

Toward a Formal, Visual Framework of Emergent Cognitive Development of Scholars

Amir Hussain · Muaz Niazi

Received: 18 February 2012 / Accepted: 6 May 2013 / Published online: 4 June 2013
© Springer Science+Business Media New York 2013

Abstract Understanding the cognitive evolution of researchers as they progress in academia is an important but complex problem; one that belongs to a class of problems, which often require the development of models to gain further understanding of the intricacies of the domain. The research question that we address in this paper is: how to effectively model this temporal cognitive mental development of prolific researchers? Our proposed solution is based on noting that the academic progression and notability of a researcher are linked with a progressive increase in the citation count for the scholar's refereed publications, quantified using indices such as the Hirsch index. We propose the use of an yearly increment of a scholar's cognition quantifiable by means of a function of the scholar's citation index, thereby considering the index as an indicator of the discrete approximation of the scholar's cognitive development. Using validated agent-based modeling, a paradigm presented as part of our previous work aimed at the development of a cognitive agent-based computing framework, we present both formal as well as visual agent-based complex network representations for this cognitive evolution in the form of a temporal cognitive level network model. As proof of the effectiveness of this approach, we demonstrate the validation of the model using historic data of citations.

Keywords Agent-based modeling · Cognitive development · Cognitive agent-based computing · Complex adaptive system · Hirsch index · Complex networks

A. Hussain
University of Stirling, Scotland, UK
e-mail: ahu@cs.stir.ac.uk

M. Niazi (✉)
Bahria University, Islamabad, Pakistan
e-mail: muaz.niazi@gmail.com

Introduction

In this era of proliferation of research publications and global availability of the Internet, it is common to use the growing citations of refereed publications of a scholar as an indicator of his or her success in academia. However, what started out as an informal exercise initially has now evolved into a somewhat formal practice of quantifying this increase in citations as a measure of notability of scholars [49]. While some have argued about the exact nature of the significance of citations of literature [2], used prudently, citations may be used as an indicator of the significance of published research [11].

While citations alone do not reflect a person's notability, a continual citation buildup can be considered as a practical demonstration of a community interest into the person's work. Researchers have recently proposed capturing a person's overall notability by using metrics such as the Hirsch index [25] among others [7, 15]. Each of these indices attempts to give an approximate measure of the productivity, notability as well as impact of the published work of a scientist or scholar. So, while on one hand, the evolution of a scholar's Hirsch index represents the person's academic growth, on the other, it can perhaps also be considered as an indication of the cognitive mental evolution of the scholar. Thus, each yearly sampling of the h-index may also be considered to represent a discrete stage in the evolutionary cognitive development of the researcher.

Contributions

Could we somehow quantify this evolution in the form of a computational model? This is the research question that we attempt to address in this paper. As a solution to

this representational problem, we first develop a formal computational model for capturing this cognitive mental development of prolific researchers. Subsequently, using validated agent-based modeling level from the cognitive agent-based computing framework [39], we develop visual representations of this cognitive evolution followed by a proof of the effectiveness of this approach using historic data from citation indices.

Specifically, the contributions in this paper can be listed as follows:

1. We propose an approach to the modeling of cognitive development involved in scientific research by proposing a novel technique of merging citation data with researcher h-index by proposing temporal cognitive level networks (TCLN), a network modeling approach allowing the visualization of evolutionary cognition of researchers by means of their research papers and associated Hirsch index.
2. We also demonstrate the benefits of TCLN including a simpler, visual representation in addition to a reduction in the utilization of computational resources used for storing and displaying the network. In a TCLN representation, nodes represent researchers, their papers, their citations in addition to a citation index demonstrating the emergent productivity of each researcher in the same representational model. It also requires a much lower order of memory footprint than existing author-paper citation network models.
3. As proof of concept, we present the design and development of an agent-based simulation model for TCLN. This model demonstrates both the calculation of Hirsch index and the representational capabilities of TCLN. Validation of the effectiveness of this approach is demonstrated by means of using historic data of researchers' evolution retrieved from the Google Scholar Index. TCLN can thus represent researchers, papers, citations as well as research productivity in the same model. We also analyze the space complexity proving a reduction in complexity from the order of $O(mn)$ for author-paper networks to the linear order $O(m)$, where " m " is the average number of papers published by a researcher and " n " is the number of researchers. The main objective of this work is to simplify the network representation such that it is easier to study the coevolution of state and topology in research networks as well as represent researcher repute and productivity.
4. For a validation of the framework and Hirsch index calculations, we design and develop an agent-based model, which is then compared with a well-known tool "Publish or Perish" [22] to retrieve historic data from the Google Scholar index. The agent-based simulation

model also demonstrates the representational power of TCLNs.

Outline

The structure of the rest of the paper is as follows: In the next subsection, first we provide background and related work. Next, we develop a formal framework and methodology using a formal specification language "Z" [24]. Subsequently, we describe the agent-based model and algorithms. This is followed by a discussion of the simulation experiments and validation exercises. Finally, we conclude the paper.

Background and Related Work

In this section, we present background and related work. We first start with the important.

Overview

An effect of the hosting of scientific communication (especially open access publications) over the Web is the emergence of citation indices [17] such as Thomson Reuters Web of Science, Scopus, Google Scholar and PubMed. Never before were we able to thereby examine a cross-sectional view of the research productivity of human civilization at any given time. While all such data are easily accessible online, the analysis of such data in terms of looking for specific patterns can be quite daunting because of the exponentially increasing size of citation corpora. As an example, a search for just the top 10 journals listed in "Computer Science, Cybernetics" category of Journal Citation Reports from the years 1995–2010 using the ISI Web of knowledge gives result of the order of 2,801 papers, 6,376 authors, 4,787 keywords and 83, 633 cited references,¹ whereas a topic search for "expert system*" returns 48,875 records.

Due to the complex nature of the citation data, simple statistical measures may not give ample useful information. Instead, an effective mechanism may be to use complex network approach, which involves the transformation of citation data into a network format. This complex network transformation is not optional since the particular network format is needed for the discovery of complex patterns. Now, there are several ways in which citation data can be represented in the form of a network. The simplest could be to create what is typically termed as a citation network [26]. In a citation network, entities such as research papers,

¹ Results based on search performed in August 2010.

authors, institutions or journals become nodes (or vertices²) and citations become the arcs (directed lines) or links. Another possibility is to develop a co-authorship network, which is based on making authors as nodes and co-authorships as lines [6, 30].

Whatever the means of development of the exact type of networks, the eventual goal is to be able to apply network analysis techniques on the resultant network. However, citation networks have certain representational and technical problems. Firstly, the size of citation corpora is growing exponentially. Representing it graphically for analysis would thus require an ever-increasing requirement of computational resources such as RAM. Typical methods using tools such as Network Workbench or CiteSpace [8] solve this problem of network display by pruning the network to display top nodes only [34]. However, this type of preprocessing can actually possibly result in the removal of papers/authors or other citation indicators which might have lesser citations on their own but might have served as central nodes in transition of ideas or connecting different subject categories. Secondly, traditional networks extracted from citation data do not represent the productivity of individual researchers because measuring productivity often requires the use of algorithms which need to take all publications into consideration and not just the top cited one.

Problems in Modeling Research

For the development of a deeper understanding of the cognitive processes requires the development of explicit models [16] starting with implicit mental models. These models also serve the purpose of enhancing our knowledge about cognitive abilities while paving the way for future usage of ideas in the autonomous mental/cognitive development in robots and animals [50].

Research can therefore be modeled as a complex adaptive system (CAS) [38] where numerous researchers at different levels of experience in performing research work together to produce ideas valuable to the particular discipline. The evaluation of these ideas is conducted by means of the publication process of refereed research articles and patents. In addition, as time progresses, community interest, measurable by means of citations indices, indicates the strength of the proposed hypotheses in the views of other peer researchers. The research process [35] is known to be tied in with numerous socio-cultural aspects as noted by Weng et al. [50]. Watts and Gilbert [48] have proposed a model of knowledge-seeking using through scientific publication. Fischer has proposed a theory of cognitive

development, called the skill theory, where cognitive development is based on discrete skill sets.

As we can note here, all of these aspects primarily tie in with the ideas of complexity, evolution and in general, CAS. As such, while research has numerous aspects and facets which can be modeled at the micro-levels, an intuitive way of modeling the emergent outcome of this process would be to examine the cognitive evolution of researchers, which is the goal of this paper. Here, the key problem addressed is how to develop a multi-faceted model comprising of computational, formal and visual representations of the cognitive mental development of researchers.

Cognitive Agent-Based Computing

Recently agent-based modeling and complex networks have been combined and used in the form of cognitive agent-based computing, a single modeling and simulation framework presented in earlier in [43] and expanded in [39]. Applications of the framework have been shown in the areas of disaster alerting systems such as [42], breast cancer decision support [45], power saving and energy harvesting for corporate networks [41], fault tolerance in large-scale transactional systems [32] and peer-to-peer networks [37]. Validated agent-based modeling using in the current paper has been presented earlier in [40].

Other examples of agents and multi-agent systems include the use of JADE agents such as by Fortino et al. [18]. Likewise, an agent-based platform for programming wireless sensor networks has been presented by Aiello et al. [1].

The Importance of Citations in Research

“Publish or Perish” [51] refers to the pressure to publish work continuously to advance in the academia. However, simply publishing articles is not enough. In general, the idea of research evaluation is based on assigning a scalar value directly or indirectly related to the citations of papers. These scalar values include the Thomson Reuters Journal Citation report Impact factors for Journals and various types of indices [7, 13, 25] proposed recently. Recent work [14] has shown that the Hirsch index is a concavely increasing function of time, asymptotically bounded by

$$T^{\frac{1}{\alpha}}$$

where T is the total number of papers in that group, and α is the exponent of Lotka's law of the system.

Impact of a Paper

Citation analysis as a measure of productivity of researchers and evaluation of research has been studied by

² In the literature, vertices and nodes are used interchangeably for graphs. Subsequently we shall only use the term “node.”

Moed [29]. The impact of a paper can be considered to be demonstrated by the interest shown in the paper by the community. One way of measuring this is to count the number of non-self citations recorded in peer-reviewed scientific literature. The use of citations as a measure of impact can be attributed to Eugene Garfield, chairman emeritus ISI, who formalized and gave the idea of scientific impact and impact factors [19–21].

Impact of a Researcher

The impact of a paper leads to the idea of measuring the impact of a scientist over the entire career. This measurement was first proposed in 2005 by Hirsch [25]. In general, the impact of a scientist can be measured by being highly cited by other authors. Egghe [15] presented a detailed overview of the interrelation of Hirsch with related indices. These indices have been considered as effective measures of a scientist's impact or research productivity.

Hirsch index, which is the most fundamental of all indices, is defined formally as follows:

“A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each.”

So, suppose an author has 6 papers: 5 of them have 4 citations and the sixth has 1 citation. Then $h = 4$. It is important to note here that the calculation of h-index is not quite simple. It involves intelligent retrieval of information from some standard indexing source (such as Google Scholar or SCI, SSCI or Scopus) and then calculation of the index. A well-known tool for the calculation of Hirsch index used extensively by researchers is Ann Harzing's “Publish or Perish” [22].

Traditional Citation Analysis Approach

Complex networks are one way of representing relationships and interactions in CAS. In the domain of citation analysis, as noted by Hummon [26], citation networks have been used to represent citations since Garfield's original paper [19]. Research article networks are acyclic because all papers point back in time making close loops rare [3, 31].

Nowadays, citation analysis [9, 27] is performed using some of the following steps:

1. First, data are retrieved from citation databases such as Science Citation Index (SCI) and the Social Sciences Citation Index (SSCI) compiled by the Institute for Scientific Information (ISI). Data are filtered for scope and correctness.
2. Next, the data are pruned for top nodes
3. Next it is represented as a directed or undirected network.

4. Now, clusters are discovered from the data and highly cited and influential papers inside the cluster are identified using some algorithm [10].
5. The network is drawn using various layouts.
6. Network is gradually modified to highlight particular structural features of interest.
7. Interrelation of citation clusters is then studied in detail.
8. Key nodes are identified using different types of centrality measures.

In Fig. 1, we can see how a simple citation network can become complex with just a few thousand nodes.

Cognitive Framework Development

In this section, we present the development of a formal agent-based modeling framework using a formal specification language [46].

The use of formal frameworks allows for exact specification of complex systems and logics as shown by d'Inverno and Luck [12] for the domain of agents and multi-agent systems. Another framework for the development of agent-based models using a similar extension/description of formal specification for wireless sensor networks sensing in complex adaptive environments has been described in Niazi and Hussain [33]. The goal of the current framework is to describe and specify the agent-based model of researchers and its network representation using mathematical symbols instead of being a purely theoretical exercise as is common in formal specification models.

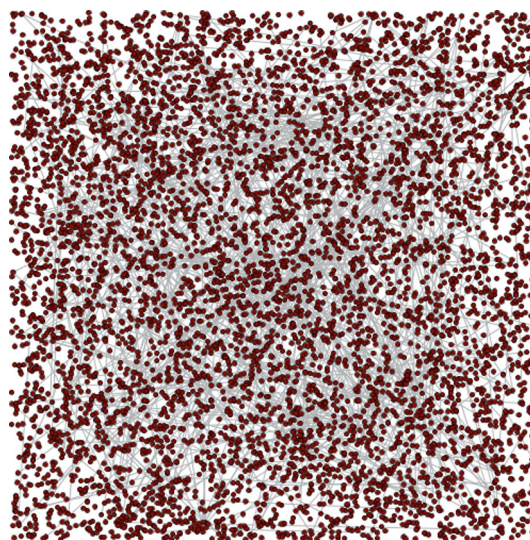


Fig. 1 Erdos992 Pajek data set (Batagelj and Mrvar [4]) drawn using Network WorkBench [47]

Framework

Before we can define a TCLN to represent state transitions over time, let us first define a network formally starting with the definition of a *Node*.

<i>Node</i>
$x, y: \mathbb{Z}$
$x \geq \text{min-}x\text{cor}$
$x \leq \text{max-}x\text{cor}$
$y \geq \text{min-}y\text{cor}$
$y \leq \text{max-}y\text{cor}$

So we have defined *Node* as containing x and y coordinates which are constrained by the limits of our Cartesian coordinate world. Next, we can define a schema for a *Line* as following

<i>Line</i>
$n1, n2: \text{Node}$

A *Line* thus contains two nodes $n1$ and $n2$ of type *Node*. Next, we can define a *Network* state schema as following:

<i>Network</i>
$\text{nodes}: \text{Node}$
$\text{lines}: \text{Line}$
$\forall x: \text{lines} \exists y1, y2 \in \text{nodes} \mid y1=x.n1, y2=x.n2$

A network consists of lines and nodes, which we are depicting by including these two in the declaration. And in the predicate section, we are saying that for each line, there are two nodes, which are both members of the *nodes* subset. Next, we need to define a *ResearchPaper* schema as follows:

<i>ResearchPaper</i>
<i>Node</i>
$r: \text{Researcher}$
$\text{citationCount}: \mathbb{N}$
$\text{citationCount} \geq 0$

Now, here we are saying that *ResearchPaper* is going to be represented as a *Node* in our network. In addition, we are saying that we are including the *Node* and a variable of type *Researcher* schema in this schema. The problem here is that we have still not defined a *Researcher* till now.

So, we also need to formally specify the *Researcher* schema as following:

<i>Researcher</i>
<i>Node</i>
$rp: \{x \mid x \in \text{ResearchPaper}\}$
$\text{totalCites}: \mathbb{N}$
$h\text{-index}: \mathbb{N}$
$\text{totalCites} \geq 0$
$x = h\text{-index}$

Each *Researcher* firstly includes a *Node* and then finite set rp of *ResearchPaper* type. There is also a total Cites member of type “ $\text{fat } N$,” which represents the total number of citations that the *Researcher* has at a given time. Here, in the predicate section, we define the total citations to be a number greater than or equal to zero. A researcher also has an h-index value, which gives the Hirsch index at any given time. In addition, we want to place the *Researcher* on the X - Y plane based on the value of the Hirsch index, which we have accomplished by the schema inclusion of the *Node* schema.

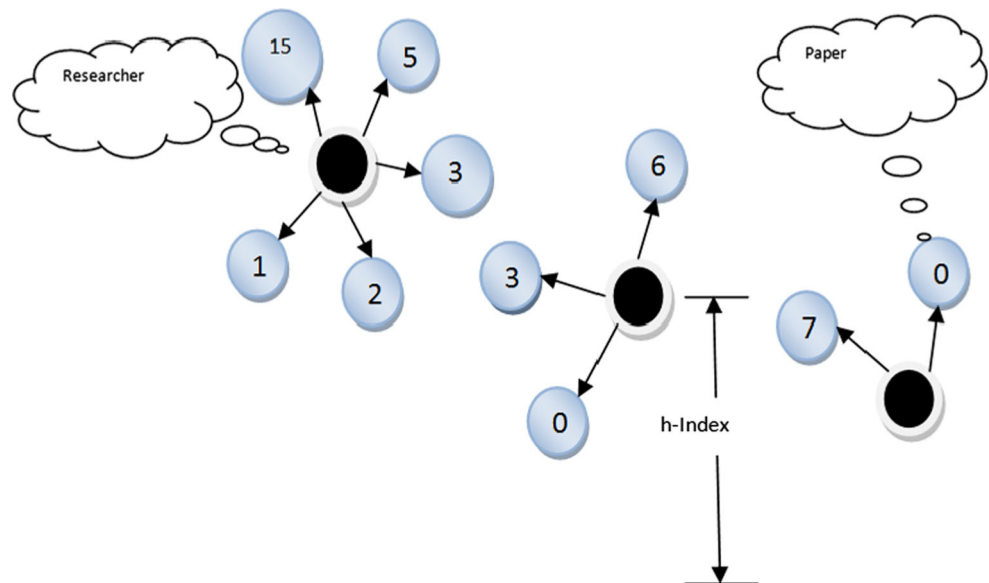
Now that we have the basic framework, we can start with the definition of our specific novel “*TemporalCitationNetwork*” schema.

<i>TemporalCitationNetwork</i>
<i>Network</i>
$\forall x \in \text{lines}$
$x.y1 \Rightarrow x.y2 \in \text{ResearchPaper} \text{ and } x.y2 \in x.y1.rp$
or
$x.y2 \in \text{Researcher} \Rightarrow x.y1 \in \text{ResearchPaper} \text{ and } x.y1 \in x.y2.rp$

This specific type of *TemporalCitationNetwork* basically sets the rules that we have a network but for all lines in this network, the ends are a *ResearchPaper* and a *Researcher* only. This network is temporal because it is dependent upon time. The entire network evolves its state as *Researchers* publish more papers and papers get more citations. Let us define these state transitions as following:

<i>PublishPaper</i>
$r: \Delta \text{Researcher}$
$p?: \text{Paper}$
$\Delta \text{TemporalCitationNetwork}$
$r.rp' = r.rp \cup \{p?\}$
$\text{lines}' = \text{lines} \cup \{l \mid l.y1=r, l.y2=p?\}$
$r.\text{TotalCites}' = r.\text{TotalCites} + p?.\text{CitationCount}$
$r.h_index' = \text{UpdateHIndex}(r)$

Fig. 2 TCLN with black nodes depicting researchers placed according to their Hirsch index and blue nodes depicting individual research papers and citations



Here, we can see that in the *PublishPaper* schema, we update the TCLN as well as the *Researcher* because of the addition of the new paper. Since this update occurs on an yearly basis, we can increase the citation count as well and subsequently use *UpdateHIndex*, an operation schema to update the Hirsch index. A graphical representation of this can be seen in Fig. 2.

Theorem 1 The use of TCLN requires space of the order of $O(m)$ as compared to “traditional” author-paper citation networks whose space requirement is of the order of $O(mn) \approx O(n^2)$, where “ n ” represents the number of average citations received per paper and “ m ” represents the average number of papers per researcher.

Proof The proof of the theorem is easy to follow. In case of traditional citation networks, each paper is connected to other papers. If we depict the papers and the papers citing them in the same network as in the case of traditional networks, then there will be inter-connections of the order of mn . On the other hand, in a TCLN, since we have reduced any such interconnections between researchers and other researchers or papers and other papers, each researcher and the published papers thus appear as a hub-spoke architecture as shown earlier in Fig. 2. Thus, the number of links to be drawn per researcher is equal to the number of papers. Hence, this proves the theorem. The implications of the theorem 1 are also validated and can be seen in the agent-based simulation model described in the next section.

Simulation Model

In this section, we present the agent-based simulation model for representation of TCLN. The model has been

developed using NetLogo, a very popular agent-based simulation toolkit [52]. In NetLogo, an agent-based model typically consists of three different types of entities:

1. Turtles (or breeds of turtles)
2. Links (or breeds of links)
3. Optionally there could be interaction based on the use of patches

Once these different entities have been defined, the key step in the model is to define the behavior of each agent. Agent-based modeling and simulation can be considered to be an advanced form of object-oriented programming. As such, it can be and is often used to represent complex concepts.

So, for our problem, as defined in the formal specification, we define the following types (*breeds*) of agents in our model:

- (A) Researcher agents
- (B) ResearchPaper agents

In the following subsections, we give an overview of these *breeds* of agents.

Researcher Agents

These agents are made up of the breed “researchers.” In NetLogo, this implies that we can subsequently address the breed separately. Now, similar to the model given in the formal specification, we enclose the following three attributes:

```
num-papers, total-citations and my-papers
```

As defined in the formal framework in Sect. 3, “num-papers” is the number of papers of the researcher. “my-papers” contains references to the ResearchPaper agents

connected to this Researcher agent. “total-citations” represents an aggregation of all citations received by all papers. This is updated in the simulation on an “yearly” basis. Now, before we define the details of the interactions, let us also define the ResearchPaper agents.

ResearchPaper Agents

These agents contain the following three variables

```
tend-to-be-cited; tendency is from 0 to 1
num-cites; Total number of citations
my-res; My author
```

“tend-to-be-cited” is a number representing the uniform probability of a paper to get citations. This is useful as an extension of the original model from the framework, because using this, the agent-based model can even represent growth of random researchers and not just real researchers based on retrieved data from indices. Being able to use randomly created set of researchers is done to allow for future extensions based on the testing of hypotheses such as those related to formation of research groups and journal editorial boards.

It would be pertinent to note here that in a TCLN, we shall show each paper as single-authored. In other words, each paper which has been co-authored will be reflected as many times as there are authors, in the model. The reason for choosing this is primarily to avoid any clutter and not show any primary structures of connectivity for optimization reasons (We refer to primary structure in a network to “direct inter-connections between nodes.” An example of primary structure includes the original citation network format). The reason for this is that the goal of our work is not just to be able to represent the coevolution of state-topology networks but also to reduce the memory space required to display the network visually (as proven earlier in Theorem 1). If this simplification is not performed, graphic rendering of an even small subsection of a

moderately complex citation network can range from a few minutes to hours, depending upon the available CPU and other hardware resources of the system. Thus, it would not be possible to observe the coevolution of state-topology visually in real time without the methodological simplification as proposed via TCLNs.

In a TCLN, with the passage of time, the key changes which can occur can be foreseen as following:

- (A) The number of researcher agents increase over time
- (B) The number of papers per researcher increases over time.
- (C) The state of each paper (label) shows the citations and these increase over time.

What this really means is that the use of TCLN can greatly simplify the network representation and thus they are capable of display significantly more complex data easily.

Algorithms and Functions

In this subsection, we present the algorithms and functions implemented in the simulation for each agent. There are two parts of our agent-based model. One can be used to represent the evolution of an actual researcher by loading data from a file. The other part is relevant to the generation of random researchers. Since the generation of the researchers is the more general case and the simulation of evolution of actual researchers is a specialized case, here, we describe the general algorithms used.

Initially, in the *setup* function, we initialize the agents. Each researcher agent is initialized with a uniform random number based on the *max-init-papers* global variable controllable by a user interface element called the “slider.” Inside the same code, the agents then call *make-papers* and *set-lab* functions. Here, we show the code in the following segment in Fig. 3. It is important to note here that we have purposefully chosen the actual code instead of other representations because of the complexity associated with

Fig. 3 Code for the make-papers function

```
to make-papers
  hatch-papers num-papers
  [
    set shape "circle"
    set color blue
    set tend-to-be-cited ((random 101) + 1) / 100
    set num-cites random 10
    let temp [my-papers] of myself ;Add self to the list of the caller's paper
    set temp fput self temp
    set label num-cites
    ask myself
    [
      set my-papers temp
    ]
    create-link-with myself
    set heading random 360
    set my-res myself
    fd 3
  ]
end
```

```

to-report h-index
  let temp-list[]
  foreach my-papers
  [
    ask ?
    [
      set temp-list fput num-cites temp-list
    ]
  ]
  set temp-list reverse sort temp-list
  let i 0
  foreach temp-list
  [
    set i i + 1
    if i > ?
    [
      report i - 1
    ]
  ]
  if length temp-list = 1
  [
    report 1
  ]
  report i
end

```

Fig. 4 Code segment for the h-index algorithm

agent-based algorithms. At times, we have observed that the native Logo code can be far more comprehensible as compared to traditional representations of code such as pseudo-code, flow charts or sequence diagrams.

Here, we can see that each researcher agent calls “hatch” to create paper agents. These agents are then given a uniform random num-cites with initial citations less than ten. These are then each inserted in a list of papers retrieved from the researcher agent. Subsequently, the list is updated again. Next, a link is created with the researcher agent, and the paper agents align themselves spatially on the sides of the researcher.

Next, we show the algorithm for the calculation of h-index in Fig. 4.

As we can see, the function starts out by first sorting the papers according to their citations. Subsequently, it performs a counting of the papers while also performing a comparison of the count with the current paper cites. If the index is greater than the paper, then one less than that is reported as the Hirsch index of the researcher. And if no exact match is found, then after the loop is completed, then the count is reported as the Hirsch index. An exception is when the number of papers is just one, when the number “1” is simply the Hirsch index.

Validation Experiments

Validation is one of the most important steps in the development of any model. Validation of agent-based models [36] can range from using collected data to ensure correlation with the real world to in-simulation validation such as using a virtual overlay multi-agent system [40].

McBurney and White [28] note that validation is of four basic types:

1. Internal validity
2. Construct validity
3. External validity
4. Statistical validity

Internal validation refers to having sound reasons to believe that a cause–effect relation is present between independent and dependent variables. Statistical validity is similar to internal validity but also checks to see whether the occurrence was not a pure chance. Construct validity refers to results supporting the theory behind the research, and external validity refers to whether the results can be generalized to another situation.

Thus, according to this classification, we perform internal, construct and statistical validation checks. However, we leave external validity for future work. Next, we show the validation of both the representational abilities of TCLNs and the calculation of the Hirsch index progression of researchers.

Validation of the Representational Abilities of TCLN Models

To validate the random researcher generation capabilities of the agent-based model, we show the creation of $n = 60$ random researchers in. Researchers are each placed on the Cartesian coordinate system, based on their Hirsch index. Each researcher, shown in black, has two labels. The first label shows the total number of published papers, and the second shows the Hirsch index. Each paper, shown in blue, has the total citations as a label. This experiment validates the representational ability of the TCLN modeling paradigm for researchers as shown in Fig. 5.

Validation of Hirsch Index Calculation

For a further validation of TCLNs as well as the validation of the calculations of the Hirsch index, we take the emergent cognitive development of a renowned researcher “Victor Lesser,” who is considered an authority in the domain of multi-agent systems. Using Publish or Perish [23], we queried Google Scholar and discovered firstly that there are a total of 649 papers listed. However, 546 of these papers have actual years properly indexed so we shall use these in our analysis of the evolution of the researcher. The first paper shows itself in the year 1968 so we start there till 1988. The detailed h-index data for 20 years for “Victor Lesser” has been plotted in Fig. 6.

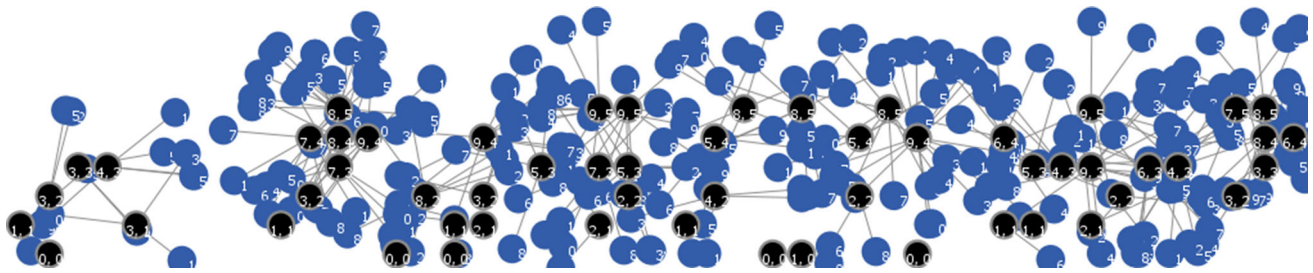


Fig. 5 Simulation view of $n = 60$ random researchers with $\text{max-init-papers} = 10$

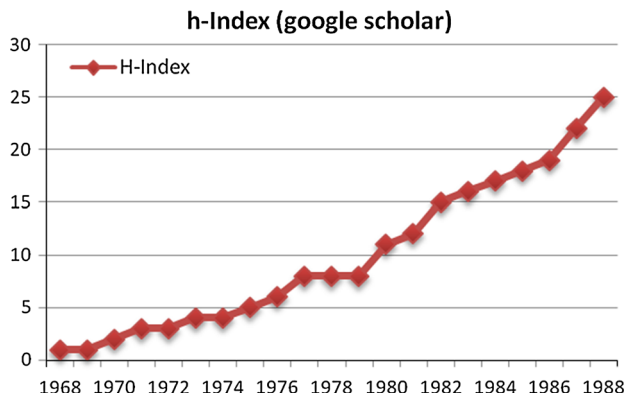


Fig. 6 h-Index plot for 20 years for “Victor Lesser” obtained using Publish or Perish program

Figure 7a, b shows 10 years of evolution of h-index of Victor Lesser as depicted using TCLN using the simulated agent-based model. The detailed results and the retrieved data (via Google Scholar index) depicted in the experiments are given in Table 1. In addition, we plot the number of nodes needed to be displayed in traditional citation networks versus TCLNs in Fig. 8.

Here, Table 1 shows two columns for the Hirsch index. One column shows the Hirsch index calculated using Google Scholar and the Publish or Perish program, and the other, the “Calc hI,” indicates the Hirsch index calculated by the algorithm alongside the evolution of the researcher. The table also shows the number of cited papers per researcher. Thus, we can see how we have validated both the calculation of the h-index and the effectiveness of TCLNs for the modeling of researcher reputation (state) coevolution with the changes in the topology using our agent-based model.

Discussion

In this section, we first discuss the different validation techniques used in the framework. Next, we note how the proposed framework can be used for evaluating cognitive development of researchers across disciplines. We also

discuss how the framework can be generalized for application in other domains.

Discussion of Validation

The first point to note here is the statistical validation of any network model. Let us first examine how statistical validation is typically performed. In statistical validation, the data are sampled and compared with well-known data sources based on standard statistical metrics. However, this type of validation does not guarantee validation of entire results. Instead the results only match averages or standard deviation. Often statistical validation entails removing the outliers.

However, when we develop a complex network model, we basically take the entire data set and develop the model around it. So, unlike statistical validation, this validation is completely truthful to the actual data since the network can simply be built only when every node and every link between the nodes have been developed and expanded in the form of the network. In complex network models, each and every data point is further treated as a node/agent and thus is represented visually/formally in the form of a graph/network. Therefore, any new measures for the measurement of any statistical data after the formation of this network would be an accurate and valid representation of the actual data.

Still, we would like to mention that the validity of any model can only be made as strong as the data it is validated against. So, for instance, if the citation indices such as Scholar/Web of Science have errors inside them, then the same errors would be reflected in any model, whether statistical or network based. However, since in the network models every node and link is of equal importance, such errors might show up with more clarity than models which aggregate the results and validate using statistical measures such as averages and curve fitting.

Generalization of the Framework

The proposed framework can be considered as a means of representing the cognitive development of researchers.

Fig. 7 Validation exercise 1: evolution of a researcher’s Hirsch index (Victor Lesser)

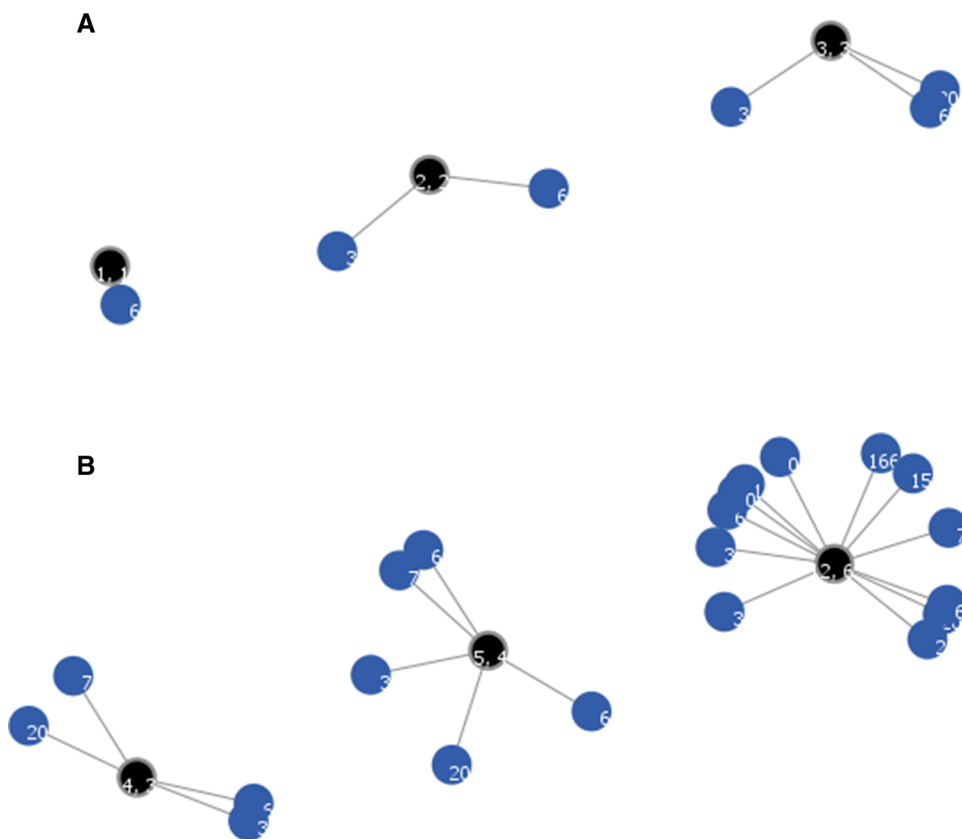


Table 1 Data for validation

Years	h-index	Citations of papers	Calc hI
1968	1	6	1
1969	1	6	1
1970	2	[6 3]	2
1971	3	[20 6 3]	3
1972	3	[20 7 6 3]	3
1973	4	[20 7 6 6 3]	4
1974	4	[20 7 6 6 3 2 0]	4
1975	5	[166 20 7 6 6 3 3 2 0 0]	5
1976	6	[166 21 20 15 7 6 6 3 3 2 0 0]	6
1977	8	[166 122 103 71 21 20 15 8 7 6 6 3 3 2 0 0 0]	8

Calc hI = Calculated h-index and h-index is the h-index from Publish or Perish software which obtains data via Google Scholar (Current as of 17 July, 2010)

Even though the framework has been demonstrated for the cognitive evolution of a Computer Science researcher, we have also shown in Sect. 5.2 that the particular network representation is also useful for any set of researchers from any disciplines because the agent-based model based on NetLogo demonstrates the application of the modeling framework for any random researchers. In other words, the framework is actually a generalized application of

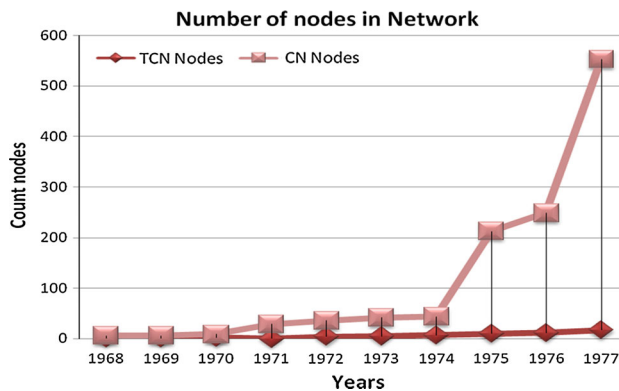


Fig. 8 Comparison of number of nodes needed to display an author-paper citation network versus a TCLN for Victor Lesser’s papers

computation techniques correlated with the Scientometric evolution of researchers aspiring from any discipline. Now, the interesting part here is that unless purely simulation-based models, the validation of the framework uses actual Scientometric data, available via Google Scholar, Scopus and Web of Science. As such, once we follow this particular modeling and simulation paradigm, it is trivial to model any researcher from any domain even in the same network. This would thus allow for the inter-disciplinary comparison of researchers using the proposed framework. Another possible generalization is by means of using a

citation index similar to the individual h-index. Individual h-index [5] allows for measuring the notability of the author but reduces the effects of co-authorships.

Proposed Extensions of the Framework

In its current form, the framework only offers a view of representing the cognitive mental development of researchers from the point of view of other researchers. However, this could easily be considered as one application of the cognitive modeling paradigm. If we were to take another index such as g-index for the same scenario, we might come up with more interesting results. Unlike the h-index, the g-index gives more weight to highly cited articles. Likewise Schreiber's method [44] of calculation of an index involves the use of fractional paper counts for multi-authored papers. The same modeling framework can be easily employed for any of these cases. In addition, looking at the actual contents of the papers that get cited will help to better understand cognitive and interaction dynamics between different researchers. To this end, we plan to apply semantic computing techniques [53] in order to analyze the content of each cited paper. In particular, an ensemble of semantic multi-dimensional scaling [54] and neural networks [55] will be employed to infer latent semantic connections between cited papers and, hence, to better model citation dynamics, e.g., by finding recurring citation patterns or by detecting papers that are actually cited as bad examples.

Conclusions and Future Work

In this paper, we have presented a formal modeling framework as well as the design and prototype implementation of an agent-based simulation model for TCLN. We have shown how the use of our proposed TCLN model reduces the order of complexity demonstrating the use of citations as well as the Hirsch index in the same citation network model. We have validated our agent-based model using a standard tool “Publish or Perish” by using data from the Google scholar. The retrieved data is used to evaluate the evolution of h-index over time for a renowned researcher in Computer Science. We also validated the representational ability of TCLN for random configurations using an agent-based model.

Although the validation study uses the growth of a Computer Science researcher using Hirsch index, it is possible to use the same framework for researchers in other domains. In addition, the TCLN representation may not be limited to the Hirsch index; future research is aimed at using the TCLN representation for modeling g-index as well as hc-index and others proposed by the research community.

Acknowledgments We would like to express thanks to Dr. Tamim Khan at Bahria University for taking time to verify the formal framework developed in the paper.

References

1. Aiello F, Fortino G, Gravina R, Guerrieri A. A java-based agent platform for programming wireless sensor networks. *Comput J*. 2011;54(3):439–54.
2. Amsterdamska O, Leydesdorff L. Citations: indicators of significance? *Scientometrics*. 1989;15(5):449–71. doi:10.1007/bf02017065.
3. Batagelj V. Efficient algorithms for citation network analysis. Arxiv preprint cs/0309023. 2003.
4. Batagelj V, Mrvar A. Pajek datasets. 2006. <http://vlado.fmf.uni-lj.si/pub/networks/data>.
5. Batista PD, Campiteli MG, Kinouchi O, Martinez AS. Is it possible to compare researchers with different scientific interests? *Scientometrics*. 2006;68(1):179–89.
6. Börner K, Maru J, Goldstone R. The simultaneous evolution of author and paper networks. *Proc Natl Acad Sci USA*. 2004;101(Suppl 1):5266.
7. Braun T, Glänzel W, Schubert A. A Hirsch-type index for journals. *Scientometrics*. 2006;69(1):169–73.
8. Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inf Sci Technol*. 2006;57(3):359–77.
9. Chen P, Redner S. Community structure of the physical review citation network. *J Informetr*. 2010;4(3):278–90.
10. Clauset A, Newman M, Moore C. Finding community structure in very large networks. *Phys Rev E*. 2004;70(6):66111.
11. Cronin B. The citation process. The role and significance of citations in scientific communication, vol. 1. London: Taylor Graham; 1984.
12. d’Inverno M, Luck M. Understanding agent systems. Berlin: Springer; 2004.
13. Egghe L. Theory and practise of the g-index. *Scientometrics*. 2006;69(1):131–52.
14. Egghe L. Dynamic h index: the Hirsch index in function of time. *J Am Soc Inf Sci Technol*. 2007;58(3):452–4.
15. Egghe L. The Hirsch-index and related impact measures. *Ann Rev Inf Sci Technol*. 2010;44:65–114.
16. Epstein J. Why model? *J Artif Soc Soc Simul*. 2008;11(4):12.
17. Falagas M, Pitsouni E, Malietzis G, Pappas G. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *FASEB J*. 2008;22(2):338.
18. Fortino G, Rango F, Russo W. Statecharts-based JADE agents and tools for engineering multi-agent systems. In: Setchi R, Jordanov I, Howlett RJ, Jain LC, editors. Knowledge-based and intelligent information and engineering systems. Berlin: Springer, Heidelberg; 2010. p. 240–250.
19. Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*. 1955;122(3159):108.
20. Garfield E. Citation analysis as a tool in journal evaluation. *Science*. 1972;178(4060):471–9.
21. Garfield E. The history and meaning of the journal impact factor. *JAMA*. 2006;295(1):90.
22. Harzing A, van der Wai R. Google Scholar as a new source for citation analysis. *Ethics Sci Environ Polit (ESEP)*. 2008;8(1):61–73.
23. Harzing A, van der Wal R. A Google Scholar h-index for journals: an alternative metric to measure journal impact in

- economics and business. *J Am Soc Inf Sci Technol.* 2009;60(1): 41–6.
24. Hayden S, Zermelo E, Fraenkel A, Kennison J. Zermelo-Fraenkel set theory. Columbus: CE Merrill; 1968.
 25. Hirsch J. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci.* 2005;102(46):16569.
 26. Hummon N, Dereian P. Connectivity in a citation network: the development of DNA theory. *Soc Netw.* 1989;11(1):39–63.
 27. Kajikawa Y, Takeda Y. Citation network analysis of organic LEDs. *Technol Forecast Soc Chang.* 2009;76(8):1115–23.
 28. McBurney DH, White T. Research methods. New York/Boston: Pearson/Wadsworth; 2006.
 29. Moed H. Citation analysis in research evaluation. Dordrecht: Kluwer; 2005.
 30. Newman M. Coauthorship networks and patterns of scientific collaboration. *Proc Natl Acad Sci.* 2004;101(Suppl 1):5200.
 31. Newman MEJ. The structure and function of complex networks. *SIAM Rev.* 2003;45(2):167–256.
 32. Niazi M, Ahmed HF, Ali A. Introducing fault-tolerance and responsiveness in web applications using SREFTIA. In: Paper presented at the proceedings of the international multicongress on computer science and information technology, Wisla, Poland, Nov 6–10, 2006. 2006.
 33. Niazi M, Hussain A. A novel agent-based simulation framework for sensing in complex adaptive environments. *IEEE Sens J.* 2010;11(2):404–12.
 34. Niazi M, Hussain A. Agent-based computing from multi-agent systems to agent-based models: a visual survey. *Scientometrics.* 2011; 1–21. doi:10.1007/s11192-011-0468-9.
 35. Niazi M, Hussain A, Baig AR, Bhatti S. Simulation of the research process. In: Paper presented at the 40th conference on winter simulation, Miami, FL. 2008.
 36. Niazi M, Siddique Q, Hussain A, Kolberg M. Verification and validation of an agent-based forest fire simulation model. In: Paper presented at the SCS spring simulation conference, Orlando, FL, USA, April 2010. 2010.
 37. Niazi MA. Self-organized customized content delivery architecture for ambient assisted environments. Paper presented at the Proceedings of the third international workshop on Use of P2P, grid and agents for the development of content networks, Boston, MA, USA. 2008.
 38. Niazi MA. Complex adaptive systems modeling: a multidisciplinary roadmap. *Complex Adapt Syst Model.* 2013;1(1):1.
 39. Niazi MA, Hussain A. Cognitive agent-based computing-I: a unified framework for modeling complex adaptive systems using agent-based and complex network-based methods. *Springer Briefs in Cognitive Computation.* Springer: Dordrecht; 2012. doi: 10.1007/978-94-007-3852-2.
 40. Niazi MA, Hussain A, Kolberg M. Verification and validation of agent based simulations using the VOMAS (virtual overlay multi-agent system) approach. Paper presented at the MAS&S 09 at Multi-Agent Logics, Languages, and Organisations Federated Workshops, Torino, Italy, 7–10 September 2009. 2009.
 41. Niazi MA, Laghari S. An intelligent self-organizing power-saving architecture: an agent-based approach. In: Computational intelligence, modelling and simulation (CIMSIM). IEEE fourth international conference on 2012. 2012. p 70–75.
 42. Niazi MA, Siddique Q, Hussain A, Fortino G. SimConnector: An approach to testing disaster-alerting systems using agent-based simulation models. Paper presented at the Federated conference on computer science and information systems, Szczecin, Poland. 2011.
 43. Niazi MAK. Towards a novel unified framework for developing formal, network and validated agent-based simulation models of complex adaptive systems. Stirling: University of Stirling; 2011.
 44. Schreiber M. A modification of the h-index: the hm-index accounts for multi-authored manuscripts. *J Informetr.* 2008;2(3): 211–6. doi:10.1016/j.joi.2008.05.001.
 45. Siddiq A, Niazi MA, Mustafa F, Bokhari H, Hussain A, Akram N, Shaheen S, Ahmed F, Iqbal S. A new hybrid agent-based modeling and simulation decision support system for breast cancer data analysis. In: Information and Communication Technologies, 2009. ICICT '09 international conference on 15–16 Aug 2009. 2009. p 134–39.
 46. Spivey JM. Understanding Z: a specification language and its formal semantics. Cambridge: Cambridge Univ Press; 1988.
 47. Team N. Network workbench tool. USA: Indiana University/Northeastern University/University of Michigan; 2006.
 48. Watts C, Gilbert N. Does cumulative advantage affect collective learning in science? An agent-based simulation. *Scientometrics.* 2011;89(1):437–63. doi:10.1007/s11192-011-0432-8.
 49. Weingart P. Impact of bibliometrics upon the science system: inadvertent consequences? *Scientometrics.* 2005;62(1):117–31.
 50. Weng J, McClelland J, Pentland A, Sporns O, Stockman I, Sur M, Thelen E. Autonomous mental development by robots and animals. *Science.* 2001;291(5504):599–600. doi:10.1126/science.291.5504.599.
 51. Clapham PJ. Publish or Perish. *Bioscience.* 2005;55(5):390–91.
 52. Wilensky U. NetLogo. Evanston, IL: Center for Connected Learning Comp.-Based Modeling, Northwestern University; 1999.
 53. Cambria E, Hussain A. Sentic computing: techniques, tools, and applications. Dordrecht, Netherlands: Springer; 2012. ISBN: 978-94-007-5069-2.
 54. Cambria E, Song Y, Wang H, Howard N. Semantic multi-dimensional scaling for open-domain sentiment analysis. *IEEE Intell Syst.* 2013. doi:10.1109/MIS.2012.118.
 55. Cambria E, Mazzocco T, Hussain A. Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining. *Biol Inspir Cogn Arch.* 2013;4:41–53.