REPEATED ACTIVE SCREENING OF NETWORKS FOR DISEASES

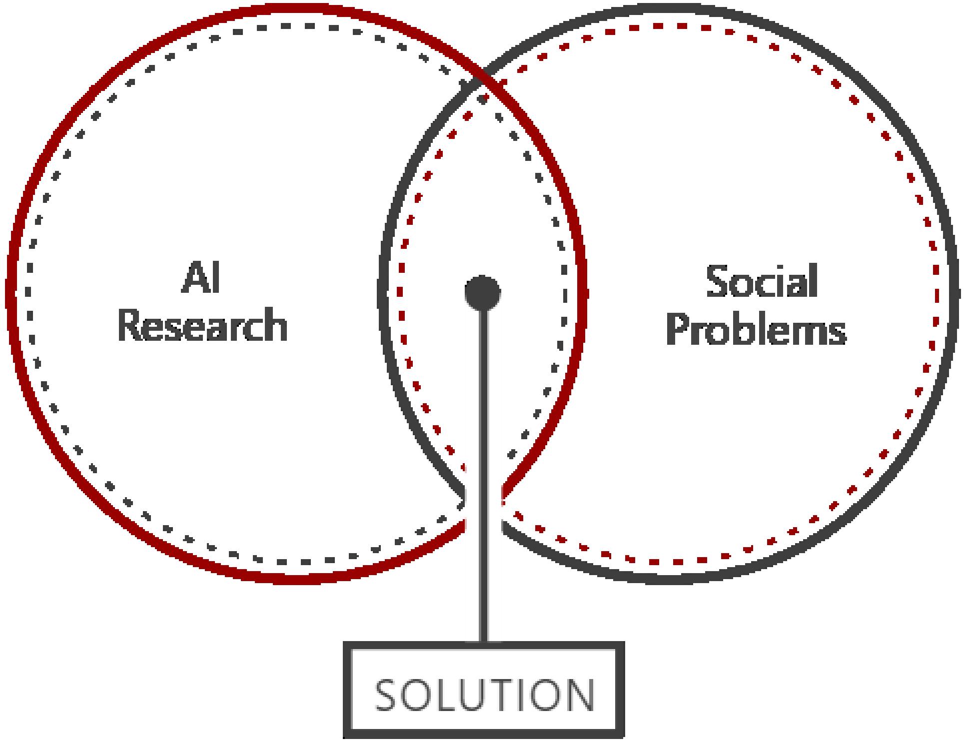
Biswarup Bhattacharya, Han Ching Ou, Arunesh Sinha *, Sze-Chuan Suen, Bistra Dilkina & Milind Tambe

University of Southern California, * University of Michigan

epiDAMIK Workshop @ KDD 2018

AI FOR SOCIAL GOOD

USC University of Southern California



INTRODUCTION

- Number of health problems – AI can be utilized
- HIV, infectious diseases, nutrition, among others

- Curable infectious diseases
  - Tuberculosis - 10M+ people affected in 2016 [WHO]
- Minimizing the number of infected individuals

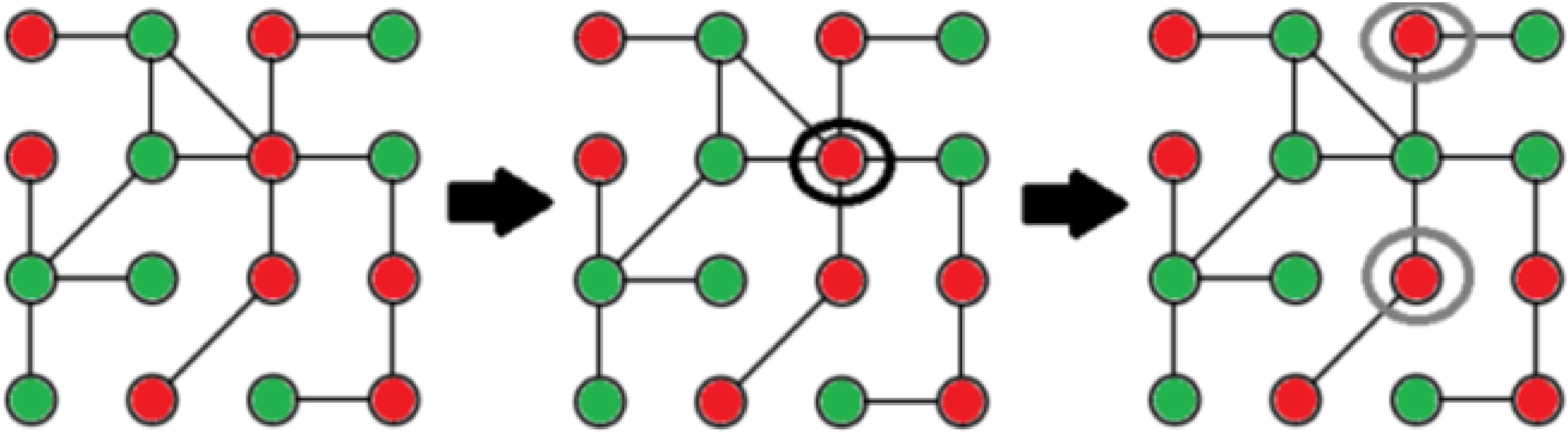
INTRODUCTION – ACTIVE SCREENING

- Individuals may not be able to seek treatment themselves
  - Distance from clinic, failure to self diagnose etc.
- Often – a matter of outreach and identification
  - Resource constraints – e.g. 1 health worker / 500 people (India)

- Problem of **Active Screening**
  - *Definition*: Individuals **sought out** by health workers and treated

- Passive Screening: Individuals seek treatment voluntarily

INTRODUCTION – ACTIVE SCREENING



- Which nodes to act on first?
- Which nodes to act on next?

PROBLEM STATEMENT

ACTS Problem

Given –

- A known network of individuals (*n*)
- Infectious disease parameters
- Limited resources (*k*)

Find – An active screening policy

To maximize – Number of healthy individuals over time

STATUS QUO

Previous works generally do not consider:

- Multiple timesteps
- Uncertainty in health states
- Latent stages
- Lack of permanent immunity

As discussed: Hard to predict infected nodes

In the field: Heuristics used – degree, high-risk societies

CONTRIBUTIONS

1. ACTS Active Screening Model
   - POMDP-like model
2. TRACE Algorithm for ACTS
   - **Synergy of 3 Key Ideas**: Greedy, eigenvalue & community
   - Practically significant results of increase in healthy population

OVERVIEW

1. Problem Modeling
   - SEIS Disease Model
   - Active Screening Model
2. TRACE Algorithm
   - Belief States & Attractiveness Score (FIRST KEY IDEA)
   - Dynamic Eigenvalue (SECOND KEY IDEA)
   - Community Formation (THIRD KEY IDEA)
3. Experiments

OVERVIEW

1. Problem Modeling
   - SEIS Disease Model
   - Active Screening Model
2. TRACE Algorithm
3. Experiments

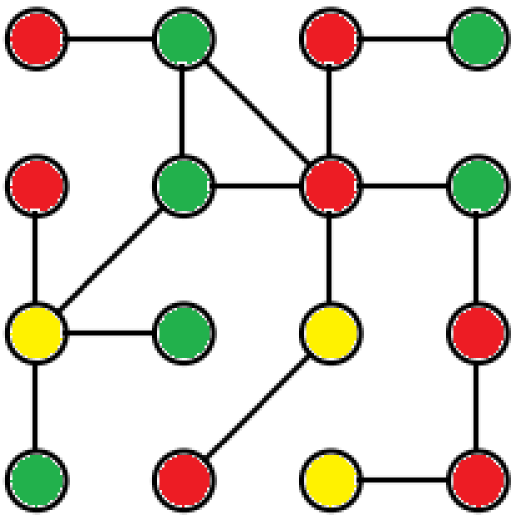
SEIS COMPARTMENTAL DISEASE MODEL

$$\text{Susceptible } (S) \xrightarrow{\alpha} \text{Exposed } (E)$$
$$\text{Exposed } (E) \xrightarrow{\beta} \text{Infected } (I)$$
$$\text{Infected } (I) \xrightarrow{c} \text{Susceptible } (S)$$

- 3 states – S (healthy), E (exposed, cannot infect others), I (infected)
- Note: *c* is the probability individuals voluntarily screen themselves
- Latent stage (E) + Lack of permanent immunity

ACTIVE SCREENING MODEL – [S], [A], T, O, Z, R

- n individuals ⇔ n nodes
- Each node's state – S, E or I
- States not readily known by us (agents)
- Action: Screen (1) or not screen (0)
  - k ($< n$) individuals to be screened at each stage



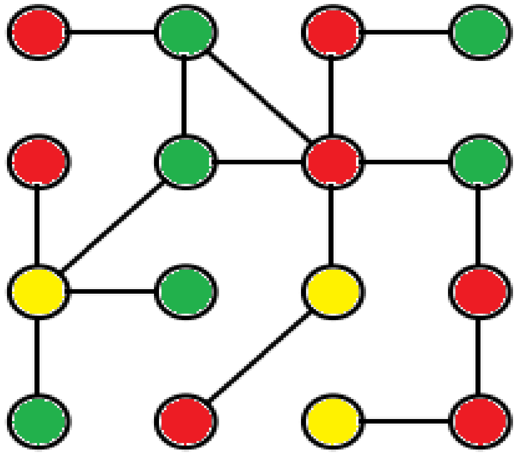
ACTIVE SCREENING MODEL – S, A, [T], O, Z, R

- Cyclic and unidirectional: S → E → I

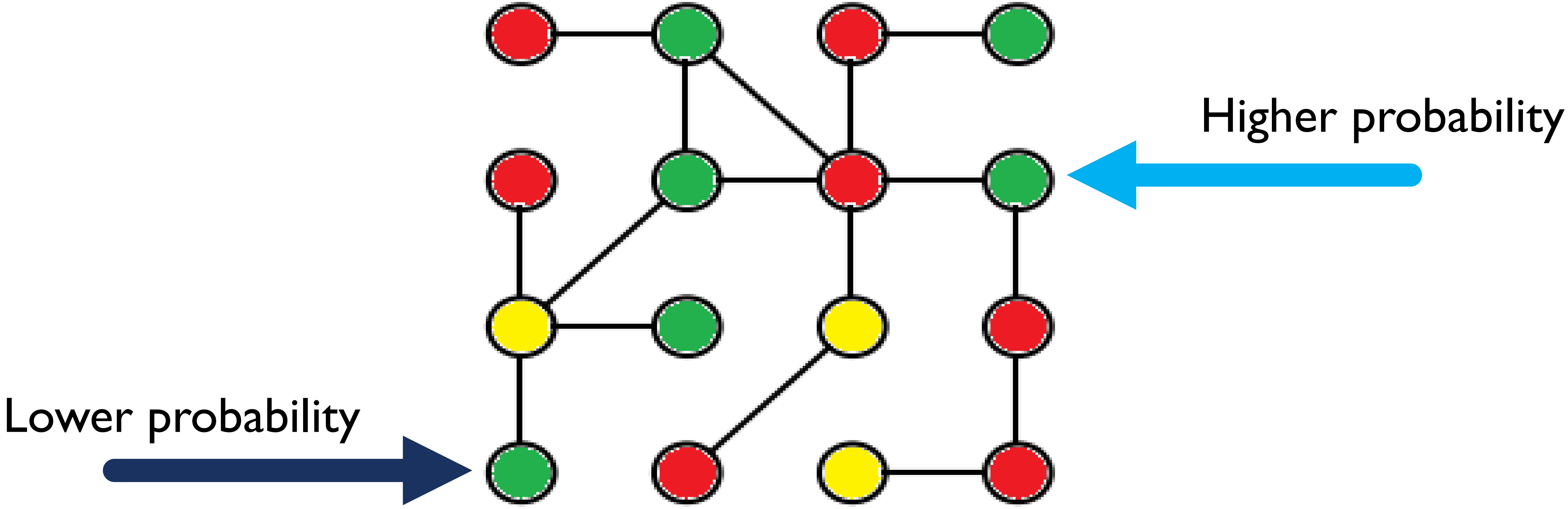
$$\text{Susceptible } (S) \xrightarrow{\alpha} \text{Exposed } (E)$$
$$\text{Exposed } (E) \xrightarrow{\beta} \text{Infected } (I)$$
$$\text{Infected } (I) \xrightarrow{c} \text{Susceptible } (S)$$

- From Row state To Column state
- *q(j)*: Number of infected neighbors of node *j*

$$T^0 = \begin{array}{c c} & \begin{array}{ccc} S & E & I \end{array} \\ \begin{array}{c} S \\ E \\ I \end{array} & \begin{bmatrix} q_j & 1-q_j & 0 \\ 0 & 1-\beta & \beta \\ c & 0 & 1-c \end{bmatrix} \end{array}, \quad T^1 = \begin{array}{c c} & \begin{array}{ccc} S & E & I \end{array} \\ \begin{array}{c} S \\ E \\ I \end{array} & \begin{bmatrix} q_j & 1-q_j & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

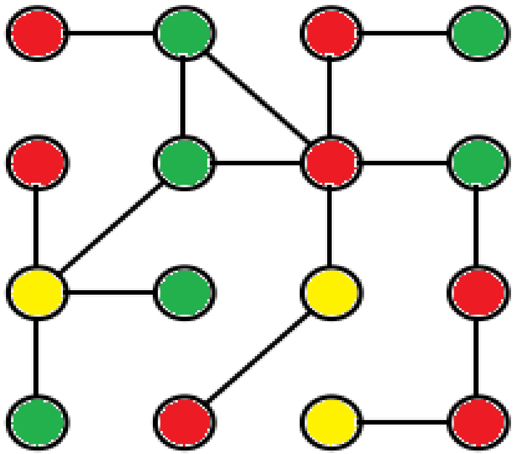


ACTIVE SCREENING MODEL – S, A, [T], O, Z, R



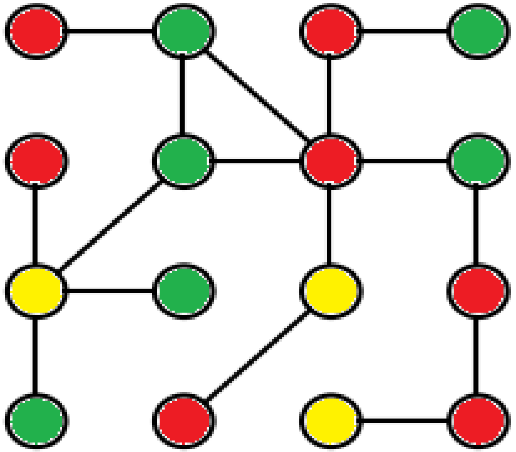
ACTIVE SCREENING MODEL – S, A, T, [O], [Z], R

- Actual health state observed on screening (a=1)
- Else, no observation



ACTIVE SCREENING MODEL – S, A, T, O, Z, [R]

- +1 for every healthy (S) individual
- In shown network, R = +6



- Objective: Maximize increase in number of disease-free half-years over no intervention

$$\sum_{t=0}^{t=T} |S|_t$$

WHY NOT POMDP? – SCALABILITY

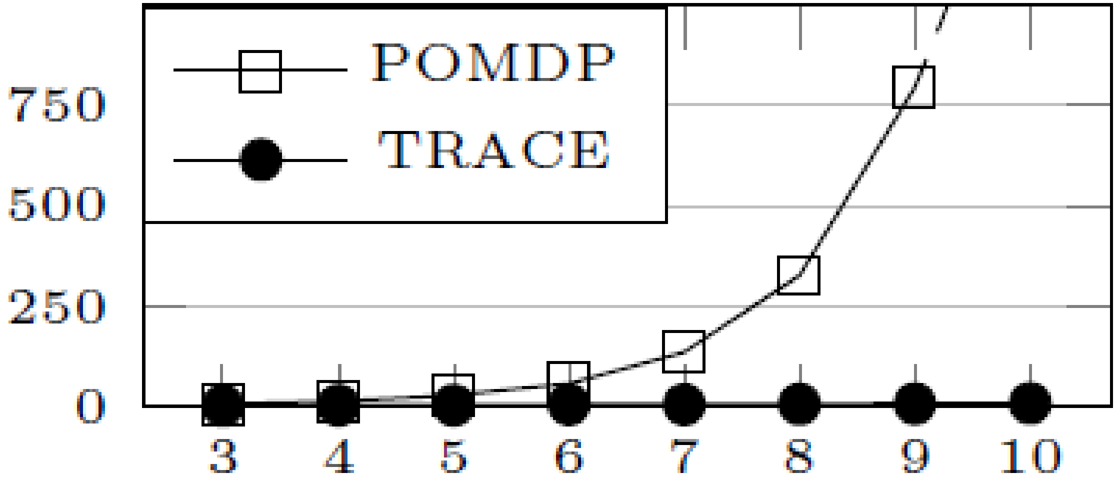


Figure: Runtime (s) v/s n

OVERVIEW

1. Problem Modeling
2. TRACE Algorithm
   - Belief States & Attractiveness Score (FIRST KEY IDEA)
   - Dynamic Eigenvalue (SECOND KEY IDEA)
   - Community Formation (THIRD KEY IDEA)
3. Experiments

TRACE ALGORITHM

- Generates an online POMDP policy
- Synergy of three approaches:
  1. Community [Hendrickson and Leland, 1995]
  2. Beliefs (Greedy)
  3. Eigenvalues [Prakash et al., 2012]

HOW TO HANDLE UNKNOWN STATES?

- Maintain beliefs

$$b_i^t = [b_{i,S}^t, b_{i,E}^t, b_{i,I}^t]$$

- Maintaining marginals good enough [Chakrabarti et al., 2008]
- Other representations – prohibitively large
- Belief update rules – similar to T matrices

HOW TO HANDLE UNKNOWN STATES?

- Maintain **beliefs** for EVERY node → O(3 x n) space

$$b_i^t = [b_{i,S}^t, b_{i,E}^t, b_{i,I}^t]$$

- Maintaining marginals good enough [Chakrabarti et al., 2008]
- Other representations – prohibitively large

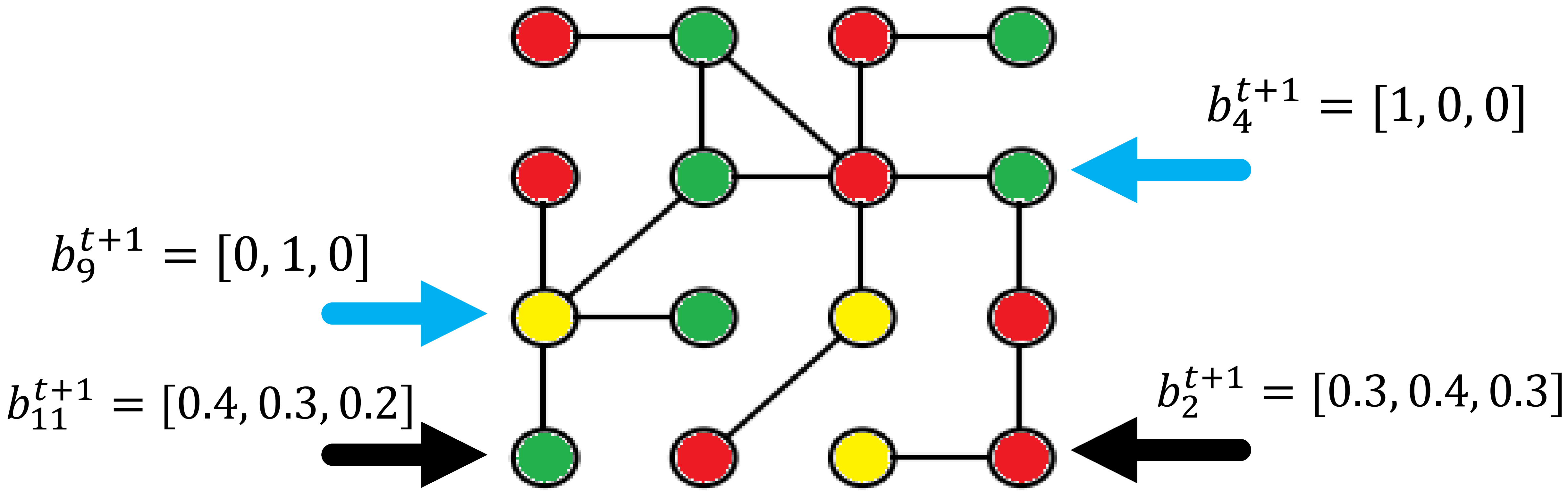
BELIEF STATES UPDATE

$$b_i^t = [b_{i,S}^t, b_{i,E}^t, b_{i,I}^t]$$

- Belief update rules ($b_i^t \rightarrow b_i^{t+1}$) – similar to T matrices
- Start with [0.33, 0.33, 0.33] belief for all
- Belief set to actual state for nodes screened in current timestep
  - E.g. Change to [1,0,0] on screening S node, [0,1,0] if E, [0,0,1] if I
- Update normally if not screened in current timestep

BELIEF STATES – EXAMPLE UPDATE (T=0 → T=1)

If nodes with **light arrows** are screened (initially all beliefs are [0.33,0.33,0.33])



FIRST KEY IDEA

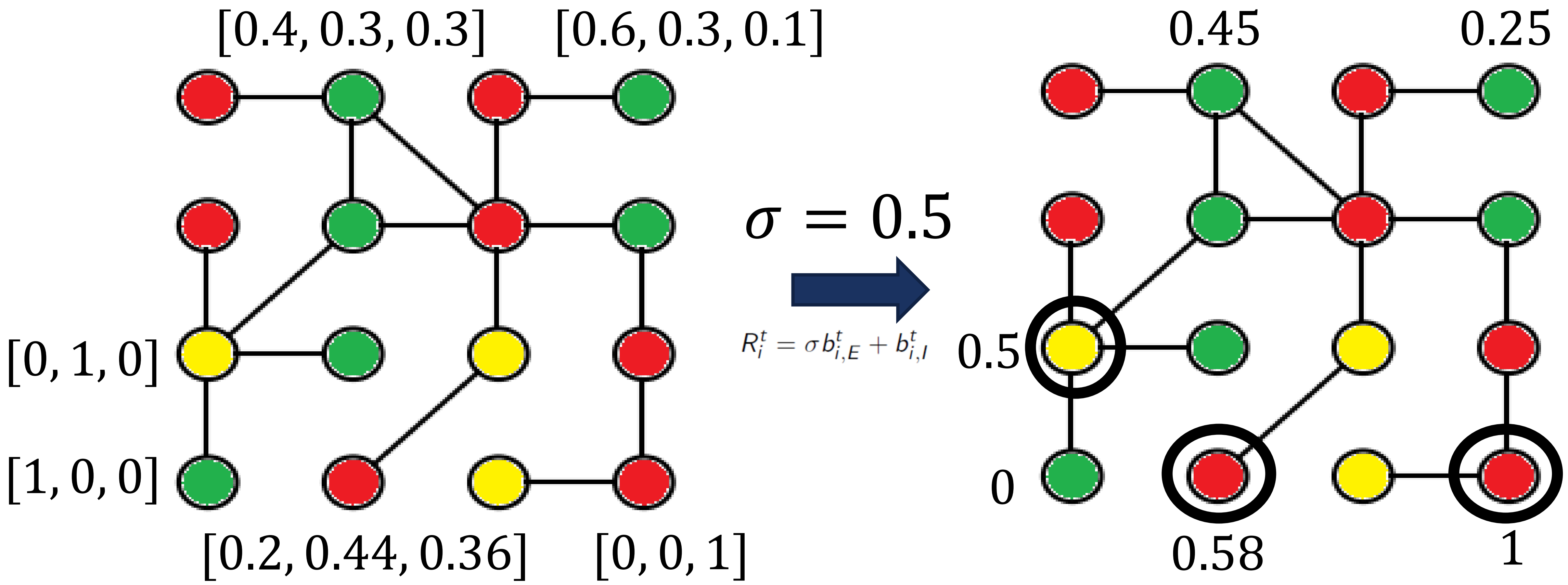
GREEDY

- Attractiveness score for every node based on **beliefs**

$$R_i^t = \sigma b_{i,E}^t + b_{i,I}^t$$

- Simply screen based on higher score
- Possibly not optimal

GREEDY SELECTION



TOWARDS OPTIMALITY

EIGENVALUES

- An epidemic dies out iff

$$\frac{\alpha}{c} < \frac{1}{\lambda_A^*} \text{ and } \beta \neq 0 .$$

[Prakash et al., 2012]

- λ_A^* = largest eigenvalue of the adjacency matrix A of a graph
- High α and/or low c make the limit harder to achieve

SECOND KEY IDEA – DYNAMIC EIGEN

$$\frac{\alpha}{c} < \frac{1}{\lambda_A^*} \text{ and } \beta \neq 0 \,.$$

- "Remove" nodes such that λ_A^* decreases → increases $1/\lambda_A^*$
- S nodes cannot infect neighbors → remove S nodes
- Our case: multiply each row by $(1 - b_{i,S}^t)$
- Iteratively remove + check ⟶ slow for large n

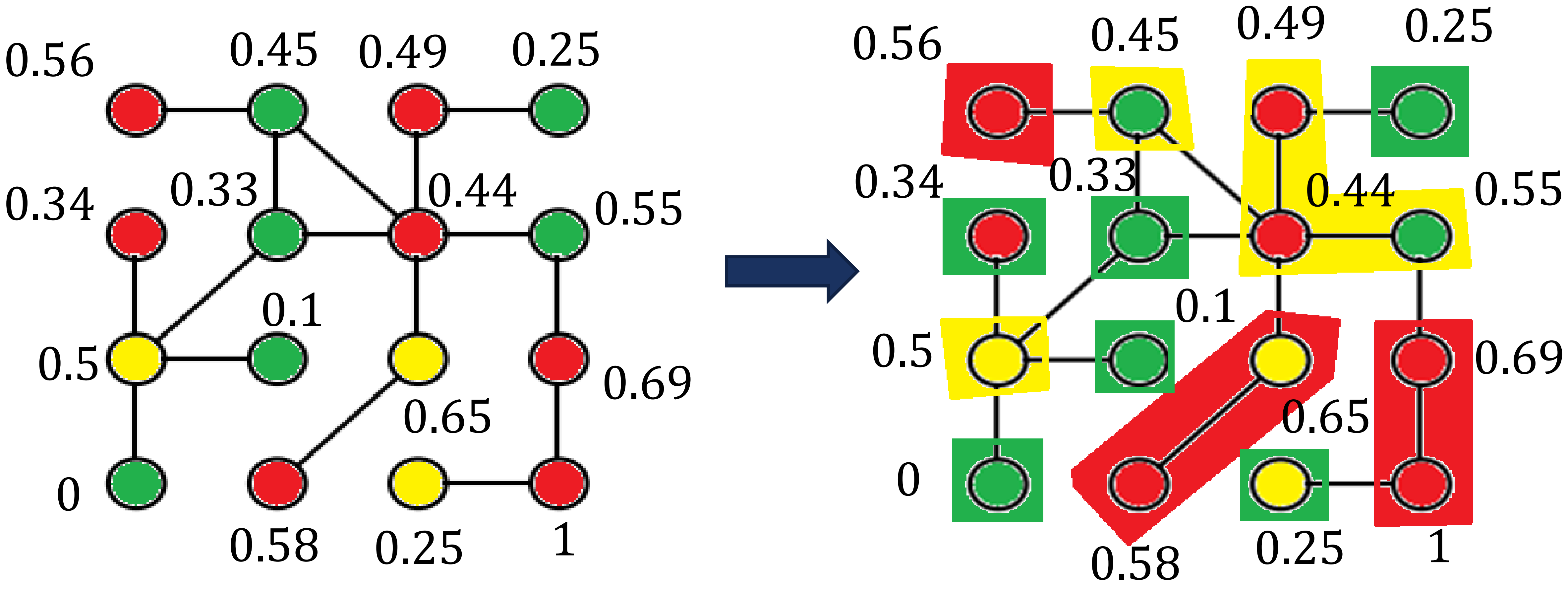
THIRD KEY IDEA – SPEEDING UP DYNAMIC EIGEN

COMMUNITIES

- Group nodes by attractiveness scores
- *Coarsening* [Hendrickson & Leland, 1995]
- Number of groups $\leq n$ $\rightarrow$ DynamicEigen scales up
- Can be proven for scale-free graphs:

number of groups = $O\left(\frac{1.5}{d} n\right)$ (d = average degree)

COMMUNITIES – GRAPH COARSENING



TRACE ALGORITHM – THREE STEPS

1. Greedy approach: Belief information [Unknown health states]
2. Community-based approach: Grouping nodes [Speeds up next step]
3. Eigenvalue-based approach: Reducing spectral radius [Optimality]

None superior by itself! (7 observations in extended version)

OVERVIEW

1. Problem Modeling
2. TRACE Algorithm
3. Experiments

NETWORKS

- 10 real-world networks in extended version (n = 75 to 16730)

- 3 in epiDAMIK submission (n = 202 to 1899)
  1. India (n = 202, [Banerjee et al., 2013]): Collected from a rural Indian village
  2. Infectious Exhibition (n = 410, [Isella et al., 2011])
  3. Irvine (n = 1899, [Opsahl & Panzarasa, 2009])

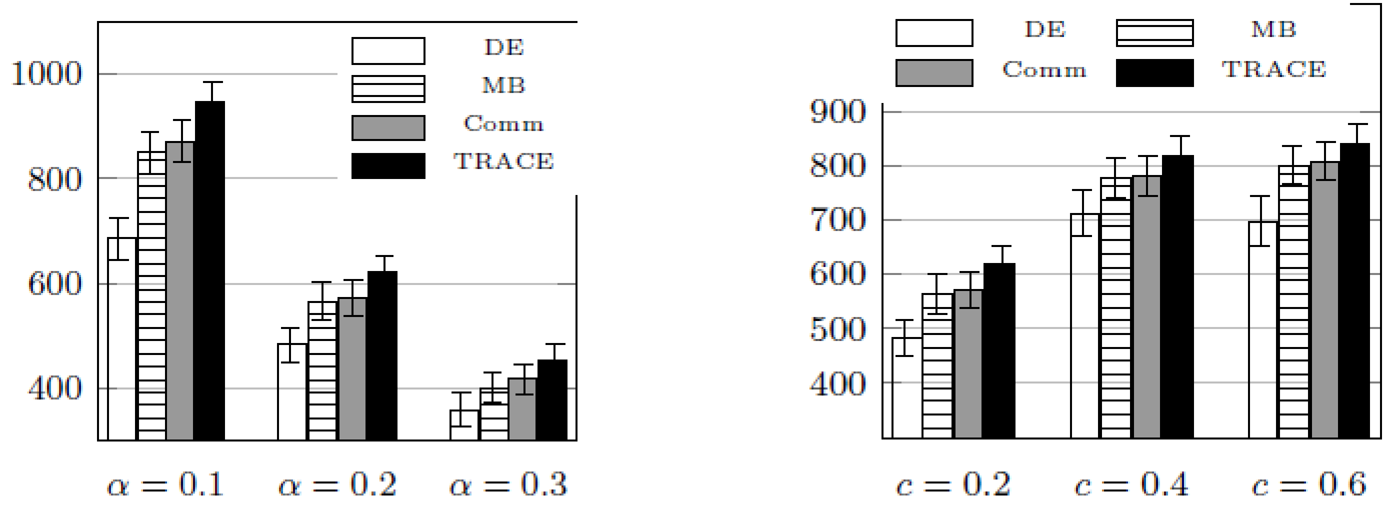
SETUP

- $\alpha = 0.1 - 0.3, \beta = 0.25, c = 0.2 - 0.6$
- Each round = 6 months
- Total simulation = 10 years
- $k = 5\%, \sigma = 0.5$
- Metric: Increase in number of disease-free half-years over no intervention

$$\sum_t |S|_{t,algo} - \sum_t |S|_{t,none}$$

RESULTS – VARYING PARAMETERS

- MB: Greedy
- DE: Just DynamicEigen without community
- Comm: 0-1 knapsack select without eigen

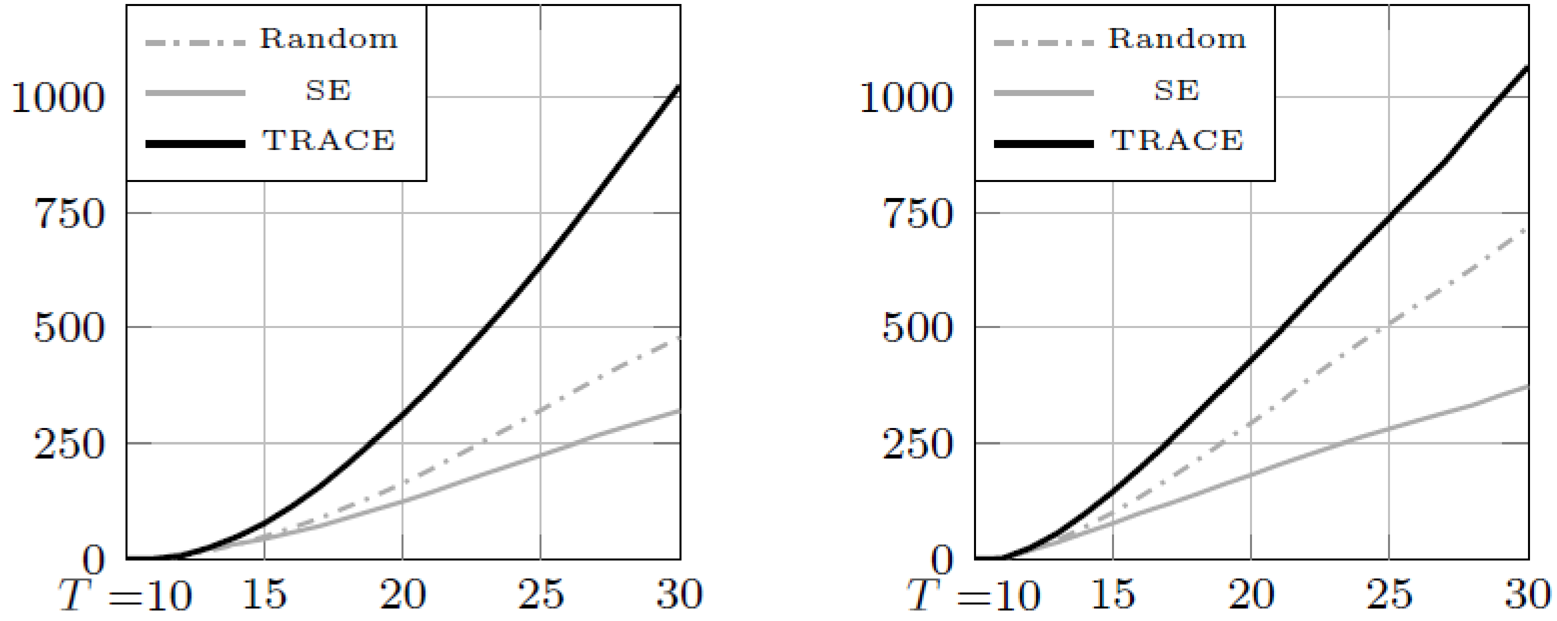


(c) Varying α ($c = 0.2$)

(d) Varying c ($\alpha = 0.2$)

Figure: Performance by TRACE components (India network)

RESULTS – OVER TIME



(a) India network (b) Exhibition network

Increase in $\sum_{t=0}^{t=T} |S|_t$ over None for varying T
$(\alpha = 0.1, \beta = 0.25, c = 0.2)$

KEY TAKEAWAYS

- ✓ **Hard problem** – Multi-round + SEIS + unknown health states
- ✓ **Belief states** to estimate the uncertain health status
- ✓ ↓ **spectral radius** ⇒ ↓ disease prevalence
- ✓ **Three approaches**: Eigenvalue, community, greedy
- ✓ TRACE → **practically significant** results

FUTURE WORK

- Addressing TB in India

- Future: Complex disease models, birth and death, costs, network uncertainty

THANK YOU!

"TRACE: Algorithmic ACTS for Preventing the Spread of Recurrent Infectious Diseases on Networks"

bbhattac@usc.edu

https://biswarupb.github.io/