

Analysis of connection behaviour of communication network flow based on semantic understanding

Jing Sun^{a,b,*}, Qing Yu^{a,b}, Muheyat Niyazbek^{a,b}, Fayuan Chu^{a,b}

^a Information Science and Engineering School, Xinjiang University, Wulumuqi, Xinjiang 830046, China

^b Key Laboratory of Multilingual Information Technology in Xinjiang Uygur Autonomous Region, Xinjiang University, Wulumuqi, Xinjiang 830046, China

ARTICLE INFO

Keywords:

Semantic network transactions
Network visualization
Visual text analysis
Text analysis method

ABSTRACT

It presents a way to build support for knowledge, analytical reasoning and explore analysis methods in visual text analysis. To do this, use automatically from the use of semantic network model parameters K- unstructured text data obtained for the neighbourhood. Semantic network analysis and network analysis methods to obtain quantitative and qualitative insights. Qualitative analysis and quantitative indicators to explore the semantic structure can support it. have discussed the basic theoretical assumptions about the text modelling to analyse the semantic network and subsequent semantic network. For a systematic overview, demonstrated the essential network elements of the qualitative meaning, which means the network analysis, to understand the meaning of a given network, support analyst doing. Possible exploration strategy. As proof of concept, use a visual survey and analysis of the semantic network. To explain the method proposed a typical analysis of the Wikipedia article, use the visual text analysis system.

1. Introduction

The idea of the semantic organization has a long history; giving a potential proper system adjusted to the logical turn of events and applications has opened the premise of demonstrating and portraying information. The semantic organization makes it conceivable to demonstrate the semantic connections in the diagram of the named hubs and edges. Diagram hypothesis and organization examination, handle the proper organization; it tends to be investigated. Zero in on additional attributes, rich range of techniques and measurements, examining the organization structure, and describing it will develop with the informal community investigation and organization science.

The figure additionally outlines the advantages offers investigation and examination of organization representation and perception. Visual examination strategy joins the investigative thinking, intelligent visual examination measure, including backing and information to construct computerized information examination. Light, visual investigation of the semantic organization can be consequently utilized as the medium from the unstructured content information of the content's visual examination. Naturally extricate the critical data from the content, including the recovery and text mining methods.

Investigators' consistent cooperation with the pursuit network

requires an intelligent framework to explain its Semantic network analysis methodology proposed method can achieve user-centered visual exploration and analysis of complex semantic network, and based on the visual text analysis system. By linear interconnections packaged into text word network, the system provides an alternative view of the text and allows the analyst to explore the semantic context structure. Compared with the linear reading only text, the system helps to get insights from the semantic network analysis. The interaction beten the quantitative and qualitative analysis of network structure supports the semantic structure of the explanation. Supported target to support analysts to perform everyday text analysis tasks, put forward the essential elements of visual analysis of an adaptation text semantic network.

In the past few decades, witnessed a network of science and technology; innovation evolved from theory to practical application of the new application. Another aspect is the increase of the drive terminal and network user diversity index and diversity of network flow information. Understand network traffic behaviour can do a lot, has become the network monitoring and management; it predicted a significant topic of network events. Therefore, a comprehensive and accurate analysis of network traffic behaviour Dawn is a prerequisite for establishing security and stability, as ll as discipline from academia and industry to attract the attention of a reliable network environment.

* Corresponding author.

E-mail address: sjtree2020@163.com (J. Sun).

<https://doi.org/10.1016/j.micpro.2020.103578>

Received 10 November 2020; Received in revised form 24 November 2020; Accepted 30 November 2020

Available online 2 December 2020

0141-9331/© 2020 Elsevier B.V. All rights reserved.

Network flow behaviour analysis is an effective way to enhance network security by ensuring behaviour, aggregating flow data, and extracting corresponding flow behaviour features at different times and geographic locations. Most traditional analysis methods look at packet payloads in traffic categorized by many such applications to write their signatures in manual network traffic data such as deep packet/payload inspection technology. Focus on analysing network traffic, including analysis, considering network traffic as a time-varying signal, using signal processing-based methods, time series analysis, machine learning, and network flow. Mining data techniques for analysing time regions and higher-order statistical characteristics.

Despite the best efforts of network applications, increasing the network's size to analyse network traffic, high network traffic, and manual analysis methods to improve its cost. I think it will take some time. On the other hand, some methods are based cannot meet the demand due to the lack of accurate information on the signal processing for coarse grains and behaviour analysis. For this reason, network management and monitoring remain a vital challenge to develop suitable tools for behaviour analysis of network traffic.

Recording network communication process beten the source and destination hosts, and each stream to generate the original data and the two hosts are connected to ports such as a timestamp, IP address, port number, protocol. It describes the behaviour pattern for connecting beten the network entity and a network application flow connection. The pattern-based method shows the connection beten the natural and effective, and complete interdependent data elements, with the only connection strength behaviour analysis and mining network flow.

2. Related works

Ongoing years have seen a flood of writing in the exploration territory of organization stream conduct examination [1]. To restrict our extension, currently audit some old-style works that are generally

identified with our network conduct examination work. In writing, diagram based organization stream conduct investigation can be partitioned into two classifications: network-wide based methodologies and host-level based methodologies.

Organization-wide put together methodologies mostly centre concerning arrange wide associated relations [2]. Proposed the strategy to fabricate logical diagrams of hub connection. They used the distinction among associated examples of organization applications for performing network characterization; they generally disregarded organization streams' generally associated relations [3]. Traffic scattering diagrams, which expand on crafted, re utilized to observe organization traffic and identify undesirable applications. Supposedly, this the principal work to profile "who talks whom" relations at the organization-wide level by utilizing a coordinated diagram model. Proposed traffic movement charts (TAGs) to derive new applications and identify worm proliferation.

Proposed traffic causality charts to dissect worldly and spatial causality of streams. High [4] precision in application ID accomplished by utilizing chart mining calculation to remove the element and a likeness measure in the component vector space of TCGs. In any case, these methodologies are direct chart models, which exclusively depend on the diagram structures.

To distinguish P2P botnets in their holding upstage, [5] proposed a chart based arrangement considered Entelechies that centres on the "social" conduct [7]. Entelechies misuses botnets' intrinsic conduct by analysing extensive and low-por streams, and it would be progressively corresponding to other botnet recognition draws near [8]. Most of the late works firmly identified with our strategy are [9]. They demonstrated that chart-based methodologies are effectively applied to interface expectations in informal online organizations, such as Twitter, by including extraction. Nonetheless, the above methodologies, which propose a few diagram measures, are not appropriate in massive scope networks based on substantially more excess data [6].

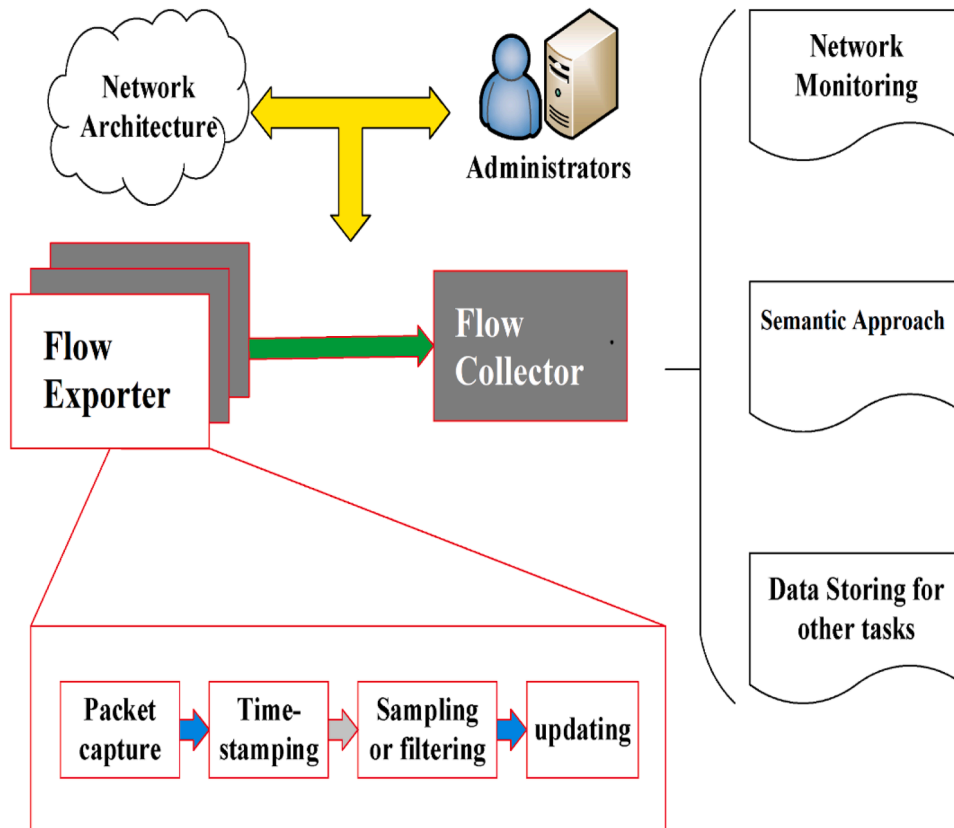


Fig. 1. Proposed architecture semantic approach.

Having level-based methodologies to investigate the associated relations beten the source and objective IP, bipartite [10] charts are regularly utilized in-stream conduct examination. Introduced a bipartite graph and its one-mode projection to study the social behaviour of Internet application traffic [11]. Similar social practices of end show that comparative coefficient grouping could be utilized to identify rising applications and peculiar traffic designs, spoken to all correspondences among customers and workers as a ighted, close bipartite chart.

Utilizing a two-stage rotating score spread calculation on the diagram, they could distinguish dubious customers in a checked organization utilizing bipartite diagrams, such as clients versus documents in a P2P framework [12]. Assessed the nature of neighbourhoods dependent on the diagram dataset's semantics and estimated the inconsistency discovery presentation with physically infused irregularities [13]. With this sort of approach, it is hard to uncover the concealed connections beten various networks of an organization.

Related work can be applied to our methodological contrasts aren't territories to be found. Characteristic language preparing, information portrayal, and network examination and representation investigation [14]. Text information-digging strategy for a robot to recoup from FIG printed data and text mode a given undertaking. Investigate the immense scope of records. In the overall case, oneself requesting augmentation direct, given the semantic request of disintegration and repetitive character portrayed dimensionality decrease, multidimensional scaling, a header check or idle.

Focus on the book model to expand the procedure using the K- next neighbourhood model for the connection extraction on the window. In future work, this model can re-definition Fi in strategy in response to the request [15]. Also, to improve the connection extraction and model building might be embraced ideas minutes in, rather than a formal plan to use models for the depiction of encyclopaedic knowledge representation, such as the Word of the net information, near related to the observation content information, suggest that display the semantic organization.

So cannot give a reasonable idea of such action and the like according to a given ontology. Prefer to display a common language to communicate" semantics; it can be identified as a representation or description of deterministic information, just as with civil society's psychological labelling methods. In the case of a given meta-information, relying on technology to improve lightight [16] comment this model, the meta-information including attachments referenced in our strategy, and still included in the first approach Fi in social network analysis focused on b survey and histological examination, expand to a meticulous dissection [17] rich semantics of the organizational structure of the system, and accept from a subjective point of view they are focused on the essential semantic investigation.

Visual inspection using the book's framework, allowing the network application survey, natural recovery semantic organization, and to sinter the dynamic perception, our approach is closely identified with the visual content of the survey [18]. Diagnosis depends on the visual content of the frame to the coded text represents two afferent logic features. Arc-shaped frame or help textbooks of a given summary outwardly [19].

Other types of perceived focus on design validation and key extraction. Generalize relevant data in the phrase's tree structure [20], so the Nets assembly relationships with customers word of more complex semantic organization planning unstructured content. Respect the visual research methods.

3. Proposed methodology

This section first describes the network flow, and based on the analysis described learning behavioural semantics of the configuration shown in Fig. 1. Finally, explain the basic concepts and applications of networking topics, and to apply a theme to worm detection.

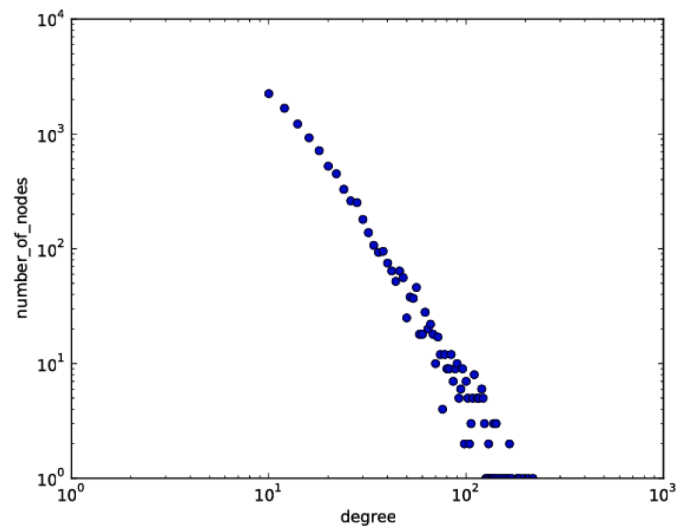


Fig. 2. For a model semantic network of distribution.

3.1. Automated network retrieval

Our hypotheses use similar Semantic approach techniques to model a generalized semantic network model and unstructured text. Focus on close analysis; build from the spelling semantic model. To this end, apply a filter; the cut termination list is not semantically related words in the partial text words. Word relations are established. Network on behalf of a search word, get relations with the extracted Word corresponding to the edge node undirected ighted graph. The node is due to the source text; the text cites the underlying source text. Calculated by the layout applied to a quantitative measurement of the network ights. Embedding algorithm using flexible spring layout to obtain visual analysis and exploration for local representation. Like other dimension reduction technique, encode the tightness of space and semantic similarity. For our earlier work, further re-create more advanced technology and define the C language needed to extract full voice analysis of the relationship beten domain analyses.

The main problem in analysing unstructured text data is ambiguity and complexity about the meaning of natural language. Another problem is the lack of automated data analysis and contextual information contextual explicit knowledge. Text sources are often heterogeneous and contain intentional meaning without context information to be addressed, relying on slang, abbreviations, metaphors, or stretching almost common sense of the world. Therefore, it is usually applied strictly to any text source ontology classification challenges. Natural language processing is very detailed, but focus on a structured approach to a simple text model and try to use it better in the first step of several NLPs network. Our method also uses the intuitive research on behalf of, and in the methodological framework and analysis of text data text interactive systems analysis, general model using any text. Driven by practical application, support tools, explore and analyse the sources.

3.2. Semantic network analysis

Propose a semantic network element analysis methodology. Our terminology describes how to analyse a semantic network structure analysis semantic networks obtained according to the given text model. Therefore, recommend helping reveal the quality to distinguish generate quantitative measurement and semantic analysis to analyse human-centered automation networks. An automated network analysis algorithm is applied to calculate the index and semantic analysis of a given statistical process based primarily on the human eye. It can also recognize and understand the purpose of exploration or analysis to infer semantic structure. This is primarily and found support for extending

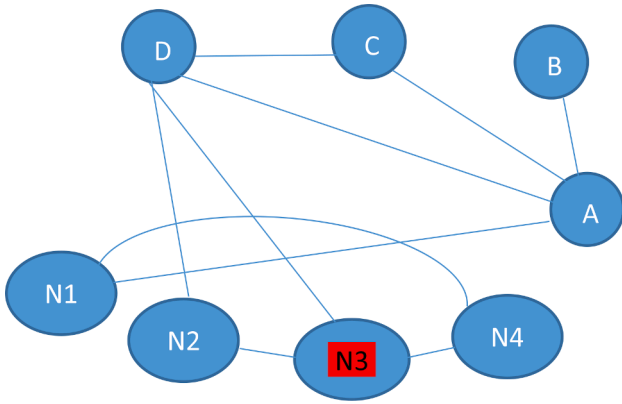


Fig. 3. Semantic network model structure.

knowledge of world quality from a given semantic network of knowledge in specific areas.

Fig. 2 define the both analysts can automatically handle large amounts of data at high speed and explore and understand the significance of the true meaning of extension analysis. Understanding these characteristics, analysts can quickly identify critical nodes and visit the local cluster. Determination of other networks, for example, help characterize the diameter or density of the network structure. Many indicators summary, can go beyond the height of the network layout to use.

3.3. Network structure

Fig. 3 define the Por-law distribution and sematic analysis to help provide clues to describe the topology of the network nodes and clusters. Display nodes associated with the distribution of the code. Understanding these characteristics, analysts can quickly identify critical nodes and visit the local cluster. Determination of other networks, for example, help characterize the diameter or density of the network structure. Many indicators summary, can go beyond the height of the network layout to use.

Qualitatively speaking, a sparse network may be indicated by a global lack of Word semantic consistency word. High density, it seems more consistent, rather than displaying a dense network of higher amounts to achieve the relationship of the Word, this is a. The number of nodes may indicate a variety of qualitative and higher present in the semantic network word language representation. In a broad sense, a qualitative network structure helps to show the outstanding child to explore.

3.4. Node

Quantitative node indicates the number of adjacent nodes, and wherein each metric is connected to the scan through the node. Words with a high degree refer to the local hub network, which usually means much eventuated associated inch 10 indicates a hub height corresponding to necessary.

Use other measures to quantify the central node's importance in a network, the core issues and global hub indication; it is usually referred to as the lor order. The central node indicates a network, which means that it has a connection with a plurality of other nodes. Local hub generally represents an abstraction from the reference event's context by excluding stop words from the network.

4. Result and discussion

To network flow connection behaviour is the root cause analysis to verify the validity of identifying the type of network traffic behaviour. A

Table 1

Semantic network analysis and exploration for qualitative and quantitative interpretation of semantic network elements.

Element	Quantity	Quality
Network structure	Statistics like degree distribution	Characteristic topological properties of the network.
Path	Length	Collocation, semantic relation, meaning
Node	Degree	The complexity of a semantic concept on a word basis. Position in a semantic field.
Hubs	Centrality measures, filtering ($d(v) \geq n$)	Centrality measures, filtering ($d(v) \geq n$)
Subgraphs	(see network structure above)	Complex semantic context is encoded in interconnected semantic fields.
Cluster	Clustering coefficient; filtering ($d(v) \leq n$)	Strongly connected components encoding specific semantic topics or complex concepts

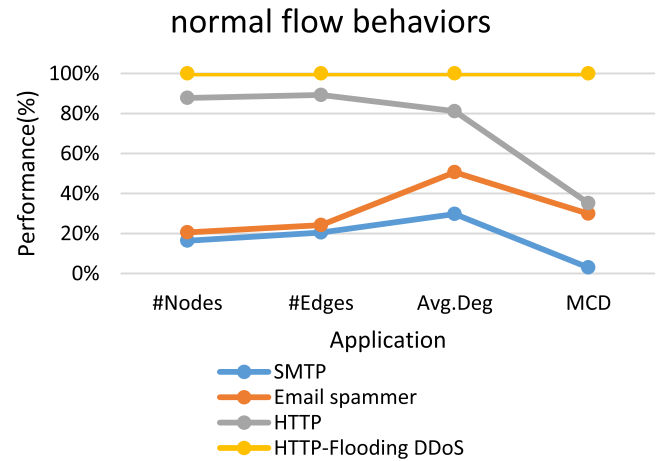


Fig. 4. Normal flow behaviour's performance in network.

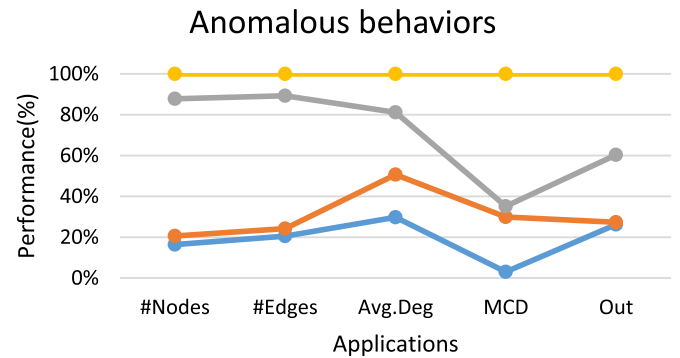


Fig. 5. Anomalous flow behaviours' performance in network.

single experiment conducted revealed the root cause of the network. Mining carriers are often used at the node level. The experiment in 5 minutes flows verification data link. Table 1. The most frequent item in the network is shown in the previous five flow source address. That the source IP address is connecting too many hosts for remote control, which would attract network monitors' attention.

Suspicious port scanning behaviours due to the connection to the destination host with the fact that only a small number of open-source port, and a lot of different destination port. Unlike the above flow characteristics, and the source port number and destination port are

Table 2

Normal flow behaviour and abnormal behaviour.

Applications	#Nodes	#Edges	Avg	MCD	In	Out	GCC	Depth
SMTP	3146	4345	2.76	3.66%	47.97	52.03	75.43	45
Email spammer	800	785	1.94	32.7 %	97.88	2.25	32.25	2
HTTP	12889	13785	2.83	6.39%	25.35	65.40	65.40	4
HTTP-Flooding DDoS	2344	2272	1.75	78.79%	94.54	78.80	78.80	3

open, as long as the exchange between the two host's large amounts of data, the last one is marked as active alpha flow. Below Figs. 4 and 5 functioning flow behaviours and abnormal behaviours as show you in Table 2.

5. Conclusion

Discussed the general approach and qualitative aspects of network analysis; this means analysing and interpreting network structure derived coverage is quantified. Show to develop a semantic network analysis using analytical assumptions and inference to visually find out more about the minute information to help analysts explore the unknown source text visually. Yes, even with the proposed approach, exchange all bills for visual text analysis systems to read and understand the text. Let's face an abstract perspective by providing alternative text sources and supporting text analysis to the semantic network model's given text sources. For abstract and model are chosen, immediately view and analyse the text word by word, without a glance, read the word is connected and within the context. In the formal structure of cooperation, use the visual text analysis, and network analysis relies on an interactive visualization of the systematic framework. This intuitive access is also provided for determining our interpretation and the specific process intended to analyse the text of advantage and reading.

Declaration of Competing Interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgement

(1) Research On Technologies of Kazakh Continuous Speech Keyword Spotting: National Natural Science Foundation of China, 61462084; (2) Research on Parallel Corpus of Uyghur-Han to be built for the Medical Service: National Natural Science Foundation of China, 61562082.

References

- [1] M. Gocek, R. Marković, J. Dolenšek, M.S. Runic, M. Marshall, A. Stožer, M. Perc, Loosening scientific disciplines' shackles with network science, *Phys. Life Rev.* 24 (2018) 162–167.
- [2] M. Jalili, M. Perc, Information cascades in complex networks, *J. Complex Netw.* 5 (2017) 665–693.
- [3] K. Xu, Z.L. Zhang, S. Bhattacharyya, Internet traffic behaviour profiling for network security monitoring, *IEEE/ACM Trans. Netw.* 16 (2008) 1241–1252.
- [4] E. Sahara Puddin, N. Jiang, Y. Jin, Z.L. Zhang, Know your enemy, know yourself: block-level network behaviours profiling and tracking, in: *IEEE2010 Global Telecommunications Conference*, 2010, 6–10 December.
- [5] A. Fahad, Z. Tari, I. Khalil, I. Habib, H. Alnuiri, Toward an efficient and scalable feature selection approach for internet traffic classification, *Comput. Netw.* 57 (2013) 2040–2057.
- [6] B. Li, J. Springer, G. Bebis, M.H. Gunes, A survey of network flow applications, *J. Netw. Comp. Appl.* 36 (2013) 567–581.
- [7] D. Helbing, D. Brockmann, T. Chadefaux, K. Donnay, U. Blanke, O. Woolley-Meza, M. Moussaid, A. Johansson, J. Krause, S. Schutte, et al., Saving human lives: what complexity science and information systems can contribute, *J. Stat. Phys.* 158 (2015) 735–781.
- [8] A. Boukhtouta, S.A. Mokhov, N.E. Lakhdari, M. Debbabi, J. Paquet, Network malware classification comparison using DPI and flow packet headers, *J. Comput. Virol. Hacking Tech.* 12 (2016) 69–100.
- [9] P. Barford, J. Kline, D. Plonka, A. Ron, Signal analysis of network traffic anomalies, in: *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement*, ACM, Marseilles, France, New York, NY, USA, 2002, pp. 71–82, 6–8 November.
- [10] T.T. Nguyen, G. Armitage, A survey of techniques for internet traffic classification using machine learning, *IEEE Commun. Surv. Tutor.* 10 (2008) 56–76.
- [11] Y. Zhou, G. Hu, GNAED: a data mining framework for network-wide abnormal event detection in backbone networks, in: *IEEE International Performance Computing and Communications Conference (IPCCC)*, 2011, pp. 17–19, November.
- [12] L.C. Freeman, S.P. Borgatti, D.R. White, Centrality in valued graphs: a measure of betweenness based on network flow, *Soc. Netw.* 13 (1991) 141–154.
- [13] Q. Yan, Y. Zheng, T. Jiang, W. Lou, Y.T. Hou, Periclean: unveiling peer-to-peer botnets through dynamic group behaviour analysis, in: *Computer Communications (INFOCOM)*, Hong Kong, China, 2015, pp. 316–324, 26 April–1 May.
- [14] L. Akoglu, H. Tong, D. Koutra, Graph-based anomaly detection and description: a survey, *Data Min. Knowl. Discov.* 29 (2015) 626–688.
- [15] T. Karagiannis, K. Papagiannaki, M. Faloutsos, BLINC: multilevel traffic classification in the dark, *ACM SIGCOMM Comput. Commun. Rev.* 35 (2005) 229–240.
- [16] Iliofotou, M., Pappu, P., Faloutsos, M., Mitzenmacher, M., Singh, S. and Varghese, G., Network monitoring using traffic dispersion graphs (TDGs). 23–26 October 2007.
- [17] Y. Jin, E. Sharafuddin, Z.L. Zhang, Unveiling core network-wide communication patterns through application traffic activity graph decomposition, *ACM SIGMETRICS Perform. Eval. Rev.* 37 (2009) 49–60.
- [18] H. Asai, K. Fukuda, P. Abry, P. Borgnat, H. Esaki, Network application profiling with traffic causality graphs, *Int. J. Netw. Manage.* 24 (2014) 289–303.
- [19] D. Markert, P. Parsch, A. Masrur, Impact of probabilistic vehicle estimates on communication reliability at intelligent crossroads, *Microprocess. Microsyst.* 78 (2020), 2020, 103262, ISSN 0141-9331.
- [20] J. Hua, N. Shunwiritu, Research on term extraction technology in the computer field based on wireless network technology, *Microprocess. Microsyst.* (2020), 103336, ISSN 0141-9331.



Jing Sun, Master of Computer Science, Lecturer, Graduated from the Xi'an Jiao Tong University in 2008. Worked in Xinjiang University. Field of research includes machine learning, Optimization algorithm, audio information processing and natural language and information processing. more than 1 papers published and 1 book published.



Qing Yu, Master of Engineering, Associate Professor. Graduated from the Xinjiang University in 2008. Worked in Xinjiang university. Her research interests include Artificial intelligence, machine translation. more than 2 papers published.



Muheyat Niyazbek, Bachelor, Associate Professor. Graduated from the Jilin Institute of Chemical Technology in 1991. Work in Xinjiang University, College of Information Science and Engineering. His research interests in natural language and information processing, computer architecture and network technology. more than 2 papers published.



Fayuan Chu, Master of Computer Science, Lecturer, Graduated from the Xi'an Jiao Tong University in 2007. Worked in Xinjiang University. His research interests include artificial intelligence, machine learning, speech recognition, nature language processing, and image processing.