Visualization of giant connected component in directed network - Preliminary study

EVA KLIMKOVA, ROMAN SENKERIK, IVAN ZELINKA, TOMAS SYSALA
Tomas Bata University in Zlin, Faculty of Applied Informatics
Nam. T.G.Masaryka 5555, 760 01 Zlín
CZECH REPUBLIC
klimkova@fai.utb.cz, senkerik@fai.utb.cz, zelinka@fai.utb.cz, sysala@fai.utb.cz

Abstract: -This paper is focused on the description as to how to find all giant connected components in directed network by means of computer technology. For experimental testing, the growing complex networks with preferential linking were used within this research. Roulette-wheel selection method was used as a preferential selection algorithm in the task of generation of complex networks.

Key-Words: - Complex network, Directed network, Giant connected component, Tree, Growing network.

1 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]. Then, it is possible to understand, where the network is vulnerable to damage and when are resistant.

For the overview of directed complex networks, the giant connected components [2] 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 descried in [3].

In this paper, it is proposed, how to find all giant connected component.

This work is an extension and continuation of previous research [4] focused on Investigation on relations between complex networks and evolutionary algorithms dynamics.

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 [5]. 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.

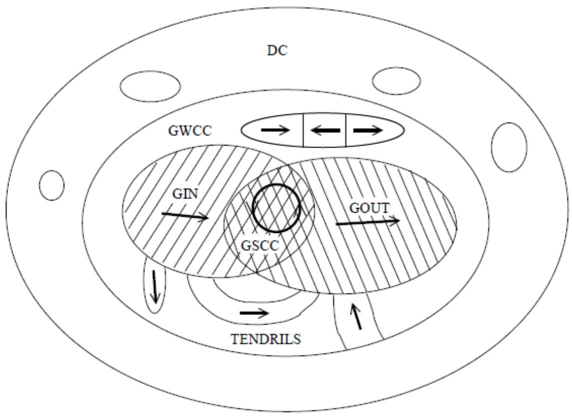

Fig. 1 General structure of directed network where the giant connected component is present.

2 Experiment design

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

However the structure of directed network can be more complex. 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.

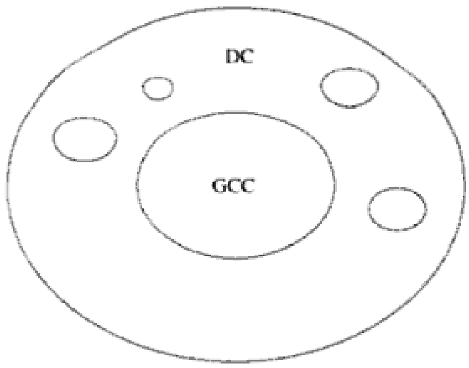

Fig. 2 General structure of undirected network.

Network where giant components are present, are depicted in Figures 6 and 8.
Note that the definitions of the GIN and GOUT in the [6], [7] differ from the new definitions presented in [3]. In the old definition, the GSCC 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.

2.1 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. 3). 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 in Fig. 5 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.
The second possible way is based on simple fact that the GSCC may not consist of only one loop. 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.

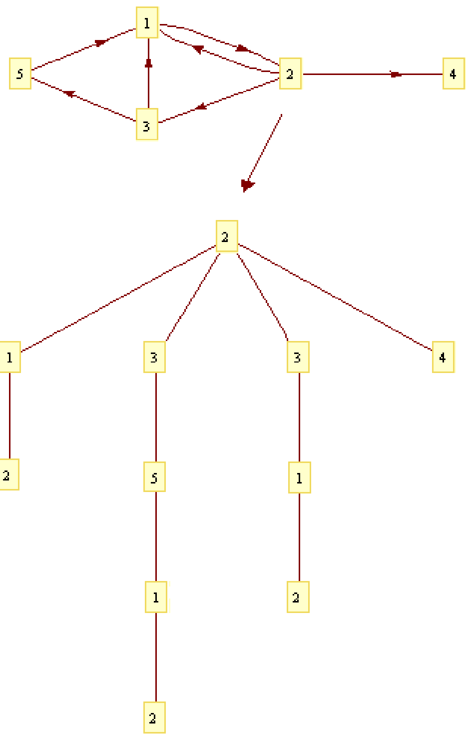

Fig. 3 Network with loop transferred into tree.

2.2 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.

3 Experimental Results

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

ISBN: 978-1-61804-004-6

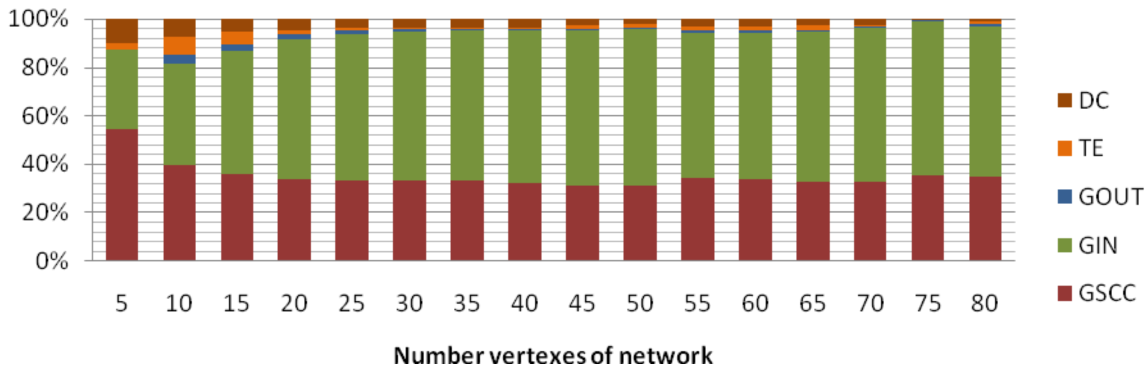

Fig. 4 The average number of all giant connected components

Table 1 Average numbers of all GCC (in percent)

N	GSCC	GIN	GOUT	TE	DC	GWCC
5	54.48	32.89	0.00	2.66	9.97	90.03
10	39.72	41.92	3.78	7.29	7.29	92.71
15	35.79	50.76	2.98	5.07	5.40	94.60
20	33.96	57.71	2.08	1.25	5.00	95.00
25	33.10	60.34	1.55	1.21	3.79	96.21
30	33.38	61.18	1.32	0.59	3.53	96.47
35	33.46	61.79	0.64	0.51	3.59	96.41
40	32.50	62.50	0.57	0.68	3.75	96.25
45	31.12	64.39	0.51	1.22	2.76	97.24
50	31.02	64.63	0.65	1.76	1.94	98.06
55	34.32	60.04	1.02	1.23	3.39	96.61
60	33.91	60.16	1.09	1.72	3.13	96.88
65	32.83	61.74	0.87	2.12	2.45	99.55
70	32.77	63.58	0.47	0.47	2.70	97.30
75	35.23	63.92	0.42	0.00	0.42	99.58
80	35.02	61.97	1.14	0.68	1.19	98.81

The results of average number of GCC (percentage) are presented in Table 1 and Fig. 4. Examples of two selected networks are shown in Figures 6 and 8. Two illustrative examples were chosen. The first one was a small network with 25 vertexes, and the second one was a large network with 75 vertexes.

3.1 Case study 1

In this case study, the detailed description and analyse for the selected example of the small network with 25 vertexes were done. For the structure of the used network, please refer to Fig. 6. Table 2 shows the average values of the GCC and histogram of GSCC distribution is shown in Fig. 5.

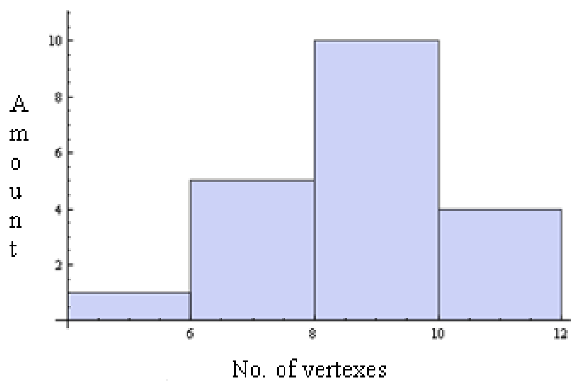

Fig. 5. Histogram of number of GSCC in the network with 25 vertexes.

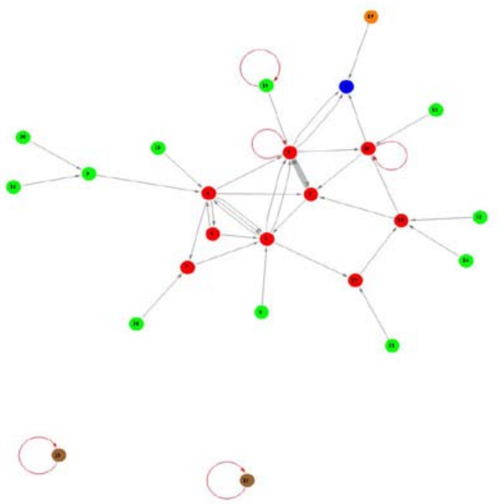

Fig. 6 Example of the small network, where GSCC (red), GIN (green) GOUT (blue), TE (orange) and DC (brown) are present.

Table 2. Average values of the GCC

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

3.2 Case study 2

Networks with seventy five vertexes were analyzed in this case study. Graphical example of the network is in Fig. 8. The average numbers of all giant connected components are present in Table 3. Histogram of GSCC distribution is shown in Fig. 7.

Table 3. Average values of the GCC

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

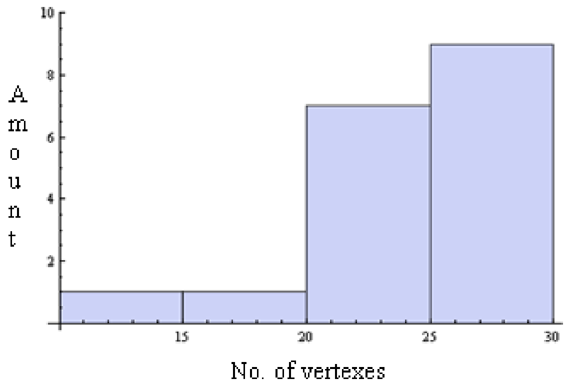

Fig. 7 Histogram of number of GSCC in the network with 75 vertexes.

4 Conclusion

This paper deals with the development of the algorithm for the visualization of all Giant Connect Components in the network. 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.

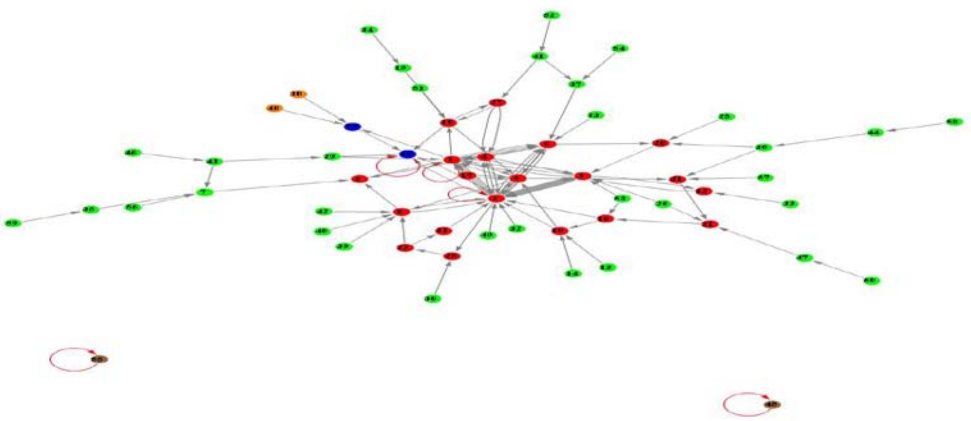

Fig. 8 Selected example of the large network, where GSCC (red), GIN (green) GOUT (blue), TE (orange) and DC (brown) are present.

ISBN: 978-1-61804-004-6

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

[3] 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.

[4] 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.

[5] G. Ergun and G. J. Rodgers, Growing Random Networks with Fitness, *Physica A*, Vol. 303, 2002, pp. 261-272.

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

[7] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications, *Physical Review E*, Vol. 64, No.2, 2001, pp. 26118-26134.