ShareTrace: Contact Tracing with Asynchronous, Parallel Message Passing on a Temporal Graph

Ryan Tatton Erman Ayday Youngjin Yoo Anisa Halimi

Case Western Restern Reserve University

1. Motivation

- Proximity-based contact tracing relies on user device interaction to estimate infection spread.
- ShareTrace is a privacy-preserving, proximity-based contact-tracing solution that estimates a person's marginal posterior probability of infection (*exposure score*) via iterative message passing on a factor graph based on their prior probability of infection (*symptom score*) and (in)direct contact with others.
- The scalability and efficiency of the original ShareTrace algorithm (*risk propagation*) [2] can be improved with asynchronous, non-iterative message passing on a temporal graph.

2. Proposed Scheme

- Reformulate the factor graph as a temporal contact graph of user nodes (Figure 1).
- Partition the temporal graph, where each subgraph is an *actor* [1] that can pass messages.
- Actors compute and non-iteratively pass messages (*risk scores*) until a stopping condition is met.
- Only propagate messages if they are likely to affect other users' exposure scores in the graph.
- *Send tolerance* γ parametrizes the trade-off between completeness and efficiency of asynchronous, non-iterative message passing. A user node u only propagates a risk score if it is sufficiently high (by a factor of γ) and sufficiently old (relative to the initial message sent by u).

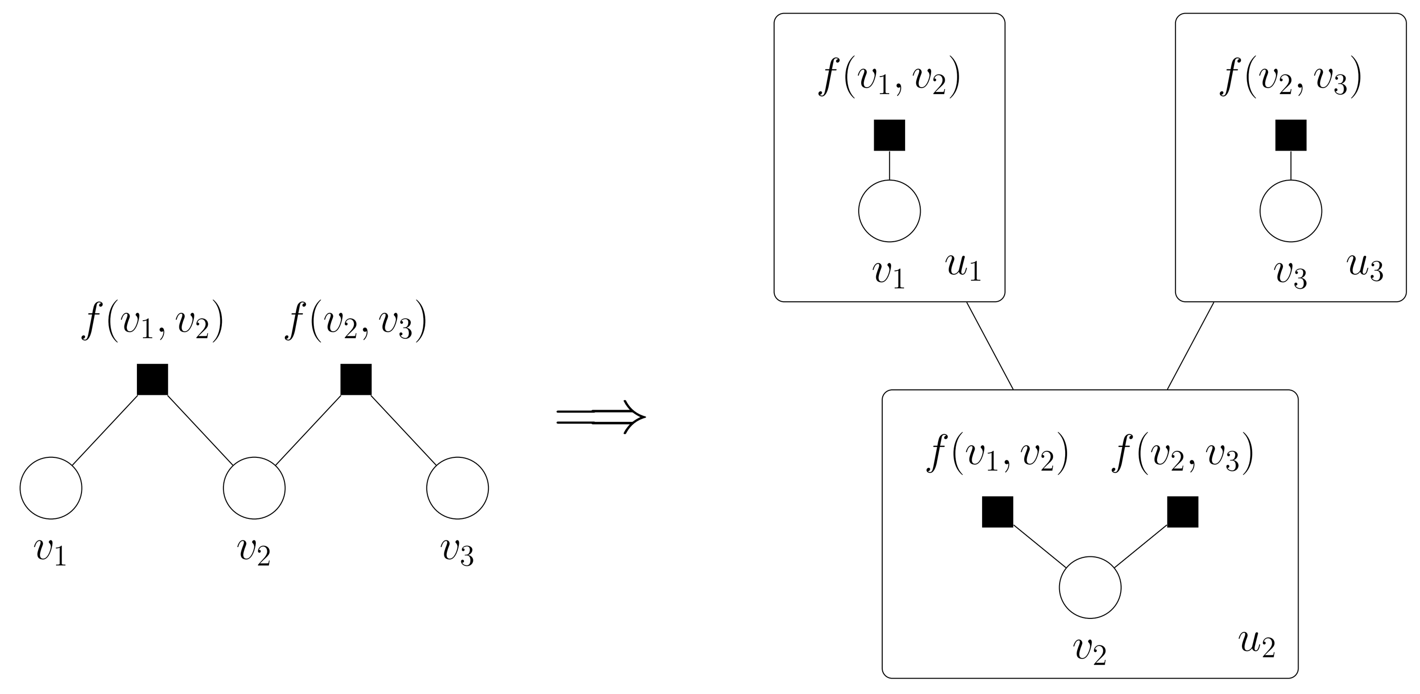


Figure 1. A factor graph of 3 variables and 2 factors (left) as a temporal contact graph of 3 users (right).

3. Message Reachability (MR)

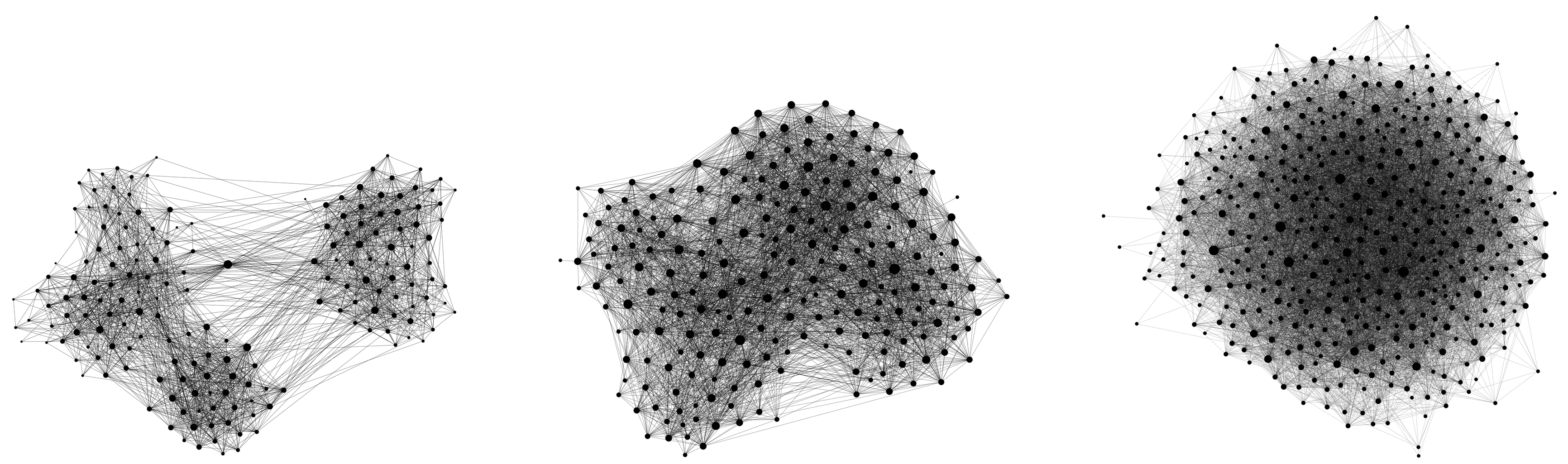
- The *estimated MR* $\hat{m}_\theta$ is a measure to characterize the propagation of risk as a dynamic process on a temporal graph [3], i.e. the number of users a user can impact with their risk.
- Parametrized by θ, which includes the transmission rate α, the send tolerance γ, the initial risk score r_u of user u, and the initial risk score r_v of some reachable user v from u:

$$\alpha^{\hat{m}_\theta(u)} r_u \leq \gamma \alpha r_v \implies \hat{m}_\theta(u) \leq 1 + \log_\alpha \left(\frac{\gamma r_v}{r_u}\right) \quad (1)$$

- The *actual MR* m_θ can be found by applying an augmented shorted-path algorithm with message passing, which accounts for the temporality of contacts.
- A user v is *message-reachable* from user u if there exists a path from u to v such that all users along the path satisfy (1) and the risk score of u is within B days of each contact.

4. Experimental Design

- Synthetic graphs included random geometric graphs (RGG) [4], benchmark graphs (LFRG) [8], and clustered scale-free graphs (CSFG) [6]. See Figure 2 regarding real-world graphs.
- METIS was used to partition the graphs with a load imbalance factor of 0.2 and set to attempt contiguous partitions and minimize the connectivity between partitions [7].
- Users were assigned "high" risk (scores above 0.5) with probability 0.2, and "low" risk otherwise.
- Default parameter values: transmission rate $\alpha = 0.8$; send tolerance $\gamma = 0.6$.
- We ran 10 trials for each synthetic (real-world) graph for MR (resp. scalability) experiments.



(a) High school (Thiers13): 180 / 2,220 (b) Workplace (InVS15): 217 / 4,274 (c) Scientific conf. (SFHH): 403 / 9,565

Figure 2. SocioPatterns contact graphs (number of users / number of contacts) [5]. Each edge represents the most recent time of contact (duration of at least 20 seconds) between two users.

5. Efficiency Results

- We evaluated the effects of transmission rate and send tolerance on risk propagation efficiency using synthetic graphs with 5,000 users and 2 actors.
- A send tolerance of $\gamma = 0.6$ optimizes for completeness and efficiency by permitting 99% of the possible user updates, while resulting in faster runtimes, $(Q_1, Q_2, Q_3) = (0.13, 0.13, 0.46)$, and fewer messages, $(Q_1, Q_2, Q_3) = (0.13, 0.15, 0.44)$, for a transmission rate of $\alpha = 0.8$.

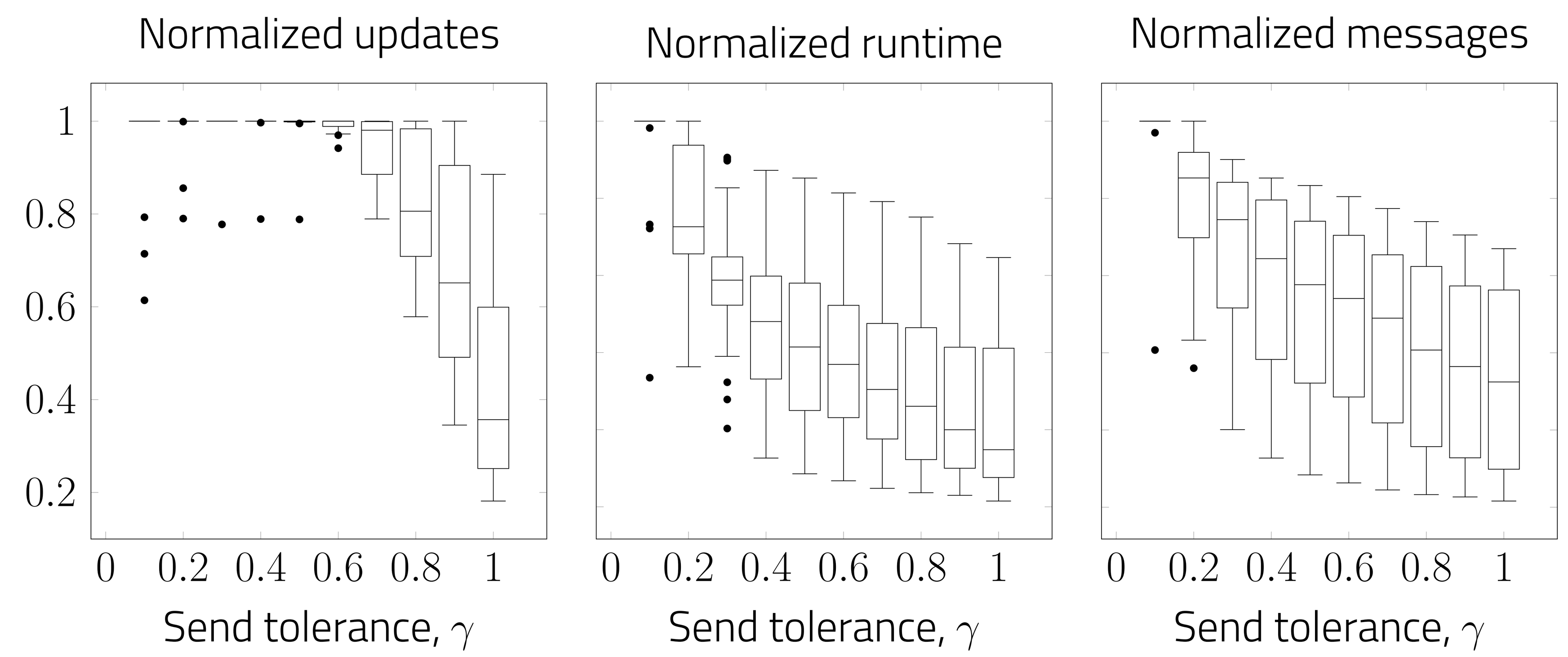


Figure 3. Effects of send tolerance and transmission rate on efficiency. All dependent variables are normalized across graphs and transmission rates. Efficiency on LFRGs were less sensitive to changes in transmission rate and send tolerance than RGGs and CSFGs, which is the cause for the large interquartile ranges.

6. Message Reachability Results

- Equation (1) was a better estimator on synthetic graphs, compared to real-world graphs (Table 1).
- With lower (higher) send tolerances (resp. transmission rates), (1) suggests higher MR; however, a message is only passed under certain conditions, so (1) tends to overestimate m_θ.

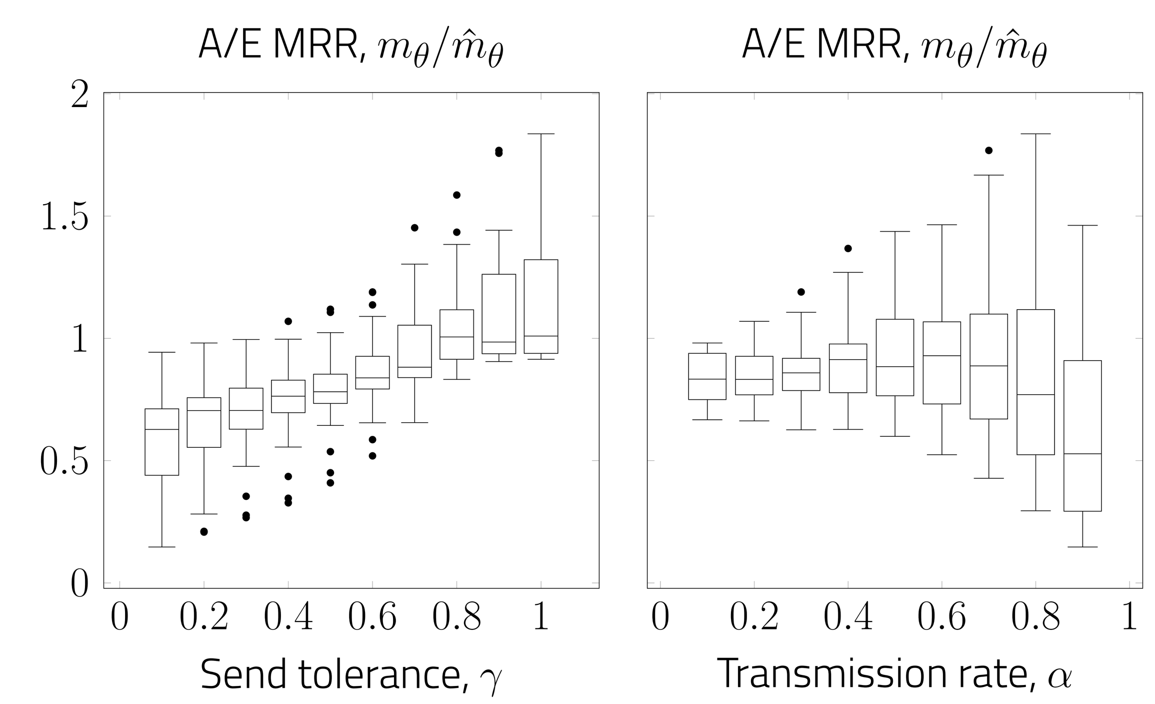


Figure 4. Effects of send tolerance and transmission rate on actual/estimated MR ratio (A/E MRR) on synthetic and real-world graphs.

Graph	$m_\theta/\hat{m}_\theta \pm 1.96 \cdot \mathrm{SE}$
Synthetic	
LFR	0.88 ± 0.14
RGG	0.74 ± 0.12
CSFG	0.90 ± 0.14
	$\mathbf{0.85 \pm 0.08}$
Real-world	
Thiers13	0.58 ± 0.01
InVS15	0.63 ± 0.01
SFHH	0.60 ± 0.01
	$\mathbf{0.60 \pm 0.01}$

Table 1. A/E MRR for synthetic and real graphs ($\alpha = 0.8$, $\gamma = 0.6$). Synthetic (real-world) ratios are averaged across parameter combinations (resp. trials).

7. Scalability Results

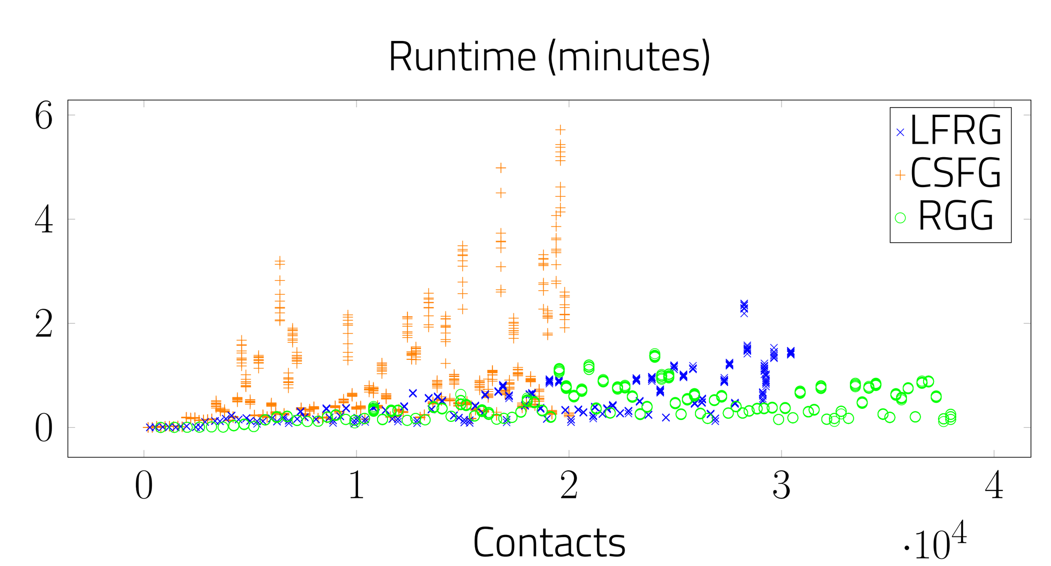


Figure 5. Runtimes of risk propagation on synthetic graphs (100–10,000 users; ~200–38,000 contacts). A linear regression fit explains ($R^2 = 0.52$) the runtime of LFRGs and RGGs with slope $m = (1.1 \pm 0.1) \cdot 10^{-3}$ seconds/contact and intercept $b = 4.3 \pm 1.6$ seconds ($\pm 1.96 \cdot \mathrm{SE}$). The runtime behavior of CSFGs requires further investigation.

Acknowledgements

Research was partly supported by the Cisco Research University Funding grant number 2800379.

References

[1] G. A. Agha. *Actors: A Model of Concurrent Computation in Distributed Systems*. MIT Press, 1986.

[2] E. Ayday, Y. Yoo, and A. Halimi. ShareTrace: An iterative message passing algorithm for efficient and effective disease risk assessment on an interaction graph. In *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2021.

[3] A. Barrat and C. Cattuto. Temporal networks of face-to-face human interactions. In *Temporal Networks*. 2013.

[4] J. Dall and M. Christensen. Random geometric graphs. *Physical Review E*, 66, 2002.

[5] M. G'enois and A. Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7, 2018.

[6] P. Holme and B. J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65, 2002.

[7] G. Karypis and V. Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing*, 20:359–392, 1998.

[8] A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78, 2008.

 ryan.tatton@case.edu