ShareTrace: Proactive Contact Tracing with Asynchronous Message Passing

Ryan Tatton

Case Western Reserve University

7 February 2025

Introduction: Types of Contact Tracing

- Digital contact tracing (DCT)
- Proximity tracing
- Decentralized DCT
  - Broadcast model
  - Message-oriented model

Introduction: Limitations of Other Approaches

- No DCT approach exists that incorporates both non-diagnostic information and indirect contacts to estimate infection risk.
- Accounting for indirect contact can substantially improve the efficacy of DCT [10].
- Cherini et al. [4] propose exchanging pseudonyms of indirect contacts, but restrict themselves to diagnostic testing.
- Gupta et al. [6] incorporate non-diagnostic information, but do not account for indirect contact.

Introduction: ShareTrace

- Accounts for both non-diagnostic information and indirect contact to estimate infection risk.
- Developed in collaboration with Dataswyft [2].
- Ayday, Yoo, and Halimi [1] designed ShareTrace to use proximity tracing for contact discovery.
  - In practice, this was infeasible, because Apple and Google's Exposure Notification API did not permit the user's ephemeral identifiers to be stored remotely in a Dataswyft Personal Data Store.

Proposed Design: Definitions

- Risk propagation
- Risk score
  - Symptom score
  - Exposure score

Synchronous Risk Propagation

RISK-PROPAGATION(S, C)

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

2: $r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

Synchronous Risk Propagation

RISK-PROPAGATION(S, C)

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

2: $r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

3: $r_i^{(n)} \leftarrow \infty$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

2: $r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

3: $r_i^{(n)} \leftarrow \infty$

4: **while** $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

$R_i^{(n-1)} \leftarrow$ top k of S_i

$r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

$r_i^{(n)} \leftarrow \infty$

while $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

    $\mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \,..\, n-1] \,\}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

1: $R_i^{(n-1)} \leftarrow$ top k of S_i

2: $r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

3: $r_i^{(n)} \leftarrow \infty$

4: **while** $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

5: $\quad \mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \,..\, n-1] \,\}$

6: $\quad \lambda_{ij}^{(n)} \leftarrow \max \{\, \alpha s_t \mid s_t \in \mu_{ij}^{(n)}, t < t_{ij} + \beta \,\}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

$R_i^{(n-1)} \leftarrow$ top k of S_i

$r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

$r_i^{(n)} \leftarrow \infty$

while $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

    $\mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \,..\, n-1] \,\}$

    $\lambda_{ij}^{(n)} \leftarrow \max \{\, \alpha s_t \mid s_t \in \mu_{ij}^{(n)}, t < t_{ij} + \beta \,\}$

    $R_i^{(n)} \leftarrow$ top k of $\{\, \lambda_{ji}^{(n)} \mid f_{ij} \in N_i \,\}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

$R_i^{(n-1)} \leftarrow$ top k of S_i

$r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

$r_i^{(n)} \leftarrow \infty$

while $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

$\quad \mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \,..\, n-1] \,\}$

$\quad \lambda_{ij}^{(n)} \leftarrow \max \{\, \alpha s_t \mid s_t \in \mu_{ij}^{(n)}, t < t_{ij} + \beta \,\}$

$\quad R_i^{(n)} \leftarrow$ top k of $\{\, \lambda_{ji}^{(n)} \mid f_{ij} \in N_i \,\}$

$\quad r_i^{(n)} \leftarrow \max R_i^{(n)}$

Synchronous Risk Propagation

$\textsc{Risk-Propagation}(S, C)$

$R_i^{(n-1)} \leftarrow$ top k of S_i

$r_i^{(n-1)} \leftarrow \max R_i^{(n-1)}$

$r_i^{(n)} \leftarrow \infty$

while $\|\mathbf{r}^{(n)} - \mathbf{r}^{(n-1)}\| > \epsilon$

$\quad \mu_{ij}^{(n)} \leftarrow R_i^{(n-1)} \setminus \{\, \lambda_{ji}^{(\ell)} \mid \ell \in [1 \,..\, n-1] \,\}$

$\quad \lambda_{ij}^{(n)} \leftarrow \max \{\, \alpha s_t \mid s_t \in \mu_{ij}^{(n)}, t < t_{ij} + \beta \,\}$

$\quad R_i^{(n)} \leftarrow$ top k of $\{\, \lambda_{ji}^{(n)} \mid f_{ij} \in N_i \,\}$

$\quad r_i^{(n)} \leftarrow \max R_i^{(n)}$

return $\mathbf{r}^{(n)}$

Experiment 1: Accuracy I

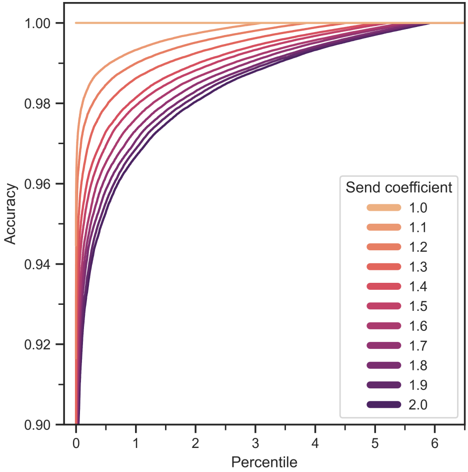

Figure: Cumulative accuracy distributions.

Experiment 1: Accuracy II

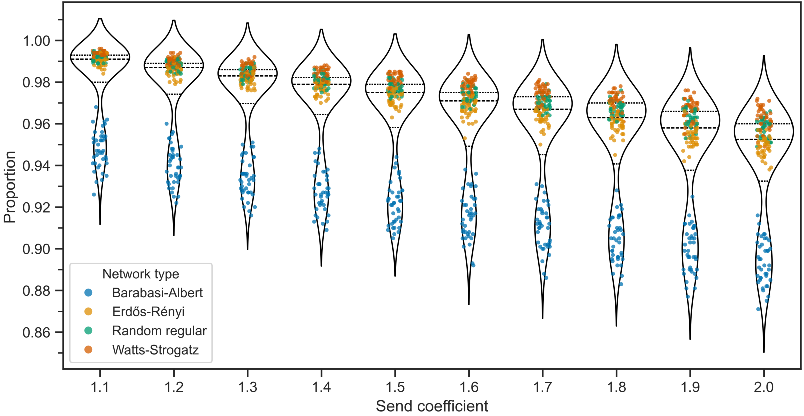

Figure: Send coefficient optimality distributions. The dashed line inside each violin marks the median. The upper and lower dotted lines inside each violin mark the upper and lower quartiles, respectively.

Experiment 1: Efficiency I

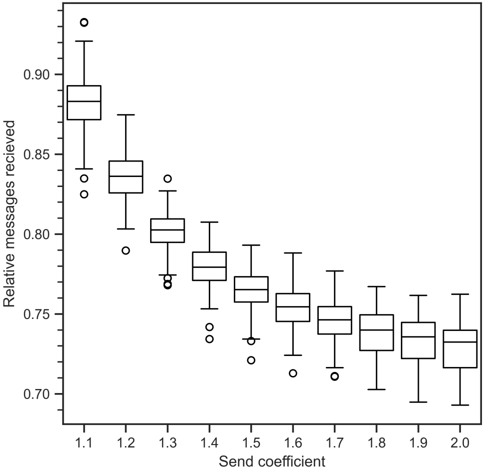

Figure: Message-passing efficiency. The send coefficient $\gamma = 1$ was used as a baseline for message-passing efficiency since it was found to be the maximum send coefficient that achieves perfect accuracy.

Experiment 1: Efficiency II

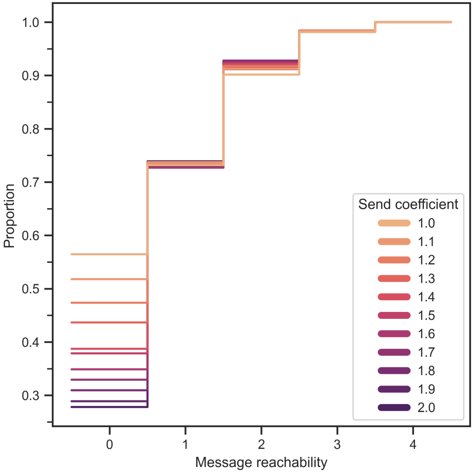

Figure: Message reachability cumulative distributions.

Experiment 1: Exploration I

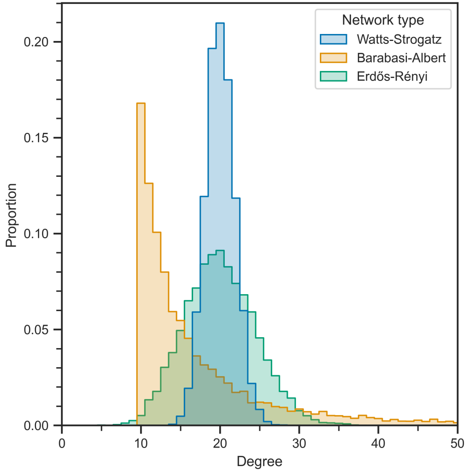

Figure: Contact network degree distributions. All vertices in random regular contact networks had a degree of 20, so the distribution was omitted to provide more visual space for the distributions of other contact networks.

Experiment 1: Exploration II

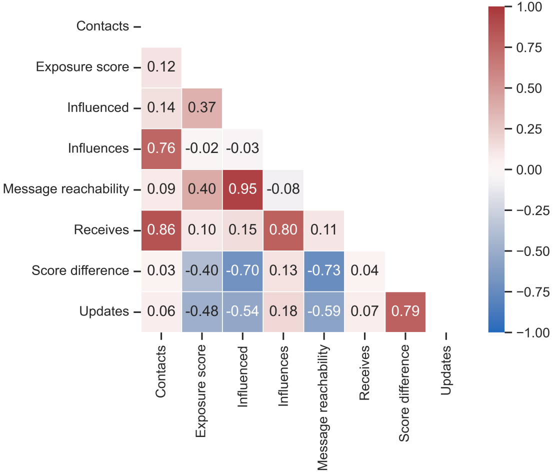

Figure: Correlation matrix of dataset attributes. Each cell is the Spearman rank partial correlation coefficient [15], controlling for the effect of the send coefficient. All coefficients are significant ($p < 0.01$), adjusting for multiple comparisons via the Holm–Bonferroni method [7].

Experiment 2: Benchmarking Hypothesis Testing

Experiment 3: Benchmarking I

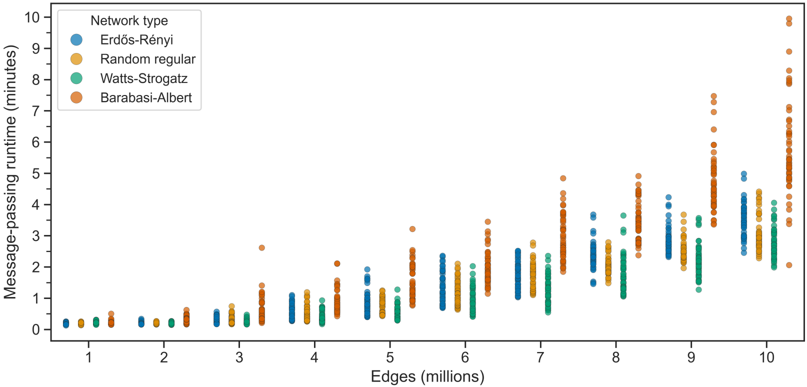

Figure: Message-passing runtimes.

Experiment 3: Benchmarking II

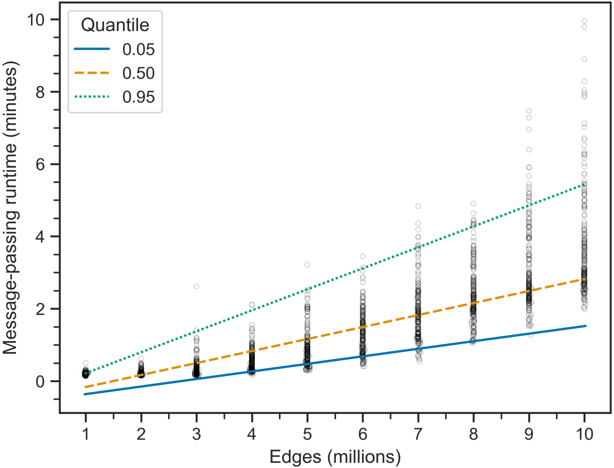

Figure: Message-passing runtimes with regression lines.

Conclusion: Future Work

- Incorporate differential privacy techniques that are designed for DCT applications that utilize risk scores [12].

Conclusion: Future Work

- Incorporate differential privacy techniques that are designed for DCT applications that utilize risk scores [12].
- Formally define the security and privacy characteristics of ShareTrace, using the framework proposed by Kuhn, Beck, and Strufe [9] to characterize the latter.

Conclusion: Future Work

- Incorporate differential privacy techniques that are designed for DCT applications that utilize risk scores [12].
- Formally define the security and privacy characteristics of ShareTrace, using the framework proposed by Kuhn, Beck, and Strufe [9] to characterize the latter.
- Conduct a simulation-based analysis of asynchronous risk propagation with COVI-AgentSim [5].

Conclusion: Future Work

- Incorporate differential privacy techniques that are designed for DCT applications that utilize risk scores [12].
- Formally define the security and privacy characteristics of ShareTrace, using the framework proposed by Kuhn, Beck, and Strufe [9] to characterize the latter.
- Conduct a simulation-based analysis of asynchronous risk propagation with COVI-AgentSim [5].
- Explore the utility and feasibility of integrating decentralized technologies [3, 8, 14, 17, 18] and self-soverign identity [11, 13] into the system design.

Prior Designs and Implementations

- "Thinking like a vertex" with Apache Giraph
- Factor subgraph actors
- Driver-monitor-worker framework
- Projected subgraph actors [16]
- Contact search

References I

[1] Erman Ayday, Youngjin Yoo, and Anisa 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. DOI: 10.1145/3459930.3469553.

[2] Erman Ayday et al. *ShareTrace: A smart privacy-preserving contact tracing solution by architectural design during an epidemic*. White paper. Case Western Reserve University, 2020.

[3] Juan Benet. *IPFS - content addressed, versioned, P2P file system*. 2014. arXiv: 1407.3561 [cs.NI].

[4] Renato Cherini et al. "Toward deep digital contact tracing: Opportunities and challenges". In: *IEEE Pervasive Computing* 22.4 (2023), pp. 15–25. DOI: 10.1109/mprv.2023.3320987.

[5] Prateek Gupta et al. *COVI-AgentSim: An agent-based model for evaluating methods of digital contact tracing*. 2020. arXiv: 2010.16004 [cs.CY].

[6] Prateek Gupta et al. "Proactive contact tracing". In: *PLOS Digital Health* 2.3 (2023), pp. 1–19. DOI: 10.1371/journal.pdig.0000199.

[7] Sture Holm. "A simple sequentially rejective multiple test procedure". In: *Scandinavian Journal of Statistics* 6.2 (1979), pp. 65–70. URL: https://www.jstor.org/stable/4615733.

References II

[8] Navin Keizer et al. "A survey on content retrieval on the decentralised web". In: *ACM Computing Surveys* 56.8 (2024). DOI: 10.1145/3649132.

[9] Christiane Kuhn, Martin Beck, and Thorsten Strufe. "Covid notions: Towards formal definitions—and documented understanding—of privacy goals and claimed protection in proximity-tracing services". In: *Online Social Networks and Media* 22 (2021). DOI: 10.1016/j.osnem.2021.100125.

[10] Francisco Pozo-Martin et al. "Comparative effectiveness of contact tracing interventions in the context of the COVID-19 pandemic: A systematic review". In: *European Journal of Epidemiology* 38.3 (2023), pp. 243–266. DOI: 10.1007/s10654-023-00963-z.

[11] Alex Preukschat and Drummond Reed. *Self-sovereign identity: Decentralized digital identity and verifiable credentials*. Manning, 2021.

[12] Rob Romijnders et al. "Protect your score: Contact-tracing with differential privacy guarantees". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 38.13 (2024), pp. 14829–14837. DOI: 10.1609/aaai.v38i13.29402.

[13] Frederico Schardong and Ricardo Custódio. "Self-sovereign identity: A systematic review, mapping and taxonomy". In: *Sensors* 22.15 (2022). DOI: 10.3390/s22155641.

References III

[14] Ruizhe Shi et al. "A closer look into IPFS: Accessibility, content, and performance". In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 8.2 (2024). DOI: 10.1145/3656015.

[15] Charles Spearman. "The proof and measurement of association between two things". In: *The American Journal of Psychology* 15.1 (1904), pp. 72–101. DOI: 10.2307/1412159.

[16] Ryan Tatton et al. "ShareTrace: Contact tracing with the actor model". In: *2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*. ©2022 IEEE. 2022, pp. 13–18. DOI: 10.1109/healthcom54947.2022.9982762.

[17] Dennis Trautwein et al. "Design and evaluation of IPFS: A storage layer for the decentralized web". In: *Proceedings of the ACM SIGCOMM 2022 Conference*. 2022, pp. 739–752. DOI: 10.1145/3544216.3544232.

[18] Carmela Troncoso et al. "Systematizing decentralization and privacy: Lessons from 15 years of research and deployments". In: *Proceedings on Privacy Enhancing Technologies* 2017.4 (2017), pp. 307–329. DOI: 10.1515/popets-2017-0056.