Identification of Scientific Collaborations: Information Model and Quality Features

Nosov Magnitogorsk State Technical University. Lenin Street, 38, Magnitogorsk City, Chelyabinsk Region, Russian Federation, 455000

**Keywords:** Publication activity, Graph, Fuzzy sets, Assessment of rankings.

**Abstract.** The paper considers the problem of constructing a system for evaluating the research activities of the scientific and pedagogical staff (SPS) in a higher educational institution. The analysis of the SPS publication activity in the Nosov Magnitogorsk State Technical University is carried out. The concept of publication collaboration and its comparison with the definition and type of the graph is presented. The research revealed some signs of false collaborations and formed the structure of the array for the collaboration of two scientific and pedagogical workers. Based on the description of the array, an adjacency matrix is constructed, the components of which are its elements. As a result of applying the theory of fuzzy sets, the structure of the linguistic variable Collaboration is determined for defining the type of collaboration and built the membership function for one of the components of the variable.

**Introduction**

In modern higher educational institutions (HEIs) one of the main problems is the organization of the management system for the activities of scientific and pedagogical staff (SPS). One of the tools to manage activities in the system of the university is the rating system. It is a rating system for the university as a whole, as well as for its large divisions (institutes/faculties), as well as departments and individual scientific and pedagogical workers. Among the main groups of activities in leading Russian universities are: educational activities, research work, international activities. As part of research activity, the publication activity of scientists is singled out, which is evaluated on the basis of indicators of scientometric systems of international and national level. The composition of such indicators is improved from year to year for an effective evaluation of the SPS work. It was demonstrated in works [1-2] that since the implementation of "Project 5-100" a positive significant effect has been observed. Universities that have received financial support, showing significant steady growth of publications. An analysis of the works [3-5] showed the urgency of the problem of estimating the SPS rating, and the review [6-8] led to the conclusion about the description of collaborations.

In the context of the article, a scientific publication collaboration is understood as the interaction of two or more scientific and pedagogical workers which are authors of scientific publications. Collaborations reflect both the simple "scientific supervisor-student" relationship and the more complex ones that link the interaction of departments, institutions and universities. Collaborations are depicted as a graph showing a large number of links between two or more nodes (authors). There are several types of collaborations: true, false and isolated. True collaborations show an objective picture of the development of internal and external relations of the SPS False collaborations are "artificially created". Artificial collaborations arise when a person is included as a co-author or a group of people who do not take part in the research. Isolated collaborations are a collective of scientific and pedagogical workers, united in a group to study a particular subject area, and do not update the composition. The aim of the research is to develop information structures and methods for establishing the type of publication collaborations.
Theory of Graphs for Determining the Type of Collaboration

In graph theory offers a basic definition: the graph \( G \) is a pair \( G = \{V, E\} \), where \( V \) – set of nodes, \( E \) – set of edges, connecting the nodes of the graph. If the edge \( e \) joins two nodes \( v_1 \) and \( v_2 \), then they are denoted by a pair \((v_1, v_2)\). If as the set \( V \) we select the set of authors of scientific publications and, as a connection between authors, the availability of general publications, we obtain a graph according to the definition given above. It can also be argued that the publication collaboration can be represented in the form of a graph \( G \), which is a digraph, a multigraph, and a pseudograph [9]. A fragment of the graph is presented to represent the publication activity (refer with Figure. 1). For this graph additional characteristics are introduced (refer with Table 1).

![Figure 1. An example of a scientific publication collaboration: a fragment of a graph for a closed group of three authors.](image)

### Table 1. Example of describing the characteristics of the edges of the graph in Figure 2.

<table>
<thead>
<tr>
<th>#</th>
<th>Edge</th>
<th>Multiplicity</th>
<th>#</th>
<th>Edge</th>
<th>Multiplicity</th>
<th>#</th>
<th>Edge</th>
<th>Multiplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>((v_1,\overline{v_1}))</td>
<td>(k_{11})</td>
<td>4</td>
<td>((\overline{v_2},v_2))</td>
<td>(k_{22})</td>
<td>7</td>
<td>((\overline{v_3},v_3))</td>
<td>(k_{33})</td>
</tr>
<tr>
<td>2</td>
<td>((v_1,\overline{v_2}))</td>
<td>(k_{12})</td>
<td>5</td>
<td>((\overline{v_2},v_1))</td>
<td>(k_{21})</td>
<td>8</td>
<td>((\overline{v_3},v_1))</td>
<td>(k_{31})</td>
</tr>
<tr>
<td>3</td>
<td>((v_1,\overline{v_3}))</td>
<td>(k_{13})</td>
<td>6</td>
<td>((\overline{v_3},v_2))</td>
<td>(k_{32})</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When organizing the storage of the graph in the form of a vertex adjacency matrix, a square asymmetric matrix with nonzero elements on the main diagonal will be obtained.

**Adjacency Matrix for Identifying False Collaborations**

The most difficult task is to determine the false collaborations, in this connection, the signs for its definition are singled out [10]:

– a large number of GRNTI codes (State rubricator of scientific and technical information) corresponding to the fields of research of one author;
– presence of more than five co-authors in articles;
– the value of the impact factor of the journal where the article is published, less than the value 0,1;
– absence or a small number of works in which the selected author is listed in the first place in the list of co-authors;
– absence or a small number of works performed in a single authorization.

Based on the signs of false collaborations, it can be said that there is a need to evaluate the contribution of each of the authors of the collaboration. An example of an array structure for the collaboration of two scientific and pedagogical workers (refer with Eq. 1).

\[
N = \{m_{ij}(n, G, M, IF, First, United)\}.
\]

(1)

where \( N \) – array of links between two authors; \( m_{ij} \) – array element \( N \); \( i, j \) – number of authors; \( n \) – article number; \( G \) – number of GRNTI codes; \( M \) – the number of co-authors; \( IF \) – the impact factor of the journal in which the articles are published; \( First \) – number of works in which the author is listed in
the first place in the list of co-authors; United – the number of works prepared by the author in a single author.

After the formation of the array N, the adjacency matrix of the graph nodes is constructed, the values of which will be the elements $m_{ij}$ of the array N. A general example of such a matrix is shown in Table 2.

Table 2. General example of the adjacency matrix.

<table>
<thead>
<tr>
<th>a1</th>
<th>a2</th>
<th>...</th>
<th>an</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>m11</td>
<td>m12</td>
<td>...</td>
</tr>
<tr>
<td>a2</td>
<td>m21</td>
<td>m22</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>mij</td>
<td>...</td>
</tr>
<tr>
<td>an</td>
<td>mn1</td>
<td>mn2</td>
<td>...</td>
</tr>
</tbody>
</table>

Each year there will be a new adjacency matrix. Thus, the complete structure of the initial data forms a hypercube. The structure of the initial data is given for their accumulation over several years (refer with Figure.2), where a1, a2, ..., an – author1, author2, ..., authorn, respectively.

Figure 2. The structure of the initial data for the accumulation of the values of the SPS rating system.

**Application of the Theory of Fuzzy Sets for Decision Making**

In the formulation of the signs, indefinite qualitative levels are used: a large number of GRNTI codes of articles by one author, absence or a small number of works in which the selected author is listed first in the list of co-authors and the absence or a small number of works performed in a single author. The presence of such signs led to the need to use the tools of the theory of fuzzy sets. In order to
determine whether the collaboration is false, a complexity-structured linguistic variable "Collaboration" and terms (refer with Figure. 2) are introduced on the basis of the above characteristics.

<table>
<thead>
<tr>
<th>Collaboration</th>
<th>True</th>
<th>Close to True</th>
<th>Close to False</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>A11</td>
<td>A12</td>
<td>A13</td>
<td>A14</td>
</tr>
<tr>
<td>M</td>
<td>A21</td>
<td>A22</td>
<td>A23</td>
<td>A24</td>
</tr>
<tr>
<td>IF</td>
<td>A31</td>
<td>A32</td>
<td>A33</td>
<td>A34</td>
</tr>
<tr>
<td>First</td>
<td>A41</td>
<td>A42</td>
<td>A43</td>
<td>A44</td>
</tr>
<tr>
<td>United</td>
<td>A51</td>
<td>A52</td>
<td>A53</td>
<td>A54</td>
</tr>
</tbody>
</table>

Figure 3. The structure of the linguistic variable "Collaboration" for determining the type of collaboration.

Figure 3 represents the following notations: G – number of GRNTI codes, M – number of article co-authors, IF – the value of the journal impact factor, First – Number of works in which the selected author is listed in the first place in the list of co-authors, United – number of articles prepared by a single author, A11-A54 – some numeric values, with what, A11, A41, A51 – exact numeric values, and the rest are numerical ranges.

Figure 2 shows the membership function of the component M.

Figure 4. The form of the membership function for a component of a linguistic variable – M.

Figure 4 represents the following notations: μ_M – value of the membership function for the component M; Mb – basic numerical variable that corresponds to a linguistic M; T – true collaboration; CT – the collaboration is close to true; CF – the collaboration is close to false; F – false collaboration.
Conclusions

1) For the proper functioning of the rating system for evaluating the activity of scientific and pedagogical staff, it is required not only to create and describe the structure by which collaborations between employees are described, but also to learn how to apply it in real conditions. The main types of collaborations were described. The signs that characterize the false collaborations of the authors of scientific articles are shown. To construct the collaborations, an array \( N \) was formed, which includes elements with a complex structure \( m_{ij} \). On the basis of the formation of the array, an adjacency matrix was constructed, the elements of which are \( m_{ij} \), considered with respect to the authors. The totality of such tables for each period of time was called the "hypercube".

2) A selection of sets for describing publication collaborations in the form of a graph was made. A lot of authors of scientific publications are defined as the vertices of the graph, and the availability of publications as edges. It is determined that the graph of the scientific publication collaboration is a pseudomultigraph. Taking into account the form of the obtained graph, the adjacency matrix is asymmetric with nonzero elements on the main diagonal.

3) An analysis of the structure of information characterizing the publications of the SPS collaborations made it possible to apply the theory of fuzzy sets. The structure of the linguistic variable Collaboration is formed and the basic numerical variable of one of the components of the variable is determined - the number of coauthors (M) and its membership function is constructed.

References