Data-Driven Learning of Contact Networks for Targeted Vaccination in Outbreaks

Sepehr Elahi **Paula Mürmann** Patrick Thiran

EPFL, Switzerland

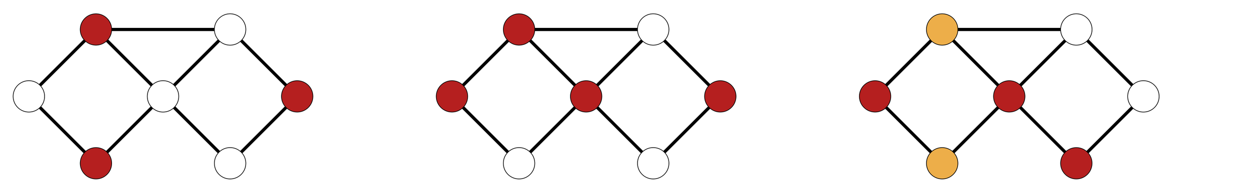
Motivation

Mathematically rigorous approach to contact network recovery and vaccination in a reinfecting epidemic model, posing two **challenges**:

- Unknown Network: The underlying contact graph showing *who can infect whom* is unknown
- Limited Resources: We can only vaccinate a small fraction of the population, so vaccinations must be precise and impactful

The Propagation Model: Graphical SIS

- Individuals represented as vertices in a graph $\mathcal{G}$ with states Susceptible or Infected
- Edges represent all *potential* infection pathways
- Recovery (I $\to$ S) with probability p_{rec}
- Infection (S $\to$ I) with probability $\propto p_{\text{inf}}$ if neighbors are infected
- Vaccinated vertices: lower probability of infection



○ Susceptible
● Infected
● Vaccinated

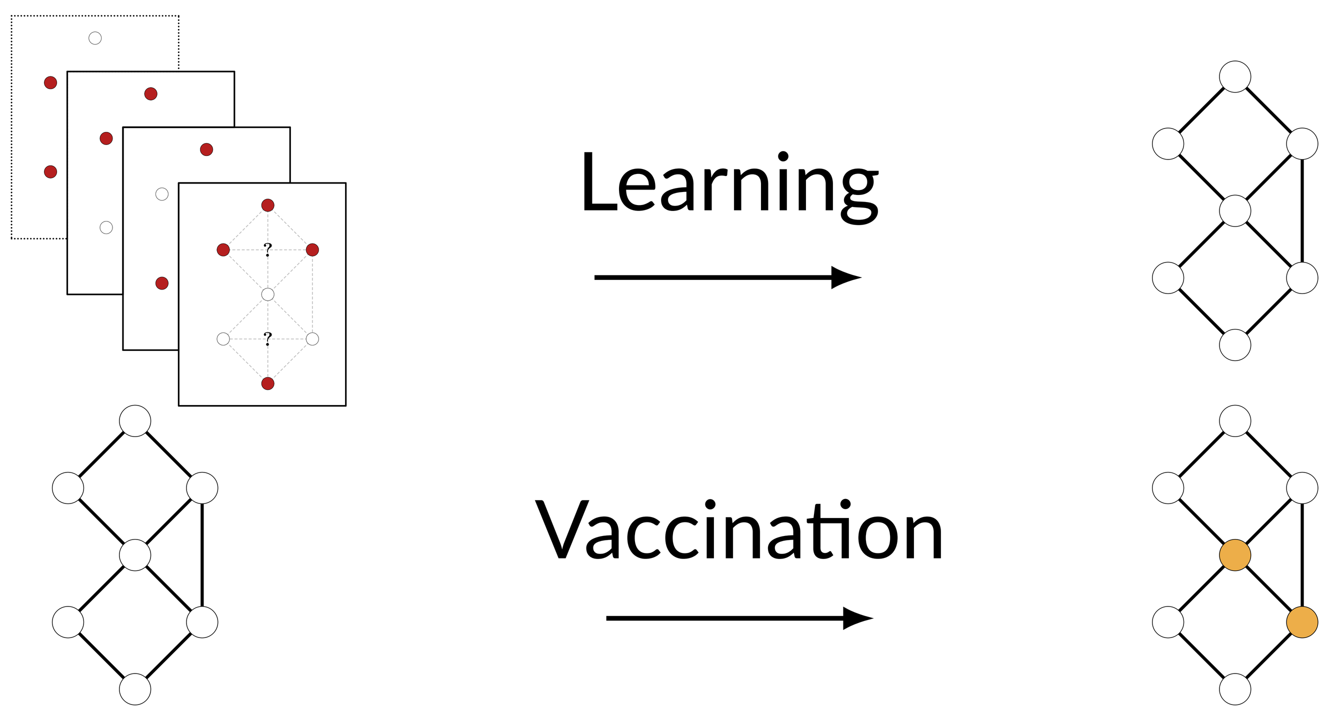
Vaccinating an Unknown Graph (VUG) Problem

Goal: **Vaccinate** K vertices to minimize the expected extinction time (i.e., the average time until all individuals have returned to state **S**)

Challenge: Infection pathways (edges) are **unknown**

Our Approach to Solving the VUG Problem

1. **Learn** the underlying graph from observations: $\hat{\mathcal{G}}$
2. Compute the K vertices to **vaccinate** using learned graph $\hat{\mathcal{G}}$



Learning the Graph: SISLearn

Observations:

- Infection states (**S** or **I**) of vertices over rounds $t \in \{1, 2, \dots, T\}$
- Edges are <u>not</u> observed

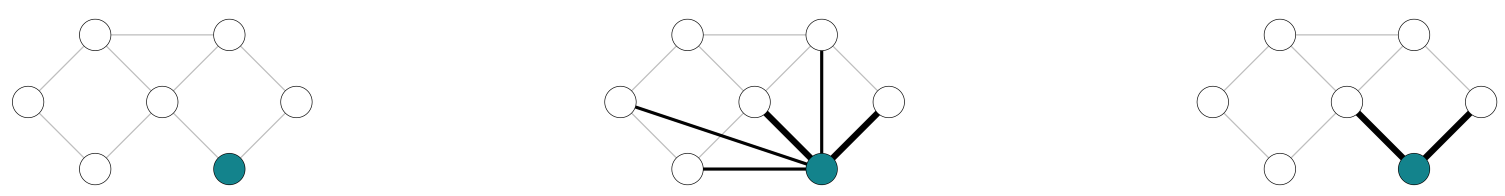
Idea: Learn **neighbors** of vertex v using a correlation indicator

Correlation Indicator: Does a vertex u **influence** the state of v?

$$\mathbb{P}(v \text{ gets infected } \mid u \text{ is infected})$$

Learning algorithm: Vertex-wise **inclusion/exclusion mechanism**

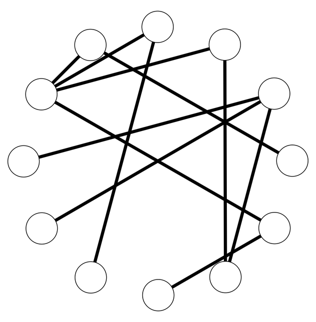
1. Build super-neighborhood from vertices with high influence on v
2. Condition on the super-neighborhood and remove all vertices that do not influence v



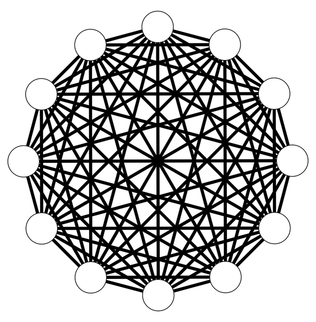
Vaccination Strategies

Observation: Minimizing extinction time $\approxeq$ minimizing **spectral radius** of the graph $\rho(\mathcal{G}) := \max\{\lambda \mid \lambda \in \text{eigval}(\mathcal{G})\}$ [1]

Spectral Radius: one-number measure of a network's connectivity (higher $\implies$ outbreak spreads easier and faster)



Sparse Graph
- low spectral radius
- fast extinction



Dense Graph
- high spectral radius
- slow extinction

Spectral Radius Minimization Problem (SRM)

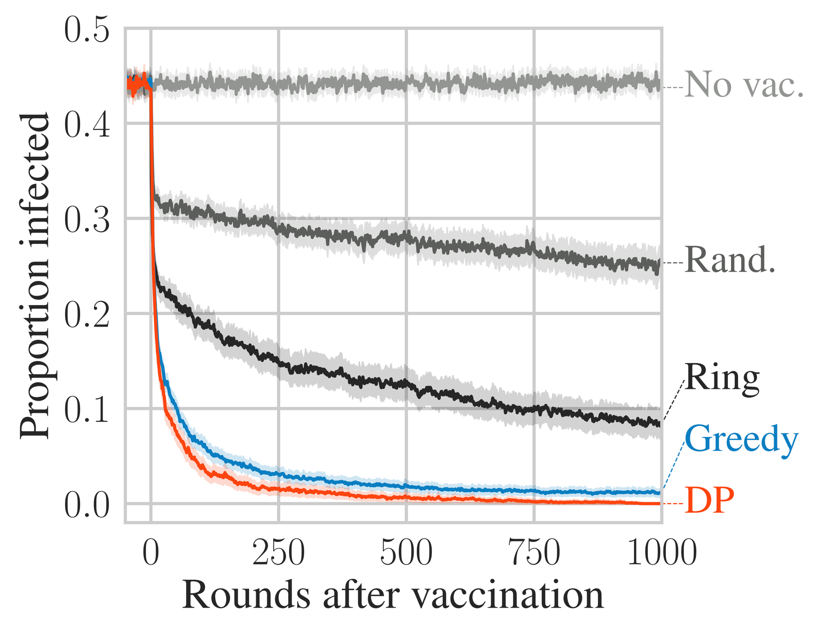
Idea: We pick K people whose vaccination optimally reduces the network's connectivity: $R^* = \text{argmin}_{R \subseteq V, |R| \leq K}\, \rho(\mathcal{G}[V \setminus R])$

Challenge: SRM is NP-hard on general graphs (naïvely takes <u>exponential time</u> to compute)

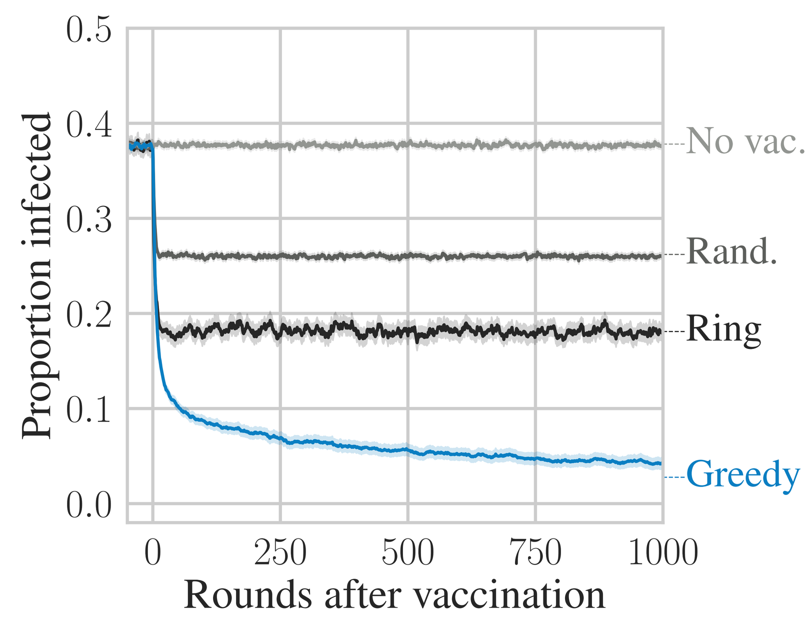
Our Two Approaches:

- Dynamic Programming (DP): Exact but slower approach (best for networks with ≤ 60 vertices)
- Greedy: Approximate but fast approach (works on networks with as many as 10'000s vertices)

Performance on Flu Outbreak Networks

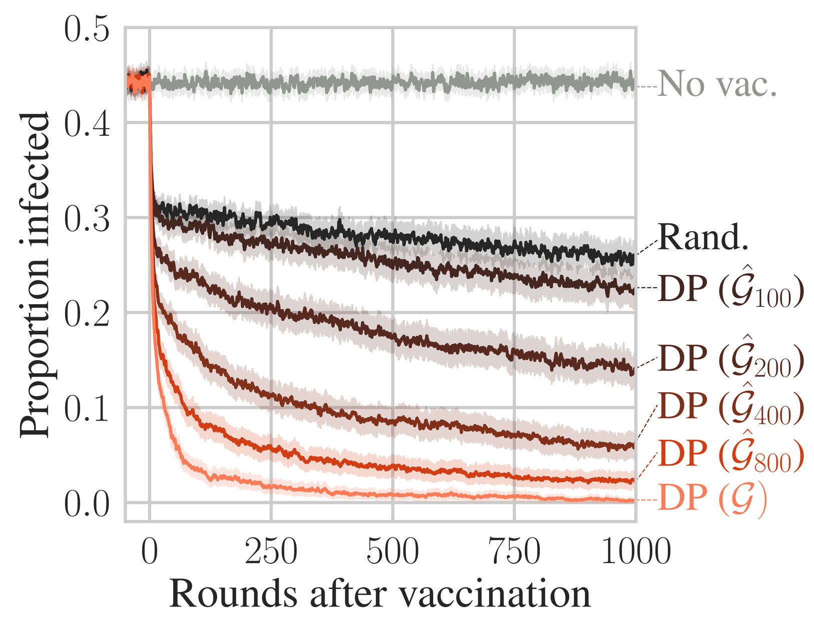


DP and Greedy vs. baselines on augmented networks from the 2009 H1N1/H3N2 outbreak in Beijing (40 vertices and 80 edges) [2], learned using `SISLearn` (↓ is better)

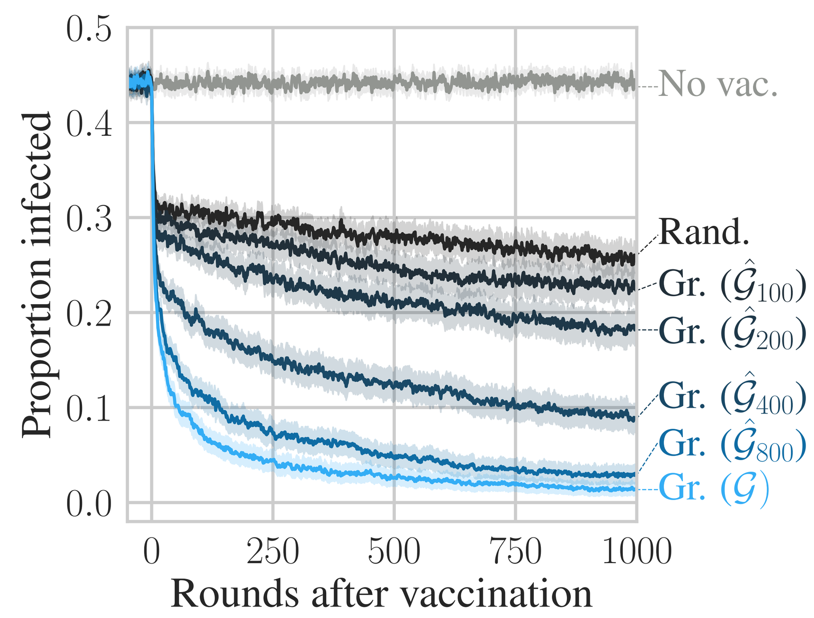


Greedy vs. baselines on augmented networks from the 2009 H1N1 outbreak in Pennsylvania (286 vertices and 818 edges) [2], learned using `SISLearn` (↓ is better)

Performance with Limited Data



DP on the learned graph ($\hat{\mathcal{G}}_{T'}$) from `SISLearn` using different numbers of rounds (T') for learning augmented 2009 Beijing H1N1/H3N2 networks (↓ is better)



Greedy on the learned graph ($\hat{\mathcal{G}}_{T'}$) from `SISLearn` using different numbers of rounds (T') for learning augmented 2009 Beijing H1N1/H3N2 networks (↓ is better)

References

[1] P. Van Mieghem, D. Stevanović, F. Kuipers, C. Li, R. van de Bovenkamp, D. Liu, and H. Wang, "Decreasing the spectral radius of a graph by link removals," *Phys. Rev. E*, 2011.

[2] J. C. Taube, P. B. Miller, and J. M. Drake, "An open-access database of infectious disease transmission trees to explore superspreader epidemiology," *PLOS Biology*, 2022.