An Open Dataset of Cyber Asset Graphs for Cybercrime Research

Supporting Information

This supplement material introduces the basic information of the CAG-CR-22 dataset (Cyber Asset Graphs for Cybercrime Research) and the representative CAGs that can be mined in the dataset. Moreover, this supplement material also provides a set of domain knowledge for CAG mining, core cyber assets identification, and critical path identification.

Please visit the following URL to obtain the CAG-CR-22 dataset.

- https://github.com/csuvis/CyberAssetGraphData

1. Data Description

The CAG-CR-22 dataset is stored in CSV format with a total uncompressed volume of 721MB. It contains a Node.csv and a Link.csv files that record the information about nodes (i.e., cyber assets) and edges (i.e., relations), respectively.

1.1 Node.csv

The Node.csv file has a size of 228M and includes 2.37 million data records. Each data record represents a node described with four fields, as shown in Table 1.1. The type field has eight possible values, which are detailed in Table 1.2. Moreover, the industry field provides the types of cybercrime activities related to a "Domain" node. The possible types of the industry field are provided in Table 1.3.

Table 1.1 Descriptions of fields of Node.csv file.

Field	Field description	Field format	Example
id	Represents node id, it is the unique identification of a node.	String	For example, "Domain_0d9f06a82e90193f68e72e53acd55e23c74afb0e3589608627e423c64d19f6db".
name	Represents node name, it is encrypted with MD5 and character invalidation for anonymization.	String	For example, "+86.533xxxxx", "0d9f06a82e.com".

type	Represents node type (i.e., cyber asset type), it has eight possible values (see Table 1.2).	String	For example, “Domain”, “IP”, “Cert”.
industry	Represents the types of cybercrime activities related to a “Domain” node. It has nine possible types (see Table 1.3). The value of this field is a formatted string to involve multiple types for a “Domain” node.	String	For example, “['A'], ['B']”, “['C', 'D']”

Table 1.2 Detailed information of the type field in Table 1.1.

Type field name	Number of records
Domain	About 2 million records
IP	About 200 thousand records
Cert	About 130 thousand records
Whois_Name	About 18 thousand records
Whois_Phone	About 2 thousand records
Whois_Email	About 4 thousand records
IP_C	About 6 thousand records
ASN	About 3 hundred records

Table 1.3 Descriptions of types of cybercrime activities for the industry field in Table 1.1

Industry type	Description
A	Online illegal pornography spread
B	online illegal gambling
C	online fraud
D	online illegal drug trafficking
E	online illegal firearm trafficking
F	hacker
G	online illegal transaction platform
H	online illegal payment platform
I	others

1.2 Link.csv

The Link.csv file has a size of 493M and includes 3.28 million data records. Each data record represents an edge described with three fields, as shown in Table 1.4. The relation field has eleven possible values to represent the type of a relation, as detailed in Table 1.5.

Table 1.4 Descriptions of fields of Link.csv file.

Field	Description	Field format	Example

relation	Represents the type of a relation. It has eleven possible types (see Table 1.5).	String	For example, “r_cert”, “r_dns_a”.
source	Represents the source node id.	String	For example, “IP_37f7ed5739b43757ff23c712ae4d60d16615c59c0818bf5f2c91514c9c695845”.
target	Represents the target node id.	String	For example, “Domain_2d3bbcec29453b6f56fb85ea28e8e5ea5fc5f5562e0f896b6b52b113a6cc1e44”.

Table 1.5 Detailed information of the relation field in Table 1.4.

Relation type	Number of records
r_cert	About 230 thousand records
r_subdomain	About 450 thousand records
r_request_jump	About 6 hundred records
r_dns_a	About 2.05 million records
r_whois_name	About 100 thousand records
r_whois_email	About 28 thousand records
r_whois_phone	About 19 thousand records
r_cert_chain	About 15 thousand records
r_cname	About 130 thousand records
r_asn	About 69 thousand records
r_cidr	About 170 thousand records

2. Examples of CAGs

A cyber asset graph (CAG) is a subgraph in the dataset. Generally, a CAG is a collection of closely related cyber assets held by the same cybercrime gang. The dataset contains numerous CAGs that may be related to many cybercrime gangs in the real world.

A CAG can be mined in the dataset based on a few given entry cyber assets. We provide five examples of CAGs mined in the dataset by experts with five sets of entry cyber assets. Table 2.1 provides the statistical information about the five CAGs. The domain knowledge used by experts for CAG mining and core cyber asset identification are summarized in Section 3.

In this section, we present the entry cyber assets and visualization result for each of the five CAGs. It is worth noting that large star-like clusters commonly exist in a CAG. To reduce the visual clutter of CAG visualization result, we only preserve a part of the isomorphic neighbor nodes of the center node of a cluster using a random sampling method. Taking the CAG 1 presented in Figure 2.1(b) as example, the center node of the left star-like cluster has 19883 original neighbor nodes. These neighbor nodes have

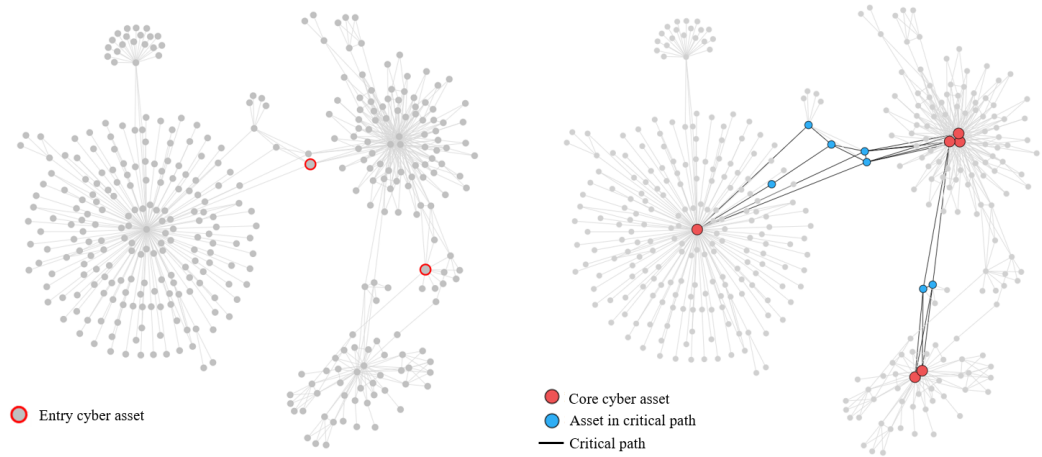
similar local structures (e.g., similar types of cyber assets and relations and similar connectivity with surrounding nodes). After sampling, only 115 nodes are preserved to show in the visualization result of the CAG.

Table 2.1 Statistical information about the five examples of CAGs.

CAG ID	Size	Number of core cyber assets	Number of critical paths
CAG 1	Small-sized	6	22
CAG 2	Medium-sized	8	732
CAG 3	Medium-sized	2	1
CAG 4	Large-sized	73	1261
CAG 5	Large-sized	50	293

2.1 CAG 1

CAG 1 is a small-sized graph mined with two given entry cyber assets. Figure 2.1(a) presents the visualization result of the CAG, including 368 nodes and 617 edges with the entry cyber assets marked by red borders. Figure 2.1(b) is the visualization of the CAG with highlighted core cyber assets and critical paths. The list of entry cyber assets and core cyber assets of CAG 1 are provided in Table 2.2 and Table 2.3, respectively.



(a) (b)

Figure 2.1 Visualization results of CAG 1. (a) visualization result with highlighted entry cyber assets; (b) visualization result with highlighted core cyber assets and critical paths.

Table 2.2 Entry cyber assets of CAG 1.

Asset id	Asset name	Asset type
Domain_c58c149eec59bb14b0c102a0f303d4c20366926b5c3206555d2937474124beb9	c58c149eec.com	Domain

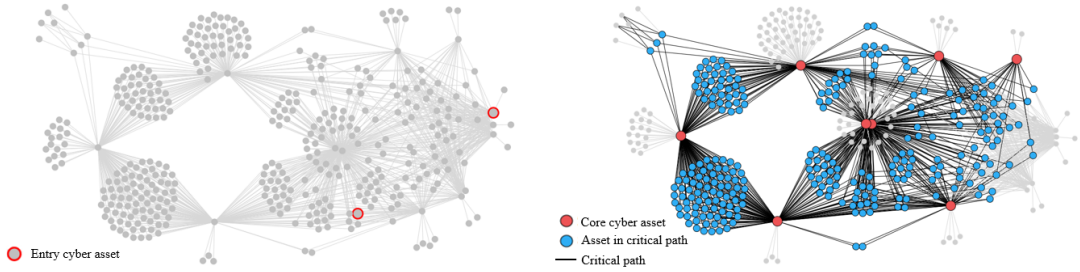
Domain_f3554b666038baffa5814c319d3053ee2c2eb30d31d0ef509a1a463386b69845	f3554b6660.com	Domain

Table 2.3 Core cyber assets of CAG 1.

Asset id	Asset name	Asset type
Whois_Name_db0925a5aeb1849fa7b41f7a29c1192d38e12e97fb6e82e72e894e3c733130ef	Linxxxxx Xu	Whois_Name
Whois_Email_5a3d16b7df3d815d5f3436bd5dd5c5e1054ee7cb74d4fd8d9efdf3af362a4a18	54498xxxxx@xxx.xxx	Whois_Email
Whois_Phone_f6974ce3fa84ae76d75b9211f3162155db77566a36c82549b66a9a3d966a928b	+86.533xxxxx	Whois_Phone
IP_38d08556e5f342ddca3d2001e92f56b2e835b43a8ff78e202ede932442cae5b2	116.206.xxx.xxx	IP
Cert_fe794a69eacd63b21245bf4eda826222fc6c5862bebf77aa05459cb308cfd063	fe794a69ea	Cert
Cert_e72592e3cf6097989d7af61181669ba7c72fe3e7059ecf79f284391665d32fe5	e72592e3cf	Cert

2.2 CAG 2

CAG 2 is a small-sized graph mined with two given entry cyber assets. Figure 2.2(a) presents the visualization result of the CAG, including 401 nodes and 1,099 edges with the entry cyber assets marked by red borders. Figure 2.2(b) is the visualization of the CAG with highlighted core cyber assets and critical paths. The list of entry cyber assets and core cyber assets of CAG 2 are provided in Table 2.4 and Table 2.5, respectively. The nodes in the visualization results of CAG 2 are not sampled.



(a) (b)

Figure 2.2 Visualization results of CAG 2. (a) visualization result with highlighted entry cyber assets; (b) visualization result with highlighted core cyber assets and critical paths.

Table 2.4 Entry cyber assets of CAG 2.

Asset id	Asset name	Asset type
IP_400c19e584976ff2a35950659d4d148a3d146f1b71692468132b849b0eb8702c	156.241.xxx.xxx	IP
Domain_b10f98a9b53806ccd3a5ee45676c7c09366545c5b12aa96955cde3953e7ad058	b10f98a9b5.com	Domain

Table 2.5 Core cyber assets of CAG 2.

Asset id	Asset name	Asset type
IP_f9b588fa3410ab89fa0e50b011c9ac8ddfa4a3125ea3df13fa4598faa5e15f8a	45.114.xxx.xxx	IP
Cert_c992a7d7f01fae6098d8f1ba358002074db1b977cceafc07c04b40e657ec0425	c992a7d7f0	Cert
IP_36b2ba5b0800d154ef3add5672b7561af9535edd92d2c3323c64880498b45a05	45.114.xxx.xxx	IP
Domain_8659e9de39a88dc208eae9c4eab0791afd0406142fd7220cac3e7793dc802a43	8659e9de39.com	Domain
Cert_d570aebdd3b0f0b4194315d8df020dc805f114401fe3c6999967f60de17b6176	d570aebdd3	Cert
Cert_a77b63d27d07fd9cc522afb93664f99d9f56f9edadf84e44ef4537748dc19141	a77b6c3d27d	Cert
Cert_1b22e6e2c9f9d7afd041a1a0ef2178dbaaf3248c4261496a382ff46520d55e71	1b22e6e2c9	Cert
IP_cd3ce4957d196a1a2871f3b850cb5ab89ffa2643033d7b05319951f2be9322e0	164.88.xxx.xxx	IP

2.3 CAG 3

CAG 3 is a medium-sized graph mined with five given entry cyber assets. Figure 2.3(a) presents the visualization result of the CAG, including 589 nodes and 1,057 edges with the entry cyber assets marked by red borders. Figure 2.3(b) is the visualization of the CAG with highlighted core cyber assets and critical paths. The list of entry cyber assets and core cyber assets of CAG 3 are provided in Table 2.6 and Table 2.7, respectively.

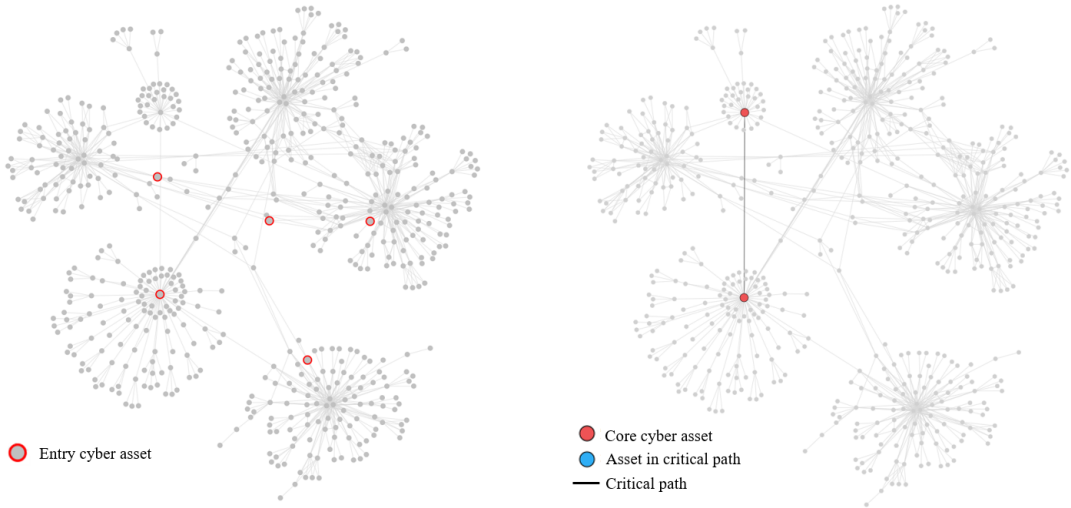


Figure 2.3 Visualization results of CAG 3. (a) visualization result with highlighted entry cyber assets; (b) visualization result with highlighted core cyber assets and critical paths.

Table 2.6 Entry cyber assets of CAG 3.

Asset id	Asset name	Asset type
Domain_24acfd52f9ceb424d4a2643a832638ce1673b8689fa952d9010dd44949e6b1d9	24acfd52f9.com	Domain
Domain_9c72287c3f9bb38cb0186acf37b7054442b75ac32324dfd245aed46a03026de1	9c72287c3f.com	Domain
Domain_717aa5778731a1f4d6f0218dd3a27b114c839213b4af781427ac1e22dc9a7dea	717aa57787.com	Domain
Domain_8748687a61811032f0ed1dcdb57e01efef9983a6d9c236b82997b07477e66177	8748687a61.com	Domain
Whois_Phone_f4a84443fb72da27731660695dd00877e8ce25b264ec418504fface62cdcbbd7	+1.971xxxxx	Whois_Phone

Table 2.7 Core cyber assets of CAG 3.

Asset id	Asset name	Asset type
Domain_8cfbc3dc32c53413f89000e1ed8c7c387032e5c0138a250085e3be964dbac32e	8cfbc3dc32.com	Domain
Domain_24acfd52f9ceb424d4a2643a832638ce1673b8689fa952d9010dd44949e6b1d9	24acfd52f9.com	Domain

2.4 CAG 4

CAG 4 is a large-sized graph mined with two given entry cyber assets. Figure 2.4(a) presents the visualization result of the CAG, including 2,354 nodes and 5,271 edges

with the entry cyber assets marked by red borders. Figure 2.4(b) is the visualization of the CAG with highlighted core cyber assets and critical paths. The list of entry cyber assets and core cyber assets of CAG 4 are provided in Table 2.8 and Table 2.9, respectively.

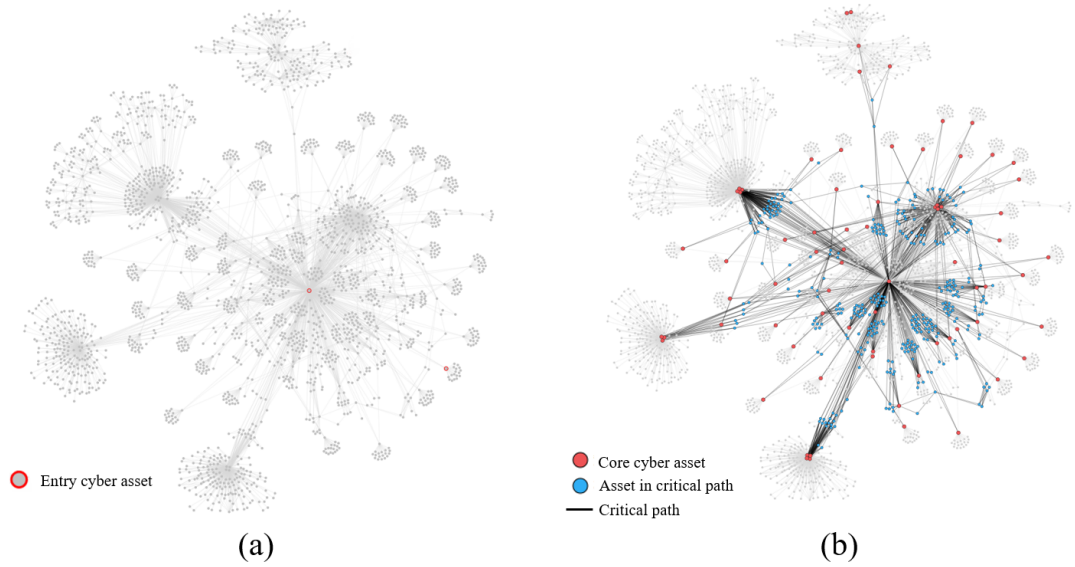


Figure 2.4 Visualization results of CAG 4. (a) visualization result with highlighted entry cyber assets; (b) visualization result with highlighted core cyber assets and critical paths.

Table 2.8 Entry cyber assets of CAG 4.

Asset id	Asset name	Asset type
IP_7e730b193c2496fc908086e8c44fc2dbbf7766e599fabde86a4bcb6afdaad66e	23.82.xxx.xxx	IP
Cert_6724539e5c0851f37dcf91b7ac85cb35fcd9f8ba4df0107332c308aa53d63bdb	6724539e5c	Cert

Table 2.9 Core cyber assets of CAG 4.

Asset id	Asset name	Asset type
IP_20cb513ab710daecf65fd60e07c536697f519553f3bf711fdbbae2a11ec57c7b	142.91.xxx.xxx	IP
Cert_6724539e5c0851f37dcf91b7ac85cb35fcd9f8ba4df0107332c308aa53d63bdb	6724539e5c	Cert
Whois_Name_d93e740c6670760fce94cd3199e7e24bae82b16739c488d8191290ef7b403e0e	duxxxxxng	Whois_Name
Whois_Email_4bc12ad46aae48bd6c12bcbf626389c6e8d7733ace2fce23162223ebf80d285b	u5834xxxxx@xxx.xxx	Whois_Email
Whois_Phone_dc202b8538a1769e0d2a76acaea73dbd157b63100970088c1d5f9dfffc3fec59	+86.133xxxxx	Whois_Phone

Whois_Phone_edd207d24f6a00614700f6438b24f816a00b0f4357136a26e3cf89162a7f4a51	+86.62xxxxx	Whois_Phone
IP_057a478eceb14c597622878ccb9824f8548ca2b1d82766ab92f3ac387c8fc650	142.91.xxx.xxx	IP
IP_9c7f3a5ef80a17c1b9c24e5c4c81f949ef3d893ae9db7ac553ab193ce5cc88a5	142.234.xxx.xxx	IP
IP_4ea0cbe0ae9dc770e66a956fbe088134aa68b55b43ad9b1f60a7c66ae661b763	142.234.xxx.xxx	IP
IP_3412e9509da22092aafd094b425d73795d333ea97c824668deaab3a059df3640	142.234.xxx.xxx	IP
IP_7165dd0b4a443b64410cc6513d0e50efe3d5a9c1dd73f41b73b59e9a409dfcf3	142.234.xxx.xxx	IP
IP_9e443b8ba2f05c0412f10439fe4c4248f6fd13d1a9d3fdaa0a680e6dec05f9e5	142.234.xxx.xxx	IP
IP_7e730b193c2496fc908086e8c44fc2dbbf7766e599fabde86a4bcb6afdaad66e	23.82.xxx.xxx	IP
IP_eb31538c3780105d5eadffd6e2e45d2a3254954c08aef1f01f4a144d09f4f7c0	23.82.xxx.xxx	IP
IP_811e6921e4d1ae943639d6b7106a7ae8c9621c4802ae6e92cc3f9e33e1f6c15d	142.234.xxx.xxx	IP
IP_360d0f5dba08d72ea4a745954ab94178b8b628c9845907c5d7e48466f8378fc4	142.234.xxx.xxx	IP
IP_be6f0363f06e36e64c83570d449198813e221309d760d578fd3c2ca85320e92e	142.234.xxx.xxx	IP
IP_9d15d67d84f9c11da315d2de4f95c301fb8b2874ff4b63b5e6cd0c335dc89c4e	23.82.xxx.xxx	IP
IP_48cc8fb4cbae8e075772452280942726e4005062f2006eba0f8a5e97307a2e44	23.82.xxx.xxx	IP
IP_4714c2a981c80adb2d80c504cad2e9a0b4cdc35d1798f759c07b9ea3aa612382	23.82.xxx.xxx	IP
IP_781d31131cbed86d08290734aac8a394037a00277b0588f16510bd2bae691ed2	23.82.xxx.xxx	IP
IP_7cb46c2a9ddcfdfabad5768916fdb49c54230a9eb41dc96e20f682313d8eeffc	142.234.xxx.xxx	IP
IP_04023c1aeceae5ac951ca1130a324cf0475181ac060f5f0203f994aee93cd869	23.82.xxx.xxx	IP
IP_6d2cb9d60f491d5384166b2c4c8837267663a385e2433ff49588f42ef44a1a42	23.82.xxx.xxx	IP
IP_ac1ccc9721107587b330cae0ef1446269a02ce708189dccca916e4d6a01bb4d1	147.255.xxx.xxx	IP
Whois_Name_2980849521de9f8397063dc0a9ccda4e2d29c6ce926e1fa293fba125e409ed25	linxxxxxfei	Whois_Name
Whois_Email_c816181a3dbdb584ac229204d2c70bda193513f80ff68f810e484dec1eb37227	linzixxxxx@xxx.xxx	Whois_Email
Whois_Phone_aa84858e660bf880b5f8608159a30b708693659150edc98c01b6763ac762fc2f	+86.136xxxxx	Whois_Phone
Whois_Phone_ef24501a5fc203424f080e57f022a15650303a4dde633eeab401cbe4ee8e2cd4	+86.85xxxxx	Whois_Phone
IP_43fb28c11ddee1a568d667fb8ebcbbf00922be6d05477d22bb87b6ba8ef1d2b6	142.234.xxx.xxx	IP

IP_d0fcc120e6549f0ea3f32a8cacef5500ec36130a5740efbe303329ab3f2e4e1e	147.255.xxx.xxx	IP
IP_7b9954624b78d63b870d3d03daea2c89620ce866901c3affc07554fde1aa5f7f	147.255.xxx.xxx	IP
IP_f6ab81a09c02772b2782b93a3ff98fd5f52960f49df7d384c545fb9ad1d3baee	147.255.xxx.xxx	IP
IP_c8c1b9572d1097792254051a91fc2b667411b1dbefd721a603c6fa99bbe1d593	23.82.xxx.xxx	IP
IP_2ae1f89e74c2bd2a9e5b5f3f5055932716116da9ea12c07647ac4a376ec97c9c	147.255.xxx.xxx	IP
IP_6a7072f17c61d8bdc5946890375a3b4c7168a7b34cb2f5559f8b95d269e881b1	23.82.xxx.xxx	IP
IP_ea400d44363c5f7bf0eeea84877353fbaf386305c0336a0367ffd9069b2982ad	147.255.xxx.xxx	IP
IP_11d326ee33c8f657ac4a745e6b20d10a537810ddc0dd2f50dece068413b1e88a	147.255.xxx.xxx	IP
IP_c6eff0c2bf2309e08475f9a236d6f8a8f3c49bb19fc8b7215fed332cbfd49767	147.255.xxx.xxx	IP
IP_4ba2451625b0f5981d9f32a67068d9ac5163eeead194e7b04c6670aab347e772	147.255.xxx.xxx	IP
IP_1b0a108effa74a5cf827da705f9ffe85f638a739afc9e9d5604e590c3c4621d3	147.255.xxx.xxx	IP
IP_c4e5b75e65cd9015613385dade65ca1cfdacc91aeba6be18dd16436bc89fc43e	23.82.xxx.xxx	IP
IP_02d188e9090ee3d867b019326c62ff8270ec2e8d090a73165e8bd3ce1af3c735	147.255.xxx.xxx	IP
IP_51a69bde990d2628f95ccdf47434d1c70dc803a48c6d50b96e64c22e478e34ed	147.255.xxx.xxx	IP
IP_8b59e4330973b564101ac61fb4aecca4926a58b7e33ac0183b3f57123e269c96	147.255.xxx.xxx	IP
IP_90a4481bcfa1e55e5a016b8470ecb713ebadba193e0f2d6b2e9c4893e1eb754c	23.82.xxx.xxx	IP
IP_ba509acc37b1a4ac6687b2f79ea6188b8dfcc09c43e3d03c0209bf7a35d74569	142.234.xxx.xxx	IP
IP_fd274aab417beb2bbd649cce02bd98f38483b40fbc781747cdeac4a83bb0a58c	147.255.xxx.xxx	IP
IP_8c412ab36e7a6d560a51eeee4d5c7129b80d50793de8e46278482fad5f99f67c	147.255.xxx.xxx	IP
IP_f20ee5251b1db59eaced2666e489877dc1a16dc6115787dd2908fe78c201751c	142.234.xxx.xxx	IP
IP_5bf45e726aa028411949919a8f4269e0bb4c20f5799244e3f40814eca8c5ab98	23.82.xxx.xxx	IP
IP_aa2457df14baa80b6aae2e1c2a1793ab629ab9c490abeed811964ce5804dd0ab	23.82.xxx.xxx	IP
IP_e9b2f6b3d2f3fc2c5b9c8d0328e8e3f6e68cf2d1bbaa847eb0a63e497dc3e92e	147.255.xxx.xxx	IP
IP_3fa35c026b1deb86be5a4968f285a32c1c3707e0deae161f9c988015503b98bd	147.255.xxx.xxx	IP
IP_2574b7925df59e907fb004023981ba66a2cc081b4be8924efa3801531b69cb26	147.255.xxx.xxx	IP

IP_f48f05535f3ee00572a97f62ab30bbfbeb69f2d2699915365ff1801ac64cbf6c	23.82.xxx.xxx	IP
IP_6fb775ff1d42248cec2de73b3c0469e81bd1c11751f9801b7b69a0bcc12deec8	23.82.xxx.xxx	IP
IP_ff4cc9d69d7a7f05c5dd17b6ebe8645113f165bd40e2c5b7a87a4b38329b93b3	23.82.xxx.xxx	IP
IP_8ab8d26901e63cc39d8a24f6596379c7d53897a92fb420f1062f6a7fef74ffdb	23.82.xxx.xxx	IP
IP_b38dfdd8637bf25278f0b9a339ab922da1a555f9bc62851c265ed2030cb9a242	23.82.xxx.xxx	IP
IP_fabfdc32b5b020765c7a804720639ded89269eae1dd50c3fa0e1bc4633888a0f	147.255.xxx.xxx	IP
Whois_Name_4084859b020fb36a6752c9ebd8e007ba6276559caa88f61a6ed70cb623a7e472	zixxxxxin	Whois_Name
Whois_Email_a914b9e5452d187e33fb40035e9bb84e965f94781efb4c314b79aed5bff83a5e	linzixxxxx@xxx.xxx	Whois_Email
Whois_Phone_78013cb7b815afd27fe315621f6a7c07b98b6ce356e1358f203adcc14c31e897	+82.25xxxxx	Whois_Phone
Whois_Name_4b453b66cf224e44221aa553fdf2ddb3fd268d0369f24b65e8a5460bb8459fc8	xiaxxxxxtao	Whois_Name
Whois_Email_63356c67b755652e57c5022f9b3979f44f33262714a3a5b22e48e41a553cf911	10293xxxxx@xxx.xxx	Whois_Email
Whois_Phone_4e0f95d9b503e761e41920f9077b7575a9672f06fc22b48e07436762b8c11cac	+86.133xxxxx	Whois_Phone
Whois_Phone_3b79ecbf3fd8117fe9da37efca11caea1e205f5ecafcd318bafbea0f6cd0af17	+86.592xxxxx	Whois_Phone
Whois_Name_40fdcd1a22c4d828cbd5486259bd72aef82cc2919d7e746bc05b67b48138e9cb	xiaxxxxxgao	Whois_Name
Whois_Email_c955367c07a39201886fba23f4190287efe4817760d500effb03c1c14dc4f56f	kpp88xxxxx@xxx.xxx	Whois_Email
Whois_Phone_6e1c815adbd2826806d8399060724141217117591bd1e7b3d86302ba8edaa8a8b	+86.132xxxxx	Whois_Phone
Whois_Email_7d373db7aae622498e8e79bb93485b161d625f82d806e45d40ea8c33bdb4ee00	wkk88xxxxx@xxx.xxx	Whois_Email
Whois_Phone_bd607f55b38dd92bc2e3450e2034954689d101a7ac7a144a30696efaa8c8db6f	+86.010xxxxx	Whois_Phone

2.5 CAG 5

CAG 5 is a large-sized graph mined with four given entry cyber assets. Figure 2.5(a) presents the visualization result of the CAG, including 1,079 nodes and 2,345 edges with the entry cyber assets marked by red borders. Figure 2.5(b) is the visualization of the CAG with highlighted core cyber assets and critical paths. The list of entry cyber assets and core cyber assets of CAG 5 are provided in Table 2.10 and Table 2.11, respectively.

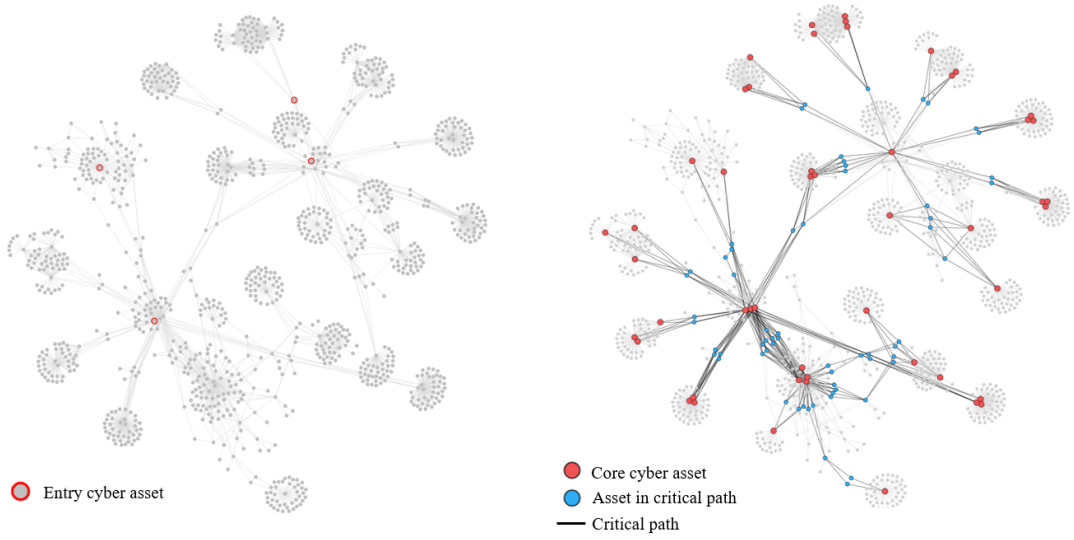


Figure 2.5 Visualization results of CAG 5. (a) visualization result with highlighted entry cyber assets; (b) visualization result with highlighted core cyber assets and critical paths.

Table 2.10 Entry cyber assets of CAG 5.

Asset id	Asset name	Asset type
Whois_Phone_fd0a3f6712ff520edae7e554cb6dfb4bdd2af1e4a97a39ed9357b31b6888b4af	+86.400xxxxx	Whois_Phone
IP_21ce145cae6730a99300bf677b83bbe430cc0ec957047172e73659372f0031b8	3.234.xxx.xxx	IP
Domain_7939d01c5b99c39d2a0f2b418f6060b917804e60c15309811ef4059257c0818a	7939d01c5b.com	Domain
Domain_587da0bac152713947db682a5443ef639e35f77a3b59e246e8a07c5eccae67e5	587da0bac1.com	Domain

Table 2.11 Core cyber assets of CAG 5.

Asset id	Asset name	Asset type
Whois_Name_af9c8790603b2045d997ea7062e2fd93c931560ae48932b95f20085663878464	jixxxxxao	Whois_Name
Whois_Email_2e7c374df8dfbeb2a499b2686e7a448539e49a3ea9bd97ece8de39d1f1a45856	laosixxxxx@xxx.xxx	Whois_Email
Whois_Phone_4939081cd8c3df7854212ca0855ddcf12a4a1ae4b7eba4c6dbdae8ae2507a03b	+86.0454xxxxx	Whois_Phone
Whois_Phone_401b35fa2f213ee5afe58d538064b40640caa69a4acbef2e0d5fa90eef5cc39c	+86.454xxxxx	Whois_Phone
Whois_Name_ea40376482fb013b3f713cb9f36dcbca1807bde5173fa57db7778f027e3ed0e5	jiaxxxxx wu	Whois_Name
Whois_Email_d7537914ce0c8d6b94c8860e2627871d80464ebad7a64c0bb796492e7adb9767	adminxxxxx@xxx.xxx	Whois_Email
Whois_Phone_fd0a3f6712ff520edae7e554cb6dfb4bdd2af1e4a97a39ed9357b31b6888b4af	+86.400xxxxx	Whois_Phone

Cert_df4c8b036186629dd62df86c1dd01de9912cafb8601d6c35ee954c1ec7204594	df4c8b0361	Cert
Whois_Name_824cd2dc385e9c8f630d60d3e673b2db5f9feaa201e6cc964c76319303599c2e	baixxxxxeng	Whois_Name
Whois_Email_a741784a50a806f82a67faccf5e24257737d5ba2525ae6a58a28edcf0946de47	20198xxxxx@xxx.xxx	Whois_Email
Whois_Phone_4d2db1fd924a7c375ef266f7469f73b92bcb34cd91b9e7c6ee3155341856935b	+86.132xxxxx	Whois_Phone
Whois_Name_313ec81fe1a2a87fca490a8070ecc54cda1251dbc4bff07826d8c6447e0b16f0	zhixxxxxang	Whois_Name
Whois_Email_d00bffb916dd17942e14e840058f853f78c3bc68cc1d52916cdee541e528f9e2	31857xxxxx@xxx.xxx	Whois_Email
Whois_Phone_64f48962fb094de63f57268d6039c2535c9dbd1262d3346b39fcbe240a74efe4	+86.158xxxxx	Whois_Phone
Whois_Name_d14901929e03d9e2c4b3a3bdca4f38ba56a069697c6664cb1fbf44e16bd0a0ce	xuexxxxx lu	Whois_Name
Whois_Email_5d30fae3931d90be89faf76350486daf55ec8d31fb5e5c3e0f1cfa3e61d97af2	xh557xxxxx@xxx.xxx	Whois_Email
Whois_Phone_b2c3242dc164d74e1eacbfc4f210e939ace8a9ca15545742eac891fd75b81d32	+86.132xxxxx	Whois_Phone
Domain_7939d01c5b99c39d2a0f2b418f6060b917804e60c15309811ef4059257c0818a	7939d01c5b.com	Domain
Cert_23bc88ac81643991c5222159985f22301e7df3cb3acabcf879a83927fae56d2e	23bc88ac81	Cert
Whois_Name_5cc3c37b39f7572b30515823086dcce0d80298e5cd35576e6dcf4189a1f30254	yaxxxxxma	Whois_Name
Whois_Email_c7a47fa03378d277f88e66df5680f566d6b8f0448d07236cd1ef14bc0a7548aa	gd222xxxxx@xxx.xxx	Whois_Email
Whois_Phone_1f0d2c9c2a15238654bbccfef619ad98d865a962b7e438756c1541c4dd8df3e7	+86.131xxxxx	Whois_Phone
IP_21ce145cae6730a99300bf677b83bbe430cc0ec957047172e73659372f0031b8	3.234.xxx.xxx	IP
Whois_Name_5a90e75628e81244b692fe1ce7dbc3ea16e3e9dfb2a6312f6ddb2f3209dd150c	minxxxxxong	Whois_Name
Whois_Email_14ff64bee5f766738802bb272f478746be42ea08a256a947ec50b2e471de0af7	enamexxxxx@xxx.xxx	Whois_Email
Whois_Phone_d4433a072ceae2c0c4ce3b827b0bb7a16a57cc9c921795b73d493eb4c6bd05d9	+86.131xxxxx	Whois_Phone
Whois_Phone_d09d0994cef3553708537f9e83b1cb339347fb529a557d0be0ff6a7961bb561d	REDACTED Fxxxxx	Whois_Phone
Whois_Name_8170a48a4ca837cbfcfe6126afa36c9ed320dd4c0d7f9af7bd8755b0d97028cd	Domaixxxxxrator	Whois_Name
Whois_Email_f425078fdb678fb1e9d47ea57aa0b9eb10a78fa472e74a2d0cf3f9ecb5cc506a	dotmexxxxx@xxx.xxx	Whois_Email
Whois_Phone_2e64889b700fd2d6f8ac7c42e30ad76e4e4f178ef8a4b286a4bff87ef4cf7fbb	+86.592xxxxx	Whois_Phone
Whois_Name_9de4f742d3fc5da51c455879cc07287e27ef81ec8e73cbf8d8b4b189c8743c69	qixxxxxhe	Whois_Name
Whois_Email_d4061b5ae4b32db680a27ed3cac5c20165be5c6178de002a807b4804c16a153b	lmangxxxxx@xxx.xxx	Whois_Email
Whois_Phone_a423b886f65d70f0816d40c090d6736016813f7d176ac887ee0a209343fd0a70	+86.176xxxxx	Whois_Phone

Whois_Name_4f9f6ee2e8509f96c2a0b0b1a3405164d248036da3fc1b42773f9922fb7971c3	junxxxxxiao	Whois_Name
Whois_Email_65a430f2e0de1c24b6c3f64e74bc176db334bf9bd8526aaefa36bb69232e65c3	huangxxxxx@xxx.xxx	Whois_Email
Whois_Phone_ea843606d2cd820d0de8c7f61e6779aefe1ccf6f0b96b8275bc2e958d1b137bd	+86.185xxxxx	Whois_Phone
Whois_Name_c5f19d1642581661b3940aa98c60018e60fc51d75c0c64e568d918426af3d371	zhaxxxxxang	Whois_Name
Whois_Email_a73d8df17dc3f6ef9be450a03c0e75f23c0eb4f69a00a69ed0c27b5f503022f9	36200xxxxx@xxx.xxx	Whois_Email
Whois_Phone_da2b70bac336fd00ebd7d366a85caed6dc9c6b666b544e12af67f24496ee4e2e	+86.156xxxxx	Whois_Phone
Whois_Name_dd05ed92c4e9a858d9af922c9046350210627b50cf84919e7bc633eeff7a49d7	xiaxxxxx wu	Whois_Name
Whois_Email_f0f809953a6581e4422b85426a23906179659a9c3b3cd884a95fbd5baa1c8ce0	c1821xxxxx@xxx.xxx	Whois_Email
Whois_Phone_3ef3a7cc6ae45bdba4449cdc610283758167c3ae9f63dc2b10974e608bb74c86	+86.130xxxxx	Whois_Phone
Domain_4e1add55e97e79c460f43466801e18df214f9a1dd88259fb6fa8bf7c39aeeb63	4e1add55e9.com	Domain
Whois_Name_f60987f4b09719f245531d3d7ff07fca3801827378fade2e7e7ae54f769c18e9	Legaxxxxxment	Whois_Name
Whois_Email_72cceabb9eee6803eaed8f7daa8ca403afcc491e457e9f0a68a22fd75098e20c	hostmxxxxx@xxx.xxx	Whois_Email
Whois_Phone_94e122e5cb723fecec23c9c1747b5eaef65471f3a54d61a98aa00d218ca05ece	+1.206xxxxx	Whois_Phone
Whois_Phone_8bf2f2c901a2ca39607935c1c4ca65685caa287feb93bb607ac0012793b37ca5	+1.206xxxxx	Whois_Phone
Whois_Name_d93c941eef173511e77515af6861025e9a2a52d597e27bf1825961c2690e66cd	Domxxxxxmin	Whois_Name
Whois_Email_fd8ba4fe69bd059e6ffe78e02e39d0d1b4dc56bb0ea034fb4d93ec75cce83483	suppoxxxxx@xxx.xxx	Whois_Email
Whois_Phone_46d7be8975e9f5690e60e65f7547fb87293b233b3fd59b6332e6c98bcb4f2702	+1.720xxxxx	Whois_Phone

3. Domain knowledge

To help data users to start-up data analysis, we provide a set of basic domain knowledge for CAG mining, core cyber asset identification, and critical path identification.

3.1 Domain knowledge of CAG mining

Domain knowledge 1. The mining range for a given entry cyber asset should be limited in the 3-hop neighborhood of the entry cyber asset.

Domain knowledge 2. CAG mining should consider the strength levels of relations. The cyber assets that are correlated to any entry cyber asset through extremely high or

high strength relations within 3 hops, moderate strength relations within 2 hops, and weak strength relations within 1 hop, are candidates that can be included in a CAG.

Domain knowledge 3. Cyber assets outside the 3-hop neighborhood of any entry cyber asset can be included in the CAG only if they are correlated to the extremely high importance cyber assets within 3 hops through extremely high strength relations.

Domain knowledge 4. Isomorphic neighbor nodes of core cyber assets are suggested to be sampled to reduce the overall size of a CAG.

Domain knowledge 5. New entry/seed cyber assets can be newly added during a CAG mining process.

3.2 Domain knowledge of core cyber asset identification

Domain knowledge 1. Cyber assets of moderate importance level are generally not considered as core cyber assets.

Domain knowledge 2. Cyber assets with more than half of the relations being weak strength are not generally considered as core cyber assets.

Domain knowledge 3. Domain cyber assets that are connected to more than two IP assets are probably deployed by CDN (content delivery network), which are not considered as core cyber assets.

3.3 Domain knowledge of critical path identification

Domain knowledge 1. The paths between two core cyber assets with a length longer than 4 hops are not considered as critical paths.

Domain knowledge 2. The shorter path/paths are more likely to be critical paths if multiple paths exist between two core cyber assets.

Domain knowledge 3. The path/paths with higher strength levels are the more likely to be critical paths if multiple paths exist between two core cyber assets.