



# NCGs: Building a Trustworthy Environment to Identify Abnormal Events Based on Network Connection Behavior Analysis

Hangyu Hu<sup>(✉)</sup>, Xuemeng Zhai, Mingda Wang, and Guangmin Hu

School of Information and Communication Engineering,  
University of Electronic Science and Technology of China,  
Chengdu 611731, China  
huhangyuuestc@gmail.com

**Abstract.** With the continuous development and wide application of various network technologies, such as the mobile, wireless and sensors network, network services are becoming more and more high-speed, diversified and complex. Also, network attacks and infrequent events have emerged, making the promotion of network anomaly detection more and more significant. In order to control and manage the networks and establish a credible network environment, it is critical to facilitate an accurate behavioral characteristic analysis for networks, proactively identify abnormal events associated with network behavior, and improve the capacity of responding to abnormal events. In this paper, we use Network Connection Graphs (NCGs) to model flow activities during network operation. After we construct a NCG in a time-bin, then we can extract graph metric features for quantitative or semi-quantitative analysis of flow activities. And we also could build a series of NCGs to describe the evolution process of network operation. During these NCGs, we have conducted dynamic analysis to find out the outlier points of graph metric features by using Z-score analysis method so that we can detect the hidden abnormal events. The experiment results based on real network traces have demonstrated that the effectiveness of our method in network flow behavior analysis and abnormal event identification.

**Keywords:** Network Connection Graphs · Graph metrics · Dynamic analysis · Outlier detection · Event identification

## 1 Introduction

Recent years have witnessed the rapid improvement of various network technologies, such as the mobile, wireless and sensors network. These network technologies aim to connect everything in the cyber world and allow everything interacts with each other [1–3]. Since the advantages of these different network types, it is attracting tremendous attention from the academia, industry and government to carry on researches about network behavior analysis. However, with the increasing scale and complexity of the Internet and increasing frequency of network abnormal events, the volume, velocity and variety of data traffic travelling on network rises at an exponential rate; anomalous

network events have the feature of erupting suddenly without known signatures, which could cause great catastrophe, thus network behavior analysis is still a problem which is important and urgently needed to be intensively explored.

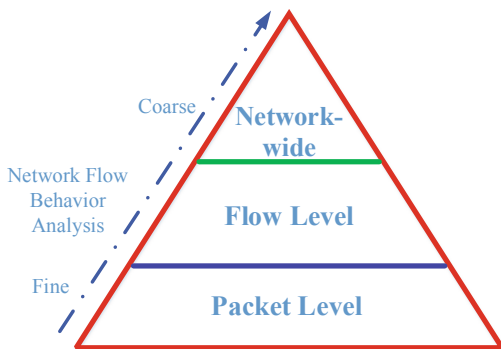
The occurrence of network abnormal events originate by many reasons, such as bad operation of network devices, network operation errors, network intrusions (such as Denial of Service (DoS) attacks, flash crowds, port scanning and worm propagation), which will damage the normal functions of the network [4–7]. Therefore, the identification of network abnormal event is becoming an indispensable and essential part of network behavior analysis. In order to control and manage the network successfully and build a trustworthy network environment, it is necessary to analyze, extract and identify the connection relationship of the flow activities in real time and accurately, actively discover the type and abnormal behavior of the flow, and improve the ability of the network system to cope with the anomalous behavior of the network.

In our recent work, we have studied the problem of precisely identifying network flow behaviors and discovering the root cause of network events among a great deal of network flow data [8]. This paper is a continuation of recent work, by drawing on the concept of knowledge graph. And we propose Network Connection Graphs (NCGs), a new method used for network flow measurement, analysis and visualization over time. Firstly we collect network traffic and preprocess flow information in real time in order to construct data model for NCGs. Depending on the purpose of study, various rules can be used to determine nodes and edges in the graph. Then we extract graph metric features. It can intuitively reflect the different interaction between hosts in the network (for example, a certain number of packet exchanges), and avoid network encryption and other issues. After that, we analyze differences of NCGs in time series to detect the outlier-points and identify abnormal events by matching the connected patterns with corresponding graph metric features. The approach has been validated by using network traces from Abilene, Internet2 backbone communication network. The results have demonstrated that the effectiveness of our approach in capturing the connected patterns of network abnormal events, and scalability in both of static graph analysis and dynamic analysis, even for evolution graph analysis. In summary, NCGs provide an easy-to-understand, effective in mining and flexible on construction means to do contribution in network flow behavior analysis.

## 2 Related Work

There are commonly two kinds of methods in detecting and identifying network abnormal events which are divided into misuse based detection method and anomaly detection method. Both of them are particularly important in network management and monitoring, but nowadays are very difficult to achieve better results because of the impact of massive network data and other security and privacy factors.

From the perspective of fine and coarse level of analysis, traditional network behavior analysis methods can be divided into three levels under different granularity, as Fig. 1 shows: (a) packet level, such as signature-based application identification and well-known classification based on port number; (b) statistical technology of flow level; (c) network-wide level, which is the most coarse analysis method, often be considered as the complement for detailed and precise analysis method.



**Fig. 1.** Network behavior analysis granularity.

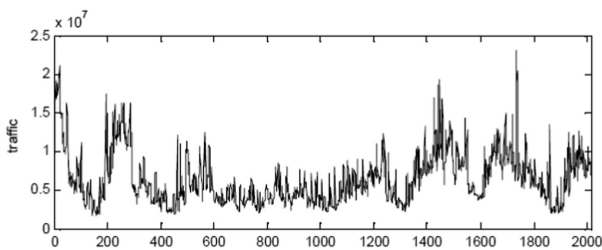
### 2.1 Packet Level Behavior Analysis

Packet level analysis method [9, 10] usually focus on the packet content including timestamp, IP addresses of source and destination hosts, port numbers opened for source and destination hosts, packet and byte counts, protocol, and TCP flags, etc. Table 1 shows the example of network packet data. The advantage of employing packet level analysis is to provide detail information about network interaction on the finest level of granularity.

The volume information of network traffic can be considered as a signal in the domain of signal processing. Figure 2 shows the volume of network traffic of one week trace with traffic every 5 min. Many researchers use time series analysis method, such as wavelet-analysis, PCA analysis, and ICA analysis.

**Table 1.** The example of network packet data

No.	Timestamp	SrcIP	DstIP	Sport	Dport	Proto.	#Bytes
1	0.00011	192.168.1.90	202.258.158.25	80	32548	6(TCP)	1540
2	0.00025	123.256.25.256	192.168.1.90	23548	80	6(TCP)	3340
3	0.000475	147.32.84.171	147.32.84.165	139	1040	6(TCP)	40
4	0.001953	147.32.84.171	147.32.84.165	139	1040	6(TCP)	44
...	.....	.....	.....	.....	.....	.....	.....



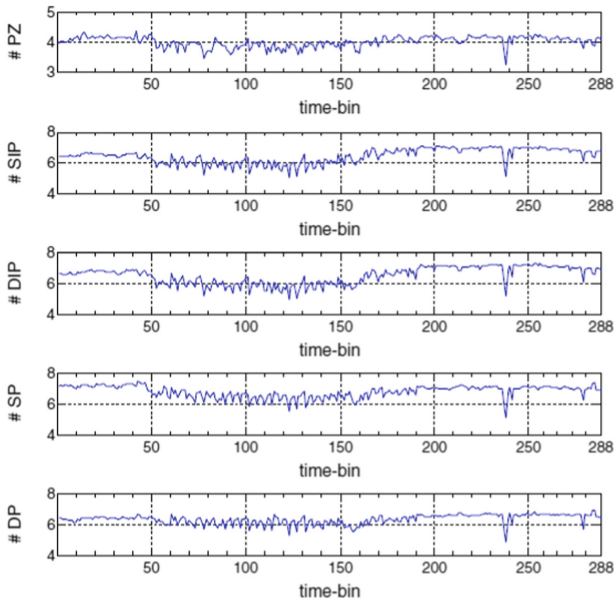
**Fig. 2.** Network traffic of one week traces (5 min for a time-bin).

## 2.2 Flow Level Behavior Analysis

Network operators lately have paid more attention on which new application or network event is flowing over the network that makes a wide range of network behavioral analyzing tasks focusing on the network flow data [11–13]. Network flows are active in host interactions so that we can make the most of connection relationship among various flows to analyze the behavioral characteristics of network flow, in particularly given a great number of IP-to-IP interaction flow data, how can we find interesting or suspicious behaviors, patterns and anomalies in real time through mining communication patterns, characterizing interaction structure, and modeling connection trends of new applications, users and other entities in the network. Thus this practical requirement has motivated a broad prospect for the development of a novel research field—network flow behavior analytics, being nowadays an interesting research domain.

## 2.3 Network-Wide Level Behavior Analysis

Network-wide behavior analysis [14, 15] which focuses on the global behavior information is including traffic, topology, and state information, globe level interaction information, link behavior information, etc. One of the network-wide level behavior analyses is shown in Fig. 3 by calculating different entropy of network-wide level properties.



**Fig. 3.** Entropy values of network-wide level properties in one day time interval. Among them, PZ denotes packet size, SIP denotes Source IP addresses, DIP denotes Destination IP addresses, SP denotes Source Ports and DP denotes Destination Port [15].

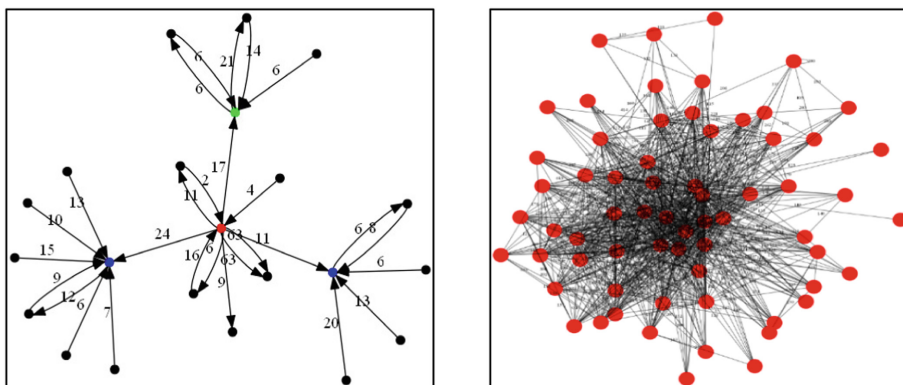
### 3 Overview of Network Connection Graphs

In this section, we introduce knowledge graphs, Network Connection Graphs (NCGs), graph metric features and discuss how these features are used to quantify NCGs and identify network abnormal events.

#### 3.1 Concepts of Network Connection Graphs

**Definition 1: Knowledge graph** is a structured semantic knowledge base to model connected relationships between various entities and reflect the knowledge information among them [16]. Its basic constituent unit is “entity-relationship-entity”, as well as the related attribute values. This type of relational information could be used to construct various graph models to forming a network of knowledge structure.

**Definition 2: A Network Connection Graph (NCG)** is a graphical representation of interactions of various devices, which can be modeled as a directed connect graph  $G = (V, E)$ , where  $V$  is the set of vertices representing IP addresses in the interaction flows and  $E$  is the set of edges representing the relationships between numerous devices in networks. Let  $n = |V|$  to represent the number of nodes and  $m = |E|$  to represent the number of edges in a graph. The NCG is different from other graph models by setting weights to edges and attributed information to nodes, which is a special case of knowledge graph applied to network flow behavior analysis.



**Fig. 4.** A visualization example of NCG concept graph and actual NCG under real network trace.

#### 3.2 Flow Knowledge Data Model Construction for NCGs

The data model for NCGs construction procedure is consists of two steps. The first step of it is to collect network flow data and pre-process these data into network flow. Since our approach does not require the detailed packet payload, we aggregate all information of network packet in transport layer header into flows based on an  $N$ -tuple flow— $\langle Time, SrcIP, DstIP, SrcPort, DstPort, Prot, Tcp\_flag \dots \rangle$ . Secondly, a graphical

representation model of NCGs. Figure 4 is drawn with employing visualization software GraphViz [17] to show a visualization example of NCG concept graph and actual NCG under Abilene network traffic trace.

### 3.3 Quantifying NCG Graph Metric Features

NCG could be regarded as a type of weighted complex network structure. By drawing on the study of network science [18], we propose two kinds of graph metric features to describe behavioral of network flow connected relations and use these features to offer the quantitative or semi-quantitative analysis for network flow behaviors.

**Basic metric features of NCG:** describes the basic parameters of network topology, such as the number of nodes and edges, the average degree of nodes and so on. The occurrence of network events would cause a sudden change of these features by increasing or decreasing of nodes and edges, such as the suddenly shut down of FTP servers will result in the reduction of this feature.

**Status metric features of NCG:** this kind of feature represents the state and compactness of nodes and edges in the global structure of network, including the node strength, the density of graph, cluster coefficient of node, cluster coefficient of the global graph, etc. Note that some network events would not cause the change of network topology, but they would have an impact on the status of nodes and edges, such as NCG basic features may change little when DNS request failure happens, but the status features may change since the connection structure would develop into bi-mesh pattern from star pattern.

These above features could contribute a lot in summarizing various network connection patterns. Figure 5 shows a brief example of graph features of different connected structures.

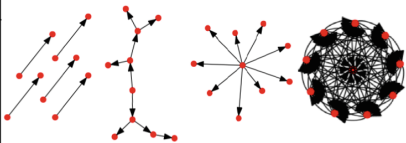
Property	Formula				
Subgraph Type	$G = \langle V, E, A \rangle$ V, Vertex set E, Edge set A, Adjacency matrix $A = (a_{ij})_{n \times n}$				
#Nodes	m	10	10	10	10
#Edges	n	5	9	9	45
Avg. Deg	$K = \frac{1}{n} * \sum_{i,j=1}^n a_{ij}$	1	1.8	1.8	9
Max. Connected Density	$MCD = \frac{Deg_{max}}{n-1} * 100\%$	0.11	0.33	1	1
In(%)	$In = \frac{m_{in}}{m} * 100\%$	50%	50%	90%	\
Out(%)	$Out = \frac{m_{out}}{m} * 100\%$	50%	10%	10%	\
In&Out(%)	$InO = \frac{m_{inO}}{m} * 100\%$	\	40%	\	100%
Max. Depth	$dep(v) = dep(u) + 1$	1	3	1	9

Fig. 5. A brief example of graph features of different connected structures.

### 3.4 Statistical Analysis for Graph Metric Features in NCGs

A NCG is corresponding to a vector space including the graph metric features, which means we could use some parameters to represent and distinguish NCG. Over the time, there are a lot of NCGs formed in each time-bin, thus as well as a series of graph metric features. However, when different network abnormal events occur, the structure changes of the NCG corresponding to the inherent characteristics are also different, which eventually result in the different changes of the graph metric features.

Therefore, we use the Z-score analysis method [19, 20] to find out the outlier-points of different graph metric features, and then identify the NCG of present moment as a suspicious state. There are several steps in this process, and also shown as in Fig. 6:

- Step1: Collecting network traffic and then aggregating packet data into network flow data;
- Step2: Constructing NCG over time to form a time series graph sequence;
- Step3: Extracting graph metric features for each a NCG of the graph sequence, and then build different sequence representations for various graph metric features; calculating the z-score for different feature sequences;
- Step4: Through deviation analysis to decide whether any graph metric features are under the anomalous state, and finally matching the features according to the inherent interactive characteristics of anomalies to identify the abnormal events.

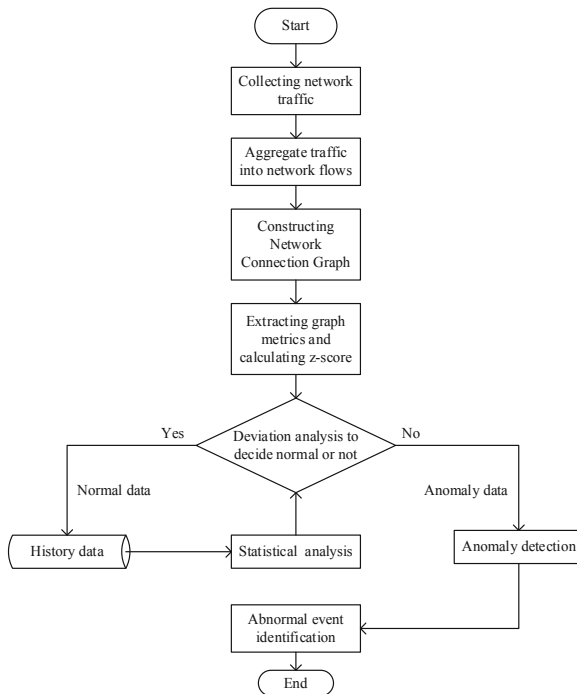


Fig. 6. Z-score based statistical analysis for Network Connection Graphs.

## 4 Experimental Results

To validate our approach in abnormal event identification, we used the sampled network flow dataset: Abilene [21], which belong to the Internet2 backbone network, connecting over 200 US universities and peering with research networks in Europe and Asia, and the topology structure [22] is shown in Fig. 7.

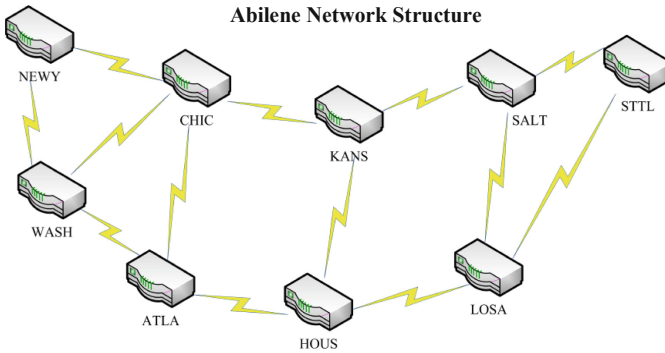


Fig. 7. Abilene network topology structure [22].

### 4.1 Static Graph Metric Feature Analysis

We have collected network flow data of Jan 11th in 2008. The trace is sampled with 5 min cycle and a NCG is generated every 5 min. We first calculate the graph metric features for different network applications in order to acquire connected characteristics of the normal flow data based on destination port number. Table 2 gives the graph metrics features for 6 kinds of well-known applications (including SMTP, DNS, HTTP, NetBIOS, eDonkey and WinMX). As shown in the table, we could draw a conclusion that the inherent interactive characteristics are different between each other.

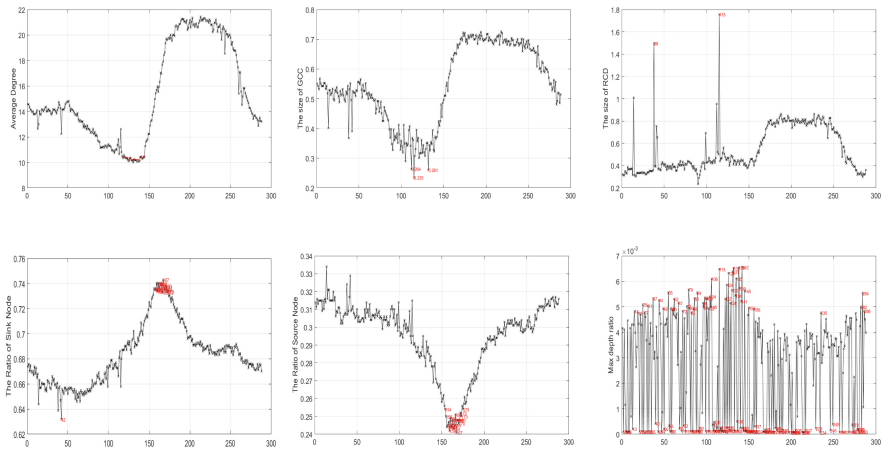
Table 2. Graph metric features for different network applications.

App.	#Node	#Edges	Avg. Deg	MCD	Directionality		GCC	Max Dep
					Sink	Source		
SMTP	3146	4345	2.76	3.66%	47.97%	52.10%	75.43%	45
DNS	9155	20265	4.43	7.14%	36.70%	63.62%	92.61%	337
HTTP	12889	13185	2.05	6.39%	25.39%	74.61%	65.40%	4
NetBIOS	10969	10523	1.92	4.46%	95.54%	4.47%	4.54%	3
eDonkey	10161	14355	2.83	0.94%	36.55%	63.47%	85.86%	295
WinMX	5966	14015	4.70	5.52%	38.55%	61.46%	98.27%	523



### 4.2 Dynamic Graph Metric Feature Analysis

Figure 8 shows the dynamic changes of NCG graph metric features in 2008-01-11 network flow trace, including Avg.Deg, MCD, directionality (sink node ratio and source node ratio), GCC and Max depth. If the feature exceeds the threshold calculated by Z-score analysis method, we will label it out. Tables 3 and 4 respectively shows the outlier-point extraction of Avg. Degree feature and MCD in 2008-01-11 network flow trace. The No. 130 dataset and No. 90 dataset is detected as suspicious state by the Z-score analysis method. Then we compare the graph metric features of this dataset with any normal dataset, with results are shown in Figs. 9 and 10.



**Fig. 8.** Dynamic changes of NCG graph features in 2008-01-11 network flow trace.

**Table 3.** Outlier-point extraction of Avg. Degree in 2008-01-11 network flow trace.

	#Nodes	#Edges	Avg. Deg	MCD	Directionality			Size of GCC	Max. Depth
					Sink	Source	Ino		
No. 130 suspicious dataset	1902	3554	3.1	14.9%	61.5%	38%	0.5%	46.48%	4
No. 153 normal dataset	2447	5344	4.7	32.3%	66.8%	32.7%	0.5%	52.8%	4

**Table 4.** Outlier-point extraction of MCD in 2008-01-11 network flow trace.

	#Nodes	#Edges	Avg. Deg	MCD	Directionality			Size of GCC	Max. Depth
					Sink	Source	Ino		
No. 90 suspicious dataset	223	565	5.9	98.6%	59.6%	40.4%	0%	23.8%	1
No. 170 normal dataset	289	617	3.4	69.8%	64%	36%	0%	34.9%	1

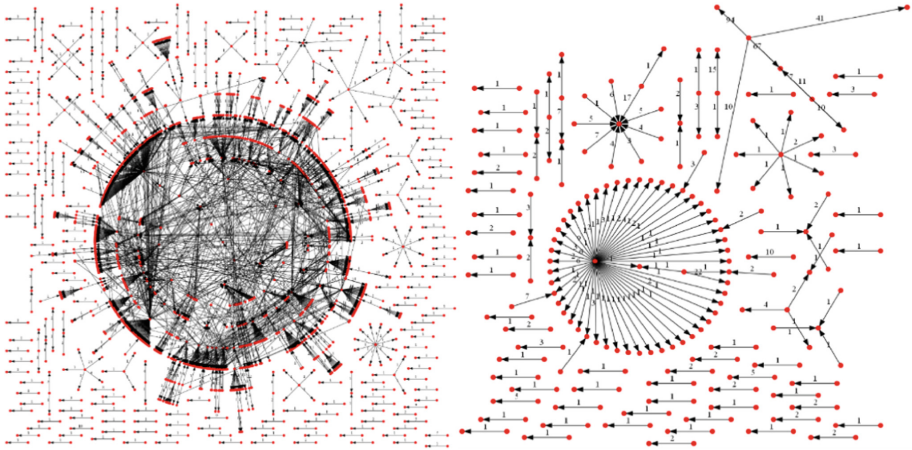


Fig. 9. NCG of No.130 suspicious dataset and NCG of No. 153 normal dataset.

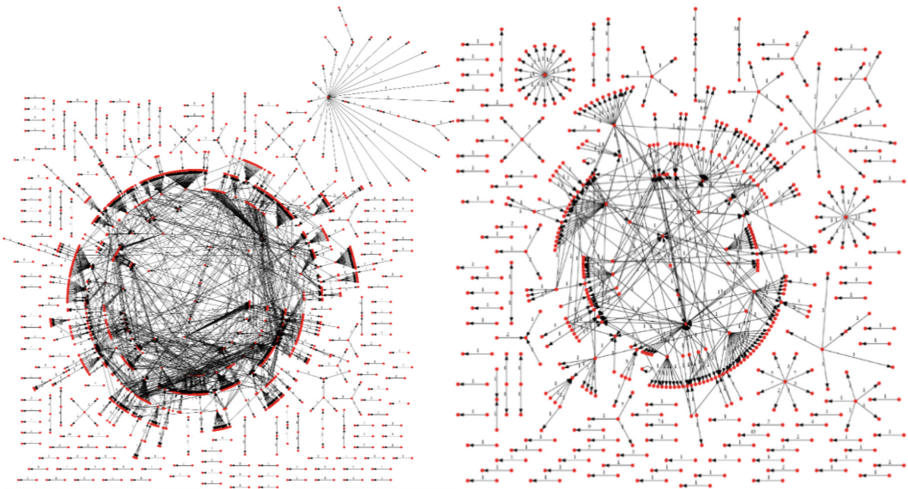
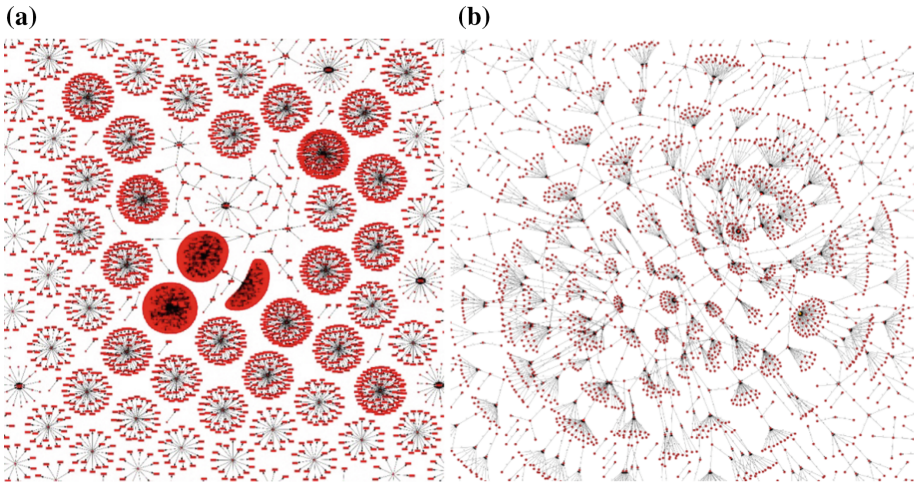


Fig. 10. NCG of No.90 suspicious dataset and NCG of No. 170 normal dataset.

### 4.3 Network Abnormal Events Visualization by Combining Manual Analysis and NCGs

We also validated our approach by manual label method in order to restore the complete connection relations of network abnormal events. Figure 11 shows the NCG visualization of DDoS attack and Worm propagation. From these two NCGs, we can conclude that their graph metric features are different from normal flows, especially compared with P2P structure based flow activities.



**Fig. 11.** NCG visualization of Abilene's trace: (a) DDoS attacks; (b) Worm propagation.

## 5 Conclusion and Prospection

In this paper, we propose the Network Connection Graphs (NCGs) as a novel tool to explore network flow activities by capturing the connected relationships of interaction flows. From the results of our experiments evaluated by real network traffic trace, combining with static analysis and dynamic analysis methods, we are effectively extract graph metric features for understanding network flow behaviors and identifying abnormal events. To our minds, our approach would contribute a lot to the foundation of next-generation communication network management and security service system development, and has implications for network security protection.

In the future work, aiming at increasing the robustness and effectiveness of our approach, we would start from sensitive analysis to improve the accuracy and performance in comparison with state of the art methods. Then we plan to investigate other dynamic analysis methods to thoroughly explore connected patterns for network events no matter normal or abnormal in order to completely characterize the network flow behaviors.

## References

1. Whitmore, A., Agarwal, A., Da Xu, L.: The Internet of things—a survey of topics and trends. *Inf. Syst. Front.* **17**(2), 261–274 (2015)
2. Qi, Y.: Information potential fields navigation in wireless Ad-Hoc sensor networks. *Sensors* **11**(5), 4794–4807 (2011)
3. Yang, X.L., Shen, P.Y., Zhou, B.: Holes detection in anisotropic sensornets: topological methods. *Int. J. Distrib. Sens. Netw.* **8** (10), 135054 (2012)
4. Zarpelão, B.B., Miani, R.S., Kawakani, C.T., de Alvarenga, S.C.: A survey of intrusion detection in Internet of Things. *J. Netw. Comput. Appl.* **84**, 25–37 (2017)

5. Zhang, H., Yi, Y., Wang, J., Cao, N., Duan, Q.: Network security situation awareness framework based on threat intelligence. *CMC: Comput. Mater. Continua* **56**(3), 381–399 (2018)
6. Habeeb, R.A.A., et al.: Real-time big data processing for anomaly detection: a Survey. *Int. J. Inf. Manag.* (2018)
7. Cheng, J., Xu, R., Tang, X., Sheng, V.S., Cai, C.: An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment. *CMC: Comput. Mater. Continua* **55**(1), 095–119 (2018)
8. Hu, H., Zhai, X., Wang, M., Hu, G.: Linked-behaviors profiling in IoT networks using network connection graphs (NCGs). In: Sun, X., Pan, Z., Bertino, E. (eds.) *ICCCS 2018*. LNCS, vol. 11067, pp. 429–439. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00018-9\\_38](https://doi.org/10.1007/978-3-030-00018-9_38)
9. Barford, P., et al.: A signal analysis of network traffic anomalies. In: *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement*. ACM (2002)
10. Zonglin, L., Hu, G., Yao, X.: Multi-dimensional traffic anomaly detection based on ICA. In: *IEEE Symposium on Computers and Communications, ISCC 2009*, pp. 333–336. IEEE (2009)
11. Karagiannis, T., Papagiannaki, K., Faloutsos, M.: BLINC: multilevel traffic classification in the dark. In: *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 4, pp. 229–240. ACM (2005)
12. Iliofotou, M., Pappu, P., Faloutsos, M., Mitzenmacher, M., Singh, S., Varghese, G.: Network monitoring using traffic dispersion graphs (TDGs). In: *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pp. 315–320. ACM (2007)
13. Jin, Yu., Sharafuddin, E., Zhang, Z.-L.: Unveiling core network-wide communication patterns through application traffic activity graph decomposition. *ACM SIGMETRICS Perform. Eval. Rev.* **37**(1), 49–60 (2009)
14. Lakhina, A., Crovella, M., Diot, C.: Characterization of network-wide anomalies in traffic flows. In: *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*. ACM (2004)
15. Zhou, Y., Hu, G., Wu, D.: A data mining system for distributed abnormal event detection in backbone networks. *Secur. Commun. Netw.* **7**(5), 904–913 (2014)
16. Wang, Q., Mao, Z., Wang, B., Guo, L.: Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* **29**(12), 2724–2743 (2017)
17. *GraphViz* (2011). <http://www.graphviz.org/>
18. Lewis, T.G.: *Network Science: Theory and Applications*. Wiley, Hoboken (2011)
19. Cheadle, C., Vawter, M.P., Freed, W.J., Becker, K.G.: Analysis of microarray data using Z score transformation. *J. Mol. Diagn.* **5**(2), 73–81 (2003)
20. He, D., Chan, S., Ni, X., Guizani, M.: Software-defined-networking-enabled traffic anomaly detection and mitigation. *IEEE Internet Things J.* **4**(6), 1890–1898 (2017)
21. <http://abilene.internet2.edu>
22. [https://en.wikipedia.org/wiki/Abilene\\_Network](https://en.wikipedia.org/wiki/Abilene_Network)