PLOS ONE

RESEARCH ARTICLE

A Theoretical Model for the Associative Nature of Conference Participation

Jelena Smiljanić[1,2], Arnab Chatterjee[3], Tomi Kauppinen[4], Marija Mitrović Dankulov[1]*

1 Scientific Computing Laboratory, Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia, **2** School of Electrical Engineering, University of Belgrade, P.O. Box 35-54, 11120 Belgrade, Serbia, **3** Condensed Matter Physics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India, **4** Aalto University School of Science, P.O. Box 11000. FI-00076 AALTO, Finland

* mitrovic@ipb.ac.rs

Citation: Smiljanić J, Chatterjee A, Kauppinen T, Mitrović Dankulov M (2016) A Theoretical Model for the Associative Nature of Conference Participation. PLoS ONE 11(2): e0148528. doi:10.1371/journal.pone.0148528

Editor: Matjaz Perc, University of Maribor, SLOVENIA

Received: November 26, 2015

Accepted: December 8, 2015

Published: February 9, 2016

Data Availability Statement: All relevant data files are available from https://figshare.com/articles/Conference_Datasets/2066907.

Funding: This work was supported by ON171017, The Ministry of Education, Science, and Technological Development of the Republic of Serbia (http://www.mpn.gov.rs/: J.S. and M.M.D.); and 675121, The European Commission (http://ec.europa.eu/index_en.htm: M.M.D.).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Participation in conferences is an important part of every scientific career. Conferences provide an opportunity for a fast dissemination of latest results, discussion and exchange of ideas, and broadening of scientists' collaboration network. The decision to participate in a conference depends on several factors like the location, cost, popularity of keynote speakers, and the scientist's association with the community. Here we discuss and formulate the problem of discovering how a scientist's previous participation affects her/his future participations in the same conference series. We develop a stochastic model to examine scientists' participation patterns in conferences and compare our model with data from six conferences across various scientific fields and communities. Our model shows that the probability for a scientist to participate in a given conference series strongly depends on the balance between the number of participations and non-participations during his/her early connections with the community. An active participation in a conference series strengthens the scientist's association with that particular conference community and thus increases the probability of future participations.

Introduction

Social data at a large scale is nowadays available over the internet. Researchers are making the best use of these data to find trends, statistics and patterns, which sometime reveal as robust features, similar to 'laws' in natural science. In recent years, a huge community of researchers [1] including mathematicians, statisticians, computer scientists, theoretical physicists, sociologists, economists, financial analysts, geographers, anthropologists, and biologists of various sub-disciplines have contributed to a larger, developing field, commonly known as 'computational social science' [2]. Empirical data, after a rigorous analysis produces information that is of immense interest for theoreticians. Statistical mechanics, which has been proved to be versatile in modeling phenomena across different areas of physics, and beyond, seems to be the most desired tool even for the above emerging discipline [3, 4].

The abundance of a new data about scientific activities such as publications, collaborations, and citations led to the emergence of a new interdisciplinary field of research about science and how science works [5]. These studies provide insights about the impact of scientists and their publications [6–8], authors' reputation and scientific success [9], patterns of collaboration and their impact on authors' reputation [10, 11], the role of cumulative advantage in career longevity [12, 13] and scientific mobility [14] among many other things. Despite the attention given to publication records and citation patterns, another integral part of modern science, scientific meetings, have so far been largely overlooked. This negligence is particularity interesting, given the pervasive role of the meetings in scientific disciplines. Scientific meetings provide arenas for a fast dissemination of the latest results, exchange and evaluation of ideas as well as a knowledge extension. However, the most important function of scientific meetings is to facilitate social contacts. They provide an opportunity and platform to extend the network of collaborators through the creation of new contacts, and to strengthen existing links by getting reacquainted with old friends.

Undoubtedly, conference participation has a very positive impact on scientific career. In addition to the opportunities they provide, attending a scientific meeting can be very costly, both in terms of time and money. Bearing in mind that the number of national and international meetings have drastically increased in the last few decades, it is clear that scientists are now pressed to make a careful selection of the meetings they will attend. Extensive studies [15–17] have shown that conference characteristics, such as the attractiveness and the reachability of the location or the choice of keynote speakers affect the decision of scientists to attend a meeting. The role of the social component in conference choice is so far unexplored, mainly due to lack of quality data. The social component, such as the association with a conference community or conference inclusiveness, are of crucial importance when it comes to whether a conference participation was beneficial or not. This is particularly evident in the case of young scientists, who are new to a community and struggle to overcome the social obstacle of an initial contact [18, 19]. One of the rare studies on conference participation [20] has shown that conferences have a stable core of regularly attending participants, regardless of the conference location and distance. Having in mind that characteristics like the attractiveness of a location and the quality of keynote speakers are fluctuating from one year to another, it is clear that social component of a conference strongly influence the scientists decision to attend the conference and their long-term participation patterns, accordingly.

The association with a conference community and conference inclusiveness, can have a strong influence on scientists persistence in participating at the specific conference. The problem of the order-parameter persistence (first-passage time), is a well studied phenomenon in non-equilibrium statistical dynamics in condensed matter systems [21]. Persistence is defined as the probability that fluctuating variable does not change the sign until time t, and for many non-equilibrium systems this probability decays with time as a power-law [21]. Here we carry out the analysis of persistence of participation patterns of more than 100000 scientists at six national and international conferences of different sizes and from different fields of science. We study the probability of total and successive number of participations, as well as the distribution of time lags between two successive participations. We find that all three measured probabilities have a shape of a truncated power law, regardless of the conference size and degree of specialization. This indicates that the probability for a participant to attend the next meeting is not constant, but rather it grows/decays with a number of participations/non-participations. This observation is directly related to the strength of the association with the conference community. We propose a microscopic stochastic model which includes this influence of balance between the number of participations and non-participations, as well as the role of conference inclusiveness, on the probability to attend the conference next year. Results of our

model show that the studied conferences have a relatively low inclusiveness, i.e. the probability for a scientist to participate in the next meeting after the first attendance. We also show that conference attendance is characterized by *positive feedback*. The growth in the total number of participations results in a stronger attractiveness of the conference community to participants, and vice versa. Longevity of scientific career of publishing in scientific journals is also characterized by a power-law distribution with an exponential cut-off [12]. Using the empirical analysis and stochastic model Petersen et al. [12] have shown that longevity and past success of scientists lead to cumulative advantage in further development of their career. Although the distribution of career longevity and conference persistence have a similar behaviour, there is a significant difference of characteristic exponents, which indicates that a different mechanism underlie these two phenomena.

This paper is structured as follows: first, we perform empirical analysis of participation patterns for six conferences. We then propose and describe the model of conference participation dynamics. Finally, we perform numerical simulations and discuss some properties of the model, and estimate the values of parameters that correspond to empirical data.

Results

Data set

For our empirical analysis we use data for six conference series in different fields of science. We collected and filtered information about abstracts presented at the American Physical Society March Meeting (APSMM), American Physical Society April Meeting (APSAM), Society for Industrial and Applied Mathematics Annual Meetings (SIAM), Neural Information Processing Systems Conference (NIPS), International Conference on Supercomputing (ICS) and Annual International Conference on Research in Computational Molecular Biology (RECOMB). All these scientific meetings are held annually, but they differ in the topic, sizes, degree of specialisation, longevity and degree of localisation (national versus international). When it comes to the meeting size it can vary from a few dozens, like ICS and RECOMB, to several thousands of participants at APSMM. Some of these meetings are on highly focused topic, NIPS, while others are designed to cover the entire scientific fields, like APSMM, APSAM and SIAM. Four of these conferences (SIAM, NIPS, ICS and RECOMB) have an international character with venues all over the world, while APSMM and APSAM are annual conferences of American Physical Society which are always held in North American cities. APSMM, SIAM and APSAM are conferences with a long tradition, while first meetings of NIPS, ICS and RECOMB have been organized during late 80s and early 90s. Detailed information about conferences and data is given in S1 File.

To be able to track participants at the conference over the years, we have labeled them based on name, affiliation and co-authors and performed author name disambiguation (see Methods for details). We are interested in studying the participation patterns of scientists starting from their first attendance at the conference series. Thus, for conferences for which the data are not available from their beginning (APSMM, APSAM and SIAM), we have filtered out the authors that may have attended the conference before the starting year in our dataset (see Methods for the details of our filtering procedure).

Empirical results

For all scientists we have the information about the years of their appearance as authors in the book of abstracts of particular a conference series. The information about the list of authors who actually attended the conference is not available for the conferences considered in this paper. Hence, as a proxy for a conference participation in a given year, we use the appearance

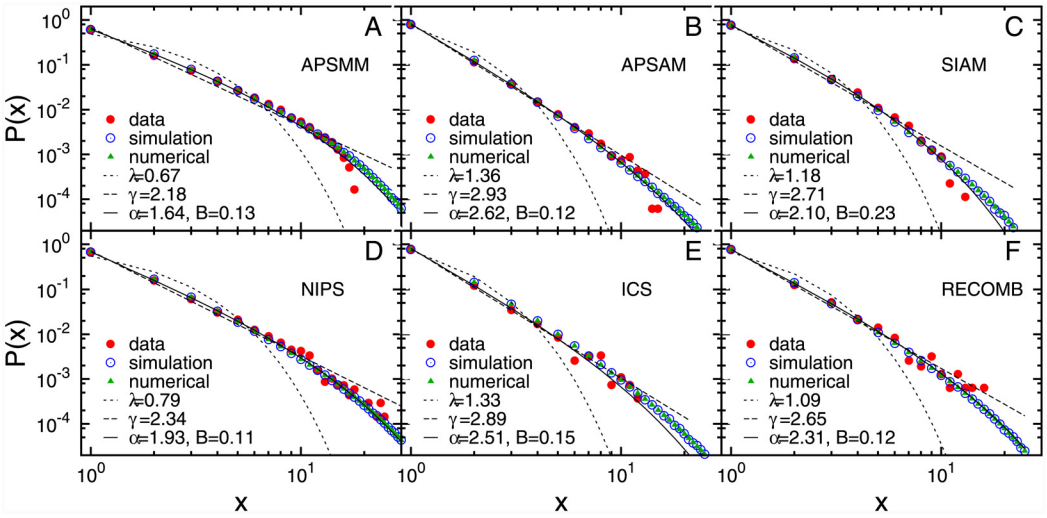

Fig 1. The total number of participations. The probability distribution of the total number of participations obtained from the empirical data (red circles), simulations (blue circles) and numerical iterative algorithm (green triangles). The full line is the best fit to truncated power law, $x^{-\alpha}e^{-Bx}$, while the dashed and dash-dot line denote the best fit to power-law distribution, $x^{-\gamma}$ and exponential distribution, $e^{-\lambda x}$, respectively.

doi:10.1371/journal.pone.0148528.g001

of a scientist as a co-author of at least one abstract in conference proceeding for that year. Not all authors that are mentioned in the book of abstracts have actually attended the conference, but one can argue that as co-authors they have actively contributed to the material presented and thus participate as a contributors in the conference [15].

First we analyse *the total number of author's participations* (the number of times an author has participated), x, at the given conference series. Fig 1, shows the probability distribution of the total number of participations, $P(x)$, averaged over all participants, for each of the six analysed conferences. The comparison of the quality of fits between exponential, power-law and truncated power-law, Fig 1, shows that all curves are very well represented by power law with exponential cut-off (see Methods), with the value of exponent $\alpha \in (1.6, 2.7)$. The disparity in the total number of participations indicates that most scientists belong to the group of occasional participants, with more than half of all participants attending a particular conference only once. For instance, the percentage of all participants that attend the conference only once is the highest for APSAM and ICS, around 81%, and the lowest for APSMM and NIPS, 63% and 68% respectively. This observation indicates that communities of all these conferences have a relatively low inclusiveness. On the other hand, it is also clear that some of the participants are very regular, attending the conference (almost) every year. These participants form the group of regular attendees whose conference participation is mainly driven by social factors, i.e. their sense of *association with the community*.

In the case of when the probability to attend a conference is constant or random, the expected distribution of total number of attendances is of exponential type. Thus, the power-law nature of the distribution of total participations strongly suggests that the probability of participation at some future conference increases with the number of previous participations. By participating frequently at a particular conference scientists not only expand, but also strengthen, their collaboration network which leads to their further engagement with the community.

We further explore the participation patterns by analysing the number of successive participations (Fig 2) and the time lag between two successive participations (Fig 3). The distributions of these quantities also exhibit the truncated power-law behaviour (see Methods). The observed

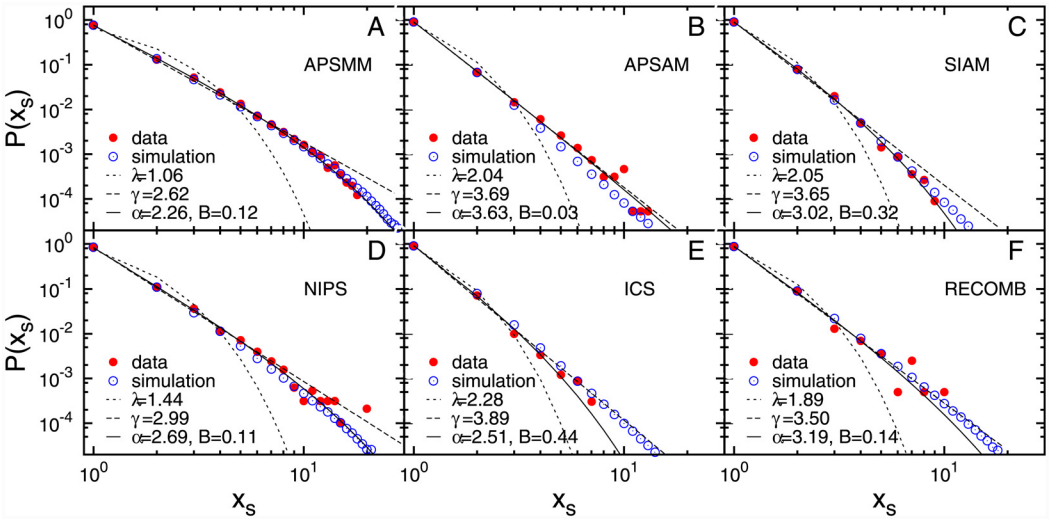

Fig 2. The number of successive participations. The probability distribution of the number of successive participations, x_s, obtained from empirical data (red circles) and numerical simulations of the model (blue circles). The full, dashed and dash-dot line are the best fit to truncated power law, power-law and exponential function respectively.

doi:10.1371/journal.pone.0148528.g002

distributions of the number of successive participations, with exponent $2 \leq \alpha \leq 4$, suggests that even frequent attendees make a pause in their participation, although these breaks are usually short, i.e. long breaks of five and more years occur with a low probability, Fig 3. A long-period of non-participation results in fading of existing collaboration ties with the community while new ones are never formed. Due to this fading, the probability to attend the meeting decreases with total number of non-participations. This indicates that conference participation of most scientists takes place in a limited period of time with a relatively short and small number of breaks.

As it was shown in Ref [12] the distribution of the journal career longevity exhibits a truncated power-law behaviour with cut-off around 10 years. The exponential cut-off in the distribution of all three measures is a consequence of the two combined finite-size effects that influence the asymptotic behaviour, the finite life time of scientist's association with one community or her/his career in one field of research or in science in general [12], and limitations of used datasets. This effect will be also observed in the distribution of conference participations. The end of a career inevitably results in a termination of participation in conferences and thus also the conference community membership. Also, used datasets have a relatively short time span (less than three decades), due to which they do not include scientists with long careers [12]. Both of these effects affect the value of the exponential cut-off, which is lower in the case of conference participation, between 4 and 9 years, compared to the one observed for the career longevity.

Model

The empirical results from six different series shown in the previous section indicate that the probability for a scientist to attend the next meeting of a conference series depends on the balance of previous participations and non-participations. Petersen et al. [12] show that Matthew (*rich get richer*) effect is responsible for the career longevity in several competitive professions, including science. They argue that it becomes easier to move forward in the career with an increasing past success of an individual, and show, using their stochastic career progressive model, that this mechanism leads to a truncated power-law distribution of the career longevity.

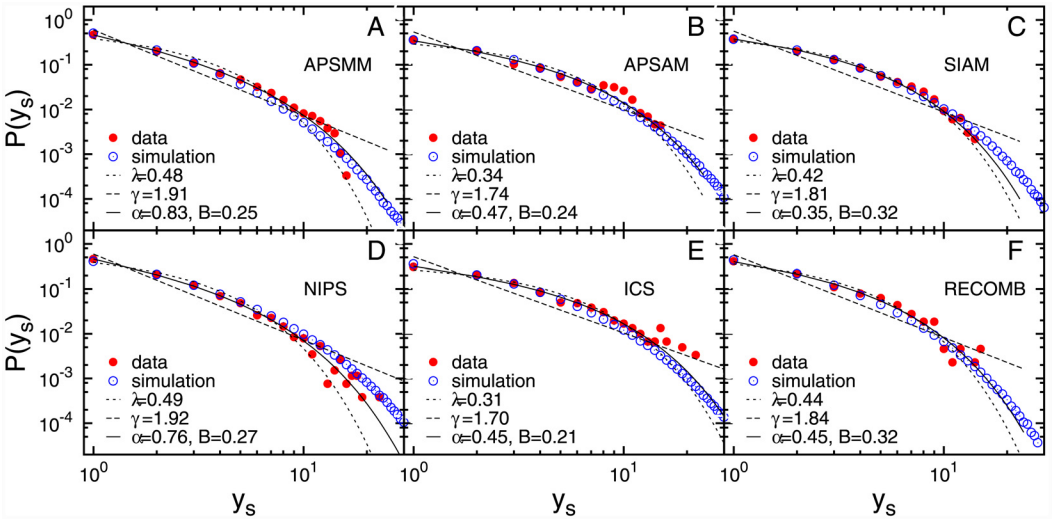

Fig 3. The time lag between the two successive participations. The probability distribution of the time lags between two consecutive conference participations y_s: empirical data (red circles) and numerical simulations data (blue circles). The lines correspond to respective fits as in Figs 1 and 2.

doi:10.1371/journal.pone.0148528.g003

In their model, they assume that the stochastic process governing career progress is similar to Poisson process, where progress is made at any given step with the rate $g(x) \equiv 1 - exp[-(x/x_c)^{\alpha}]$, where $1/x_c$ is a hazard rate corresponding to random career ending while the parameter α is the same as power-law exponent in the pdf of career longevity. Using this model for $\alpha < 1$ they were able to obtain truncated power-law distributions for career duration in several professions.

The empirical results of conference participation patterns suggest that the probability for a scientist to participate in a conference is not constant or random, but that it rather grows with the number of participations. This is reflected in the increase of proportion of authors who are going to attend the conference next year with total number of previous conference attendance (see Figure A in S1 File). Higher number of participations of a scientist at the conference results in better connections with the community and thus higher probability that the author will participate in the following conference. But unlike career longevity, where the length of the waiting times between two successive steps in the career does not influence the progress rate, the probability for conference participation is strongly influenced by the number and length of pauses (Figure B in S1 File). The longer the scientists are absent from the community the weaker are their connections and lower are the probabilities to participate in the following events. For this reason and the fact that the pdf obtained from the model proposed in Ref [12] exhibits a truncated power-law only for the exponents $\alpha < 1$ Petersen et al. model [12] cannot be applied for modelling conference participation dynamics.

We propose a new stochastic model for conference attendance dynamics which can explain our empirical findings. Our model is based on a 2-bin generalized Pólya process [22–24] and random termination time of a career. As opposed to the Petersen model where the progress rate depends only on the current position of scientist in his/her career, the 2-bin generalized Pólya incorporates dependence on the balance between participations and non-participations. Let x stands for the total number of participations at the conference, y stands for the number of conferences an author has not participated since she/he appeared at the conference for the first time and t is the number of events held, $t = x + y$. All authors start with $x = 1$ and $y = 0$. According to our model, the probability that a scientist with x total number of participations and y

number of non-participations will appear at the next conference is given by

$$g(x,y) = \frac{x^p}{x^p + (y+y_0)^p} = \frac{z^p}{1+z^p} \ , \quad (1)$$

where $z = \frac{x}{y+y_0}$ measures the balance between participations and non-participations, parameter p is the *exponent* of the model, and y_0 determines the initial balance value. The probability that a scientist will not attend the next conference is equal to $1 - g(x, y)$. Depending on the exponent p, the function g can correspond to positive ($p > 1$) or negative feedback ($p < 1$) [22]. When $p = 1$ and $y_0 = 0$, the Eq 1 is equivalent to the equation for a Pólya-Eggenberg problem [25]. As we shall see in the following section, the value of the parameter p for all conferences is larger than one, suggesting that the conference participation dynamics is characterized by the *positive feedback*: scientists who participate in the conference frequently and make less and shorter pauses have a stronger association with the conference community and thus have a higher probability to participate in the following events. The value of the parameter y_0 determines the probability of a scientist to attend the next event after her/his first occurrence at the conference. According to our model this parameter is the same for all scientists attending one conference series, thus it can be interpreted as a measure of the conference community inclusiveness.

Evolution equation. The probability $P(x, t)$ for the author to have x conference participations after t conferences since his/her first participation is equal to the probability to attend the next conference $g(x - 1, t - x)$ times the probability of already attending $x - 1$ conferences at time $t - 1$ plus the probability of skipping the next conference $1 - g(x, t - 1–x)$ times the probability of already attending x conferences at time $t - 1$:

$$P(x,t) = \frac{(x-1)^p}{(x-1)^p + (t-x+y_0)^p} P(x-1,t-1) + \frac{(t-1-x+y_0)^p}{x^p + (t-1-x+y_0)^p} P(x,t-1). \quad (2)$$

The probability distribution $P(x)$ of the number of total conference attendances for a particular conference series is obtained by summing $P(x, t = T)$ over all possible T:

$$P(x) = \sum_{T=1}^{\infty} P(x, t=T) P(T) \ , \quad (3)$$

where T denotes the duration of a scientist's membership in the community. In our case, we assume that the duration of a scientist's membership in a conference community can be terminated at any year after his/her first appearance with probability H, which gives the distribution of time intervals

$$P(T) = H(1-H)^{T-1}. \quad (4)$$

Numerical simulation results

Since the analytical solution of Eq 3 cannot be obtained, we estimate the model parameters y_0, H and p using numerical simulations (see Methods). The best estimates of the model parameters for each of the six conferences are given in Table G in S1 File. As shown in Figs 1, 2 and 3, the model with the properly chosen parameters nicely reproduces the behaviour of participants at six conferences, for all three measured quantities.

For all six conferences the estimated value of parameter p is greater than 1, which suggests that the positive feedback mechanism underlies the conference participation dynamics. This means that the probability for a scientist to attend the next year event grows superlinearly with the balance between the number of participations and pauses (z). The value of the parameter y_0

together with the value of p determines the probability for a scientist to participate in the conference next year after his/her first participation, i.e. the initial inclusiveness of the conference community. Table H in S1 File shows the estimated value of the initial inclusiveness for all six conferences. The APSMM has the highest probability, around 25%, for newcomers to attend the conference next year, while APSAM has the lowest, 9%. One could assume that the size and diversity of topics of a conference have an essential influence on conference inclusiveness, but according to our results this is not the case. The ordering of the conferences according to size, Table H in S1 File, and their initial inclusiveness do not correlate. APSAM is the second largest conference but has the lowest inclusiveness, while the RECOMB as the smallest conference is ranked as third and has the inclusiveness of 15%. Further, it follows from our results that the diversity of topics covered by the conference does not have a significant effect on the return probability of newcomers. Although the first ranked conference according to inclusiveness, APSMM, covers the widest range of topics among considered conferences, the APSAM and SIAM, which are also considered general conferences, have a lower inclusiveness than NIPS and RECOMB. This suggests that the conference inclusiveness is influenced by some other factors, which are not related to the size, degree of specialisation or localisation (national and international), but rather to social structure and openness of the conference community toward newcomers.

We solve Eq 3 numerically using an iterative method (see SI for more details) and compare it with simulation results. Fig 1 shows an excellent matching between results obtained using the iterative algorithm and numerical simulations for the estimated values of parameters.

Discussion and conclusion

The goal of this paper has been to investigate the conference participation patterns and propose a simple stochastic model of conference participation dynamics. The motivation behind this is to better understand the mechanisms that underlie the repeated participation in the same conference series and explore whether the conference series topic, size, degree of specialisation, longevity and degree of localisation (national and international) influence the participation probability and inclusiveness of the specific community. Our study is based on empirical analysis and modelling of authors participation at six different conference series in the last three decades: APSMM, APSAM, SIAM, NISP, ICS and RECOMB. We note here that it would be important to verify our findings with the data from other conferences.

The set of considered conferences is very heterogeneous. Although they differ in size, topic and topic diversity, national structure of participants and conference longevity, they are characterized with similar participation patterns. The distributions of the total number of participations for all six conferences exhibit the same, truncated power-law, behaviour with values of exponent α between 1.6 and 2.7. A similar behaviour is also observed for the distributions of the number of successive participations and the duration of pauses between them. The observed statistical evidence strongly imply that the dynamics of conference participation is governed by universal forces which are independent of the specific conference features or the scientific field. This and the fact that conferences often have a stable core of attending participants [20] suggests that these have social origins and that social factors, such as the association with a conference community and its inclusiveness, strongly influence the probability for a scientist to attend the future meetings and their participation patterns at the specific conference series, accordingly.

The observed truncated power-law behaviour of the distributions of participations indicates that the probability for a scientist to participate in the next year conference is growing(decreasing) with the balance between the number of participations and pauses. To further explore this

we proposed a stochastic model based on 2-bin generalized Pólya process which incorporates the dependence on the ratio between number of participations an pauses. Our model shows that the positive feedback mechanism underlies the conference participation dynamics. The probability for a scientist to attend a conference grows superlineary with the number of participations, while the frequent pauses have the opposite effect. The scientists who are able to overcome the initial obstacles and create social ties with the conference community by frequent participation at the beginning have a higher probability to attend the conference in the following years. A frequent participation strengthens the scientist's association with a conference community which further increases the probability for future participations. On the other hand, scientists with a small number of initial participations have a low probability to participate in the following conference, thus small number of participations, and eventually stop attending the conference. The initial inclusiveness of the specific conference community has the main influence on early participation patterns. As we showed, this inclusiveness does not depend on the size, degree of specialisation or topic of the conference, but rather on the openness of the community toward newcomers.

Our analysis indicates that social factors, such as the association with the community and the community inclusiveness are the main driving forces of conference participation dynamics. In general the community/group cohesion and the ability to attract and retain newcomers and other members influence the dynamics of their participation in group activities [26]. On the other hand, a member's engagement in group activities strengthens ties to other group/community members, and contributes to the creation of the bonding capital, while the ties of non-attendees dissolve and weaken with time [27]. Conference communities are just one example of these systems, thus we expect to observe the similar group participation patterns in other types of social communities, both online and offline. Further investigations and studies of other social systems will reveal and characterize the connection between a social network structure and group inclusiveness, and participation dynamics in group activities.

Methods

Data filtering Identification of the different authors may involve a few issues. On one hand, an author may use different spelling variants to sign his first and middle name. On the other hand, the author's name may be related to several different authors, thus using only the initials of the last name and first name increases additionally error rates in disambiguating the author names. In our data sets, data from NIPS and RECOMB conferences did not require additional cleaning, while for the SIAM and ICS data, we have used python fuzzy partial string matching of author's first and middle names, which gave a high accuracy. For APSMM and APSAM conferences, where data are highly heterogeneous, we have used a method described in [28] to disambiguate the author names. This method considers pairs of names that match on last name and first name initials. Then it groups the authors based on their affiliation and co-authors. Because the same affiliation could be formatted differently, the two affiliations were considered the same if their fuzzy token set ratio was higher than 50%.

The sources and detailed description of the data are given in Tables A, B and C in S1 File. For NIPS, ICS and RECOMB, we have complete data from their very beginning. Remaining data sets required filtering out the authors with a high probability of attending conference before the starting year in our dataset, Y_0. Therefore, for APSMM, APSAM and SIAM we have isolated authors with the first recorded year of conference attendance, smaller than $Y_0+\langle\tau\rangle$, where $\langle\tau\rangle$ is the average waiting time between a consecutive conference attendance for all the authors who took part at the conference during the $[Y_0, Y_f]$ period. This way we excluded between 10% (APSMM and SIAM) and 25% (APSAM) authors from our analysis.

Functional fits We have used the maximum-likelihood fitting method [29] to fit three different functions to the probability distributions of the total number of participations, the number of and the time lags between two successive participations: exponential function $e^{-\lambda x}$, power-law function $x^{-\gamma}$ and truncated power-law $x^{-\alpha}\, e^{-Bx}$. It follows from the comparison of fits of these three functions to empirical data that the truncated power-law is the best fit for the probability distribution of all three measured quantities, see Figs 1, 2 and 3. In order to compare these three fits we calculate the log likelihood ratio, $\mathcal{R}$, and π-value (see Ref [29]) which compares the fits to the power-law with exponential cut-off with the pure power-law for the distribution of total number of participations (Table D in S1 File) and the number of successive participations (Table E in S1 File). In the case of nested distributions, the negative value of $\mathcal{R}$ indicates that the larger family of distributions, in this case the truncated power-law, is a superior model. When the value of $\mathcal{R}$ tends to 0, one can use π-value. The small π-value suggests that the smaller family of distributions, in this case power-law, can be ruled-out. Both the log likelihood ratio and the π-value indicate that the truncated power-law is a superior model compared to pure power-law for both distributions. A similar procedure can be applied for the comparison between truncated power-law and exponential fits, but since from the visual inspection it is clear that the distributions do not follow the exponential fits, we have omitted these results. The comparison between exponential and the power-law with exponential cut-off fit, given in Table F in S1 File, indicates that the power-law distribution with exponential cut-off fit is better than exponential fit for the distribution of the time lags. For all six conferences, the power-law with exponential cut-off distribution gives the best fit for all three empirical distributions.

Parameter estimation We simulate the model for $N = 100000$ different authors. Starting from $x = 1$ and $y = 0$ at $t = 1$, an author will appear at the next conference with probability $g(x, y)$ or skip it with the probability $1 - g(x, y)$. The author can terminate his/her membership in the community at each time step with the probability H. In order to estimate the values of parameters p, y_0 and H, we calculate the distribution of total number of attendances x, from the simulations and compare it to the empirical distribution using Kullback-Leibler Distance [30]. We perform the simulations for several different sets of parameter (y_0, H, p) to determine which combination of parameter values makes the model optimally close to the empirical data. For each parameter set the results are averaged across 100 simulations.

Supporting Information

S1 File. Supplementary Information: A theoretical model for the associative nature of conference participation. Proportion of conference participants g with x conference attendances who are going to attend the conference next year (**Figure A**). Proportion of conference participants ρ with n missed conferences after x-th conference attendance who are going to skip the conference next year, but will take part at some future conference from the observation period (**Figure B**). Pages on the web from which we downloaded conference data (**Table A**). Summary of the conference data. Columns 2 and 3 indicate for each conference the year in which data we have collected begin (Y_0) and end (Y_f). The total number of different participants at the conference during that period of time is given in column 4 (**Table B**). The number of participants at the conference per year (**Table C**). Log likelihood ratio $\mathcal{R}$ and the π-value compare the fit to the power-law with the fit to the power-law with an exponential cutoff for the probability distribution of number of conferences at which each author appears (**Table D**). Log likelihood ratio $\mathcal{R}$ and the π-value compare the fit to the power-law with the fit to the power-law with an exponential cutoff for the probability distribution of the number of successive participations at the conference (**Table E**). Log likelihood ratio $\mathcal{R}$ and the π-value compare the fit to the

exponential with the fit to the power-law with an exponential cutoff for the probability distribution of the time lag between two consecutive conference participations (**Table F**). The optimal parameter values of the model for each conference (**Table G**). Stagnancy rate $1 - g(1, 0)$ at $t = 1$ for each conference and exponent α of power-law with an exponential cutoff distribution fit with the corresponding conference order (**Table H**).
(PDF)

Acknowledgments

Numerical simulations were run on the PARADOX supercomputing facility at the Scientific Computing Laboratory of the Institute of Physics Belgrade.

Author Contributions

Conceived and designed the experiments: JS AC TK MMD. Analyzed the data: JS. Wrote the paper: JS AC TK MMD. Collected the empirical data: JS.

References

1. Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, et al. Life in the network: the coming age of computational social science. Science. 2009; 323(5915):721–723. PMID: 19197046
2. Cioffi-Revilla C. Computational social science. CompStat. 2010; 2(3):259–271.
3. Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. Rev Mod Phys. 2009; 81:591–646. doi: 10.1103/RevModPhys.81.591
4. Sen P, Chakrabarti BK. Sociophysics: An Introduction. Oxford University Press, Oxford; 2014.
5. Schweitzer F. Scientific networks and success in science. EPJ Data Science. 2014; 3:35. doi: 10.1140/epjds/s13688-014-0035-8
6. Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: Toward an objective measure of scientific impact. Proc Natl Acad Sci USA. 2008; 105(45):17268–17272. doi: 10.1073/pnas.0806977105 PMID: 18978030
7. Radicchi F, Fortunato S, Markines B, Vespignani A. Diffusion of scientific credits and the ranking of scientists. Phys Rev E. 2009; 80(5):056103. doi: 10.1103/PhysRevE.80.056103
8. Wang D, Song C, Barabási AL. Quantifying long-term scientific impact. Science. 2013; 342(6154):127–132. doi: 10.1126/science.1237825 PMID: 24092745
9. Sinatra R, Wang D, Deville P, Song C, Barabasi AL. Scientific impact: the story of your big hit. In: APS Meeting Abstracts; 2014. p. 17004.
10. Petersen AM, Riccaboni M, Stanley HE, Pammolli F. Persistence and uncertainty in the academic career. Proc Natl Acad Sci USA. 2012; 109(14):5213–5218. doi: 10.1073/pnas.1121429109 PMID: 22431620
11. Azoulay P, Graff Zivin JS, Wang J. Superstar Extinction. QJ of Econ. 2010; 125(2):549–589. doi: 10.1162/qjec.2010.125.2.549
12. Petersen AM, Jung WS, Yang JS, Stanley HE. Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. Proc Natl Acad Sci USA. 2011; 108(1):18–23. doi: 10.1073/pnas.1016733108 PMID: 21173276
13. Petersen A, Penner O. Inequality and cumulative advantage in science careers: a case study of high-impact journals. EPJ Data Science. 2014; 3(1):24. doi: 10.1140/epjds/s13688-014-0024-y
14. Deville P, Wang D, Sinatra R, Song C, Blondel VD, Barabási AL. Career on the Move: Geography, Stratification, and Scientific Impact. Sci Rep. 2014; 4:4770. doi: 10.1038/srep04770 PMID: 24759743
15. Borghans L, Romans M, Sauermann J. What makes a good conference? Analysing the preferences of labour economists. Labour Economics. 2010; 17(5):868–874.
16. Mair J, Thompson K. The UK association conference attendance decision-making process. Tourism Management. 2009; 30(3):400–409.
17. Witt SF, Sykes AM, Dartus M. Forecasting international conference attendance. Tourism Management. 1995; 16(8):559–570.
18. Garfield E. Making contacts at conferences. The Scientist. 1988; 2:294–295.

19. Garfield E. Making contacts at conferences—a problem for the young scientist. Essays of an Information Scientist. 1977; 3:668–672.
20. Van Dijk J, Maier G. ERSA Conference participation: does location matter? Papers in Regional Science. 2006; 85(4):483–504. doi: 10.1111/j.1435-5957.2006.00102.x
21. Majumdar SN. Persistence in nonequilibrium systems. Current Science. 1999; 77(3):370–375.
22. Drinea E, Frieze A, Mitzenmacher M. Balls and bins models with feedback. In: Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms. SODA 02; 2002. p. 308–315.
23. Pemantle R. A survey of random processes with reinforcement. Probab Surveys. 2007; 4:1–79. doi: 10.1214/07-PS094
24. Oliveira RI. The onset of dominance in balls-in-bins processes with feedback. Random Structures & Algorithms. 2009; 34(4):454–477. doi: 10.1002/rsa.20261
25. Johnson NL, Kotz S. Urn Models and Their Application: An Approach to Modern Discrete Probability Theory. Wiley; 1977.
26. Friedkin NE. Social cohesion. Annu Rev Sociol. 2004; 30:409–425. doi: 10.1146/annurev.soc.30.012703.110625
27. Sessions LF. How offline gatherings affect online communities. Information, Communication & Society. 2010; 13(3):375–395. doi: 10.1080/13691180903468954
28. Wu J, Ding XH. Author name disambiguation in scientific collaboration and mobility cases. Scientometrics. 2013; 96(3):683–697. doi: 10.1007/s11192-013-0978-8
29. Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. SIAM Review. 2009; 51(4):661–703. doi: 10.1137/070710111
30. Burnham KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Science & Business Media; 2002.