Topological structure analysis in directed network

E. Klimkova, R. Senkerik, and I. Zelinka

Abstract — This paper is focused on the description, as to how to represent the network topology. It is very important to know the network topology and to understand its properties. This work describes how to find all the Giant Connected Component in directed network. The growing complex networks with preferential linking were used for experimental testing within this research. Roulette-wheel selection method was used as a preferential selection algorithm in the task of generation of complex networks.

Keywords— Complex network, Directed network, Giant connected component, Growing network, Tree.

I. INTRODUCTION

DIRECTED networks can be found in both nature and in man-made systems, i.e., network of citations of scientific papers, communication network, network of collaboration, telephone call graph, neural network, network of metabolic reaction etc.

It is important to understand to the topological structure of networks and its changes under external action [1 - 4]. Then, it is possible to understand, where the network is vulnerable to damage and when are resistant. The first random graphs were studied in the sixties [5, 9]. Later, studies about dynamic of growing networks [10 - 14] were performed. Nowadays, it is possible to study and easily simulate the real network [15 - 20]. For example, the model of small world networks contains many real networks such as the World Wide Web [21].

For the overview of directed complex networks, the giant connected components [3] are used. General structure of directed network, where the giant connected component is present, is depicted in Fig. 1. How to calculate the sizes of all giant connected components of a directed graph is described in [22].

In this paper, it is proposed, how to find all giant connected component. This work is an extension and continuation of previous research focused on Investigation on relations between complex networks and evolutionary algorithms dynamics [24].

Manuscript received June 29, 2011.

E. Klimkova is with the Department of Informatics and Artificial Intelligence, Faculty of Applied Informatics, Czech Republic, e-mail: klimkova@fai.utb.cz
R. Senkerik is with the Department of Informatics and Artificial Intelligence, Faculty of Applied Informatics, Czech Republic, email: senkerik@fai.utb.cz
I. Zelimka is with the Technical University in Ostrava, email: ivan.zelinka@vsb.cz

This paper is focused on the description, as to how to find all giant connected components in directed network by means of computer technology. Growing complex networks with preferential linking were used for testing [25]. Roulette-wheel selection was used as a preferential linking algorithm.

The structure of the paper is following: Firstly, the term Giant connected component is explained, and then a problem design is proposed. The following sections are focused on the description of used complex network and a visualization algorithm. Results and conclusion follow afterwards.

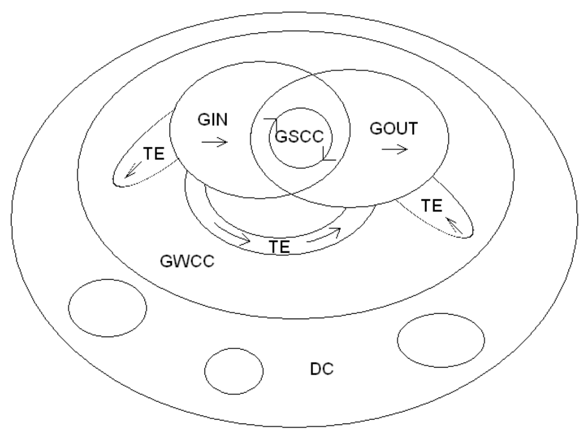

Fig. 1 General structure of directed network where the giant connected component is present.

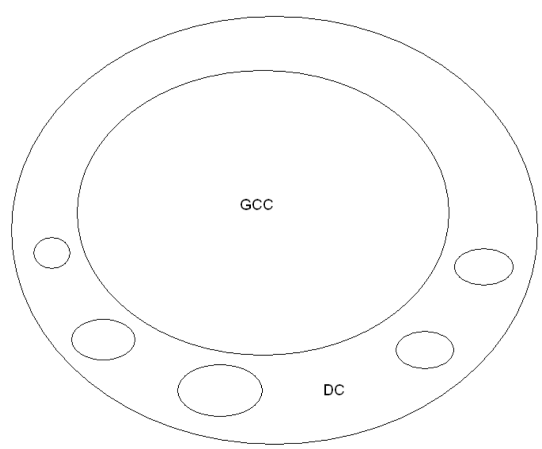

Fig.2 General structure of undirected network.

II. Experiment design

In undirected network, Giant Connected Component (GCC) and Disconnected Component (DC) are present Fig. 2.

However the structure of directed network can be more complex (See Fig. 1). In case that directness of edges is not present, the network consists of Giant Weakly Connected Component (GWCC) and Disconnected Components (DC).

After the projection of edges orientation, the GWCC is composed from the Giant Strongly Connected Component (GSCC), the Giant Out-Component (GOUT), the Giant In-Component (GIN) and the Tendrils (TE).

The giant strongly connected component is the set of vertices attached each by each with a directed path. The giant out-component is the set of vertices approachable from the GSCC by a directed path. The giant in-component contains all vertices from which the GSCC is approachable. The tendrils are the vertices which have no access to the GSCC and are not reachable from it. In particular, it indeed includes something like “tendrils” going out of GIN or coming in the GOUT but also there are “tubes” going from the GIN to GOUT without passage through GSCC and numerous clusters which are only “weakly” connected. Network where giant components are present, are depicted in Fig. 3.

Please note, that the definitions of the Giant In-connected Component and Giant Out-connected Component given in [25] differ from the new definitions presented in [22]. In the old definition, the Giant Strongly Connected Component is included into both GIN and GOUT, so the GSCC is the interception of the GIN and GOUT. The new definition was introduced for the sake of brevity and logical presentation.

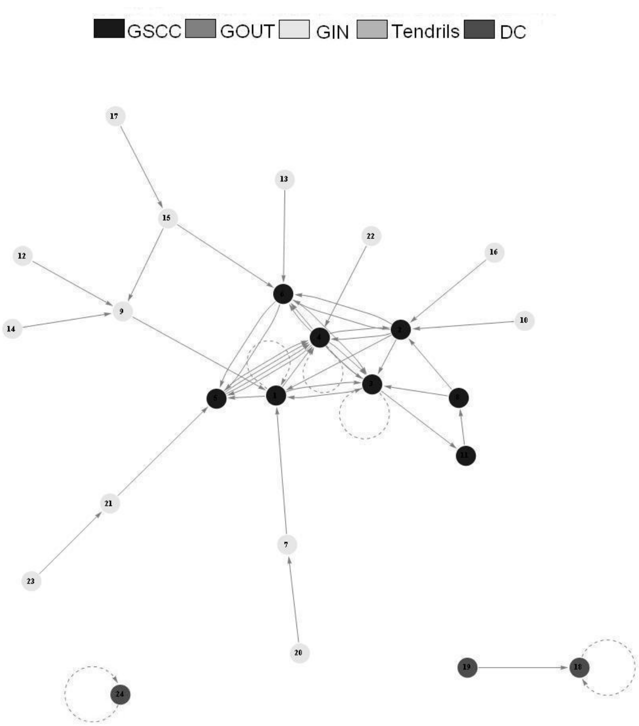

Fig. 3 Network where GCC is present. S – GSCC, I – GIN, O – GOUT, T – Tendrils, D – DC.

A. Numerical method

Number of all components can be derived by numerical method. Z – transform was used, for undirected network (1), and for directed network (2) [22].

$$\Phi(x) \equiv \sum_k P(k) x^k \tag{1}$$

$$\Phi(x,y) \equiv \sum_{k_i k_o} P(k_i, k_o) x^{k_i} y^{k_o} \tag{2}$$

Where k is total number of connection, k_i and k_o is incoming and outgoing connection, obviously $k = k_i + k_o$. Degree distribution $P(k)$ can be derived from $P(k) \equiv P^{(w)}(k) = \sum_{ki} (k_i, k-k_i)$ [6]. If all edges are within the considered network, average in and out degree distribution is based on $\partial_x \Phi(1,y) \mid_{x=1} = \partial_y \Phi(y,1) \mid_{y=1} \equiv z^{(d)}$ Accordingly average degree distribution take the form $z = 2z^{(d)}$ Degree distribution undirected network in Z-transformation is equal $\Phi^{(w)}(x) = \Phi(x,x)$. Of that relationship is derived pattern $\Phi_1^{(w)}(x) = \Phi^{(w)\prime}(x)/z$. GWCC is determined if the relationship is valid $\Phi_1^{(w)\prime}(1) > 1$ [22].That is similar like Molloy and Reed criterium (3) [7, 8]

$$\sum_k k(k-2)P(k) > 1. \tag{3}$$

Number of Giant weekly connect component W is given by (4) and (5) [22]

$$W = 1 - \Phi^{(w)}(t_c) \tag{4}$$

$$x_c = \Phi_1^{(w)}(t_c)\,. \tag{5}$$

Number of Giant In-component and Giant Out-component can be derived like the previous form. Z-transformacion of out-degree distribution is equal $\Phi_1^{(o)}(y) \equiv \partial_x \Phi(x,y) \mid_{x=1} / z(d)$. In-degree distribution is obtain similarly $\Phi_1^{(i)}(x) \equiv \partial_y \Phi(x,y) \mid_{y=1} / z(d)$. The GIN and GOUT exist if $\Phi_1^{(i)\prime}(1) = \Phi_1^{(o)\prime}(1) = \partial^2_{xy} \Phi(x,y) \mid_{x=1,y=1} / z(d). > 1$. This implies (6) [22]

$$\begin{aligned} &\sum_{k_i k_o} (2k_i k_o - k_i - k_o) P(k_i, k_o) = \\ &2\sum_{k_i k_o} k_i (k_o - 1) P(k_i, k_o) = 2\sum_{k_i k_o} k_o (k_i - 1) P(k_i, k_o). \end{aligned} \tag{6}$$

Then the equations are derived (7)

$$\begin{aligned} x_c &= \Phi_1^{(i)}(x_c) \\ y_c &= \Phi_1^{(o)}(y_c). \end{aligned} \tag{7}$$

Relative size of Giant In-connected Component and Giant Out-Connected Component is given by (8)

$$I = 1 - \Phi(x_c, 1)$$
$$O = 1 - \Phi(1, y_c). \tag{8}$$

Relative size of Giant Strongly Connected Component can be derived from (7), similarly Giant In-component and Giant Out-component [6]. If incoming edges k_i and outgoing edges k_o are statistically independent then the probability that all incoming edges come from finite in-components is equal to $x_c^{k_i}$. Then $1 - x_c^{k_i}$ is the probability that vertex has the infinite in component. Probability that the vertex has the infinite out-component is equal to $1 - y_c^{k_o}$. If in and out-component are infinite, probability that the vertex belong to GSCC is $(1 - x_c^{k_i})(1 - y_c^{k_o})$. Then the probability, that the vertex is in GSCC, is given by (9) [15]

$$\sum_{k_i k_o} (1 - x_c^{k_i})(1 - y_c^{k_o}) P(k_i, k_o) \,. \tag{9}$$

Relative size of Giant strongly connected component is equal to (10) [15]

$$S = \sum_{k_i k_o} (1 - x_c^{k_i})(1 - y_c^{k_o}) P(k_i, k_o) =$$
$$1 - \Phi(x_c, 1) - \Phi(1, y_c) + \Phi(x_c, y_c). \tag{10}$$

If the relative size of *W*, *S*, *I* and *O* component is known, Relative size of Tendrils is (11) [15]

$$T = W + S - I - O. \tag{11}$$

When joint distribution of in-and out-degrees induce, $P(ki, ko) = P^{(i)}(k_i) P^{(o)}(k_o)$ (8), then *S* is equal *IO*. In another case such as the factorization S is impossible. Then, $x_c = y_c = 1$ and *I*, *O*, and *S* concurrently approaches zero [22].

B. Description of used method

GSCC is very important part of the network. In general, it is the core of directed network. The challenging task is how the GSCC can be identified by means of computer technology.

One possible way is to transform the network into a tree. Within this approach, one vertex is selected as the root of tree and links are transformed into routes (See Fig. 4). From the first vertex leads a lot of routes in the case of huge network, therefore some insignificant edges may be omitted. GSCC can be presented as a loop; therefore the example of insignificant route is the 2→4. In this case, GSCC include vertexes 2,3,5,1, thus the routes 2→3→1→2 and 2→1→2 may be omitted. For the detailed description of developed algorithm, please refer to pseudo-code depicted in Fig. 5.

In addition, several loops may be found in larger networks. If there is a connection between several loops, GSCC includes these loops otherwise larger loop is intended as the GSCC.

If the GSCC is addressed, other components can be found. Vertexes belonging into the GIN must be linked to the core of the network and there exists the road from the core to GOUT. Vertexes belonging to the GWCC and does not belong to the above three categories, are Tendrils.

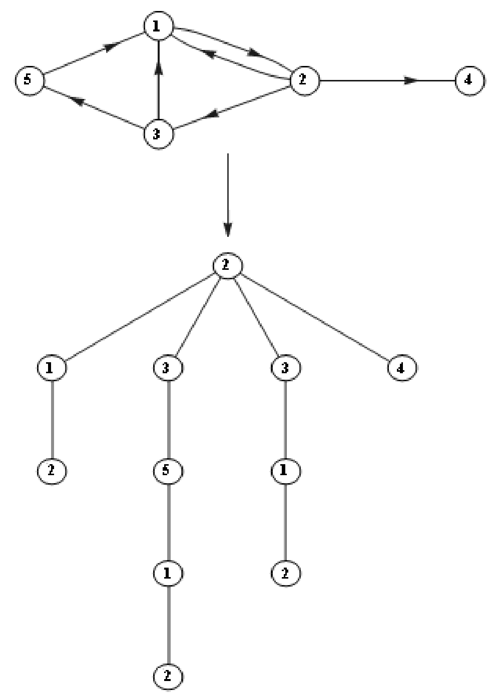

Fig. 4 Network with loop transferred into tree.

C. Generation of complex networks examples

For the testing of the developed algorithm, the growing complex networks with preferential linking were used. New vertex connections were chosen by means of Roulette-wheel selection. Developed algorithm was tested on the networks with five to eighty vertexes.

III. Experimental Results

This paper consists primarily of two illustrative case studies focused on the detailed visualization of complex networks by means of Giant Connected Component.

The experiment was repeated twenty times for each dimension of the network to confirm the robustness and efficiency of developed algorithm. For the experiment, desktop PC with single-core, 1.81 GHz CPU and 2 GB RAM was used.

The results of average number of Giant connected Component are presented in Table I. Fig. 6 – 8 contains detailed graphical analysis of the results presented in Table I. Average number of all Giant Connected Components is present in Fig. 6, where is clearly visible which components are growing and which not. In Fig. 7, there is depicted the average numbers of Giant Out-connected Component, Tendrils and Disconnected Component in percentage in special graph, due to the very low appearance of theses component in the network. Other components (GSCC GIN and GWCC) are present in special graph depicted in Fig. 8.

Examples of two selected networks are shown in Fig. 10 – 11 and 13 – 14. The first case study was a small network with 25 vertexes, and the second case study was a large network with 75 vertexes.

```
AdMat = Import[Adjacency Matrix]
GMat ← AdMat
index ← Dimensions [GMat]
Delete multiple edges [GMat]
Delete loop to itself [GMat]

in ← vertexes with only in edges
out ← vertexes with only out edges
Delete in [index]
Delete out [index]  * index have vertices with both
in and out edges *

* make tree *
while index
          find neighbor index[i]
          tree [x] ← {index[i], neighbor}
                         *tree = {vertex, parent}*
          i ++

* walking the tree*
t ← tree[1]

Procedure Walking [t] (
     If  t [1]= tree [i,2]
            then stack [x] ← {vertex, depth,  parent}
                 queue [x] ← {vertex, depth}
          t ← First [queue]
     If depth > Length[index] or Length[queue] = 0
                   then End
                   else call Procedure Walking [t])

* find loop *
root ← tree[1]
parent ← root
loop ← root[1]
while h < Last[stack[2]]
          If stack[i] = {h, parent}
                   then child [x] ← stack [i]
          loop[y] ← child [x]
          If Last[loop] = root[1]
                   then loopA ← loop

gscc ← loopA
```

Fig. 5 Pseudo-code of developed algorithm.

Table I Numbers of Giant Strongly Connected Component, Giant In-component, Giant Out-component, Tendrils, Disconnected Component and Giant weakly connected Component.

N	GSCC	GIN	GOUT	TE	DC	GWCC
5	2,72	1,64	0,00	0,13	0,50	4,50
10	3,97	4,19	0,38	0,73	0,73	9,27
15	5,37	7,61	0,45	0,76	0,81	14,19
20	6,79	11,54	0,42	0,25	1,00	19,00
25	8,28	15,09	0,39	0,30	0,95	24,05
30	10,01	18,35	0,40	0,18	1,06	28,94
35	11,71	21,63	0,22	0,18	1,26	33,74
40	13,00	25,00	0,23	0,27	1,50	38,50
45	14,00	28,98	0,23	0,55	1,24	43,76
50	15,51	32,32	0,33	0,88	0,97	49,03
55	18,88	33,02	0,56	0,68	1,86	53,14
60	20,35	36,10	0,65	1,03	1,88	58,13
65	21,34	40,13	0,57	1,38	1,59	64,71
70	22,94	44,51	0,33	0,33	1,89	68,11
75	26,42	47,94	0,32	0,00	0,32	74,69
80	28,02	49,58	0,91	0,54	0,95	79,05

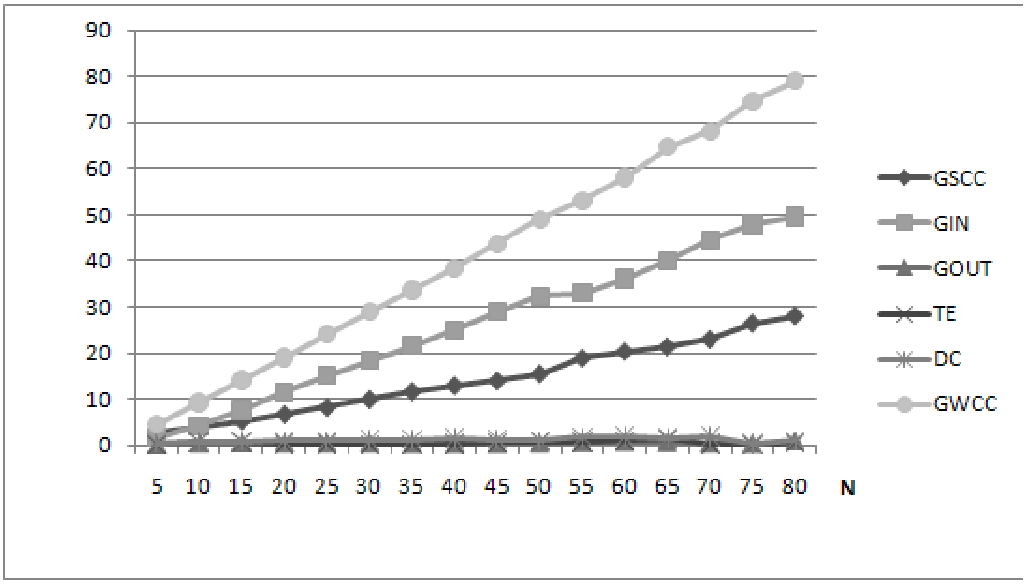

Fig. 6 Average number of Giant Connect Component in studied network.

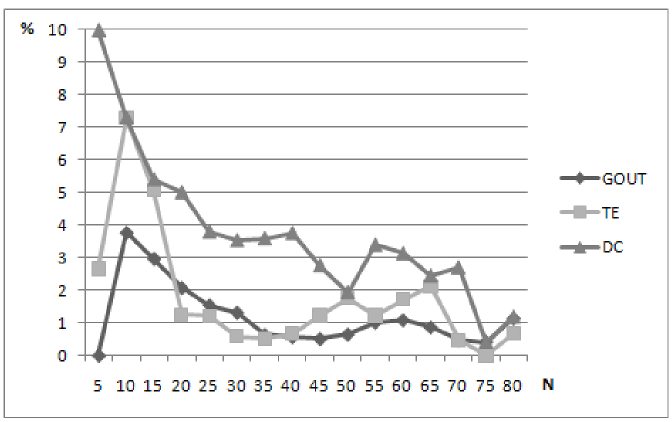

Fig. 7 Average numbers of GOUT, Tendrils and DC in percentage.

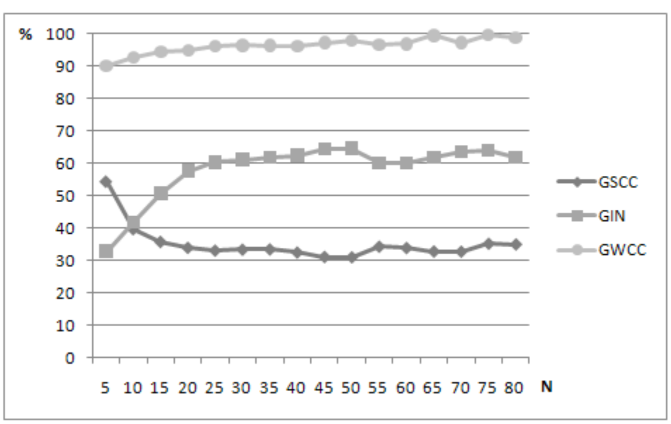

Fig. 8 Average numbers of GSCC, GIN and GOUT in percentage.

A. Case study 1

In this case study, the detailed description and analyze for the selected example of the small network with 25 vertexes were performed. Histogram of Giant Strongly Connected Component distribution is shown in Fig. 9. For the structure of the used network, please refer to Fig. 10. 3D model of network is present in Fig. 11. These examples represent the graphical output of the algorithm described above. In the small growing networks there occur more GOUT, TE and DE component, than in large networks. Giant Strongly Connect Component has on average 41 % of the total vertices in the network. Table II shows the average values of all the Giant Connected Component. From these results, it is obvious, that typically the GSCC and GIN are present in the network, and the percentage occurrence of these two components together is 88%. Other components appear in the network with a low probability.

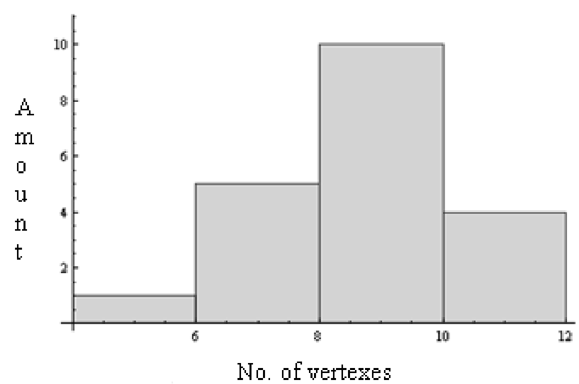

Fig. 9. Histogram of number of GSCC in the network with 25 vertexes.

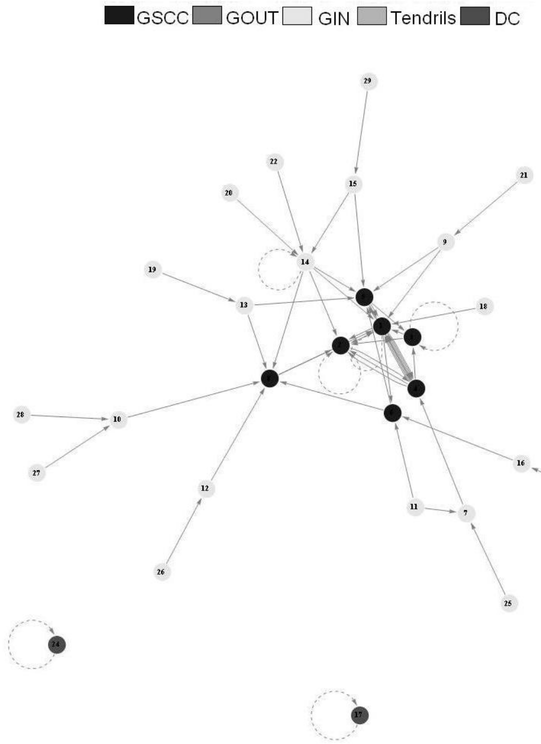

Fig. 10 Example of the small network, where Giant Strongly Connected Component, Giant In-component, Giant Out-component, Tendrils and Disconnected Component are present.

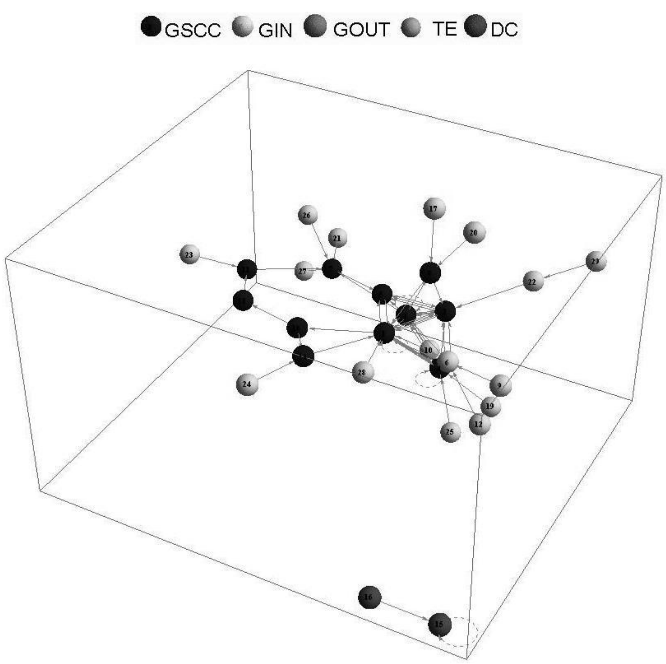

Fig. 11 3D model of example of the small network, where Giant Strongly Connected Component, Giant In-component, Giant Out-component, Tendrils and Disconnected Component are present.

B. *Case study 2*

Networks with seventy five vertexes were analyzed in this case study. Histogram of Giant Strongly Connected Component distribution is shown in Fig. 12. Graphical example of the network is in Fig. 13. 3D model of network is present in Fig. 14. These examples represent the graphical output of the algorithm described above. It is obvious, that there are many Giant Strongly Connect Components and Giant In-components present in the large network. Together these two components represent 97.31% of the network. Other components are present in the network with a very low probability. Further detailed information is in Table III.

Table II. Average values of the Giant Connected Component

N=25	Average	%
GSCC	8.15	33.96
GIN	13.85	57.71
GOUT	0.50	2.08
TE	0.30	1.25
DC	1.20	5.00
GWCC	22.80	95.00

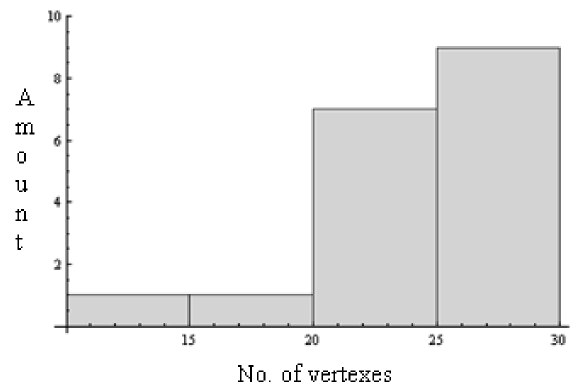

Fig. 12 Histogram of number of GSCC in the network with 75 vertexes.

Table III. Average values of the Giant Connected Component

N=75	Average	%
GSCC	24.25	32.77
GIN	47.05	63.58
GOUT	0.35	0.47
TE	0.35	0.47
DC	2.0	2.70
GWCC	72.0	97.30

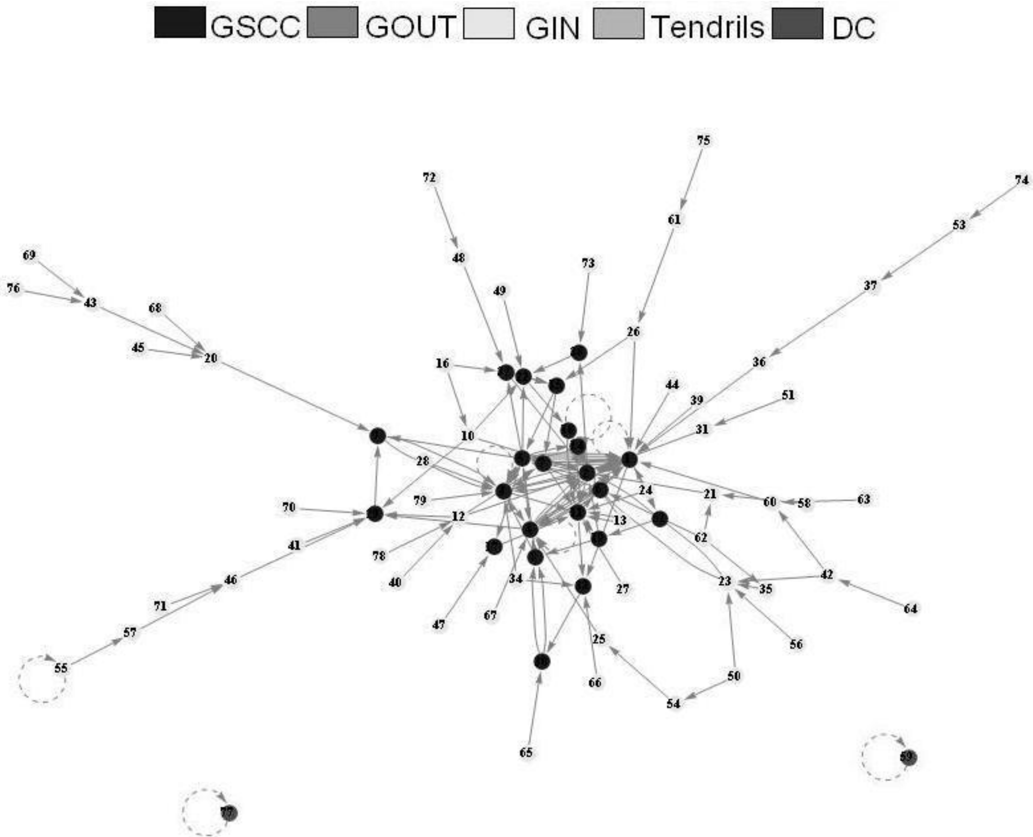

Fig. 13 Example of the large network, where all Giant Connected Component are present.

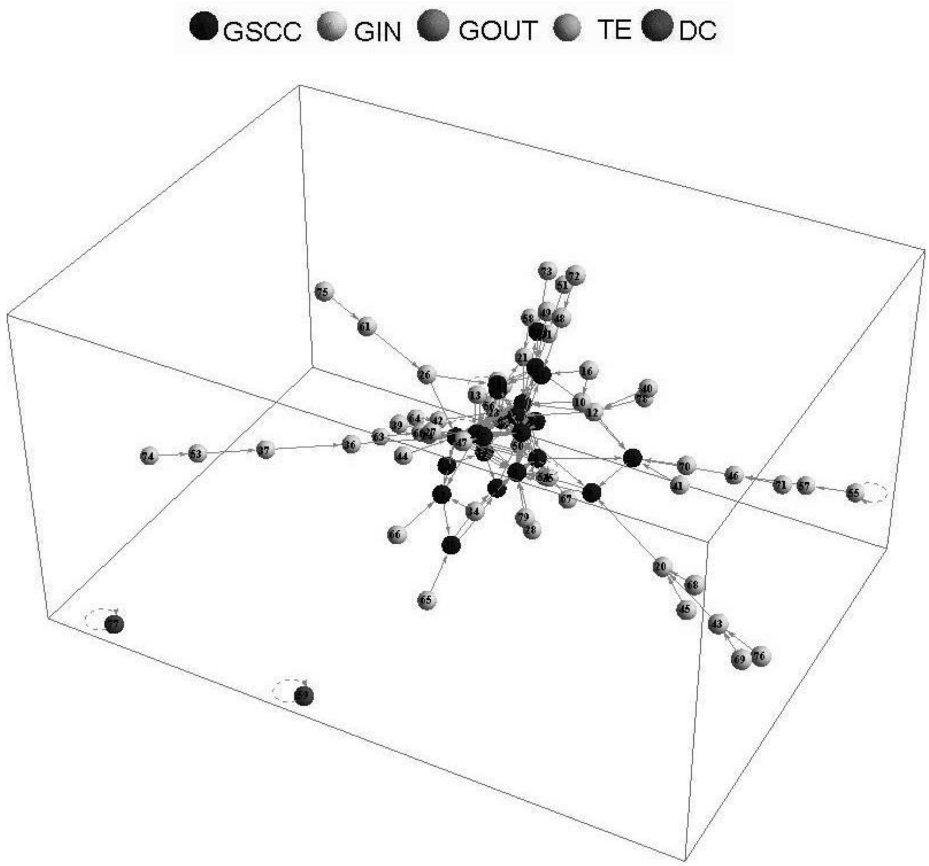

Fig. 14 3D model of example of the large network, where all Giant Connected Component are present.

I. Conclusion

This paper deals with the development of the algorithm for the visualization of all Giant Connected Components in the network. It was explained, how all the GCC can be calculated by means of mathematical methods. In this paper the method as to how to determine the number of GCC and how to graphically illustrate them by means of computer technology was described. The experiment was repeated twenty times for each dimension of the network to confirm the robustness and efficiency of developed algorithm.

For the experiments a small network with 25 vertexes and a large network with 75 vertexes were used.

As demonstrated, this method is very simple to implement and very easy to use. Furthermore, importance of this research is growing every day. Complex networks can be found in many scientific fields, but also in nature. It is important to understand the structure of networks, especially where the networks are prone to faults.

Acknowledgment

This work was supported by the grants: Internal Grant Agency of Tomas Bata University under the project No. IGA/40/FAI/11/D, grant of Ministry of Education of the Czech Republic MSM 7088352101, grant of Grant Agency of Czech Republic GACR 102/09/1680 and by European Regional Development Fund under the project CEBIA-Tech No. CZ.1.05/2.1.00/03.0089.

References

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, „Complex Networks: Structure and Dynamics“, *Physics Reports*, Vol. 424, 2006, pp. 175-308.

[2] M.E.J. Newman, "The structure and function of complex networks", *SIAM Review*, vol. 45, no. 2, 2003, pp. 167-256.

[3] S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks*, Oxrofd University press, 2003

[4] S. Bornholdt, H. G. Shuster, *Handbook of Graphs and Networks,* WILEY-VCH, 2003.

[5] P. Erdos, A. Rényi, "On the strength of connectedness of a random graph", *Acta Mathematica Academiae Scientiarum Hungaricae*, vol. 12, no. 1-2, 1961, pp. 261-267.

[6] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 64, no. 2 II, 2001, pp. 261181-261187.

[7] M.Molloy, B. Reed,"A Critical Point for Random Graphs with a Given Degree Sequence". *Random Structures and Algorithms* vol. 6, 1995, pp.161-180.

[8] M. Molloy, B. Reed, "The Size of the Largest Component of a Random Graph on a Fixed Degree Sequence". *Combinatorics, Probability and Computing*, vol. 7, 1998, pp. 295-306.

[9] N. Ikeda, "Control of Network Structure by an External Field on Random Walkers", *6th WSEAS International Conference on Non-Linear Analysis*, 2007.

[10] J. Mendes, A. Samukhin, S. Dorogovtsev, "Anomalous Percolating Properties of Growing Networks", *Phys. Rev. E 64*, 066110, 2001.

[11] P. Bialas, & A. K. Oleś, "Correlations in connected random graphs", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, no. 3, 2008.

[12] P.L. Krapivsky, G.J. Rodgers, & S.Redner, , "Degree distributions of growing networks", *Physical Review Letters*, vol. 86, no. 23, 2001, pp. 5401-5404.

[13] P.L. Krapivsky, & S. Redner, "Organization of growing random networks", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 63, no. 6 II, 2001, pp. 066123/1-066123/14.

[14] P. Adesso, A. Gravetti, M. Lohgo, F. Postiglione, "The Influence of Network Topological Models on the Prediction of End-to-End Loss Probabilities", *5th WSEAS Int. Conf. on Multimedia*, 2005

[15] A. Barabási, & R. Albert, "Emergence of scaling in random networks", *Science*, vol. 286, no. 5439, 1999, pp. 509-512.

[16] R. Albert, H. Jeong, & A. Barabási, "Error and attack tolerance of complex networks", *Nature*, vol. 406, no. 6794, 2000, pp. 378-382.

[17] R. Albert, H. Jeong, A.-L. Barabási, "Diameter of the world wide web " *Nature* vol. 401, 1999, pp. 130-131.

[18] S. Mocanu, S. Taralunga, "Immunization Strategis for Networks with Scale-Free Topology", *5th WSEAS Int. Conf. on Non-Linear Analysis*, 2006.

[19] R. Dobrescu, "Modeling complex biological systems using Scale Free Networks", *5th WSEAS Int. Conf. on Non-Linear Analysis*, 2006.

[20] S.N. Dorogovtsev, & J.F.F. Mendes, "Scaling properties of scale-free evolving networks: Continuous approach", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 63, no. 5 II, 2001, pp. 561251-5612519.

[21] D. J. Watts, S. H. Strogatz, "Collective Dynamics of 'Small-world' Networks", *Nature*, vol. 393 Jun. 1998, pp. 440-442.

[22] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, "Giant strongly connected component of directed networks", *Physical Review E*, Vol.64, No.2, 2001, pp. 25101-25104.

[23] C. Cooper, A. Frieze, "The size of the largest strongly connected component of a random digraph with a given degree sequence", *Combinatorics Probability and Computing*, vol. 13, no. 3, 2004, pp. 319-337.

[24] I. Zelinka, D. Davendra, V. Snasel, R. Jasek, R. Senkerik, Z. Oplatkova, "Preliminary Investigation on Relations Between Complex Networks and Evolutionary Algorithms Dynamics", *proceedings of CISIM 2010*, Poland, Cracow, October 8 – 10, 2010.

[25] G. Ergün, & G.J. Rodgers, "Growing random networks with fitness", *Physica A*: *Statistical Mechanics and its Applications*, vol. 303, no. 1-2, 2002, pp. 261-272.

E. Klimkova was born in Czech Republic, and went to the Tomas Bata University in Zlin in Czech Republic, where she studied telecommunication systems and obtained her MSc. degree in 2010.
She is a Ph.D. student at the same university. Her research is concerned in evolution dynamic of complex network. Her e-mail address is: klimkova@fai.utb.cz

R. Senkerik was born in the Czech Republic, and went to the Tomas Bata University in Zlin, where he studied Technical Cybernetics and obtained his MSc degree in 2004 and Ph.D. degree in Technical Cybernetics in 2008.
He is now a lecturer at the same university (Applied Informatics, Cryptology, Artificial Intelligence, Mathematical Informatics). His email address is: senkerik@fai.utb.cz

I. Zelinka was born in the Czech Republic, and went to the Technical University of Brno, where he studied Technical Cybernetics and obtained his degree in 1995. He obtained Ph.D. degree in Technical Cybernetics in 2001 at Tomas Bata University in Zlin.
He is now a Professor at the Technical University in Ostrava, Czech Republic. His specialization is artificial intelligence.
Email address: ivan.zelinka@vsb.cz .