

# Hetero-functional Graph Resilience of the Future American Electric Grid

Thompson, D. J. Schoonenberg, W. C. H. Farid, A. M.

Thayer School of Engineering at Dartmouth college, Hanover, NH, USA

As climate change takes hold in the 21<sup>st</sup> century, it places an impetus to decarbonize the American electric power system with renewable energy resources. There is a broad technical consensus<sup>1-3</sup> that these renewable energy resources can not be integrated alone but rather require a whole host of profound changes in the electric grid's architecture; including meshed distribution lines, and energy storage solutions. One question that arises is whether these three types of mitigation measures required by decarbonization will also serve as adaptation measures when the climate changes and extreme weather phenomena become more prevalent. Consequently, this paper presents a structural resilience analysis of the American electric power system that incrementally incorporates these architectural changes in the future. Building upon a preliminary study<sup>4</sup>, the analysis draws on an emerging *hetero-functional graph theory*<sup>5</sup> based upon the inter-connectedness of a system's *capabilities*. The hetero-functional graph analysis confirms our formal graph understandings from network science<sup>6-8</sup> in terms of cumulative degree distributions and traditional attack vulnerability measures. The paper goes on to show that hetero-functional graphs relative to formal graphs more precisely describe the changes in functionality associated with the addition of distributed generation and energy storage as the grid evolves to a decarbonized architecture. Finally, it demonstrates that the addition of all three types of mitigation measures enhance the grid's structural resilience; even in the presence of disruptive random and targeted attacks. The paper concludes that there is no structural trade-off between grid sustainability and resilience enhancements and that these strategic goals can be pursued simultaneously.

While renewable energy can be integrated centrally at a utility-scale, one of its primary benefits is that it can empower end-consumers with distributed generation (DG) in the form of rooftop solar photovoltaics (PV), small-scale wind turbines, and even run-of-river hydro power. The potential for power back-flow in an electric distribution system designed for one-way outward flow requires a migration from a radial to a meshed topology. In the meantime, these DG resources are intermittent and often require complementing energy storage solutions. These three additions represent fundamental changes to the system architecture of the American Electric Power System (AEPS).

Such architectural changes can have a profound impact on the system's *resilience* in terms of its own ability to withstand disruptions; be they natural, artificial, or intentionally nefarious<sup>6-8</sup>. To address such questions quantitatively, the network science community has used graphs to mathematically represent the *form* of a system<sup>9</sup>. The graph's nodes are made to correspond to the grid's power plants, substations, and consumers while the graph's edges are made to correspond to the grid's power lines. For clarity, we refer to such a mathematical model as a *formal graph* (FG)<sup>9</sup>. Consequently, system resilience can be quantitatively studied in terms of successive node or edge failures. While such a simple graph model can address the resilience improvements caused by a migration towards meshed distribution networks, and the addition of *new nodes* that represent solar PV and energy storage solutions, it is ill-equipped to address the integration of such distributed energy resources on existing nodes as in the case of solar panels on rooftops and batteries at homes, substations, and centralized generators. In effect, such additions (as shown later) do not numerically change the formal graph, and consequently, have no effect on the value of the associated resilience measure.

In contrast, the model-based systems engineering (MBSE) community recognizes that a formal graph representing system form is merely a subset of system architecture and that a more comprehensive description of architecture must also describe: 1.) a set of functions that the system performs and 2.) the allocation of those functions to the elements of form<sup>9,10</sup>. While the MBSE literature normally describes system architecture using graphical models (e.g. UML & SysML)<sup>10</sup>, hetero-functional graph theory has developed to translate these models into their quantitative equivalents<sup>5</sup>. Consequently, hetero-functional graphs are able to explicitly and quantitatively describe the incorporation of new functionality onto existing formal nodes as in the case of rooftop solar and home batteries.

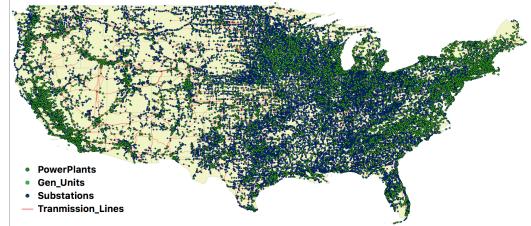


Fig. 1. A GIS Visualization of the American Electric Power System.

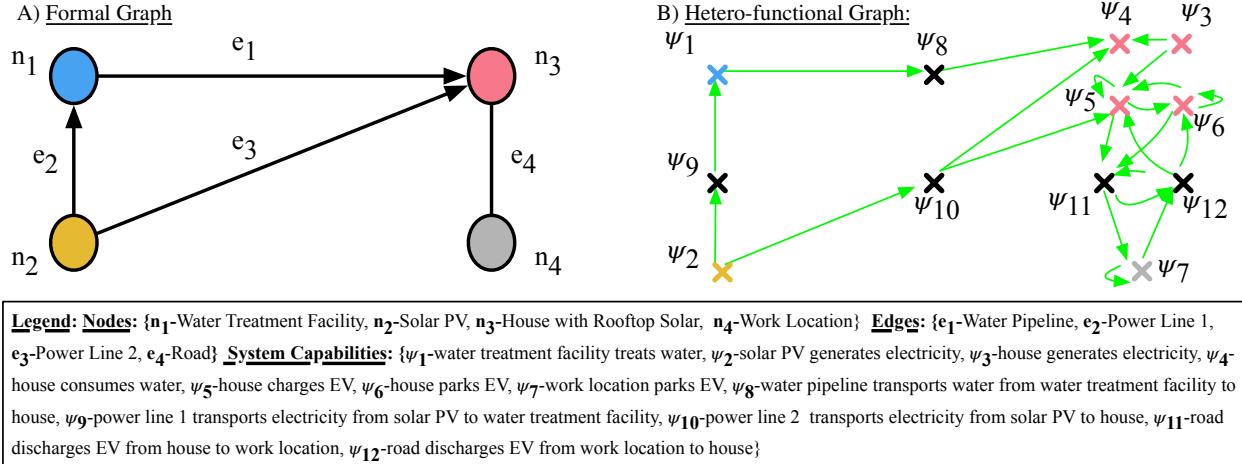


Fig. 2. A Visual Comparison of a Formal Graph (FG) and a Hetero-functional Graph (HFG) Model of the Same Hypothetical System.

The simple example in Fig. 2 illustrates the differences between a formal graph and a hetero-functional graph (HFG). The formal graph (in Fig 2a.) shows a system composed of four nodes: a water treatment facility, a solar PV panel, a house with rooftop solar, and a work location. These are connected by four edges: a water pipeline, two power lines and two roads. Fig 2b. shows the associated hetero-functional graph following the methods described in the methods section. Instead of four nodes that represent point-like facilities, the hetero-functional graph now has 12 nodes that represent **system capabilities**. The water treatment facility, solar PV panel, and work location appear unchanged between the two graphs because they each have only one capability. In contrast, the house with rooftop solar provides four capabilities in the HFG. This multiplicity of capabilities assigned to a single facility forms the basis upon which to investigate the effect of DG on electric power system resilience. Thirdly, the edges in the formal graph now appear as transportation capabilities (**nodes**) in the HFG. Finally, the directed edges in the HFG indicate the logical sequences of these capabilities such that if one were to follow them a “story” of capabilities would emerge. (i.e. The water treatment facility treats water ( $\psi_1$ ) and then the water pipeline transports the water from the water treatment facility to the house ( $\psi_8$ )). It is important to recognize that because the FG and HFG have different quantities of nodes and edges, the associated values of graph measures also differ.

Despite these apparent differences, FGs and HFGs, when studied in the context of the AEPS, demonstrate remarkable similarities in their degree distributions. The Platts Map Data Pro data set<sup>11</sup> was used to conduct the analysis. As shown in Fig. 1, it consists of a GIS layer with 13,568 power plants, 34,649 generation units, 78,880 substations, and 104,329 transmission lines. A FG graph adjacency matrix is readily extracted from this GIS data and the methods section describes the construction of the associated HFG. Both the FG and HFG, as shown in Fig. 3, confirm the network science result<sup>8,7</sup> of a cumulative degree distribution with exponential decay law  $P(k \leq K)ce^{-\alpha k}$ . They have exponential coefficients of  $\alpha_{FG} = 0.44$  and  $\alpha_{HFG} = 0.27$  respectively. This result suggests that the underlying single-scale small-world socio-technical dynamic of preferential attachment of transmission lines appears in both network models. The larger exponential coefficient in the HFG arises because it always has more nodes and edges than its FG counterpart. Nevertheless, the presence of small-world structure in both graphs indicates similar behavior with respect to resilience and attack vulnerability.

In order to verify this hypothesis, the FG and HFG models of the AEPS were subjected to nodal attacks and then assessed with respect to their average degree centrality, size of largest cluster, and average size of isolated cluster. (See Supp. Materials for details). Fig. 4a shows the random attack vulnerability of the FG and HFG with respect to in-degree and out-degree centrality measures. The parity of in-degree and out-degree centrality in the case of the FG is caused by its undirected nature, while in the case of the HFG, it is caused by the two transportation capabilities assigned to each power line combined with the single capability assigned to power plants and substations. Both the FG and HFG show that the average degree centrality degrades linearly with the fraction of randomly removed

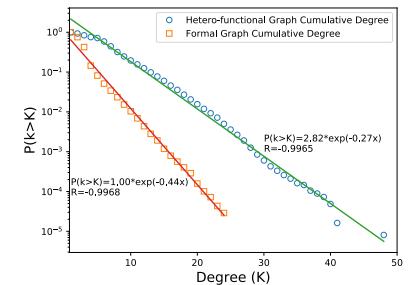


Fig. 3. The formal and hetero-functional graphs of the AEPS have an exponential-tail cumulative degree distribution indicating a single-scale small-world network.

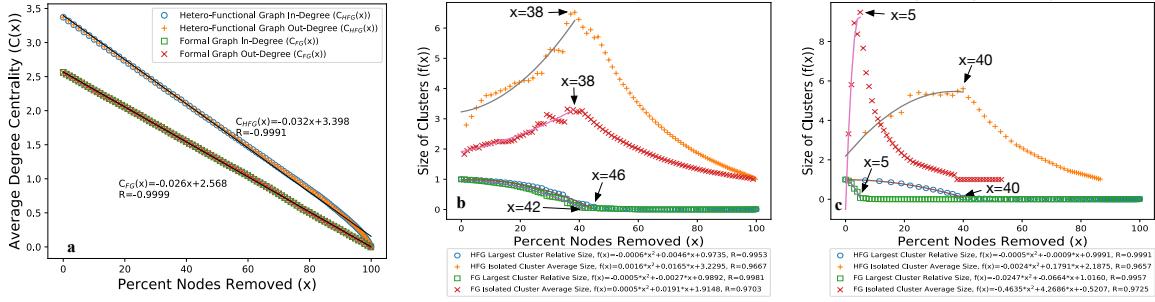


Fig. 4. The FG and the HFG of the AEPS demonstrate similar behavior in: **a** their average degree centrality when subjected to successive random attacks, **b** the size of their largest cluster and average size of their isolated clusters when subjected to random attacks, and **c** the size of their largest cluster and average size of their isolated clusters when subjected to targeted attacks.

nodes. The slopes of their best-fit lines are -0.026 and -0.032 respectively, and their regression coefficients are -0.9999 and -0.9991 respectively. This intuitive phenomena occurs because random attacks do not discriminate on the basis of the nodal-degree. Consequently, the successive removal of two sets of nodes do not change the average number of edges lost. Fig. 4b confirms the network science result<sup>6</sup> that the size of the largest cluster of the FG of an electric power system degrades quadratically under random attacks before reaching an inflection point (at  $x=42\%$  in this case). Fig. 4b also confirms that the average size of isolated clusters first grows to a relative peak (at  $x=38\%$  in this case) before sharply falling again. These two behaviors also appear in the associated HFG with a largest cluster inflection point of  $x=46\%$  and a peak average size of isolated cluster at  $x=38\%$ . Fig. 4c also shows that the FG and HFG confirm the results of the published literature<sup>6</sup> regarding the structural response to targeted attacks. This time, the size of the largest cluster reaches an inflection point at 5% and 40% for the FG and HFG respectively. Also, the peak average size of the isolated cluster occurs at 5% and 40% respectively. In short, all four of these structural responses to nodal attacks in the FG are matched closely in the associated HFG. The differences in the labelled critical points stem from the larger number of nodes and edges in the HFG. Collectively, the results in Fig. 4 show that the FG and HFG can be used interchangeably to study traditional attack vulnerability measures.

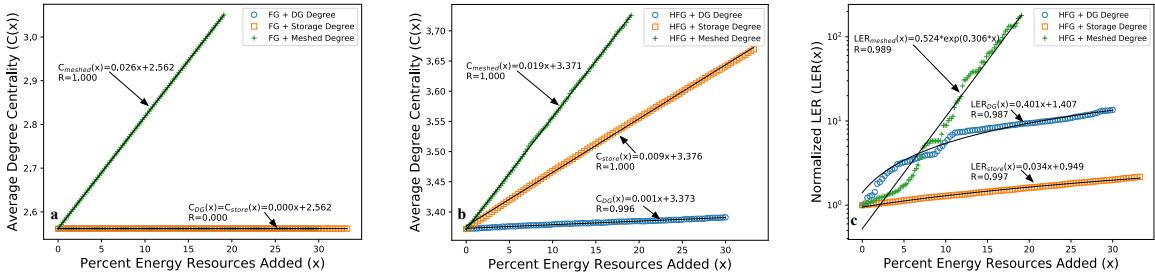


Fig. 5. The FG **a** and the HFG **b** of the AEPS demonstrate differing behavior in: their average degree centrality when subjected to successively adding architectural and functional improvements. The FG **a** is only able to capture the improvement of adding meshed power lines while the HFG **b** measure improvements from the additions of distributed generation and storage as well. The Latent Engineering Resilience of the HFG measures the potential improvements from adding all three forms of additions to the AEPS.

Such a conclusion serves as the foundation upon which to investigate the AEPS as it migrates towards a decarbonized system architecture. More specifically, and as detailed in the methods section, the AEPS was subjected to incremental additions of distributed generation, energy storage resources, and meshed power distribution lines. While the last of these is clearly a change in the AEPS formal topology, the first two represent fundamental changes in the AEPS' **functionality** without a commensurate change in formal topology. Indeed, from the perspective of an electric power utility or grid operator, the FG does not change when end-users add rooftop solar and batteries to their buildings. Fig. 5a shows the effect of adding these resources on the average degree centrality of the FG. As expected, the addition of new meshed power distribution lines increases the average degree centrality linearly with a slope of 0.026 and a regression coefficient of 1. In contrast, the addition of distributed generation and energy storage resources has **no effect** on the average degree centrality of the FG because the underlying adjacency matrix remains **entirely unchanged**. Such a result calls into question either the adequacy of the FG as a **model** or the adequacy of degree centrality as a resilience **measure**. After all, an end-user with newly installed distributed generation or energy storage would continue to have some form of electric power service even if they were entirely disconnected from the rest of the grid.

In contrast, Fig. 5b shows that the degree centrality of the HFG responds as the AEPS migrates towards a decarbonized system architecture. Again, as expected, the addition of new meshed power distribution lines increases the average degree centrality linearly with a slope of 0.019 and a regression coefficient of one. Furthermore, the introduction of energy storage and distributed generation resources now increases the average degree centrality linearly with a slope of 0.009 and 0.001 respectively and a regression coefficient of one and 0.996 respectively. Although, the topology of the underlying formal graph experiences no change during the addition of these two types of resources, the HFG incorporates both new capability nodes as well as connecting edges. More specifically, each new distributed generation resource adds a new capability node and a new edge. In the meantime, each storage resource adds a new capability node but adds at least four new edges. These results show that the HFG is more adequate than the FG as a model when system architecture is changing its functionality and not just its formal topology.

The results of Fig. 5b, however, understate the “resilience value” of distributed generation. Again, from a practical perspective, an end-user would not differentiate between electricity supplied from the grid or that supplied from distributed generation. In contrast, a degree centrality measure only values the resilience that comes from greater grid connectivity. As an alternative, Fig. 5c calculates “latent engineering resilience” (LER)<sup>12</sup> measure as the AEPS responds to the same architectural changes found in Fig. 5b. As explained in the methods section, the LER measures was specifically developed to calculate the number of viable service paths in hetero-functional graphs. As expected, Fig. 5c shows that the LER grows exponentially (with a coefficient of  $\alpha = 0.306$  and regression coefficient of 0.989) as meshed distribution lines are added. Such lines will exponentially increase the number of available service paths. In the meantime, the LER grows linearly with addition of DG with a slope of 0.401 and a regression coefficient of 0.987. Each new DG resource makes use of the network topology to introduce a relatively large but proportional number of service paths. Finally, the LER grows linearly with a slope of 0.034 and a regression coefficient of 0.997 as energy storage resources are added. Unlike DG, energy storage resources still require generation resources in order to contribute a service path and so their resilience enhancing effect is contingent upon generation and distribution capabilities. The results in Fig. 5 show that the HFG relative to a FG more precisely describes the future evolution of the AEPS’ architecture. Furthermore, the quantification of these resilience improvements is more accurately measured using a LER measure based upon service paths than simply a network centrality measure. Finally, and most importantly, the addition of DG and energy storage resources in combination with meshed distribution lines enhance the AEPS’ transition to a highly resilient and decarbonized system architecture. In other words, from an architectural perspective, there is no trade-off between grid sustainability and resilience enhancements and that these strategic goals can be pursued simultaneously.

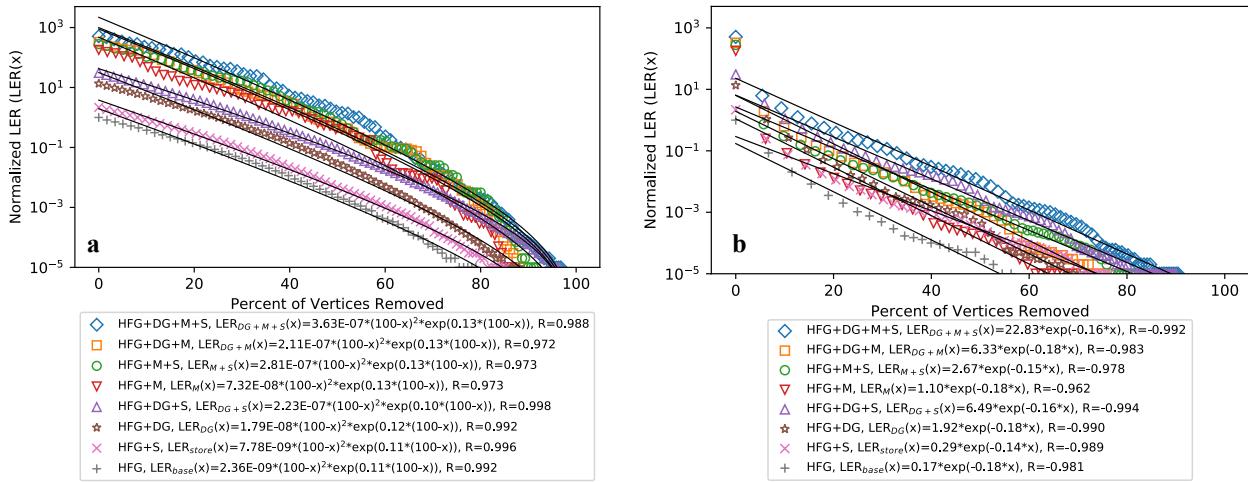


Fig. 6. Eight increasingly sustainable grid architectures are studied in terms of their resilience to random and targeted attacks. These include all combinations of 33% of generators with additional storage (S), 30% of substations with additional distributed generation (DG), and 20% additional meshed distribution lines (M). **a** shows the response to random attacks. **b** shows the response to targeted attacks based upon highest degree centrality.

This conclusion is further investigated in Fig. 6. More specifically, the LER of eight increasingly decarbonized architectures of the AEPS is studied in presence of random and targeted attacks. As detailed in the methods section, the eight investigated architectures include all combinations of: 30% of substations with additional distributed

generation (DG), 33% of generators with additional storage (S), and 20% additional meshed distribution lines (M). Irrespective of the choice of decarbonized architecture, these systems respond differently to random and targeted attacks. In the case of random attacks (Fig. 6a), the  $LER \propto (100 - x)^2 * e^{-\alpha(100-x)}$  while in the case of targeted attacks (Fig. 6b), the LER more closely follows  $e^{(-\alpha x)}$ . Targeted attacks on the basis of highest degree centrality remove an exponentially decreasing number of edges as can be inferred from Fig. 3. Each of these edges, in turn, contribute to an exponential number of paths; resulting in an overall exponential effect. In contrast, the random attacks combine the exponential loss of paths with the parabolically decreasing size of the largest cluster (shown in Fig. 4b). Consequently, and as expected, successive targeted attacks more effectively diminish the grid's LER than random attacks do. In both cases, the LER measure is able to precisely differentiate between all three types of architectural changes. In agreement with the results from Fig. 5c, the architectures with meshed distribution lines, as a group, exhibit the greatest resilience. This group is followed by the architectures with distributed generation which is in turn followed by the architectures with energy storage. All of these cases report higher resilience values than the baseline system representing the AEPS in its present form. In other words, these results confirm that the evolution of the AEPS' to a decarbonized architecture composed of distributed generation, energy storage, and meshed distribution lines will simultaneously enhance its resilience.

This paper has presented a structural resilience analysis of the American electric power system as it evolves towards a decarbonized architecture consisting of distributed generation, energy storage, and meshed distribution lines. To conduct the analysis, it relied on hetero-functional graphs which were shown to confirm our formal graph understandings from network science<sup>6-8</sup> in terms of cumulative degree distributions and traditional attack vulnerability measures. Such hetero-functional graphs more precisely describe the changes in functionality associated with the addition of distributed generation and energy storage as the grid evolves to a decarbonized architecture. Finally, it demonstrates that the addition of all three types of mitigation measures enhance the grid's structural resilience; even in the presence of disruptive random and targeted attacks. Consequently, there is no structural trade-off between grid sustainability and resilience enhancements and that these strategic goals can be pursued simultaneously.

## REFERENCES

- [1] S. O. Muhanji, A. Muzhikyan, and A. M. Farid, "Distributed Control for Distributed Energy Resources: Long-Term Challenges & Lessons Learned," *IEEE Access*, vol. 6, no. 1, pp. 32 737 – 32 753, 2018. [Online]. Available: <http://dx.doi.org/10.1109/ACCESS.2018.2843720>
- [2] A. M. Annaswamy, M. Amin, C. L. Demarco, T. Samad, J. Aho, G. Arnold, A. Buckspan, A. Cadena, D. Callaway, E. Camacho, M. Caramanis, A. Chakrabortty, A. Chakraborty, J. Chow, M. Dahleh, A. D. Dominguez-Garcia, D. Dotta, A. M. Farid, P. Flikkema, D. Gayme, S. Genc, M. G. i. Fisa, I. Hiskens, P. Houpt, G. Hug, P. Khargonekar, H. Khurana, A. Kiani, S. Low, J. McDonald, E. Mojica-Navia, A. L. Motto, L. Pao, A. Parisio, A. Pinder, M. Polis, M. Roozbehani, Z. Qu, N. Quijano, and J. Stoustrup, *IEEE Vision for Smart Grid Controls: 2030 and Beyond*. New York NY: IEEE Standards Association, 2013. [Online]. Available: <http://www.techstreet.com/ieee/products/1859784>
- [3] J. Stoustrup, A. M. Annaswamy, A. Chakrabortty, and Z. Qu, *Long-Term Challenges for Future Electricity Markets with Distributed Energy Resources*. Berlin, Heidelberg: Springer, 2017. [Online]. Available: <http://dx.doi.org/10.1007/978-3-319-98310-3>
- [4] D. Thompson, W. C. Schoonenberg, and A. M. Farid, "A Hetero-functional Graph Analysis of Electric Power System Structural Resilience," in *IEEE Innovative Smart Grid Technologies Conference North America*, Washington, DC, United states, 2020.
- [5] W. C. Schoonenberg, I. S. Khayal, and A. M. Farid, *A Hetero-functional Graph Theory for Modeling Interdependent Smart City Infrastructure*. Berlin, Heidelberg: Springer, 2018. [Online]. Available: <http://dx.doi.org/10.1007/978-3-319-99301-0>
- [6] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, Jul 2000.
- [7] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the north american power grid," *Physical review E*, vol. 69, no. 2, p. 025103, 2004.
- [8] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the national academy of sciences*, vol. 97, no. 21, pp. 11 149–11 152, 2000.
- [9] E. Crawley, B. Cameron, and D. Selva, *System architecture: strategy and product development for complex systems*. Prentice Hall Press, 2015.
- [10] S. Friedenthal, A. Moore, and R. Steiner, *A Practical Guide to SysML: The Systems Modeling Language*, 2nd ed. Burlington, MA: Morgan Kaufmann, 2011.
- [11] Platts, "Platts energy map data pro," S&P Global Platts, Tech. Rep., 2017. [Online]. Available: <http://www.platts.com/products/map-data-pro>
- [12] A. M. Farid, "Static Resilience of Large Flexible Engineering Systems: Axiomatic Design Model and Measures," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–12, 2015. [Online]. Available: <http://dx.doi.org/10.1109/JSYST.2015.2428284>

## I. METHODS

### A. Constructing A Hetero-functional Graph

A hetero-functional graph, like any other graph, is constructed by identifying a set of nodes and connecting them with edges. This is a two step process that we describe simply here and illustrate using the example provided in Fig. 2. The interested reader is referred to the hetero-functional graph theory text<sup>5</sup> for a more elaborate exposition.

Unlike formal graphs where the nodes represent elements of form like power plants, substations, and consumers, the nodes in a hetero-functional graph are *system capabilities*. A capability is the feasible allocation of a given *system process* (or function) to a given *system resource* as an element of form. The system processes  $P$  describe the functionality of the system in terms of a transitive verb followed by an operand (e.g. “generate + electric power”). These processes are often classified as (in-place) transformation processes  $P_\mu$  and transportation processes  $P_\eta$ . The system resources  $R$  describe the formal composition of the system in terms of nouns. For example, in Fig. 2, the set of system resources  $R = \{\text{Water Treatment Facility, Solar PV, House with Rooftop Solar, Work Location, Water Pipeline, Power Line 1, Power Line 2, Road}\}$ . (Note that  $R = V \cup E$  where  $V$  and  $E$  are the nodes and edges of a formal graph.) In the meantime, the system processes  $P = \{\text{treat water, generate electricity, consume water, charge EV, store EV, transport water from water treatment facility to house, transport power from solar PV to water treatment facility, transport power from solar PV to house, discharge EV from house to work location, discharge EV from work location to house}\}$ . The feasible allocation of system process to system form is captured in a *system knowledge base*  $J_S$ .

**Definition 1. System knowledge base:** A binary matrix  $J_S$  of size  $\sigma(P) \times \sigma(R)$  whose element  $J_S(w, v) \in \{0, 1\}$  is equal to one when  $e_{wv} \in \mathcal{E}$  (in the SysML sense) exists as a system process  $p_w \in P$  being executed by a resource  $r_v \in R$ . The  $\sigma()$  operator returns the size of a set.

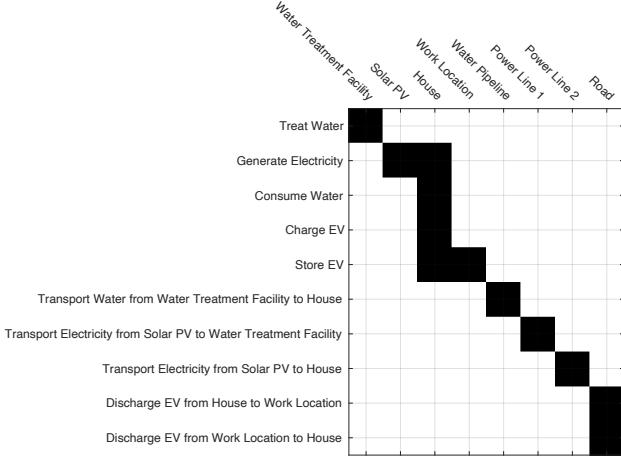


Fig. 7. The hetero functional system Knowledge Base for System Depicted in Fig. 2.

Note that the system knowledge base itself constructs a bipartite graph between  $P$  and  $R$ . The system knowledge base associated with the system depicted in Fig. 2 is shown in Fig. 7 as a monochrome image. The capabilities of a given resources are explicitly reflected in the associated column of  $J_S$ . Consequently, the set of capabilities  $\mathcal{E} = \{1.) \text{ water treatment facility treats water, 2.) solar PV generates electricity, 3.) house consumes water, 4.) house generates electricity, 5.) house charges EV, 6.) house stores EV, 7.) work location stores EV, 8.) water pipeline 1 transports water from water treatment facility to house, 9.) power line 1 transports electricity from solar PV to water treatment facility, 10.) power line 2 transports electricity from solar PV to house, 11.) road “discharges” EV from house to work location, 12.) road “discharge” EV from work location to house\}$ . These capabilities make up the nodes of the hetero-functional graph.

Once the capabilities of the system have been identified as the nodes of a hetero-functional graph, they can be connected with edges so as to form a hetero-functional adjacency matrix  $A_\rho$ .

$$A_\rho = (J_S \ominus K_S)^V (J_S \ominus K_S)^{VT} \ominus K_\rho \quad (1)$$

where  $(\cdot)^V$  is shorthand for  $\text{vec}(\cdot)$ ,  $\ominus$  denotes Boolean subtraction of matrices,  $K_S$  is the *system constraints matrix*, and  $K_\rho$  is the *system sequence constraints matrix*.  $K_S = \mathbf{0}$  when all capabilities are functional.

**Definition 2. System sequence constraints matrix:** a square binary constraints matrix  $K_\rho$  of size  $\sigma(R)\sigma(P) \times \sigma(R)\sigma(P)$  whose elements  $K(\chi_1, \chi_2) \in \{0, 1\}$  are equal to one when string  $z_{\chi_1 \chi_2} = e_{w_1 v_1} e_{w_2 v_2} \in Z$  is eliminated.

$K_\rho$  is calculated by identifying constraints that impede the logical sequence  $z_{\chi_1 \chi_2}$  between an ordered pair of capabilities  $e_{w_1 v_1} e_{w_2 v_2}$ . Five types of constraints are possible on  $e_{w_1 v_1} e_{w_2 v_2}$ :

- 1) When  $w_1$  and  $w_2$  indicate transformation processes but  $v_1 \neq v_2$ .
- 2) When  $w_1$  indicates a transformation process and  $w_2$  indicates a transportation process but  $v_1$  is not the origin of  $w_2$  as a transportation process.
- 3) When  $w_1$  indicates a transportation process and  $w_2$  indicates a transformation process but  $v_2$  is not the destination of  $w_1$  as a transportation process.
- 4) When  $w_1$  and  $w_2$  indicate transportation processes but the destination of the former is not equivalent to the origin of the later.
- 5) When  $w_1$  and  $w_2$  are not permitted by the functional reference architecture of the system. For example, in electric power systems, the generation of electric power systems is followed by any number of transportation processes which is followed by the consumption of electric power.

Finally, it is often useful to use a projection operator  $\mathbb{P}$  to eliminate the empty rows and columns in  $A_\rho$ .

$$\tilde{A}_\rho = \mathbb{P} A_\rho \mathbb{P}^T \quad (2)$$

From these steps, the hetero-functional adjacency matrix  $\tilde{A}_\rho$  corresponding to Fig. 2 is shown in Fig. 8a below. It is contrasted with the associated formal graph in Fig. 8b.

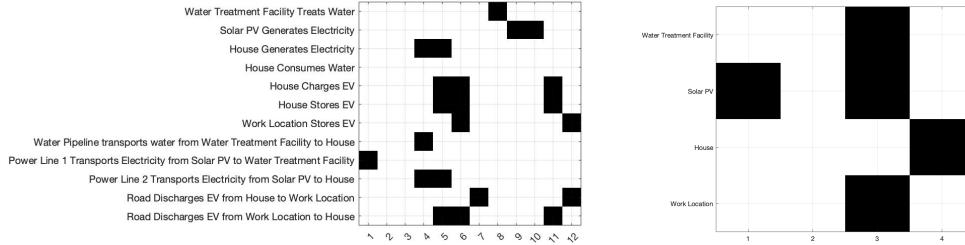


Fig. 8. The (Projected) Hetero-functional Graph Adjacency Matrix (a) of the System Depicted in Fig. 2. The Formal Adjacency Matrix (b) of the System Depicted in Fig. 2.

### B. Constructing a Hetero-functional Graph of the American Electric Power System

The Platts Map Data Pro data was used to create the hetero-functional graph of the American electric power system in three steps:

- 1) The Platts Map Data Pro data was converted into a corresponding KML file.
- 2) The KML file was processed into an XML file compatible with the hetero-functional graph theory toolbox<sup>13</sup>.
- 3) The hetero-functional graph theory toolbox provides the functionality to automatically calculate the hetero-functional adjacency matrix  $\tilde{A}_\rho$ .

The Platts Map Data Pro is a proprietary Geographic Information Systems (GIS) database that contains 38 different GIS data sets of North American energy infrastructure and its associated markets. This paper focuses on the transmission system and thus uses four GIS layers to create the hetero-functional graph: the 1.) Power plants, 2.) Generation units (i.e. individual generation facilities within power plants), 3.) Substations, and 4.) Transmission Lines. Each of these North American layers were cropped to return elements from the United States' electric grid. This GIS data can be straightforwardly exported as a KML file using the in-built functionality of a capable GIS editor (e.g. QGIS as an open-source solution, and ArcGIS as a leading commercial software).

The KML files extracted from the Platts Map Data Pro require several steps of data processing to produce a single XML file that is compatible with the hetero-functional graph-theory toolbox<sup>13</sup>.

- 1) First, any resources marked as canceled, retired, or shutdown are removed from the KML file.
- 2) Second, any resources with duplicate GPS locations are merged such that only one resource exists per GPS location. The meta-data for this resource is adopted from one of the other resources giving preference to power plants, then generation units and then substations as a last option.
- 3) Third, the remaining individual generation units are classified as power plants. These three steps yield a consistent set of formal nodes.
- 4) Fourth, any transmission lines without a well-defined origin or destination formal node are removed.
- 5) Fifth, the remaining resources are organized into a formal graph<sup>14</sup>. All isolated and sub-component nodes and edges not part of the largest connected component are identified. These isolated nodes and clusters are removed from the system.
- 6) Sixth, the system process for each type of resource is inferred. Power plants “generate electric power”, substations “consume electric power”, and transmission lines “transport electric power from origin to destination” and “transport electric power from destination to origin.” Fig. 9b shows the knowledge base of an electric power system where the four types of system capabilities mentioned above are instantiated only once. Fig. 9a shows the associated SysML activity diagram<sup>9</sup>.
- 7) Lastly, the electric power system functional reference architecture shown in Fig. 9a is encoded in the XML file as three valid pairs of system processes; 1.) transmission follows generation, 2.) transmission follows transmission, and 3.) consumption follows transmission. Any other pairs of system processes are invalid and impose constraints in the system sequence constraints matrix  $K_p$ . Fig. 9c shows the hetero-functional graph.

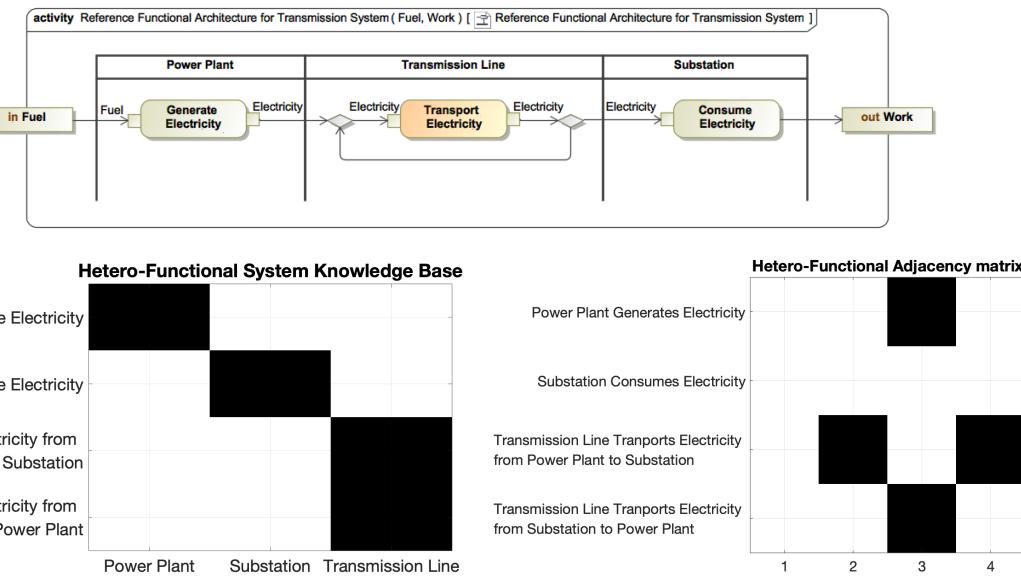


Fig. 9. The example electric power system functional reference architecture a with a single instantiation of each system process pairing. b gives the system knowledge base of the example system. c gives the hetero-functional adjacency matrix of the example electric power system.

The hetero-functional graph theory toolbox<sup>13</sup> is then used to produce the formal and hetero-functional graphs from the XML file described above. The resulting formal graph is composed of 69,386 formal nodes (i.e. power plants and substations) and 105,826 formal edges (i.e. transmission lines). As stated in Section A of the methods, these nodes and edges constitute the 175,212 resources in the system knowledge base while the system processes are defined as shown in Fig. 9b. The resulting system knowledge base  $J_S$  has a size of  $4,814,416,999 \times 175,212$ . Once  $J_S$  is formed, the hetero-functional adjacency matrix is formed using Equation (1).  $K_p$  is formed in a pairwise fashion observing the five types of constraints identified in Section A of the methods. The first four types of constraints are checked numerically, while the last is drawn from the functional pairs in Step 6 above. The resulting hetero-functional adjacency matrix  $A_p$  has a size of  $(8.435 \times 10^{14}) \times (8.435 \times 10^{14})$ . Equation (1) is then used to eliminate empty rows and columns. The projected hetero-functional adjacency matrix  $\tilde{A}_p$  of the American electric grid has a size  $370,220 \times 370,220$  with 1,709,691 capabilities as filled elements. From this point, it was straightforward to calculate the cumulative degree distribution of AEPS’ hetero-functional and formal graphs.

### C. Comparing the Attack Vulnerability of Scale-free Formal and Hetero-functional Graphs

This paper used nodal attacks to compare the attack vulnerability of formal graphs and hetero-functional graphs. In nodal attacks, a set of nodes are identified either at random or by a targeting heuristic such as greatest degree centrality. These nodes are subsequently removed from their respective graphs. In addition to the identified nodes being removed, all edges connected to the removed nodes are also removed<sup>15</sup>. The same physical intuition was maintained between the hetero-functional graph and formal graph during attacks. Rather than attacking a given system capability (as a node in a hetero-functional graph), an entire system resource (as a node in a formal graph) was attacked. All of the associated system capabilities (i.e. the filled elements in the associated column of the knowledge base) were removed. Because the formal edges had been removed in the case of the formal graph, their system capabilities were also removed. For example, if node  $n_3$  in Fig. 2 is attacked, then edges  $e_1$ ,  $e_3$ , and  $e_4$  are removed as well. In the hetero-functional graph, first capabilities  $\psi_3 - \psi_6$  would be removed to reflect the loss of  $n_3$  and then capabilities  $\psi_8, \psi_{10}, \psi_{11}, \psi_{12}$  would all be removed to reflect the loss of edges  $e_1$ ,  $e_3$ , and  $e_4$ .

The nodal attacks were applied as both random and targeted attacks. Rather than attacking a single node at a time, for computational simplicity, each attack iteration removed one percent of the initial node count. In the case of a random attack, the one percent of formal graph nodes were randomly selected and then removed. The corresponding capabilities in the hetero-functional graph were then removed as well. In the case of a targeted attack, the greatest degree centrality heuristic was used to identify the targeted nodes. This one percent of nodes were then removed from the formal graph and the corresponding capabilities were then removed from the hetero-functional graph; thus maintaining the same physical intuition. The most central nodes were then reevaluated and attacked again.

Fig 4a was developed using a random attack and measuring the average degree distribution<sup>15</sup> of each graph at each attack iteration. Fig. 4b was developed using a random attack and measuring the component sizes of each graph at each attack iteration. The largest weakly connected component size<sup>15</sup> was measured and divided by the total remaining nodes in each graph for the largest cluster relative size. The sizes of the remaining connected component (excluding the largest connected component) were then averaged together to get the isolated cluster average size at each attack iteration. Fig 4c utilizes the same measures as Fig 4b. However, a targeted attack based upon greatest degree centrality heuristic was applied to target central nodes.

### D. Predicting the Structural Resilience of the AEPS' Migration to a Decarbonized System Architecture

In Fig 5, the resilience effects of adding the architectural improvements of distributed generation, meshed transmission, and storage were measured. Distributed generation was added randomly to 30% of all the substations; which amounted to 19,828 substations gaining such capabilities. When adding meshed distribution lines, each node with only a single connection to another node was connected to the nearest node to which it was not already connected. An additional 20% of transmission lines were thus added to the AEPS; which amounted to 16,953 additional transmission lines. Storage was added to every buffer that had the potential for generating electricity. These nodes include power plants and substations that were designated to receive distributed generation capabilities. Storage was thus added to 33% of all buffers which amounted to 23,088 additional energy storage resources.

Each type of improvement was added to the AEPS over a series of 100 iterations. The additions of distributed generation, meshed transmission, and storage were analyzed individually. Fig. 5a and Fig. 5b measure the average degree centrality<sup>15</sup> of the formal graph and hetero-functional graph respectively. After each iteration of adding an architectural improvement to the formal graph and the corresponding capabilities to the hetero-functional graph, the average degree centrality was measured for both graphs. It is notable that since both distributed generation and storage are functional additions, the formal graph sees no structural change. Therefore, it sees no measurable change in resilience from distributed generation or storage<sup>4</sup>. However, because the nodes in the hetero-functional graph are system capabilities, there is a measurable structural change in the average degree centrality of the hetero-functional graph as distributed generation and storage are added. Fig. 5c follows the same architectural improvements as Fig. 5a and 5b while measuring the hetero-functional graph's Latent Engineering Resilience (LER)<sup>12</sup>. The LER was normalized by the original base case AEPS with no architectural improvements. The addition of distributed generation and storage report linear increases in the LER while the addition of meshed transmission lines results in an exponential increase in the LER. Intuitively, as lines are added to the AEPS each step of a service delivery path has increasingly more path options; thus exponentially growing the number of deliverable service paths.

### E. Assessing the Resilient Response of Several Decarbonized System Architectures of the AEPS to Random and Targeted Attacks

Fig 6a and Fig 6b measure the LER of the AEPS with different architectural improvements under random and targeted attacks respectfully. Just as previous figures maintained an equivalent physical intuition when attacking hetero-functional and formal graphs that intuition was maintained in Fig 6. Physical buffer resources (i.e. formal nodes) in corresponding formal graphs were selected for random or targeted attacks and all of the associated capabilities of the buffer and its connected edges were removed from the hetero-functional graph. Attacks were applied to one percent of the formal graph nodes at a time in both random and targeted attacks. By maintaining the same physical intuition, removing one percent of nodes being from a formal graph could result in removing more than one percent of the hetero-functional graph nodes. This phenomena is especially noticeable in Fig. 6b as the initial targeted attack results in removing over five percent of the nodes from the hetero-functional graphs. When applying targeted attacks in Fig. 6b, the greatest degree centrality heuristic was used to identify the targeted formal nodes.

After each attack iteration, the LER of the remaining hetero-functional graph was re-evaluated. In all cases, the LER was normalized by the initial unimproved base case AEPS. Thus, the base case takes an initial value of 1 while the architecturally improved AEPSs measure an initial LER value greater than 1. Under random attacks the LER gradually decreases following the regression  $LER = \alpha(100 - x)^2 e^{\beta(100-x)}$ . However, the targeted attack causes a large initial drop paired with a faster decrease in the LER following the regression  $LER = \alpha e^{\beta x}$ . The initial LER drop is caused because the attack of one percent of attacked formal nodes removes over five percent of the nodes in the hetero-functional graph. Using degree centrality as the targeting method, buffers with the most transmission lines are removed first. The first attack therefore removes power plants, substations, and the largest number of transmission lines in a single attack. As buffers connected to the most transmission lines are targeted, there are increasingly fewer path options for services to be delivered. Thus, similar to the addition of meshed transmission lines in Fig. 5c which yields an exponential increase in the LER, targeting buffers by degree centrality yields an exponential decay in the LER.

### F. Data Availability

The data that support the findings of this study are commercially available from Platts but restrictions apply to the available dataset, which were used under license for the current study, and are not publicly available. Data are however available from the authors upon reasonable request with express written consent of Platts.

### METHODS RESOURCES

- [13] P. Hegde, W. C. Schoonenberg, D. Thompson, and A. M. Farid, “The Hetero-functional Graph Theory Toolbox,” *scientific reports*, vol. 1, no. 1, p. 11, 2020.
- [14] J. Leskovec and R. Sosić, “Snap: A general-purpose network analysis and graph-mining library,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, p. 1, 2016.
- [15] M. Newman, *Networks: An Introduction*. Oxford, United Kingdom: Oxford University Press, 2009. [Online]. Available: <http://books.google.ae/books?id=LrFaU4XCsUoC>

### II. AUTHOR CONTRIBUTIONS

Dakota Thompson was the lead author and primary analyst of the quantitative results. Wester C. H. Schoonenberg was the secondary analyst to the quantitative results. Amro M. Farid designed the research scope and questions and guided conduct of the research to a conclusion.

### III. ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Dakota Thompson at Dakota.J.Thompson.Th@Dartmouth.edu. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).