Information Technology & Tourism, Vol. 8 pp. 15–29

1098-3058/06 $60.00 + .00

www.cognizantcommunication.com

COMPLEX SYSTEMS, INFORMATION TECHNOLOGIES, AND TOURISM: A NETWORK POINT OF VIEW

RODOLFO BAGGIO

Master in Tourism and Economics, Bocconi University, Milan, Italy
School of Tourism and Leisure Management, The University of Queensland, Australia

There is a growing interest in complexity science as a framework for understanding social and economic systems. This article aims at presenting this approach giving a brief overview of the complexity framework and illustrating some of the methods in order to allow the reader to gain a deeper appreciation of this perspective. The role of information management and information technology in tourism, emphasized on numerous occasions, is examined in this context. It is argued that this framework can offer tools and techniques able not only to better understand the general state from a theoretical point of view, but can also provide practical guidance in specific situations. As an example, the structure of the community of websites belonging to Italian travel agencies is analyzed.

Key words: Information technology; Tourism; Network structure; Complex systems; Web structure

Introduction

Almost every paper, book, or article published in the last few years, dealing with the relationships between Information and Communication Technology (ICT) and the world of tourism starts by stating the importance of ICTs in this context.

The characteristics of tourism make it a very difficult subject to be defined with a reasonable accuracy. Terms such as complex, dynamic, networks, information intensive, and others are very often used to describe the area, so that a characterization such as *complex dynamical information network business* is very likely to be generally accepted without major discussions.

These words bring to a physicist's mind a series of concepts and considerations, especially if we think of a physicist involved in the information technology arena.

Some of these considerations may be quite useful to attempt to define a general framework in which the role and the functions of ICTs are better explained.

In recent years a number of ideas, concepts, and techniques typical of physics have been applied to different disciplines such as biology,

Address correspondence to Rodolfo Baggio, Master in Tourism and Economics, Bocconi University, via Sarfatti, 25, 20136 Milan, Italy. Tel: +39 02 5836 5447; Fax: +39 02 5836 5439; E-mail: rodolfo.baggio@uni-bocconi.it

economy, and sociology. Some of the results of this interdisciplinarity are interesting and have helped in improving the general understanding of these fields. In particular, in the economic area of study, physicists are having an increasingly important role, so that a new discipline has been born: *econophysics*.

The basic idea is far from being new (see, e.g., Mirowski, 1989); it may be dated back to the birth of what we today call physics, but what in present times has changed is the attitude. After Hobbes's attempts (*The Leviathan*, 1651) to use the Galilean laws of motion to derive an ideal configuration of society, or David Hume's (*An Enquiry Concerning Human Understanding*, 1748) hopes to build up a science of mind and society in the image of Newton's theories of the solar system, to mention a few examples, we think in a different way today.

Instead of mechanically applying laws and theorems to the economy, our feeling today is that the best use of what we commonly identify as *scientific method* is in trying to predict the performance of a system (a society or an economy, in our case) by deriving this prediction from a set of initial conditions and a series of mathematical models. In such a way we hope to be able to obtain a reasonable account of the system, to determine what choices lead to what consequences, and to make reasonable forecasts of its future behavior.

The nonlinearity of the relationships among the numerous components of social and economic environments and the difficulty in stating simple laws to study the behavior of such systems has brought to a number of methods and techniques assembled in what is known as *science of complexity*.

The aim of this article is to present this approach and to give a brief overview of the complexity framework and some of its methods in order to allow the reader to acquire a deeper appreciation of this perspective.

The rest of this article is organized as follows. The next two sections give an overview of the framework of complex adaptive systems (CAS) and of the tools used to study them, focusing on the topological analysis of the underlying networks. The fourth section present a current model of the World Wide Web based on network theories. The fifth section contains a preliminary investigation of Italian travel agencies used as an example of application of these methods. Finally, the last section presents some concluding remarks and discusses possible future directions.

Complexity

The natural language concept of complexity has several meanings, usually related to the size and the number of components in a system. There is still no universally accepted definition, nor a rigorous theoretical formalization, of complexity. Nonetheless, it is currently a much investigated research topic.

Intuitively we may characterize a complex system as "a system for which it is difficult, if not impossible to reduce the number of parameters or characterizing variables without losing its essential global functional properties" (Pavard & Dugdale, 2000).

Basically, we consider a system complex if its parts interact in a nonlinear manner. There are rarely simple cause and effect relationships between elements and a small stimulus may cause a large effect, or no effect at all. The nonlinearity of the interactions among the system's parts generates a series of specific properties that characterize the complexity of its behavior.

It is important to highlight the difference between complicated and complex. A complicated system is a collection of a number, often very high, of elements whose collective behavior is the cumulative sum of the individual ones. In other words, a complicated system can be decomposed in subelements and understood by analyzing each of them. On the contrary, a complex system can be understood only by analyzing it as a whole, almost independently by the number of parts composing it.

For example, as Amaral and Ottino (2004) note:

> A Boeing 747-400 has more than 3×10^6 parts. In complicated systems, such as the Boeing, parts have to work in unison to accomplish a function. One key defect (in one of the many critical parts) brings the entire system to a halt. This is why redundancy is built into the design when system failure is not an option. More importantly, com-

> plicated systems have a limited range of responses to environmental changes. . . . a Boeing without its crew is not able to do much of anything to adjust to something extraordinary. (p. 147)

On the other hand, a “simple” object made of only two elements—a double pendulum, a pendulum hanging from another pendulum—is well known to any physics student for its totally unpredictable, chaotic behavior (under the basic Newtonian laws of motion). A “simple” school of fishes, composed of a few dozen elements, is able to adapt its behavior to the external conditions without apparent organization but following a few very easy rules regarding local interaction, spacing, and velocity (Reynolds, 1987).

Generally, as Bar-Yam (1997) notes, a complex system is a mesoscopic structure, composed of a number that is not too low nor too high (but even this distinction is rather confused).

A special class of complex systems is the one composed by those that influence and are influenced by the external environment and in which the interactions among the elements are of a dynamic nature. In a *complex adaptive system*, this is the term used to denote this type of system, the parts: “interact with each other according to sets of rules that require them to examine and respond to each other’s behaviour in order to improve their behaviour and thus the behaviour of the system they comprise” (Stacey, 1996, p. 10).

For a complex adaptive system (CAS) the main characterizing features may be summarized as follows:

- **Nondeterminism.** It is impossible to anticipate precisely the behavior of such systems even knowing the function of their elements. The dependence of system’s behavior from the initial conditions is extremely sensitive and appears to be extremely erratic; the only predictions that can be made are probabilistic.
- **Presence of feedback cycles (positive or negative).** The relationships among the elements become more important than their own specific characteristics and the feedback cycles can influence the overall behavior of the system.
- **Distributed nature.** Many properties and functions cannot be precisely localized; in many cases there are redundancies and overlaps; it is a distributed system.
- **Emergence and self-organization.** A number of emergent properties are not directly accessible (identifiable or foreseeable) from an understanding of its components. Very often, in a CAS, global structures emerge over a critical threshold of some parameter. Typically a new hierarchical level appears that reduces the complexity. In continuing the evolution, the system evolves, increasing its complexity up to the next self-organization process.
- **Limited decomposability.** The dynamic structure is studied as a whole. It is difficult, if not impossible, to study its properties by decomposing it into functionally stable parts. Its permanent interaction with the environment and its properties of self-organization allow it to functionally restructure itself.
- **Self-similarity.** It implies that the system considered will look like itself on a different scale, if magnified or made smaller in a suitable way. The self-similarity is evidence of a possible internal complex dynamics. The system is at a critical state between chaos and order, a condition that has been also called a self-organized critical state. A self-similar object, described by parameters N and z, sees a power law relationship between them: $N = z^k$. The best known of these laws is the rank-size rule, which describes population in cities, word frequencies, and incomes. The so-called Zipf’s Law has the general form $P(r) = Kr^{-q}$, where $P(r)$ is the size of the event (population in the case of cities), r is its rank in descending order of size, and K is a scaling constant. A power law means that there is no “normal” or “typical” event, and that there is no qualitative difference between the larger and smaller fluctuations.

Examples of complex adaptive systems include: the patterns of birds in flight or the interactions of various life forms in an ecosystem; the behavior of consumers in a retail environment; people and groups in a community; the economy; the stock-market; the weather; earthquakes; traffic jams; the immune system; river networks; zebra stripes; sea-shell patterns; and many others.

Complexity, a multidisciplinary concept derived from mathematics and physics, applies well to the world of economics. The interactions of human calculations in decision making add a layer of complexity that may not exist in other disciplines. As Saari (1995) writes, "even the simple models from introductory economics can exhibit dynamical behavior far more complex than anything found in classical physics or biology" (p. 222).

To describe this approach, we may identify a number of features of an economy that present difficulties for the "linear" mathematics usually employed in economics (derived from J. Holland, the father of genetic algorithms, and quoted in Arthur, Durlauf, & Lane, 1997, introduction):

- **Dispersed Interaction.** What happens in the economy is determined by the interaction of many dispersed, possibly heterogeneous, agents acting in parallel. The action of any given agent depends upon the anticipated actions of a limited number of other agents and on the aggregate state these agents co-create.
- **No Global Controller.** No global entity controls interactions. Instead, controls are provided by mechanisms of competition and coordination between agents. Economic actions are mediated by legal institutions, assigned roles, and shifting associations. Nor is there a universal competitor—a single agent that can exploit all opportunities in the economy.
- **Cross-cutting Hierarchical Organization.** The economy has many levels of organization and interaction. Units at any given level—behaviors, actions, strategies, products—typically serve as "building blocks" for constructing units at the next higher level. The overall organization is more than hierarchical, with many sorts of tangling interactions (associations, channels of communication) across levels.
- **Continual Adaptation.** Behaviors, actions, strategies, and products are revised continually as the individual agents accumulate experience—the system constantly adapts.
- **Perpetual Novelty.** Niches are continually created by new markets, new technologies, new behaviors, new institutions. The very act of filling a niche may provide new niches. The result is ongoing, perpetual novelty.
- **Out-of-Equilibrium Dynamics.** Because new niches, new potentials, new possibilities, are continually created, the economy operates far from any optimum or global equilibrium. Improvements are always possible and indeed occur regularly.

It is easy to recognize in this list the main features characterizing a complex adaptive system.

Intuitively, the tourism sector, as economic activity, shares many of these characteristics. The theoretical work in this field is still in its infancy and just a handful of researchers have started to consider the complex systems approach as a more effective framework for understanding of the many and different phenomena (Farrell & Twining-Ward, 2004; Faulkner, 2000; McKercher, 1999).

Much is still to be done, but the hopes are those of being able to understand, for example, how crises, disasters, or turbulent changes may influence the sector, or why, after crises such as 9/11, the tourism sector is able to show a rapid and almost unexpected recovery (World Tourism Organization [WTO], 2002).

McKercher's (1999, p. 429) model, for example, looks very promising, even if a little "neglected." The main components of the system are:

- The *Traveler*, who is the essential player in tourism, for without people traveling no tourism would occur.
- *Communication vectors* used to connect the traveler to the destination.
- *Considerations* or factors that influence the effectiveness of the Communication vectors used.
- *Destination or Internal tourism community* consisting of all businesses involved in tourism at the destination.
- *External tourism* agencies (public and private sector) that try to influence tourism.
- *Other tourism-related externalities*, such as alternative tourism destinations that affect a destination's ability to attract travelers.
- *Non-tourism-related externalities*, or macroenvironmental forces, such as changing political, economic, or social conditions, war, natural disaster, that affect people's ability to travel.
- *Outputs from the system* both desired and undesired.

- *Rogues or Chaos makers* who can push a system to the edge of chaos.

The model tries to describe the operation of complex tourism systems in order to allow a representation of the elements that influence tourism on a wide range of possible scales: national, regional, local, and, possibly, at an enterprise level. In fact, even if the number of actors influencing the system changes at each level, the relationships between the different elements are similar.

In this way, the author tends to provide a better framework able to explain, for example, the failure of many well-designed, controlled, and sustainable tourism development plans. By using this approach we may be also able to better understand the impact of the usage, and the uses, of technologies like the Internet on parts of the system and how the system may react to more developments. The general picture can be summarized in the following way.

Let us consider a destination as a representative tourism subsystem. It is a system whose elements can be thought as belonging to a number of groups (such as those proposed by McKercher, 1999). These groups are connected with relationships that exhibit a nonlinear dynamic behavior, producing outcomes that cannot be simply explained "summing up" the individual characteristics. In the evolution of the system it is possible to find many of the features indicated by Arthur, Durlauf, and Lane (1997) as portraying a complex adaptive system (CAS).

A central property of a CAS is the possible emergence of unforeseen properties or structures: the self-organization is one of the most striking features characterizing a complex system. A consequence of this is the robustness or resilience of the system to perturbations (or errors): the system is relatively insensitive and has a strong capacity to return to a stable behavior, in absence of external inputs. For tourism destinations, and for the tourism sector in general, this property is the one that has been exhibited on several occasions after crises and disruptions. Recent events such as the 9/11 and Madrid terrorist attacks, the Bali bombings, the SARS epidemics, the Iraq war, and others have greatly affected the sector. However, almost unexpectedly, the recovery to preevent levels (see, e.g., the tourist arrivals statistics provided by WTO) has been accomplished in a relatively short period of time, typically a few months. No simple "linear" model could have explained fully this behavior.

The Role of Information Management in Tourism

Information technology has profound implications for the tourism industry and is being used extensively in a great variety of functions. Contemporary ICTs have radically altered the way in which information is conveyed throughout the industry and to customers, with the effect of bridging the distance between all the components of this market.

The use and the uses of ICTs are driven by the interaction between complex demand requests and the rapid expansion and sophistication of new products. One of the obvious results is that tourism is one the most important application areas on the WWW.

"In few other areas of activity are the generation, gathering, processing, application and communication of information as important for day-to-day operations as they are for the travel and tourism industry" (Poon, 1993, p. 154).

The importance of IT for tourism basically resides in the information-intensive nature of the business (Buhalis, 2000; Werthner & Klein, 1999). If this is true, then any technology usable to effectively manage information acquires an enormous value. These arguments are those usually brought to stress the value of ICTs in the tourism field. With the framework discussed in this article, a different argument may be brought.

Does information have a role in a complex system? In order to understand fully the effects of information in a complex system we should be able to define a metric, a way of measuring complexity. We have seen that the concept is still not well characterized and there is no common definition for it. It is no surprise, then, to find a vast number of proposed measures of complexity. *The Hypertext Bibliography of Measures of Complexity* by B. Edmonds (available at: http://bruce.edmonds.name/combib/) contains more than 300 references, and it is updated only to 1997.

We may think of complexity as the amount of independent pieces of information needed to specify a system, and of organization as the characterization of the degree of the interrelations among the components in terms of number, scope, and dynamics. In this way we may define complexity, as Murray Gell-Mann says (1994), as: "The length of the shortest message that will describe a system . . . employing language, knowledge, and understanding that both parties share" (p. 34).

Complexity can therefore be seen as a measure of how difficult it is to describe the regularities of a structure in complete detail. Mathematically, it may be expressed as the difference between a maximally compressed Shannon information content and the information content of the elements that are truly random (Gell-Mann & Lloyd, 2004).

From an economic point of view, it has been shown that in complex systems like multicellular organisms or insect colonies, efficient information sharing can result in sensible advantages both in reducing individual efforts and in improving group organization, even for cases in which the information acquisition has a cost (Lachmann, Sella, & Jablonka, 2000).

The relationship between information and complexity, obviously, is not linear. An excess of information leads to an increase of complexity. The information overload (Heylighen, 2002) may create a number of problems, basically consisting in increased instability and randomness of the connections among the elements of the system. And we all know, from a personal point of view, that a high level of *data smog* leads to anxiety, stress, alienation, and potentially dangerous errors of judgment, withholding overall economic productivity.

Managing a system essentially means finding a control structure able to produce order by inducing relations among its elements, which has the effect of reducing the indistinguishability of the states of the system.

By moving along the evolutionary path of a dynamic system, from complete order to random behavior, complexity is maximized in the middle. Processing the right amount of information is a necessary condition for reducing complexity and therefore managing the system effectively.

If we convey this line of reasoning into the tourism sector, it easy to see not only the importance of managing information, but the inevitable necessity of it.

The Analysis of Complex Systems: Network Theory

A number of tools have been developed in recent years to cope with the task of describing a complex system. In their review, Amaral and Ottino (2004) identify three main classes of tools belonging to three areas well known to physicists and mathematicians: nonlinear dynamics, statistical physics, and network theory. The last one has recently had much attention and has provided helpful results in a great number of fields.

It is possible to describe many complex systems in terms of networks of interacting elements: roads, railroads, airlines, electric circuits, oil pipelines, social and business groups, telecommunication connections, chemical reactions, and food webs, just to give a few examples. The complexity of many systems is rooted in the combined web defined by the system's components and their interactions.

The variety in types is countless, but all networks are characterized by a set of nodes and by the connections between the nodes. The interactions of nodes through the connections lead, in many cases, to a global behavior of the network, which is not observed in the single elements of the network: the emergent behavior of a complex system. The collective properties of dynamic systems composed of a large number of interconnected parts are strongly influenced by the topology of the underlying network.

The mathematical models of network structures have been developed in graph theory. A graph is a generalization of the concept of a set of dots (vertices, nodes), connected by links (edges, arcs). These, depending on the specific situation, may or may not have a direction (the graph is directed or undirected). In a directed graph it is possible to track a route from some vertex to another, but not in the opposite direction. Links may be associated with numeric values (they may represent distances, costs, energies, information exchanges, etc.) called weights. The term graph, by itself, is usually assumed to mean a simple graph, in which at most a single edge exists between any two vertices (directed or undirected, weighted or not).

More formally a graph is defined as a pair $G = (V,E)$, where V is a set of vertices, and E is a set of pairs of distinct vertices that are members of V: $E = \{(u,v)\,|\,u,v \in V\}$. The elements of E are called arcs or edges.

In the last few years, a number of researchers have shed light on some topological aspects of many kinds of social and natural networks (the WWW, power grids, collaboration networks, networks of words, metabolic networks, economic agents).

The first mathematical model, which has been used for many years to describe many kinds of networks, is due to Erdös and Rényi (1959, 1960). For them (the ER model) a network is a set of N nodes connected by k links that are chosen randomly, with probability p, from the $N(N-1)/2$ possible links.

The random distribution of the nodes degrees k (the number of connections each node has) follows a Poisson law with a peak $\langle k \rangle$:

$$P(k) \approx \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

The average node degree $\langle k \rangle$ (the mean in the Poisson distribution above) is a basic aspect of a graph and may be used as distinctive characteristic of a network as typical instance of a statistical ensemble. Other quantities commonly used to characterize a graph are:

- clustering coefficient C, that measures how close the neighborhood of each vertex comes on average to being a complete subgraph (part of the graph in which every node is connected to all the others, also called clique). The clustering coefficient is a measure of the connectivity of the neighborhood of a node; it captures the local density of a graph. For a particular node, it is defined as the ratio of the number of edges in the neighborhood of that node to the total number of possible edges in the neighborhood. The overall clustering coefficient of a graph is the average of the clustering coefficient of all the nodes (with a degree greater than one);
- average length L of a path between any two vertices. A path is a sequence of vertices $v_1, v_2, \ldots v_k$, with the property that for any $i(1 \leq i < k)$ there is an edge between v_i and $v_i + 1$;
- diameter D: the maximal shortest distance (path) between any pairs of vertices.

At the end of the 1990s, empirical investigations (Barabási & Albert, 1999; Watts & Strogatz, 1998) confirmed that, in many cases, ER random graphs are quite different from real world networks. Following these, numerous investigations have been done and a great variety of theoretical works have been published (see the reviews by Albert, & Barabási, 2002; Dorogovtsev & Mendes, 2002; Newman, 2003).

The distribution $P(k)$ of the nodes degrees can be used to classify the networks into three broad classes (Amaral, Scala, Barthélémy, & Stanley, 2000):

- **single-scale networks:** in which $P(k)$ exhibits exponential or Gaussian tails. This class contains the random ER graphs and the small world (SW) networks described by Watts and Strogatz (1998). These are characterized by large clustering coefficients and short average path lengths. Their degree distribution still follows a poissonian law;
- **scale-free networks:** $P(k)$ has a power-law distribution: $P(k) \sim k^{-\gamma}$. The distribution is largely uneven, there is no characteristic mean nodal degree (the mean of a poissonian ER distribution), but some (few) nodes act as very connected hubs, having a very large number of ties, while the majority of nodes have a small number of links. Scale-free (SF) networks, introduced by Barabási and Albert (1999), are dynamic networks. They grow with the addition of new nodes and new links that are not distributed randomly, but follow specific mechanisms; the most commonly invoked is a preferential attachment in which a new node has a higher probability to attach to one of the most connected ones;
- **broad-scale networks:** for which the degree distribution has a mixed behavior, a power law regime followed by a sharp cutoff (exponential or Gaussian decay) of the tail.

Both SF and SW networks, very common structures among the real-world networks, show peculiar characteristics such as (Newman, 2003):

- robustness: stability of the system to random removal (or failure) of randomly chosen elements; but also
- fragility: high sensitivity to targeted attacks to the most connected hubs;
- low internal friction: extent and speed of disease (viruses, but also messages, fads, beliefs, etc.) transmission are greatly improved with respect to a random ER network, in some cases it is shown that there are no critical thresholds at all for these phenomena.

We may expect that the topology of the tourism actors network, like many other social and economic networks, exhibits structures like the ones discussed above. If a social network has a scale-free structure (see, e.g., the review by Newman, 2003), which is characterized by a power law distribution of the nodal degrees, and we know (Gisiger, 2001; Guiderà, Danon, Díaz-Guilera, Giralt, & Arenas, 2003; Ostling, Harte, Green, & Kinzig, 2003) that a power law very often implies self-similarity (also called fractal) properties, it is natural to expect that any system of tourism operators has a similar arrangement.

A simple example can confirm this idea. Figure 1 shows the structure of the connections among a number of components of an alpine Italian tourist district (basic data are from Denicolai & Gramegna, 2002).

The nodes of the network represent a mixture of single operators (hotel associations, groups of intermediaries, etc.) and groups, and the number of elements is too small to be able to derive a significant degree distribution. Nonetheless, a *scale-free* structure is rather evident, some nodes have a high number of connections, while the majority of the others have few.

Being part of a more general economic system, and exhibiting almost the same structure, our network is a good example of those self-similarity characteristics that are typical of a *complex system.*

The World Wide Web as Complex Network

The World Wide Web (WWW) is, without any doubt, the most important and significant phenomenon of the last few years; its impact on almost all forms of our social, economical, and personal life is enormous. The WWW remains basically uncontrolled; any organization, institution or single person can create a website, of any size, with relatively little effort. This growth, without any clear rule or centralized control, produces a huge, self-organizing, complex system.

Furthermore, for its nature, the WWW is easily measurable in terms of nodes and connections, and it is no surprise that from the study of the Web topology originated the first seminal works that founded what today is called the *science of networks* (Barabási & Albert, 1999; Faloutsos, Faloutsos, & Faloutsos, 1999; Watts & Strogatz, 1998).

The WWW can be described in terms of a large directed graph whose vertices are the html documents and whose edges are the hyperlinks connecting one document to another; its topology determines its connectivity.

The study of the Web as a graph is not only intriguing per se, but it is also of great importance in providing "practical" answers to significant problems. First of all is the one we may denote as *visibility*, which can be easily translated into the problem of finding functional algorithms for crawling (Deo & Gupta, 2001), searching, and community discovery (Gibson, Kleinberg, & Raghavan, 1998; Newman & Girvan, 2004). These problems translate directly into an economic problem, for the definite value acquired by the links in easing the tracking down of a website. The connections (hyperlinks) can be seen as a pseudo–monetary unit (Walker, 2002).

The current model of the WWW is mainly due to the research by Broder et al., (2000), Dill et al. (2001), and Flake, Lawrence, Giles, and Coetzee (2002).

The bow-tie model (as it is commonly called) sees the Web as a self-organizing, self-similar structure, basically divided into three main components:

- a core of strongly connected nodes (SCC), accounting for almost 28% of the web pages in the sample studied, all joined with bidirectional links;
- a set of pages (IN) mainly connected in a mo-

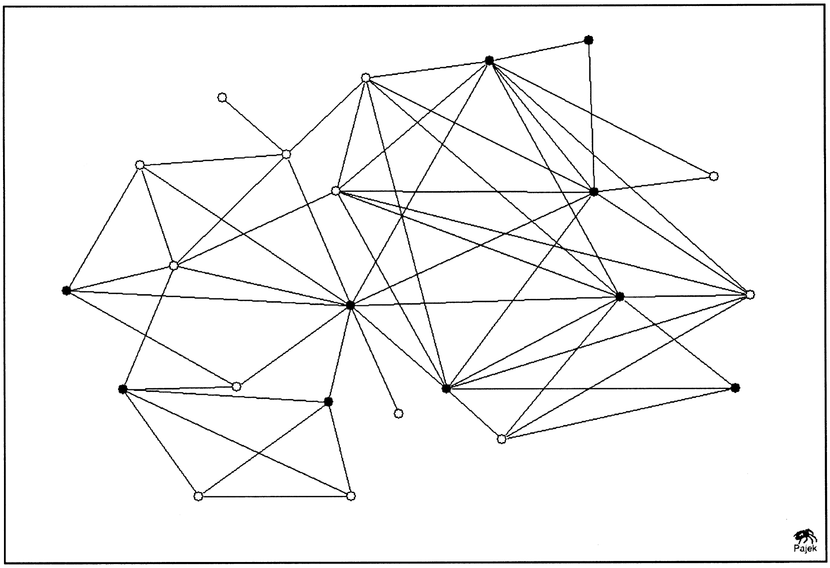

Figure 1. The connections among the operators in an Italian alpine tourism district (after Denicolai, & Gramegna, 2002, network drawn with Pajek). Filled circles represent groups of operators, open circles are single organizations.

nodirectional way to the SCC. The pages in this component (almost 21% of the sample) have outgoing links that reach the SCC, but they are virtually unreachable by other parts of the web; and

- a similar sized (21%) set of pages (OUT) reachable by the pages contained in SCC, but whose links are mainly inward bound (i.e., there is always a path from SCC to OUT pages, there is no direct connection from OUT to SCC or IN).

The picture is completed by two more sets: TENDRILS composed by pages (again 21%) providing paths from IN pages or to OUT pages without passing through the SCC elements. The TENDRILS contain pages that cannot reach the SCC, and cannot be reached from the SCC. It is possible for a TENDRIL hanging off from IN to be hooked into a TENDRIL leading into OUT, forming a TUBE—a passage from a portion of IN to a portion of OUT without touching SCC. To complete the model, there are pages forming a sort of islands disconnected from the other main components (DCC). The general picture, a bow-tie like graph, is shown in Figure 2.

The supposed high global connectivity of the Web appears to be not so high; only in 24% of the cases it is possible to find a path between any two nodes chosen at random (Broder et al., 2000). The average distance between nodes is ~16; it goes down to 6.8 if we consider the undirected connections in SCC.

The degree distribution, on the average, follows

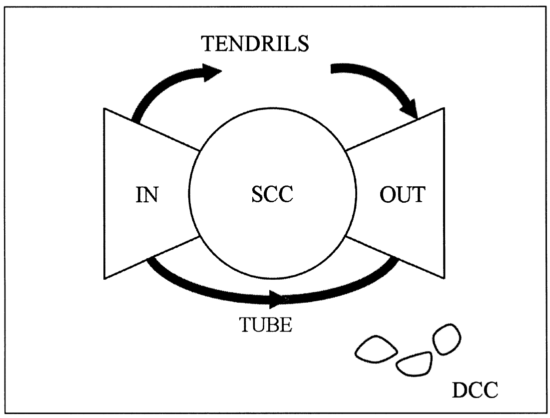

Figure 2. The bow-tie model of the Web.

a power-law $P(k) \sim k^{-\gamma}$. The values of the exponent γ are: $\gamma = 2.1$ for the indegree distribution (incoming links to a page) and $\gamma = 2.72$ for the outdegree distribution (links outgoing from a page).

In social network analysis, outdegree indicates expansiveness and indegree indicates popularity. The difference (although small) in the measured distributions may be interpreted as a relatively higher tendency to receive links because of own popularity rather than opening towards the external world of web pages (a steeper power law means that the number of websites with outdegree = k_{out} decreases much faster).

These values, or values very close to these, have been confirmed by several different empirical studies, which confirmed also the value for the web diameter $D \sim 7$, and the average degree of a node in the Web $\langle k \rangle = 7$ (Pastor-Satorras & Vespignani, 2004).

It is rather clear that the positioning of a website in one of the most connected components is of paramount importance if an organization wants to obtain good reachability and visibility in order to better achieve its objectives.

This is more significant if we think that the next generation of crawlers, spiders, or other automated searching tools will use data about the link topology and adopt "link-based" browsing methodologies. They will make it possible to enable navigation approaches that effectively use general structural information (Deo & Gupta, 2001).

One technique that is presently explored is the possibility to identify a *community* [i.e., a group of websites (or web pages) among which there are relatively strong and direct ties, typically because of common arguments and interests]. Different methods have been proposed to detect such communities (Castellano, Cecconi, Loreto, Parisi, & Radicchi, 2004; Newman & Girvan, 2004). Almost all of them are based on some measurement of characteristics (centrality, betweenness, clustering, etc.) of densely connected subgroups of nodes in the WWW network that may be identified as communities. Almost all of the algorithms exclude from consideration nontopological information and try to use the subgraph parameters to define the strength of the community from a quantitative point of view.

An Example: The Network of Italian Travel Agency Websites

The ideas, the models, and the methods discussed so far can have, as already told, a practical outcome in many different situations.

As an example let us consider the travel agencies (TA). Even if TA may have different configurations and perform different types of business, they may be broadly defined as belonging to a set of intermediaries offering services and serving functions both to tourism producers (hotels, transport companies, attractions, etc.) and to travelers.

In the "real" world travel agencies compete fiercely to obtain a good positioning in the market. The competition among them is high and the determinants for success are numerous: type and quality of the products and services, collaboration with other organizations, marketing strategies, pricing strategies, etc.

Among these, physical location has a great importance. For an agency shop, physical position is an important element; having well-designed windows on a very busy street is commonly presumed to be a positive determinant in attracting significant numbers of customers.

The advent of the Internet, and its huge diffusion, has generated the same positioning and visibility problem. Being easily found among billions of websites and being considered useful and usable by the possible customer are the main characteristics determining not only the success of a company in the virtual environment, but possibly influencing also the performance in "real" life (Teerling & Huizingh, 2004).

Contents and services provided, the design and the appeal of the shop windows, have been analyzed, evaluated, and discussed in several investigations (see, e.g., Frew, 2004; Frew, Hitz, & O'Connor, 2003). This series of works has also established a set of criteria and methodologies for both evaluating a website and designing and arranging the elements that can be regarded as success factors.

One more issue can be raised. Is the placement good? Is the shop window appearing on a busy street? In other terms we may ask if a website, or better the community of websites belonging to a certain group of tourism operators, is well positioned on the Web in order to favor a good visibility and reachability.

A sample of 700 Italian travel agencies websites has been analyzed. By using Google, without doubt the most used, complete, and popular search engine, the number of incoming links for each website has been derived. Care has been taken in cleaning the results from links coming to a website's pages from pages in the same domain.

The resulting frequency distribution is shown in Figure 3. The plot (Fig. 3) on a log–log paper clearly shows a power-law behavior with a few websites showing a great number of inbound links and most having only a few.

The relationship can be written: $P(k) \sim k^{-\gamma}$, with an estimate for the exponent $\gamma = 1.29 \pm 0.07$.

One more quantity, connected with the "positioning" of a website, that can be easily determined is the PageRank. Used in Google as one of the elements contributing to the order in which query results are presented to the user, PageRank (PR) represents a popularity rating to each page. This definition is recursive and it assigns each page a fraction of the rank of each value pointing to it, weighted by the outdegree (the number of

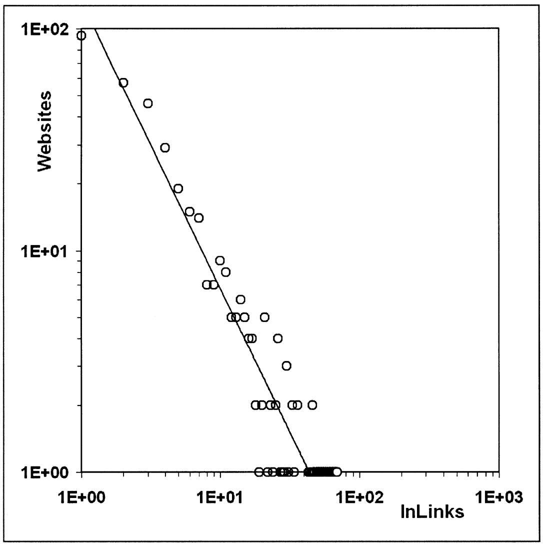

Figure 3. The indegree distribution of a sample of Italian travel agencies websites (axes are both logarithmic).

outgoing links). A website PR is the sum of the PRs of the website pages. To calculate the Page Rank for a page, all of its inbound links are taken into account.

Suppose there are only three pages: A, B, C. B and C are linking A; their PageRanks are $PR(B)$, $PR(C)$ and they have a total of $d_{out}(B)$ and $d_{out}(C)$ links.

$PR(A)$ will be $(1 - p)$ is a constant expressing the probability of randomly browsing one of the two linked pages):

$$PR(A) = \left(\frac{1-p}{2}\right) + p\left(\frac{PR(B)}{d_{out}(B)} + \frac{PR(C)}{d_{out}(C)}\right)$$

Formally, for a page i, $PR(i)$ is determined (Page, Brin, Motwani, & Winograd, 1999) by the equation:

$$PR(i) = \frac{1-p}{n} + p \cdot \sum_{j \in In(i)} \frac{PR(j)}{d_{out}(j)}$$

where 1, 2, . . . n are the pages on the Web, $d_{out}(i)$ is the outdegree of a web page, $In(i)$ the set of pages pointing to page i, and $p(0 < p < 1)$ a decay factor (i.e., the probability with which a user proceeds with a random walk during his browsing). This expression represents a system of linear equations and $PR(i)$, the unique solution vector, is the eigenvector of the matrix giving the probability of a random walk.

It has been shown (Pandurangan, Raghavan, & Upfal, 2002; Pastor-Satorras & Vespignani, 2004) that, in the SCC component of the WWW, the PageRank follows a power law distribution with an exponent ~2.1.

The PR of websites and pages can be derived by consulting Google, where the values are given on a 1 to 10 scale.

The distribution of PR for the Italian travel agencies websites considered above is shown in Figure 4, where the histogram shows the distribution of the PR scored by Google for the websites considered.

As it can be clearly seen, the values are not distributed according to a power law, as we would expect, but they are spread in a random way. The normal curve superimposed on the distribution

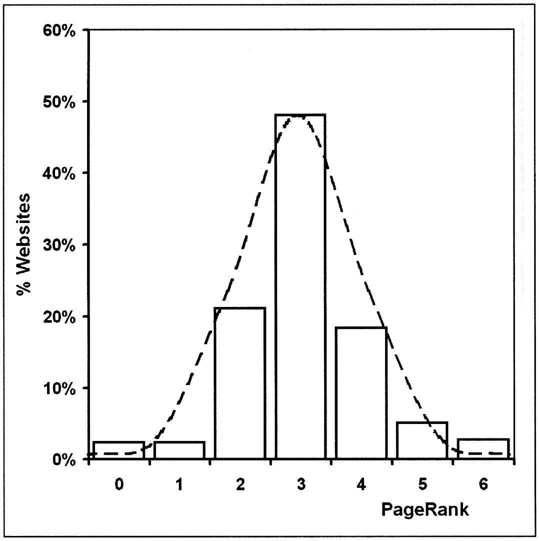

Figure 4. The PageRank distribution of a sample of Italian travel agencies websites (the dotted line represents a generic normal distribution).

shown in Figure 4 matches the data almost perfectly.

The analysis of the measurements in our sample allows us to draw some interesting conclusions. As far as the form of the degree distribution is concerned, the results are in agreement with all the measurements known (Albert & Barabási, 2002; Newman, 2003); the inlinks to the websites examined follow a power law distribution (Fig. 3). As our sample is a subnetwork of the WWW, this result is a clear confirmation of the self-similar structure of the Web.

The exponent of the inlink distribution is, however, quite different from the one usually found in the literature for similar studies (Albert & Barabási, 2002; Newman, 2003); the distribution appears to be much flatter ($\gamma = 1.29$) and more uniform than the one that may be considered typical, at least of the Web SCC component ($\gamma = 2.1$).

This result may lead us to conclude that the topological characteristics of the Italian travel agents community are not compatible with our knowledge of the main connected component of the Web graph.

One confirmation of this explanation comes from the PageRank distribution. The obvious randomness (Fig. 4) may be interpreted as a clear sign of a sparse arrangement of our sample on the WWW.

It is obviously difficult to assign to our community a precise location on an ideal map of the Web (Fig. 2); nonetheless, we may confidently say that the most probable one is outside the main SCC component. The most probable position for the sample analyzed is in one of the disconnected components of the bow-tie model of the Web. In other words, this community is less connected to the general WWW network and in essence to the general "social" network of the Web.

The consequences of this situation can be critical for the operators involved. Staying in components of the Web not "well connected" to the rest of the virtual world implies an intrinsic difficulty in being found, and much lower website traffic levels, thus harming the overall performance even for a single organization (both online and offline) and losing to a competitive set of similar organizations that, on the contrary, have reached a better group visibility. Elsewhere it has also been shown (Antonioli & Baggio, 2004) that the websites offered online by Italian tourism intermediaries have a low amount of content and services, and perform poorly in qualitative evaluations by users.

The combined effects of these two elements—visibility on the Web and "quality" of the online offering—have important outcomes for an organization.

With an ever-growing proportion of Internet users seeking tourist information on the Net, and willing to buy tourist products online, providing modest service and putting their websites in poorly connected areas of the WWW increases the risk of losing valuable customers such as the ones coming from the wealthy, Internet-intensive countries (US, Northern Europe, Japan, etc.). For an intermediary, this situation may have serious effects and may generate a general distrust that can eventually affect also the economic performance of the "real" world.

Italian travel service providers will have to completely change their attitude towards new information and communication technologies, and especially towards notions concerning the connectivity of a website. The ever-increasing number of proposals of automated searching tools, based on the topology of the Web or of its subnetworks, strongly suggests effective actions in order to improve the connectivity of these websites to the rest of the Web and strong efforts to reach the SCC component.

Even with the obvious limitation of being just an initial estimate, the technique used can give new insight to the analysis of the relationships among the websites of a community of tourism operators. By using these techniques, coupled with the more "traditional" analysis of website content and functions characteristics, it is possible to add a less subjective perspective and give more practical indications to achieve the maximum benefit from the usage of ICTs.

Concluding Remarks

It is nowadays widely recognized that information management and the technologies that support it are of paramount importance for the travel and tourism sector. The existence of a theoretical framework in any field of art or science recognizes the importance of that field, and is essential for an understanding of the foundations. More than that, theories and models are crucial not only per se, but also for the implications they may have for the areas of practice, public appreciation, and knowledge.

This article contains a first attempt at defining such a scheme for what concerns the field of tourism and the role of IT in the framework of the *science of complexity*. Inside this framework we may use the methods and the techniques of the analysis of complex systems to gain better knowledge and to find practical solutions to problems.

We have applied these techniques to an initial analysis of the connectivity of a particular subgroup of tourism websites, those belonging to Italian travel agencies, showing the implications the results may have for the future development of this business.

One simple consequence of our results is the suggestion to improve the connectivity of a website by implementing (or revising) an effective linking strategy. An important point is that this strategy should incorporate the idea of forming communities of similar organizations, in order to facilitate the formation of a group that could more easily be recognized as such.

This result may have a greater importance if we consider the elements constituting a tourism destination. In this respect, a well-connected set of sites, with a great level of inter- and cross-linking could be better identified by the Internet users, thus helping to establish and improve the awareness that is such an important element of the success on the market. Future research will address these topics.

The techniques presented in this article may also have more applications.

By comparing the "real" network with its "virtual" counterpart it will be possible to derive hints to improve efficiency and effectiveness of the investments tourism operators are today almost forced to make in the field of information and communication technologies.

By studying the tourist community network topology and comparing it with more general social network communities it will be possible to find optimization algorithms in order to adapt it to better meet the expectations of both tourists and industry. For example, an optimization of the network of the information exchanges among the operators of a destination can be especially useful for a DMO wishing to implement policies or giving marketing or operational indications.

Tools, conceptual outlines, and methods, although not yet fully clearly and rigorously defined, are already available and have also proved to offer important indications in different fields.

Acknowledgments

The author wishes to thank Hannes Werthner and Magda Antonioli Corigliano for the valuable discussions and support. The author also wishes to thank the anonymous referees for their helpful suggestions.

Biographical Note

Rodolfo Baggio holds a degree in Physics from the University of Milan, Italy. He is a lecturer at Bocconi University where he teaches courses in Computer Science and coordinates the Information and Communication Technologies area in the Master of Tourism and Economics Programme. After having worked for the IT industry, he is now involved in a number of educational and research projects on the relationships between IT and the travel and tourism sector.

He is conducting his Ph.D. research at the School of Tourism and Leisure Management of The University of Queensland, Australia.

References

Albert, R., & Barabási, A-L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics, 74*, 47–91.

Amaral, L. A. N., & Ottino, J. M. (2004). Complex networks—augmenting the framework for the study of complex systems. *European Physical Journal B, 38*, 147–162.

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small world networks. *Proceedings of the National Academy of Sciences USA, 97*, 11149–11152.

Antonioli Corigliano, M., & Baggio, R. (2004). Italian tourism on the Internet—new business models. In K. Weiermair & C. Mathies (Eds.), *The tourism and leisure industry—shaping the future* (pp. 301–316). New York: The Haworth Press.

Arthur, W. B., Durlauf, S., & Lane, D. (Eds.). (1997). *The economy as an evolving complex system II.* Reading, MA: Addison-Wesley [introduction available at: http://www.santafe.edu/arthur/Papers/ADLIntro.html] (last accessed August, 2004)

Barabási, A-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509–512.

BarYam, Y. (1997). *Dynamics of complex systems.* Reading, MA: Addison-Wesley.

Broder, A. Z., Kumar, S. R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J. L. (2000). Graph structure in the web. *Proceedings of the 9th WWW Conference/Computer Networks, 33*(1–6), 309–320.

Buhalis, D. (2000). Tourism and information technologies: Past, present and future. *Tourism Recreation Research, 25*(1), 41–58.

Castellano, C., Cecconi, F., Loreto, V., Parisi, D., & Radicchi, F. (2004). Self-contained algorithms to detect communities in networks. *European Physical Journal B, 38*, 311–319.

De Liso, N., & Filatrella, G. (2002). Econophysics: The emergence of a new field? *Economia Politica, 19*, 297–332.

Denicolai, S., & Gramegna, A. (2002). *Relazioni interorganizzative nei distretti turistici montani* (Working paper, year X, No. 6, December). Università degli Studi di Pavia'Facoltà di Economia, Dipartimento di Ricerche Aziendali.

Deo, N., & Gupta, P. (2001). Graph-theoretic web algorithms: An overview. In T. Böhme & H. Unger (Eds.), *Lecture notes in computer science* (Vol. 2026, pp. 91–102). Berlin: Springer-Verlag.

Dill, S., Kumar, S. R., McCurley, K., Rajagopalan, S., Sivakumar, D., & Tomkins, A. (2001). Self-similarity in the web. *Proceedings of the 27th International Conference on Very Large Data Bases* (pp. 69–78) (and *ACM Transactions on Internet Technology, 2*(3), August 2002, 205–223).

Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics, 51*, 1079–1187.

Erdös, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae (Debrecen), 6*, 290–297.

Erdös, P., & Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5*, 17–61.

Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Computer Communication Review, 29*, 251–263.

Farrell, B. H., & Twining-Ward, L. (2004). Reconceptualizing tourism. *Annals of Tourism Research, 31*(2), 274–295.

Faulkner, B. (2000). *The future ain't what it used to be: Coping with Change, Turbulence and Disasters in Tourism Research and Destination Management.* Griffith University, Professorial Lecture Series No. 6, 17 August [available online: http://www.gu.edu.au/ins/collections/proflects/faulkner00.pdf] (last accessed July, 2004)

Flake, G. W., Lawrence, S., Giles, C. L., & Coetzee, F. M. (2002). Self-organization of the Web and identification of communities. *IEEE Computer, 35*(3), 66–71.

Frew, A. J. (Ed.) (2004). *Information and Communication Technologies in Tourism 2004, Proceedings of the International Conference in Cairo, Egypt.* Wien: Springer.

Frew, A. J., Hitz, M., & O'Connor, P. (Eds.). (2003). *Information and Communication Technologies in Tourism 2003, Proceedings of the International Conference in Helsinki, Finland.* Wien: Springer.

Gell-Mann, M. (1994). *The Quark and the jaguar.* New York: W. H. Freeman.

Gell-Mann, M., & Lloyd, S. (2004). Effective complexity. In M. Murray Gell-Mann & C. Tsallis (Eds.), *Nonextensive entropy—interdisciplinary applications* (pp. 387–398). Oxford, UK: Oxford University Press.

Gibson, D., Kleinberg, J., & Raghavan, P. (1998). Inferring Web communities from link topology. *Proceedings 9th ACM Conference on Hypertext and Hypermedia. ACM Press* (pp. 225–234).

Gisiger, T. (2001). Scale invariance in biology: Coincidence or footprint of a universal mechanism? *Biological Reviews of the Cambridge Philosophical Society, 76*, 161–209.

Guimerà, R., L. Danon, L., Díaz-Guilera, A., Giralt, F., & Arenas, A. (2003). Self-similar community structure in a network of human interactions. *Physical Review, E, 68*, 065103(R).

Heylighen, F. (2002). *Complexity and information overload in society: Why increasing efficiency leads to decreasing control.* [Online at: http://pespmc1.vub.ac.be/Papers/Info-Overload.pdf] (last accessed October, 2004)

Lachmann, M., Sella, G., & Jablonka, E. (2000). On the advantages of information sharing. *Proceedings of the Royal Society of London, B, 267*, 1287–1293.

McKercher, B. (1999). A chaos approach to tourism. *Tourism Management, 20*, 425–434.

Mirowski P. (1989). *More heat than light: Economics as* *social physics, physics as nature's economics.* Cambridge, UK: Cambridge University Press.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*, 167–256.

Newman, M. E. J., & Girvan, M., (2004). Finding and evaluating community structure in networks. *Physical Review, E, 69*, 026113.

Ostling, A., Harte, J., Green, J. L., & Kinzig, A. P. (2003). A community-level fractal property produces power-law species-area relationships in nature. *Oikos, 103*, 218–224.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the Web.* Stanford Digital Library Project, Working Paper SIDL-WP-1999-0120, Stanford University, CA.

Pandurangan, G., Raghavan, P., & Upfal, E. (2002). Using PageRank to characterize Web structure. *Proceedings of the 8th Annual International Conference on Computing and Combinatorics* (pp. 330–339) [available online at: http://citeseer.ist.psu.edu/pandurangan02using.html] (last accessed June, 2004]

Pastor-Satorras, R., & Vespignani, A. (2004). *Evolution and structure of the Internet—a statistical physics approach.* Cambridge, UK: Cambridge University Press.

Pavard, B., & Dugdale, J. (2000). The contribution of complexity theory to the study of socio-technical cooperative systems. *Proceedings of Third International Conference on Complex Systems*, Nashua, NH, May 21–26 [available online: http://www.sv.cict.fr/cotcos/pjs/] (last accessed July, 2004)

Poon, A. (1993). *Tourism, technology and competitive strategies.* Oxon: CAB International.

Popper, K. (1959). *The logic of scientific discovery.* London: Hutchinson, (First edition 1935: *Logik der Forschung.* Wien: Springer)

Reynolds, C. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21*, 25–34.

Saari, D. G. (1995). Mathematical complexity of simple economics. *Notices of the American Mathematical Society, 42*(2), 222–230.

Stacey, R. (1996). *Complexity and creativity in organizations.* San Francisco: Berrett-Koehler.

Teerling, M. L., & Huizingh, E. K. R. E. (2004). *How about integration: The impact of online activities on store satisfaction and loyalty.* Research Report 04F08, University of Groningen, Research Institute SOM (Systems, Organisations and Management).

Walker, J. (2002). Links and power: The political economy of linking on the Web. *Proceedings of the 2002 ACM Hypertext Conference* (pp. 72–73). Baltimore, MD: ACM Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small world' networks. *Nature, 393*, 440–442.

Werthner, H., & Klein, S. (1999). *Information technology and tourism—a challenging relationship.* Wien: Springer.

World Tourism Organization. (2002). *Tourism proves as a resilient and stable economic sector.* [Online at http://www.world-tourism.org/newsroom/Releases/more_releases/june2002/data.htm] (last accessed July, 2004)