

A Novel Method for Determining Research Groups from Co-Authorship Network and Scientific Fields of Authors

Jan Pisanski

University of Ljubljana, Faculty of Arts

E-mail: Jan.Pisanski@ff.uni-lj.si, <http://oddelki.ff.uni-lj.si/biblio/oddelek/osebje/pisanski.html>

ORCID 0000-0002-3060-111X

Mark Pisanski

Tomaž Pisanski (corresponding author)

University of Primorska, FAMNIT

E-mail: Tomaz.Pisanski@upr.si, https://en.wikipedia.org/wiki/Tomaz_Pisanski,

ORCID 0000-0002-1257-5376

Keywords: co-authorship network, scientific field, maximal spanning tree, induced subnetwork, pruning of networks, pathfinder network, MST, line-cut

Received: March 9, 2020

Large networks not only have a large number of vertices but also have a large number of edges. Although such networks are generally sparse they are usually difficult to visualise, even locally. This paper considers the case where large weights on edges represent similarity between the corresponding end-vertices. We follow two main ideas in this paper. The first one is network pruning, that is removal of edges that makes the resulting network more manageable while keeping the main characteristic of the original network. The other idea is to partition the network vertex set in such a way that the induced connected components represent groups of network elements that fit together. Furthermore, we assume that the vertices of the network are labeled by types. Here we apply our approach to co-authorship network of researchers in Slovenia in order to identify research groups, finding group leaders and the degree of interdisciplinarity of the group. For the network pruning phase we use a MST-pathfinder network and for vertex partition appropriate line-cuts. Each cluster is assigned a distribution of types. In this case, the types correspond to scientific fields, also known as research interests of authors. A measure of interdisciplinarity of research group is derived from such a distribution.

Povzetek: Velika omrežja nimajo le mnogo vozlišč, ampak imajo tudi mnogo povezav. Čeprav so običajno taka omrežja redka, so nepregledna in jih je težko prikazati na sliki, tudi lokalno. Ta prispevek obravnava omrežja, pri katerih velike vrednosti uteži na povezavah pomenijo podobnost pripadajočih krajišč. V prispevku sledimo dvema glavnima idejama. Prva je kleščenje omrežja, to je odstranitev manj pomembnih povezav, zaradi česar je nastalo omrežje bolj obvladljivo, hkrati pa se ohrani glavna značilnost prvotnega omrežja. Druga ideja je razdeliti vozlišča omrežja tako, da inducirane povezane komponente predstavljajo skupine omrežnih elementov, ki se med seboj prilegajo. Poleg tega predpostavljamo, da so vozlišča omrežja označena s tipi. V tem prispevku uporabljamo naš pristop k omrežju soavtorstev raziskovalcev v Sloveniji z namenom identifikacije raziskovalnih skupin, iskanja voditeljev skupin in stopnje interdisciplinarnosti skupine. Za fazo kleščanja omrežja uporabljamo usmerjevalno omrežje (MST-pathfinder network), za vozliščno razbitje pa ustrezne reze povezav. Vsaki skupini je dodeljena porazdelitev tipov. Mero interdisciplinarnosti raziskovalne skupine izpeljemo iz takšne porazdelitve. V tem primeru tipi predstavljajo znanstvena področja, oz. raziskovalne interese avtorjev.

1 Introduction

In contemporary research community scientists collaborate within formal or informal research groups. Identifying such groups from data available in various bibliometric networks is an interesting challenge. In this note we propose a method that uses the co-authorship network on the one hand and declared scientific field of authors that can be extracted from some bibliographic databases, on the other.

We propose a theoretical model that uses a network, i.e. graph with weights on edges and labels, called types, on its vertices. We may view labels as scientific fields or sub-fields. Our approach is quite general and can be applied to any weighted network with types. In this paper we apply it to co-authorship networks. Note that scientific fields are sometimes called research interests.

The method consists of two steps. In the first step the original co-authorship network is pruned in order to reduce

the number of edges and increase the number of components, in our case producing research groups. In this step line-cuts are determined. In the second step a collection of induced monotype subnetworks is pruned by applying MST-pathfinder algorithm to further reduce the number of edges while keeping the same connectivity. Our original contribution is combination of both methods and the use of symmetric predicate in the first step; see Algorithm 3. Note that the idea of using MST, pathfinder and MST-pathfinder has been used extensively in the past in variety of contexts of bibliographic and other research[6, 8, 11, 22, 23].

This rough general approach may be refined in several different ways. We present in detail only one such refinement and discuss some others in the conclusion. In general, bibliographic networks are very large and allow for a variety of methods for data mining [15], however in this pilot study we focus our attention on a relatively small data set. The data is restricted to Slovenian researchers and is taken from Slovenian bibliographic system SICRIS. Moreover, only researchers that are co-authors of mathematicians are considered.

2 Pruning of co-authorship network

2.1 Co-authorship network

For basics in graph theory, the reader is referred to [4], for network theory, see for instance [3].

Let V be a list of authors from some bibliographic database. We say that $u, v \in V$ are adjacent: $u \sim v$, if u and v are co-authors of a common work from the corresponding database. Sometimes we restrict our attention to certain types of works or certain types of co-authorships. Usually only scientific works are considered and the co-authorship graph is computed from a two-mode network WA composed of pairs (w, a) , works and authors for each co-author a of work w . Since binary relation \sim is irreflexive¹ and symmetric it defines a simple graph $G = (V, \sim)$ that we call the *co-authorship graph*. Let $E = \{uv \in V^2 | u \sim v\}$ denote the set of unordered pairs of adjacent vertices of G . Instead of $G = (V, \sim)$ we may use notation $G = (V, E)$ to denote the same graph. The graph may be weighted where the weights w on the edges represent the number of joint papers between the two authors. In this way a *network* $N = (V, E, w)$ is obtained. Let $w(u, v)$ denote this weight. Sometimes, we may consider the weight of co-authorship differently for different number $n(w)$ of co-authors of work w . Let $W(u, v)$ denote the collection of works co-authored by u and v . For any work w let $n(w)$ denote the number of authors of w . Then

$$w(u, v) = |W(u, v)|.$$

¹Sometimes one may use also loops at each vertex. The weight associated with a loop may depend on the method that the co-authorship graph is constructed. If it is obtained by multiplication of two-mode networks [4] it represent the number of works for a given author. In the fractional approach it may represent the total contribution of an author. Loops are removed if we follow Newman's approach.

In a fractional approach [2] the weight $f(u, v)$ is defined as:

$$f(u, v) = \sum_{w \in W(u, v)} \frac{2}{n(w)^2}.$$

In case of Newman's normalization the weight is:

$$f(u, v) = \sum_{w \in W(u, v)} \frac{2}{n(w) \cdot (n(w) - 1)}$$

A network N is a weighted graph $N = (V, E, w)$, where $w : E \rightarrow \mathbb{R}$ is the *weight function*. In our case it is positive and the value 0 means there is no edge between u and v .

A graph $G = (V, \sim)$ is transformed into the network $N = (V, E, a)$, where $a(u, v) = 1$ for all adjacent pairs of vertices $u \sim v$. The same bibliographic database can produce at least three types of networks for the weight functions a, w, f , defined above:

1. (V, E, a) , the *binary case*,
2. (V, E, w) the *standard case*, and
3. (V, E, f) , the *fractional case*.

Algorithm 4 Prune the network $N = (V, E, w)$,
 $n = |V|, m = |E|$

- 1: Partition the edge set E into subsets E_i with equal weights: $E_i = \{e \in E | w(e) = w_i\}$.
 - 2: Order the parts in descending order of weights $w_1 > w_2 > \dots > w_k$
 - 3: **for** $u \in V$ **do**
 - 4: $S_u = \{u\}$
 - 5: $F = \emptyset$
 - 6: **for** $i = 1, 2, \dots, k$ **do**
 - 7: $F_i = \emptyset$
 - 8: **for** $e = uv \in E_i$ **do**
 - 9: Let S_u, S_v be the corresponding sets.
 - 10: **if** $S_u \neq S_v$ **then**
 - 11: Append e to F_i .
 - 12: **for** $e = uv \in F_i$ **do**
 - 13: **if** $S_u \neq S_v$ **then**
 - 14: $S_u = S_v = S_u \cup S_v$
 - 15: Extend F by F_i
 - 16: **return** subnetwork $Pr(N) = (V, F, w)$.
-

There is another aspect that we have not considered in this paper. Namely, the weight of an edge $e = uv$ between two authors u and v may depend also on the total number of papers authored by each of the two authors. In this case we may modify the network to allow loops and define $w^*(u, v) = w(u, v)$ for $u \neq v$ and let $w^*(u) = w^*(u, u)$ denote the total number of papers having u as an author. Note that in general w^* cannot be computed directly from w since we have no information about the single-authored papers. In this case the best way to compute w^* is to multiply WA^T by WA , where WA represents a two-mode network work-author. The theory of two mode networks and

their applications to bibliographic data can be found, for instance in [3].

2.2 Pruning networks

In the analysis of large networks, dense networks present a challenge. Usually one tends to partition the set of vertices and investigate the induced networks on such parts. In [3] one may find a variety of concepts that are useful in such analysis, e.g. cuts, islands, etc. Nevertheless, such subnetworks may be dense again and the role of particular vertices is not clearly visible. For this reason we prune the original network $N = (V, E, w)$ by appropriately selecting a subset of important edges $E' \subset E$. If w' denotes the restriction of w on E' , the pruned subnetwork $N' = (V, E', w')$ is obtained.

In case of co-authorship networks large weights indicate close collaboration between authors. When considering research groups one may assume strong collaboration within each group. Hence, in such a case a natural approach to pruning would be to remove all edges of lesser weights, while keeping the same connected components. A possible solution is given by the well-known maximum cost spanning tree. More precisely, in case of a disconnected network the resulting graph is a maximum cost spanning forest.

However, the problem with a maximum cost spanning forest is that, in case when several edges have the same weights, the forest may not be unique. We use a Kruskal-like algorithm that produces a unique pruned network. Algorithm 1 is almost identical to the MST-pathfinder algorithm of [14] and produces the *pathfinder network* $Pn(\infty, n - 1)$; for discussion and various aspects see also [19, 5, 20].

It is not hard to see, that the following is true:

Theorem 1. *The network $N'(V, F, w)$ is uniquely determined from the original network. If all weights are positive, the connected components are the same as in the original network.*

Theorem 2. *If all weights in $N(V, E, w)$ are distinct, Algorithm 1 produces the (unique) maximum cost spanning forest. On the contrary, if all weights are the same no edge is discarded.*

Moreover, we easily compute the running time of Algorithm 1.

Theorem 3. *The time complexity of the above algorithm is $\mathcal{O}(m \log m)$.*

In fact, the time complexity is the same as for Kruskal's algorithm [10]. The sorting and partitioning takes $\mathcal{O}(m \log m)$ steps. There are two loops, each with $\mathcal{O}(m)$ steps, and the time complexity for the UNION-FIND is of lesser order.

By applying this pruning method strong ties among the nodes remain visible.

3 Line-cuts

For further refining the network $N(V, E, w)$ one may choose a parameter $t > 0$, the threshold, or cut parameter and prune the edges with weights less than t . In this way the network $N_t(V, E_t, w)$ is obtained, where

$$E_t = \{e \in E | w(e) \geq t\}.$$

The choice of parameter t depends on our aims. There are several obvious goals. For instance:

1. We may choose maximal value of t that guarantees at least a prescribed number of connected components, say κ .
2. An alternative is to insist that all components have at most prescribed number of vertices, say ν .

We present the basic pruning algorithm; see Algorithm 2. It produces essentially a *line-cut*, see for instance [3]. The only difference is that we keep isolated vertices.

Algorithm 5 Prune the network $N = (V, E, w)$, given threshold parameter t . Connected components of the resulting network are called *line-cuts*.

```

1:  $F = \emptyset$ 
2: for  $e \in E$  do
3:   if  $w(e) \geq t$  then
4:     Append  $e$  to  $F$ .
5: return subnetwork  $Pr(N, t) = (V, F, w)$ .
```

In Python, Algorithm 2 can be reduced to a single statement:

$$F = [e \text{ for } e \in E \text{ if } w(e) \geq t]$$

4 Pruning networks with vertex types

Let us assume we are given a finite number of types, or colors T , a network $N(V, E, w)$ and a mapping $c : V \rightarrow T$. The structure $N(V, E, w, T, c)$ will be called a *weighted network with vertex types*. When pruning network with vertex types, a connected component consisting of vertices of a single type will be called *monotype*. Additionally, we will refer to the number of types used in a connected component as its *type number*. The maximum of type numbers of network components is called the type number of the network. In particular, we are interested in networks of low type number, preferably with monotype networks.

Parameters of pruning may be adjusted in such a way that a monotype network is obtained. For networks with vertex types, in addition to the two goals described in Section 3, a third goal may be considered.

- One may insist that all connected components are monotype, or more general that each component has at most δ types (colors).

The following basic algorithm (Algorithm 3) for a given network with types removes all edges that have endpoints of different types, or more generally, when they satisfy a symmetric predicate P .

Algorithm 6 Prune the network with vertex types $N = (V, E, w, T, c)$ with given threshold parameter t and (symmetric) predicate $P : T^2 \rightarrow \{\top, \perp\}$. Connected components of the resulting network are called *monotype line-cuts*.

```

1:  $F = \emptyset$ 
2: for  $uv = e \in E$  do
3:   if  $P(c(u), c(v))$  and  $w(e) \geq t$  then
4:     Append  $e$  to  $F$ .
5: return subnetwork  $Pr(N, t, P) = (V, F, w, T, c)$ .
```

As we mentioned above the predicate P usually is true if both endpoints are of the same type. However, other options are possible. Namely we may have a similarity imposed on the predicates and P signifies that two types are sufficiently similar.

We need an algorithm to analyse the network with vertex types; see Algorithm 4. Using these numbers we may select different parameters and re-run this algorithm to reduce the size of the maximal component or alternatively limit the number of different components. We may also insist that all components be composed of a single type.

5 Interdisciplinarity of research groups and leaders of research groups

For a given network with vertex types one may perform basic statistics on it. Namely, one may compute absolute frequencies of types on the vertex set.

$$f(x) = |\{v \in V | c(v) = x\}|$$

$$f_i(x) = |\{v \in V_i | c(v) = x\}|$$

or relative frequencies

$$\phi(x) = f(x)/n$$

$$\phi_i(x) = f_i(x)/n_i$$

where $n = |V|$ and $n_i = |V_i|$.

We consider two measures:

$$r(N) = \max\{\phi(x) | x \in T\}$$

$$s(N) = |\text{supp } \phi| = |\{x \in T | \phi(x) > 0\}|$$

and for each component:

$$r(V_i) = \max\{\phi_i(x) | x \in T\}$$

$$s(V_i) = |\text{supp } \phi_i| = |\{x \in T | \phi_i(x) > 0\}|$$

Both measure the diversity of research interests in a research group. If $r(V_i) < 0.5$ there is no dominant discipline. If $r(V_i) = 1$, the group is totally homogeneous.

Algorithm 7 Analyse network with vertex types $N = (V, E, w, T, c)$.

```

1: Partition  $V$  into connected components  $V_1, V_2, \dots, V_k$ 
2: Let  $d = \max\{|V_j|; j = 1, 2, \dots, k\}$ 
3: for  $j = 1, 2, \dots, k$  do
4:    $b(j) =$  number of different types in  $V_j$ .
5: Let  $\gamma = \max\{b(j); j = 1, 2, \dots, k\}$ 
6: return number of connected components  $k$ , order of maximal connected component  $d$ , and maximal number of types  $\gamma$  in any component.
```

One way to define a leader of a research group is to determine the vertex of maximal degree in the corresponding network, or even better the sum of weights of edges to the neighbouring vertices. There are two parameters that we are interested in. Let m be the number of edges of network N and let d be the maximal degree attained at vertex x . Let d' be the second largest degree. Then x can be defined as a leader of the research group, while *dominance* is the quotient d/m and *absolutism* is defined by expression $1 - d'/d$. Note that it would be also interesting to explore the *diversity index* [21] in this context. However, we will address all of these in a future work.

6 Example

The data used in our experiments was taken from COBIS-S/SICRIS [18] for the works indexed in Scopus [17]. Only papers, where at least one author was a mathematician, were considered. Co-authors that were not registered as researchers in Slovenia were not included. Scientific fields alias research interests, used in Slovenia have three levels. On Level 1 we have:

- 1 Science
- 2 Engineering
- 3 Medicine
- 4 Biotechnology
- 5 Social Sciences
- 6 Humanities
- 7 Interdisciplinary

The next table shows division of Science on Level 2.

- 1.01 Mathematics
- 1.02 Physics
- 1.03 Biology
- 1.04 Chemistry
- 1.05 Biochemistry and Molecular Biology
- 1.06 Geology
- 1.07 Comp. Intensive Methods and Appl.
- 1.08 Environment Protection
- 1.09 Pharmacy

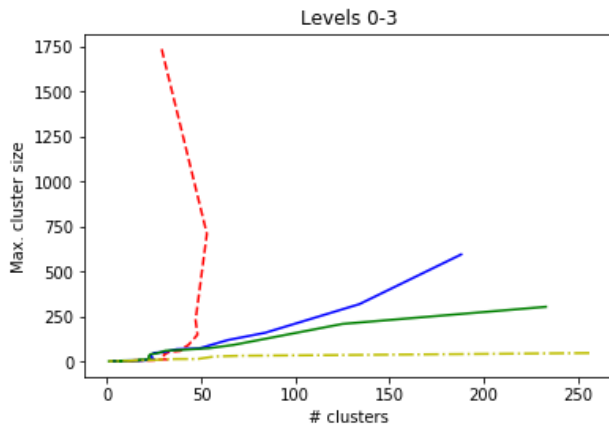


Figure 1: Line-cuts for threshold values $t = 0, 1, 2, \dots$ for four different predicates, each depending on the level $\ell = 0, 1, 2, 3$. Each predicate depends on the interpretation of equality between two research types. Red – $\ell = 0$, blue – $\ell = 1$, green – $\ell = 2$ and yellow – $\ell = 3$.

Finally, the division of Mathematics in the Level 3 is indicated here:

- 1.01.01 Analysis
- 1.01.02 Topology
- 1.01.03 Numerical and Computer Mathematics
- 1.01.04 Algebra
- 1.01.05 Graph Theory
- 1.01.06 Probability and Statistics

Level ℓ may be interpreted as the length of the research interest code that is used to test equality: for $\ell = 0$, the string is not used at all, for $\ell = 1$ only the first characters are compared, for $\ell = 2$, the first four characters are compared, while for $\ell = 3$ all seven characters are compared. Different levels can be associated with the suitable choice of predicate P in Algorithm 3. Let P_ℓ denote the predicate applicable to level ℓ . For instance, for $u = 1.01.01$ and $v = 1.01.04$ $P_2(u, v) = \top$ while $P_3(u, v) = \perp$.

Here we give an example of a pruned research group network. We intend to perform a thorough analysis on more complete data set elsewhere.

Figures 2 and 3 depict the same research group. The network in Figure 3 is tree-like and is obtained from the

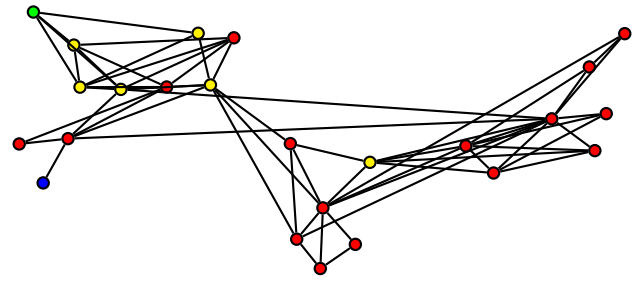


Figure 2: One of several research groups determined by the line-cut for $t = 8$.

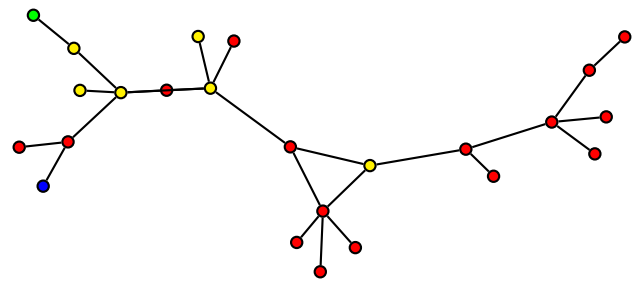


Figure 3: The research group of Figure 2, pruned by the MST-pathfinder network. Red – Graph Theory, Yellow – Algebra, Blue – Numerical Mathematics, Green – Mathematics

network of Figure 2 by MST-pathfinder method. Vertex colors denote vertex types: red: 1.01.05, yellow: 1.01.04, blue: 1.01.03 and green 1.01.00. In the database some researchers were assigned research interest at level 2, e.g. 1.01 (Mathematics). For consistency, we expanded that to level three as 1.01.00. Note that the research group in Figure 3 is composed of two subgroups, one predominantly interested in graph theory and the other in algebra. There is a central triangle connecting the two subgroups.

7 Conclusion

Co-authorship graphs and networks are important in the study of research structure and dynamics; see for instance [7, 9, 12, 1]. Their practical value has first been recognised by specialised systems, such as MathSciNet and Zb-Math; see [16, 24]. Including them in more general bibliographic systems such as SICRIS [18] would be beneficial for most users. Potential applications are plenty. In this paper we presented only one aspect of such applications. In a recent paper [13] a completely different application is sought, namely, organising talks at a conference in such a way that speakers with similar topics are scheduled at different times.

The data that was available to us has also authors with UNKNOWN research interest. In this preliminary study we considered it as a separate research interest. It would be interesting to repeat the study with some flexibility and con-

sider the function: $c : V \rightarrow T \cup \{\text{UNKNOWN}\}$.

Clearly line-cuts refine the vertex partition and apply only within a component. Note that in general one could take different thresholds in different components. In case we intend to have components with given maximal size ν , then indeed different threshold values may be used. In or future more comprehensive work we intend to address some further extensions and applications of the MST-pathfinder method as well as some of the parameters that we have introduced.

Acknowledgement

We would like to thank Vladimir Batagelj for useful advice and fruitful discussion. The work of both referees improved the presentation considerably and is gratefully acknowledged. Work of Jan Pisanski is supported in part by the ARRS grants P5-0361 and J5-8247, while work of Tomaž Pisanski is supported in part by the ARRS grants P1-0294, J1-7051, N1-0032, and J1-9187.

References

- [1] T. Bartol, G. Budimir, P. Južnič, K. Stopar. Mapping and classification of agriculture in Web of Science: other subject categories and research fields may benefit. *Scientometrics*, vol. 109 (2016) no. 2, pp. 979-996. <https://doi.org/10.1007/s11192-016-2071-6>
- [2] V. Batagelj. On Fractional Approach to Analysis of Linked Networks, *Scientometrics* (2020). <https://doi.org/10.1007/s11192-020-03383-y>
- [3] V. Batagelj, P. Doreian, A. Ferligoj, N. Kejžar. *Understanding large temporal networks and spatial networks: Exploration, pattern searching, visualization and network evolution*, (2014) John Wiley & Sons.
- [4] J.A. Bondy and U.S.R. Murty. *Graph theory*, (2008) Graduate Texts in Mathematics, 244. Springer, New York.
- [5] C. Chen. Science mapping: a systematic review of the literature. *Journal of Data and Information Science*, vol. 2 (2017) no.2, pp1-40. <https://doi.org/10.1515/jdis-2017-0006>
- [6] C. Chen, S. Morris. Visualizing evolving networks: Minimum spanning trees versus pathfinder networks. In *IEEE Symposium on Information Visualization 2003* (2003), pp. 67-74. <https://doi.org/10.1109/INFVIS.2003.1249010>
- [7] A. Ferligoj et al. Scientific collaboration dynamics in a national scientific system. *Scientometrics*, vol. 104, (2015), no. 3, pp. 985–1012. <https://doi.org/10.1007/s11192-015-1585-7>
- [8] M. Gallivan, M. Ahuja. . Co-authorship, homophily, and scholarly influence in information systems research. *Journal of the Association for Information Systems*, vol. 16 (2015) no. 12, 2. <https://doi.org/10.17705/1jais.00416>
- [9] L. Kronegger, F. Mali, A. Ferligoj, and P. Doreian. Collaboration structures in Slovenian scientific communities. *Scientometrics*, vol. 90 (2012), no.2, pp. 631–647. <https://doi.org/10.1007/s11192-011-0493-8>
- [10] J.B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*. vol.7 (1956) no.1, pp. 48–50. <https://doi.org/10.1090/S0002-9939-1956-0078686-7>
- [11] T. Jacobsen, R.L. Punzalan, M.L. Hedstrom. Invoking “collective memory”: Mapping the emergence of a concept in archival science. *Archival Science*, vol. 13(2013)no. 2-3, pp. 217-251.
- [12] S. Pečlin, P. Južnič, R. Blagus, MČ. Sajko, J. Stare. Effects of international collaboration and status of journal on impact of papers. *Scientometrics*, vol. 93 (2012) no. 3, pp. 937-948. <https://doi.org/10.1007/s11192-012-0768-8>
- [13] J. Pisanski, T. Pisanski. The use of collaboration distance in scheduling conference talks. *Informatika : an international journal of computing and informatics*, vol. 43 (2019) no. 4, pp. 461–466, <https://doi.org/10.31449/inf.v43i4.2832>.
- [14] A. Quirin O. Cordón, V. P. Guerrero–Bote, B. Vargas–Quesada, F. Moya–Anegón. A quick MST-based algorithm to obtain Pathfinder networks $(\infty, n - 1)$, *Journal of the American Society for Information Science and Technology* vol. 59 (2008) no. 12, pp.1912–1924. <https://doi.org/10.1002/asi.20904>
- [15] J. Leskovec, A. Rajaraman, and J. Ullman. *Mining of Massive Datasets* (2014), Cambridge University Press. <https://doi.org/10.1017/CBO9781139924801>
- [16] MathSciNet: <https://mathscinet.ams.org/mathscinet/index.html>
- [17] Scopus: <https://www.scopus.com/home.uri>
- [18] SICRIS: <https://www.sicris.si/public/jqm/cris.aspx?lang=eng>
- [19] A. Vavpetič, V. Batagelj, V. Podpečan. An implementation of the Pathfinder algorithm for sparse networks and its application on text network, In M. Bohanec (ed.), *12th International Multiconference Information Society*, vol. A (2009) pp. 236–239.

- [20] HD. White. Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, vol. 54 (2003) no. 5, pp. 423-434. <https://doi.org/10.1002/asi.10228>
- [21] Diversity index, Wikipedia, https://en.wikipedia.org/wiki/Diversity_index
- [22] H. Yang, HJ. Lee. Research trend visualization by MeSH terms from Pubmed. *International journal of environmental research and public health*, vol. 15 (2018) no. 6, 1113. <https://doi.org/10.3390/ijerph15061113>
- [23] SY. Yu. Detecting collaboration patterns among iSchools by linking scholarly communication to social networking at the macro and micro levels. *LIBRES: Library and Information Science Research Electronic Journal* vol. 23 (2013) no. 2, pp. 1–13.
- [24] zbMATH:
<https://zbmath.org/>

