Supplementary Materials for

How New Concepts Become Universal Scientific Approaches – Insights from Citation Network Analysis of Agent-based Complex Systems Science

Christian E. Vincenot

Correspondence to: Christian@Vincenot.biz

DOI: 10.1098/rspb.2017-2360

This PDF file includes:

Figs.S1 to S5
Tables S1 to S3
Captions for Movie S1
Extended Methodology

Other Supplementary Materials for this manuscript includes the following:

Movie S1

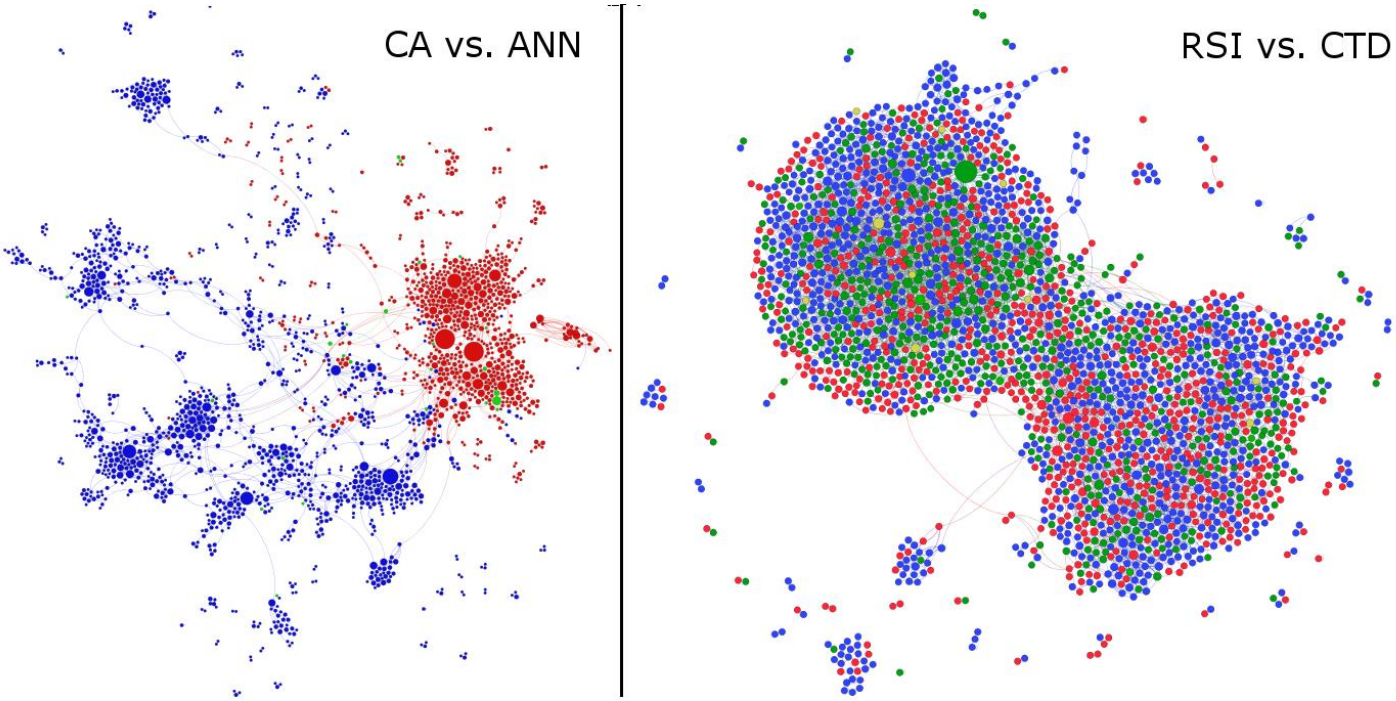


Fig. S1. Illustration of disjunct and fused disciplines. For the sake of comparison with the ABM-IBM case study, visualizations with the same methodology of typical disjunct and fused disciplines, namely Cellular Automata (CA; red) vs. Artificial Neural Network (ANN; blue) (left) and the study of Repetitive Strain Injury (RSI; red) vs. Cumulative Trauma Disorder (CTD; blue) (right), is provided. Green nodes represent publications assimilated to both corpora under study. Node size is linearly dependent on in-degree. Node location is computed based on attraction by neighbors (i.e., nodes connected through citations). The 2000 most cited papers of each community were retrieved from Scopus.

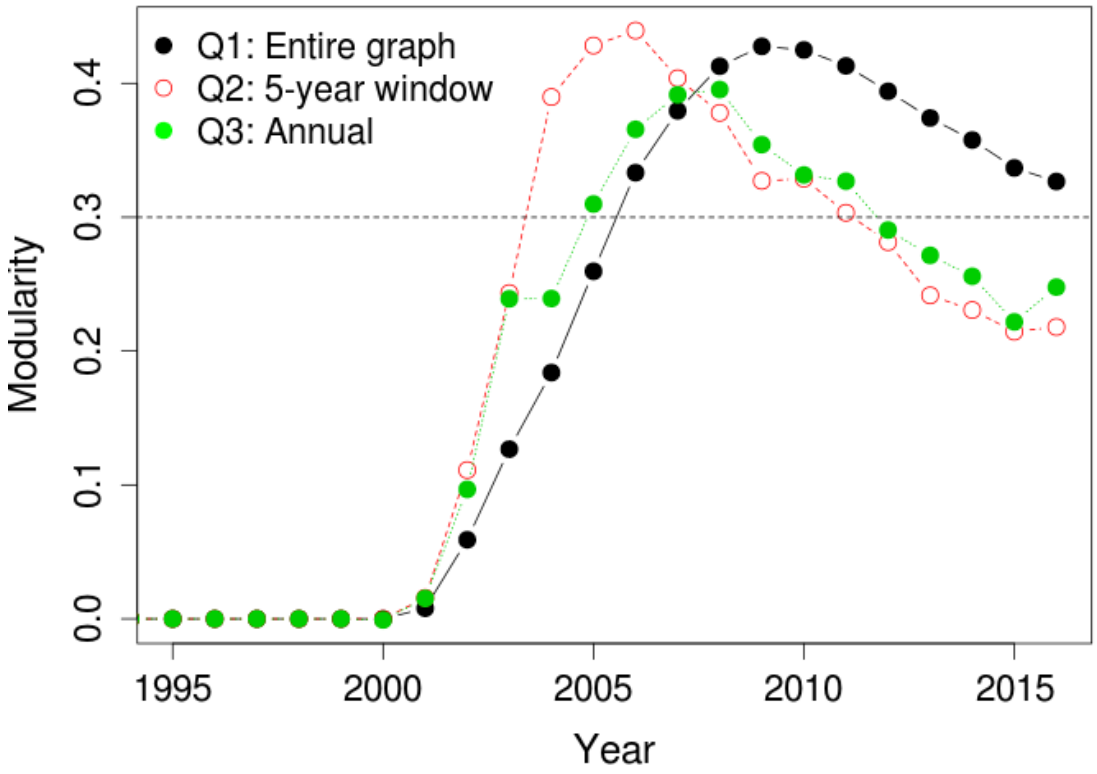


Fig. S2. Evolution of community clustering within ACS expressed in terms of modularity of the citation graph. Three measures of modularity are reported: Q_1 (black dots), the classical modularity; Q_2 (red circles), the modularity of the graph produced when only papers published within a window of 5 years are considered; Q_3 (green dots), the modularity of the graph produced when only papers published in a given year as well as publications cited therein are considered. When taking into account all papers in the calculation (Q_1), modularity declined only slowly from 2010 onwards owing to the cumulated clustering of existing publications, which generated inertia in the metric.

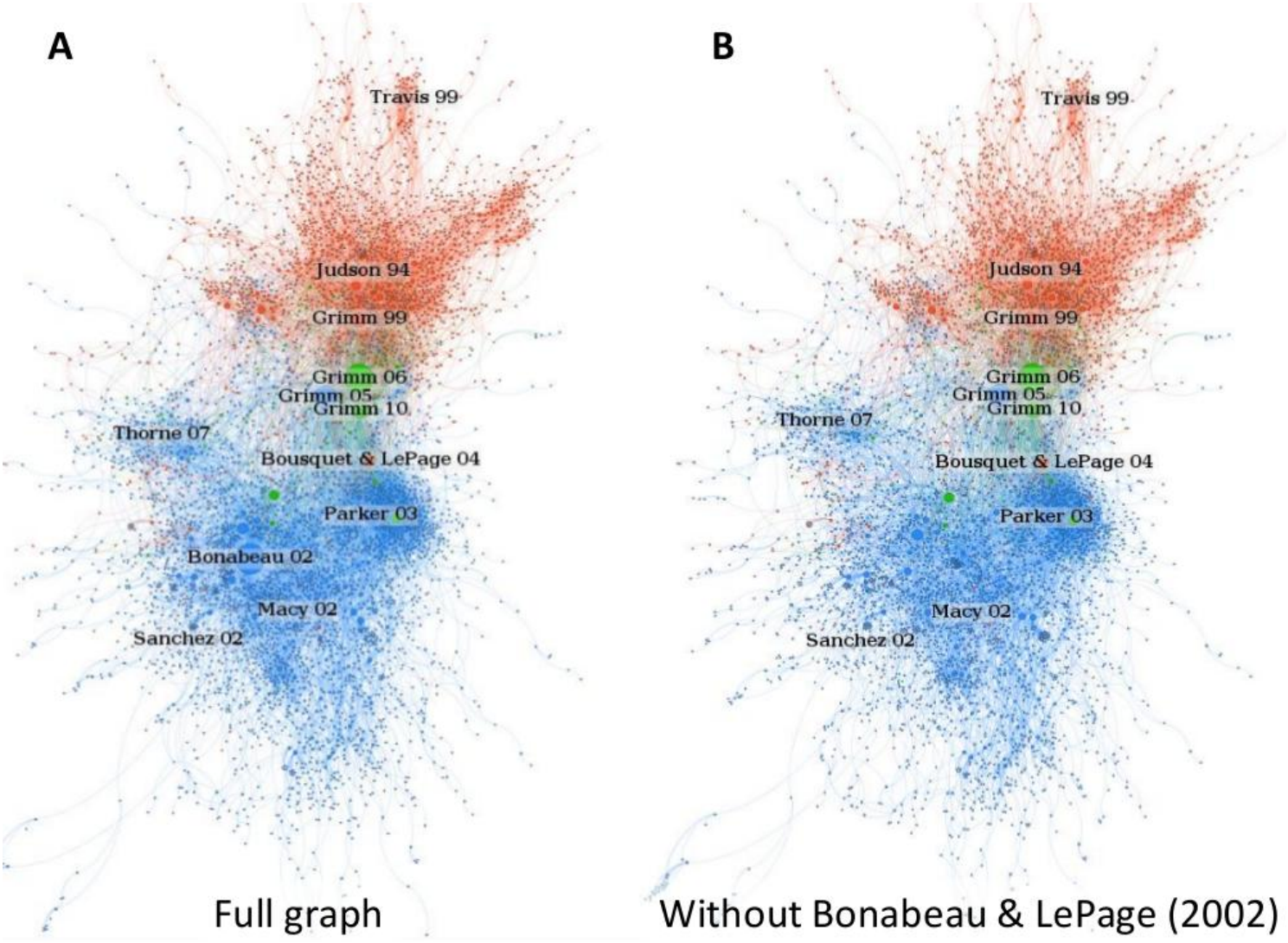


Fig. S3. Lack of influence of global paper citation count on the fusion of the ABM and IBM corpora. In a similar fashion to Fig. 5, the arrangement of the graph following the ForceAtlas 2 algorithm is represented here (A) with and (B) without the most cited paper in our dataset, namely Bonabeau and LePage (2002) (named in the graph "Bonabeau 02"). The layout—more particularly the distance between the ABM and IBM corpora—is unchanged.

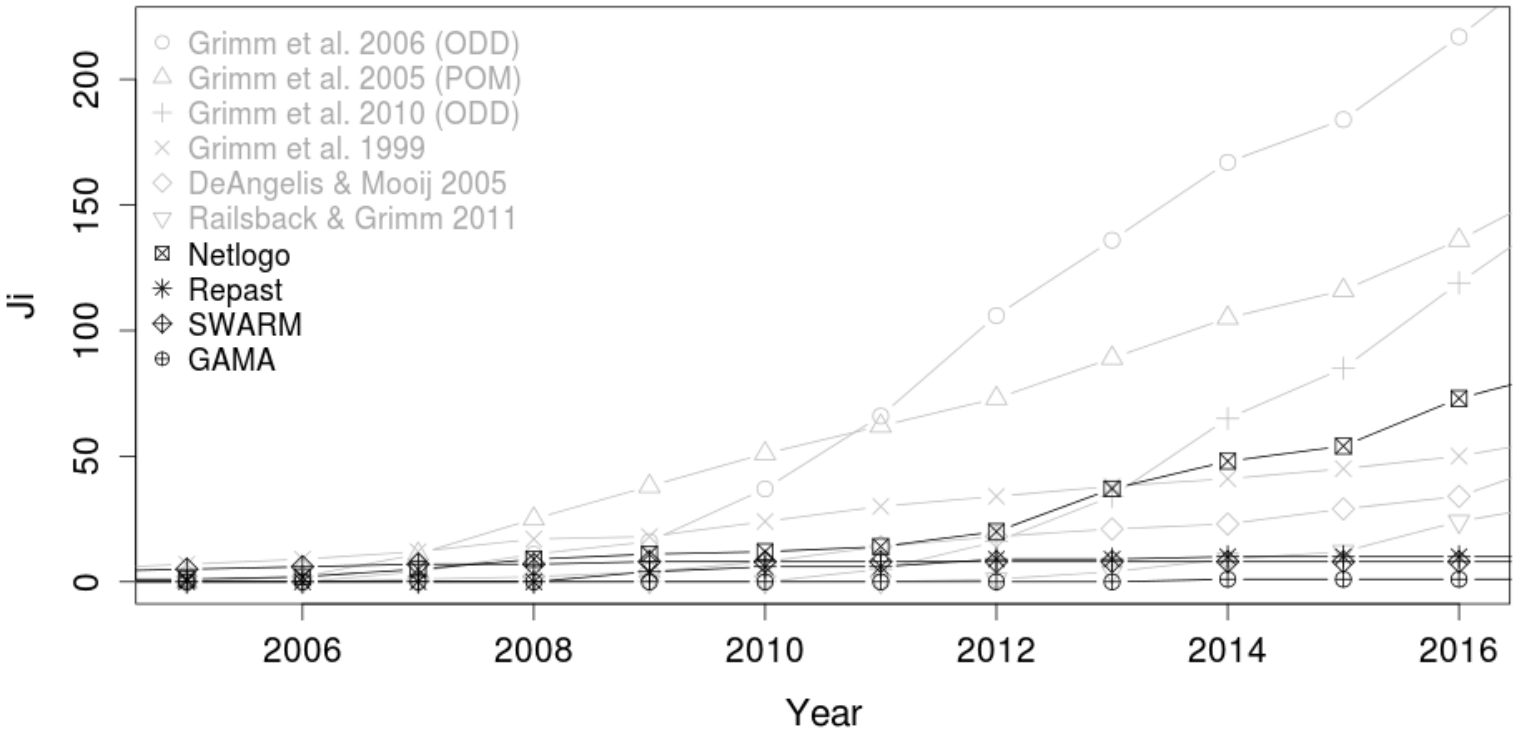


Fig. S4. Evolution of the importance of software in the unification of ACS measured with the metric J_i, which represents the minimum between the number of citations stemming from the ABM and the IBM corpora (see *Extended Methodology* below). The plot presents values for four common agent-based modeling platforms (in black) compared to the six key publications (in gray) identified as most important in fostering the fusion between ABM and IBM.

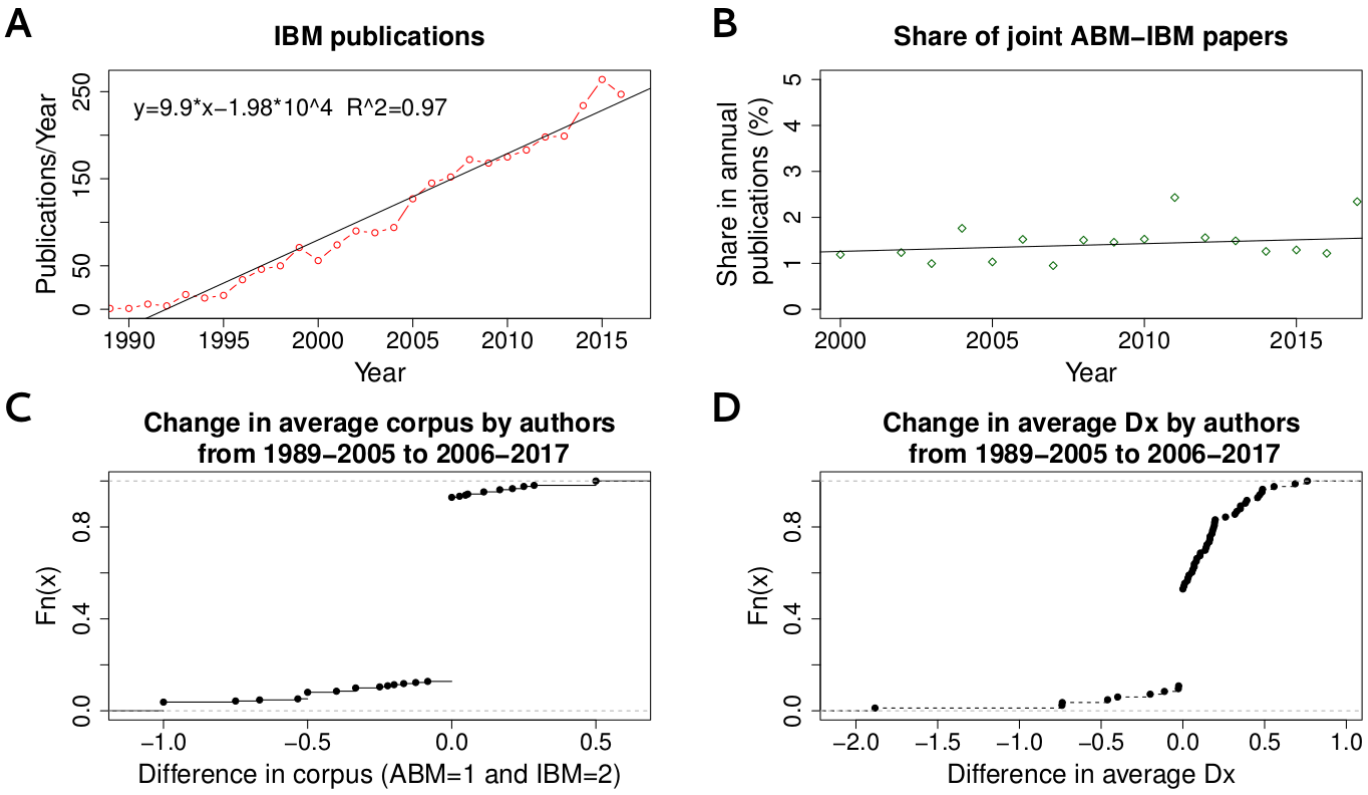


Fig. S5. Arguments to reject the alternative hypothesis that would explain observations by a simple terminological switch (i.e., replacement of the term "IBM" by "ABM" in the literature). (A) Instead of declining, the number of IBM publications per year continued growing steadily. (B) The share of joint ABM-IBM papers published annually (i.e., number of joint ABM-IBM papers divided by the total number of papers published every year) has not increased. (C) A large majority of authors (80% of the 211 tracked) have not even once changed terminology. The plot displays the cumulative distribution function Fn(x) of the change in corpus for authors. The latter was calculated for each author as an average of the corpus values assigned to his publications. Zero thus represents no change in terminology between the periods 1989-2005 and 2006-2017, whereas positive and negative values indicate changes towards a more frequent use of the terms "IBM" and "ABM" respectively. (D) Most of the authors (56%) cited more frequently papers from the other corpus. The plot shows the cumulative distribution function Fn(x) of the difference in hetero-citation balance Dx for authors between 1989-2005 and 2006-2017. To avoid ambiguities in the interpretation, this calculation was limited to authors who have never changed terminology between these two periods.

Table S1. Top 10 most cited publications in both ABM and IBM corpora. "IBM->" and "ABM->" represent the number of times the paper was cited by IBM and ABM papers respectively. Citations by joint ABM-IBM nodes are ignored. Papers are ranked based on the metric J_i, which represents the minimum between the number of citations from the IBM and ABM corpora. Note that only citations within this network are considered. Moreover, this metric accounts for citations of a given paper by other publications and not citations of other publications by the given paper. The six key papers responsible for the unification of ACS (see *The Role of Key Methodological Papers*) appear in boldface type.

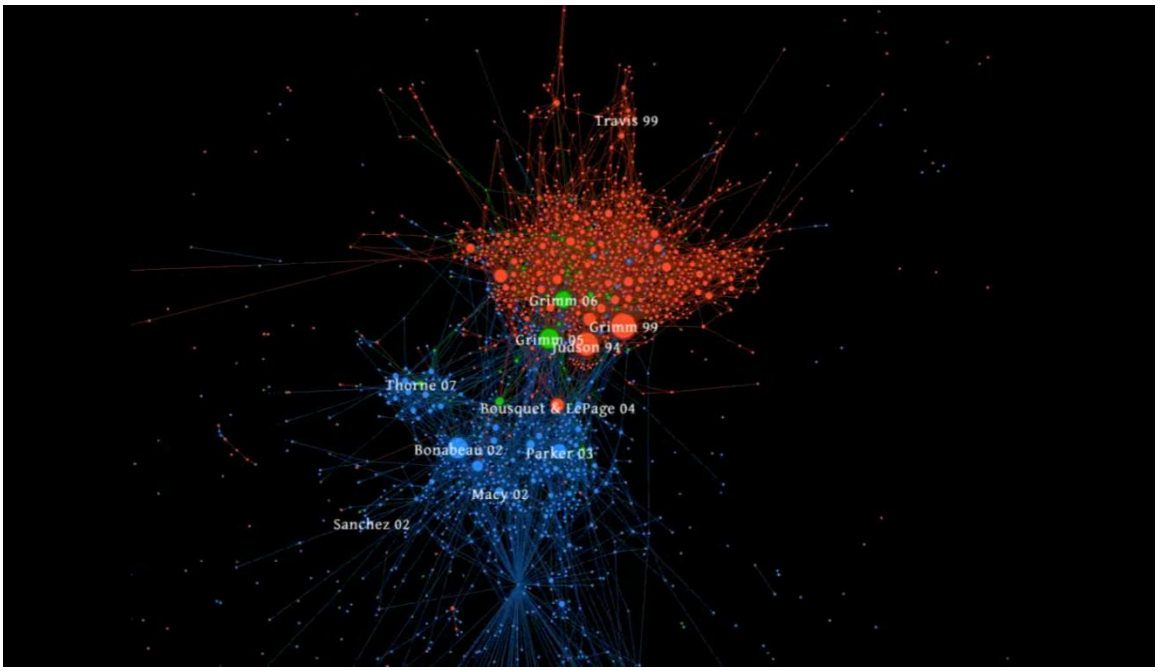
Authors	Reference	Year	Corpus	IBM->	ABM->	J_i
Grimm et al.	**Ecological Modelling 198:115–126**	**2006**	**Both**	**198**	**291**	**198**
Grimm	**Science 310:987–991**	**2005**	**ABM**	**132**	**208**	**132**
Grimm et al.	**Ecological Modelling 221:2760–2768**	**2010**	**Both**	**116**	**255**	**116**
Grimm	**Ecological Modelling 115:129–148**	**1999**	**IBM**	**147**	**42**	**42**
DeAngelis and Mooij	**Annu. Rev. Ecol. Evol. Syst. 36:147–68**	**2005**	**IBM**	**78**	**36**	**36**
Railsback and Grimm	**Agent-based and individual-based modeling: A practical introduction (Book)**	**2011**	**Both**	**26**	**106**	**26**
Kreft, Booth, and Wimpenny	Microbiology 144:3275-87	1998	IBM	54	20	20
Topping et al.	Ecological Modelling 167:65-82	2003	Both	17	24	17
Bousquet and LePage	Ecological Modelling 176:313–332	2004	IBM	15	122	15
Railsback	Ecological Modelling 139:47–62	2001	IBM	34	14	14

Table S2. Top 25 most cited publications within the network. The six key papers responsible for the unification of ACS (see *The Role of Key Methodological Papers*) appear in boldface type.

Authors	Reference	Year	Corpus	Deg^{--}
Bonabeau	Proceedings of the National Academy of Sciences 99:7280–7287	2002	ABM	648
Grimm et al.	**Ecological Modelling 198:115–126**	**2006**	**Both**	**526**
Grimm et al.	**Ecological Modelling 221:2760–2768**	**2010**	**Both**	**394**
Grimm	**Science 310:987-991**	**2005**	**ABM**	**367**
Parker et al.	Annals of the Association of American Geographers 93:314–337	2003	ABM	234
Grimm	**Ecological Modelling 115:129–148**	**1999**	**IBM**	**204**
Macy and Willer	Annual Review of Sociology 28:143–166	2002	ABM	187
Railsback, Lytinen, and Jackson	Simulation 82:609-623	2006	Both	169
North and Macal	Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation (Book)	2007	ABM	167
North et al.	ACM Transactions on Modeling and Computer Simulation 16:1-25	2006	ABM	166
Macal and North	Journal of Simulation 4:151–162	2010	ABM	159
Matthews et al.	Landscape Ecology 22:1447–1459	2007	Both	156
Macal and North	Proceedings of the Winter Simulation Conference, 2005	2005	ABM	153
Bousquet and LePage	Ecological Modelling 176:313–332	2004	IBM	144
Railsback and Grimm	**Agent-based and individual-based modeling: A practical introduction (Book)**	**2011**	**Both**	**144**
DeAngelis and Mooij	**Annu. Rev. Ecol. Evol. Syst. 36:147–68**	**2005**	**IBM**	**129**
Macal and North	Proceedings of the Winter Simulation Conference, 2009, 86-98	2009	ABM	119
Rahmandad and Sterman	Management Science 54:998–1014	2008	ABM	113
Judson	Trends in Ecology & Evolution 9:9–14	1994	IBM	107
Windrum, Fagiolo, and Moneta	Journal of Artificial Societies and Social Simulation 10 n. 2	2007	ABM	103
Luke et al.	Simulation 81: 517-527	2005	ABM	98
Brown et al.	Journal of Geographical Systems 7:25–47	2005	ABM	85
Brown et al.	International Journal of Geographical Information Science 19:153–174	2005	ABM	83
Kreft et al.	Microbiology 147:2897-912	1998	IBM	79
Rand and Rust	International Journal of Research in Marketing 28:181–193	2011	ABM	79

Table S3. Top 25 publications in betweenness centrality (B.C.). The six key papers responsible for the unification of ACS (see *The Role of Key Methodological Papers*) appear in boldface type.

Authors	Reference	Year	Corpus	B.C.
DeAngelis and Mooij	**Annu. Rev. Ecol. Evol. Syst. 36:147–68**	**2005**	**IBM**	**232374**
Grimm et al.	**Ecological Modelling 198:115–126**	**2006**	**Both**	**159658**
Grimm et al.	**Ecological Modelling 221:2760–2768**	**2010**	**Both**	**80035**
Railsback, Lytinen, and Jackson	Simulation 82:609-623	2006	Both	74592
Heath, Hill, Ciarallo	Journal of Artificial Societies and Social Simulation 12:9	2009	ABM	40408
Railsback and Grimm	**Agent-based and individual-based modeling: A practical introduction (Book)**	**2011**	**Both**	**32227**
Heckbert, Baynes, and Reeson	Annals of the New York Academy of Sciences 1185:39–53	2010	ABM	29913
van Nes, Lammens, and Scheffer	Ecological Modelling 152:261-278	2002	IBM	29816
Lammens, van Nes, and Mooij	Freshwater Biology 47:2435–2442	2002	IBM	27926
Grimm	**Ecological Modelling 115:129-148**	**1999**	**IBM**	**27618**
An	Ecological Modelling 229:25–36	2012	ABM	26380
Bousquet and LePage	Ecological Modelling 176:313–332	2004	IBM	25434
Strand, Huse, and Giske	American Naturalist 159:624-44	2002	IBM	23706
Caron-Lormier et al.	Ecological Modelling 212:522–527	2008	IBM	21866
Grimm et al.	Ecological Modelling 115:275–282	1999	IBM	20429
Railsback	Ecological Modelling 139:47-62	2001	IBM	20213
Matthews et al.	Landscape Ecology 22:1447-1459	2007	Both	19104
Hellweger and Bucci	Ecological Modelling 220:8–22	2009	Both	18070
Nikolai and Madey	Journal of Artificial Societies and Social Simulation 12 n.2	2009	Both	18041
Smajgl et al.	Environmental Modelling & Software 26:837–844	2011	ABM	17523
Filatova et al.	Environmental Modelling & Software 45:1-7	2013	ABM	17432
Parrott and Kok	Complexity International 7:1–19	2000	Both	15886
Macal and North	Proceedings of the Winter Simulation Conference, 2009, 86-98	2009	ABM	15498
Crooks and Heppenstall	Agent-Based Models of Geographical Systems, 85-105	2012	ABM	15096
Macal and North	Proceedings of the Winter Simulation Conference, 2006, 73-83	2006	ABM	14550



Movie S1. Genesis and evolution of the science of agent-based complex systems seen from the perspective of its citation network from 1990 to January 2016. The dynamic graph shows the temporal dynamics of the agent-based modeling (blue vertices) and individual-based modeling (red vertices) corpora. Vertices are positioned based on the ForceAtlas2 layout algorithm without vertex overlapping.

This video can be viewed on the Royal Society's website following the link provided in the manuscript. A high-resolution version can be downloaded from:
http://dx.doi.org/10.5061/dryad.19nr2

Extended Methodology

Bibliographic Dataset

The raw bibliographic dataset supporting this study was built upon Scopus records. With permission of Elsevier, this file was uploaded on Dryad and can be downloaded from: http://dx.doi.org/10.5061/dryad.19nr2

Dynamic network visualization

The citation network was visualized in Gephi 0.8.2 (61). Orphan vertices (i.e., null degree vertices) were removed and the graph was arranged following the ForceAtlas2 algorithm without spatial overlapping, under which vertices simply repulse each other while edges attract the vertices that they link (62). Vertex radius was set to be linearly proportional to in-degree, such that highly cited publications would appear as larger vertices. The reduced graph (7,340 vertices) was rendered with dynamic connectivity (i.e., time-tagged edges; e.g. 40) using publication dates as timeline. Changes in the citation network through time could be observed in this manner.

Based on this visualization procedure, papers potentially initiating the fusion of ABM and IBM research could be identified by visually inspecting the graph. Vertices influential in the unification of ACS were tentatively discriminated by their size coupled with their location in the area of space where the clusters of ABM and IBM vertices overlapped. This first-stage visual inspection was, however, rigorously complemented by the computation quantitative metric J_i (see *Materials and Methods* and Table S1) for each node of the graph.

Note that the metrics described in the manuscript (modularity, hetero-citation indices, Ji, etc) were independent of the ForceAtlas 2 layout algorithm, which served only a visualization purpose.

Baseline cases

This metric was validated with corpora that were either clearly separated or clearly integrated. Corpora representative of the first scenario related to “Cellular Automata” (CA) and “Artificial Neural Network” (ANN), which are radically different modeling approaches, whereas the second situation was tested by building corpora based the synonymous terms “Repetitive Strain Injury” (RSI) and “Cumulative Trauma Disorder” (CTD) (59).

Significance of the trend in modularity and assortativity

We performed simulations to assess whether the post-2005 trend in modularity and assortativity reflected effective changes in hetero-citation balance or was merely the result of chance (the null hypothesis). Parametric bootstrap replications produced 40 graphs in which vertex outdegree (i.e. the number of references from each publication) remained untouched compared to the original graph, but the hetero-citation balance after 2005 was artificially kept constant at the same value as observed in average in 1987-2005. This was done by computing the hetero-citation balance for the latter period and by reconnecting outgoing edges from post-2005 vertices to maintain this value constant. Herero-citation share $S(V_i)$ was recalculated for each vertex V_i following the formula obtained by reversing Equation 3:

$$S(V_i) = \frac{n_Y(V_i) \cdot (D(V_i) + 1)}{n_{X \cup Y}(V_i)}$$

Modularity and assortativity timeseries were then computed on each graph to produce an envelope of trajectories for the null hypothesis scenario.

Importance of software

Software is frequently referenced through websites, repositories, conference papers, or grey literature, which are all not indexed. They were therefore absent from Scopus records on which the aforementioned analysis was based. To palliate this limitation, citations to four common modeling platforms were extracted from the "Cited References" field in our dataset and values of the J_i metric were reconstructed. References to Netlogo, RePast, SWARM, and GAMA were identified based on the citation format recommended for each of them (32-34, 58).

Validation of the interpretation and author-based analysis

The use of the terms "individual-based" in some scientific research fields (e.g., ecology) and "agent-based" in others (e.g., computer science) became the basis for this exploration of how ACS research has been integrated. The method presented here, which allowed for the quantitative study of community dynamics, was thus based on the assumption that terminology reflects community membership. Consequently, our conclusion on a unification of ACS based on decrease in modularity and increase in hetero-citation in ABM and IBM publications could be challenged and our results interpreted alternatively as simple change in terminological usage by authors. One could indeed object that the more popular term "ABM" is simply replacing the older term "IBM" in the literature, thereby engendering our observations. This competing hypothesis was rejected based on a *reductio ad absurdum* argument relying on four postulates:

(i) A terminological replacement of "IBM" by "ABM" would most probably result in a decrease in IBM publications
(ii) As corollary of (i), the number of joint ABM-IBM papers would increase sharply
(iii) A terminological switch would, by definition, consist in authors first using systematically one term and later on starting to use the other term.
(iv) The hetero-citation of authors who use consistently one term only should not change significantly

The remarkably constant increase in number of IBM and joint ABM-IBM papers proved the two first postulates false (Fig. S5A and S5B, respectively).

Author-based analysis was carried out to address the remaining statements. First, the sequence of corpora of publications published by each author was constructed. By coding ABM and IBM publications as values of 1 and 2 respectively (here, joint ABM-IBM publications were ignored), the trend in terminology used by each author in his publications was identified, and further analysis was restricted to authors with at least two publications in each of 1989-2005 and 2006-2017 (n=211 with 2131 publications). We note that this approach is highly sensitive and prone to overestimating changes, as neither contribution importance (i.e., position in the papers' authors list) nor a cut-off value to discriminate significant changes in terminology were considered. Consequently, an "IBM

author" who contributed only in a minor way (e.g., 15th author out of 20) to a single ABM publication was regarded as having switched terminology. Still, terminological changes between 1989-2005 and 2006-2017 were infrequent (Fig. S5C), not significant ($p=0.24$, asymptotic paired permutation test), and invalidated assertion (iii).

Lastly, the hetero-citation balance of all authors (i.e., average of the hetero-citation balance of their publications) who had not changed terminology was calculated over the above-mentioned periods (Fig. S5D). Publications not referencing any other work (i.e., leaf vertices) were removed from authors' publications lists, and joint ABM-IBM publications were coded with a value of 1.5 (i.e., between 1 and 2, representing ABM and IBM, respectively). The increase in hetero-citation in 63 of the 112 authors tracked contradicted point (iv).

The foregoing demonstration based on strong inference disproved that the present observations are due to changes in terminology. We incidentally note that this alternative hypothesis, if it had proven right, while casting doubt on the interpretation of the mechanisms at play (i.e., cross-fertilization of the ABM and IBM literature), would not have refuted our conclusion (i.e., emergence of ACS), as a terminological consolidation would in fact per se be a sign of coalescence of the communities.