



OPEN

Analysis of dynamic networks based on the Ising model for the case of study of co-authorship of scientific articles

V. Andrea Hurtado-Marín¹, J. Dario Agudelo-Giraldo^{2✉}, Sebastian Robledo^{3,4} & Elisabeth Restrepo-Parra¹

Two computational methods based on the Ising model were implemented for studying temporal dynamic in co-authorship networks: an interpretative for real networks and another for simulation via Monte Carlo. The objective of simulation networks is to evaluate if the Ising model describes in similar way the dynamic of the network and of the magnetic system, so that it can be found a generalized explanation to the behaviours observed in real networks. The scientific papers used for building the real networks were acquired from WoS core collection. The variables for each record took into account bibliographic references. The search equation for each network considered specific topics trying to obtain an advanced temporal evolution in terms of the addition of new nodes; that means 3 steps, a time to reach the interest of the scientific community, a gradual increase until reaching a peak and finally, a decreasing trend by losing of novelty. It is possible to conclude that both methods are consistent with each other, showing that the Ising model can predict behaviours such as the number and size of communities (or domains) according to the temporal distribution of new nodes.

The modeling of social phenomena with physical bases has allowed an affirmative answer to whether there are or not laws that govern them¹. The concept has gained strength in recent years thanks to two crucial facts: on the one hand, the access to large databases, turning society into a huge sample for research²; and on the other, to advances in the study and characterization of complex networks³. In principle, it has been assumed that social groups are systems composed of interacting agents capable of taking a stance in the face of an external disturbance, in a similar way to a many-body system in statistical physics⁴⁻⁶. Different types of algorithms have been proposed to detect communities in the framework of physics, Potts model in⁷ and label propagation algorithm (LPA)^{8,9} are some examples. In particular, algorithms under Ising model also has been widely implemented^{6,10,11}.

One of the most significant current discussions in social networks is understanding their dynamics through the time. For instance in¹², it has been shown that the addition of the social dynamics decreases the network influence on the decision of their members, and in¹³ is highlighted the importance of considering the collaboration dynamic among academics to understand the evolution of a research topic. Therefore, the metrics of temporal evolution are the crucial mechanisms behind it. Trying to solve this problem, in this work is proposed the Ising model for detecting evolution patterns in social communities.

Under the precept that by their nature, there are different types of networks, and that they diverge in a large number of behaviors¹⁴, this research proposes to study co-authorship of scientific articles around specific topics or areas. A co-authorship network is made up of scientists with close social contacts whose interactions represent collaborations or cooperation¹⁵⁻¹⁷. Although in recent studies, these contacts have evidenced the formation of the scientific communities, most of the works developed lack dynamic analysis^{6,15,18-21}, or explicitly despite the variability of weight in each link^{1,22-24}. For this reason, one of the objectives of our study has been to look for predictable patterns or behaviors that can arise from these dynamic interactions. Specifically, the formation of

¹PCM Computational Applications, Universidad Nacional de Colombia, Manizales, Colombia. ²Departamento de Física y Matemáticas, Universidad Autónoma Manizales, Manizales, Colombia. ³ECOSOL Research Group, Universidad Católica Luis Amigó, Manizales, Colombia. ⁴Centro de Bioinformática y Biología Computacional de Colombia, BIOS, Ecoparque Los Yarumos, Manizales, Colombia. ✉email: josed.agudelog@autonoma.edu.co

scientific communities is analyzed taking into account the appearance and the redistribution of new nodes and the weight of the links.

The studies required a comparative analysis between the physical description of real networks and spins simulations. The objective of this work was to observe if the Ising model describes the temporal dynamic of real networks and from this give a generalized explanation of the behaviors observed by mean of simulation. The real networks procedure consisted of applied an improved version of the algorithm presented by Son et al. for flow networks⁶. According to a recent classification presented by El-Moussaqui et al. for identification of communities in complex networks²⁵, this method has a divisive and hierarchical approach, reason by which is denominated here as DHM. For the magnetic simulations, the Ising model with mean field discrimination was implemented using the Monte Carlo method (MCM). The correspondence between the two methods is analyzed by observables such as internal energy and the number and size of communities (or domains).

Some other works have taken advantage of different models and computational tools for improving the analysis of collaborative networks. For instance, in²⁶ is used a bipartite graph approach blended to joint in the structure as researchers attitudes and interests as the community's recognition. In²⁷ for examining disambiguating names in references based upon the topological and the hierarchical descriptions, in²⁸ an analysis of patterns that relate ranking, first and last authors, specific roles, and others characteristics, indicate that must take into account the distinct types of scientific contributions, and in²⁹ coauthors characteristics are obtained applying learning models with training.

Methodology

As mentioned above, two developments were necessary to add to the analysis process: an experimental one, obtained from real co-authorship networks with the authors as nodes, and the other by simulation under equivalent interaction situations, but between spin moments. The procedure for each of them is described below:

A. Ising model applied to the interpretation of co-authorship networks via DHM. In the construction of the network, each node represents an author and each link a scientific collaboration. The scientific papers were obtained from the Core Collection of Web of Science (WoS), one of the most relevant bibliographic databases widely used in academia^{30,31}. Core collection consists of ten sub-datasets (eight citation indexes and two chemical indexes), with information from over 18,000 high impact journals, over 180,000 conference proceedings, and over 80,000 books³². Four co-authorship networks were built with the records of specific queries by topic and title, and limiting the document just to scientific articles. The research topics and title were: econophysics, magnetoresistance—nanoparticles, Monte Carlo—thin films, and Monte Carlo—core shell nanoparticles.

These four networks were selected from a large number of searches, three of them based upon our expertise in condensed matter physics. The condition of selection was an advanced temporal evolution in terms of the addition of new nodes. That means the presence of 3 steps: a time to reach the interest of the scientific community, a gradual increase until reaching a peak, and finally, a decreasing trend by losing of novelty. Temporal evolutions can be fitted with an extreme value distribution presented by Eq. (1) (as is present in the “Results and discussion” section). In this distribution, N_o is a number that varies according to the total number of nodes (or spins), t_p is the time for which the function has a maximum and β is a scale parameter. It should be clarified at this point that dynamic networks can present different types of distribution for the entry of new nodes. For this reason, it was ensured that both the analysis of the real networks and the Monte Carlo simulation were characterized by the shape of distribution of new nodes.

$$N(t_s) = N_o * \frac{1}{\beta} \exp\left(\frac{t_s - t_p}{\beta}\right) \exp\left(-\exp\left(\frac{t_s - t_p}{\beta}\right)\right). \quad (1)$$

The variables for each record were: ID (assigned by WoS), DOI (digital object identifier), authors, year of publication and bibliographic references. For the articles present within the bibliographic references with DOI, Crossref was used to obtain the information of their authors and their year of publication. A minimum standard was set to label authors with the initials of the first name and the first surname (all written in capital letters), thus reducing the typographic differences between journals. The software for the construction and identification of communities of the co-authorship networks was compiled in R.

The networks built are dynamic (changing over time). For this reason, the year variable was transformed into time steps (t_s). For each of the steps, a static network was built, taking into account the nodes and links that emerged in earlier times. Additionally, in the present work, the weight of the links was weighted based on the number of documents written by pairs. To identify the scientific communities, the procedure proposed in⁶ was used, which was applied to each of the connected components of the static network. Initially, the two most influential nodes within the network that generate a community structure are identified, denoted as s and t . Node s is the highest degree and node t is the next highest degree node with which this structure is generated. The algorithm to identify the communities is based on the existing analogy between a flow network and a ferromagnetic random field Ising model³³ whose Hamiltonian is given by:

$$H = -\frac{1}{2} \sum_{i,j=1}^n J_{ij} \sigma_i \sigma_j - \sum_{i=1}^n B_i \sigma_i, \quad (2)$$

where each node i represents a spin σ_i in the magnetic system, J_{ij} is the exchange constant between the spins and B_i is the field on the spin σ_i . The exchange constant J_{ij} between two spins σ_i and σ_j is equal to the weight of the link. The weight is obtained by the number of interactions between two authors; that is to say, the number of

scientific documents that were published together. For this reason, only ferromagnetic couplings ($J_{ij} \geq 0$) were considered. The steps for each component are: (i) build a flow network³³, (ii) calculating the maximum flow³⁴ and (iii) identify the community structure. A flow network contains the same nodes and links as the original network plus two additional nodes called source and sink denoted as s and t . In the algorithm proposed by Son et al. in⁶, all the possibilities of pair of nodes s and t were tested. But the authors propose that these nodes could be identified from the nodes of higher degree, to reduce the computational cost.

The magnetic field distribution used is given by Eq. (3), for two nodes s and t ; this is equivalent to imposing the boundary condition $\sigma_s = +1$ and $\sigma_t = -1$, which induces frustration in the system. In the flow network the weights in the links are given by $c_{ij} = 2J_{ij}$ ⁶ (also denominated as capacity). The infinite value of the local field applied to the spins σ_s and σ_t , within the computational simulations, was calculated as $B_s = \sum_{i,j=1}^n c_{ij}$ and

$$B_t = - \sum_{i,j=1}^n c_{ij}, \text{ respectively.}$$

$$B_i = \begin{cases} +\infty, & \text{para } i = s \\ -\infty, & \text{para } i = t \\ 0, & \text{para } i \neq s, t \end{cases} \quad (3)$$

To identify the communities, the criterion established by the maximum flow—minimum cut theorem is used. The algorithm consists of subdividing the sample continuously until the division does not have a community structure. Each iterative subdivision always leads to identifying two sets C_s and C_t , with the two most influential nodes. To generate a subdivision, it must be satisfied that $\frac{\ln(D_{st})}{\ln(n)} > 1$, where D_{st} is the product between the number of nodes that belong to sets C_s and C_t , and n is the number of internal nodes by subdivision. At the end iterative subdivisions, when $\frac{\ln(D_{st})}{\ln(n)} < 1$, the set of n nodes correspond to a community. These point in the same direction of s . More information on this procedure can be found at^{6,33}. The communities should have more than 2 nodes. Nodes that do not belong to a community have a spin randomly assigned. The local field value for each community is assigned according to B_s and B_t .

B. Ising model simulated via MCM. The computational program for the magnetic modeling via Monte Carlo was thought from preliminary results obtained for the co-authorship networks. The spins were randomly located within a box of length $L = 16$ muc, where muc are magnetic cell units (relative to positions in a cubic crystalline system). The minimum distance between moments was set at 1 muc. The addition of new moments to the system obeyed a location probability given by $P = \exp(0.5(1 - r_{ij}))$, whereby comparing it with a random number, the positions closest to that of the last spin added were privileged.

Additionally, each simulation was divided into time steps to introduce new spins to the system, according to the distribution presented in Eq. (1), with $N_0 = 2458$ (maximum number of spins) and $\beta = 6$. This is in accordance with the fact that new authors (nodes) appear in the co-authorship networks over time (dynamic network). At each time, the system was simulated under a canonical assembly (NVT) at 60 K. In the preliminary results, the temperature did not show a significant influence. The time variable ranged from 1 to 40 steps. The final-time density was established at 0.6 spins per cell.

The Metropolis algorithm was implemented to generate the state fluctuations. Periodic boundary conditions were not taken into account. The number of Monte Carlo steps was set at 100,000 with a cutoff for calculating observables at 50,000 per time step. FORTRAN95 was the software used to compile the code.

The Hamiltonian used in the realization of this model is defined by Eq. (4), where J_{ij} was established at ± 10 meV. A value of $r_{cut} = 1.5$ muc was selected because, in physics publications, the number of authors in each article is approximately 4 on average. Therefore, choosing a larger cut-off radius would involve other types of networks in which the interactions are more numerous. Interactions with neighbors that are less than 1.25 muc apart were considered ferromagnetic ($J_{ij} > 0$), while at further distances, the interactions were antiferromagnetic ($J_{ij} < 0$). The mean field per site ($B_{mean,i}$) was used to identify domain boundaries. In particular, if $-1 < B_{mean,i} < 1$, σ_i is considered a frustrated spin, without domain. This phenomenon is also observed in co-authorship networks in which there are authors who do not belong to a defined scientific community.

$$H = - \sum_{\langle i,j \rangle} J_{ij} \sigma_i \sigma_j. \quad (4)$$

Results and discussion

Table 1 presents information on the search equation for each network. The first four columns are information on the date of the query, the real time period, the number of articles found and the number of references. The next three columns correspond to information extracted from each network: the transformation of the period into the number of time steps, the total number of nodes (N_n) and the total number of links (N_l).

Figure 1 shows the temporal evolution of the number of new nodes and links for each network. As can be inferred, a new research topic takes time to reach the interest of the scientific community; later, there is an exponential increase in the number of new authors until reaching a peak time (t_p) and, finally, the topic loses its appeal or novelty, showing a decreasing trend. Each of the curves is shown an adjustment to the distribution established by Eq. (1). This figure shows that networks first reach a maximum in new nodes than in new links. These results were also observed in the simulated behavior. Figure 2 presents the temporal distribution of new spins and new interactions for $t_p = 27$. As mentioned previously, the shape of the distribution is fixed with the same one with which the co-authorship networks were adjusted (Eq. (1)). The fact that the links come from a collective behavior causes any transition phase in them or in their observable derivatives to present this delay.

Search equation	Date of query	Real time period	Number of papers	Number of references	Time steps	N_n	N_l
TITLE: (econophysics)	29/02/2020	1930–2019	102	1751	66	2282	5242
TITLE: (magnetoresistance) AND TITLE: (nanoparticles)	29/02/2020	1953–2019	68	1463	54	4514	20,955
TITLE: (monte carlo) AND TITLE: (thin films)	29/02/2020	1930–2019	211	3840	73	8221	31,726
TOPIC: (monte carlo) AND TITLE: (core shell nanoparticles)	11/03/2020	1950–2019	58	1351	57	3528	15,534

Table 1. Search information for each network. WoS source.

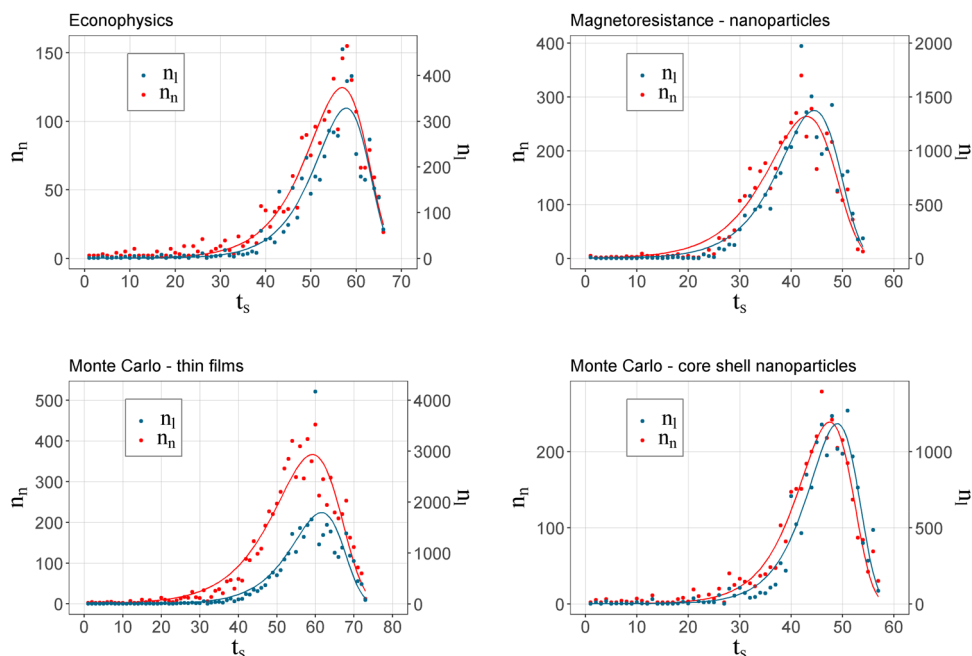


Figure 1. Number of new nodes and new links as a function of time for the different co-authorship networks.

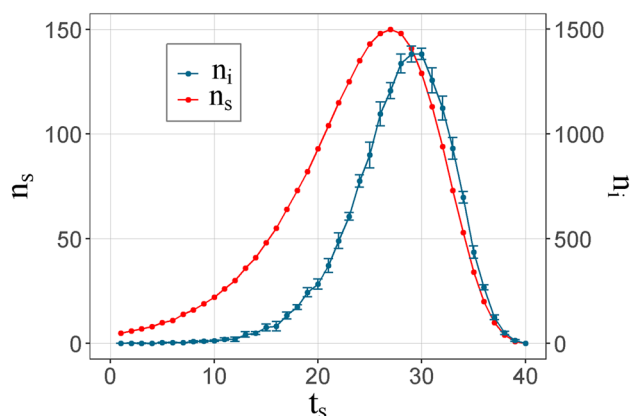


Figure 2. Temporal distribution of the insertion of new spins and the appearance of interactions for simulations with $t_p = 27$ in Monte Carlo method.

In Fig. 3, the energy (U) is presented as a function of time for the different networks. The values are normalized to the minimum value. U decreases with time due to the entry of new nodes whose interactions mostly contribute with negative interactions (Eq. (2)). The inflection point was obtained by fitting the exponential curve of Eq. (5), but decreasing. The fit is represented by a blue line on the graph. It is expected by the correspondence between energy and links, that these curves present a time at the inflection point, called t_c , close to the t_p of the links. The values for t_c , along with the peak times of nodes and links, are presented in Table 2. In the language of physics, this behavior is associated with the percolation threshold of the system³⁵. Below the percolation threshold,

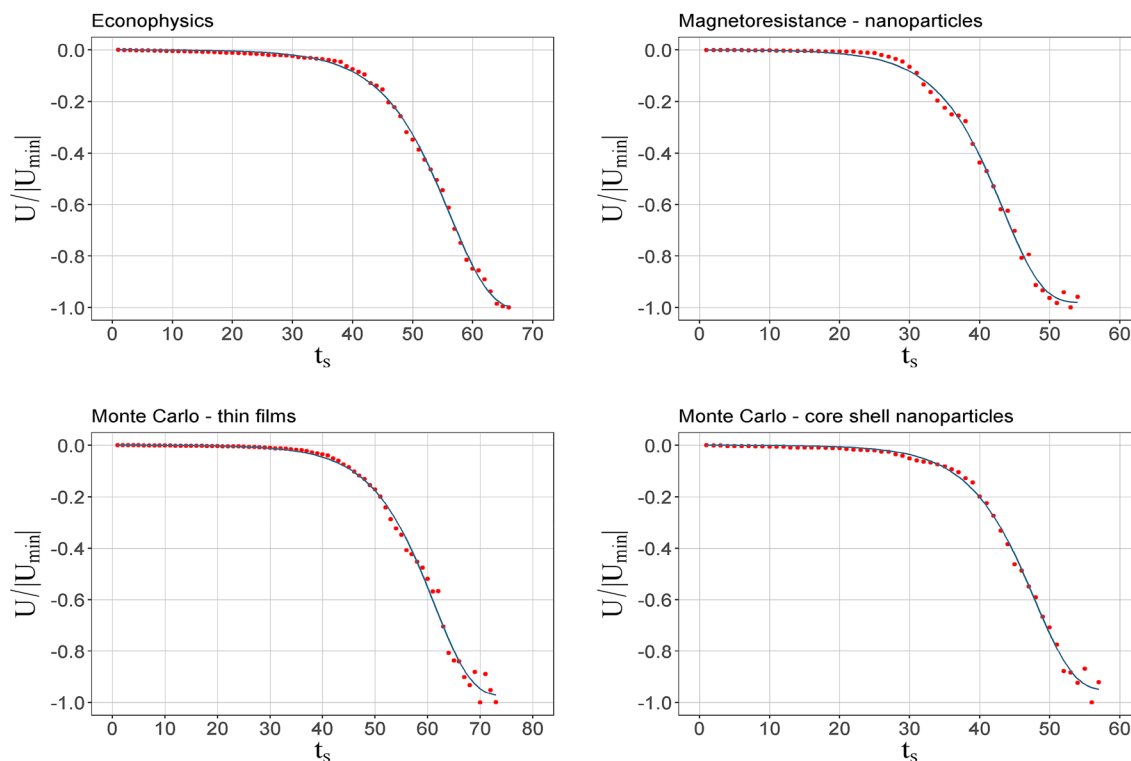


Figure 3. Normalized energy as a function of time for the co-authorship networks. The blue line is an adjustment to get the inflection point.

Search equation	Nodes t_p	links t_p	Energy t_c	Number of communities, t_c
TITLE: (econophysics)	56.9 ± 0.2	57.8 ± 0.3	56.2 ± 0.1	55.7 ± 0.2
TITLE: (magnetoresistance) AND TITLE: (nanoparticles)	43 ± 0.3	44.5 ± 0.3	43.4 ± 0.2	40.9 ± 0.2
TITLE: (monte carlo) AND TITLE: (thin films)	59.2 ± 0.3	61.6 ± 0.5	61.3 ± 0.2	55.3 ± 0.2
TOPIC: (monte carlo) AND TITLE: (core shell nanoparticles)	47.5 ± 0.2	49 ± 0.2	47.9 ± 0.2	45.2 ± 0.2

Table 2. Summary of characteristic times in co-authorship networks. The difference between the times t_c of energy and t_c of the number of communities show that they are not the product of the same state transition.

the probability of finding a connected path between two separated spins in the system is very low; because of this reason, there cannot be any long-range magnetic ordering. Above the percolation threshold, the probability of finding connected paths increase, so exchange interaction broadcast further. In the language of networks, this inflection point means a transition from a state in which the system presents a fluctuating topology in terms of the distribution of nodes in the communities to a state with well-defined communities.

$$y = A * \left(1 - \exp \left(-\exp \left(\frac{x - C}{B} \right) \right) \right). \quad (5)$$

To complement, Fig. 4 presents the energy as a function of time for different t_p in the spin distribution. The values are normalized to the minimum value. As in co-authorship networks, U decreases over time. This means that the number of interactions is mainly ferromagnetic, since there are a greater number of spins (nodes) in the domains (communities), than in the borders where antiferromagnetic interactions occur. t_c was found by fitting the curves to Eq. (6) and setting the second derivative equal to zero.

$$U = \frac{-\arctan(At_s + B) + C}{D}. \quad (6)$$

Since increasing t_p decreases the density of nodes and links in the first-time steps, it can be specified that the change of state given by t_c requires a minimum density. In the inset, in Fig. 4, it can be seen that a linear relationship can represent these two variables.

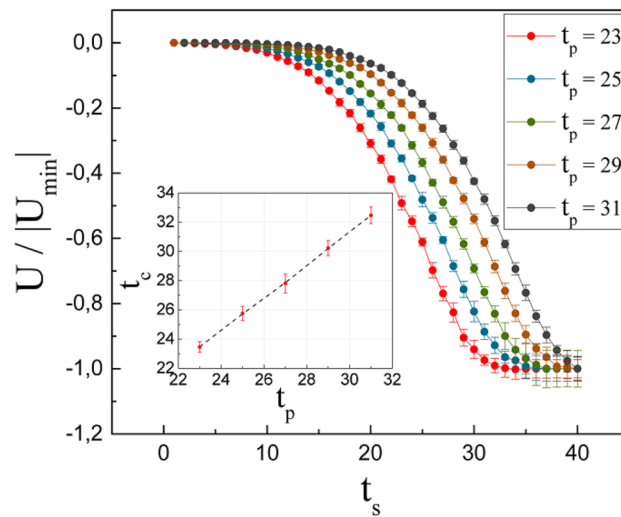


Figure 4. Normalized energy as a function of time for different values of t_p . Inset: t_c as a function of t_p .

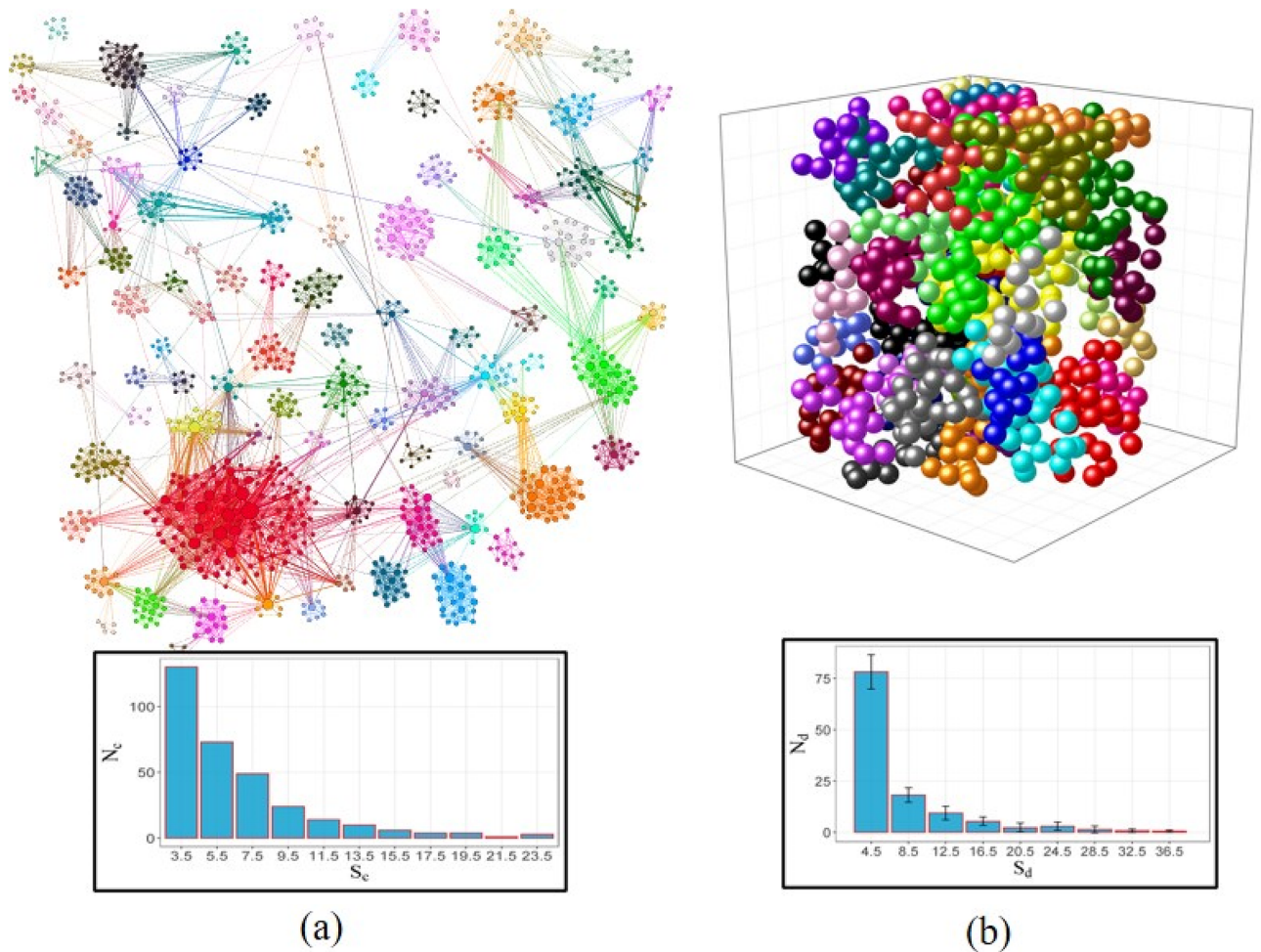


Figure 5. Images of (a) giant component of the core shell nanoparticles network, showing the communities that have more than 4 nodes and (b) magnetic domains that have more than 10 spins, for a distribution of simulated atoms with $t_p = 27$. The histograms inserted in (a) and (b) represent the size distribution of the communities and domains, respectively.

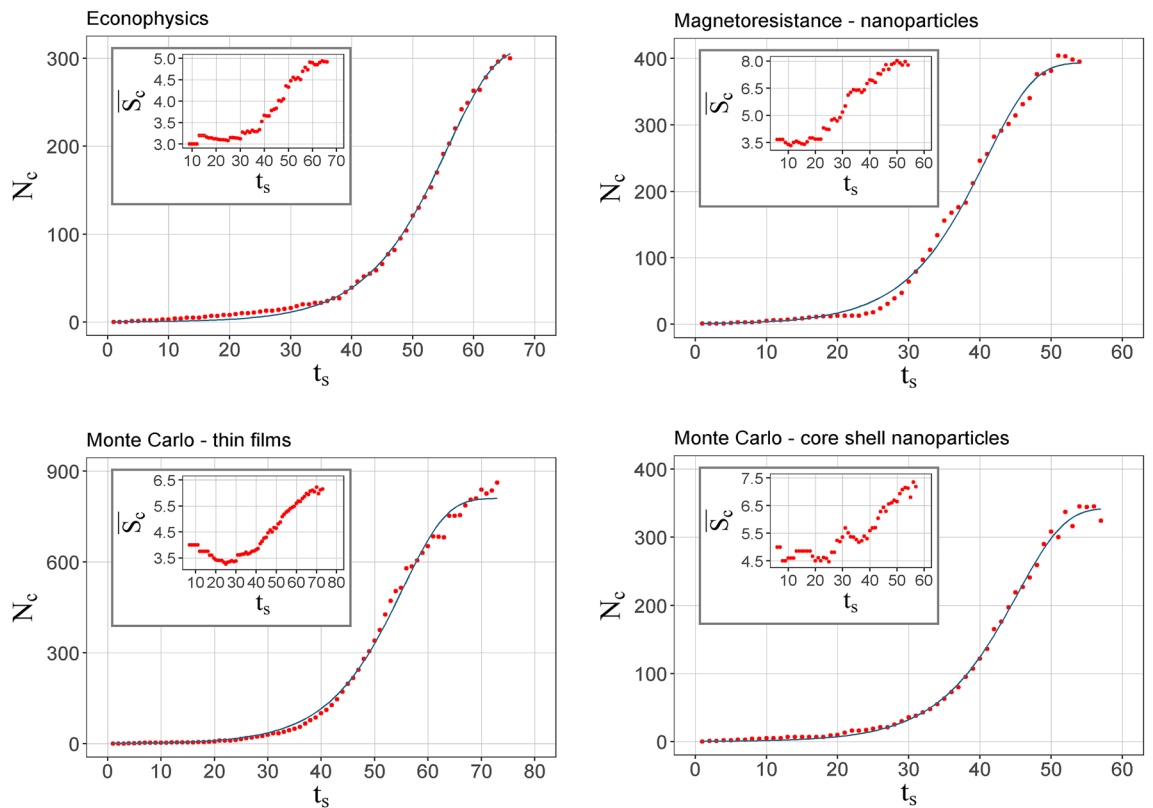


Figure 6. Number of communities as a function of time for the co-authorship networks. The inserted figures correspond to the temporal evolution of the average size of the community, respectively.

Figure 5 presents descriptive images of the final samples and their communities (or domains), particularly in a) for the giant component of the core shell nanoparticles network with communities of more than 4 nodes, and b) for the distribution of simulated atoms for $t_p=27$, there the domains that have more than 10 spins are shown. The internal histograms show the distribution of sizes, S_c and S_d for communities and domains, respectively.

It is remarkable to say that, at difference of other methods where a community is assigned to all nodes, even if it is not coherent, spin models are characterized because allows to determine the nodes that do not belong to a community. In the simulations, the boundaries of magnetic domains are formed from frustrated spins, which can be identified if mean field per site is low.

Figure 6 shows the number of communities (N_c) as a function of time for the different networks. In the internal insets of each figure, the temporal evolution of the average community size (\bar{S}_c) in the number of nodes is presented. It was assumed that the formation of communities begins in the time in which at least 3 are observed. Given this condition of presence \bar{S}_c begins to decline. This is likely because as nodes are added to the system, links are formed that favor the segmentation of the study. Then \bar{S}_c increases by adding new nodes with the publication of new articles; however, over time the dominant effect is the formation of new communities of small sizes, as can be seen in the histogram in Fig. 5a. This previous analysis allows us to interpret the temporal curves of the number of communities. The adjustment was made to each of the trends using Eq. (5). In these curves, it can be observed that at first, there are no communities due to the publication of articles with only two authors at this time. Subsequently, there is a slow growth given the small number of new nodes. Then, the number of community experiences accelerated growth as a result of a higher rate of new nodes and collaborations. However, from the time step t_c , at the inflection point, the growth rate is reduced. For each of the networks, the value for t_c is below the values of t_p for the new nodes and links (Table 2). In other words, the inflection point occurs before the system presents the largest number of new nodes and links. Additionally, the t_c taken from energy is also above that of the domains. Finally, in the last steps of time, N_c tends to stabilize due to the few authors that transit between communities in these times. In principle, the distribution established for the entry of new nodes could play a crucial role in predetermining the inflection point. This is very important, if that inflection point is known in real time, subsequent events can be predicted, including when the maximum number of new nodes will occur per year and an estimate of that number.

The simulations also made it possible to analyze these trends in detail. Figure 7a presents the number of domains as a function of time for different t_p in the n_i distributions. Very similar trends are observed between one method and another, networks and simulation. In the first step the number of domains is zero due to the low number of spins and the competition dominated by temperature on a very low interaction energy in the system. As time passes, domains begin to form at high rates, more stable due to the increase in FM links, mainly. This process ends in an approximately constant value for N_d since, in the last steps, very few spins enter, actually affecting the system. Also, in Fig. 7a, the increase in t_p shows an increase in time at the inflection point. The

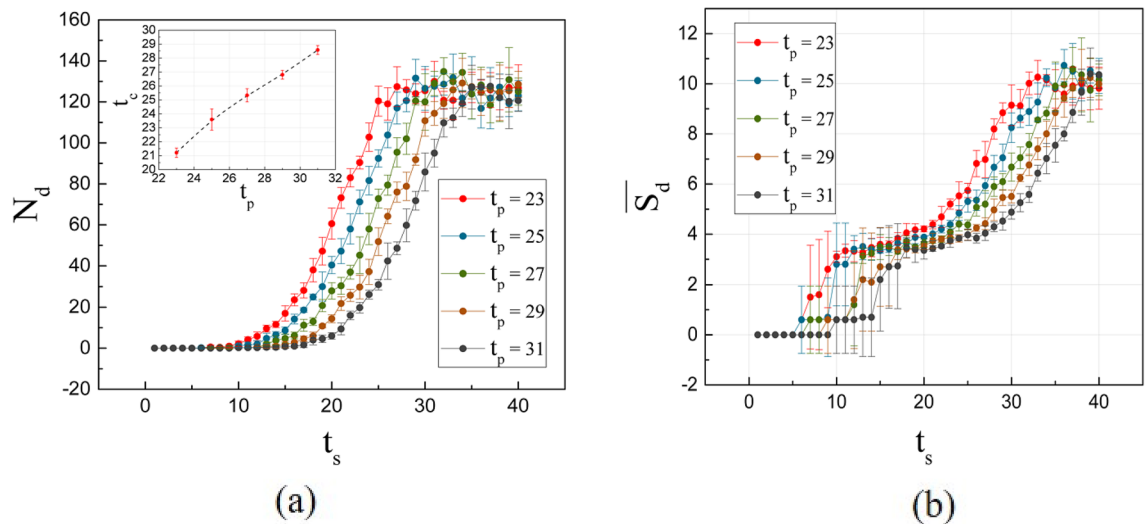


Figure 7. (a) Number of domains as a function of time for different values of t_p in Monte Carlo simulations. Inset: t_c as a function of t_p , obtained from these trends. (b) Average domain size as a function of time for different values of t_p .

relationship can be approximated to linearity, as seen in the inset figure. These times show to be lower than those observed for t_c in the energy and in t_p both for the nodes and the links, which agrees satisfactorily with the previous analysis of the co-authorship networks. For this reason, it cannot be considered that it is the same transition.

For the curves of the average domain size (\bar{S}_d) as a function of t_s , in Fig. 7b), 4 stages can be analyzed. The first corresponds to the first-time steps where great instability is observed. This is due to the fact that the temperature prevents the formation of stable domains due to the low density of links. In terms of co-authorship, this indicates that various working groups are being formed, but strong links have not yet been formed to last over time. Note how the error bars show the great randomness of the links. In the second stage, the average domain size tends to stagger. In particular, at $t_p = 31$ in two steps. These stable sections were also observed in the co-authorship analysis (see Fig. 6). Apparently, they can occur anywhere on the curve during community size growth. In the third stage, the average domain size increases as time passes; in the middle of this stage it is observed in the graph of the number of domains that N_d tends towards stability, approximately. This fact is a sign that the new spins that are added to existing domains. In the last stage \bar{S}_d tends to stability. Again, the speed with which the system goes through these 4 stages is higher for distributions with shorter peak times, the result of higher initial densities at lower times.

Conclusions

A methodology based on Ising-type magnetic models was implemented that made it possible to describe the behavior of dynamic and weighted co-authorship networks. The temporal distribution of new nodes was identified as a key pattern in networks. The results of real co-authorship networks and computational simulations showed a high correspondence. Observables such as internal energy and the number and size of communities or domains, verified the effectiveness of the modeling implemented in both methods. The inflection point of energy as a function of time showed a change of state which requires a minimum density of links. The inflection point of the number of communities showed a shorter time than the transition in energy and the peak times of the distribution of nodes and links. Given these results, it can be highlighted that co-authorship network systems can be predicted following the time trend established by a magnetic exchange interaction model.

The study provides others considerable contributions: First, this goes beyond of the most popular methods for community detection in co-authorship networks³⁶, applying a methodology to identify them. Second, this examines the changing features of several social academic networks through time. Here we go beyond the traditional analysis of static networks³⁷. It is essential to highlight this because social connections influence the emergence and fall of an academic theme³⁸. Thus, for new researchers, it is not only important to identify a topic, but also its current academic social community. Finally, we propose a new method to find the tipping point in social networks taken from the field of magnetism. This contributes to understand the research topic's momentum improving the researcher's decisions about the time and energy spent in the process.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. OriginPro 8, RStudio and Gephi were used for image processing^{39–41}.

Received: 3 October 2020; Accepted: 8 February 2021

Published online: 11 March 2021

References

- Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**(2), 591–646. <https://doi.org/10.1103/PhysRevLett.109.248109> (2009).
- Barabási, A. L. *Bursts: The Hidden Patterns Behind Everything We Do, from Your E-mail to Bloody Crusades* 1st edn. (Dutton, 2010).
- Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**(1), 47–97. <https://doi.org/10.1088/1478-3967/1/3/006> (2002).
- Fernandez, M. A., Korutcheva, E. & de la Rubia, J. A 3-states magnetic model of binary decisions in sociophysics. *Phys. A* **462**, 603–618. <https://doi.org/10.1016/j.physa.2016.06.017> (2016).
- Galam, S., Gefen, Y. & Shapir, Y. Sociophysics: A new approach of sociological collective behaviour. I. mean-behaviour description of a strike. *The Journal of Mathematical Sociology* **9**(1), 1–13. <https://doi.org/10.1080/0022250X.1982.9989929> (1982).
- Son, S. W., Jeong, H. & Noh, J. D. Random field Ising model and community structure in complex networks. *Eur. Phys. J. B* **50**, 431–437. <https://doi.org/10.1140/epjb/e2006-00155-4> (2006).
- Reichardt, J. & Bornholdt, S. Detecting fuzzy community structures in complex networks with a Potts model. *Phys. Rev. Lett.* **93**(21), 218701 (2004).
- Raghavan, U. N., Albert, R. & Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **76**(3), 036106 (2007).
- Han, J., Li, W., Su, Z., Zhao, L. & Deng, W. Community detection by label propagation with compression of flow. *Eur. Phys. J. B* **89**(12), 272 (2016).
- Stauffer, D. Social applications of two-dimensional Ising models. *Am. J. Phys.* **76**, 470 (2008).
- Peixoto, T. P. Network reconstruction and community detection from dynamics. *Phys. Rev. Lett.* **123**, 128301 (2019).
- Herrera, M., Armelini, G. & Salvaj, E. Understanding social contagion in adoption processes using dynamic social networks. *PLoS ONE* **10**, e0140891 (2015).
- Lalli, R., Howey, R. & Wintergrün, D. The dynamics of collaboration networks and the history of general relativity, 1925–1970. *Scientometrics* **122**, 1129–1170 (2019).
- Barabási, A. L. & Pósfai, M. *Network Science* 1st edn. (Cambridge University Press, 2016).
- Newman, M. E. J. & Park, J. Why social networks are different from other types of networks. *Phys. Rev. E* <https://doi.org/10.1103/PhysRevE.68.036122> (2003).
- Newman, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E* **64**(1), 016131 (2001).
- Barabási, A. L. *et al.* Evolution of the social network of scientific collaborations. *Phys. A* **311**(3–4), 590–614. [https://doi.org/10.1016/S0378-4371\(02\)00736-7](https://doi.org/10.1016/S0378-4371(02)00736-7) (2002).
- Newman, M. E. J. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U.S.A.* **98**(2), 404–409 (2001).
- Zhou, W. & Zhao, S. Community detection based on gravitational coefficient in collaboration network. *Mod. Phys. Lett. B* **34**(14), 2050143 (2020).
- Pelacho, M., Ruiz, G., Sanz, F., Tarancón, A. & Clemente-Gallardo, J. Analysis of the evolution and collaboration networks of citizen science scientific publications. *Scientometrics* **126**, 1–33 (2020).
- Zhou, W. & Zhao, S. Community detection based on gravitational coefficient in collaboration network. *Mod. Phys. Lett. B* **34**, 2050143 (2020).
- Newman, M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* **64**(1), 016132 (2001).
- Kumar, S. Co-authorship networks: A review of the literature. *Aslib J. Inf. Manage.* **67**(1), 55–73. <https://doi.org/10.1108/AJIM-09-2014-0116> (2015).
- Orman, G. K., Labatut, V., Plantevit, M. & Boulicaut, J.-F. Interpreting communities based on the evolution of a dynamic attributed network. *Soc. Netw. Anal. Min.* **5**, 1–22 (2015).
- El-Moussaoui, M., Agouti, T., Tikniouine, A. & El Adnani, M. A comprehensive literature review on community detection: Approaches and applications. *Procedia Comput. Sci.* **151**, 295–302 (2019).
- Carusi, C. & Bianchi, G. Scientific community detection via bipartite scholar/journal graph co-clustering. *J. Inform.* **13**, 354–386 (2019).
- Amancio, D. R., Oliveira, O. N. & Costa, L. D. F. On the use of topological features and hierarchical characterization for disambiguating names in collaborative networks. *Europhys. Lett.* **99**(4), 1–6 (2013).
- Corrêa, E. A., Silva, F. N., da Costa, L. F. & Amancio, D. R. Patterns of authors contribution in scientific manuscripts. *J. Inform.* **11**(2), 498–510 (2017).
- Xie, Z. Predicting the number of coauthors for researchers: A learning model. *J. Inform.* **14**(2), 101036 (2020).
- Li, H. & Liu, W. Same same but different: Self-citations identified through scopus and web of science core collection. *Scientometrics* **124**(3), 2723–2732 (2020).
- Liu, W. A matter of time: Publication dates in web of science core collection. *Scientometrics* **126**, 849 (2020).
- Liu, W. The data source of this study is web of science core collection? Not enough. *Scientometrics* **121**(3), 1815–1824 (2019).
- Hartmann, A. K. & Rieger, H. *Optimization Algorithms in Physics* 1st edn. (Wiley-VCH, 2002).
- Goldberg, A. V. & Tarjan, R. E. A new approach to the maximum-flow problem. *J. Assoc. Comput. Mach.* **35**(4), 921–940 (1988).
- Arbeláez-Echeverri, O. D., Agudelo-Giraldo, J. D. & Restrepo-Parra, E. Atomistic simulation of static magnetic properties of bit patterned media. *Physica E* **83**, 486 (2016).
- Huang, Z. *et al.* Visualizing complex networks by leveraging community structures. *Phys. A* **565**, 125506 (2021).
- Abbasi, A., Altmann, J. & Hossain, L. Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *J. Inform.* **5**, 594–607 (2011).
- Kastrin, A. & Hristovski, D. Scientometric analysis and knowledge mapping of literature-based discovery (1986–2020). *Scientometrics* <https://doi.org/10.1007/s11192-020-03811-z> (2021).
- “Origin(Pro). Version 8. OriginLab Corporation.
- RStudio Team. *RStudio: Integrated Development for R* (RStudio, PBC, 2020).
- Bastian, M., Heymann, S. & Jacomy, M. *Gephi: An Open Source Software for Exploring and Manipulating Networks*, Vol. 3, 361–362 (2009).

Acknowledgements

Authors are grateful to Universidad Nacional de Colombia, which, through the Honor Degree scholarship, has been of support for the realization of this research.

Author contributions

J.D.A.-G. and S.R. conceived of the presented idea as thesis of V.A.H.-M. Like advisor, E.R.-P. verified the analytical methods and encouraged V.A.H.-M. to investigate in social physics area. All authors developed the

modelling. V. A. H.-M. performed the computations. All authors discussed the results and contributed to the final manuscript.

Funding

The funding was provide by Dirección de Laboratorios of the Universidad Nacional de Colombia sede Manizales , Ministerio de Ciencia y Tecnología e Innovación of Colombia and also by Dirección de Investigaciones of the Universidad Autonoma de Manizales.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.D.A.-G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021