nature human behaviour

SUPPLEMENTARY INFORMATION
https://doi.org/10.1038/s41562-018-0346-z

In the format provided by the authors and unedited.

Social influence maximization under empirical influence models

Sinan Aral* and Paramveer S. Dhillon*

MIT Sloan School of Management, Cambridge, MA, USA. *e-mail: sinan@mit.edu; dhillon@mit.edu

Social Influence Maximization Under Empirical Influence Models

Supplementary Information

Sinan Aral and Paramveer S. Dhillon
e-mail:{sinan,dhillon}@mit.edu
MIT Sloan School of Management
100 Main Street, Cambridge, MA 02142, USA

Contents

1 Supplementary Methods

1.1 Models of Empirical Influence Maximization

Assume we are given a graph G(V, E) on a set of V nodes and E edges with a neighborhood matrix ($\mathbb{N}$) (which gives the list of all the neighbors of a given node 'i' in the graph as $\mathbb{N}(i)$).

Our main premise in this paper is that influence and susceptibility to influence are two separate concepts as has been shown by previous empirical work (Aral and Walker, 2012) and should be modeled separately and we are interested in seeing the impact of modeling the non-uniform joint distribution of influence and susceptibility on the outcome of social influence maximization. Is the final influence spread different? If yes, by how much and is it economically significant? Are the chosen seeds any different? If yes, can we characterize that difference based on some structural properties or some latent properties of the nodes?

Towards that end, we assume that each node 'i' has associated with it, two p-dimensional parameters $\Lambda_i = \{\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{ip}\}$ and $\Theta_i = \{\theta_{i1}, \theta_{i2}, \ldots, \theta_{ip}\}$ which represent that node's latent influence and susceptibility to influence respectively on p different dimensions. Modeling the correlation structure between the instantiations of these two parameters for an ego and its alters gives rise to eight possible configurations:

1. **AAC** (**A**ssortative Influence, **A**ssortative Susceptibility, **C**omplements Influence-Susceptibility) $\rho_{\Lambda_i \Lambda_{j\in\mathbb{N}(i)}} > 0, \rho_{\Theta_i \Theta_{j\in\mathbb{N}(i)}} > 0, \rho_{\Lambda_i \Theta_i} > 0 \quad \forall i \in V$
2. **AAS** (**A**ssortative Influence, **A**ssortative Susceptibility, **S**ubstitutes Influence-Susceptibility) $\rho_{\Lambda_i \Lambda_{j\in\mathbb{N}(i)}} > 0, \rho_{\Theta_i \Theta_{j\in\mathbb{N}(i)}} > 0, \rho_{\Lambda_i \Theta_i} \leq 0 \quad \forall i \in V$
3. **ADC** (**A**ssortative Influence, **D**isassortative Susceptibility, **C**omplements Influence-Susceptibility) $\rho_{\Lambda_i \Lambda_{j\in\mathbb{N}(i)}} > 0, \rho_{\Theta_i \Theta_{j\in\mathbb{N}(i)}} \leq 0, \rho_{\Lambda_i \Theta_i} > 0 \quad \forall i \in V$
4. **ADS** (**A**ssortative Influence, **D**isassortative Susceptibility, **S**ubstitutes Influence-Susceptibility) $\rho_{\Lambda_i \Lambda_{j\in\mathbb{N}(i)}} > 0, \rho_{\Theta_i \Theta_{j\in\mathbb{N}(i)}} \leq 0, \rho_{\Lambda_i \Theta_i} \leq 0 \quad \forall i \in V$

5. **DAC** (**D**isassortative Influence, **A**ssortative Susceptibility, **C**omplements Influence-Susceptibility) $\rho_{\Lambda_i \Lambda_{j\in\mathbb{N}(i)}} \leq 0, \rho_{\Theta_i \Theta_{j\in\mathbb{N}(i)}} > 0, \rho_{\Lambda_i \Theta_i} > 0 \quad \forall i \in V$

6. **DAS** (**D**isassortative Influence, **A**ssortative Susceptibility, **S**ubstitutes Influence-Susceptibility) $\rho_{\Lambda_i \Lambda_{j\in\mathbb{N}(i)}} \leq 0, \rho_{\Theta_i \Theta_{j\in\mathbb{N}(i)}} > 0, \rho_{\Lambda_i \Theta_i} \leq 0 \quad \forall i \in V$

7. **DDC** (**D**isassortative Influence, **D**isassortative Susceptibility, **C**omplements Influence-Susceptibility) $\rho_{\Lambda_i \Lambda_{j\in\mathbb{N}(i)}} \leq 0, \rho_{\Theta_i \Theta_{j\in\mathbb{N}(i)}} \leq 0, \rho_{\Lambda_i \Theta_i} > 0 \quad \forall i \in V$

8. **DDS** (**D**isassortative Influence, **D**isassortative Susceptibility, **S**ubstitutes Influence-Susceptibility) $\rho_{\Lambda_i \Lambda_{j\in\mathbb{N}(i)}} \leq 0, \rho_{\Theta_i \Theta_{j\in\mathbb{N}(i)}} \leq 0, \rho_{\Lambda_i \Theta_i} \leq 0 \quad \forall i \in V$

where $\rho_{x,y}$ represents the Pearson correlation between x, y and $\mathbb{N}(i)$ represents the list of all the neighbors of a given node 'i' in the graph.

Throughout our analysis our main objective is to do an apples-to-apples comparison with the extant state-of-the-art. So, we begin with the basic Independent Cascade (IC) and Linear Threshold (LT) models and precisely retain the standard assumptions of the influence maximization problem i.e. 1) greedy optimization, 2) the size of seed-set k, 3) and discrete time-dynamics. We do so to easily distinguish our contribution and to make sure that no confounds or other covarying facets of the problem can explain our results. Below, we show how to adapt the IC and LT influence propagation models to handle our models of Empirical Influence Maximization (EIM).

1.1.1 Adaptation of IC and LT Models

The two standard influence propagation models used in Influence Maximization research are:

- **Independent Cascade (IC) Model:** Cascade models were first introduced by (Goldenberg et al., 2001a,b) in the context of marketing. In the simplest of these models, the **Independent Cascade** (IC) model, each edge (u, v) in the graph is associated with a probability p_{uv} known as the "influence" probability. Given the "influence" probabilities and an initial set of active nodes, as each node u becomes active, it is given a single chance to activate each inactive neighbor w independently with probability p_{uw}. Typically, the edge probabilities are assumed to be fixed to some constant value e.g. 0.1 or they are assumed to be distributed uniformly at random $\mathcal{U}(0, 1)$ (which is supposed to indicate lack of knowledge) (Kempe et al., 2003; Chen et al., 2009; Wang et al., 2012).

- **Linear Threshold (LT) Model:** In the LT model each node v has a latent threshold δ_v and for every $u \in \mathbb{N}(v)$ (where $\mathbb{N}(\cdot)$ is the set of neighbors in the graph), (u, v) has a non-negative weight $w_{u,v}$ such that $\sum_{u\in\mathbb{N}(v)} w_{u,v} \leq 1$. Given the thresholds and an initial set of active nodes, the process unfolds deterministically in discrete time steps. At time 't', an inactive node v becomes active if $\sum_{u\in\mathbb{A}(v)} w_{u,v} \geq \delta_v$, where $\mathbb{A}(v)$ is the set of active neighbors of v up to time-step 't-1'. Once activated, a node stays active and the process terminates when no more activations are possible. In order for the resulting influence maximization problem to be submodular, the latent thresholds $\delta_v \sim \mathcal{U}(0, 1)$. The edge weights $w_{u,v}$ are either assumed to be distributed $\mathcal{U}(0, 1)$ or they are fixed inversely proportional to the in-degree of the node (Kempe et al., 2003; Chen et al., 2009; Wang et al., 2012).

As can be seen both the above class of models i.e. IC and LT assume the edge propagation probabilities to be specified in advance—either fixed to some constant value or chosen using some structural property of the network. However, the EIM models that we propose have two parameters per node. So, we generate edge propagation probabilities using the influence and susceptibility parameters of an ego 'i' and its alter 'j' for the edge pointing from 'i' towards 'j' as $p_{ij} = \frac{\Lambda_i^\top \cdot \Theta_j}{\|\Lambda_i\|\|\Theta_j\|}$. In other words, our model modifies an alter's (node 'j') susceptibility over and above its baseline susceptibility (Θ_j) by an amount linear and multiplicative in the ego's (node 'i') influence (Λ_i). There are several other conceptualizations possible with other functional forms e.g. functional forms that are additive or exponentiated and our analysis holds for them also. We chose to report the linear-multiplicative conceptualization as it's the simplest and most intuitive and also takes into account some of the boundary cases e.g. if an alter's susceptibility is a "hard" zero then no matter what the influence of the ego is, the final edge propagation probability will be zero.

1.1.2 Submodularity of the solution

(Kempe et al., 2003) showed that influence maximization under IC and LT models is submodular, hence a greedy algorithm would lead to $1 - \frac{1}{e}$ approximation guarantee. However, it remains to be shown that submodularity of influence maximization is preserved under our empirical models of influence maximization —in fact as we see below, the submodularity of our procedure follows from (Kempe et al., 2003).

First, let's consider the case of the IC model. Its proof of submodularity (as proved by (Kempe et al., 2003)) just requires that the edge propagation numbers be probabilities i.e. $0 \leq p_{u,v} \leq 1$ and that they are prespecified and do not change during the unfolding of the influence maximization process. It can easily be seen that both these conditions are clearly satisfied for our EIM models—the edge propagations ($p_{ij} = \frac{\Lambda_i^\top \cdot \Theta_j}{\|\Lambda_i\|\|\Theta_j\|}$) are probabilities since $\lambda_{i1}, \lambda_{i2}, \ldots, \theta_{j1}, \ldots, \theta_{jp} \in [0, 1]$ and they never change dynamically during the unfolding of the influence maximization process.

Next, let's consider the LT model. Once again, (Kempe et al., 2003) proved that its submodular. The proof of LT model's submodularity relies on the fact that (as with the IC model), the edge propagation numbers be probabilities i.e. $0 \leq w_{u,v} \leq 1$ and that they are prespecified and do not change during the unfolding of the influence maximization process. In addition to that it also requires that the node specific thresholds δ_v are also distributed $\mathcal{U}(0, 1)$ and that they can not be fixed. Hence, in our adaptation of the LT model, we keep the node thresholds distributed $\mathcal{U}(0, 1)$ but fix the edge propagation numbers based on the EIM models.

Hence, our adaptations of both the IC and LT models preserve the submodularity of the solution.

1.1.3 Baseline Models

There have been a variety of models that have been proposed in previous work to assign the edge propagation probabilities on the underlying graph. The most commonly used heuristics are assigning the probabilities randomly, fixing them at some arbitrary level e.g. 0.1 or choosing them based on some structural properties of the network e.g. inversely proportional to the in-degree of the node.

In this work we benchmark the performance of our models against two of the most commonly used heuristics. First, we compare against the model which chooses the influence and susceptibility parameters both uniformly at random from the interval [0,1] and then takes the normalized inner product of the corresponding influence and susceptibility parameters ($= \frac{\Lambda_i^\top \cdot \Theta_j}{\|\Lambda_i\| \|\Theta_j\|}$) to get the final edge propagation probabilities. We call this model "Random" in the Results section.

Next, we compare against the model which assigns the edge propagation probabilities based on the heuristic of picking a fixed number 0.1 for all the edges for the IC model. For the LT model, the same heuristic of picking a fixed number reduces to choosing the edge weights based on the inverse of the in-degree of the node, since each node's incoming weights sum to 1. So, even in the case of LT model all the edge weights are constant but they are all equal to the inverse of the in-degree of the node. In the results section, we call this Baseline, "Constant" for the IC model and "Inv-Deg" for the LT model.

The IC model does not perform normalization of edge weights like the LT model does, which ensures that each node's incoming edge weights sum to 1. So, it becomes imperative to ensure that the sum of all the edge weights in the IC model is comparable for all the EIM and Baseline models. We ensure this by normalizing all the edge weights in all the graphs for the IC model to sum upto $0.1 \times number\ \ of\ \ nodes$.

1.2 Generating graphs with Correlations

In this section we describe the data generating process to generate the influence and susceptibility parameters for the graphs satisfying the above (dis)-assortativity assumptions. While performing our labeling we do not alter the structure of the graphs in any way—we just label it with influence and susceptibility parameters. Algorithm 1 describes the procedure. It is one such generative process and there are other possibilities.

The iterative graph labeling procedure outlined in Algorithm 1 borrows ideas from (Ugander and Backstrom, 2013) and (Raghavan et al., 2007) and labels the graph with a binary labeling 0/1 for both a node's influence and susceptibility. The labels can be thought to correspond to "Low" and "High" types for both influence and susceptibility. It gives us four possibilities—an individual node can be:

- Low Influence, Low Susceptibility.
- Low Influence, High Susceptibility.
- High Influence, Low Susceptibility.
- High Influence, High Susceptibility.

Now, it can readily be seen how the eight models of empirical influence maximization map onto these four possibilities. For example, the graph whose influence and susceptibility parameters follow the correlations of the model **AAC** (Assortative Influence, Assortative

Susceptibility, Complements Influence and Susceptibility) will have majority of nodes (chosen by a prespecified parameter $fracComplements$ and fixed at 0.75 for our simulation results.) with their own influence and susceptibility parameters as complements i.e. either both "High" or both "Low". It will also have more nodes that are connected to nodes whose influence parameter is same type as their—both "High" or both "Low", since this model has assortative influence. The same holds for the susceptibility parameter for the AAC model.

The algorithm for the graph labeling (Algorithm 1) is really straightforward. It starts by initializing the graph with majority ($fracComplements = 0.75$) of nodes having the desired correlation between their own influence and susceptibility, for instance in the case of AAC, we start with 75% of the nodes having their own influence and susceptibility as complements. Next, we want to enforce the assortativity/disassortativity patterns with the node's neighbors. So, depending on whether we want to enforce assortativity (or disassortativity) we assign the relevant influence/susceptibility label as the majority (or minority) label observed in that node's neighborhood respectively [1]. Once, we have the new labels for all the nodes we calculate the number of nodes that would move from "Complements" to "Substitutes" and vice-versa with the new labeling. Now, we could relabel all those nodes with their new labels, however in order to preserve stability of the labeling as there is the possibility of degenerate labelings i.e. all nodes of one type, we take a more conservative approach and move exactly the same number of nodes from "Complements" to "Substitutes" as we do from "Substitutes" to "Complements" as this roughly preserves the proportions of "Complements" and "Substitutes" in the network[2] . This process is repeated for a few iterations (typically 10) or till when there is no change in the objective function.

Note that in Algorithm 1 we have the same fixed sequence of random coin flips for all the 8 models, but distributed on the graph in different ways according to whether it's AAC or ADC etc. In other words, while using Algorithm 1 if one were to flip a coin while generating labels for the 8 different models you would get the same sequence of heads/tails. Hence, the only difference in the edge propagation probabilities in the different graphs arises due to the variation in assortativity/disassortativity assumptions.

The Figure 4 in the main paper shows the graphs generated by different models AAS, ADS and DDS and its easy to discern how the different correlation patterns interact with the structure of the network and ultimately lead to different set of optimal seeds.

Real-valued Influence and Susceptibility Parameters: The graph labeling generated by Algorithm 1 is binary (0/1), so we need a mechanism to generate real-valued parameter values for Influence and Susceptibility parameters. We accomplish that simply by conditioning on the type of the node i.e. "High" (1) or "Low" (0) and drawing samples from "well separated" Beta distributions.

We sample real-valued numbers for "High" and "Low" types from well separated Beta Distributions with means 0.75 and 0.25 respectively (shown in Supplementary Figure 1). We tried other parameter values of the Beta Distribution but the results were not significantly different. We chose Beta Distribution mainly due to convenience as its random variates

[1] This is akin to the community detection algorithms (Raghavan et al., 2007) which assign the majority neighborhood community label to a node.

[2] https://code.facebook.com/posts/274771932683700/large-scale-graph-partitioning-with-apache-giraph/

Algorithm 1 Generating Graphs with Correlated Binary Influence & Susceptibility Attributes

procedure CORRELATED-LABEL-PROPAGATION($\mathbb{G}, numIterations, fracComplements, setting$)
▷ fracComplements is the fraction of nodes whose influence and susceptibility labels are complements i.e. 0,0 or 1,1.
  if setting == (AAC ‖ ADC ‖ DAC ‖ DDC) **then**
    fPositive ← fracComplements.
  if setting ==(AAS ‖ ADS ‖ DAS ‖ DDS) **then**
    fPositive ← 1-fracComplements.
  for each node $i \in$ V($\mathbb{G}$) **do**
    Assign initial labels probabilistically i.e.
    $[i.inf \leftarrow 0, i.susc \leftarrow 0, i.group \leftarrow C]$, with prob. fPositive/2.
    $[i.inf \leftarrow 1, i.susc \leftarrow 1, i.group \leftarrow C]$ with prob. fPositive/2.
    $[i.inf \leftarrow 0, i.susc \leftarrow 1, i.group \leftarrow S]$ with prob. (1-fPositive)/2.
    $[i.inf \leftarrow 1, i.susc \leftarrow 0, i.group \leftarrow S]$ with prob. (1-fPositive)/2.

  for j ∈ 1:numIterations **do**
▷ Iteratively propagate the labels based on the labels of their neighbors.
    $objectiveFn_j$ =0
    Arrange the nodes V($\mathbb{G}$) in a random order $O_j(\mathbb{G})$.
    for each node $k \in O_j(\mathbb{G})$ **do**
      if (setting == (AAC ‖ AAS) **then**
        $k.inf_j \leftarrow$ Majority Influence-Label among neighbors of k.
        $k.susc_j \leftarrow$ Majority Susceptibility-Label among neighbors of k.
      if setting == (DAC ‖ DAS) **then**
        $k.inf_j \leftarrow$ Minority Influence-Label among neighbors of k.
        $k.susc_j \leftarrow$ Majority Susceptibility-Label among neighbors of k.
      if setting == (ADC ‖ ADS) **then**
        $k.inf_j \leftarrow$ Majority Influence-Label among neighbors of k.
        $k.susc_j \leftarrow$ Minority Susceptibility-Label among neighbors of k.
      if setting == (DDC ‖ DDS) **then**
        $k.inf_j \leftarrow$ Minority Influence-Label among neighbors of k.
        $k.susc_j \leftarrow$ Minority Susceptibility-Label among neighbors of k.
      if ($k.inf_j$ ==0 & $k.susc_j$ ==0) ‖ ($k.inf_j$ ==1 & $k.susc_j$ ==1) **then**
        $k.group_j \leftarrow$ C.
        $objectiveFn_j$ += Net change in # of Assort edges.
      if ($k.inf_j$ ==0 & $k.susc_j$ ==1) ‖ ($k.inf_j$ ==1 & $k.susc_j$ ==0) **then**
        $k.group_j \leftarrow$ S.
        $objectiveFn_j$ += Net change in # of Disassort edges.
    switchersCS ← # nodes ($group_{j-1}$ == C & $group_j$ == S).
    switchersSC ← # nodes ($group_{j-1}$ == S & $group_j$ == C).
    changeLabelsPreserveBalance ← $min(switchersSC, switchersCS)$.
  ▷ For all the nodes whose labels are candidates to be changed, change the labels of only the top $changeLabelsPreserveBalance$ nodes.
  binaryGraphLabeling ← $\mathbb{G}(max_j(objectiveFn_j))$ ▷ Select the graph labeling with the highest objective function value.
  return binaryGraphLabeling.

have support over [0,1] hence can be interpreted as probabilities and its flexibility to model different shapes of the density function. The algorithm is described in Algorithm 2.

Algorithm 2 Generate Final Graph Labeling

procedure GENERATE-FINAL-LABELING($binaryGraphLabeling$, μ_0, μ_1, σ_0, σ_1)
  for each $p \in binaryGraphLabeling$ **do**
    ▷ Conditionally sample from Beta Distribution.
    if p==0 **then**
      p ← Beta(μ_0, σ_0). ▷ $\mu_0 = 0.25$, $\sigma_0 = 0.1$. It corresponds to shape parameter $r = 4.4375$ and scale parameter $\alpha = 13.3125$
    if p==1 **then**
      p ← Beta(μ_1, σ_1). ▷ $\mu_0 = 0.75$, $\sigma_0 = 0.1$. It corresponds to shape parameter $r = 13.3125$ and scale parameter $\alpha = 4.4375$
  finalGraphLabeling ← binaryGraphLabeling.
  return finalGraphLabeling.

1.2.1 Alternative Methods for generating graphs with correlations

The correlated label propagation procedure that we used for graph generation gives robust results across different random draws, but there are other statistical methods for generating graphs with correlated attributes like the ones that we desire. One obvious way to generate such correlations among the attributes is by specifying the adjacency matrix of the graph as the precision matrix of a multivariate normal distribution and drawing random samples from that. However, we noticed that allowing the sparsity pattern of the precision matrix to be given by the graph adjacency matrix constrains the magnitudes and the directions of correlations severely for the datasets in our sample and we were not able to get reasonable sized correlations. Moreover, it is computationally challenging to generate such draws in the first place owing to having to compute the singular value decomposition of a huge graph adjacency matrix.

Attributed Graph Models (AGM) (Pfeiffer III et al., 2014) is another approach for generating graphs with correlated attributes, however unlike our setting they generate both graphs and attributes. So, in order to test the robustness of our findings we experimented with Preferential Attachment graphs on 10000 vertices with two attributes generated by AGM.

We used a Naïve Bayes model for attributes as suggested by the authors (`http://nld.cs.purdue.edu/agm/`), but we conditioned a node's attributes on its neighbors' labels to get the 8 assortative/disassortative patterns that we desired. The proposal distribution for the degree distribution was the Fast Chung-Lu (FCL) distribution. The results are shown in Supplementary Figures 2, 3, 4 5, 6, 7. As can be seen, the relative performance of various methods is the same as for the graphs generated using correlated label propagation, which attests the robustness of our results.

2 Supplementary Note 1

2.1 Data

We show results on 12 (6 synthetic and 6 real-world) datasets and where the influence/susceptibility attributes were generated via correlated label propagation. All the real-world datasets are the ones that are commonly used in previous work in the influence maximization and were downloaded from `http://snap.stanford.edu`. We preprocessed the datasets to keep only the largest weakly connected component of vertices, removed all the repeated/loopy edges and made the graphs undirected. The summary statistics of all the datasets are shown in Supplementary Table 1.

2.1.1 Diagnostics: Correlation patterns in generated labelings.

After generating the influence and susceptibility parameters, we perform diagnostics to check the observed correlations between the parameter values and if they actually follow the patterns specified by the 8 EIM models. For instance, for the AAC model, the correlation between a node's influence parameter and its neighbor's influence parameter $\rho(\lambda_i, \lambda_{j\in\mathbb{N}(i)})$, the correlation between a node's susceptibility parameter and its neighbor's susceptibility parameter $\rho(\theta_i, \theta_{j\in\mathbb{N}(i)})$ and the correlation between a node's influence parameter and its susceptibility parameter $\rho(\lambda_i, \theta_i)$ should all be positive (where ρ represents Pearson Correlation).

Supplementary Figures 8, 9 and 10 show these correlation patterns for all the 12 datasets for 10 i.i.d random draws of the parameters. As can be seen the correlation patterns broadly follow the specification of the various Empirical Influence Maximization models.

2.2 Influence Maximization Algorithm

Once we have generated the above graphs, the influence maximization procedure is straightforward. We use the recently proposed Two Phase Influence Maximization (TIM) algorithm (Tang et al., 2014) for influence maximization under both the adapted IC and LT models. The number of seeds was set to 100 and the epsilon parameter set to 0.1 as suggested by the authors of the paper. TIM uses reverse reachability and reverse influence sampling to significantly improve the computational complexity of finding the optimal seed-set that will maximize the influence spread. We did not use the Cost Effective Lazy Forward (CELF) (Leskovec et al., 2007) or CELF++ (Goyal et al., 2011) for the influence maximization algorithm as being simulation based, they do not scale to networks of sizes that we consider in this paper.

2.3 Results

Supplementary Figures 11- 58 show the influence spread, seed characteristics and comparison of different EIM models for both IC and LT influence propagation models for all the 12 datasets.

Supplementary References

Sinan Aral and Dylan Walker. Identifying influential and susceptible members of social networks. *Science*, page 1215842, 2012.

Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 199–208. ACM, 2009.

Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223, 2001a.

Jacob Goldenberg, Barak Libai, and Eitan Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 9(3):1–18, 2001b.

Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*, pages 47–48. ACM, 2011.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429. ACM, 2007.

Joseph J Pfeiffer III, Sebastian Moreno, Timothy La Fond, Jennifer Neville, and Brian Gallagher. Attributed graph models: Modeling network structure with correlated attributes. In *Proceedings of the 23rd international conference on World wide web*, pages 831–842. ACM, 2014.

U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3), 2007.

Youze Tang, Xiaokui Xiao, and Yanchen Shi. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, SIGMOD '14, pages 75–86, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2376-5. doi: 10.1145/2588555.2593670. URL `http://doi.acm.org/10.1145/2588555.2593670`.

Johan Ugander and Lars Backstrom. Balanced label propagation for partitioning massive graphs. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 507–516, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1869-3. doi: 10.1145/2433396.2433461. URL `http://doi.acm.org/10.1145/2433396.2433461`.

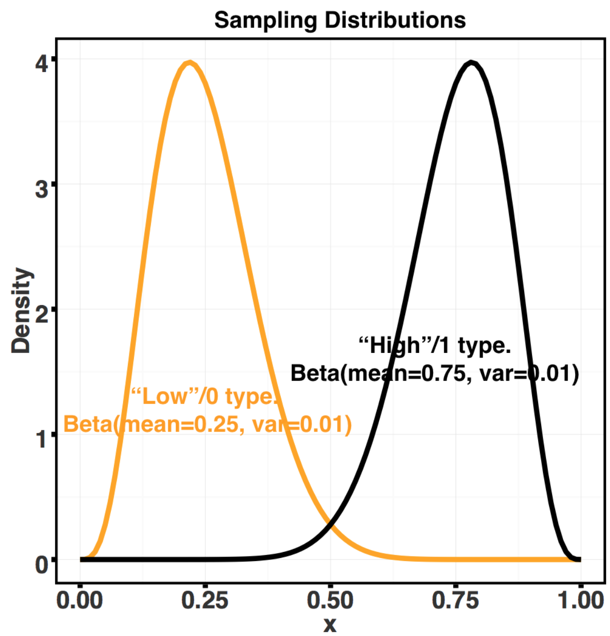
Chi Wang, Wei Chen, and Yajun Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25 (3):545–576, 2012.

Supplementary Tables

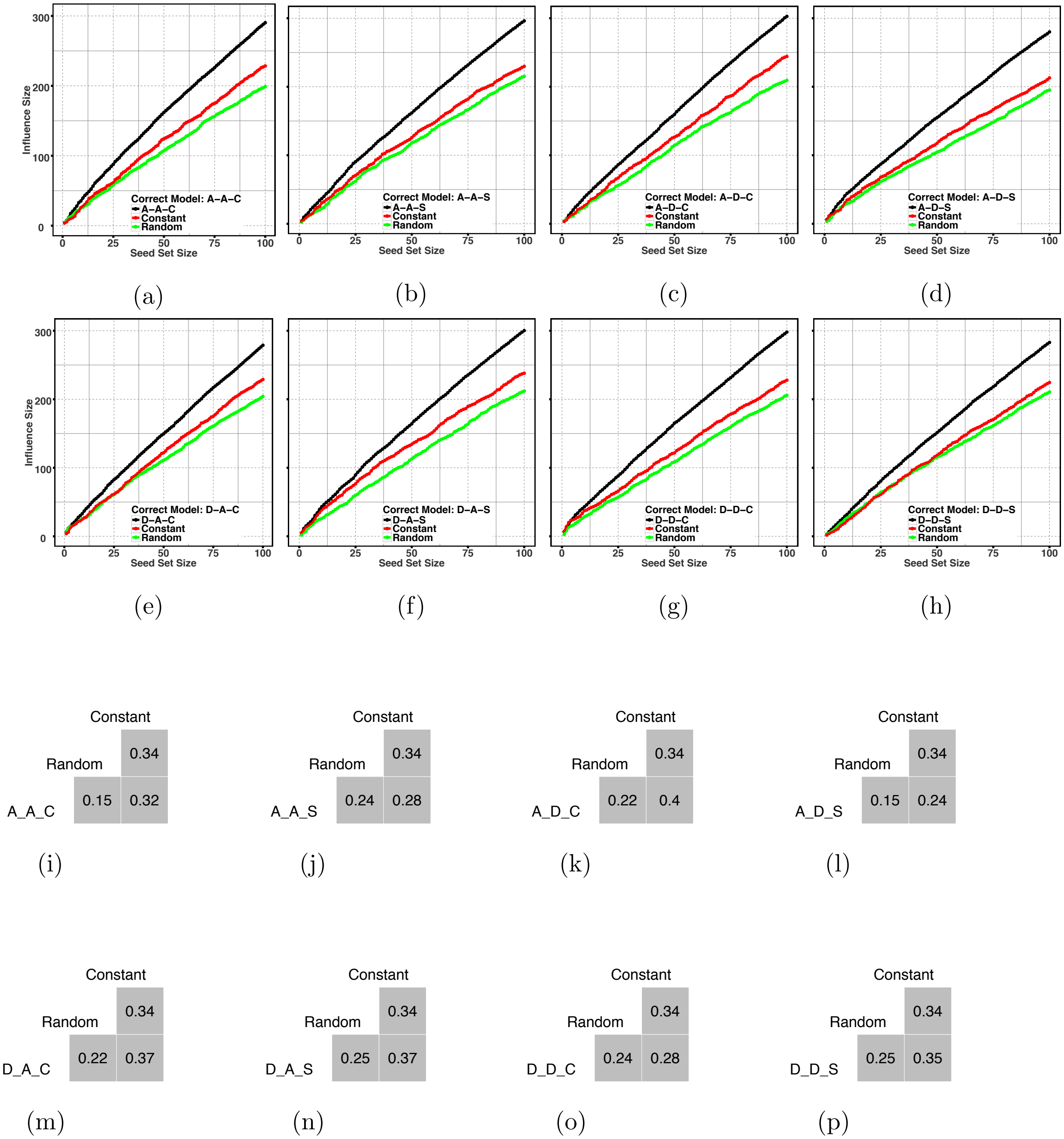
Dataset	URL	#Nodes/#Edges	Degree (Mean/Min./Max.)	Closeness Centrality (Mean/Min./Max.)	Transitivity (Frac. closed Δ)
1. Small World ($p_{rewire} = 0.05$) (SWP05)	N/A	10000/50000	10/6/14	0.16/0.14/0.18	0.49
2. Small World ($p_{rewire} = 0.1$) (SWP1)	N/A	10000/50000	10/4/16	0.19/0.16/0.21	0.35
3. Small World ($p_{rewire} = 0.15$) (SWP15)	N/A	10000/50000	10/4/17	0.20/0.18/0.22	0.24
4. Small World ($p_{rewire} = 0.2$) (SWP2)	N/A	10000/50000	10/4/18	0.21/0.19/0.23	0.17
5. Preferential Attachment (Exponent=0.5) (PAP5)	N/A	10000/9999	2/1/20	0.08/0.04/0.14	0
6. Preferential Attachment (Exponent=1) (PA1)	N/A	10000/9999	2/1/51	0.09/0.05/0.16	0
7. Collaboration network Arxiv High Energy Physics- Phenomenology (CA-HEPPH)	`http://snap.stanford.edu/data/ca-HepPh.html`	11204/117619	21/1/491	0.22/0.11/0.33	0.66
8. Collaboration network Arxiv High Energy Physics- Theory (CAHEPTH)	`http://snap.stanford.edu/data/ca-HepTh.html`	8638/24806	5.8/1/65	0.17/0.08/0.25	0.28
9. Citation network Arxiv High Energy Physics- Phenomenology (CITHEPPH)	`http://snap.stanford.edu/data/cit-HepPh.html`	34401/420784	24.5/1/846	0.24/0.12/0.33	0.15
10. Citation network Arxiv High Energy Physics- Theory (CITHEPTH)	`http://snap.stanford.edu/data/cit-HepTh.html`	27400/352021	25.7/1/2468	0.24/0.11/0.37	0.12
11. Epinions (Who trusts whom)	`http://snap.stanford.edu/data/soc-Epinions1.html`	75877/405739	10.7/1/3044	0.24/0.1/0.38	0.07
12. Slashdot November 2008	`http://snap.stanford.edu/data/soc-Slashdot0811.html`	77360/469180	12.1/1/2539	0.25/0.13/0.38	0.02

Supplementary Table 1: Summary Statistics of various datasets.

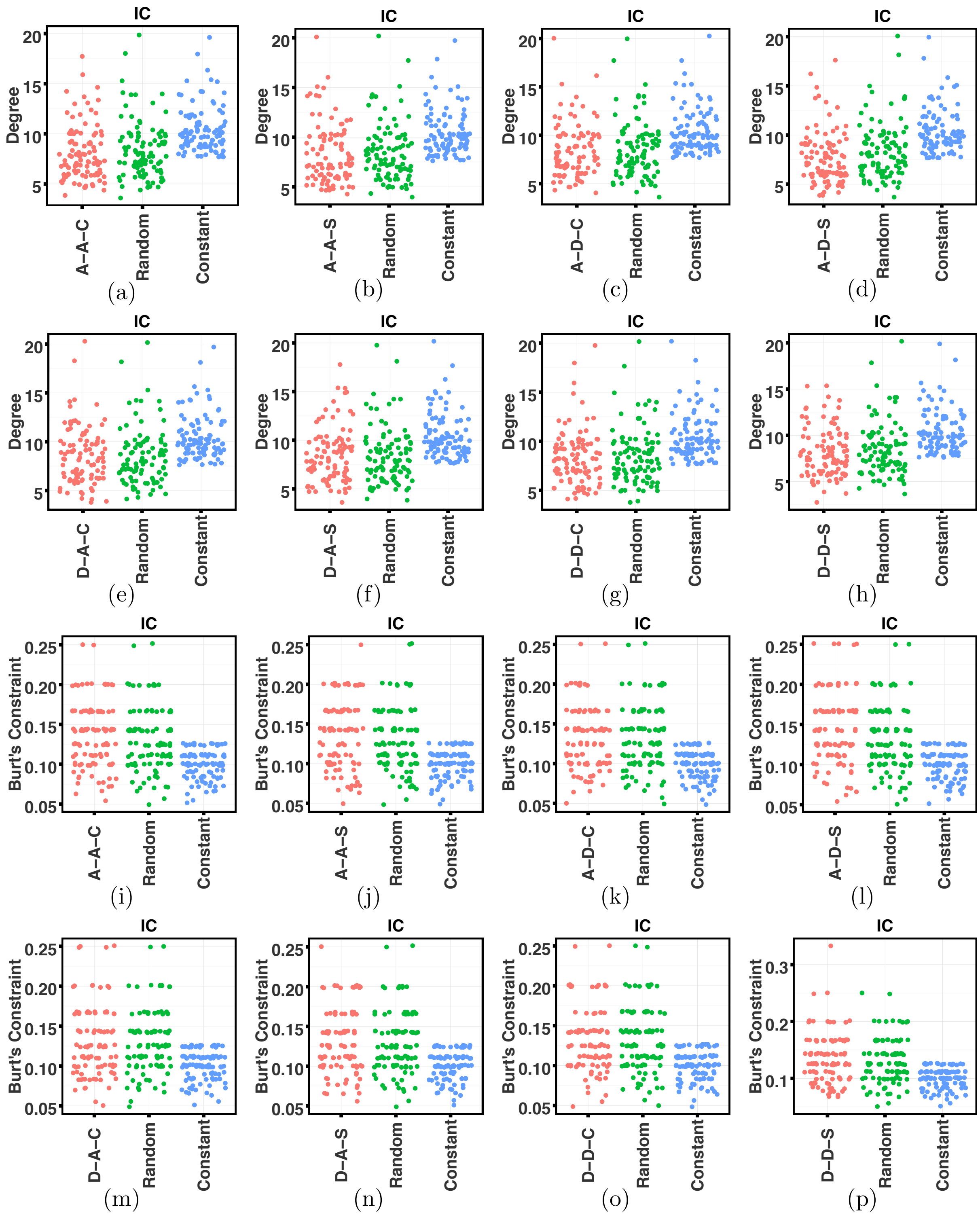
Supplementary Figures



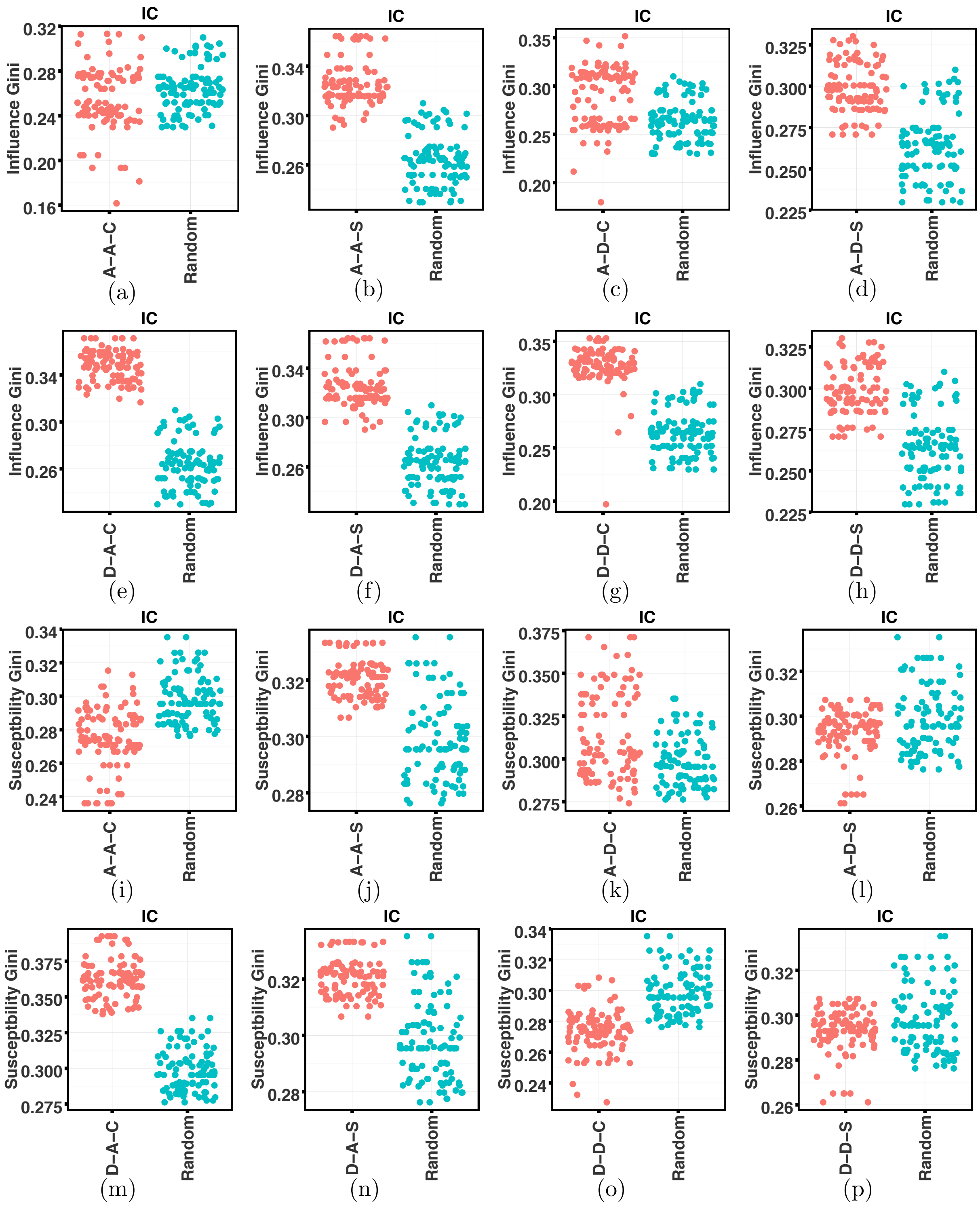
Supplementary Figure 1: Density plot of the Beta Distributions.



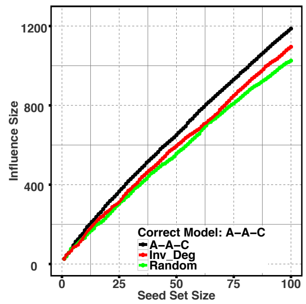
Supplementary Figure 2: Influence spread as a function of the seed set size and seed overlap for graphs generated using AGM. Influence propagation model is IC with p=0.1. **Network**: Preferential Attachment, Nodes=10000, Edges = {AAC=49970, AAS=49968, ADC=49975, ADS=49973, DAC=49970, DAS=49963, DDC=49966, DDS=49970}. **Note:** AGM generates both graphs and attributes, therefore, the graphs generated under different assumptions have different number of edges. Hence, the baselines i.e. Random and Constant are also different for the different assortative/disassortative assumptions.



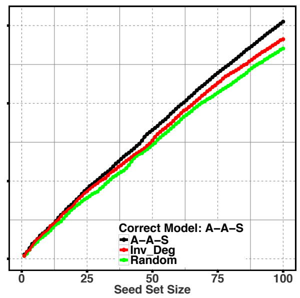
Supplementary Figure 3: Distribution of degree and Burt's constraints of the seed nodes chosen for graphs generated using AGM. Influence propagation model is IC with p=0.1. **Network**: Preferential Attachment, Nodes=10000, Edges = {AAC=49970, AAS=49968, ADC=49975, ADS=49973, DAC=49970, DAS=49963, DDC=49966, DDS=49970}. **Note:** AGM generates both graphs and attributes, therefore, the graphs generated under different assumptions have different number of edges. Hence, the baselines i.e. Random and Constant are also different for the different assortative/disassortative assumptions.



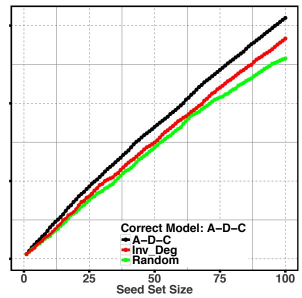
Supplementary Figure 4: Distribution of the Gini coefficient for the influence and susceptibility parameters of the seeds nodes chosen for graphs were generated using AGM. Influence propagation model is IC with p=0.1. **Network**: Preferential Attachment, Nodes=10000, Edges = {AAC=49970, AAS=49968, ADC=49975, ADS=49973, DAC=49970, DAS=49963, DDC=49966, DDS=49970}. **Note:** AGM generates both graphs and attributes, therefore, the graphs generated under different assumptions have different number of edges. Hence, the baselines i.e. Random and Constant are also different for the different assortative/disassortative assumptions.



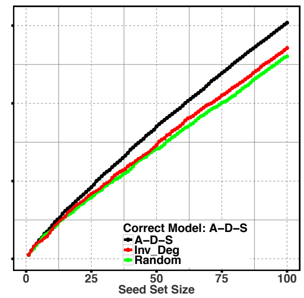
(a)



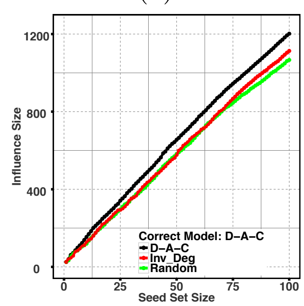
(b)



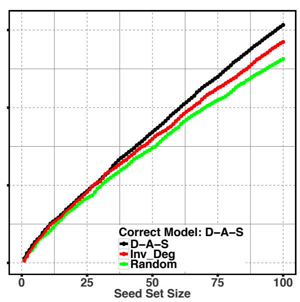
(c)



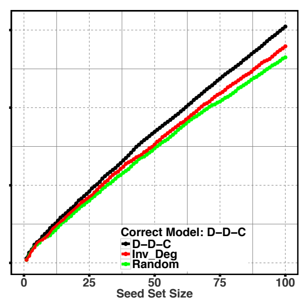
(d)



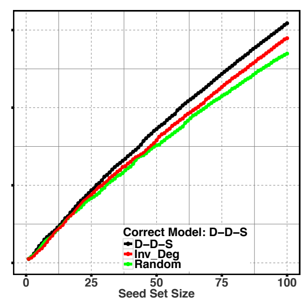
(e)



(f)

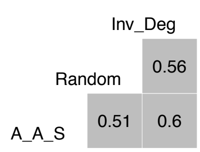


(g)

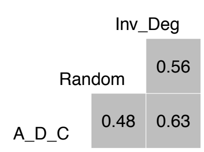


(h)

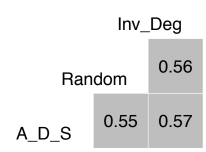
(i)



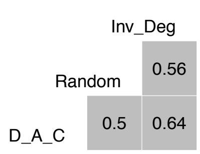
(j)



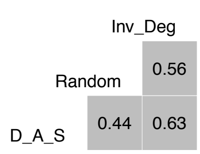
(k)



(l)

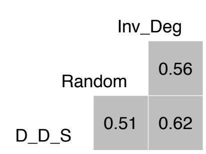


(m)



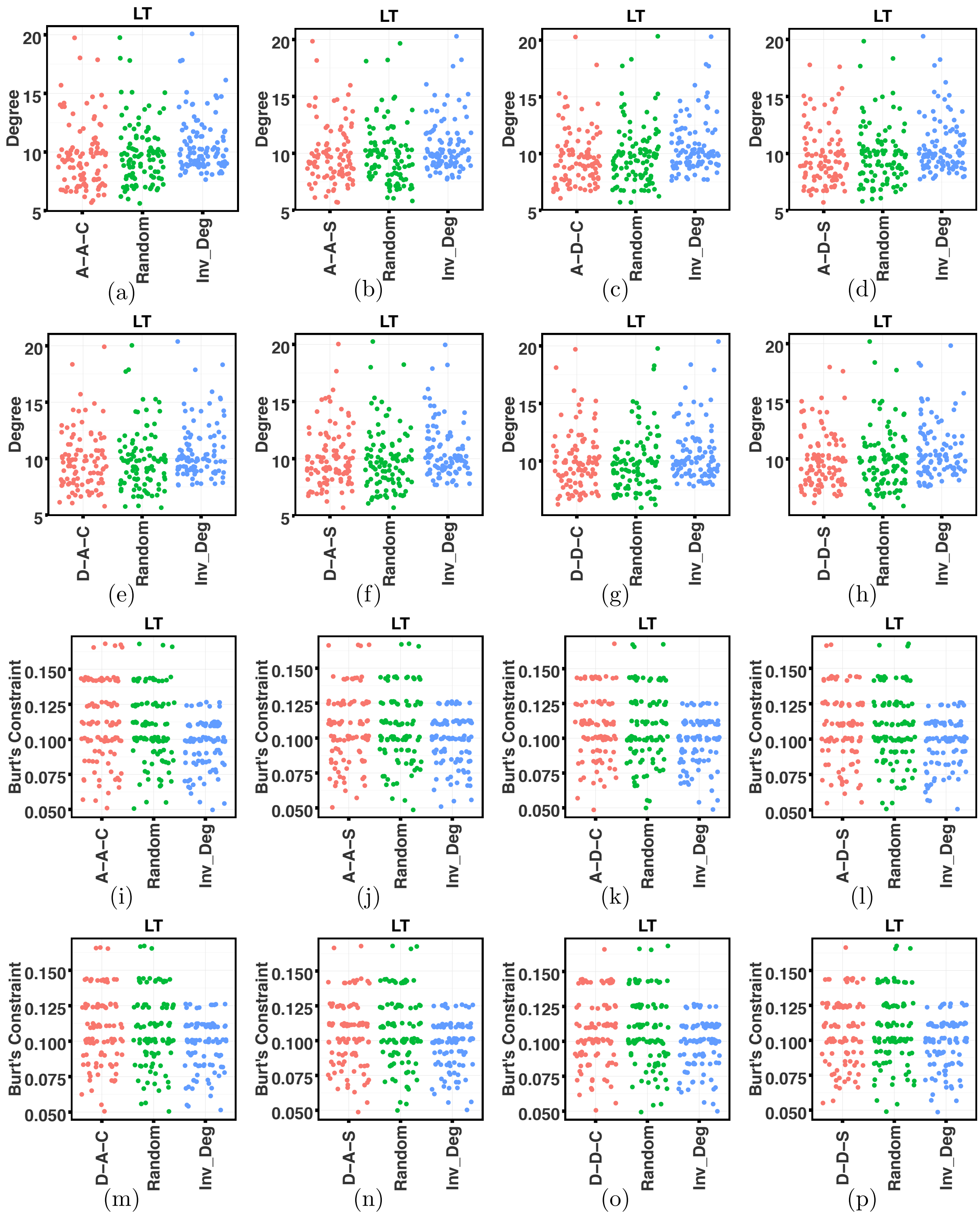
(n)

(o)

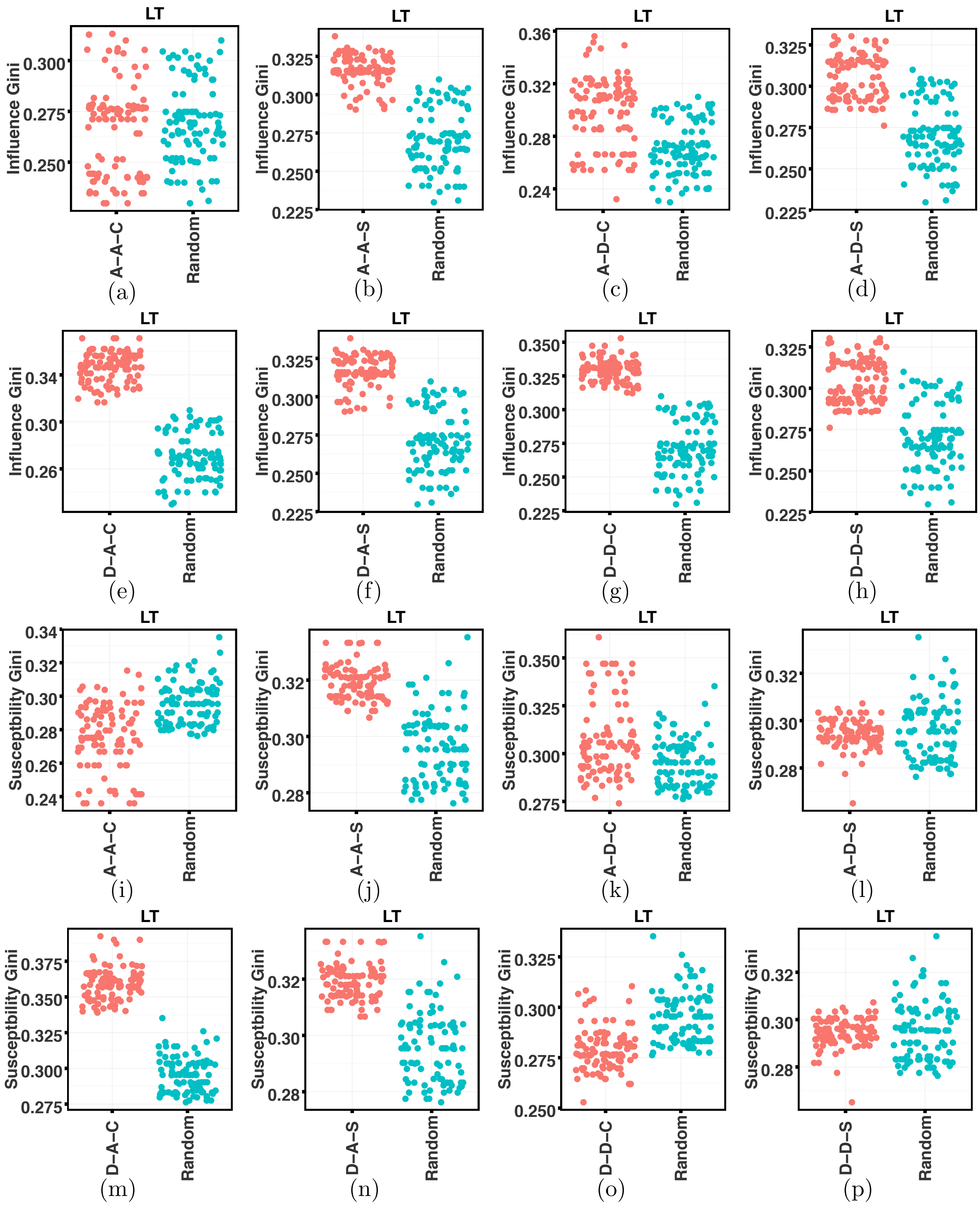


(p)

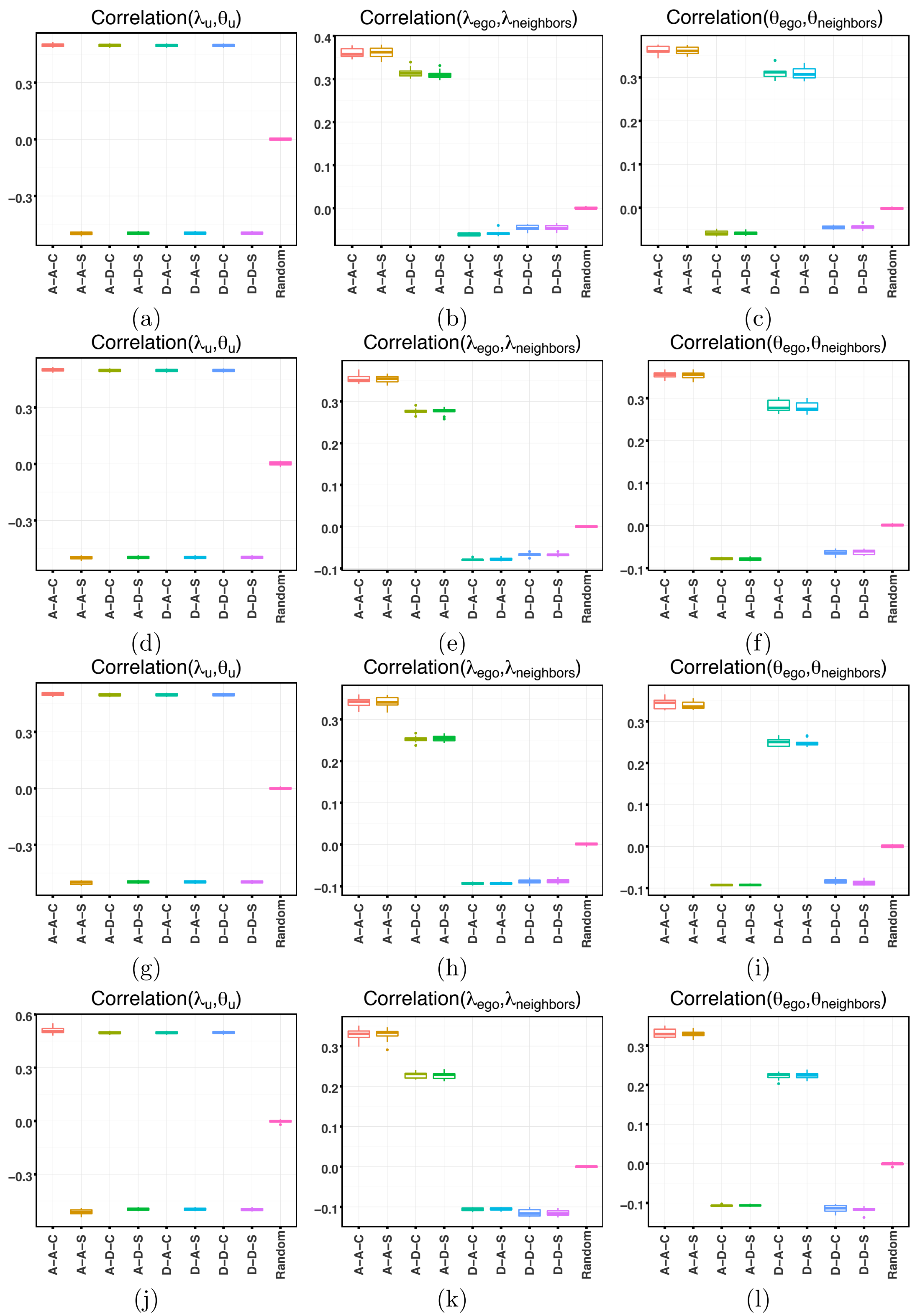
Supplementary Figure 5: Influence spread as a function of the seed set size and seed overlap for graphs generated using AGM. Influence propagation model is LT. **Network**: Preferential Attachment, Nodes=10000, Edges = {AAC=49970, AAS=49968, ADC=49975, ADS=49973, DAC=49970, DAS=49963, DDC=49966, DDS=49970}. **Note:** AGM generates both graphs and attributes, therefore, the graphs generated under different assumptions have different number of edges. Hence, the baselines i.e. Random and Inverse Degree (Inv_Deg) are also different for the different assortative/disassortative assumptions.



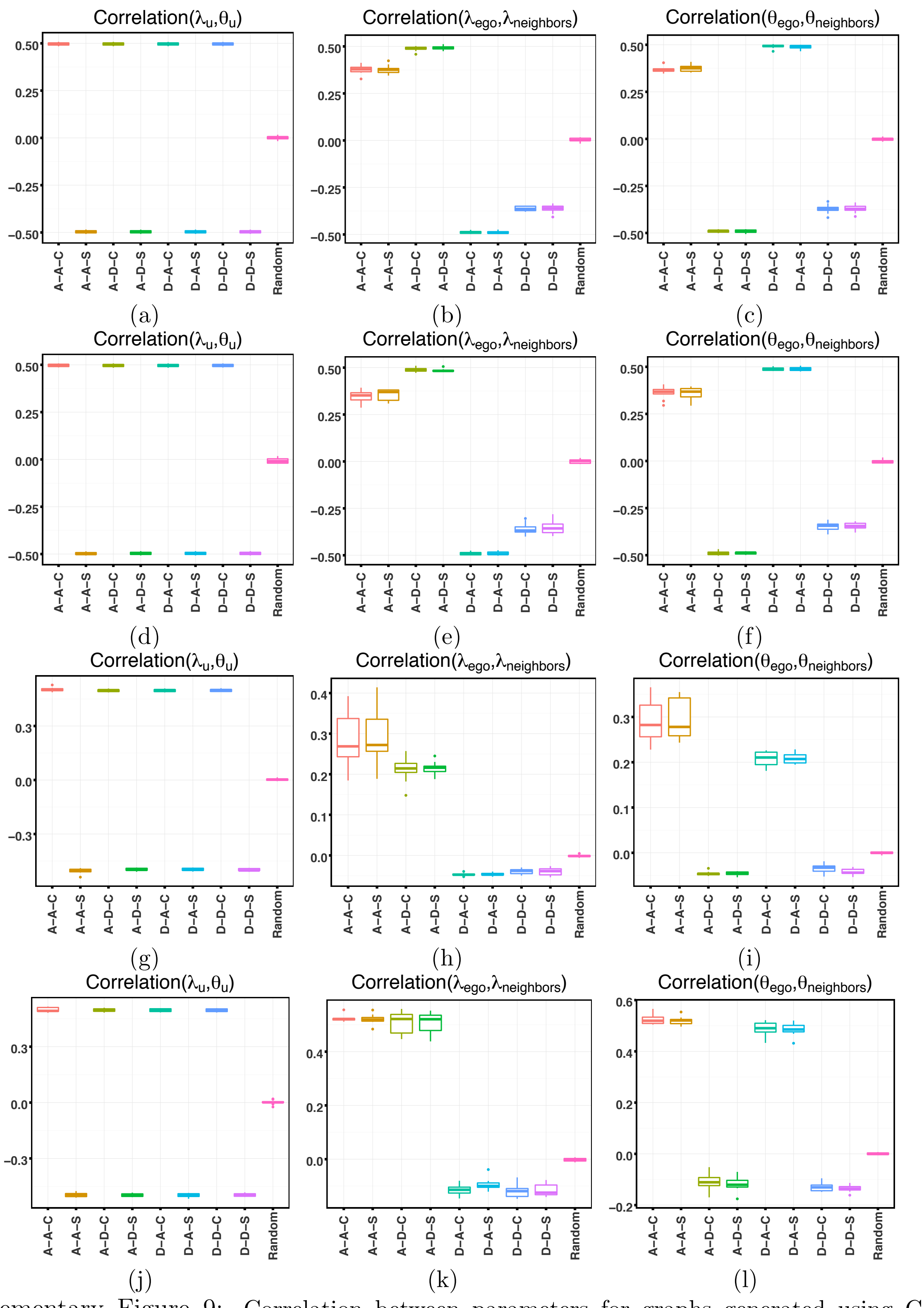
Supplementary Figure 6: Distribution of degree and Burt's constraints of the seed nodes chosen for graphs generated using AGM. Influence propagation model is LT. **Network**: Preferential Attachment, Nodes=10000, Edges = {AAC=49970, AAS=49968, ADC=49975, ADS=49973, DAC=49970, DAS=49963, DDC=49966, DDS=49970}. **Note:** AGM generates both graphs and attributes, therefore, the graphs generated under different assumptions have different number of edges. Hence, the baselines i.e. Random and Inverse Degree (Inv_Deg) are also different for the different assortative/disassortative assumptions.



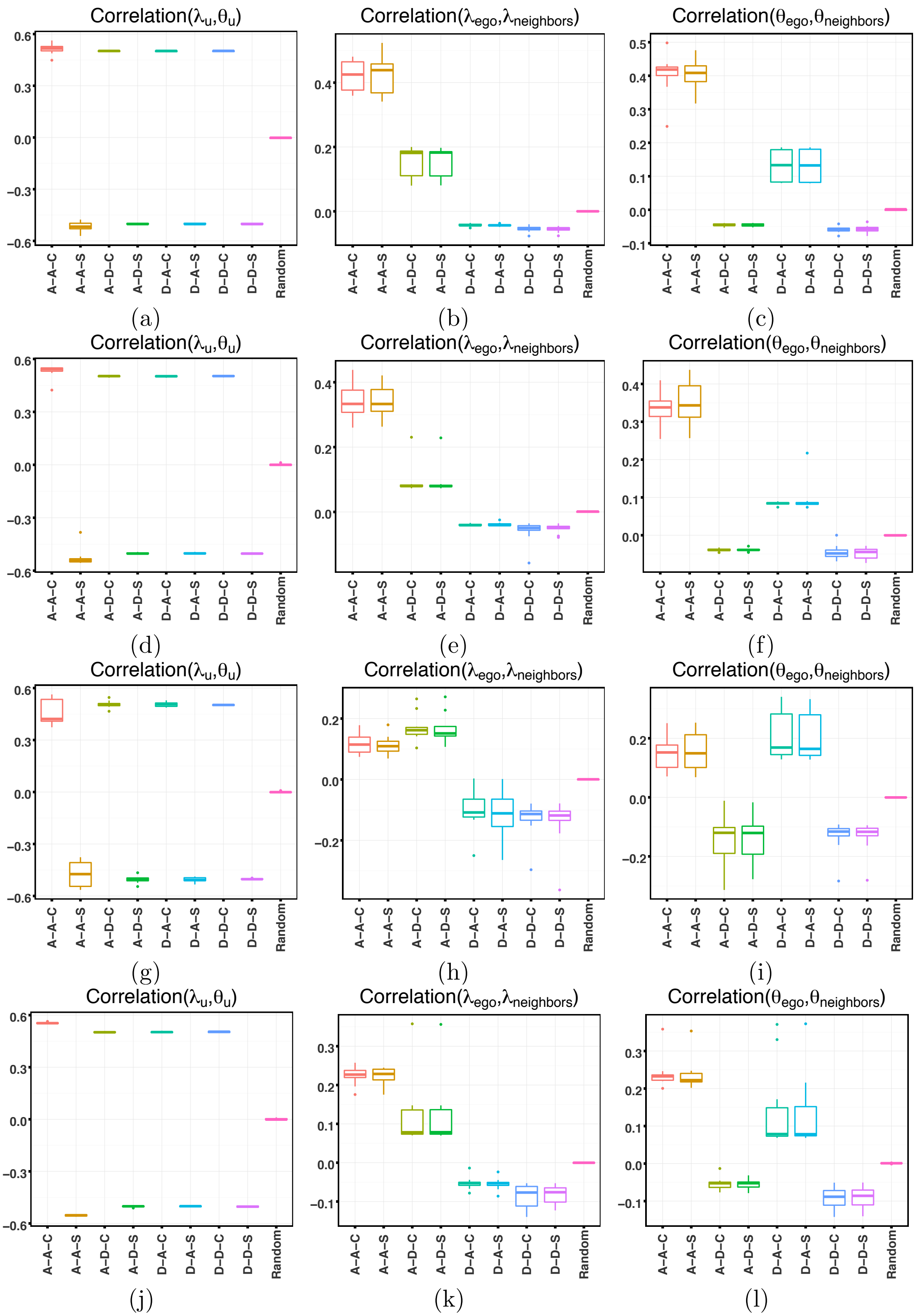
Supplementary Figure 7: Distribution of the Gini coefficient for the influence and susceptibility parameters of the seeds nodes chosen for graphs were generated using AGM. Influence propagation model is LT. **Network**: Preferential Attachment, Nodes=10000, Edges = {AAC=49970, AAS=49968, ADC=49975, ADS=49973, DAC=49970, DAS=49963, DDC=49966, DDS=49970}. **Note:** AGM generates both graphs and attributes, therefore, the graphs generated under different assumptions have different number of edges. Hence, the baselines i.e. Random and Inverse Degree (Inv_Deg) are also different for the different assortative/disassortative assumptions.



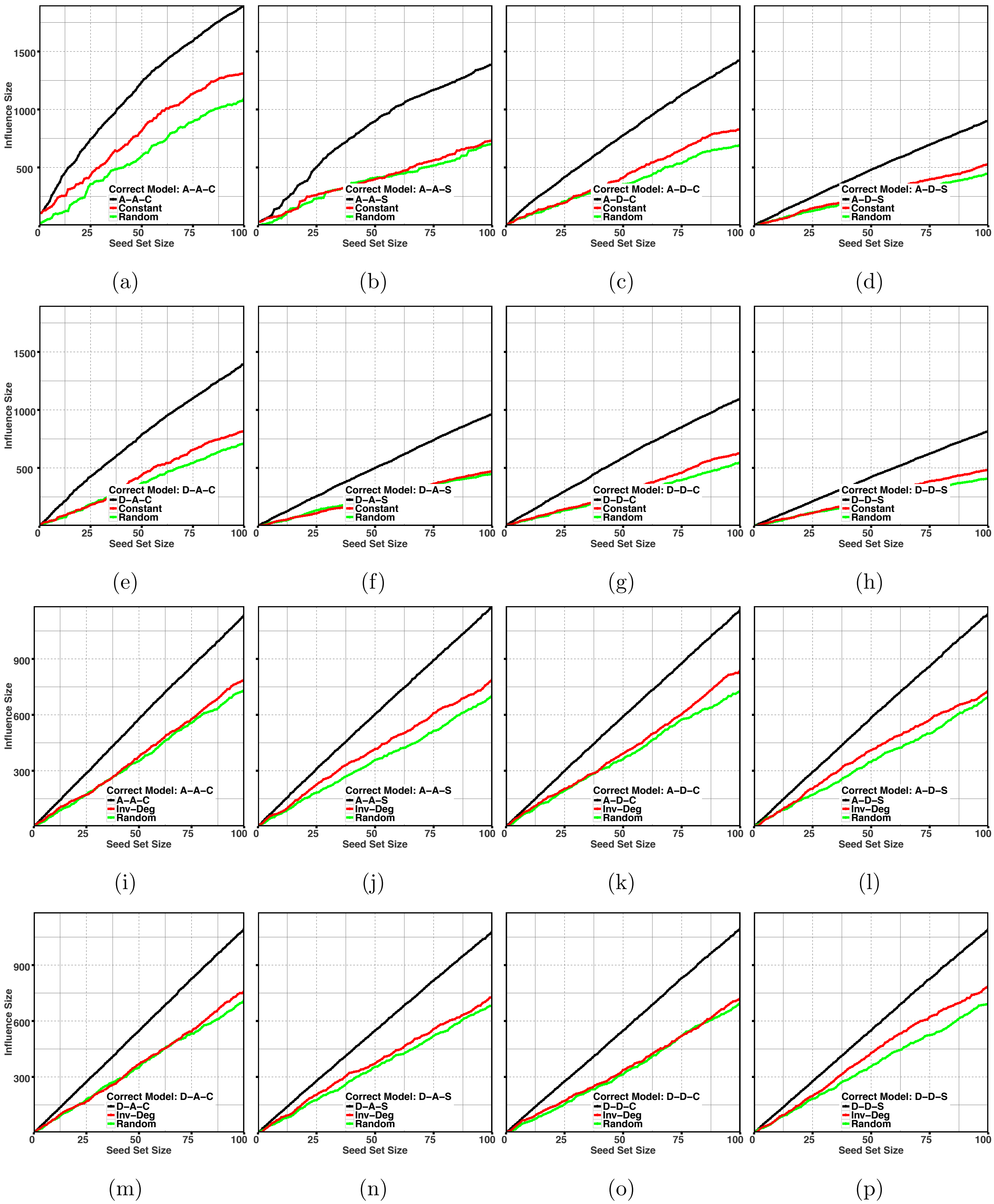
Supplementary Figure 8: Correlation between parameters for graphs generated using Correlated Label Propagation (Algorithm 1). **Networks: Panels (a-c) Small World** $p_{rewire} = 0.05$**, (d-f) Small World** $p_{rewire} = 0.1$**, (g-i) Small World** $p_{rewire} = 0.05$**, (j-l) Small World** $p_{rewire} = 0.2$. Note that all the plots show 10 random i.i.d draws of the parameter values.



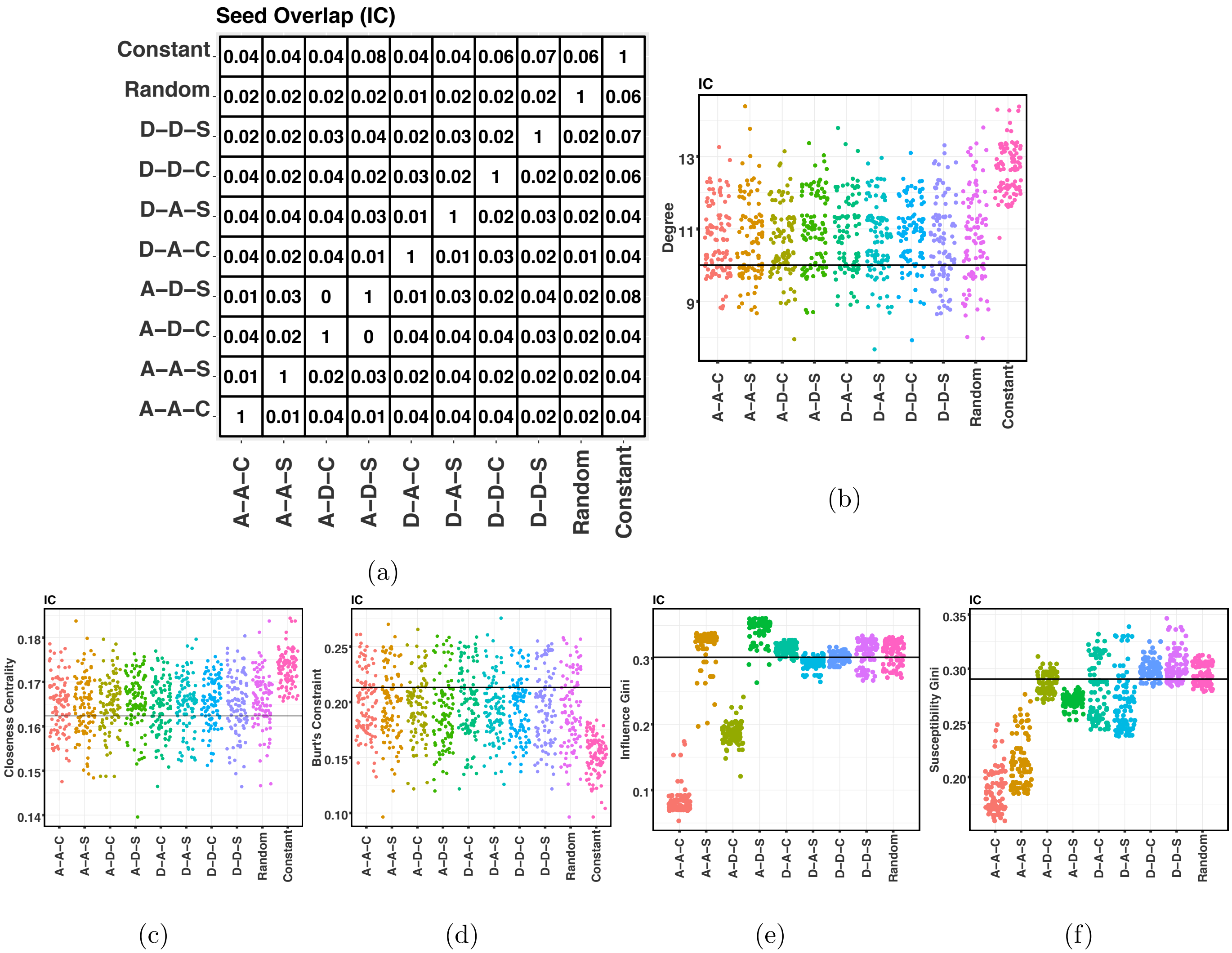
Supplementary Figure 9: Correlation between parameters for graphs generated using Correlated Label Propagation (Algorithm 1). **Networks: Panels (a-c) Preferential Attachment** $exponent = 0.5$**, (d-f) Preferential Attachment** $exponent = 1$**, (g-i) Collaboration Network Arxiv HEP-PH (High Energy Physics - Phenomenology (j-l) Collaboration Network Arxiv HEP-TH (High Energy Physics - Theory).** Other details are the same as the last figure.



Supplementary Figure 10: Correlation between parameters for graphs generated using Correlated Label Propagation (Algorithm 1). **Networks: Panels (a-c) Citation Network Arxiv HEP-PH (High Energy Physics - Phenomenology, (d-f) Citation Network Arxiv HEP-TH (High Energy Physics - Theory, (g-i) Slashdot Friend Network November 2008, (j-l) Epinions "Trust" Network.** Other details are the same as the last figure.



Supplementary Figure 11: Influence spread as a function of the seed set size. **Network: Small World, Nodes=10000, Edges =50000,** p_{rewire} = 0.05. Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are “Random” and “Constant” (all edges=0.1) and for the LT model are “Random” and “Inverse Degree”.

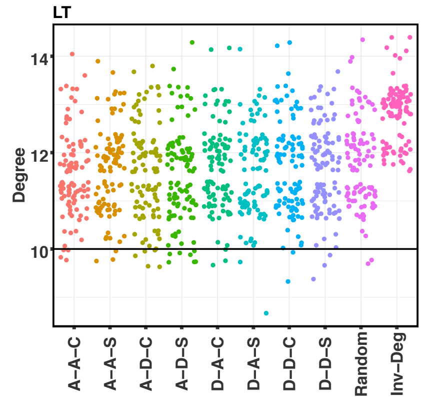


Supplementary Figure 12: Various seed characteristics for the IC model. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.05$. (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.

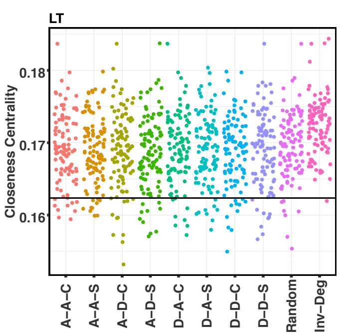
Seed Overlap (LT)

	A–A–C	A–A–S	A–D–C	A–D–S	D–A–C	D–A–S	D–D–C	D–D–S	Random	Inv–Deg
Inv–Deg	0.18	0.16	0.18	0.19	0.16	0.15	0.16	0.16	0.17	1
Random	0.06	0.06	0.08	0.09	0.06	0.07	0.09	0.08	1	0.17
D–D–S	0.05	0.1	0.05	0.11	0.04	0.12	0.02	1	0.08	0.16
D–D–C	0.1	0.06	0.1	0.06	0.11	0.05	1	0.02	0.09	0.16
D–A–S	0.05	0.12	0.04	0.11	0	1	0.05	0.12	0.07	0.15
D–A–C	0.1	0.05	0.13	0.03	1	0	0.11	0.04	0.06	0.16
A–D–S	0.04	0.14	0.01	1	0.03	0.11	0.06	0.11	0.09	0.19
A–D–C	0.12	0.03	1	0.01	0.13	0.04	0.1	0.05	0.08	0.18
A–A–S	0.01	1	0.03	0.14	0.05	0.12	0.06	0.1	0.06	0.16
A–A–C	1	0.01	0.12	0.04	0.1	0.05	0.1	0.05	0.06	0.18

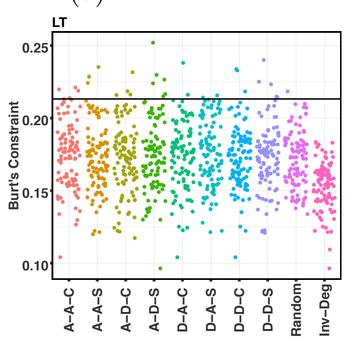
(a)



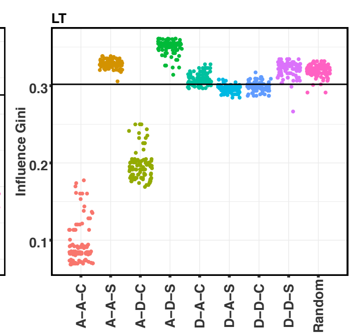
(b)



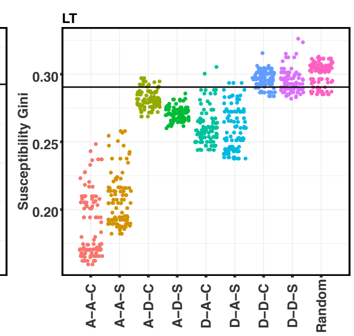
(c)



(d)

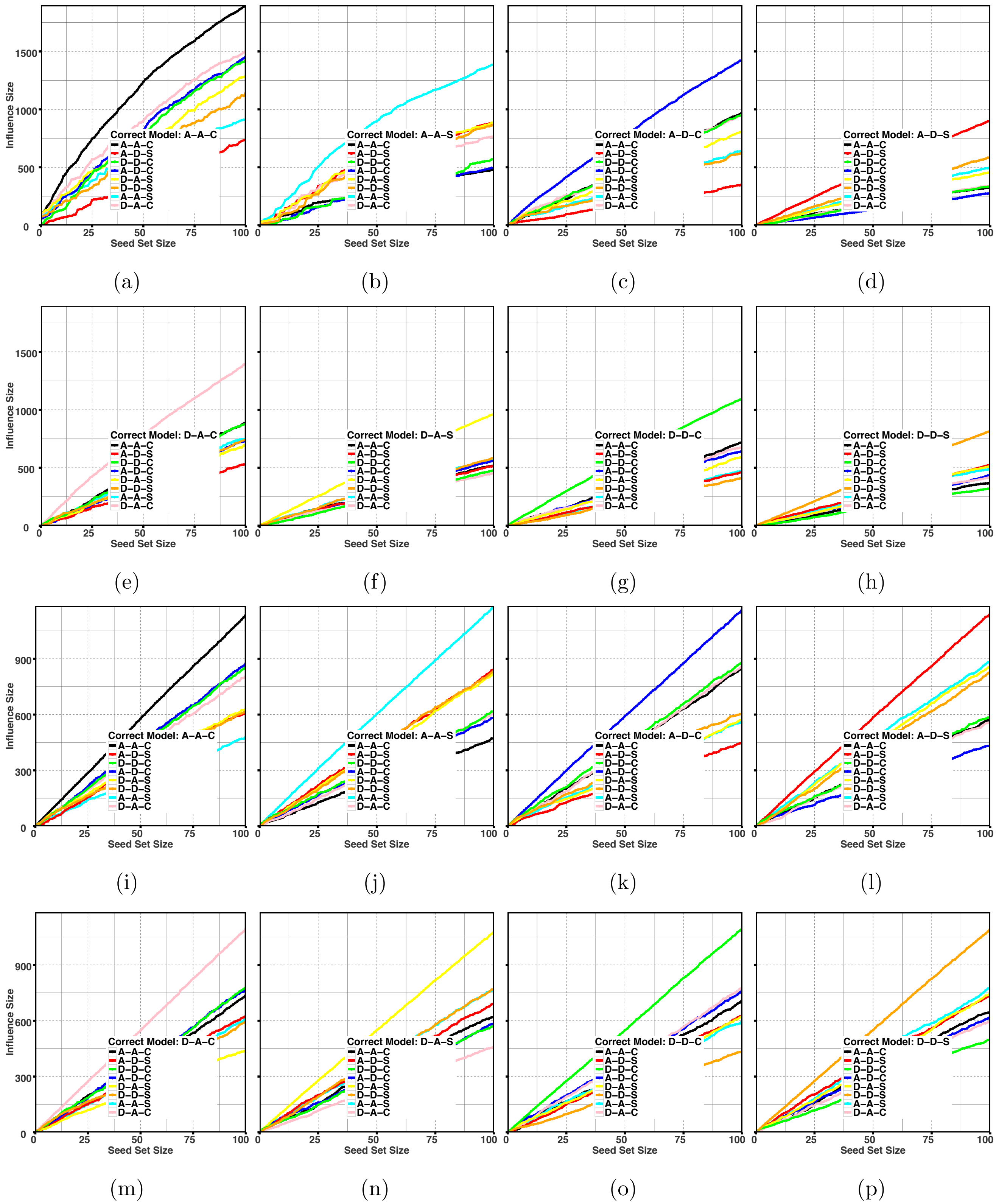


(e)

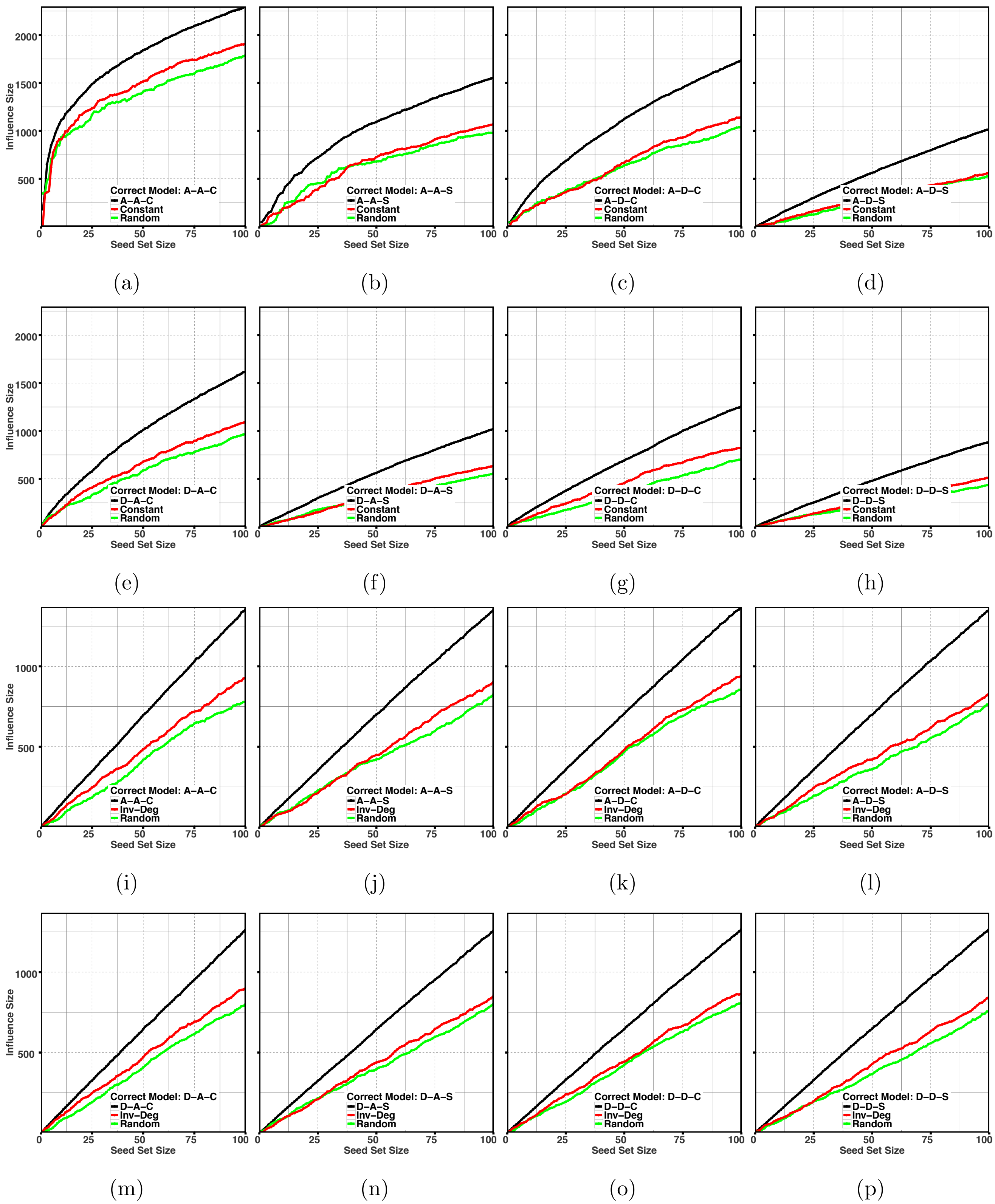


(f)

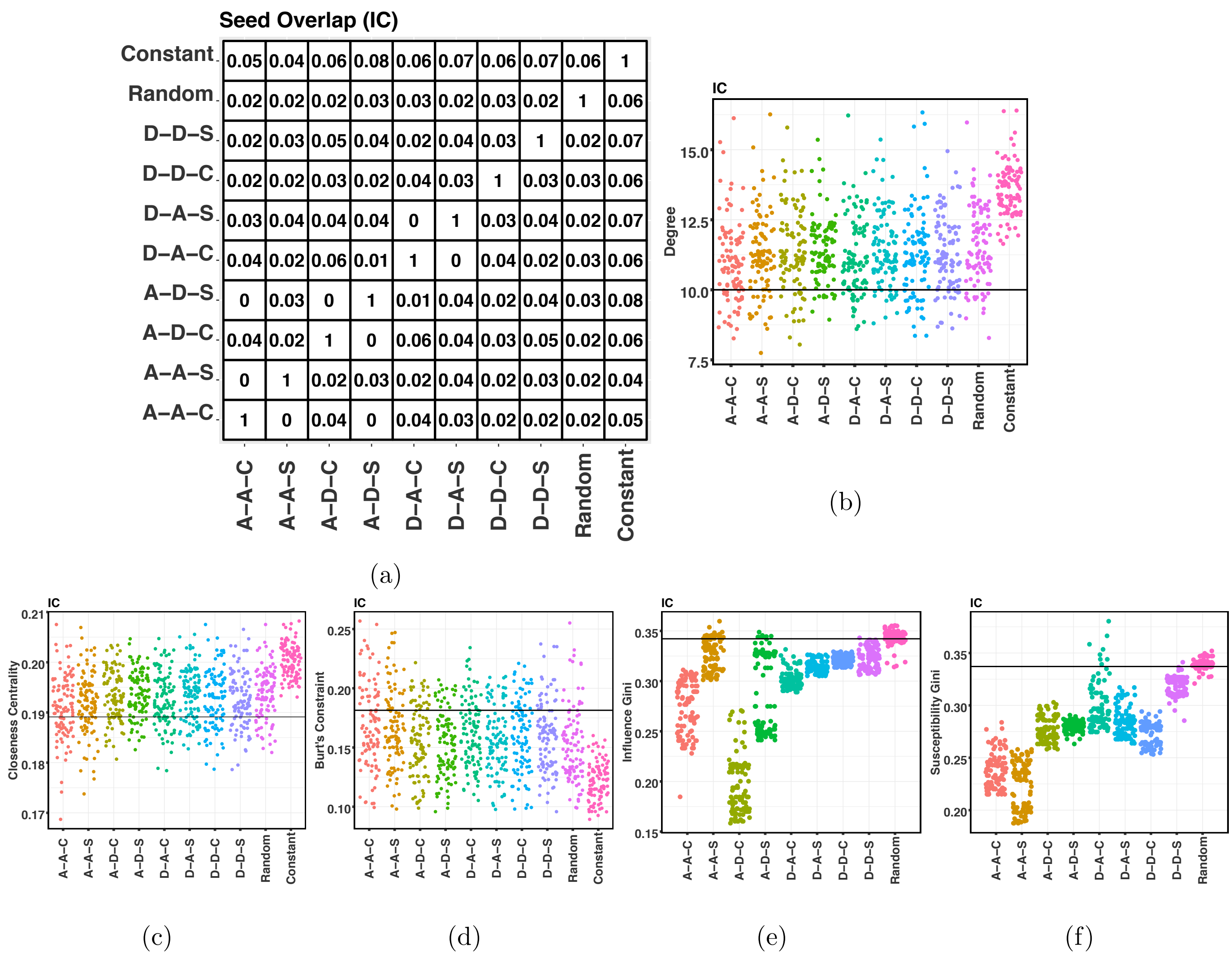
Supplementary Figure 13: Various seed characteristics for the LT model. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.05$. (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



Supplementary Figure 14: Influence spread as a function of the seed set size. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.05$. Panels (a-h): IC Model, Panels (i-p): LT Model.



Supplementary Figure 15: Influence spread as a function of the seed set size. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.1$. Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are "Random" and "Constant" (all edges=0.1) and for the LT model are "Random" and "Inverse Degree".

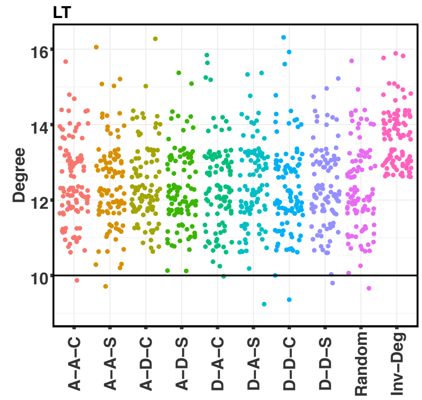


Supplementary Figure 16: Various seed characteristics for the IC model. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.1$. (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.

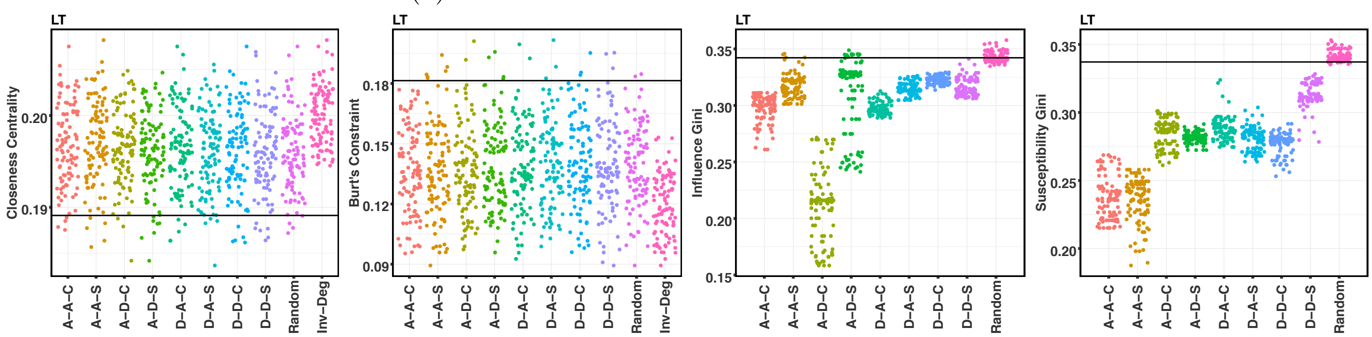
Seed Overlap (LT)

	A–A–C	A–A–S	A–D–C	A–D–S	D–A–C	D–A–S	D–D–C	D–D–S	Random	Inv–Deg
Inv–Deg	0.2	0.17	0.2	0.19	0.19	0.17	0.19	0.17	0.18	1
Random	0.09	0.09	0.07	0.09	0.07	0.09	0.07	0.07	1	0.18
D–D–S	0.09	0.09	0.07	0.12	0.05	0.11	0.05	1	0.07	0.17
D–D–C	0.11	0.07	0.13	0.05	0.14	0.06	1	0.05	0.07	0.19
D–A–S	0.06	0.11	0.05	0.13	0.01	1	0.06	0.11	0.09	0.17
D–A–C	0.13	0.04	0.14	0.04	1	0.01	0.14	0.05	0.07	0.19
A–D–S	0.04	0.15	0.01	1	0.04	0.13	0.05	0.12	0.09	0.19
A–D–C	0.15	0.04	1	0.01	0.14	0.05	0.13	0.07	0.07	0.2
A–A–S	0.02	1	0.04	0.15	0.04	0.11	0.07	0.09	0.09	0.17
A–A–C	1	0.02	0.15	0.04	0.13	0.06	0.11	0.09	0.09	0.2

(a)

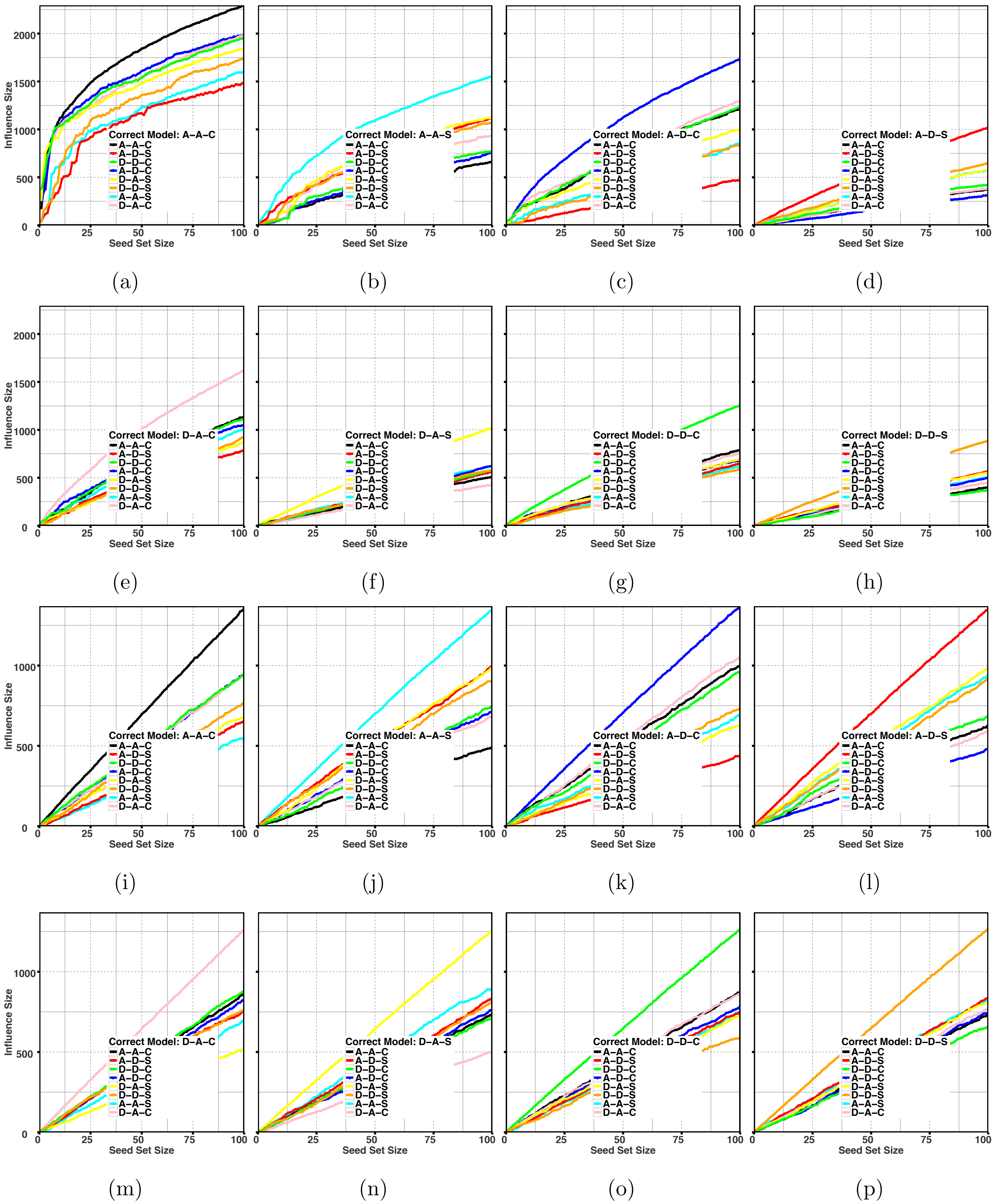


(b)

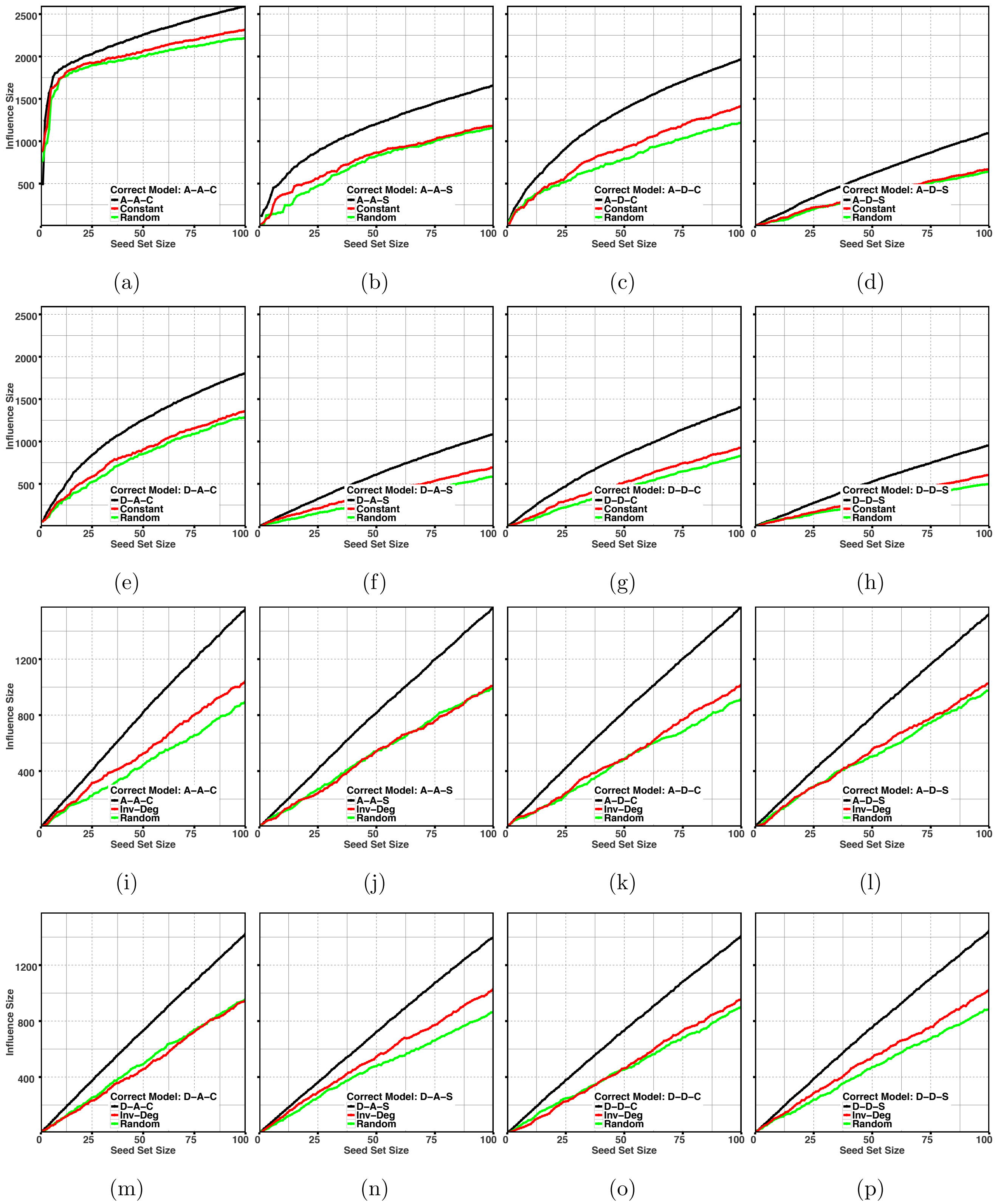


(c) (d) (e) (f)

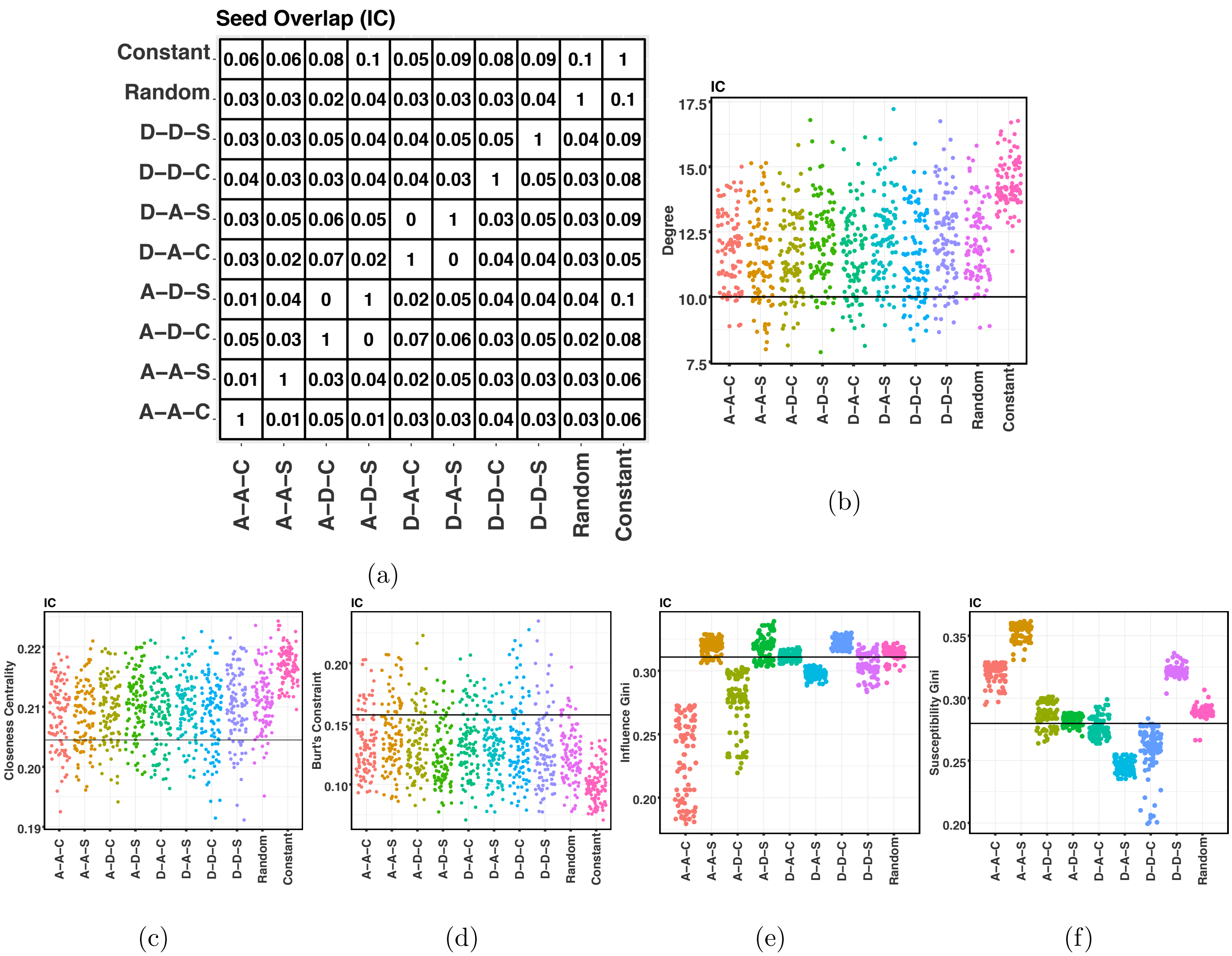
Supplementary Figure 17: Various seed characteristics for the LT model. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.1$. (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



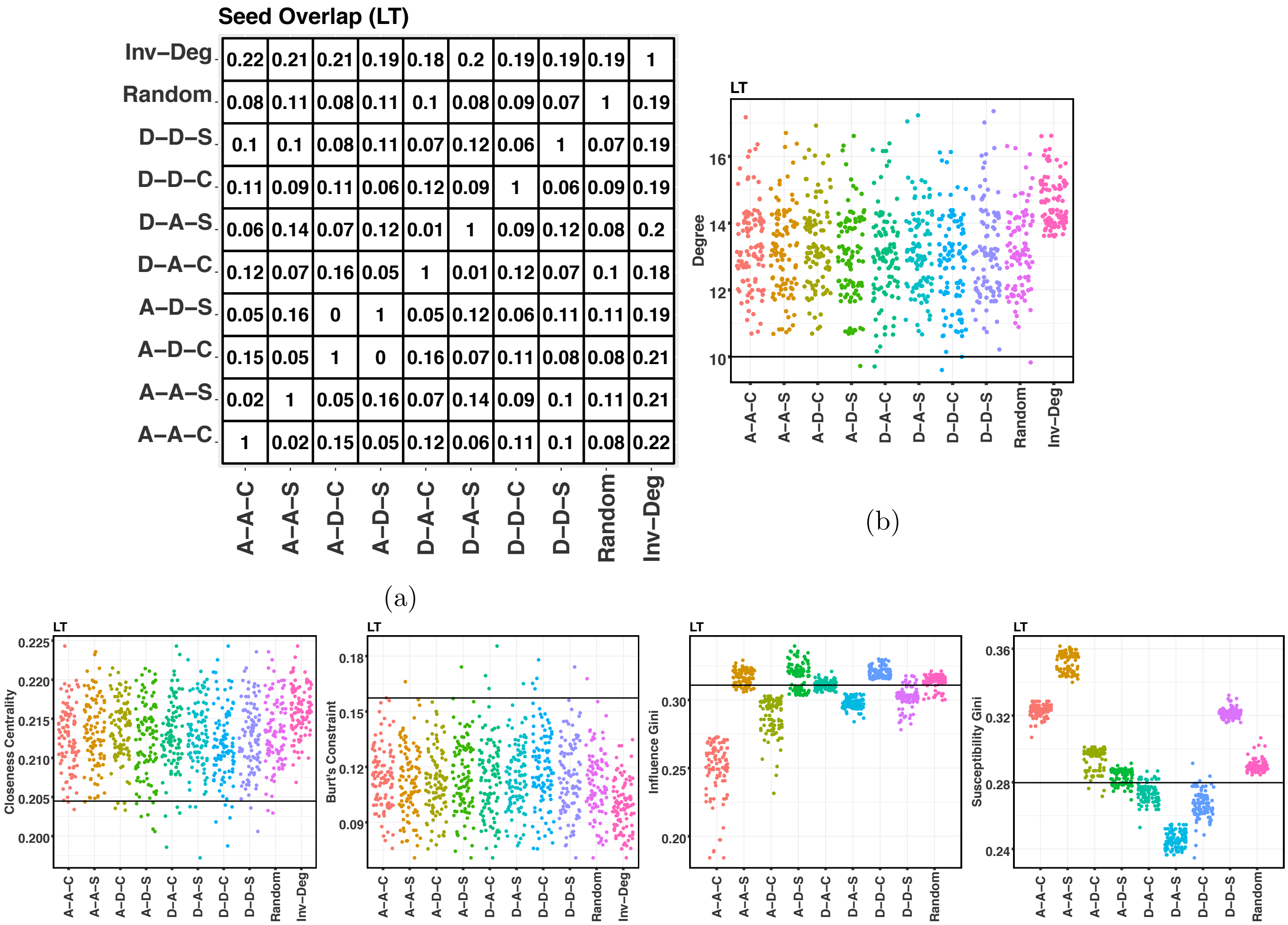
Supplementary Figure 18: Influence spread as a function of the seed set size. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.1$. Panels (a-h): IC Model, Panels (i-p): LT Model.



Supplementary Figure 19: Influence spread as a function of the seed set size. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.15$. Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are “Random” and “Constant” (all edges=0.1) and for the LT model are “Random” and “Inverse Degree”.



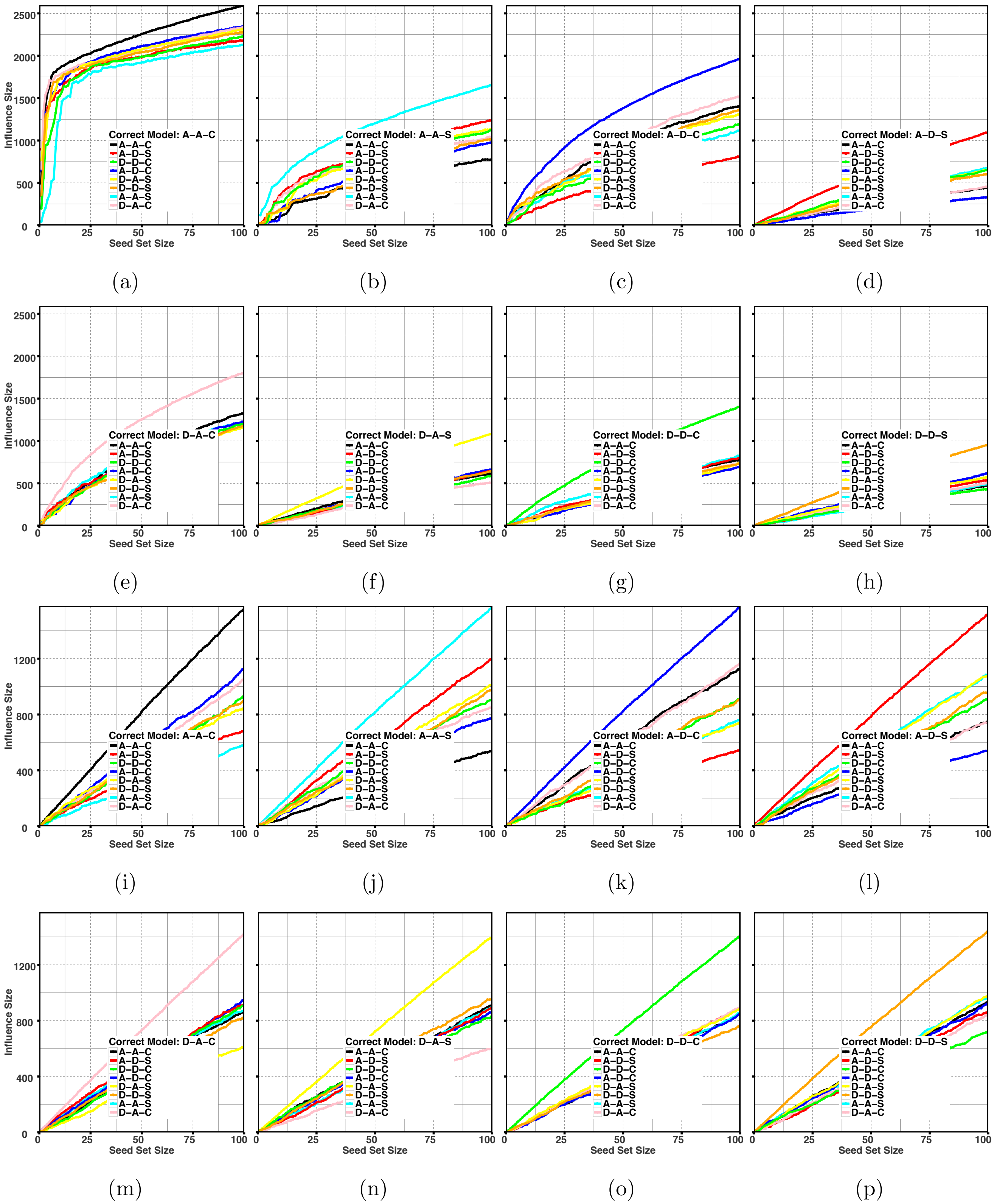
Supplementary Figure 20: Various seed characteristics for the IC model. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.15$. (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



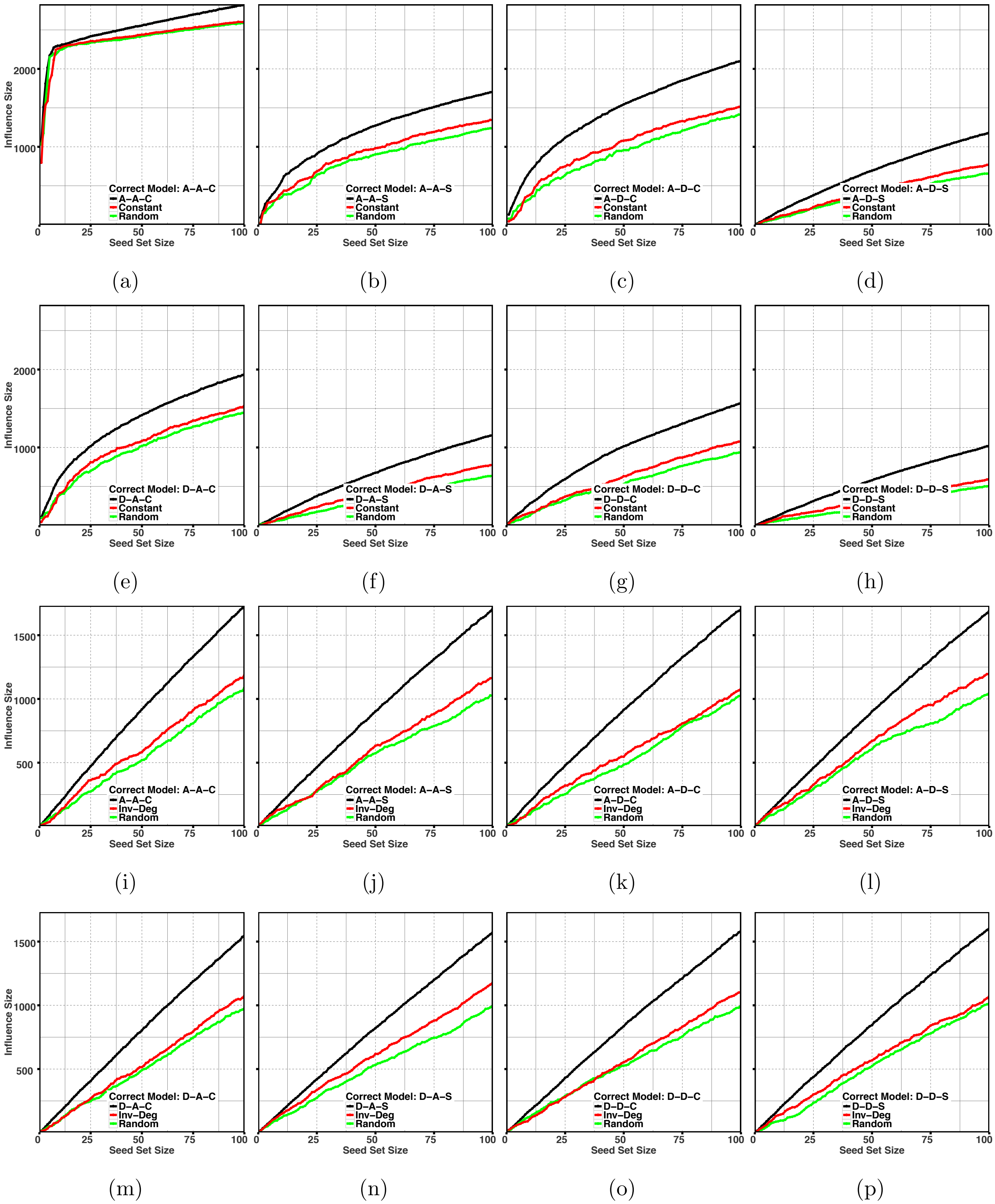
(a) (b)

(c) (d) (e) (f)

Supplementary Figure 21: Various seed characteristics for the LT model. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.15$. (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



Supplementary Figure 22: Influence spread as a function of the seed set size. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.15$. Panels (a-h): IC Model, Panels (i-p): LT Model.

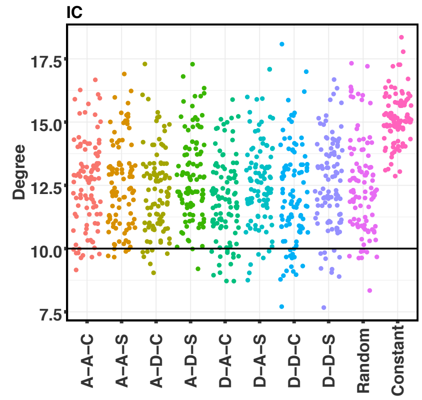


Supplementary Figure 23: Influence spread as a function of the seed set size. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.2$. Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are "Random" and "Constant" (all edges=0.1) and for the LT model are "Random" and "Inverse Degree".

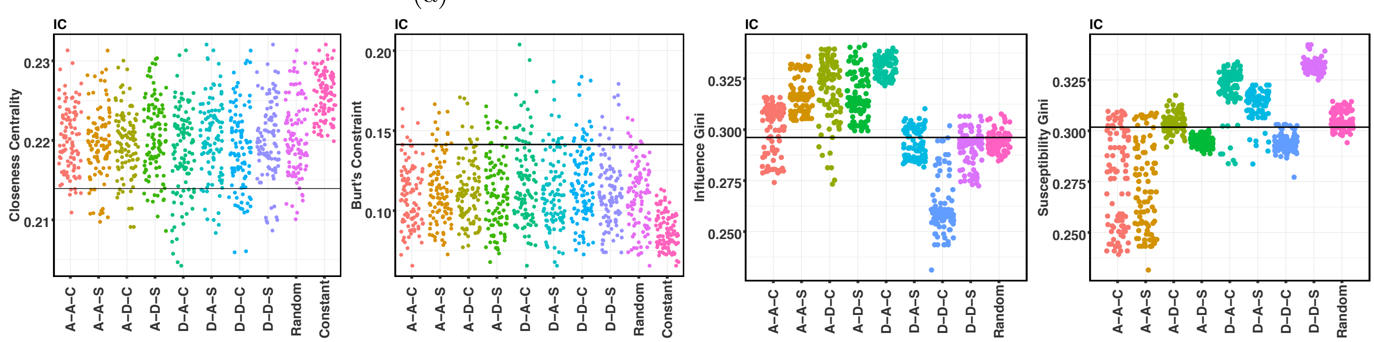
Seed Overlap (IC)

	A–A–C	A–A–S	A–D–C	A–D–S	D–A–C	D–A–S	D–D–C	D–D–S	Random	Constant
Constant	0.07	0.09	0.09	0.13	0.1	0.1	0.08	0.1	0.09	1
Random	0.03	0.02	0.03	0.04	0.02	0.04	0.03	0.03	1	0.09
D–D–S	0.03	0.04	0.05	0.07	0.02	0.05	0.03	1	0.03	0.1
D–D–C	0.04	0.03	0.04	0.05	0.03	0.04	1	0.03	0.03	0.08
D–A–S	0.04	0.06	0.04	0.08	0	1	0.04	0.05	0.04	0.1
D–A–C	0.03	0.04	0.09	0.02	1	0	0.03	0.02	0.02	0.1
A–D–S	0.03	0.06	0	1	0.02	0.08	0.05	0.07	0.04	0.13
A–D–C	0.04	0.02	1	0	0.09	0.04	0.04	0.05	0.03	0.09
A–A–S	0.01	1	0.02	0.06	0.04	0.06	0.03	0.04	0.02	0.09
A–A–C	1	0.01	0.04	0.03	0.03	0.04	0.04	0.03	0.03	0.07

(a)



(b)



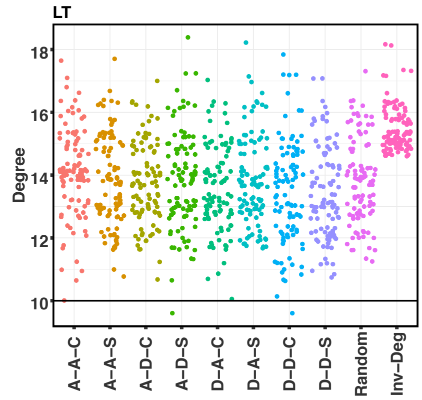
(c) (d) (e) (f)

Supplementary Figure 24: Various seed characteristics for the IC model. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.2$. (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.

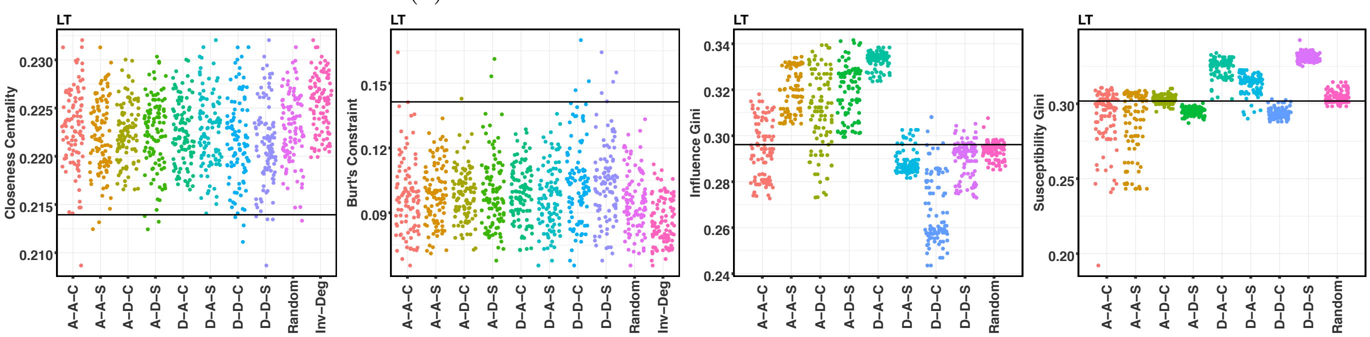
Seed Overlap (LT)

	A–A–C	A–A–S	A–D–C	A–D–S	D–A–C	D–A–S	D–D–C	D–D–S	Random	Inv–Deg
Inv–Deg	0.25	0.23	0.25	0.24	0.24	0.21	0.22	0.22	0.22	1
Random	0.1	0.11	0.12	0.09	0.08	0.1	0.09	0.08	1	0.22
D–D–S	0.11	0.09	0.1	0.11	0.09	0.12	0.08	1	0.08	0.22
D–D–C	0.12	0.11	0.1	0.11	0.12	0.08	1	0.08	0.09	0.22
D–A–S	0.09	0.14	0.08	0.16	0.01	1	0.08	0.12	0.1	0.21
D–A–C	0.14	0.07	0.16	0.07	1	0.01	0.12	0.09	0.08	0.24
A–D–S	0.05	0.19	0.01	1	0.07	0.16	0.11	0.11	0.09	0.24
A–D–C	0.19	0.06	1	0.01	0.16	0.08	0.1	0.1	0.12	0.25
A–A–S	0.01	1	0.06	0.19	0.07	0.14	0.11	0.09	0.11	0.23
A–A–C	1	0.01	0.19	0.05	0.14	0.09	0.12	0.11	0.1	0.25

(a)

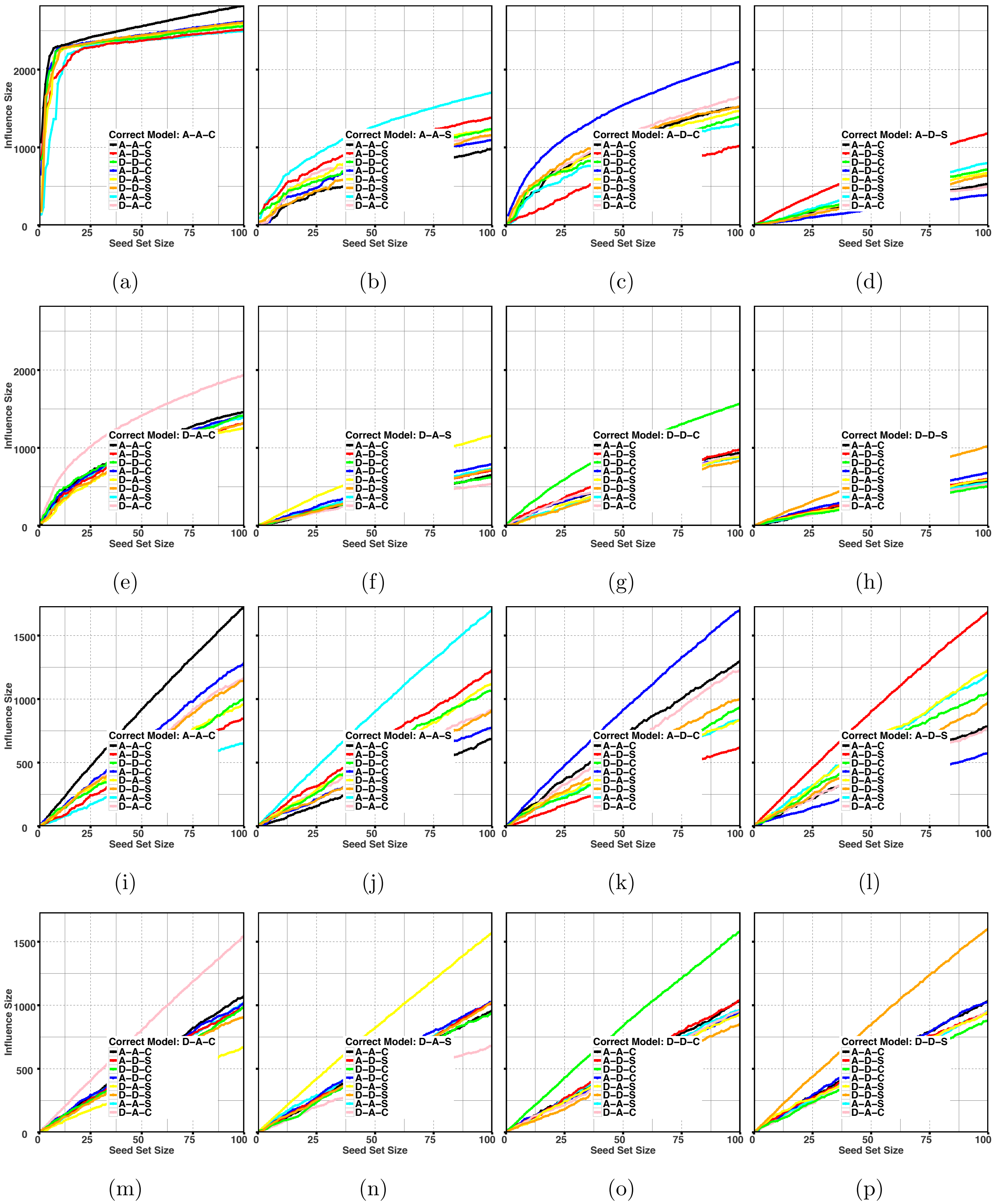


(b)

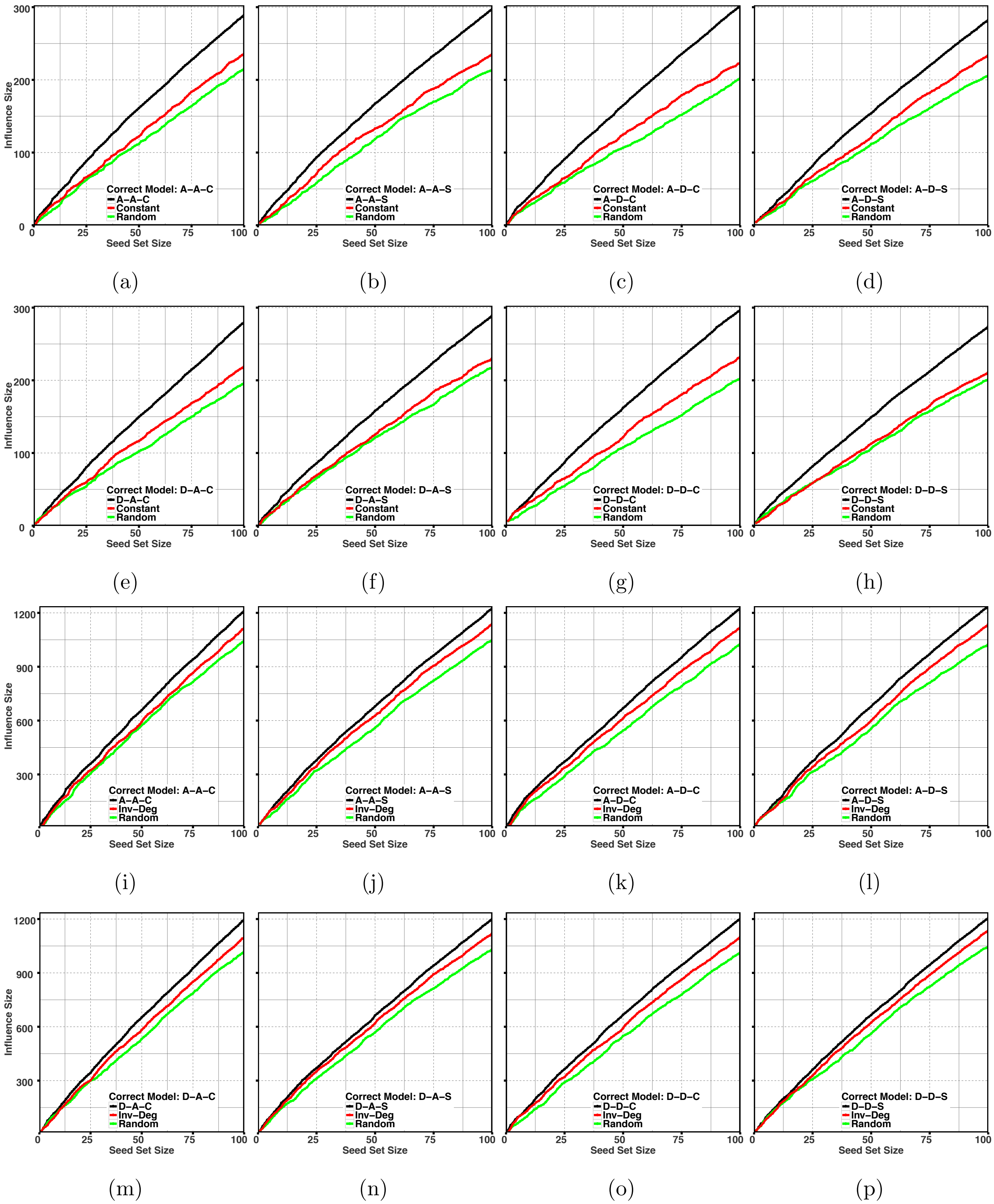


(c) (d) (e) (f)

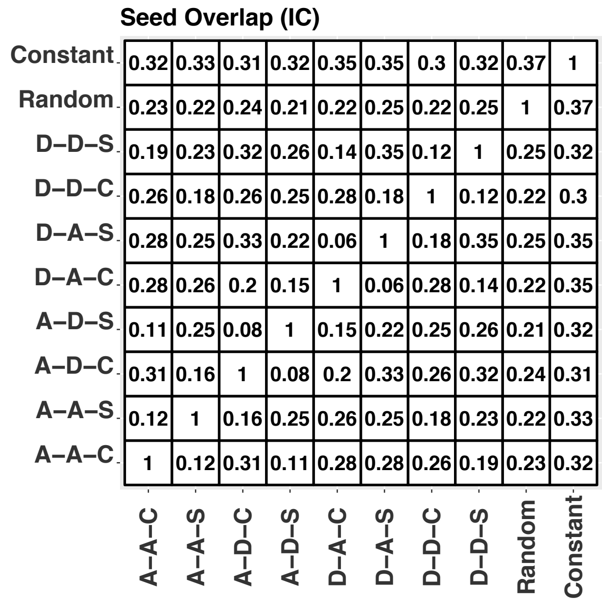
Supplementary Figure 25: Various seed characteristics for the LT model. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.2$. (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



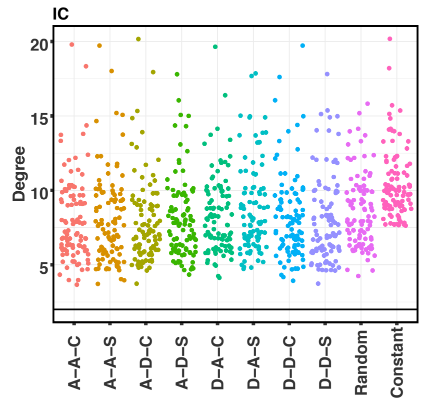
Supplementary Figure 26: Influence spread as a function of the seed set size. **Network: Small World, Nodes=10000, Edges =50000,** $p_{rewire} = 0.2$. Panels (a-h): IC Model, Panels (i-p): LT Model.



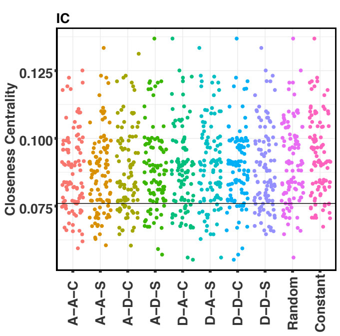
Supplementary Figure 27: Influence spread as a function of the seed set size. **Network: Preferential Attachment, Nodes=10000, Edges =9999, Exponent = 0.5, Zero Appeal =2.** Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are "Random" and "Constant" (all edges=0.1) and for the LT model are "Random" and "Inverse Degree".



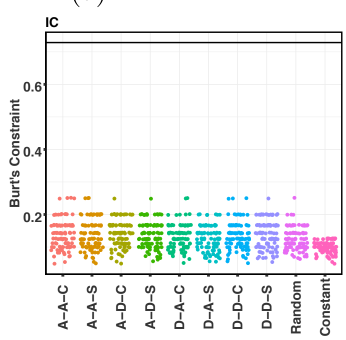
(a)



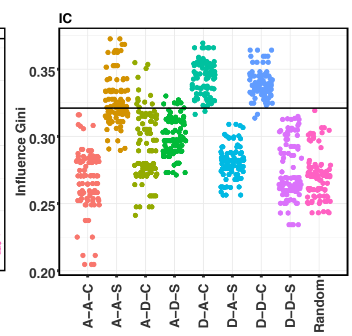
(b)



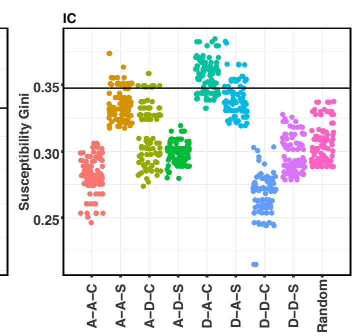
(c)



(d)

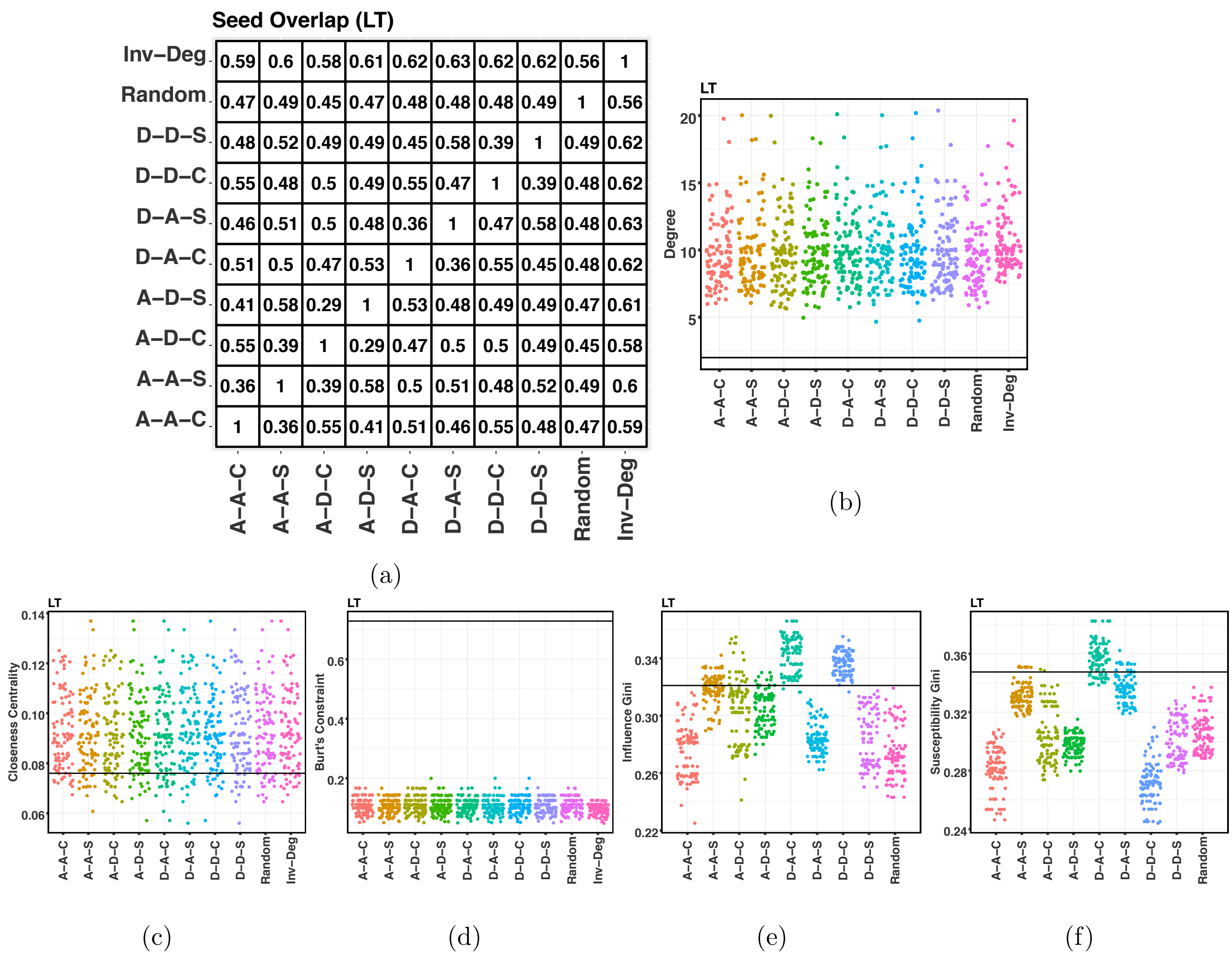


(e)

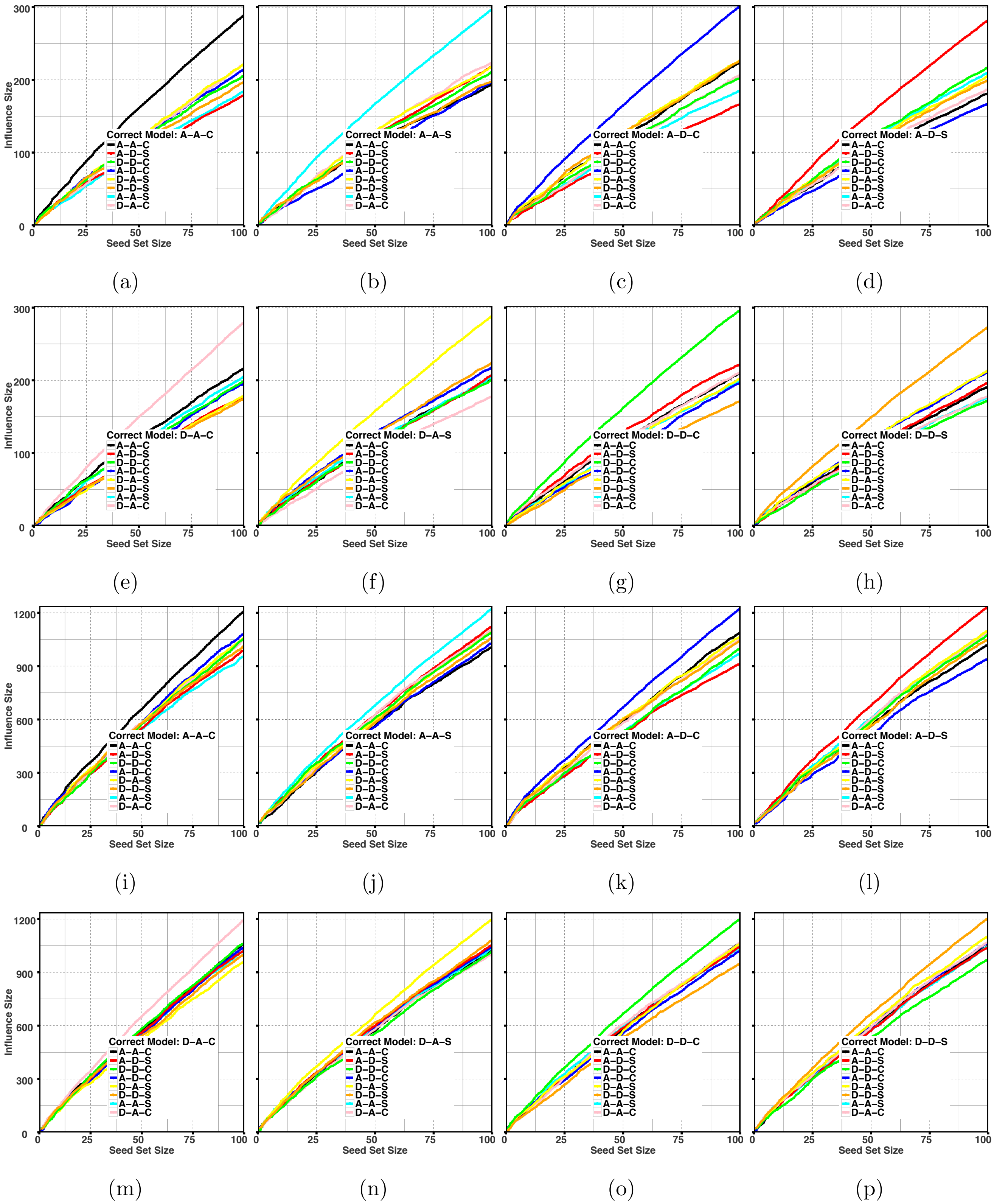


(f)

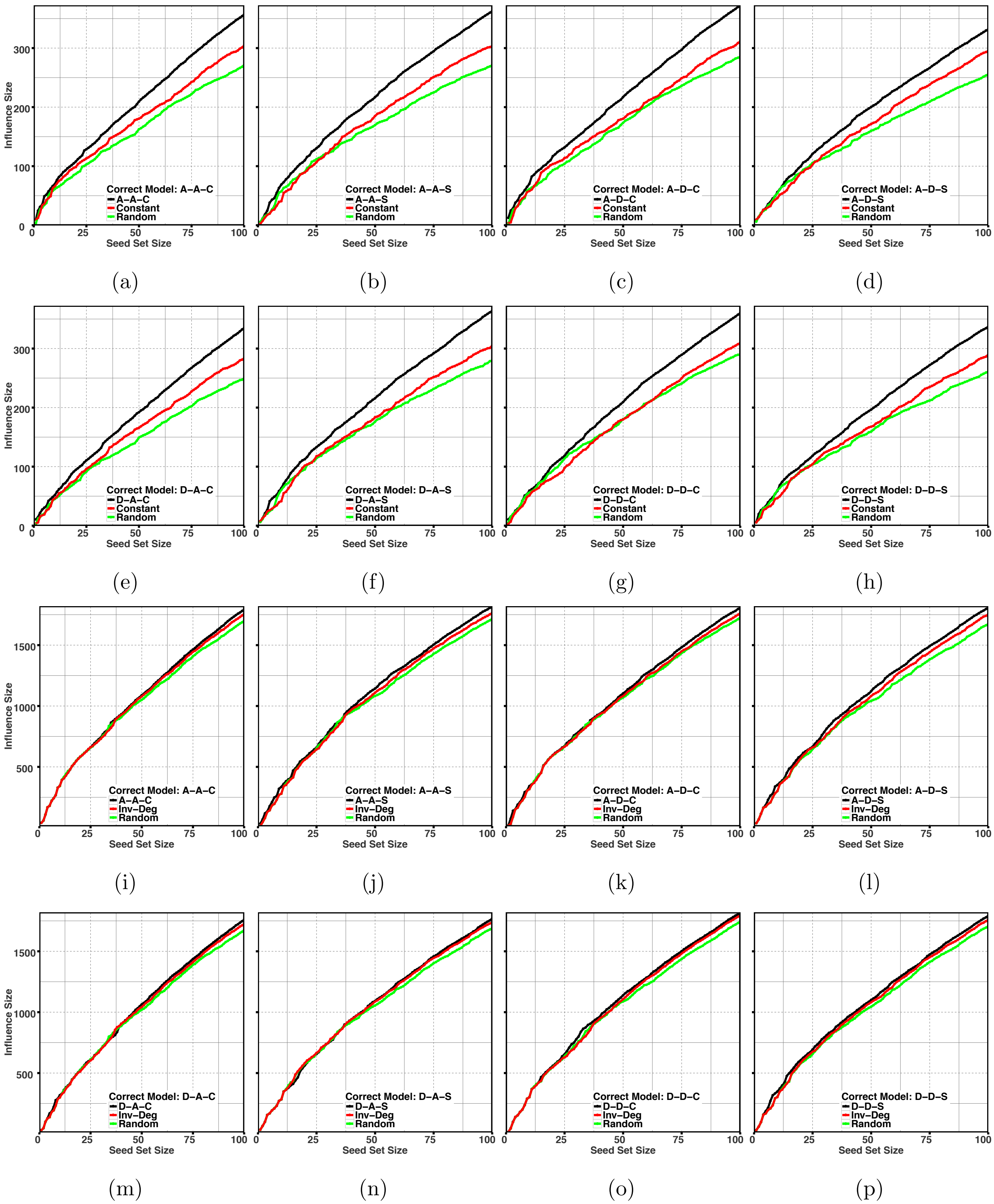
Supplementary Figure 28: Various seed characteristics for the IC model. **Network: Preferential Attachment, Nodes=10000, Edges =9999, Exponent = 0.5, Zero Appeal =2.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



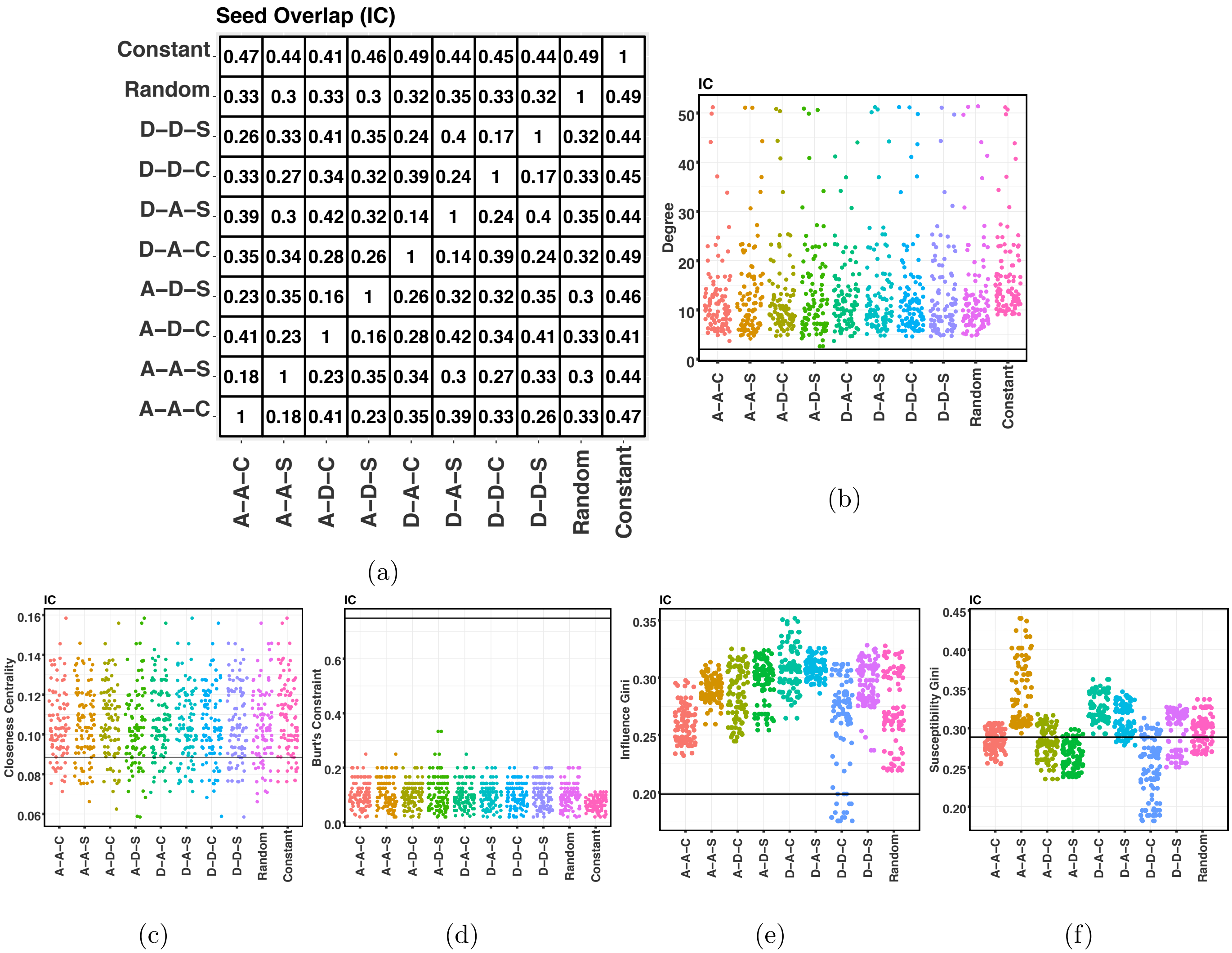
Supplementary Figure 29: Various seed characteristics for the LT model. **Network: Preferential Attachment, Nodes=10000, Edges =9999, Exponent = 0.5, Zero Appeal =2.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



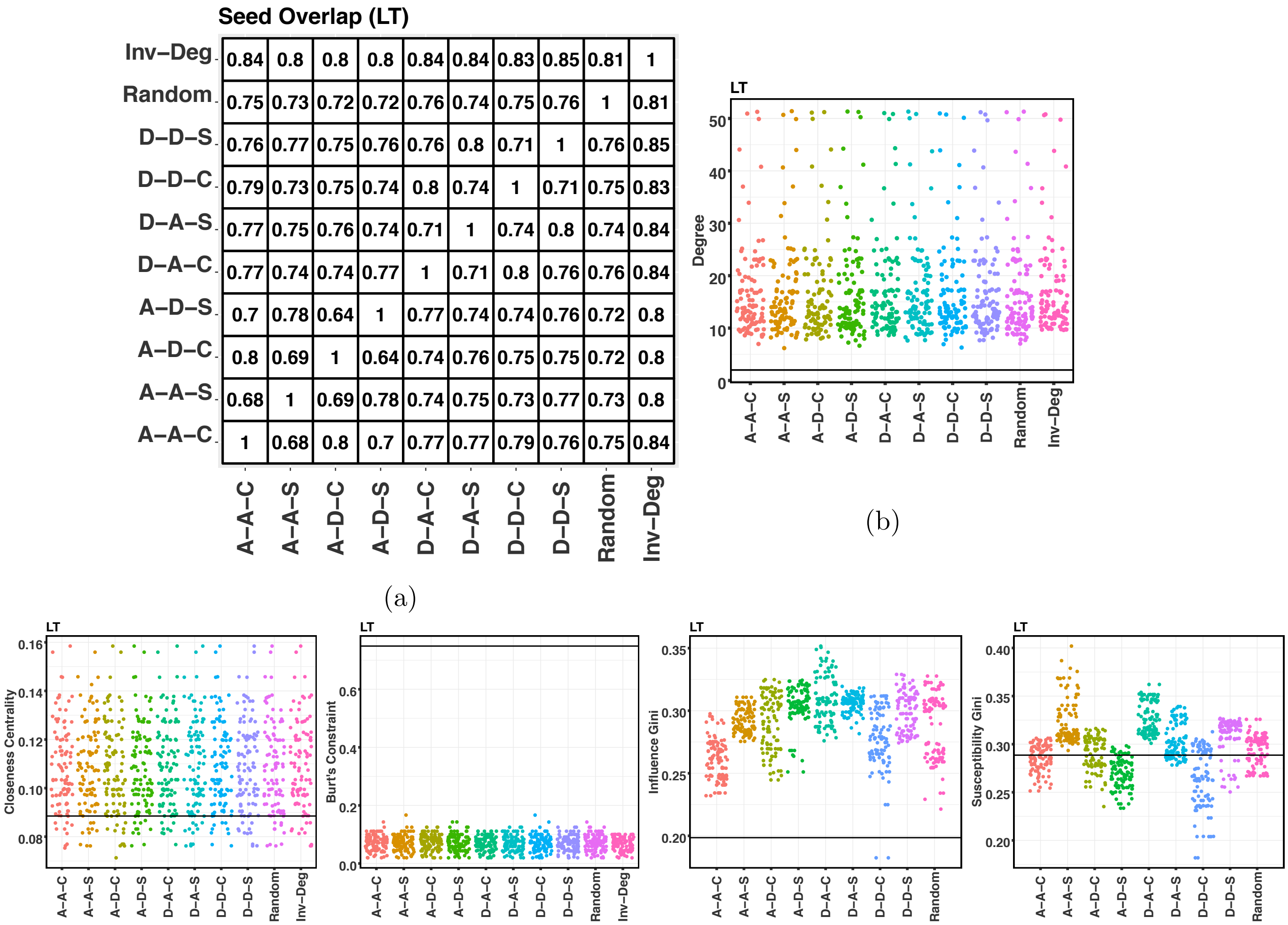
Supplementary Figure 30: Influence spread as a function of the seed set size. **Network: Preferential Attachment, Nodes=10000, Edges =9999, Exponent = 0.5, Zero Appeal =2.** Panels (a-h): IC Model, Panels (i-p): LT Model.



Supplementary Figure 31: Influence spread as a function of the seed set size. **Network: Preferential Attachment, Nodes=10000, Edges =9999, Exponent = 1, Zero Appeal =2.** Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are “Random” and “Constant” (all edges=0.1) and for the LT model are “Random” and “Inverse Degree”.

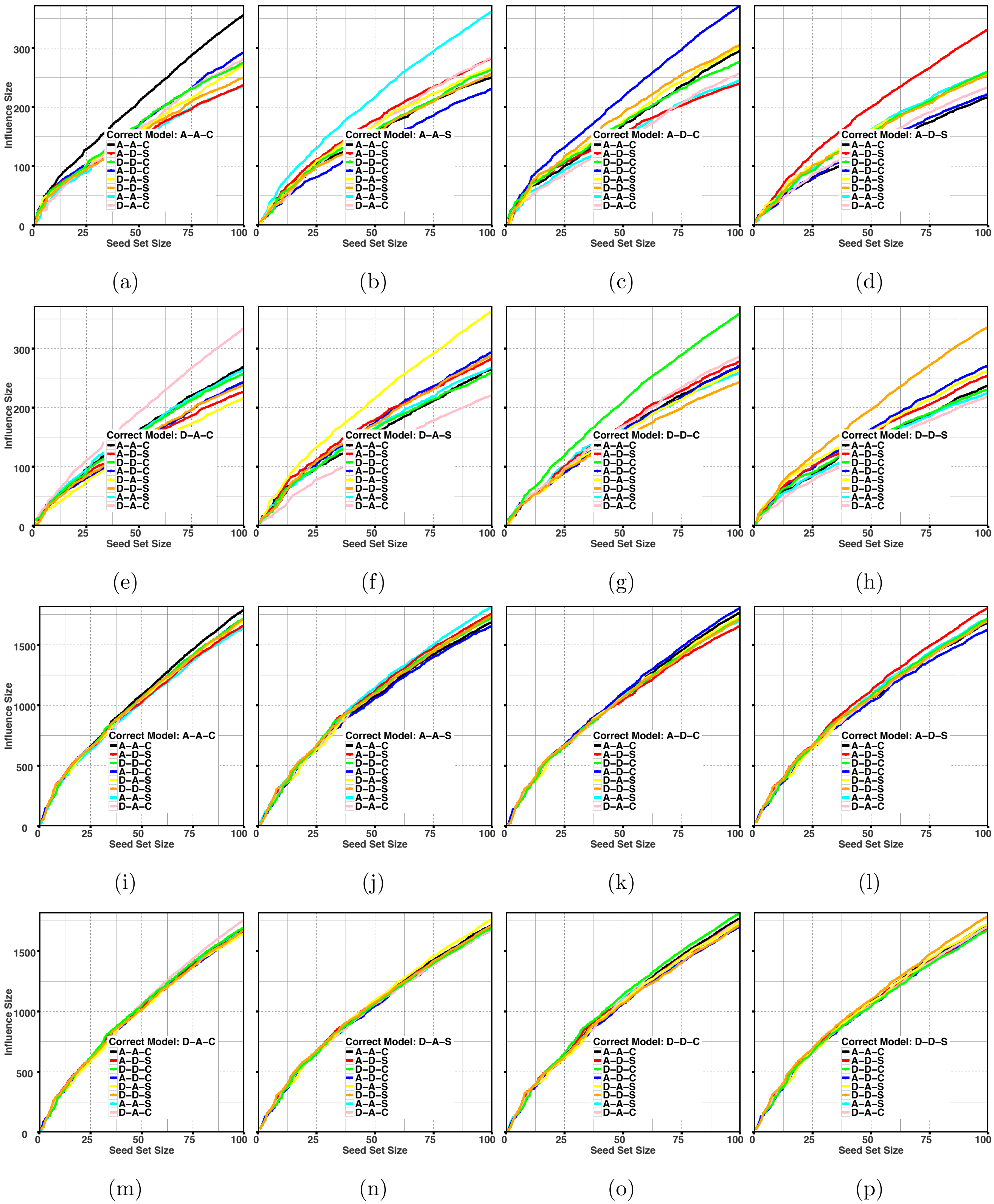


Supplementary Figure 32: Various seed characteristics for the IC model. **Network: Preferential Attachment, Nodes=10000, Edges =9999, Exponent = 1, Zero Appeal =2.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.

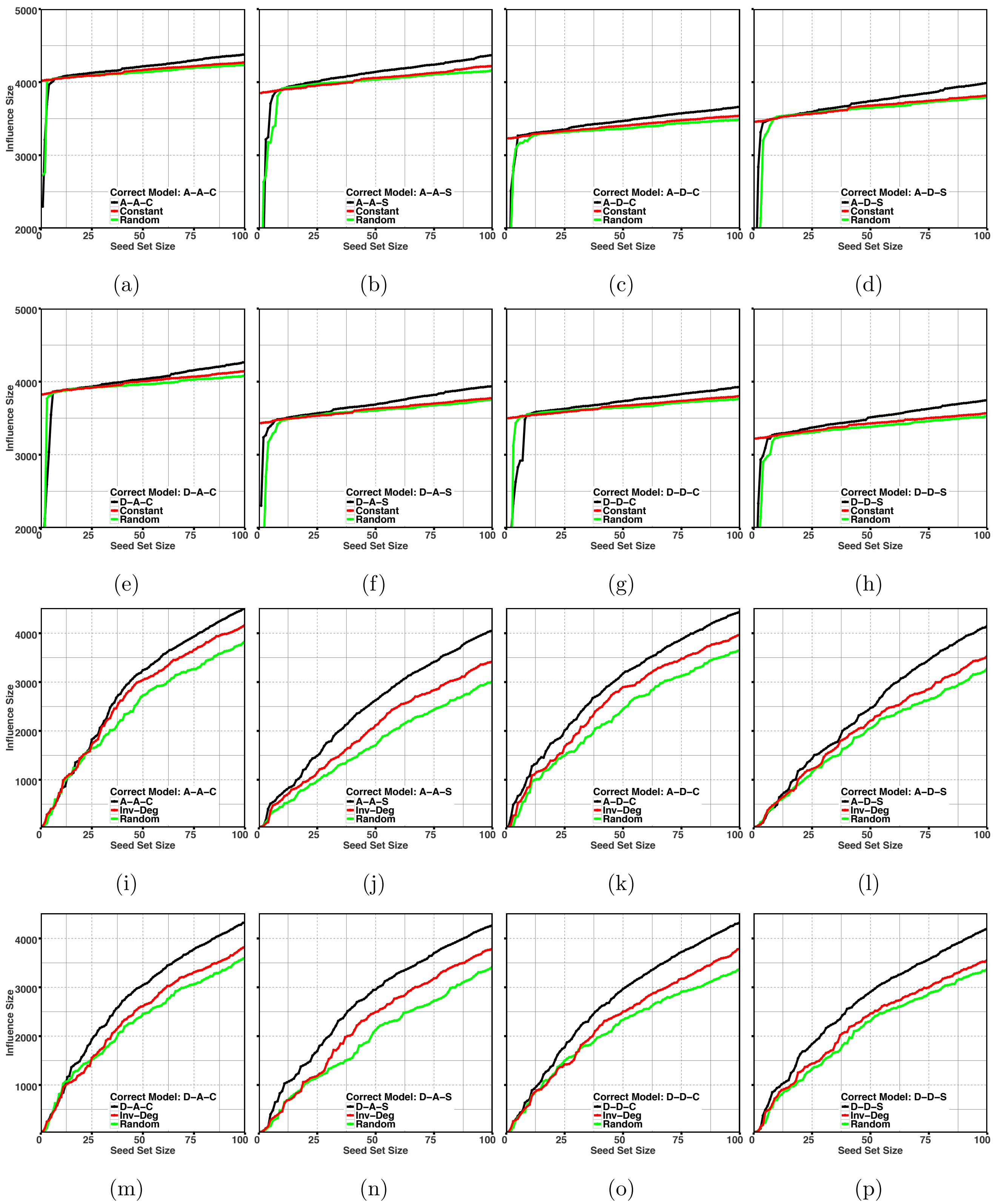


(a) (b) (c) (d) (e) (f)

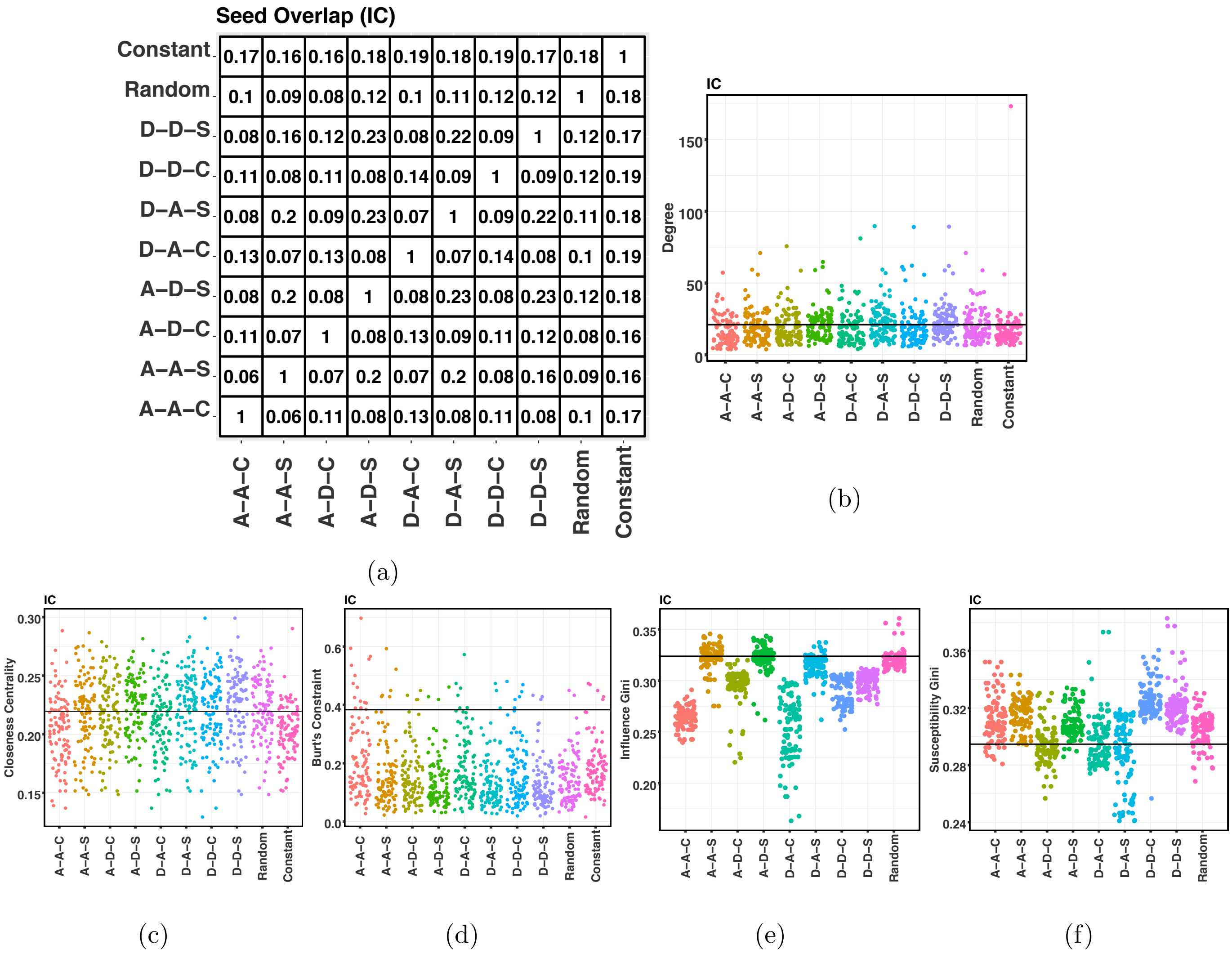
Supplementary Figure 33: Various seed characteristics for the LT model. **Network: Preferential Attachment, Nodes=10000, Edges =9999, Exponent = 1, Zero Appeal =2.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



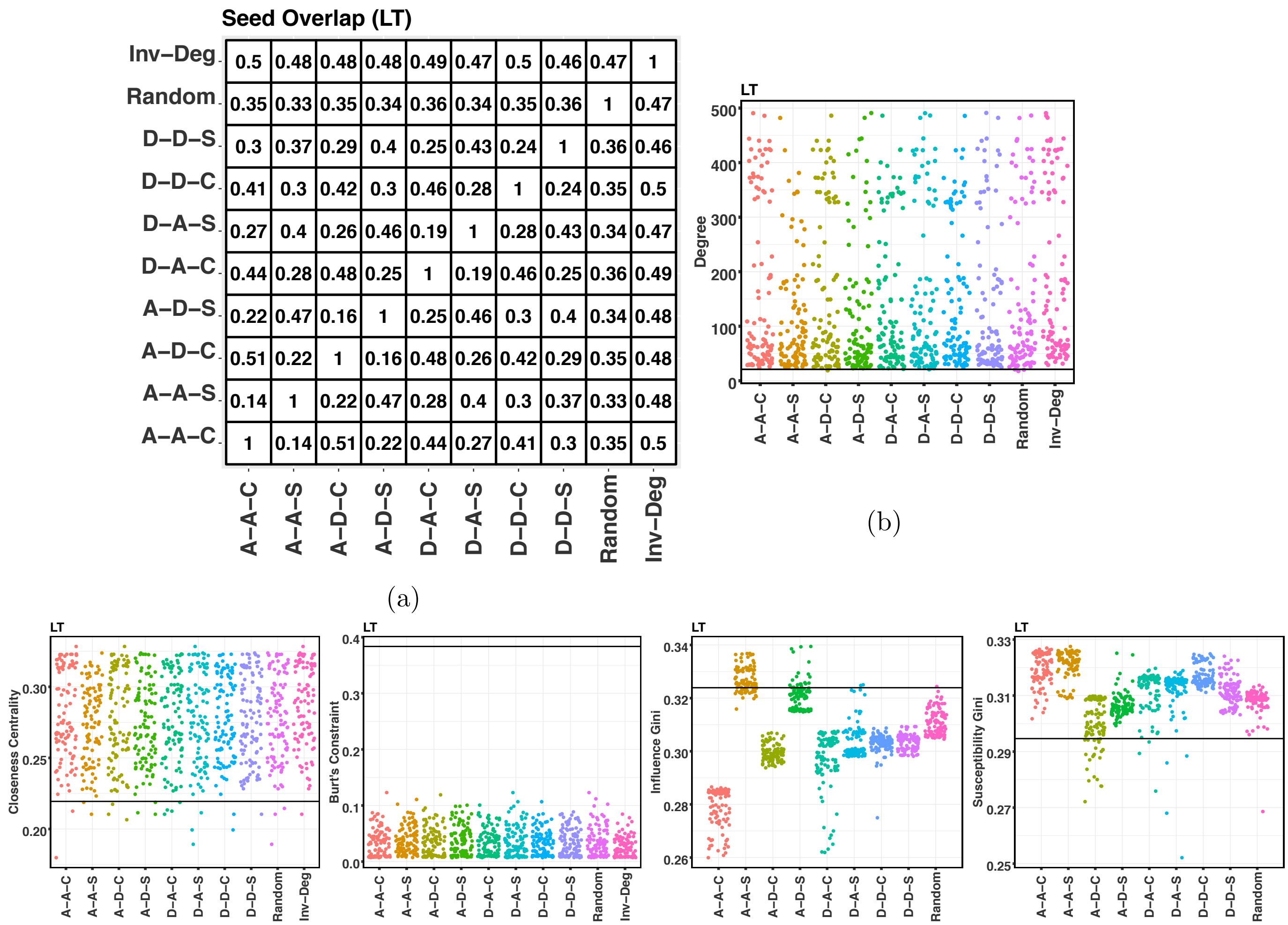
Supplementary Figure 34: Influence spread as a function of the seed set size. **Network: Preferential Attachment, Nodes=10000, Edges =9999, Exponent = 1, Zero Appeal =2.** Panels (a-h): IC Model, Panels (i-p): LT Model.



Supplementary Figure 35: Influence spread as a function of the seed set size. **Network: Collaboration Network Arxiv HEP-PH (High Energy Physics - Phenomenology), Nodes=11204, Edges =117619.** Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are “Random” and “Constant” (all edges=0.1) and for the LT model are “Random” and “Inverse Degree”.



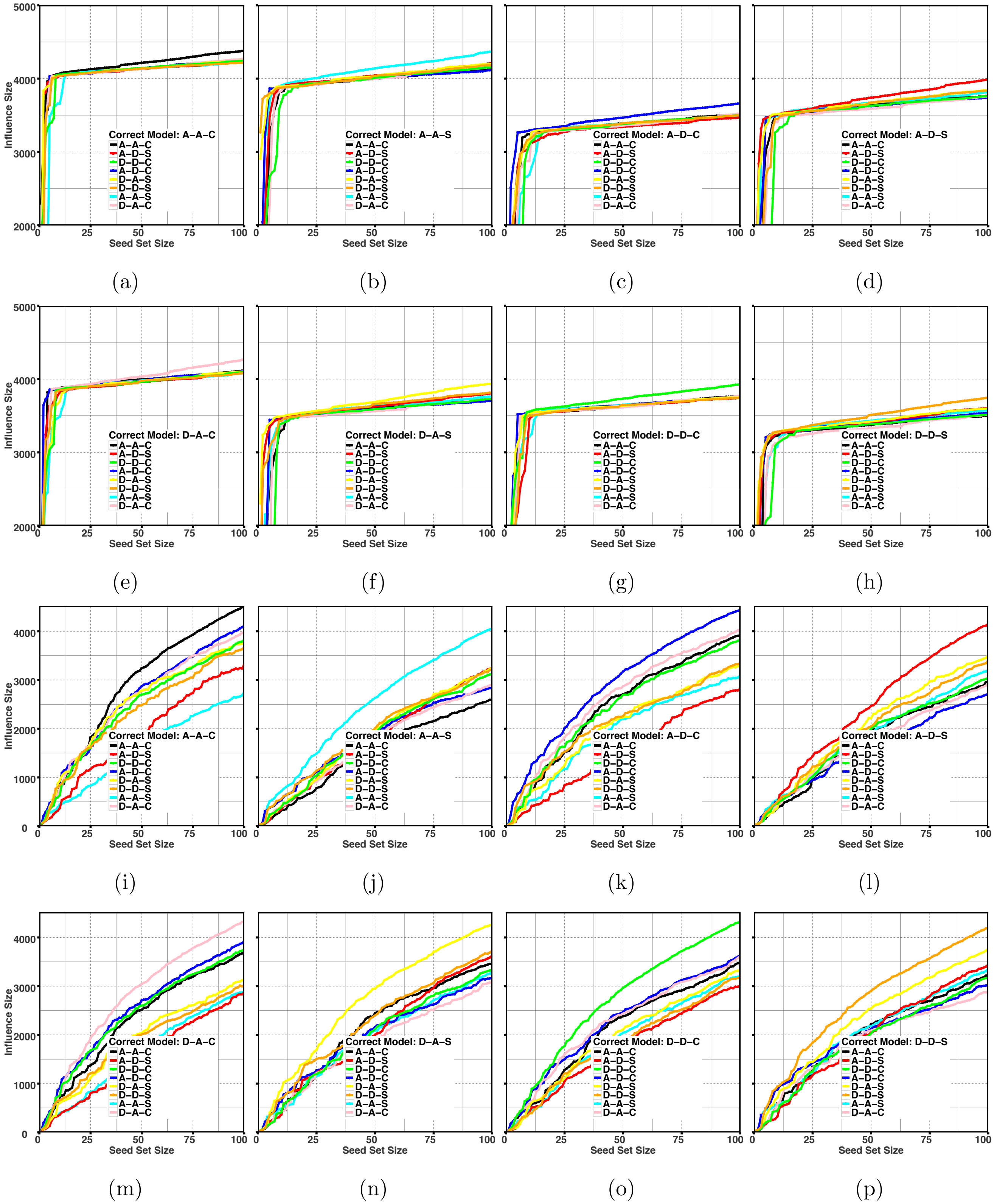
Supplementary Figure 36: Various seed characteristics for the IC model. **Network: Collaboration Network Arxiv HEP-PH (High Energy Physics - Phenomenology), Nodes=11204, Edges =117619.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



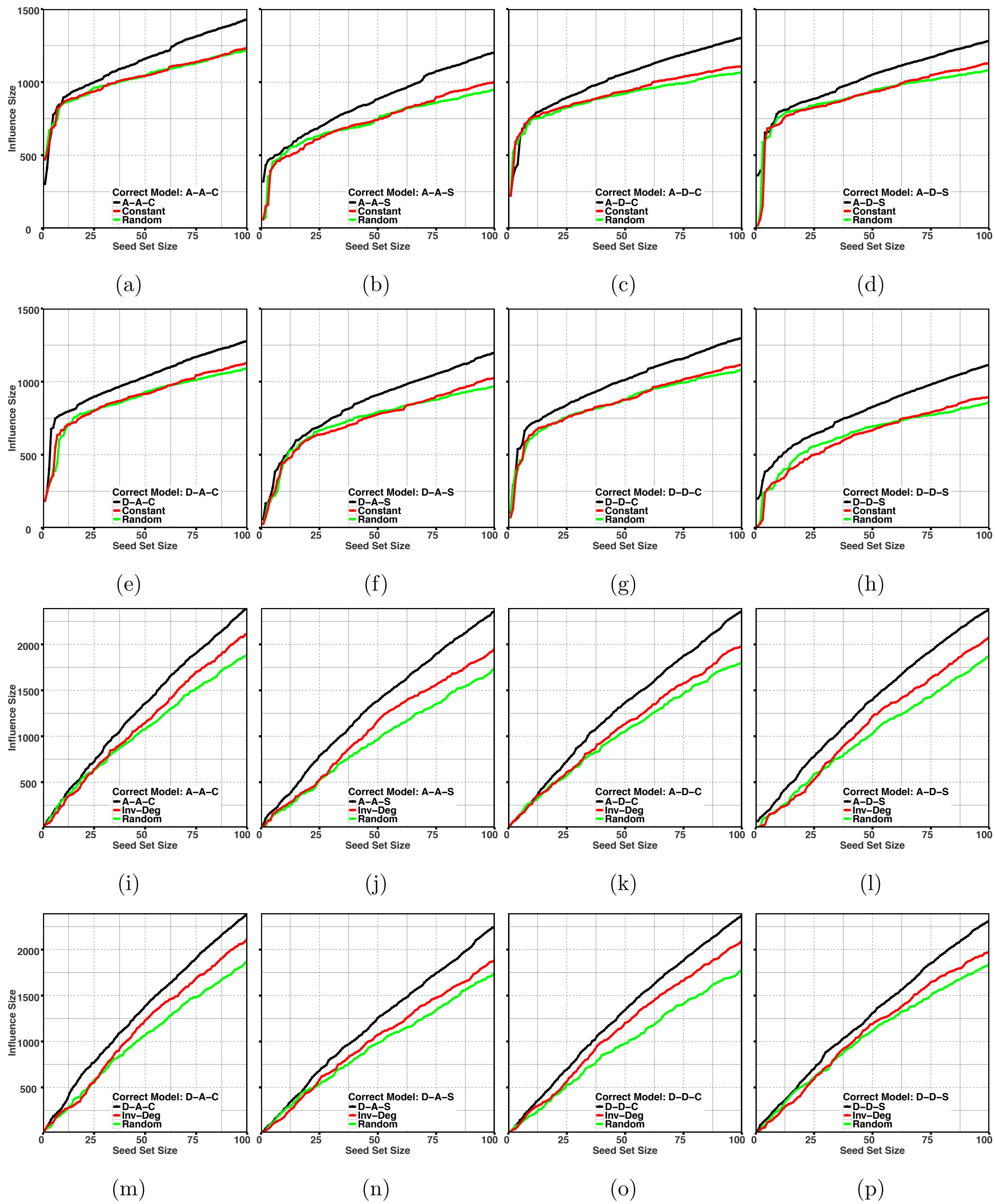
(a) (b)

(c) (d) (e) (f)

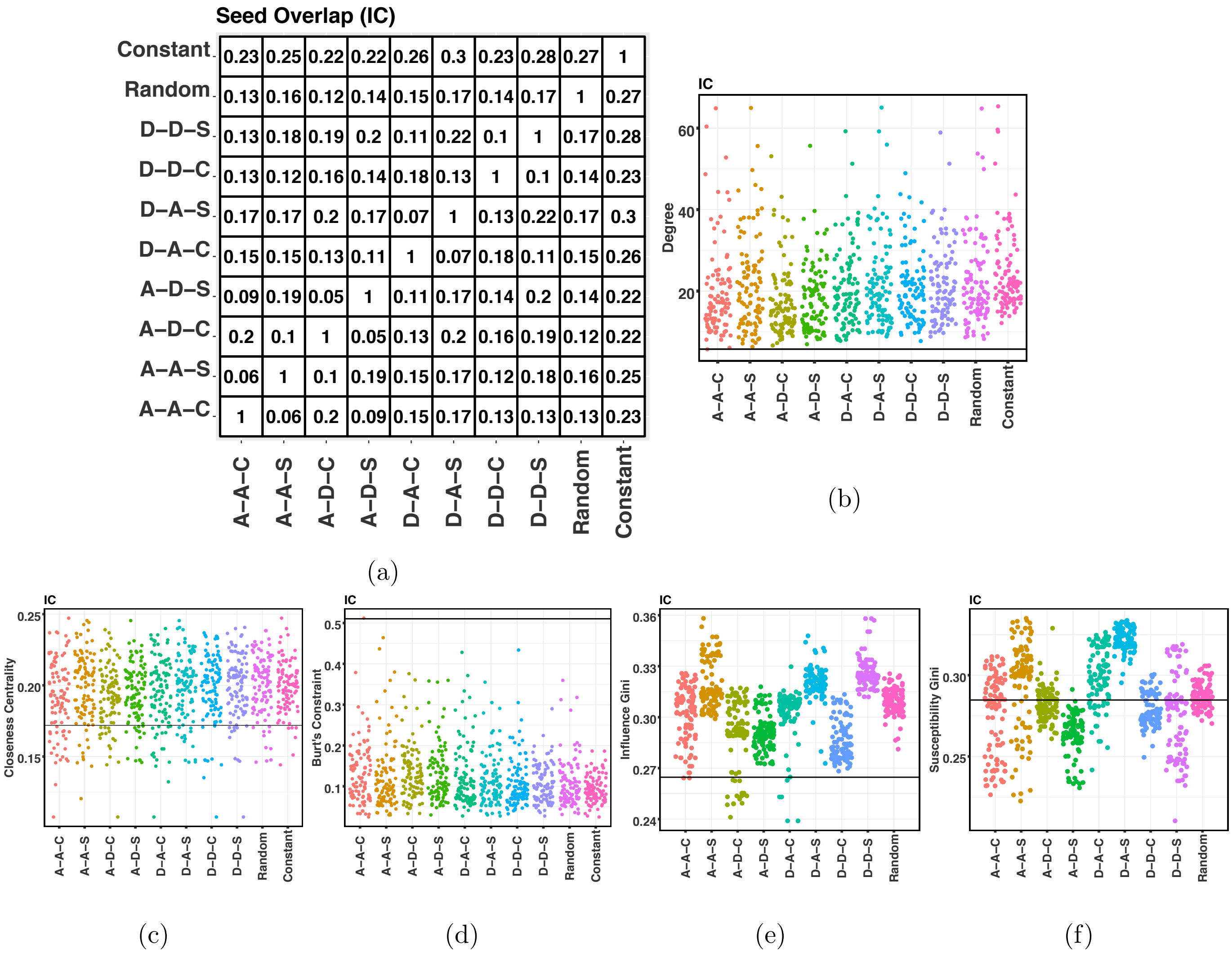
Supplementary Figure 37: Various seed characteristics for the LT model. **Network: Network: Collaboration Network Arxiv HEP-PH (High Energy Physics - Phenomenology), Nodes=11204, Edges =117619.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



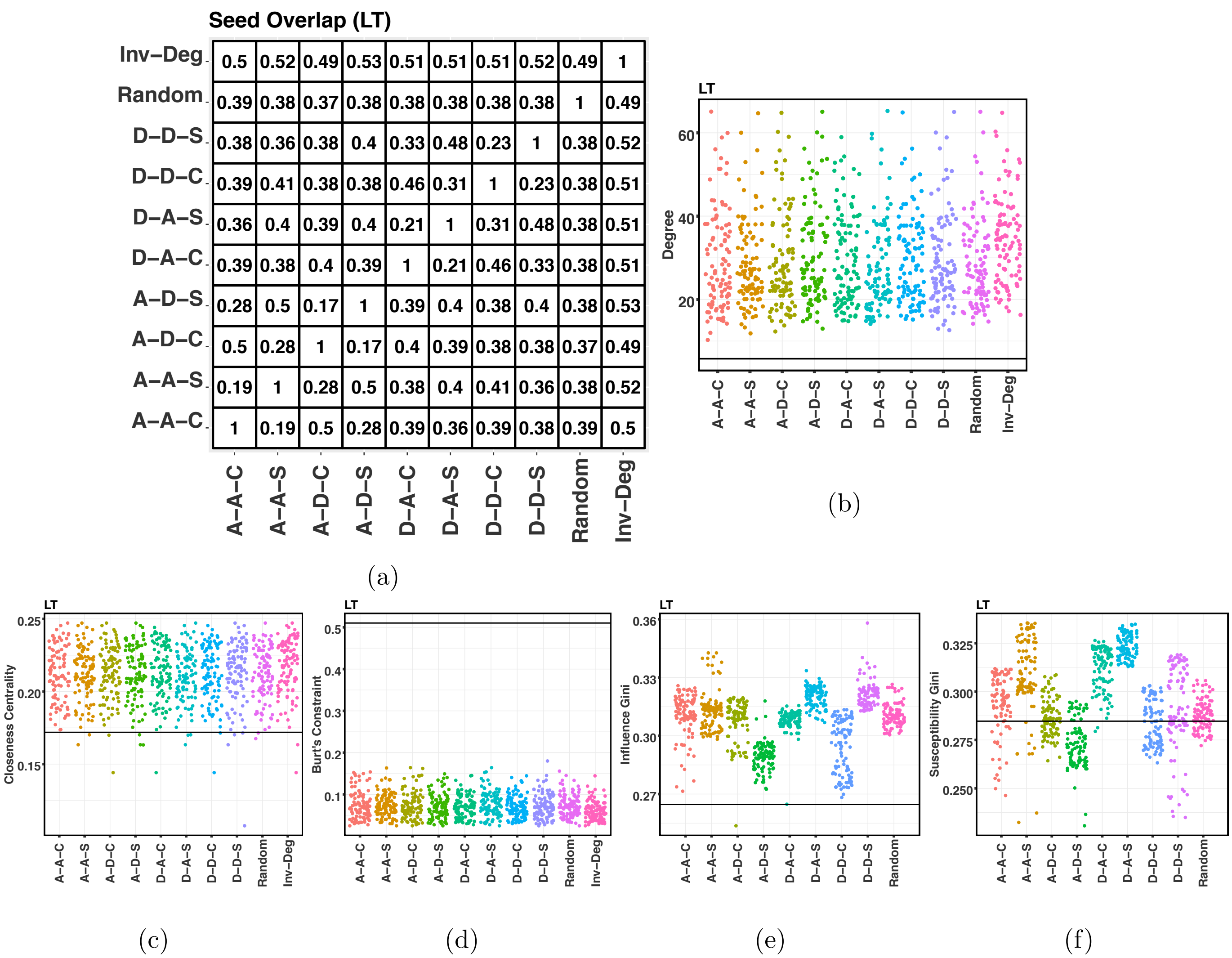
Supplementary Figure 38: Influence spread as a function of the seed set size. **Network: Network: Collaboration Network Arxiv HEP-PH (High Energy Physics - Phenomenology), Nodes=11204, Edges =117619.** Panels (a-h): IC Model, Panels (i-p): LT Model.



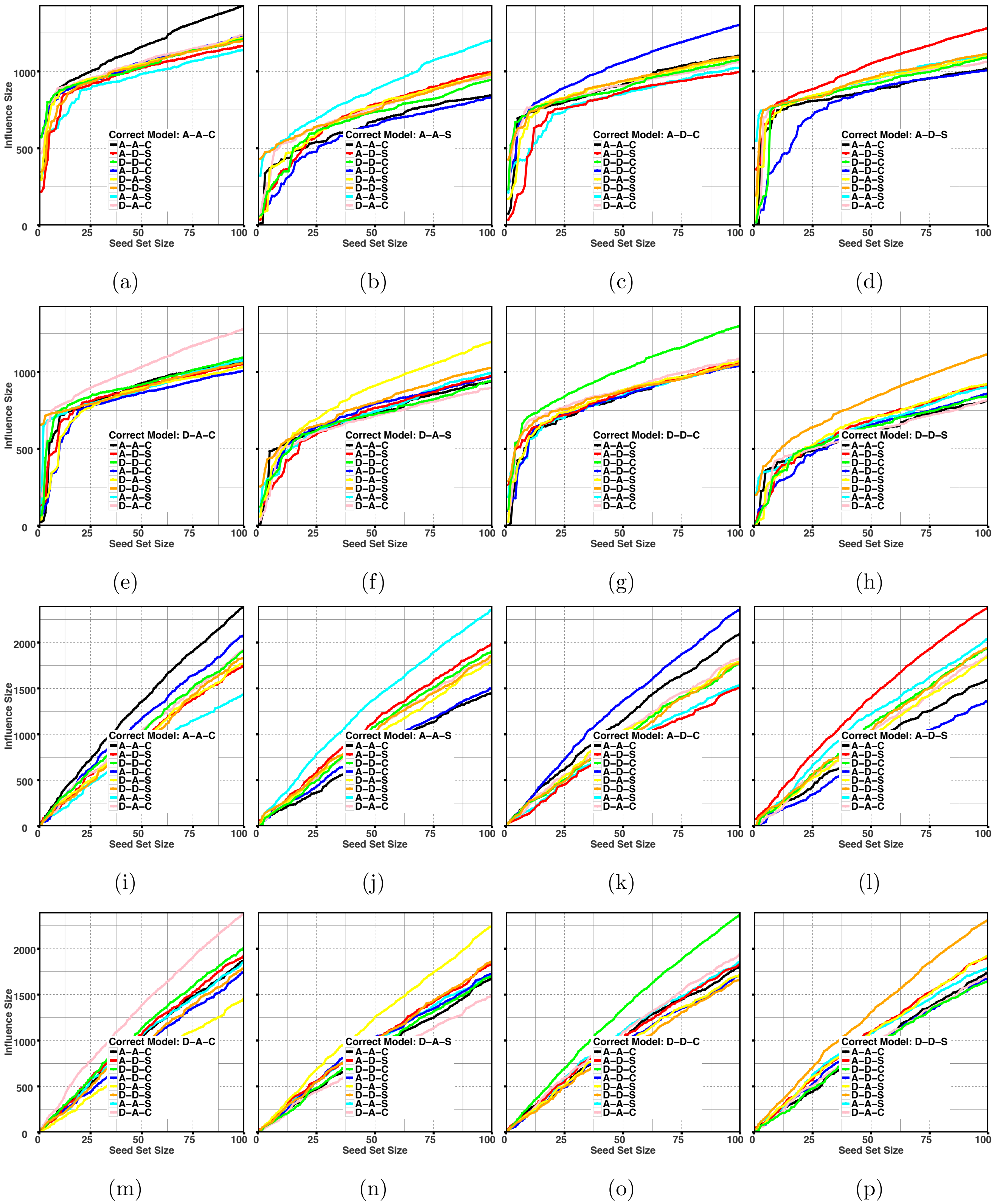
Supplementary Figure 39: Influence spread as a function of the seed set size. **Network: Collaboration Network Arxiv HEP-TH (High Energy Physics - Theory), Nodes=8638, Edges =24806.** Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are "Random" and "Constant" (all edges=0.1) and for the LT model are "Random" and "Inverse Degree".



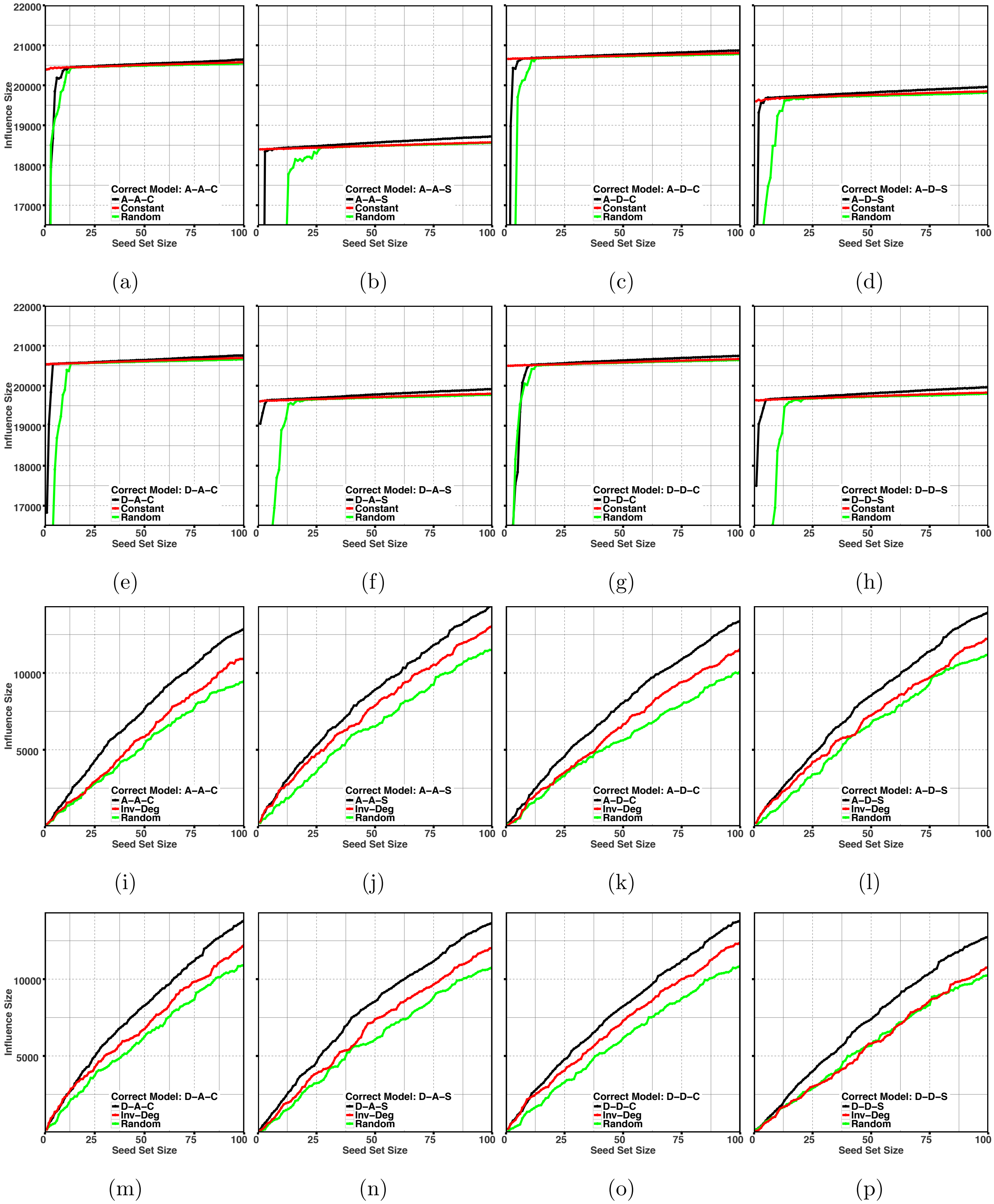
Supplementary Figure 40: Various seed characteristics for the IC model. **Network: Collaboration Network Arxiv HEP-TH (High Energy Physics - Theory), Nodes=8638, Edges =24806.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



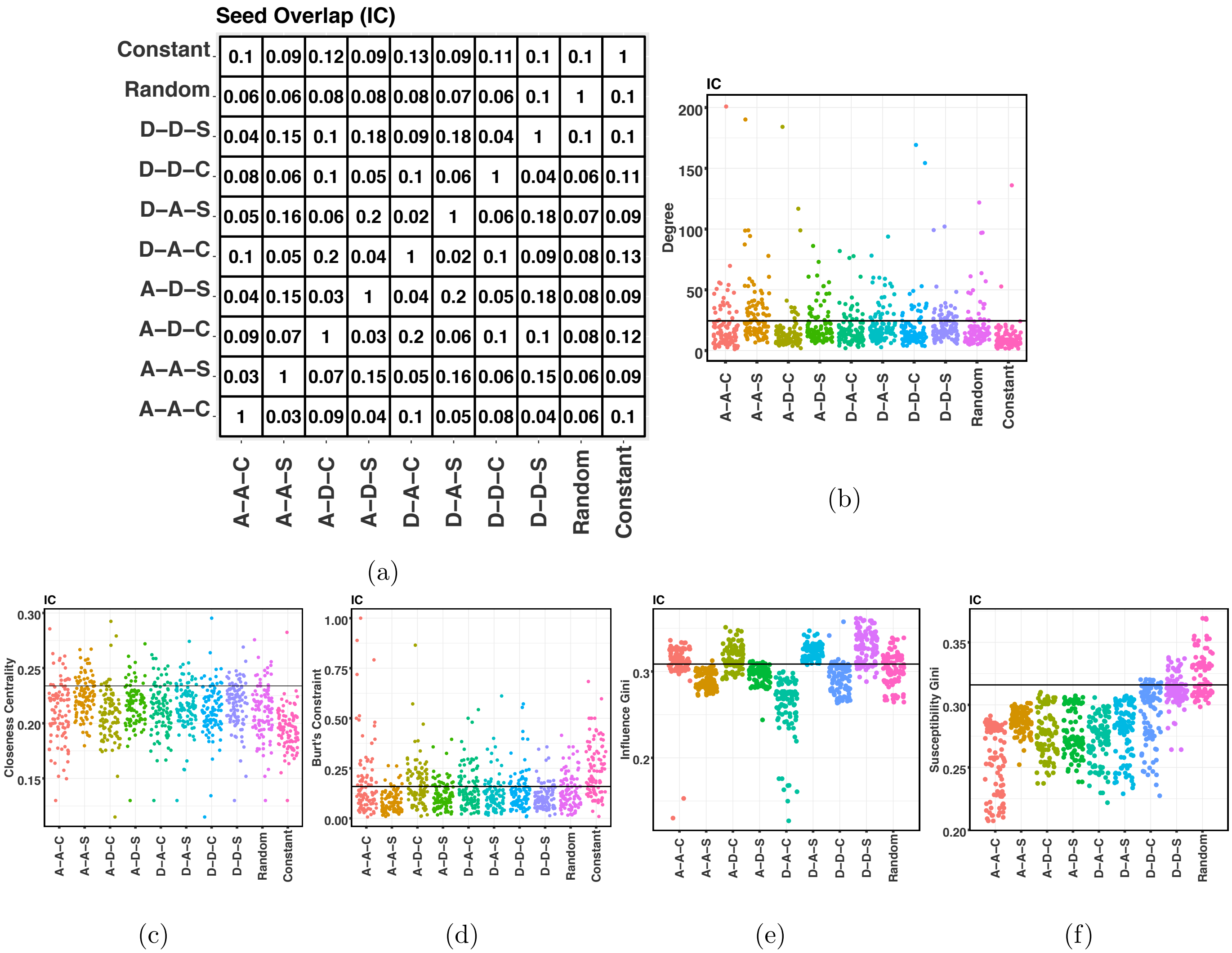
Supplementary Figure 41: Various seed characteristics for the LT model. **Network: Collaboration Network Arxiv HEP-TH (High Energy Physics - Theory), Nodes=8638, Edges =24806.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



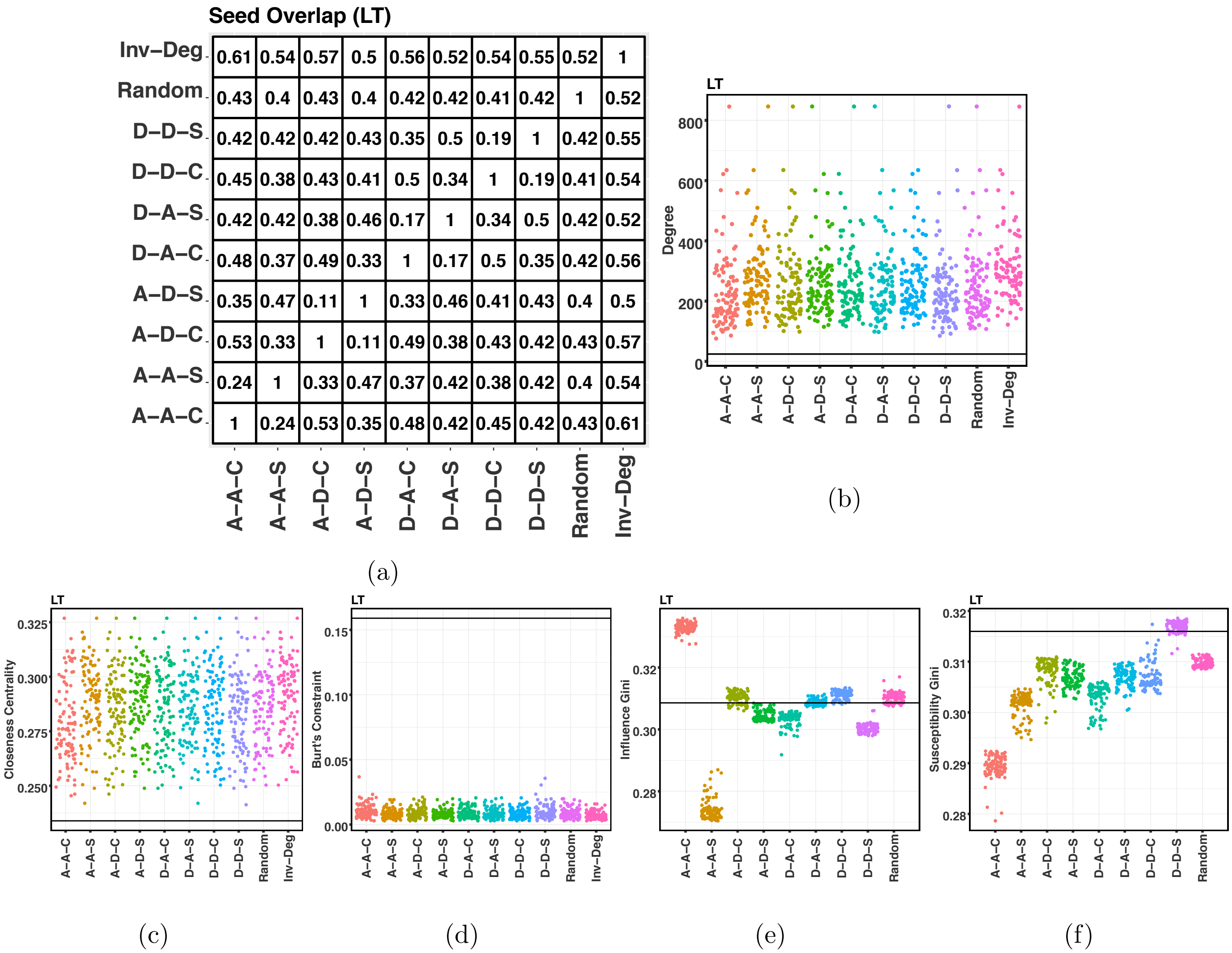
Supplementary Figure 42: Influence spread as a function of the seed set size. **Network: Collaboration Network Arxiv HEP-TH (High Energy Physics - Theory), Nodes=8638, Edges =24806.** Panels (a-h): IC Model, Panels (i-p): LT Model.



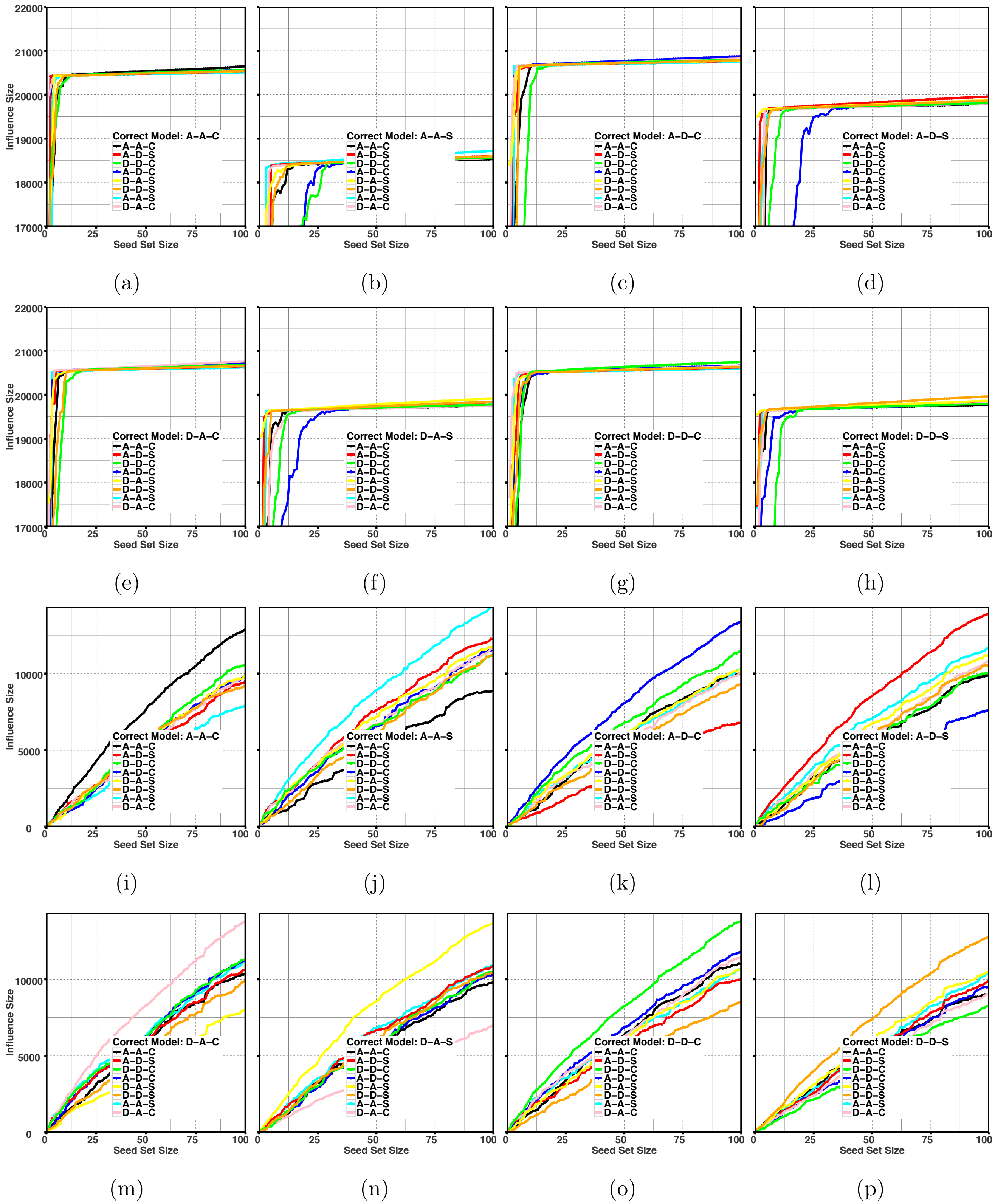
Supplementary Figure 43: Influence spread as a function of the seed set size. **Network: Citation Network Arxiv HEP-PH (High Energy Physics - Phenomenology), Nodes=34401, Edges =420784.** Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are "Random" and "Constant" (all edges=0.1) and for the LT model are "Random" and "Inverse Degree".



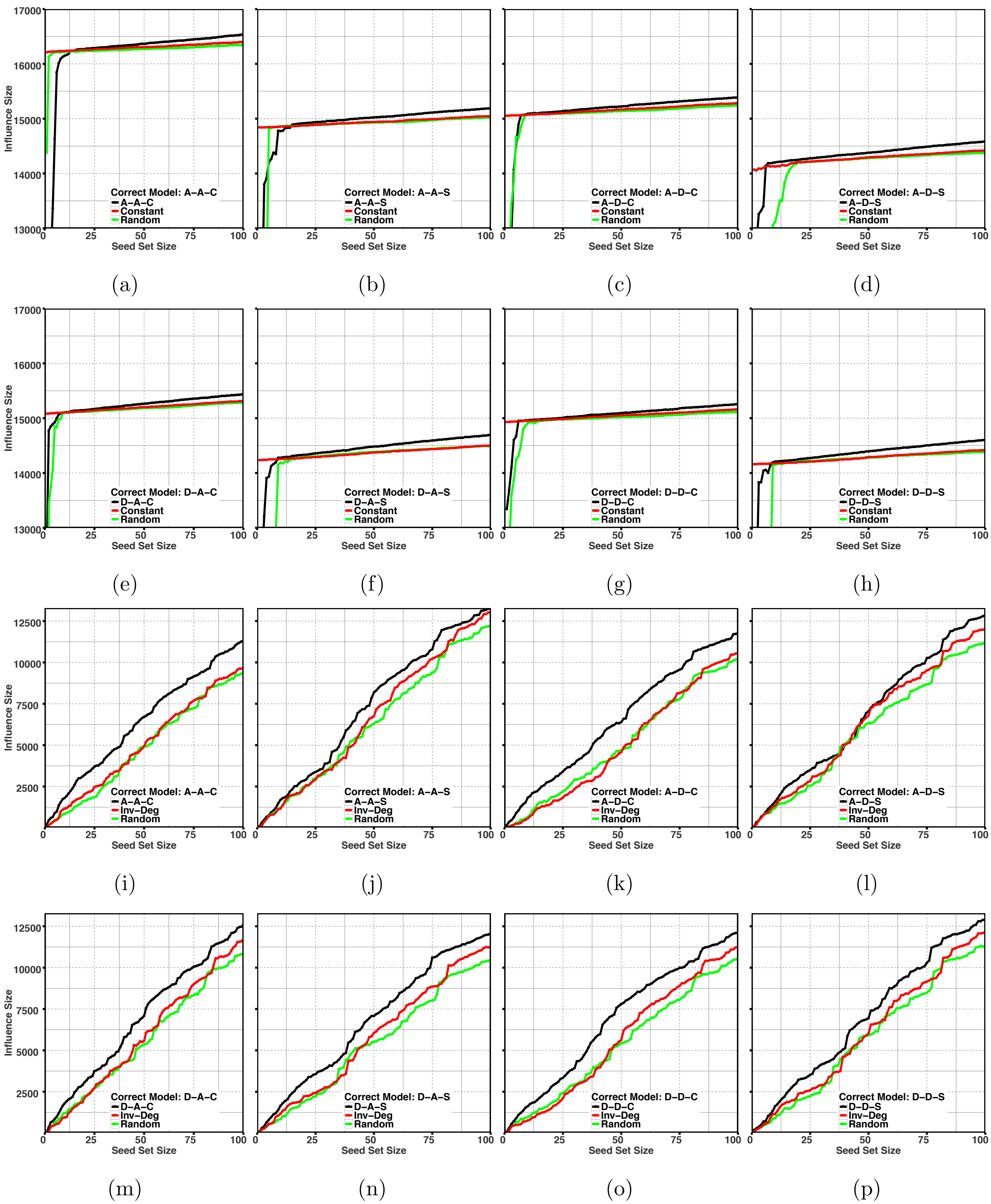
Supplementary Figure 44: Various seed characteristics for the IC model. **Network: Citation Network Arxiv HEP-PH (High Energy Physics - Phenomenology), Nodes=34401, Edges =420784.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



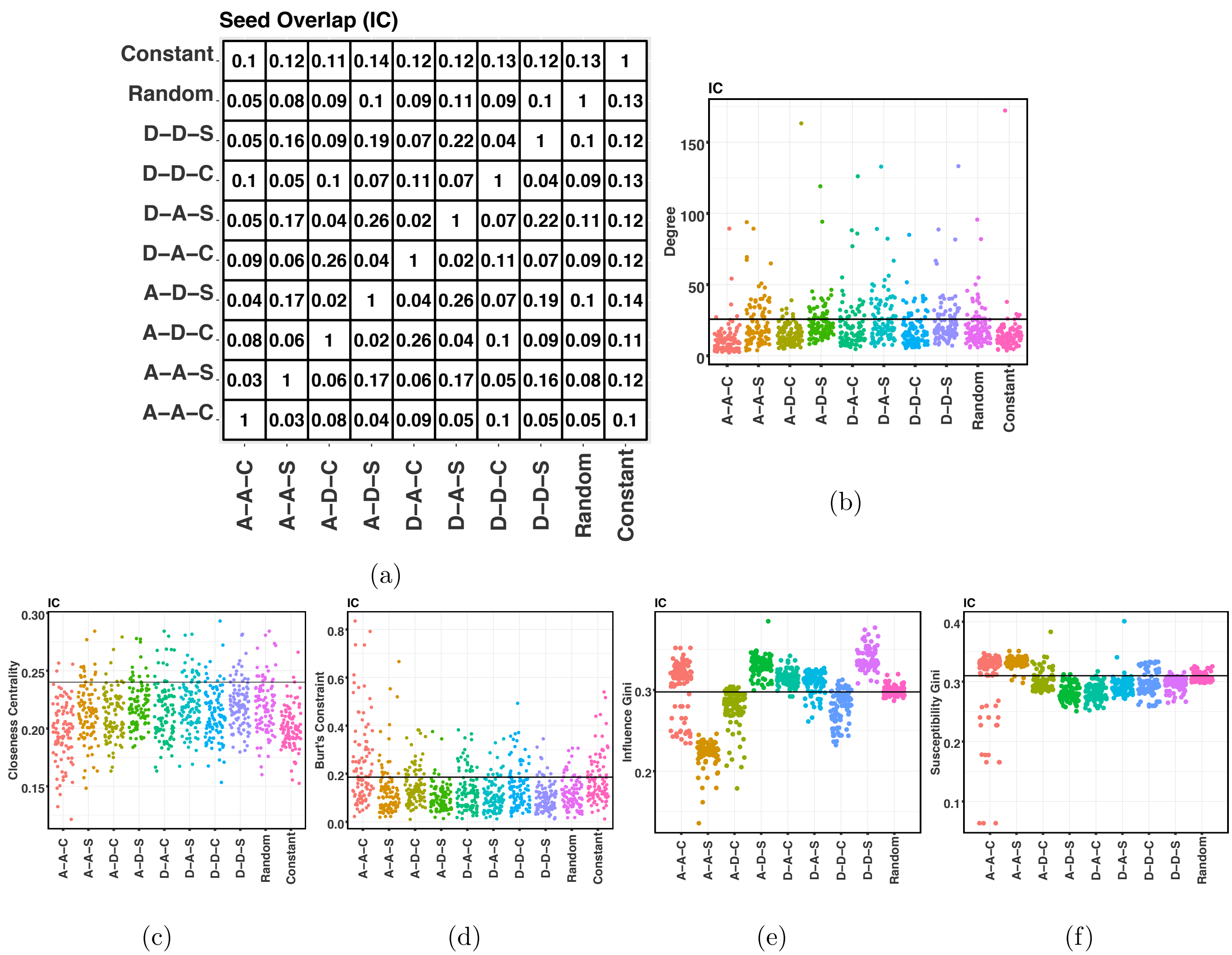
Supplementary Figure 45: Various seed characteristics for the LT model. **Network: Citation Network Arxiv HEP-PH (High Energy Physics - Phenomenology), Nodes=34401, Edges =420784.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



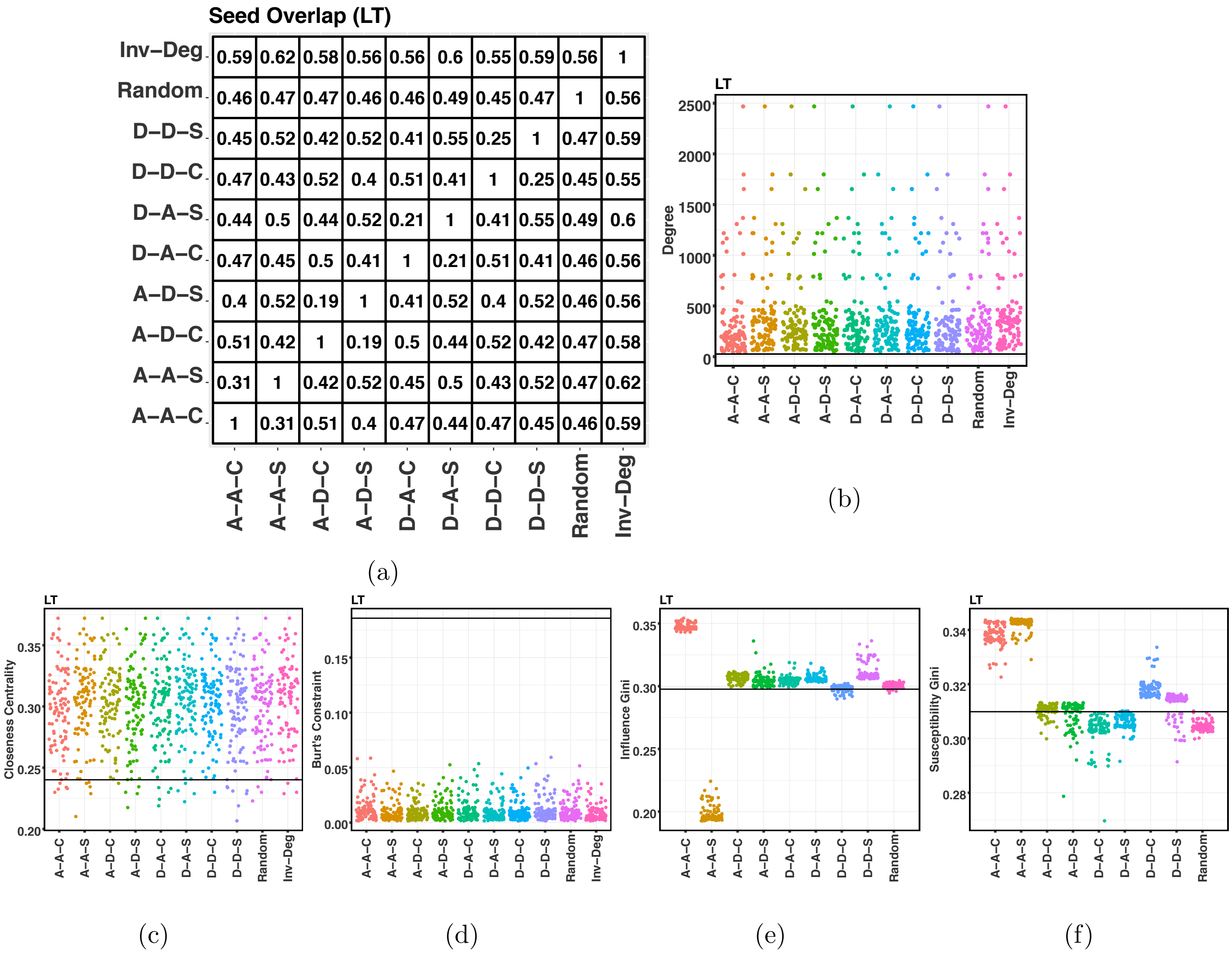
Supplementary Figure 46: Influence spread as a function of the seed set size. **Network: Citation Network Arxiv HEP-PH (High Energy Physics - Phenomenology), Nodes=34401, Edges =420784.** Panels (a-h): IC Model, Panels (i-p): LT Model.



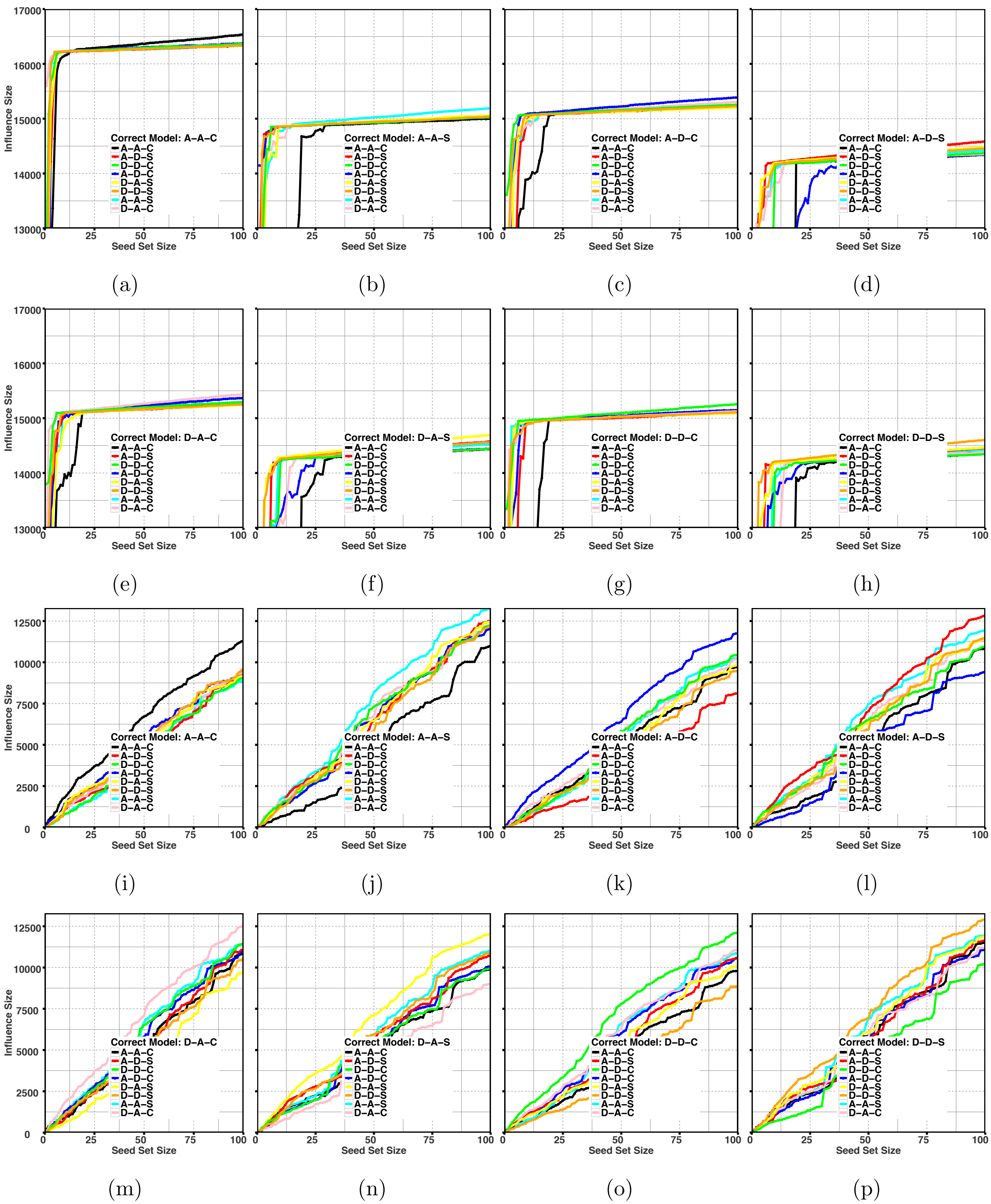
Supplementary Figure 47: Influence spread as a function of the seed set size. **Network: Citation Network Arxiv HEP-TH (High Energy Physics - Theory), Nodes=27400, Edges =352021.** Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are “Random” and “Constant” (all edges=0.1) and for the LT model are “Random” and “Inverse Degree”.



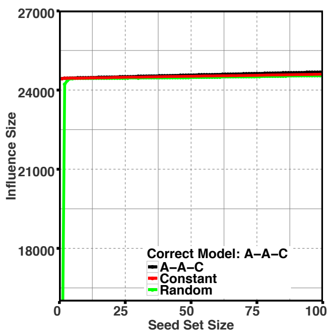
Supplementary Figure 48: Various seed characteristics for the IC model. **Network: Citation Network Arxiv HEP-TH (High Energy Physics - Theory), Nodes=27400, Edges =352021.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



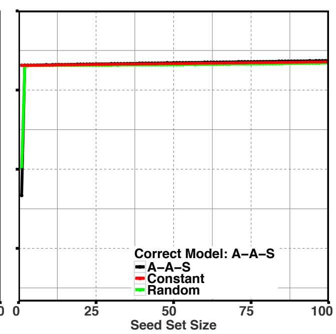
Supplementary Figure 49: Various seed characteristics for the LT model. **Network: Citation Network Arxiv HEP-TH (High Energy Physics - Theory), Nodes=27400, Edges =352021.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



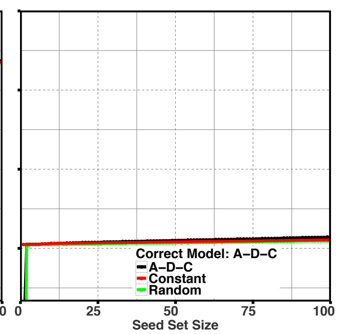
Supplementary Figure 50: Influence spread as a function of the seed set size. **Network: Citation Network Arxiv HEP-TH (High Energy Physics - Theory), Nodes=27400, Edges =352021.** Panels (a-h): IC Model, Panels (i-p): LT Model.



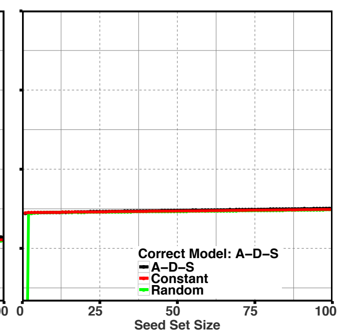
(a)



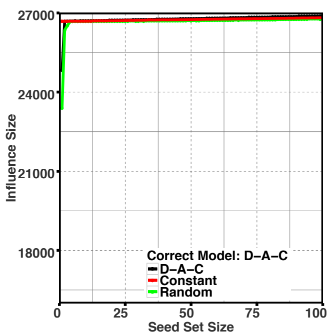
(b)



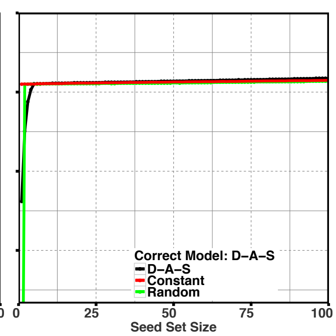
(c)



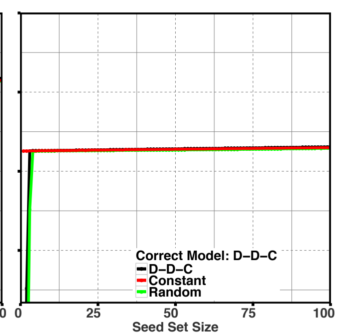
(d)



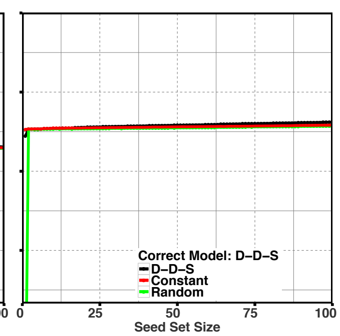
(e)



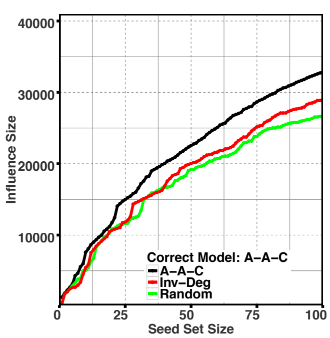
(f)



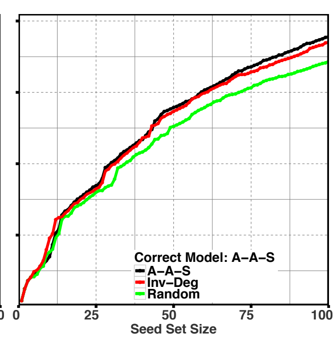
(g)



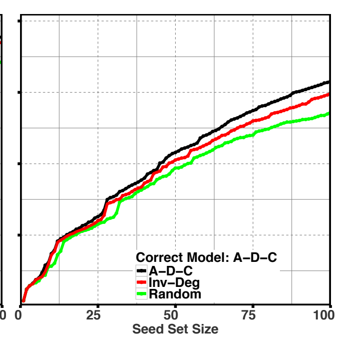
(h)



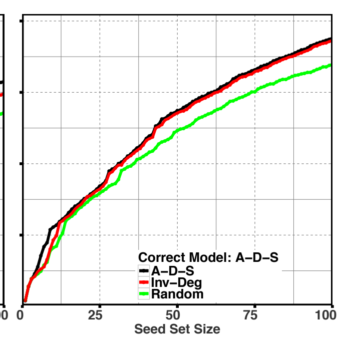
(i)



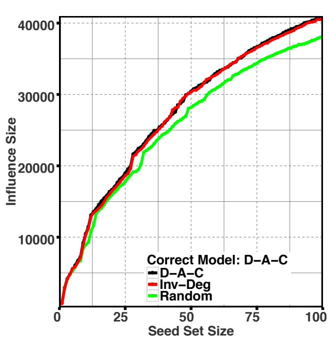
(j)



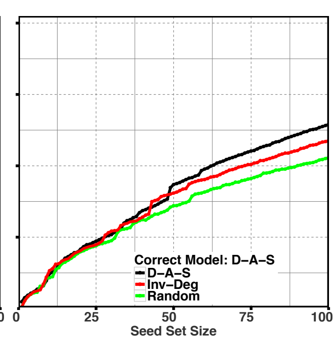
(k)



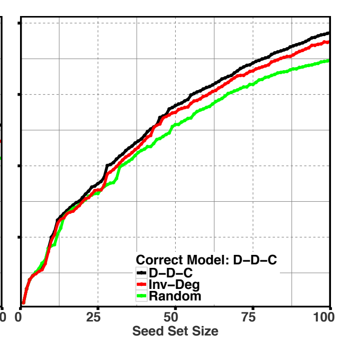
(l)



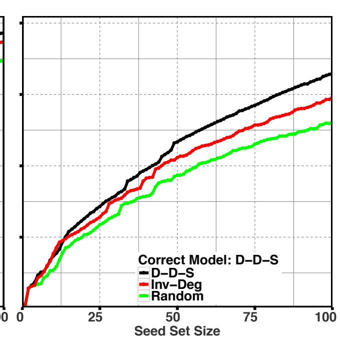
(m)



(n)

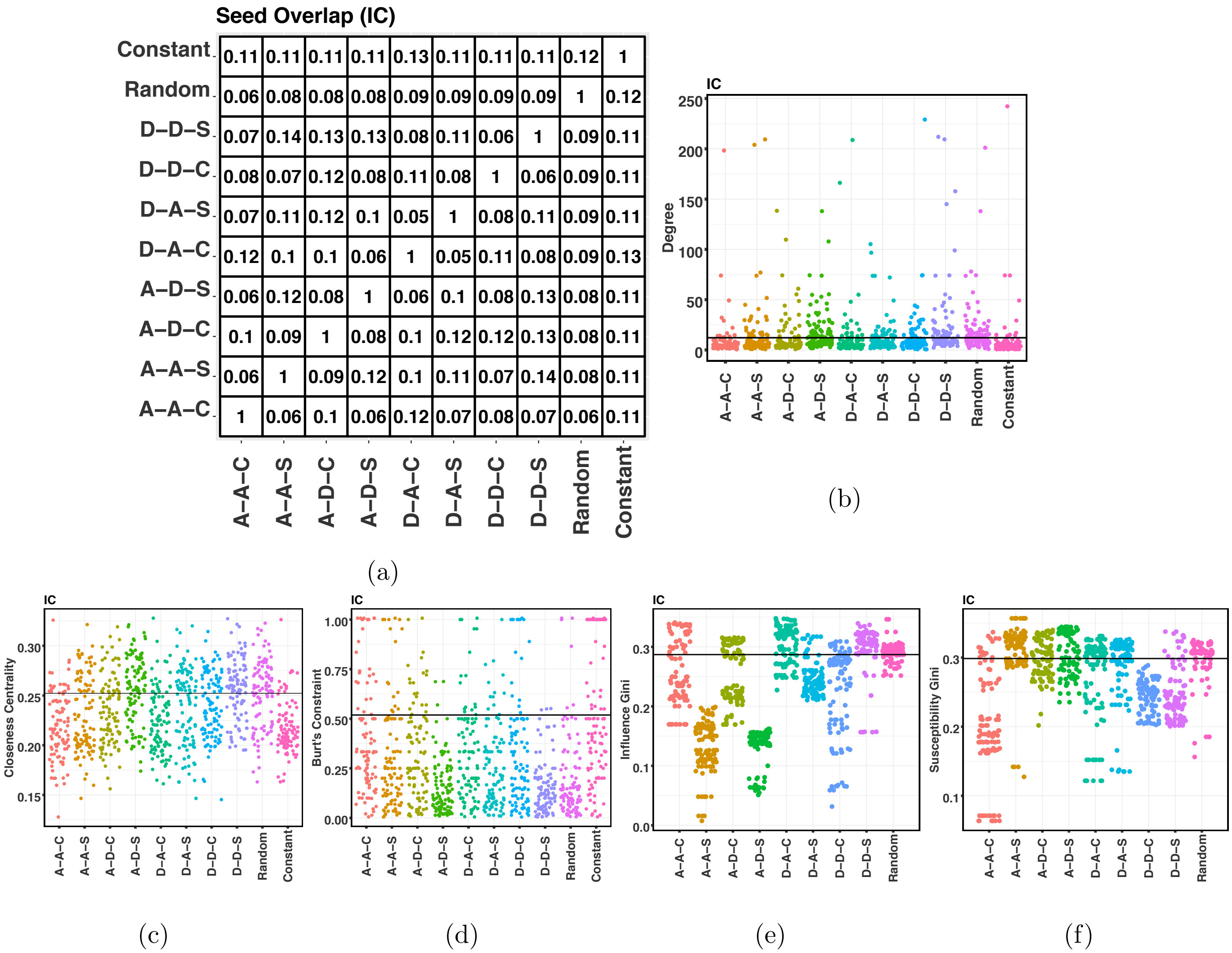


(o)

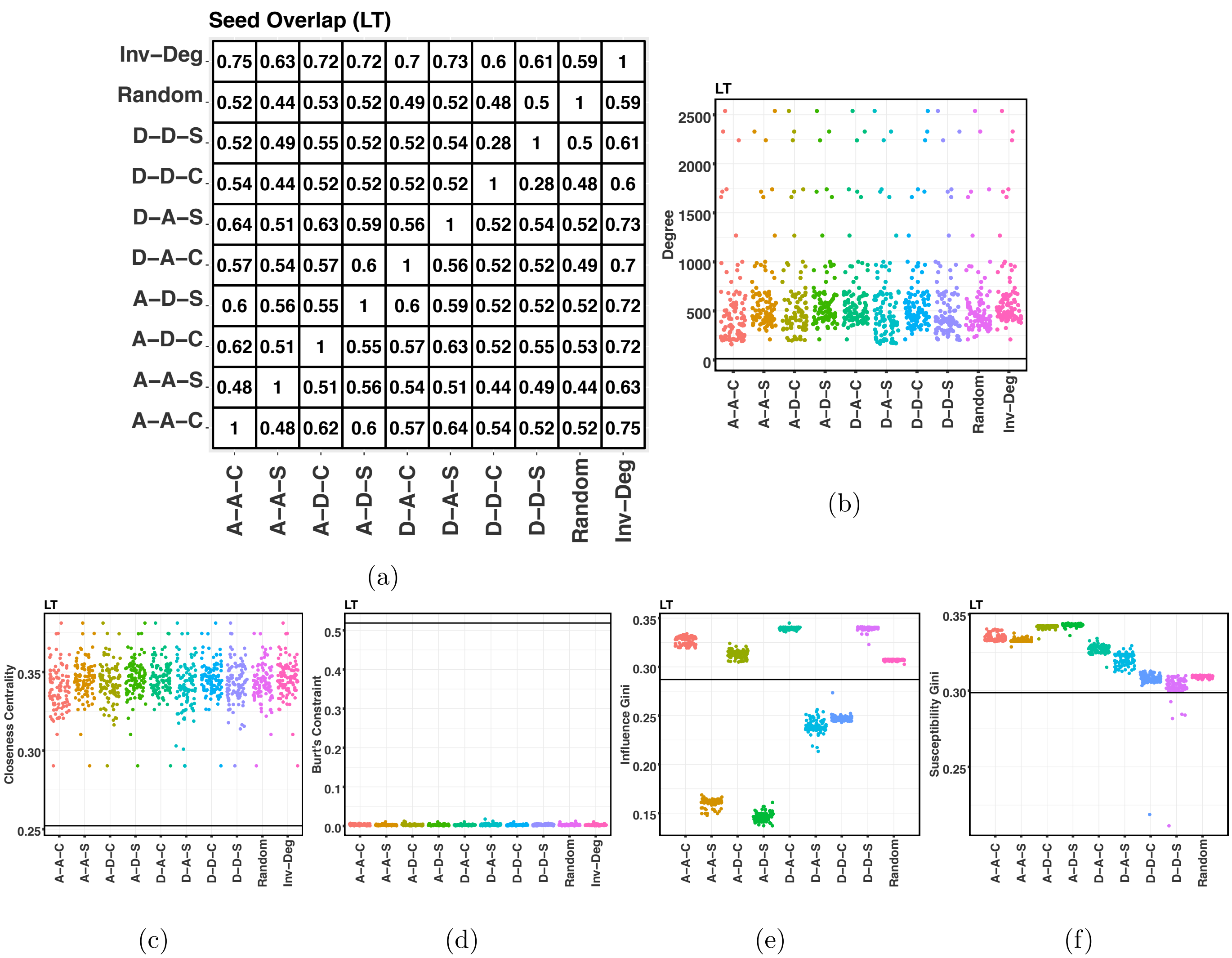


(p)

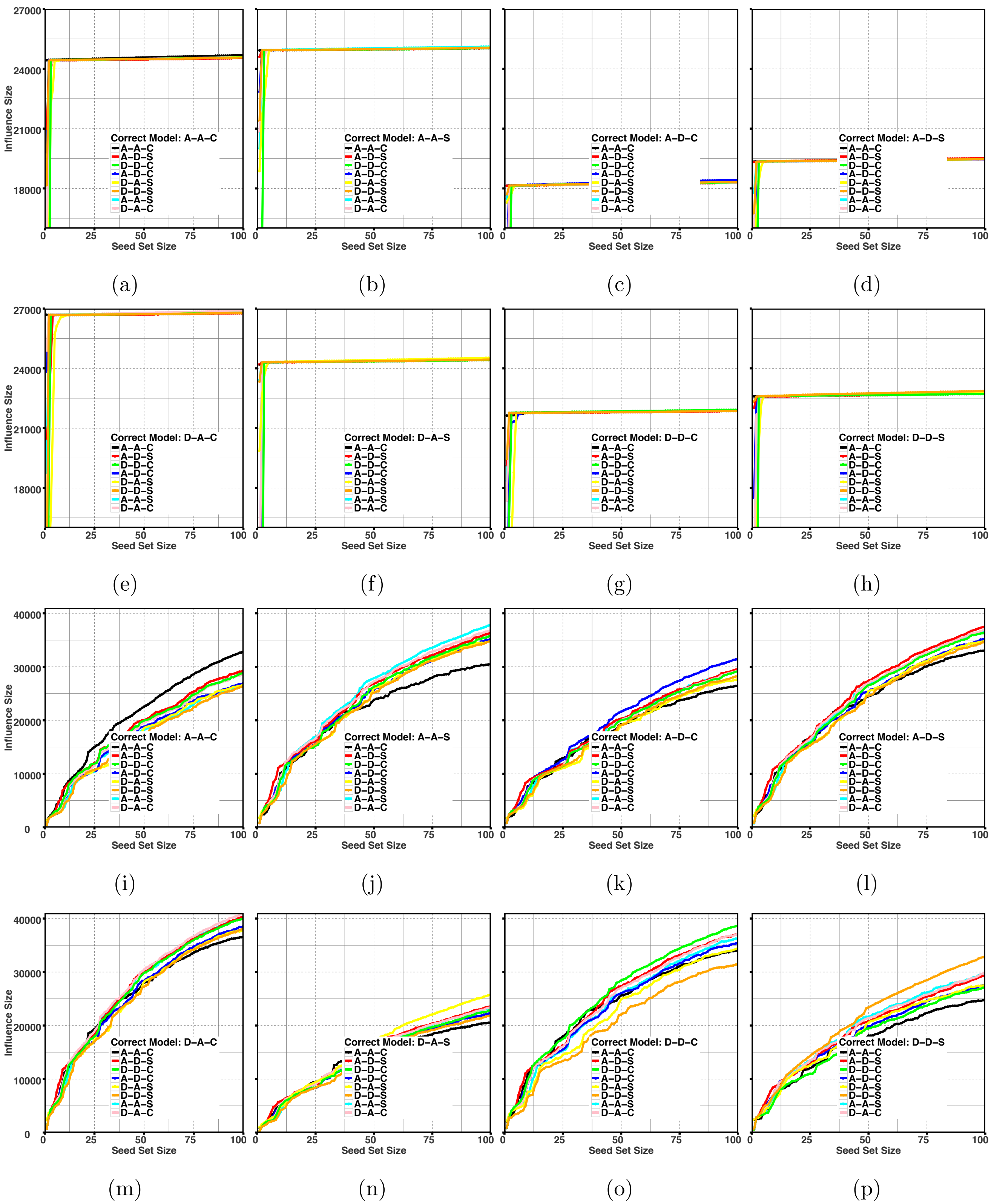
Supplementary Figure 51: Influence spread as a function of the seed set size. **Network: Slashdot Friend Network November 2008, Nodes=77360, Edges =469180.** Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are “Random” and “Constant” (all edges=0.1) and for the LT model are “Random” and “Inverse Degree”.



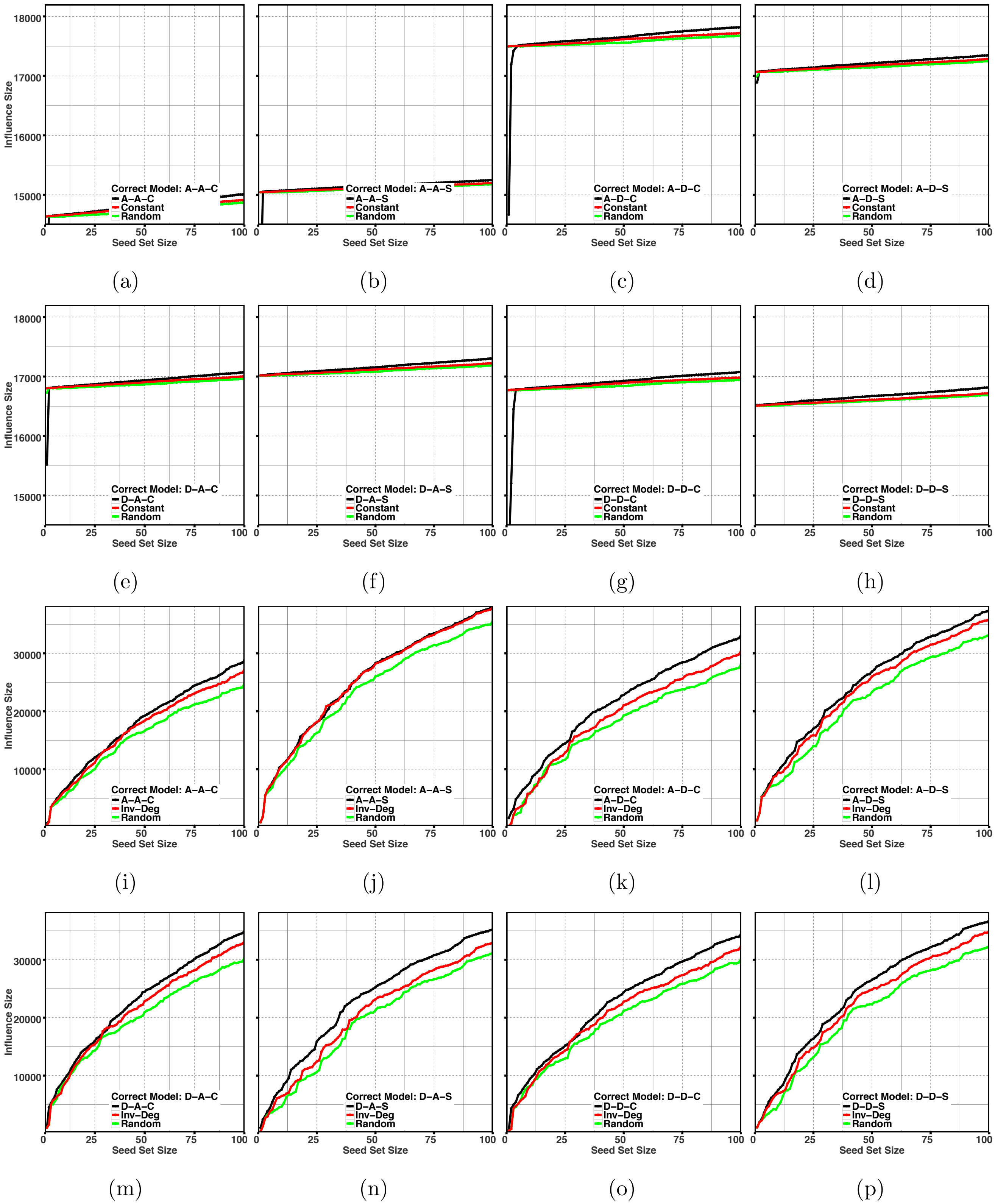
Supplementary Figure 52: Various seed characteristics for the IC model. **Network: Slashdot Friend Network November 2008, Nodes=77360, Edges =469180.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



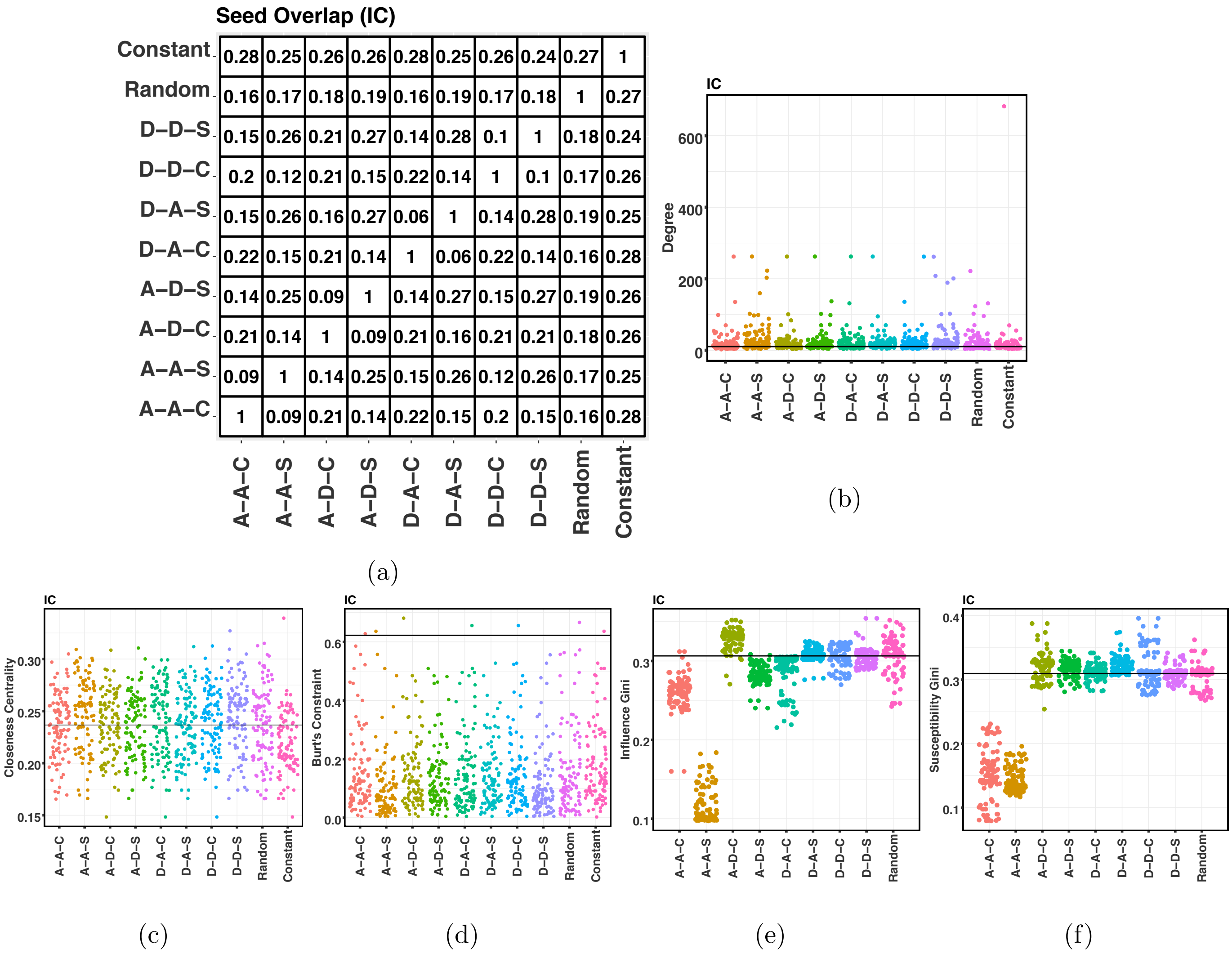
Supplementary Figure 53: Various seed characteristics for the LT model. **Network: Slashdot Friend Network November 2008, Nodes=77360, Edges =469180.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



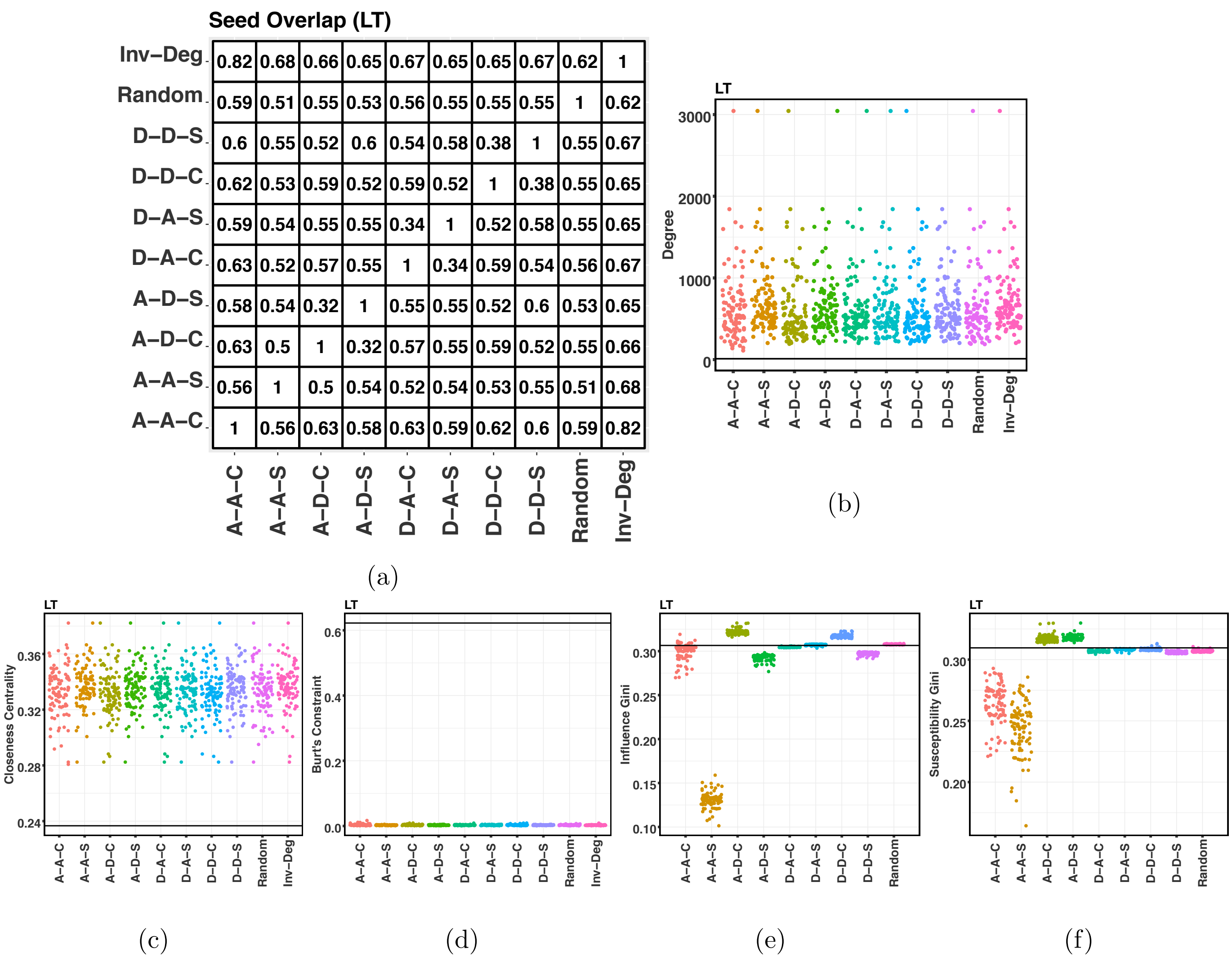
Supplementary Figure 54: Influence spread as a function of the seed set size. **Network: Slashdot Friend Network November 2008, Nodes=77360, Edges =469180.** Panels (a-h): IC Model, Panels (i-p): LT Model.



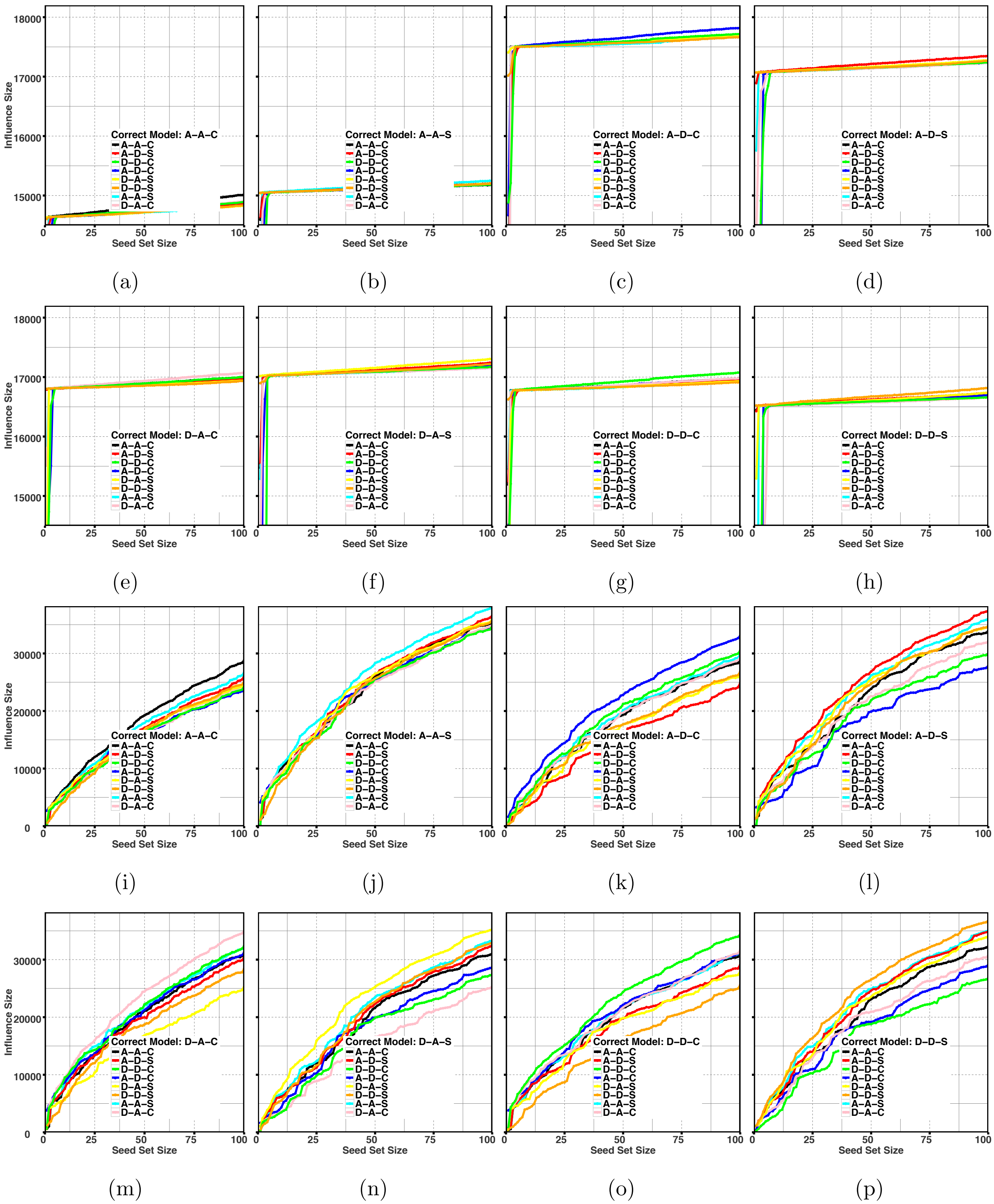
Supplementary Figure 55: Influence spread as a function of the seed set size. **Network: Epinions "Trust" network , Nodes=75877, Edges=405739.** Panels (a-h): IC Model, Panels (i-p): LT Model. The Baseline edge propagation probabilities for the IC model are "Random" and "Constant" (all edges=0.1) and for the LT model are "Random" and "Inverse Degree".



Supplementary Figure 56: Various seed characteristics for the IC model. **Network: Epinions "Trust" network , Nodes=75877, Edges=405739.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



Supplementary Figure 57: Various seed characteristics for the LT model. **Network: Epinions "Trust" network , Nodes=75877, Edges=405739.** (a) shows the mean fractional seed set overlap (averaged over 10 random draws) between the different models. (b, c, d) show Degree, Closeness Centrality and Burt's Constraint of the selected seeds. The horizontal black line in (b, c, d) show the mean Degree, Closeness Centrality and Burt's Constraint of the entire network respectively. (e, f) show the Gini coefficient of the Influence (e) (and Susceptibility (f)) parameters of a seed node and the sub-network induced by its friends-of-friends. The horizontal black line in (e, f) shows the mean Influence and Susceptibility Gini for the entire network for the "Random" model.



Supplementary Figure 58: Influence spread as a function of the seed set size. **Network: Epinions "Trust" network , Nodes=75877, Edges=405739.** Panels (a-h): IC Model, Panels (i-p): LT Model.