ShareTrace: Contact Tracing with the Actor Model

Ryan Tatton Erman Ayday Youngjin Yoo Anisa Halimi

Case Western Restern Reserve University

1. Motivation

- Proximity-based contact tracing relies on mobile-device interaction to estimate infection spread.
- ShareTrace is a privacy-preserving, proximity-based contact-tracing solution that estimates a user's infection risk (*exposure score*) via 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, synchronous ShareTrace algorithm (*risk propagation*) [2] can be improved with asynchronous message passing on a temporal contact network.

2. Proposed Scheme

- Reformulate the factor graph as a contact network (Figure 1) and partition the network into subnetwork *actors* [1] that pass messages (i.e., risk scores) asynchronously until convergence.
- Only propagate risk scores if they are likely to affect other users' exposure scores in the network.
- The *send coefficient* γ parametrizes the accuracy-efficiency trade-off of asynchronous message passing: an actor only propagates a risk score if it is sufficiently high (by a factor of γ) and sufficiently old (relative to its initial message).
- Guaranteed convergence for $\alpha < 1$ and $\gamma > 0$.

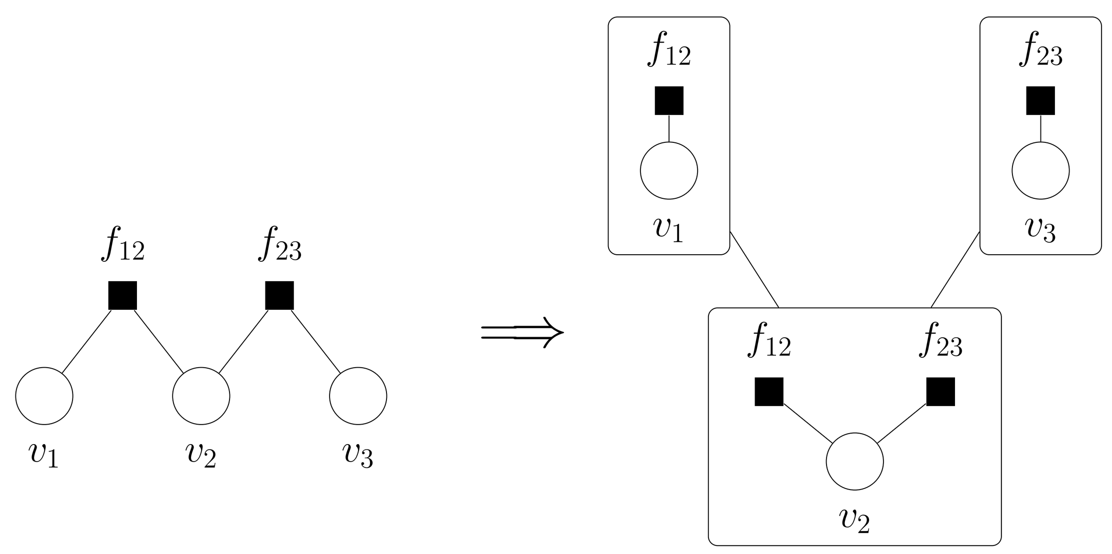


Figure 1. A factor graph of 3 variables and 2 factors (left) as a contact network of 3 users (right).

3. Message Reachability (MR)

- A *time-respecting path* is a sequence of nondecreasing contacts in a contact network; vertex v is *temporally reachable* from vertex u if a path from u to v exists.
- Generalize temporal reachability to account for message-passing semantics: *MR from vertex u to vertex v* along the shortest path P where $f(u,i,j,v) = 1$ if all constraints are satisfied and $f(u,i,j,v) = 0$ otherwise:

$$m(u,v) = \sum_{(i,j)\in P} f(u,i,j,v) \qquad m(u) = \max\{m(u,v) \mid v \in V\}$$

- Let $H(x)$ be the *Heaviside function* and enumerate the path vertices $[0\,..\,|P|-1]$.

$$m(u) = \max_P \left\{ \sum_{(i,j)\in P} H(t_{ij} + \beta - t_u) \cdot H(\alpha^i \cdot s_u - \gamma \cdot s_i) \cdot H(t_i - t_u) \right\}$$

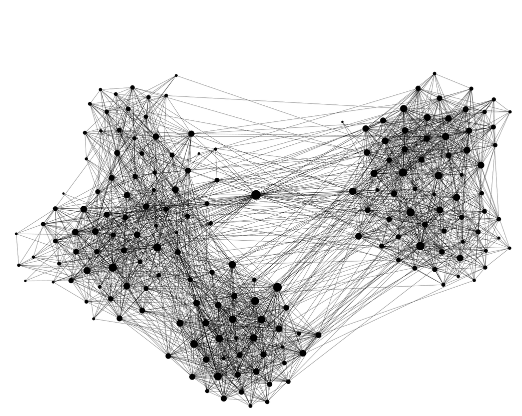
where s_i (t_i) is the initial risk score (time) of user i, t_{ij} is the latest contact time between users i and j; and β is a time buffer to account for reporting delay.
- *Estimated MR of vertex u to vertex v* relaxes the temporality constraints:

$$\hat{m}(u,v) = \log_\alpha \left\{ \gamma \cdot \frac{s_v}{s_u} \right\} \qquad \hat{m}(u) = \max\{\hat{m}(u,v) \mid v \in V\} \tag{1}$$

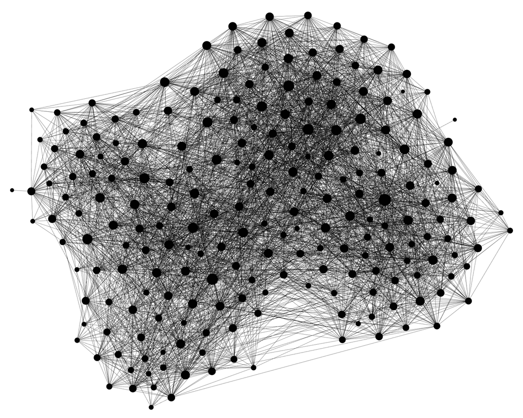
- MR quantifies communication complexity, individual risk, and population risk.

4. Experimental Design

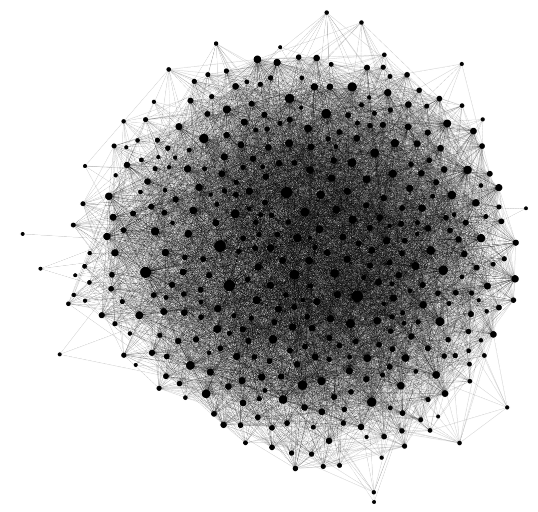
- Synthetic networks included random geometric graphs (RGG) [3], benchmark graphs (LFRG) [7], and clustered scale-free graphs (CSFG) [5]. Figure 2 visualizes real-world networks.
- METIS was used to partition the networks with a load imbalance factor of 0.2, set to attempt contiguous partitions, and minimize inter-partition connectivity [6].
- Users were assigned "high" risk (scores above 0.5) with probability 0.2.
- Default parameter values: transmission rate $\alpha = 0.8$; send coefficient $\gamma = 0.6$.
- 10 trials for each synthetic (real-world) network for MR (scalability) experiments.



(a) High school: 180 / 2,220



(b) Workplace: 217 / 4,274



(c) Scientific conference: 403 / 9,565

Figure 2. SocioPatterns contact networks (number of users / number of contacts) [4]. Each edge represents the most recent time of contact (duration of at least 20 seconds) between two users.

5. Efficiency Results

- Synthetic networks with 5,000 users and 2 actors were use to evaluate the effects of transmission rate and send coefficient on risk propagation efficiency.
- A send coefficient of $\gamma = 0.6$ optimizes for accuracy 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 $\alpha = 0.8$.

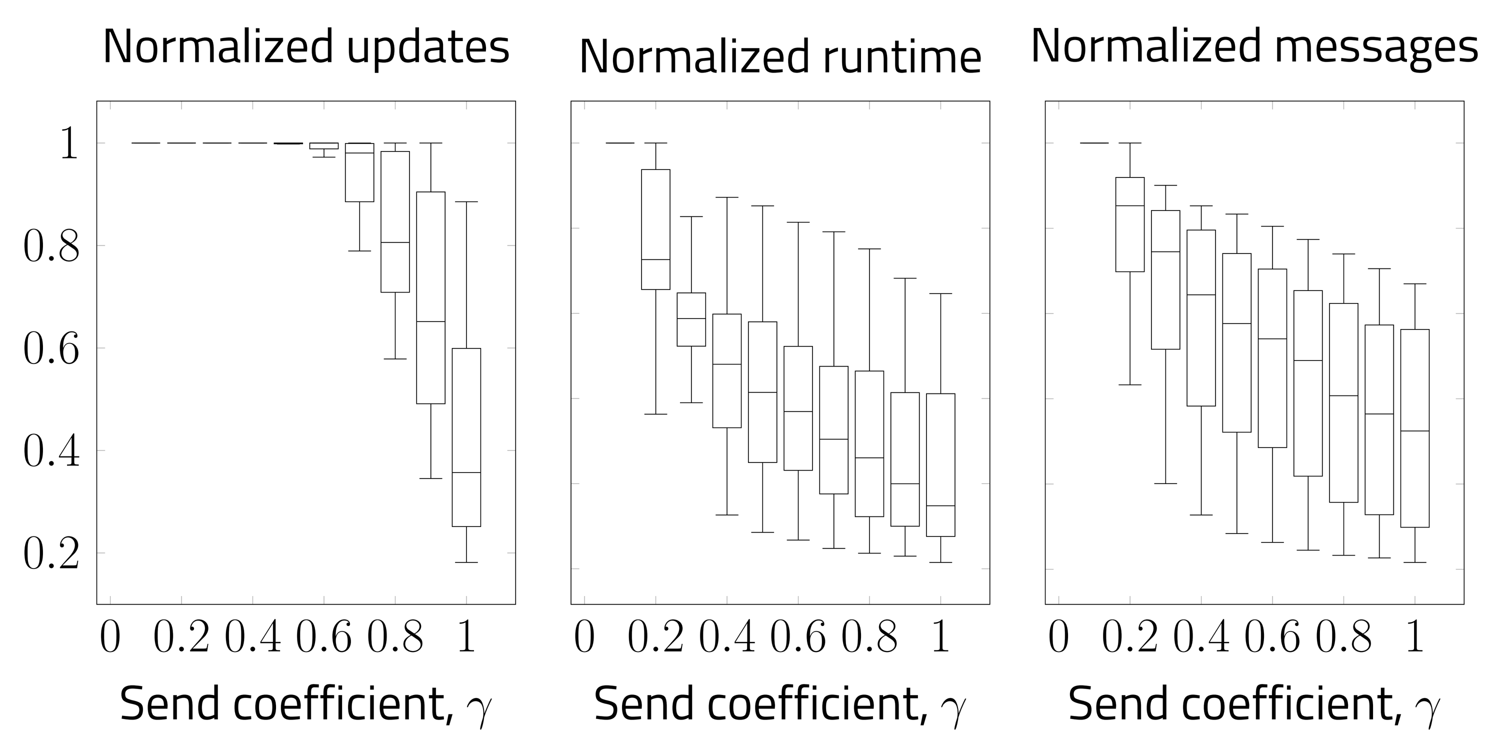


Figure 3. Effects of send coefficient and transmission rate on efficiency. All dependent variables are normalized across networks and transmission rates. Efficiency on LFRGs were less sensitive to changes in transmission rate and send coefficient than RGGs and CSFGs, which is the cause for the large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

6. Message Reachability Results

- Equation (1) was a better estimator for synthetic networks than real-world networks (Table 1).
- With lower (higher) send coefficients (transmission rates), (1) suggests higher MR; however, a message is only passed under certain conditions, so (1) tends to overestimate reachability.

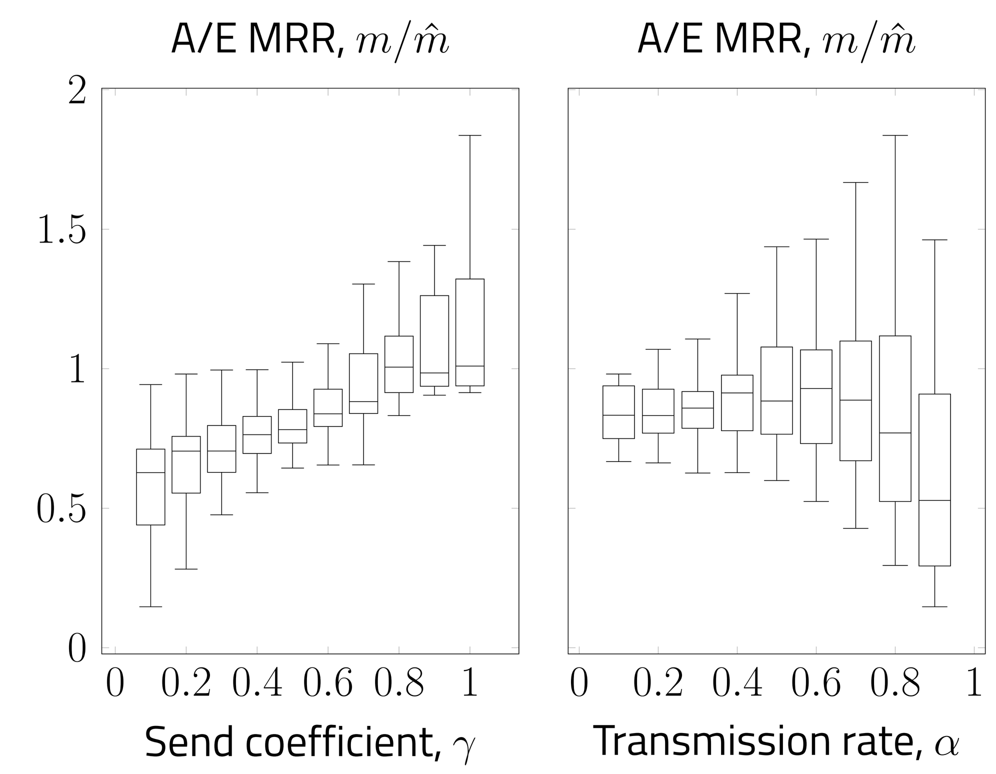


Figure 4. Effects of send coefficient and transmission rate on actual/estimated MR ratio (A/E MRR) on synthetic and real-world networks.

Network	$m/\hat{m} \pm 1.96 \cdot \mathrm{SE}$
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 networks ($\alpha = 0.8, \gamma = 0.6$). Synthetic (real-world) ratios are averaged across parameter combinations (trials).

7. Scalability Results

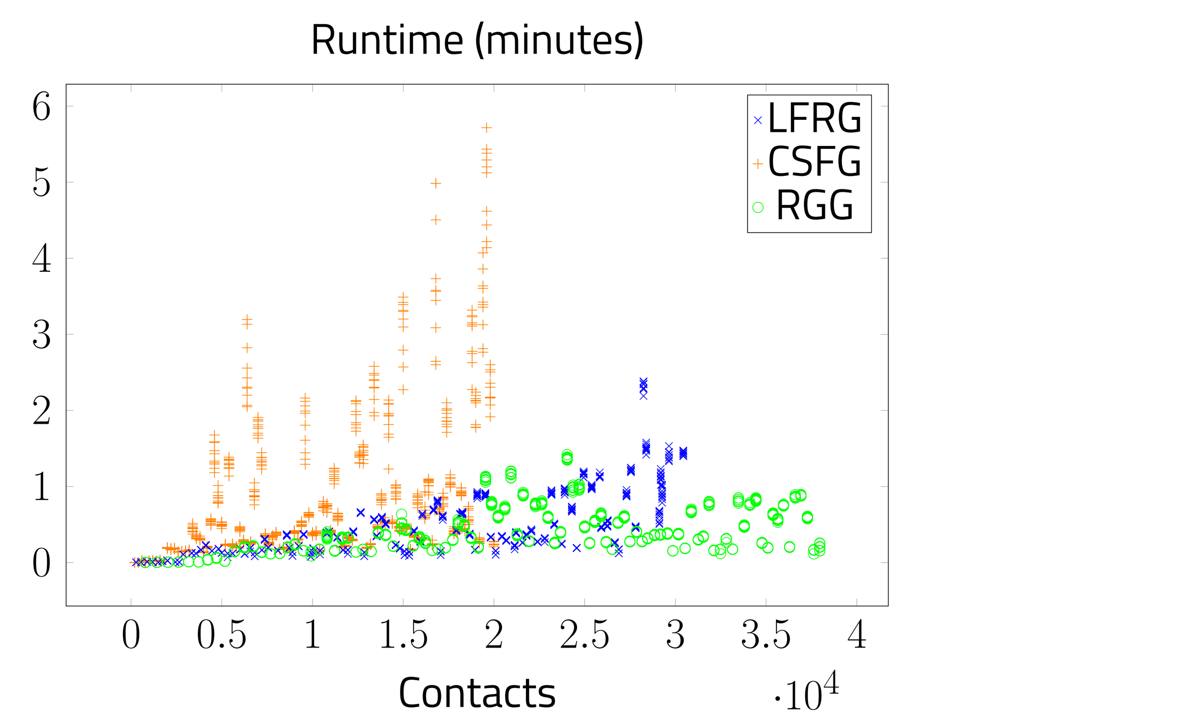


Figure 5. Runtimes of risk propagation on synthetic networks (100–10,000 users; ~200–38,000 contacts). A linear regression fit explains ($R^2 = 0.52$) the runtime of LFRGs and RGGs with slope $m = (1.1 \pm 0.1) \cdot 10^{-3}$ seconds/contact and intercept $b = 4.3 \pm 1.6$ seconds ($\pm 1.96 \cdot \mathrm{SE}$). 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] J. Dall and M. Christensen. Random geometric graphs. *Physical Review E*, 66, 2002.

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

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

[6] 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.

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