Maturation of the eigenvector of the centrality of the hypergraph of co-authors

Nasyrov Ruslan
1st year student,
Faculty of Applied Mathematics and Computer Science
Moscow Institute of Physics and Technology
Russian Federation, Dolgoprudny
E-mail: nasyrov.rr@phystech.edu
Musatov Daniil
Candidate of Physical and Mathematical Sciences,
Department of Discrete Mathematics, MIPT
Russian Federation, Dolgoprudny

May 19, 2021

1 Finding the eigenvector of the centrality of a hypergraph and its projections

We will consider hypergraphs whose edges are scientific articles, and whose vertices are authors who participated in writing articles in the field of Computer Science.

Research plan:

1. Uploading data from the source.
2. Creating a hypergraph.
3. Calculating hypergraph projections.
4. Calculating the centrality vectors of these projections.

1.1 Uploading data

Uploading data from the source datasets[6] was produced using the library json[7].

1.2 Creating a hypergraph

A hypergraph is created from a list of edges using the networkx library[5]. Its characteristics:

1. Vertexes: 363043
2. Hyper-edges: 435135
3. Maximum hyper-edge: 427 vertices
4. Average edge size: 4
5. Connectivity components: 22912
6. Maximum component size: 298870
7. Average component size: 16.

2 Analysis of the centrality vector maturation

2.1 Hypothesis

Let's assume that there is a maturation of the centrality vector of the hypergraph with an increase in the projection number and there is a correlation between the limit vector and the vector of different projections, and check this.

Analysis Plan:

1. Calculation of the correlation between the limiting centrality vector and the centrality vectors of the graph projections.
2. Plotting the correlation graph.
3. Conclusion.

Consider the correlation of the top-n vertices of the limit vector and the projection vector at

1. n = 100
2. n = 1000
3. n = 363043 (all vertexes)

To do this:

1. Let's number the vertices in the limit vector in descending order of their centrality.
2. Take the first n vertices, so we get the limit vector.
3. For each projection, we try to form a vector in the following experimental way: if the vertex v in the k-th projection occurs among the top-n vertices of both the limit vector and the vector of its projection, then in the resulting vector it is assigned the previously calculated centrality, otherwise - 0.

2.2 Projection 100 (all edges with more than 100 vertices were removed)

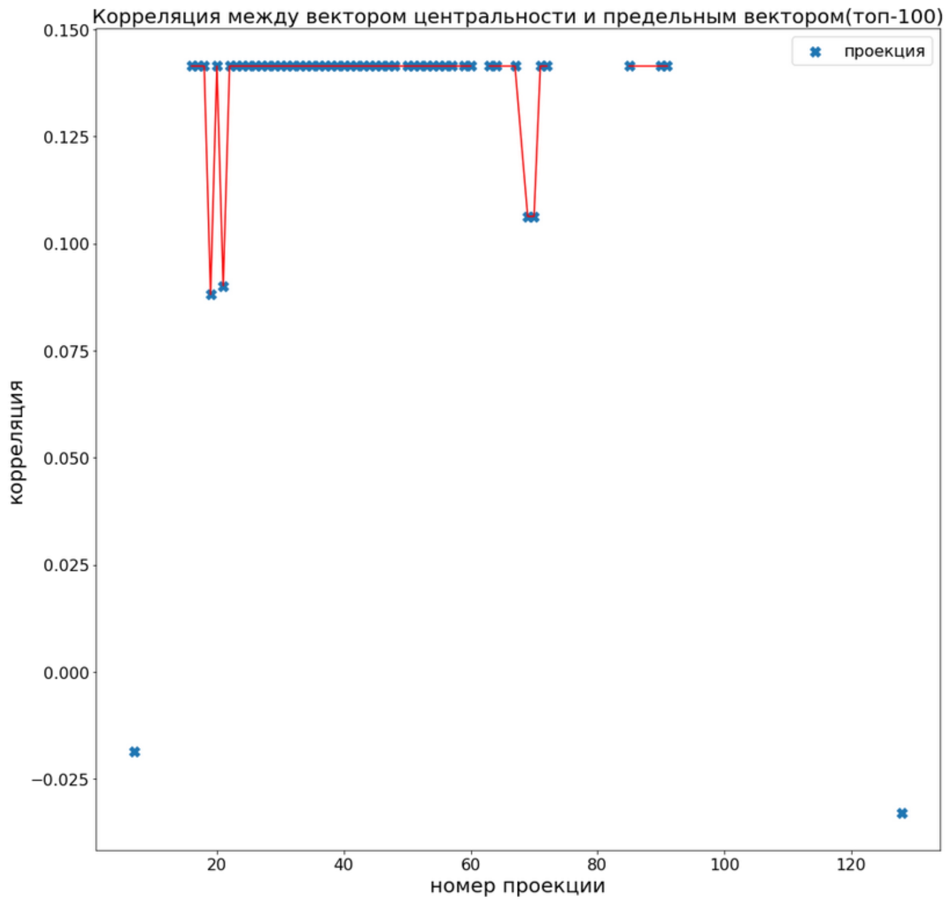

Figure 1: Plot of correlation between vectors of centrality of projections of top-100 vertexes. On the horizontal axis - the projection number, on the vertical-the correlation value.

We see that the projection vectors are weakly correlated with the limit vector. In addition, on the early projections (numbered 20-90), the correlation with the limit vector is almost the same(about 0.140). The absence of points for some projection means that the top 100 vertices of this projection do not intersect with the top 100 of the limit vector. Accordingly, in this case, it makes no sense to talk about correlation.

2.3 Projection 1000

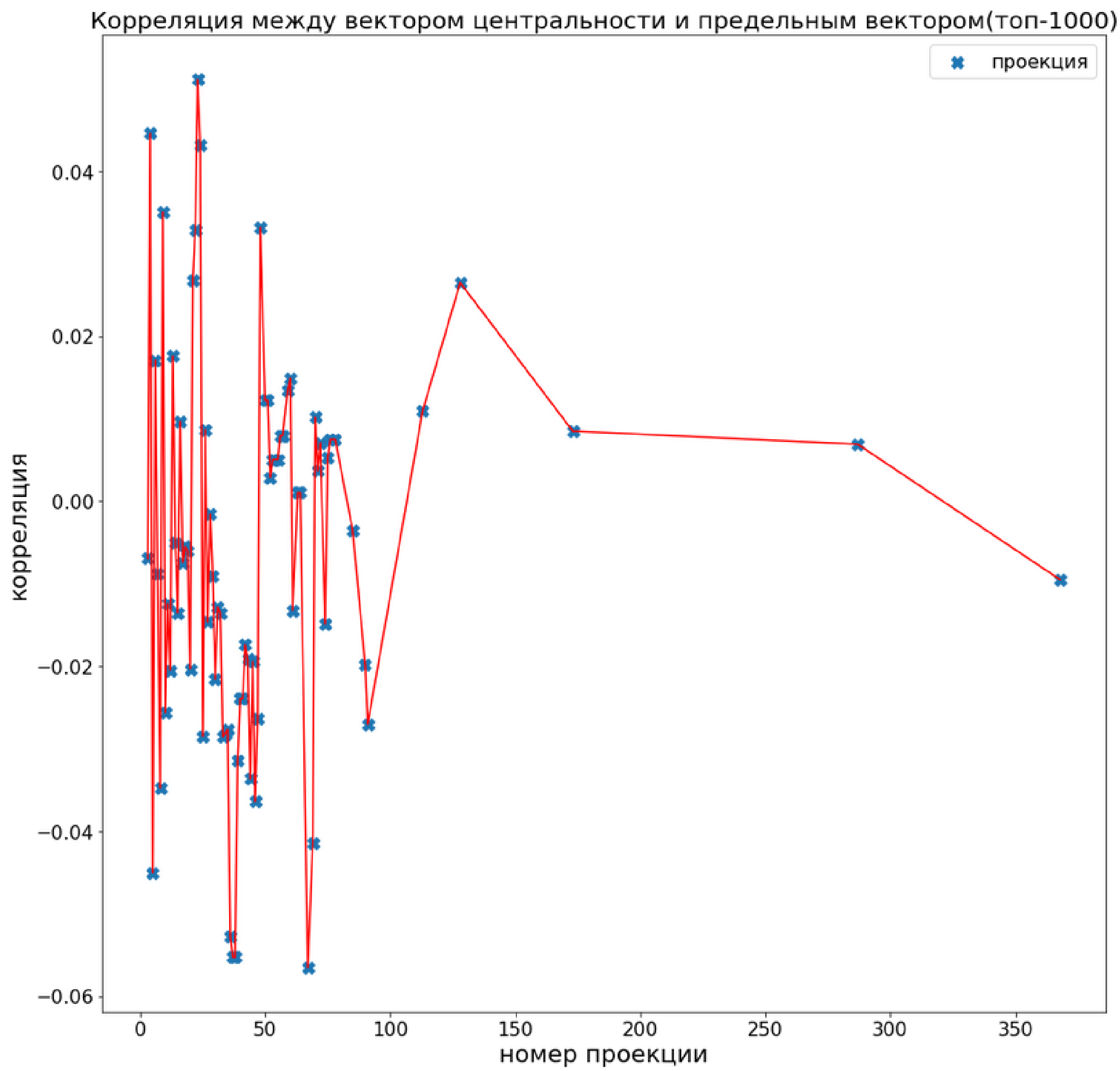

Figure 2: Plot of correlation between vectors of centrality of projections of top-1000 vertexes. On the horizontal axis - the projection number, on the vertical-the correlation value.

We see: there is no correlation.

2.4 All edges are involved

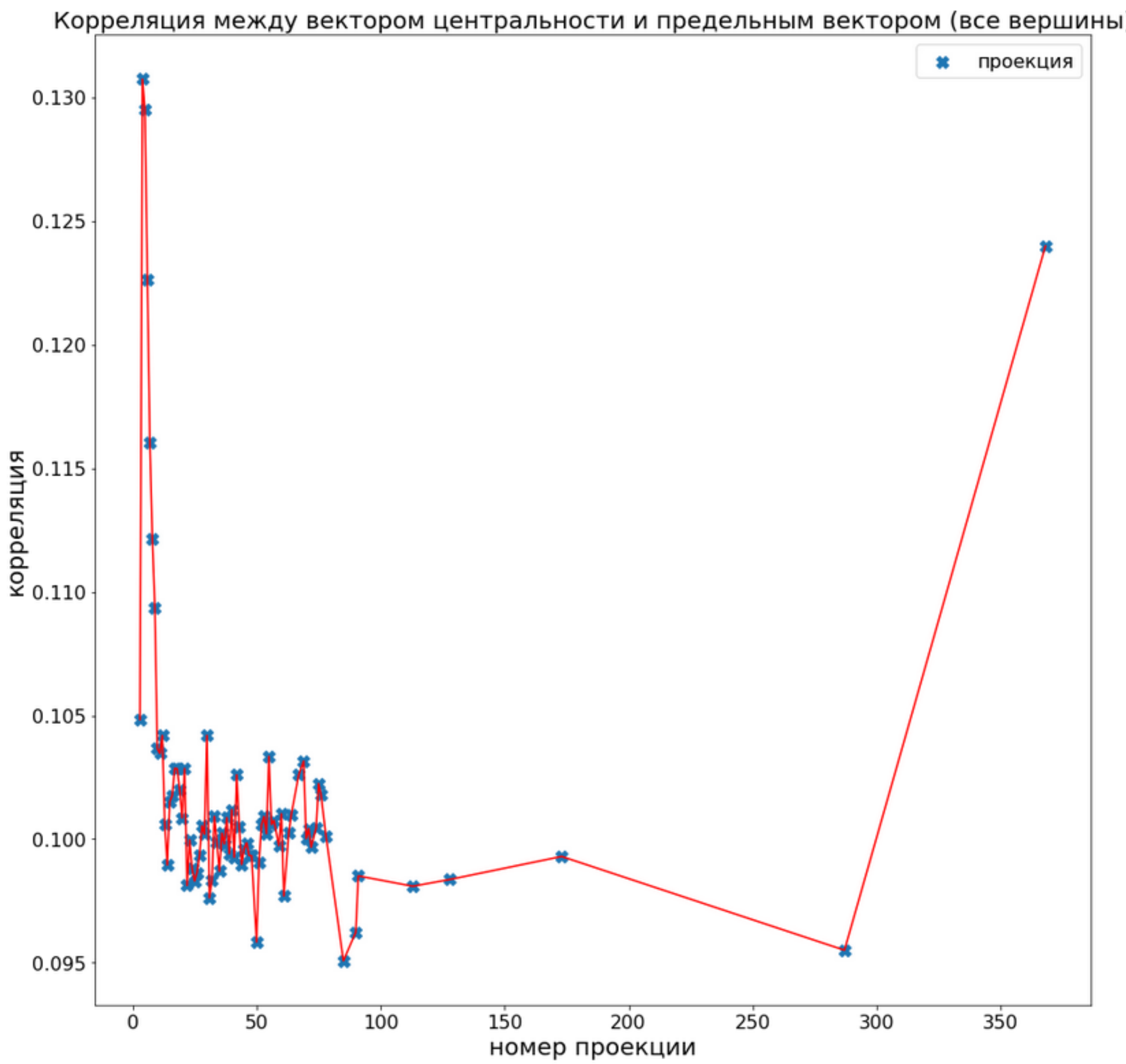

Figure 3: Plot of correlation between vectors of centrality of projections of all vertexes.

We see that there is almost no correlation.

2.5 Conclusion

The centrality vector of the hypergraph changes strongly when large hyperrebs are added, and weakly when small hyperrebs are added(in the early projections). At the top 1000 and for all vertices, the correlation jumps strongly on the initial projections. No maturation of the hypergraph centrality vector is observed for different projections.

3 Result

As part of the research, a program was written that allows analyzing the maturation of the centrality vector for an arbitrary hypergraph. With its help, the maturation of the centrality vector for the hypergraph of articles in the direction of Computer Science is analyzed. Based on the results obtained, we can conclude that the maturation of the hypergraph centrality vector is not observed and our hypothesis has not been confirmed.

References

[1] A.R. Benson, "Three hypergraph eigenvector centralities", SIAM Journal on Mathematics of Data Science 1(2) (2019), 293312.

[2] B. Lemmens and R. Nussbaum, "Nonlinear Perron-Frobenius Theory", CambridgeTracts in Mathemtics 189 (2012). Cambridge University Press.

[3] Maurice G. Kendall, "The treatment of ties in ranking problems", Biometrika Vol. 33, No. 3, pp. 239-251. 1945.

[4] matplotlib.org

[5] networkx.org

[6] github.com/mattbierbaum/arxiv-public-datasets

[7] json.org

[8] scipy.org

[9] medium.com/@jeffreystewart/semantic-data-master