Supplemental material: Influence maximization on temporal networks

1 Network characteristics

The datasets used in the study represent temporal networks with interactions between two entities of the network, either directed or undirected. Each interaction is represented with a time stamp indicating the time when it occurred. We divide these datasets into slices of equal length W in time. After dividing, we only consider those slices/layers that have more than 10% of all nodes in the dataset active for the given layer, i.e., have at least one interaction in the layer. To give an example, the "Hypertext, 2009" dataset has been divided into slices of length 14400 seconds. The division gives 15 slices in total. 4 of these 15 slices have less than 10% of all nodes active. These 4 slices are removed, and we end up with a temporal network with 11 layers. All the information for the other networks regarding the number of total, removed, and remaining slices can be found in Table S1. Once the sparse layers are removed, we end up with temporal networks that have layers with densities shown Figure S1.

Dataset	W	total	removed	T
Email, dept. 1	2880000	25	7	18
Email, dept. 2	2880000	25	7	18
Email, dept. 3	2880000	25	7	18
Email, dept. 4	2880000	25	7	18
High school, 2011	14400	19	8	11
High school, 2012	14400	51	30	21
High school, 2013	14400	26	12	14
Hospital ward	14400	25	5	20
Hypertext, 2009	14400	15	4	11
Primary school	7200	17	6	11
Workplace	28800	35	15	20
Workplace-2	28800	35	15	20

Table S1: List of the empirical datasets used to construct temporal networks. From left to right, we report: the name of the dataset, the length W of the temporal window used to slice the data (time is reported in seconds), the total number of slices obtained by dividing the dataset, the number of removed slices that had less than 10% of all nodes active in it, and the number T of network layers resulting after slicing and cleaning data.

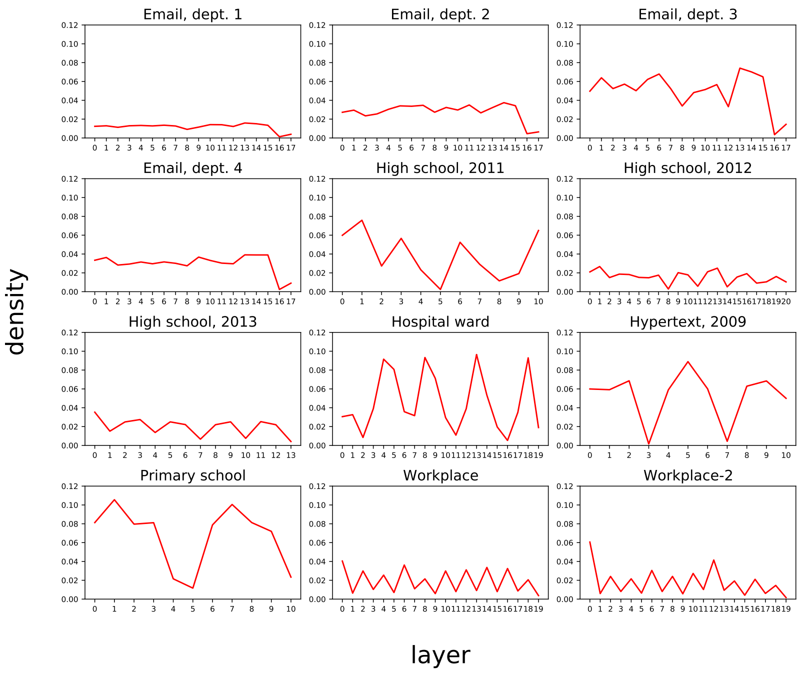

Figure S1: **Density values of each layer in the temporal networks.** Each panel represents a single network, showing the evolution of density between layers. The density is defined as $2E_t/(N(N-1))$, where E_t is the number of edges in layer t, and N is the total number of nodes in the network, including all those that had at least one interaction in the whole temporal network.

2 Critical threshold

2.1 Finding the critical threshold

In order to analyze the problem of influence maximization in different dynamical regimes, we need to estimate the critical value λ_c of the spreading probability λ of each temporal network. We remark that $\lambda_c = \lambda_c(\mu)$, i.e., λ_c is a function of the recovery probability μ. To estimate λ_c, we start the spreading process from each active node in the first layer of the temporal network. The process is repeated 500 times for each node and averaged over all realizations and all nodes. The numerical simulations are done for a range of λ values, the λ value that maximizes the ratio of standard deviation and mean of outbreak sizes is found using brute-force search, and the value found gives the critical threshold for a temporal network and a given μ value.

2.2 Effect of shuffling layer order on critical threshold

In our analysis, we observe that the temporal ordering of the layers has a significant effect on the dynamics. The value of the critical threshold changes with the ordering, and also the influential spreaders identified differ depending on the layer ordering. In Figures S2-S12 we show the effect of shuffling the order of the layers on the critical threshold value of the temporal network. In most cases, we observe that the true value of the critical threshold is lower than the mean and median of the critical thresholds calculated from multiple realizations of the temporal network with randomly ordered layers. This pattern suggests that the interaction dynamics in a temporal network are not happening randomly, and there is a correlation between the interaction patterns of layers closer to each other in the ordering.

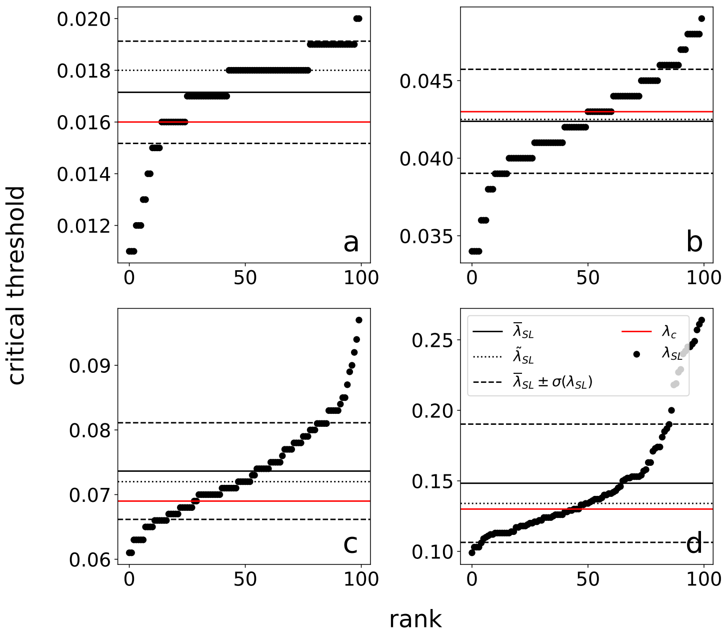

Figure S2: **Sensitivity of the spreading outcome to network dynamics.** (a) Best estimates of the critical spreading probability λ_{SL} for randomized versions of the "Email, dept. 1" temporal network. SIR recovery probability is $\mu = 0$. We display horizontal lines identifying the average $\bar{\lambda}_{SL}$ (full black line), the region corresponding to one standard deviation away from the mean [$\bar{\lambda}_{SL} \pm \sigma(\lambda_{SL})$, dashed black lines], the median value $\tilde{\lambda}_{SL}$ (dotted black line), and the actual critical value λ_c measured on the non-randomized version of the network (red full line). (b) Same as in panel a, but for $\mu = 0.25$. (c) Same as in panel a, but for $\mu = 0.5$. (d) Same as in panel a, but for $\mu = 1$.

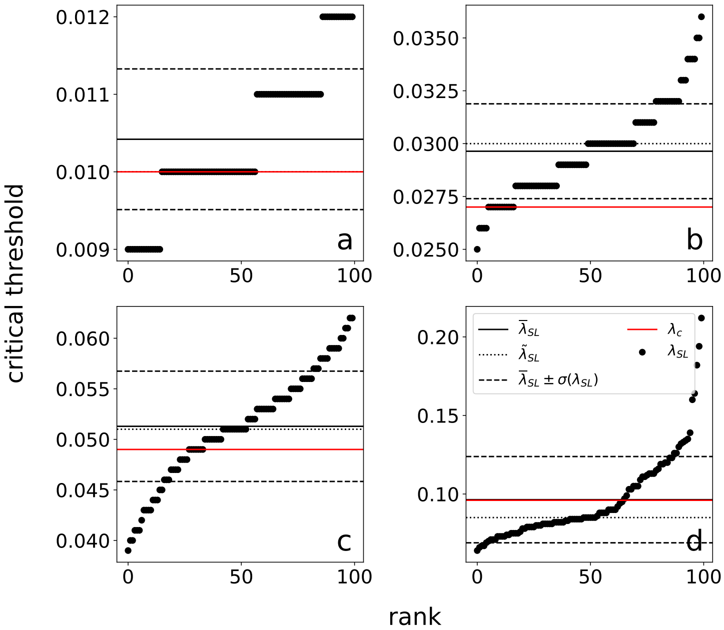

Figure S3: Same as Fig. S2, but for "Email, dept. 2" network.

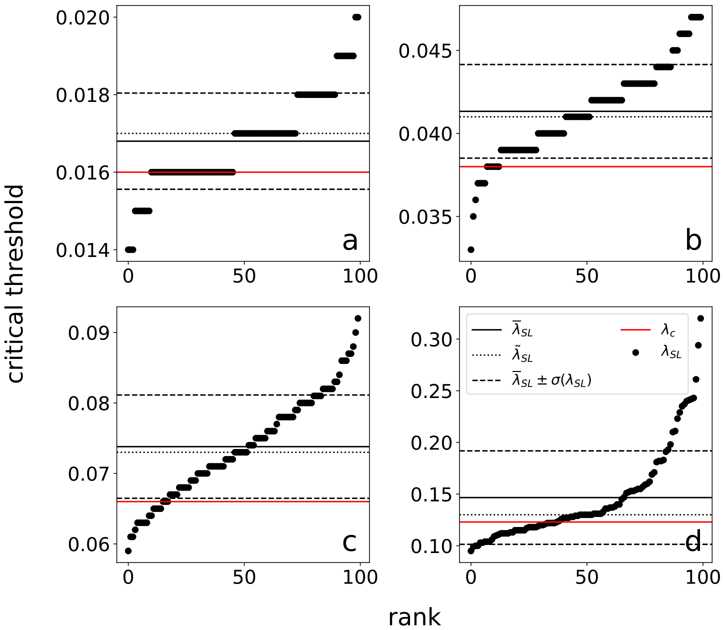

Figure S4: Same as Fig. S2, but for "Email, dept. 3" network.

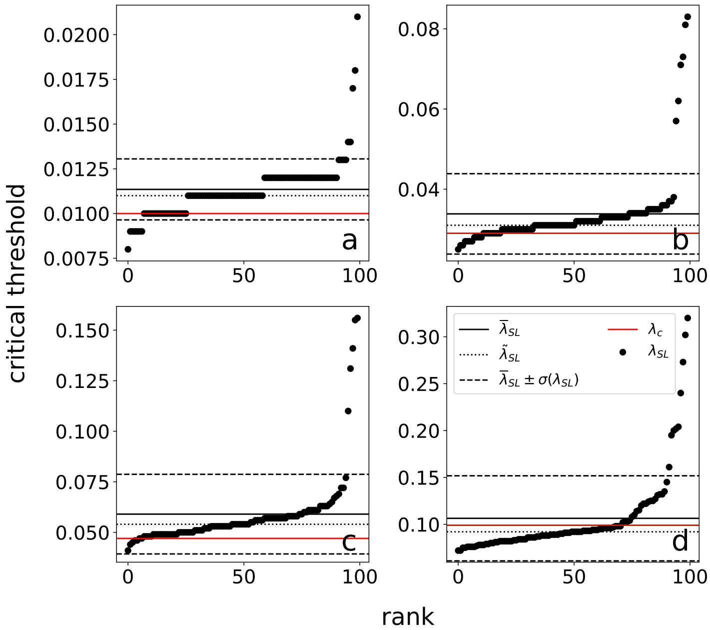

Figure S5: Same as Fig. S2, but for "Email, dept. 4" network.

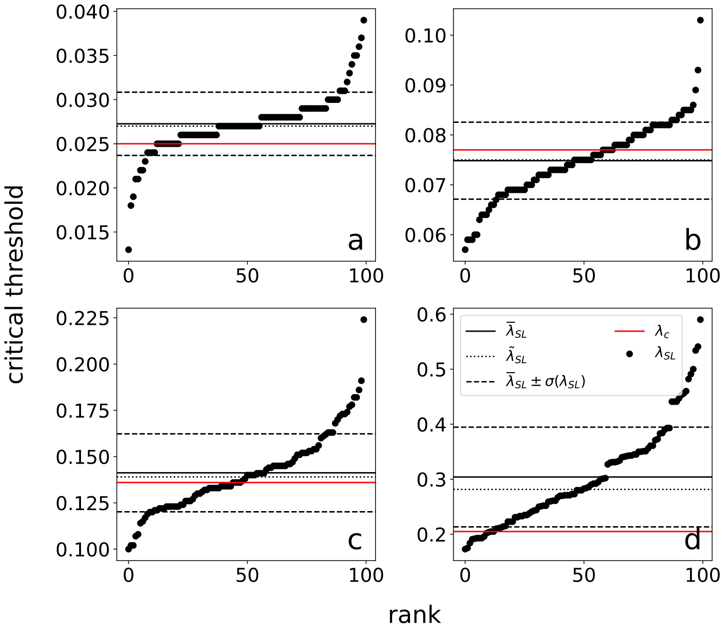

Figure S6: Same as Fig. S2, but for "High school, 2012" network.

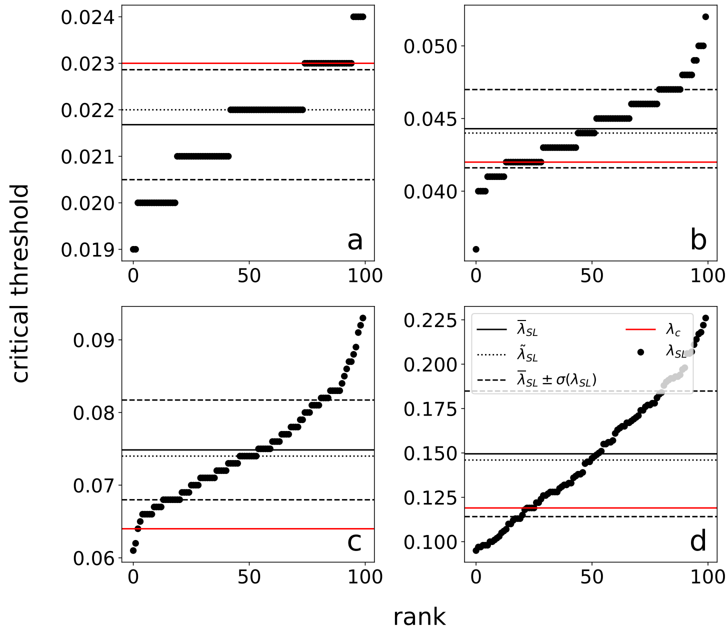

Figure S7: Same as Fig. S2, but for "High school, 2013" network.

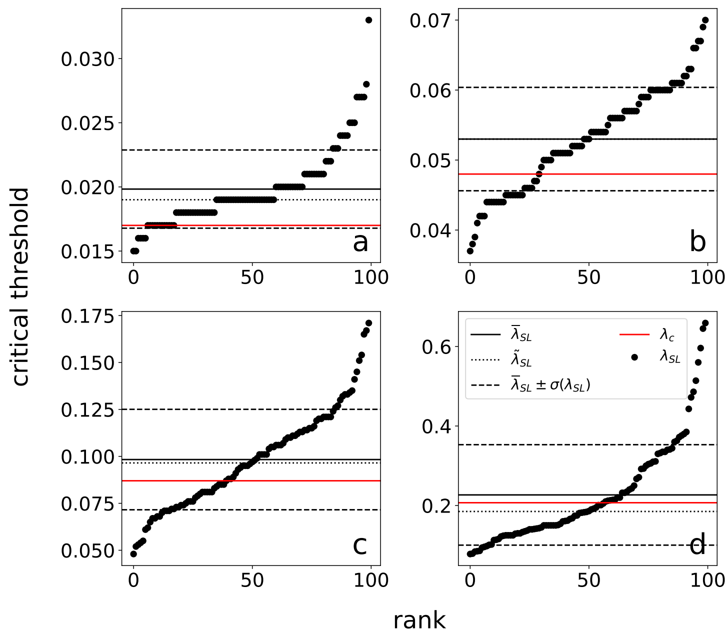

Figure S8: Same as Fig. S2, but for "Hospital ward" network.

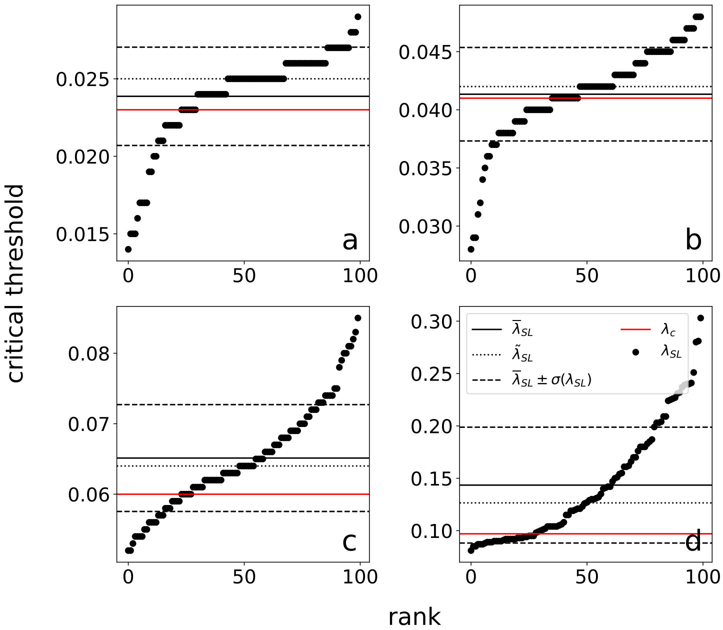

Figure S9: Same as Fig. S2, but for ”Hypertext, 2009” network.

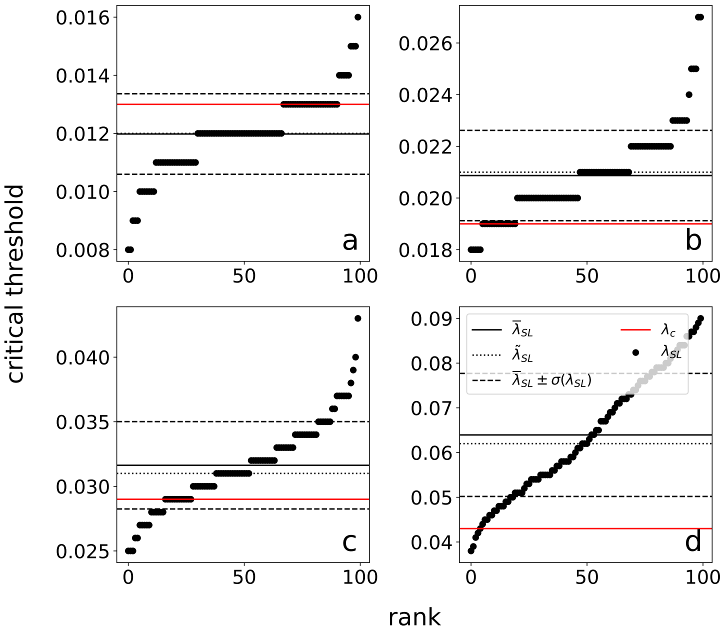

Figure S10: Same as Fig. S2, but for ”Primary school” network.

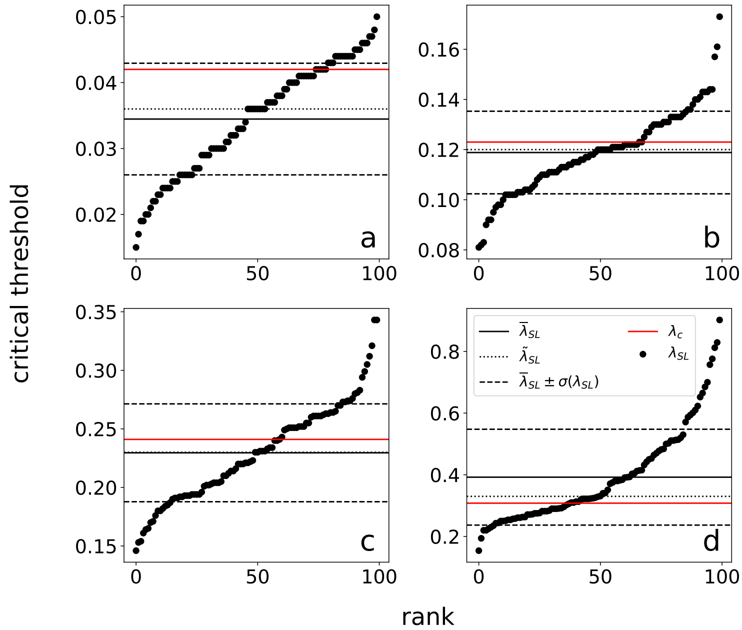

Figure S11: Same as Fig. S2, but for "Workplace" network.

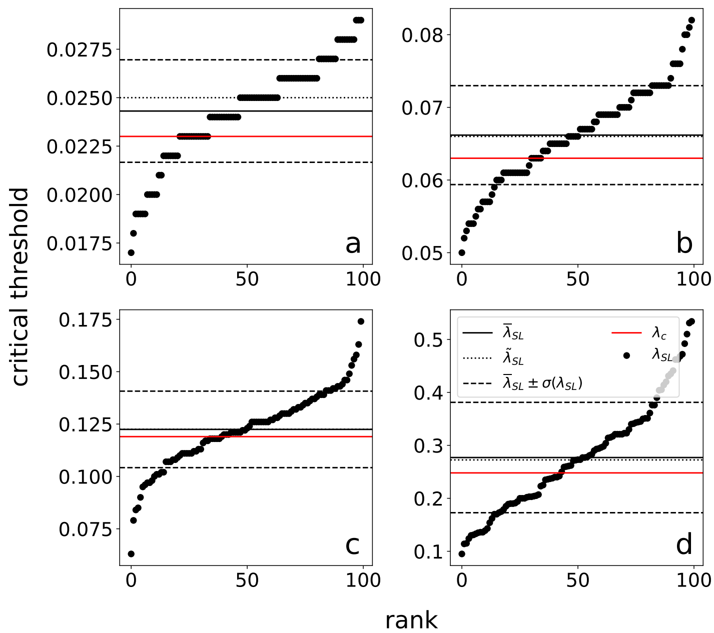

Figure S12: Same as Fig. S2, but for "Workplace-2" network.

3 Influence maximization

3.1 Accounting for time horizon

We consider three methods that use a static network as the only topological input for the identification of the influential spreaders. The first uses the first layer (FL), the second uses a randomly selected layer (RL), and the last uses the aggregated version of the temporal network (ST) to find influential spreaders. When the SIR model is employed on a static network, traditionally the model stops when there are no more infected nodes remaining. In our SIR model on temporal networks, the spreading happens until the end of the time horizon (i.e., number of layers in a temporal network). In the three methods previously mentioned, when $\mu > 0$, there is no information on time horizon (when $\mu = 0$, the time horizon is known because if we let the SIR model run until there are no more infected nodes, the spreading would end when all nodes are infected, which is not informative for selecting influential spreaders). To observe the effect of knowing the time horizon on these three methods, we implement FL-T, RL-T, and ST-T. All these methods run for T (number of layers) steps to find influential spreaders, rather than waiting for no nodes to be left in the infected state. The results can be observed in Figure S13. From the figure, it can be observed that adding the information of time horizon to the methods has no significant effect on their performance.

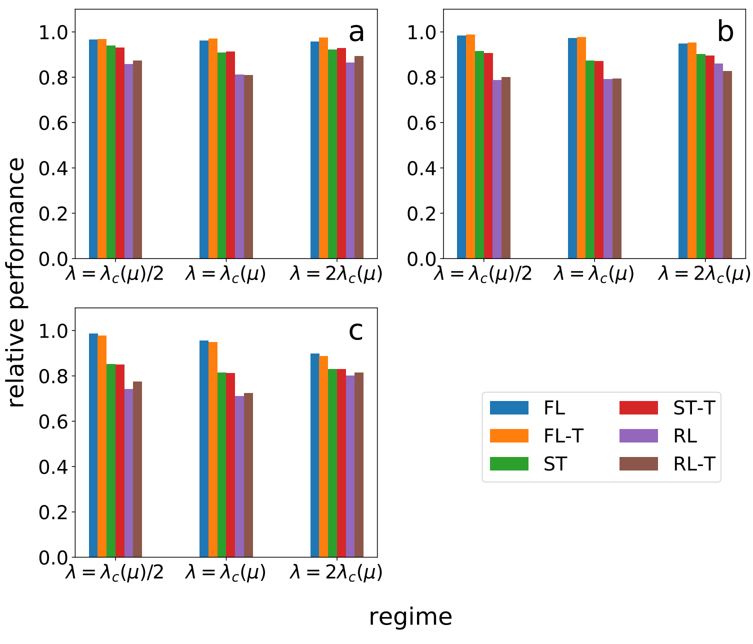

Figure S13: (a) Similar to Fig. 5 of the main paper, but for different methods. The figure serves to analyze the effect of the information of time horizon on the identification of influential spreaders. Different dynamical regimes are studied with recovery probability $\mu = 0.25$ fixed. (b) Same as in panel a, but for $\mu = 0.5$. (c) Same as in panel a, but for $\mu = 1$.

4 Results for tests of performance

Here we present the average value of the outbreak size as a function of the relative size of the seed set, for all networks, under different values of the spreading and recovery probabilities. The figures are similar to the Figure 4 of the main text, but here all the methods analyzed in the paper are included in the plots.

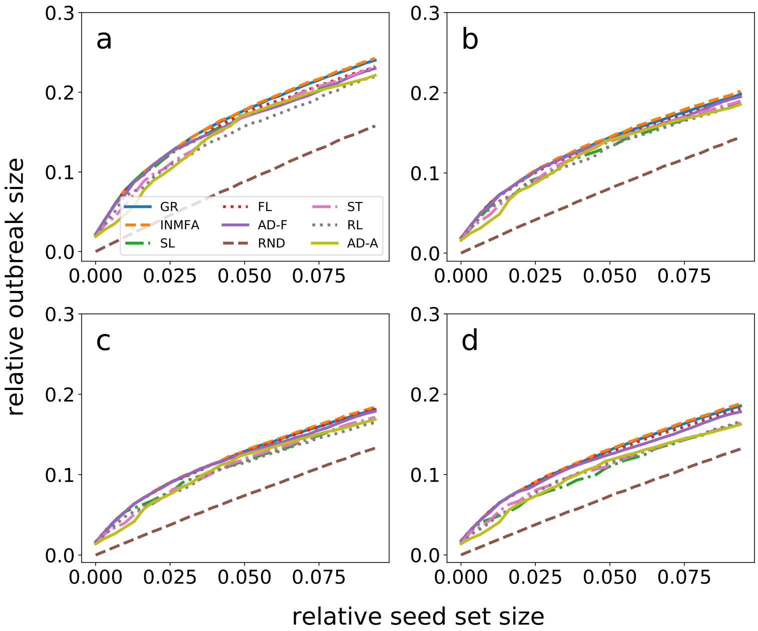

Figure S14: **Identification of influential spreaders in temporal networks.** (a) Average value of the relative size of the outbreak, i.e., $\langle O(\mathcal{X})\rangle$, as a function of the relative size of the seed set, i.e., $|\mathcal{X}|/N$. The network analyzed is "Email, dept. 1". Spreading dynamics is subcritical, i.e., $\lambda = 0.5\lambda_c(\mu)$, and the recovery probability is $\mu = 0$. (b) Same as in panel a, but for $\mu = 0.25$. (c) Same as in panel a, but for $\mu = 0.5$. (d) Same as in panel a, but for $\mu = 1$.

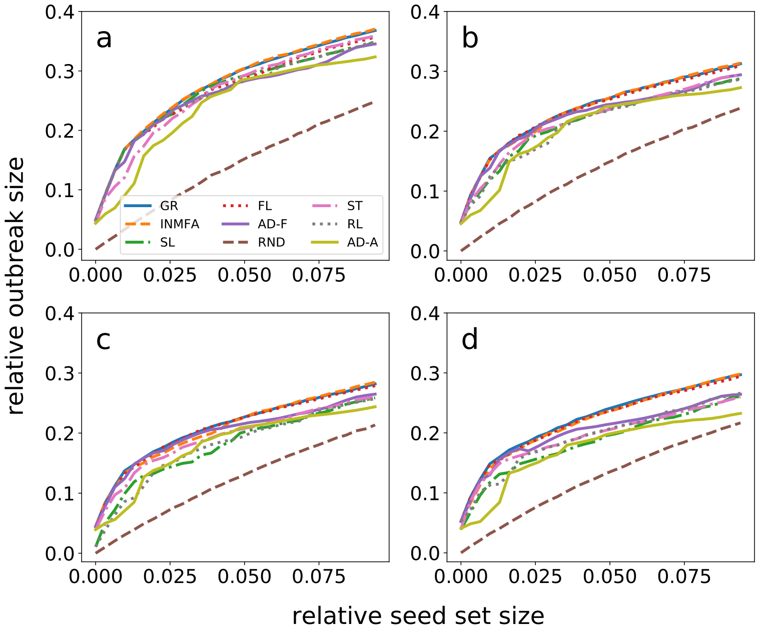

Figure S15: Same as Fig. S14, but for ”Email, dept. 1” network in critical regime.

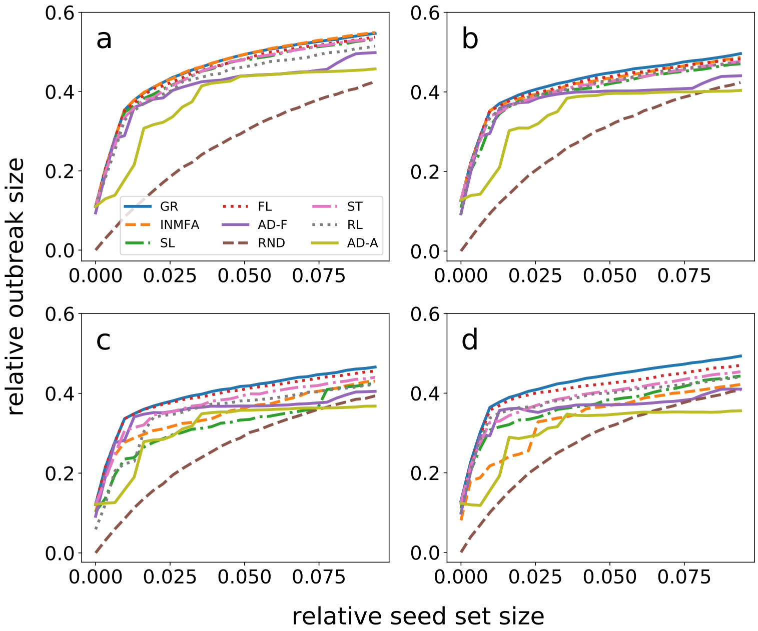

Figure S16: Same as Fig. S14, but for ”Email, dept. 1” network in supercritical regime, i.e., $\lambda = 2\lambda_c(\mu)$.

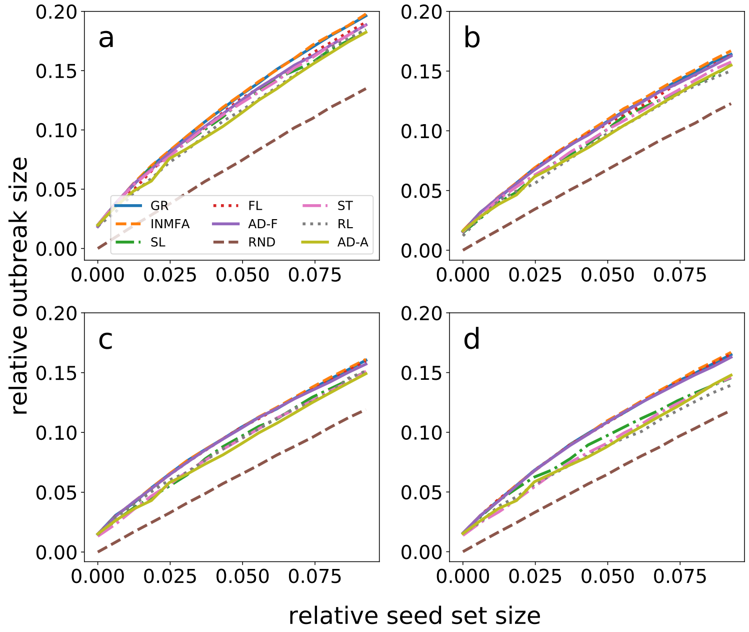

Figure S17: Same as Fig. S14, but for ”Email, dept. 2” network in subcritical regime.

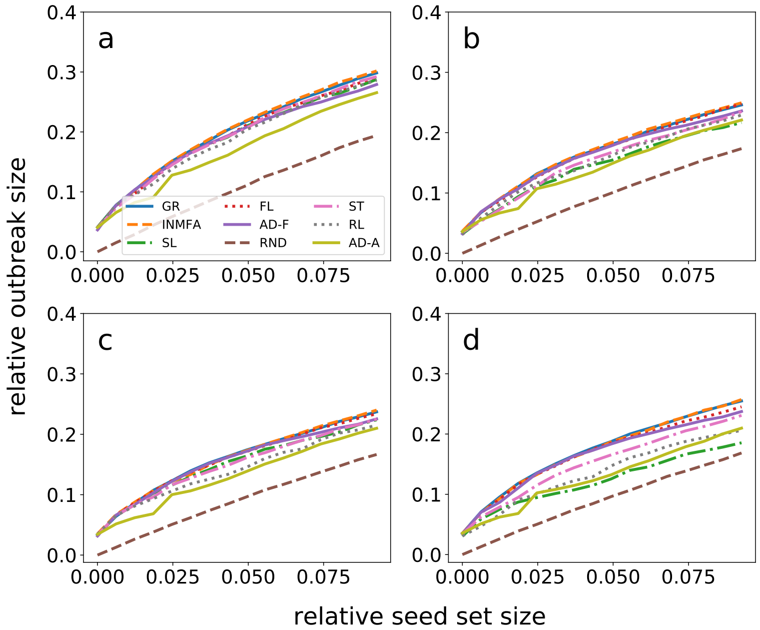

Figure S18: Same as Fig. S14, but for ”Email, dept. 2” network in critical regime.

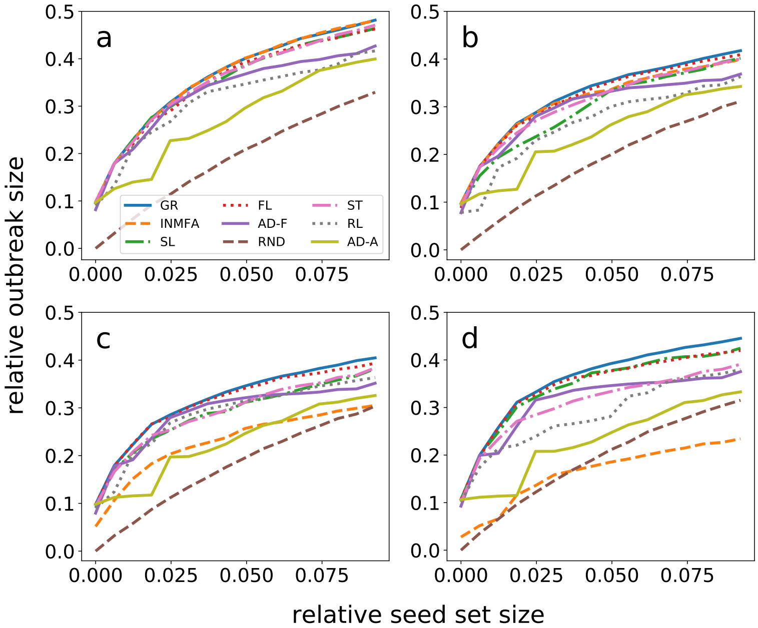

Figure S19: Same as Fig. S14, but for "Email, dept. 2" network in supercritical regime.

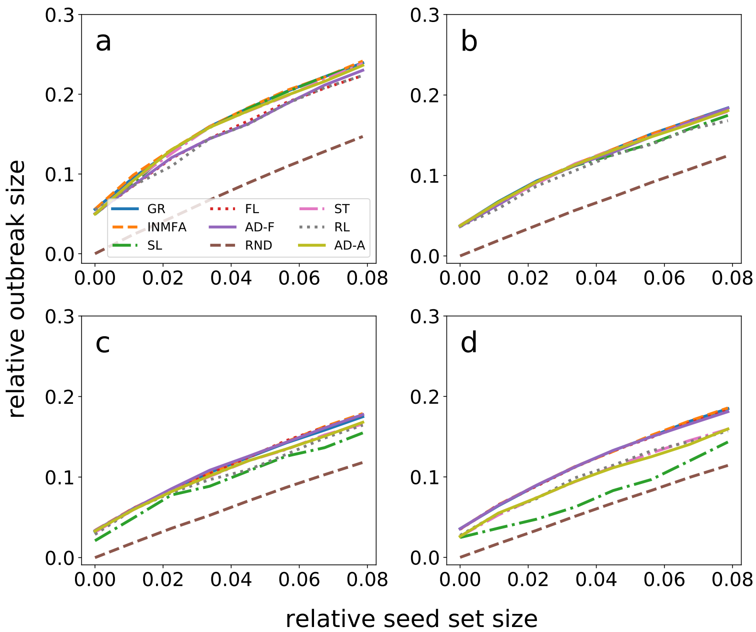

Figure S20: Same as Fig. S14, but for "Email, dept. 3" network in subcritical regime.

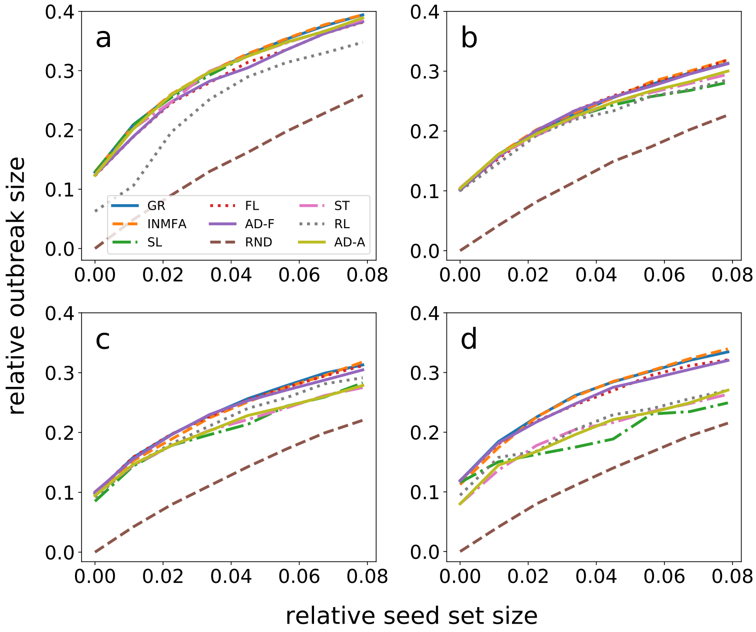

Figure S21: Same as Fig. S14, but for "Email, dept. 3" network in critical regime.

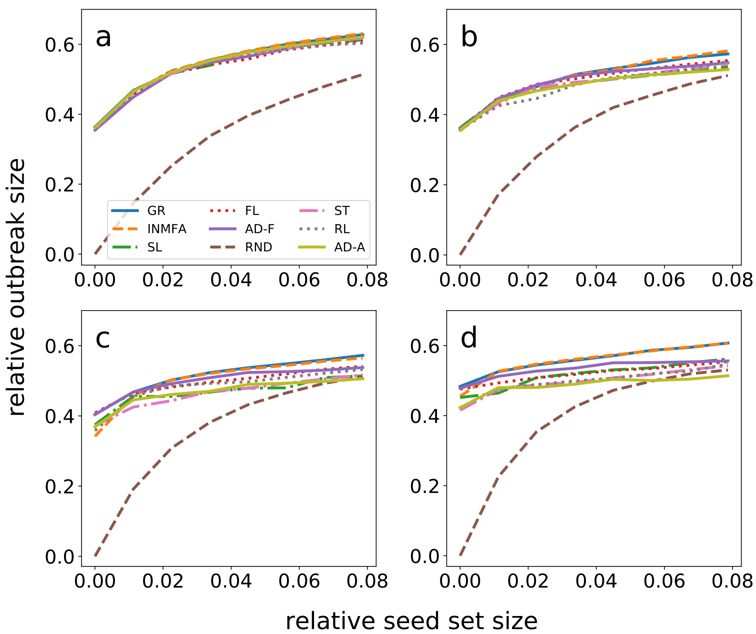

Figure S22: Same as Fig. S14, but for "Email, dept. 3" network in supercritical regime.

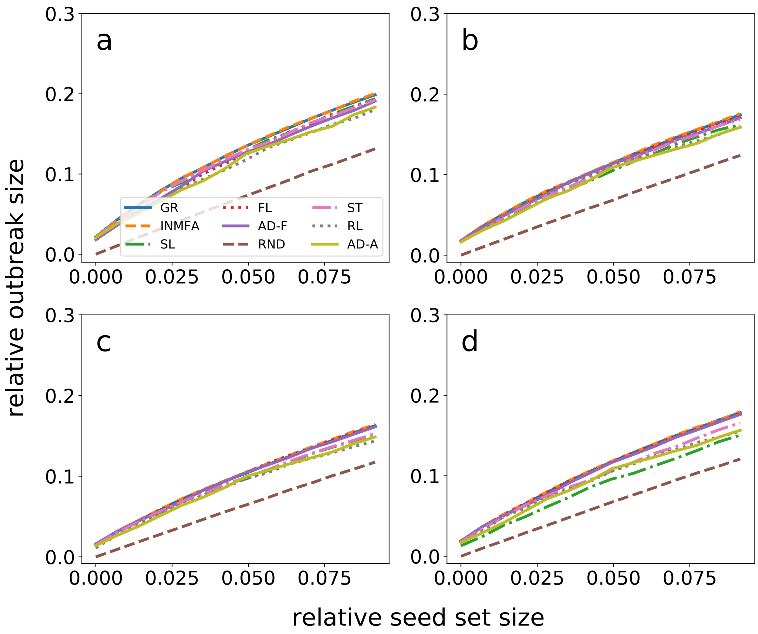

Figure S23: Same as Fig. S14, but for "Email, dept. 4" network in subcritical regime.

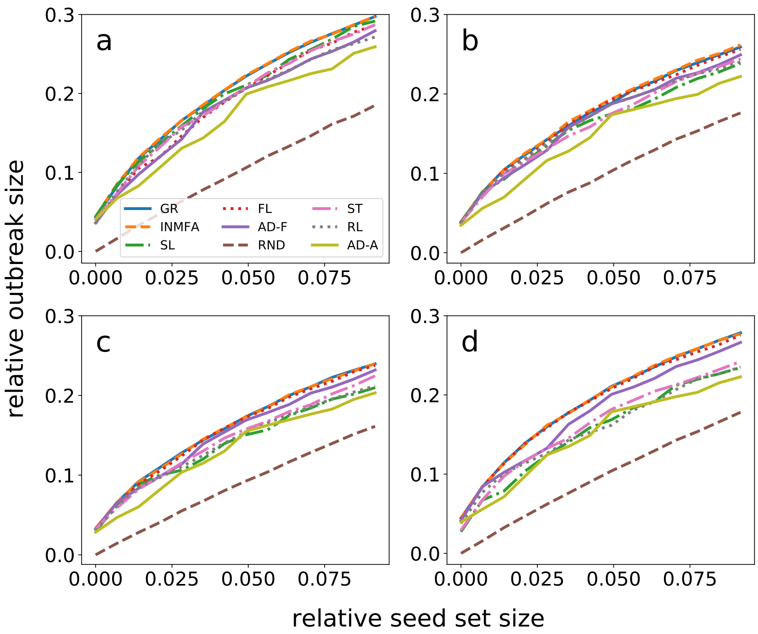

Figure S24: Same as Fig. S14, but for "Email, dept. 4" network in critical regime.

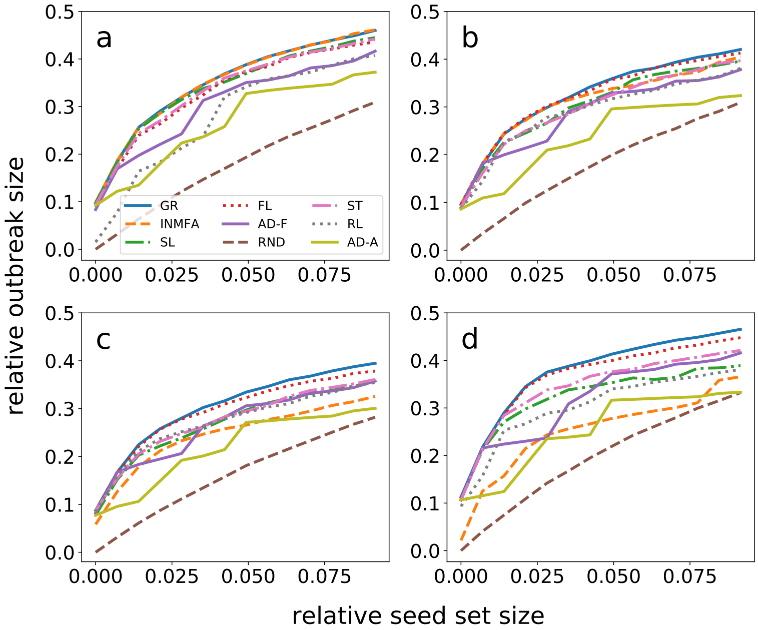

Figure S25: Same as Fig. S14, but for "Email, dept. 4" network in supercritical regime.

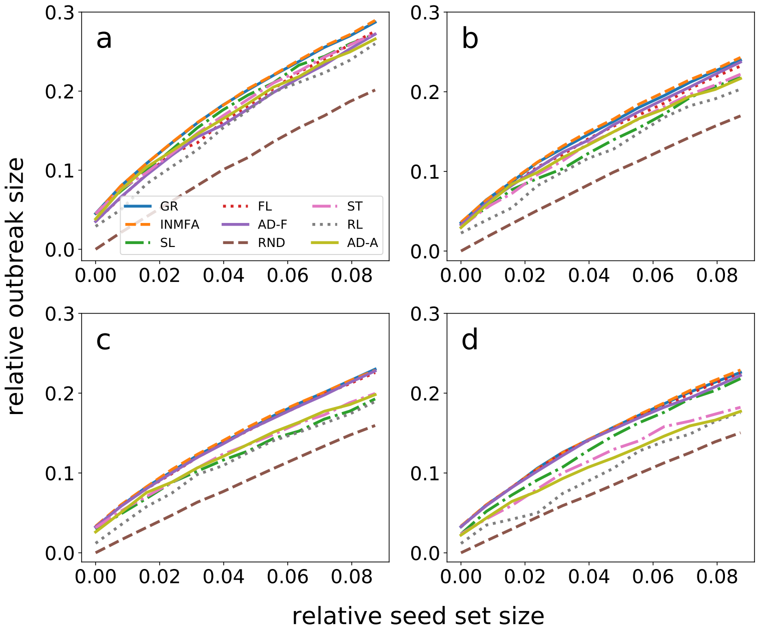

Figure S26: Same as Fig. S14, but for "High school, 2011" network in subcritical regime.

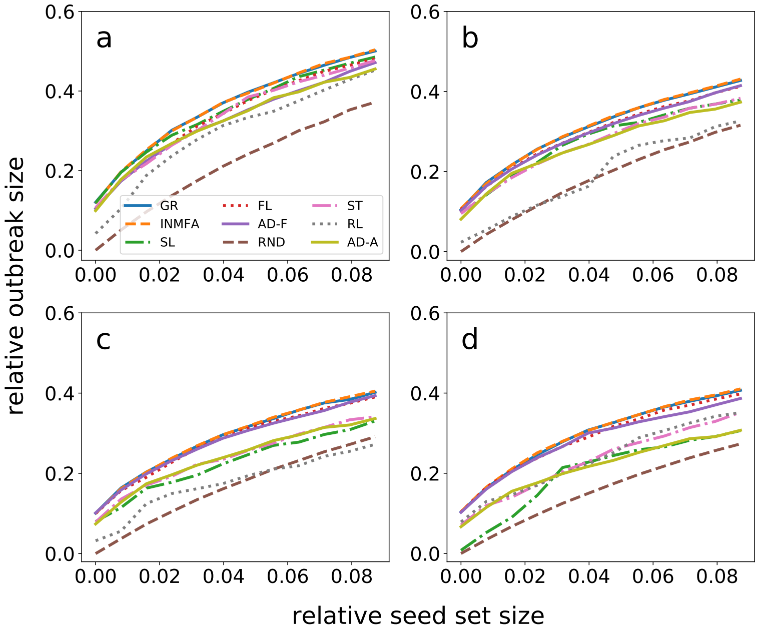

Figure S27: Same as Fig. S14, but for "High school, 2011" network in critical regime.

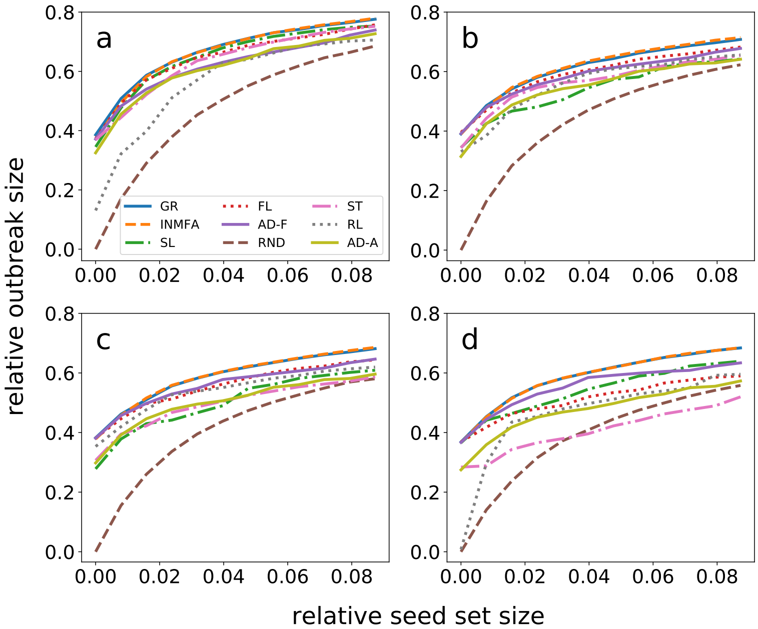

Figure S28: Same as Fig. S14, but for "High school, 2011" network in supercritical regime.

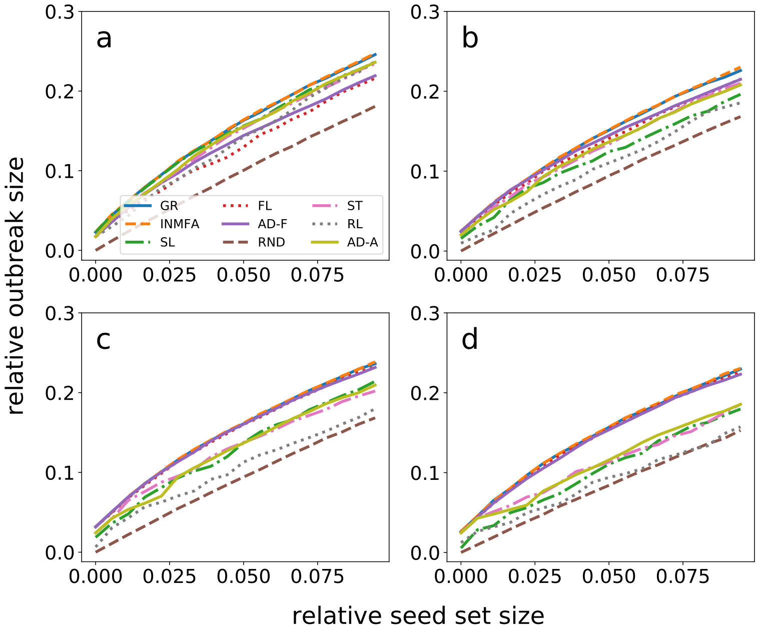

Figure S29: Same as Fig. S14, but for "High school, 2012" network in subcritical regime.

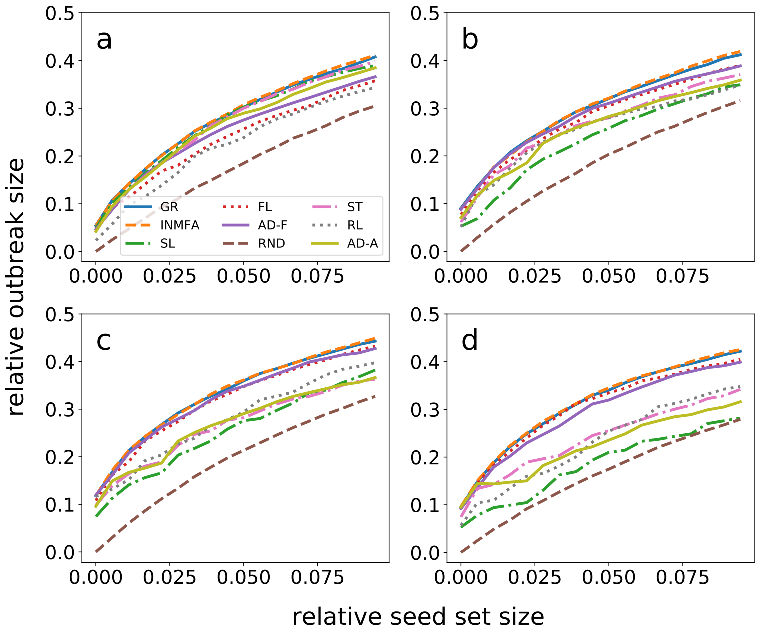

Figure S30: Same as Fig. S14, but for "High school, 2012" network in critical regime.

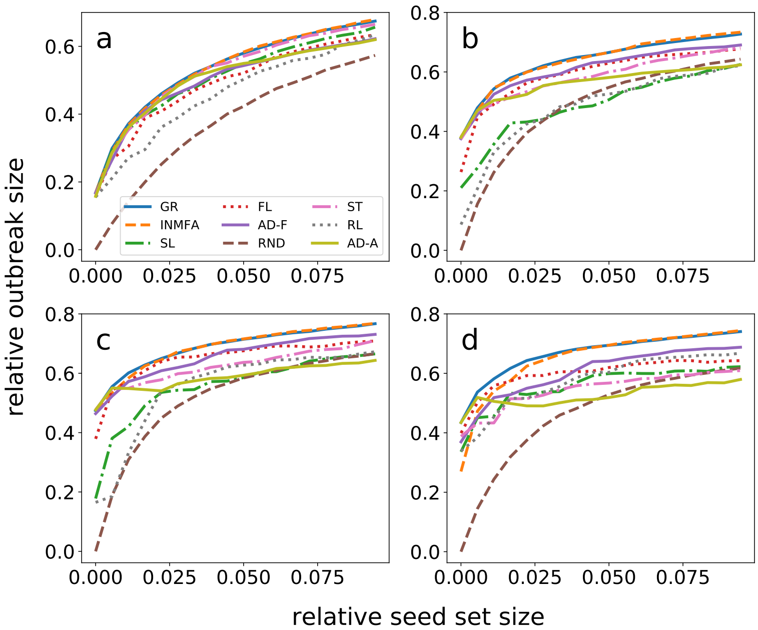

Figure S31: Same as Fig. S14, but for "High school, 2012" network in supercritical regime.

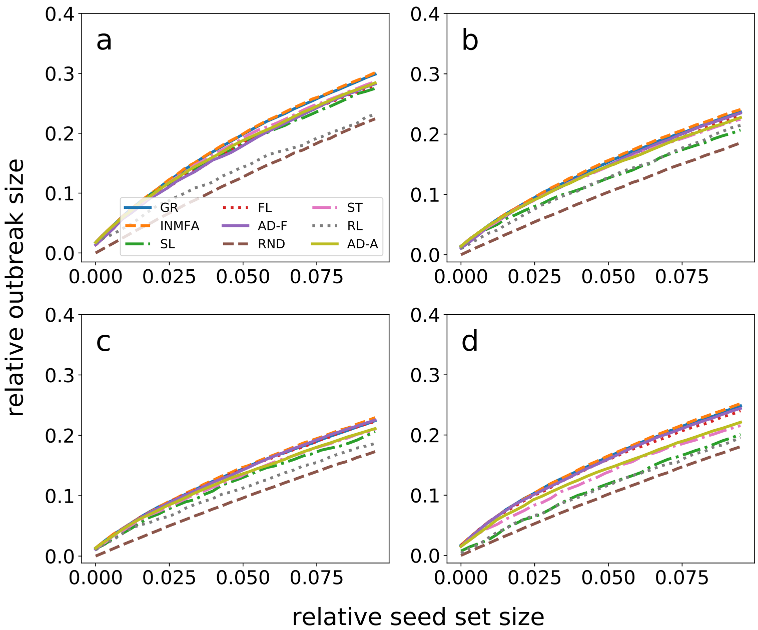

Figure S32: Same as Fig. S14, but for "High school, 2013" network in subcritical regime.

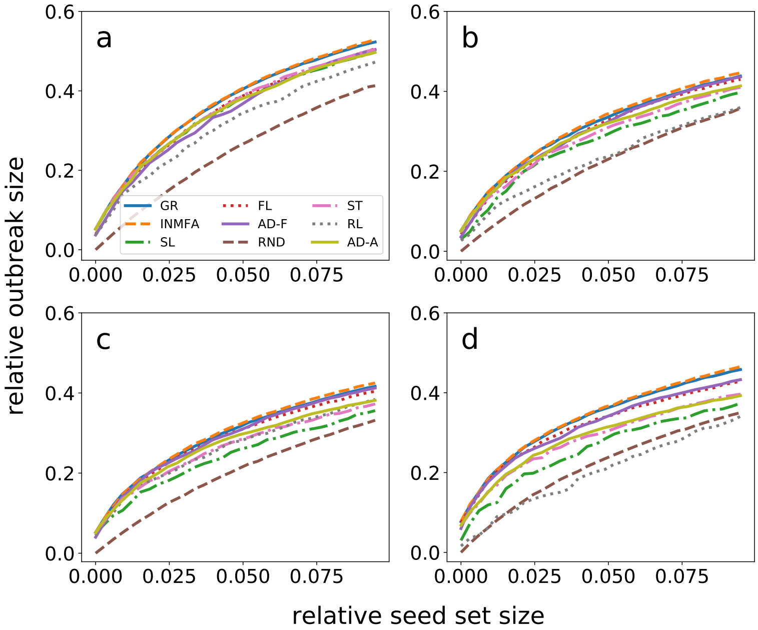

Figure S33: Same as Fig. S14, but for "High school, 2013" network in critical regime.

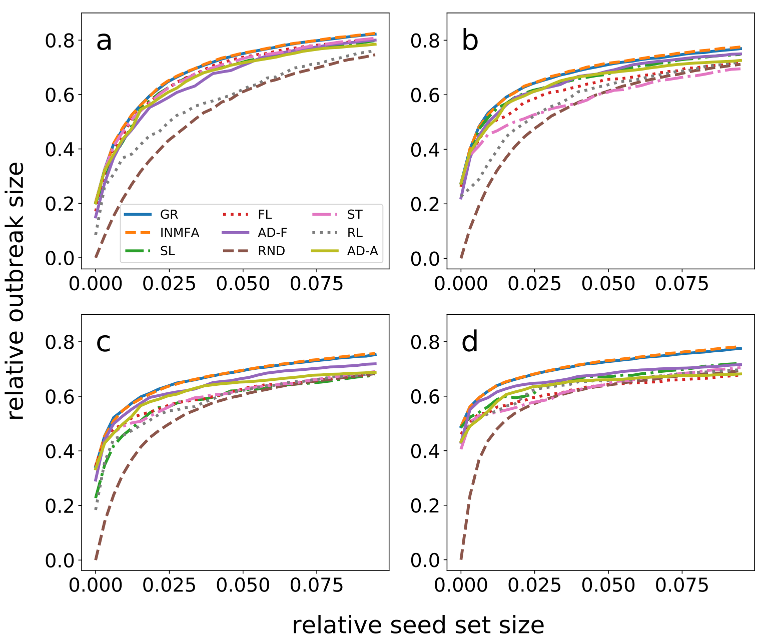

Figure S34: Same as Fig. S14, but for "High school, 2013" network in supercritical regime.

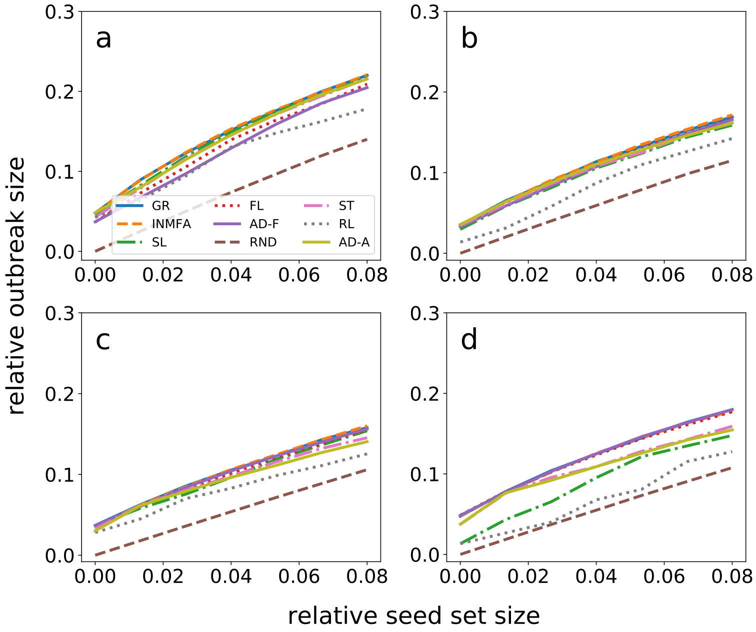

Figure S35: Same as Fig. S14, but for "Hospital ward" network in subcritical regime.

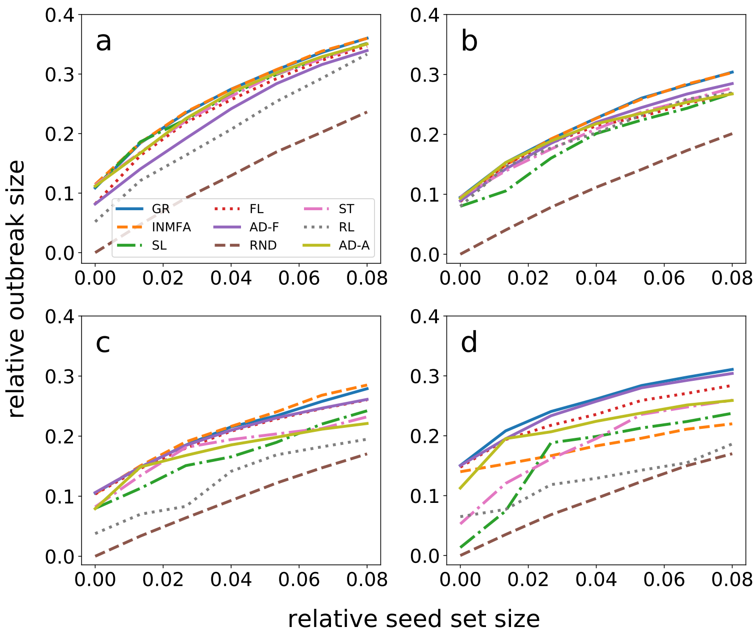

Figure S36: Same as Fig. S14, but for "Hospital ward" network in critical regime.

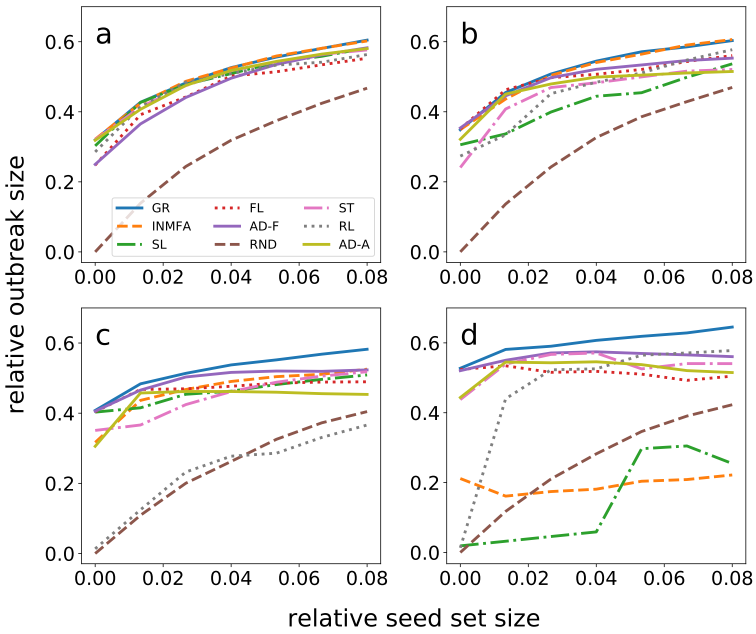

Figure S37: Same as Fig. S14, but for ”Hospital ward” network in supercritical regime.

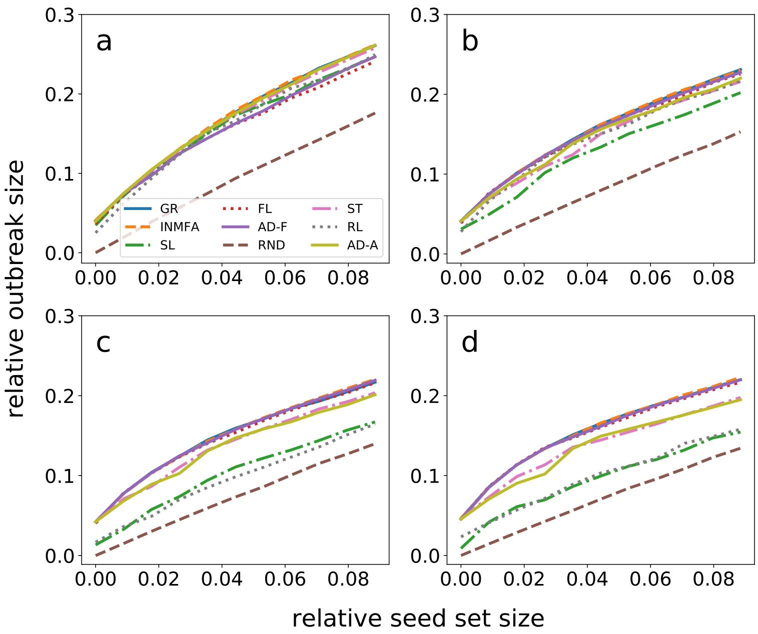

Figure S38: Same as Fig. S14, but for ”Hypertext, 2009” network in subcritical regime.

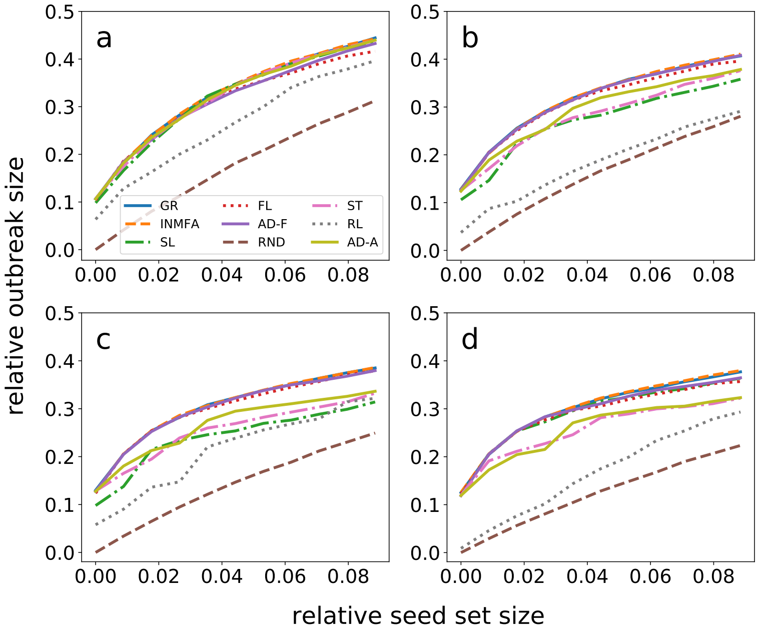

Figure S39: Same as Fig. S14, but for "Hypertext, 2009" network in critical regime.

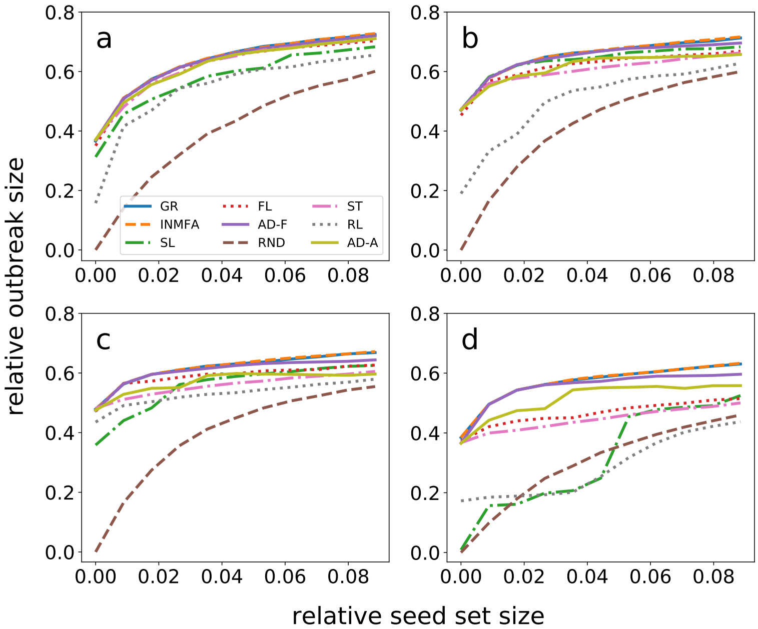

Figure S40: Same as Fig. S14, but for "Hypertext, 2009" network in supercritical regime.

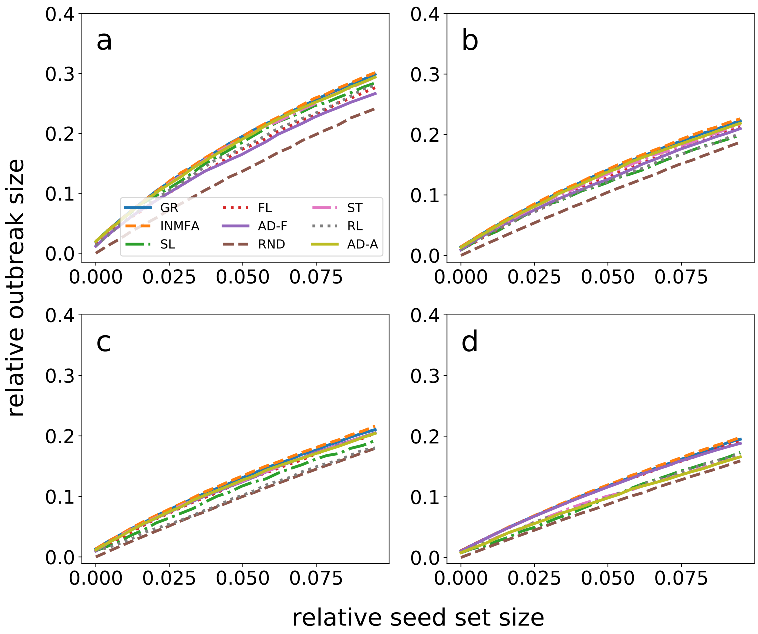

Figure S41: Same as Fig. S14, but for ”Primary school” network in subcritical regime.

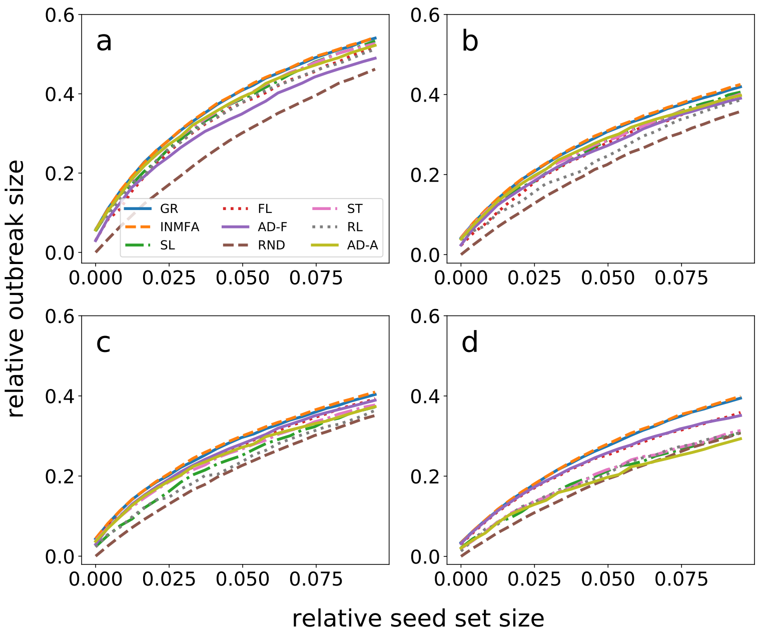

Figure S42: Same as Fig. S14, but for ”Primary school” network in critical regime.

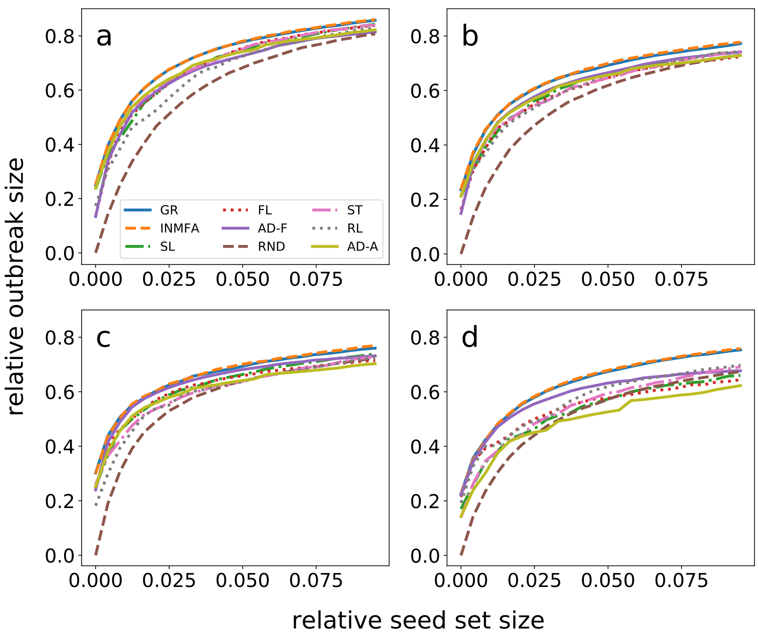

Figure S43: Same as Fig. S14, but for ”Primary school” network in supercritical regime.

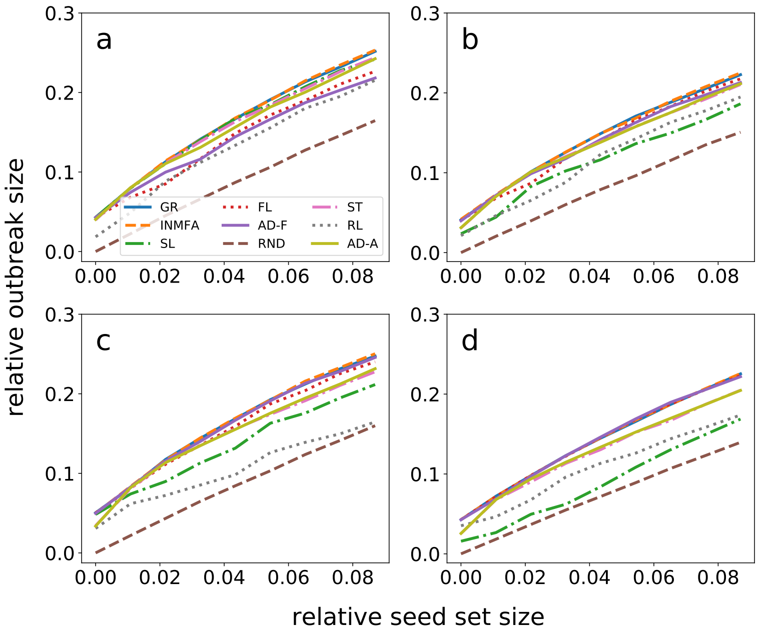

Figure S44: Same as Fig. S14, but for ”Workplace” network in subcritical regime.

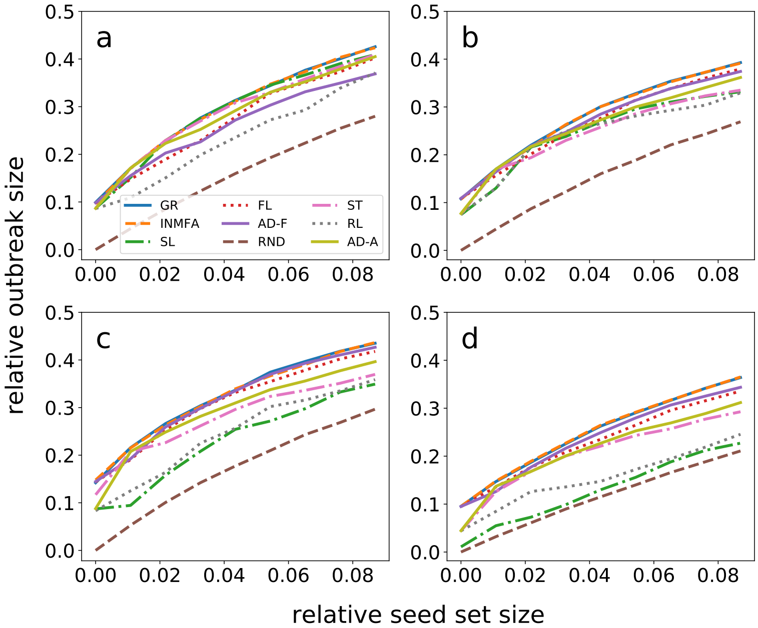

Figure S45: Same as Fig. S14, but for ”Workplace” network in critical regime.

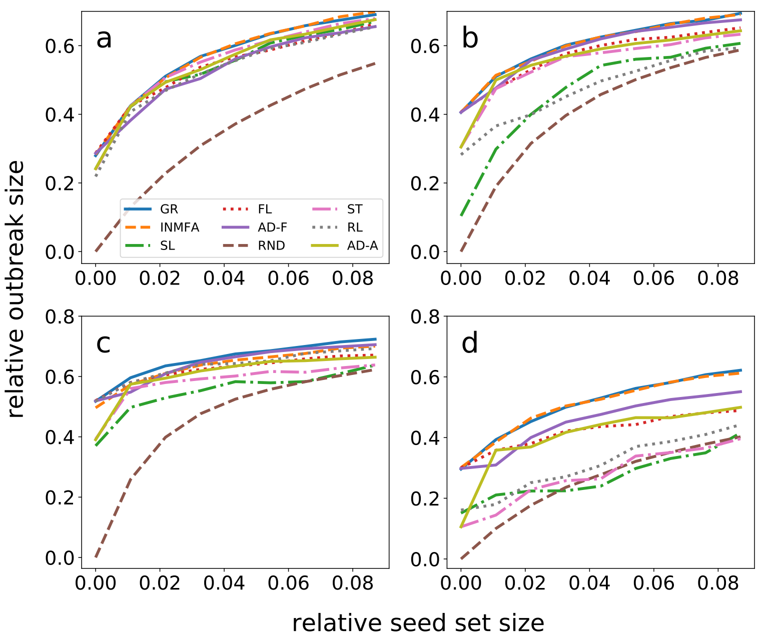

Figure S46: Same as Fig. S14, but for ”Workplace” network in supercritical regime.

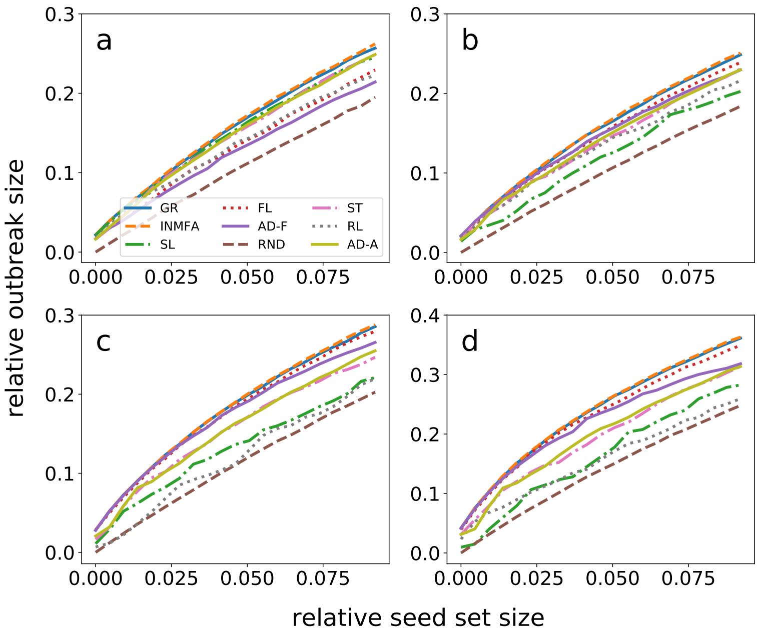

Figure S47: Same as Fig. S14, but for ”Workplace-2” network in subcritical regime.

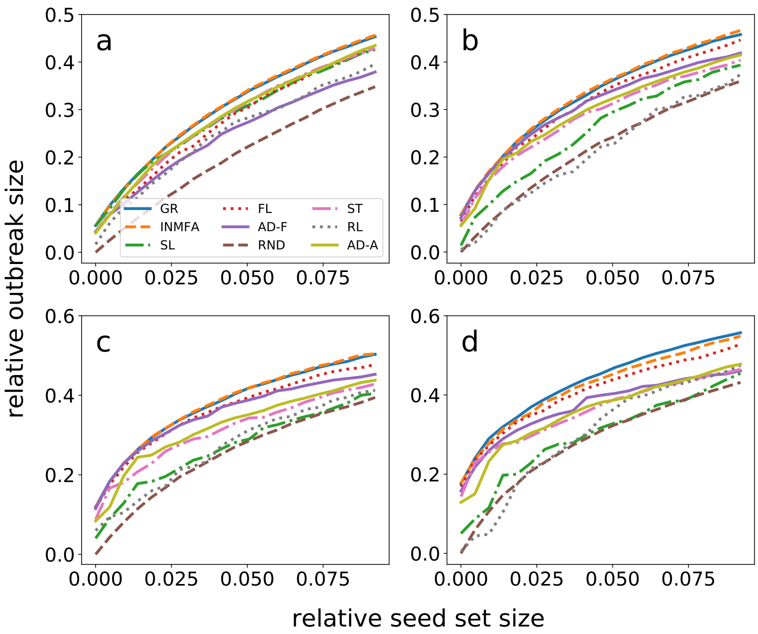

Figure S48: Same as Fig. S14, but for ”Workplace-2” network in critical regime.

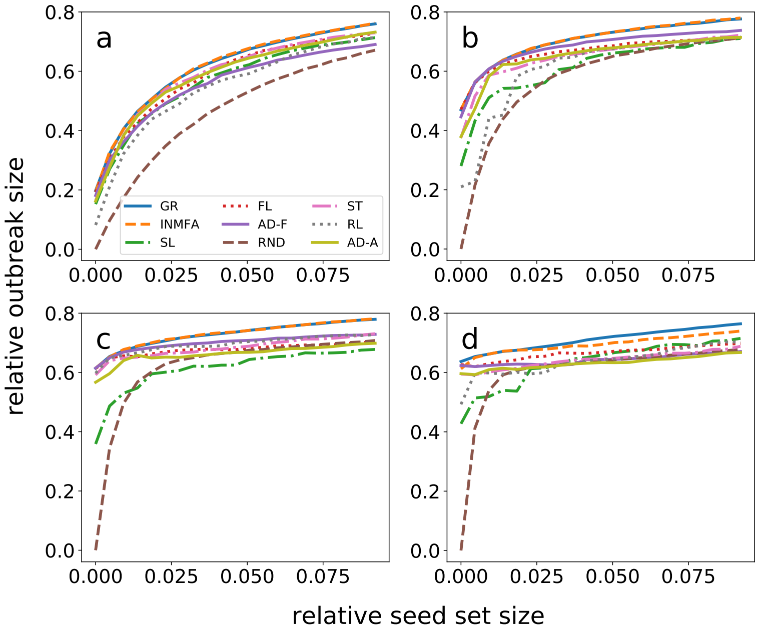

Figure S49: Same as Fig. S14, but for "Workplace-2" network in supercritical regime.