On Complexity and the Prospects for Scientific Advancement

Luciano da Fontoura Costa*1📖

^{1}Universidade de São Paulo, Instituto de Física de São Carlos, São Carlos, SP, Brasil.

With the onset of areas such as complex systems, network science, and artificial intelligence, efforts have been invested in modeling science itself. In the present work, we report a related approach to modeling the influence of the complexity of knowledge on the respective prospects for scientific advancement. More specifically, we focus on the question of how much the topological complexity of the knowledge network can influence the prospects for scientific advancement. Once the knowledge has been represented as a complex network, we consider one of its subnetworks, the nucleus, as representing the currently known portion of that network. The relative number of nodes adjacent to the nucleus, and the ratio between this quantity and the quantity of edges interconnecting the nucleus with the remainder of the network, are taken as quantifications of the potential for scientific advancement and the efficiency with which these advances can take place. Subsequent nucleus sizes are considered in both a simpler network (Erdos-Renyi) and a more complex model (Barabasi-Albert). The results surprisingly tended to vary little between these two models, suggesting that the complexity of the knowledge network may have little effect on the prospects for scientific advancement as modeled in the present approach.
Keywords: Complexity, complex networks, scientific advancement, network science.

'*Ma la città non dice il suo passato, lo contiene come le linee d'una mano...*'

Italo Calvino, Le città invisibili.

1. Introduction

It took a long time to be discovered and, once discovered, it took some time to be believed. But, yes, the paintings in pre-historic caves were authentic indeed.

Dating back to as early as 30.000 years, *rock art* as found in Chauvet and other locations worldwide, is fascinating in so many aspects. Revealing a near perfect balance of abstraction and realism, these paintings represent one of the very few available direct links with our remote ancestors. Though the paintings tell us little about important questions such as language, other forms of art as music, or the religious beliefs at the time, they *do* conceive a message to us ... and what an impressive message it is! For one simple reason, at least: these paintings represent the earliest evidence of the importance of *modeling* for human beings.

As a matter of fact, rock art possesses virtually every feature of scientific models. First, they provide a simplified, though effective, *representation* of real world entities and actions. Their traces concentrate on the most intrinsic features of the objects (e.g. a rhino tusk,

or the graceful moves of antelopes), which are skillfully emphasized. Then, we have that rock art possibly contributed to a better *understanding* of the represented entities, in the sense that the painting, realized in dark, hardly accessible depths of cave chambers, could not have been accomplished without intentional preliminary conceptualization of the entities being portrayed. Last, but not least, rock art has an intrinsic simulation and *predictive* role. For instance, the hunting scenes that have come to us may have provided the opportunity for mentally rehearsing strategies and possible events in challenging hunting expeditions necessary to secure food for their families and communities in those long gone times.

Yet, precocious and precious as it was, the human predisposition to modeling continued to blossom in an inexorable manner. It is one of the hypothesis of the present work that performing some kind of modeling, even if unconscious, constitutes an intrinsic characteristic of humans, especially in the medium and long terms. Or, in other words, human beings are, intrinsically, *modeling agents* not only of their environment, but also of other individuals (e.g. [1]). We could go as far as understanding much of the human history—not only in science and technology but also in politics, agriculture, and so many other activities—as being, to a good extent, the *history of modeling.*

Though several elements substantiate this surprising approach, one is of particular importance: the *prediction* ability endowed to humans by the modeling activity. As such, many of the human decisions along history have

---

* Correspondence email address: luciano@ifsc.usp.br

been based on the available respective models, even if performed unawarely. For instance, much thought was invested in foreseeing the result of actions such as founding a new city, planning crops, and organizing defenses. All these actions relied critically in intrinsic modeling activities, in which the most important elements, as well as their properties and interrelationships, were carefully taken into account, so that possible outcomes of specific decisions could be predicted as precisely as possible. As a consequence, modeling was progressively acknowledged as a sound approach, and the ultimate consequence was the consolidation of science and technology as we know them nowadays.

With continuous advances in statistics, pattern recognition, and artificial intelligence, the own modeling approach became the subject of its attention. In other words, people got interested in better understanding, through meta-modeling and simulation, the modeling approach characterizing science.

As an example, any body of knowledge can be represented as a complex network, in which each node represents a portion of knowledge while the respective interconnections express interrelationships between those pieces of knowledge (e.g. [2–8] and references therein). A specific example of this approach are the citation networks, where each node corresponds to a scientific article, while the links indicate citations between articles, giving rise to a *directed* graph or complex network. Other types of interrelationships are possible between the pieces of knowledge represented as nodes, including content similarity and pre-requisite conditioning. Observe also that the piece of knowledge represented as a node in this type of complex network approach can vary substantially along scales, ranging from single concepts to more complete subparts such as an abstract, a theorem, a procedure, individual models, complete articles, or even books and theories.

While knowledge can be represented as a complex network, the act of scientific investigation can be understood as the means to unveiling this network in an as effective as possible manner (e.g. [2, 7, 9–11]). Observe that part of the knowledge network is typically known already when these explorations begin. For simplicity's sake, this portion of the network, which corresponds to a subgraph or subnetwork, will henceforth be referred to as the *nucleus* of a given research program. The objective of these researches can then be understood as means of complementing the nucleus (e.g. [2]) by moving into yet unexplored adjacent pieces of knowledge (neighboring nodes). Therefore, the adjacency between the nucleus and the nodes connected to it provides a basis for conceptualizing the potential scientific advancement with respect to the currently available knowledge.

The effectiveness of these explorations can be estimated by using distinct metrics that take into account the aimed results. For instance, if one is interested in expanding the nucleus as fast as possible, the number of new nodes discovered after a given number of steps can be taken as respective metric. If what is desired instead is to connect two separate concepts (nodes), the distribution of distances between these two respective nodes can be taken into account. In the present work, we will consider the number of adjacent links and nodes reachable from the current nucleus as possible quantifications of the *prospect* or *potential* for scientific advancement implied by the interconnections between nucleus and an the overall knowledge network.

There are at least two interesting aspects that can directly impact the current potential for scientific advancement as conceptualized above: (a) the topological characteristics of the subnetwork corresponding to the already known body of knowledge (i.e. the nucleus), which is greatly influence by the manner in which the nucleus is chosen; and (b) the topology of the overall knowledge network itself (e.g. uniformly random or scale free). Several interesting questions are respectively implied. For instance, it would seem that a more sparse nucleus can promote scientific opportunity. If so, would this tendency be similar in complex knowledge networks with simpler or more complex topologies? This is one of the main questions addressed in the present work, reflecting the current interest in complexity and complex systems (e.g. [4, 12–16]).

As we develop our study, we will see that that the relative number of nodes (n/N), where N stands for the total number of nodes in the knowledge network, and links (e) adjacent to the nucleus provide particularly interesting quantifications of the respective prospects for scientific advancement. In particular, we will consider the ratio $s = n/e$ as a particularly informative indication of the efficiency in which scientific advancement takes place. Observe that this index is favored by large n, characterizing that more pieces of knowledge can be reached from the nucleus through e links.

The ratio s can be interpreted as a measurement of *efficiency* or *facilitation*, as it corresponds to the average number of interconnections between the nucleus nodes and their respective still unexplored neighbors. In other words, if the nodes adjacent to the nucleus are accessible through alternative manners, they will be more likely to be discovered at any given time. As an example, node 5 in Figure 1 can be accessed both from nodes 3 and 6. Therefore, relatively large values of s can be taken as being favorable for scientific advancement.

We shall start by presenting the main hypothesis and representations, and introducing the two adopted indices for quantification of the degree of scientific prospect, and follow by presenting simulation results considering uniformly random networks (more specifically the Erdős-Rényi model) and a scale free model (the Barabási-Albert model), which are respectively understood as representing a simpler and a more complex structure. We hope for good prospects in our developments.
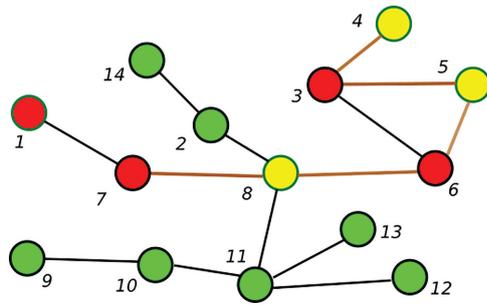
Figure 1: A given overall knowledge network with $N = 14$ nodes and a possible nucleus (red nodes), corresponding to the portion of the overall network that is currently known. In the present work, we quantify the *prospect* for discovery established by such configurations in terms of the number of nodes n and links e not yet known that are adjacent to the nucleus (shown in yellow) and which can, therefore, be discovered while exploring from the nucleus (through links shown in orange). More specifically, we will give special attention to the ratio $r = n/N$ as a relative indication of the prospect for scientific advancement and a ratio $s = n/e$ taken as a measurement of the efficiency of the scientific advancement (observe that $0 \leq s \leq 1$). In the case of the example in this figure, we have $r = 3/14$ and $s = 3/5$. How will the potential for scientific discovery defined by specific types of nucleus vary with respect to the topological structure and complexity of the nucleus and overall knowledge network?

2. Hypotheses, Representations and Measurements

Reflecting the fact that all the universe may be fully interrelated (e.g. Bell's theorem), especially through fields of asymptotic decay, the overall network can be, in principle, assumed to constitute a connected component. However, this characteristic may not be required from the nucleus, as not every piece of available knowledge is taken as being interconnected while performing some particular research program, which can also be developed possibly independently by different research groups or along distinct periods of time.

In order to study the influence of complexity on the prospects for scientific advancement, we will resource to two well-known distinct theoretical models of complex networks for representing the overall knowledge network (e.g. [17]). The 'simpler' model will correspond to a uniformly random network, more specifically the Erdős-Rényi model. The Barabási-Albert model will be adopted for representing a more 'complex' topology knowledge network. Throughout this article, all networks to be compared will have the same number N of nodes and average degree $\langle k \rangle$. The reason for the quotations is that complexity remains an elusive and somehow subjective concept (e.g. [16]). For simplicity, we will be restricted to a fixed undirected network with size of $N = 1000$ nodes.

Though several different approaches could have been used for specifying the nucleus, we will limit our research to a simple but effective method. More specifically, a set

Figure 2: The adjacency matrix K of the network in Fig. 1. The columns corresponding to the nucleus nodes (i.e. 1, 3, 6, 7) are marked in red. The rows corresponding to still unexplored nodes covered by the red columns (i.e. 4, 5, 8) are marked in yellow. The number of connections between the nucleus and the nodes to be discovered can be obtained by adding the entries along the respective red rows intersected with the red columns.

of C nodes will be chosen *uniformly* among the N nodes of the overall knowledge network. For generality's sake, we take the relative number $c = C/N$ into account to express the nucleus size in relative terms. Then, in order to infer how the prospects for scientific advancement change as the overall network becomes more understood, we will vary c between 1 and N, therefore defining respective signatures of the considered measurements in terms of the free variable c.

For instance, the number of discoverable nodes for a nucleus of size c will be henceforth expressed as $n(c)$. A possible manner to identify these nodes, illustrated in Figure 2, corresponds to finding, along the columns of the respective adjacency matrix K corresponding to the nucleus nodes, all those nodes have not yet been explored. Observe that the degree of each of the network nodes can be readily obtained by adding the entries along the respective adjacency matrix column. The average degree $\langle k \rangle$ of a network therefore corresponds to the average of its nodes degrees. The number of connections between the nucleus and the nodes to be discovered can be readily calculated by adding along the respective red rows intersected with the red columns.

Interestingly, the results obtained in this work revealed little variation of the two adopted indices ($r(c)$ and $s(c)$) with respect to distinct nucleus of the same size in the two considered network models.

In order to perform our simulations, we will consider $R = 20$ realizations of each experiment configuration, which will allow us to develop our comparative discussion in terms of the average ± standard deviation of each of the two considered measurements $r(c)$ and $s(c)$.

3. Scientific Prospect in 'Simple' and 'Complex' Knowledge Networks

Figure 3 depicts the results obtained for $r(c)$ and $s(c)$ for $N = 1000$ and average degree $\langle k \rangle = 6, 20$ and
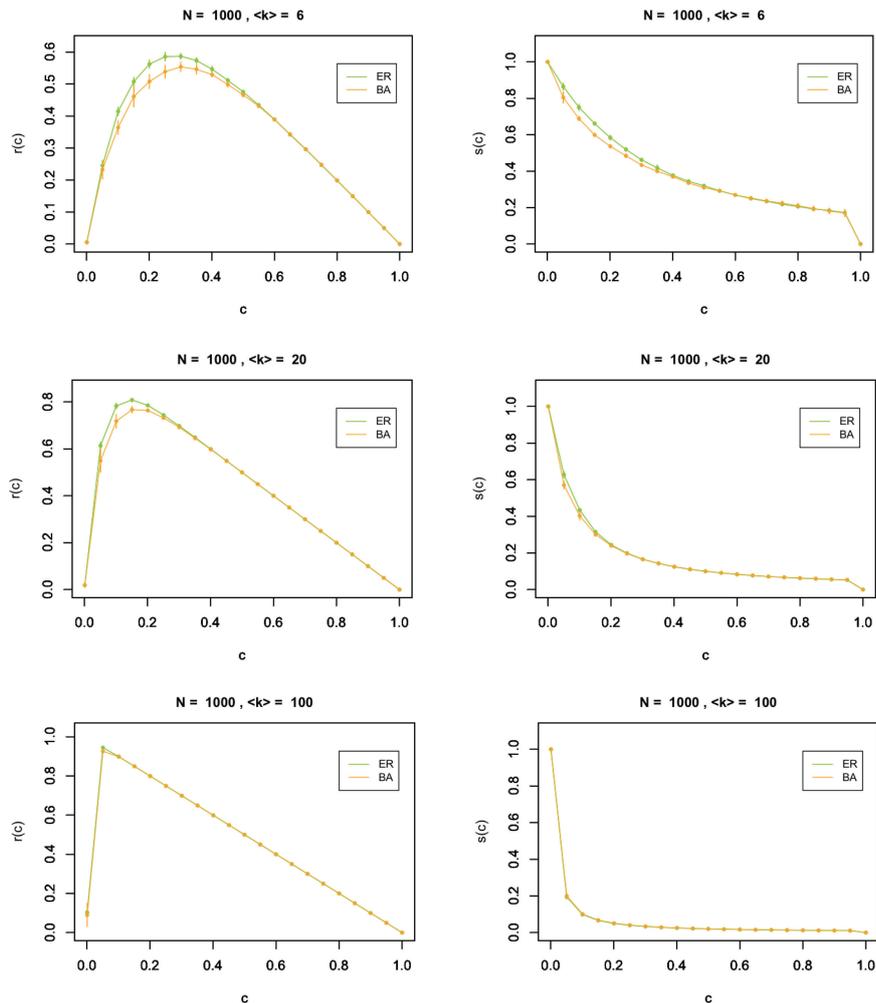
Figure 3: The two parameters $r(c)$ and $s(c)$ quantifying the current scientific prospects (left-hand side) and the efficiency (right-hand side) in terms of the relative size of the nucleus c. These results were obtained for an overall knowledge network with $N = 1000$ and $\langle k \rangle = 6, 20$ and 100, for $R = 200$ realizations. The nodes constituting the nucleus were drawn with uniformly random probability among the $N = 1000$ nodes of the overall network. Interestingly, quite similar results have been obtained for both the 'simpler' (ER) and more 'complex' (BA) network, suggesting that the intrinsic complexity of the overall knowledge network had little impact on both $r(c)$ and $s(c)$. In addition, a peak of scientific opportunity is observed for $r(c)$ in all cases, which is followed by a regime of linear decrease. This peak tends to be displaced to the left with increasing average degrees. The last point in each of the $s(c)$ curves does not follow from the simulations and serves only as a reference. See text for a discussion of other interesting aspects implied by these results.

100 considering a 'simpler' (ER) and a 'more complex' (BA) network. A total of $R = 20$ realizations of this configuration were performed in order to estimate the average $\pm$ standard deviation of the two indices shown by the curves in Figure 3.

Several interesting, and even surprising, aspects are suggested by the simulation results shown in Figure 3. First and foremost, we have that the intrinsic topology of the overall knowledge network had little influence on the two indices $r(c)$ and $s(c)$, yielding similar curves for both ER and BA networks throughout. This would suggest that the topological 'complexity' of the overall considered network tended to have little effect on both the prospects and efficiency defined by any size of nucleus. This is surprising, as it could be expected that

the degree heterogeneity of the BA model would have somehow impacted the number of reachable nodes from the nucleus (e.g. through the presence of hubs).

The important point here is that the several nodes in the nucleus effectively provide an effective probing even of heterogeneous structures such as a BA network. In other words, given a nucleus with not too small size, its nodes will be likely to encompass hubs as well as nodes with very small degree, providing a representative sampling of the network topology (e.g. average node degree) while reflecting the average degree to a good extent. Figure 4 provides an illustration of this possible effect respectively, for simplicity's sake, to a small BA network: a similar number of yet unexplored nodes are accessible from different choices of nuclei.
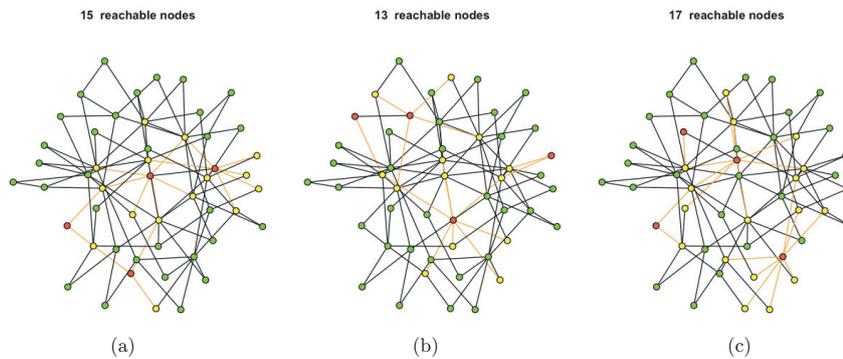
Figure 4: Illustration of the basic result in which the number o nodes reachable (yellow) from the nodes in the nucleus (red) does not change substantially with different uniform choices of node to constitute the nucleus. The novel and non-reachable nodes are shown in green. Though this interesting property could be expected in the case of uniformly random network models such as ER, it is relatively surprising to be observed in a network presenting higher levels of degree heterogeneity, as is the case with the BA model. The network in this example is a realization of the BA network with $N = 30$ nodes, $C = 4$ nodes, and $\langle k \rangle = 2$. This property is believed to underly several of the obtained results.

The above identified effect is believed to account for several of the here presented results. Therefore, situations in which the nodes composing the nucleus are chosen in non-uniform manner, e.g. according to some preferential rule (e.g. favor higher or lower degrees, clustering coefficient, etc.), are expected to lead to differentiation of the way in which the two considered types of networks are explored. For instance, if the nucleus nodes are chosen proportional to their degree in a BA network, we will have a markedly enhanced prospect to scientific advancement as compared to a more uniform topology as in ER networks.

The little influence of complexity on $r(c)$ and $s(c)$ has been verified, at least for the adopted configurations, to be irrespective of the average degree of the knowledge network. At the same time, the curves of $r(c)$, which are here understood to reflect the prospects for scientific advancement given a specific nucleus, presented a maximum peak in every case. After reaching the peak, all obtained signatures tended to undergo a regime of monotonic linear decay with c, which can actually be verified to readily converge to a straight line with inclination -1.

The overall unfolding of $r(c)$ in terms of c can be conceptualized in terms of the following three main events: (i) the number of reachable nodes n increases quick and steadily with c up to a maximum peak, (ii) subsequently enters a saturation regime, followed by (iii) a linear decrease. In other words, we have a period of fast increasing prospects for scientific advancement, followed by a longer period of more moderate, nearly linear decrease of prospects. This has been observed, at least for the adopted configurations, for both ER and BA structures irrespectively of node degree $\langle k \rangle$ and overall number of nodes N.

All obtained signatures are also characterized by substantially small standard deviations (short error bars

in the graph), corroborating that the obtained values vary little with different choices of the nucleus. As could be expected, the dispersion is typically larger in the BA than in the ER mode. Observe also that larger dispersion is observed especially for smaller average degree, in the case of the BA model at values of c immediately before the peak. In the case of the largest average degree ($\langle k \rangle = 100$), a particularly large standard deviation was observed for $C = 1$, which can be understood as a moderate manifestation of the degree heterogeneity characteristic of scale free network models, including hubs.

In order to better appreciate the change of the value and position of the peak of scientific opportunity defined by the $r(c)$ signatures in every case, the averages of these two values have been estimated for average degree $\langle k \rangle$ varying from 1 to 100 for both the ER and BA configurations with $N = 1000$. These results are presented in Figure 5.

Again, both considered network models led to quite similar behavior regarding the prospects for scientific advancement as modeled in the current work.

Interestingly, the peak values of $r(c)$ quickly saturate as $\langle k \rangle$ increases. This means that the prospects for scientific advancement tend to vary little with $\langle k \rangle$, except for a relatively short initial range of this parameter. Once in the saturation regime, further increases in the overall network connectivity have little effect on the peak of the $r(c)$ curve.

The positions of the peaks of $r(c)$, shown in the right-hand curves in Figure 5, which corresponds to the value C of $\langle k \rangle$ for which the peak of $r(c)$ is observed, tend to decrease quick and steadily from a relatively small initial value of $c = 9$ to even smaller values of 2 or 3. Indeed, for $\langle k \rangle$ above 20, the peak of scientific prospect takes place for nuclei with just C = 2 or 3 nodes. It is also interesting to observe that the decrease of the efficiency
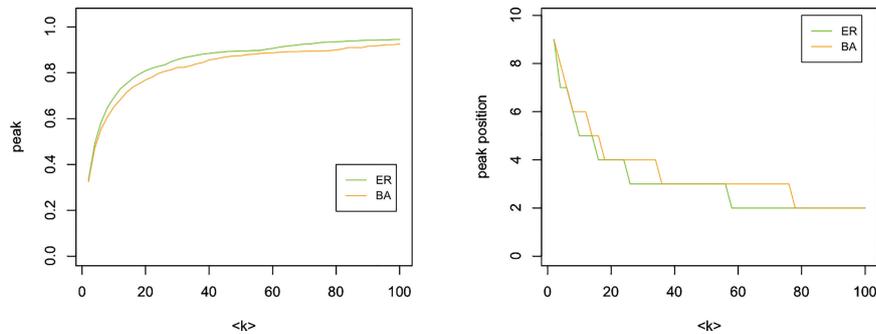
Figure 5: The values and positions of the peaks defined by the $r(c)$ curves for average degree $\langle k \rangle$ varying from 0 to 100 and considering the ER and BA models with $n = 1000$ nodes, for $R = 50$ realizations of each configuration. The peak position is shown in terms of C, i.e. the number of nodes in the respective nucleus where the peak of $r(c)$ is observed.

parameter with c (Fig. 3) tends to be more accentuated for larger values of average degree.

4. Concluding Remarks and Prospects

Science, which depends critically on the development of models, has progressed a long way since its remote beginnings crystalized in rock art, to the point that it became the subject of its own investigation. We got interested in studying and modeling science not in order to better understand this unique human initiative, but also to try to enhance it and make its pursuit even more effective.

Along the history of science, the simpler problems tended to be solved first, so that we the standing problems concentrate among those that are among the most challenging and complex. This trend has motivated new areas in complex systems, such as non-linear dynamic systems, network science and artificial intelligence, aimed at paving the way to new scientific advancements. These concepts and methods have been applied more and more broadly, including the representation of knowledge as complex networks. A particularly interesting question then arises: how would the topology of knowledge networks, and in particular their 'complexity', impact the prospects for scientific advancements defined by a given already known portion, or nucleus?

This question has been addressed in the present work with the help of network science. More specifically, we considered two well-known theoretical models of complex networks, namely the Erdős-Rényi and Barabási-Albert approaches, representing respectively 'simpler' and 'more complex' topologies. Then, the available body of knowledge was assumed to be represented as a subgraph of these overall knowledge networks, which was called *nucleus*. The number n of yet undiscovered nodes that are adjacent (and therefore potentially accessible) to this nucleus, as well as the number e of respective links defining these adjacencies, were then taken into account in order to defined two indices, namely $r(c) = n/N$ and

$s(c) = n/e$ that were understood as possible indications of the current potential for scientific advancements and their respective efficiency.

Despite the simplicity of the approach developed in this work, several interesting, and even intriguing, results have been derived. First, we have the surprising observation that the obtained curves of $r(c)$ and $s(c)$ vary little between the ER and BA model (considering the adopted configurations), also presenting an overall qualitative tendency which is, to a large extent, irrespective of the node average degree and network size. These results seem to lead to the surprising implication that the topological complexity underlying the overall knowledge network may have little impact on the prospects for scientific advancement and related aspects.

It should be nevertheless observed that it has been argued [13] that the complexity of a network goes beyond heterogeneity of node degree distribution, also encompassing the distribution of many other topological properties that are required in order to obtain a more complete mapping from the network properties into the respective measurement-based characterization. Indeed, a good deal of the similarity of the dynamics obtained in ER and BA structures stems from the fact that these models were assumed to have the same average node degree, which ended up not only influencing the performed dynamics, but also effectively inferred by the uniformly random choice of nodes to compose the nucleus. As such, the performed dynamics ultimately depended strongly on topological properties shared by the two types of considered networks.

More generally, we can postulate that this type of insensitivity of a given dynamics to two or more topological network models can be possibly accounted by at least the two following issues: (i) the dynamics depends strongly on topological features shared by the considered overall networks; and/or (ii) the considered types of networks differ only with respect to a relatively restrict set of properties, such as average node degree, that are not differentiated among the considered overall networks.

In the present case, the similarity of the obtained dynamics can be to a good extent explained by the fact that the two considered models distinguished one another in complexity mainly with respect to their degree distribution while sharing the same average value. Another important aspect is that the nodes composing the nucleus were chosen uniformly along time time.

It would be interesting to study if the trends suggested in the present work would hold for networks presenting complexity also of other topological measurements as well as for other types of exploratory dynamics. It could be expected that non-small world networks presenting nodes with modularized topological features may reveal a stronger influence of the overall network structure on the prospects for complementing distinct types of nucleus, especially when the nodes in the nucleus are chosen in non-uniform manners.

In addition, in all the considered situations, the prospects for scientific advancement as modeled by the present approach exhibit a characteristic unfolding pattern corresponding to a quick increase of $r(c)$ with c, up to a maximum peak (saturation), followed by a longer period of more moderate, nearly linear decrease. The efficiency parameter presented a continuously decreasing trend, nearly exponential, characterized by a quick reduction of efficiency followed by stabilization at very small values. These signatures tended to directly accompany the unfolding of $r(c)$, with the initial interval of fast increase of scientific prospects coinciding with the interval in which the efficiency $s(c)$ decreases more steeply. The duration and sharpness of the unfolding along each of these regimes was found to depend strongly on the average node degree of the knowledge network.

Other interesting results include the relatively little dispersion observed around the average of $r(c)$ and $s(c)$ in all considered situations, which corroborates that the specific topological features of the nuclei had little influence on those two measurements. We also investigated the dependency of these features respectively to the average degree $\langle k \rangle$, having observed that the maximum peak value tends to increase quickly to approximately $0.9N$ with $\langle k \rangle$, while the peak position decreases steeply to very small nucleus sizes of 2 or 3 in both considered network models.

It is observed that the interpretation of the concepts and measurements adopted in the present work respectively to the overall topics of prospects for scientific advancement are intrinsically specific to the respectively adopted approach and parameter configurations. In other words, it is by no means meant that the prospects for scientific advancement in the real-world necessarily follow the model or results reported here. However, it is expected that these results can shed some light on instigating related questions, perhaps providing subsidies for better understand the dynamics of science.

It is also hoped that the present work can serve as a didactic example of how network science renders itself as a relatively simple and effective means of developing a simple model of an important problem, with potentially interesting results that can be further investigated.

The presently reported approach and respective results paves the way to a number of possible extensions. For instance, it would be interesting to consider other network topologies, such as Watts-Strogatz, modular, and geographic network models, as well as networks incorporating weights and directionality. It would also be interesting to verify the effect of choosing the nucleus nodes in a non-uniform way, such as by using rules preferential to some topological properties of the nodes. Better understanding about the observed phenomena can also be obtained by developing analytical expressions for $r(c)$ and $s(c)$.

It would also be promising to adopt diverse types of nucleus topology, such as strings of interconnected nodes, trees, or cliques. A related question would be to try to identify, through optimization, the nuclei of a given size that leads to the smallest or largest number of reachable nodes.

Also of interest is to verify if real-world networks lead to signatures of $r(c)$ and $s(c)$ that are different from a respective null model (e.g. ER network with the same number of nodes and average degree). In this respect, this methodology could provide an interesting statistical test about the sensitivity of a given network to different choices of nucleus, providing complementary indication about the heterogeneity of the topological features of the overall network and each given type of nucleus.

Adddional possibilities consist in considering the reported approach and results to problems other than scientometrics. Indeed, it would be interesting to apply the proposed methodology and measurements to characterize and better understand complex systems involving the interaction between a subnetwork (nucleus) with the remainder of the system, such as it is often verified in biological molecules interactions, epidemics, social networks, ecology, linguistics, urbanism, economics, neuronal networks, philogenetics, among many other possibilities. Given that the proposed approach provides information about the interface between the nucleus and the overall network, it also has potential for applications in the analysis of the robustness of networks. Another interesting problem is, given an overall network, how can specific nuclei be designed so that specific considered dynamics progress in similar ways in distinct networks, including those presenting rather distinct complexity?

The approach reported in this work can also be understood as implying an interesting respective paradigm, namely the study of dynamic processes in networks that are insensitive to specific topological features of the overall network. This kind of research could contribute to devising approaches that can be used to eventually circumvent the complexity of some topological and dynamical structures while providing possible means for their control.

Some of the above possibilities are currently being developed and future results could be expected.

Acknowledgments

References

[1] L.F. Costa, Modeling: The human approach to science, Researchgate, 2019, available in: https://www.research gate.net/publication/333389500_Modeling_The_Human_ Approach_to_Science_CDT-8, accessed in 01/10/2020.

[2] L.F. Costa, Phys. Rev. E. **74**, 026103 (2006).

[3] J.B. Bunoro and L.F. Costa, Rev. E. **82**, 016103 (2010).

[4] L.F. Costa and F.A. Rodrigues, Europhysics Letters **85**, 083106 (2009).

[5] F.N. Silva, B.N. Travencolo, M.P. Viana and L.F. Costa, J. Phys. A. **43**, 325202 (2010).

[6] H.F. Arruda, F.N. Silva, C.H. Comin, D.R. Amancio and L.F. Costa, Information Sciences **421**, 4154 (2017).

[7] T.S. Lima, H.F. Arruda, F.N. Silva, C.H. Comin, D.R. Amancio and L.F. Costa, Chaos **28**, 48001 (2009).

[8] H.F. Arruda, F.N. Silva, C.H. Comin, D.R. Amancio and L.F. Costa, Physica A, **515**, 641 (2019).

[9] A. Rzhetsky, J. Foster, I. Foster and J. Evans, PNAS **112**, 4569 (2015).

[10] J. Foster, A. Rzhetsky and J. Evans, American Sociological Review, **80**, 875 (2015).

[11] F. Shi, J. Foster and J. Evans, Social Networks 46, **265** (2015).

[12] F. Heylighen, What is complexity?, available in: http://pespmc1.vub.ac.be/COMPLEXI.html, accessed in 05/05/2019.

[13] L.F. Costa, What is a complex network? available in: https://www.researchgate.net/publication/324312765_ What_is_a_Complex_Network_CDT-2, accessed in 01/ 10/2020.

[14] M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos*, (Simon and Schuster, New York, 1993).

[15] S. Kauffman, *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*, (Oxford University Press, 1996).

[16] L.F. Costa, Quantifying complexity, Researchgate, 2019. available in: https://www.researchgate.net/publication/ 332877069_Quantifying_Complexity_CDT-6, accessed in 01/10/2020.

[17] L.F. Costa, F.A. Rodrigues, G. Travieso and P.R. Villas Boas, Advances in Physics **265**, 167 (2007).