Mapping the Network of Environmental Exposure Using Mixed Graphical Models

Keywords: environmental exposure, KONEHS, mixed graphical models, exposure mixtures, complex networks

Saif Nasr[1], Hyung Doo Kim[2], Inho Hong[1]*

[1]Graduate School of Data Science, Chonnam National University, Gwangju, Korea.
[2]Department of Occupational and Environmental Medicine, Hanyang University Guri Hospital, Guri, Republic of Korea.
*ihong@jnu.ac.kr

Extended Abstract

Motivation. Environmental health studies often examine individual pollutants as independent risk factors for health outcomes. However, exposures typically occur as complex mixtures shaped by shared behavioral and structural determinants [1], reflecting common lifestyle patterns and environmental contexts. Treating them as independent may obscure important structural relationships within exposure systems and limit understanding of how multiple exposures interact simultaneously. Here, this study asks:
(1) How are exposures interconnected within a population?
(2) Do they form modular communities reflecting shared pathways?
(3) Which exposures occupy central positions within the system?
Addressing these questions will advance mixture research by adopting a systems-oriented perspective on environmental exposures and providing a more comprehensive view of population-level exposure dynamics.

Approach and Methodology. We analyze mixed-type exposure data from the Korean National Environmental Health Survey (KONEHS), which includes biomarker measurements, lifestyle behaviors, dietary variables, and demographic characteristics from 4239 individuals. The dataset comprises continuous, binary, and categorical variables representing heterogeneous exposure domains.
As a preliminary step, we estimated a Gaussian Graphical Model (GGM) [2] to explore the conditional dependency structure among exposure variables through sparse partial correlation estimation. This prototype network provides an initial structural approximation of exposure interdependencies and allows visualization of potential clustering patterns and variations in node connectivity.
Building upon this exploratory framework, we extend the analysis using Mixed Graphical Models (MGM) [3], which enable simultaneous estimation of conditional relationships across mixed variable types within a unified high-dimensional setting. Unlike GGM, MGM accommodates non-Gaussian and categorical variables while preserving sparse network structure. Following network estimation, community detection and centrality analysis [4][5] will be applied to identify modular substructures and structurally important exposures. Finally, demographic characteristics will be examined in relation to the identified communities to contextualize the exposure network within population subgroups (see Figure 1a).

Results. The preliminary GGM network (Figure 1b) reveals structured conditional dependencies among exposure variables, with heterogeneous connectivity and identifiable clustered substructures. Notably, biomarker measurements of related heavy metals tend to cluster together, while lifestyle and dietary variables form separate but connected modules, reflecting shared exposure pathways. These patterns indicate that exposures are organized into domain-specific communities rather than operating independently. The MGM analysis will extend this framework to mixed variable types and provide a comprehensive representation of exposure interactions.

Conclusions and Outlook. This study proposes a systems-oriented framework for modeling environmental exposure mixtures using Mixed Graphical Models. By conceptualizing exposures as interconnected networks rather than independent risk factors, the approach provides a structural perspective on exposure mixtures. Network estimation, community detection, centrality analysis, and demographic interpretation will provide insights into the exposure mixtures originating from individual demography, lifestyle, and environment.

This work was supported by KOITA grant funded by MSIT (KOITA20250002-83) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00242528, 2025S1A5C3A0100835411).

References

[1] C. Billionnet et al. “Statistical approaches for studying the health effects of exposure to multiple environmental chemicals.” In: *Environmental Health Perspectives* 120(5) (2012), pp. 623–630.
[2] J. Friedman, T. Hastie and R. Tibshirani. “Sparse inverse covariance estimation with the graphical lasso.” In: *Biostatistics* 9(3) (2008), pp. 432–441.
[3] J. M. B. Haslbeck and L. J. Waldorp. “Structure estimation for mixed graphical models in high-dimensional data.” *arXiv preprint arXiv:1510.05677* (2015).
[4] D. Borsboom and A. O. J. Cramer. “Network analysis: An integrative approach to the structure of psychopathology.” In: *Annual Review of Clinical Psychology* 9 (2013), pp. 91–121.
[5] S. Epskamp et al. “Estimating psychological networks and their accuracy: A tutorial paper.” In: *Behavior Research Methods* 50(1) (2018), pp. 195–212.

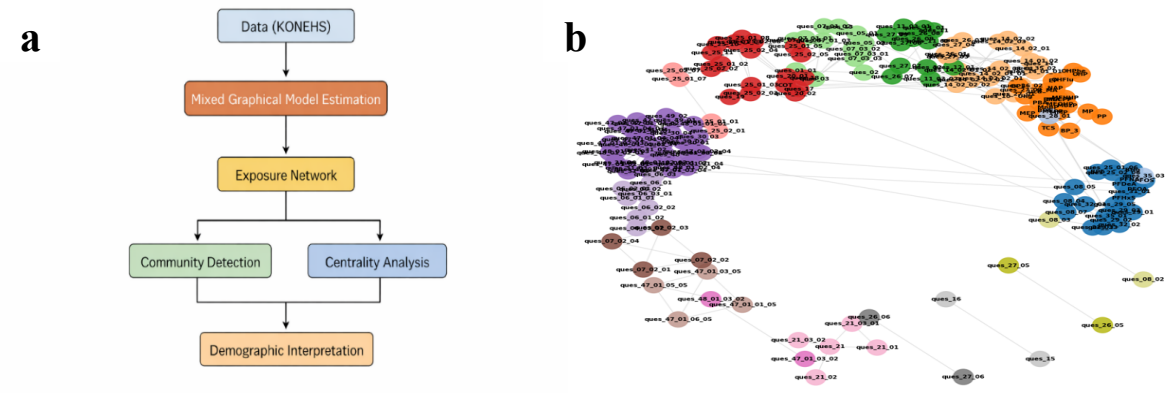

Figure 1. **(a)** Network-based analytical pipeline for modeling environmental exposure mixtures. (**b)** Preliminary exposure network estimated using a Gaussian Graphical Model (GGM)