SKG representation approaches combining topological connections and interpretable vectors that participate in the program of geometrically equivariant graph neural networks whose mathematical properties are recalled in the survey by [18].

1.2. Main Findings

Our findings include the first test of a new conceptual framework for SKG, hereafter called GRAPHYP, specifically designed for the representation of cognitive communities and their cliques, distinguished by their search practices. Instead of delivering a single AI-processed answer to a research question, SKG GRAPHYP represents the landscape of conflicting (or adversarial) answers, which a query on a research question might reveal. For that purpose, GRAPHYP adopts new approaches in output representation and interpretability, with a novel methodology, called MSCCC, for designing “Manifold Subnetworks of Cliques in Cognitive Communities”. GRAPHYP captures documentary selections from SKG users, allowing classification of their search paths in a given research field. Users are detected from the manifold of their search practices and classified in cliques inside “cognitive communities” from the search history of their logs of scientific documentation.

The manifold of practices is expressed from metrics of differentiated uses of documentary resources by triplets of nodes shaped in graph subnetworks, with the following three parameters that we will use to build the graph: mass, intensity, and attention.

We selected parameters that could handle a large amount of data without distortion and which could altogether be detailed or summarized. Mass and intensity, as basic descriptors of users and items, allowed us to ground our approach on analytic literature on behaviors of information retrieval; attention, with our quantitative approach (see note 1 and Section 3) added to the first two parameters “a form of feature augmentation” in many directions, and we fully share that position from the information retrieval literature [3].

Subnetworks of cliques of documentary practices are themselves structured in flows of data geometrically oriented by an entity alignment. GRAPHYP thus exploits a new stage in the interpretability of SKGs, with detection of the variety of documentary routes.

The rest of this paper is organized as follows: Section 2 reviews background and other works. Section 3 analyzes GRAPHYP’s manifold subnetworks of cliques in cognitive communities and its graph matching representations. Section 4 conducts a discussion and Section 5 presents the summary and perspective of our research.

2. Background and Other Works

Methodologies for SKG design are now confronted by a simple scientific context: digital analytics of scholarly content progresses slowly and is reported as being in an early phase only (see, e.g., the Open Research Knowledge Graph wiki: <https://gitlab.com/TIBHannover/orkg/orkg-frontend/-/wikis/home> 2 August 2022), while the disconnection between research on information behaviors and research on information systems development has been recently highlighted in a comprehensive review [19]. In this context, however, research resources are available, in adjacent scientific spheres, and addressing the considerable structural diversity of digital community networks [20]. A wide range of suitable technologies are available, to express the adversarial learnings of academic dispute, with entity alignment techniques [21], with dynamic community discovery [6], and with architectures of message-passing graph subnetworks [22] (see also <http://snap.stanford.edu/class/cs224w-2020/slides/13-communities.pdf> 2 August 2022) that express the magnitude of manifolds [23]. User knowledge in informational search sessions has been modeled for the task of predicting knowledge gain and knowledge state of users in Web search sessions [24] with limitations due to the limited availability of search session data.

Headings of the surrounding environment of our research could be reviewed as follows.

2.1. Background: Information Space of Cliques in Cognitive Communities

The approach to the information space of cognitive systems applied in SKG GRAPHYP is influenced both by symmetry in information, driven by new concepts of differential geometry in graphs [25], and by ubiquity in networks, which is carried by descriptive applications of graphlets in multiplex [26] that inspired our design of cliques.

Within this general scientific framework, the background of SKG GRAPHYP with its methodology for designing “Manifold Subnetworks of Cliques in Cognitive Communities” (MSCCC), is described below in its scientific environment, while following the successive stages of its operation. We outline differences of our approach with existing citation counts methodologies.

- **Cognitive community detection**

A first step in MSCCC is to detect “cognitive communities” from the search history of logs on scientific documentation about a research question; we build digital structures, communities, that ingest the raw material of documentary resources, and we class those communities in dedicated subnetworks of nodes, according to their practices. Technologies are now available for discovery and classification of entities in a wide range of dynamic, complex, structured networks of nodes, representing communities of practice

in the SKGs [6] within dedicated subnetworks. In GRAPHYP modeling, the identity and content of communities are processed from the “raw material” of the analysis of search history queries.

Cognitive community detection from query analysis and data from search sessions can be captured from multi-source documents, through the extraction of knowledge instances such as entity, relation, attribute [16] with the SKG functioning as a semantic network. To exploit the raw material of queries, complementary logs [27] are processed for an “explicit” semantic ranking via knowledge graph embedding. For instance, a comprehensive study of log analysis of researchers to Semantic Scholar’s search engine [28] has been recently achieved. Applications are developed on Semantic Scholar Academic Graph with an easy-to-use JSON Archive (<https://api.semanticscholar.org/corpus> and <https://api.semanticscholar.org/corpus/download/> 2 August 2022). Transition from raw material to structured communities requires a machinery and steps. The first step is to find original cognitive processes in various contexts of information retrieval, in directions that highlight what could be called “networks of cognitive relevance”. This raises the question of how to express the multiplicity of cognitive practices from metrics of cognitive relevance of the differentiated uses of documentary resources. In SKG GRAPHYP, we use triplets of nodes shaped into graph subnetworks, with three parameters, mass, intensity, and attention, which measure the documentary behaviors of the users on the items (see below, 3.2 Subnetworks of cognitive communities).

Community detection also assumes a data quality assessment approach among the alternatives, and it is proposed to use a classifier as a judge [29]. New designs in that direction can be mentioned with generative adversarial networks (Vanilla GAN) and adversarial autoencoders [29].

- **Building cliques in a co-authorship or citation graph and in GRAPHYP: comparative features**

Cliques in a co-authorship or citation graph differ from cliques in GRAPHYP with respect to their modeling and their purpose; in both cases counts are submitted to drawbacks of self-citation and controversies, with differing impacts on the quality of measured results.

Networks of co-authorship [30] result from the linking of mutually connected groups of authors; cohesion indices may be introduced in a clique approach, by redefining the network density using, for instance, a variance density index. In that context, the identification of the characteristics and the impact of self-citation results from the possible correlations of the affiliation of the researchers to the same institutions and/or their collaborations—or lack of collaboration—for articles published in the journals of the research domain. Adversarial classification of citations, (negative citation, disputing citation, etc.) has also been studied by citation networks. Recently, without mention of cliques, the first large scale study of 4 million papers investigated the “expression of disagreement in scientific literature” [2], by analyzing disagreement between two papers as well as statements indicative of disagreements inside the communities; it explored a data framework of in-text citations. The impact of disagreement on self-citation, and on citation impact, was also analyzed in depth.

In our approach of cliques, GRAPHYP’s conceptual framework differs, by a systemic capture of the whole search history of users and items for meaningful keywords illustrating a research question. More fundamentally we propose a modeled information retrieval record. While modeling “adversarial documentation cliques” in a cognitive community, GRAPHYP does not provide a record of cliques of scientific opinions or findings on the same research question; we share the opinion that this could be quite an embarrassing challenge for an AI, as outlined by the authors of the above-mentioned paper on “disagreements in scientific literature” [2], who state, “When it comes to defining scientific disagreement, scholars disagree”. Far from proposing a scientific assessment of the data captured on the search history of researchers, our objective in GRAPHYP is more accessible and pragmatic. We model documentary profiles, in a classifier of research experience, leaving it open to

interpretation by all interested members of the cognitive community related to the specific research object.

Moreover, we integrate the need for a model-based information retrieval, “understanding the user”, into logging and interpreting the user’s interactions [31] as well as modeling the user’s behaviors [32].

User models for information retrieval and their algorithmic integration into the search process have long given rise to model-based information retrieval, built on top of user models [33], and more recently to efficient techniques for representational learning in information retrieval [34], with new, more structured and modeled “content de-scriptors” [35]. At a minimum GRAPHYP proposes a comprehensive profile of search experiences individualizing cliques inside communities, as a new kind of a “document descriptor”; re-using logs of documentary connection on a queried keyword captures a representation of search experiences of the others users and allows positioning explorations of a keyword. Modeling of cliques in GRAPHYP is purely informative and does not induce scientific conclusions; it allows all cliques to be analyzed, by representing adversarial documentary practices, whatever they may be. GRAPHYP thus helps the researcher to optimize the path she chooses. without favoring a so-called “best” privileged documentary path, or even any so-called “efficient” or “accurate” clique.

- **Manifold of the cognitive communities featured in subnetworks**

A second step in the methodology of MSCCC involves finding a modeling of manifolds (of the search results) in the form of structured subnetworks.

The implementation of such algorithms (known as manifold learning) in many files of information processing is currently attracting attention [36] and has recently become a popular field for its ability to propose efficient solutions in subdivisions of networks in a graph: “Think globally, fit locally”, as expressed by the associated slogan. Manifold learning currently refers to the user’s specific experience in various real-world practices (e.g., biomedicine, radiology), the information of which is embedded into alternative subnetworks, between which the user must select one, or even a network of subnetworks. It is observed that the modeling of the manifold representations of subnetworks actualizes the ideas of C.E. Shannon on symmetry in information; actualization here means describing information with systems of cognitive variety, whose representations are structured with symmetry, as analyzed below.

A new family of manifold mapping updates representations of cliques in a context where knowledge variety has been defined as a “causal pathway” to information, requiring its own descriptive geometry. The authors of the concept of causal pathway, [17], coming from the field of differential geometry, base their approach on the assumption of the endemic diversity of the fundamental concepts of information, “any information space could be viewed as in part representing a “causal pathway” embedded within some culture, but included are semantic, dynamic principles seeking to incorporate states of experience, properties often lacking in traditional cognitive theories”.

In such an approach, graphs are supposed to have their own added value in representing the manifold of cognitive information, with the application of a specific expressive symmetry in the construction of information. [17] refers to C.E. Shannon, with the concept of “Rate Distortion Manifolds”, and raises the “necessary conditions gauging the reliability of a source entropy rate relative to a channel capacity”, which are defining a research program for information representation, within which our own research question is housed with SKG GRAPHYP representation of manifolds of cliques. In their research program, the authors observe, “introducing simplicial methods to analyze the underlying combinatorial structure of the manifold, we may recover graph-theoretic models as suited to the navigation through various types of information highways, systems of coding, symbolic dynamics and complexity”. Their reference to Shannon suggests that they foresee an extension of the well-known theoretical condition of symmetry to a representation of “symmetrized manifolds”, and thus that Shannon’s theory could find an extension to a represented symmetry of (cognitive) information systems. This powerful concept is at work

today, in approaches such as the use of hypergraph manifold regularization, to maintain consistent relationships between the original data, transformed data, and soft labels [37].

- **Consistent manifold of search history of cognitive communities**

In our MSCCC methodology, representing cognitive manifolds requires suitable homogeneous data. SKG users' search history is a rich data stream to build such consistent cognitive manifolds and drive relevance in the SKGs. As observed by [36], "Automatic inference schemes based on document content and user activity can be used to estimate such constituents of relevance". The parsing of documentary logs on a research question helps to build "semantic paths" to knowledge [15], as coined also by the first comprehensive survey dedicated to cognitive graph applications [38]. User data and user-generated content are thus an endemic component for shaping the, "naturally created community information on content sharing platforms to infer potential tags and indexing terms", with the aim "to mitigate the vocabulary gap between content and query" [39].

A large corpus of knowledge is available on retrieval methodology in user interaction frameworks (see, e.g., "Proceedings of the Workshop on Understanding the User-Logging and Interpreting User Interactions in Information Search and Retrieval": https://ir.weis.de/anthology/volumes/2009.sigirconf_workshop-2009uiir/ 2 August 2022) and the breadth of manifold definition of cliques has been demonstrated in networks of manifold-valued data [23].

- **Graph matchings of subnetworks of cognitive communities**

Subnetworks of documentary practices are themselves structured in flows of data geometrically oriented by an entity alignment, which could make it possible to detect, in a controversy, a preferred community identified in a self-supervised mapping of adversarial representation of all observable communities.

A novel functionality is required to build up the semantic adversarial representation of cognitive communities of knowledge. Proposed in the SKG GRAPHYP, this functionality is necessary to identify, within a manifold of communities, typical adversarial positions and controversies. This implies firstly representing each community in a specific modeling of all "possible" manifolds, within a dedicated subnetwork. Tailored subnetworks are now accessible in graph neural networks (GNN) with a topologically aware message-passing scheme based on substructure encoding [40], which successfully overcomes the former inability of GNN to detect and count graph substructures, via subgraph isomorphism counting. This illustrates one among many benefits of geometric deep learning models [25].

- **Meta learning experience in the analysis of documentary routes**

A cornerstone concept in this context is that of clustering and embedding of neural manifolds [41] recently developed as a meaningful and interpretable feature space, adding a geometric constraint to make the clusters identifiable.

Meta-learning in SKGs could thus take the future path of "Cognitive Graphs" [39]. The analysis of sense-making in search activities develops in directions that highlight what might be called "networks of cognitive relevance" of identified cliques. This is explicitly the case with the attractive idea of "cognitive orbits" studied from man-machine interactions in a low-dimensional cognitive manifold and leading to "cognitive topological structures" [42]. A neighbor approach lies in the idea of learning interaction kernels for agent systems on Riemannian manifolds [43]; networks have something to say and might make sense in orienting alternatives in cognitive processes.

At last, "adaptive interfaces" of the same family as GRAPHYP—knowledge learning from each other—are being developed today in the field of human-machine interactions: these interfaces propose seamless interactions with users during online operations, with closed-loop adaptation of the interface, driven by the user's known movement intention [44].

2.2. Other Works on Information Space of Cognitive Communities: “Learning from Predecessors”

Pre-trained language models currently receive much attention, while pre-existing models are considered questionable [45]. We have previously mentioned the studies of the “cognitive information space” analyzing a particular cognitive situation [17]. Learning from the information retrieval of predecessors has been widely studied for a long time in the fields of Web search procedures. However, the need for global modeling in an ecosystem of self-supervised approaches [46] persists and could fruitfully benefit from adjacent works (<https://www.connectedpapers.com/main/5a00ab293237c4038b9e902adb3fce11ca9e801d/A-%22Searchable%22-Space-with-Routes-for-Querying-Scientific-Information/graph> 2 August 2022) in fields where indirect communication mediated by modifications of the environment is intensively studied. The interdisciplinary field on “learning from predecessors” has developed rich content applied to graph heuristics, within which we position our future work on hypertexts and graphs; this includes the multifaceted study of texts in philology of graphs [47], social stigmergic cognition [48], or even the stimulating idea of knowledge graphs as rhizomes in the sense of G. Deleuze (<https://www.synaptica.com/knowledge-graphs-and-their-punk-rhizomatics/> 2 August 2022), as well as new philosophical approaches to the hyperlinking of texts, structured from an integrated “Topology of mind” [49].

3. Approach. SRK GRAPHYP: Detection and Search of Cognitive Communities in Geometric Adversarial Information Routes

In Section 3, we apply this background context to the design of SKG GRAPHYP, for the detection and search of cognitive communities by modeling geometric adversarial information routes. The objective is to analyze and assess knowledge construction operations as revealed by the search history. We observe that processes of knowledge-building, captured by the search history, raise specific modelable characteristics, delineating and capturing different ways of obtaining research results (The versions of the construction of knowledge depend on the distributed relations of knowledge and their contexts of use; however, for the same context, there may be various hypotheses on the interpretations of the facts and therefore on the possible outcomes. The experimentation itself can follow varied protocols according to varying hypotheses, and thus the same initial knowledge leads either to alternative ways to reach alternative outcomes, or to the same outcomes reached by alternative ways. Comparing ways to outcomes is one source of research assessment and one strategic way to exploit search history in GRAPHYP.) among which science has to choose and to optimize its access routes to results. In this field, research impact assessment is challenging in that it must establish a relevant demarcation between the directions traced by scientific outcomes and their routes, and identify the associated semantic changes in the scientific vocabulary, as well as the related practices in data stewardship. These tasks are included in the purpose of the SKG GRAPHYP.

3.1. Design of the SKG GRAPHYP: Building the Information Space of a Cognitive System

(i) General modeling and problem set-up

Purpose: The purpose of GRAPHYP is to represent the plurality of research answers that can be categorized from the captured diversity of search histories on a scientific subject, thus representing the adversarial versions modeled in cliques, of the answers given to a research question. Examples of plurality of scientific approach that one may want to trace through alternative documentary practices abound in all areas that develop controversies: genetic contacts between Sapiens and Neanderthal; vaccine alternatives in the face of a pandemic; responses to climate change; origins of dark matter, etc.

Method: We followed recent approaches to scientific and technological information-oriented “semantics-adversarial” and “media-adversarial” using cross-media retrieval methods, to find “effective subspaces” of cliques [50].

Originality: Our approach is close to the recently featured category of systems of “model-based information retrieval”, which articulate indexing, retrieval, and ranking,

built from a single corpus; GRAPHYP is close to that multipurpose group of multi-task learners for multiple information retrieval tasks [45].

Output and uses: the purpose of GRAPHYP is to represent MSCCC corresponding to the type of SKG described in Section 2 as generative adversarial networks (GAN; [51]). It models antagonistic subgraph features of cliques and articulates underlying topological distribution of graph structures. This modeling is proposed at different scales and levels of granularity in a “generative” architecture, adjustable to needs, with an improved training ability [52]. It can thus be used to discover new graph structures and to generate evolving graphs.

(ii) Design of SKG GRAPHYP

The graph design of SKG GRAPHYP is a crown bipartite graph with connected edges (<https://mathworld.wolfram.com/UtilityGraph.html> 2 August 2022) featured with distance-transitivity (<https://mathworld.wolfram.com/Distance-TransitiveGraph.html> 2 August 2022). A first sketch of this modeling has already been described in [46].

SKG GRAPHYP achieves its purpose by positioning search communities in a “searchable space” where all users gathered in cliques of searching communities share the same keywords. The geometry of GRAPHYP allows two functions:

- it allows each clique in its community to be positioned in the searchable space, according to the characteristics of its search history;
- it assists a clique inside a community in navigating on the graph, to reach the position of neighboring cliques in the same community, linked by the same characteristics of search goals («Search goals» as a generic term, encompasses similar queries, keywords, or groups of URLs).

3.2. Subnetworks of Cognitive Communities: Detection and Integration in the SKG GRAPHYP

The basic subnetwork unit in the SKG GRAPHYP must be relevant, simple, and duplicable, to adapt to the constraints of MSCCC modeling. As shown in Section 2, GRAPHYP’s subnetworks are in line with equivariant subgraph aggregation networks recently described and perform well on multiple graph classification benchmarks [22].

We will now examine (i) cognitive community definition and detection and then (ii) entity alignment.

(i) Cognitive community and its cliques: Definition and detection

- Positioning of cognitive communities in the SKG geometry of a bipartite crown hypergraph

User preferences are expressed by choices that the search history records in a modeled matrix of possible choices. This modeled matrix is designed by a mini-max system of topological locations of cognitive communities, between two extreme preferences within the framework of GRAPHYP. Between these extrema, recording typical search sessions helps to model the full range of information-seeking needs of user communities for similarly formulated search sessions. GRAPHYP models all the possible choices made from the same question. Our graph construction borrows the unfrequented methodology of a “map equation” formulating a data flow network [53]; first described in [54], it infers that programmed, “links in a network induce movement across the network and result in system-wide interdependence”. Any clique in a cognitive community will be defined hereunder by co-integration in the SKG of triplets of values of the parameters recording the search history, with the metrics of mass, intensity, and attention.

- Basic identification of the retrieval profile of a cognitive community on its search route: mass and intensity of nodes in search history

Let us develop in operational terms the methodology initially sketched in [46] which has been transformed here in terms of calculation bases. In order to record the “search

routes" of any clique in a community for a given keyword, routes that may differ and are intended to be compared, we write:

$$Q_n = f(N_n; K_n) \quad (1)$$

where Q_n is the number of searches related to a given topic. Altogether, Q expresses the quantitative weight of any community as measured by the documentary usages of this community for its cognitive purpose of answering a research question. In addition, Q could compile different queries provided they belong to the same perimeter of research question.

Let us also consider that for each search Q , we identify a parameter of mass which corresponds to a number N of users, and a parameter of intensity K which corresponds to the number of items (URL, documents, articles) constituting the search outcomes, among a corpus of items related to the keyword or the group of related keywords. We can consider the N users of K items as a community of users of the same query route Q (alternatively, we could represent Q search session results by another expression of preferences, i.e., not a user/item approach, but an item/item approach, where N items and K items are mixable in communities of preferences where we consider that this mix of publications characterizes comparable sets of search sessions). The positioning of any distinct route can be expressed for a given search within the limits of a system of typical search sessions.

- Recording dynamics of search sessions: a third node measuring the value of a parameter of attention

Here we propose a new method for recording the dynamics of GRAPHYP, which clarifies and differs from [46]. Let us calculate the mean values of N and K on the whole set of search sessions; we can normalize the presentation of all search sessions, as located above or below the mean ratio N/K . Additional information on that ratio is given by its dynamics at the scale of the whole set of analyzed search sessions, as well as by its value in any triplet. In fact, with any recorded value of the N/K ratio, the ratio of attention is an associated index of dynamics in documentation behavior, which expresses that N/K preferences could be conversely recorded either from an abruptly changing behavior or from a steadily increasing or decreasing behavior in reading articles (in our example). "Attention" is thus a behavioral component of the observed retrieval experience that measures the "permanence" or conversely the "rupture" in search experience rhythm; this third parameter can be "stable" or "erratic". When combined with intensity (which can measure few or many documents) and mass (which can represent a large or a small number of users), attention thus integrates a useful additional parameter of stability/disruption in search practice.

By adding a parameter of attention, we change our function Q of two variables N , K , into a triplet where the third term linking N and K represents this expression of stability/disruption of behaviors, and allow us to address the multiple factors bound to attention (mechanisms, types of co-attention, intra-attention, and all documentary behaviors to measure a "real-valued hint" in information retrieval [3]). For instance, we could practice community detection of the readers of a usual group of chemistry articles "before" and "after" the publication of a new important article and we would thus notice if this additional publication "accelerated"—or not—the readings in peripheral related domains.

Let us measure the value of that third term by a ratio calculated from a value of normalization, expressed from the mean value of N/K . We can consider a fraction α/β where α is the numerator calculated from N mean value and β is the denominator derived from the mean value of K . This fraction α/β will vary, consequently, with any recorded group of reader and article values. (An alternative procedure could be to note α the coefficient of increase of N and β the coefficient of increase of K when Q varies by one unit when an additional query on the same search is recorded.)

The value of this fraction brings a specific element of dynamic analysis. It expresses the degree of attention emerging in cliques of the cognitive community's documentary practice and makes it possible to measure the "stability" or "disruption" of the behaviors of users on items of a search session, when the quantities N and K increase or decrease between

searches when N and K values are computed on the whole set of a group of searches for the purpose of detecting cliques in the community of searches on the same keyword. Therefore, attention contributes to measure changes in the dynamic of search of documents met on routes of search.

The fraction α/β provides thus a dynamic index of the variations recorded in the practices and controversies of various cliques in scientific communities, revealing quantitatively how “strong” or “weak” they could be. This approach towards sensitiveness to frequency may also indirectly help detect differences between cliques in communities approaching the same concept by homonyms. (For the same new category of items, several communities could be «neutral» to a change of publishing orientation, while others could be reactive.)

(ii) **Entity alignment of cognitive communities in GRAPHYP modeling**

- Networking search sessions and detection of cliques in cognitive communities in the SKG

We know that, with the added mix of users and items that it measures, the search experience profile tends to be “stable” when (α/β) , as this fraction is approaching its mean value on the whole set of recorded corresponding search sessions, and “unstable” in any of the other cases, when the recorded value deviates from the mean value. The method is here related to the graph assortativity approaches described in Wolfram Assortativity (<https://reference.wolfram.com/language/ref/GraphAssortativity.html> 2 August 2022).

With the three above described parameters, two triplets could be shaped to represent a formal graph-based representation of paths between two limits fixed to the expression of preferences of the users. With the design of Figure 1 hereunder, we could position between these two limits the six non-contradictory solutions that combine values of parameters that could be connected between those two limits. It shapes the following bipartite crown graph with connected nodes, representing six typical networks of nodes; those subnetworks in GRAPHYP represent the modelizable cliques of any cognitive community in our conceptual framework.

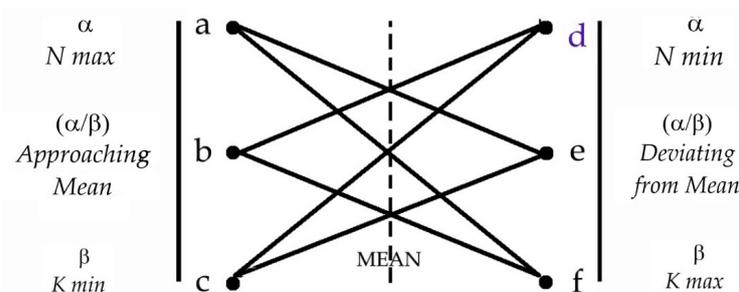


Figure 1. Mapping search experience in GRAPHYP: a–f, as the six typical modelizable cliques combining triplets of nodes, included in a cognitive community [46].

Each letter here materializes the head of a network of three nodes: a, b, c, d, e, f, are all characterized by the combination of two other nodes (aef, bdf, etc.). Figure 1 shows a complete representation of which of the six typical positions of cliques contained in possible search experiences could be modeled between the two triplets of nodes designed in SKG GRAPHYP. It provides a tool for classification of observed search sessions Q in a series of searches on a given keyword, according to the user and item choices.

Going into the details of the mathematical design of GRAPHYP would be too technical for this article. We refer the reader to the Supplementary Materials (see below of the article) and to the article (Fabre, 2019) [46] for a first draft of GRAPHYP development. The weighting of graph subnetwork matches is discussed in the Supplementary Materials with the help of an example developed in Section 3.3 hereunder.

In Section 3.3, we test the parameters of GRAPHYP from data on readings in the context of scientific impact assessment.

3.3. Tests of GRAPHYP's Triplet Adjustment: Mapping of Retrieval Profile

We successfully tested the robustness and significance of node triplets integrated in the image of community behaviors. The data came from real-world search history records, in a test panel of our triplets for approximately 10 million search sessions; we are developing in the meantime the analysis of click metrics of scholarly content—a topic currently in full development—in order to map pictures of retrieval profiles [55].

In order to identify the nature of the data required to implement GRAPHYP in the context of a digital library, we made a prototype using access log files from OpenEdition.org platforms that contain more than 200,000 papers from 592 journals in Humanities and Social Sciences. These log files were collected by the Web analytics platform Matomo (<https://matomo.org/about/> 2 August 2022) and then filtered to eliminate bots and requests for documents other than scientific papers. User IDs were associated to the entry logs according to the anonymized IP address and sessions IDs were estimated based on the recorded timestamps. This means that in the following we call *session* a sequence of clicks (content requests) performed by an anonymous user in a limited time.

We can assume that each session so defined corresponds to a specific need, even if we do not know the original queries made by the users (the Web referrers have not been communicating queries for a few years now and the incoming links are usually too vague to be useful). In most cases, the readers come directly from a Web search engine and then directly follow the links without using the internal search engine of the platform. Since the queries are unknown, our aim is to compare general reader behaviors, regardless of their precise information needs.

The process is as follows. First, we subdivide all the logs into blocks of sessions according to a chronological criterion (again, if we had the queries, we could build the blocks according to thematic criteria). Then, for each user session in each block, we estimate the values of K as the number of articles read (Figure 2). This makes it possible to estimate, in each block of sessions, the number of readers N who have read K articles and thus the mean values of the different N and K for this block (Figure 3). Depending on the construction method of the blocks, these values correspond to the different formulations of the queries or to their variety in a time range as well as to the readers' profiles.

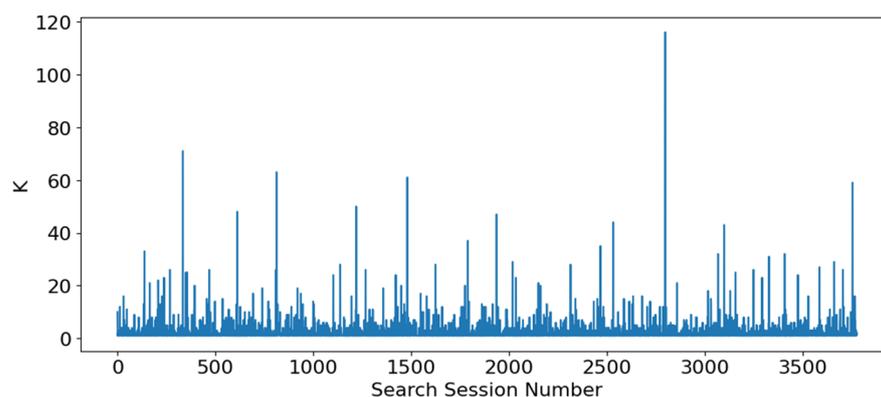


Figure 2. K values for 4000 sessions: during some sessions, many articles are read (K) but in most cases, only a few articles are retrieved. On average, 2.7 articles are downloaded, 75% of the sessions are for three papers or less, some sessions exceed 100 items.

Once this is done, it remains to determine the GRAPHYP type for each block of sessions according to the values of N , K , and α/β . In a practical way, some thresholds help to identify the tendencies towards the min/max values of N and K and the deviation from stability of the values of the ratio α/β above or under the mean value of the sample. For example, we can choose to assign type f if the means of N and K for the considered block are very high (much greater than the means of N and K of all blocks) and α/β close to

mean value. Gathering the sequence of types highlights the typical routes of search for the given logs.

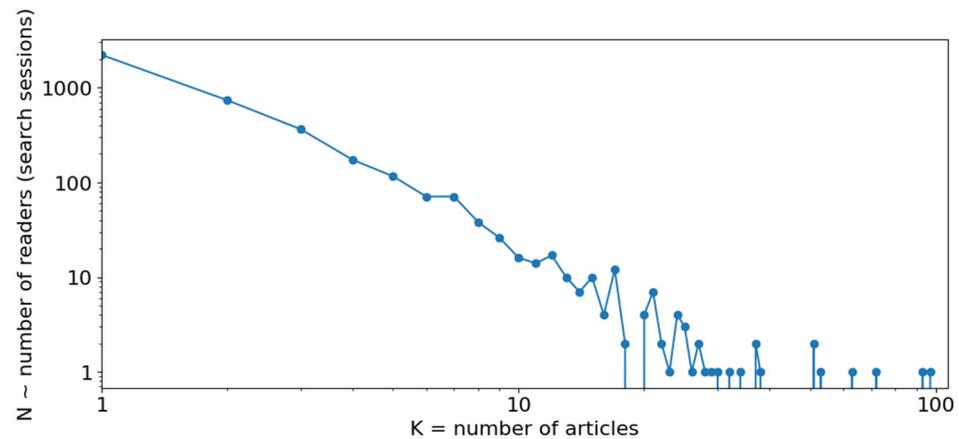


Figure 3. *N* and *K* values for one block of user sessions: quite a few people read more than ten articles while most readers read only one or two articles.

By considering 100 blocks of entry logs ordered chronologically and corresponding to 371,906 search user sessions and 1 million documents read, 42% of the sessions are of type *b* and 34% of type *c*. Even if over a period of time we can estimate that the current events and news in the world influence the requests in the same direction (hot topics) and because we do not know the actual user queries, we cannot really cluster them thematically to create more coherent blocks. As a consequence, we combine here in the same block very different user needs and the result was expected: the majority behavior is stable over time. There are always readers who are looking for specific articles and only read those, and other readers who go deeper into a current topic by reading many articles. Enlarging the time-window of the logs would allow detection of more evolutions of behaviors.

3.4. Subnetwork Analysis and Comparisons of Cliques in Cognitive Community

(i) **Comparing cliques inside a cognitive community on adversarial search route**

The following figure is an illustration of a clique in a cognitive community.

Six typical search routes for cliques in a cognitive community are represented in Figure 4 in order to illustrate how GRAPHYP allows exploration of cliques according to their typical neighbor search sessions.

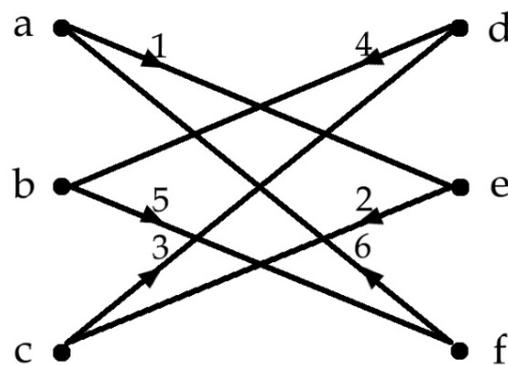


Figure 4. Exploring neighboring cliques in a cognitive community.

GRAPHYP shows an undirected weighted graph structure (each edge is bi-directional and receives a distinct weight from the nodes linked inside a network), which is connected (one can reach any node from all other nodes inside identified paths) and builds a minimum

spanning tree as a subgraph containing all nodes, connected here with the minimum possible number of edges.

(ii) Representing cliques in cognitive communities at various scales

Cooperative or connecting routes can be identified by the GRAPHYP data structure, which involves adjusting the size of the selected community to an optimized scale. This in turn raises questions of the optimal calibration of subgraphs in the perspective recently called by Michael Bronstein, that of “latent graph learning” and the associated methodology for counting isomorphisms of subgraphs [40].

As illustrated by Figure 5 below, GRAPHYP with nodes A, B, C, etc., is built by the addition of graphs of the same shape, and open thus to a self-similar substitution (<https://tilings.math.uni-bielefeld.de/glossary/selfsimilar-substitution/> 2 August 2022); this self-similarity characteristic of GRAPHYP [46] allows knowledge extraction at any scale and allows operating scalability from perimeters of information processed from operations of addition, subtraction, multiplication and division, which, to the best of our knowledge, has not yet been outlined in graph architectures.

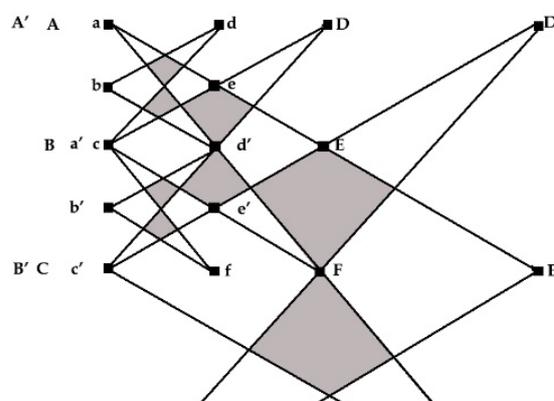


Figure 5. Scalable scales of cliques and cognitive communities.

4. Discussion

With GRAPHYP, we have designed an automated representation of the content consulted, exploiting what is known as “predecessor information”. It required a non-trivial modeling of adversarial subnetworks of cliques in cognitive communities through explainable paths of reasoning, to let the user choose, in a self-supervised attitude, a version of the knowledge best adjusted to his/her own hypotheses. As the analysis of cognitive structures is entering a new phase, driven by geometric graphs that are learning cognitive manifolds, the GRAPHYP toolkit appears as an illustration at the intersection of those complementary evolutions: *it is a graph designed not to represent information, but to model information representation*. However, as noted in Section 1, SKGs still remain insufficiently effective in scholarly communication, and the main contribution of our study is to show a conceptual framework promising directions of development for a more complete representation of scientific documentary choices, facilitating the exploration of all possible routes of knowledge discovery and allowing scientists as well as their readers to trace more systematically the documentary itineraries of scientific choices.

The first tests carried out and presented in Section 3.3 used query logs whose search terms themselves were not known. More complete tests exploiting a detailed semantic analysis of query logs are in preparation.

5. Conclusions and Further Works

Two main conclusions emerge from this paper.

The first is that it is important that the design of the SKG takes diversity of cognitive information into account. We have introduced here a new SKG that is consistent with the adversarial nature of the cognitive process in scholarly communication.

The second is that scientists shall derive unexplored benefits from representations of diversity of cliques in any field of knowledge and in cognitive communities. We develop an approach to “alternatives” in the results of search activities, as a new ancillary support to the discovery strategy, fully exploiting the potential of digital information architecture and its ability to address the design of diversity through exploitation of search history.

As other more general remarks, we propose three observations from the GRAPHYP experience.

The SKG is a structured interaction. Knowledge is not embedded in the graph, and the graph itself does not create knowledge. The KG is only a vehicle between two physical realities that it seeks to connect. In this sense, the KG is a mixed composite facility, a place of interaction, open to scientific interpretation of an architecture of documentary features represented on a geometric knowledge graph.

The diversity of information is the fuel of rich interactions. “Heterosis”, the sharing of diversity, remains a fruitful means of enrichment; GRAPHYP’s final outputs thus combine the strengths of mutual information and of multiverse services.

The contextualized diversity in SKGs allows users and items to come closer.

For future work, we plan to compare our approach with existing solutions from other conceptual approaches in the field. Further works will focus on the study of the modeling of SKGs with alternated relative symmetries in information representations in graphlets architectures of cliques, applied to the analysis of hypertext links of scientific results on the Web.

Supplementary Materials: A notebook describing the test procedure (Section 3.3), is available as a supplementary file on GitHub: <https://github.com/pbellot/graphyp>. We have published two preprints for the SGK community that accompany this paper and contain similar results. The two preprints can be found under these links: <https://hal.archives-ouvertes.fr/hal-03365118> (accessed on 5 October 2021); <https://doi.org/10.48550/arXiv.2205.01331> (accessed on 3 May 2022).

Author Contributions: Conceptualization, R.F.; Investigation, R.F.; Methodology, R.F.; Supervision, R.F. and O.A.; Writing—original draft, R.F. and O.A.; Writing—review & editing, O.A., P.B., J.S. and D.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable, the study does not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Manghi, P.; Mannocci, M.; Osborne, F.; Sacharidis, D.; Salatino, A.; Vergoulis, T. New trends in scientific knowledge graphs and research impact assessment. *Quant. Sci. Stud.* **2021**, *2*, 1296–1300. [CrossRef]
- Lamers, W.S.; Boyack, K.; Larivière, V.; Sugimoto, C.R.; van Eck, N.J.; Waltman, L.; Murray, D. Meta-Research: Investigating disagreement in the scientific literature. *eLife* **2021**, *10*, e72737. [CrossRef]
- Tay, Y.; Luu, A.; Hui, S.C. Multi-Cast Attention Networks for Retrieval-based Question Answering and Response Prediction. *arXiv* **2018**, arXiv:1806.00778v1. [CrossRef]
- Nedioui, M.A.; Moussaoui, A.; Saoud, B.; Babahenini, M.C. Detecting communities in social networks based on cliques. *Phys. A Stat. Mech. Appl.* **2020**, *551*, 124100. [CrossRef]
- Fried, Y.; Kessler, D.; Shnerb, N. Communities as cliques. *Sci. Rep.* **2016**, *6*, 35648. [CrossRef] [PubMed]
- Rossetti, G.; Cazabet, R. Community Discovery in Dynamic Networks: A Survey. *ACM Comput. Surv.* **2019**, *51*, 1–37. [CrossRef]
- Jaradeh, M.Y.; Singh, K.; Stocker, M.; Auer, S. Triple Classification for Scholarly Knowledge Graph Completion. In Proceedings of the 11th on Knowledge Capture Conference, Virtual Event, 2–3 December 2021; pp. 225–232. [CrossRef]
- Hund, A.; Wagner, H.-T.; Beimborn, D.; Weitzel, T. Digital innovation: Review and novel perspective. *J. Strateg. Inf. Syst.* **2021**, *30*, 101695. [CrossRef]
- Zitt, M.; Lelu, A.; Cadot, M.; Cabanac, G. Bibliometric Delineation of Scientific Fields. In *Springer Handbook of Science and Technology Indicators*; Glänzel, W., Moed, H.F., Schmoch, U., Thelwall, M., Eds.; Springer: Cham, Switzerland, 2019; pp. 25–68. [CrossRef]
- Jaradeh, M.Y.; Oelen, A.; Farfar, K.E.; Prinz, M.; D’Souza, J.; Kismihók, G.; Stocker, M.; Auer, S. Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. In Proceedings of the 10th International Conference on Knowledge Capture, Marina Del Rey, CA, USA, 19–21 November 2019; pp. 243–246. [CrossRef]

11. Lin, Y.; Evans, J.A.; Wu, L. New directions in science emerge from disconnection and discord. *J. Informetr.* **2022**, *16*, 101234. [[CrossRef](#)]
12. Milojević, S. Quantifying the cognitive extent of science. *J. Informetr.* **2015**, *9*, 962–973. [[CrossRef](#)]
13. Hogan, A.; Blomqvist, E.; Cochez, M.; d’Amato, C.; de Melo, G.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S.; et al. Knowledge Graphs. *Synth. Lect. Data Semant. Knowl.* **2021**, *22*, 1–237. [[CrossRef](#)]
14. Nayyeri, M.; Müge Çil, G.; Vahdati, S.; Osborne, F.; Rahman, M.; Angioni, S.; Salatino, A.; Recupero, D.R.; Vassilyeva, N.; Motta, E.; et al. Trans4E: Link Prediction on Scholarly Knowledge Graphs. *arXiv* **2021**, arXiv:2107.03297v1. [[CrossRef](#)]
15. Dstandau, M.; Fekete, J.-D. The missing path: Analysing incompleteness in knowledge graphs. *Inf. Vis.* **2021**, *20*, 66–82. [[CrossRef](#)]
16. Chen, Z.; Wang, Y.; Zhao, B.; Cheng, J.; Zhao, X.; Duan, Z. Knowledge Graph Completion: A Review. *IEEE Access* **2020**, *8*, 192435–192456. [[CrossRef](#)]
17. Glazebrook, J.F.; Wallace, R. Rate Distortion Manifolds as Model Spaces for Cognitive Information. *Informatica* **2009**, *33*, 309–345.
18. Han, J.; Rong, Y.; Xu, T.; Huang, W. Geometrically Equivariant Graph Neural Networks: A Survey. *arXiv* **2022**, arXiv:2202.07230. [[CrossRef](#)]
19. Huvila, I.; Enwald, H.; Eriksson-Backa, K.; Liu, Y.-H.; Hirvonen, N. Information behavior and practices research informing information systems design. *J. Assoc. Inf. Sci. Technol.* **2021**, *73*, 1043–1057. [[CrossRef](#)]
20. Easley, D.; Kleinberg, J. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*; Cambridge University Press: Cambridge, UK, 2010. [[CrossRef](#)]
21. Munne, R.F.; Ichise, R. Entity alignment via summary and attribute embeddings. *Log. J. IGPL* **2022**, jzac021. [[CrossRef](#)]
22. Bevilacqua, B.; Frasca, F.; Lim, D.; Srinivasan, B.; Cai, C.; Balamurugan, G.; Bronstein, M.M.; Maron, H. Equivariant Subgraph Aggregation Networks. *arXiv* **2022**, arXiv:2110.02910. [[CrossRef](#)]
23. Chakraborty, R.; Bouza, J.; Manton, J.H.; Vemuri, B.C. ManifoldNet: A Deep Neural Network for Manifold-Valued Data with Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 799–810. [[CrossRef](#)]
24. Yu, R.; Tang, R.; Rokicki, M.; Gadiraju, U.; Dietze, S. Topic-independent modeling of user knowledge in informational search sessions. *Inf. Retr. J.* **2021**, *24*, 240–268. [[CrossRef](#)]
25. Bronstein, M.M.; Bruna, J.; Cohen, T.; Veličković, P. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv* **2021**, arXiv:2104.13478. [[CrossRef](#)]
26. Dimitrova, T.; Petrovski, K.; Kocarev, L. Graphlets in Multiplex Networks. *Sci. Rep.* **2020**, *10*, 1928. [[CrossRef](#)] [[PubMed](#)]
27. Agosti, M.; Crivellari, F.; Di Nunzio, G.M. Evaluation of Digital Library Services Using Complementary Logs. In Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Boston, MA, USA, 19–23 July 2009.
28. Xiong, C.; Power, R.; Callan, J. Explicit Semantic Ranking for Academic Search via Knowledge Graph Embedding. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 1271–1279. [[CrossRef](#)]
29. Ghoggh, B.; Ghodsi, A.; Karray, F.; Crowley, M. Generative Adversarial Networks and Adversarial Autoencoders: Tutorial and Survey. *arXiv* **2021**, arXiv:2111.13282. [[CrossRef](#)]
30. Grilo Rosa, M.; Sousa Fadigas, I.; Tamanini Andrade, M.; Barros Pereira, H. Clique Approach for Networks: Applications for Coauthorship Networks. *Soc. Netw.* **2014**, *3*, 80–85. [[CrossRef](#)]
31. Buscher, G.; Gwizdka, J.; Teevan, J.; Belkin, N.J.; Bierig, R.; van Elst, L.; Jose, J. SIGIR 2009 workshop on understanding the user: Logging and interpreting user interactions in information search and retrieval. *ACM SIGIR Forum* **2009**, *43*, 57–62. [[CrossRef](#)]
32. Clarke, C.L.A.; Freund, L.; Smucker, M.D.; Yilmaz, E. SIGIR 2013 workshop on modeling user behavior for information retrieval evaluation. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 28 July–1 August 2013; p. 1134. [[CrossRef](#)]
33. Heidari, M.; Zad, S.; Berlin, B.; Rafatirad, S. Ontology Creation Model based on Attention Mechanism for a Specific Business. In Proceedings of the 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 21–24 April 2021; pp. 1–5. [[CrossRef](#)]
34. Lin, J.; Ma, X. A Few Brief Notes on DeepImpact, COIL, and a Conceptual Framework for Information Retrieval Techniques. *arXiv* **2021**, arXiv:2106.14807. [[CrossRef](#)]
35. Lin, J. A proposed conceptual framework for a representational approach to information retrieval. *ACM SIGIR Forum* **2022**, *55*, 1–29. [[CrossRef](#)]
36. Roweis, S.T.; Saul, L.K. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)] [[PubMed](#)]
37. Shao, S.; Xu, R.; Wang, Z.; Liu, W.; Wang, Y.; Liu, B. DLDL: Dynamic label dictionary learning via hypergraph regularization. *Neurocomputing* **2022**, *475*, 80–88. [[CrossRef](#)]
38. Chen, M.; Tian, Y.; Wang, Z.; Xu, H.; Jiang, B. A Comprehensive Survey of Cognitive Graphs: Techniques, Applications, Challenges. *Preprints* **2021**, 2021080155. [[CrossRef](#)]
39. Eickhoff, C. Contextual Multidimensional Relevance Models. Ph.D. Thesis, TU Delft, Delft, The Netherlands, 14 October 2014. [[CrossRef](#)]
40. Bouritsas, G.; Frasca, F.; Zafeiriou, S.; Bronstein, M.M. Improving Graph Neural Network Expressivity via Subgraph Isomorphism Counting. *arXiv* **2006**, arXiv:2006.09252. [[CrossRef](#)] [[PubMed](#)]

41. Tripuraneni, N.; Jin, C.; Jordan, M.I. Provable Meta-Learning of Linear Representations. In Proceedings of the 38th International Conference on Machine Learning, PMLR, Virtual Event, 18–24 July 2022; Volume 139, pp. 10434–10443.
42. Cheng, R.; Liu, C.; Meng, S. A Study of Cognitive Orbits Based on Man-machine Interactions. *Open Cybern. Syst. J.* **2015**, *9*, 2694–2702.
43. Maggioni, M.; Miller, J.J.; Qiu, H.; Zhong, M. Learning Interaction Kernels for Agent Systems on Riemannian Manifolds. *arXiv* **2021**, arXiv:2102.00327v3. [[CrossRef](#)]
44. Rizzoglio, F.; Casadio, M.; De Santis, D.; Mussa-Ivaldi, F.A. Building an adaptive interface via unsupervised tracking of latent manifolds. *Neural Netw.* **2021**, *137*, 174–187. [[CrossRef](#)]
45. Metzler, D.; Tay, Y.; Bahri, B.; Najork, M. Rethinking search: Making domain experts out of dilettantes. *ACM SIGIR Forum* **2021**, *55*, 1–27. [[CrossRef](#)]
46. Fabre, R. A searchable space with routes for querying scientific information. In Proceedings of the 8th International Workshop on Bibliometric-Enhanced Information Retrieval (BIR 2019), Cologne, Germany, 14 April 2019; pp. 112–124. Available online: <http://ceur-ws.org/Vol-2345/paper10.pdf> (accessed on 2 August 2022).
47. Weber, T. A Philological Perspective on Meta-scientific Knowledge Graphs. In *ADBIS, TPD and EDA 2020 Common Workshops and Doctoral Consortium*; Springer: Cham, Switzerland, 2022. [[CrossRef](#)]
48. Marsh, L.; Onof, C. Stigmergic epistemology, stigmergic cognition. *Cogn. Syst. Res.* **2008**, *9*, 136–149. [[CrossRef](#)]
49. Logan, R.K.; Pruska-Oldenhof, I. *A Topology of Mind: Spiral Thought Patterns, the Hyperlinking of Text, Ideas and More*; Springer: Cham, Switzerland, 2022; 244p. Available online: <https://link.springer.com/book/9783030964351> (accessed on 2 August 2022).
50. Li, A.; Du, J.; Kou, F.; Xue, Z.; Xu, X.; Xu, M.; Jiang, Y. Scientific and Technological Information Oriented Semantics-adversarial and Media-adversarial Cross-media Retrieval. *arXiv* **2022**, arXiv:2203.08615. [[CrossRef](#)]
51. Liu, W.; Chen, P.; Yu, F.; Suzumar, T.; Hu, G. Learning Graph Topological Features via GAN. *IEEE Access* **2019**, *2019*, 2898693. [[CrossRef](#)]
52. Liu, L.; Zhang, Y.; Deng, J.; Soatto, S. Dynamically Grown Generative Adversarial Networks. In Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), Virtual Event, 2–9 February 2021; pp. 8680–8687. Available online: <https://ojs.aaai.org/index.php/AAAI/article/download/17052/16859> (accessed on 2 August 2022).
53. Rosvall, M.; Bergstrom, C.T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1118–1123. [[CrossRef](#)] [[PubMed](#)]
54. Rosvall, M.; Axelsson, D.; Bergstrom, C.T. The map equation. *Eur. Phys. J. Spec. Top.* **2010**, *178*, 13–23. [[CrossRef](#)]
55. Fang, Z.; Costas, R.; Tian, W.; Wang, X.; Wouters, P. How is science clicked on Twitter? Click metrics for Bitly short links to scientific publications. *J. Assoc. Inf. Sci. Technol.* **2021**, *72*, 918–932. [[CrossRef](#)]