LLM-Based Community Surveys for Operational Decision Making in Interconnected Utility Infrastructures

Adaeze Okeukwu-Ogbonnaya[1], Rahul Amatapu[2], Jason Bergtold[1], and George Amariucai[1]

[1] Kansas State University, Manhattan, Kansas
[2] The University of Kansas, Lawrence, Kansas

Abstract. We represent interdependent infrastructure systems and communities alike with a hetero-functional graph (HFG) that encodes the dependencies between functionalities. This graph naturally imposes a partial order of functionalities that can inform the sequence of repair decisions to be made during a disaster across affected communities. However, using such technical criteria alone provides limited guidance at the point where the functionalities directly impact the communities, since these can be repaired in any order without violating the system constraints. To address this gap and improve resilience, we integrate community preferences to refine this partial order from the HFG into a total order. Our strategy involves getting the communities' opinions on their preferred sequence for repair crews to address infrastructure issues, considering potential constraints on resources. Due to the delay and cost associated with real-world survey data, we utilize a Large Language Model (LLM) as a proxy survey tool. We use the LLM to craft distinct personas representing individuals, each with varied disaster experiences. We construct diverse disaster scenarios, and each simulated persona provides input on prioritizing infrastructure repair needs across various communities. Finally, we apply learning algorithms to generate a global order based on the aggregated responses from these LLM-generated personas.

Keywords: Hetero-Functional Graph · Interdependent Infrastructure · Ranking · LLM Personas · Repair Prioritization · Community Preferences

1 Introduction

Natural disasters often damage multiple infrastructures simultaneously, creating complex challenges for prioritizing repairs. Traditional approaches rely heavily on technical system constraints and expert judgment, which may overlook the urgent needs of vulnerable communities. Recent studies show that purely technical prioritization often fails to address immediate local needs[11][13][27], which underscores the importance of incorporating social consideration into recovery planning[33][12].

To address this, there is growing interest in integrating community perspectives into infrastructure restoration to ensure fair allocation of limited resources[5][10].

However, directly collecting community input through surveys can be costly and slow, especially immediately after disasters. Large Language Models (LLMs) present a new opportunity by generating synthetic personas with diverse demographic and socioeconomic profiles. They can approximate community preferences at scale, complementing traditional technical assessments.

In this paper, we propose a novel framework that combines the Hetero-functional Graphs (HFGs) model of interdependent infrastructures to derive partial technical repair orders, LLM-generated personas to simulate diverse community preferences over repair priorities, and a neural network-based pairwise ranking model with chainization, converting these inputs into a global prioritization.

We demonstrate this approach using a simplified infrastructure model as a proof of concept, analyze sensitivity to prompt variations, and discuss limitations and future works.

2 Literature Review

Hetero-functional graphs (HFGs) are widely used to model interdependent infrastructure systems in smart cities, capturing dependencies across power, water, and other networks [29][31][24]. While prior work applied HFGs to analyze system robustness under failures, their use in guiding repair prioritization during disaster recovery remains largely unexplored. Our work extends HFGs into this domain by using them to derive partial technical orders of restoration.

Large Language Models (LLMs) have recently been explored for simulating stakeholder opinions and community-like feedback. For instance, Dolant and Kumar [9] proposed a multi-agent LLM framework where simulated personas engage in decision discourse under disaster scenarios, while Li et al. [21] demonstrated how LLM-generated personas could approximate public preferences in contexts like election forecasting, also highlighting potential biases. Shi et al. [30] used multi-perspective LLM debates to mimic diverse community viewpoints. Other studies, such as Xie et al. [32] and Chen et al. [4,3], applied LLMs in emergency planning, integrating user-centric preferences into decision support systems. While these works show LLM potential for approximating community input, concerns remain about fairness and systemic biases [21,30]. Our study addresses this by performing sensitivity analyses on prompt formulations to detect shifts in generated preferences.

Infrastructure prioritization studies often combine technical and social factors through weighted decision models. For example, rankings may incorporate asset condition, cost-benefit, and efficiency alongside equity or vulnerability scores [23,6,7,8,19]. However, technical factors typically dominate, with social metrics playing a secondary role or being reflected only indirectly through regulator or stakeholder weighting [28,19,1]. Some research shows that explicitly including social vulnerability can alter priorities, such as favoring school access or walkability [23,8,7], but few directly gather community preferences. In contrast, we use LLMs to simulate community perspectives where no technical total order exists, that is, when multiple repairs are equally feasible from a system standpoint, allowing community-like priorities to resolve ties.

Finally, we build on advances in order learning from pairwise comparisons. Matrix completion approaches [26], nuclear-norm regularization for partial orders [16], and graph-based methods like GNNRank [18] all recover global rankings from limited data. Lee and Kim's chainization technique [20] augments partial orders with pseudo-comparisons to learn full rankings, which we adapt here to generate comprehensive infrastructure repair sequences from LLM-derived pairwise preferences.

3 Methodology

In this section, we describe our approach to modeling interdependent infrastructures with a hetero-functional graph (HFG), simulating community preferences using Large Language Models (LLMs), learning pairwise priorities via a neural network with chainization, and testing sensitivity to prompt variations.

3.1 Hetero-Functional Graph (HFG) Model

We model the infrastructure system and interdependencies across three communities using an HFG, where nodes represent functionalities (generation, transport, storage, consumption) and directed edges capture dependencies. Figure 1 shows the bottom part of the HFG in our model. The graph induces a partial order; for example, consumption nodes like "Consume Water in Residential Area 1" can be restored only after upstream repairs, but are independent of each other, leaving ambiguity when repair crews are limited. To resolve this, we integrate community preferences to rank among technically equivalent actions. The HFG data is available at [25].

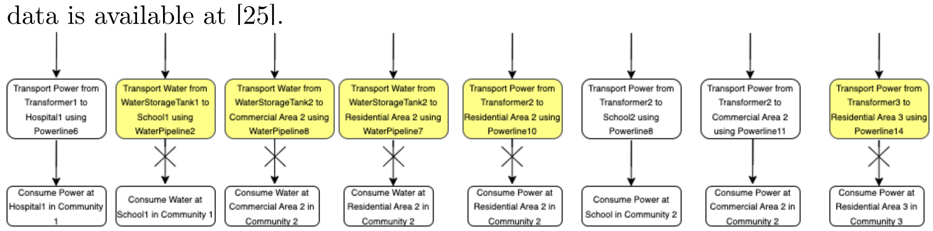

Fig. 1: A subset of the hetero-functional graph (HFG) showing dependencies across power and water systems. The final layer consists of consumption functionalities that can be repaired in any order. Arrows represent functional dependencies.The highlighted functionalities are damaged while the links marked × are paths requiring repair before the associated community-level services (bottom layer) can be restored.

3.2 Simulating Community Preferences with LLMs

We generate 200 synthetic personas using GPT-3.5-turbo via LangChain with few-shot prompts, each defined by 44 attributes. These include demographic, socioeconomic, and linguistic characteristics modeled after American Community

Survey (ACS) categories, along with additional variables such as disaster experience and geographic class. This was at a temperature of 0.9 through the OpenAI API on April 2nd, 2025, between 2:00 PM and 4:00 PM Central Time. We assign each persona disaster scenarios describing damage across three communities with different social vulnerability scores (SVS)[14], asking them to choose between pairs of infrastructure repairs. The prompts include the full functional status of all communities, and a choice between two repair options (e.g., water in schools vs. power in residential areas). We use the GPT-3.5-turbo model with a temperature of 0.7 and generated responses on April 3, 2025, between 2:00 PM and 4:00 PM Central Time. This process produces all unique pairwise comparisons among damaged functionalities, totaling 36 questions aggregated over 7,200 persona responses, which yields normalized soft labels for each question represented as probability vectors $[p_1, p_2]$, which indicates the fraction of personas who preferred each option. Both community-aware and community-unaware datasets are produced to test whether identity affects choices. In the community-aware dataset, each persona knows their assigned community, while in the community-unaware dataset, community labels are removed from prompts. All data, including prompt templates, are publicly available at [25].

3.3 Feature Extraction & Data Processing

Each repair option is encoded with features listed in Table 1, including normalized SVS. The SVS is originally defined on a 1–10 scale in our scenario specifications to illustrate community vulnerability levels, but we normalize it to a 0–1 scale before feeding it into the neural network. We represent each pairwise question as two feature vectors and a probability label $[p_1, p_2]$. In total, we aggregate 36 distinct pairwise questions drawn from 7,200 persona responses.

Table 1: Infrastructure Repair Option Features

Feature	Description
Infrastructure Type	One-hot: Water, Power
Facility Type	One-hot: Hospital, Residential, Commercial, School
Community ID	One-hot: Community 1, 2, 3
Geographical Class	One-hot: Urban, Suburban, Rural
Social Vulnerability Score (SVS)	Continuous (0 to 1)

3.4 Pairwise Comparator and Chainization Model

We train a comparator neural network to predict preferences between pairs of repair options. Each option is processed independently by a shared feedforward branch with two hidden layers of size 64 and ReLU activations. The resulting embeddings are concatenated and passed to a final linear layer, which produces logits for the two choices, followed by a softmax. We train using a hybrid loss, that is, cross-entropy for strong preferences ($P_1 \neq P_2$), and KL-divergence toward $[0.5, 0.5]$ when the soft labels indicate near ties. Optimization uses Adam with a batch size of 8. The learning rate 1×10^{-3} was selected after testing alternatives (5×10^{-4} and 1×10^{-4}) via six-fold cross-validation, yielding the lowest mean validation loss. We apply early stopping with a patience of six bad epochs to avoid overfitting.

After training, we apply Chainization where we use the model predictions to form a directed graph $G = (V, E)$ where nodes are repair options and edges ($option2 \rightarrow option1$) indicate that option 1 is preferred over option 2. We use PageRank to derive a global priority order, since, unlike topological sorting, it handles cycles by assigning importance scores.

3.5 Random Subset Sampling and Fine-Tuning with Pseudo-Pairs

To reduce survey burden, we train the comparator using the same hybrid loss on random subsets of 20%, 30%, 40%, 50%, 60%, and 75% of the pairwise data, then infer an initial global ranking. We generate pseudo-pairs from this ranking: if A outranks B, we add (A, B) with a hard label $[1.0, 0.0]$. We then fine-tune on the combined original samples and pseudo-pairs, again using the same hybrid loss (cross-entropy and KL), which gives us final pairwise predictions.

3.6 Validation and Sensitivity Analysis

We compare global rankings from community-aware vs. community-unaware personas to see if identity shifts priorities. Additionally, we assess prompt sensitivity by regenerating persona responses and repeating the full modeling pipeline for six prompt variations: (1) reworded questions rephrased with alternative criticality language, (2) reworded requests for reasoning style, (3) empathetic framing emphasizing human impact, (4) modified instruction layout into explicit step-wise format, (5) choice-first structure presenting the decision before justification, and (6) removing SVS information entirely to test vulnerability influence [22,2,15,17]. We compute Kendall τ correlations and Top-k overlaps with the baseline.

4 Results

We report results on the impact of community awareness on prioritization, the effect of limiting the number of preference queries using chainization and pseudo-pairs, and on sensitivity to prompt variations.

4.1 Impact of Community Awareness in Simulated Preferences

We compare rankings from community-aware vs. community-unaware personas to test if explicit identity shifts priorities. Table 2 shows the final repair ranking, which is identical for both the community-aware and community-unaware personas. While we expected in-group bias, our result suggests that personas weighted other factors more heavily.

We observe that repairs to schools and residential areas dominate the top ranks. Water in a school in Community 3 is consistently ranked first, followed by water in Community 1's school. Most residential repairs follow, with power typically prioritized over water within the same community. The only commercial repair ranks last, which suggests lower perceived urgency. Figure 2 illustrates that even when aware of identity, Community 3 personas chose their own community only 41.6% of the time, community awareness had little impact on repair preferences and that broader vulnerability awareness likely drove the decision-making.

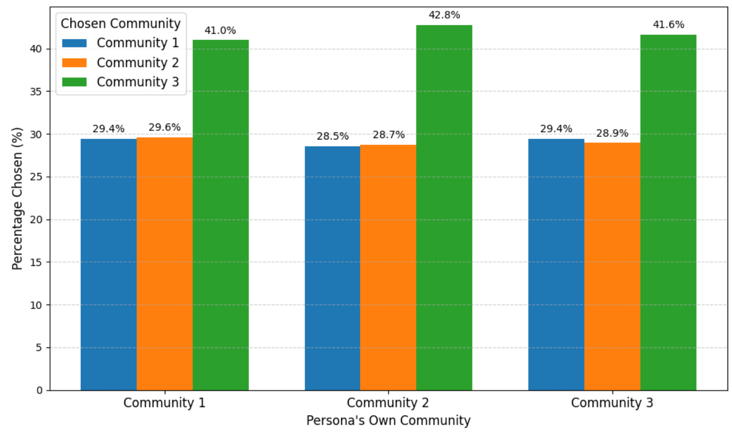

Fig. 2: Community Preferences by Community-Aware Personas

4.2 Accuracy with Limited Pairwise Comparisons

We next test if we can recover global rankings from fewer pairwise inputs. Figure 3 shows mean Kendall's τ correlations (averaged over 100 runs) between inferred and full rankings across sampling fractions. Using only 20–30% of the comparisons yields $\tau < 0.6$, but above 50% we achieve $\tau > 0.7$, indicating that good rankings can be obtained with fewer queries, reducing survey burden.

Table 2: Final Global Repair Prioritization (Same for Community-Aware and Community-Unaware Datasets)

Rank	Repair Option
1	Repair Water School in Community 3
2	Repair Water School in Community 1
3	Repair Power Residential in Community 2
4	Repair Power Residential in Community 3
5	Repair Water Residential in Community 3
6	Repair Power Residential in Community 1
7	Repair Water Residential in Community 2
8	Repair Water Residential in Community 1
9	Repair Water Commercial in Community 2

4.3 Sensitivity to Prompt Variations

We assess robustness by rerunning the entire pipeline under six prompt variants. Table 3 shows item ranks, while Figure 4 summarizes Kendall's τ correlations and Top-k overlaps with the baseline. We see that prompt variants yield moderate agreement (τ from 0.39 to 0.72), with only the SVS-removed prompt achieving high correlation. We observe that Top-5 overlaps remain high across cases, suggesting top priorities are stable while lower ranks shift.

More so, our results indicate that SVS strongly influences persona prioritization and outweighs community affiliation in repair choices. This raises questions

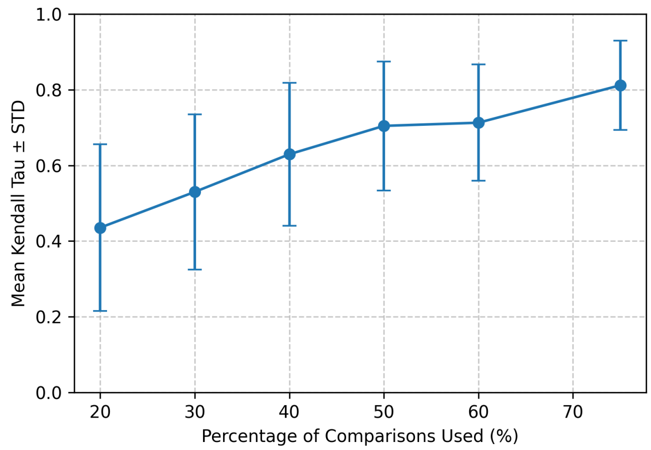

Fig. 3: Kendall's τ vs. percentage of comparisons used.

about whether the LLM separates vulnerability from affiliation when reasoning, or treats them as interchangeable signals of need. Additionally, our sensitivity analysis shows that the repair rankings remain highly dependent on the prompts provided, with SVS as a dominant factor.

Table 3: Global Repair Prioritization Across Prompt Variations

Repair Option	Instruction Format	Reworded Question	Reworded Reasoning	Shifting Tone	Choice First	Without SVS
Repair Water School in Community 3	1	1	1	1	3	2
Repair Water School in Community 1	4	2	6	2	5	5
Repair Power Residential in Community 2	6	3	5	6	4	1
Repair Power Residential in Community 3	3	8	2	5	1	3
Repair Water Residential in Community 3	2	7	3	3	2	4
Repair Power Residential in Community 1	8	6	9	8	8	6
Repair Water Residential in Community 1	7	5	7	7	7	8
Repair Water Residential in Community 2	5	4	4	4	6	7
Repair Water Commercial in Community 2	9	9	8	9	9	9

5 Conclusion

This work presents a new approach to infrastructure repair prioritization by combining hetero-functional graphs, community preferences generated through large language models, and a machine learning model trained on pairwise comparisons. Our method addresses the common challenge of deciding what to repair first when technical dependencies allow for multiple valid options. By simulating

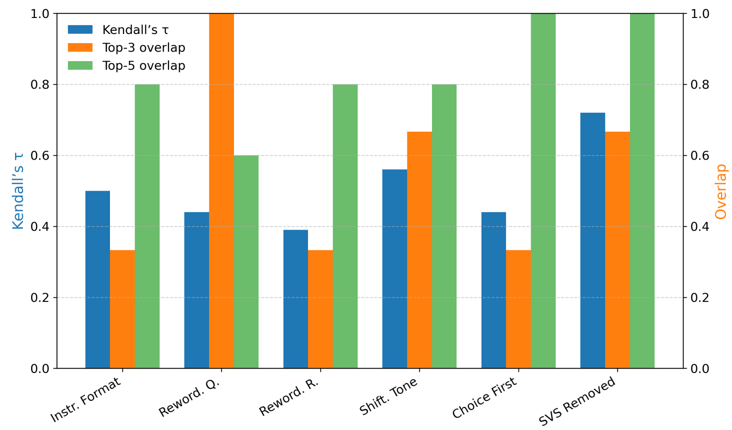

Fig. 4: Prompts Variations Overall Agreement with Original using Kendall's τ and Top-K overlaps.

community input with synthetic personas and applying chainization, we generate rankings that reflect both system constraints and community priorities.

We find that LLM-generated personas provide consistent preferences across many disaster scenarios. Even when personas are unaware of their community identity, their rankings align with priorities such as protecting vulnerable populations and restoring essential services. Our results show that full-data rankings can be approximated using a reduced set of comparisons, lowering the burden of data collection. And we expect that with actual survey data, the selection will likely be more accurate. We also tested how prompt changes affect the outcomes and found that while lower-ranked items shift, top priorities remain stable across variations.

There are some limitations. The generated preferences may not fully reflect how real people behave under stress. Also, the infrastructure model is simplified and includes a fixed list of repair tasks. Our method also combines all infrastructure types into a single priority list, which may not reflect how repairs are typically carried out by specialized teams. However, the approach can still support broader planning efforts such as resilience investment and upgrade decisions, where cross-sector comparisons are useful. Also, we do not compare against classical expert heuristics or existing survey-based methods, or run explicit ablations to isolate the impact of pseudo-pair fine-tuning. Testing on larger, more realistic systems remains unexplored. We also do not consider social cognitive factors such as measurements of community connectedness and personality traits.

Future research should apply this method to more complex systems, validate results with real stakeholder input, explore better ways to select informative comparisons, and compare to traditional prioritization approaches. It will also be important to separate infrastructure types, and incorporate ablation studies. Despite these limitations, our unified ranking approach offers a practical framework for recovery and planning in disaster settings.

Acknowledgments

This research was supported by the U.S. National Science Foundation (NSF) under Grant No. 2148878. We thank Amin Enderami, Elaina Sutley, and Eugene Vasserman for their contribution to this work.

References

1. Baur, R., Le Gauffre, P., Sægrov, S.: Multi-criteria decision support for annual rehabilitation programmes in drinking water networks. Water Science and Technology: Water Supply **3**(1-2), 43–50 (2003)
2. Chatterjee, A., Renduchintala, H.K., Bhatia, S., Chakraborty, T.: Posix: A prompt sensitivity index for large language models. arXiv preprint arXiv:2410.02185 (2024)
3. Chen, M., Tao, Z., Tang, W., Qin, T., Yang, R., Zhu, C.: Enhancing emergency decision-making with knowledge graphs and large language models. International Journal of Disaster Risk Reduction **113**, 104804 (2024)
4. Chen, S., Luo, G., Chen, L.: Intelligent urban emergency response: Integrating large language models, multi-objective optimization, and reinforcement learning. In: 2024 IEEE Smart World Congress (SWC). pp. 1469–1476. IEEE (2024)
5. Chen, Y., Ji, W.: Public demand urgency for equitable infrastructure restoration planning. International Journal of Disaster Risk Reduction **64**, 102510 (2021)
6. Das, R., Nakano, M.: A multi-criteria decision-making model using socio-technical attributes for transportation bridge maintenance prioritisation. International Journal of Construction Management **23**(4), 579–585 (2023)
7. Dell'Anna, F., Berta, M., Bottero, M., Mallia, G., Morgese, V.: Multicriteria-decision support for master plan scheduling: urban regeneration of an industrial area in northern italy. Construction Management and Economics **42**(5), 476–501 (2024)
8. Dhansinghani, R., Ibrahim, A., Kannoth, A., Miller, C., Nguyen, L., Pham, S., Bailey, R.: Design of a prioritization methodology for equitable infrastructure planning. In: 2022 Systems and Information Engineering Design Symposium (SIEDS). pp. 315–320. IEEE (2022)
9. Dolant, A., Kumar, P.: Agentic llm framework for adaptive decision discourse. arXiv preprint arXiv:2502.10978 (2025)
10. Domingue, S.J., Emrich, C.T.: Social vulnerability and procedural equity: Exploring the distribution of disaster aid across counties in the united states. The American Review of Public Administration **49**(8), 897–913 (2019)
11. Doorn, N., Gardoni, P., Murphy, C.: A multidisciplinary definition and evaluation of resilience: The role of social justice in defining resilience. Sustainable and Resilient Infrastructure **4**(3), 112–123 (2019)
12. Drakes, O., Tate, E., Rainey, J., Brody, S.: Social vulnerability and short-term disaster assistance in the united states. International Journal of Disaster Risk Reduction **53**, 102010 (2021)
13. Eisenman, D.P., Cordasco, K.M., Asch, S., Golden, J.F., Glik, D.: Disaster planning and risk communication with vulnerable communities: lessons from hurricane katrina. American journal of public health **97**(Supplement_1), S109–S115 (2007)
14. Enderami, S.A., Sutley, E.: Social vulnerability score: a scalable index for representing social vulnerability in virtual community resilience testbeds. Natural Hazards **120**(7), 6457–6480 (2024)

15. Gandhi, V., Gandhi, S.: Prompt sentiment: The catalyst for llm change. arXiv preprint arXiv:2503.13510 (2025)
16. Gunasekar, S., Koyejo, O.O., Ghosh, J.: Preference completion from partial rankings. Advances in Neural Information Processing Systems **29** (2016)
17. He, J., Rungta, M., Koleczek, D., Sekhon, A., Wang, F.X., Hasan, S.: Does prompt formatting have any impact on llm performance? arXiv preprint arXiv:2411.10541 (2024)
18. He, Y., Gan, Q., Wipf, D., Reinert, G.D., Yan, J., Cucuringu, M.: Gnnrank: Learning global rankings from pairwise comparisons via directed graph neural networks. In: international conference on machine learning. pp. 8581–8612. PMLR (2022)
19. Karasneh, S., Moqbel, S.: Priority-based decision model for rehabilitation of water networks using fahp. Water Conservation & Management **8**(1) (2024)
20. Lee, S.H., Kim, C.S.: Order learning using partially ordered data via chainization. In: European Conference on Computer Vision. pp. 196–211. Springer (2022)
21. Li, A., Chen, H., Namkoong, H., Peng, T.: Llm generated persona is a promise with a catch. arXiv preprint arXiv:2503.16527 (2025)
22. Mizrahi, M., Kaplan, G., Malkin, D., Dror, R., Shahaf, D., Stanovsky, G.: State of what art? a call for multi-prompt llm evaluation. Transactions of the Association for Computational Linguistics **12**, 933–949 (2024)
23. Mohamadiazar, N., Ebrahimian, A., Azizinamini, A.: Equitable prioritization of bridge rehabilitation projects using a spatial multi-criteria decision support framework. Transportation Research Record p. 03611981241254110 (2024)
24. Munikoti, S., Lai, K., Natarajan, B.: Robustness assessment of hetero-functional graph theory based model of interdependent urban utility networks. Reliability Engineering & System Safety **212**, 107627 (2021)
25. Okeukwu-Ogbonnaya, A., Amatapu, R.: CommRank-HFG. `https://github.com/adaezy/CommRank-HFG` (2025)
26. Park, D., Neeman, J., Zhang, J., Sanghavi, S., Dhillon, I.: Preference completion: Large-scale collaborative ranking from pairwise comparisons. In: International Conference on Machine Learning. pp. 1907–1916. PMLR (2015)
27. Paton, D., Johnston, D.: Disasters and communities: vulnerability, resilience and preparedness. Disaster Prevention and Management: An International Journal **10**(4), 270–277 (2001)
28. Pramesti, N.P., Priyosulistyo, H., Aminullah, A., Koesmargono, A.: The bridge maintenance factors model: a pls-sem approach. Age **22**, 18 (2021)
29. Schoonenberg, W.C., Khayal, I.S., Farid, A.M.: A Hetero-functional graph theory for modeling interdependent smart City infrastructure. Springer (2019)
30. Shi, L., Liu, H., Wong, Y., Mujumdar, U., Zhang, D., Gwizdka, J., Lease, M.: Argumentative experience: Reducing confirmation bias on controversial issues through llm-generated multi-persona debates. arXiv preprint arXiv:2412.04629 (2024)
31. Thompson, D.J., Farid, A.M.: A hetero-functional graph structural analysis of the american multi-modal energy system. Sustainable Energy, Grids and Networks **38**, 101254 (2024)
32. Xie, Y., Jiang, B., Mallick, T., Bergerson, J.D., Hutchison, J.K., Verner, D.R., Branham, J., Alexander, M.R., Ross, R.B., Feng, Y., Levy, L.A., Su, W., Taylor, C.J.: A rag-based multi-agent llm system for natural hazard resilience and adaptation (2025), `https://arxiv.org/abs/2402.07877`
33. Zhai, W., Peng, Z.R., Yuan, F.: Examine the effects of neighborhood equity on disaster situational awareness: Harness machine learning and geotagged twitter data. International Journal of Disaster Risk Reduction **48**, 101611 (2020)