# DYNAMICS OF THE MATH-NET.RU CITATION GRAPH

Andrey Pechnikov[1], and Anthony Nwohiri[2]

[1] Institute of Applied Mathematical Research, Karelian Research Centre RAS, Petrozavodsk, Russia
[2] University of Lagos, Lagos, Nigeria

E-mail: [1] pechnikov@krc.karelia.ru
[2] anwohiri@unilag.edu.ng

A citation graph is a directed graph whose vertices denote authors of scientific publications, while the arc $(i, j)$, connecting two vertices, exists if and only if author $i$ has at least one reference in his papers (possibly in co-authorship) citing the paper of author $j$ (also possibly in co-authorship). The weight of an arc is equal to the number of publications by author $j$ cited in all publications by author $i$. It follows from the definition that a citation graph has no isolated vertices. It may contain loops, i.e., arcs of the form $(i, i)$, corresponding to self-citation.

Let $G_t(V, E, W)$ be a citation graph, where $V, E$ and $W$ are sets of vertices, arcs and weights of arcs, respectively, and $t$ is the time at which the graph is built. The time of the earliest publication can be assumed to be 0.

We set a time axis with fixed moments $\{t_0, t_1, t_2, ...t_n\}$, where $t_0$ is the publication date of the oldest paper, and tn is the publication date of the newest paper. In this case, graph sequence $G_{t1}, G_{t2}, \ldots G_{tn}$ can be constructed. Here graph $G_{t1}$ is a citation graph, where the set of vertices are articles published during the time interval $[t_0, t_1]$. The set of arcs and their weights is constructed on the basis of the citations that appeared among the published papers during this period. We transit from graph $G_{ti}$ to graph $G_{t(i+1)}$ by adding new vertices and arcs that appeared during the period $(i, i + 1]$.

For constructions of this kind, of great interest are the problems of determining heredity-preserving transition operations, where certain general properties and characteristics are preserved at transition from the previous to the next graph [1].

Our study was based on data obtained from Math-Net.Ru (http://www.mathnet.ru), a well-known web resource containing a rich collection of full-text archives of leading Russian mathematical journals and author information [2].

For the purposes of the study, we selected data that met a number of conditions. Based on previously obtained results [3] on the aging structure of citations and their half-life of 8 years, we chose a study period of 2011 to 2020. The general characteristics of the original data were: 16,925 authors, 28,672 papers, and 108 mathematical journals.

For $t_0 = 2011$, a sequence of citation graphs $G_{2012}, G_{2013}, \dots G_{2020}$ was constructed and their main properties and characteristics were investigated. Since loops do not affect these properties and characteristics, citation graphs henceforth in this paper shall mean loopless graphs.

The change in the number of vertices, arcs (without multiplicity) and multiple arcs (weights) is linear from $|V| = 1,584, |E| = 2,673, |W| = 3893$ in 2021 to $|V| = 16,454, |E| = 61,553, |W| = 143,324$ in 2020.

Obviously, graph connectivity is one of the interesting properties of a graph. In our case, graphs are not simply connected, much less – strongly connected, they contain a large number of strongly connected components (SCCs). The change in the total number of SCCs is logarithmic. In 2012, there were 188 SCCs, stabilizing by 2019 and 2020 at 1,030 and 1,038, respectively.

At the same time, the maximal strongly connected component (MSCC) behaves as follows. For the first 4 years, it slowly and almost linearly grows from 10 to 129 vertices, and then also almost linearly, but sharply grows to 4,112 vertices in 2020, growing by about 800 vertices per year.

The proportion of the MSCC vertices grows from 0.006 in 2012 to 0.25 in 2020.

Of particular interest is the behavior of two MSCC characteristics: diameter and modularity. For the first 6 years, the diameter grows as a second-degree polynomial, starting from 2 and reaching 37, the maximum value. This is followed by a gradual decrease to 23 in 2020.

The modularity coefficient of MSCC [4] grows rapidly from 0.045 to 0.436 for the first 4 years, then it makes a sharp jump, exceeding 0.8, a value that has been almost stable for the past 5 years.

The number of "dangling" vertices (vertices with only incoming or only outgoing links) changed linearly from 1,092 in 2012 to 8,972 in 2020. Their proportion in the graph in the same period slowly decreased from 0.689 to 0.54.

So, with a linear increase in the number of vertices and arcs in the graphs, we have an initial period (equal to four years) of "slow accumulation" of the MSCC, after which it grows linearly, but rather quickly. At the same time, the proportion of MSCC vertices with respect to the cardinality of the set of all vertices increases.

However, starting from the 5th year, the MSCC modularity coefficient becomes stable, while the MSCC diameter, having reached its maximum, tends to decrease. One explanation could be the following:

1. a new paper in the current year is very likely to reference articles included in the MSCC, given the accumulated authority of the MSCC;

2. there will be no links to a new paper for some time, and it will be "dangling";

3. at the same time, for papers published in previous years, the probability of references to them increases, which leads to a decrease in the MSCC diameter.

The stability of the modularity coefficient follows from the diversity in the mathematical fields of study. Vertices belonging to the same domain are more closely related to each other than vertices from different domains. Therefore, the MSCC grows at the expense of single references linking various areas of mathematics that are much more strongly connected within themselves.

# References


[1] Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D.U. Complex networks: Structure and function. Phys Rep. 2006. Vol. 424. 175–308.

[2] Chebukov D., Izaak A., Misyurina O., Pupyrev Yu., Zhizhchenko A. Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today. Lecture Notes in Comput. Sci. 2013. Vol. 7961. 344–348.

[3] Pechnikov A., Chebukov D., Nwohiri A. On Some Journal Citation Properties: Math-Net.Ru as a Case. CEUR Workshop Proceedings. Proceedings of the 23nd Conference on Scientific Services Internet. 2022. Vol. 3066. 80-89.

[4] Newman M.E., Girvan M. Finding and evaluating community structure in networks. Physical Review E. 2004. Vol. 69(2). P 026113.