EasyChair Preprint

№ 769

A "Searchable" Space with Routes, for Querying Scientific Information

Renaud Fabre

February 5, 2019

A "Searchable" Space with Routes, for Querying Scientific Information

Renaud Fabre

Laboratoire Paragraphe (EA349), Université Paris 8, Saint-Denis, France
renaud.fabre01@gmail.com

Abstract. Accuracy of scientific and technological information (STI) retrieval is on probation while its exploration expands: continuous additions of contents, comments and data structures create rich but uneasy conditions for selections of items by users of the same keywords. Moreover, users of information are confronted to a "hidden face" of searchable space: their own choices that could be recorded from their selections on lists of proposed answers to a query and could help mapping their navigation on recorded "routes" of search, are not open to consultation, and remain either unveiled or even unavailable. At best, users benefit from useful recommendations but, as data on their own preferences and choices is not shared, networked information for global navigation remains fuzzy. In this position paper, our standpoint is that search of scientific information generates data on various needs and users, and search would then take benefit from an efficient re-use of users generated contents, in addition to any current metrics. First, with examples including our own data, we suggest that search practices in STI domains could be designed as an addition of sub-communities of users, sharing interest for their own selections of items. Second, we propose a data structure modeling interaction between those sub-communities and networking routes between them, in order to give a common room to "searchability" of a keyword domain. We thus propose a design of alternative paths to answers, which could better fit search architecture with observed needs for information retrieval in scientific research.

Keywords: knowledge management · recommender systems · query and keyword modeling · complex systems applications · information retrieval

1 Introduction

The research agenda [19] of National Academies of Sciences of the United States observed recently that digital scientific and technical information (STI) is a "complex system" that has to improve its architecture and its accuracy. In this context, that position paper raises questions about optimization of "searchability" of STI, which refers to efficient access to relevant information by users.

"Searchable" is "capable of being computationally searched", within an independent community of search, open to serendipity [8], navigating in an unknown space: “Web search is governed by a unified hidden space, and each involved element such as query and URL has its inborn position” [4]. STI community of search ask for accuracy of delivered items, as highlighted by reviews of “observational” approaches to information retrieval [12], and performs a rich semantic search [1]. In this community, “user’s initial search aims and intentions evolve as new information is encountered” [20].

Main issue is of this position paper is to observe that users of STI are building together a community of information retrieval, and that this community remains confronted to a “hidden face” of its common searchable space; available output of search never proposes user’s generated contents mapping communities of users of selected contents issued from the same keyword. Those data, which are not open to consultation, remain either unveiled or even unavailable. Lack of those data affects specifically search in STI: in uses for scientific research, a simple keyword could give access to a very wide range of structured discussions and interpretations, each delivered with their own specific “version” and “vocabulary”. Without the help of data identifying specific sub-communities, search could appear as a lonely walk in a forest of fuzzy homonyms.

However, current research hardly study that situation, as only “few approaches take advantage from searches performed previously by users” [13]. Fortunately, recent approaches on interactive information retrieval and user’s behavior highlights that search in STI covers a large variety of distinct ways and needs [15].

From that new strategic standpoint, in a nutshell, this article tackles the following research question: Why and How could STI search users could benefit from each other’s search sessions?

Our main contribution is, first, to characterize STI search communities, with examples including our own data, illustrating variety of search needs and behaviors characteristics, existing under cover of a common queryword. Second, with help of a bi-partite graph developed for information retrieval, we build a modeling of groups of users performing search sessions, and formalize their links.

Result is a data structure mapping all recordable categories of sub-communities of search (users-items groups), that, altogether, shape the global search process surrounding a keyword. A “compass” helps orienting navigation, and create conditions for complete “searchability” of a queryword, with modeled alternative paths to answers. In that way, we propose a search architecture devoted to fit with observed needs of information retrieval in STI contexts.

2 Part 1. Communities of Users in STI Search

STI search strategies depends on community behaviors of users that could be identified from structured “relations between the topics and the use of documents” [5], and could develop various alternative strategies of search and uses [9].

To characterize communities of users and items in STI searches, we identify two levels of community building: one level is the one of behaviors that could be shared by all users in any STI search session; second level is STI uses which are specific to distinct scientific communities. Both levels seem to interact in community building of STI search.

2.1 STI Uses of Specific Communities: Differences Between Practices

Differences in STI practices is sourcing differences of search attitudes. We take an example from data on STI practices extracted from a survey of research data management achieved at the CNRS in 2014 [18]. Results are part of a nationwide survey on scientific information and documentation including 432 directors of French public research laboratories.

With an high reponse rate (30%) this survey was based on 91 questions send to 1250 directors of laboratories. Data reveals significant differences in STI management in any fields of production and uses of scientific information. To give an example, adoption of Research Data Management (RDM), " we can distinguish three groups: (1) laboratories from nuclear and particle physics and from social sciences and humanities appear globally more advanced regarding RDM than other disciplines; (2) laboratories from the three domains ecology and environment, informatics and earth sciences and astronomy have dedicated resources and make their data available; (3) laboratories in the field of physics appear aware of the challenge". Regarding to OA, results reveal also significant differences of management of STI practices: those results have all direct impacts on the ways and purposes of search activities and on searched items.

The same kind of result is supplied by an other national survey (COPIST): here the CNRS examined STI practices and management at national scale for all research organisations. Results include answers of 106 research and higher education institutions and shows, altogether, a strong will to share their pracices and large differences between current organisation. Detailed results on search tools and bibliometrics are available[1] and shows significant differences expressed by users on their uses and their will to share their uses and their documents[2].

2.2 Behaviors of STI Search: Variety of Needs for Search and Search Results

As a whole, as researchers are interested in new coming ideas and emerging results, a "cold start" biased answer to a query, is a permanent threat: a new article's content has, by definition, not yet be produced elsewhere. Search is innovation-oriented, as it tries to find out items which could be identified altogether, as "typical" and "original": a versatile answer to a query could thus sometimes be estimated as fruitful.

The ultimately searched answer could be found on unpredictable instrumental bases:[3], and "multiple search strategies" have been experienced whith STI

[1] http://www.cnrs.fr/dist/z-outils/documents/copist-premiers-resultats.pdf, p. 55 and further

[2] *Ibid*, pp. 11 and 19

[3] P. Feyerabend, *Contre la méthode*, Paris Seuil 1975

context [7]. The same conclusion exists for search tactics: 29 separate ways for search have been decribed [2] and reference chasing remains an “open problem intersecting bibliometrics and information retrieval” [6]. It is then clear that “search is not research” and has its own separate ways wich are numerous and variable [10]. On their side researchers, as users of STI search tools, express a stress of information overload [11], while reviews of literature on recommender systems [21], shows that global performance is questionable. As remarked already, systemic approaches to search are hardly developed on “*collaborative query management system*” or “search and browse interaction” [16], which could be developed, from analyses of parsing and logs on publisher’s knowledge bases (PKB) and from systematic querying of search logs.

At least, variability of behaviors, of practices, of reception of search are interacting an create altogether, the need of a common data structure.

3 Part 2: Modeling a “Keyword Eco-System” Linking Search Behaviors

This second part proposes modeling a kind of an “ecosystem” for a keyword: it describes sub-groups of search and the routes that are open to navigation between them, in “interactive information retrieval process” [7], while studying evolutions of search needs with evolutions of knowledge [15].

The data content of any switched user-item group, informs on “how researchers are searching” in a sub-community, as each “cluster of search” of user-items shared information could differ at any steps of discovery processes and change from one scientific sub-community to the others, or with advancement of scientific ideas from one standpoint, as it is the case in great research infrastructures using the same kind of approach for a large variety of experiments. Main goal is to “re-rank results from the user’s intent” by formalizing an analyzed interaction between queries and results retrieved, and “lower the cognitive burden” of search [22].

Graphs are since long mobilized fo a large variety of tasks related to knowledge representation. Bi-partite graph characteristics [17] are exploited in this article, with a typical *crown-graph* figure, familiar to neural networks approaches, but used here in a novel direction, suggested by information analysis [14] allowed by their typical geometry.

3.1 Definition 1: “Route of Querying”

Any user’s route is figured by a path of vertex, built on three parameters of data routinely recorded for any query on a given keyword: a number of retrieved documents (or URL), b number of users having selected those documents, and c index of correlation intensity of the link between users and documents. This last measure expresses frequency and dynamics of change of matching between each recorded number of users and number of documents, observed in queries belonging to the same group.

Let us consider that our goal is to record “routes of querying” for a given keyword, and that those routes, which of course could differ, are designed to be compared [3].

Let us then write:

$$Q_n = f(N_n, K_n)$$

where Q_n is a number of queries produced by users of a search engine. Let us also consider that for each query Q, we can record a number of users N and a number of items (URL, documents, articles) K recording the same search on a keyword, at the same time among the same corpus of items: we will call this sub-community a community of users of the same “route” of querying.

For any distinct route, we could express the limits of its system of possible queries [13] for a given keyword. Let us pose two limits of substitution between “users” and “items” for any query choice: one limit is where a few users ask for many items and one other limit is when many users will consult a very few items. We then could write these limits of possible substitution between users and items, N and K:

$$Q_1 = f(N_{\max,K_{\min}} \quad \text{or} \quad Q_1 = f(N_{\min,K_{\max}}$$

Or, in a general form:

$$\begin{gathered}[\text{max, min}\\ Q_n = \text{fN}, K\\ [\text{max, min}\end{gathered}$$

3.2 Definition 2: Variation of Queries’ Content

Let us now write α the coefficient of increase of N and β the coefficient of increase of K when Q varies by one unit, that is to say when a new additional query for a given keyword or to a new period of time of the same query is recorded (for instance: check of users interested in items of a domain before and after publication of a famous article). Those coefficients of increase will allow to measure “stability” or instability of the function, according to variations of coefficients of increase or decrease of quantities n and k between two queries of the same route. We will then record whether or not, users and/or items have varied in a correlative way, between queries Q_1 and Q_2.

We assume here that the data structure of any query could vary from one to the other, and that this variation could be expressed by an observable and recordable index of change in the relation between number of users and number of items.

We could then write:

$$\begin{gathered}\alpha \qquad\qquad\quad \beta\\ Q_n = f\ [\text{nmax}, \text{min kmax}, \text{min}\end{gathered}$$

Let us note that value “min” or “max” of n and k give data on the measured quantity of those variables to “produce” Q, but let us note also that, from one unit Q to the other (from one query to the other) the coefficient of increase or of decrease between value "min" or "max" of n and k, gives an additional data on quantity of those variables in production of Q. This additional data gives a value of "stability" or "instability" of the process of querying, which is a dynamic index of behavior. For any unit n and k of production of Q, variation of quantities measured by α and β is a critical information about Q.

3.3 Definition 3: Dynamic Networks of Communities (Users and Items)

We know in fact that with regard to the slope of a set of queries, with also the mix of user and items that it measures, the slope tends to be a stable one when $\alpha + \beta \leqslant 1$ or $(\alpha + \beta) \to 1$: in that last case, transition between Q_{n-1} and Q_n is steady because, as measured by $\alpha + \beta$, variation of $(n + k)$ is still of the same order as the variation of Q. There is no significant increase or decrease in the number of units (users and items) recorded between those two or more successive queries.

Conversely, there could be also an alternative situation in which $(\alpha + \beta) \pm \infty$ that is to say that for each variation of one unit of Q, we have an unstable variation of characteristics of Q, indicated by an unstable variation of $(\alpha + \beta)$, which will increase or decrease abruptly: relations between users and items will change strongly. In that last case, between units Q_{n-1} and Q_n, the variation of the quantities x and y would be unstable, which means that the query's couple of user and item will change significantly. This last situation could indicate that a threshold is reached in the effectiveness of the query.

In fact, in that last case, why continuing to allocate (n, k) to Q if $(\alpha + \beta)$ tends toward minus infinity $-\infty$? In that last case in fact, combination (n, k) becomes over "performing" which is questionable when it's time to decide a new query Q_{n+1} on the same characteristics of n and k. With such characteristics of a query, it could be more interesting to change n or k than to reproduce the same quantities in next query. More generally, for transition between Q_{n-1} and Q_n, choice of stability in n and k could be "*justified"* if we observe that stability prevails with $(\alpha + \beta)$ tends to 1 and "*questionable*" when instability prevails with $\alpha + \beta$ tends to $\pm\ \infty$.

On long series of queries, there will be "learning" phenomena which will give sense to expression of limits to variation of the purposed investigation on a long set of queries covering large fields. In that perspective, it is possible to fix arbitrary limits of variations to a given set of queries, with the "best" and the "worse", like for instance:

$$\text{Best} = N_{\text{max}}, (\alpha + \beta) \to 1, K^{\beta}_{\text{min}}$$

And

$$\text{Worst} = N_{\text{min}}, (\alpha + \beta) \pm \infty, K^{\beta}_{\text{max}}$$

If we have selected those two opposite limits to queries belonging to an STI data search, we could write our limits with the following framework:

"Best"	"Worst"
N_{max}	N_{min}
$(\alpha + \beta) \to 1$	$(\alpha + \beta) \pm \infty$
β	β
K_{min}	K_{max}

3.4 Definition 4: "Compass" for Positioning of Communities Routes

On the basis of preceeding expression, the two triplets could give shape to a formal graph-based presentation of *paths existing between those two limits,* and having the characteristic to position all the *non contradictory solutions* existing between those two limits with three parameters. The following bi-partite Hamiltonian graph with connected nodes, offers that representation.

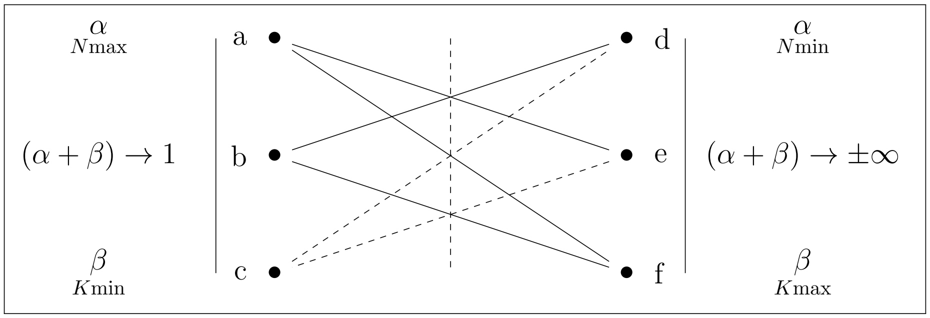

Fig. 1. Compass GRAPHYP fixing nodes and "directions" of routes of search

Figure 1 shows a complete representation of all the typical intermediary situation between our two limits, which gives us a tool for classification of observed queries Q in series of searches on a given keyword, according to users and items choices. Structured by GRAPHYP, this set of query search typical position has two main characteristics, which will be detailed below, with the help of Figure 2.

NODE (Query)	TYPICAL DIRECTION OF SEARCH (Users, Items)
a	$\mathbf{N}\alpha_{\ \mathbf{max}}$, $\mathbf{K}\beta_{\ \mathbf{max}}$, $(\alpha + \beta) \to \pm\ \infty$
b	$\mathbf{N}\alpha_{\ \mathbf{min}}$, $\mathbf{K}\beta_{\ \mathbf{max}}$, $(\alpha + \beta) \to \mathbf{1}$
c	$\mathbf{N}\alpha_{\ \mathbf{min}}$, $\mathbf{K}\beta_{\ \mathbf{min}}$, $(\alpha + \beta) \to \pm\ \infty$
d	$\mathbf{N}\alpha_{\ \mathbf{min}}$, $\mathbf{K}\beta_{\ \mathbf{min}}$, $(\alpha + \beta) \to \mathbf{1}$
e	$\mathbf{N}\alpha_{\ \mathbf{max}}$, $\mathbf{K}\beta_{\ \mathbf{min}}$, $(\alpha + \beta) \to \pm\ \infty$,
f	$\mathbf{N}\alpha_{\ \mathbf{max}}$, $\mathbf{K}\beta_{\ \mathbf{max}}$, $(\alpha + \beta) \to \mathbf{1}$

Fig. 2.

As observed in **Figure 2**, GRAPHYP expresses the whole ensemble of *non-contradictory positions that structured routes of search could occupy during any search* on a given keyword in a data base. This has consequences, defined by three characteristics on operating rules of GRAPHYP.

Characteristic 1: GRAPHYP functions like a compass as it supplies all possible directions, and "locates" the recorded one of a query Q. This set of possible "positions" during search gives an overview of the proposed modeling of the searchable space of STI data base.

Characteristic 2: The second feature of GRAPHYP modeling of search is that it allows recording and comparisons of "behaviors" of querying search. It allows users to *learn from their past recorded behavior* as well as from the recording of other users of the same base, if data is made accessible to all users.

In those two ways, the "compass" has the functionality to "orient" and "locate" search attitudes as being recorded as individual dynamic behaviors.

Characteristic 3: GRAPHYP allows also expressing, by summation of individual attitudes, the **global trend of a community** of users according to characteristics of their uses.

One could at least select arbitrarily one node, between a and e, to fix there an **optimum point**, and measure the distance of a set of observed routes toward that selected node. Mapping of routes of STI search becomes possible, like in air or sea routes. Cooperative or conflicting routes could be identified by the data structure of GRAPHYP.

3.5 Navigation Rule 1: A "Progressive Learning System" (PLS)

GRAPHYP creates opportunity to reach output (STI data search results) while identifying alternative ways to reach it for a given input (query). In that way, *learning profile* of users, the way they access to items and which items could be thus compared and located locally and globally. GRAPHYP could be used as a kind of a browser for uses of a STI community.

Like any compass, GRAPHYPER secures navigation's parameters at any scale (see infra), owing to its built-in characteristics of information processing and transfer, to its functions of positioning previously recorded routes, and to its potential of recording and recommending paths of navigation. In that way *GRAPHYP belongs to learning systems combining intuitive choices and automatic path definition at any operational scale.*

Another notion, interaction of routes, could be found in the specific context of "mutual reachability" of nodes is one central issue of this part of this article.

From this starting point, with the processing resources supplied by GRAPHYP, one could describe various functionality of forecast.

3.6 Navigation Rule 2: Fixing Affinity of Communitie's Routes

Let us take all nodes (a, b, c, d, e, f) as possible starting points for definitions of "goals" of routes of search at time T. Any other node could be designed as "objectives" to be reached on a route built from the starting point of a given node. We could then record "search profiles" corresponding to circulation from one node to the other on a recorded route.

From design of Figure 3 above, one could observe that a continuous circulation from node to node, let appear a common edge in all the pairs corresponding to any node: node *a* has a common edge with node *e*, *b* has *a* common edge with *f*, etc. From this standpoint, we could remark that it is continuously possible to have a circular "travel" inside GRAPHYP, from edge to edge, while coming back to starting node by the way of the complementary edge of that node.

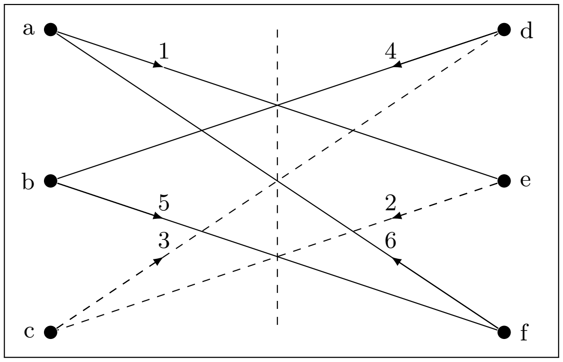

Fig. 3. Recording and Positioning of Nodes on Routes of Data Search in GRAPHYP

3.7 Navigation Rule 3: Mapping New Routes between Communities

For instance, as shown on Figure 3, node a could allow an exploration of GRAPHYP's other nodes which will represent six steps on one way, designed here. An other six-step way could be practiced from the same node *a* starting from its other edge. There are at least 12 possible steps which could be explored from any node of that eulerian circuit, and GRAPHYP offers then 72 possibilities of identification of characteristic "search profile" which could be modeled and recorded in an STI data base.

In that ways could be identified numerous characterized positions of exploration of a data base, and this identification of conditions of navigation: it offers users and their community precise records of explored and unexplored ways to discovery.

One could remark that those explorations of "possible" routes, could be managed altogether with any classical probabilist approach, with which the possibilist approach could combine in any purposed uses.

3.8 Navigation Rule 4: Fixing a Search Profile

We could also apply the design of Figure 3 to propose another significant feature of traceability of the searchable space of STI, which is recording search profiles, which allows to "read" characteristic steps of a route of navigation in a data base. This result could be produced when projecting on x, y axes as shown on Figure 4.

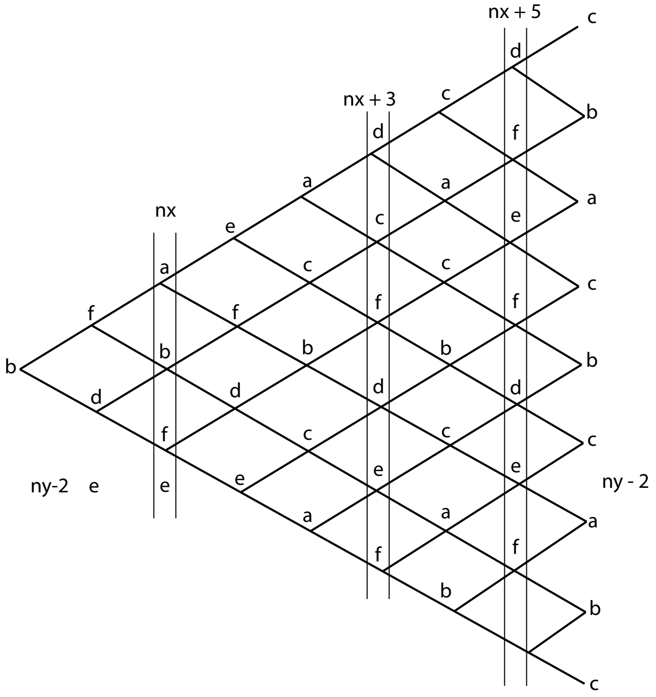

Fig. 4. Recording and positioning of nodes on routes of data search in graphyp

Based on the design sketched by Figure 4, one could, for any query applied to this grid of structured data, (in our example: $nx + 3$ and $nx + 5$, $ny2$) it becomes feasible to identify the origin, and so the "kinship" of any position located on GRAPHYP recorded data base. One could then *evaluate all diverging or converging solution surrounding any observed node located on the grid.* For any observed network, there exists an "**induced network**" which applies the property of "mutual reachability" of connected nodes in a network.

Mapping of routes of documentation using those ideas, should help confronting strategies of discovery.

3.9 Navigation Rule 5: Mapping Distances Between Communities

Exploration of neighbor routes is an important feature of the searchable space, as it informs on its depth and treewidth and, as far as we know, there are hardly no developments in mapping of this aspect of searchable space of STI.

Reduction of uncertainty on size and direction of future routes to discovery could benefit from representations of ways which have been followed and/or abandoned (Figure 4). On that base, exploration of neighbor coming routes could benefit from tree couples of edges which shapes this node:

$$a = (e, f) \quad b = (d, f) \quad c = (d, e) \quad d = (b, c)$$
$$c = (a, c) \quad f = (a, b)$$

We face here the grid of nodes which are linked by a specified common edge, resulting from the position of each node on the grid. We could then propose the mapping of a grid of *correlated neighbor routes*. The grid observed on Figure 5, structured whith same nodes as in grid of Figure 4, gives opportunity of projection that could be studied in a further works.

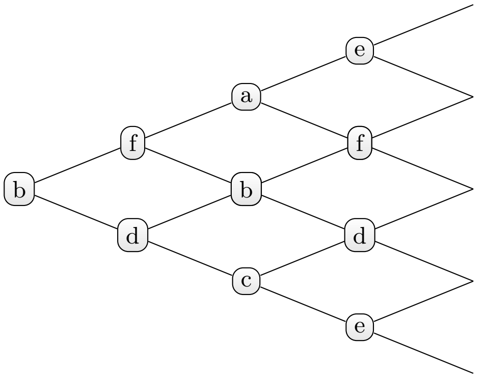

Fig. 5. Grid of Correlated Neighbor Routes

3.10 Navigation Rule 6: Fractal Scalability of Searchable Space

A final property of the bipartite graph that gave birth to GRAPHYP is self-similarity which is designed on the following Figure 6.

The graph with nodes A,B, C. . . .is built on a larger dimension of the preceeding one and the addition of those “compasses” on a self-similar basis, allows building of architectures of information of the same type of information processing at any scale, and the application to the operating frame of GRAPHYP, of basic operation of addition, subtraction, multiplication, division. Scalability of the system could thus be established as a final feature of traceability of that system. We began already further works on that characteristic of the data structure.

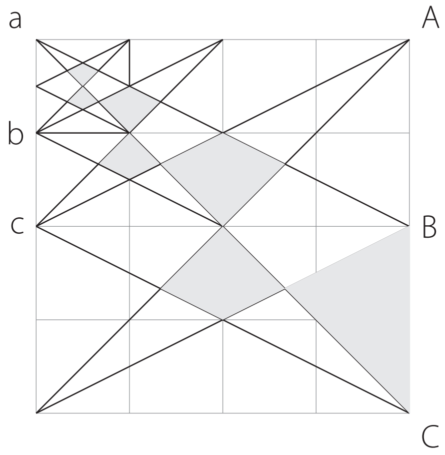

Fig. 6. Self Similarity of the Bipartite Graph Graphyp

References

1. Bast, H., Buchhold, B., Haussmann, E.: Semantic search on text and knowledge bases. Found. Trends Inf. Retr. **10**(2-3), 119–271 (Jun 2016). https://doi.org/10.1561/1500000032
2. Bates, M.J.: Information search tactics. Journal of the American Society for Information Science **30**, 205–214 (July 1979)
3. Bendersky, M., Wang, X., Metzler, D., Najork, M.: Learning from user interactions in personal search via attribute parameterization. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. pp. 791–799. WSDM '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3018661.3018712, http://doi.acm.org/10.1145/3018661.3018712
4. Bing, L., Zheng-Yu, Li, N.P., Lam, W., Wang, H.: Learning a unified embedding space of web search from large-scale query log. Knowledge-Based Systems, Elsevier (2018), available online
5. Cabanac, G., Chevalier, M., Chrisment, C., Julien, C.: Organization of digital ressources as an original facet for exploring the quiescent information capital of a community. Int J Digit Libr **11**, 239–261 (2010)
6. Cabanac, G.: What is the primordial reference for ...?—redux. Scientometrics **114**(2), 481–488 (Feb 2018). https://doi.org/10.1007/s11192-017-2595-4, https://doi.org/10.1007/s11192-017-2595-4
7. Carevic, Z., Lusky, M., van Hoek, W., Mayr, P.: Investigating exploratory search activities based on the stratagem level in digital libraries. CoRR **abs/1706.06410** (2017). https://doi.org/10.1007/s00799-017-0226-6, http://arxiv.org/abs/1706.06410
8. Conrad, L.Y., Moeller, P.D.: Search, serendipity, and the researcher experience. The Serials Librarian **72**(1-4), 190–193 (2017). https://doi.org/10.1080/0361526x.2017.1292744

9. Fabre, R.: New Challenges for Knowledge: Digital Dynamics to Access and Sharing. Wiley (2016). https://doi.org/DOI:10.1002/9781119378112
10. Feyerabend, P.: Contre la méthode. Seuil, Paris (1975)
11. Gibney, E.: How to tame the flood of literature: Recommendation services claim to help researchers keep up with the most important papers without becoming overwhelmed. Nature **513**(7516) (2014)
12. Gregory, K., Groth, P., Cousijn, H., Scharnhorst, A., Wyatt, S.: "searching data: A review of observational data retrieval practices". ArXiv (2017)
13. Gutierrez Soto, C.: Exploring the Reuse of Past Search Results in Information Retrieval. Ph.D. thesis, Université Paul Sabatier, Toulouse III (2016)
14. Haynes, P.S., Alboul, L., Penders, J.: Dynamic graph-based search inunknown environments. Journal of Discrete Algorithms (6 july 2011), online
15. Kacem, A., Mayr, P.: Users are not influenced by high impact and core journals while searching. GESIS (2018), bIR 2018 Workshop on Bibliometric-enhanced Information Retrieval
16. Khoussainova, N., Alazinska, M., Gatterbauer, W., Chul Kwon, Y., Suciu, D.: A case for a collaborative query management system. Cidr (2009), u. of Washington, Seattle
17. Kolda, S.G.: Measuring and modeling bipartite graphs with community structure. Journal of Complex Networks **vol 5**, 581–603 (2017)
18. Schöpfel, J., Ferrant, C., Andre, F., Fabre, R.: Research data management in the french national research center. Data Technologies and Applications (2018). https://doi.org/10.1108/DTA-01-2017-0005
19. The National Academies of Science, E., Medicine: Communicating science effectively, a research agenda. Consensus study report (2017)
20. Singh, V.: Ranking and query refinement for interactive data exploration. Procedia Computer Science **125**, 550–559 (2018). https://doi.org/doi.org/10.10 16
21. Villegas, N., Sanchez, C., Daz-Cely, J., Tamura, G., Icesi, U., Colombia: Character-izing context-aware recommender systems: A systematic literature review. Knowledge-Based Systems **140**, 173–200 (2018)
22. Yan, Z., Zheng, N., Ives, Z.G., Talukdar, P.P., Yu, C.: Active learning in keyword search-based data integration. The VLDB Journal **24**, 611–631 (October 2015)