Referee: 1

Comments to the Author

Please find attached

In this paper, the authors propose two optimization-based methods to identify the most lethal disease cliques from comorbidity graphs, derived from large-scale electronic health records (EHRs). The first method uses a mixed-integer linear programming (MILP) formulation, while the second method extends the Bron–Kerbosch algorithm for clique enumeration. The paper provides a detailed computational study using data from 10.6 million patient records to evaluate the effectiveness of both approaches.

However, several major issues and limitations need to be addressed 1)Technical Novelty: The combinatorial optimization approach, while helpful, lacks significant innovation and requires stronger justification, especially in the context of prior work on EHR data analysis 2) Comorbidity Modeling: The use of cliques to model comorbidities simplifies complex disease interactions and does not fully capture the distinctions of multiple disease relationships 3) Epidemiological Modeling: The assumption of longitudinal data in defining mortality is not fully compatible with the EHR dataset used, which may limit the accuracy of the findings.

* The approach offers advantages in handling large datasets, providing exact and incremental solutions, and offering a flexible framework applicable to real-world healthcare settings. However, the technical novelty of the combinatorial optimization method is quite limited. Network formulations from EHR data have been explored extensively in previous literature. The authors need to provide stronger justification for their methodological choices to demonstrate how this work advances the state of the art.
* The way comorbidities are modeled in the graph is not sufficiently justified. The edge set in the comorbidity graph is defined as containing an edge if and only if diseases u and v frequently co-occur in EHR data. However, the paper does not clearly define what constitutes “frequent co-occurrence” or provide a threshold or rationale for determining this frequency. This lack of clarity makes it difficult to assess the robustness of the graph structure.
* The authors use cliques (subsets of pairwise adjacent vertices) to define comorbid diseases, but this approach oversimplifies comorbidity relationships. In reality, comorbidities often involve complex interactions among multiple diseases that may not neatly fit into the structure of pairwise adjacent vertices. Diseases may have indirect or complex associations that vary across different patients. This modeling choice should be reconsidered or more thoroughly justified.
* The paper’s definition of mortality seems to assume longitudinal data, which is not fully satisfied in the context of the EHR data used in this study. The data may not capture long-term outcomes, as it represents a snapshot of patient encounters rather than continuous patient tracking over time. The authors need to address this limitation, perhaps by adjusting their definition of mortality or discussing how it impacts the study's conclusions.
  + Our definition of mortality in this study includes two key outcomes:
    - In-hospital mortality (expirations within the hospital)
    - Hospice enrollment.

Our EHR dataset captures longitudinal data that tracks patient encounters over time, providing a continuous record of patient progress until death either in-hospital or under hospice care. Therefore, our use of the EHR data reflects a longitudinal perspective, rather than merely a snapshot of isolated encounters. We believe this approach is appropriate for addressing mortality within the context of this study.

* In the MILP formulation, the specific upper-bound value bbb for clique size is not well justified. Why is it limited to smaller sizes (e.g., 4)? A sensitivity analysis of different values could enhance the understanding of how clique size influences mortality rates and provide better insights into the interactions among diseases.
* Although the Bron–Kerbosch algorithm is well-established for clique enumeration, the manuscript lacks a discussion of the potential drawbacks of using it in healthcare applications. For example, this algorithm may overlook important disease interactions that do not form strict cliques but still have high mortality rates. The authors might consider exploring alternative methods, such as fault-tolerant clique relaxations, to capture these non-clique interactions.
* The study provides interesting computational results, but the clinical relevance of the identified cliques is not discussed in detail. For example, how do these disease combinations align with current clinical knowledge about disease progression? A deeper analysis of the medical significance of these findings would add practical value to the study.
  + On Page 20: “The results in Table 3 have not revealed new insights of medical significance, as the diseases featured in the five highest mortality rate cliques for each value of parameter b are well-known within the medical community to be the most lethal comorbidities [Add some references and discussions]. However, the results do provide a verification of clinical insights through an analysis of large-scale EHR data. Furthermore, the fact that the proposed approach is not detecting spurious comorbidities among its top-K cliques is encouraging, as the detection of spurious clusters is a common concern in any graph-based data mining approach. The procedure followed in preparing the dataset and the design of the modified BK algorithm seem to be offering a practically effective and useful framework for mortality rate analysis of comorbidities.”
  + On pages 20-21: “From the results in Table 4, the individual mortality rate linked to secondary malignancies is the highest among all the (remaining) diseases, although it is much lower at 0.24. Notably, when paired with septicemia it corresponds to the highest mortality rate of 0.52, more than doubling mortality rate of secondary malignancies and more than the sum of the individual mortality rates [Add some references and discussions]. These two diseases continue to remain among the top-two highest mortality rate cliques when we increase b to three and then to four, with the respective rates increasing to 0.58 and then to 0.63.”
  + On page 21: “We can also see from the results in Tables 5 and 6 that this experiment offers information complementary to that reported in Table 4. Through this experiment we can recognize the disease(s) to be most wary of when a patient subgroup is already diagnosed with an existing clique of diseases, which is the main purpose behind investigating marginal mortality rates, to identify lethal comorbidities that are specific to the patient population under consideration.”
    - Expand discussion regarding Tables 5 & 6 results based on medical references to see if there are new insights discovered?
* The authors mention that the MILP approach struggles with scalability, but they do not offer alternatives or recommendations for dealing with larger datasets beyond the proposed Bron–Kerbosch modification. Could the authors explore alternative decomposition or heuristic methods to improve scalability for the MILP formulation?

Minor Issues:

* Table Representation: In Tables 1 and 2, which summarize the computational results, the authors could include confidence intervals or statistical metrics to provide better insight into the stability of their results.
* Discussion on Missing Data: The manuscript briefly mentions the exclusion of patients without recorded diagnoses. Could the authors discuss how missing and incorrect data might affect the validity of their results and suggest possible mitigation strategies?
  + Include some narratives in a Limitation section?
  + The data cleaning process excluded patients with missing diagnosis codes, age, race, gender, and discharge status (indicating mortality). Missingness is a common issue in EHR data. Additionally, our dataset only includes patients from hospitals using Cerner EHR systems. Information may be missing if patients fail to comply with follow-ups, treatments, or seek care at institutions using other EHR systems. Although our results align with existing literature, external validation with diverse datasets would further enhance the generalizability and robustness of our models.
* Future Research Directions: The authors propose several potential improvements, such as integrating temporal comorbidity graphs. This is an exciting direction, but more specific details or preliminary results could enhance the argument for this future work.