Chapter 13
Modeling Innovation

Roberto Serra, Marco Villani and David Lane

13.1 Why Model Innovation, Which Models of Innovation?

The innovation theory (briefly, I_2T), which has been developed in the ISCOM project and which is presented in this book (Chapters 9 and 10), is based on the analysis of different case studies, spanning different time periods and different kinds of products, from the introduction of printing in the Renaissance, to key new technologies introduced in the 19th and 20th centuries, up to present-day ongoing innovation efforts.

This theory is qualitative (although fairly rigorous), and it does not claim to be able to provide either predictions or quantitative descriptions of the phenomena that it addresses. The theory makes statements concerning the core entities, relationships and processes that are necessary to understand a wide range of different phenomena and aims at uncovering some of their non-obvious features.

A natural question is whether modeling can be of any help in the development and refinement of such a theory. Indeed, quantitative models are often used to compare the theoretically expected behavior with the one observed, and, wherever quantitative predictions are possible, to forecast the behavior of the interesting variables.

But if the theory is inherently qualitative, what contribution can models provide? Of course, this question has been asked several times in the development of the social sciences, and it has received interesting, albeit partial, answers.

In the present context, models of particular interest are those that allow one to describe the behavior of a large collection of agents, thus bridging the gap between the level of a few human agents or organizations, which is addressed by I_2T, and the level of the behavior of a system comprising many more such agents. In order to be precise, let us stress that we concentrate here on dynamical models, which allow one to observe how the system changes in time.

This is by no means the only kind of model that might initiate a dialogue with the present theory. For example, a different kind of model could be also considered,

R. Serra (✉)
University of Modena and Reggio Emilia, Modena and Reggio Emilia, Italy

D. Lane et al. (eds.), *Complexity Perspectives in Innovation and Social Change*, Methodos Series 7, DOI 10.1007/978-1-4020-9663-1_14,

aimed at a sophisticated description of the interaction among two or few agents. However, in the development of our research, we attributed higher priority to exploring the behaviors of a collection of interacting agents, as suggested in, for example, Epstein and Axtell (1996), Gilbert and Terna (2000), and Axelrod and Tesfatsion (2006). The reason is that I_2T is endowed with several positive feedback mechanisms, and it is well known that in such cases, counterintuitive behaviors can be observed at a system level. Therefore, the behavior of a large group of agents cannot be easily inferred from that of a pair, or a few of them. Modeling can thus provide a major contribution to unfold the large-scale consequences of the theory.

The behavior of a collection of agents observed in the model can then provide information to the theorists about the consequences of their assumptions. Confirmation of expected behaviors is an interesting result. Even more interesting, there may be unanticipated or surprising results, which can lead to a re-thinking of the theory itself.

However, the interpretation of the results of the simulation is complicated by the fact that the model cannot simply be considered "the theory in action:" in order to make it manageable, it is indeed necessary to introduce simplifications and make choices that are not dictated by the theory.

Therefore, when there are surprising results at the aggregate level, it may be unclear whether these are the consequences of the simplifications introduced, or whether they represent a real outcome of the theory itself. It is therefore necessary to perform an in-depth analysis to disentangle the contribution of the (largely arbitrary) choices that are model-related from genuine unanticipated consequences of the theory.

While bridging the gap between the levels of a few agents to that of a collection of many of them appears to be the most important motive to resort to modeling, there are other important reasons that make modeling useful to develop a broad and rigorous perspective on innovation. They can be briefly summarized as follows:

- The development of a formal model requires that, in the restricted universe defined by the model itself, the claims of the theory be precisely stated. Therefore ambiguities and imprecisions can be more easily spotted and amended, thus sharpening the rigor of the theory
- The models developed provide a toy universe where one can easily experiment with the consequences of new suggestions, thereby leading to the possibility of exploring a wide set of alternatives
- The observations of the way in which the models evolve and of the different behaviors of the agents can provide a useful way to describe and communicate the theory itself and its major characteristics.

Let us now consider the type of models that are better suited for the purposes stated above. "Dynamical Model" is a general term that encompasses several alternative approaches, including differential equations and their time-discrete analogs, stochastic differential equations, cellular automata and many others. We have chosen to focus our work on so-called agent-based models (ABMs), which are sometimes referred to as multi-agent systems, in particular in the Artificial Intelligence literature.

The main reasons for this focus are that in these models, it is possible:

- to deal with agents capable of sophisticated information processing,
- to endow agents with adaptive and learning capabilities that not only modify their behavior in time, but also change the rules which underlie this behavior,
- to introduce agents' goals and beliefs in a straightforward way, and
- to take fully into account the heterogeneity of agents, therefore escaping the constraints of the "representative agent" of traditional economic modeling.

ABMs are well suited, therefore, to relate hypotheses concerning the microscopic behavior of individual agents to emergent collective phenomena in a population composed of many interacting agents.

While models based on differential equations are too often quickly dismissed by devotees of ABMs as inadequate, we have recently argued that their expressive power is greater than what is usually assumed (Serra & Villani, 2006). However, in the present case we prefer to resort to ABMs because the reference theory requires that agents possess sophisticated ways to handle information about the world they inhabit, and because both intentionality and heterogeneity are thought to be important.

Attractive as agent-based models may be, they allow the modeler too much freedom with respect to the design of agents and their interactions. ABMs often have too many variables and parameters, but their worst shortcoming is that there is no theory of their behavior comparable to, say, that of differential equations. Therefore, the effects of the choice of the parameters cannot be foreseen or evaluated based on such a theory. Moreover, as observed above, these models usually include strong nonlinearities that are known to be able to give rise to unexpected outcomes.

Indeed, algorithmic models like ABMs may be highly arbitrary, so there is an *embarras de richesses*, and conceptual guidelines are needed to constrain the set of allowable models. Two major considerations may help to provide such constraints.

The first is that it is necessary to look for model behaviors that are robust with respect to different perturbations, which may involve changes in parameter values and, to some extent, in the choice of the functions that describe the agents and their behaviors. If simulation results match real world observations (which may be limited and imprecise, as so-called "stylized facts") only for a limited set of parameter values, e.g. for a very particular kind of decision rule, then suitable reasons should be found to justify these choices, before claiming that the model has anything to do with the real system it is supposed to simulate.

The second methodological consideration is that a model may be constrained strongly by its relationship with a theory of the social process that it is supposed to describe. In this way, the model should concentrate on those aspects that the theory identifies as the most relevant, dropping out or drastically simplifying a set of related, but (according to the theory) less important aspects. This approach is similar, in some sense, to that of classical hard science, and it is the one we take in this paper.

The theory provides the basic entities of the model and describes the kinds of relationships among them. The model represents a simplified universe inhabited by (some of) the theoretical entities, engaging in some of the kinds of relationships predicated by the theory. Simplification is necessary to deal with manageable systems: we do not look for an all-encompassing model, but, rather, we foresee a family of models that capture different features of the theory. The model described here is meant to represent the first member of this family.

The plan of the chapter is the following. In Section 13.2, we will briefly summarize those aspects of I_2T that are the more relevant to our purposes, and, in particular, we will focus on the indications and constraints that they provide for models consistent with the theory. In the following Section 13.3, we will outline a description of one such model, which we call I_2M (ISCOM Innovation Model). Although we made some efforts to keep the model as simple as possible, the need to include a strong relationship with the theory led to a model with a variety of features, and explaining them in detail would take up the whole chapter. (Readers interested in more details about the model may consult Lane, Serra, Villani, & Ansaloni, 2005).

A particular problem is the setting of the initial state of the model. Our approach, inspired by the notion of "exaptive bootstrapping" (see Chapters 1 and 14, this volume), is described in Section 13.4 below. In the same section, we also discuss the reference parameter values, chosen based on extensive simulations (which however cannot even approach a complete sampling of the large parameter space) and provide some details about the interactions between recipes and artifact space. Section 13.5 describes how the model responds to various kinds of perturbations, linking up with general interest in "avalanche size distributions" for complex systems (see for example Kauffman, 1993; Bak, 1996).

The penultimate two sections are dedicated to an exploration of some interesting behaviors of the model. In Section 13.6, we discuss the influence of the agents' propensity to build long-lasting relationships on the probability of successfully realizing innovation projects. Section 13.7 investigates the influence of the initial condition (i.e. the state of the system at the beginning of the simulation) on formation of topological structures in artifact space. The final section sums up the results obtained so far with the model and provides some directions for further research.

13.2 Requirements from the Theory

We now examine the major constraints that the theory imposes on the model. We will not attempt here to summarize the I_2T, which is extensively discussed in Chapters 1, 9 and 10 of this volume, but we will limit ourselves to emphasizing the implications of the theory's main assumptions on our modeling efforts.

The first claim of the theory is that innovation processes involve transformations of relationships among agents, among artifacts, and between agents and artifacts. This notion underlies the concepts of an "agent-artifact space" and, in particular, that of "reciprocality," which essentially claims that artifacts mediate interactions between agents, and vice versa, and, therefore, both agents and artifacts are necessary for a proper understanding of the behavior of market systems and of innovation. Despite first appearances, this is neither a minor nor an obvious assumption, since it is entirely conceivable to focus just on artifacts ("technological trajectories" have been given much credit in the past) or on agents alone, as is generally the case for neoclassical theories of innovation or epidemiology-based theories of innovation diffusion (see Chapter 11 for further discussion on this point).

One straightforward consequence of this claim is that it excludes the possibility simply to project agent-artifact space onto one of its two constitutive subspaces, ignoring the dynamics in the other subspace.

In I_2T, artifacts are given meanings by the agents that interact with them, and different agents take on different roles. The meaning of artifacts cannot be understood without taking into account the roles that different agents can play. Thus, artifacts may be given different meanings by different agents or by the same agent at different times.

Innovation, in the sense of I_2T, is not just novelty, but also a transformation in the structure of agent-artifact space, which unleashes a cascade of further changes. The innovation may involve the introduction of a new artifact, but also a change in relationships with other agents, or even a new interpretation of an existing artifact. In a sense, this theory can be seen as a theory of the interpretation of innovation, where "interpretation" actually means attribution of functionalities to artifacts and of roles to agents.

According to the theory, a new interpretation of artifact functionality typically arises in the context of so-called generative relationships. By interacting, a few agents come to invent and share a new interpretation, based on the discovery of different perspectives and uses of existing or expected artifacts. The generative potential of a relationship may be assessed in terms of the following criteria:

- heterogeneity: the agents are different from each other, they have different features and different goals, but the heterogeneity is not so intense as to prevent communication and interaction;
- aligned directedness: the agents are all interested in operating in the same region (or in neighboring regions) of agent-artifact space; and
- mutual directedness: the agents are interested in interacting with each other.

Moreover, agents must be allowed to interact and to take joint actions.

These are the aspects of the theory that appear to be most relevant for the modeling effort. This theory of innovation is highly sophisticated in describing the interactions between different players in innovation processes, and it cannot be entirely mapped onto a specific computer-based model. Therefore, the modeling activity aims at developing models that are based on abstractions of some key aspects of this qualitative theory.

The basic requirements for a model aiming at a dialogue with the theory, according to the principles expressed above, are then the following:

1. both agents and artifacts must be present,
2. the artifacts' meanings must be generated within the model itself: since I_2T claims that new meanings are generated through interactions among agents and artifacts, it would be inappropriate here to resort to an external oracle to decide *a priori* which meanings are better than others,
3. the roles of agents must also be generated within the model: indeed the theory claims that new roles are also generated through interactions among agents and artifacts,
4. agents must interact with artifacts and with other agents: interacting only with artifacts would prevent the possibility of describing agent-agent relationships,
5. an agent should be able to choose the other agents with whom to start a relationship; in general, an agent will be able to handle a finite number of relationships at a time, and it will choose a subset of the other agents as its partners. Agents must be allowed to dissolve a disappointing or unsatisfactory relationship to look for a better one, and
6. agents must have intentionality: they may be interested in certain types of artifacts or in entering a relationship with other agents.

13.3 An Outline of the I_2M Model

Let us now briefly introduce the main features of the model I_2M, which has been developed on the basis of the above principles. We will limit ourselves here to a very concise overview of the model, and we refer the interested reader to (Lane et al., 2005) for a more complete and detailed account.

In I_2M, agents can "produce" artifacts, which, in turn, can be used by other agents to build their own artifacts. An agent can produce several artifacts for the same agent, and it can produce one type of artifact for different customer agents.

Each agent has a set of recipes that allow it to build new artifacts from existing ones. Agents can try to expand the set of their recipes by applying genetic operators either to their own recipes or, by cooperating with another agent, to the joint set of the recipes of both. Moreover, each agent has a warehouse or store, where its products are put, and, from which, its customers can take these products.

The meaning of artifacts in this model is just what agents do with them, while the role of agents is defined by which artifacts they produce, with whom, and for whom. The role of agents is also partly defined by the social network they are embedded into, as explained in the following paragraph.

If we look at the network of agents, there is a directed link of the "strong tie" type between A and B if there is a customer-supplier relationship between the two. There are also weak ties between two agents ("acquaintances"), which refer to the fact that agent A knows something about agent B (e.g., its products). Note that other types of networks can also be considered; for example, the artifact network (where two artifacts are linked if there is an agent which uses one of them as an input to produce the other) and also a heterogeneous network, where both agents and artifacts are explicitly represented. In this latter case, most relationships are between an artifact and an agent: there is no direct link between two artifacts and no strong tie between two agents, since each of these ties is mediated by a direct tie with an artifact, but there may be links between two agents of the weak-tie kind.

The value that an agent (say, A) gives to its relationship with another agent, B, is summarized in a single numerical variable (the "vote"), which is composed of the sum of two terms. The first term is increased or decreased based on the history of supplier/customer interactions between A and B, while the second term takes into account the results of previous co-operations in developing new projects, if any. A parameter, which can be changed in different simulations, determines the relative weight of these two terms.

A key point is the structure of artifact space. It is required that the space has an algebraic structure, and that suitable constructors can be defined to build new artifacts by combining existing ones. For reasons discussed elsewhere, we have adopted a numerical representation for artifacts and the use of mathematical operators instead of e.g., binary strings, λ-calculus or predicate calculus. Therefore, the agents are "producers" of numbers by applying mathematical operators to other numbers, and recipes are defined by a sequence of operators and the numerical arguments on which these operators act.

Each agent is also endowed with a numerical variable (called its "strength") that measures how successful it has been so far. Strength increases proportionally to the number of artifacts sold and decreases proportionally to the number of active recipes. Note, however, that strength, in the present version of the model, cannot be interpreted as "money" since it is not conserved in the interactions between two agents (if A and B interact, and ΔS_A and ΔS_B represent the change in strength of the two agents due to this interaction, it may well happen that $\Delta S_A \neq -\Delta S_B$).

As far as innovation is concerned, an agent can invent new recipes or eliminate old ones. In the present version of the model, no new agents are generated, while agents can die because of lack of inputs or of customers.

The model is asynchronous: at each time-step, an agent is selected for update, and it tries to produce what its recipes allow it to do. Therefore, for each recipe, it looks for the input artifacts, and, if they are present in the stocks of its suppliers, it produces the output artifact and puts it into its stock (the supplier stocks are, of course, reduced). Production is assumed to be fast, i.e. it does not take multiple steps.

Besides performing the usual "buy-and-sell" dynamics, when its turn comes, an agent can also decide to innovate. Innovation is a two step-process. In the first step, the agent defines a goal, i.e. an artifact that it may add to the list of its products, while the second step concerns the attempt to generate a recipe that yields the goal as output.

In the goal-setting phase, an agent chooses one of the existing types of artifacts (which, recall, is a number M) and then either tries to imitate it (i.e. its goal is equal to M) or it modifies it by a jump (i.e. by multiplying M by a random number in a given range). It has been verified that imitation alone can lead to a sustainable structure for agent-artifact space, in which, however, innovation eventually halts, which need not be the case when (some) agents innovate via jumps (Lane et al., 2005).

After setting its goal, an agent tries to reach it either by using the operators of one of its recipes with different inputs or by combining two recipes to generate a new one with genetic algorithms. In this phase, the agent can decide to cooperate with another agent, sharing the recipes of both, to reach a common goal. The propensity of an agent to cooperate is ruled by a parameter, while for most simulations we discuss, choice of the partner is made with a probability distribution biased in favor of those agents to whom the choosing agent has assigned a high vote.

13.4 Typical Model Behaviors

13.4.1 Initialization

Setting up initial conditions "by hand" is time demanding, so we developed an automatic "initial condition builder." First, raw materials are introduced; afterwards, agents are added one at a time. The randomly generated recipes of the newly added agent can use only already existing artifacts as inputs, producing in this way (possibly) new artifacts. The networks that result from such a process have the property that older artifacts are likely to be connected to more agents than new ones.

Networks generated according to the above procedure have the common feature that the unused artifacts (in particular, those produced by the most recently added agent, many of which are new to the system) cannot provide any source of activation to their owners, and, therefore, the agent that produces them cannot survive, causing a "domino effect" that generally destroys the entire network. To avoid this collapse, we can either provide an external reward region or introduce a modification of the dynamics (e.g. the possibility of changing a provider) that leads to self-sustaining rings, as shown in Fig. 13.1.

13.4.2 Time Behavior of Some Key Variables

We have performed many experiments with this model, in order to test the model and to understand some of its behaviors. In particular, we find that

1. Systems without innovation collapse (Fig. 13.1a) unless: there is an external market (Fig. 13.1b) or a self-sustaining loop is already present within the system (Fig. 13.1c);
2. Imitation alone is unable to introduce a significant number of novelties; and
3. The simultaneous presence of imitation and jump actions allows a strong increase of diversity in the resulting artifact space.

To provide some hints about the temporal behavior of the model, we do not present statistical results, but rather show typical model behaviors. During the simulation shown below, agents perform innovations that are either "social" (by linking up with new suppliers) or "technological" (by generating new goals and recipes).

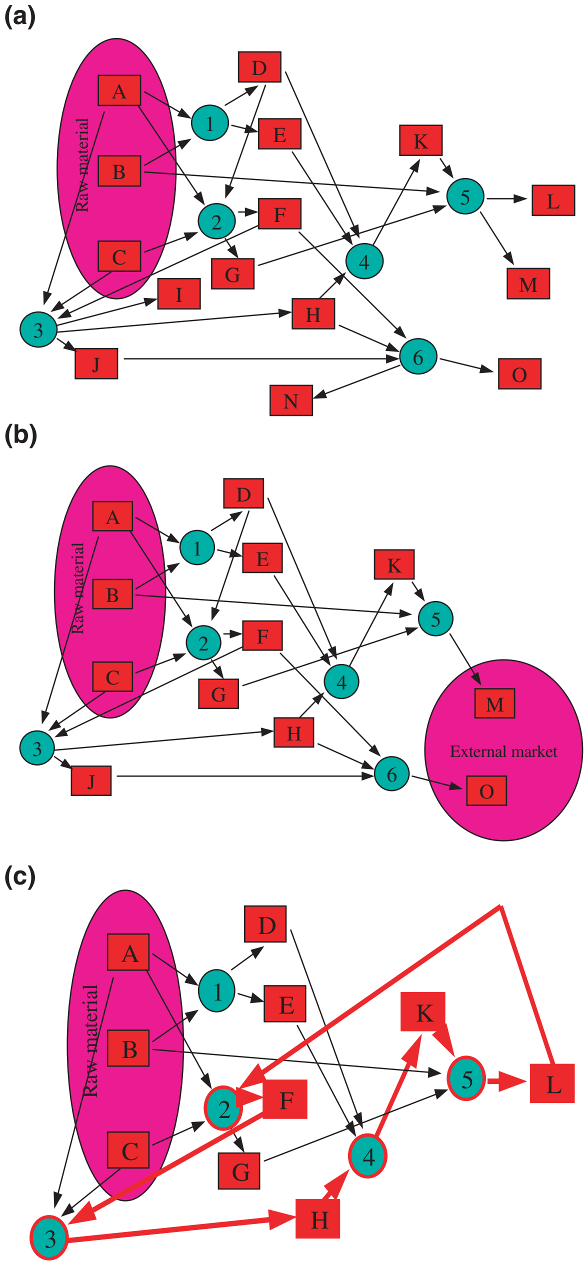

Fig. 13.1 (**a**) The initial unstable situation, and (**b**), (**c**) two possible stable configurations. In (**b**), the agents' survival is guaranteed by the presence of an external market (artifacts N and L disappear), whereas in (**c**), it is guaranteed by the presence of a self-sustaining cycle (highlighted in *red* – artifacts M, O and J disappear)

Each agent knows only the artifacts produced by those agents with whom it has a tie, weak or strong.

We have examined different innovation procedures. In particular, when innovation in recipes is performed by applying genetic operators to the set of existing recipes, we have considered both the case of “lonely” agents, who modify only their own recipes, and “collaborative” agents, who share their recipes in genetic search of new ones.

A typical example is shown in Fig. 13.2, where we can observe the temporal behavior of some variables. The parameters describing the characteristics of the run are collected in Table 13.1.

Figure 13.2a shows the number of artifacts that are present in the system, in particular, the total number of artifacts, the number of artifacts produced, and the number of artifacts used. The difference between the total number of artifacts and that of artifacts used is the number of still unused innovations.

For these counts, the same artifact type (i.e., a number), if produced by different agents, counts as different artifacts. In contrast, Fig. 13.2b shows the number of different artifact *types* present in the system. Clearly, many artifacts types are produced by more than one agent.

Figure 13.2c, d show the temporal behavior of some aspects of the artifact space, that is, the diameter (the difference between the highest and the lowest number

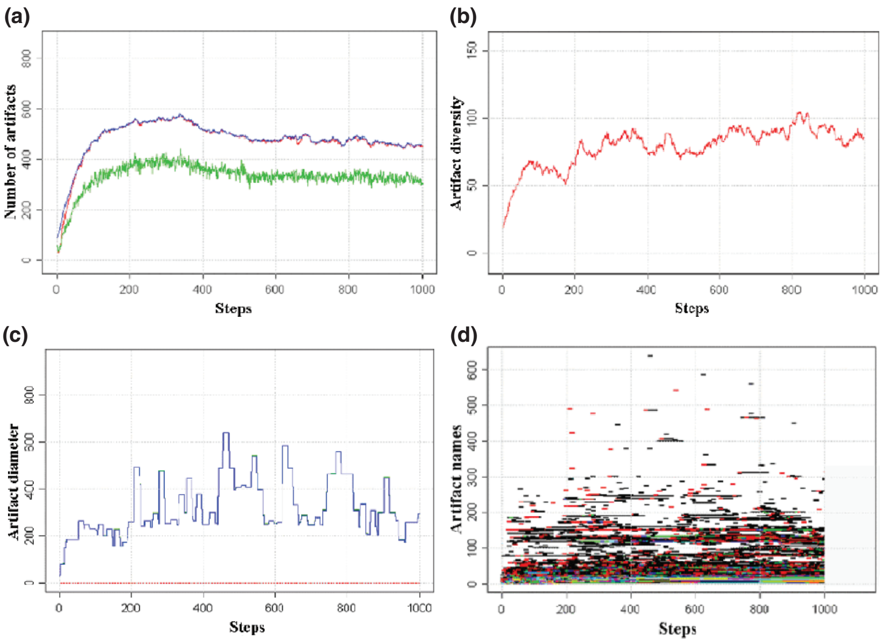

Fig. 13.2 The temporal behavior of some aspects of the artifact space. (**a**) The total number of artifacts present inside the system (*blue line*: total number of artifacts; *red line*: total number of produced artifacts; *green line*: total number of used artifacts). (**b**) The number of different “names” present inside the system (an index of the diversity present inside the system). (**c**) The diameter of the artifact space (the difference between the highest and the lowest “names” presents on the system). (**d**) a representation of the evolution of the artifact space (for each step there is a dot if the corresponding value is realized by at least an artifact, the different colors representing the different numbers name’s realizations)

Table 13.1 The parameters describing the characteristics of the run described in Section 13.4.2

Parameter	Value and/or description
Initial number of agents	40
Initial number of recipes for each agent	2
Number of allowed inputs for each recipe	2,3
Number of raw materials	2
Innovation probability during each step	0.2
Jump probability (once an agent initiate the innovation phase)	0.7
Jump range	[0.7, 6]
Innovation strategy	Each agent tries by itself; in case of failure, it tries again collaborating with another agent
Agents' knowledge	Each agent "knows" the artifacts: – of its providers – of its acquaintances
Number of random acquaintances for each agent	5
Vote strategy	Based upon the past successful joint projects

present on the system) and the "history" of the artifacts (where for each step there is a dot if the corresponding value is realized by at least one artifact, the different colors representing the different numbers of artifacts that realize the "name").

By observing Fig. 13.2, it is possible to monitor the expansion of artifact space, which is followed by a more stable phase. The peak in the total number of artifacts does not correspond to periods of maximum diversity, and the space occupation is not uniform (there are deserted areas close to zones with high density of artifacts).

Other variables monitor the agents' behavior: for example, Fig. 13.3 shows the fraction of the system known and the number of recipes owned by each agent. It is possible to observe the emergence of different paths: despite the fact that, in this experiment, the agents share the same parameter values, in the end there are agents that know a very small fraction of the system and agents that are aware of more than 60% of the artifacts, and there are agents that own less than five recipes while some have more than fifteen. This high heterogeneity implies that not all the agents are allowed to develop to the same level and suggests the presence of particular structures inside the system.

13.4.3 Goals

In I_2M, agents have a "goal" in artifact space, i.e. a new artifact they aim at producing, either by themselves or in cooperation with another agent. The goal (G) itself is determined by the following method:

- First, one of the existing artifacts is chosen at random (heuristically, this is a way to sample the set of existing artifacts), let us call it the intermediate goal (IG).

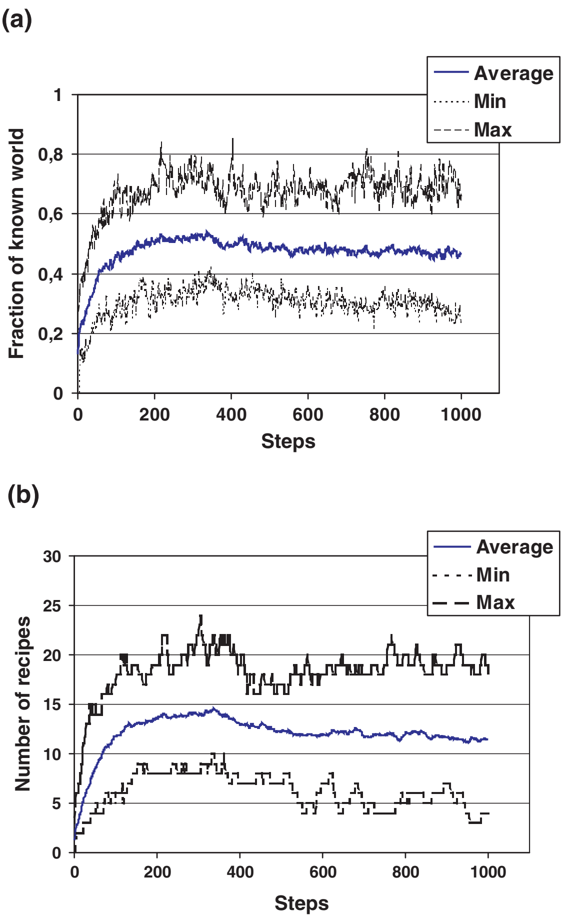

Fig. 13.3 (**a**) The fraction of the system known by each agent; (**b**) the number of recipes owned by each agent. The plots report for each run step the average, the minimum and the maximum of the interested variable

- Second, the IG is modified by a "jump." Recall that artifacts are numbers: the number corresponding to the IG is multiplied by a random number belonging to a given range R (briefly: G = J(IG)).

Resorting to randomness simulates the definition of a goal in a system where it is hard to predict whether a particular artifact will be successful or not, and where all the properties of an artifact "in use" cannot be established *a priori*, solely on the basis of engineering design considerations.

An important parameter is the goal persistence, which measures how long an agent will stick to its goal if it has been so far impossible to reach it. We can assume that each agent has the probability p_m of maintaining its own goal, ranging from 0 (changing the goal each time the agent has to innovate) to 1 (always keeping the same goal until it is successfully produced). Qualitatively, we can regard p_m as a measure of agent flexibility, i.e. its propensity to change its objectives (of course, it is actually a decreasing function of "flexibility").

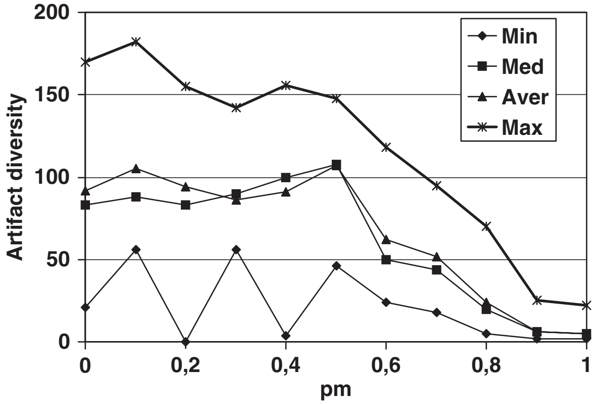

Fig. 13.4 Diversity of artifacts present in the system (average, median, minimum, and maximum out of 10 runs) as a function of the agent probability p_m of maintaining the goal. If p_m is equal to 1, the agents try to realize the same goal until they succeed in reaching it

The effects of the agents' different innovation propensities are evident by observing Fig. 13.4, which shows the diversity of artifacts in the system after 2000 steps of simulation.

If the probability of changing objective is sufficiently high, the system is able to maintain a sustained growth of diversity (Fig. 13.5); otherwise, diversity levels off and remains always lower than in the previous case. A threshold seems to appear around $\boldsymbol{p_m} \approx 0.5$. This effect is likely due to the fact that the perseverance in attempting to create an artifact that is hard or impossible to achieve with the available operators constrains some agents to continue in unsuccessful attempts, and, more importantly, prevents them from introducing other innovations.

13.4.4 Recipe Structure

In the experiments described so far, the recipes can use only two kinds of operator, that is, addition and subtraction, and have two or three input artifacts. The recipe structure is very simple, but is it enough to produce the increase of diversity of Fig. 13.5b, where the environments at step 50 and at step 1000 are very different.

To better appreciate this phenomenon, we made some experiments where the recipes have initially the same proportion of 2, 3, 4, 5, 6, or 7 inputs – in order to simplify the exposition; in the following, we indicate the number of inputs of each recipe as the recipe "size."

Not all kinds of recipes have the same survival probability: in fact, recipes unable to find all the desired inputs at the same time will, after a while, disappear, and obtaining seven different inputs at the same time is more difficult than finding only two or three.

Without any particular constraint, after a while we should observe an inhomogeneous distribution where short recipes overwhelm longer recipes: in effect, in absence of goals, this distribution is the final one (Fig. 13.6b). However, as just reported, we have a particular constraint: the agents do not try to generate new

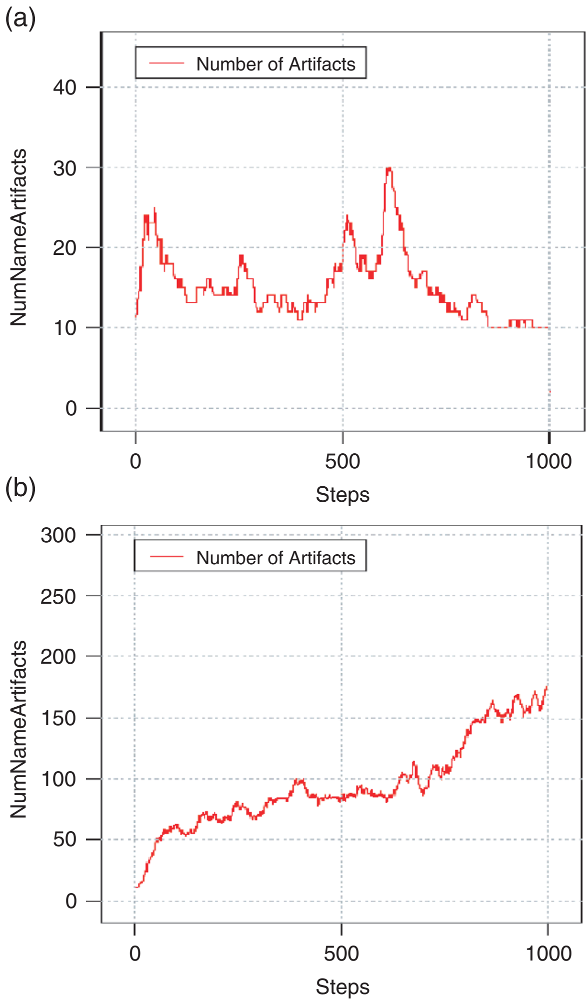

Fig. 13.5 The artifact diversity for a system with high p_m (**a**) and low p_m (**b**) versus time. Note the change of scale on the graphs

recipes at random but rather they aim at specific points in artifact space. It is not necessarily the case that simple recipes can reach the agents' desired goals!

In fact, the runs where the agents have goals reveal a more complex history. At the starting point, more or less all the kinds of recipes have the same number of realizations (Fig. 13.6a), whereas at step 2000, a distribution where the short recipes overwhelm the longer ones begins to emerge. But as the artifact space becomes more fragmented by increasing the deserted areas interposed with zones having high density of artifacts, simple recipes are not able to fulfill all the requirements – for example, to be able to build a new artifact belonging to an area separated from all the available inputs areas by one or more deserted zones. In this case, recipes with more than two inputs can be useful: at step 5000, the agents more frequently use recipes with four or five inputs (Fig. 13.6c).

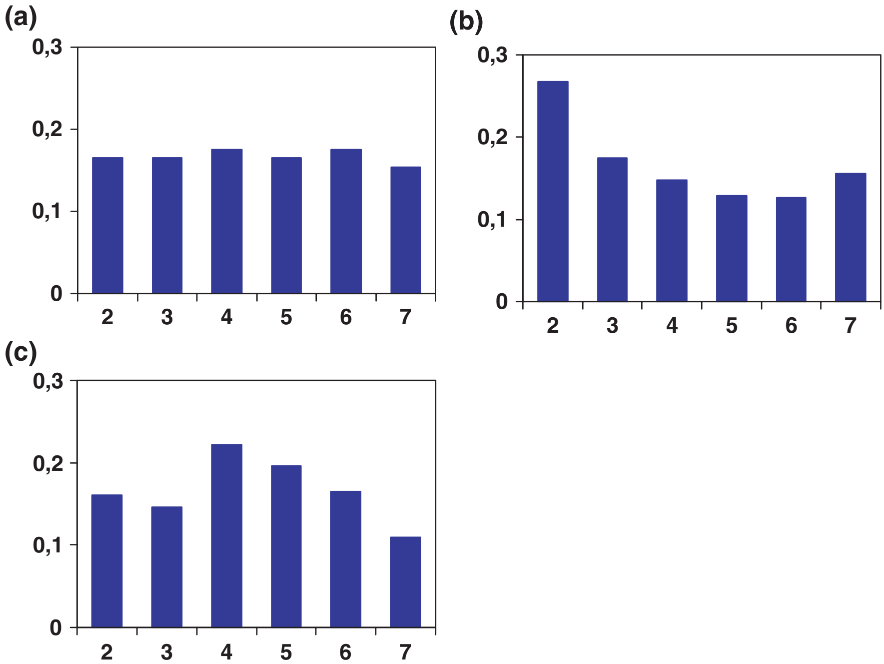

Fig. 13.6 The distribution of recipe size during a typical run where the allowed number of inputs spans from two to seven. (**a**) Initial distribution; (**b**) the final the distribution of recipe size in systems without goals; (**c**) the final distribution of recipe size in systems with goals

13.5 Distribution of Avalanches

13.5.1 Introduction

The model simulations show systems that change rapidly in time: artifacts continuously come to life and die, new functionalities are added to the already existing ones, agents can augment their own recipes or rapidly decline and eventually expire. Some questions naturally arise: how robust is the system? Are there bursts? What is the distribution of the size of avalanches of change?

Two different kinds of perturbations can be considered:

- internal perturbations, which can be observed by looking at the time plot of some relevant variable, e.g., the total number of artifacts. These are due to events that occur in the time evolution of the system. For example, it may happen that a recipe is eliminated that was used by other agents, and these latter are no longer able to produce some of their products. If the situation does not recover rapidly, this may lead to the loss of other recipes and perhaps to the death of one or more agents; and

- external perturbations, which can be generated from outside, for example by removing an agent or a recipe at a given time step. In this case we compare the evolution in the perturbed vs. unperturbed case.

To observe the system behavior, we can consider several variables. In the following discussion, we consider mainly the total number of artifacts as indicator of the "health state" of the system.

13.5.2 Internal Perturbations

The internal perturbations are the "natural fluctuations" of our system. By observing several runs we can often see that the system approaches a quasi steady state level of the number of artifacts, but we can observe also that there are continuous oscillations around this level, which have very different dimensions. Very often, the deviations are small, but sometimes they are quite large. The suspicion arises that we are observing a sort of "punctuated equilibrium," where perturbations of all scales coexist. Big avalanches can have their cause inside the system, without the necessity of external interference. The succession of big avalanches, interposed by more numerous avalanches of minor magnitude, could give birth to a power law distribution for the fluctuations' magnitude, as it seems to have been observed in several natural phenomena ranging from earthquakes to species extinctions (Bak, 1996).

To observe the magnitude of internal perturbations, we consider 20,000 steps of the evolution of systems with 40 agents. During the simulations, we observe the temporal distance among the peaks (see Fig. 13.7a) and the crises' magnitude (that is, the distance between adjacent maxima and minima – see Fig. 13.7b). The series do not span many orders of magnitude, and, therefore, one cannot claim that the tails surely follow power law behaviors, but the shapes do not contradict such a hypothesis. For the variables shown in the figures the exponents of the best approximating power law are -1.5 and -2.8, respectively.

13.5.3 External Perturbations

Another interesting feature of these systems is their robustness when they are subjected to external perturbations. Again, we consider mainly the total number of artifacts as indicator of the "health state" of the system.

Each experiment is composed of a pair of model runs. Each pair starts with the same initial condition and the same seed of the random number generator: therefore, the two runs are identical as long as they proceed without perturbations. At a predetermined step, a perturbation occurs: starting from that step, the two trajectories diverge. We tested several time points to perturb the system; here we present experiments where the perturbation is introduced at step 800, after the peak that characterizes the greater part of our runs.

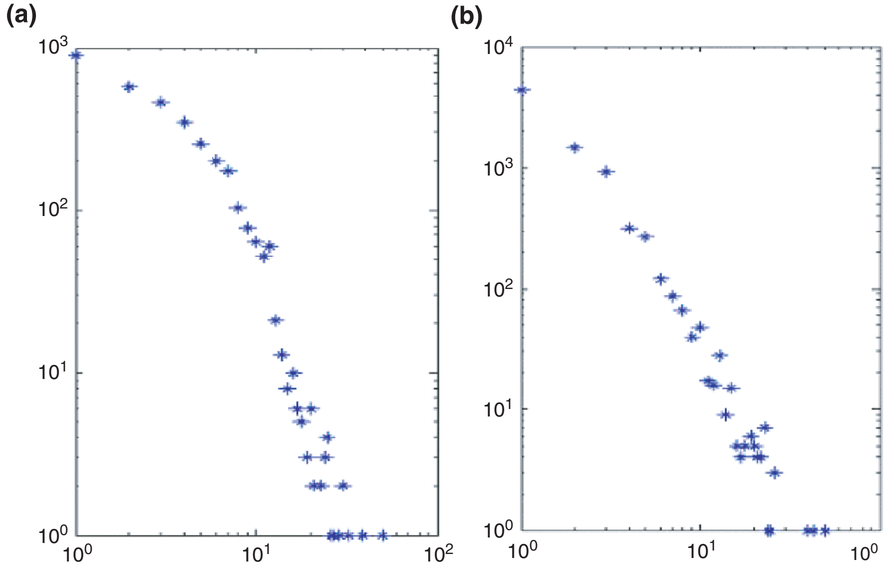

Fig. 13.7 (**a**) The distribution of temporal distances among the peaks of the total number of artifacts of a 40 agents system, during 20,000 steps (log-log scale). Note the power law tail, which has an exponent near -1.5. (**b**) The distribution of crisis' magnitude of the total number of artifacts of a 40 agents system, during a 20,000 step evolution (log-log scale). Note the power law tail, which has an exponent near -2.8

It is possible to choose various measures of the difference between the two systems. In the following, we use the final area interposed between the perturbed and the un- perturbed trajectories (see Fig. 13.8a). Three main kinds of perturbation are considered:

- the deletion of a randomly chosen agent,
- the deletion of a randomly chosen recipe, and
- a different sequence of random numbers (which functions as a control for the magnitude of the differences resulting from the first two cases).

We did 1,000 experiments (each experiment being a pair of runs) for each kind of perturbation. The following phenomena are observed above the level of system noise:

a. the deletion of a randomly chosen agent is an event the system remembers; nevertheless after a while the perturbed system is able to reach a new equilibrium,

   – but the deletion of the first agent introduced into the system is a dramatic event: the perturbed system is not able to reach a new equilibrium and slowly moves to an increasing distance from the unperturbed system;

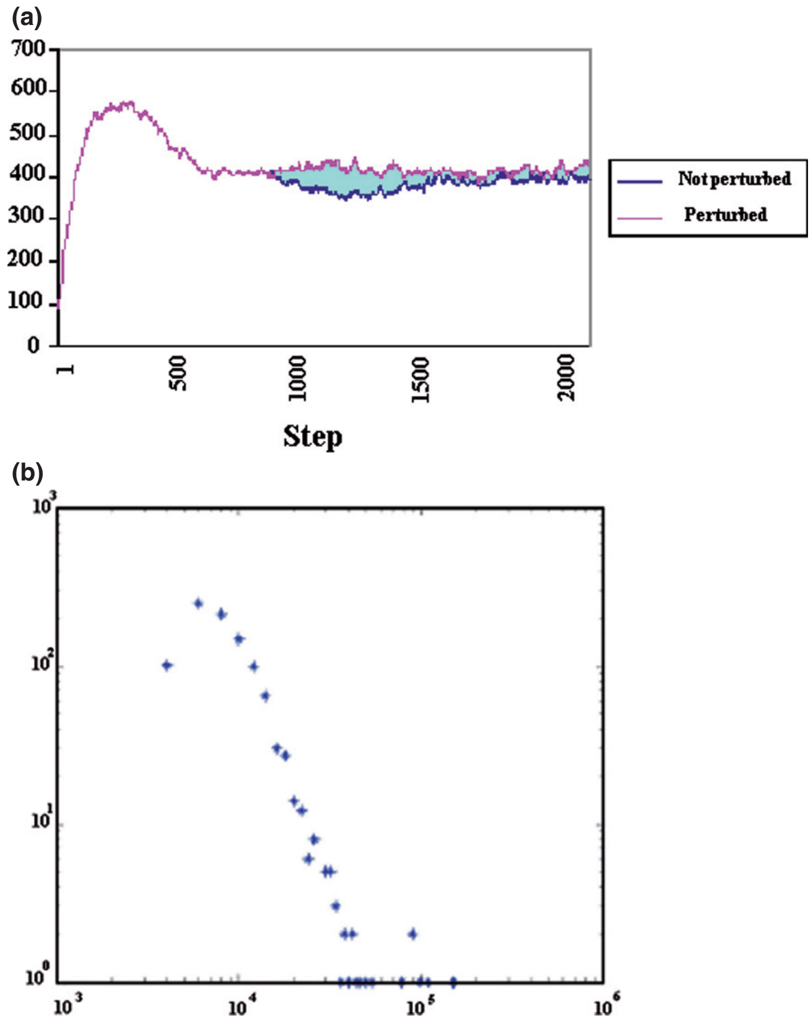

Fig. 13.8 (**a**) The final area interposed between a pair of perturbed and unperturbed trajectories; and (**b**) the final distribution of the final areas interposed between 1,000 pairs of perturbed and unperturbed trajectories (log-log plot)

b. the recipe deletion is an event that does not effectively damage the perturbed system, which rapidly recovers the previous level; or
c. a mere change of the random sequence does not effectively damage the perturbed system, which rapidly returns to the previous level (the results are similar to case (b), thus confirming that the recipe deletion has a small effect)

The fact that removal of an agent has bigger consequences than the removal of a recipe could be explained by remembering that one agent holds several recipes (in fact, the big avalanches are caused by the elimination of the agents that have a number of recipes higher than the average), and that the deletion of an agent cannot be recovered. Even more interesting, it is clear that some agents are more important than other ones (the observation of point a, above): this aspect will be analyzed in depth in Section 13.6. On the other hand, the comparison between the deletion of a recipe and the change in the seed of random generator indicates that the deletion of a recipe is an almost negligible event (points b and c, above): a recipe can be easily rebuilt by the system.

Finally, we can observe that the final distribution of the measures (Fig. 13.8b) shows again a long tail, roughly linear in the log-log scale.

13.6 Stable Relationships

13.6.1 Introduction

A key issue of the set of theories upon which I_2M is built is the agents' ability to create (stable) relationships that foster innovations. In conditions of ontological uncertainty (Lane & Maxfield, 2005), the theory states that successful relationships should be based on the generative potential of the partners. That is, an agent should create preferential relations not with arbitrary agents, but only with the subset of agents that, at that moment, show high potential for generativity in relationships with the focal agent.

In I_2M the agents can maintain two kinds of relationships:

- an agent can "know" another agent (i.e. the first agent knows the existence and the outputs of the second one); this knowledge, if not subsequently enforced by means of a artifact exchange, has the duration of only one step, and
- an agent can have artifact exchanges with another agent (being one of its providers).

The second kind of relation is supported by the agent's recipes, the presence of which guarantees the temporal stability of the relationship. Therefore, exchange of artifacts involves stronger relations than those due to "acquaintances."

These acquaintances allow the agent to locate its own goal and to choose the partner to realize it; therefore, it is useful that advantageous acquaintances be maintained while the less convenient ones be discarded. But how should an agent recognize advantageous acquaintances?

13.6.2 The Importance of Past History

We considered different situations, where the choice of the partners is either random or based upon the agent's history. This latter alternative is implemented as a vote, which each agent gives to each other agent it knows; the vote is a function of the history of the relationships between the two. In general, it is the sum of two terms, one related to the history of buy-and-sell relationships, the other dependent on the history of attempts to develop a joint project (i.e. to share the inputs and recipes in order to reach a common goal).

Let us consider the model results when the vote is only influenced by the projects the two agents did together in the past. In this case, the vote dynamics is very simple:

$$V\,(t+1) = V\,(t) + \Delta_t - \lambda V\,(t)\,, \tag{13.1}$$

where

$$\Delta_t = \begin{matrix} +2 & \textit{if there is a successful join t project at time t} \\ 0 & \textit{otherwise} \end{matrix}\,. \tag{13.2}$$

Unknown agents are given $V(t) = 0$, and occasional acquaintances (agents just known by chance) are given an initial $V(t) = \varepsilon$, with $\varepsilon < 1$. The third term of the equation is a forgetting term, which lowers the vote of those agents that no longer engage in partnerships. When an agent has to choose a partner to collaborate with the intention of realizing its goal, the choice is probabilistic, based on the votes.

The effects of this vote attribution mechanism is shown in Fig. 13.9, where we can see the vote given by a certain agent to the other agents and the relative table of presence/absence of stable collaborations (a stable collaboration being a relationship in which the vote is higher than Δ_t for hundreds of steps). Note that stable collaborations can arise, be interrupted, and later start again. The vote and the partnership mechanisms are able to establish what we can interpret as reciprocal trust (as manifested in reciprocal high votes).

13.6.3 First Results

Now we can compare the results obtained by the voting system with the results obtained by randomly selecting partners. The system has several variables, not all clearly implicated in this change of strategy: the major effects are evident on the variables more strongly involved in relationship processes. For example, it is possible to compute the number of projects (i.e. the innovations that reach the predefined goal by combining the recipes of two different agents) realized by each agent and, subsequently, to plot and compare the resulting distribution. The major effects of the introduction of the voting system are visible in Fig. 13.10a (in which the distributions are calculated by cumulating the results from 30 runs).

For this figure, note that:

1. The median and the average of the two distributions are similar; and
2. The distribution corresponding to the voting system has a more dispersed shape, indicating the existence of a subset of agents able to create more projects than the other one, at the price of blocking the creative action of another, more numerous subset.

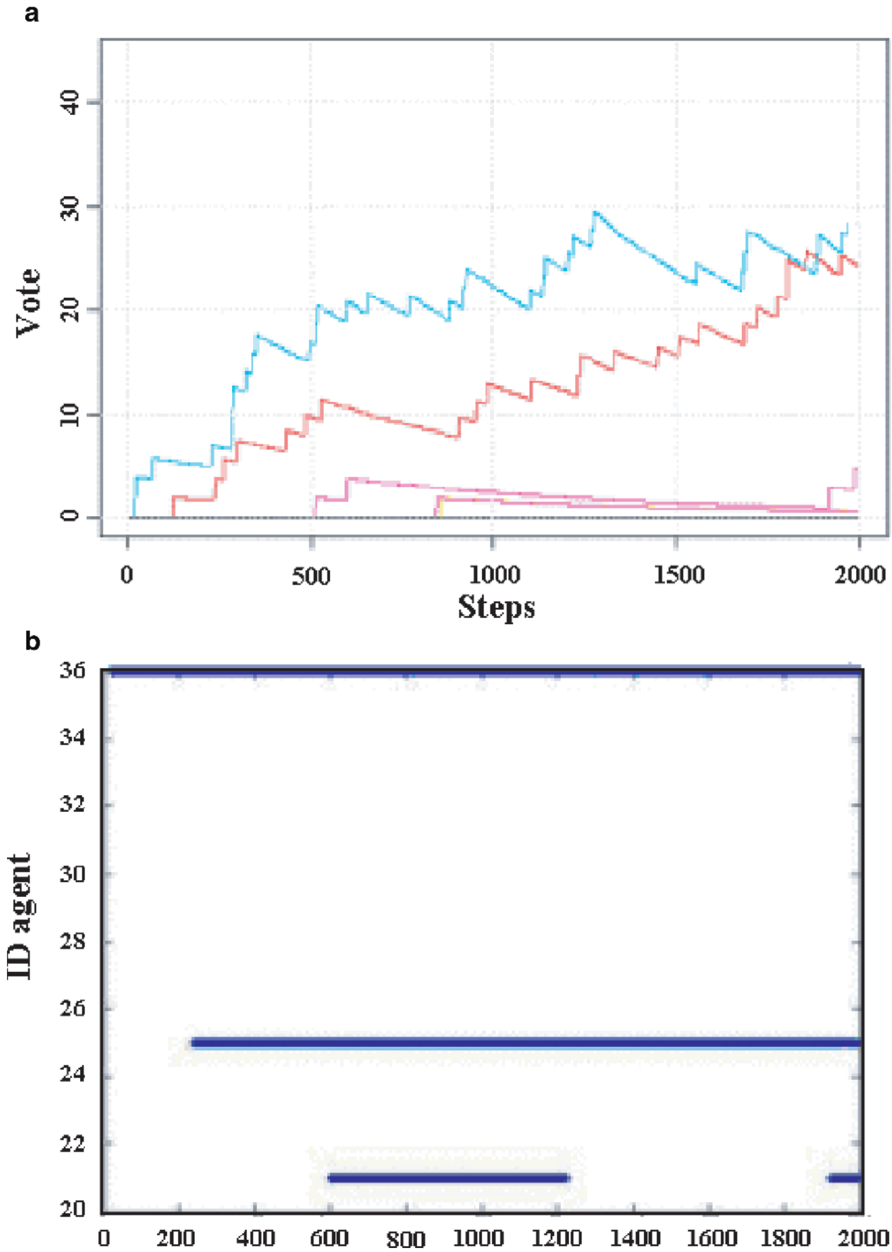

Fig. 13.9 (**a**) The vote given by agent 11 to the other agents during a 2,000 step run; and (**b**) the relative table of presence/absence of stable collaborations (the *blue lines* corresponding to the number of the collaborating agents)

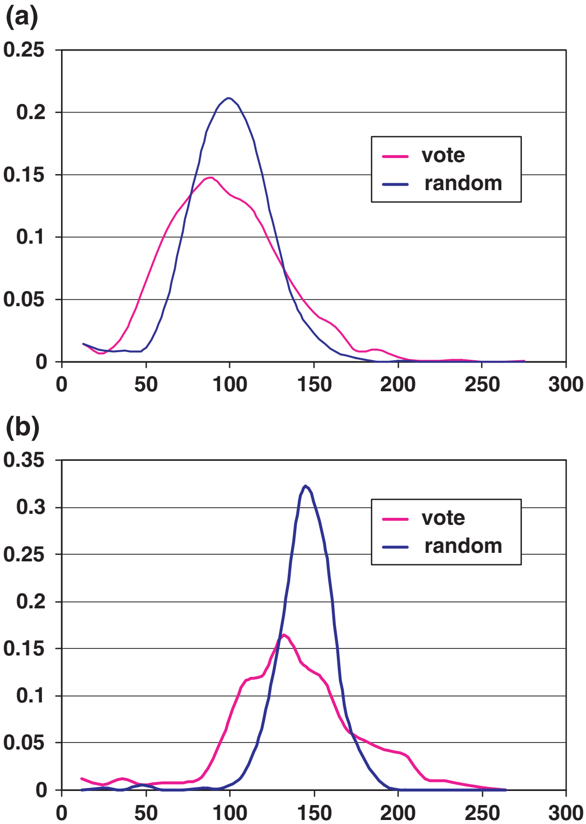

Fig. 13.10 Distribution of the number of projects (i.e. the innovations that reach the predefined goal by combining the recipes of two different agents) realized by each agent, with (*red line*) or without (*blue line*) a voting system. (**a**) Summation of 30 model runs with standard parameters; and (**b**) summation of 30 model runs with rationally limited agents

This second characteristic is strengthened significantly by conditions of ontological uncertainty. The agents of Fig. 13.10b suffer from great reasoning limitations (the performance of their "genetic engines" have seriously deteriorated, passing from a population of 80 individuals and 40 generations to a scheme with 4 individuals and 2 generations). It is possible to observe that:

- To realize their goals, the agents have to make more frequent use of partnerships (the averages of the two distributions increase greatly);
- The distribution of agents randomly choosing their partners is higher and proportionately narrower than in the previous situation;
- The more intense recourse to joint innovation processes enhances the advantages as well as the disadvantages of utilizing the characteristics of the voting system. In particular, the fraction measuring the area between the right tails of the distributions doubles, passing from 7% to 13% (at the price of a similar increase of the area between the distributions' left tails, which passes from 13% to 25%). In other words, the number of agents able to exploit the characteristics of the voting system increases at the price of an increased number of agents badly connected and debilitated.

It is worth observing that, in accordance with the statements of the I_2T, the systems where relationships matter and ontological uncertainty plays an important role foster the existence of particularly connected agents, able to exploit the generative nature of their relationships.

13.7 Structures in Artifact Space

In I_2M, different entities are present, and among these entities, several kinds of interactions take place. Generally, if we have to deal with entities and binary interactions, it is possible to schematize the corresponding systems by means of graphs (networks); several I_2M interactions have a binary shape, and, therefore, in our model, it is possible to define various interesting networks.

In this section, we focus our attention on a particular network that is able to reveal some details regarding the structure of the artifact space: the artifact type network. The artifact type network is a graph where two different numbers are linked if there is at least one recipe within the system that uses an artifact corresponding to the first number in order to build an artifact corresponding to the second number. In this case, the link is directed from the first number to the second one. By means of network analysis, it is possible to address questions regarding the global structure of the systems, by abstracting from the details of each single interaction.

In particular, we are interested in observing in our system the existence of so-called "technological waves," that is, the sequential emergence of strongly interrelated sets of artifacts. A real example of these waves is the succession of the technologies related to "maritime trade" (the epoch of great maritime discoveries), "manufacturing coal and steel" (first industrial revolution), "manufacturing electricity and automobiles" (second industrial revolution), "informational and communication activities" (the contemporary period) (Chapter 8, this volume). The presence of these waves is important because it reveals the possibility of a perpetual novelty, where interacting systems of new kinds of artifacts can substitute already existing ones. With this aim, it is possible to distinguish several interesting questions, as for example:

- Are there technological waves present in our system?
- If so, under what conditions can they appear?

To investigate these questions, we have to identify particular subsystems within the whole artifact network. There are several network analysis techniques able to identify, at least approximately or partially, such kind of organizations: for example, strong component analysis (Scott, 2000), k-core analysis (Scott, 2000), and the percolation of k-cliques (Palla, Derényi, Farkas, & Vicsek, 2005). In the experiments we have discussed so far in this chapter, all these analyses identify only one organization, which emerges during the early simulation steps and is never replaced by other systems (see Fig. 13.11a for an example). This is the typical outcome when the starting materials are very near each other: a very stable world, in which newcomers

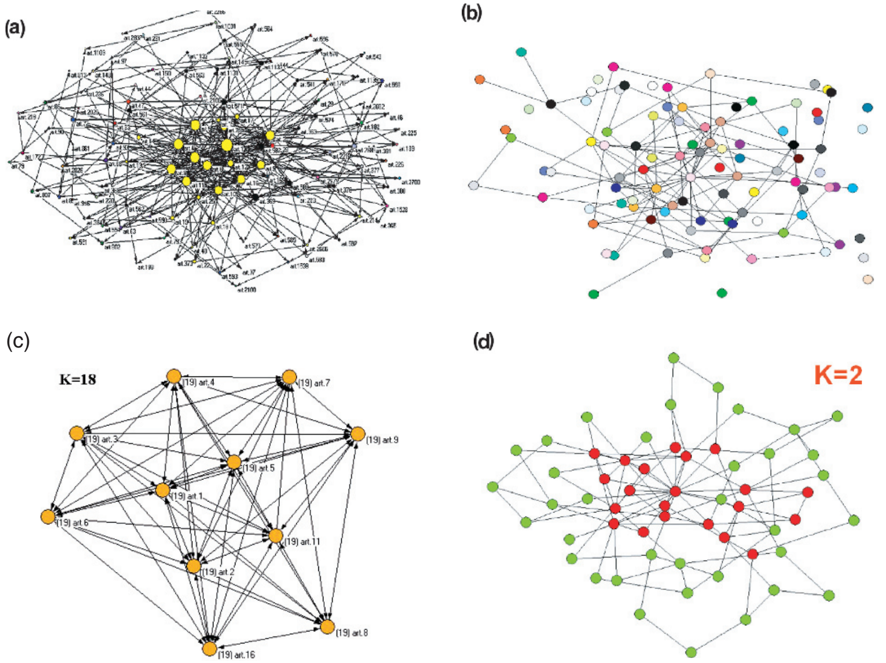

Fig. 13.11 The outcome of several analyses performed on the “names” network. (**a**) Strong component analysis: a huge strong component predominates, embracing almost all the nodes, whereas all other components are formed only by one or two nodes. The vertex sizes are proportional to the number of paths that go through the nodes. (**b**) The same as (**a**), but in presence of two clusters of raw materials: there are no strong components. Each color indicates a different strong component (additionally, the visualization tool has a limited number of colors that are repeated for different components: in effect each strong component on the picture is formed by only one node (**c**) K-core analysis; there is only one core for each value of k, up to the very high value of 18. The identified nodes are the same nodes that prevail in (**a**). (**d**) The same as (**c**), but in presence of two cluster of raw materials: we find a core only for values of k lower than 3 (in (**b**) the maximum values of k is 18)

do not replace, in a wholesale way, the basic structure constructed by already existing artifacts.

However, preliminary simulations of I_2M highlight the possibility that in addressing these issues, it is important to consider the diversity of the starting materials (the raw materials of our system). When these artifacts are not agglomerated in only one cluster, but, in contrast, show more than one preferential position, the final situation is quite different (see Fig. 13.11b, c).

Even more interesting is the evolution of the outcome of the communities identified by means of the percolation of k-cliques. The communities identified by this kind of analysis are the organizations closest to the definition of “technological waves,” and it is possible to observe the subsequent replacement of several communities in time. Last, but not least, we can observe the existence of more than two

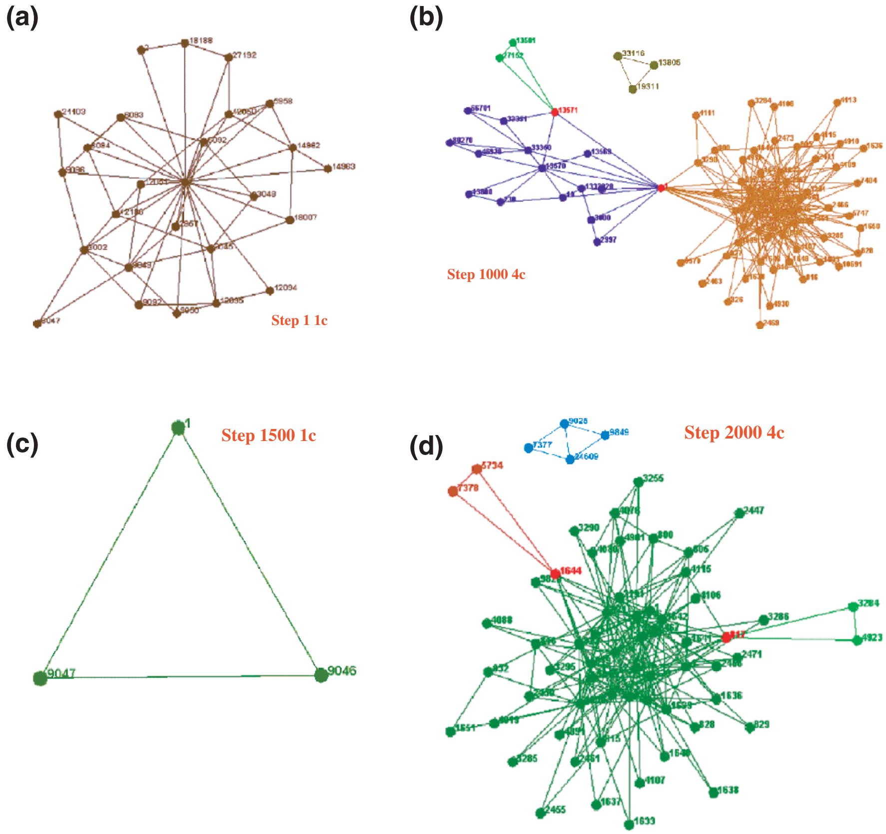

Fig. 13.12 The evolution of the outcome of the communities identified by means of the percolation of 3-cliques. It is possible to observe the subsequent replacement of several communities in time, including the presence of a crisis (**c**) and a subsequent recovery period (**d**). Last but not least, we can observe on occasion the coexistence of more than two communities, also if the initial clusters of row materials are only two: it is a clue toward the existence of non linear effects

communities, even when there are only two initial clusters of raw materials (see Fig. 13.12), indicating the possible existence of strongly non-linear effects.

13.8 Conclusions

After having presented some of its main features, we now consider what kind of model I_2M is. There are, of course, very many meanings of the word model (we learned from our colleagues in biology that for them even a rat or a mouse may be a "model" of a human organism!), but we will confine ourselves to mathematical or computer-based models, to which I_2M belongs.

We can compare it to some selected models, highlighting common features as well as differences, although by no means will we provide an exhaustive review or classification of these various models.

Let us first consider the Alchemy system (Fontana & Buss, 1994),which is loosely inspired by ideas concerning the origin of life and the role of self-maintaining collections of molecules. Alchemy is a kind of "abstract chemistry," where there are functions instead of molecules. The formalism of lambda calculus captures a kind of structure-function relationship that parallels that of molecules.

In the course of time, interesting self-organizing phenomena can take place in the Alchemy reaction dynamics. In particular, cycles are formed, which are composed of functions that catalyze each other's production. When one observes the interactions that take place among these high level entities, one finds algebraic structures whose behavior can be described without making reference to the low level ones that give rise to them.

Spontaneous cycle formation also is observed in I_2M, as described in the previous sections. Apart from technical differences, both models take a very abstract view of the phenomena that stimulated their conception. In a sense, Alchemy's viewpoint is more abstract, as it gets rid of a number of the fundamental features of the biological world that inspired it. On the other hand, it is also more focused on a specific (abstract) kind of question, i.e. those that relate to the formation and interaction of high-level entities. Neither of the models aims at providing meaningful comparisons with actual quantitative data, but, rather, they provide proof-of-principle that interesting emergent phenomena can take place when many microentities interact.

A completely different model is that of a chain of macroscopic harmonic oscillators, where one of them has a mass, M, much larger than that of the others (which have all the same mass, m). In this case, it is possible, assuming that the law of classical mechanics hold, to write a set of exact equations of motion. If we suppose that the initial conditions are not precisely known, the deterministic description is substituted by a law (Liouville equation), which governs the time evolution of the probability distribution of the positions and velocities of all the particles. Let us now suppose that we are only interested in the behavior of the heavy particle, whose motion is slower than that of the others. We can then project the Liouville equation onto the subspace of the variables that describe the heavy particle. This projection takes the form of a perturbative expansion (Serra, Zanarini, Andretta, & Compiani, 1986), in which the leading term (Fokker-Planck equation) describes the heavy oscillator in the same way as a Brownian particle. This approximation is valid on a time scale that is shorter than that of the slow oscillations and fast with respect to that of the light oscillators.

The model provides an approximation to the behavior of the slow oscillator, which is based on an established theory. Moreover, estimates of the accuracy of the approximation, as well as correction terms, are also provided.

Both models (the oscillator chain and I_2M) are based on a theory, but the former is quantitative, while the innovation theory is qualitative and verbal. It is worth noticing that, in both cases, unanticipated consequences of the theory can be observed: this is obvious in the I_2M case, but the emergence of Brownian motion in the oscillator chain also can be considered an unanticipated behavior. We observe that, in both cases, comparison with experimental data is not particularly relevant, although for different reasons. In particular, for the oscillator chain the reliability of the underlying theory provides a firm foundation for the model-based claims.

As a final example of a different kind of model, consider random Boolean models of genetic networks (Kauffman, 1993). These are based upon a description of the phenomenon (the regulation of the expression of various genes in the cell), but, in order to get a manageable model, many simplifications are introduced. Some of them represent approximations to the observed variables (like the use of Boolean functions for gene expression levels or the neglecting of protein dynamics while resorting to a "gene only" model), while other simplifications represent lack of knowledge by introducing randomness (e.g., the choice of the connections among genes is done at random, and also the Boolean functions are chosen at random).

In the case of gene regulation network models, it is possible to draw some general conclusions, like those that concern the relationship between number of different cell types and genome size (Kauffman, 1993), or the distribution of the size of perturbations in expression levels induced by knock-out (Serra, Villani, & Semeria, 2004; Serra, Villani, Graudenzi, & Kauffman, 2007). By comparing the model results with actual observed data, it is thus possible to justify *a posteriori* the simplifications that have been introduced.

Even here, the basic description of the mechanisms of gene regulation may be considered a partial theory of the phenomenon, and the model is rooted in this theory. Similar to I_2M, the model's purpose is that of unfolding the systemic consequences of assumptions concerning the individual gene behavior. The main difference between the two cases is that experimental data are available for random Boolean networks. However, it is interesting to notice that, for about 30 years, the only data of this type have been that concerning the relationship of different properties (cell cycle length, number of cell types) with genome size, and that this is a highly questionable measure, given the uncertainties about the number of actual genes. In spite of this limitation, the model has been widely used as an approximation to the functioning of real cells and as a tool to explore some candidate generic features (e.g., dynamical criticality of real cells).

Briefly, we can observe that the three models considered above, as well as the innovation model, are based on a set of hypotheses concerning the microscopic interactions and are used to explore the global properties that emerge from these interactions. In a sense, every agent-based model of social or economic systems shares this property, and, therefore, is based on an implicit theory of the phenomena at hand. However, most models of this kind do not present a clear distinction between what the theory claims and what the model is and shows. A major part of our effort in the I_2M case is that of providing an explicit discussion of the relationship between model and theory.

The usefulness of this model will ultimately rest on its capability to activate a dialogue with the theory, in order to improve the latter. From this perspective, a major difficulty comes from the existence of several specific mechanisms, which are necessary to make the model work, but which can affect its outcomes. Therefore, one of the major aims of future research will be to disentangle those contributions from others that come from basic theoretical assumptions.

It is also clear that the model still lacks some of the features of the theory: notably, the choice of the partner and the setting of the goal, which might be made more sophisticated than the naïve mechanisms so far introduced. In addition, the drive to innovation is questionable, since our agents are natural born innovators, and are not led to look for novelties due to specific reasons rooted in their particular circumstances. Therefore, the model could be modified by incorporating these characteristics. However, a higher priority should be attributed to a simplification of the model itself, making it free from perhaps unnecessary complications in handling the customer-supplier relationships, to make it more comprehensible without hiding its key innovation mechanisms.

References

Axelrod, R., & Tesfatsion, L. (2006). A guide for newcomers to agent-based modeling in the social sciences. In L. Kenneth, K. L. Judd, & L. Tesfatsion (Eds.), *Handbook of computational economics*, *volume.2: Agent-based computational economics* (pp. 1647–1658). Amsterdam, The Netherlands: North-Holland.

Bak, P. (1996). *How nature works*. New York, NY: Springer.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Cambridge, MA: Massachusetts Institute of Technology Press.

Fontana, W., & Buss, L. W. (1994). What would be conserved if 'the tape were played twice'? *Proceedings of the National Academy of Sciences*, *91*, 757–761.

Gilbert, N., & Terna, P. (2000). How to build and use agent-based models in social science. *Mind and Society*, *1*, 57–72.

Kauffman, S. A. (1993). *The origins of order*. Oxford, UK: Oxford University Press.

Lane, D., & Maxfield, R. (2005). Ontological uncertainty and innovation. *Journal of Evolutionary Economics*, *15*, 3–50.

Lane, D.A., Serra, R., Villani, M., & Ansaloni, L. (2005). A theory-based dynamical model of innovation processes. *ComPlexUs*, *2*, 177–194.

Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature 435*, 814.

Scott, J. P. (2000). *Social network analysis: A handbook*. (2nd ed.). London, UK: Sage Publications Ltd.

Serra, R., & Villani, M. (2006). Agents, equations and all that: on the role of agents in understanding complex systems. In M. Schaerf, & M. O. Stock (Eds.). *Reasoning, action and interaction in AI systems and theories*. *Springer Lecture Notes in Computer Science 4155*, 159–175.

Serra, R., Villani, M. Graudenzi, A., & Kauffman, S. A. (2007). Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. *Journal of Theoretical Biology*, *249*, 449–460.

Serra, R., Villani, M., & Semeria, A. (2004). Genetic network models and statistical properties of gene expression data in knock-out experiments. *Journal of Theoretical Biology*, *227*, 149–157.

Serra, R., Zanarini, G., Andretta, M., & Compiani, M. (1986). *Introduction to the physics of complex systems*. Oxford, UK: Pergamon Press.