Identifying Influential Individuals in a Causal Network Through Random Effects Framework

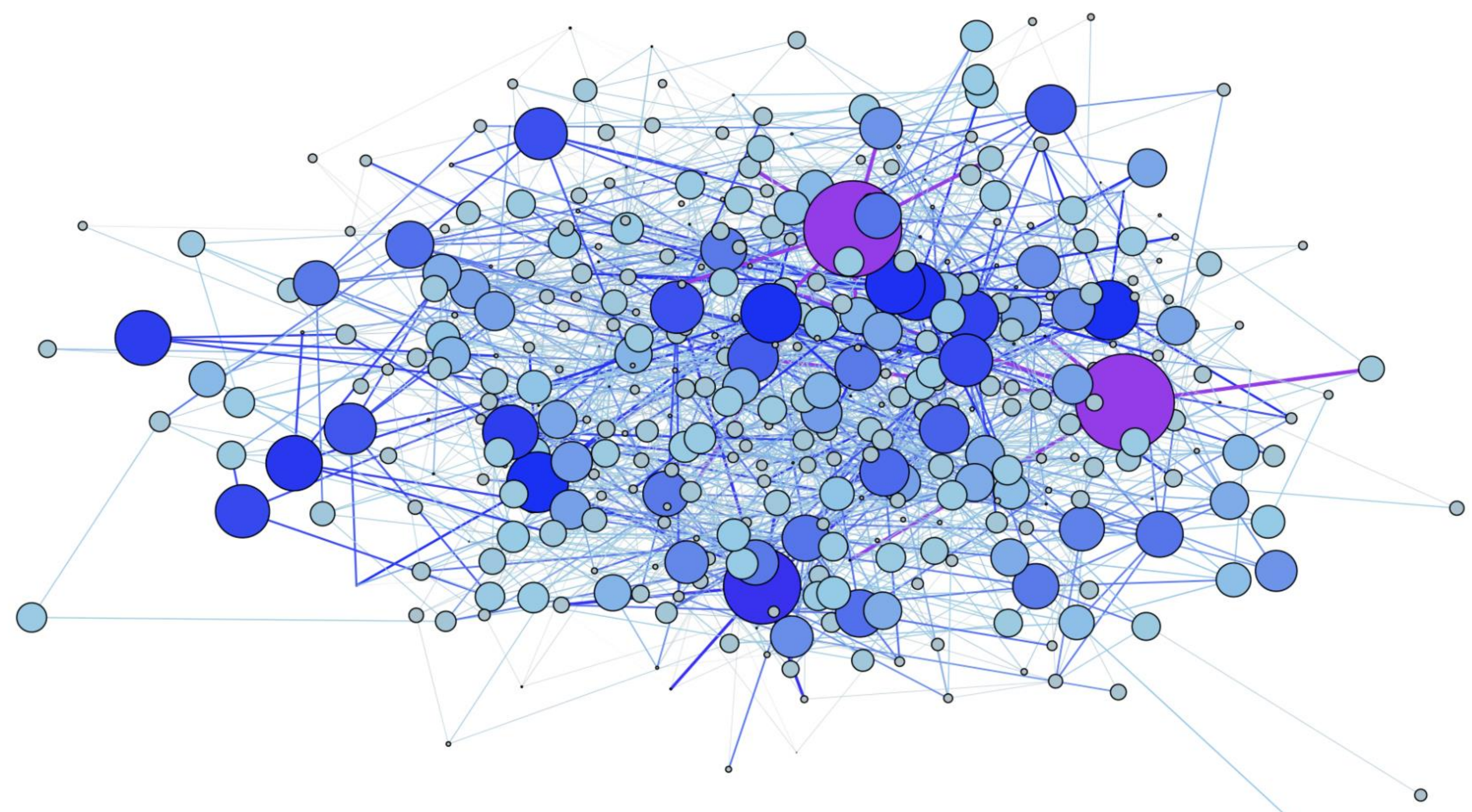
Noam Tetelman, under the supervision of Prof. Daniel Nevo, TAU

Background

When analyzing a social network, how can interventions aimed at specific groups within the network impact the network as a whole? For example, how the fact that 30% of the network is vaccinated affects the unvaccinated individuals? In causal inference, this phenomenon is referred to as **network interference.** This work aims to study those indirect effects by identifying the **"most influential"** individuals, focusing on causal methods often overlooked in general network analyses.

Current Work – Influence isn't solely determined by network degree

We follow the causal definition of influentials as **individuals whose treatment has the most substantial effect throughout the entire network** [1,2]. Under specific model assumptions and within a random effects framework, we demonstrate the ability to estimate the most influential individuals as a **personal characteristic**, either independently or in combination with their network degree.



A simulated social network in which larger vertices represent individuals with greater social influence as a trait (larger b_i).

Notations and Assumptions

- $Y_i(\mathbf{z}),\ i = 1, \dots, N$ - the potential outcome of subject i, had the network treatment vector was: $\mathbf{z} = (\mathbf{z}_1, .., \mathbf{z}_N)$
- $(Y_i, Z_i, \mathbf{X}_i)$ - the observed outcome, treatment and covariates values of that subject i (respectively).
- $\mathbf{A}$ - the adjacency matrix (known)
- $\boldsymbol{\epsilon} \sim N\left(\vec{0}_N, \Sigma(\vec{\sigma_\epsilon})\right)$, dependent RVs, $\vec{\sigma_\epsilon}$ are the variance parameters.
- $\mathbf{b} \sim N(\vec{0}_N, \sigma_b^2 I_{NXN})\ iid$ - b_i is the random effect associated with subject i and will be used to assess its influence.
- $\forall i, j$: ϵ_i and b_j are independent.
- We also assume *conditional exchangeability:* $\forall \mathbf{z}$: $\mathbf{Y}(\mathbf{z}) \perp \mathbf{Z} \mid \mathbf{X}$ and *consistency:* $\mathbf{Y}(\mathbf{Z}) = \mathbf{Y}$.

The Potential Outcome Mixed Model

Under the assumptions, we adapt the model proposed by Lee et al. (2023):

$$Y_i(\mathbf{z}) = \delta^T \mathbf{X}_i + \alpha z_i + \sum_{j=1, j\neq i}^{N} \left(b_j + \gamma\right) A_{ij} z_j + \epsilon_i$$

Influence Measures

- b_i - the influence of subject i on its neighbors as a personal characteristic.
- $degree_i * (b_i + \gamma)$- how treating vs. not treating subject i, influences the entire network outcomes.

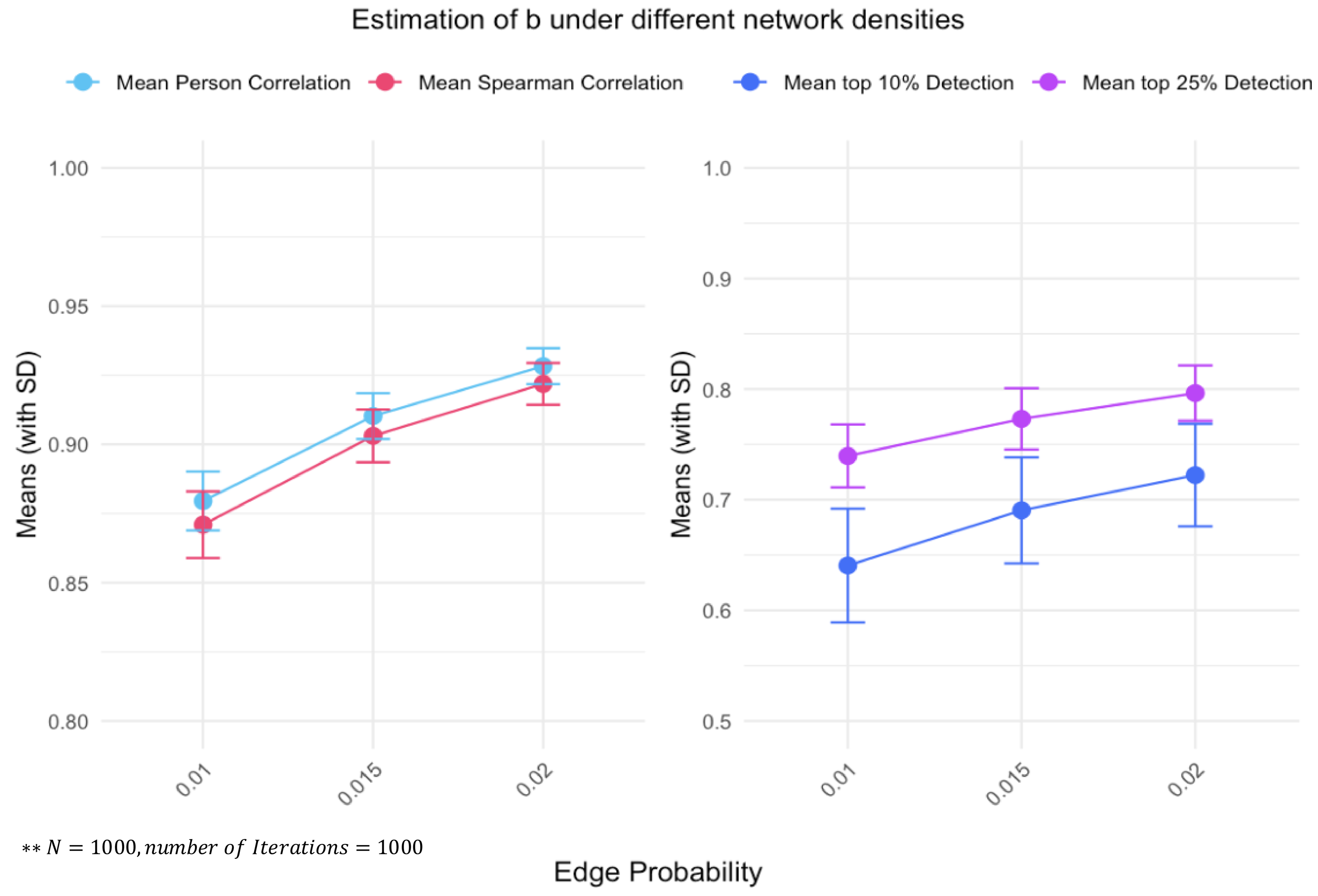
Estimation Method

- The model parameters: $\vec{\sigma_\epsilon}$, σ_b^2, δ^T, α, γ are estimated using maximum likelihood estimation and restricted maximum likelihood (REML) estimation procedures.
- $\mathbf{b}$ is estimated by $\widehat{\mathrm{E}(\mathbf{b} | \mathbf{Y}, \mathbf{Z}, \mathbf{A}, \mathbf{X})}$ where the parameters are replaced with their estimators.
- The total influence of subject i is estimated by: $\mathrm{degree_i}(\widehat{b_i} + \widehat{\gamma})$

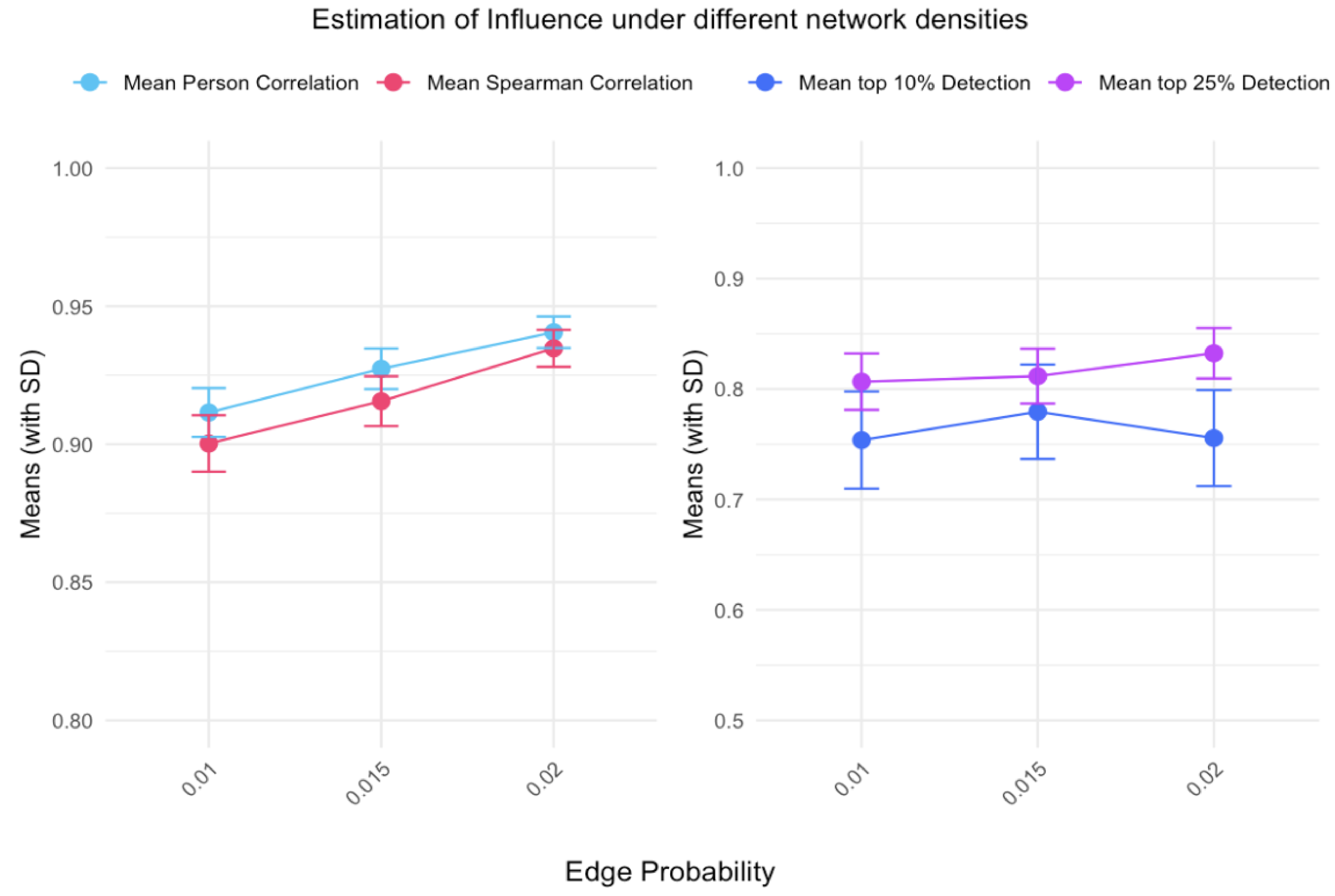
Simulations Results

The correlation of both estimators ($\mathbf{b}$ and total influence of treated individuals) with the real values, achieved an average > 0.85, across all three networks. Additionally, we successfully identified, on average, over 80% of the top 25% treated influencers (and over 75% of the top 10%).

Estimation of b under different network densities



Estimation of Influence under different network densities



Future work

Utilize this approach within generalized linear mixed models (GLMM) and assess its performance through a real-world data application.

References

[1] Smith, S.T., Kao, E.K., Shah, D.C., Simek, O. & Rubin, D.B. (2018) Influence estimation on social media networks using causal inference. In 2018 IEEE Statistical Signal Processing Workshop (SSP). IEEE. pp. 328–332). IEEE.

[2] Lee, Y., Buchanan, A. L., Ogburn, E. L., Friedman, S. R., Halloran, M. E., Katenka, N. V., Wu, J., & Nikolopoulos, G. K. (2023). Finding influential subjects in a network using a causal framework. Biometrics, 79(4), 3715–3727 .