An agent-based model of interdisciplinary interactions in science

Juste Raimbault [1,2,3,*]

[1] **CASA, University College London**

[2] **UPS CNRS 3611 ISC-PIF**

[3] **UMR CNRS 8504 Géographie-cités**

* **juste.raimbault@polytechnique.edu**

[1] *CASA, University College London*
[2] *UPS CNRS 3611 ISC-PIF*
[3] *UMR CNRS 8504 Géographie-cités*

Correspondence*:
CASA, UCL, 90 Tottenham Court Road, London W1T 4TJ, United Kingdom
*juste.raimbault@polytechnique.edu

ABSTRACT

An increased interdisciplinarity in science projects has been highlighted as crucial to tackle complex real-world challenges, but also as beneficial for the development of disciplines themselves. This paper introduces a parcimonious agent-based model of interdisciplinary relationships in collective entreprises of knowledge discovery, to investigate the impact of scientist-level decisions and preferences on global interdisciplinarity patterns. Under the assumption of simple rules for individual researcher project management, such as trade-offs between invested time overhead and knowledge benefit, model simulations show that individual choices influence the distribution of compromise points between emergent level of disciplinary depth and interdisciplinarity in a non-linear way. Different structures for collaboration networks may also yield various outcomes in terms of global interdisciplinarity. We conclude that independently of the research field, the organization of research, and more particularly the local balancing between vertical and horizontal research, already influences the final positioning of research results and the extent of the knowledge front. This suggests direct applications to research policies with a bottom-up leverage on the interactions between disciplines.

Keywords: Interdisciplinarity; Agent-based modeling; Scientific collaborations; Model exploration; Model calibration

1 INTRODUCTION

The role of interdisciplinary projects in science has been highlighted as crucial for the development of complexity approaches and an effective tackling of real-world issues. Many aspects of knowledge production have a role in enhancing interdisciplinary collaborations. [1] study the circular relationship between diversity and innovation, and show that underrepresented groups have a higher likelihood of successfully innovate in science. [2] use an agent-based model to study the co-evolution between knowledge diffusion and the structure of knowledge. Each discipline has its own view on interdisciplinarity, as for example [3] unveil an asymmetry between social and hard sciences in the credit given to other disciplines within interdisciplinary projects. Other social or political factor are to be taken into account when investigating the disciplinary structure of science: access to funding has for example a strong impact on the efficiency of knowledge production [4]. [5] show that the discrepancy between disciplines is intrinsic to the type of knowledge produced, as they suggest that paradigms are more likely to persist in “low-power” sciences. The organisation of research is also an important factor, and teams and single authors produce different aspects of the common knowledge [6]. [7] model probables trajectories according to the type of research environment. The link between open access, which is a driver of increased collaborations and potentially increased interdisciplinarity, and the quality of research, is investigated by [8].

Interdisciplinarity in itself has extensively been studied by quantitative studies of science. [9] show that interdisciplinary papers perform better in terms of citation on the long run than mainstream papers. [10] investigate the interdisciplinarity of scientists themselves and how it evolved in time, and show that more scientists have switched between topics recently. [11] provide empirical evidence for an optimal intermediate level of interdisciplinarity in terms of research impact.[12] study within the particular context of an interdisciplinary summer school the propensity of mixing within interdisciplinary projects, and find evidence consistent with random mixing. [13] show that randomness has an important role in determining individual trajectories success in physics.

Following [14], agent-based modeling is a privileged approach to simulate the behavior of scientists. [15] use an agent-based model to simulate the impact of a workflow to process data under different collaboration scenarios. [16] simulate citation dynamics, and more particularly the consequence of introducing a performance index on citation patterns. Agent-based modeling has extensively been used for the evaluation of peer review practices. [17] surveys 46 simulation studies of peer review with numerous applications. [18] empirically calibrates an agent-based model of peer review for more than 100 journals, and provides a tool to evaluate systems of peer reviews. [19] describes a theoretical model involving various actors of science. Agent-based models are more broadly used to study social dynamics such as group organisation in [20].

Various works have dealt with microscopic modeling of knowledge production, among which for example the Nobel game introduced by [21] which investigates the balance between falsification of previous theories and the elaboration of new theories. [14] also proposed an agent-based model of science, consistently with the perspectivist approach developed in [22]. We develop here a simple agent-based model of scientific research focusing on the interplay between disciplinary and interdisciplinary research. The rationale relies on the basic assumption that scientists can choose when starting a new project between interdisciplinary collaboration and a work within their discipline. How can the choice patterns at the micro-level influence the overall interdisciplinarity level ? The model is voluntary parcimonious to test if even many simplification some structural effects still hold.

2 AN EMPIRICAL ANALYSIS OF INTERDISCIPLINARITY PATTERNS

In order to give empirical support to the modeling choices for the ABM, we first study the properties of a large scientific corpus. We propose to use the Arxiv citation network, which represents a significant proportion of physics and computer science. An open dataset providing parsed authors and citations is made available by [23]. This allows constructing a citation network with $|V| = 1,396,261$ nodes (papers) and $|E| = 6,849,633$ citation links. This corresponds to $1,506,500$ unique authors which we disambiguated by concatenating first name and last name.

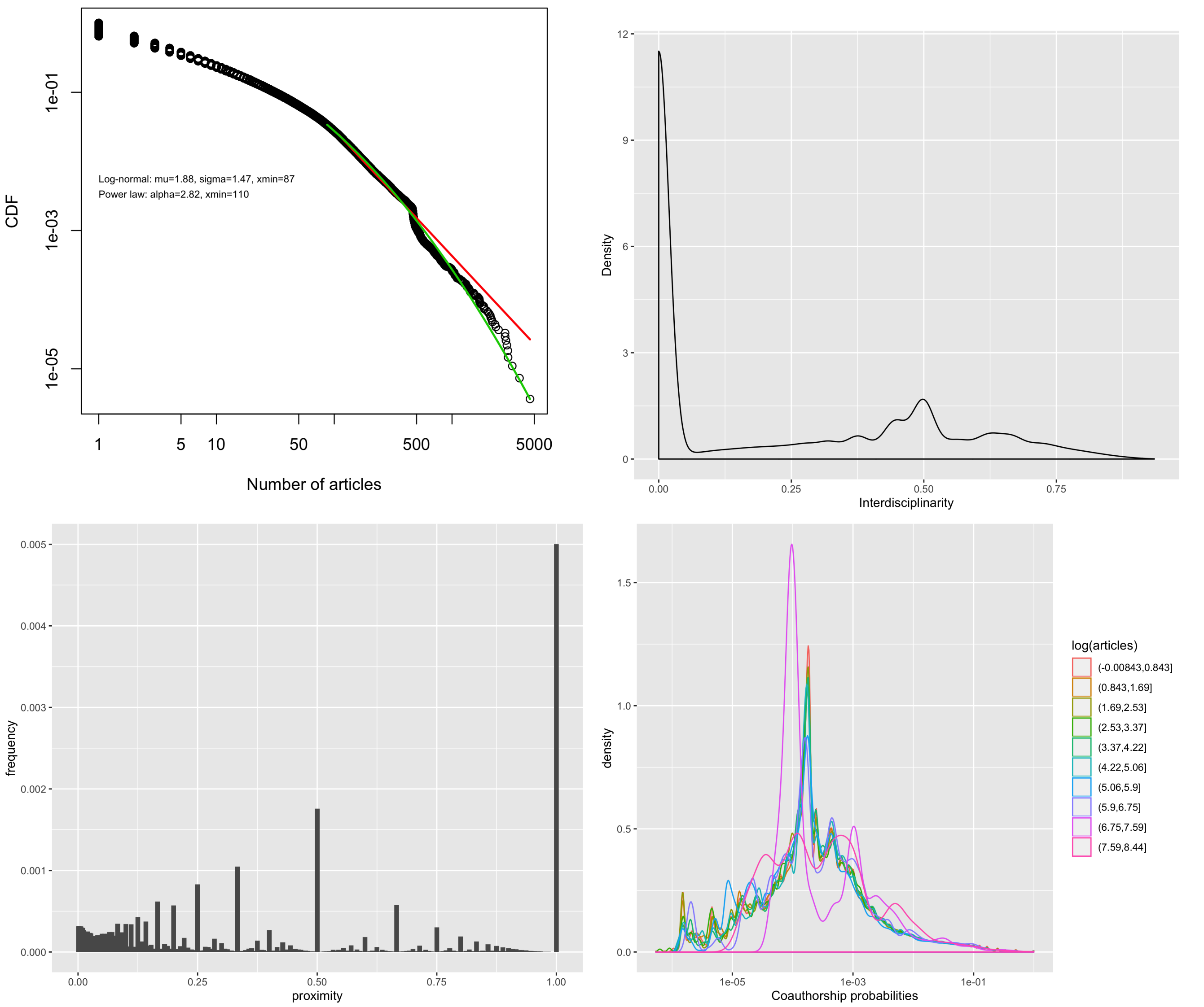

Figure 1. Collaborations and interdisciplinarity within the Arxiv dataset. *(Top left)* Cumulative distribution function of the number of articles per author (these were disambiguated using first and last name only, statistics may not be accurate). We compare a log-normal and a power-law fit. *(Top right)* Distribution of interdisciplinarity per author, computed as an Herfindhal index of probabilities within endogenous citation communities. *(Bottom left)* Distribution of positive author proximities, defined as cosine similarity between authors probability distribution within citation communities. *(Bottom right)* Distribution of co-authorship probabilities, conditioned by the number of articles.

We measure interdisciplinarity following the approach of [24]. Therefore, we construct endogenous disciplines within the citation network, and thus proceed to a community detection in the citation network, using a Louvain community detection algorithm. We obtain therein a modularity of 0.78 and 38 communities with a size larger than 1000. Working with these main endogenous citation communities (which can be interpreted as scientific fields of citation practice), we construct probabilities for authors to belong to each community. These are computed as $p_{ik} = N_{ik}/N_i$ for author i and community k, were N_{ik} is the number of articles authored within this community and N_i the total number of articles authored. This allows computing a cosine proximity between authors defined as $s_{ij} = \vec{p}_i \cdot \vec{p}_j$, and also an interdisciplinarity measure as an Herfindhal diversity index given by $h_i = 1 = \sum_k p_{ik}^2$. Finally, we also study co-authorship probabilities $c_{i \to j}$ defined as the probability for author i to co-author with author j knowing that the author has written a paper (the matrix is thus non symmetric).

We show in Figure 1 the empirical results obtained. The number of papers by author is close to a power-law with an exponent of 2.82, although a log-normal law seems to better fit the data. Regarding interdisciplinarity of authors, although a large majority of authors are mono-disciplinary, we find a secondary peak at 0.5 and a non negligible proportion of authors spanning the indicator range up to very high values of 0.8. This confirms the relevance of our model with an active interdisciplinarity. When studying cosine similarity between authors using their probabilistic description within communities, we find a broad range of values, also witnessing a high diversity (knowing that most authors are at a 0 proximity, since the plot is conditional for readability). Co-authorship probabilities follow rather symmetrical distributions with fat tails on a log-scale, consistently when conditioning on the number of papers authored. This is consistent with the power-law assumed for the propensity for interdisciplinarity for authors.

3 AN AGENT-BASED MODEL OF INTERDISCIPLINARITY

3.1 Rationale

Many dimensions and processes are at play to shape collaborations between scientists and more broadly between scientific disciplines. These include for example social networks, governance and funding issues, or knowledge proximity (which can occur on various knowledge domains, from methodological to empirical or theoretical). Our rationale is to propose an agent-based model grasping some of this complexity from the bottom-up focusing on scientist behavior, but simple enough so that it can be systematically explored. We include thus in the model two basic antagonist processes, namely a propensity to collaborate mostly determined by knowledge proximity, and some resources constraints (time, funding) which affect negatively the possibility to collaborate. Working with scientists outside one's field has indeed a high cost, from finding common ground and research questions to a possible construction of integrated knowledge [25].

To determine the propensity to collaborate, we follow the heuristic of the Probabilistic Niche Model (PNM) used in ecology to determine trophic interactions between species [26]. This approach was successfully used in a social science model by [27].

3.2 Model description

Agents are N scientists A_i, characterized by a probability distribution $d(x)$ representing their disciplinary positioning in an abstract way: research is summarized by a one dimensional variable $\mathbb{R}$, and the disciplinary positioning on this axis is given by the distribution. The model is setup with normal distributions of width σ with an average distributed uniformly in $[0;1]$. Scientists also have a time budget per day, that we will summarize as a future timetable $T(t_0) : t > t_0 \mapsto p(t) \in \mathcal{P}$ where $\mathcal{P}$ is the space of scientific projects. The central feature of the model is the utility function $U(d_i, d_j)$ determining an abstract utility for scientist i to collaborate with j for a given project. It will be a function of the disciplinary overlap $o = \int_x d_i(x) \cdot d_j(x) dx$ and different assumptions on the form of this cost function can be tested. We take a linear cost in the overlap and a varying benefit, expressing the fact that researchers have different strategies regarding their interdisciplinary positioning. This way, we have $U(d_i, d_j) = o/i^{\alpha} - o$, assuming a fat-tail distribution of individual preferences for interdisciplinarity, given by a power law of parameter α. A discrete choice formulation gives the probabilities for a scientist i to choose among j collaborators by

$$p_j = \exp\left(\beta U(d_i, d_j)\right) / \sum_k \exp\left(\beta U(d_i, d_k)\right) \tag{1}$$

The social network of relations between scientists can be initialized with the following options: as a fixed scale-free social network, with a correlation parameter ρ

The temporal evolution of the model goes as follows: (i) one scientist with no current activity is picked up at random, and starts a project with one of its potential collaborators taken as its neighbors in the network that have free time, chosen with the probability p_j. The project has a random uniform duration and timetables are updated accordingly; (ii) current projects are updated and finished if necessary. The outcome of the model if measured by average depth across project, defined for one project as the overlapping areas between distribution, and average interdisciplinarity measured by total area covered.

4 RESULTS

The model is implemented in NetLogo [28] and explored with the OpenMole model exploration software [29]. Source code and results are available on the open git repository of the project at `https://github.com/JusteRaimbault/Perspectivism`. Processed empirical data and simulation data used in the paper are available on the dataverse repository at `https://doi.org/10.7910/DVN/GMQ5A8`.

The parameter space explored corresponds to $\alpha, \beta, \sigma, \rho$. We fix the number of agents to $N = 100$.

4.1 Stochasticity of model outputs

A first numerical experiment is conducted to assess the stochasticity of model outputs, and how much repetitions are needed for a robust estimation. We randomly sample 100 points in the parameter space with a Latin Hypercube Sampling, and run 500 repetitions of the model for each point.

The Sharpe ratio (absolute average relative to standard deviation) estimated for each parameter point has for the depth indicator a minimal value of 3.81 and a median of 18.55. For this indicator, even when noise is at his highest value, the average indicator value is 4 times larger. For the interdisciplinarity indicator, minimum is at 61.02, which means that noise is negligible for this output.

4.2 Global sensitivity analysis

To summarize the influence of model parameters on outputs, we proceed to a Global Sensitivity Analysis as introduced by [30].

4.3 Grid exploration

We run a basic grid exploration of the parameter space, both with random and small-world social networks, for parameters α, β, σ with 50 repetitions of the model for each parameter points, corresponding to 158,400 model runs. Figure 2 shows indicators variation on a given subspace and the corresponding Pareto front between depth and interdisciplinarity. We show a second order influence of preference hierarchy α and non-linearity of model behavior as a function of all parameters. Convergence properties are reasonable with this number of repetitions. Large individual disciplinary width σ causes the choice parameter β to have no influence, whereas low values give an increasing interdisciplinarity and a decreasing depth as a function of β. Random behavior ($\beta = 0$) leads to a constant depth of projects. When examining the Pareto front between the two contrary objectives, the optimal points occur for intermediate β when σ is fixed, suggesting non-trivial behavioral optima at a fixed disciplinary configuration. These first exploration show

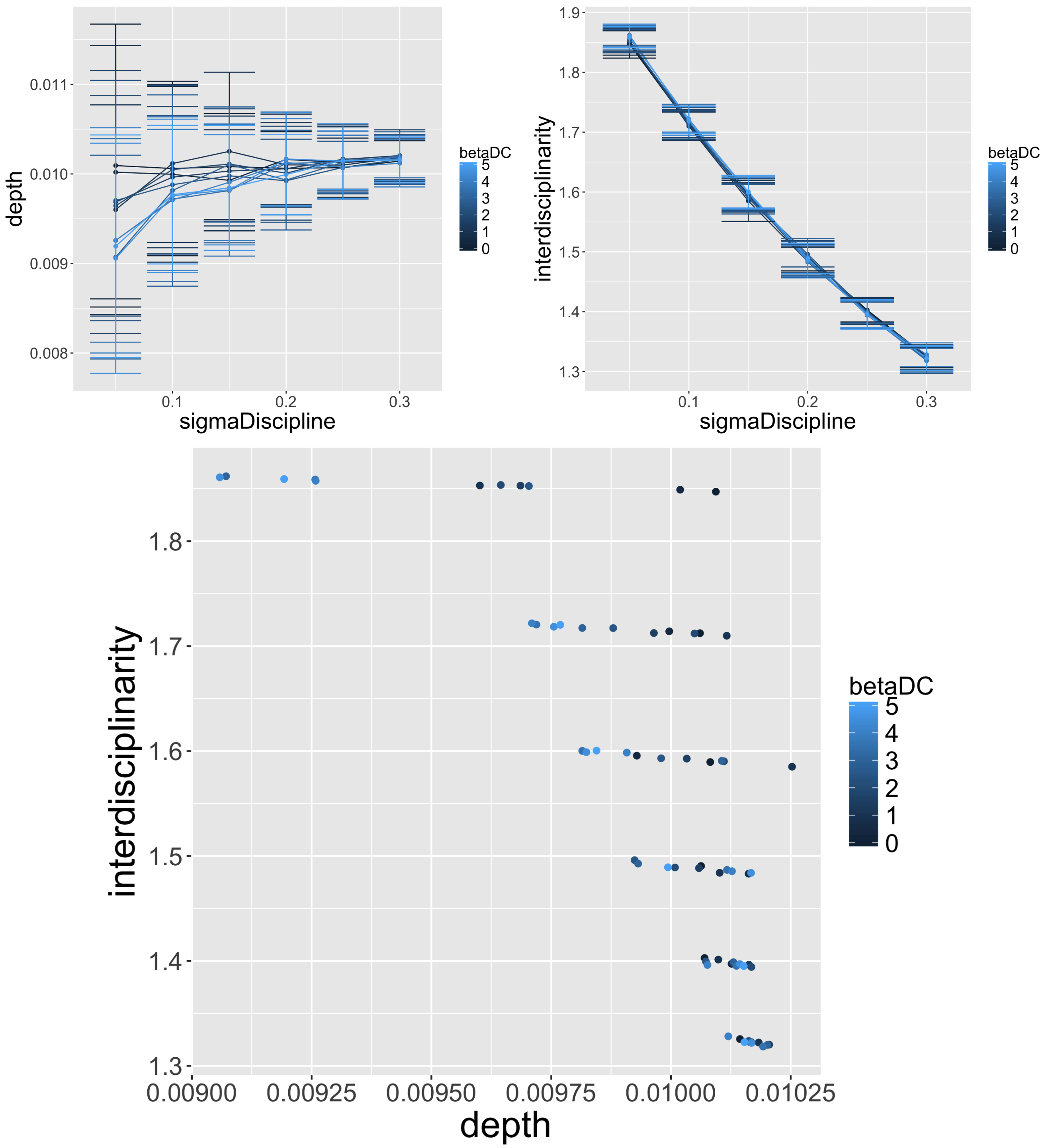

Figure 2. Patterns of interdisciplinarity from model simulations. We show measures of depth and interdisciplinarity (top row) at fixed $\alpha = 0.5$ and network structure, for varying discrete choice parameter β as a function of individual extent σ. On the bottom, the Pareto front of average point between these two objectives.

the complex dynamics of interdisciplinarity even with simple interaction rules and network structure, and suggests further applications such as the exploration of policies by changing network structure or studying in a more refined way the influence of α. Preliminary non-systematic model experiments, in particular changing the type of network structure, suggest that it may also have significant effect on model outcomes.

4.4 Model optimization

The way collaborations are established and indicators are computed, one can expect the level of interdisciplinarity and the depth of projects to be contradictory. To understand the compromise between the two, we run a multi-objective optimization procedure on the model. We use the NSGA2 algorithm (Non-dominated Sorting Genetic Algorithm II) [31] which is an optimization heuristic well-suited for multi-objective problems. The algorithm is integrated into the OpenMOLE software in a way to handle stochasticity and distribute computations with an island scheme [29]. We take a population of $\mu = 200$ and run the algorithm for $10,000$ generations with islands of 15 minutes.

We obtain a linear compromise Pareto front as shown in Fig. **??**.

4.5 Model calibration

5 DISCUSSION

5.1 Implications of results

5.2 Model limitations and possible extensions

Possible refinements of the model, towards a less stylized and more behavioral and micro-based model, could for example include the introduction of time budgets, simultaneous projects and dynamical time investment for scientists. The assumption of two-person projects is also strongly constraining, and relaxing it would require the extension of depth and interdisciplinarity measures that is not necessary straightforward. Furthermore, the absence of learning and of evolution of the social network when completing a project suggests a short time scale of application: further refinements should include dynamics of individual distributions and of individual relationships.

5.3 Perspectivism and Model Coupling

Beyond the simplifying opposition between fully constructivist and realistic approaches to science, several alternatives have been developed, among which Perspectivism [22] is a way to tackle most of the issues opposing these two by taking an agent-based approach to the production of scientific knowledge. The main feature of this viewpoint is to consider each scientific enterprise as a single perspective, in which an agent aims at understanding an aspect of the real world (the ontology) with the mean of a medium, which is considered as a model. Constituted disciplines thus contains more or less compatible perspectives. The explicitation of this approach has been done by [32] to embed it into knowledge domains, as a generalization of knowledge domains introduced by [33].

We postulate that this approach to science may be a powerful tool to foster interdisciplinary collaborations, if used in a reflexive way in the construction of projects. [34] propose a similar framework. More precisely, we suggest to apply an "Applied Perspectivism", in the sense of an explicit perpectivist positioning within a given collaboration, and associated guidelines and protocols for collaboration. This would imply a high-level of reflexivity for each agent implied, a mapping of the different layers of the enterprise and the positioning of each agent regarding the domains of knowledge. This way, in the particular case of model coupling, the explicitation of positioning and of the structure of each knowledge implied should ease interactions. As Banos points out [35], transversal work must alternate with deeper investigations in each discipline, in a kind of "virtuous circle" [36]. Fostering a synergy between complementary knowledge is the core aspect more important than interdisciplinarity in itself [37]. This raises the issue of, before individual researcher particularities, how a given collective structure of scientific knowledge production should balance between these disciplinary and interdisciplinary knowledge. It is clear that this question is deeply endogenous to each studied subject, and even each particular approach taken, but within the applied knowledge framework described above, we have reasons to believe that certain structural properties may be rather general. Indeed, each discipline is expected to bring components for each knowledge domain, and the co-evolving perspective is built on their interrelations. This paper proposed to investigate basic aspects of this issue, by means of agent-based modeling.

This work aimed at providing quantitative evidence of the feasibility of the epistemological point of view described above and inform potential implementation for some of its processes, more precisely how can certain level of coupling of perspectives (or overlap of ontologies) may be achieved given specializations of scientists and a given dynamic of interaction.

6 CONCLUSION

In conclusion, we show with a simple model that the individual choices produce an emerging structure of the research front, suggesting that applied perspectivism requires a careful tuning of research structure and researcher behaviors since Pareto-optimal configurations correspond to non-trivial parameter points. Future developments should include more realistic behavioral assumption, and a formalisation of the applied perspectivism approach to include it in the agent-based model.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

JR designed the study, coded the model, conducted the experiments, and wrote the paper.

FUNDING

The authors acknowledge the funding of the EPSRC project number EP/M023583/1.

ACKNOWLEDGMENTS

Results obtained in this paper were computed on the vo.complex-system.eu virtual organization of the European Grid Infrastructure (http://www.egi.eu). We thank the European Grid Infrastructure and its supporting National Grid Initiatives (France-Grilles in particular) for providing the technical support and infrastructure.

DATA AVAILABILITY STATEMENT

The datasets analyzed and generated for this study can be found in the dataverse repository at `https://doi.org/10.7910/DVN/GMQ5A8`.

REFERENCES

[1] Hofstra B, Kulkarni VV, Munoz-Najar Galvez S, He B, Jurafsky D, McFarland DA. The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences* **117** (2020) 9284–9291. doi:10.1073/pnas.1915378117.

[2] Jang J, Ju X, Ryu U, Om H. Coevolutionary characteristics of knowledge diffusion and knowledge network structures: A ga-abm model. *Journal of Artificial Societies & Social Simulation* **22** (2019).

[3] Urbanska K, Huet S, Guimond S. Does increased interdisciplinary contact among hard and social scientists help or hinder interdisciplinary research? *PloS one* **14** (2019).

[4] Gross K, Bergstrom CT. Contest models highlight inherent inefficiencies of scientific funding competitions. *PLoS biology* **17** (2019).

[5] Akerlof GA, Michaillat P. Persistence of false paradigms in low-power sciences. *Proceedings of the National Academy of Sciences* **115** (2018) 13228–13233.

[6] Pavlidis I, Petersen AM, Semendeferi I. Together we stand. *Nature Physics* **10** (2014) 700.

[7] Rouse WB, Lombardi JV, Craig DD. Modeling research universities: Predicting probable futures of public vs. private and large vs. small research universities. *Proceedings of the National Academy of Sciences* **115** (2018) 12582–12589.

[8] van Vlokhoven H. The effect of open access on research quality. *Journal of Informetrics* **13** (2019) 751 – 756. doi:https://doi.org/10.1016/j.joi.2019.04.001.

[9] Thurner S, Liu W, Klimek P, Cheong SA. The role of mainstreamness and interdisciplinarity for the relevance of scientific papers. *arXiv e-prints* (2019) arXiv:1910.03628.

[10] Zeng A, Shen Z, Zhou J, Fan Y, Di Z, Wang Y, et al. Increasing trend of scientists to switch between topics. *Nature communications* **10** (2019) 1–11.

[11] Larivière V, Gingras Y. On the relationship between interdisciplinarity and scientific impact. *Journal of the Association for Information Science and Technology* **61** (2010) 126–131.

[12] [Dataset] Brown J, Murray D, Furlong K, Coco E, Dablander F. A breeding pool of ideas: Analyzing interdisciplinary collaborations at the complex systems summer school (2020). doi:10.31235/osf.io/e3z4v.

[13] Pluchino A, Burgio G, Rapisarda A, Biondo AE, Pulvirenti A, Ferro A, et al. Exploring the role of interdisciplinarity in physics: Success, talent and luck. *PloS one* **14** (2019).

[14] Giere RN. An agent-based conception of models and scientific representation. *Synthese* **172** (2010) 269–281.

[15] Shafiee ME, Berglund EZ. Agent-based modelling approach to evaluate the effect of collaboration among scientists in scientific workflows. *Journal of Simulation* **13** (2019) 1–13.

[16] Bornmann L, Ganser C, Tekles A, Leydesdorff L. Does the h-index reinforce the matthew effect in science? the introduction of agent-based simulations into scientometrics. *Quantitative Science Studies* **1** (2020) 331–346. doi:10.1162/qss_a_00008.

[17] Feliciani T, Luo J, Ma L, Lucas P, Squazzoni F, Maruvsic A, et al. A scoping review of simulation models of peer review. *Scientometrics* **121** (2019) 555–594.

[18] Kovanis M, Porcher R, Ravaud P, Trinquart L. Complex systems approach to scientific publication and peer-review system: development of an agent-based model calibrated with empirical journal data. *Scientometrics* **106** (2016) 695–715.

[19] Shneiderman B. Twin-win model: A human-centered approach to research success. *Proceedings of the National Academy of Sciences* **115** (2018) 12590–12594.

[20] Dionne SD, Sayama H, Yammarino FJ. Diversity and social network structure in collective decision making: Evolutionary perspectives with agent-based simulations. *Complexity* **2019** (2019).

[21] Chavalarias D. What's wrong with science? *Scientometrics* (2016) 1–23.
[22] Giere RN. *Scientific perspectivism* (University of Chicago Press) (2010).
[23] [Dataset] Clement CB, Bierbaum M, O'Keeffe KP, Alemi AA. On the use of arxiv as a dataset (2019).
[24] Raimbault J. Exploration of an interdisciplinary scientific landscape. *Scientometrics* **119** (2019) 617–641.
[25] Frodeman R. *Sustainable knowledge: A theory of interdisciplinarity* (Springer) (2013).
[26] Williams RJ, Anandanadesan A, Purves D. The probabilistic niche model reveals the niche structure and role of body size in a complex food web. *PloS one* **5** (2010) e12092.
[27] Raimbault J, Broere J, Somveille M, Serna JM, Strombom E, Moore C, et al. A spatial agent based model for simulating and optimizing networked eco-industrial systems. *Resources, Conservation and Recycling* **155** (2020) 104538.
[28] Tisue S, Wilensky U. Netlogo: A simple environment for modeling complexity. *International conference on complex systems* (Boston, MA) (2004), vol. 21, 16–21.
[29] Reuillon R, Leclaire M, Rey-Coyrehourcq S. Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Generation Computer Systems* **29** (2013) 1981–1990.
[30] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. *Global sensitivity analysis: the primer* (John Wiley & Sons) (2008).
[31] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation* **6** (2002) 182–197.
[32] Raimbault J. An applied knowledge framework to study complex systems. *Forthcoming in CSDM2017 proceedings. arXiv:1706.09244 at https://arxiv.org/abs/1706.09244* (2017).
[33] Livet P, Müller JP, Phan D, Sanders L, Auatabu T. Ontology, a mediator for agent-based modeling in social science. *Journal of Artificial Societies and Social Simulation* **13** (2010).
[34] Ellemers N, Fiske ST, Abele AE, Koch A, Yzerbyt V. Adversarial alignment enables competing models to engage in cooperative theory building toward cumulative science. *Proceedings of the National Academy of Sciences* **117** (2020) 7561–7567. doi:10.1073/pnas.1906720117.
[35] Banos A. *Pour des pratiques de modélisation et de simulation libérées en géographie et SHS*. Ph.D. thesis (2013).
[36] Banos A. Knowledge accelerator'in geography and social sciences: Further and faster, but also deeper and wider. *Urban Dynamics and Simulation Models* (2017) 119–123.
[37] Leydesdorff L, Ivanova I. The measurement of "interdisciplinarity" and "synergy" in scientific and extra-scientific collaborations. *Available at SSRN* (2020).