Received: 30 October 2023 | Revised: 14 December 2023 | Accepted: 21 January 2024

IET Cyber-Physical Systems: Theory & Applications

DOI: 10.1049/cps2.12087

ORIGINAL RESEARCH

Leveraging graph clustering techniques for cyber-physical system analysis to enhance disturbance characterisation

Nicholas Jacobs[1] | Shamina Hossain-McKenzie[1] | Shining Sun[2] | Emily Payne[3] | Adam Summers[4] | Leen Al-Homoud[2] | Astrid Layton[3] | Kate Davis[2] | Chris Goes[1]

[1]Cyber Resilience R&D, Sandia National Laboratories, Albuquerque, New Mexico, USA

[2]Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA

[3]J. Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, Texas, USA

[4]Electric Power Research, Sandia National Laboratories, Albuquerque, New Mexico, USA

Correspondence

Shamina Hossain-McKenzie.
Email: shossai@sandia.gov

Funding information

Sandia National Laboratories, Grant/Award Number: EHS LDRD 229324

Abstract

Cyber-physical systems have behaviour that crosses domain boundaries during events such as planned operational changes and malicious disturbances. Traditionally, the cyber and physical systems are monitored separately and use very different toolsets and analysis paradigms. The security and privacy of these cyber-physical systems requires improved understanding of the combined cyber-physical system behaviour and methods for holistic analysis. Therefore, the authors propose leveraging clustering techniques on cyber-physical data from smart grid systems to analyse differences and similarities in behaviour during cyber-, physical-, and cyber-physical disturbances. Since clustering methods are commonly used in data science to examine statistical similarities in order to sort large datasets, these algorithms can assist in identifying useful relationships in cyber-physical systems. Through this analysis, deeper insights can be shared with decision-makers on what cyber and physical components are strongly or weakly linked, what cyber-physical pathways are most traversed, and the criticality of certain cyber-physical nodes or edges. This paper presents several types of clustering methods for cyber-physical graphs of smart grid systems and their application in assessing different types of disturbances for informing cyber-physical situational awareness. The collection of these clustering techniques provide a foundational basis for cyber-physical graph interdependency analysis.

KEYWORDS

critical infrastructures, cyber-physical systems, data analysis, decision making, directed graphs, graph theory, hardware-in-the loop simulation, power system security

1 | INTRODUCTION

Clustering analysis has played a prominent role in data analysis for a multitude of domains and has a long history of providing insight into unexpected relationships and interdependencies. Clustering is an exploratory technique using algorithms to sort large datasets into groups defined by statistical similarities. The goal is to identify useful patterns or relationships within data that are otherwise not readily apparent. The resultant clusters or groupings of data can be achieved through a variety of measures and models; each cluster's significance depends on the application area, data availability and type, and understanding of clustering goals and needs [1]. With these considerations, clustering analysis is an extremely effective tool. For example, one can cluster different data objects that exhibit similar behaviour when certain events occur or understand

which data objects have no association or dependence on one another.

In power systems applications, clustering analysis has often been used to provide insight into relationships between different grid components (e.g. generators, lines, loads), impact of different disturbances (e.g. faults, outages), and similar, localised behaviours (e.g. modes, stability margins) [2]. These types of clustering analyses focus on power system data such as nodal/edge connections and time-series measurements.

Yet the power grid and other critical infrastructure systems are becoming increasingly cyber-physical, rapidly evolving with smart grid technologies, wide-area monitoring capabilities, and advanced automation. This strengthening cyber-physical mutuality does not limit cyber disturbances to the cyber system (e.g. communication network) and physical disturbances to the physical system (e.g. power system)—disturbances can propagate between systems [3]. Similarly, operational system changes can affect both cyber and physical domains; cyber-physical situational awareness (CPSA) is needed for both normal operation and during disturbances [4].

An improved cohesive cyber-physical mathematical modelling notion of the system that links its state with its data is needed [5], as described in different applications including cyber-physical state-based modelling and analysis of large-scale power system infrastructure. To help improve CPSA and provide a more quantitative assessment of cyber-physical system relationships, we study *graph clustering* techniques to apply to cyber-physical data. This paper focuses on grid cyber-physical graph models and the application of several different types of clustering methods. These methods are compared for their diagnostic value in assessing cyber-physical interdependencies during different types of disturbances. Understanding where the clustering methods overlap and where they have gaps for assessing these disturbances will provide next steps in achieving more comprehensive CPSA.

Specifically, this paper will focus on grid cyber-physical graph models and the application of several different types of clustering methods. These methods will be compared for interpreting cyber-physical interdependencies during different types of disturbances. Understanding the overlap and gaps between these different clustering methods for assessing these disturbances will provide next steps in achieving more comprehensive CPSA. This comprehensive CPSA can greatly enhance cyber-physical system (CPS) security with its risk characterisation which can improve CPS vulnerability analysis, attack detection, and adaptive response techniques. Additionally, deeper insight into nodal and edge relationships in the cyber-physical graph can be achieved that can provide criticality rankings, identification of important interfaces in the CPS, and paths for cascading disturbances.

The main contribution of this paper is to develop a foundational basis for graph-based CPS interdependency analysis that uses graph clustering techniques to characterise CPSs during cyber-, physical-, and cyber-physical disturbances. This approach is developed through discussion of prior and existing research in Section 2, development of an emulation experiment with various disturbance scenarios in Section 3, presentation of clustering methods of interest and their results from the emulation scenarios in Section 4, comparison of the clustering results and insight into the CPS interdependencies in Section 5, and conclusions and future work in Section 6.

2 | BACKGROUND

In this section, we highlight previous work on the use of different clustering techniques and graph analysis research as applied to power grids and cyber-physical power systems. In refs. [6–8], the authors consider power grids as graphical network models in which both the cyber and physical networks are considered. In each piece of literature, the authors define and test a different mathematical or theoretical approach to define cyber-physical resiliency of power systems, using graph theory concepts.

For example, in ref. [6], the authors build a dependency graph of the power system by monitoring system calls and traffic between the different system components. These are modelled as weighted nodes in a Bayesian network. The authors rank physical contingencies using graph theory techniques based on power system topology, which include vertex centrality measures such as closeness centrality (to rank generator outage contingencies) and edge betweenness centrality (to rank line outage contingencies). In ref. [7], authors construct a Competitive Markov Decision Process (CMDP) model based on the power system and cyber topology information and relative cyber-physical interconnections. In ref. [8], the authors treat the power system as a graph and define both physical and cyber resiliency metrics based on the power system topology and the communication network, which are assumed to be isomorphic.

In ref. [9], Baranwal and Salapaka work on improving power system management through use of clustering techniques to enhance voltage control in large power systems, which can help in controlled islanding applications that can limit large-scale blackouts. As with refs. [6–8], the authors view the power system as a weighted graph network, with the nodes denoting the buses. The difference in this work is that the edges between the nodes are quantified as the "electrical similarity" between components. This similarity is defined as the "influence" of the two buses on the rest of the network, which is quantified as the measure of system-wide voltage fluctuations resulting from reactive power injections from the two buses. Since this electrical similarity is inherently physical, it can be calculated using power flow equations. Once the graph is developed, control strategies are then developed for voltage control and islanding case studies. Similar work is done in ref. [10], in which the authors focus on controlled islanding applications by using spectral clustering of the power system and performing N-k line contingency analyses with statistical correlations studied as well. In both refs. [9, 10], it is important to note that the authors studied the power system as a purely physical system, instead of cyber-physical, which is the focus of our paper.

In ref. [11], the authors implement and compare different clustering techniques to help identify demand days in the system, where the goal would be to cluster similar energy demands with respect to the time periods selected. The different clustering techniques studied include k-centres, k-means, k-medians and k-medoids, with k-medoids showing the best clustering results. In ref. [12], the authors focus on spectral graph theory and hierarchical clustering techniques to study and analyse islanding in hybrid energy systems (HESs). The integration of renewables has made the traditional power system more complex, thus requiring more complex contingency analyses to ensure overall system resiliency and reliability. Other applications of clustering can include the use of spectral clustering using topology information and electrical grid data to implement "locational marginal pricing, phasor measurement unit (PMU or synchrophasor) placement, and power system protection," according to work done in ref. [13]. Another important application of clustering in power systems is the creation of synthetic power network graphs which accurately represent real power system topology and system characteristics, as developed by ref. [14]. In ref. [15], the authors propose a method of dividing a power network into different zones based on the "electrical distance" between the elements in the system, which is defined by the authors as the "absolute value of the inverse of the system admittance matrix." Other work in the literature includes further analysis and studies related to hierarchical spectral clustering applications [16] and electrical and topological structure of electric grids, where an algorithm is developed that generates an accurate weighted "minimum distance graph" that is similar in properties and structure to real-life power grids, as done in ref. [17].

Building on the existing literature in this field, this paper will focus on comparing different cyber-physical clustering techniques to analyse the inter-dependencies between the physical network and its components and the cyber network and its components.

3 | CYBER-PHYSICAL EMULATION EXPERIMENT DESIGN

3.1 | WSCC 9-bus emulation

The WSCC 9-bus system is a simple approximation of the Western System Coordinating Council (WSCC) to an equivalent system with 9 buses and 3 generators [18]; the oneline diagram is pictured in Figure 1. Additionally, a corresponding synthetic cyber network was created for the WSCC 9-bus system, described in detail in ref. [19], and is shown on the right side of Figure 1. The power system oneline diagram will be referred to as the physical model and the network diagram will be referred to as the cyber model. Together, they comprise the cyber-physical model of this power system, encompassing the integrated power system operational physics, sensing, computation, and communication to and control of power system components [20].

A combined, directed graph of the WSCC 9-bus cyber-physical model is generated using graph-theoretic, power system and network observability techniques. The graph is pictured in Figure 2a. Additional details on this observability-based approach are provided in ref. [21]. To simplify identification, we will use a set of letters and acronyms to denote system components as follows: for physical components, bus (B), load (L), and generator (G), and for cyber components, relay (R), switch (SW), human machine interface (HMI), and control centre (CC). This will allow easier notation for specific components in the results, such as bus 1 being represented as B1 and relay 10 being shown as R10. The graphs of cyber and physical networks are based on their configurations and power flow direction. The integration of cyber and physical networks depends on digital protective relays and other control and sensor devices that bridge the cyber-physical domains.

Digital protective relays have both communication and control capabilities and thus link the cyber and physical networks. They deliver data among cyber networks and control

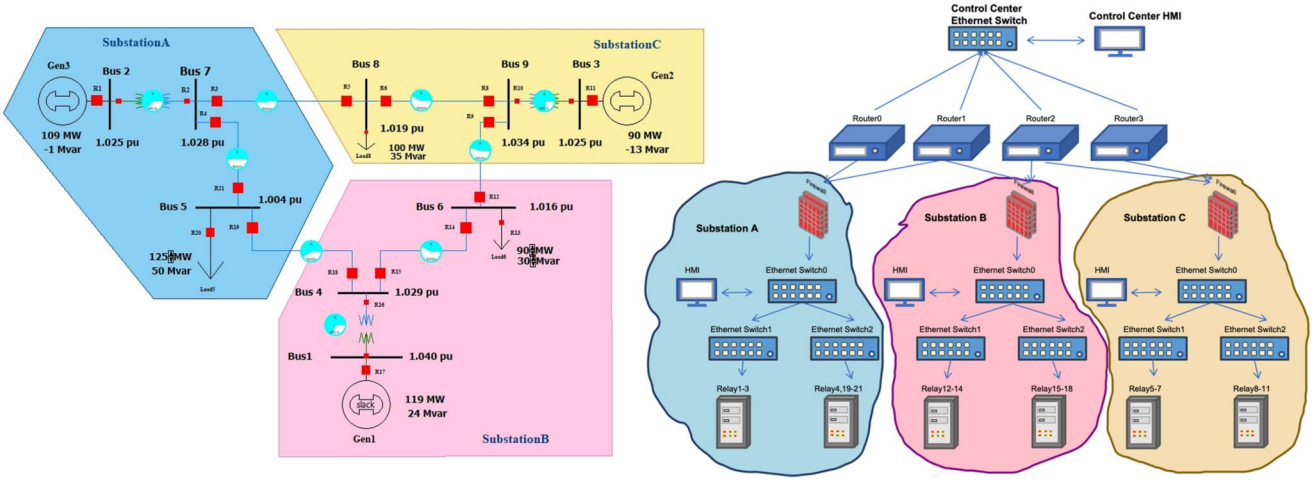
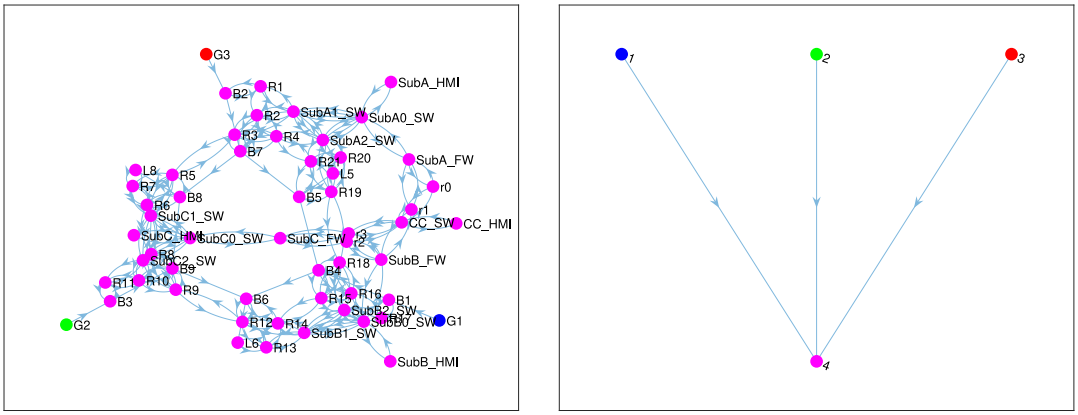


FIGURE 1 Oneline diagram of WSCC 9-bus physical power system with labelled relay placement and corresponding cyber network.



(a) (b)

FIGURE 2 Graph components and condensation graph for full system. (a) Graph components for full CPS. (b) Condensation graph for full CPS.

physical devices to maintain the stability of physical networks and provide system safety. For each substation in the WSCC 9-bus case, there are relays protecting nearby components; thus there is a directed connection from relay to bus or relay to load that serves to integrate the cyber and physical networks. The relay locations are shown in Figure 1. Details for this approach are provided in our prior work [21]. This results in a model that can be readily applied to compare the clustering methods described in this work.

3.2 | Cyber-physical disturbance scenarios

To compare between the different techniques studied in this work, we examined how different modifications to the cyber and physical graph connections affect the results gathered. As each technique is used to examine different aspects of the graph structure, this can result in differences in the intuition gained about the system. For this, we will use the WSCC 9-bus power system use case along with a hypothetical cyber network as described in ref. [22], with three substations (A, B, C). The scenarios that will be studied involve changes in the CPS graph structure used for clustering analysis. They are as follows:

1. Generator 1 and line 6-9 outage
2. DoS against load 5 and 6 relays (R13 and R20)
3. DoS + line and generator outage
4. Double line outage: line 7–5, line 6–4
5. DoS against control centre switch (cc_sw) and substation C switch SubC1_SW
6. Double line outage: line 7–5, line 6–4 and DoS against control centre switch (cc_sw) and substation C switch SubC1_SW

In Scenario 1, generator 1 and line 6–9 are disconnected, which requires load shedding to prevent over-current faults. Scenario 2 looks instead at a cyber attack on the control network that is aimed at preventing the relays at loads 5 and 6 from responding to any load shedding commands. Scenario three is the combination of both. These scenarios can be represented as modifications to the CPS graph through node or edge deletion. In Scenario 4, a double line outage occurs and no mitigation is deployed. Scenario 5 studies a distributed DoS (DDoS) at two locations and Scenario 6 studies the DDoS in conjuction with the double line outage.

The emulation comprises two key elements. A real-time digital simulator (RTDS) enables streaming C37.118 data from PMUs in the RTDS WSCC 9-bus model. SCEPTRE^{TM}, a Sandia industrial control system (ICS) emulation tool, enables modelling of ICS cyber/control networks and implementation of actual communication protocols such as Modbus and DNP3. The details of this emulation, scenarios, and implementation method are described in more detail in ref. [22]. The physical disturbance data sets, consisting of bus frequencies, are collected from 8 different PMUs in the WSCC 9-bus model. The cyber disturbance data sets, consisting of round-trip times (RTTs), are collected from 3 different relays in each Substation A, Substation B, and Substation C. The tools pyPMU and iperf3 were used to collect the PMU C37.118 and RTT data collection, respectively [23, 24].

4 | CLUSTERING METHODS OF INTEREST AND RESULTS

In our examination of various clustering methods for analysing the structure of cyber-physical interconnections, a few different methods were identified. Throughout this study we will focus on what insights are gained about the system structure, what information is required, and the complexity for each technique. These results will inform our discussion on similarities and differences between the various methods and aid our understanding of how changes to CPS appear and affect the structure of the system under the scenarios described in Section 3.2. The methods examined in this work will consist of examining the graph components and condensation graphs, K-Shell cores, graph modules generated by examining the system as a bipartite graph related to producer-consumer relationships, and clustering from graph embedding techniques.

Each of these methods can be used to analyse various structural properties of the CPS and assist in understanding system interconnections.

All the techniques studied here are applied on a cyber-physical system graph, constructed as described in ref. [21]. A graph G is defined as a set of vertices V and a set of edges E which connect those vertices. There are both directed and undirected graphs, where the edges E either contain inherent information on direction or not. In all cases in this paper, we will be using directed graphs, where each edge $e_i \in E$ is a set of vertices $\{v_i, v_j\}$ which shows interaction from v_i to v_j but not the other way around which would require a corresponding edge $e_j \in E$ with $\{v_j, v_i\}$. We will use the terms vertex and node interchangeably throughout this paper.

4.1 | Graph analysis

We begin our analysis with fundamental graph analysis methods and metrics. Using this foundation we can better study the relationships and information gleaned from the various graph decomposition approaches and related techniques. Graph analysis is used in many applications to better understand system structure, including path analysis, such as for network routing protocols and search algorithms [25], community detection [26], change detection and distance metrics [27], and many other areas of interest. For more complete background on graph theory and techniques used for analysing graph structures, see [28].

There are several structural properties of graphs that could be useful in analysing important dependencies in the system across both the cyber and physical domains. One example is ranking importance of different nodes or edges in the graph for the purpose of keeping the graph connected. There are a variety of metrics that could be used for such a purpose, such as node centrality, commonly used paths, and cut sets, among others. One such method that examines this part of the graph structure is K-Shell decomposition which uses the number of edges attached to each node, otherwise known as node degree, as a way to examine how central each node is. This is further described in Section 4.2.

An additional application of interest is in better visualising complex networks and inferring which nodes and edges are critical to keep the overall CPS connected and operational. By examining graph components and condensation graphs, we can deconstruct overall structure of the graph and how it can be decomposed into subgraphs where nodes are highly connected to each other but may be mostly separate from the rest of the graph.

For this analysis, the structure of a graph is examined through the existence of paths to and from each node. In an undirected graph, where there is no direction associated with each edge, the graph is said to be connected if every node has a path to every other node. In directed graphs, this concept needs to be extended to account for the direction of the edges, where weakly connected graphs mean every node has a path either to or from that node to the rest of the graph, while strongly connected graphs require paths in both directions from every node. Often, only subgraphs are connected instead of the entire structure, and these subgraphs are called graph components. A representation of the graph that uses the connections of the graph components is called a condensation graph.

4.1.1 | Results from WSCC 9-bus emulation scenarios

In Figure 2 we can see how a graph can be decomposed using graph components into sections which are each strongly connected. In the case of the full CPS graph for the WSCC 9-bus system, the majority of the system is highly interconnected in a central component, with only the 3 generators in the system being separate. This shows how the generators and their connections to the rest of the system are critical, which matches existing intuition about power systems where generation capacity and placement are very important.

The graph structure itself changes for each of the disruption scenarios described in Section 3.2. As certain nodes or edges in the graph are deleted or modified due to various system failures, that can result in potential issues or significant changes in the graph structure. Figures 3 and 4 show how the graph structure changes for each scenario.

The condensation graphs for the WSCC 9-bus system show some fairly interesting but simple aspects of the graph. As noted previously, the entire cyber-physical graph is strongly connected except for the generators in the power system, which due to the direction of power flow are represented only as supplying power to the power system. Because of this, each generator ends up being in a graph component on its own. When G1 is disconnected, the number of graph components drops from 4 to 3, showing a very simple way to easily isolate, identify, and measure when important structural changes, such as losing a generator, occur in the CPS. In the DOS scenario, the loads at buses 5 and 6 are disconnected from communications-enabled relays controlling any load shedding that could be performed, and this makes them no longer strongly connected to the rest of the system. In other words, the loads themselves still are connected to the power system but cannot interact with the control network, which splits those nodes out of the strongly connected central component in the graph. This is easily seen by examining the condensation graphs in Figures 4b and 4c. In scenarios 5 and 6 where the control centre switch is DOSed, the control centre is disconnected from the rest of the system, which means that observability of system conditions will be lost from the control centre, as shown in Figures 8e, 4e, 8f, and 4f.

As shown here, graph components and condensation graphs help to isolate structural changes to CPS graphs under various cyber-physical attack or failure scenarios. These results can help in visualising such changes and identify when additional issues could arise, such as disconnections of portions of the graph or where certain edges could become critical in keeping the graph connected.

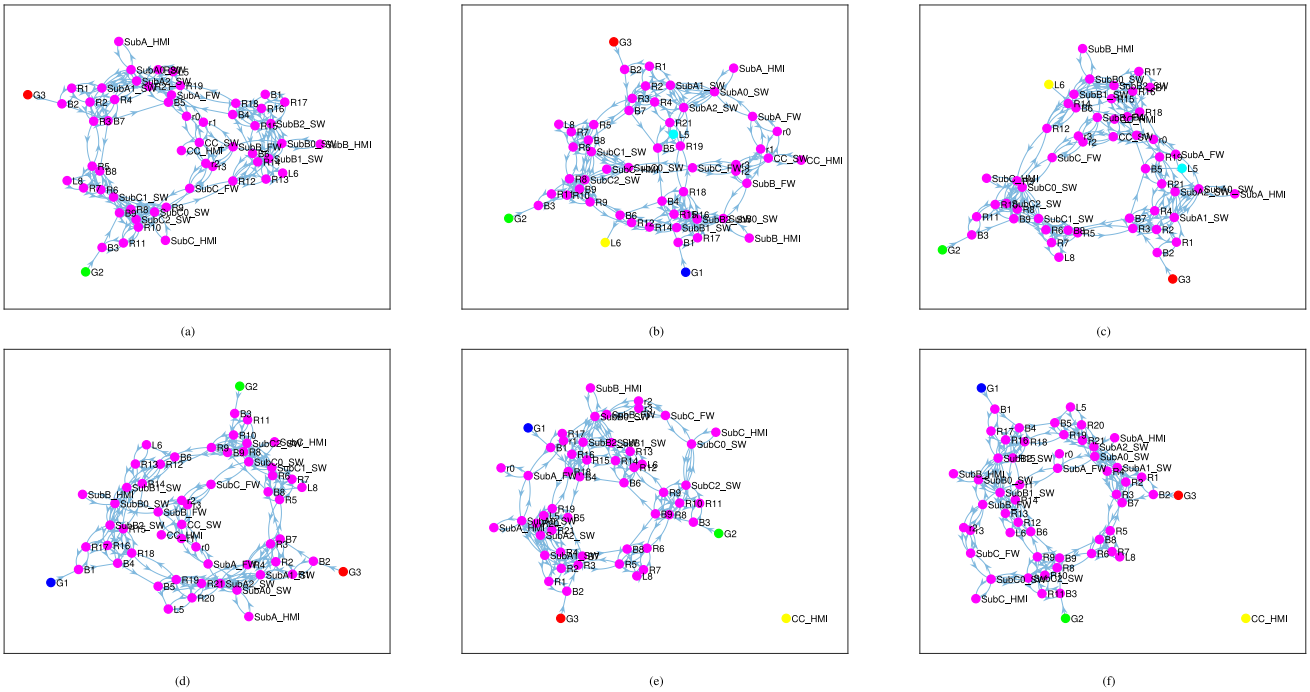


FIGURE 3 Graph components and condensation graph for scenarios. (a) Scenario 1 graph components, (b) Scenario 2 graph components, (c) Scenario 3 graph components, (d) Scenario 4 graph components, (e) Scenario 5 graph components, (f) Scenario 6 graph components.

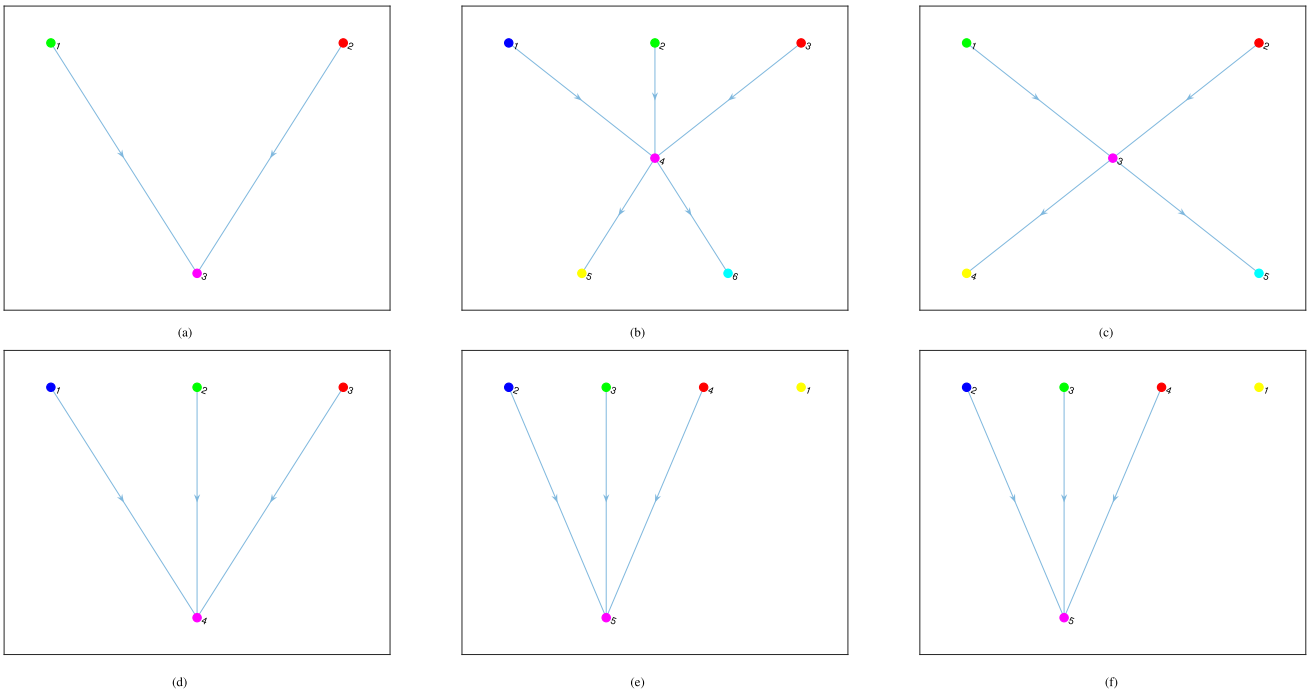


FIGURE 4 Condensation graphs for scenarios. (a) Scenario 1 condensation graph, (b) Scenario 2 condensation graph, (c) Scenario 3 condensation graph, (d) Scenario 4 condensation graph, (e) Scenario 5 condensation graph, (f) Scenario 6 condensation graph.

4.2 | K-Shell decomposition

In the analysis of complex networks, understanding node importance and hierarchical structure is paramount. A fast technique that offers insight into this hierarchical nature is the K-Shells decomposition method [29]. At its core, this technique operates on the principle of iteratively peeling away the least connected nodes. Each shell is described by the number of edges into and out of each node in the graph, which is the node degree. The 1-shell is the shell where each node has a degree of one, the 2-shell a degree of two, and so forth. This layered structure reveals the hierarchy of the nodes [30]. This

technique is often applied to epidemiology [30], social networks [30] and, more recently, cybersecurity [31]. Here, we present the unweighted K-Shells methods on a cyber-physical network.

The K-Shell decomposition of a network is a method used to characterise the structural importance or resilience of nodes in a network.

1. Structural Importance: Nodes in higher-numbered shells tend to be more central in the network. They are more deeply embedded in the network, making them crucial for network connectivity. This does not consider that a node like a generator is central to the physical network but not the cyber network. In this network configuration all the generators are in shell one.
2. Resilience: From a network resilience perspective, removing nodes in the higher shells would typically have a more profound impact on the network's connectivity than removing those in the lower shells. In other words, nodes in the higher shells are often more critical for maintaining network connectivity. This is observed in scenarios 5 and six in which adverse events affect both the cyber and physical systems.
3. Granular Node Classification: Instead of a single centrality measure, K-Shell gives a granular classification. For example, all nodes in the 1-shell have similar importance, all in the 2-shell have a higher importance, and so on. While this may be true for social networks, this does not necessarily apply to power systems.
4. Hierarchy: In many real-world networks, most nodes will belong to lower shells, and only a few nodes will belong to the highest shells. This hierarchy often reflects various roles or functions of nodes in the network. For instance, in social networks, higher-shell individuals might be hubs.
5. Robustness: In terms of network robustness, nodes in higher shells can be considered more robust because they are connected to more nodes that are also well-connected. Conversely, nodes in the outermost shells (lower-numbered shells) are the most vulnerable to disconnection.

In summary, the K-Shell decomposition provides a hierarchical way to classify nodes based on their connectivity and the overall structure of the network. This classification can be instrumental in various applications where understanding the structural importance or resilience of nodes is crucial.

4.2.1 | Results from WSCC 9-bus emulation scenarios

The K-Shell analysis results for each scenario are presented in Figure 5; the colour interpretation of the nodes for each scenario is provided in Table 1.

The pre-scenario K-Shell analysis finds that the original WSCC 9-Bus CPS interface is made up of three shells. Shell 1 contains 7 nodes, Shell 2 contains 20 nodes, and Shell 3 contains 30 nodes. The first disturbance scenario does have a physical impact to the CPS. This results in a change of the shell structure, as seen in Figure 5a. Generator 1 has been disconnected from the system and Shell 1 now contains 6 nodes. The second and third scenarios result in L5 and L6 moving from Shell 2 to Shell 1 (Figures 5b and 5c). Scenario four results in physical change to the system. However, it does not affect the K-Shell output (Figure 5d). Scenarios five and six have the greatest impact to the K-Shell output as seen in Figure 5e,f. In scenarios five and six, where multiple nodes are modified, we see the appearance of a zero shell where the Control Centre HMI is disconnected from the network, as can be seen in Figure 5e,f.

4.3 | Bipartite graphs

Mutualistic networks in nature, such as plant-pollinator or plant-herbivore models, have been found to be resilient to unexpected disturbances [32–34]. Ecologists have attributed this to their unique hierarchical network structure, which also supports highly specialised actors [35]. Ecological analyses of mutualistic networks use bipartite graphs and clustering analyses to model and quantify interaction structures as related to network function and to allocate conservation efforts amongst species [34, 35]. A bipartite model groups actors into two sets and highlights interactions between the two groups. A modularity analysis is then used to identify generalist and specialist species, interaction clusters or modules between species, and general interaction distribution patterns across all species. These modules may be thought of as similar to the clusters identified in the other techniques investigated here. Modularity (Q_N) quantifies the level to which components are clustered into modules based on interactions. The Newman algorithm [36, 37] from by Zuo [38], produces a value from zero to one for Q_N (Equation (1)) where e_{ii} is the percentage of edges in module i, and a_i is the percentage of edges with at least one end in module i, with zero indicating no clustering or modules based on interaction patterns and one indicating clearly defined clusters with no interactions crossing clusters. A modularity analysis also produces a visual depiction of the clusters, using colours to highlight within module interactions. The level of modularity that can be achieved is controlled somewhat by the overall connectance (Equation (2)) of the interface, or the ratio of actual interactions or edges (L) to total possible edges in the graph. A higher connectivity (closer to one) limits the level of modularity that can be achieved while a lower level (closer to zero) makes modularity more likely, but not guaranteed. A bipartite network graph is relevant in cyber-physical systems as it uniquely focuses on the interface between the cyber and physical components, helping to identify any patterns that are impacting performance.

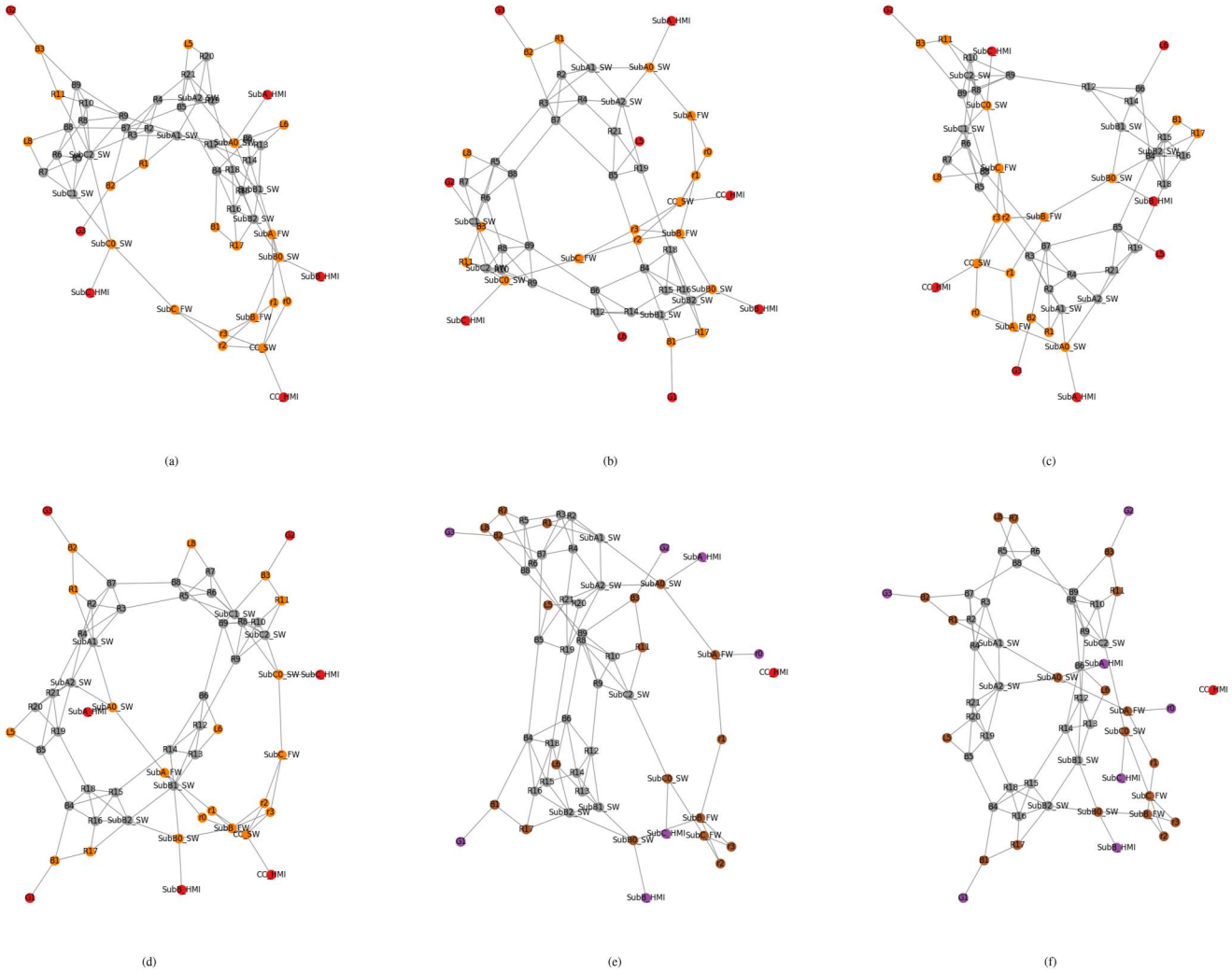


FIGURE 5 K-Shell analysis results for different scenarios. (a) Scenario 1 K-Shells, (b) Scenario 2 K-Shells, (c) Scenario 3 K-Shells, (d) Scenario 4 K-Shells, (e) Scenario 5 K-Shells, (f) Scenario 6 K-Shells.

TABLE 1 K-Shell result colour interpretation for each scenario.

Scenario	Colour	Order of K-Shells
1	Red, orange, grey	1,2,3
2	Red, orange, grey	1,2,3
3	Red, orange, grey	1,2,3
4	Red, orange, grey	1,2,3
5	Red, purple, brown, grey	0,1,2,3
6	Red, purple, brown, grey	0,1,2,3

$$Q_N = \max\left(\sum_{i=1}^{k}\left(e_{ii} - a_i^2\right)\right) \tag{1}$$

$$C = \frac{L}{N(N-1)} \tag{2}$$

4.3.1 | Results from WSCC 9-bus emulation scenarios

This work builds on that of ref. [39] to examine how understanding modules describing the cyber-physical (CP) interface can inform our understanding of disturbance (removal of nodes and edges) propagation between cyber and physical parts of the network. As the entire CP power system is highly complex, understanding what impact the interface design has on the overall network response is highly beneficial. The modularity analysis finds that the WSCC 9-Bus CPS interface is made up of 12 modules, seen in Figure 6 and detailed in Table 2. Three modules contain one of 3 buses (B7, B9, and B4), each having a dedicated set of 3 relays. Three modules have a set of 2 relays per bus (B8, B6, and B5) and the remaining 6 modules are made up of one bus or load connected to one relay. The modularity for the original WSCC 9-Bus CPS and disturbance scenario 1 is $Q_N = 0.898$. This is a high modularity (very close to

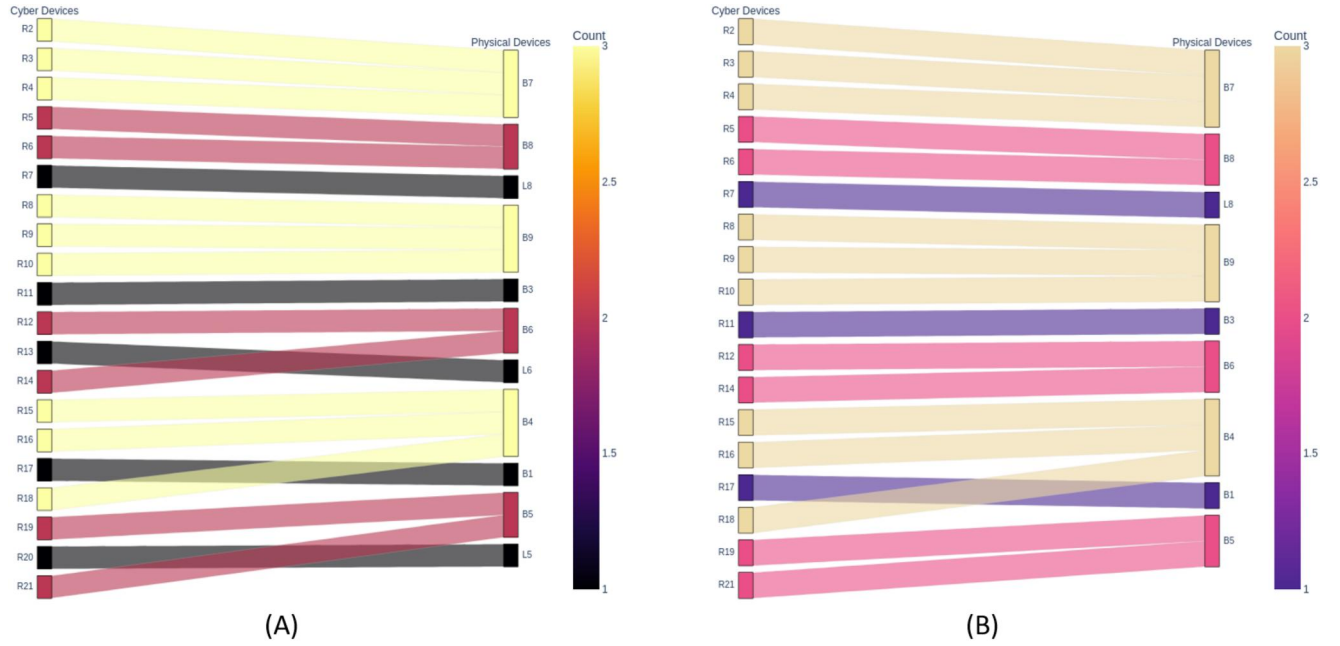


FIGURE 6 Modules found within the cyber-physical interface of WSCC 9-Bus case study for (a) the original, Scenario 1, and Scenarios 4–6; and (b) for Scenarios 2 and 3. There are no connections at the interface that fall outside of a module.

TABLE 2 Modules for WSCC 9-bus CPS system.

Module	Physical components	Cyber components
1	B7	R2, R3, R4
2	B9	R8, R9, R10
3	B4	R15, R16, R18
4	B8	R5, R6
5	B6	R12, R14
6	B5	R19, R21
7	B2	R1
8	L8	R7
9	B3	R11
10	L6	R13
11	B1	R17
12	L5	R20

Note: During disturbance scenarios 2 and 3, modules 10 and 12 are taken out, leaving 10 of the original 12 modules remaining.

Abbreviations: B, bus; L, load; R, relay.

one), a result of the low levels of connectivity ($C = 0.0834$, Equation (2)) at the interface. Mutualistic networks in nature have been found to have low levels of modularity, prompting the question: What would a biologically inspired CPS interface look like and how would it perform - both under normal circumstances and during disturbances? Biological ecosystems have been found to be characteristic of high levels of performance in both of these cases, most likely a result of long periods of reformation to be able to both grow and develop as well as survive [40].

The first disturbance scenario does not directly impact the interface at all and leaves the 12 modules intact. The second disturbance scenario takes out two of the 1:1 modules: load 6 to relay 13 (module 10 in Table 2) and load 5 to relay 20 (module 12 in Table 2). The third disturbance scenario, which combines both the generator and line outage with a DoS, results in the same 10 modules as the second scenario—losing modules 10 and 12. The fourth, fifth, and sixth disturbance scenarios investigated here (a double line outage of B7-B5 and B6-B4, a DoS against the control and SubC1 switches, and the combination of the two) do not directly impact the cyber-physical interface and therefore maintain all 12 original modules listed in Table 2. The modularity for the second and third disturbances, the only two that directly impacted the interface, drops slightly to $Q_N = 0.881$, down from 0.898. This is still an extremely high modularity (very close to one), still mostly a result of the low levels of connectivity at the interface (0.1 for the second and third disturbance). It should be noted that the disturbance scenarios being tested here are not full disturbance propagation scenarios and therefore we are unable to truly see the impact on the interface of cyber disturbances propagating across the interface to impact the physical network components (which would further impact the modules at the interface) and physical disturbances propagating across the interface to impact the cyber network components. Future work will expand to include these simulations.

4.4 | DeepWalk based technique

DeepWalk is a type of graph embedding technique which extracts the information based on a graph's topology [41]. The DeepWalk method generates random walks on the graph, and uses a random path traversing approach to learn the node interdependencies. The term "random walk" is interchangeable with Markov Chain when the walks are time-homogeneous and finite state spaces are considered [42]. In the DeepWalk algorithm, random walk at each step following an outgoing edge is uniformly random, while a Markov chain could generate next state based on diverse distribution functions [43]. The DeepWalk approach draws inspiration from natural language processing, in which proximate nodes express similar meanings and tend to cluster together. Similarly, within a certain amount of time, adversaries may target the most relevant connecting components in the power grid during DOS events. Random walk on a directed graph starts with every single vertex in the graph and visits its connected nodes in the path [44]. Figure 7 shows a simple illustration of a random walk with a starting point 'G1'. The sequence of nodes could be represented as shown in the sample random walk in Figure 7. We considered the random walk process as the potential access path each time the DeepWalk generated.

To defend a power system against advanced adversaries, defenders must seek to anticipate the corresponding components that might be affected. This can help stakeholders to provide a preventive and proactive response, minimise damage and protect the system. The random walk process simulates the attacking path and estimates the likelihood of the starting node given all the passing nodes. The authors in [41] propose a direct method by estimating the likelihood of a starting vertex given all the possible passing vertices during the random walks. The estimated likelihood equation can be denoted as: $Pr(v_i|(v_1, v_2, \ldots, v_{i-1}))$.

The technique is an unsupervised feature learning technique, in which each node has the same possibility of being chosen during the random walk step. It then captures interdependencies between the nodes by using latent variables through the skip-gram method. The skip-gram approach will gather highly pertinent nodes in n-dimensional vectors space. The clustering results of a DeepWalk-based approach will be accomplished with principal component analysis (PCA) and k-means. We are assuming that nodes in the same cluster have high interdependencies.

4.4.1 | Results from WSCC 9-bus emulation scenarios

In this paper, we are setting the number of DeepWalk-based clustering groups to be 6. The random walk conducted by each node is set to be 100,000 times. Figure 8 illustrates the graph representation of DeepWalk based clustering results of scenario 2. The full clustering results are listed in Table 3.

Due to the stochasticity of the DeepWalk process, we noticed discrepancy in the cluster results of test scenarios. Certain nodes appears in different clusters across these scenarios. However, it is notable that there are high similarities across these cluster results. To facilitate the cross-analysis process, we found most of the nodes resided at least three times in the six scenarios, which implies the recurring pattern among the clusters.

For example, {G1,B1,B4,R15,R16,R17,R18,SubB2_SW, SubB_HMI,SubB_FW} components are considered an identical cluster for most are found in Cluster 1, Table.3. Any node that happens to be attacked may lead to a following DOS attack within the same cluster. For the first scenarios involving Generator 1 and the line 6-9 outage, people are seeking strategic reactions for isolating affected components, shifting focus on the potential next targets and assessing system losses. The DeepWalk based clustering results behave like attack-forecasting results and provide insights to proactive security approaches. The operators may take advantage of the clustering results to make a proactive response, prevent a cascading failure, and reduce losses.

Assumption is made that the possibilities of each node being attacked during a single random walk in different scenarios are identical. During attack events, adversaries tend to achieve the most damage with a constrained time frame. Based on our findings, it seems that higher correlation between certain nodes is related to being more susceptible to the effects of each incident. This observation suggests that nodes within the same cluster are more likely to be attacked at the same time. This will be investigated further in the future, and different risk levels will be incorporated into the clustering process.

5 | COMPARISON OF GRAPH CLUSTERING TECHNIQUES

Each graph clustering technique examined in this work decomposes the structure of a graph to expose important properties and relationships between nodes in the graph. These analyses can be performed to examine node importance and used as a selection tool in a network to identify nodes that are

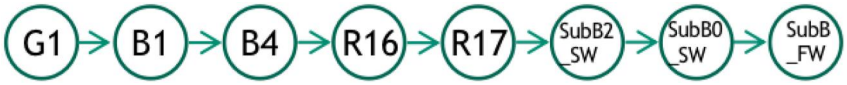


FIGURE 7 Random Walk with Starting Point 'G1' and Eight as the Walk Length. In this experiment, it was considered as a simulated access path.

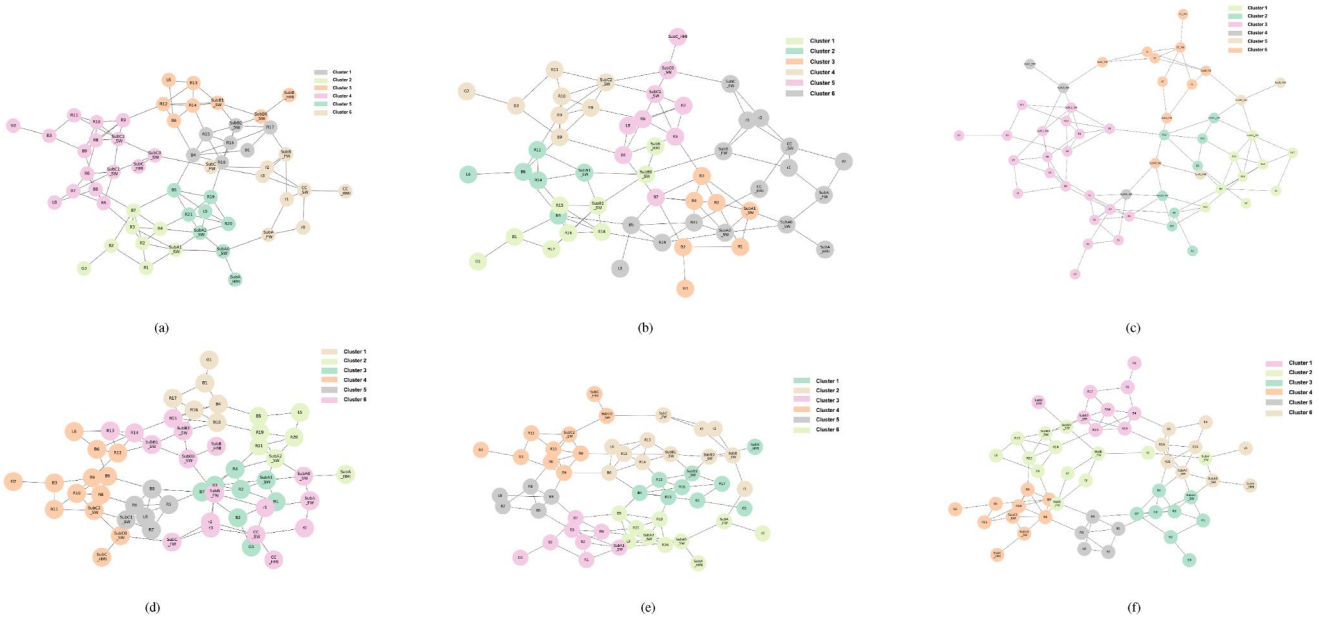


FIGURE 8 Graph Representation of DeepWalk based Clustered Results. (a) Generator 1 and line 6-9 outage, (b) DoS against R13 and R20, (c) DoS + Line and G1 outage, (d) Double Lines Outage, (e) DoS against CC_SW and SubC1_SW, (f) Double Lines Outage and DoS against CC_SW and SubC1_SW.

highly interconnected. The results of these analyses directly impact, and should be embedded within, the risk assessment, management, and mitigation solutions for the system. According to [45], these interdependency clusters can help meet three main objectives to achieve cyber-physcial security: strengthen energy sector cybersecurity preparedness, coordinate cyber incident response and recovery, and provide research results that inform the development of more resilient energy delivery systems. In this context, several key observations are noted:

- **Graph Analysis**. General graph analysis is a well-established foundation with which to begin an interdependency analysis for cyber-physical systems. It provides a "ground-truth" baseline that can be transferred into more advanced applications such as comparisons with, or further development of, graph decomposition and analysis strategies. A basic understanding of different graph concepts and components is useful in clustering cyber-physical systems and interdependency analysis. For example, graph components can come in several varieties, such as strongly or weakly connected. It is also important to consider whether different variations in graph components apply for either undirected or directed graphs, as this impacts the use cases where it may be relevant. Examples are biconnected components and graph cliques. In our results, condensation graphs are seen to be a useful mechanism for visualising important edges in the graph and decomposing the complex graph structure into a condensed and logically verifiable structure. Another observation is that since CPS are highly interconnected, graph components may be almost the entire graph. For example, in Section 4.1.1, the three generators are identified as individual components, which exposes them as important, and it matches our intuition about power systems. Additionally, in Scenarios 5 and 6, the control centre in the communications network becomes split from the rest of the CPS, which is evident in Figures 8e and 8f.

- **K-Shell Decomposition**. K-Shell decomposition is a technique that uses the node degrees to cluster nodes in the graph. Node degree is specifically for undirected graphs, so an undirected graph is required for K-Shell. Directed graphs split the concept into *in-degree* and *out-degree*, for incoming and outgoing edges, respectively. K-Shell can be seen to help identify and rank the importance of nodes based on how many connections they have. A node in a higher K-Shell has more edges coming into or out of the node, so it may be more central in the graph. It may thus have more built-in redundancy for that node's function. Another interpretation is that more pathways through that node may be enabled. Conversely, nodes in the 1-shell have only one edge, so disconnecting that edge would disconnect that node from the graph. This 1-shell connection is reminiscent of the relationship of a *critical controller* or a *critical state variable* on the physical side, as in refs. [46, 47]. This points to its potential utility in assessing and improving the cyber-physical controllability and observability of the system. In our system, that is observed to happen in scenarios 5 and 6 when the control centre HMI, which was in the 1-shell, gets fully disconnected from the rest of the CPS.

TABLE 3 Results of the DeepWalk methods.

No. Cluster	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
1	{B1,B4,R15,R16, R17, R18, SubB2_SW}	{G1,B1,R15,R16,R17, R18,SubB2_SW, SubB_HMI, SubB0_SW}	{B1,B4,L6,R15,R16, R17,R18,R19, SubB2_SW}	{G1,B1,B4,R16, R17,R18}	{G1,B1,B4,R15, R16, R17,R18, SubB2_SW, SubB_HMI}	{G1,B1,B4,R15,R16, R17,R18, SubB2_SW, SubB_HMI}
2	{L6,B6,R12,R13, R14, SubB0_SW, SubB1_SW, SubB_HMI}	{L6,B4,B6,R12,R14, SubB1_SW}	{L5,B5,B6,R12,R14, R21, SubA2_SW, SubB1_SW}	{B5,L5,R19,R20,R21, SubA_HMI, SubA2_SW}	{L6,B6,R12,R13, R14,r1,r2, r3, SubB0_SW, SubB1_SW, SubB_FW, SubC_FW}	{L6,B6,R12,R13,R14, r1,r2,r3, SubB0_SW, SubB1_SW, SubA_FW, SubB_FW, SubC_FW }
3	{G3, B2,B7,R1,R2,R3, R4,SubA1_SW}	{G3,B2,R1,R2,R3,R4, SubA1_SW}	{G2,G3,B2,B3,B7,B8, B9, L8,R1,R2,R3, R4,R5,R6,R7, R8, R9,R10,R11, SubC1_SW, SubC2_SW}	{G3,B2,B7,R1,R2, R3, R4,SubA1_SW}	{G3,B2,B7,R1,R2, R3,R4, SubA1_SW}	{G3,B2,B7,R1,R2,R3, R4, SubA1_SW}
4	{G2,L8,B3,B8,B9,R5, R6, R7,R8,R9,R10, R11, SubC0_SW, SubC1_SW, SubC2_SW, SubC_HMI}	{G2,B3,B9,R8,R9,R10, R11,SubC2_SW}	{SubC0_SW, SubC1_SW, SubC_HMI}	{G2,L6,B3,B6,B9,R8, R9,R10,R11, SubC0_SW, SubC2_SW, SubC_HMI}	{G2,B3,B9,R8,R9, R10,R11, SubC0_SW, SubC2_SW, SubC_HMI}	{G2,B3,B9,R8,R9, R10,R11, SubC0_SW, SubC2_SW, SubC_HMI}
5	{B5,L5,R19,R20,R21, SubA0_SW, SubA2_SW, SubA_HMI}	{B7,B8,L8,R5,R6,R7, SubC0_SW, SubC1_SW, SubC_HMI}	{SubB0_SW, SubA_HMI, SubB_HMI}	{L8,B8,R5,R6,R7, SubC1_SW }	{B8,L8,R5,R6,R7}	{B8,L8,R5,R6,R7}
6	{r0,r1,r2,r3,SubA_FW, SubB_FW, SubC_FW, CC_HMI,CC_SW}	{L5,B5,R19,R21,r0,r1, r2,r3, SubA_FW, SubB_FW, SubC_FW, SubA0_SW, SubA2_SW, SubA_HMI, CC_HMI,CC_SW}	{r0,r1,r2,r3,SubA_FW, SubB_FW, SubC_FW, SubA0_SW, CC_HMI, CC_SW}	{R13,R14,R15,r0,r1,r2, r3, SubA_FW, SubB_FW, SubC_FW, SubA0_SW, SubB0_SW, SubB1_SW, SubB2_SW, SubB_HMI, CC_HMI,CC_SW}	{L5,B5,R19,R20,r0, SubA_FW, SubA0_SW, SubA2_SW, SubB1_SW, SubA_HMI}	{L5,B5,R19,R20,R21, r0, SubA0_SW, SubA2_SW, SubA_HMI }

- **Bipartite Graphs**. Bipartite network graphs help to identify patterns arising due to the cyber-physical interface connections. A modularity analysis can aid in identifying interaction patterns and level of clustering (termed modules) to enable tying those to network functioning. The number of original modules identified in the fully functioning CPS network here are seen to decrease as a result of disturbance scenarios 2 and 3, which are the only ones that directly impact components at the CP interface. The utility of this method for screening whether disturbances have (or do not have) a cyber-physical (cross-layer) property can be seen directly here. Additional simulations are needed to use the approach to understand disturbance propagation across in the interface. The highly modular findings for the interface of the WSCC-9 bus case study suggests that an interesting future optimisation problem could be to design the modularity of the interface closer to what is seen in mutualistic biological ecosystems, making a less modular and more hierarchical interface to increase interface robustness. The results here prompt the further exploration of modularity and interface design, including its quantification in other larger and realistic power systems configurations. We do also see some overlap with the clusters found by the DeepWalk method, for example, module 1, made up of B7, R2, R3, and R4, overlaps is contained within cluster three of the DeepWalk. Similarly module 9, made up of B3 and R7, is contained within cluster 4 and modules 1 and 3 are both contained within cluster 1. There is less agreement however between the modules and the k-shells, for example, modules 3 and 9 do not line up with the k-shells but module 11 does. That there is some agreement between the clustering techniques is promising to the combined use of multiple techniques to understand different facets of these highly complex cyber-physical networks.

- **DeepWalk-based Technique**. In the DeepWalk-based technique, a random walk step serves as a stochastic simulation block. This step generates access paths for the DeepWalk-based clustering technique. By simulating the stochastic access paths from each node in the graph, the method examines common path traversals through the graph. In this instance, the random walks are unweighted since there is no additional information in the graph concerning edge weights and probability of each edge being used. Through Table 3, we can observe that there are high interdependencies between the components. Attacks to anyone in the cluster will potentially affect other nodes in the same cluster. We could conclude that during a cyber event when certain devices aer already being controlled, the likelihood of nodes residing in the same cluster being affected will increase. Since risk is defined by likelihood multiplied by impact [48], the risk level of the nodes in the same cluster will rise when either node is affected by the adversaries. This can help stakeholders to better identify as well as predict attacks, to prevent cascading failures and take corrective actions.
  Further, the results can tell us how those risk levels change or remain constant during specific disturbance scenarios. Clustering results could also serve a role in predictive analysis to help with anticipating incidents and making decisions. It suggests that identifying and applying other control elements that are in the same cluster as the compromised element(s) for use in risk mitigation and response may have positive impacts on the system's operational reliability. It suggests some next steps such as high-fidelity cyber-physical testbed emulation to further explore and validate this hypothesis. For example, the DeepWalk-based technique results suggest that the following actions could be performed by operators after the DOS attacks occur (which isolate the affected systems). Since later stages of a multi-stage attack, such as physical impact, typically require a successful intrusion and earlier steps first, if operators can anticipate the consequences and the possible next-step targets during an early stage, they should be able to take appropriate actions more quickly and maintain better control of the system. In summary, the clustering results can help identify the potential available and successful proactive response actions.

From each of these methods, different parts of the structure of a CPS graph can be isolated and identified. Many of the differences between these techniques arise from the specific information that each technique uses concerning the graph structure. Table 4 provides a summary of the different input requirements and output types of each technique. Each method also has certain limitations. First, in examining graph components and condensation graphs, only the paths to and from nodes matter. Hence, we learn nodes' connections to other nodes. If there is a distinct separation between portions of the graph, this can be used to identify important edges, and we can better visualise the structure using a condensation graph. However, if the graph has many edges and everything is connected, then looking at graph components does not help decompose the structure any further. For K-Shell decomposition, it is based solely on node degree, so it is useful for ranking nodes based on the number of connections of the node, but that is all. For bipartite graphs, disturbances that affect only the cyber side or only the physical side are not captured.

Future work will look more at impact propagation, going beyond the snapshots considered here. For this, we will need to look at the time domain disturbance data as well as the relationship between some of the key characteristics identified in this paper: node criticality, CPS clusters, and the differences in those clusters, especially as the scenarios and systems change over time. Then, we can observe more case-specific and scenario-specific information that is contained in each of the

TABLE 4 Comparison of input requirements and output types for each graph clustering technique.

	Input	Output
Graph analysis	Combined CPS graph	Decomposed full CPS clusters
		Number of clusters
		Graph hierarchy
		Important connections
K-shell decomposition	Combined CPS graph	Decomposed full CPS clusters
	K (number of edges)	Node centrality
Bipartite graphs	Partial CPS graph (only CPS interfaces)	Decomposed interface clustersc
		Number of clusters
		Topological interface dependencies
		Cluster coherence - modularity
Deep walk-based technique	Combined CPS graph	Decomposed full CPS clusters
	K (number of clusters)	Likelihood Estimation for node vulnerability
		Access paths

method's results. Hence, there is value in expanding to larger cases and to other disturbance scenarios where we can gain further insight into the interdependency analysis methods as well as validate our current hypotheses on the meaning of the different clusters, shells, components, and interfaces.

6 | CONCLUSIONS AND FUTURE WORK

In this paper, we examined various decomposition and clustering techniques to study the structure and interconnections in CPS graphs. This is useful for a variety of cyber-physical analysis applications, such as:

1. Studying interdependencies to identify potential cascading failures or attack paths
2. Ranking graph nodes and edges as critical or non-critical
3. Better visualising graph structure and interconnections
4. Identifying important interfaces between the cyber and physical portions of the graph
5. Examining how the graph structure changes during failure scenarios.

Several different techniques were studied in a representative cyber-physical graph with using the WSCC 9-bus benchmark power system model and an associated communication network. These scenarios consist of different combinations of generator and line outages alongside DoS attacks. The results of applying each technique to these scenarios helped to identify several application areas where the techniques studied in this paper assisted in identifying important structural elements of the cyber-physical graph. Each of the techniques were compared and were found to highlight different structural characteristics of the cyber-physical systems and provide insight into different node and edge criticality.

The results of this paper are foundational in developing a comprehensive approach for interpreting and studying cyber-physical system interdependencies for a multitude of applications and scenarios. By comparing and assessing the results of multiple clustering techniques for cyber-physical graphs, we can understand the structural relationships in cyber-physical systems and how these relationships change or remain the same for different disturbances. While this approach currently only examines the structural properties and connections between components in the CPS, it does provide a basis for assessing cyber-physical system interdependencies for which the natural next step is to assess the dynamic, time-series cyber-physical relationships. With both structural and temporal understanding of cyber-physical interactions, comprehensive CPSA can be achieved that helps inform decision-makers for planning and response strategies.

In future work, we will add to the structural CPS clustering analysis toolset with time- and data-dependent analysis. This includes diving deeper into the relationship between two cyber-physical nodes and assessing their data exchange to understand what variations occur during different disturbances. For example, if a relay cyber node is communicating open/close commands to a generator bus physical node, this will involve a stream of communications that transmit the control command and also periodically check generator bus status. We will also update the cyber-physical graph model to include weighted edges and nodes that differentiate component and edge criticality, as well as closely examine how different types of attacks and behaviours may appear as heterogenous interactions in the system. Since the current work contained here is not modelling dynamical behaviour but rather exposing structural properties and important connections in the graph, the nature of the links is abstracted and will be investigated further in future work. The overall structural and data-dependent interdependency analysis approach will provide a comprehensive toolset to assess cyber-physical systems for a variety of disturbances.

AUTHOR CONTRIBUTIONS

Nicholas Jacobs: Conceptualisation; Formal analysis; Investigation; Methodology; Project administration; Visualisation; Writing – original draft; Writing – review & editing. **Shamina Hossain-McKenzie**: Conceptualisation; Funding acquisition; Methodology; Project administration; Writing – original draft; Writing – review & editing. **Shining Sun**: Formal analysis; Investigation; Methodology; Visualisation; Writing – original draft. **Emily Payne**: Formal analysis; Investigation; Methodology; Visualisation; Writing – original draft. **Adam Summers**: Formal analysis; Investigation; Methodology; Visualisation; Writing – original draft. **Leen Al-Homoud**: Investigation; Writing – original draft. **Astrid Layton**: Conceptualisation; Methodology; Supervision; Writing – original draft; Writing – review & editing. **Kate Davis**: Conceptualisation; Investigation; Supervision; Writing – original draft; Writing – review & editing. **Chris Goes**: Data curation; Software.

ACKNOWLEDGEMENTS

We would like to thank the entire InterGraph-CPS LDRD team, including Michael Livesay, Erin DeCarlo, Rachid Darbali-Zamora, Jack Flicker, and Daniel Bauer. Additionally, we'd like to thank Steve Scott and his invaluable help in improving the paper flow and writing. The team at Texas A&M University would like to thank lab-mate Luis Rodriguez for his assistance in running the ecological modularity analysis. This material is based upon work supported by the Sandia Laboratory Directed Research and Development Project # 229324.

CONFLICT OF INTEREST STATEMENT

This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Shamina Hossain-McKenzie https://orcid.org/0000-0002-3085-8193

REFERENCES

1. Jain, A.K.: Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. 31(8), 651–666 (2010). https://doi.org/10.1016/j.patrec.2009.09.011
2. Ibrahim, M.S., Dong, W., Yang, Q.: Machine learning driven smart electric power systems: current trends and new perspectives. Appl. Energy 272, 115237 (2020). https://doi.org/10.1016/j.apenergy.2020.115237
3. Boyaci, O., et al.: Spatio-temporal failure propagation in cyber-physical power systems (2022)
4. Hossain-McKenzie, S., et al.: Towards the characterization of cyber-physical system interdependencies in the electric grid (2023)
5. Davis, K.R., et al.: A cyber-physical modeling and assessment framework for power grid infrastructures. IEEE Trans. Smart Grid 6(5), 2464–2475 (2015). https://doi.org/10.1109/TSG.2015.2424155
6. Vellaithurai, C., et al.: CPIndex: cyber-physical vulnerability assessment for power-grid infrastructures. IEEE Trans. Smart Grid 6(2), 566–575 (2014). https://doi.org/10.1109/tsg.2014.2372315
7. Clark, A., Zonouz, S.: Cyber-physical resilience: definition and assessment metric. IEEE Trans. Smart Grid 10(2), 1671–1684 (2017). https://doi.org/10.1109/tsg.2017.2776279
8. Venkataramanan, V., et al.: Cp-tram: cyber-physical transmission resiliency assessment metric. IEEE Trans. Smart Grid 11(6), 5114–5123 (2020). https://doi.org/10.1109/tsg.2020.2996137
9. Baranwal, M., Salapaka, S.: Clustering and supervisory voltage control in power systems. Int. J. Electr. Power Energy Syst. 109, 641–651 (2019). https://doi.org/10.1016/j.ijepes.2019.02.025
10. Rocchetta, R.: Enhancing the resilience of critical infrastructures: statistical analysis of power grid spectral clustering and post-contingency vulnerability metrics. Renew. Sustain. Energy Rev. 159, 112185 (2022). https://doi.org/10.1016/j.rser.2022.112185
11. Schütz, T., et al.: Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis. Renew. Energy 129, 570–582 (2018). https://doi.org/10.1016/j.renene.2018.06.028
12. Wu, J., et al.: Spectral graph clustering for intentional islanding operations in resilient hybrid energy systems. IEEE Trans. Ind. Inf. 19(4), 5956–5964 (2022). https://doi.org/10.1109/tii.2022.3199240
13. Hogan, E., et al.: Towards effective clustering techniques for the analysis of electric power grids (2013)
14. Hu, J., Sankar, L., Mir, D.J.: Cluster-and-Connect: an algorithmic approach to generating synthetic electric power network graphs. IEEE, 223–230 (2015)
15. Blumsack, S., et al.: Defining power network zones from measures of electrical distance. IEEE, 1–8 (2009)
16. Sánchez-García, R.J., et al.: Hierarchical spectral clustering of power grids. IEEE Trans. Power Syst. 29(5), 2229–2237 (2014). https://doi.org/10.1109/tpwrs.2014.2306756
17. Hines, P., et al.: The topological and electrical structure of power grids. IEEE, 1–10 (2010)
18. Al-Hinai, A.S.: Voltage Collapse Prediction for Interconnected Power Systems. West Virginia University (2000)
19. Hossain-McKenzie, S., et al.: Adaptive, cyber-physical special protection schemes to defend the electric grid against predictable and unpredictable disturbances. IEEE, 1–9 (2021)
20. U.S. National Science Foundation: Cyber-physical Systems: Enabling a Smart and Connected World (2023)
21. Jacobs, N., et al.: Cyber-physical observability for the electric grid (2020)
22. Hossain-McKenzie, S., et al.: Harmonized Automatic Relay Mitigation of Nefarious Intentional Events (HARMONIE) - Special Protection Scheme. SPS (2022). https://doi.org/10.2172/1890265
23. Šandi, S., Krstajić, B., Popović, T.: pyPMU — open source python package for synchrophasor data transfer (2016)
24. ESnet / Lawrence Berkeley National Laboratory. Iperf3: A TCP, UDP, and SCTP Network Bandwidth Measurement Tool. (2023)
25. Bertsekas, D., Gallager, R.: Data Networks, 2nd ed. Prentice-Hall, Inc., USA (1992)
26. Yang, Z., Algesheimer, R., Tessone, C.J.: A comparative analysis of community detection algorithms on artificial networks. Sci. Rep. 6(1), 30750 (2016). https://doi.org/10.1038/srep30750
27. Wills, P., Meyer, F.G.: Metrics for graph comparison: a practitioner's guide. PLoS One 15(2), e0228728 (2020). https://doi.org/10.1371/journal.pone.0228728
28. Deo, N.: Graph Theory with Applications to Engineering and Computer Science. Dover Books on MathematicsDover Publications (2016)
29. Carmi, S., et al.: A model of Internet topology using k-shell decomposition. Proc. Natl. Acad. Sci. 104(27), 11150–11154 (2007). https://doi.org/10.1073/pnas.070117510430
30. Garas, A., Schweitzer, F., Havlin, S.: A k-shell decomposition method for weighted networks. New J. Phys. 14(8), 083030 (2012). https://doi.org/10.1088/1367-2630/14/8/083030
31. Yao, J., et al.: Node importance evaluation method for cyberspace security risk control (2021)
32. Bascompte, J., Melián, C.J.: Simple trophic modules for complex food webs. Ecology 86(11), 2868–2873 (2005). https://doi.org/10.1890/05-0101
33. Guimera, R., Sales-Pardo, M., Nunes Amaral, L.A.: Module identification in bipartite and directed networks. Phys. Rev. 76(3), 036102 (2007). https://doi.org/10.1103/physreve.76.036102
34. Olesen, J.M., et al.: The modularity of pollination networks. Proc. Natl. Acad. Sci. USA 104(50), 19891–19896 (2007). https://doi.org/10.1073/pnas.0706375104
35. Kadoya, T., Gellner, G., McCann Kevin, S.: Potential oscillators and keystone modules in food webs. Ecol. Lett. 0(0), 1330–1340 (2018). https://doi.org/10.1111/ele.13099
36. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. USA 103(23), 8577–8582 (2006). https://doi.org/10.1073/pnas.0601602103
37. Leicht, E.A., Newman, M.E.J.: Community structure in directed networks. Phys. Rev. Lett. 100(11), 118703 (2008). https://doi.org/10.1103/PhysRevLett.100.118703
38. Zhuo, Z.: Community Detection by Maximizing Modularity - Python Implementaion of Newman Spectral Method. python library (2018). https://github.com/zhiyzuo/python-modularity-maximization
39. Hossain-McKenzie, S., et al.: Towards the characterization of cyber-physical system interdependencies in the electric grid (2023)
40. Ulanowicz, R.E.: Growth and Development: Ecosystems Phenomenology. Springer-Verlag (1986)
41. Perozzi, B., Al-Rfou, R., Skiena, S.: DeepWalk. ACM (2014)
42. Benson, A.R., Gleich, D.F., Lim, L.: The spacey random walk: a stochastic process for higher-order data. CoRR, 02102 (2016). abs/1602
43. Wisconsin Madison oU, C.S.: ADVANCED ALGORITHMS,Lecture15: Random Walks and Markov Chains. 787
44. Konstantopoulos, T.: Markov chains and random walks. Lecture notes (2009)
45. Multiyear Plan for Energy Sector Cybersecurity. Tech. Rep., Department of Energy, Office of Electricity Delivery and Energy Reliability; (2018)
46. Bobba, R.B., et al.: Detecting False Data Injection Attacks on Dc State Estimation. Preprints of the First Workshop on Secure Control Systems. CPSWEEK (2010)
47. Hossain-McKenzie, S., et al.: Analytic corrective control selection for online remedial action scheme design in a cyber adversarial environment. IET Cyber-Physical Systems: Theory Applications 2(4), 188–197 (2017). https://doi.org/10.1049/iet-cps.2017.0014
48. Stoneburner, A.G., Feringa, A.: Risk Management Guide for Information Technology Systems. National Institute of Standards and Technology, Technology Administration (2002)

How to cite this article: Jacobs, N., et al.: Leveraging graph clustering techniques for cyber-physical system analysis to enhance disturbance characterisation. IET Cyber-Phys. Syst., Theory Appl. 1–15 (2024). https://doi.org/10.1049/cps2.12087