Scientific Visualization, 2023, volume 15, number 1, pages 17 - 28, DOI: 10.26583/sv.15.1.02

Investigation of the Properties of First Nearest Neighbors' Graphs

A.A. Kislitsyn A.A.[1], Yu.N. Orlov[2], M.V. Goguev[3]

Keldysh Institute of Applied Mathematics of RAS, Miusskaya sq., 4,
Moscow, 125047, Russia

[1] ORCID: 0000-0003-2388-0496, alexey.kislitsyn@gmail.com
[2] ORCID: 0000-0002-1356-5137, ov3159f@yandex.ru
[3] ORCID: 0000-0001-5365-070X, goguev.mv14@physics.msu.ru

Abstract

In this study we present a benchmark of statistical distributions of the first nearest neighbors in random graphs. We consider distribution of such graphs by the number of disconnected fragments, fragments by the number of involved nodes, and nodes by their degrees. The statements about the asymptotic properties of these distributions for graphs of large dimension are proved. The problem under investigation is to estimate the probability of realization of a certain structure of the first nearest neighbors graph depending on the distribution function of distances between the elements of the studied set. It is shown that, up to isomorphism, the graph of the first nearest neighbors does not depend on the distance distribution. This fact makes it possible to conduct numerical experiments on the construction of basic statistics based on a uniform distribution of distances and obtain tabulated data as a result of numerical modeling. We also discuss the approximation of the distribution of graph vertices by degrees, which allows us to estimate the proportion of randomness for a particular structure resulting from clustering elements of a certain set by the nearest neighbor method. The asymptotic analysis of the fragment distribution is discussed.

Keywords: Nearest neighbors graph, graph statistical structure, distribution of nodes degree, asymptotical distribution.

1. Introduction

In the present paper we study statistics of the so-called first nearest neighbors graphs (next FNNG). The main goal of this investigation is to analyze the adequacy of the nearest neighbor heuristics as a machine learning model. Usually, such a problem arises in many tasks of classification, where we need to recognize some attributes of the system, which presumably determine the state of this system. Machine learning methods for FNNG models are described in sufficient detail, for example, in works [1, 2]. Practical heuristics using the first (as well as second, third, etc.) nearest neighbors for clustering problems are described in [3]. In the work of [4], the dependence of the number of nearest neighbors on the dimension of the embedding space is studied. Various algorithms for searching nearest neighbors are described in [5]. In the work [6] some approximation methods for FNNG searching were discussed. Thus, work in this area continues to develop actively, and a number of tasks remain unresolved, despite the long period of study.

The essence of the problem we are working on is that for reliable machine learning it is required to have sufficient number of samples from the same attribute distribution. However, the structures of the nearest neighbors corresponding to various samples from such a distribution may differ significantly. With regard to FNNG, the question arises as to how the structure of connections between elements of a set observed in practice is characteristic of sets of the type being studied. To carry out such an analysis, it is necessary to formalize the concept of sample statistics for FNNG. This problem is not so simple as it may seem. On the one hand, if it is necessary to find the probability that a certain family of graphs has a given property, then the problem is solved methodically simply: graphs of this family are generated and their proportion for which the studied property is fulfilled is determined. After this determination, the convergence can be analyzed in the sense of the central limit theorem. We emphasize that both the family and the property are defined in terms of graph parameters: it can be the number of edges, the average density of connections, the number of disconnected subgraphs, etc. But if we are not interested in the graph itself, but in its interpretation as a structure of a physical system, then the situation becomes much more complicated. For example, let's take a certain number of fiction texts belonging to certain authors and build a graph of nearest neighbor connections for them. The question arises: how should the structure of the resulting graph be treated? Let us suppose, that we have three disconnected subgraphs with node amount equals N_1, N_2, N_3. Is this structure characteristic of this particular group of authors and texts, or is it in some sense typical of writers in general? If a graph with five disconnected fragments is obtained for other texts by the same authors, should it be considered as a significantly different system or as a random variation of the first one?

Since in practice, as a rule, one specific set of objects for which data were collected is analyzed, the question of the variability of established empirical patterns becomes of practical importance. Different samples of the same type of objects will generate, generally speaking, different connection graphs. In this sense, the relationship graph is random, since the selection of objects into the analyzed group is random. The theory of random graphs is developed in great detail in the direction of "probability theory". Starting from the classical work [7], random graphs have been studied in the context of solving many applied problems: see, for example, [8 - 11].

The asymptotic properties of random graphs were also analyzed within the framework of specific probabilistic models in relation to various ontologies, social and transport networks and similar objects (see [12 - 14]) However, in the direction of "mathematical statistics" there are significantly fewer constructive models and results in graph theory. This is due to the fact that "graph sampling" does not take place by itself, because the graph simply visualizes some property of the studied group of objects. So, the concept of sampling is addressed specifically to a group of objects that should be homogeneous in the studied property in the statistical sense. But since the studied property is revealed just as a result of analyzing the graph structure, it is impossible to form groups of objects with the desired property in advance. Then, having considered the distribution of parameters that are considered key for a particular sample in a particular model of the system, it is possible to multiply samples with the found empirical distribution of parameters, and then study the statistics of graphs corresponding (according to the assumption of the model) to various states of the system under study. Since the sample of parameters is finite, the corresponding sample distribution function fluctuates from sample to sample. Then it is necessary to study the variability of the graph family structure when changing the distribution of system parameters.

The main problem is that the analysis of one particular graph does not allow us to estimate the probability of its realization as an element from a certain set of random graphs, since the desired set is not formalized. A rare exception is the nearest neighbor graph, which is based on a matrix of pairwise distances between points. In this case, the distance distribution function is known, which makes it possible to find out which graphs correspond to it.

We formulate the main tasks of statistical analysis of FNNG, which are of interest for the classification and identification of elements from some finite set.

First, it is necessary to determine the realization probability of a certain graph structure with N nodes in the form of a set Q of disconnected fragments. In the case of nearest neighbors, the number Q can be varied from 1 (the graph is connected) to $[N/2]$ (the graph consists of pairs of mutually closest nodes or, according to the parity of N, consists of pairs and one triple).

Secondly, since the distribution of a graph by degrees of nodes depends on the number of these nodes themselves, it is necessary to construct a distribution of a connected fragment V by the number of nodes depending on the total amount of nodes N in the graph and the number of fragments Q.

Thirdly, it is interesting to study distribution of nodes in the fragment by their node degrees P with known values N, Q and V up to isomorphism.

Thus, for a given total number of nodes N, we will be interested in the distribution $G_N(Q)$ of graphs by the number Q of disconnected fragments, the distribution $U_N(Q,V)$ of these fragments by the number of nodes V, and the distribution $H_N(Q,V,P)$ of a given graph by nodes degree P. It is quite natural that independent samples of distances from the same distribution will generate, generally speaking, non-isomorphic graphs. Then we can find out which structures are most likely. And as a result, it is possible to separate by studying the constructed distributions whether we are dealing in each case with a typical situation or with a rare deviation, which may be caused by a special reason that was not considered when forming the similarity model of objects.

The asymptotic behavior of sample distributions for high-dimensional graphs is also of our interest. So we suggest a new method of FNNG generation for high dimension metric space. Instead of generation of random coordinates of multidimensional vector we generate a random symmetric matrix of distances between objects in our space. Thus we can construct a lot of samples of such matrices and investigate the distribution function of corresponding graph structure. The matrix order is equal to the number of points N. If N is sufficiently large, only the graph analysis can visualize a typical mutual structure of random objects. So the statistical properties of the distributions, mentioned above, enable us to get an idea of typical FNNG structure and to estimate the corresponding probability of realization of such a graph.

2. The main properties of Nearest Neighbor Graphs

We will assume that for each point of the set under study there is a unique nearest neighbor. In our study we follow to a standard definition of FNNG as an oriented graph, connecting two points A and B in a metric space if and only if $B = \arg\min_X \rho(A, X)$, where ρ is a distance function between two given points.

If the number of elements in the set is relatively small, then the problem of calculation of graph distribution according to the number of disconnected fragments and distribution of nodes according to the number of incoming edges can be solved by direct enumeration. For a large graphs, such enumeration process may have a significant computational complexity, and it is not obvious in advance how accurately the problem of collecting statistics should be solved.

If we proceed from the fact that the distances between the elements of the set are known, then we can assume that the physical system for which the FNNG structure is analyzed is determined by the distribution function of these distances. Then it would be possible to study FNNG families corresponding to given distributions of distances between points of the set. The question arises: how much does the probability of realizing one or another FNNG structure depend on the type $F(l)$ distribution of distances between points?

Proposition 1. The probability of realizing the FNNG structure does not depend on the distribution of distances between points.

Proof. As it is known from the general course of probability theory, if F is a distribution function of random variable ξ, then a random variable $\eta = F(\xi)$ has a uniform distribution on $[0;1]$. Based on this statement, algorithms for generating a sample $\{x_k\}$ with a given continuous density of the distribution function $f(x)$ are constructed, for which $F(x)$ is sample distribution function. To do this we can generate uniform distributed set $\{y_k\}$ on $[0;1]$, after that from the equation $y_k = F(x_k)$ the set of solutions $\{x_k\}$ can be numerically found. This series is a sample subset from $F(x)$ by construction. Let $\{y_k\}\big|_1^n$ be a uniformly distributed set of distances, where $n = N(N-1)/2$, and N is an amount of numbers in the set. This set corresponds to a structure of certain FNNG. Due to the monotony of the function $y_k = F(x_k)$ the inverse function is also monotonic with the same direction of monotony as the forward mapping. Hence, the ordering of the quantities in the samples $\{y_k\}\big|_1^n$ and $\{x_k\}\big|_1^n$ is the same. Therefore, the graphs of the corresponding nearest neighbors also coincide. So, the probabilities of FNNG structure realization do not depend on the distribution of distances between nodes. **Proposition 1 is proved.**

From this it follows that in numerical analysis it is enough to collect statistics of FNNG for a uniform distribution of distances, without discussing yet whether it is possible to implement such realization for points of Euclidean space at all (on a straight line, for example, it is impossible). This will allow tabulating the probabilities of the realization of a particular structure.

Suppose that for any N, all available structure realizations of each nearest neighbor graph are listed up to isomorphism. A structure is a specific collection of fragments with a given number of nodes.

Let the amount of such structures be equal to $M(N)$, the probability of realization of each of them is known (for example, from a computational experiment) and is equal to $A_m(N),\ m = 1,2,\dots,M$. The value of $M(N)$ is a number of ways, which the connected set of N points can be divided into Q subsets, $Q = 1, 2,\dots,[N/2]$, with a number of nodes $V_1, V_2,\dots, V_Q$ and with the condition $\sum_{k=1}^{Q} V_k = N$. Ideally, this is the solution to the problem of a particular graph realization. The distribution of degrees of nodes for each structure is known by construction, we denote it $h_m(N,P),\ 0 \le P \le N-1$. This distribution is considered precisely within the particular structure, so that $\forall m, N:\ \sum_{P=0}^{N-1} h_m(N,P) = 1$. Then the distribution of a random graph by nodes degree is $F(N,P) = \sum_{m=1}^{M} A_m(N) h_m(N,P)$.

The problem is that the determination of $A_m(N)$ is associated with a complete search and has exponential complexity, which becomes for large N an insurmountable obstacle. It follows from the above that the number of possible FNNG structures up to isomorphism is related to the number $R(N)$, ways of splitting a given natural number N into a sum of natural numbers. An asymptotic estimate of the number of partitions is given by the Hardy – Ramanujan formula [15], which approximately has the form:

$$R(N) \approx \frac{1}{4\sqrt{3}N} \exp\left(2\pi\sqrt{N/6}\right).$$

The number of disconnected fragments is determined by the following sentence.

Proposition 2. Amount $M(N)$ of different fragments, in the form of which the graph of the first nearest neighbors of N points can be realized, asymptotically equal to the principal part of the derivative of the number of partitions $R(N)$ by the number of nodes:

$$M(N) \approx \frac{\pi R(N)}{\sqrt{6N}}.$$

Proof. Number of structures $M(N)$ is an amount $R(N)$ of the representation of the number N in form of sum of natural numbers minus the number of impossible graph structures in case of this problem. Isolated points, of which there may be one, two, etc., are impossible. Let's denote this number of invalid structures by $K(N)$. Then $M(N) = R(N) - K(N)$. Note now that the number of different structures is taken into account up to isomorphism. This means that there is $M(N-1)$ options when there is one isolated node, $M(N-2)$ options for two isolated vertices, etc. Hence, the number $K(N)$ is presented as sum $K(N) = \sum_{k=2}^{N-1} M(k)$. Then from here and from equality $M(N) = R(N) - K(N)$ it follows that $R(N) = \sum_{k=2}^{N} M(k)$.

Thus,

$$M(N) = R(N) - R(N-1).$$

From formula (1) for $R(N)$ it follows that $M(N) = R(N)\left(1 - \sqrt{\frac{N}{N-1}} \exp\left(a\sqrt{N-1} - a\sqrt{N}\right)\right)$, where $a = 2\pi/\sqrt{6}$. Let's obtain the main part by N in this expression. We have the following transformations:

$$\sqrt{\frac{N}{N-1}} \exp\left(a\sqrt{N-1} - a\sqrt{N}\right) = \frac{1}{\sqrt{1-1/N}} \exp\left(\frac{-a}{\sqrt{N-1}+\sqrt{N}}\right) =$$

$$= \left(1 + \frac{1}{2N} + o\left(\frac{1}{N}\right)\right) \exp\left(-\frac{a}{2\sqrt{N}} \frac{1}{1-1/(4N)+o(1/N)}\right) = 1 - \frac{a}{2\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right),$$

from which it follows that the main part $M(N)$ is equal to $M(N) \approx \frac{aR(N)}{2\sqrt{N}}$. Since here $a/2 = \pi/\sqrt{6}$, we get the result (2). **Proposition 2 is proved**.

The illustration of this proposition is presented in Fig. 1 (a, b). As it is known [16], the problem of natural numbers decomposition is closely connected with asymptotical semi-circle Wigner distribution. These figures correspond to situation, where the amount [N/2] of nodes is presented as a set of disconnected graphs. One can see from one million numerical experiments for generating FNNG structures with uniform distribution distances for cases $N = 200$ and $N = 1000$, that the distribution of probability of the fragment numbers tends to Wigner distribution.

The exponential growth in the number of ways to divide a connected graph into fragments leads us to the fact that it is necessary to abandon a detailed consideration of the graph structure and move on to its more enlarged characteristics. Let $G_N(Q)$ be a probability of the nearest neighbor graph realization in the form of Q fragments and $U_N(Q,V)$ is a distribution of these fragments by the number of nodes, i.e. it is the probability that in the graph of N nodes, which make up into Q fragments, there is a structure with a given amount of nodes V.

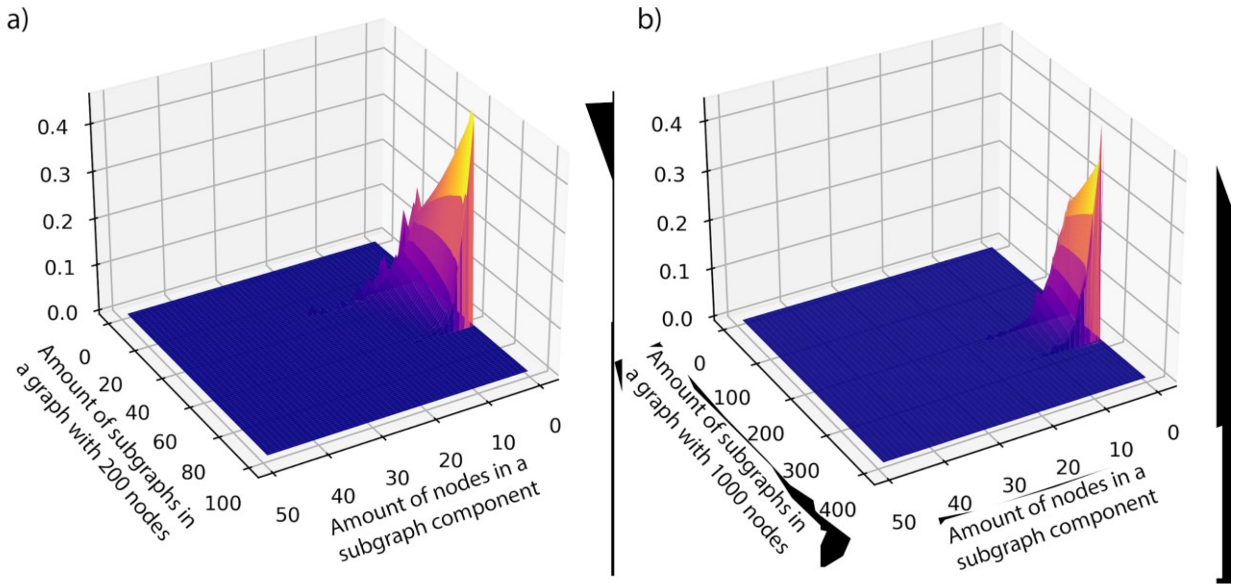

Fig. 1. The distribution of FNNG structures on subgraphs and nodes:
(a-left) N = 200; (b-right) N = 1000

To describe this graph-family we will need the key object – the number of non-isomorphic connected graphs with a given number of nodes. Let 's denote this value $Y(V)$. It can be used to express the number of other non-isomorphic structures. If, for example, the graph consists of two fragments with the number of nodes in them equal to V_1 and V_2, then the number of non-isomorphic graphs in such a family is equal to $Y(V_1)Y(V_2)$. Due to the independence of FNNG from the type of distribution of distances between nodes, the probability of each graph within this family is the same and equals

$$p(V_1,\dots,V_Q)=1/\prod_k Y(V_k),\ \ k=1,2,\dots,Q;\ \ \sum_k V_k = N.$$

We emphasize that the distribution of nodes by degrees depends on the configuration, which, although equally probable within a given partition on fragments with a given number of nodes, is naturally different for different partitions. Therefore, the distribution of nodes by their degree corresponding to a certain configuration of Q fragments (regardless of the specific realization of these fragments) in the form of a certain node distribution, is not uniquely determined. It itself has some distribution $B_N(Q,H,P)$: this is the probability that in this configuration the degree P has the frequency of realization H. In practice, the probability $B_N(Q,H,P)$ is estimated based on the results of experiments during which the realization of Q fragments was obtained. In each experiment the number of $m_j(P)$ nodes, that have a degree P, can be found, so that the corresponding empirical frequency is equal to $H_j(P) = m_j(P) / N$. The set $\{m_j(P)\}$ forms an empirical distribution $B_N(Q,n,P)$ of node number n, with a degree P in a given configuration Q.

If we now consider random realization of FNNG on N nodes, then the distribution of its vertices by degrees has the form

$$\Phi_N(n,P) = \sum_Q G_N(Q) B_N(Q,n,P), \quad \sum_n \Phi_N(n,P) = 1 .$$

Each realization is a selection of pairs $(n_0,0),(n_1,1),\ldots,(n_k,k),\ldots$ with conditions $\sum_k n_k = N,\ \sum_k kn_k = N$: the number of nodes is equal to *N*, the sum of the degrees of the nodes is also equal to *N*.

3. The results of numerical experiments

To build a benchmark of FNNG statistics the set of graphs with the node number from 100 to 1500 with step 100 was modeled. Statistics were collected based on the results of one million graph realization of each size.

Further, based on the results of the simulation, we can answer the question, what is the probability $f_P(n)$ that a random FNNG contains a given number n of nodes with given degree P by incoming edges. For example, the distribution $f_0(n)$ with $P = 0$ is presented in Fig. 2 for $N = 300$.

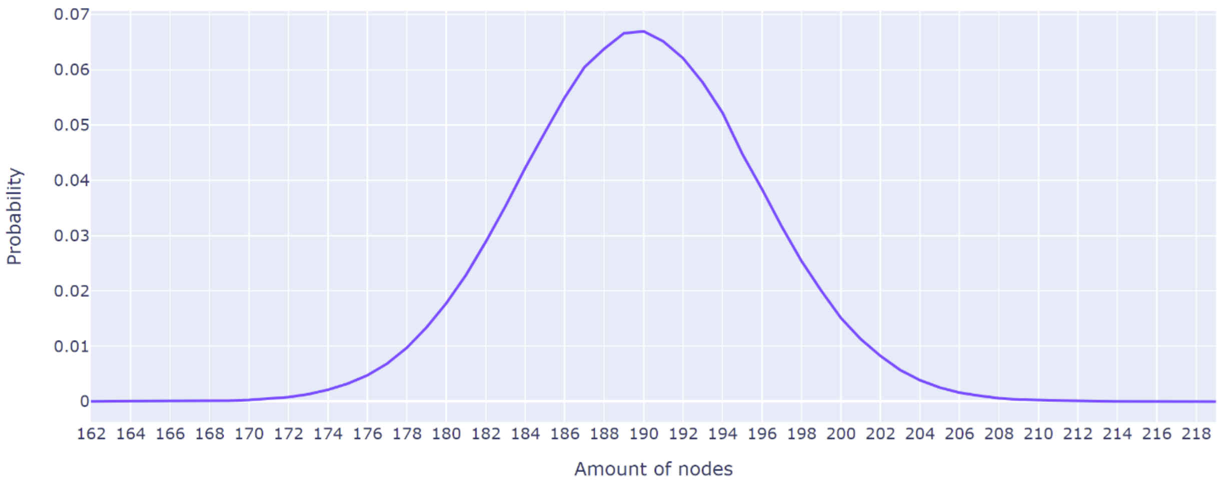

Fig. 2. Distribution of the number of nodes with degree $P = 0$ for *N*= 300

Similarly, it is possible to construct distributions nodes by the first degree, second, etc. If we arrange these distributions on the same plane in the form of a “surface” $F(P,n)$, which is a set of distributions $f_P(n)$, then we get the phase space of possible realizations FNNG. At the same time, not every trajectory is possible $(P,n(P))$, but only one for which the conditions are satisfied $\sum_P n(P) = N,\ \sum_P Pn(P) = N$.

Example of distribution $F(P,n)$ is shown in Fig. 3 for $N = 1000$.

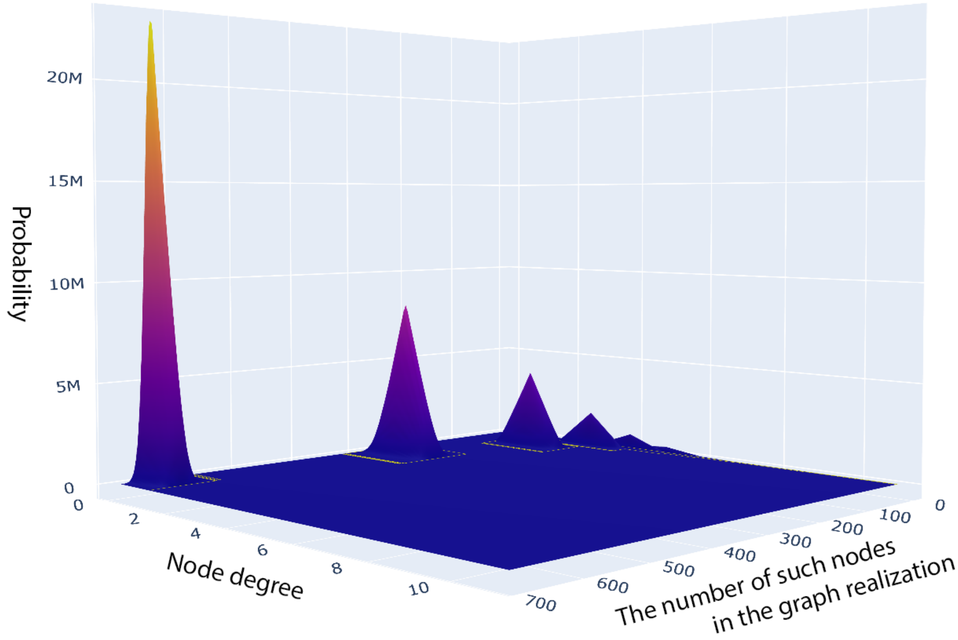

Fig. 3. Distribution of nodes by degrees for N = 1000

Surfaces $F_N(P,n)$ represent a numerically obtained reference base for analysis of FNNG. For example, for $N=100$ the most likely realization of a random FNNG has a probability of approximately $3\cdot 10^{-4}$. On the other hand, analyzed in [17] task of recognizing the authors of literary texts of particular one hundred authors leads to the line "degree of nodes - number of nodes" in the form $(0;34),(1;40;),(2;18),(3;8)$. It follows from the benchmark data that the probability of such a random graph is $3\cdot 10^{-8}$, it is ten thousand times less than the probability of a typical nearest neighbor graph. This indicates that this particular sample is formed, in any case, by objects with a high degree of correlation, and a similar distribution should not be expected with another sample of this type. If the graph was typical, it would not make sense to look for the reasons of such connections. For this task, it is possible to discuss the reasons of the appearance of specific clusters. Note also that the number of disconnected FNNG fragments in the recognition task from [16] turned out to be equal 23, which is close to the most likely $\left(25=\frac{1}{2}(N/2)\big|_{N=100}\right)$ according to the Wigner limit distribution.

An important result of the numerical analysis is that the empirical distributions $f_P(n)$ turned out to be very close to normal. For example, for the density of the distribution of the number of nodes of degree 0 in a graph from N nodes with the determination above 0,999, we get an approximation (normalized by the number of nodes)

$$f_0(n)\approx B_0N\exp\left(-A_{00}\left(n-\mu_0N\right)^2\right),\ B_0=0{,}0667,\ A_{00}=0{,}014,\ \mu_0=0{,}63.$$

For nodes with the first degree, the distribution has the form

$$f_1(n)\approx B_1N\exp\left(-A_{11}\left(n-\mu_1N\right)^2\right),\ B_1=0{,}0587,\ A_{11}=0{,}010,\ \mu_1=0{,}26.$$

Distributions for other degrees of nodes have a similar form. Note now that these distributions are bound by the condition $\sum_k kn_k=N$, which mean that they are correlated. For example, the joint distribution of degrees 0 and 1 has the form (see Fig. 4):

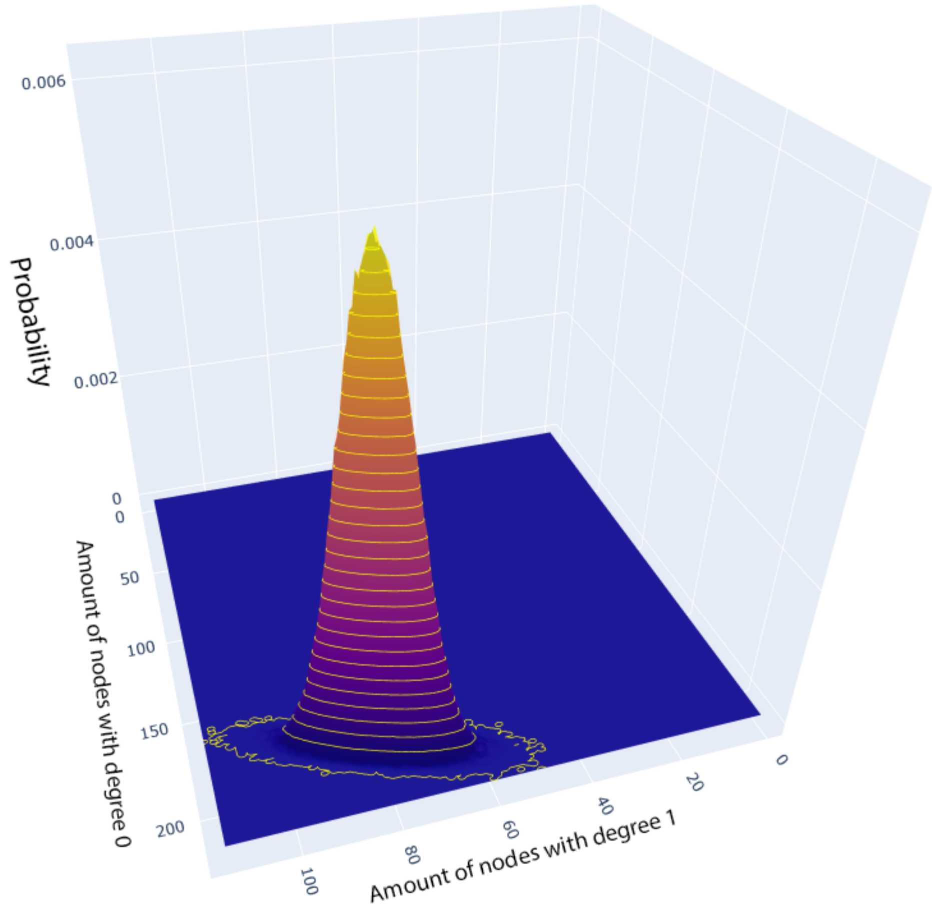

Fig. 4 – Joint distribution of nodes by degrees 0 and 1 for N = 300

The main part of this distribution has the form of an ellipse stretched along the diagonal. This means that the empirical joint distribution of the number of nodes by degrees can be represented by a multidimensional normal distribution:

$$f\left(n_0, n_1, \ldots, n_p\right) = \frac{N}{\left(2\pi\right)^{p/2}\sqrt{\det C}} \exp\left(-\frac{1}{2}\sum_{i,j=0}^{p} C_{ij}^{-1}\left(n_i - N\mu_i\right)\left(n_j - N\mu_j\right)\right),$$

Where $\det C$ is a determinant of the covariance matrix, and C_{ij}^{-1} is an element of the inverse covariance matrix.

The fact that the distribution of degrees of nodes is Gaussian follows from the local limit theorem of Moivre-Laplace applied to the polynomial distribution of vertices by degrees. Indeed, if μ_j is a probability that the node in FNNG has the degree j, then the probability that n_j nodes from N have a degree j, equal to $C_N^{n_j}\left(\mu_j\right)^{n_j}\left(1-\mu_j\right)^{N-n_j}$, what has Gaussian asymptotics in the limit of large N. So, the distribution of the form (8) is quite expected for correlated normally distributed random variables and confirms the correctness of the computational experiment. A nontrivial computational fact turned out to be that the probabilities μ_j, appearing in (8), themselves have a normal distribution for sufficiently large values N. According to the results of the analysis of graphs with the number of nodes N from 300 to 1500, based on a million experiments, the following values were obtained for each variant of the number of nodes μ_j, which are independent on N:

$$\mu_0 = 0{,}632;\ \mu_1 = 0{,}264;\ \mu_2 = 0{,}080;\ \mu_3 = 0{,}019;\ \mu_4 = 0{,}0036;\ \mu_5 = 0{,}00059;\ \mu_6 = 0{,}00008$$

These quantities have a Gaussian approximation with determination 0,999:

$$\mu_j \approx 3{,}7268 \exp\left(-0{,}1074\left(j+4{,}01\right)^2\right),\ j = 0,1,\ldots .$$

As a result, it turns out that both the normalization of the distribution (8) and the sum of the average values of the degrees equal to *N*.

4. Conclusion

This study is devoted to the development of a method for statistical analysis of graph structures using the example of nearest neighbor graphs. The aim of the work, which is still far from complete, is to construct a theory of the sampling method applying to graphic structures. In particular, here we consider the problem of estimating the probability that the observed graph has a typical structure in terms of the number of disconnected subgraphs and the distribution of nodes degree. In case of FNNG, these results will allow more correct interpretation of pattern recognition using this heuristic method.

These results are very useful for practical analysis of any concrete structure of objects system under the base supposition, that these objects are non-correlated with each other. If the number of nodes in each separate sub-graph corresponds to the mode of distribution function of the nodes and also the distribution of nodes powers is approximately equal to theoretical Gaussian distribution with the given accuracy, then we may confirm, that these objects are independent. But in the opposite case we can estimate the probability of this graph realization and conclude, that the observed structure corresponds to certain connections between objects.

In addition, the study of the frequency of occurrence of certain FNNG structures is closely related to the classical problems of combinatorial geometry. For example, an important task is to estimate the dimension of the embedding space of problem parameters by the matrix of distances between them. The maximum number of nearest neighbors for any of the points under study can be used to estimate the space dimension from below. Note that this problem has not been solved for spaces of arbitrary dimension. We also point out the connection between FNNG statistics and set splitting and Voronoi diagrams, which are encountered in various applications of discrete mathematics. It seems that numerical experiments in these directions could lead to the appearance of practically useful heuristics.

Another task area arising in the statistical analysis of FNNG structures - is the construction of asymptotic formulas for dependent distributions. Such solutions would make it possible in some cases to circumvent the practical difficulties associated with a complete enumeration of options that have exponential complexity.

References

1. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. – Springer, 2001. – 764 p.
2. *Chen G. H., Shah D.* Explaining the success of nearest neighbor methods in prediction. Foundations and Trends in Machine Learning, 10(5-6):337–588, 2018.
3. *Miller G. L., Teng S., Thurston W. and Vavasis S.A.* Separators for Sphere-Packings and Nearest Neighbor Graphs. Journal of the ACM, Vol. 44, No. 1, 1997, pp. 1–29.
4. *Eppstein D., Paterson M., Yao F.* On Nearest-Neighbor Graphs. Discrete & Computational Geometry, 1992, pp. 1-20.

5. *Baranchuk D., Persiyanov D., Sinitsin A. and BabenkoA*. Learning to routein similarity graphs. Proc. of the InternationalConference on Machine Learning, pp. 475–484, 2019.

6. *Kiana Hajebi, Yasin Abbasi-Yadkori, Hossein Shahbazi and Hong Zhang*. Fast Approximate Nearest Neighbor Search with k-Nearest Neighbor Graph. Proc. of 22-d International Joint Conference on Artificial Intelligence, 2009. P. 1312-1317.

7. *Erdos P., Renyi A*. On random graphs. Publ. Math. Debrecen, 1959. V. 6. P. 290-297.

8. *Kolkin V. F*. Sluchaynie graphi (in Russian). – M.: FizMatLin, 2004. – 256 c.

9. *Bollobas B*. Random Graphs. – Cambridge University Press, 2001. – 520 p.

10. *Janson S., Luczak T., Rucinski A*. Random graphs. – New York: Wiley, 2000. – 333 p.

11. *Raygorodski A. M*. Modelirovaniesluchaynihgrafoviihprimenenie. TrudiMFTI, 2010. T. 2. № 4. C. 130-140.(in Russian)

12. *Raygorodski A. M*. Modeli Interneta (in Russian). – Dolgoprudniy, «Intellekt», 2013. – 64 p.

13. *Barabasi L.-A., Albert R*. Emergence of scaling in random networks. Science, 1999. V. 286. P. 509-512.

14. *Leskovec J., Chakrabarti D., Kleinberg J., Faloutsos C., Gharamani Z*. Kronecker graphs: an approach to modeling networks. J. Machine Learning Research, 2010. V. 11. P. 985-1042.

15. *Hardy G.H., Ramanujan S*. Asymptotic Formulae in Combinatory Analysis. Proc. London Math. Soc., 1918, vol. 17, pp. 75-115.

16. *Vershik A. M., Kerov S. V*. Asymptotika mery Plansherelia symmetricheskoi gruppy predelnaia forma tablits Yunga (in Russian). Doklady AN USSR, 1977. V. 223 № 6. P. 1024-1027.

17. *Kislitsyn A. A., Orlov Yu.N*. Investigation of the nearest neighbor graph statistics (in Russian). Keldysh Institute Preprints. 2021. № 85. 23 p.