Analysing the effectiveness of backward contact tracing in epidemic control

Mahi Patel (201801039)* and Riddhi Tanna (201801427)†
Dhirubhai Ambani Institute of Information & Communication Technology,
Gandhinagar, Gujarat 382007, India
SC437, Complex Networks

In this project, we aim to analyse the effectiveness of forward + backward contact tracing over forward contact tracing. Using simulation models and mathematical expressions, we visualise by what factor the effectiveness increases and how the epidemic chain breaks faster on combining both forward and backward contact tracing.

I. INTRODUCTION

Contact tracing is an essential tool in controlling the spread of epidemics. It involves tracing people that have come in contact with an infected person and isolating them to break the chain of transmission. Predominantly, forward contact tracing is used to identify potential infections, i.e., tracing to whom the infection may have spread. We explore the efficiency of backward contact tracing (i.e., tracing from whom the infection may have spread). This method proves to be more efficient than forward contact tracing in terms of controlling the size of the epidemic since it fully leverages the statistical biases that arise from heterogeneity in contacts.

We can model the network of people using nodes that represent people and an edge between two nodes that represents face-to-face contact between the two nodes (people) [1].

II. BIAS DUE TO FORWARD TRACING

When a node is infected, its neighbours can potentially be infected through the edges. Then, a node with more edges is more likely to appear as a neighbour. Thus, it has a higher chance of being infected. This phenomenon follows from the friendship paradox which states that your friends are likely to have more friends than you. Let $p(x)$ be the degree-distribution of a node with x contacts. Hence, a neighbouring node with $x^{'}$ contacts would have a degree-distribution proportional to $x^{'}p(x)$ (following the friendship paradox). This bias is leveraged during forward contact tracing where we trace the contacts of randomly sampled infected nodes [1].

III. BIAS DUE TO BACKWARD TRACING

Using the simple branching process model, we check the efficiency of backward contact tracing and forward

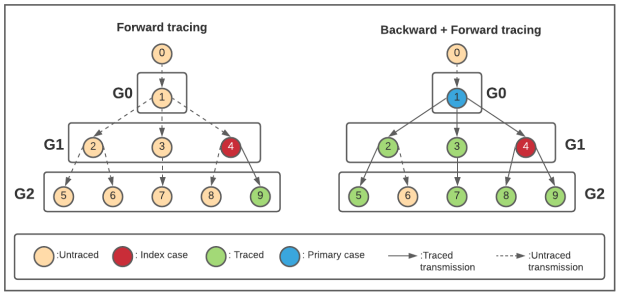

FIG. 1: Schematic Representation of Forward tracing and forward + backward tracing

contact tracing. An index case is discovered by symptom-based surveillance. The transmission chains connected to the index case are numbered such that the index case belongs to Generation 1 (G_1). In backward tracing, the parent of the index case belongs to Generation 0 (G_0) and is identified as the primary case. We forward trace from the primary case to other G_1 contacts.

The transmission chains are modeled by a branching process (as followed in [2]) where $p(x)$ represents the probability mass function of the number of secondary transmissions by a single case. So, $p(x)$ is higher for the primary case since the probability of identifying a primary case is proportional to the number of contacts of the node. Hence, the number of offsprings of the primary case can be modeled by $p(x|G_0) = \frac{xp(x)}{E(x)}$ where $E(x) = \sum xp(x)$.

The average number of G_1 cases identified by backward tracing (including the index case) is $E(x|G_0) = \sum x^2p(x)E(x) = E(x^2)E(x) = R(1+v^2)$, where $R = E(x)$ is the reproduction number and v is the coefficient of variation (the standard deviation of x divided by its mean). So, the higher the overdispersion (v), the higher is the number of G_1 cases traced through backward tracing of the index case. Whereas in the case of forward tracing, the mean number of identified cases is R, independent of the degree of overdispersion. $p(x)$ is assumed to follow a negative binomial distribution with an overdispersion parameter k. The mean number of G_1 cases identified by backward tracing is:

$$E(x|G_0) = R(1+v^2) = 1 + R(1+\frac{1}{k}) \qquad (1)$$

* 201801039@daiict.ac.in
† 201801427@daiict.ac.in

where k is inversely proportional to v^2.

IV. METHODOLOGY

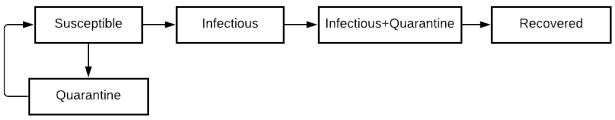

FIG. 2: Five-Compartment Model for Epidemic Simulation

A basic simulation model is a modified SIR model with five compartments (Fig. 2): Susceptible, Infected, Infected-Quarantined, Quarantined and Recovered. A susceptible node gets infected. An infected node may recover or die. Under both circumstances, they move to the recovered compartment. As seen in II, G_2 cases are forward traced through the index case successfully, and hence the relative reduction in infectiousness due to quarantine is c. As seen in III, primary case (G_0) is identified through backward tracing successfully with a probability b which results in identification of more offsprings (G_1 cases). Each subsequent case is identified successfully with a probability q. The overdispersion parameter is k and the reproduction number is R. In the case that a G_1 case is not traced, it may also be independently identified through symptom-based surveillance with a probability d. We run a simulation for 50 time steps and record our observations in Section V.

A good measure of effectiveness of a method can be the number of cases we manage to avert [2]. Hence, we measured the effectiveness of contact tracing with respect to the relative reduction in the total number of G_3 cases. We assume that by the time we trace a G_2 case, we may be able to isolate it but we may not be able to avert the infection.

Hence, following equation 1, in the case of forward tracing, the average number of G_2 cases traced are

$$< G_2\_traced >= Rq(1 + Rd(1 + \frac{1}{k})) \tag{2}$$

and the estimated number of G_3 cases averted are

$$< G_3\_averted >= R^2qc(1 + Rd(1 + \frac{1}{k})) \tag{3}$$

In the case of forward and backward tracing, number of G_1 cases identified are

$$G_1\_identified = (1 - (1 - d)(1 - bq))R(1 + \frac{1}{k}) \tag{4}$$

and the average number of G_2 cases traced are

$$< G_2\_traced >= Rq(1 + (1 - (1 - d)(1 - bq))R(1 + 1/k)) \tag{5}$$

Hence, the estimated number of G_3 cases averted is

$$< G_3\_averted >= R^2qc(1 + (1 - (1 - d)(1 - bq))R(1 + \frac{1}{k}) \tag{6}$$

V. OBSERVATIONS

Fig. 3 and Fig. 4 show two snapshots of our simulation at time $t = 25$ and $t = 50$. We can see that at $t = 25$, there are more number of infected nodes in the case of forward tracing than in the case of forward + backward tracing. Similarly, at the end of the simulation at $t = 49$, lesser number of nodes have been infected in the case of forward + backward tracing than in the case of forward tracing. Hence, the size of the epidemic is smaller in the case of forward + backward tracing.

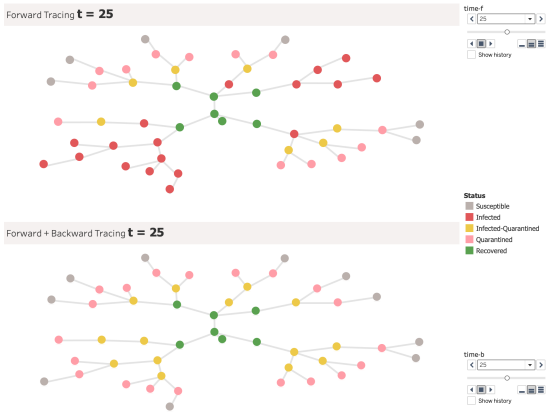

FIG. 3: Snapshot of epidemic simulation at t = 25 for both cases

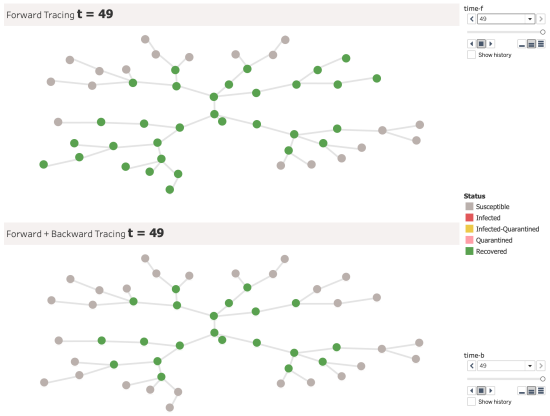

FIG. 4: Snapshot of epidemic simulation at t = 49 (end of simulation) for both cases

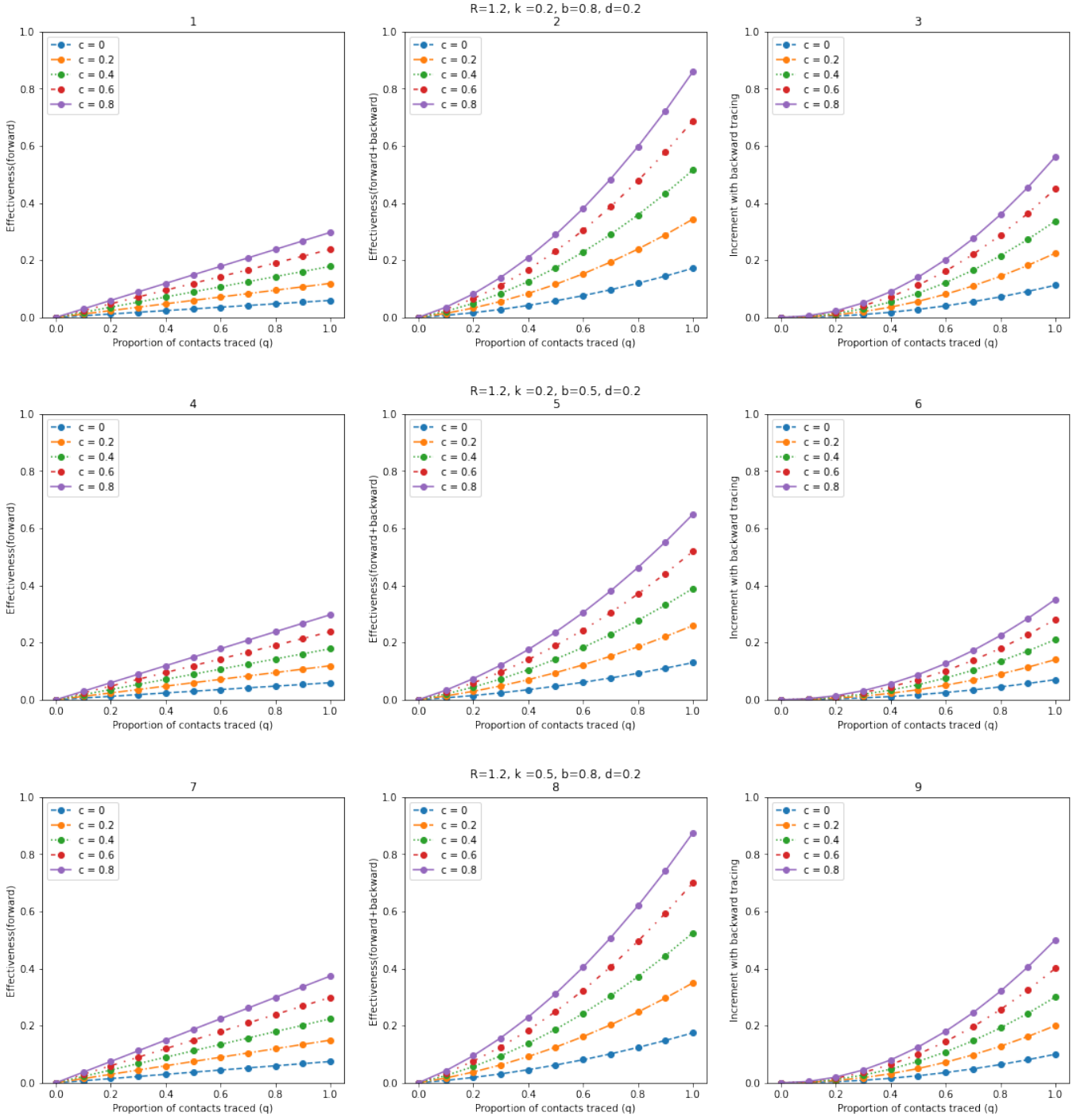

FIG. 5: Plot of the estimated effectiveness (i.e proportion of G_3 cases averted) of forward and forward + backward contact tracing against proportion of traced contacts(q) for different parameter values. Left panel (1, 4, 7): Effectiveness of forward tracing alone; Middle panel (2, 5, 8): Effectiveness of both forward + backward tracing; Right panel (3, 6, 9): Incremental effectiveness of using backward tracing along with forward tracing. Colours represent the relative reduction in infectiousness of G_2 cases if traced and put in quarantine(c).

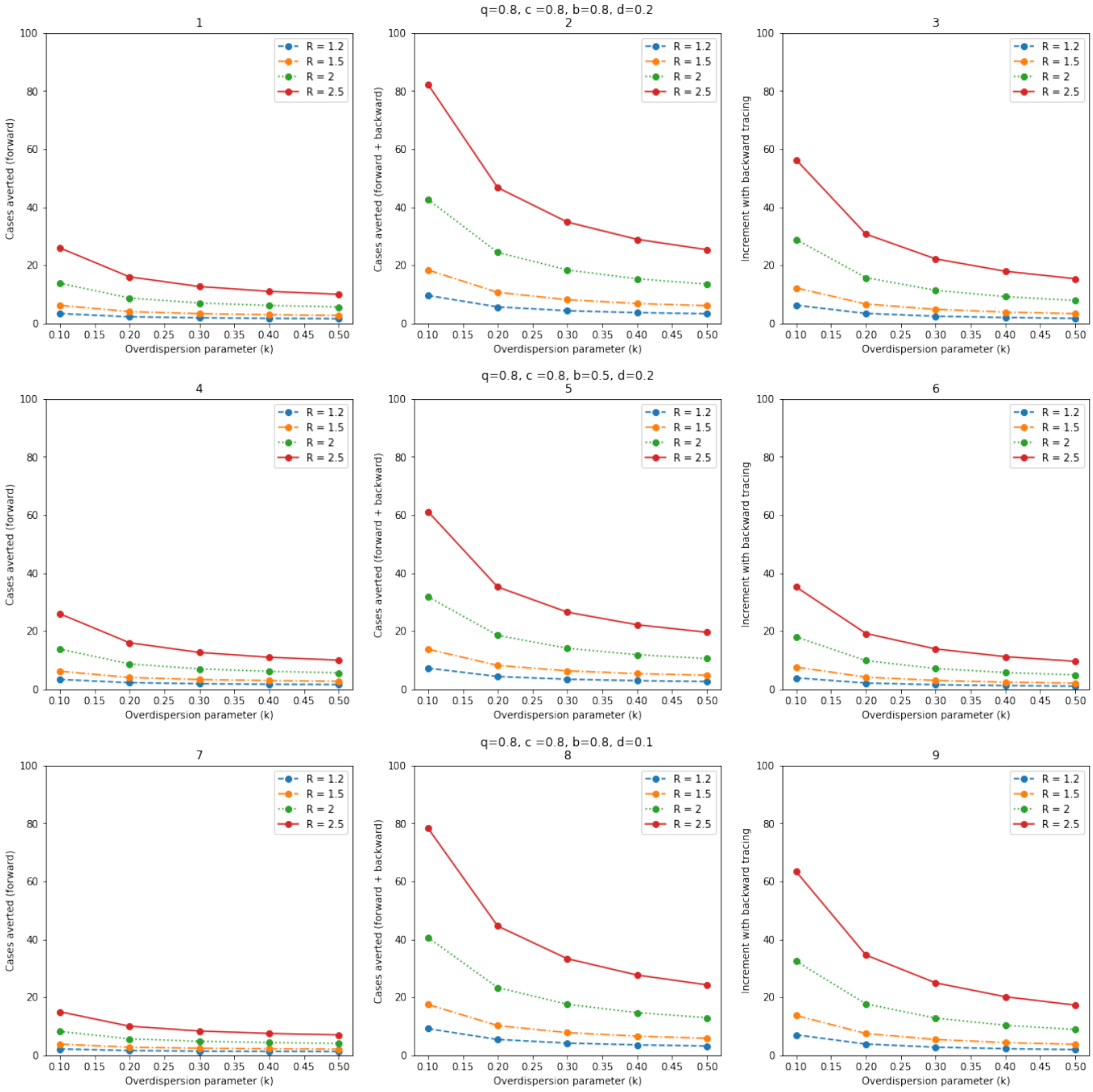

FIG. 6: Plot of the number of G_3 cases averted of forward and backward contact tracing against overdispersion parameter(k) for different parameter values. Left panel (1, 4, 7): Cases averted by forward tracing alone; Middle panel (2, 5, 8): Cases averted by both forward + backward tracing; Right panel (3, 6, 9): Increment in cases averted by using backward tracing along with forward tracing. Colours represent different reproduction numbers(R).

In Fig. 5, we observe that combining both forward and backward contact tracing increases the effectiveness of contact tracing by a factor of almost 2 as q approaches. Smaller k and larger b result in aversion of larger number of cases by backward tracing, thus increasing the effectiveness. With an increase in the value of c, the effectiveness increases.

In Fig: 6, we observe that combining both forward and backward contact tracing increases the number of total cases averted in G_3. With the increase in value of k, number of cases averted reduces, lowering the effectiveness.

VI. DISCUSSION

Through our model and simulations, we intend to verify the effectiveness of forward + backward contact tracing over forward contact tracing. We see that the former applies forward tracing after backward tracing to the parent case to identify multiple G_1 cases. There is an overheard of identifying the primary case (G_0) in the former on top of what happens in the latter. In reality, increased back tracing has the risk of contacts providing misleading information due to recall bias. This may reduce the overall effectiveness of back tracing. Using backward tracing may also help in identifying super-spreading hubs due to the leveraged statistical biases as mentioned in Section III. Hence, following the proposed model, applying forward tracing to those hubs will further avert G_3 cases.

A caveat to note is that the observations discussed above hold true only when k is very small [2].

VII. CONCLUSION

We see that a promising strategy in reducing the size of an epidemic is by implementing a mix of backward and forward tracing. Our results show that this strategy can be almost 2-3 times more effective than the traditional forward contact tracing. Our results provide further evidence for this approach by quantifying the possible benefit of backward tracing, especially when the offspring distribution is highly variable (i.e. when k is small).

[1] Kojaku, S., Hébert-Dufresne, L., Mones, E. et al. "The effectiveness of backward contact tracing in networks" *Nat. Phys. 17*, 652–658 (2021).

[2] Akira Endo, Quentin J Leclerc, Gwenan M Knight, et al. "Implication of backward contact tracing in the presence of overdispersed transmission in COVID-19 outbreak" *Welcome Open Research*, (2020).

[3] M. E. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical review E*, vol. 64, no. 2, 2001.

[4] GitHub Repository for Data Preprocessing to create Network Charts in Tableau (Link)