Collective intelligence: Analysis and modelling

Erika Suárez Valencia
Universidad del Valle
Cali, Colombia
erika.suarez@correounivalle.edu.co

Víctor Andres Bucheli Guerrero
Universidad del Valle
Cali, Colombia
victor.bucheli@correounivalle.edu.co

Roberto Zarama
Universidad de los Andes
Bogotá, Colombia
rzarama@uniandes.edu.co

Ángel García Baños
Universidad del Valle
Cali, Colombia
angel.garcia@correounivalle.edu.co

Abstract

Purpose - This paper focuses on the underpinning dynamics that explain collective intelligence.

Design/methodology/approach - Collective intelligence can be understood as the capacity of a collective system to evolve towards higher order complexity through networks of individual capacities. We observed two collective systems as examples of the dynamic processes of complex networks—the wiki course PeSO at the Universidad de Los Andes, Bogotá, Colombia, and an agent-based model inspired by wiki systems.

Findings - The results of the wiki course PeSO and the model are contrasted with a random network baseline model. Both the wiki course and the model show dynamics of accumulation, in which statistical properties of non-equilibrium networks appear.

Research limitations/implications - Our work is based on network science. We analyzed data from two kinds of networks: the wiki course PeSO and an agent-based model. Limitations due to the number of computations and complexity appeared when there was a high order of magnitude of agents.

Practical implications - Better understanding can allow for the measurement and design of systems based on collective intelligence.

Social implications - In the context of higher education systems, the question remains of how to design teaching methodologies to develop collective thinking.

Originality/value - The results show how collective intelligence emerges from cumulative dynamics.

Keywords Knowledge management; Collective intelligence; Agent-based models; Network science.

Paper type Research paper

1. Introduction

The production of knowledge in contemporary societies has experienced a phase of transition away from individualism towards collectivism. For instance, the science of past centuries was characterized by eponymy (Merton, 1968), a situation that brought about some famous first author disputes: Newton versus Leibnitz on differential equations or Darwin versus Wallace on evolutionary theory. In these cases, discussions about and between first authors were frequent and open. Today, acknowledgement of the individual scientist could be not so important.

The Internet has allowed for new forms of social interaction and organization. In this context, the aim of this paper is to understand new forms of knowledge production. In the new socio-technical system, ideas flow and interchanges grow over time in a way that has not been seen before. Along with this expansion have come questions about plagiarism and other forms of misconduct (although we do not develop on this topic here). Furthermore, collective behavior on the Internet has given rise

to projects such as Wikipedia, a collectively developed encyclopedia which is the most complete in existence, and crowdsourcing, where many ideas from many participants are organized together to solve complex problems. Such examples show how alternative forms of organization around knowledge may appear. In these new forms of organization, interactions and interchanges grow, and the probability of new outcomes emerges.

In this paper, we study collective intelligence through two complementary approaches. On the one hand, we analyze the structural properties of networks obtained from data from the Wiki course PeSO at the Universidad de Los Andes, Bogotá, Colombia. On the other hand, a mathematical model of collective intelligence based on wiki systems is proposed. This work is based on networks, specifically equilibrium and non-equilibrium networks (Dorogovtsev and Mendes, 2013).

The results show similar behavior for both the model and the PeSO wiki course. The model reproduces the curves of clustering coefficient and average path length over time in a similar way to small world networks. The empirical networks (the model and the PeSO wiki course) are compared with a random baseline. The results are consistent, revealing small world network properties as a to study collective intelligence. We observe collective intelligence as a complex adaptive system. Thus, the measurement of collective intelligence can be based on the differences between a random process and self-organized criticality (Tang and Bak, 1988; R Chialvo, 2004; Sornette, 2006; Barrat *et al.*, 2008; Dorogovtsev and Mendes, 2013).

This document is organized into four sections. The first section presents our motivation for measuring collective intelligence. The second section describes the empirical analysis and the proposed model. The third section shows the results. The final section presents the discussion and proposals for future work.

2. Motivation: Towards the measurement of collective intelligence

2.1 Motivation

We assume that many people nowadays are interconnected via the Internet, and that the resulting interactions and networks allow for the development of projects of collective intelligence. In fact, around the world there are $2.8e10^9$ interconnected people producing information, outcomes and knowledge, such as Linux (operating system), Wikipedia (open encyclopedia), Open Government (in the US), crowdfunding (funding network) and crowdsourcing (networks to solve complex problems based on knowledge).

Collective intelligence can be understood as the capacity of a group of people to collaborate in order to achieve goals in a complex context (Heylighen, 2013). Collective intelligence is distributed within a network where each interaction continually aggregates value. It is coordinated in real time, developed through the effective mobilization and reciprocity of competencies (Lévy, 1994). In addition, collective intelligence can be seen as the capacity of a human community to evolve towards higher order complexity thought, problem solving and integration through collaboration and innovation (Pór, 1995). In this paper, we propose an operative definition of collective intelligence, based on which it is possible to compute measurements and modeling. Thus, collective intelligence is defined—more or less successfully—as the capacity of a collective system to evolve towards higher order complexity through networks of individual capacities.

2.2 Research proposal

We observed the collective production of knowledge over time by building up a co-authorship network through the Wiki-ITRB (http://wiki.uniandes.edu.co/PESO/tiki-index.php). The Wiki-ITRB is one activity in the course 'Organizational System Thinking' or PeSO (its Spanish acronym), offered at the Universidad de Los Andes, Bogotá, Colombia. The activity was developed between 2011 and 2012.

The purpose of the Wiki-ITRB is to collectively write ITRB (Informe Técnico de Revisión Bibliográfica—Technical Reports of Literature Reviews) documents. ITRB documents propose one question for a given topic, and students then include arguments, author positions and opinions about the proposed question. The activity aims to encourage students to acquire the competencies to write argumentative documents. Based on the theory of collective intelligence, we designed a collaborative document schema via a wiki platform. Students participated in the writing and modification of several documents, with references, arguments, corrections, etc. Each student could promote, eliminate and/or edit a text or document. Finally, each student decided to be the author of a subset of documents, which she or he edited and evaluated.

The participation of students in the Wiki-ITRB is stored up over time; this allows for the building up of a network from the aggregation of connected authors via co-authored documents. We constructed a dynamic network through the extraction of subnets of documents over time. We were therefore able to evaluate the dynamics of structural network properties.

3. Empirical analysis and modeling

Classical random networks (Erdős and Rényi, 1959) are constructed with connections between randomly selected pairs of vertices. By contrast, small world networks are characterized as being highly clustered, and small path lengths (Watts and Strogatz, 1998). For the networks in our study, we computed clustering coefficient and average path length. Both have behaviors that are totally different in equilibrium networks (random networks) and non-equilibrium networks (Dorogovtsev and Mendes, 2013). We assume that the mechanisms underpinning the networks of wiki systems are not random, but are rather mechanisms that self-regulate collective production.

In this paper, we study the network of co-authored Wiki-ITRB pages based on clustering coefficient and average path length. At the same time, the proposed agent-based model was studied using the same measurements. The wiki systems were therefore studied based on independent structural features. Random graphs, built according to the Erdős–Rényi (ER) model, exhibit a small average shortest path length (varying typically as the logarithm of the number of nodes) along with a small clustering coefficient. Small world models have a small average shortest path length, but at the same time a clustering coefficient significantly higher than expected for a random model.

We propose an agent-based model to understand collective intelligence in a socio-technical system. This is a model organized by a non-linear combination of agents (Wolfram, 2002; Flake, 1998) Thus we propose that collective behavior can be modeled as non-linear relations among editors. In this paper, we focus on the study of non-equilibrium networks and their structural properties as a measurement of collective intelligence, as explained above.

3.1 Agent-based modeling

The aim of the agent-based model is to understand the evolution of wiki systems in order to gain a better understanding of collective intelligence. The agents are people and documents, where people have an agent edition capacity, which indicates how many documents they can edit (and not the number of modifications they are able to do in a single document). Documents have a probability of being selected, and in terms of the accumulation of total edits, this affects (in a similar way to votes) their probability of being selected in the next iteration of the model.

The parameters analyzed are the number of agents, agent edition capacity and simulation time (represented as steps in the execution of the model). For each parameter, one network of co-author editions was constructed and measurements of clustering coefficient and average path length obtained.

Our study of collective intelligence is made through the accumulation of editions for each document and its influence on the documents' probability of being selected and taken into account for future edition by an agent. Thus, the evolution of a network of co-author editions on the basis of previous editions is presented. The model's reinforcement loops perform in a similar way to other complex systems, such as brains, ant colonies, etc. (Wolfram, 2002; Flake, 1998). In this sense, agent

edition capacity is like the computing capacity of a node that belongs to a network that presents behavior more intelligent than each individual node.

3.1.1 Assumptions

- Agent edition capacity is a natural value and all agents have the same capacity; for instance, when agent edition capacity = 2, this means that one agent can edit 2 documents.
- Each agent edits documents according to their agent edition capacity; the greater their edition capacity, the more documents they can edit.
- The agent selected at each step is chosen in a uniformly random way.
- The edition of documents positively affects their probability of selection in the future. Therefore, documents with more editions are more likely to be edited again in the next round.

3.1.2 Description of the model

The model produces a network of agents or a co-author editions network, where an edge connects two agents who have made editions to the same document. The network is constructed as follows:

```
total-editions = 0

for i = 1 to t
  iteration

link all agents that edited same document
```

Pseudocode for one iteration:

```
agent = choose one random agent

for i = 1 to k
  document = choose one document based on probability
  add agent to document's list of editors
  increment by 1 editions in document
  increment by 1 total editions

for each document in documents
  probability = (editions + 1) / (total-editions + total-docs)
```

k = agent edition capacity and t = simulation time.

The documents' probabilities are updated as follows: at the beginning of the simulation, every document has the same probability, 1 per total number of documents. After each iteration, as can be seen in the pseudocode, the probability of one document is calculated taking into account the edition made to it and the total editions made in the system.

3.1.3 Experimental design

The probability of connection between two agents, given a determined number of agents, depends on the simulation time and the agent edition capacity. We observed the structural properties of the co-author editions network according to three assigned parameters: total agents, time simulation and agent edition capacity.

Each parameter was evaluated as follows: total agents between 10^1 and 10^3. For each number of agents, the time simulation was evaluated from 2 times to 10 times the number of agents. Agent edition capacity was evaluated from 1 to 10. Each simulation was run 80 times, thus the measurements presented below correspond to the average over 80 simulations.

Experiments were performed in NetLogo 5.0.5 (Wilensky, 1999) with an implemented extension to export the resulting graph to graph6 format (https://github.com/erikasv/NetLogo-graph6), and the analysis was performed in Mathematica 9 (Wolfram Research, Inc., 2012).

4. Results

Aside from the produced results, the model was developed as a framework to study networks constructed from interaction rules at a micro level. It allows for the study of collective intelligence based on a network science approach.
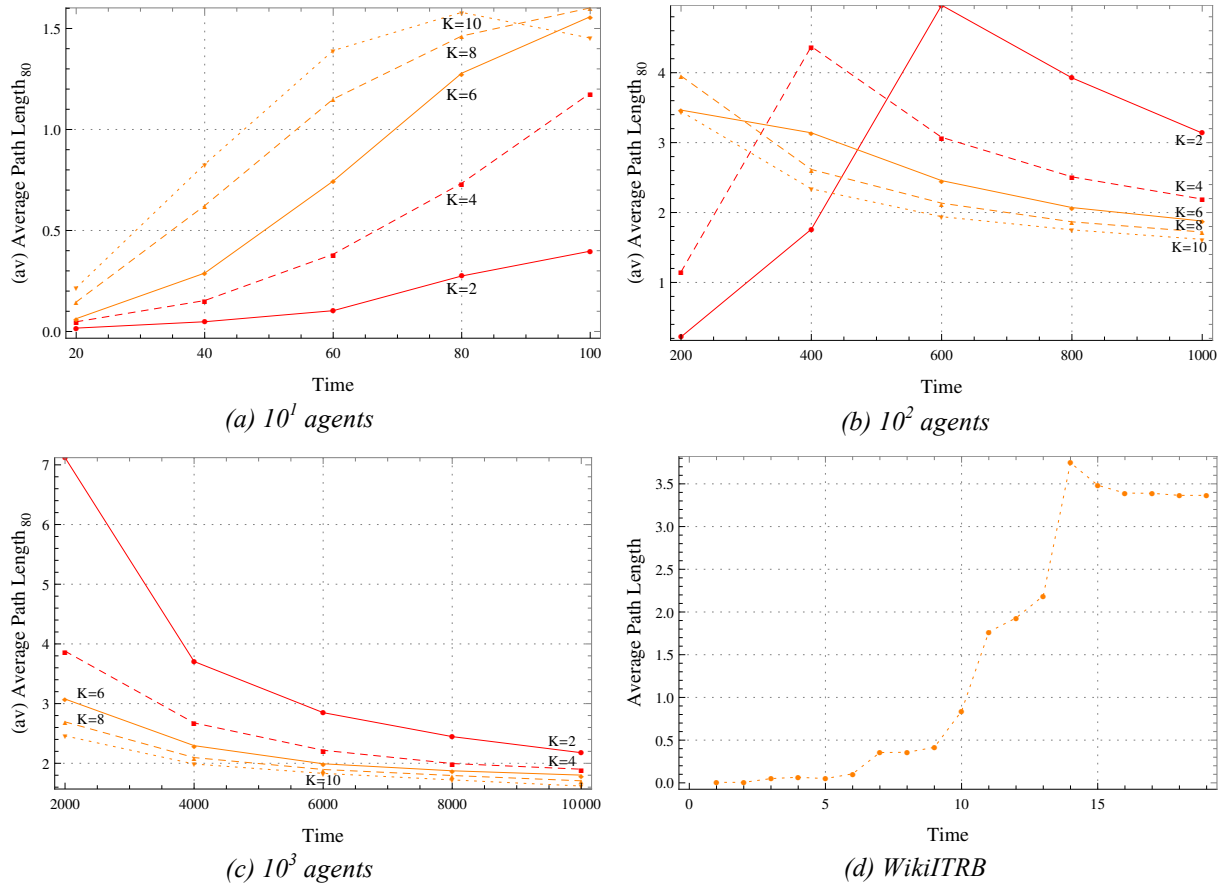
Figure 1 shows the clustering coefficient for each number of agents and the Wiki-ITRB. In a, b and c, the results show how the curve of the average clustering coefficient evolved over 80 simulations. Figure 1 d. shows the dynamic of higher values of clustering coefficient for the Wiki-ITRB. The clustering coefficient in wiki systems (model and Wiki-ITRB) demonstrates the same behavior and is consistent with the clustering coefficient in small world networks.
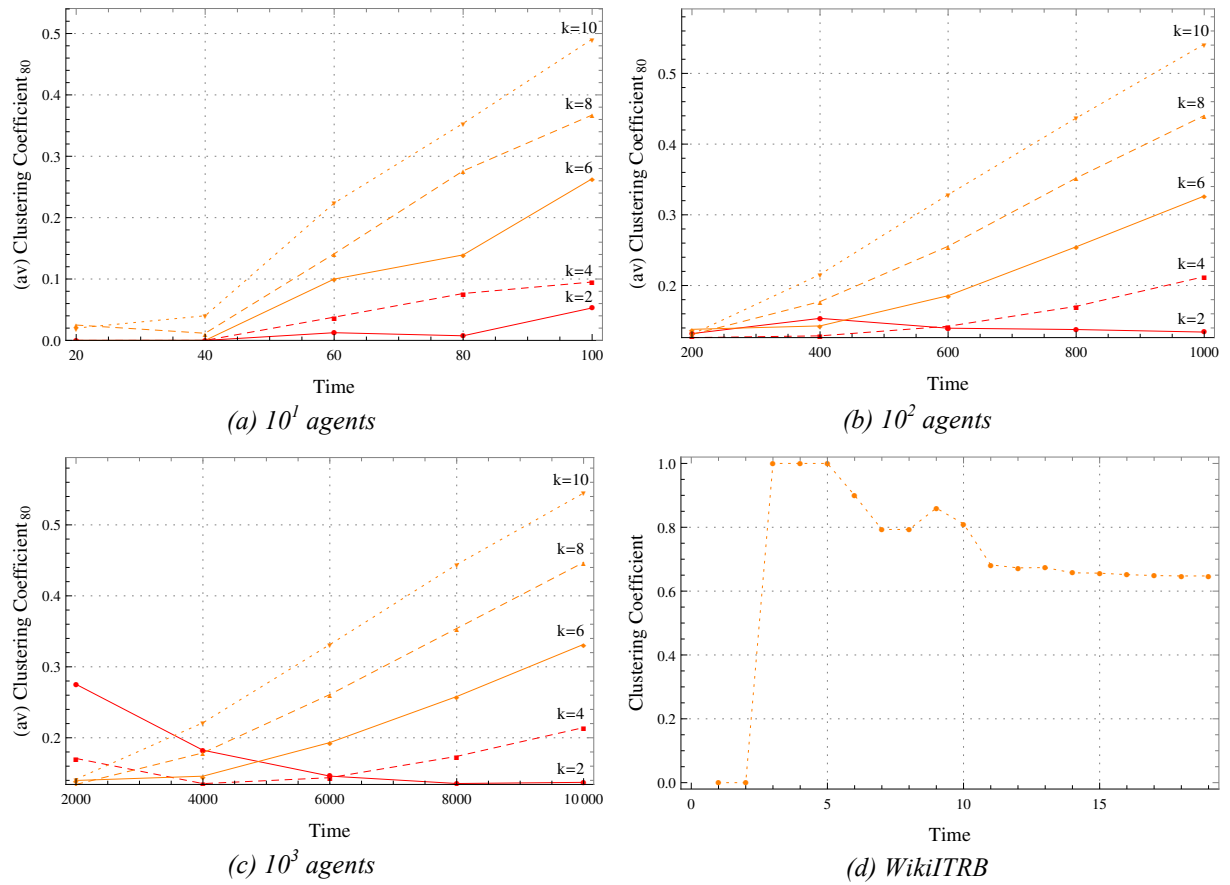

Figure 1. Values of the clustering coefficient through time units. Figures a, b and c correspond to model executions with 10^1, 10^2 and 10^3 agents respectively, and Figure d corresponds to the Wiki-ITRB. In Figure 1, along the x axis are time units and the y axis shows the clustering coefficient. Figures a, b and c shows the clustering coefficient of the resulting network after running the model. Each line in the graphic corresponds to one value of agent edition capacity (k).

Figure 1 shows that when both time (t) and k increase, the clustering coefficient also increases. Regardless of the number of agents, all graphics resulting from the model are very similar to the one of the Wiki-ITRB. While k increases, there is monotonic growth over time; however, this behavior is not present for low values of k. Thus, when k >= 3 and t >= 5 times the total number of agents, behavior is expected to be consistent. This shows how the values of individual edition capacity (k) and the time simulations (t) are relevant in the design of measurements or design systems based on collective intelligence. The implications of this result are discussed below.

Figure 2 shows the average path length (apl) for each number of agents and the Wiki-ITRB. In a, b and c, the results show how the curve of the average apl over 80 simulations evolved for the model and Figure 2 d. for the Wiki-ITRB. The apl in wiki systems demonstrates the same behavior. The results showed in Figure 1 and 2 are consistent with (Ingawale et al., 2009).

Figure 2. Values for average path length through time units in model executions with 10^1 (a), 10^2 (b) and 10^3 (c) agents. In this figure, average path length is represented along the y axis. Contrary to the clustering coefficient, the average path length decreases when t and k increase, which is consistent with the behavior of small world networks.
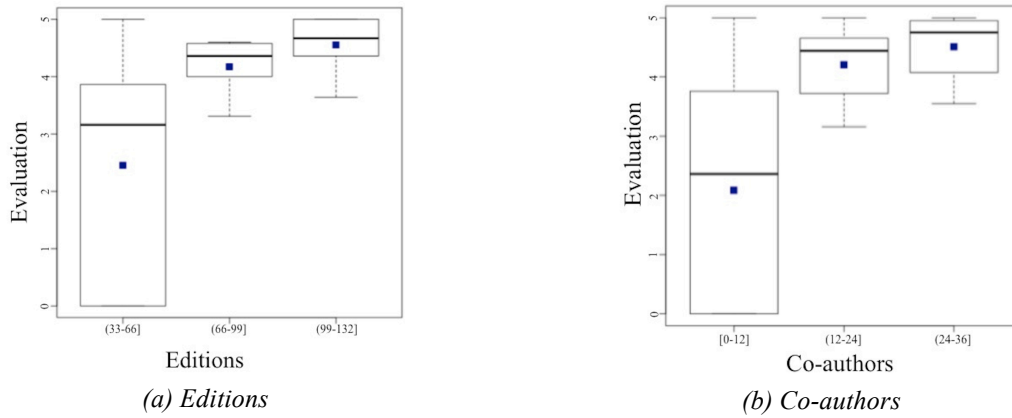
Figure 3. Values of evaluation (scale 1-5) (a), average evaluation and standard deviation of evaluation documents by bins of number of editions. In a similar way, b shows the average evaluation and standard deviation of evaluation documents by bins of number of co-authors of wiki pages.

In Figures 1 and 2 it can be seen that the clustering coefficient and average path length are saturated by high time values. The time for simulations for higher values therefore implies that a lot of agents have connections and the small world structure disappears. In-depth study of this behavior and the umbrals for the measurements is proposed for future work.

Figure 3 shows the evaluation of each document in the Wiki-ITRB, and demonstrates how the documents with more editions or more co-authors receive a better evaluation (scale 1-5). These results suggest that individual edition capacity and time are relevant for the acquisition of emergent properties such as those of small world networks, where group composition is self-organized. At the same time, the accumulation of a number of editions has an effect on the qualifications of documents. In sum, it suggests that collective intelligence is related to the accumulation dynamics of editions, thus with better documents there are more editions and more agents working on them. This constrains the evolution of co-editor networks and the structural properties of small world networks appear.

5. Discussion and future work

Collective intelligence can be understood as the capacity of a collective system to evolve towards higher order complexity through networks of individual capacities. We observed two collective systems in terms of a dynamic process in complex networks—the Wiki course PeSO and an agent-based model based on wiki systems. The results from both the course and the model were contrasted with a random network baseline model. Both the course and the model show dynamics of accumulation, in which statistical properties of non-equilibrium networks appear. The proposed model reproduces the behavior observed in the PeSO course; this behavior is also described for small world networks (Watts and Strogatz, 1998). From this, we interpret that collective intelligence emerges from cumulative dynamics.

Two measurements have been observed: clustering coefficient and average path length. Both had consistent values in terms of individual edition capacity (k) and time units (t), where neither could be too large or too small. As the results show, when values are too large, the outcome of the simulation are complete graphs, and when values are too small, the graphs are not connected, which means that either it is a random system or it is too simple. For this set of values, however, there is enough complexity to replicate phenomena observed in real systems. Thus, we show how the process of accumulation of editions can be seen as being self-organized system.

The results presented here demonstrate how collective intelligence emerges from cumulative dynamics. This provides a better understanding of how to measure and design systems based on collective intelligence. In the context of higher education systems, one example of this is Wiki-ITRB activity (the PeSO wiki course); however, it would be necessary to develop a better understanding of collective intelligence in order to implement this strategy. An open question for further investigation is how teaching methodologies in higher education can be designed to develop collective thinking.

Intelligence is hard to define in a rigorous way, but it is related to the perception, adaptation and even modification of an environment, for the purpose of survival and reproduction (Dawkins, 1986). Systems that are completely organized and where nothing changes (or that only follow deterministic rules) are rigid; they cannot adapt to differing or complex environments. On the other hand, systems that are completely random have, by definition, no memory, thus the system cannot learn from similar past situations and react appropriately. In an intermediate point are chaotic systems, which can adapt better to extreme events (Langton, 1990) and can generate self-organized structures. In (Kauffman, 2000) it is mentioned that the complexity of the universe is due to the fact that it is not ergodic, i.e. all possible configurations have not been explored. This could explain the loss of the small world phenomenon in our study when the time units in the simulation were too large. At the same time, individual capacity cannot be too low (values of k < 3) otherwise the system becomes quite similar to a random network. The study of these umbrals is proposed for future work.

We understand wiki systems as resulting from a cumulative process, whereby the accumulation of editions goes towards the development of wiki pages. Thus, the more editions there are, the better the wiki page (more visible, more votes and/or more edited); furthermore, the more editions a wiki page receives, the more editions it is likely to receive in the future. This reinforcement cycle of the wiki system transforms a random network into a small world network of co-authors or co-editors. The model presented here is thus an accumulative system, where there is no loss of information (no loss of nodes or edges). It is proposed that in future work, a less accumulative system should be examined.

To better understand collective intelligence, we propose that future work focuses on the in-depth study of the proposed model: distribution of agent edition capacity (k) and the computation of other measurements such as Small-World Characteristic Q and robustness.

Acknowledgements

References

Barrat, A, Barthelemy, M & Vespignani, A 2008, *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge.

Dawkins, R 1986, *The blind watchmaker*, New York, Norton.

Dorogovtsev, SN & Mendes, JF 2013, *Evolution of networks: From biological nets to the Internet and WWW*, Oxford University Press, New York, NY.

Erdős, P & Rényi, A 1959, 'On random graphs I', *Publicationes Mathematicae Debrecen*, vol 6, pp. 290-297.

Flake, GW 1998, *The computational beauty of nature: computer explorations of fractals, chaos, complex systems, and adaptation*, MIT Press, Cambridge.

Heylighen, F 2013, 'Self-organization in Communicating Groups: the emergence of coordination, shared references and collective intelligence', Springer Berlin Heidelberg, Berlin.

Ingawale, M, Dutta, A, Roy, R & Seetharaman, P 2009, 'Small worlds of Wikipedia: Implications for growth quality and sustainability of collaborative knowledge networks', AMCIS, vol 1, no. 1, p. 439.

Kauffman, SA 2000, *Investigations*, Oxford University Press, USA.

Langton, CG 1990, 'Computation at the Edge of Chaos: Phase Transitions and Emergent Computation', *ninth annual international conference of the Center for Nonlinear Studies on Self-organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks on Emergent computation*, Phys. D In CNLS.

Lévy, P 1994, *L'intelligence collective: pour une anthropologie du cyberspace*, La Découverte, Paris.

Merton, RK 1968, 'The Matthew effect in science', *Science*, vol 159, no. 3810, pp. 56--63.

Pór, G 1995, 'The quest for collective intelligence', *Community building: Renewing spirit and learning in business*, pp. 271-280.

R Chialvo, D 2004, 'Critical brain networks', *Physica A: Statistical Mechanics and its Applications*, vol 340, no. 4, pp. 756-765.

Sornette, D 2006, *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*, Springer Science & Business, Berlin.

Tang, C & Bak, P 1988, 'Critical exponents and scaling relations for self-organized critical phenomena', *Physical Review Letters*, vol 60, no. 23, pp. 2347–2350.

Watts, DJ & Strogatz, SH 1998, 'Collective dynamics of'small-world'networks', *Nature*, vol 393, no. 6684, pp. 409-10.

Wilensky, U 1999, *NetLogo*, http://ccl.northwestern.edu/netlogo, Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

Wolfram Research, Inc. 2012, *Mathematica*, 90th edn, Wolfram Research, Inc. , Champaign, Illinois.

Wolfram, S 2002, *A New Kind of Science*, Wolfram Media, Champaign, IL.