

# Modeling citations and cartels

**Minhyuk Park**

University of Illinois Urbana-Champaign

**Haotian Yi**

University of Illinois Urbana-Champaign

**Ian Chen**

University of Illinois Urbana-Champaign

**Tandy Warnow**

University of Illinois Urbana-Champaign

**George Chacko**

`chackoge@illinois.edu`

University of Illinois Urbana-Champaign

---

## Article

### Keywords:

**Posted Date:** April 2nd, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-9270520/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Modeling citations and cartels

Minhyuk Park<sup>1</sup>, Haotian Yi<sup>1</sup>, Ian Chen<sup>i</sup>, Tandy Warnow<sup>1</sup>, and George Chacko<sup>1,\*</sup>

<sup>1</sup>Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign, Urbana, IL, 61801

\*chackoge@illinois.edu

## ABSTRACT

Citation data are widely used to inform researchers and administrators. Interpreting the results of citation analysis is non-trivial on account of uncertainty around motivations for citation, varied disciplinary practices, data quality, standards for quality, and the lack of ground truths. *In-silico* approaches that harness autonomous behavior and randomness offer an option in this regard. We have developed an agent-based model of citation that allows us to parsimoniously model citation dynamics at article and author levels. We characterize the model through baseline simulations, sensitivity, and ablation analysis. We also model the behavior of cartels that seek to manipulate citations. We observe an early advantage to cartels that is lost over time implying that, on balance, cartel membership may offer less than its members hope for but may still be incentivizing for some who enroll in such membership.

## Introduction

Citation data enjoy widespread use by experts and non-experts, and are used to study productivity, impact, collaboration, community structure, and misconduct among other topics of inquiry<sup>1-5</sup>. Interpreting citation studies is challenging given the uncertainty around motivations for citation, subjective views on what ought to be cited, disciplinary norms, data quality and incommensurability between studies. While a unifying theory of citation<sup>6,7</sup> that explains citation behavior seems elusive, a patchwork of historical theories of citation is extant, which can be roughly grouped into normative and rhetorical categories<sup>8-12</sup>. Although individually plausible, none of these theories is well-supported by experimental results at scale; likely because of the challenge in quantifying and distinguishing between overlapping motivations imputed to researchers.

From a different perspective, mathematical models of citation<sup>13,14</sup> that incorporate randomness offer explanatory promise and simulations using agent-based models (ABMs)<sup>15-19</sup> that leverage autonomous behavior in addition to randomness represent a complementary avenue for model-based reasoning and discovery. Herein, we present the Scalable Agent-based Simulator for Citation Analysis with Sampling and Authors (SASCA-ReSA), an ABM that parsimoniously models the dynamics of referencing and accumulation of citations. SASCA-ReSA is less parsimonious than earlier models we have developed, PyABM, SASCA, and SASCA-ReS<sup>19-21</sup>, yet it is distant from being saturated.

In SASCA-ReSA, the input to a simulation is a citation network—represented as a directed graph with articles as nodes and citations as edges. The output of a simulation is a larger citation network created through the addition of new nodes in time-steps of a year. The number of nodes added in each time-step and the duration of the simulation is specified by a user. Notable in SASCA-ReSA is the addition of *author attributes* in the form of a corresponding author who may contribute multiple articles to the network and the count of authors for each article. These new features enable modeling author reputation and scale of collaboration as factors influencing citation counts.

Nodes in SASCA-ReSA networks are of two types, seeds and agents. All nodes have *attributes* that attract citations from other agents. Seed nodes are present at the beginning of a simulation and can receive but do not make citations. Each new node added during a simulation is an autonomous agent that makes citations to other nodes. Agents are assigned a quota of citations to make and a ‘*phenotype*’ that defines their choice of nodes to cite. Based on their phenotype, agents score eligible nodes and make weighted random selections to satisfy their citation quotas. Once an agent completes its quota, it remains a part of the network and can receive citations from other agents throughout the simulation.

The attributes of a node are its year of publication, its fitness (a measure of its quality), *in\_degree* (count of citations received), *out\_degree* (citation quota), the reputation of its corresponding author (*ar*), and the count of authors (*na*) contributing to it. Fitness is assigned to each node by sampling from a power-law distribution ranging from 1:1000, *na*, the count of authors contributing to each article is assigned through sampling from a real-world distribution and *ar* is defined as the H-index<sup>22</sup>, computed in-graph for an author. *In\_degree* and *ar* are updated at each time step.

An agent’s phenotype consists of weights that sum to 1 for *in\_degree* (*pa\_w*), fitness (*fit\_w*), author reputation (*ar\_w*), and author count (*na\_w*). Additionally, an agent’s affinity for target proximity to itself in the network is specified by a parameter  $\alpha$ , and recency, the propensity of authors to cite newer papers, is enforced through a separate protocol described in<sup>21</sup>.

At network level, the distribution of articles per author fits a Lotkas’s Law distribution<sup>23</sup> Further details of SASCA-ReSA are provided in Methods and a Github site<sup>24</sup> where the software implementation of the model is archived as open source.

SASCA-ReSA enables the study of strategies of individual researchers as well as the effects of these strategies on network structure. In the following sections, we largely focus on the former. We first examine the model using a one factor at a time (OFAT) approach<sup>25</sup>. Next, we examine the effect of planting a small number of agents with defined attributes and phenotype into a simulation followed by an ablation study. Finally, we use SASCA-ReSA in a simple model of the behavior of citation cartels<sup>4,26</sup>. We use agent-based modeling to stimulate thought and reasoning and do not claim identical citation behavior in the real-world. Conclusions drawn from ABMs are pertinent to the model environment and only *suggest* corresponding behavior. We do not explore the question of whether citation analysis to evaluate research is meritorious (our view is that it is not) nor do we support ‘gaming the metrics’<sup>27</sup> in any shape or form.

## Results

**Baseline Data** SASCA-ReSA was used to generate baseline data using a standard protocol<sup>19</sup> that we have used during the development of SASCA-ReSA and its precursor models. In this protocol, a seed network is allowed to grow by 3% each year

**Table 1.** Distributional statistics of agent in-degree across 10 baseline replicates (701,570 agents per replicate). The median is stable across replicates (SD = 0.0) while variability increases at the extreme tail, with the maximum in-degree varying by  $\pm 11,761$  across replicates.

| Statistic          | Mean     | SD       |
|--------------------|----------|----------|
| Median             | 12.0     | 0.0      |
| 75th quartile      | 27.6     | 0.5      |
| 90th percentile    | 59.0     | 0.4      |
| 99th percentile    | 301.4    | 3.4      |
| 99.9th percentile  | 1,706.5  | 44.4     |
| 99.99th percentile | 8,987.1  | 403.1    |
| Maximum            | 72,280.7 | 11,760.7 |

for 30 years through the iterative addition of agents. The seed network<sup>19</sup> is derived from a citation network centered around the cell biology literature concerning extracellular vesicles<sup>28</sup> but can be replaced by any other citation network or random graph. The final network consists of 1,193,102 nodes (491,532 seed nodes and 701,570 agents). During these baseline experiments, the agent background was randomized, that is, agent phenotypes were randomized with respect to  $pa\_w$ ,  $fit\_w$ ,  $ar\_w$ , and  $na\_w$  all four summing to 1 for each agent, while fitness,  $\alpha$ , and  $out\_degree$  were assigned to each agent through sampling from synthetic or real-world distributions (Methods).

The citation count accumulated by agents correlated most strongly with publication year (the earlier an agent is created the more chance it has to receive citations) and  $out\_degree$  (Fig. 1), which is consistent with results from our preceding models. Progressively weaker correlations were seen with respect to the number of authors, initial author reputation, fitness and  $\alpha$ . We also note consistency among replicates with a sharp increase in variability after the 99th percentile of  $in\_degree$  (Table 1).

We also examined the median values of the publication year, fitness, number of authors, initial  $ar\_w$ , and  $\alpha$  (Fig. 2). Agents were grouped into tiers consisting of all agents, the top 1%, 0.1%, and 0.01% by in-degree and pooled across 10 baseline replicates to consider 7,015,700 agent-observations. Publication year decreases monotonically toward the extreme tail, reflecting an advantage for agents that are created early. Fitness increases with in-degree tier, rising from a median of 1 for all agents to 27 at the top 0.01%, a trend that is consistent with observations from previous models<sup>19–21</sup>. Interestingly, initial  $ar$  shows a non-monotonic pattern, peaking at the top 1% (median 7) before declining at higher thresholds, which may suggest that, at the extreme tail, fitness and early publication are stronger determinants of  $in\_degree$  than author reputation.  $\alpha$  is more or less invariant across tiers, consistent with its weak Spearman correlation (Fig. 1).

**OFAT Analysis** To examine the effects of individual variables on citations received by agents, we designed conditions in which either a single attribute or phenotype weight of an agent was increased to near maximum possible value while the remaining attributes and weights were held relatively constant (Table 2).

We measured baseline data with either randomized or static agent backgrounds. For the static background, agent weights for preferential attachment ( $pa$ ), fitness,  $num\_authors$  ( $na$ ), and  $author\_reputation$  ( $ar$ ) were set to 0.25 each. Three replicates were conducted for each condition and each condition was applied to all agents in the simulation.

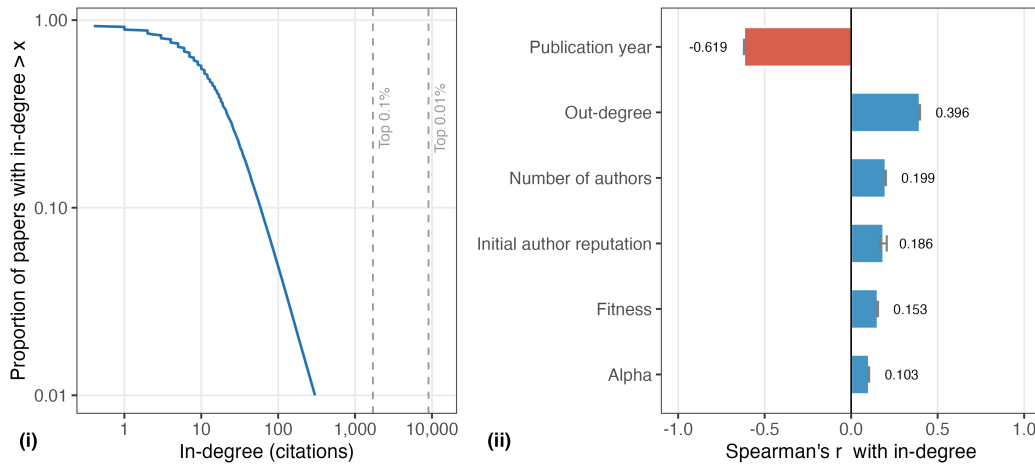


**Figure 1.** Baseline data from 10 simulations with randomized agents (i) Left panel: In\_degree distribution Roughly 7.03% (SD=0.037%) of agents received zero citations and are excluded from the plot (ii) Right panel: Spearman's Rank correlation of agent attributes with final in\_degree. Error bars shown are +/- SD.

|                           |            |        |          |           |
|---------------------------|------------|--------|----------|-----------|
| Publication year          | 2001       | 1994   | 1993     | 1990      |
| Fitness                   | 1          | 2      | 12       | 27        |
| Out-degree                | 35         | 61     | 66       | 74        |
| Number of authors         | 6          | 8      | 9        | 9         |
| Initial author reputation | 2          | 7      | 3        | 2         |
| Alpha                     | 0.500      | 0.547  | 0.531    | 0.493     |
|                           | All agents | Top 1% | Top 0.1% | Top 0.01% |

**Figure 2.** Agent attributes by tiered in-degree. Median values of six agent attributes — publication year, fitness, out\_degree, number of authors, initial author reputation (*iar*), and  $\alpha$  — computed for all agents and for the top 1%, 0.1%, and 0.01% of agents by in\_degree, pooled across 10 baseline replicates. Each row is independently normalised to its own range. Publication year decreases monotonically toward the extreme tail, reflecting the temporal accumulation of citations. Fitness increases with in-degree tier, rising from a median of 1 for all agents to 27 at the top 0.01%. Out\_degree increases monotonically from 35 to 74 across tiers, consistent with its role as the second strongest correlate of in-degree. Initial *ar* shows a non-monotonic pattern, peaking at the top 1% (median 7) before declining at higher thresholds, suggesting that at the extreme tail fitness and early publication are stronger determinants of citation accumulation than author reputation.  $\alpha$  varies very little across tiers.

For attributes, the effect of either high out\_degree (249) or high  $\alpha$  (0.95) was studied in randomized or static agent backgrounds. We did not perform an OFAT for fitness, since setting all agents to an equally high fitness is not interpretable. Instead fitness is addressed in the ablation experiment described in a following section.

When studying the impact of sharply increasing one element of agent phenotype, each weight was set to 0.91 while the remaining weights were assigned 0.03. For example W1, the preferential attachment (PA) dominant condition has the weight for preferential attachment set to 0.91 and the weights for fitness, author count, and initial author reputation set to 0.03 each.

The results show very little difference between random or static baseline configurations. While we report the maximum in\_degree achieved, we consider it less reliable than the 99.9th or 99.99th percentile since peak values can vary between replicates of the same experiment. For example, for the baseline data summarized in Table 1, the maximum ranges from 54,438

**Table 2.** OFAT experiment: in-degree distributional statistics across 11 conditions (701,570 agents per replicate, 3 replicates per condition, with statistics pooled across replicates). BSL\_rand and BSL\_stat are baseline conditions with phenotype weights either randomized or set to 0.25 respectively. W1–W4 are weight conditions in which one citation weight is set to 0.91 and the remaining three to 0.03. W5a and W5b fix out\_degree at 249 for all agents with equal or randomized weights respectively ; W6a and W6b fix  $\alpha$  at 0.95 with either equal or randomized weights. The top 0.1% and top 0.01% thresholds are defined relative to the baseline in-degree distribution ( $\geq 1,707$  and  $\geq 9,018$  citations respectively).

| ID                                                           | Condition            | Median | Mean  | p99.9 | p99.99 | Max     | % Agents $\geq$ p99.9 <sub>bsl</sub> | Agents $\geq$ p99.99 <sub>bsl</sub> |
|--------------------------------------------------------------|----------------------|--------|-------|-------|--------|---------|--------------------------------------|-------------------------------------|
| <i>Baseline conditions</i>                                   |                      |        |       |       |        |         |                                      |                                     |
|                                                              | BSL_rand             | 12     | 33.1  | 1,699 | 9,124  | 100,883 | 0.10                                 | 0.010                               |
|                                                              | BSL_stat             | 12     | 33.0  | 1,717 | 8,901  | 79,011  | 0.10                                 | 0.010                               |
| <i>Weight conditions</i>                                     |                      |        |       |       |        |         |                                      |                                     |
| W1                                                           | pa_wt dominant       | 10     | 32.9  | 2,469 | 18,064 | 146,764 | 0.14                                 | 0.026                               |
| W2                                                           | fit_w dominant       | 12     | 33.6  | 2,263 | 12,471 | 170,058 | 0.14                                 | 0.017                               |
| W3                                                           | na_w dominant        | 12     | 32.8  | 1,596 | 8,193  | 128,514 | 0.09                                 | 0.009                               |
| W4                                                           | ar_w dominant        | 9      | 32.6  | 1,624 | 4,548  | 24,660  | 0.09                                 | 0.002                               |
| <i>Attribute conditions: high out_degree (249)</i>           |                      |        |       |       |        |         |                                      |                                     |
| W5a                                                          | Equal weights (0.25) | 68     | 191.7 | 9,332 | 63,017 | 459,428 | 1.13                                 | 0.104                               |
| W5b                                                          | Randomized weights   | 68     | 192.4 | 9,367 | 62,479 | 438,259 | 1.12                                 | 0.106                               |
| <i>Attribute conditions: high <math>\alpha</math> (0.95)</i> |                      |        |       |       |        |         |                                      |                                     |
| W6a                                                          | Equal weights (0.25) | 14     | 32.5  | 1,333 | 6,329  | 41,392  | 0.07                                 | 0.006                               |
| W6b                                                          | Randomized weights   | 14     | 32.5  | 1,300 | 6,501  | 53,430  | 0.07                                 | 0.006                               |

to 94,147 across 10 replicates — a range of nearly 40,000 with a standard deviation (SD) of 11,761. In contrast, the SD at the 99.9th percentile is roughly 44.

Among the weight conditions, W1, the *pa\_w* dominant condition, has the heaviest tail at the 99.9th percentile (2,469), where a small number of papers accumulate extreme citation counts while the median citation count drops slightly. W2, the *fit\_w* dominant condition, is slightly less heavy-tailed than W1 with the top 0.1% also at 0.14 suggesting that in a *fit\_w* dominant environment, the best articles are well cited. W3, the author count (*na*) weight dominant condition, exhibits mild compression below baseline suggesting that the number of authors on a paper does not concentrate citations very effectively or frequently. W4, the author reputation (*ar\_w*) dominant condition, is the most compressed, with p99.9 of 1,624 and p99.99 of 4,548 which we interpret as an indication of limits to the effect of author reputation. The W5 experiment shows a dramatic increase in citations, which is to be expected since the total number of citations being made is much higher than the other conditions although it is also clear that these citations are not evenly distributed. The W6 experiment, high  $\alpha$  condition, shows citations slightly below baseline at p99.9 (1,300-1,333), indicating a mild compression effect of high  $\alpha$ . The W5 and W6 experiments are consistent with our earlier findings<sup>19</sup> and prior observations on ‘high-referencing’ and fractional counting<sup>29,30</sup>.

We infer from the OFAT experiment that out\_degree is the most influential factor in terms of driving citations in our model system. Among the weight conditions, *pa\_w* and *fit\_w* amplify the heavy tail seen in baseline runs while *na\_w* and *ar\_w* suppress it. However, these effect sizes are modest. High  $\alpha$  mildly compresses the distribution. Thus, we infer that prolific citing is a major driver of counts with citation preferences playing a secondary role in shaping the tail but not the bulk of the distribution. A logical follow-up question is whether a small number of agents with specific attributes can accumulate disproportionate citations in the background of a randomized environment. We attempt to explore this question below using planted nodes.

## Planted Nodes

We examined the effect of planting a small number of agents of fixed phenotype in the larger randomized background. This experiment was performed in two ways: either simultaneously planting 5 agents in year 1 of the simulation (P5) or a staggered planting of one agent in each of the first 5 years of the simulation (PS5). A reference phenotype, derived from data generated during development, was used for comparison (Table 3).

Two scenarios were explored: a fitness dominant scenario in which fitness was set high and the number of authors was low and a social dominant scenario in which fitness was low and the number of authors was high. In each case, out\_degree was either set to 249 (unrestricted) or 35 (restricted). Three replicates were conducted for each experiments for both the P5 and PS5 series of experiments. Averaged results are shown in Fig. 3 and Table 4, which suggest that PS5 substantially outperforms P5

**Table 3.** Fixed phenotype parameters for planted nodes in P5 and PS5 experiments. Fitness, num\_authors, and out\_degree are assigned fixed values at node creation rather than drawn from distributions. Weights sum to 1.0 in all conditions. The reference phenotype serves as a comparison for the fitness-dominant and social-dominant strategies. Fitness tier f1 = 1–10, f2 = 10–100, f3 = 100–1000. The social-dominant num\_authors value of 1,000 reflects large consortium authorship and is consistent with values observed in baseline simulations.

| Strategy         | Out_degree   | fitness | num_authors | out_degree | pa   | fit  | na   | ar   | $\alpha$ | Fitness tier |
|------------------|--------------|---------|-------------|------------|------|------|------|------|----------|--------------|
| Reference        | Unrestricted | 200     | 100         | 249        | 0.25 | 0.20 | 0.35 | 0.20 | 0.5      | f3           |
| Fitness-dominant | Restricted   | 200     | 6           | 35         | 0.20 | 0.65 | 0.10 | 0.05 | 0.5      | f3           |
| Fitness-dominant | Unrestricted | 200     | 6           | 249        | 0.25 | 0.55 | 0.15 | 0.05 | 0.5      | f3           |
| Social-dominant  | Restricted   | 5       | 1,000       | 35         | 0.20 | 0.10 | 0.60 | 0.10 | 0.5      | f1           |
| Social-dominant  | Unrestricted | 5       | 1,000       | 249        | 0.20 | 0.10 | 0.60 | 0.10 | 0.5      | f1           |

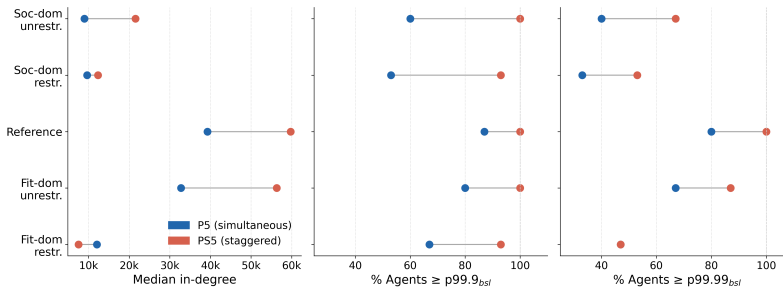


**Figure 3.** Comparison of simultaneous (P5) and staggered (PS5) planting strategies across five conditions. Each panel shows a metric of citation performance for 75 planted nodes (5 conditions  $\times$  5 nodes  $\times$  3 replicates), with values averaged across 3 replicates per condition, under simultaneous entry in year 1 (P5, blue) and staggered entry across years 1–5 (PS5, orange). From left to right: median in-degree, percentage of planted nodes exceeding the baseline top 0.1% threshold ( $\geq 1,707$  citations), and percentage exceeding the baseline top 0.01% threshold ( $\geq 9,018$  citations). PS5 outperforms P5 in the majority of conditions and metrics, with the largest gains in the fitness-dominant unrestricted and reference conditions. Restricted conditions (out\_degree=35) show more modest and variable gains from staggering. The reference condition achieves 100% of planted nodes above both thresholds under PS5.

across almost all conditions and metrics. A possible explanation is that staggered planting may reduce competition for citations in the early stages of a simulation.

Interestingly, the reference phenotype is the best performer in both series. In PS5 all 5 nodes are in the top 0.1% and 0.01%. Fitness-dominant unrestricted is a close second, which suggests that high fitness combined with high out\_degree favors success in this otherwise randomized background. Out\_degree is important: the contrast between the restricted (od=35) vs unrestricted (od=249) is large in both series and restricted conditions consistently perform worse than their unrestricted counterparts. Social-dominant conditions trail fitness-dominant conditions throughout, suggesting that a social-dominant strategy may be less reliable. The temporal effect of the PS5 experiments is non-monotonic and may reflect stochastic effects.

While the reference phenotype outcompetes the more formally designed ones, the findings in both cases are consistent with each other. In summary, fitness and out\_degree are the two most influential attributes for a planted node seeking high citation counts, and staggered entry (PS5) outperforms simultaneous planting (P5)s.

## Ablation Analysis

Since the reference phenotype performed the best in the planted node experiments, we used it in an ablation experiment to infer the relative contributions of attributes and weights to citation counts achieved under both P5 and PS5 designs.

The ablation analysis was designed as follows. Following the design of the planted node experiment five nodes are planted according to the P5 or PS5 protocols. Three replicates were performed for each condition (15 planted nodes per condition). Tier 1 ablations (A1–A4) target phenotype and consist of setting one weight to zero and redistributing its value proportionally across the weight settings of the reference phenotype. Tier 2 ablations (A5–A7) set one node attribute to its neutral value while holding weights fixed (Fig. 4).

Fitness and out\_degree are the two critical attributes. A5 (fitness ablated to 1) and A7 (out\_degree reduced to 35) produce the largest in\_degree drops across all metrics in both P5 and PS5. In PS5, median drops from 33,137 (reference) to 3,546 for

**Table 4.** Planted node experiment results for simultaneous (P5) and staggered (PS5) entry strategies across five conditions (5 planted nodes per condition per replicate, 3 replicates per condition, 15 planted nodes per condition total). Median in-degree, percentage of planted nodes exceeding the baseline top 0.1% threshold ( $\geq 1,707$  citations) and top 0.01% threshold ( $\geq 9,018$  citations), and number of rank-1 finishes (across 15 nodes  $\times$  3 replicates) are reported for each strategy. P5 nodes are planted simultaneously in year 1 and PS5 nodes are planted in staggered fashion, years 1-5. Values are pooled across 3 replicates.

| Condition        | P5 (simultaneous) |          |           |        | PS5 (staggered) |          |           |        |
|------------------|-------------------|----------|-----------|--------|-----------------|----------|-----------|--------|
|                  | Median            | Top 0.1% | Top 0.01% | Rank 1 | Median          | Top 0.1% | Top 0.01% | Rank 1 |
| Fit-dom restr.   | 12,044            | 67%      | 47%       | 0      | 7,517           | 93%      | 47%       | 0      |
| Fit-dom unrestr. | 32,774            | 80%      | 67%       | 1      | 56,297          | 100%     | 87%       | 3      |
| Reference        | 39,265            | 87%      | 80%       | 3      | 59,753          | 100%     | 100%      | 3      |
| Soc-dom restr.   | 9,661             | 53%      | 33%       | 0      | 12,321          | 93%      | 53%       | 0      |
| Soc-dom unrestr. | 8,968             | 60%      | 40%       | 1      | 21,557          | 100%     | 67%       | 1      |

A5 and 17,013 for A7. Top 0.01% falls to 33% for A5 in both strategies. These two attributes are necessary for extreme citation accumulation.

A4 (remove *ar* weight) is the best-performing condition in both P5 and PS5, with PS5 median reaching 65,672 and 100% of nodes in the top 0.01%. We propose that the *ar* weight in the reference phenotype diverts citation preference away from fitness and num\_authors targets, suppressing performance. This is consistent with the OFAT finding that AR-dominant conditions produce the most compressed distributions.

Removing other weights has little negative effect. A1 (remove *pa\_w*), A2 (remove *fit\_w*), and A3 (remove *na\_w*) all perform comparably to or better than A0 in both strategies, suggesting that attributes are more influential than individual weights and that reference weight configuration could be further optimised. Num\_authors as an attribute is non-critical when fitness is high. A6 (num\_authors ablated to 6) performs well — PS5 median 17,524, top 0.1% 100% — confirming that fitness and out\_degree together are sufficient for high citation accumulation regardless of team size. The PS5 series generally outperforms P5 across ablation conditions, consistent with the planting experiments. A5 and A6 are very similar identical between strategies, suggesting that staggering provides less benefit when a key attribute is ablated. In A7, out\_degree=35 is the empirical median rather than a true floor value, so A7 may understate the importance of out\_degree.

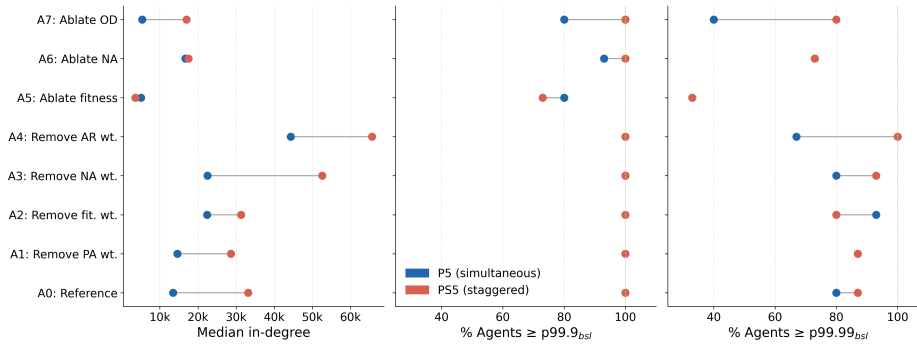


**Figure 4.** Ablation experiment results for simultaneous (P5) and staggered (PS5) planting strategies across eight conditions. Each panel shows a metric of citation performance for planted nodes under the reference phenotype with one attribute or weight ablated at a time (15 planted nodes per condition, averaged across 3 replicates). From left to right: median in-degree, percentage of planted nodes exceeding the baseline top 0.1% threshold ( $\geq 1,707$  citations), and percentage exceeding the baseline top 0.01% threshold ( $\geq 9,018$  citations). P5 nodes are planted simultaneously in year 1 (blue); PS5 nodes are planted in staggered years 1–5 (orange). A5 (ablate fitness, fitness=1) and A7 (ablate out\_degree, out\_degree=35) produce the largest performance drops across all metrics, identifying fitness and out\_degree as the two load-bearing attributes of the reference phenotype. PS5 generally outperforms P5, consistent with the planting experiment results.

## Modeling Citation Cartels

Except for extreme cases, studying citation cartels can be challenging on account of benign citation behavior that may generate cartel-like patterns of citation. In the artificial world of agent-based models, however, it is possible to explicitly design agents

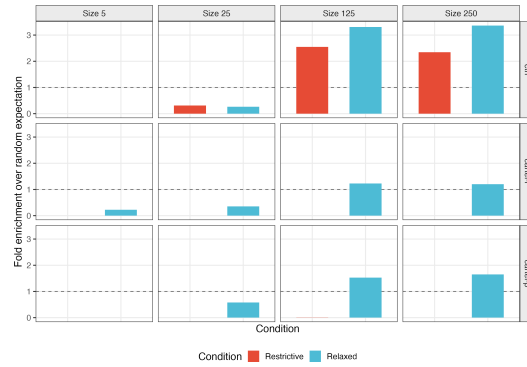


**Figure 5.** Cartel agents at elite levels of performance. Articles published by cartel authors (cartel-r and cartel-p) underperform compared to the control group. The figure shows the fraction of cartel articles in the top 1% relative to expected value for three counterfactual conditions across five cartel sizes. Fold enrichment is computed as the ratio of  $\text{fraction\_observed} / \text{fraction\_expected}$ .

with defined behaviors<sup>31</sup>. In using SASCA-ReSA to model cartel behavior, we assumed the existence of a group of researchers who collude to cite each other’s work in order to boost their citation counts for mutual benefit. Additional assumptions made were that members of this group did not enjoy strong author reputations and that they published frequently at a uniform rate—one article per year for each member of the cartel. Thus, each cartel member publishes multiple articles and those articles act as agents when making citations according to the rule set conferred upon them. This cartel model does not formally consider peer-review, journal reputation, researcher affiliations or prior collaborations, which can also be the targets of collusive activity<sup>32</sup>. Note that we use the term *cartel member* to refer to authors who are in a cartel, *cartel agent* to refer to articles published by these members, and *background agent* to identify any agent whose corresponding author is not in a cartel.

Our experimental design consisted of planting a single cartel within our standard simulation. The articles published by cartel members were assigned a fitness of 1 to reflect low quality and the number of authors for each cartel was uniformly set to 6. Finally each cartel agent directed 50% of its citation quota to articles by other cartel members.

Three counterfactual scenarios are represented in the design: *cartel-p* where cartel agents use 50% of their quota to cite each other’s papers based on attributes and phenotype, *cartel-r* where cartel agents spend 50% of their quota by randomly citing articles by other cartel members, and *ctrl* where cartel agents behave indistinguishably from background agents in the simulation. Cartels varied in size from 5, 25, 125, and 250 and experiments were run under either of two settings: *restrictive* where cartel agents had fitness=1, num\_authors=6, out\_degree=35, alpha=0.5, and phenotype weights of 0.25 each, and *relaxed* where cartel members had fitness=1 and num\_authors=6. but phenotype weights, out\_degree, and alpha were randomized as for all other agents.

When analyzing cartel simulations only agent nodes were considered. We first examined the representation of cartel agents in the top 1% of all agents by in\_degree (Fig. 5), which shows that cartel agents (either cartel-r or cartel-p) perform worse than the control group under conditions tested. Next, we examined the kinetics of citation accumulation for cartel members during the thirty year simulation period (Fig. 6). An early advantage under both restrictive and relaxed conditions is seen for cartel members, which is lost over time. As cartel size increases, the control group actually outperforms the cartels in terms of citations received per year around the mid-point of the simulation. We also computed the fraction of cartel agents that accumulated citations at elite levels in the three counterfactual scenarios (Fig. 6). The results suggest that while cartel-p and cartel-r agents underperform compared to the control group at the elite level (top 1% of in\_degree), they do enjoy an advantage in the early years following publication. In the zero-sum game of our simulation, this advantage is enjoyed at the expense of non-cartel members.

These data suggest that cartels do provide a modest early advantage to their members, but one that does not translate into elite performance at the end of these simulations; thus, in this sense, cartel members underperform compared to the control group. Moreover, the early advantage is at the expense of non-members in our simulations, and suggests a similar pattern in the real world. These data are consistent with Secchi’s preceding study using a simpler agent-based model at a smaller scale<sup>31</sup>. The presence of a cartel in our simulations did not alter the citation patterns of the background agents in any of the conditions we tested (Supplementary Fig. S1 and Supplementary Tables S1, S2).

Last, we performed community search experiments (Fig. 7), using our implementation<sup>33</sup> based on *k*-cores<sup>34</sup>, where a *k*-core is a maximal connected set of nodes where every node is adjacent to at least *k* other nodes in the set; hence the “coreness” of a set of nodes is the largest *k* so that the set is a *k*-core.

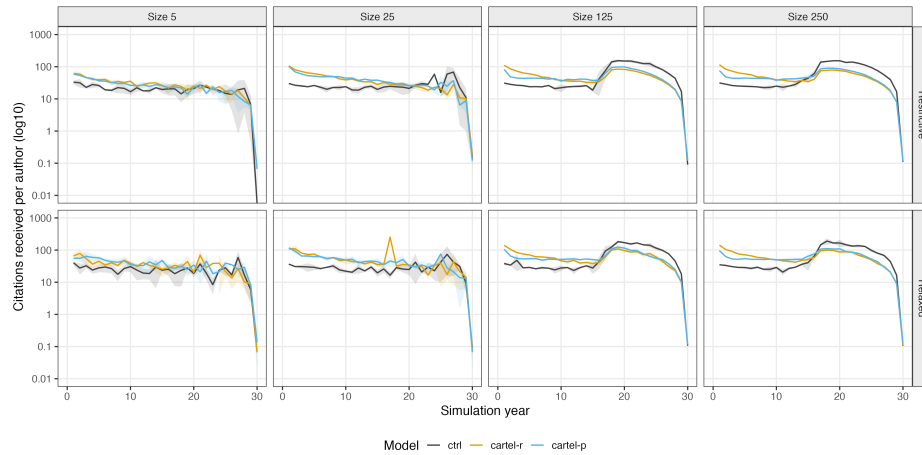


**Figure 6.** Cartel authors have an initial advantage that decreases over time. The figure shows average citations received per cartel author during a 30 year simulation in a randomized agent background. Authors designated as cartel members were selected from the same scientific age and low initial author\_reputation. Cartel members publish one article per year and cartel membership was identical across simulations for each cartel size. Counterfactual scenarios: (i) *ctrl*; agents (articles) of cartel members cite identical to background agents, (ii) *cartel-r*; cartel members cite each other randomly, and (iii) *cartel-p*; cartel members cite each other based on attributes and phenotype. For *cartel-r* and *cartel-p* scenarios, at most 50% of the citations made a cartel agent are made to cartel members and the remaining to nodes outside the cartel according to attribute and phenotype as in *ctrl*. *restrictive*: cartel members have fitness=1, num\_authors=6, out\_degree=35, alpha=0.5, and phenotype weights of 0.25 each. *relaxed*: cartel members have fitness=1 and num\_authors=6, but phenotype weights, out\_degree, and alpha are randomized as for all other agents. Each data point is the mean of citations per cartel author across 3 replicates; ribbons show +/- 1 SD. The figure is constructed from 72 independent simulations.

In this method, which is distinct from clustering the entire network, the input is a set  $A$  of nodes and the output is a set  $B$  of nodes such that  $A \subseteq B$  and  $B$  is a  $k$ -core for the largest possible  $k$ . Applying this technique to the control group, we obtained extremely large (all above 1 million nodes) and sparse communities ( $k \leq 6$ ). In contrast, the *cartel-p* or *cartel-r* group produced smaller and denser sets of nodes, with  $8 \leq k \leq 31$ . The communities around the *cartel-p* nodes grew in size with the cartel size, but this was not true for the *cartel-r* nodes. This striking result illustrates the possible effects of cartels on the larger citation network they are a part of.

## Discussion

ABMs are useful for exploring citation dynamics because they permit modeling under controlled conditions both at the individual and population levels, but are also limited by their artificial nature. SASCA-ReSA is the most complex of the ABMs we have developed for citation analysis, and adds authors as a model feature. This added feature comes with a computational cost: SASCA-ReSA, at present, imposes a performance penalty of roughly five times compared to SASCA-ReS. The penalty restricts the scale of use both in terms parallel simulations and the size of the resultant network. At present, this impacts our ability to explore SASCA-ReS and SASCA-ReSA together across a broad range of network sizes to examine converging and diverging trends.

In today's environment where the demand for resources exceeds their supply, where peer recognition is an ingredient of success, and where scientists are judged by quantitative metrics, citation counts can be viewed as all-important. An opportunity for questionable success is the manipulation of citations by cartels. Our simple representations of cartel behavior suggest limited advantages to cartel membership. Insular research communities may also display cartel-like behavior although we expect that this would be to a lesser degree. Nevertheless, distinguishing between cartels and innocent insularity remains an open question. Furthermore, we note that beyond increasing in number and size, it is very likely that cartels have already evolved towards greater sophistication in their activities and more difficult to detect. Thus, a more intensive modeling and examination of cartel behavior is needed, varying cartel sizes further and the proportion of citations made intra-cartel is a logical next step.

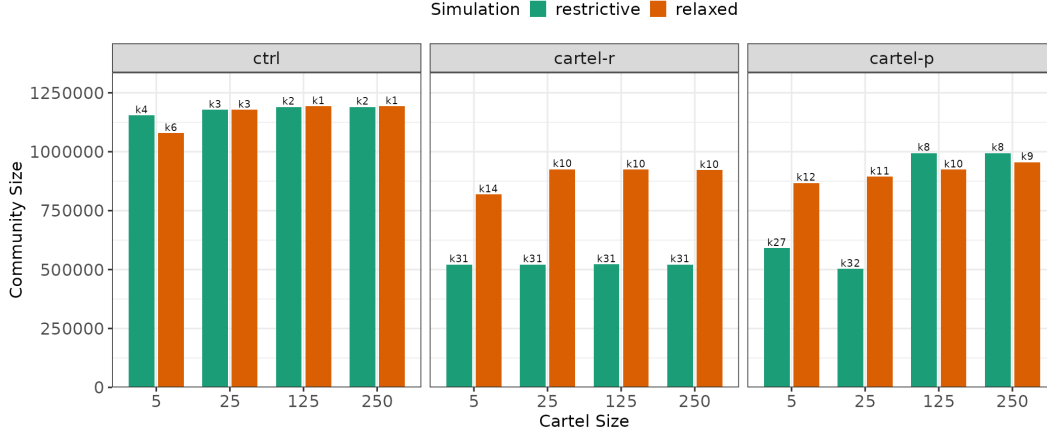


**Figure 7.** Articles by cartel members form dense community structures relative to the control group. A k-core based community search was performed using the set of cartel agents as query nodes.

## Methods

### Data

Input data used for the experiments such as the seed set, recency distribution, out-degree distribution, and number of authors distribution are available at <sup>35</sup>.

### Model

We first provide a brief overview of SASCA-ReS and then present SASCA-ReSA, which builds on SASCA-ReS.

**SASCA-ReS** is an agent-based simulator that generates a citation network. In each time-step (year), new agents are created that cite other nodes in the network. Each node  $v$  in the network has attributes, which are its fitness, publication year, and in-degree; of these, only in-degree can change over time. We denote these by  $\text{fitness}(v)$ ,  $\text{year}(v)$ , and  $\text{in\_degree}(v)$ . An agent in SASCA-ReS is generated by random selection of another node in the network, which is termed the generator for the agent. When an agent  $A$  is created, it is assigned a quota of references to make, a value for  $\alpha$ , and a phenotype, which provides values in  $[0, 1]$  for  $pa\_w$ , and  $fit\_w$ . The value for  $pa\_w$  and  $fit\_w$  are the relative weights given to preferential attachment and fitness, which are attributes of a given node (publication) in the network; hence  $pa\_w + fit\_w = 1$  is enforced. The value for  $\alpha$  determines the target fraction of the agent's quota of citations that go to the 1-hop neighborhood of its generator. Therefore,  $\alpha$  is independent of  $pa\_w$  and  $fit\_w$ .

Every agent cites its generator. With a small probability defined by the model, an agent may also cite a paper from the same year and a small percentage of its citation quota is reserved for fully random citations to nodes outside of the 2-hop neighborhood. After these, the agent makes citations to fulfill the remainder of its quota. The value for  $\alpha$  specifies the target fraction of the remaining quota that is made to the 1-hop neighborhood of the agent's generator, hence  $1 - \alpha$  specifies the target fraction of the remaining quota that is made to the nodes that are distance 2 away from the generator.

Once the number of citations to be made to both distance-1 and distance-2 neighborhoods of the generator is determined, the agents in each neighborhood are first sampled to get a set of at most 10,000 nodes. These nodes are then binned by their publication year, and the agent allocates its citations to different publication year bins in order to fit a user-provided recency distribution<sup>21</sup>. Once the set of nodes to cite, defined by the publication year bins and the number of citations within each bin, has been determined, each node is then scored based on the following process. When an agent scores a given node  $v$ , a node attribute score is computed first for each node attribute (fitness and in-degree).

We provide the preferential attachment score of node  $v$  (denoted by  $pa\_s(v)$ ) and the fitness score of node  $v$  (denoted by  $fit\_s(v)$ ) in Equation 1:

$$pa\_s(v) = \frac{\text{in\_degree}(v)^{\gamma+1}}{\sum_{x \in V} (\text{in\_degree}(x)^{\gamma+1})} \quad \text{and} \quad fit\_s(v) = \frac{\text{fitness}(v)^{\gamma+1}}{\sum_{x \in V} (\text{fitness}(x)^{\gamma+1})} \quad (1)$$

The score assigned to node  $v$  by the agent can be written as the weighted sum of these two scores, and is given in Equation 2.

$$\text{score}(v) = pa\_w \cdot pa\_s(v) + fit\_w \cdot fit\_s(v) \quad (2)$$

Note that  $pa_w + fit_w = 1$  and both  $pa_s(v)$  and  $fit_s(v)$  are normalized to sum up to 1 over all nodes in the candidate set (which is a subset of the nodes in the network). Hence the scores over all nodes in the candidate set naturally add up to 1, which means they can be interpreted as probabilities of being selected.

**SASCA-ReSA** extends SASCA-ReS through the inclusion of author features. It includes as node attributes: number of authors ( $na$ ) of each article and the reputation ( $ar$ ) of the corresponding author. The reputation of an author is computed using the in-graph H-index. The number of authors per paper is based on a randomly assigned number, taken from a user-specified distribution. Correspondingly, the agents also have two additional elements in their phenotype:  $ar_w$  and  $na_w$ . We require  $pa_w + fit_w + ar_w + na_w = 1$ . Hence, the scoring formula becomes

$$score(v) = pa_w \cdot pa_s(v) + fit_w \cdot fit_s(v) + ar_w \cdot ar_s(v) + na_w \cdot na_s(v) \quad (3)$$

Each publication at the time of creation is assigned a single author ID. The decision to either assign an existing author ID to this new publication or create a new author ID is made in order to come as close as we can to the distribution given by Lotka's law<sup>23</sup>, which states that the number of authors with  $k$  articles ( $n_k$ ) is roughly equal to  $\frac{n_1}{k^2}$ . However, only those authors whose first publication within the last 30 years are eligible to publish another publication.

## Software

Code for SASCA-ReSA is available at<sup>24</sup>.

## Funding

Work on this manuscript was partially supported by US NSF Grant No: OAC-2402559, the Illinois Computes project, and the Illinois-Inspire Partnership.

## Contributions

All authors reviewed the manuscript before submission. MP developed the SASCA-ReSA implementation. TW, and GC planned the experiment(s). MP and GC analysed results. HY and IC developed code and generated data for the community search experiments. MP, TW, and GC interpreted the results of experiments, wrote and edited the manuscript.

## Additional information

### Use of Artificial Intelligence

The authors declare the interactive use of assistive AI<sup>36</sup> for analysis of results and preparation of figures and tables for the manuscript. The text of the manuscript was written, and approved by the authors who acknowledge full responsibility for the content and conclusions presented.

### Reproducibility

The data presented in this manuscript and the scripts used to generate data and results are freely available.

### Competing Interests

The authors declare no competing interests.

## References

1. Ioannidis, J. P. A., Baas, J., Klavans, R. & Boyack, K. W. A standardized citation metrics author database annotated for scientific field. *PLOS Biol.* **17**, e3000384, DOI: [10.1371/journal.pbio.3000384](https://doi.org/10.1371/journal.pbio.3000384) (2019).
2. Boyack, K. W. *et al.* Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS ONE* **6**, e18029, DOI: [10.1371/journal.pone.0018029](https://doi.org/10.1371/journal.pone.0018029) (2011).
3. Šubelj, L., van Eck, N. J. & Waltman, L. Clustering scientific publications based on citation relations: A systematic comparison of different methods. *PLOS ONE* **11**, e0154404, DOI: [10.1371/journal.pone.0154404](https://doi.org/10.1371/journal.pone.0154404) (2016).
4. Fister, I., Fister, I. & Perc, M. Toward the discovery of citation cartels in citation networks. *Front. Phys.* **4**, DOI: [10.3389/fphy.2016.00049](https://doi.org/10.3389/fphy.2016.00049) (2016).
5. Safón, V., Docampo, D. & Cram, L. Screening articles by citation reputation. *Quant. Sci. Stud.* **6**, 405–420, DOI: [10.1162/qss\\_a\\_00355](https://doi.org/10.1162/qss_a_00355) (2025).

6. Zuckerman, H. Citation analysis and the complex problem of intellectual influence. *Scientometrics* **12**, 329–338, DOI: [10.1007/bf02016675](https://doi.org/10.1007/bf02016675) (1987).
7. Leydesdorff, L. Theories of citation? *Scientometrics* **43**, 5–25, DOI: [10.1007/bf02458391](https://doi.org/10.1007/bf02458391) (1998).
8. Merton, R. K. The Matthew effect in science: The reward and communication systems of science are considered. *Science* **159**, 56–63, DOI: [10.1126/science.159.3810.56](https://doi.org/10.1126/science.159.3810.56) (1968).
9. Cronin, B. The need for a theory of citing. *J. Documentation* **37**, 16–24 (1981).
10. Kaplan, N. The norms of citation behavior: Prolegomena to the footnote. *Am. documentation* **16**, 179–184 (1965).
11. Cozzens, S. E. What do citations count? The rhetoric-first model. *Scientometrics* **15**, 437–447, DOI: [10.1007/bf02017064](https://doi.org/10.1007/bf02017064) (1989).
12. Small, H. G. Cited documents as concept symbols. *Soc. studies science* **8**, 327–340 (1978).
13. Simkin, M. V. & Roychowdhury, V. P. A mathematical theory of citing. *J. Am. Soc. for Inf. Sci. Technol.* **58**, 1661–1673, DOI: [10.1002/asi.20653](https://doi.org/10.1002/asi.20653) (2007).
14. Golosovsky, M. & Solomon, S. Growing complex network of citations of scientific papers: Modeling and measurements. *Phys. Rev. E* **95**, DOI: [10.1103/physreve.95.012324](https://doi.org/10.1103/physreve.95.012324) (2017).
15. Gilbert, N. A simulation of the structure of academic science. *Sociol. Res. Online* **2**, 91–105, DOI: [10.5153/sro.85](https://doi.org/10.5153/sro.85) (1997).
16. Weisberg, M. & Muldoon, R. Epistemic landscapes and the division of cognitive labor. *Philos. Sci.* **76**, 225–252, DOI: [10.1086/644786](https://doi.org/10.1086/644786) (2009).
17. Harnagel, A. A mid-level approach to modeling scientific communities. *Stud. Hist. Philos. Sci. Part A* **76**, 49–59, DOI: [10.1016/j.shpsa.2018.12.010](https://doi.org/10.1016/j.shpsa.2018.12.010) (2019).
18. Bao, H. & Teplitskiy, M. A simulation-based analysis of the impact of rhetorical citations in science. *Nat. Commun.* **15**, DOI: [10.1038/s41467-023-44249-0](https://doi.org/10.1038/s41467-023-44249-0) (2024).
19. Chacko, G. *et al.* An agent-based model of citation behavior. *Appl. Netw. Sci.* **11**, DOI: [10.1007/s41109-025-00766-z](https://doi.org/10.1007/s41109-025-00766-z) (2026).
20. Park, M. *et al.* Very Large Scale Simulations of Network Growth with the Scalable Agent-based Simulator for Citation Analysis with sampling (SASCA-s) (2025). In press, 14th International Conference on Complex Networks and their Applications (2025).
21. Park, M., Yi, H., Warnow, T. & Chacko, G. Modeling the global citation network using the scalable agent-based simulator for citation analysis with recency-emphasized sampling (SASCA- ReS), DOI: [10.5281/zenodo.17789558](https://doi.org/10.5281/zenodo.17789558) (2025). Evaluated manuscript published on MetaRoR (2026), DOI: [10.70744/MetaROR.296.1.ea](https://doi.org/10.70744/MetaROR.296.1.ea).
22. Hirsch, J. E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci.* **102**, 16569–16572, DOI: [10.1073/pnas.0507655102](https://doi.org/10.1073/pnas.0507655102) (2005).
23. Lotka, A. The frequency distribution of scientific productivity. *J. Wash. Acad. Sci.* **16**, 317–323 (1926).
24. Park, M., Warnow, T. & Chacko, G. SASCA-ReSA. <https://github.com/illinois-or-research-analytics/SASCA-ReSA> (2026). Github repository.
25. ten Broeke, G., van Voorn, G. & Ligtenberg, A. Which sensitivity analysis method should i use for my agent-based model? *J. Artif. Soc. Soc. Simul.* **19**, DOI: [10.18564/jasss.2857](https://doi.org/10.18564/jasss.2857) (2016).
26. Kojaku, S., Livan, G. & Masuda, N. Detecting anomalous citation groups in journal networks. *Sci. Reports* **11**, DOI: [10.1038/s41598-021-93572-3](https://doi.org/10.1038/s41598-021-93572-3) (2021).
27. Biagioli, M. & Lippman, A. (eds.) *Gaming the Metrics: Misconduct and Manipulation in Academic Research* (The MIT Press, 2020).
28. Wedell, E., Park, M., Korobskiy, D., Warnow, T. & Chacko, G. Center–periphery structure in research communities. *Quant. Sci. Stud.* **3**, 289–314, DOI: [10.1162/qss\\_a\\_00184](https://doi.org/10.1162/qss_a_00184) (2022).
29. Small, H. & Sweeney, E. Clustering the science citation index ® using co-citations: I. a comparison of methods. *Scientometrics* **7**, 391–409, DOI: [10.1007/bf02017157](https://doi.org/10.1007/bf02017157) (1985).
30. Leydesdorff, L. & Opthof, T. Scopus's source normalized impact per paper (snip) versus a journal impact factor based on fractional counting of citations. *J. Am. Soc. for Inf. Sci. Technol.* **61**, 2365–2369, DOI: [10.1002/asi.21371](https://doi.org/10.1002/asi.21371) (2010).
31. Secchi, D. *A Simple Model of Citation Cartels: When Self-interest Strikes Science*, 23–32 (Springer Nature Switzerland, 2023). Advances in Social Simulation, DOI: [10.1007/978-3-031-34920-1\\_3](https://doi.org/10.1007/978-3-031-34920-1_3).

32. Richardson, R. A. K., Hong, S. S., Byrne, J. A., Stoeger, T. & Amaral, L. A. N. The entities enabling scientific fraud at scale are large, resilient, and growing rapidly. *Proc. Natl. Acad. Sci.* **122**, DOI: [10.1073/pnas.2420092122](https://doi.org/10.1073/pnas.2420092122) (2025).
33. Yi, H., Wei, I. C., Warnow, T. & Chacko, G. K-Core Based Community-Search. [https://github.com/AnoldGH/community\\_search](https://github.com/AnoldGH/community_search) (2026). Github repository.
34. Fang, Y., Cheng, R., Chen, Y., Luo, S. & Hu, J. Effective and efficient attributed community search. *The VLDB J.* **26**, 803–828, DOI: [10.1007/s00778-017-0482-5](https://doi.org/10.1007/s00778-017-0482-5) (2017).
35. Park, M., Warnow, T. & Chacko, G. SASCA-ReSA Data Docs. <https://github.com/illinois-or-research-analytics/SASCA-ReSA/tree/main/docs> (2026). Data documentation.
36. Anthropic. Claude (sonnet 4.5/4.6). <https://www.anthropic.com> (2025). Large language model, accessed 2025–2026.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SASCAReSAnsrsuppl.pdf](#)