COLLECTIVE INTELLIGENCE PROCESSES AND THEIR INFLUENCE ON THE DYNAMICS OF INFORMATION DIFFUSION ON THE WEB

Leonardo Caporarello and Luca Ongaro
Learning Lab, SDA Bocconi, Bocconi University, Via Bocconi 8, 20136 Milan, Italy

Keywords: Collective intelligence, Self-organization, Internet, Social web, Information network, Information overload.

Abstract: The so-called Social Web produced a paradigm shift toward a distributed model of information production and diffusion, as well as an exponential growth of the total amount of information circulating in digital format. The environment in which this information diffuses is a distributed global network of individuals and organizations acting at the same time as information producers and consumers. The emergent dynamics of self-organisation of this network fall into the realm of collective intelligence, and their study is essential in order to understand their influence on the pattern of information diffusion and cultural development.

1 INTRODUCTION

The evolution of the Internet toward the so-called Social Web (Gruber, 2008) produced several well-known important implications on human cultural development and on access to information (O'Reilly, 2005). The number of weak social ties and information sources that an individual can manage dramatically increased, as connections are no more geographically bounded. Moreover, the easiness of information transfer on the Web is continuously rising up. Also, new technologies and pattern of use of the Web increasingly allowed for a gradual change of paradigm toward a distributed production of information, with no formal distinction between producers and consumers of information.

Thus, the circulation of ideas at a global level, and the possibility to contribute to the creation and management of knowledge in a distributed fashion is determining the astonishing rate of Web content production. The global amount of information produced and stored in digital formats encountered a tenfold increase in the last ten years, with a +60% year-to-year growth rate (Gantz et al., 2007). In absence of a mechanism of information filtering, we could expect the problem of selecting the relevant information, and discriminating it from "noise", to become increasingly complex. This problem is often referred to with the expression "information overload". Some authors propose that the exposure to a large amount of diverse information on the Web, with varying degree of relevance, do produce some measurable negative effects on the ability to concentrate on tasks and on the exercise of deep cognitive thought in the consumption of information (Carr, 2008).

Some research evidences show that Internet users are growing at a rate of 100 millions new users per year (ITU Database, 2010), and that people are spending more and more time on-line. This leads us to think that the perceived benefit associated to the social Web outweighs the cost introduced by problems related to information overload. Due to the decentralized nature of the Web, any mechanism that mitigates the issue of information overload is expected to come from a distributed effort.

According to the above considerations, our research question is whether emergent and non-supervised distributed mechanisms, falling within the realm of collective intelligence (Bonabeau, 2009), effectively influence the dynamics of information diffusion on the Web, balancing out information overload. Moreover, if such a mechanism is found, we intend to identify its main implications on cultural development at both macro and micro levels.

Caporarello L. and Ongaro L..
COLLECTIVE INTELLIGENCE PROCESSES AND THEIR INFLUENCE ON THE DYNAMICS OF INFORMATION DIFFUSION ON THE WEB.
DOI: 10.5220/0003445905010504
In *Proceedings of the 1st International Conference on Cloud Computing and Services Science* (CLOSER-2011), pages 501-504
ISBN: 978-989-8425-52-2

2 THEORETICAL BACKGROUND

In order to investigate the research question, a theoretical background on which to base our analysis is needed. In particular, it is necessary to provide:

- A formal definition of self-organizing processes, emergent phenomena and collective intelligence
- A model explaining the nature of information and the dynamics of distributed information production and cultural development
- A model capable of effectively describe the dynamics of information diffusion on the Internet

A brief discussion introducing such definition and models is thus required to elaborate on the research question.

2.1 A Definition of Collective Intelligence

Collective intelligence is a cognitive process emerging at a systemic level from the interaction at the micro-level between many distributed agents (Levy, 1997). It represents a subset of the broader concepts of emergent phenomena and mass behavior, both involving surprising ordered patterns produced by distributed agent-based systems, generally non predictable from the study of a single agent. Collective intelligence is thus an emergent phenomenon involving some kind of decision-making happening at the system level.

There are numerous examples of collective intelligence in nature, such as the foraging behavior of ant colonies or the complex schooling behavior of fish without an individual leader. Collective intelligence processes are common in human societies too, an example of primary importance being the market, where the price and allocation of goods and assets is not the outcome of a central decision, but rather of the combined judgements of many independent individuals.

2.2 Memetics: A Model of Distributed Cultural Development

Memetics is a theory aimed at explaining how distributed and generally non-supervised social systems can produce and develop ideas, culture and also more trivial phenomena like fashion and trends. It was first proposed by Richard Dawkins in the final chapters of the book The Selfish Gene (1976), and subsequently developed by other authors (i.e. Dennet, 1992). The basic idea is to apply the theory of evolution to culture, postulating the existence of stable units of information, called *memes*, which play a role analogue to the one of genes in biological evolution.

In order to trigger an evolutionary process, an environment needs to provide at least a mechanism of replication capable of producing copies of entities, a mechanism of variety generation producing entities with novel characteristics, and a process of selection increasing the chances of replication of entities that better satisfy a "fitness" requirement. Memetics identify the entities being replicated with *memes* and the replication mechanism with human communication: each time a transfer of information happens between two people, by means of explicit communication or by observation of actions or behavior, ideas and their building blocks (the *memes*) are replicated, so that at the end of the process new copies of the original ideas are created and hosted in the brain of the information receiver. Imperfections of communication, subjective interpretation of ideas and recombination with other concepts all create variation in the memetic pool. Finally, a selection mechanism is introduced by the fact that some ideas are more likely to be spread than others. The "fitness" criteria of *memes* is not necessarily bounded to advantages for the humans that carry them, but still ideas that provide a benefit for societies that adopt them are more likely to be preserved and spread, while those that are harmful tend to be discarded in the long run.

Memetics makes an attempt to explain how complex cultural systems can arise from the interaction of large communities of humans, and the theory can be easily applied to cultural development in the Internet era. The highly codified nature of information on the Web and the fact that it can be efficiently copied, stored and transferred make the metaphors introduced by Memetics even more applicable.

3 THE INFORMATION NETWORK: A MODEL OF INFORMATION DIFFUSION ON THE WEB

The Web is essentially a networked structure of resources and hyperlinks between them, and as such

can be easily represented as a directed graph, the pages being the vertices and the links being the directed edges. The evolution of the Internet as a more and more social phenomena, with sites and content often representing the direct emanation of individuals or organizations being at the same time producers and consumers of information, enables us to apply the directed graph metaphor at a higher level of analysis. The social Web, in fact, makes information production and publication widely accessible, and widespread technologies like bookmarking of sites, RSS feeds, social networking applications and blogging platforms make it easy to keep track of interesting sources of information. Thus, it is possible to model the social Web as a directed graph in which the vertices or nodes are the people and organizations producing and consuming information (to which we will generally refer as agents), while the arcs or directed links express the relationship "A follows B" or "B is a source of information for A", A and B being two nodes.

The act of starting to follow a certain source of information is generally a deliberate decision, and from the moment in which that link is created there is a directed flow of information happening between those two nodes. Links in this model express also a flow of information, and possibly an influence exercised by one node on another, but generally not a "friendship" relationship, nor any other kind of mutual personal relationship between the two nodes. For these reasons this model, which we call here "information network", should not be confused with a social network.

The information flowing through the network can be thought as composed of conceptual units, or *memes*, and the information network is ultimately the environment in which the kind of evolutionary cultural processes described by Memetics take place. Nodes are in fact acting as repositories of *memes*, and units of information are replicated and transferred through the arcs of the information network. Information that is more "fit" will have more chances to be propagated from link to link, with nodes acting as information relays, while information with low value for nodes in a certain region of the network will tend to be blocked. In this process, information is often processed, and the *memes* of which it is composed are recombined and selected.

It is especially interesting to analyse the dynamics of evolution of an information network. There is in fact an upper bound to the number of nodes that a single node can follow, since following too many sources would cause a node to be unable to make use of all the information it receives. Thus, any node will try to maximise the value of the information it receives by choosing to follow those nodes that produce or diffuse information that is particularly interesting and valuable from its particular perspective. The value of information is not absolute, but different from node to node, as different people have different capability to make use of specific information, as well as different taste. Furthermore, agents acting as nodes in the network have bounded rationality, and most importantly a "partial horizon", since they cannot observe the whole information network, and thus they cannot choose a globally optimal solution to their problem of selecting the right sources/nodes. Instead, they progressively change their set of sources with steps of improvement, following new nodes that provide information of high value when they get to know them from their current sources, and ceasing to follow the least interesting ones. As a result of the distributed effort of each node trying to optimize its set of sources, the whole information network evolves its structure, shaping the paths through which different kind of information diffuses. If the structure of the information network is found to evolve toward an ordered and non-random equilibrium, that same equilibrium is by definition the outcome of a collective intelligence process, since - as said - no single node has the authority or the possibility to globally shape the network, and the global structure is the product of bounded decision taken by nodes at a micro level.

4 RESEARCH DIRECTIONS

The authors propose a quest for emergent collective intelligence processes on the Web that may shape the way in which information diffuses and culture evolves. The research will involve an analysis of existing "natural experiments", on which to make quantitative measurements and test the proposed model of information network. One promising research methodology would be to make use of web crawlers to collect data on relevant indicators of users' behavior in the information network, as well as measures of the efficiency of the information routing. In other words, a first aim will be to measure how good Internet communities are at collectively self-organize to improve the signal-to-noise ratio and to deliver information quickly to the users who can benefit most from it, while filtering out non-relevant or flawed information.

A second research direction will be the analysis of the dynamics of evolution of information networks. For this purpose it will be particularly important to build realistic agent-based simulation models, in order to test the impact of various factors on the pattern of self-organization of the whole network.

Hopefully, the analysis of emergent collective intelligence processes in information networks will provide practical implications and applicable considerations on how to positively influence their evolution toward an optimal configuration. The knowledge resulting from this kind of research may thus contribute to the field of distributed information systems design.

REFERENCES

Bloom, H., 2001. Global Brain: The Evolution of Mass Mind from the Big Bang to the 21st Century, *Wiley Books*

Bonabeau, E., 2009. Decisions 2.0: The Power of Collective Intelligence, *Sloan Management Review*, 50(2)

Border, A., 2006. Collective intelligence: a keystone in knowledge management, *Journal of knowledge Management*, 10(1)

Carr, N. G., 2008. Is Google Making Us Stupid?, *The Atlantic, Jul/Aug 2008 issue*

Dawkins, R., 1976. The Selfish Gene, *Oxford University Press (Book)*

Dennet, D. C., 1992. Consciousness Explained, *Back Bay Books*

Gantz et al., 2007. The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011, *ITC White Paper*, http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf

Gruber, T., 2008. Collective knowledge systems: Where the Social Web meets the Semantic Web, *Web Semantics: Science, Services and Agents*, 4(13)

Levy, P., 1997. Collective Intelligence: Mankind's Emerging World in Cyberspace, *Perseus Books, Cambridge MA*

O'Reilly, T., 2005. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html