Hypernetwork approach to determination of occupational accident risks

Mehmet Ali Balcı[1] and Ömer Akgüller[2]

[1] Department of Mathematics, Faculty of Science, Muğla Sıtkı Koçman University, Muğla, Turkey
[2] Department of Mathematics, Faculty of Science, Muğla Sıtkı Koçman University, Muğla, Turkey
E-mail: mehmetalibalci@mu.edu.tr[1], oakguller@mu.edu.tr[2]

Abstract

Economic losses resulting from occupational accidents cause many researchers and policy makers to be interested in this issue. It is important to present effective mathematical methods as well as conventional statistical methods in the analysis of categorical occupational accident data. In this study, a mathematical method about edge estimation on a social network created with hyper-graphs for work accident analysis and prediction is presented. This method treats the process of the random walker evolving on the network with distributed order fractional derivative. The study was carried out numerical calculations in the years 2013–2014 in the dataset occupational accidents occurred in Turkey. The results obtained on a sample give prediction of possible occupational accidents characteristics.

2000 Mathematics Subject Classification. **05C65**. 26A33, 60G22, 05C82.
Keywords. hypergraphs, fractional calculus, edge prediction, network analysis.

1 Introduction

Occupational accidents that occur globally every year cause economic losses such as injuries, deaths, material and environmental damages. Therefore, the analysis and prediction of occupational accidents are important for both safety managers and policy makers. There are several approaches to the prediction and analysis of occupational accidents, such as traditional statistical analysis [1, 2, 3], questionnaire-based qualitative analysis [4, 5, 6], data-driven machine learning [7, 8, 9], and structured data analysis [10, 11, 12].

Apart from the results of the methods used in the analysis of occupational accidents, the main purpose of this study is to try to predict occupational accidents with the discovery of unforeseen and unexpected new structures. Standard occupational accident approaches all ignore the uncertainty of accidents. Mathematically, uncertainty is studied with fundamental theories such as hypergraph theory, fuzzy set theory, rough set theory, and soft set theory [13, 14, 15, 16]. Such soft computing approaches in occupational accident data have yielded effective results in determining the factors affecting occupational accidents [17].

Many factors affect the occurrence of work accidents. Some of these can be listed as carelessness of human behavior, inexperience at work, insufficient training in safety, unsafe conditions and psychological state of workers. Occupational accident data are vectors that contain descriptive information such as the number of injured people, the environment where the accident occurred, the age of the worker, the marital status of the worker, the level of education, the number of working days of the worker, and the number of days lost as a result of the accident. It is more difficult to deal with categorical variables than with continuous variables. For this reason, graphs are used as a structural data analysis approach in our study.

Tbilisi Mathematical Journal Special Issue (8 - 2021), pp. 51–76.
Tbilisi Centre for Mathematical Sciences.

Received by the editors: 26 May 2020.
Accepted for publication: 28 November 2020.

In mathematics, a combinatorial network representing social relations between a specific group is called a social network. Within an organization of individuals, social network structure encodes the interaction amongst members. Recently, social network approaches have been studied intensively in order to characterize occupational accidents [18, 19, 20, 21, 22]. The idea behind this approach is that workers in the same job sector are affected positively or negatively by other workers. When a sector's social network is established, it is possible to determine the most effective individuals in the communication process in this network with various mathematical and statistical measurements. In this study, we first present a hypergaph formation method for structured occupational accidents data. This method uses parametrization of each category. Then we perform link prediction process on the filtration of simple graph embedding of such hypergraphs. The link prediction model we present is a novel approach and uses fractional evolution of random walker in a network. Fractional calculus is generalization of integer order operations to fractional ones. In fractional derivative a memory function is employed. Then, by using similarity, we determine the possible occupational accidents.

The paper is organized as follows: Section 2 is presented as two parts. In first part, Section 2.1, we present basic of hypergraphs and link prediction. Also, two different graph embeddings of hypergraphs are given. In second part, Section 2.2, we introduce novel link prediction algorithm by considering the fractional evolution of a random walker on a graph. In this section, we give details how the memory function is employed. Moreover, a numerical scheme for solving such link prediction problem is presented. In Section 3, we give detailed computational results on 20132014 Occupational Accidents Data of Turkey. Finally, in Section 4, we present detailed discussion and conclusions.

2 Method

2.1 Hypernetworks and edge prediction

A natural way to express structural data sets is using graphs composed of vertices and edges. In mathematics, hypergraphs are the generalization of the graph structure in such way that allowing edges contain more than two vertices, and they are explicit data structure to represent hypernetworks. In last decades, hypergraphs and their applications have been studied extensively [23, 24, 25, 26]. In particular, a hypergraph $H = (V, E)$ is consisting of tuple where V is set of vertices and E is set of hyper-edges. Each hyper-edge $e \in E$ may contain arbitrarily many vertices. Hence, $e \subset P^V$. This definition of a hypergraph lead us to conclude that a hypergraph can be considered as a set system. If the elements of $e \in E$ do not have any order, then we say that H is an undirected hypergraph. Throughout this study, we only consider undirected hypergraphs. For more detailed definitions and theorems on directed and undirected hypergraphs, we refer readers to [27]. We also need to note that, an undirected simple graph is a special case for a hypergraph in which the size of an edge is restricted to two. Since hypergraphs are the generalization of simple graphs, they encode higher order relations on data sets.

In order to determine hypergraph statistics, we need more definitions such as degrees. The degree distributions are strong indicators on network analysis. Since there is no restriction on a hyperedge cardinality, we can define two type of degrees and their distributions on hypernetworks. In a hypergraph $H = (V, E)$; the degree $d^V(v)$ of $v \in V$ is the number of hyper-edges containing v, whereas the degree $d^E(e)$ of a hyper-edge $e \in E$ can be defined as the number of vertices belongs to e. By counting how many vertices and edges have each degree, we can form the vertex degree distribution $P_{d^V(v)}(k)$, defined by

$$P_{d^V(v)}(k) = \text{fraction of vertices in } H \text{ with the degree } k \tag{2.1}$$

and the hyperedge degree distribution $P_{d^E(e)}(k)$, defined by

$$P_{d^E(e)}(k) = \text{fraction of hyperedges in } H \text{ with the degree } k. \tag{2.2}$$

We refer readers to [28] for more details on degree distributions for hypergraphs. For a hypergraph $H = (V, E)$, the sequence of vertices and hyper-edges

$$v_1, e_1, v_2, e_2, \ldots, v_{k-1}, e_k, v_k$$

is called a hyper-walk between the vertices $v_1, v_k \in V$ for $1 \leq i \leq k-1$.

In a hypergraph setting, there are several hyper-edge prediction measures used in literature. Let $N(v)$ denote the set of neighbours of $v \in V$ in $H(V, E)$. If C is all the set of absent hyper-edges in H, then we use the following features for each possible hyper-edge candidate for $c \in C$ with the score function x as

1. Common Neighbours (CN) [29] : $x(c) = \bigcap_{v_i \subseteq c} N(v_i)$
2. Jaccard Coefficient (JC) [30]: $x(c) = \dfrac{\bigcap_{v_i \subseteq c} N(v_i)}{\bigcup_{v_i \subseteq c} N(v_i)}$
3. Adamic-Adar Index (AA) [31]: $x(c) = \sum_{v_j \in \bigcap_{v_i \subseteq c} N(v_i)} \dfrac{1}{\log |N(v_j)|}$
4. Katz Index (KI) [29]: $\sum_{k=1}^{\infty} \beta^k \left| \mathcal{L}_{ij}^{<k>}, \right|$, where $\mathcal{L}_{ij}^{<k>}$ is the set of all hyper-walks with length k connecting v_i and v_j, and β is a free parameter.

With the expansion of the usage disciplines of simple graph theory, many algorithms are defined on these structures. Although hypergraphs contain higher order information that simple graphs cannot encode, it can be said that technical and methodological developments on hypergraph theory are not as remarkable as simple graph theory. There are several methods for embedding high-order relationship information contained in hypergraphs into simple graphs. The most common embedding method is to bipartite graphs which are subclasses of simple graphs such that the set of vertices decomposed into two disjoint sets in a way that none of two vertices are within the same set are adjacent [27]. For a hypergraph $H = (V, E)$, its bipartite graph representation $G_B = (V_1 \cup V_2, E_B)$ is tuple with $V_1 = V$, $V_2 = E$, and E_B is the set of dyadic relations of hyper-edge inclusions. The second embedding method we mention in this study is related to the connectivity of the vertices through hyper-edges. Let $H = (V, E)$ be a hypergraph. Since each $e_i \in E$ involves arbitrarily many vertices, there exist such a simple graph $G_{C_i} = (e_i, E_{C_i})$ such that each vertices in e_i are adjacent. Therefore, the Boolean sum of $G_C = G_{C_1} \bigoplus \cdots \bigoplus G_{C_m}$ emerges as a simple graph representation of H regrading to hyper-connectivity of vertices [32].

In Figure 1, we present an example for a hypergraph and its simple graph representations. In this example, the hypergraph $H = (V, E)$ is given with edge sets as $e_1 = \{v_1, v_2, v_5, v_6\}$, $e_2 = \{v_6, v_8, v_9\}$, $e_3 = \{v_2, v_3, v_4\}$, and $e_4 = \{v_7, v_{10}\}$, where $V = \{v_1, v_2, \ldots, v_{10}\}$. The bipartite embedding of H is a simple graph G_B with the vertex set $V_B = V \cup E$ and the elements of E_B are formed by inclusion. Moreover, the embedding of H regarding to hyper-connectivity of vertices is a simple graph $G_C = G_{C_1} \bigoplus \cdots \bigoplus G_{C_4}$.

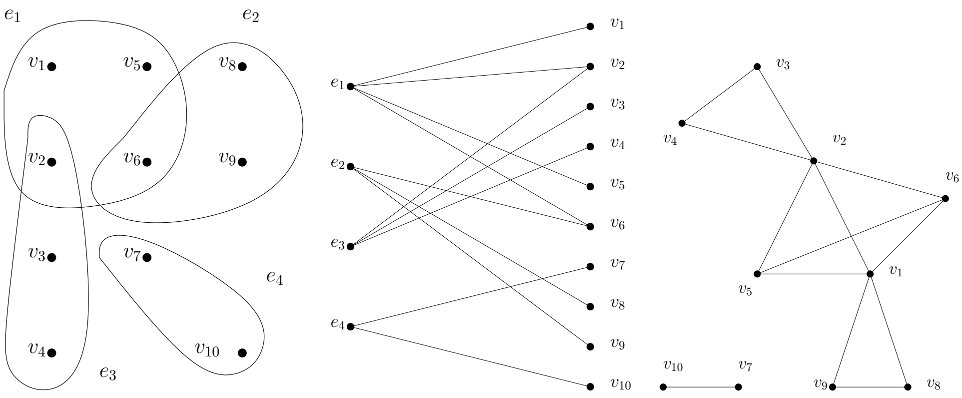

FIGURE 1. Hypergraph $H = (V, E)$ is on the left, the simple graph $G_B = (V_B, E_B)$ in in the middle, and the simple graph $G_C = (V, E_C)$ is on the right

While performing edge prediction task on $H = (V, E)$, the indices CN, JC, AA, and KI give the score of the hyper-edge tendency to be included in E. In order to keep bipartite topology of G_B embedding of H, such task should be performed as edges included in between e_i and v_j. Otherwise, the topology of H will change by emerging multi-hyperedges or loops in vertices. The motivation behind the G_C embedding of H is representing the hypergraph structure as a clique-complex. Hence, as the predicted hyper-edges being included in E, the updated embedding will still remain as a clique complex. However, such complexes are not regular. Therefore, the hyper-edge prediction on H becomes clique prediction process on G_C.

2.2 Communicability centrality with distributed power-law memory

In network analysis, centrality of a vertex is and indicator which measure importance of a vertex in a network [33, 34, 35]. There are several centrality measures in literature that are used for different purposes. In a network, an important quantity for determining communication centrality of a vertex is the communicability function [36, 37], defined by for vertices i and j as

$$Comm_{ij} = \sum_{k=0}^{\infty} \frac{A_{ij}^k}{k!} = \exp(A)_{ij} = \sum_{\mu=1}^{|V|} \exp(\lambda_n)\varphi_\mu(i)\varphi_\mu(j), \tag{2.3}$$

where A is the adjacency matrix of the network represented by $G = (V, E)$ and right-hand side emerges from the spectral decomposition of each term of the Taylor expansion of A_{ij}^k.

The communicability function defined in 2.3 can be visualized in different way. Now, let us consider a random-walker distribution on G described with $\vec{\varphi}(t)$, where t is the variable of time.

Then, the matrix differential equation

$$\frac{d\vec{\varphi}(t)}{dt} = A\vec{\varphi}(t) \tag{2.4}$$

gives the time evaluation of the random-walker $\vec{\varphi}(t)$. The analytical solution of the Equation 2.4 is $\vec{\varphi}(t) = \exp(At)\vec{\varphi}(0)$. Hence, $\exp(A)$ is the time evolution operator.

In a social network, the evaluation of a random walker can be affected from regulations done by institutions. Hence, we need to include memory of a random-walker evaluating on G. Now, let us assume that the endogenous variable $Y(t)$ is dependent on the history of the change of $\vec{\varphi}(\tau)$ in G for $\tau \in [0,t]$. Then, the endogenous variable can be defined by

$$Y(t) = \int_0^t M(t,\tau)\vec{\varphi}(\tau)d\tau, \tag{2.5}$$

where $M(t,\tau)$ is the kernel of the Volterra type operator and called the memory function. Furthermore, if we assume homogeneity in time, then

$$\frac{\partial M(t,\tau)}{\partial t} + \frac{\partial M(t,\tau)}{\partial \tau} = 0. \tag{2.6}$$

The solution of the Equation 2.6 is $M(t,\tau) = M(t-\tau)$.

In this study, we assume that the evaluation of the random-walker has power-law fading memory; that is,

$$M(t,\tau) = \frac{1}{\Gamma(n-\alpha)}\frac{m}{(t-\tau)^{\alpha-n+1}}, \tag{2.7}$$

where $\alpha > 0$ is the power-law fading parameter, $m(\alpha) \in \mathbb{R}$, $n = [\![\alpha]\!] + 1$, and Γ is the Gamma-function.

If we substitute memory function given in Equation 2.7 into the Equation 2.5, then the endogenous variable $Y(t)$ becomes

$$Y(t) = m\left(D^{\alpha}_{0+}\vec{\varphi}(t)\right) \tag{2.8}$$

with

$$D^{\alpha}_{0+}\vec{\varphi}(t) = \frac{1}{\Gamma(n-\alpha)}\int_0^t \frac{\vec{\varphi}^{(n)}(\tau)d\tau}{(t-\tau)^{\alpha-n+1}}, \tag{2.9}$$

where $\vec{\varphi}^{(n)}(\tau) = \dfrac{\partial^n \vec{\varphi}}{\partial \tau^n}$. The Equation 2.9 is called left-sided Caputo fractional derivative of $\vec{\varphi}(t)$.

The memory function we employ into the evaluation process of a random-walker makes the walker remember recent historical changes. However, such memory can not be the same for each step of the evaluation in social systems. Hence, if we distribute α parameter on time interval, with the cumulative distribution function $p(\alpha)$ let us to have the memory function as

$$M(t-\tau) = \int_0^1 \frac{p(\alpha)(t-\tau)^{n-\alpha-1}}{\Gamma(\alpha-n)}\, d\alpha. \tag{2.10}$$

Then, it is possible to define Caputo fractional derivative of distributed order on $[0,1]$ as

$$D^{[0,1]}_{0+}\vec{\varphi}(t) = \int_0^1 p(\alpha) D^{\alpha}_{0+}\vec{\varphi}(t)\, d\alpha, \tag{2.11}$$

where $\int_0^1 p(\alpha)\ d\alpha = 0$. Therefore, the evaluation of the random-walker with distributed order power-law memory emerges as

$$D_{0+}^{[0,1]}\vec{\varphi}(t) = A\vec{\varphi}(t). \tag{2.12}$$

The solution of the Equation 2.12 yields us to obtain random-walker distribution on G with adjacency matrix A. Then, it is possible to determine communicability centrality with distributed power-law memory of a particular vertex. Since the solution of the Equation 2.12 directly varies respect to the definition of $p(\alpha)$, we present a fast discretization scheme for the numerical solution. The numerical solution respect to finite differences of distributed order fractional differential equations are intensively studied in [38, 39, 40, 41, 42].

The communicability centrality with distributed power-law memory of a particular vertex depends on the ij-th entity of the vector $\vec{\varphi}(t)$. Since the matrix A encodes the neighborhood data on G, the ij-th of the vector $\vec{\varphi}(t)$ satisfies

$$D_{0+}^{[0,1]}\vec{\varphi_{ij}}(t) = A_{ij}\vec{\varphi_{ij}}(t). \tag{2.13}$$

Now, let us obtain the numerical solution of the distributed order equation 2.13. In order to obtain discretization scheme, we discretize the interval $[0,1]$, in which the order α is changing and $\Delta\alpha_k$ are the grid steps. We shall note that such discretization does not need to be equidistant; however, in this study, we give the discretization scheme respect to equidistant grid steps. Then we have

$$\begin{aligned} D_{0+}^{[0,1]}\vec{\varphi_{ij}}(t) &= p(\alpha)D_{0+}^{\alpha}\vec{\varphi_{ij}}(t)\ d\alpha \\ &\approx \sum_{k=1}^{n} p(\alpha_k)\left(D_{0+}^{\alpha}\vec{\varphi_{ij}}(t)\right)\Delta\alpha_k. \end{aligned} \tag{2.14}$$

In [43], approximation to the right-sided Caputo derivatives are given with a matrix

$$F_n^{\alpha} = \frac{1}{\tau^{\alpha}}\begin{pmatrix} \omega_0^{\alpha} & 0 & 0 & 0 & \cdots & 0 \\ \omega_1^{\alpha} & \omega_0^{\alpha} & 0 & 0 & \cdots & 0 \\ \omega_2^{\alpha} & \omega_1^{\alpha} & \omega_0^{\alpha} & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \omega_{n-1}^{\alpha} & \ddots & \omega_2^{\alpha} & \omega_1^{\alpha} & \omega_0^{\alpha} & 0 \\ \omega_n^{\alpha} & \omega_{n-1}^{\alpha} & \ddots & \omega_2^{\alpha} & \omega_1^{\alpha} & \omega_0^{\alpha} \end{pmatrix}, \tag{2.15}$$

where $t = j\tau$, $j = 0, 1, \ldots, n$, and

$$\omega_j^{\alpha} = (-1)^j \binom{\alpha}{j}.$$

By employing the matrix given in Equation 2.15 and the approximation given in Equation 2.14, we can have the discretized scheme for the Equation 2.13 as

$$\sum_{k=1}^{n} p(\alpha_k)F_n^{\alpha}\Delta\alpha_k = A_{ij}\vec{\varphi_{ij}}. \tag{2.16}$$

Then, the solution of Equation 2.16 is $n \times n$-matrix

$$\hat{\varphi} = \sum_{k=1}^{n} p(\alpha_k) \Delta\alpha_k A^* F_n^{\alpha}, \tag{2.17}$$

where A^* is Moore-Penrose matrix of the adjacency matrix.

3 Results on occupational accident data of Turkey

3.1 Data and network construction

In this study, the raw data of occupational accidents occurred in Turkey in 2013–2014 are used. Turkish Social Security Agency (SGK) has permitted to be shared the raw data with an official permission of number 99604924/910/4422955 and date 27/08/2015. The entries containing missing information were cleared from the raw data and then 432090 pieces of data were obtained. The sample data set is randomly chosen to be 10000 to perform analyses. In the selection of the sample data set, 18 sectors with the highest number of occupational accidents were selected. The selected sectors are expressed in 4-digit $NACE$ codes in this study. The web-site "https://opendata.eulerhermes.com" is referred to the readers for relevant explanations of the sectors and $NACE$ codes.

The entries of occupational accident information are taken as Number of Working Days, Age, Gender, Marital Status, Loss of Work Days, Vocational Education, Occupational Safety Training, Education Status, Number of Persons in Accident. In [44], authors briefly introduced a network formation method for similar data set by using similarity of each data entry. Their method also includes the information about the place of the occupational accident. In this study, we follow a novel way to determine hypernetwork structure. First, data sets for each $NACE$ code are divided into subgroups. Different parameters used when dividing into subgroups are given in Table 1.

Number of Working Days ($t \equiv$ days)					
$0 \leq t < 400$	$400 \leq t < 1000$	$1000 \leq t < 2000$	$2000 \leq t < 3000$	$3000 \leq t < 4000$	$t \geq 400$
Age ($t \equiv$ years)					
$18 \leq t < 25$	$25 \leq t < 30$	$30 \leq t < 35$	$35 \leq t < 40$	$40 \leq t < 45$	$t \geq 45$
Working Days Loss ($t \equiv$ days)					
$0 \leq t \leq 1$	$1 < t \leq 3$	$3 \leq t < 5$	$5 \leq t < 8$	$8 \leq t < 10$	$t \geq 10$
Number of Persons in the Accident					
1		1-3		>3	
Educational Status					
Elementary School	Secondary School		High School	University / Graduate	
Gender					
Male			Female		
Martial Status					
Married		Bachelor / Bachelorette		Other	
Vocational Training					
Yes			No		
Occupational Safety Education					
Yes			No		

Table 1. The parameters list for network formation

The hypernetwork representation of each $NACE$ sector data is composed of hyper-edges which are parameters for sub-grouping, and occupational accidents as vertex set. Then, a weighted simple-graph representation G_C is used for link prediction procedure.

3.2 Computational results

In this section, we first present the results of hypernetworks of each $NACE$ sector data. Hyper-edge distributions are presented in Figure 2 and vertex degree distributions of respect G_B. are presented in Figure 3.

From the distributions of H and G_B, it is straightforward that the G_C representations of H of each $NACE$ sector which encode the connectivity of agents in complex environment emerge as complete graphs. Complete graphs are combinatorial objects in which every pair of distinct vertices is connected by a unique edge. Hence, link prediction on G_C will be a null operation. However, this property assure us about the hypothesis of this study which depends on the information flow amongst agents in context of occupational accidents. In order to perform link prediction, we need to filter each G_C by using very known method called minimum spanning trees. A minimum spanning tree (MST) filtration of G_C is a subgraph that connects all the vertices together, without any cycles and with the minimum possible total edge weight. In this study, for the weighting scheme, we follow the cosine similarity within each $NACE$ sector data set. Then, the link prediction process is applied to MST structure. The higher similarity scores on MST measures the highest possibility of occupational accident to appear.

In order to explain this filtration process, let us consider "Repair of electrical equipment" sector with $NACE$ code 3314. The sample size of this data set is 156. Its distance matrix based on cosine similarity and MST filtration which can be considered as hierarchical clustering are given in Figure 4.

In order to perform link prediction on MST filtrations of G_C graphs, we use novel similarity measure communicability centrality with distributed power-law fading as explained in Section 2.2. The solution method described in Equation 2.17 has discretized scheme with n-steps. Since A_{ij} denotes the adjacency matrix of MST, we consider n to be equal to size of the vertex set of MST. Moreover, the adjacency matrices of MSTs are sparse. Therefore, in order to guarantee the numerical solution of Equation 2.13, we use Moore-Penrose inverse of A_{ij}. Besides, as the random walker decisions affected by the institutional regulations, the distribution of those regulations amongst agent are considered in two ways, namely normal and logistic distributions. In normal distribution case, the random walker's power-law fading memory is distributed without making strong assumptions. However, in later case which is logistic distribution, the power-law fading memory of random walker is distributed respect to that agents have also a learning process. In this present study, we perform our analyses minimally affected by the diffusion process, that is we only consider the fractional order differentiation to be $\alpha = 0.5$. In Figures 5–24, we present the numerical solution of Equation 2.13 for each $NACE$ sectors, with different distributions and $\alpha = 0.5$.

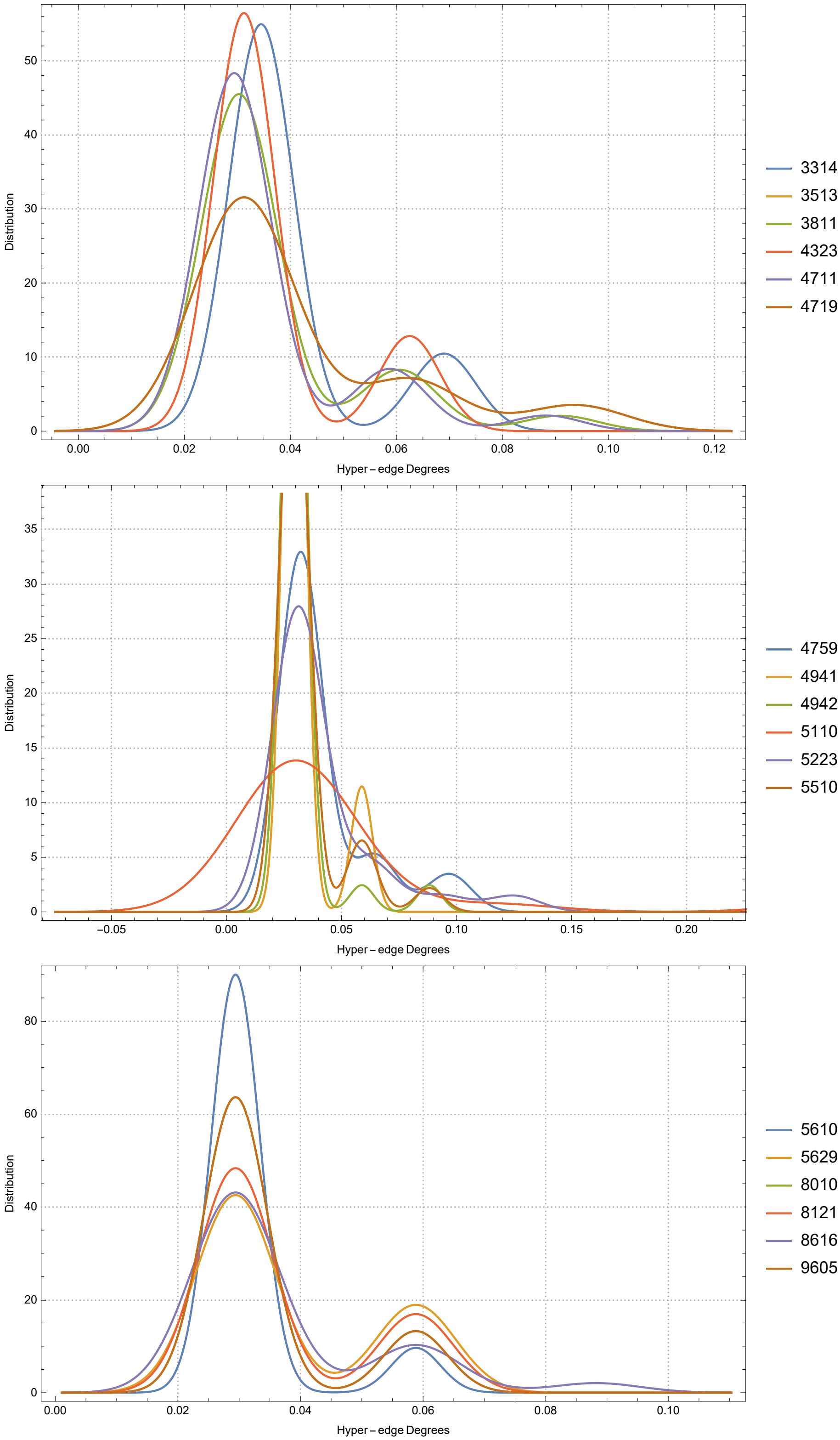

FIGURE 2. Hyper-edge degree distributions of each $NACE$ sector.

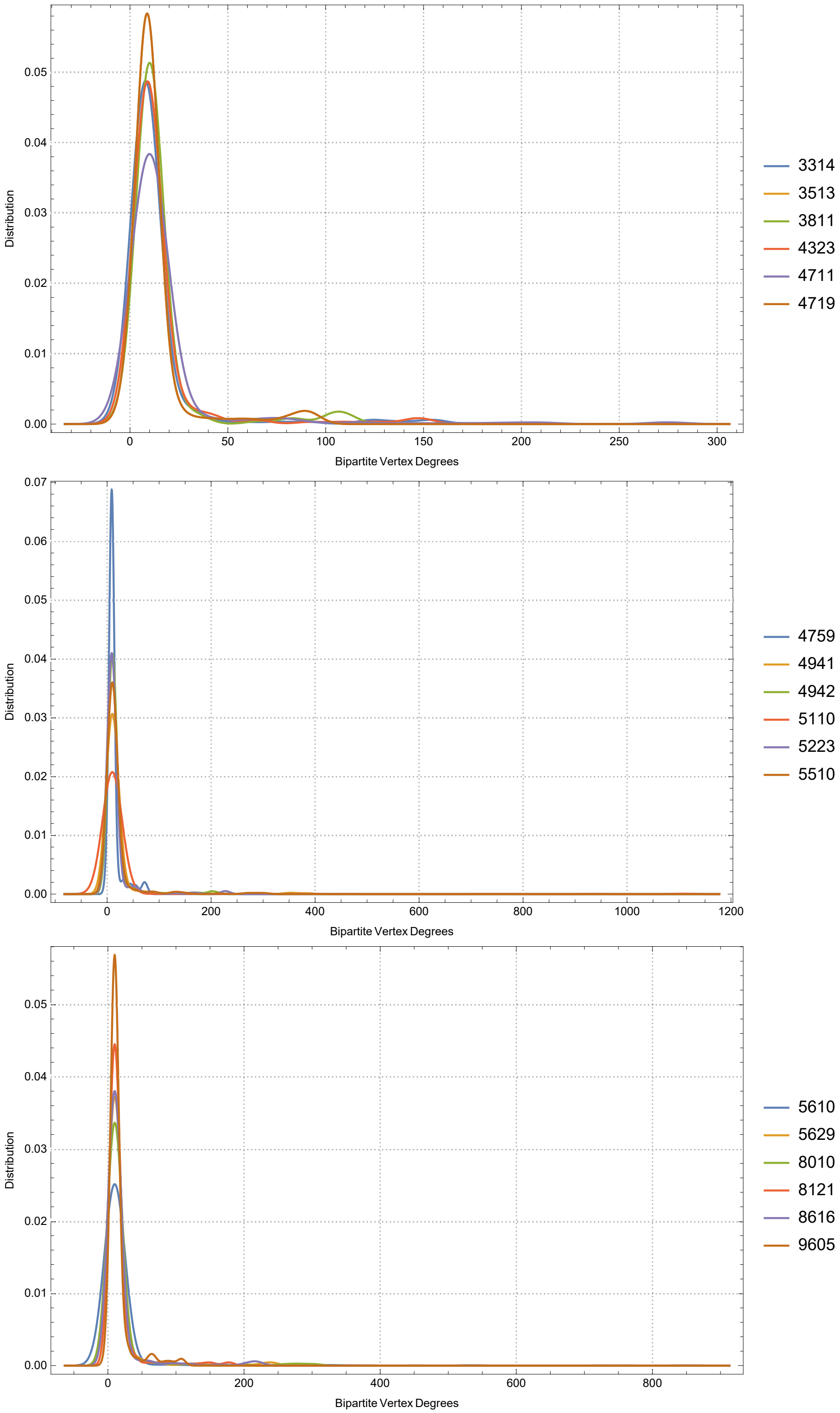

FIGURE 3. Vertex degree distributions of bipartite graphs G_B of each $NACE$ sector.

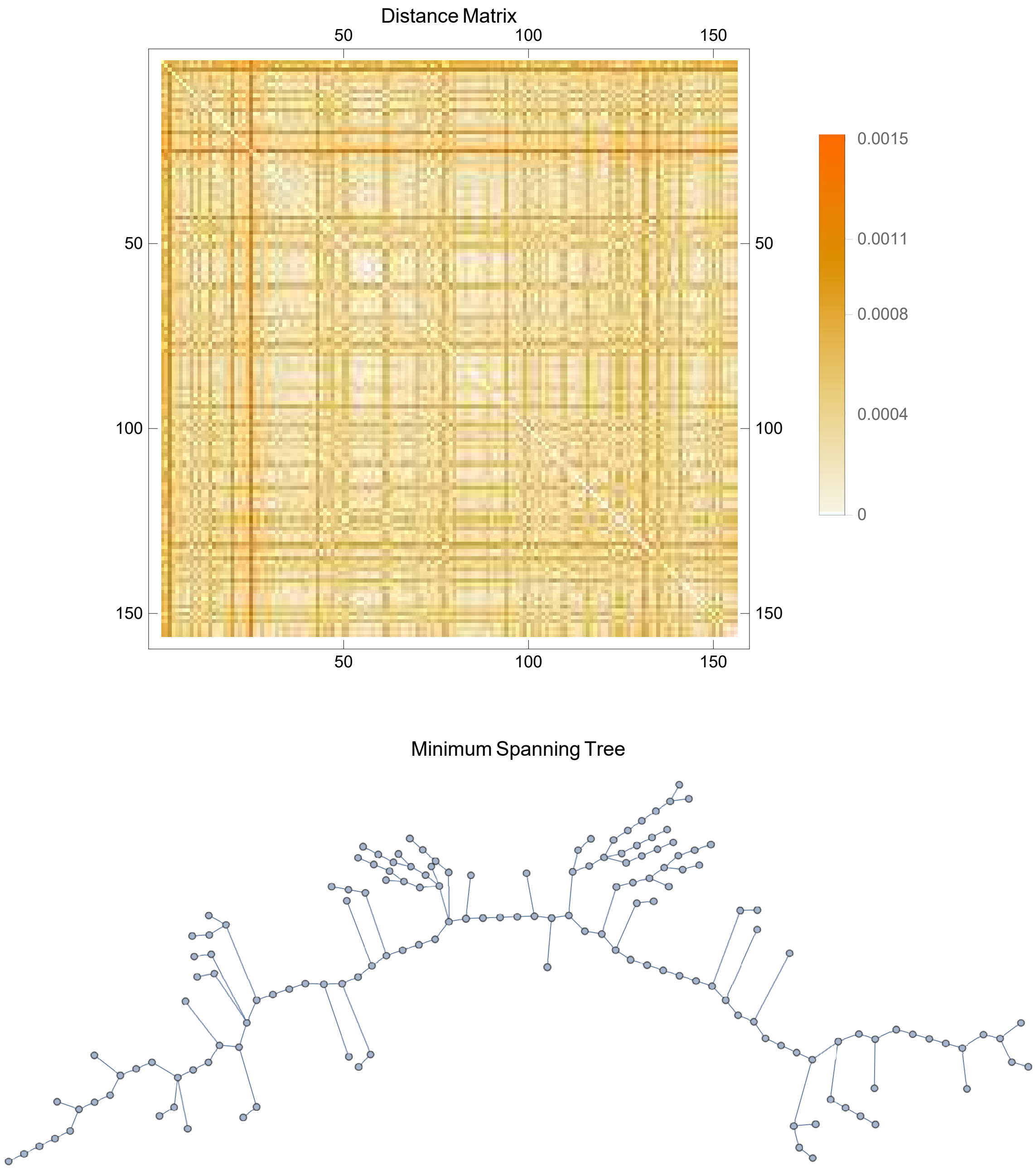

FIGURE 4. Filtration of G_C representation of 3314

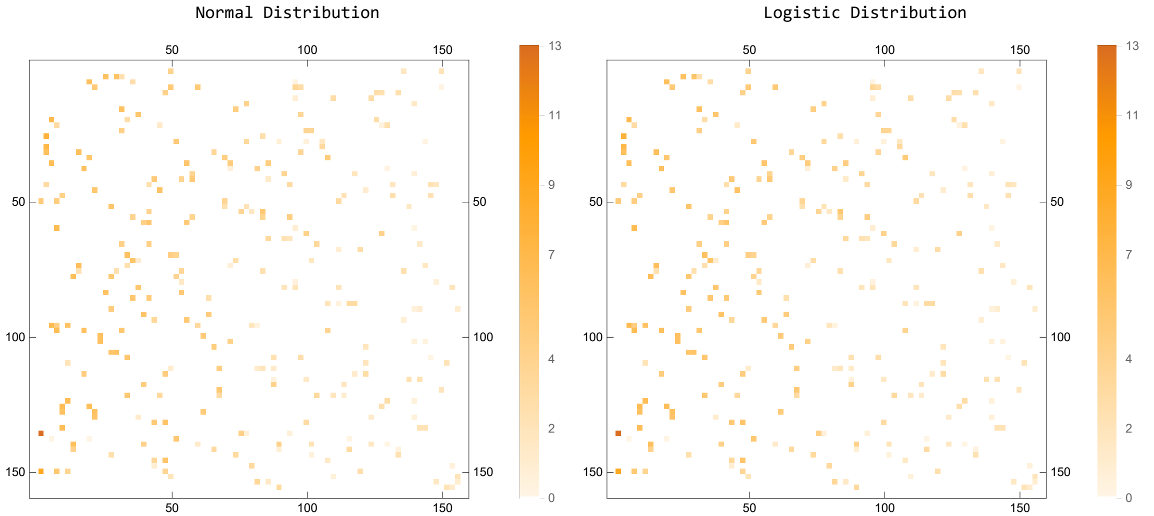

FIGURE 5. Distributed order communicability matrix of $NACE$ sector 3314 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

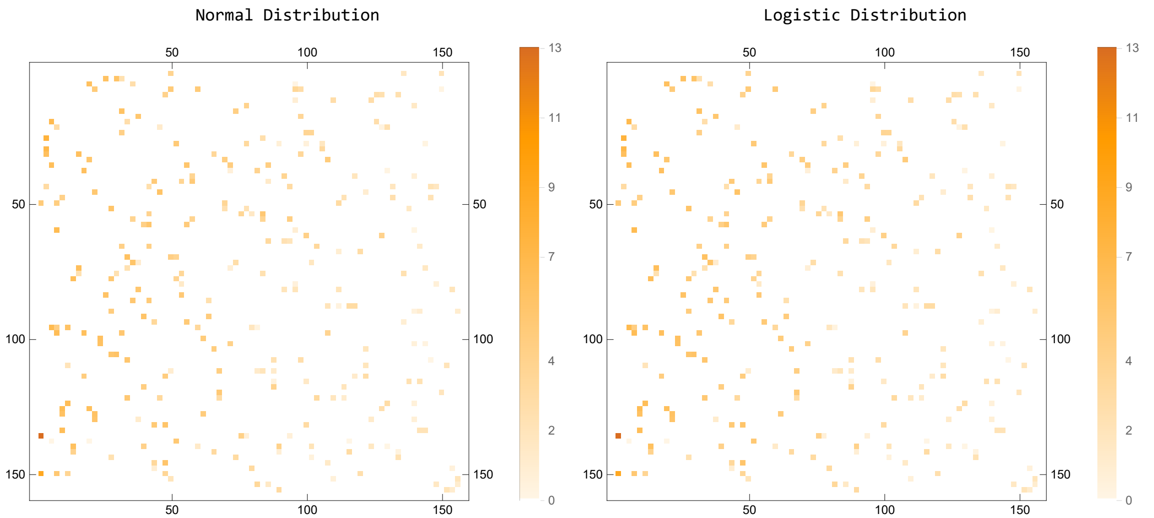

FIGURE 6. Distributed order communicability matrix of $NACE$ sector 3513 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

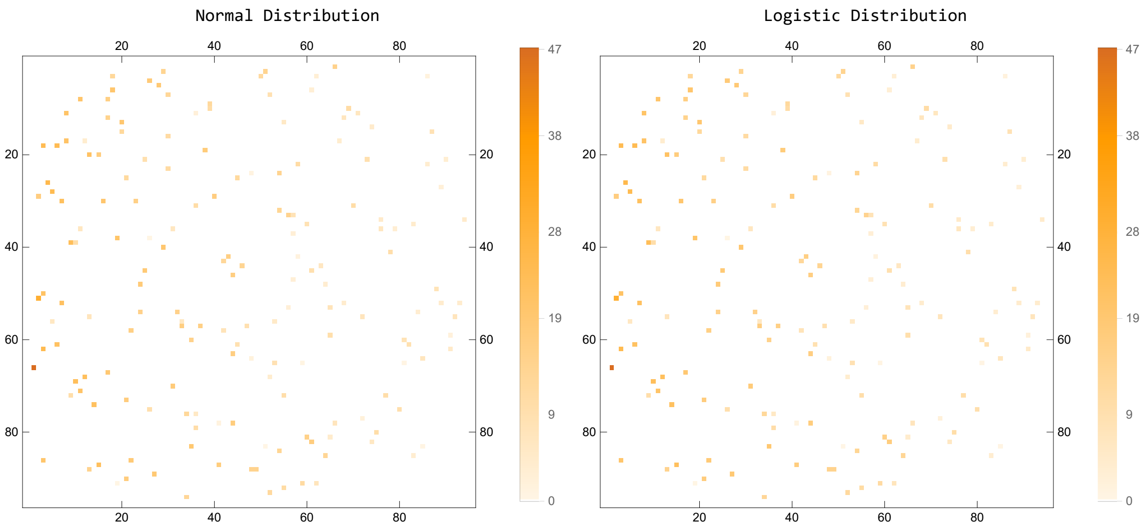

FIGURE 7. Distributed order communicability matrix of $NACE$ sector 3811 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

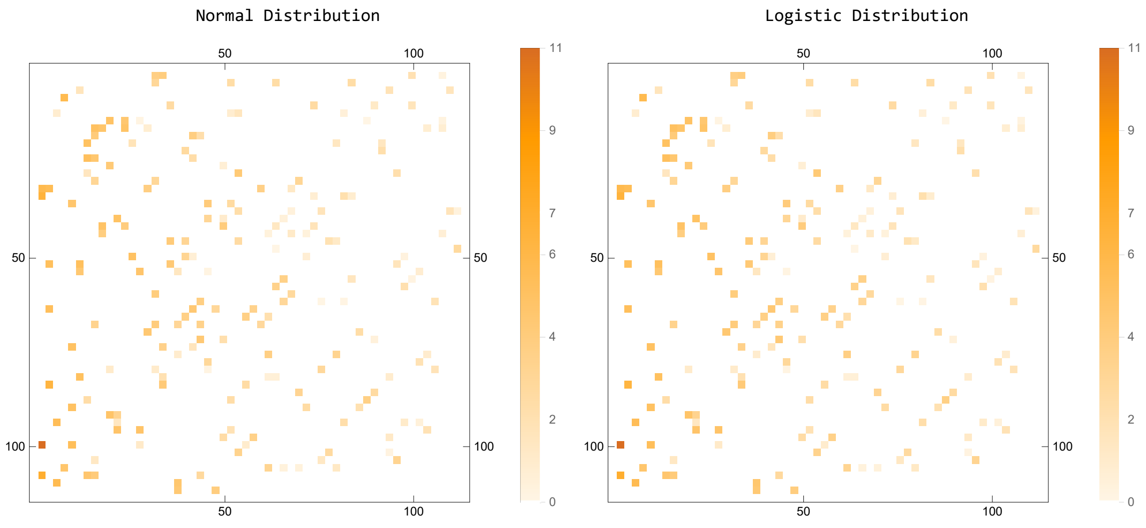

FIGURE 8. Distributed order communicability matrix of $NACE$ sector 4323 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

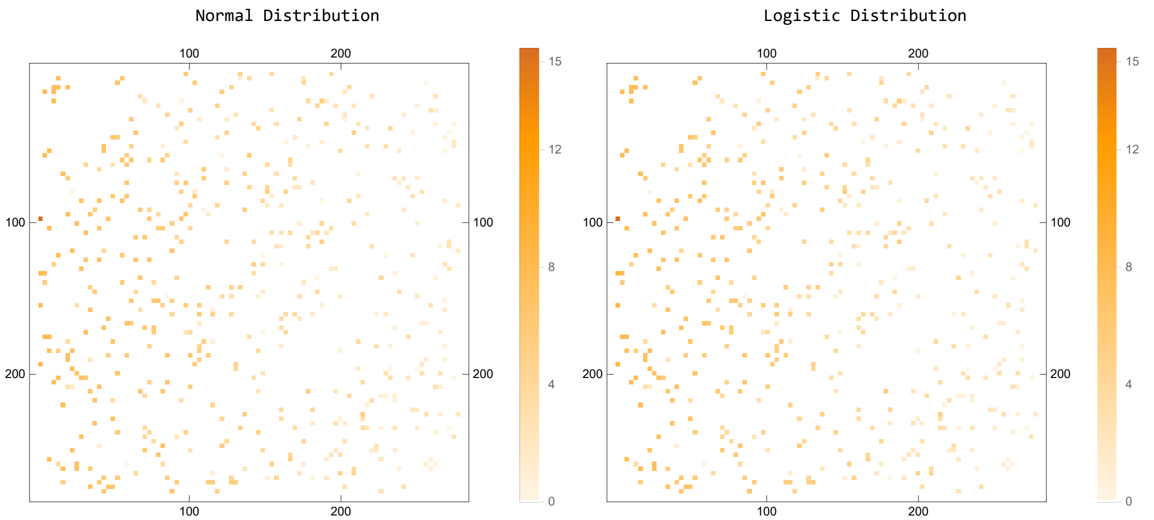

FIGURE 9. Distributed order communicability matrix of $NACE$ sector 4711 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

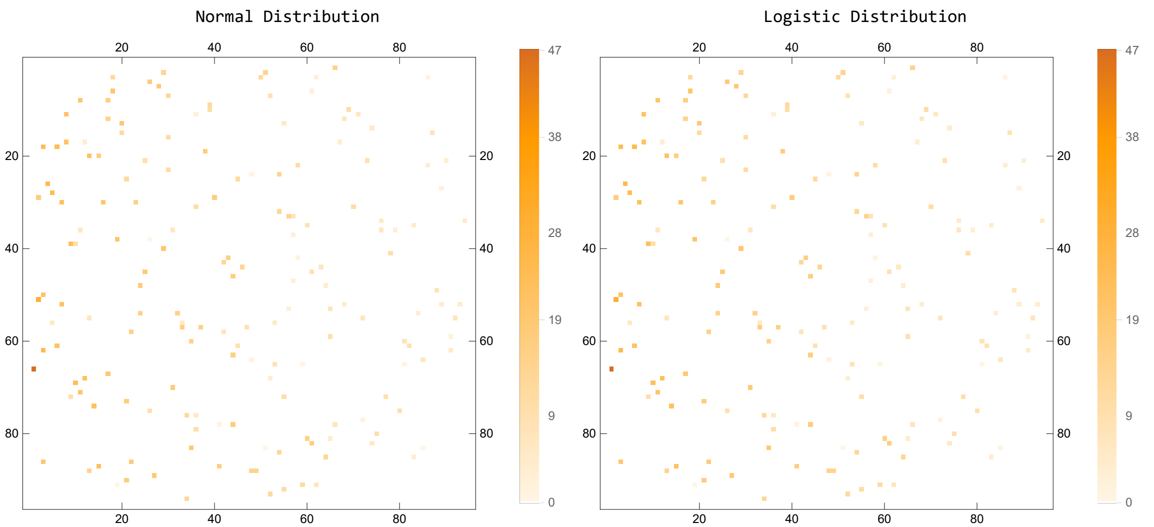

FIGURE 10. Distributed order communicability matrix of $NACE$ sector 4719 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

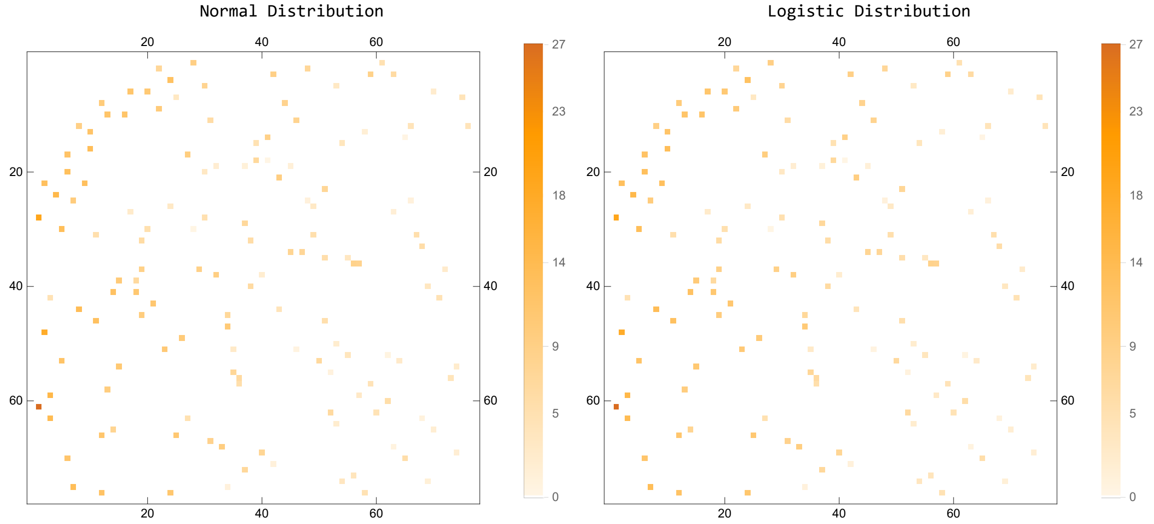

FIGURE 11. Distributed order communicability matrix of $NACE$ sector 4759 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

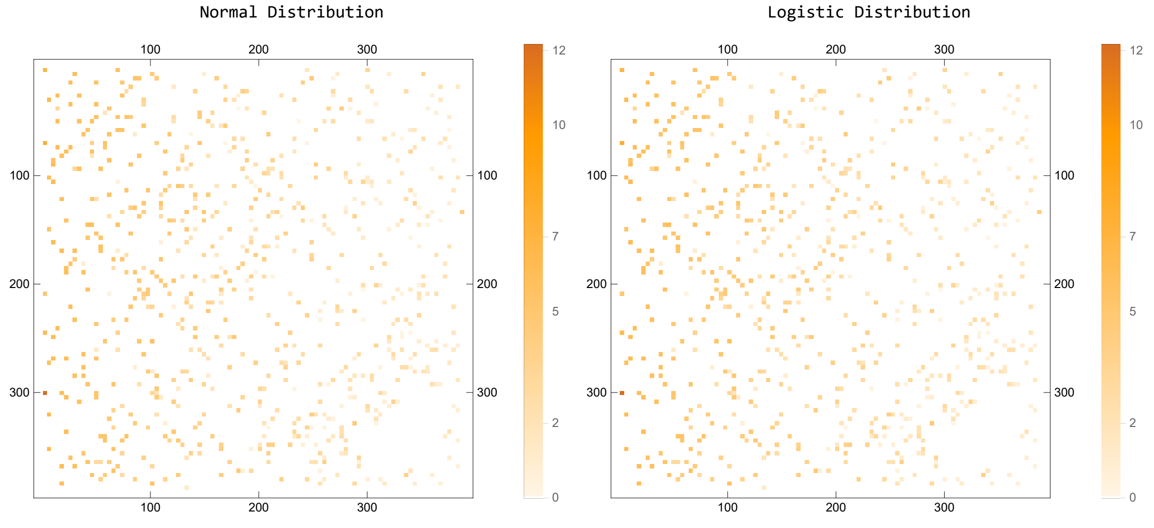

FIGURE 12. Distributed order communicability matrix of $NACE$ sector 4941 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

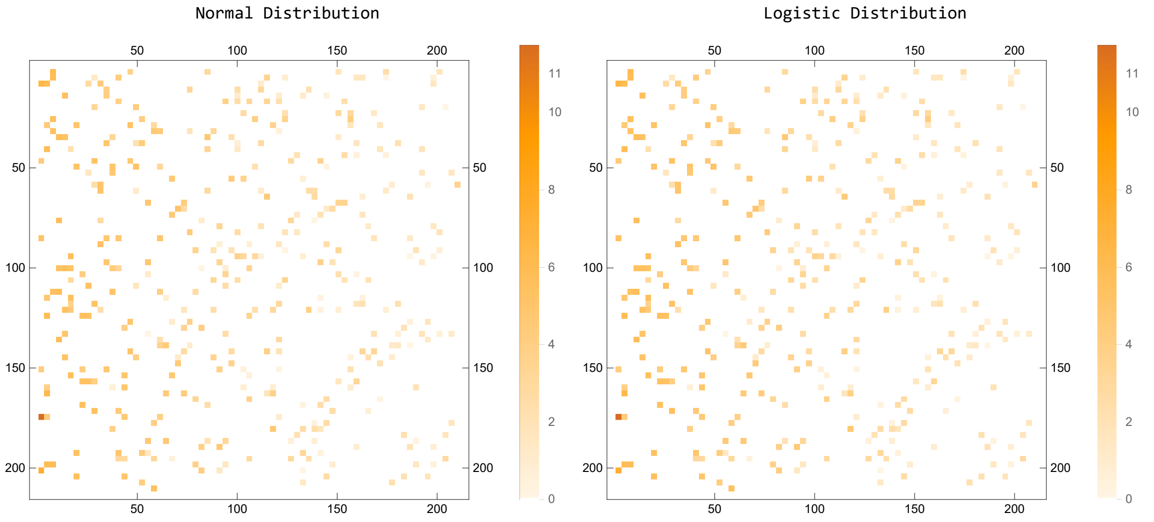

Figure 13. Distributed order communicability matrix of $NACE$ sector 4942 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

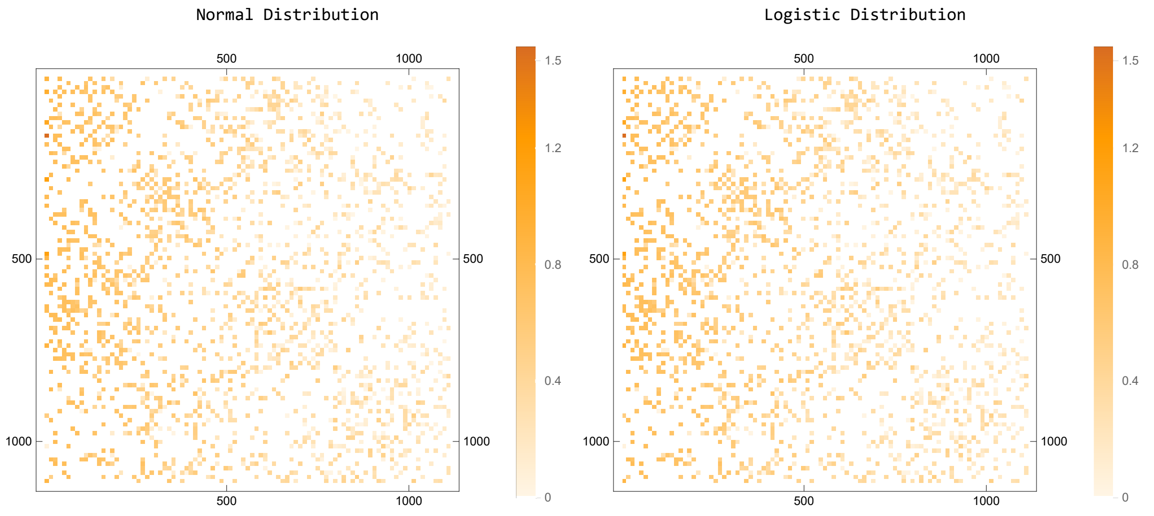

Figure 14. Distributed order communicability matrix of $NACE$ sector 5110 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

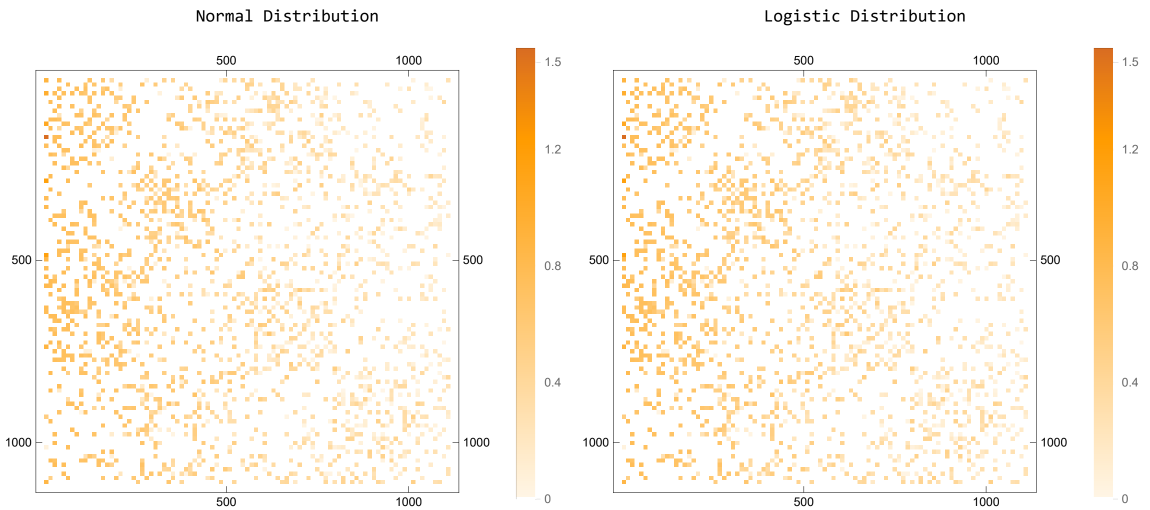

FIGURE 15. Distributed order communicability matrix of $NACE$ sector 5223 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

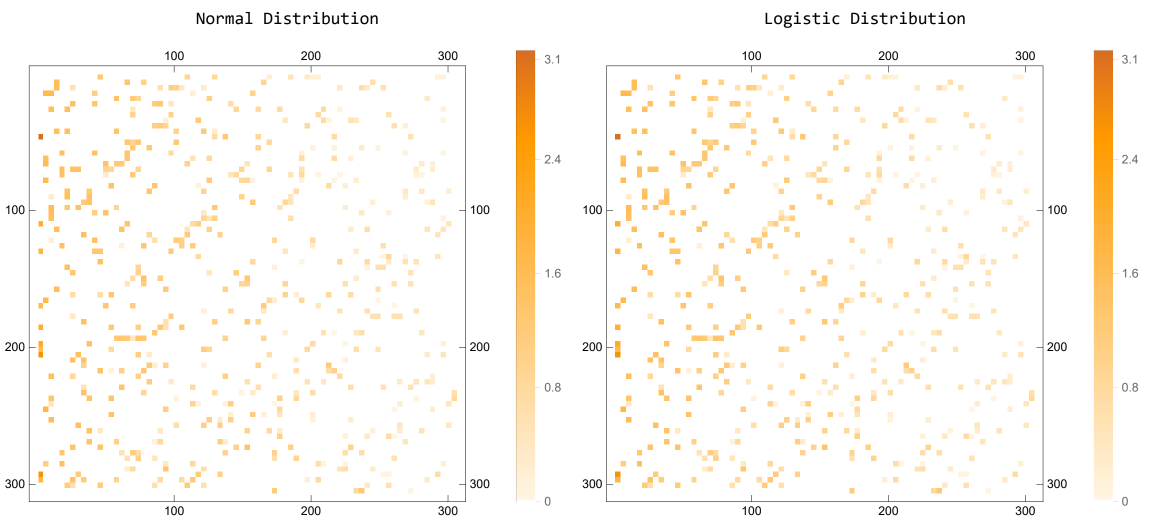

FIGURE 16. Distributed order communicability matrix of $NACE$ sector 5510 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

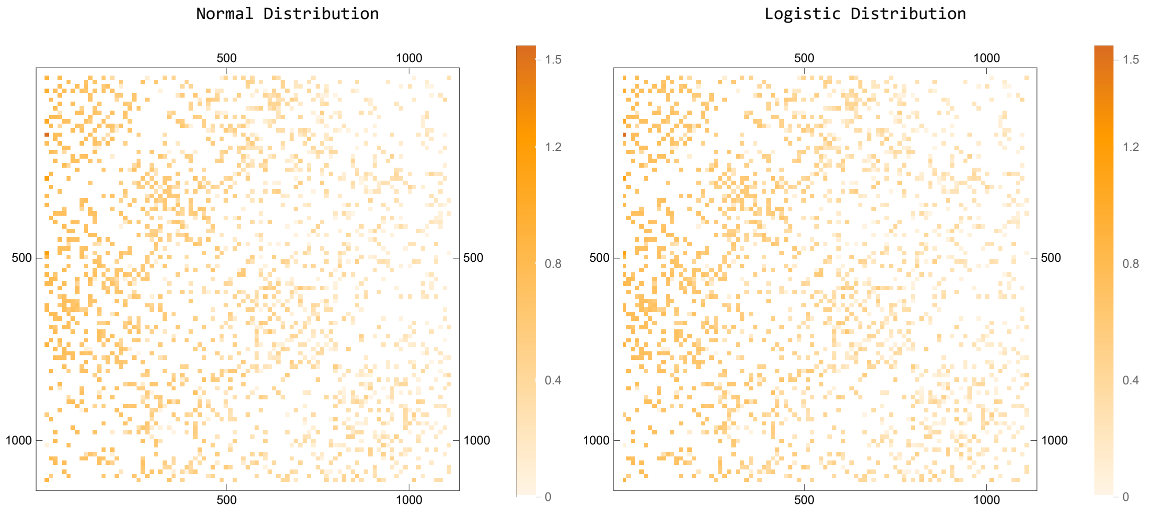

Figure 17. Distributed order communicability matrix of $NACE$ sector 5223 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

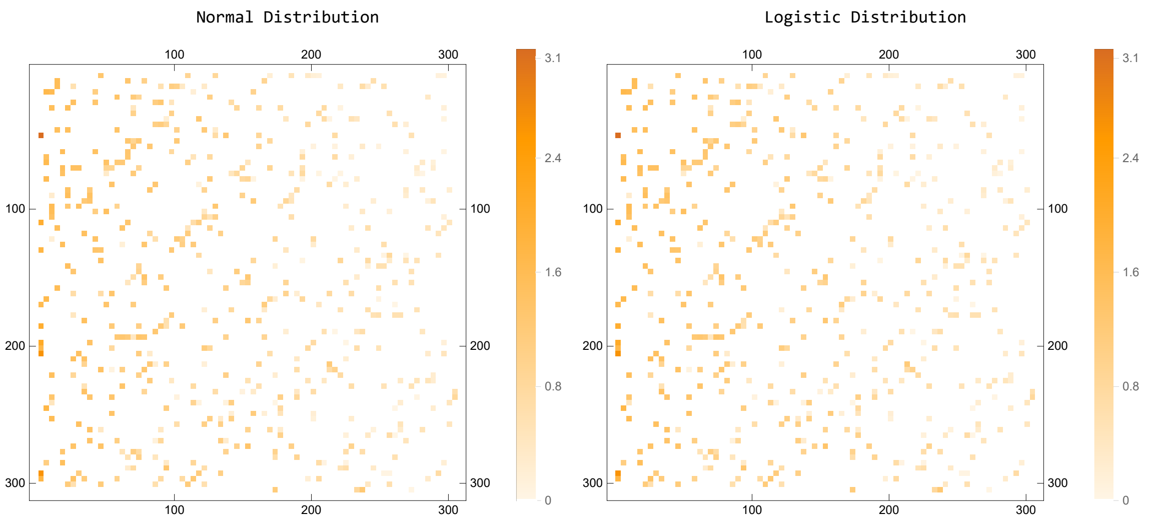

Figure 18. Distributed order communicability matrix of $NACE$ sector 5510 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

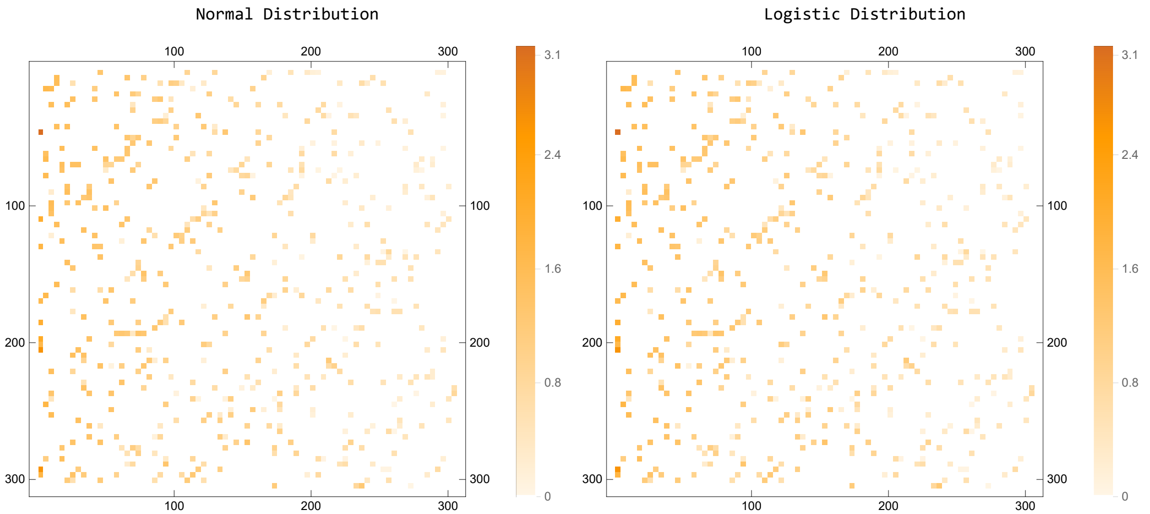

FIGURE 19. Distributed order communicability matrix of $NACE$ sector 5610 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

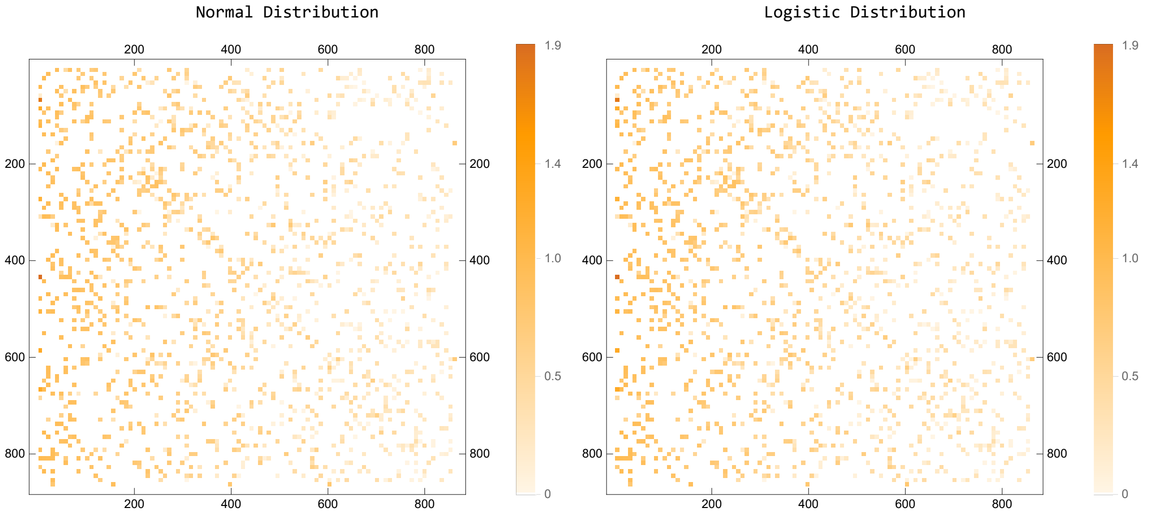

FIGURE 20. Distributed order communicability matrix of $NACE$ sector 5629 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

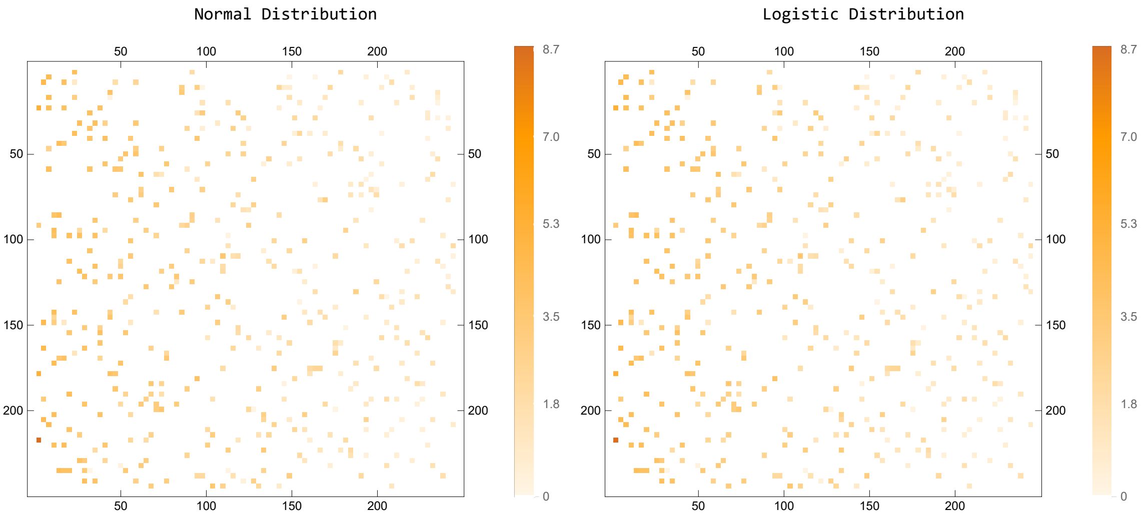

FIGURE 21. Distributed order communicability matrix of $NACE$ sector 8010 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

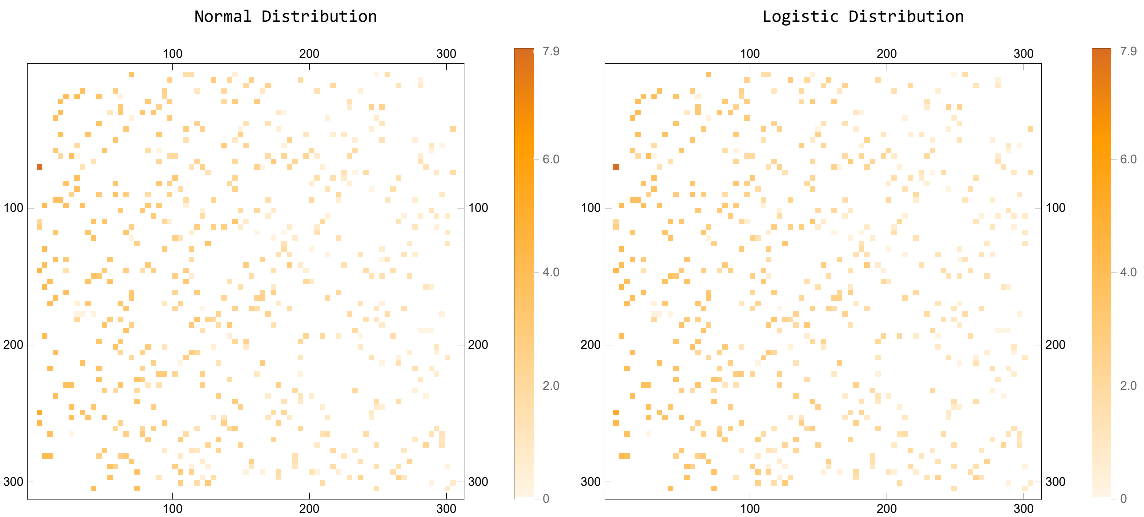

FIGURE 22. Distributed order communicability matrix of $NACE$ sector 8121 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

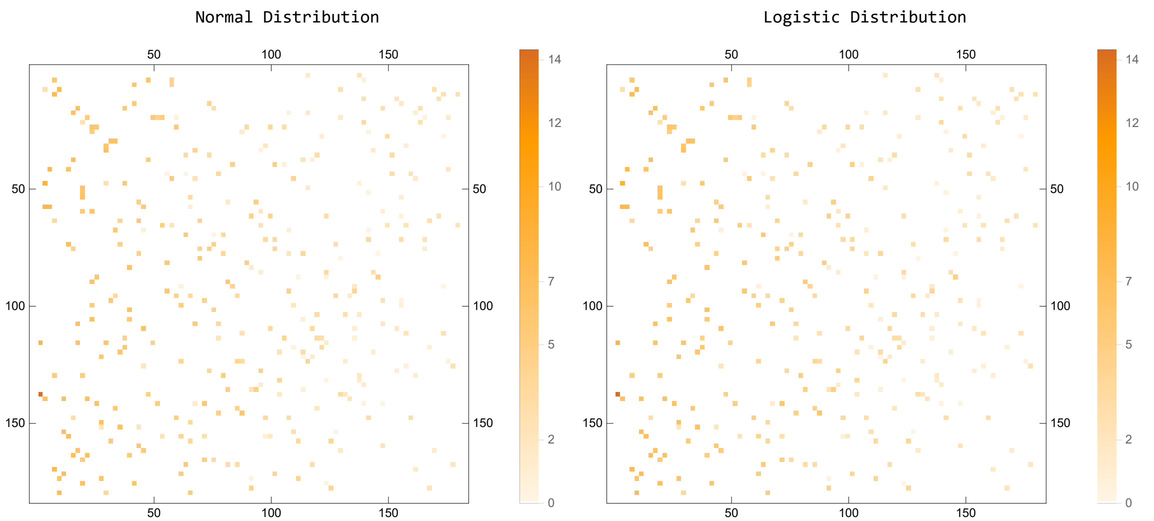

FIGURE 23. Distributed order communicability matrix of $NACE$ sector 8616 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

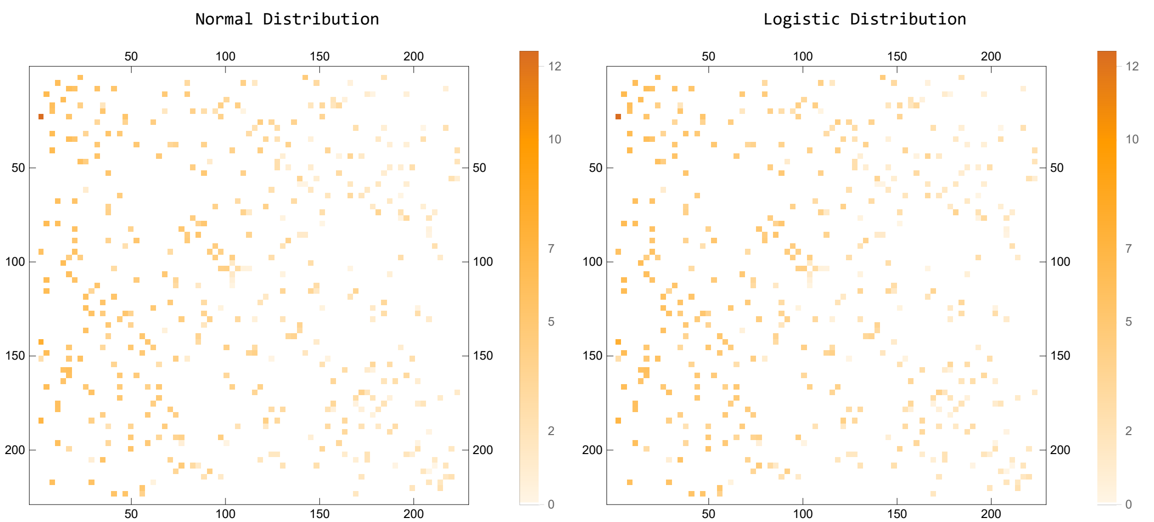

FIGURE 24. Distributed order communicability matrix of $NACE$ sector 9605 with respect to Normal and Logistic distributions with order $\alpha = 0.5$

4 Discussion and conclusion

For decision makers and policy makers working on occupational health and safety, how to improve workers' safety behavior is a focus. Analyzing the categorical data of workers who had an occupational accident by using effective mathematical methods other than conventional statistical methods yields very effective results. In this study, we present a novel mathematical method for the study of different sectors selected a significant sample of data in work-related accidents occurred in Turkey between the years 2013-2014 were used.

In order to obtain the network structure of workers who had an occupational accident using a categorical data set, firstly the data were parameterized. The hyper-network model, where each parameter corresponds to a hyper-edge, is established with a data set of 18 different sectors. Topological characterization of the created hyper-networks can be observed in hyper-edge degree distributions. The resulting hyper-edge degree distributions are quite similar to each other. The most striking differentiation in terms of hyper-edge degree distributions was observed to be the "Removal Sector" with 4942 $NACE$ code. When the data of this sector are examined, it is seen that it differs from the data of other sectors analyzed in the study as an occupational accident characteristic. Hypergraphs have several simple graph embeddings. The first embedding we use in this study is the bipartite graph showing the vertices where the hyper-edges match. When we look at the peak order distributions in the G_B bipartite representations of $H = (V, E)$ hyper-graphs created by different $NACE$ sectors, it is observed that the frequency of the peaks in communication with each other is high. This density indicates that the G_C graphs encoding the communication characterization of the vertices are complete graphs. Thus, the hypothesis that there are short links between the occupational accident information of workers working in a sector is realized.

One of the main contribution of this paper is to present a new link-prediction method based on distributed order fractional derivatives. Distributed order derivatives are fractional derivatives that have been integrated over the order of the derivative within a given range; that is, the memory of agents in a system is distributed. As explained before, the institutional regulations can be considered to be power-law fading memory within a social environment. Hence, as we are predicting link formation in MST filtration of G_C embedding, we assume that evolution of a random walker on MST is expressed with a type of fractional matrix equation. Furthermore, we give such memory to random walker to be distributed on MST, then we express such evolution as distributed order fractional matrix equation. Such distributions are chosen to be normal and logistic distributions for numerical analysis. The solution of such equation is performed by finite difference scheme. Since MST is a discrete space for the evolution domain, finite difference scheme is performed respect to the size of vertex set of MST. In numerical computations, we choose fractional order derivative to be $\alpha = 0.5$. For further studies, such order can be chosen in arbitrary and even varying values to also study diffusion characteristics of random walker. The main goal of link prediction process is to determine absent links in a graph. The distributed order measure we present shows the most possible edges tend to be included in MST. As the new edges are included, MST will turn to be G_C. Numerical results show that the highest tendencies are in "Collection of non-hazardous waste" sector with 3811 $NACE$ code, "Co operative store retail" sector with 4719 $NACE$ code, and "Retail Sale Of Electrical Household Appliances, Furniture, Lighting Equipment And Other Household Articles In Specialized Stores." sector with 4759 $NACE$ code for both normal and logistic distributions of the power-law fading kernel. Moreover, from Figures 5–24, it can be observed that the distribution of the power-law fading memory do not effect the link prediction values significantly.

Rather that the similar researches on social network analysis of occupational accidents, we analyze more than one sector. By introducing this new link prediction method based on distributed order fractional derivatives, this analyze can be performed on special sectors with bigger data for further studies.

Acknowledgement

This paper has been granted by the Muğla Sıtkı Koçman University Research Projects Coordination Office through Project Grant Number: (19/081/10/1)

References

[1] J. Takala, *Global estimates of fatal occupational accidents*, Epidemiology **10**(5) (1999) 640–646.

[2] B.Y. Jeong, *Characteristics of occupational accidents in the manufacturing industry of South Korea*, International journal of industrial ergonomics **20**(4) (1997) 301–306.

[3] C.W. Cheng, S.S. Leu, C.C. Lin, and C. Fan, *Characteristic analysis of occupational accidents at small construction enterprises*, Safety Science **48**(6) (2010) 698–707.

[4] K. Suzuki, T. Ohida, Y. Kaneita, E. Yokoyama, T. Miyake, S. Harano, Y. Yagi, E. Ibuka, A. Kaneko, T. Tsutsui and M. Uchiyama, *Mental health status, shift work, and occupational accidents among hospital nurses in Japan*, Journal of occupational health, **46**(6) (2004) 448–454.

[5] J. Pietilä, T. Räsänen, A. Reiman, H. Ratilainen and E. Helander,*Characteristics and determinants of recurrent occupational accidents*, Safety Science **108** (2018) 269–277.

[6] P. Kines, J. Lappalainen, K.L. Mikkelsen, E. Olsen, A. Pousette, J. Tharaldsen, K. Tmasson and M. Trner, *Nordic Safety Climate Questionnaire (NOSACQ-50): A new tool for diagnosing occupational safety climate*, International Journal of industrial Ergonomics **41**(6) (2011) 634–646.

[7] S. Sarkar, S. Vinay, R. Raj, J. Maiti and P. Mitra, *Application of optimized machine learning techniques for prediction of occupational accidents*, Computers & Operations Research **106** (2019) 210–224.

[8] S. Sarkar and J. Maiti, *Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis*, Safety Science **131** (2020) 104900.

[9] K. Kang and H. Ryu, *Predicting types of occupational accidents at construction sites in Korea using random forest model*, Safety Science **120** (2019) 226–236.

[10] L. Sanmiquel, M. Bascompta, J.M. Rossell, H.F. Anticoi and E. Guash, E., *Analysis of occupational accidents in underground and surface mining in Spain using data-mining techniques*, International journal of environmental research and public health **15**(3) (2018) 462.

[11] S. Gerassis, J.E. Martn, J.T. Garca, A. Saavedra, and J. Taboada, *Bayesian decision tool for the analysis of occupational accidents in the construction of embankments*, Journal of construction engineering and management **143**(2) (2017) 04016093.

[12] M. Shafique and M. Rafiq, *An overview of construction occupational accidents in Hong Kong: A recent trend and future perspectives*, Applied Sciences **9**(10) (2019) 2069.

[13] Q. Wang and Z.T. Gong, *Structural centrality in fuzzy social networks based on fuzzy hypergraph theory*, Computational and Mathematical Organization Theory **26**(2) (2020) 236–254.

[14] M. Sarwar, M. Akram and N.O. Alshehri, *A new method to decision-making with fuzzy competition hypergraphs*, Symmetry **10**(9) (2018) 404.

[15] G. Cattaneo, G. Chiaselotti, D. Ciucci and T. Gentile, *On the connection of hypergraph theory with formal concept analysis and rough set theory*, Information Sciences **330** (2016) 342–357.

[16] H.A. Nabwey, *A Methodology Based on Rough Set Theory and Hypergraph for the Prediction of Wart Treatment*, International Journal of Engineering Research and Technology **13**(3) (2020) 552–559.

[17] Ö. Akgüller, *Discrete Ricci curvature-based statistics for soft sets*, Soft Computing (2020) 1–14.

[18] K.B. Min, S.H. Song and J.Y. Min, *Topic Modeling of Social Networking Service Data on Occupational Accidents in Korea: Latent Dirichlet Allocation Analysis*, Journal of medical internet research **22**(8) (2020) e19222.

[19] H. Abbasianjahromi and A. Etemadi, *Applying social network analysis to identify the most effective persons according to their potential in causing accidents in construction projects*, International Journal of Construction Management **19**(3)(2019) 1–14.

[20] R. Alsamadani, M. Hallowell and A.N. Javernick-Will, *Measuring and modelling safety communication in small work crews in the US using social network analysis*, Construction management and economics **31**(6) (2013) 568–579.

[21] S. Li, P. Yang, X. Wu, G. Wang and M. Fan, *Unpacking the Black Box: Understanding the Effect of Social Network Characteristics on Safety Behaviors of Construction Workers* Complexity (2020) 8836682

[22] B. Pandit, A. Albert, Y. Patil and A. Al-Bayati, *Fostering safety communication among construction workers: Role of safety climate and crew-level cohesion*, International journal of environmental research and public health **16**(1) (2019) 71.

[23] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, *Hypergraph neural networks*, In Proceedings of the AAAI Conference on Artificial Intelligence **33** (2019, July) 3558–3565.

[24] F. Luo, L. Zhang, X. Zhou, T. Guo, Y. Cheng and T. Yi, *Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification*, IEEE Geoscience and Remote Sensing Letters **17**(6) (2019) 1082–1086.

[25] S. Zhang, Z. Ding and S. Cui, *Introducing hypergraph signal processing: theoretical foundation and practical applications*, IEEE Internet of Things Journal **7**(1) (2019) 639–660.

[26] M. Mao, J. Lu, J. Han and G. Zhang, *Multiobjective e-commerce recommendations based on hypergraph ranking*, Information Sciences **271** (2019) 269–287.

[27] A. Bretto, *Hypergraph theory. An introduction*, Mathematical Engineering Cham: Springer (2013)

[28] V. Zlatic, G. Ghoshal, and G. Caldarelli, *Hypergraph topological quantities for tagged social networks*, Physical Review E **80**(3) (2009) 036118.

[29] D. Li, Z. Xu, S. Li, and X. Sun, *Link prediction in social networks based on hypergraph*, In Proceedings of the 22nd International Conference on World Wide Web (2013, May) 41–42.

[30] H. Tao, Z. Li, Z. Wu, and J. Cao, *Link communities detection: an embedding method on the line hypergraph*, Neurocomputing **367** (2019) 46–54.

[31] S.E. Yoon, H. Song, K. Shin, and Y. Yi, *How Much and When Do We Need Higher-order Informationin Hypergraphs? A Case Study on Hyperedge Prediction*, In Proceedings of The Web Conference (2020, April) 2627–2633.

[32] M. A. Balcı, S.P. Atmaca and Ö. Akgüller, *Hyperpath Centers*, In Advanced Computational Methods for Knowledge Engineering (2016) 129–137.

[33] P.V. Marsden, *Egocentric and sociocentric measures of network centrality*, Social networks **24**(4) (2002) 407–422.

[34] S.P. Borgatti, *Centrality and network flow*, Social networks **27**(1) (2005) 55–71.

[35] N.E. Friedkin, *Theoretical foundations for centrality measures*, American journal of Sociology **96**(6) (1991) 1478–1504.

[36] E. Estrada and N. Hatano, *Communicability in complex networks*, Physical Review E **77**(3) (2008) 036111.

[37] E. Estrada, D.J. Higham and N. Hatano, *Communicability betweenness in complex networks*, Physica A: Statistical Mechanics and its Applications **388**(5) (2009) 764–774.

[38] J.T. Katsikadelis, *Numerical solution of distributed order fractional differential equations*, Journal of Computational Physics **259** (2014) 11–22.

[39] S. Mashayekhi and M. Razzaghi, *Numerical solution of distributed order fractional differential equations by hybrid functions*, Journal of Computational Physics **315** (2016) 169–181.

[40] G.H. Gao, H.W. Sun and Z.Z. Sun, *Some high-order difference schemes for the distributed-order differential equations*, Journal of Computational Physics **298** (2015) 337–359.

[41] J. Li, F. Liu, L. Feng and I. Turner, *A novel finite volume method for the Riesz space distributed-order advectiondiffusion equation*, Applied Mathematical Modelling **46** (2017) 536–553.

[42] X.Y. Li and B.Y. Wu, *A numerical method for solving distributed order diffusion equations*, Applied Mathematics Letters **53** (2016) 92–99.

[43] S.S. Ezz-Eldien and A.A. El-Kalaawy, *Numerical simulation and convergence analysis of fractional optimization problems with right-sided Caputo fractional derivative*, Journal of Computational and Nonlinear Dynamics **13**(1) (2018) 011010

[44] G. Tuna and M. Kurt, *Graph Communities to Analyze the Occupational Accidents: An Evidence from The Statistics of Turkey 20132014*, Gazi University Journal of Science **30**(4) (2017) 373–393.