# ShareTrace: Contact Tracing with the Actor Model

Ryan Tatton[1] Erman Ayday[1] Youngjin Yoo[2] Anisa Halimi[3]

[1]Department of Computer and Data Sciences
Case Western Reserve University

[2]Deparment of Design and Innovation
Case Western Reserve University

[3]IBM Research Europe

IEEE HealthCom, October 2022

# Motivation

- Proximity-based contact tracing relies on mobile-device interaction to estimate infection spread [2, 6, 8, 16, 18, 19, 23, 24]
- ShareTrace is a privacy-preserving, proximity-based approach that estimates a user's infection risk based on their symptoms, direct contacts, *and* indirect contacts
- Previous work [3] evaluted efficacy, but not scalability or efficiency

# Definitions

- *Risk score* $(r, t)$: a timestamped probability where $r \in [0, 1]$ and $t \geq 0$
  - *Symptom score*: prior infection probability; only considers user symptoms [21]
  - *Exposure score*: posterior infection probability; accounts for (in)direct contact
- *Factor graph* $G = (\mathcal{V}, \mathcal{F}, \mathcal{E})$: a bipartite graph [15]
  - $\mathcal{V}$: variable vertices, where a variable vertex $v \in \mathcal{V}$ represents a user
  - $\mathcal{F}$: factor vertices, where a factor vertex $f(u, v) \in \mathcal{F}$ represents a contact
  - $\mathcal{E}$: edges incident between the sets $\mathcal{V}$ and $\mathcal{F}$
- *Message* $m_{u \to v} = \{(r, t), \ldots\}$: a set of risk scores sent from vertex $u$ to vertex $v$

# Proposed Scheme

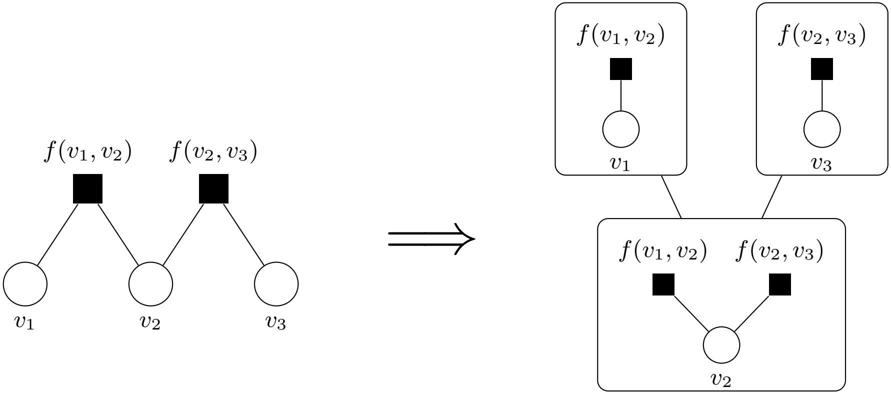



Figure: Apply one-mode projection on the factor graph (left) to get the contact network (right)

# Proposed Scheme

- Reformulate the factor graph as a *temporal contact network* [12, 14]:

$$\mathcal{C} = \{(u, v, t) \mid u, v \in \mathcal{V}; u \neq v; t \geq 0\}$$

- Partition the contact network into $K$ subnetwork *actors* $G_k \in G$ [1, 4]
- Send each actor its initial risk scores (i.e., previous exposure and symptom scores)
- Actors non-iteratively pass messages (*risk scores*) until convergence
- Assumption: risk transmission is incomplete; use a *transmission rate* of $\alpha = 0.8 \in (0, 1)$ [11]

# Proposed Scheme

- Only propagate messages if they are likely to change the risk score of other users in the network
- *Send coefficient* $\gamma$ parametrizes the trade-off between accuracy and efficiency of non-iterative message passing
- A vertex $u$ only sends a message $m_{u \to v}$ if
  - it is close in value to its initial message: $r_{u \to v} \geq \gamma \cdot r_0(u)$; and
  - it is at least as old as its initial message: $t_{u \to v} \leq t_0(u)$
- Ensures convergence when $\alpha < 1$ and $\gamma > 0$

# Message Reachability

- In temporal network theory, a *time-respecting path* (TRP) is a sequence of nondecreasing contacts.
  - Vertex $v$ is *temporally reachable* from vertex $u$ if there exists a TRP from $u$ to $v$
- Message-passing on a temporal network may not require temporal reachability
- *Message reachability* from vertex $u$ to vertex $v$ is the number of edges along the shortest path $\mathcal{P} = u \to v$ that satisfy the message-passing constraints,

$$m(u,v) = \sum_{(i,j)\in\mathcal{P}} f(u,i,j,v),$$

where $f(u,i,j,v) = 1$ if all constraints are satisfied and $f(u,i,j,v) = 0$ otherwise

# Message Reachability: Risk Propagation

Let $H(x) = \mathbf{1}_{x \geq 0}$ be the *Heaviside step function*. Then

$$m(u) = \max_{\mathcal{P}} \left\{ \sum_{(i,j) \in \mathcal{P}} f_c(u,i,j) \cdot f_r(u,i) \cdot f_t(u,i) \right\}$$

where users are enumerated $0, 1, \ldots, |\mathcal{P}| - 1,$ ; and

- Contact-time constraint: $f_c(u,i,j) = H(t_{ij} + \beta - t_0(u))$
  - *Time buffer* $\beta \geq 0$ accounts for incubation and delayed symptom reporting
- Risk-score value constraint: $f_r(u,i) = H(\alpha^i \cdot r_0(u) - \gamma \cdot r_0(i))$
  - Risk score values exponentially decrease as they propagate from their source user
- Risk-score time constraint: $f_t(u,i) = H(t_0(i) - t_0(u))$
  - Older risk scores are less likely to be propagated

# Message Reachability: Risk Propagation

- We can estimate $m(u)$ by relaxing the temporality constraints:

$$\alpha^i \cdot r_0(u) - \gamma \cdot r_0(i) = 0 \overset{\text{solve for } i}{\Longrightarrow} \hat{m}(u, v) = \log_\alpha \left\{ \gamma \cdot \frac{r_0(v)}{r_0(u)} \right\}$$

- (Estimated) message reachability of a vertex $u \in \mathcal{V}$:

$$m(u) = \max\{m(u, v) \mid v \in \mathcal{V}\}$$
$$\hat{m}(u) = \max\{\hat{m}(u, v) \mid v \in \mathcal{V}\}$$

# Message Reachability: Significance

- Generalizes temporal reachability to account for message-passing semantics
- Quantifies the communication complexity of a given algorithm
- RP: higher message reachability $\implies$ more users at risk
- RP: higher send coefficient $\gamma \implies$ higher communication cost (optimize!)

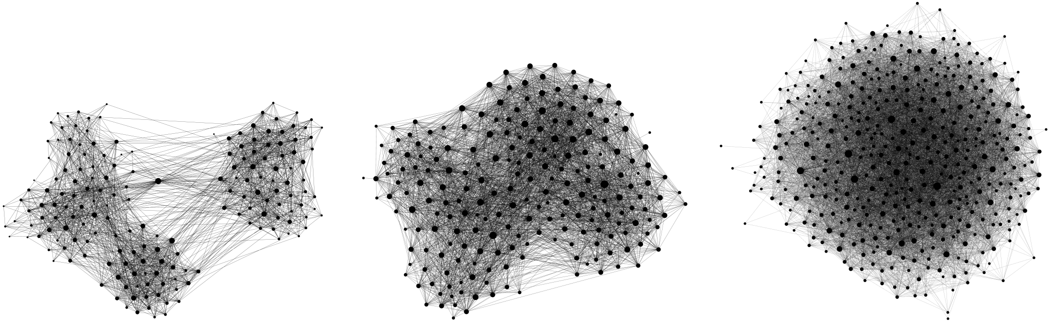
# Evaluation

- What is the optimal value for the send coefficient?
- How do the transmission rate and send coefficient affect accuracy and efficiency?
- How accurate is estimated message reachability?

$$\textit{Message-reachability ratio}: \quad \mathrm{mrr}(u) = \frac{m(u)}{\hat{m}(u)}$$

- What is the runtime performance of risk propagation?
- Synthetic graphs:
  - Random geometric graphs (RGGs) [7]
  - Benchmark graphs (LFRGs) [17]
  - Clustered scale-free graphs (CSFGs) [13]

# Evaluation: SocioPatterns Networks



(a) High school: 180/2,220 [9] (b) Workplace: 217/4,274 [10] (c) Conference: 403/9,565 [10]

Figure: SocioPatterns contact networks (number of users / number of contacts). Edges represents the most recent time of contact (duration $\geq$ 20s).

# Results: Efficiency

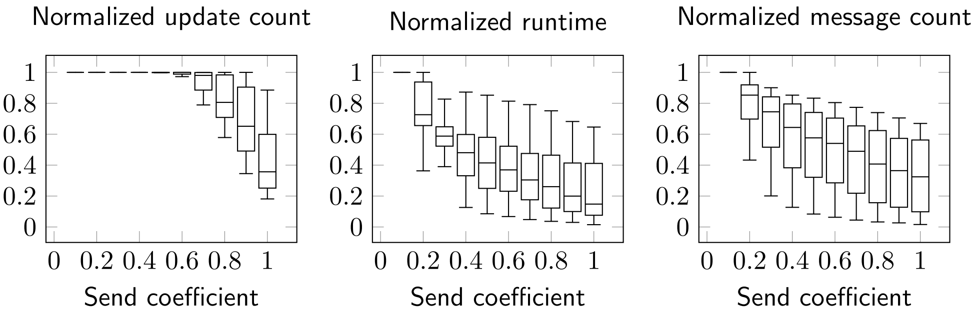



Figure: Effects of send coefficient and transmission rate on efficiency. All dependent variables are normalized across networks and transmission rates. *Update count* is the number of users whose exposure score was different from their initial score; a higher normalized value indicates better accuracy. *Message count* is the number of messages sent by actors; a lower count indicates lower communication overhead. A send coefficient of $\gamma = 0.6$ optimizes for accuracy and efficiency by permitting 99% of the possible user updates.

# Results: Message Reachability

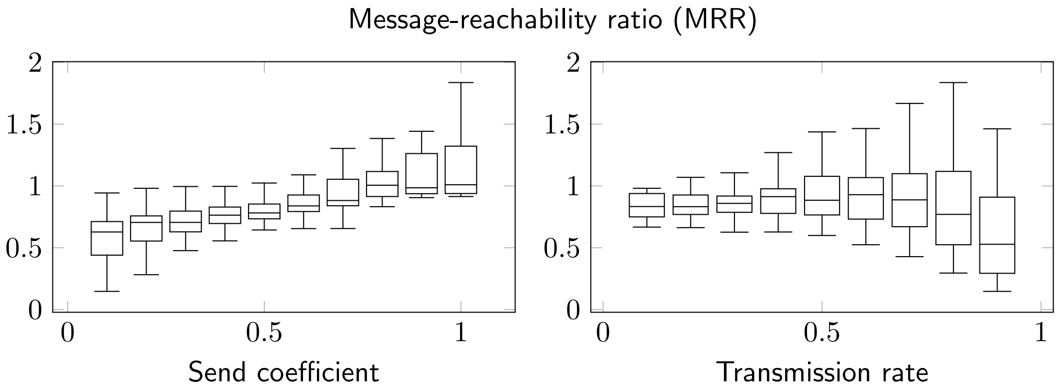



Figure: Effects of send coefficient and transmission rate on the MRR. Independent variables are grouped across networks. MRR $\geq 1$ ($< 1$) indicates overestimation (understimation).

# Results: Message Reachability

| | $\mathrm{mrr}(u) \pm 1.96 \cdot$ SE | | |
|---|---|---|---|
| *Synthetic* | LFR | RGG | CSFG |
| **0.85 ± 0.08** | 0.88 ± 0.14 | 0.74 ± 0.12 | 0.90 ± 0.14 |
| *Real-world* | High school | Workplace | Conference |
| **0.60 ± 0.01** | 0.58 ± 0.01 | 0.63 ± 0.01 | 0.60 ± 0.01 |

Table: Message-reachability ratio for synthetic and real-world contact networks ($\alpha = 0.8$, $\gamma = 0.6$). Synthetic (real-world) ratios are averaged across parameters (runs).

# Results: Scalability

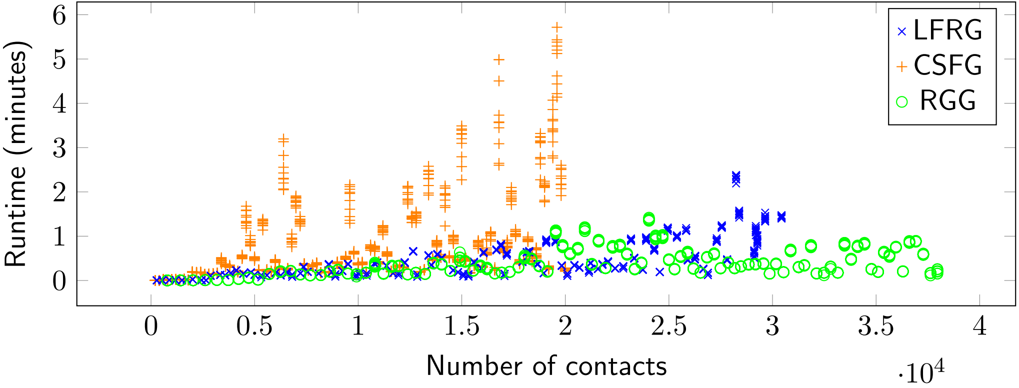



Figure: Runtime of risk propagation on synthetic networks. A linear regression fit explains ($R^2 = 0.52$) the runtime of LFRGs and RGGs with slope $m = (1.1 \pm 0.1) \cdot 10^{-3}$ s/contact and intercept $b = 4.3 \pm 1.6$s ($\pm 1.96 \cdot$ SE).

# Future Work

- Formulate risk propagation as an *online* algorithm.
  - Aligns with the message-passing semantics of risk propagation
  - Accounts for synchronization delays between a user's device and actor
  - Better scalability and fault-tolerance with Akka
  - Self-soverign identity (i.e., Web3) [22]
  - Mobile crowdsensing [5]
- How does message passing affect concurrency and network topology [20]?
- How do different risk-score and contact distributions affect risk propagation?

# References I


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

[2] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. Jha. A survey of COVID-19 contact tracing apps. *IEEE Access*, 8, 2020.

[3] 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 *Proc. 12th ACM Con. Bioinformatics, Comput. Biology, Health Inform.*, BCB 2021, 2021. doi: 10.1145/3459930.3469553.

[4] H. Baker and C. Hewitt. Laws for communicating parallel processes. Technical report, Massachusetts Institute of Technology, 1977.

[5] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry. A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities. *IEEE Commun. Surv. Tut.*, 21(3), 2019.

[6] H. Cho, D. Ippolito, and Y. W. Yu. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs, 2020. e-print: arXiv:2003.11511.

[7] J. Dall and M. Christensen. Random geometric graphs. *Phys. Rev. E*, 66, 2002. doi: 10.1103/PhysRevE.66.016121.

[8] A. B. Dar, A. H. Lone, S. Zahoor, A. A. Khan, and R. Naaz. Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. *Comput. Sci. Rev.*, 38, 2020.

[9] J. Fournet and A. Barrat. Contact patterns among high school students. *PLoS ONE*, 9, 2014. doi: 10.1371/journal.pone.0107878.

# References II


[10] M. G'enois and A. Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Sci.*, 7, 2018. doi: 10.1140/epjds/S13688-018-0140-1.

[11] L. Hamner, P. Dubbel, I. Capron, A. Ross, A. Jordan, J. Lee, J. Lynn, A. Ball, S. Narwal, S. Russell, D. Patrick, and H. Leibrand. High SARS-CoV-2 attack rate following exposure at a choir practice – Skagit County, Washington, March 2020. *MMWR Surveill. Summ.*, 69, 2020. doi: 10.15585/mmwr.mm6919e6.

[12] P. Holme. Modern temporal network theory: a colloquium. *Eur. Phys. J. B*, 88(9), 2015.

[13] P. Holme and B. J. Kim. Growing scale-free networks with tunable clustering. *Phys. Rev. E*, 65, 2002.

[14] P. Holme and J. Saramäki. Temporal networks. *Phys. Rep.*, 519(3), 2012.

[15] F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inf. Theory*, 47, 2001. doi: 10.1109/18.910572.

[16] C. Kuhn, M. Beck, and T. Strufe. Covid notions: Towards formal definitions – and documented understanding – of privacy goals and claimed protection in proximity-tracing services. *Online Soc. Netw. Media*, 22, 2021.

[17] A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E*, 78, 2008.

[18] F. Lucivero, N. Hallowell, S. Johnson, B. Prainsack, G. Samuel, and T. Sharon. COVID-19 and contact tracing apps: Ethical challenges for a social experiment on a global scale. *J. Bioeth. Inq.*, 17, 2020.

# References III


[19] T. Martin, G. Karopoulos, J. Hernández-Ramos, G. Kambourakis, and I. N. Fovino. Demystifying COVID-19 digital contact tracing: A survey on frameworks and mobile apps. *Wirel. Commun. Mob. Comput.*, 2020, 2020.

[20] N. Masuda, J. C. Miller, and P. Holme. Concurrency measures in the era of temporal network epidemiology: a review. *J. R. Soc. Interface*, 18(179), 2021.

[21] C. Menni, A. M. Valdes, M. B. Freidin, C. H. Sudre, L. H. Nguyen, D. A. Drew, S. Ganesh, T. Varsavsky, M. J. Cardoso, J. S. El-Sayed Moustafa, A. Visconti, P. Hysi, R. C. E. Bowyer, M. Mangino, M. Falchi, J. Wolf, S. Ourselin, A. T. Chan, C. J. Steves, and T. D. Spector. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.*, 26, 2020. doi: 10.1038/s41591-020-0916-2.

[22] A. Preukschat and D. Reed. *Self-Soverign Identity*. Manning Publications, Shelter Island, NY, USA, 2021.

[23] R. Raskar, I. Schunemann, R. Barbar, K. Vilcans, J. Gray, P. Vepakomma, S. Kapa, A. Nuzzo, R. Gupta, A. Berke, D. Greenwood, C. Keegan, S. Kanaparti, R. Beaudry, D. Stansbury, B. B. Arcila, R. Kanaparti, V. Pamplona, F. M. Benedetti, A. Clough, R. Das, K. Jain, K. Louisy, G. Nadeau, V. Pamplona, S. Penrod, Y. Rajaee, A. Singh, G. Storm, and J. Werner. Apps gone rogue: Maintaining personal privacy in an epidemic, 2020. e-print: arXiv:1411.5553.

[24] H. Wen, Q. Zhao, Z. Lin, D. Xuan, and N. Shroff. A study of the privacy of COVID-19 contact tracing apps. In N. Park, K. Sun, S. Foresti, K. Butler, and N. Saxena, editors, *Secur. Priv. Commun. Netw.*, volume 335 of *Lect. Notes Inst. Comput. Sci., Soc. Inform. Telecomm. Eng.*, 2020.