Simulating tourists' behaviour using multi-agent modelling

Corniglion, Sébastien and Turnois, Nadine

Université de Nice Sophia Antipolis, Institut d'Administration des Entreprises, Laboratoire CRIFP EA 1195

19 May 2011

Simulating tourists' behaviour using multi-agent modelling

Sébastien Corniglion
Université de Nice Sophia Antipolis
Institut d'Administration des Entreprises
Laboratoire CRIFP EA 1195
Pôle Universitaire St Jean d'Angely
24, Avenue des Diables Bleus
06357 Nice Cedex 4
FRANCE
Email: sebastien.corniglion@unice.fr

Pr Nadine Tournois
Université de Nice Sophia Antipolis
Institut d'Administration des Entreprises
Laboratoire CRIFP EA 1195
Pôle Universitaire St Jean d'Angely
24, Avenue des Diables Bleus
06357 Nice Cedex 4
FRANCE
Email: nadine.tournois@unice.fr

***Abstract*—We discuss who should be in charge of providing data relevant to marketing segmentation for the tourism industry. We describe the difficulties of using the most commonly found consumer behavioural models within an information system, and oppose them to a novel approach in marketing segmentation, based on outgoings analysis. We use agent-modelling techniques, based on cellular automaton rules and stochastic processes to implement our model and generate sales data. We then present our algorithm to identify similarly behaved tourists, showing that the commonly used *"nationality"* variable for segments discrimination is not efficient. We conclude with some test runs results discussion and possible further research tracks.**

***Index Terms*—Simulation, Stochastic processes, Cellular automata, Tourism, Business, Public Policy Issues, Management techniques, Marketing, Market segmentation, Customer behaviour model.**

I. Introduction

This paper presents the outcome of a multi-agents simulation, designed to produce data relevant to the simulated behaviour of tourists. The aim of this generator is two-fold: to produce artificial sales data from the tourists' trail of transactions throughout their stay, for which no actual data is easily available, and, to emphasise the existence of homogeneous groups of *"tourists' behaviour"*, in a market segmentation sense. Discounting the fact that the analysis of tourism activity is still rooted in splitting tourists into well-known variables, such as nationality or revenues, we formed the hypothesis that, although homogeneous behaviour groups may exist, the individuals composing them are not necessarily homogeneous. Hence, traditional models might not be leveraging enough decision-making abilities for devising tourism policies. In the second part of this article, we will discuss who should carry the duty of knowledge gathering with regard to the necessary variables to be fed into a marketing process. In the third section, we will review the most common behaviour models found in the literature, discuss their suitability for the tourism industry and whether they can be computerised. In the fourth part, we will present the simulation techniques used and our behaviour model. In the fifth section, we will describe the algorithmic core of the generator, and particularly how it decides whether a tourist will visit a shop and proceed to buy. The sixth part will detail how the similar "behaviour groups" are discovered, the seventh section will describe the preliminary results obtained and the concluding eighth discusses further possible research tracks.

II. Who does ?

If the tourism industry would follow the prominent definition given by Philip Kotler in his millennium edition of *Marketing Management*, it would be likely to raise more interrogations than actual actionable answers. *"The marketing process consists of analysing market opportunities, researching and selecting target markets, designing marketing strategies, planning marketing programs, and organising, implementing and controlling the marketing effort" (p.86).* Further, Kotler adds *"(and to) transform marketing strategy into marketing programs, marketing manager must make basic decisions on marketing expenditure, marketing mix and marketing allocation" (p.87)* [1]. So why do these clear definitions cause an issue for a touristic destination? Certainly because they have been conceived for addressing to self-contained organisations, whereas the nature of a destination is to be composed by a myriad of different actors, of all economical and financial sizes, only joined together by a common goal: increasing their revenues [2]. Hence, when Kotler proposes to analyse, select, design, plan, etc., the first question one may ask is ***"who does ?"*** A brief look at tourism stakeholders shows:

- Privately owned organisations, *i.e. restaurants, hotels, ...*
- Public sector of various levels, *i.e. local councils, regional assemblies, ministry(ies)*
- Public-Private joint-ventures, *i.e. some museums, historical venues, ...*

Such a diverse structure naturally brings the question of governance, as discussed in [3]. The nature of capitalist, *"free-market"* economies tends to self-justify that a *(highly)* competitive market such as tourism should regulate itself, under fundamentals such as the *supply and demand law* [4], [5] and notion of *perfect competition* [6]. However, the presence of the public sector in the market induces a breach into this perfect competition ideal. Interventions of the *public hand*, nemesis of Smith's *"Invisible Hand"*, *i.e* funding and/or subsidising a museum, distort the competition for privately-owned entities. Subsequently, the neo-classic *free-market* theory would promote abolishing such an unfair competition. However, the financial and following economical events between the 2008-2010 period have favoured a return in grace of the *public hand*. Consequently, some counter-arguments from the academic world have been greatly rewarded for refuting the classic theory, notably using the concept of *"information asymmetry"*, which formalises the studies of decisions in transactions where one party has more or better information than the other [7]–[9]. When observing the tourism ecosystem, this *"asymmetry"* seems to apply to the tourists, away from home and their comfort zone, sometimes from their own country, language and culture. They could form the *"ignorant"* party, who lacks information when agreeing to the terms of a transaction, hence perfectly fitting the definition of the *"adverse selection"*. However, when dealing with marketing a touristic area, we believe the productive actors are the *"ignorant"* party, as nobody, from independent shops to the whole value chain, owns a global view on the tourists purchasing habits, making market segmentation and consequently, target marketing very difficult, if not *impossible*. This situation reminds of the definition of *Constrained Pareto Optimality* [10], [11], where a potential planner may not be able to improve a market outcome, even in a situation of inefficiency. Nevertheless, we will try to demonstrate by the means of computer-aided, agent-based simulation [12], that if an *economically benevolent party* possesses this *global view* and provides it for free to all the *interested stakeholders*, a *Pareto improvement*, making at least one individual better off, without making any other worse off, could be achieved for the value chain. Being economically benevolent is probably not quite reasonable to assume, however the ***public sector***, whom would also potentially gain from a increased touristic activity, seems in our eyes, the best actor to perform the duty. In this paper, we suggest that the sought *Pareto improvement* could be achieved *if and only if* the ***public hand*** **<u>does</u> act** in providing necessary knowledge to all the market's stakeholders for a more efficient marketing policy.

III. Are customer behaviour models IS-ready ?

A. A review of behaviour models

According to Solomon [13], consumer behaviour is *"the study of the processes involved when individuals or groups select, purchase, use, or dispose of products, services, ideas, or experiences to satisfy needs and desires"*. Belch [14] defines it as *"the process and activities people engage in when searching for, selecting, purchasing, using, evaluating, and disposing of products and services so as to satisfy their needs and desires"*. These definitions are qualitative and focus on explaining the previous motivations resulting in a certain behaviour. Consequently, the prominent behaviour models found in the literature provide an analytic framework based on the same motivations. Andreason [15] focused on the importance of information in the decision-making process, but fails in considering attitudes in relation to repeat purchase behaviour. Bettman [16] confirmed this importance of information, but implied that customers would rarely choose a complex decision-making strategy, and would tend to select the simplest one. Blackwell et al. [17] relaxed the decision-process stage, as their model does not constraint behaviour analysis from going through all possible stages. However, in spite of allowing representation of routine and/or complex behaviour, it fails in describing why different types of personalities can produce different decision-making. Goodall [18], [19] includes the notion of destination's image influence as a factor not only influencing the selecting stage of the destination, but also affecting the behaviour of tourists in general. Mansfeld [20] uses a concept borrowed from sociology, namely the *"value-streched"*, which focuses on social classes and their various expectations in *"success values"*. His works focuses also on the choice of destination, which, under his model, variates between some class success expectations and the actual possible choice for an individual. Nicosia [21] proposes to centre on relationships between an organisation and its potential customers, through a dual-way influencing connection. However, it lacks details in explaining internal factors afferent to the personality of the consumer, or how he/she would express his/her attitude towards the products. Middleton [22] provides a model tailored for tourism, based on the *"stimulus-response"* paradigm, and concentrates on explaining the *"buyer characteristics and decision process"*. It emphasises the importance of communication *(the stimuli)* and allows feedback by looping the purchase outputs (*the response*) in a *"post-purchase and post-consumption feelings"* state, re-entering into the decision process. All these models have been built to study and explain behaviour *a posteriori*, once surveys and questionnaires, generally *in situ*, using a typical Likert scale structure [23]. They are not readily usable in an knowledge discovery IS system, as the variables they describe cannot be assimilated to "computer science" variables. Actually, they are a guidance to the researcher in tailor-making questionnaires, hence their openness to interpretations do not provide *(and do not claim to)* a stable data model, fit for being instantiated in an information system's database. In order to build the latter, the standard information system lifecycle applies [24], and requires to understand the business and the product's lifecycle to be computerised.

B. Lifecycle of the "tourism product"

Swarbrooke [25] describes the product as *"complex and multi-layered"*, as it is not only constituted by tangible items *(food, accommodation)* and intangible items *(easiness of road access, cleanliness of the place)*, but also varies in timespan and monetary value. Customers (the tourists) are away from home, hence their behaviour may be influenced and changed from their usual lifestyle [18], [26], [27]. They are looking for an overall *experience*, which includes the pre-trip phase *(anticipation)*, the *global* consumption during the stay, and concludes with the *memory* of the stay, after returning home. The actual production process of the "tourism product" involves many different missions: the shop-owner will aim to make his window as attractive as possible, whereas the local authority will make sure the extra traffic generated will be bearable for both the tourists and the locals. Moreover, the consummer-tourist also influences the production process, as an individual's behaviour may influence the experience of fellow tourists. This is particularly true in the case of a "complainer" who may trigger a complaining atmosphere around him. Nonetheless, all these entities share a same goal in increasing touristic revenues, as discussed by Moutinho [2]. Also, and like many others, a new challenge is facing the tourism industry: the Internet. Bonn, Furr & Susskind [28] and Luo, Feng, & Cai [29] have shown that tourists who searched the Internet to select their destination, but also to gather information before departure, had a tendency to spend more during their stay, compared to other sources of information. The most successful online community for tourism, namely TripAdvisor®, has been shown by Wang & Fesenmaier [30] to be a powerful platform for interaction between peers by notably propagating satisfaction and impacting the revenues of the commented locations. All these elements underline the pace needed for the industry to adapt its offers in order to be as close as possible to its customers' needs, recall our concerns on who provides the needed information and call for a market segmentation, defined by Venugopal & Baets [31] as *"a process of dividing a market into distinct groups of tourists who might require separate experience or marketing service mixes"*. Target marketing, described by Kotler [1], then follows, aiming to develop products and marketing programmes tailored for the found segments. However, we believe the standard process of selecting a model, organising an empirical study and analysing the results, which will mostly concentrate on explaining the inner causes for tourists' behaviour, in order to eventually transform them into actual marketing actions, is both structurally unsuited and too slow for the tourism industry. Hence, we propose to concentrate first on constructing a clear picture of the tourists' buying habits by analysis of their most revealing data: their outgoings. We believe the ability to know the habits segments is more important for immediate marketing action than explaining what caused these habits. This idea was approached by Bloom in [32], where he presented a novel segmenting approach, opposing linear and non-linear analyses techniques. He remarked that most behaviour segmenting analyses are performed using linear methods, where a given segment is characterised by a function affecting different weights to each entry feature *(i.e. variable)*. In spite of the actual tourist behaviour being indeed expressed by these features, they are linked by non-linear relationships, as they interact between each other, *i.e.* length of stay multiplied by the money spent *on a daily basis*. The latter approach allowed to increase by almost twice the segments' coefficient of determination (R^2), using *data mining* techniques. It led us to believe that using classification techniques from the computer science field of *data mining* would allow to describe homogeneous groups of similar buying habits. This type of data analysis is much quicker to set up and run, and would provide profiling elements to tourism policy-makers, who could then adapt their offers to the market needs. Using data mining techniques for the tourism industry has been thoroughly reviewed by Law et al. [33] but was called for further work, mining this field being described at *"infancy stage"*. Out of 174 papers selected from a 28-year span, only 14 were using mining techniques, all solely based on forecasting tourists arrival by nationality, which is not quite the same as identifying segments of similar behaviour, unless taking the hypothesis that the single variable of *"nationality"* is a determinant to create clusters of behaviour. This idea is indeed commonly found through many studies on tourism, being public or private sector-funded. However, we desired to challenge this widely-accepted postulate by choosing, as discussed above, the tourists' outgoings as main criterion of discrimination. Hence, before running our framework with actual sales data, which would require partnerships with many of the industry stakeholders and involve legal aspects for individuals' privacy, we have proceeded in generating artificial sales data, using agent-based computer simulation. Investigating for new management science theories using computer-based simulations have become increasingly popular in the last five years, especially thanks to the formalising work of Davis, Eisenhardt and Bingham [12], who propose a framework for developing such tools, detailed in the next section, and on which we based our simulator.

IV. Simulation, simulators and models

A. Numerical simulations applied to management science

The common lack of empirical data, evoked by Zott [34], is one major argument in favour of creating relevant, however artificial data. Nevertheless, numerical simulations in the management science fields had long been thought to be either too simplistic as they would remove important correlations or too basic as they would only replicate a well-know phenomenon, as discussed by Chattoe [35]. More recently, new agent-based techniques have arisen, where a software could make an *agent* (actor) evolve independently from all the others in either controlled, semi-controlled or random fashion. With respect to these advances, Davis et al. [12] discussed above, have provided an extensive review on computer-based simulation methods used for developing and evaluating new management science theories. Their paper distinguishes five types of simulation, four being targeted for a specific research question:

- *System dynamics*, which focuses on a whole system behaviour with complex timing and causality and looks for what causes system instability.

- *NK fitness landscape*, which focuses on the effectiveness and pace of a system to reach an optimal point and looks for the conditions of a high-performing strategy.
- *Genetic algorithms*, which focuses on adapting a population of agents towards the best type of agent and looks to explain the factors of this best agent.
- *Cellular automaton*, which focuses on creating ruled micro-interactions for observing macro-consequences and looks to explain how, when and how fast a pattern emerges.
- *Stochastic processes*, which focuses on random variations of the input variables and observes how this randomness impacts the agents population.

The stochastic processes have the particularity of not being as specific, but in our opinion, they come as a great help to computerise the most challenging aspect of simulating human behaviour: randomness, or at least "controlled randomness". A stochastic process can be left totally random, or be contrived to a *probability distribution*. In our eyes, using the properties of cellular automata, but loosening their micro-interactions rules via a contrived stochastic process, create a decent proposal to depict human behaviour. In consequence, we chose to use the NetLogo environment[1], which allows implementing cellular automata-type rules at "agent" and "world" level as well as providing a random engine for stochastic processes. The latter is qualified as a *"pseudo-random"* generator, for a computer being a deterministic machine, which is not qualified by definition to provide real randomness. If the discussion of the quality of a pseudo-random algorithm is out of the scope of this article, it is to note that NetLogo provides two types of these:

i. a general *random* functionnality, based on the *Mersenne-Twister* algorithm [36], well accepted as a high-quality random generator, notably passing the DieHard tests [37],
ii. other random functionalities, contrived to probabilistic distributions, namely *gamma, normal, exponentional and Poisson*.

The lack of real-world randomness involved by pseudo-random generators can in fact be turned into a precious feature for running repeatable trials: these algorithms have to be *"seeded"* with a value before they start generating numbers and from this *"seed"* depends the order of the generated numbers. Although seeded by default with the computer clock value, the seed value can be forced into a *hard-coded* one, insuring a simulator to obtain the exact same order of numbers generated for every run. Although we did not use this feature for our work, adding a simple line of code would allow reviewers to repeat the stochastic aspects of our simulator in the exact same conditions.

B. Agents model

Our work deals with two types of agents: the tourists and the "shops", where a shop is to be understood as *any place where a tourist can spend money*. We defined tourist and shop-specific variables, for which each instantiated individual would carry its specific values:

a. ***Tourist***
   i. *Age*
   ii. *Country of origin*
   iii. *Type (i.e. single, couple or family)*
   iv. *Revenue available*
   v. *Length of stay*
   vi. *List of visited shops*
   vii. *Total money spent*
   viii. *List of affinity factors to types of shops*
   ix. *List of money spent per types of shops*

b. ***Shop***
   i. *Type of shop*
   ii. *Total number of visits*
   iii. *List of sensibility factors to countries*
   iv. *List of sensibility factors to age groups*
   v. *List of sensibility factors to types of tourists*
   vi. *Average amount of money spent per visit*

1) stochastic aspects: The simulation distributes randomly initial locations, nationalities, age and duration of stay for the tourists and for the shops' locations.

2) Cellular automaton aspects: For the sake of clarity of results, we have limited the types of shops to seven, namely: *hotel, restaurant, bar, museum, take-away, souvenirs, art gallery*, and same for the tourist's nationalities, limited to eight: *GB, USA, Italy, Spain, Russia, Netherlands, Benelux and France*. Revenues are ruled by indicating an average daily expenditure per nationality. This element could also be left to random, but we believed that using easily obtainable expert insight would make the generated data closer to reality. With regards to the shops, the simulator will apply two rules. First, the sum of hotels, bars and restaurants must hold between $\frac{3}{5}$ and $\frac{4}{5}$ of the total numbers of shops. This assumption has been verified in almost all touristic venues we have been observing from our local area, which is one of the most active in Europe. Then, each type of shop (bar, restaurant, ...) is initialised to an average amount of money spend per tourists' visits, which could also be gained from expert knowledge. However, this variable is adjusted through the simulation process, when a tourist buys in the shop. The different variables of *"affinity factors"* designate dictionaries of real numbers in $[0; 1]$ which can be assimilated to a probability expressing the chance for an instance to be attracted to a target. For example, the list of affinity factors of shops in a tourist-agent, using types of shops as keys, shows the probability of a given tourist to be attracted to a particular type of shop. Finally, all the factors are initialised to zero and will evolve every time a tourist will visit and buy in a given shop *(same applies for the shops' sensibilities*[2]).

[1] http://ccl.northwestern.edu/netlogo/

[2] If the meaning of the tourists' affinity factors are self-explanatory, please note that in terms of management science, the shops' sensibility factors could be assimilated to the notion of ***reputation*** amongst their respective categories

V. SIMULATION'S DECISION CORE

With the agent-model established, we propose the decision core, which focuses in making tourists visit and buy in a shop, to be designed with the same duality between cellular automaton and stochastic processes. The latter allows the unpredictability of human behaviour, while the former forces the decision core to observe some commonly-observed behaviour generalities.

A. Decision: stochastic aspects

1) Tourists: The tourists are "free" of movements around the streets, which are modelled to be pedestrian-only. When reaching a crossing, a tourist has exactly $\frac{1}{4}$ chances to head toward one of the four cardinal points. If a tourist-agent arrives in front of a shop, the core will observe its affinity factors and apply *(see table I)*:

a. *all affinity factors are zero-valued (notably at the start of the simulation)*: the core refers to an interactively-set global constant $VB1$ in $[0;1]$ expressing the probability of not visiting a shop when the tourist-agent has no affinity,
b. *affinity factors are different from zero*: the core refers to an interactively-set global constant $VB2$ in $[0;1]$ expressing the probability of not visiting a shop when the tourist-agent has an affinity for the shop's type.

Hence, the *"dice is rolled"* each time a tourist-agent passes by a shop, no matter its affinities, giving it one out of $VB1$ or $VB2$ chances for visiting. The corresponding pseudo-code is described in algorithm 1. A subsequent purchase would also *not* necessarily affect the tourist-agent's affinity for the type of shop *(it might not have appreciated the shop's experience)*, and the same method of global constant applies in the core for deciding whether it will increase the relevant affinity factor by the money spent[3]. Finally, when a tourist-agent leaves a shop, its new heading direction will also be left to an *equidistributed* random choice.

2) Shops:

a. Once a purchased is settled, the shop feature *Average amount of money spent per visit* will not automatically be affected. An interactively set constant FB *(see table I)*, allows to provide the core with the probability of updating this feature. This allows representation of *influential* individuals, *i.e.* those who would diffuse their experience to the community (TripAdvisor®, ...) .

B. Decision: cellular-automaton aspects

The rules implemented for the tourist and shop-agents contrive their stochastic movements and buying behaviour, in order to fit a real-world generalities.

1) Tourists:

a. $\boldsymbol{\alpha}$ ***condition***: a tourist can only settle one hotel payment during its whole stay. If an hotel has not been chosen by the decision core during the tourist-agent existence, the core will force the presentation of *all the hotels* to the agent through the decision algorithm before its "departure"[4],
b. $\boldsymbol{\beta}$ ***condition***: with regard to the previous rule, if a tourist-agent has access to multiple shops at once *and* has already paid its hotel bill, the latter is removed from the list of possible attractive shops,
c. $\boldsymbol{\lambda}$ ***condition***: a tourist may enter a shop but not necessarily buy anything, as the possible purchase is randomly generated in the interval [0; *Average amount of money spent per visit*] of the chosen shop. This rule allows the natural fact of visiting a shop but not being ***able*** to buy,
d. $\boldsymbol{\mu}$ ***condition***: the possible purchase value is generated randomly in the previously mentioned interval, but contrived to the *exponential distribution*.

2) Shops:

a. $\boldsymbol{\sigma}$ ***condition***: once a purchase is settled, the three sensitivity factors of a shop-agent are raised by the minimised value of the purchase value *(see §V-A1)*, but affected by weights as follows: 0.4 for sensitivity to *age*, 0.4 for *country* and 0.2 for *tourist type*,
b. $\boldsymbol{\tau_1}$ ***condition***: visits in shop-agents other than *restaurants and bars* may only happen between 9am and 7pm,
c. $\boldsymbol{\tau_2}$ ***condition***: visits in shop-agents instances of *restaurants or bars* may happen between the intervals 8am to midnight and midnight to 3am.

Algorithm 1 Visit ($aTourist, aSetOfShops$)

```
Ensure: aTourist is in reaching distance of aSetOfShops
Ensure: (α, β, τ1, τ2) conditions
  for all shops in aSetOfShops do
    Add all shop's sensitivities to aTourist characteristics
    Gather all aTourist's affinities for current shop's type
  end for
  if (All shop's sensitivities = 0 and aTourist's attractions = 0)
  then
    Attrib Random theChosenShop from aSetOfShops
    Roll dice for not visiting theChosenShop using VB1
  else
    Attrib shop with max sensitivities to aTourist in sShop
    if aTourism has any non-zero affinity factor then
      Attrib shop with max attractiveness for aTourist in tShop
    end if
    if sShop != tShop then
      Attrib theChosenShop with
Ensure:     Roll dice 1/4 chances for choosing sShop
    end if
    Roll dice for not visiting theChosenShop using VB2
  end if
  if aTourists visits theChosenShop then
Ensure:   λ condition
    if aTourists buys in theChosenShop then
Ensure:     σ and μ conditions
    end if
  end if
```

[3]increasing affinity factors is performed using a increasing and continuous minimisation function, imaging its results in $[0;1]$, namely $f(x) = \frac{x}{x+10}$

[4]Please note that under this rule, a tourist may not be paying an hotel, which we thought to be representative of people visiting a city without necessarily staying there

Table I
SIMULATION GLOBAL CONSTANTS

Constant	Description
D_MAvgAbsPurch	Maximum deviation allowed around the mean average absolute deviations of purchases vectors
D_MAvgAbsFact	Maximum deviation allowed around the mean average absolute deviations of affinities vectors
D_PearsPurch	Minimum value allowed for the Pearson product-moment correlation coefficients of purchases vectors
D_PearsFact	Minimum value allowed for the Pearson product-moment correlation coefficients of affinities vectors
VB1	Probability of not visiting a shop for a tourist with zero valued affinity factors
VB2	Probability of not visiting a shop for a tourist with non-zero valued affinity factors
LB	Probability of not liking *(i.e. not increasing affinity)* a shop for a tourist after a purchase
FB	Probability of not providing community feedback *(i.e. not increasing shop's sensitivities factors)* of shop by a tourist
VS	Number of vertical streets
HS	Number of horizontal streets
STN	Number of "single" tourists
CTN	Number of "couple" tourists
FTN	Number of "family" tourists

VI. AN APPROACH FOR IDENTIFYING TOURISTS GROUPS OF SIMILAR BEHAVIOUR

As we discussed in section III, our two main concerns with traditional customers' behaviour models were:

a. their inability to be quickly integrable in an information system,
b. their inner motives, which are designed to explain why the customer behaved the way he/she did, which we understand to be of utmost importance, but only after identifying groups of similar behaviour.

Moreover, we evoked the recurrent habit of designing tourism studies by starting to split the tourists using their nationalities. On the contrary, we define a ***homogeneous behaviour group*** as *a set of people, with* ***no regard for their age, nationality, family profile or revenues****, who spend their holiday money in the* ***same types of shops****, and for each type,* ***within analyst-chosen levels of monetary value***. Using this definition, our simulation computes groups of similar behaviour by calculating the mean average absolute deviation of tourists' affinity factors for types of shops as well as their actual purchase values, and, the *Pearson product-moment correlation coefficients* between the two sets of factors vectors. Between a tourist and its potential similar peers, the following must hold *(see table I)*:

a. the ***mean average absolute deviations*** of the *affinity factors* and *purchases values* lists must be of the same dimensions, $\pm$ interactively set global constants, $D\_MAvgAbsFact$ and $D\_MAvgAbsPurch$ respectively,
b. ***Pearson product-moment correlation coefficients*** of the *affinity factors* and *purchases values* vectors must be of the same dimensions, $\pm$ interactively set global constants, $D\_PearsFact$ and $D\_PearsPurch$ respectively.

The *Pearson coefficients* are only used as a tool for ensuring that the values in the vectors conveying the affinity factors and purchases are ordered in the same way, as the sole usage of the mean average absolute deviation could not provide this guarantee when producing equivalent values. Currently, the simulation has been implemented to establish the behaviour groups at the end of days 3, 5, 7, 10 and 14. The process of linking a tourist to its similarly behaved peers is detailed by pseudo-code in algorithm 2. As we will detail with the results in the next section, the interesting aspect of this model is to allow some elasticity on the behaviour similarity definition, by varying the deltas of the four compared variables. In fact, two combinations could be of potential interest in market segmentation:

Algorithm 2 Link Similar Peers

```
Ensure: aTourist has non all-zeros affinity factors
  myAvgAbsDevFact ← MeanAvgAbsDev(aTourist.Affinity_Factors)
  myPersonFact ← PersonCorr(aTourist.Affinity_Factors)
  myAvgAbsDevPurc ← MeanAvgAbsDev(aTourist.Purchases)
  myPersonPurc ← PersonCorr(aTourist.Purchases)
  otherTourists ← all others aTourist and not have zero aff factors
  for all currTourist in otherTourists do
    currAvgAbsDevFact ←
    MAvgAbsDv(currTourist.Affinity_Factors)
    currPersonFact ← PersonCorr(currTourist.Affinity_Factors)
    currAvgAbsDevPurc ←
    MAvgAbsDv(currTourist.Purchases)
    currPersonPurc ← PersonCorr(currTourist.Purchases)
    if abs(myAvgAbsDevFact - currAvgAbsDevFact) <=
    D_MAvgAbsFact
    and abs(myAvgAbsDevPurc - currAvgAbsDevPurc) <=
    D_MAvgAbsPurch
    and abs(myPersonFact - currPersonFact) <= D_PearsFact
    and abs(myPersonPurc - currPersonPurc) <=
    D_PearsPurch then
      Link aTourist with currTourist
    end if
  end for
```

a. a low admitted delta for both of the two mean average absolute deviations (affinities and purchases) would lead to close financial dimension of purchases, as well as similar attractions for the very same types of shops[5],
b. a low admitted delta for the affinities mean average absolute deviations, but higher for the purchases, would lead to

[5]Please note the *Pearson coefficients* have to be set in the exact same high values for these hypothesises to hold.

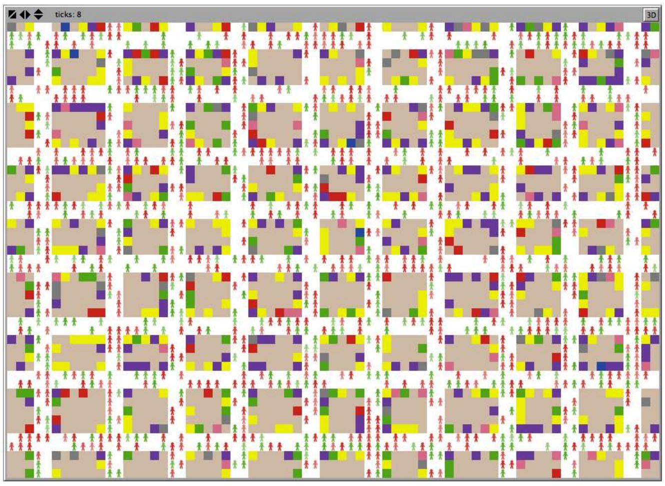

Figure 1. Simulated tourists, shops and city

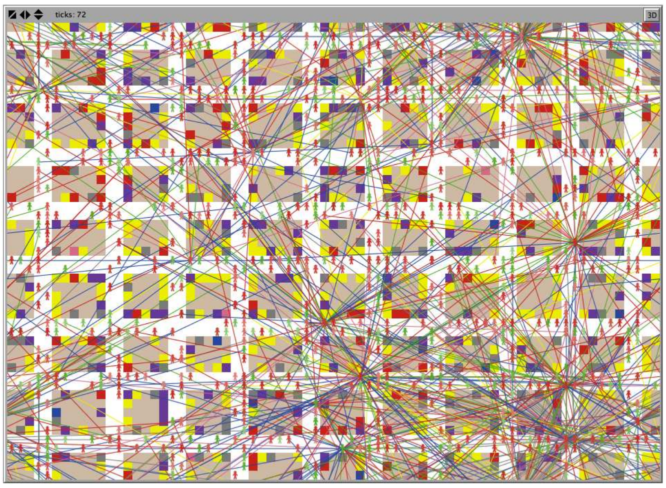

Figure 2. Graphical identification of tourists with similar behaviour

Table II
GLOBAL CONSTANTS VALUES FOR TEST RUNS

Constant	Test Run 1	Test Run 2
D_MAvgAbsPurch	**0.05**	**0.7**
D_MAvgAbsFact	0.05	0.05
D_PearsPurch	0.85	0.85
D_PearsFact	0.85	0.85
VB1	0.6	0.6
VB2	0.3	0.3
LB	0.4	0.4
FB	0.9	0.9
VS	10	10
HS	8	8
STN	300	300
CTN	400	400
FTN	350	350

similar attractions for the very same types of shops, but in a looser financial dimension[5].

VII. RESULTS OBTAINED

The simulation has been implemented using *NetLogo 4.1.1* and offers a graphical interface for running trials. It allows watching tourists moving hour per hour until they 'depart' *(see figure 1)*, entice them to buy with respect to the algorithm seen in *section V* and records every purchase made. On certain simulated days, the program will compute the behaviour groups, with regard to the model defined in *section VI*, and the user can graphically see links appearing between tourists *(see figure 2)*. As the simulator decision-core is influenced by the global constants *(see sections V, VI and table I)*, we used two sets of constants for running our trials as described in table II. These two runs have respectively produced 7,867 and 9,152 lines of data conveying tourists' purchases, tourists' and shops' variables evolution, all globally spanning through respectively fifteen and seventeen simulated days. We discussed in the previous section that changing the acceptable delta of the affinities mean average absolute deviations between tourists would show similar buying behaviour segments, but with different financial dimension. However, we wanted to make sure of the intra-heterogeneousness of the linked tourists, with respect to their inner attributes *(nationality, type, length of stay, ...)*. Hence, for preserving stability in comparing the results, we only modified the global constant *D_MAvgAbsPurch* between the two test runs. For the same reason, the distribution of the tourists' nationalities has been contrived to the same proportions for the two runs, as shown in table III. Tables IV and V confirm our hypothesis held no matter the financial dimension chosen. In fact, setting a larger delta for *D_MAvgAbsPurch* allows an even better stability of the ratio between the total number of tourists linked for a similar behaviour and the subset formed by the ones of different nationalities. We have also observed the actual linked individuals, as shown in tables VI and VII. If their numbers vary greatly with regard to the chosen financial dimension, it is remarkable that the most important period for observing the largest numbers of group leaders spans between three and five days regardless. When marketing a destination, it would imply that the available offer must "hit" the tourists as early as this time interval, to make sure of influencing the group leaders. One could wonder whether the group leaders were the ones who did affect the shops' sensitivities factors, hitting the probability set by the constant ***FB*** *(see table I)*. However, we were not yet able to formally verify this hypothesis, as at the time of the test runs, the simulator was not designed to record whether a tourist was changing a shop's factors. We can however conclude with these emerging patterns:

a. the heterogeneousness intra-group is very high with regard to usual variables, and particularly the *"nationality"* **confirming our primary hypothesis**,
b. the number of groups is large in the first three days of simulation time, and then suffers a drop of almost a half each time the simulation hits the market traditional lengths

Table III
TOURISTS' NATIONALITIES DISTRIBUTIONS

Nationality	Test Run 1	Test Run 2
Benelux	105	105
France	209	209
Italy	158	158
Netherlands	105	105
Russia	53	53
Spain	105	105
United Kingdom	210	210
United States	105	105

Table IV
ANALYSIS OF SIMILARLY BEHAVED TOURISTS FOR TEST RUN 1

Day	A. Behaviour links	B. With tourists of s. nationalities	C. With tourists of diff. nationalities	Ratio $\frac{C}{A}$
3	910	132	**778**	**0.85**
5	498	99	**399**	**0.80**
7	331	65	**266**	**0.80**
10	120	35	**85**	**0.71**
14	7	1	**6**	**0.86**

Table V
ANALYSIS OF SIMILARLY BEHAVED TOURISTS FOR TEST RUN 2

Day	A. Behaviour links	B. With tourists of s. nationalities	C. With tourists of diff. nationalities	Ratio $\frac{C}{A}$
3	9,241	1,351	**7,890**	**0.85**
5	5,798	841	**4,957**	**0.85**
7	3,863	607	**3,256**	**0.84**
10	1,232	216	**1,016**	**0.82**
14	85	12	**73**	**0.86**

Table VI
ANALYSIS OF BEHAVIOUR-CONNECTED TOURISTS FOR TEST RUN 1

Day	A. One link	B. Between 2 and 4 links	C. Between 5 and 9 links	D. Above 10 links
3	32%	**36%**	18%	**14%**
5	**41%**	39%	16%	**4%**
7	**50%**	37%	13%	**0%**
10	**57%**	38%	5%	**0%**
14	**100%**	0%	0%	0%

Table VII
ANALYSIS OF BEHAVIOUR-CONNECTED TOURISTS FOR TEST RUN 2

Day	A. One link	B. Between 2 and 4 links	C. Between 5 and 9 links	D. Above 10 links
3	**10%**	20%	15%	**55%**
5	**15%**	21%	17%	**47%**
7	**18%**	23%	18%	**41%**
10	**22%**	**31%**	19%	28%
14	**51%**	41%	8%	**0%**

of stay (5, 7 and 10 days). Past the 14-day milestone, the drop is above ten times from the 10-day one,

c. a phenomenon of group leaders is clearly observable at any time, as very few tourists get highly connected to a large number of individuals.

VIII. CONCLUSION

This article presents a software designed to artificially mimic the spending behaviour of tourists in a virtual city. It also describes a different behavioural model from the ones traditionally found in the literature, as it is targeted for decision and policy-making rather than sociometric analysis. Our main hypothesis was that usual market segmentation performed for the tourism industry, chiefly using the tourists' nationalities as discriminant, was not accurate. This has been proven thanks to both including stochastic processes in the multi-agent simulation and our expenditure-based segment model. The elements uncovered in the results exploration also support our introductory claim that if a benevolent party had access to the discussed data and would provide it for free to the economically active stakeholders, the leveraged knowledge would help increasing their marketing segmentation, enhancing the general destination's offer, hence contributing to a global *Pareto improvement*. However, this tool is only a first step toward more research investigations. On an marketing research aspect, it would still be desirable to know whether:

1) statistical criteria other than the mean average absolute deviation and the Pearson product-moment correlation coefficients should be used to qualify a behaviour group,
2) the same preoccupation may apply for explaining the "group-leaders" phenomenon,
3) widening the simulation to include tourists with existing *i.e. non-zero* affinity factors **and** larger geographical space than a single city would trigger them to travel to find their likings,
4) on the contrary, the same tourists with limited abilities to travel would see their existing behaviour "change" to *accept* the local offer.

Answering these questions would lead, in our eyes using the presented behaviour model as a base for explaining the resulting segmentation with the traditional, qualitative models we reviewed in section III. In computer science, finding patterns of knowledge in large data volumes can be achieved using data mining techniques. Mining tourism data still being a relatively new task, as mentioned in [33], the common issue of "feature selection", *i.e. selecting relevant variables from the whole available set in order to optimise the results of the mining algorithm*, will certainly be encountered. Another problem lies in the integration of domain expertise in the knowledge discovery task, as discussed in [38]. In future work, we will present a step-by-step framework for the market segmentation of the tourism industry based on:

1) a swarm intelligence algorithm such as the *harmony search* [39] or the more recent *charged system search* [40], used in conjunction with domain knowledge features such as econometrics, for automating the feature selection

process *(inspired from [41])*,

2) a clustering algorithm for discovering homogeneous market segment, *i.e. homogeneous behaviour groups*,
3) a supervised classification algorithm to explain the formation of such groups **and** their leaders.

ACKNOWLEDGMENT

This research has been funded by the French *"Association Nationale pour la Recherche Technique" (ANRT)* and the *SICTIAM*, through the *CIFRE* grant n°1007/2008.

REFERENCES

[1] P. Kotler, *Marketing management*, 10th ed. Prentice-Hall Inc., 2000.

[2] L. Moutinho, *Strategic management in tourism*. CABI Publishing, 2000, ch. Trends in tourism.

[3] F. Gerbaux and E. Marcelpoil, "Governance of mountain resorts in france : the nature of the public-private partnership," *Revue de géographie alpine*, vol. 94, no. 1, pp. 20–31, 2006. [Online]. Available: http://www.persee.fr/web/revues/home/prescript/article/rga_0035-1121_2006_num_94_1_2381

[4] A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, 1776.

[5] A. Cournot, *Recherches sur les principes mathématiques de la théorie des richesses*, 1838.

[6] W. Nicholson, *Microeconomic Theory: Basic Principles and Extensions*, 9th ed. South-Western College Pub, April 2005.

[7] G. Akerlof, "The market for "lemons": Quality uncertainty and the market mechanism," *The Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488–500, August 1970.

[8] A. M. Spence, *Market signaling: informational transfer in hiring and related screening processes*, ser. Harvard economic studies. Cambridge: Harvard University Press, 1974, vol. 143.

[9] J. Stiglitz and B. Greenwald, "Externalities in economies with imperfect information and incomplete markets," *The Quarterly Journal of Economics*, vol. 101, no. 2, pp. 229–264, May 1986.

[10] J. E. Stiglitz, "Pareto efficient and optimal taxation and the new new welfare economics," *National Bureau of Economic Research Working Paper Series*, vol. 2189, December 1988 1988. [Online]. Available: http://www.nber.org/papers/w2189.pdf

[11] J. Hammond, "Four characterizations of constrained pareto efficiency in continuum economies with widespread externalities," *Japanese Economic Review*, vol. 46, no. 2, pp. 103–124, June 1995.

[12] J. Davis, K. Eisenhardt, and C. Bingham, "Developing theory through simulation methods," *Academy of Management Review*, vol. 32, no. 2, pp. 480–499, 2007.

[13] M. Solomon, *Consumer Behaviour*. Prentice-Hall, 1996.

[14] G. E. Belch, "Belief systems and the differential role of the self-concept," *Advances in Consumer Research*, vol. 5, pp. 320–325, 1978.

[15] Andreason, *Attitudes and Consumer Behavior: A Decision Model in New Research in Marketing*. Preston, 1965.

[16] J. Bettman, *An Information Processing Theory of Consumer Choice Reading*, ser. Advances in Marketing. MA Addison-Wesley, 1979.

[17] R. Blackwell, P. Miniard, and J. Engel, *Consumer Behaviour*, 10th ed. South-Western College Pub, November 2005.

[18] Goodall, *Marketing in the tourism industry: The promotion of destination regions*. Groom Helm, 1988, ch. How tourists choose their holidays: An analytical framework.

[19] ——, *Understanding holiday choice*, ser. Progress in tourism, recreation and hospitality management. Belhaven, 1991, vol. 3, pp. 58–77.

[20] Y. Mansfeld, "The "value stretch" model and its implementation in detecting tourists' class-differentiated destination choice," *Journal of Travel and Tourism Marketing*, vol. 4, pp. 71–92, 1995.

[21] F. Nicosia and N. Robert, "Consumer behavior: Toward sociology of consumption," *Journal of Consumer Research*, 1976.

[22] V. Middleton, *Marketing in Travel and Tourism*. Butterworth-Heinemann, 1994, ch. A Stimulus-Response Model of Buyer Behavior, pp. 104–112.

[23] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 1–55, 1932.

[24] J.-L. Cavarero, "Lapage: un modèle et un outil d'aide à la conception de systemes d'information," Ph.D. dissertation, Université de Nice-Sophia Antipolis, 1979.

[25] J. Swarbrooke and S. Horner, *Consumer behaviour in tourism*. Oxford: Butterworth-Heinemann, 1999.

[26] Y. Mansfeld, "From motivation to actual travel," *Annals of Tourism Research*, vol. 19, pp. 399–419, 1992.

[27] C. Cooper, J. Fletcher, D. Gilbert, and S. Wanhill, *Tourism: Principles and practice*. Pitman Publishing, 1993.

[28] M. Bonn, H. Furr, and A. Susskind, "Using the internet as a pleasure travel planning tool: An examination of the sociodemographic and behavioral characteristics among internet users and non-users," *Journal of Hospitality & Tourism Research*, vol. 22, pp. 303–317, March 1998.

[29] M. Luo, R. Feng, and L. Cai, "Information search behavior and tourist characteristics: The internet vis-à-vis other information sources," *Journal of Travel & Tourism Marketing*, vol. 17, pp. 15–25, Febrruary-March 2004.

[30] Y. Wang and D. R. Fesenmaier, "Towards understanding members' general participation in and active contribution to an online travel community," *Tourism Management*, vol. 25, pp. 709–722, June 2004.

[31] V. Venugopal and W. Baets, "Neural networks and statistical techniques in marketing research: A conceptual comparison," *Marketing Intelligence and Planning*, vol. 12, pp. 30–38, July 1994.

[32] J. Bloom, "Tourist market segmentation with linear and non-linear techniques," *Tourism Management*, vol. 25, pp. 723–733, June 2005.

[33] R. Law, H. Mok, and C. Goh, "Data mining in tourism demand analysis: A retrospective analysis," in *Advanced Data Mining and Applications*, ser. Lecture Notes in Computer Science, H. L. X. L. J. Z. O. Alhajj, Reda; Gao, Ed. Springer Berlin / Heidelberg, 2007, vol. 4632, pp. 508–515.

[34] C. Zott, "Dynamic capabilities and the emergence of intraindustry differential firm performance: insights from a simulation study," *Strategic Management Journal*, vol. 24, no. 2, pp. 97–125, 2003.

[35] E. Chattoe-Brown, "Just how (un) realistic are evolutionary algorithms as representations of social processes?" *Journal of Artificial Societies and Social Simulation*, vol. 1, 1998.

[36] M. Matsumoto and T. Nishimura, "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator," *ACM Trans. Model. Comput. Simul.*, vol. 8, pp. 3–30, January 1998. [Online]. Available: http://doi.acm.org/10.1145/272991.272995

[37] G. Marsaglia, "DIEHARD: a battery of tests of randomness," *See http://stat.fsu.edu/ geo/diehard.html*, 1996.

[38] L. Brisson and M. Collard, *How to Semantically Enhance a Data Mining Process?*, ser. Lecture Notes in Business Information Processing. Springer Berlin Heidelberg, 2009, vol. 19, ch. 3, pp. 103–116.

[39] W. G. Zong, *Music-Inspired Harmony Search Algorithm Theory and Applications*, ser. Studies in Computational Intelligence. Springer Berlin / Heidelberg, 2009.

[40] A. Kaveh and S. Talatahari, "A novel heuristic optimization method: charged system search," *Acta Mechanica*, vol. 213, pp. 267–289, 2010, 10.1007/s00707-009-0270-4. [Online]. Available: http://dx.doi.org/10.1007/s00707-009-0270-4

[41] E. Alexandre, L. Cuadra, and R. Gil-Pita, "Sound classification in hearing aids by the harmony search algorithm," in *Music-Inspired Harmony Search Algorithm*, ser. Studies in Computational Intelligence, Z. Geem, Ed. Springer Berlin / Heidelberg, 2009, vol. 191, pp. 173–188.