Business models for the Web: an analysis of top successful web sites

R. Garigliano, L. Mich, P.L. Novi Inverardi
Business models for the Web: an analysis of top successful web sites

R. Garigliano, L. Mich, P.L. Novi Inverardi
DISA Working Papers
The series of DISA Working Papers is published by the Department of Computer and Management Sciences (Dipartimento di Informatica e Studi Aziendali DISA) of the University of Trento, Italy.

Editor
Ricardo Alberto MARQUES PEREIRA ricalb.marper@unitn.it

Managing editor
Roberto GABRIELE roberto.gabriele@unitn.it

Associate editors
Michele ANDREAUS michele.andreaus@unitn.it Financial and management accounting
Flavio BAZZANA flavio.bazzana@unitn.it Finance
Pier Franco CAMUSSONE pierfranco.camussone@unitn.it Management information systems
Luigi COLAZZO luigi.colazzo@unitn.it Computer Science
Michele FEDRIZZI michele.fedrizzi@unitn.it Mathematics
Andrea FRANCESCONI andrea.francesconi@unitn.it Public Management
Loris GAIO loris.gaio@unitn.it Business Economics
Umberto MARTINI umberto.martini@unitn.it Tourism management and marketing
Pier Luigi NOVI INVERARDI pierluigi.novinverardi@unitn.it Statistics
Marco ZAMARIAN marco.zamarian@unitn.it Organization theory

Technical officer
Mauro MION mauro.mion@unitn.it

Guidelines for authors
Papers may be written in English or Italian but authors should provide title, abstract, and keywords in both languages. Manuscripts should be submitted (in pdf format) by the corresponding author to the appropriate Associate Editor, who will ask a member of DISA for a short written review within two weeks. The revised version of the manuscript, together with the author’s response to the reviewer, should again be sent to the Associate Editor for his consideration. Finally the Associate Editor sends all the material (original and final version, review and response, plus his own recommendation) to the Editor, who authorizes the publication and assigns it a serial number.

The Managing Editor and the Technical Officer ensure that all published papers are uploaded in the international RepEc public-action database. On the other hand, it is up to the corresponding author to make direct contact with the Departmental Secretary regarding the offprint order and the research fund which it should refer to.

Ricardo Alberto MARQUES PEREIRA
Dipartimento di Informatica e Studi Aziendali
Università degli Studi di Trento
Via Inama 5, TN 38122 Trento ITALIA
Tel +39-0461-282147 Fax +39-0461-282124
E-mail: ricalb.marper@unitn.it
Business models for the Web: an analysis of top successful web sites

R. Garigliano, L. Mich, P.L. Novi Inverardi
Department of Computer and Management Sciences, University of Trento, Italy

Abstract
To investigate successful web business models, an original multidimensional framework is defined and applied to a large number of web sites. The framework – named BM*Web – combines issues already present in existing schema describing business models, with innovative aspects that have not previously been taken into account in those combinations or which are now viewed in a new light. Results of the application of BM*Web to the 500 top list of Alexa (at a specific time) highlight an articulated picture where more than one success profile exists and not all of them include a web community, although a strong relationship exists between community and success under some conditions. The identification of features that characterize the most successful business models for the Web could be used to define guidelines for company management, once the appropriate profile for a company has been recognised.

1. Introduction
The concept of business model has been investigated intensively by many authors, especially after the inception of the Internet and the Web with their enormous potential for organizations and users. Research on the definitions and classifications on web-based business models, also called e-business models, addresses questions related to the scope of a business model and its main components, usually attempting to define a taxonomy (see for example, (Timmers 1998; Rappa 2009; Bartelt and Meyer 2001) or even an ontology (Osterwalder 2004) that allows the full description of existing and (possible) future models. Not many studies focused on the characteristics of successful business models, among them (Malone et al. 2006) and (Pereira and Fife 2000).

To investigate web-based business models, we propose an original framework based on four dimensions (variables), identified using a linguistic analysis of the concepts involved in the original question ‘what makes a web business successful?’ The analysis, in a brief form, works as follows. ‘Business’ means organised production/exchange for profit: this entails features of the form of market (who deals with whom) and the nature of income. ‘Web business’ means that the features to be looked into must be those which differentiate web businesses from online ones, and the way in which they do so. At the same time, features which by necessity belong to all web activities are clearly of no interest, since they provide no division in the concept. ‘Successful’ in the business context means something related to company value (usually some multiplier of revenues or profits): however, in the web context, a reasonable proxy for value has been some form of traffic measure (users, unique visits, length of access etc.), since this usually acts as a multiplier on both the income form and the market value (the latter in cases where the model has yet to produce a profit stream).

Such linguistic analysis corresponds to the following questions:

1) Who is the business for and what are the roles played by the actors of the business transactions or relationships?
2) Does the web site technology satisfy customers’ needs in an innovative way or is it a copy of an off-line business? And then, if the web site represents a new way to satisfy customers’ needs, are these needs organized or rendered explicit in a new, creative way?
3) Which types of incomes are there on the web site?
4) Which community levels are there on the web site, that is, if there is a web community, what is exchanged among users and also how are their interactions controlled?

Each variable was defined in a first version of the framework – named BM*Web - according to objective and client-side evaluable parameters or values. The first version of the BM*Web framework was applied to a set of about 100 web sites in order to check its feasibility (Garigliano and eTourism Research Group 2006). Taking into account the insights obtained with that study, a second version of the framework is proposed here, in which the four variables are defined in an unambiguous way.

For the research described in this paper, the framework has been applied to a set of web sites, identified by extracting from the free list of Global Top 500 web sites given by Alexa (www.alexa.com, 10 June 2007) those in English or with an English version, without pornography or illegal content; for portals, a procedure to take off the most relevant sections was applied. The final set was then integrated with a few web extra sites to be able to validate the framework on all its variables.
The goal of the research was twofold: (1) at the theoretical level the new version of the four dimensional framework had to be validated for its applicability; (2) results of the application of the framework had to be analysed to find out critical success factors characterising successful business on the web. Some preliminary results were described in a conference paper (Garigliano et al. 2008).

The structure of the paper is as following: next section introduces related work to establish the context in which the BM*Web framework was defined; section three describes our approach. The application of the BM*Web framework on the web sites and main results of the study are reported in section 4; finally, section 5 presents directions for future work and open questions.

2. Related Work
Research on web-based business models addresses questions ranging from the definition of business models to the changes due to the inception of the Web and its impact on the Economics fundamentals.

Papers published on business models on the Web (or e-business models) after the .dot boom and bust that started from the late '90 share two main goals: (a) to investigate the concept of business model and (b) to investigate the changes due to the Web and its impact on the Economics fundamentals.

As regard the first question, a definition that fits most of the approaches is that given in (Feng Li 2007) “a business model is an architecture for product, service and information flows, including a description of the various business actors and their roles, a description of the potential benefits for the various business actors, and a description of the sources of revenue”. Basic concepts and definitions to address the challenges for web-based business models are described for example in (Meier and Stormer 2009).

For the second issue, positions range from those that believe that Internet and the Web are dramatically changing business rules and economics theory (e.g., Merrifield (2000) and Wood (2000) and most recently Tapscott and Williams (2006)), introducing expressions like ‘new economy’, ‘now economy’ and ‘wikinomics’ (the last one stressing the so-called Web 2.0 business models, based on the online contributions and collaboration of companies and users), to those that assume that emerging e-business models are an evolution of existing ones that can and do co-exist with them (Markides 2008).

The ultimate question for companies is ‘if and how to evolve their business model’ - which are the driving forces for emerging and successful business models, which organisational, financial and technical changes would be more effective - and answers require information to support such an important decision (Afuah and Tucci 2002), (Finnegan and Hayes 2008) as there is not a one size fit all business models (Roberts and Toleman 2007). Among recent contributions, we can cite (Weiss and Amyot 2005) and (Debelak, 2003).

Existing categorizations for web-based business models are founded on a set of criteria of differentsizeand complexity. Some of them are organised according to a limited number of elements, e.g., the type of business area, the technology used in the web site, etc. Many authors underline organizational aspects, e.g. the first papers on e-business models proposed sector- or products/services-based classifications; but these classifications are often unable to take into account emerging web-based business models. This kind of classifications adopt a ‘top down’ style in which existing e-businesses are named accordingly to off-line equivalent models or to the most relevant superficial aspects in the business; for example in (Rappa 2009) there are 9 generic business models, that are then articulated in 41 different sub-categories, according to the companies’ value proposition and the revenues. Most articulated schemas do integrate different sets of parameters introducing other economics concepts, for example connecting business models to the Porter’s activities of the value chain (e.g., (Zeng and Huang 2004)).

Other authors have investigated specific online business models, suggesting different components for their identification and evaluation. For example, in (Shin and Park 2009), the authors propose to analyze business process and customer value to introduce variants in auction business model. For our purposes – analysis of large sets of online business models - these approaches would not be feasible because they lack generality and do require data on companies that cannot be extracted from their websites.

A comparison among the most relevant contributions is given in (Pateli and Giaglis 2004). For our goal it is important to underline here those that refer to the four variables of the BM*Web framework, namely: (1) Market of reference; (2) Income; (3) Needs vs. technology; (4) Web Community exchanges and controls.

As regard to the first variable, Market of reference, or markek forms, there are commonly used values related to a classification that consider 3 or 4 classes of markets: Business to Consumers (B2C), Business to Business (B2B), Consumer to Consumer (C2C), Consumer to Business (C2B) (this last is rarely used). However, acronyms made up of only two parts, sometimes adapted to underline government actors, G2B or B2G, are not able to distinguish business models that imply more specific roles of businesses, especially in relation to web communities, so that three roles acronyms have been introduced more recently (www.wordspy.com).

Among the parameters or values used to describe a business model, financial aspects are taken into consideration by all authors as they are deeply related to the business model concept. The simplest definition of business model is “how you planned to make money” (Lewis 2000). For our approach, we have defined the Income variable according to an almost shared set of values; the most critical point was to adopt a level of detail adequate to address the trade-off between being able to evaluate such variable from the client-side and to get useful information.
The third variable, Needs vs. technology, is defined in respect to how the technology satisfies known or unknown users’ needs or requirements; it is important to remind that innovation and changes due to ICT and Internet have been investigated since the first generation of Information Systems and there are many researches and studies in this area; trends also related to Web-based Information Systems and the strategic roles assumed by web sites to organisations have also fostered the birth of a new discipline, Web engineering (Ginige and Murugesan 2001). As regarding the relationship between innovation and business model, Timmers (1999) was one of the first authors that classified business models according to innovation and functional integration. However, none of the existing model does explicitly relate the role of technology to the ‘satisfaction’ of expressed or unexpressed needs of the customer. In this way it can be analyzed also for companies that could not be directly involved in the study. Finnegam and Hayes (2008) describe the innovation related to a specific e-business model as “the extent to which processes can be performed via the Internet that were not previously possible”. They also distinguish internal and external processes affected by the application of Web-based functions. However, analysis of these aspects could not be accomplished without direct involvement of the companies and would also been time and resource consuming.

The last variable introduced in our approach, that is the presence of a Web community - most often referred as virtual or online community - is described as relevant in many classification schema (among them, (Timmer 1998; Rappa 2009; Weill and Vitale 2001)). However a common trend is to consider it as one of the elementary e-business models, or as a business model in itself. This assumption does not allow the analysis of the large variety of successful community-based companies in different sectors, with different goals, with different roles for the participants and with mechanism to manage all the issues related to trust, security and knowledge sharing.

Tappcott and Williams (2006) identified seven new models of mass collaboration that are completely changing scenarios for large and small companies according to their motto, ‘collaborate or perish’, towards the creation of the collaboration economy: peer pioneers, ideagoras, prosumers, new alexandrian, platform of participation, global plant floor, wiki workplace. In all these models web technologies are used to change the role of participants – companies and customers: however, authors do not analytically explain the differences among the identified models. Also, there are not systematic studies about the role of web communities for existing businesses. It is worth naming here the Forum One Communication (2007) that started to analyse the community ROI (Return on Investment); according to the last survey, only 22% of respondents (companies whose business model is based on one or more web communities) had clear ROI Model, but establishing a ROI model was a priority for most of them in the near term. Other statistics gathered by the Forum One Communication support the economics advantages of web communities, e.g for community users vs. non-community users it stands out that they: (a) remain customers 50% longer; (b) spend 54% more; (c) visit nine times more often; (d) have four times as many page views; while in customer support, live interaction costs 87% more per transaction on average than forums and other web self-service options; and cost per interaction averages $12 via the contact center versus $0.25 via self-service options (http://redplasticmonkey.wordpress.com/2007/05/08/online-community-roi).

An approach complementary to the one we propose in this paper is given in (Samavi et al. 2008). Authors introduce an ontology to represent business models according to goals, intentions, roles and rational of a strategy, allowing to compare different strategies and obtaining data to support top management decisions. The BM*Web framework instead has been applied to gain some insights to the success factors for online business models offering the managers information about company online strategies.

With a similar goal, defining a framework for assisting decision makers assess the potential of e-business models, in (Finnegam and Hayes 2008), a rich conceptual framework for the classification of e-business models is proposed. However, the framework has not been empirically tested to validate its structure.

### 3. Analysis scheme for business models on the web: BM*Web

On the Web it is possible to identify a very large number of measured variables, e.g. the ones required by models for quality of websites (usability, contents etc.), or those resulting from a ‘type of goods’ analysis. However, features which can distinguish online and offline models, as well as traditional and new online models, are considered the most useful in this study.

A principle of choosing discriminants which are simple and coherent has been followed. Four variables have been introduced, with the following properties:

- able to mark the features which distinguish models on the Web from those off-line;
- based on observable criteria, in order to guarantee the repeatability of the analysis;
- observable directly from the analysis of the web sites (client side) without a need to contact directly the companies involved (so that it is possible to apply the model to a large number of web sites).

The definition process for the schema is as follows: iteratively, starting from a set of values for each variable, an initial version of the model is applied to a set of 100 sites, by several markers operating independently; then, in the light of problems of different interpretation or for the need of adjusting the variable definitions, the process
is repeated until a stable version is reached. This process has been followed first by two operators in an initial phase on a smaller set of online companies, and then by three different operators on the full set.

The variables and values of BM*Web applied to a large scale study are as follows:

1) **Market of reference**: it distinguishes 7 values to specify the role of the online company, especially in its relationship with customers or other companies. This is done by analysing the goods and services flow. As well as the traditional categories B2B, B2C, C2C, C2B, the model considers also situations in which the company plays an intermediation role: B2(B)2C, C2(B)C, C2(B)2B.

2) **Income**: it indicates the form of revenues of the site. It is assumed that a business must have at least one form of revenue to survive. The BM*Web schema divides this field into 5 values: subscription (a payment to use service over a period of time); intermediation (a percentage of the value of the exchange kept by the site in order to provide support such as security, privacy etc.); advertising (in any form: banners, pop-up, overlies, integrated advertising such as premium ranking for search engines, affiliations and partnerships); single payment (a one-off payment to buy a single item or service); and other (donations, secondary income from support to another activity, membership tax, start-ups expecting to be acquired). For each site, all visible forms of income have been categorised, both principal and secondary income forms.

3) **Need vs. technology**: this variable has been introduced to verify if the web technologies are used in a site to satisfy users’ needs in an innovative way, or if the online offer is a direct transposition of the offline model, plus the standard web facilities. This aspect was already present in the work by Timmers in 1998: however, at this it is now added an analysis to find out if the user’s need was already explicitly recognised in the offline environment, or if it appeared because of the Web. This produces an interplay between two factors: technological effects and users’ needs. Three cases are recognised: the online business is a copy of the offline one; it is a new answer to needs which were already recognised; or it is a new answer to new needs.

4) **Community**: communities are a relatively recent development on the net and they have been included in the model because they appear to be significant in relation to the business model on the Web. To have a community, it is necessary that there is an active space, supported by the interaction between participants. The type of exchange is classified as follows:

- Information, in all formats: text, images, audio, video etc.;
- Commercial, economic transactions;
- Complex virtual reality experiences, in which as well as exchanging information, physical goods and money, members exchange social experiences, virtual goods etc.

Also this feature represents the level at which such exchanges are monitored, moderated or controlled. Three levels of control are considered, which can be exercised by members and/or by staff: minimal, light or specific.

- Minimal control: the legal obligations are stated, but there is no attempt to enforce them unless there are specific complaints;
- Light control: there is an attempt by automatic software, members of staff or members to monitor most of the content, but no guarantee that the monitoring takes place or that appropriate remedial actions are taken immediately;
- Specific: structures are in place (either by staff or senior members) that guarantee the application of standards, with appropriate remedial actions. Usually this requires hierarchies of members, with different monitoring powers, and explicit sanctions.

Changes in the final schema with respect to the initial one are:

- Intermediation as a market type is recognised only if the money flow goes somehow through the company (even only as a guarantee).
- In the Income variable, affiliation has been moved from ‘other’ to advertising, in order to reduce the initial list which was too detailed with statistically insignificant elements.
- For all the variables, objective criteria have been defined which would allow a different team to repeat the experiment on the same or different data.

Alltogether the BM*Web framework is now constituted by 4 variables and 19 values (Table I).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>Subscription (SUB), Intermediation (INT), Advertising (ADV), Single payment (S.P.), Other</td>
</tr>
</tbody>
</table>
4. The study
Part 4 describes the way in which the BM*WEB framework has been applied and tested against a substantial amount of real data. This process is described following the stages of the analysis.

- In 4.1 the construction of the sample is explained. The process is described in details, as the sample is of crucial importance, and many important choices had to be made.
- In 4.2 the BM*WEB framework has been applied to the data. This has been an iterative process, as the framework itself has been adjusted in the process when the initial attempt at application showed vagueness or imprecision in the original definition.
- In 4.3 to 4.5 the results of the final application of the scheme are shown. They are subdivided in three parts: descriptive analysis, multivariate analysis and illustrative values analysis.

4.1 The construction of the sample
In order to obtain a successful web sites sample on which a qualitative and quantitative analysis should be carried out (without the distortions due to an arbitrary selection process), it has been decided to use a nonprobability sampling schema of sites classified according to their popularity (in which the units have a common feature relevant to the research goals). Non probabilistic samples are needed when not all the features of the population – in this case the totality of web sites - are known; they can be used in order to perform statistical analysis of an explorative type (since the units do not have the same probability of being selected, as it is the case in random samples, they do not allow inferences on the general population). The sampling process took place in summer 2007: the 10th June the list of the 500 most visited sites has been downloaded from Alexa: in September 2007 the list of web sites to be analysed with the BM*Web framework has been completed. The choice of a third party agent, Alexa, has allowed us to obtain the list for free: Alexa builds the list of the most visited sites from the surfing habits of millions of users who have chosen to download and install on their computers the Alexa toolbar. The toolbar registers the web sites visited daily by the user and the number of pages for each site, and then sends this information to the central Alexa collection point.

These data are aggregated for each web site in order to obtain the following elements (see ‘official online guide’ to the interpretation of ranking in Alexa.com, http://www.alexa.com/site/help/traffic_learn_more):

- **Reach**: number of individual users who accessed each site, expressed as a percentage of the total number of Alexa users;
- **PageViews**: number of visited pages, in average, by each individual user.

Alexa’s list is thus based on a value, called **Traffic Rank**, which is obtained by combining **Reach** and **PageViews** as a geometric average of the daily values for the last three months (so that casual swings due to external causes, e.g. a server crash, are excluded). A certain position in the list represents both the number of visitors and the weight of their usage.

There are some distortions effects due to the way the Alexa’s list is compiled:

1) A major presence in the first 500 sites of Chinese and South-American domains, due to the fact that Alexa users are not distributed uniformly all over the world.
2) The rank obtained as a geometric average of Reach and PageView underestimate pages visited in sites (e.g. Community sites) which are mainly composed by dynamic pages (i.e. pages generated when they are visited). Alexa’s toolbar counts such visits as one, even if a user has loaded many such pages, spending a long time on the site.

Some criteria to correct or limit such distortions effects have been applied, as well as filters to eliminate sites not suitable for the present study. In particular:

- Only web sites in English have been kept, in order to be able to read the content (since the method requires that all information should be extracted from the site itself, client-side). This has implied that all local version of global sites (e.g. National versions of google) as well as sites exclusively in chinese, spanish or portuguese have been eliminated (they were also over-represented for reasons explained above).
- Web sites with pornographic or illegal content have been eliminated.
- Web sites which do not express some form of business have been eliminated.
From the original 500 sites, 163 single sites and 8 portals survived. To these, 17 sites have been added which did not occur in the top 500 (mainly for the distortion reasons mentioned above), but which are representative of important trends, especially in the western world. The portals - yahoo, msn, google, microsoft, aol, apple, indiatimes, rediff – have been subdivided in coherent sub-portals in order to carry out the analysis (since applying the schema to a whole portal would have diluted the data, adding only noise). In order to identify the sub-portals, a different site has been used, since Alexa does not provide separate rankings for subpages within a domain. Quantcast.com, the site selected, provides more detailed data, but only for the USA market. In particular, for portals it provides data about the destination to the sub-portals down to the first decimal. However, even this data does not distinguish between real sub-portals and general portal utilities, e.g. the login page. Thus, for each portal the following procedure has been established: a maximum of 3 sub-portals have been kept (with the goal of keeping only the most representative), as long as they had an access percentage above 7.5% according to Quantcast. Following this method, e.g., Apple has only 2 sub-portals (store and info), which are representative of the two main business activities, sales and assistance. For some sites, such as Indiatimes, Quantcast data has not been considered representative on its own, since it is geared towards the USA market. In that and similar cases, the subdivision proposed by Quantcast has been multiplied by the general numbers given by Alexa in order to obtain a better representation.

Table II. Sub-portals selected on the base of data on the percentage of users from the whole portal (QuantCast.com) and from the number of associated links (Alexa.com)

<table>
<thead>
<tr>
<th>Sub-portal</th>
<th>Access in % (Quantcast.com)</th>
</tr>
</thead>
<tbody>
<tr>
<td>yahoo.com</td>
<td>mail.yahoo.com 48.8</td>
</tr>
<tr>
<td></td>
<td>search.yahoo.com 40.4</td>
</tr>
<tr>
<td></td>
<td>shopping.yahoo.com 18.5</td>
</tr>
<tr>
<td>msn.com</td>
<td>msnbc.com 18.9</td>
</tr>
<tr>
<td></td>
<td>hotmail.com 17.2</td>
</tr>
<tr>
<td></td>
<td>search.msn.com 8.0</td>
</tr>
<tr>
<td>google.com</td>
<td>google.com search engine 24.9</td>
</tr>
<tr>
<td></td>
<td>maps.google.com 20.4</td>
</tr>
<tr>
<td></td>
<td>book.google.com 7.5</td>
</tr>
<tr>
<td>microsoft.com</td>
<td>update.microsoft.com 31.6</td>
</tr>
<tr>
<td></td>
<td>office.microsoft.com 15.1</td>
</tr>
<tr>
<td></td>
<td>support.microsoft.com 12.9</td>
</tr>
<tr>
<td>aol.com</td>
<td>web.mail.aol.com 24.7</td>
</tr>
<tr>
<td></td>
<td>members.aol.com 17.5</td>
</tr>
<tr>
<td></td>
<td>search.aol.com 13.0</td>
</tr>
<tr>
<td>apple.com</td>
<td>store.apple.com 21.8</td>
</tr>
<tr>
<td></td>
<td>info.apple.com 10.5</td>
</tr>
<tr>
<td>indiatimes.com</td>
<td>economictimes.indiatimes.com ***</td>
</tr>
<tr>
<td></td>
<td>cricket.indiatimes.com ***</td>
</tr>
<tr>
<td></td>
<td>photogallery.indiatimes.com ***</td>
</tr>
</tbody>
</table>

The final list of sites to analyse reached a final dimension of 200 units, with 180 from single sites and 20 from sub-portals (Table II).

4.2 Application of the schema BM*Web

Once the list of selected sites has been completed, the next step has been the observation of the framework BM*Web variables, made up by 4 variables with 19 values, for each of the 200 web sites in the sample.

The schema has been applied during November and December 2007. It is useful to remark that, as all information has to be gleaned from a site inspection, the work could take a minimum of 20 minutes per site, to an upper limit of several hours. This was repeated three times (for each analyst individually), plus all the time required to solve differences.

The variables used for the model are both qualitative and quantitative. In particular, the variables used for Needs vs. Technology, and Community (both exchange and control) are of qualitative (categorical) type.

Market and Income had been initially considered as qualitative, by assigning 1 to the prevailing mode. In fact, originally there were no quantitative variables in the schema, due to the fact that, even where they existed (e.g. for the quantification of Income), they were not available from a site inspection.

However, the use of dichotomous types (presence-absence) for the variables Market and Income turned out to be a distorting simplification. This was a direct consequence of the registration method, according to which the presence of a variables was awarded a value 1. It follows that, for sites with multiple market or income modalities, the presence, albeit marginal, of each subvariable was registered in any case as a 1. In the frequencies analysis, this fact practically causes the assigning of a very high weight to those subvariables which actually are just marginal in the web market. In order to remedy this problem, it has been decided to assign to each variable the total value of 1, which, if the are two or more values, is then divided among them. It was not possible to
assign the precise weight, due to the requirement of extracting the information from the sites only. The procedure below has thus been followed:

- for the variable Market of reference, the analysts have followed the observation of the web site and especially of the pages dedicated to the market types. So, in cases when two types of market modes have been observed, the main one has been given the value 2/3 and the secondary one the value 1/3.
- for the variable Income, since it wasn’t possible in general to establish from the site which channel was the more profitable, the analysts have followed the equality assumption. So, if e.g. a site supported 3 income channels, the value 1/3 has been assigned to each of them.

The analysis model used on the Alexa Dataset uses thus both nominal and ordinal categorical variables and quantitative ones. In details:

- The values related to the Market of reference (B2B, B2C, C2(B)2B, B2B2C, C2B, C2C, C2(B)2C), to the income type (Subscription, Intermediation, Advertising, Single Payment, Other) are of quantitative type, expressed as ratios. The total sum of the subvariables for Market and Income is always equal to 1.
- The variable Needs vs. technology is of categorical nominal type, since it allows three modalities (values): Copy of business offline, New answer to recognised needs and New answer to new needs.
- The values Employees’ control and Members’ control, related to Community, are of ordinal type, since each of them allow three modalities corresponding to the increasing level of control: minimum, light and specific.
- Finally, some values related to Community (information, commercial, complex) are of dichotomous type, since they allow only two modalities: ‘presence’ (yes) and ‘absence’ (no).

4.3 Results: A first description of the collected data

The frequencies analysis, or first level analysis, for the four variables, has been applied first to the 200 sites dataset (DBA200) and then to a sub-set of 118 sites with a community element (DBA118), in order to observe a possible community effect and other relationships among the variables of the BM*Web framework.

4.3.1 Market. In the analysis of the 7 values belonging to the Market variable, the value C2(B)2B has never been observed. This means that, in the Dataset Alexa, there are no sites operating in that particular market segment. Furthermore, the C2B value, which was observed once only in the whole DSA, does not appear at all in DBA118. For these reasons, both these values are ignored in the following analysis.

Following on, the data related to sums and averages are used, since they are not informative if used individually. This way, using pie graphs, the relative distribution of values can be shown; also, an immediate comparison can be made, both among individual values and between the whole DBA200 and the DBA118.
Observing the graph in Figure 1, covering the whole DSA, it is clear that the traditional market forms, B2C (49%) and B2B (10.3%), characterised by an upfront offer from the company, represent almost 60% of all the market forms on the web. Market forms with a direct interaction among users, such as C2C (30.8%), or with intermediation, C2(B)2C (3.6%) and B2(B)2C (6%), turn out to be still a minority of about 40%, among the most successful sites according to Alexa. However, this result is overturned considering only the sites from the DBA118, which have some community aspect (Figure 2).

While B2C goes down from 49% to 32.7% and B2B goes down from 10.3% to 6.2% (and thus their combined weight represents only 39%), C2C goes up more that 21%, moving from 30.8% to 52%, while C2(B)2C gains 2% and B2(B)2C loses 2%.

The existence of a community shows thus a significant connection with the market forms related to interaction among users. However, this should not be considered as a direct consequence of the model of analysis itself, since e.g, the presence of a forum, which qualify a site has having some form of community, does not imply at all that there should be an economic transaction, either direct or mediated, among the forum users.

Examining in details each individual value, it is possible to evince the following information:

- B2B: in the Alexa dataset (or DB200) only 35 sites (17.5%) show the presence of an offer addressed to other companies, while in the DBA118 (sites with community), this form of market is only 12%.
- B2C: this is the only market value in the whole DSA for which the median has a positive value of 0.67. This means that 50% have B2C as their main market of reference: 124 sites from DBA200 have this form of market (62%), although this value decreases for the DBA118 to 52%.
- C2(B)2C: only 11 sites on the DBA200 (5.5%) have a market form in which there is a consumer to consumer exchange, with an intermediation of the site. The absolute value remains almost the same in the DBA118, however the relative frequency goes up to 8.5%, as it is shown in the graph 2.2.
- B2B2C: this form (in which the site is acting as intermediary between another company and the consumer) is present in 19 sites from DBA200 (9.5%), while there are only 10 (8.5%) in the DBA118.
- C2B: this form has been observed only in the famous “Milliondollarhomepage” site (one of those which we added to the list as representative of some interesting trends).
- C2C: this form is the most popular in the DBA200 after B2C, as it is present in 82 sites, for a relative frequency of 41%. In the DBA118, it is almost the same number (81), while the relative frequency goes up to 68.6%. The median value for this set, 0.667, it is important as it confirms the observation that the market form C2C is the main one for half of the 118 community sites.

The Market variable is a multiple choice, so that it is possible for a site to show more than one form of market. This is useful to show a community effect on market diversification. In fact, in the DBA200 66.5% of sites has one market form only, 30.5% has two and only 3% has three. In the DBA118, on the other hand, the corresponding percentages are: 55.93% (10% less); 39% (8.5% more) and 5% (2% more). Thus, the presence of community is related to an increase of the market diversification from 35.5% to 44%.

4.3.2 Income. With respect to the Income variable, the results are presented in Figures 3 and 4. Observing Figure 3, it is obvious the main role of advertising in the DBA200 set as a form of revenues on the web, since it represents 46.6% of the whole value. Advertising in the DBA118 goes up to 51.8% (Figure 4). Such increments corresponds to a decrease in ‘single payment’, which goes down from 17.6% to 15%.2%, and ‘other’ (which contains less common forms of income), which goes down from 11.3% to 8.9%. The value for advertising is not due only to the relative facility of setting up such revenue channel and to the numerous forms that it can take. In fact, if we consider Figure 4 for the 118 community sites, the growth to 51.8% of advertising indicates that, faced with a large audience connected to the community, the company can achieve larger advertising revenues which are aimed at the community. The remaining two forms of income, subscription and intermediation, do not appear to be affected by the presence or otherwise of a community.
Analysing for each individual value of the income variable, the following appears:

- **Subscription**: it is observed for the DBA200 in 79 sites (39.5%), while for DBA118 the value is 50 (42.4%).
- **Intermediation**: this value is the least frequent in the Income variable: it is found, in DBA200, for 18 sites (9%). In DBA118, the value is 11 (9.3%). Also, intermediation is rarely the unique form of income.
- **Advertising**: this is the most common form of revenue, although it is of course difficult to ascertain which percentage of income it actually represents for the sites which use it as well as other forms. For DBA200, it is observed for 141 sites (70.5%). In DBA118, it appears for 93 sites (79%).
- **Single payment**: in DBA200, it is present in 69 sites (34.5%). In DBA118 it is present in 42 sites (35.5%). However, the effective weight goes down to 15.2%, which points out that, in community sites, single payment is mainly a secondary source of income.
- **Other**: this category contains all other minor forms of income. In DBA200, it is present in 31 sites (15.5%); in DBA118 it is present in 16 sites (13.6%).

It is also observed that the presence of community seems to induce companies to rely on multiple income channels: 51.50% of sites in DBA200 have a unique form of income, versus 45% in DBA118 (Figure 5).

4.3.3 Needs vs Technology. This variable, of nominal type, has three values, as previously described (see specification of schema parameters). The values of Needs vs. technology appear in Figure 6 for the DBA200 and in Figure 7 for the DBA118.
The comparison between these two graphs is interesting: for the DBA200, it can be seen that the distribution is essentially of bimodal type. In fact, the modality ‘copy of offline business’ appears in 72 cases over 200, while the modality ‘new answer to new need’ appears in 84 cases over 200. On the other hand, the modality ‘new answer to existing needs’ appears in only 44 cases out of 200. From this data it follows that a successful web company needs to offer either a service already existing offline or something completely new.

However, analysing Figure 7, related only to community sites, the picture changes. Here we have a unimodal distribution, since 64 sites out of 118 represents a new service for new needs. The other two modalities appear with 28 and 26 cases respectively. This data clearly shows that the presence of community not only is a result of innovation, but itself helps to generate completely new services related to the interaction needs of the users.

4.3.4 Community. As mentioned before, of the 200 DBA200 sites, 118 (59%) have a community (DBA118). 106 sites from the DBA118 have a community of type ‘information’, 7 of type ‘commercial’ and 5 of type ‘complex’ (Figure 8).

The large majority of ‘information’ communities can be explained by considering the ease of creation and management of this type of community, and its adaptability to the company’s goals and its target audience.
The type of community control is analysed for obvious reasons only for the 118 sites which do have a community (Figure 9 and Figure 10).

In both cases the legal minimum level of control is the majority. It is exercised by staff in 57.6% of cases, and by members in 66.9%. This is probably due to the fact that most communities are of type Information: while this type is easiest to set up, it is also the most difficult to monitor and control, due mainly to the volume of interactions, and the various ways in which they may be problematic (especially when picture, audio, video come into context). Light forms of control can be observed for 33.9% of staff type, and 27.1% of members type. Advanced forms of control can be observed for 8.5% of staff type and 5.9% of members type.

4.4 Results: Multivariate analysis

After the descriptive analysis, a multivariate one has been produced which put in association all the variables. Through this procedure the BM*Web framework has produced three different profiles for successful online businesses.

As the schema requires 4 variables divided into 19 values, it is necessary to use appropriate statistical techniques in order to correlate such a large number of values and to produce meaningful graphs. To analyse quantitative variables a principal components analysis (PCA) has been used. For qualitative variables, Multiple Correspondences Analysis (MCA) has been used. In both cases, variables are represented as vectors and the angle between the vectors determines the connection relationships (Greenacre 1993).

Market and Income. Principal results of the multivariate analysis applied to Market and Income are given in Figure 11 and Figure 12. In order to better understand the interactions between Market and Income, firstly the variable Needs vs. technology, and then the Community ones, will be projected onto the graph. They will be added as supplementary explicative variables, that is, they will not modify the existing projection. On a technical point, it should be noted that in a principal component analysis with quantitative and qualitative variables, the latter do not take part in the construction of the axes. However, as an explicative element, it can be added in the interpretation.

Starting from the first quadrant of graph in Figure 11, it can be observed that the value B2C is not positively associated with any form of income, while it is negatively associated with other market forms, as it would be expected. It is particularly indicative the strong negative value, -0.647, for the relation with C2C. This data show how companies which operate mainly in a B2C market have very different features from those in a C2C market, since in the former case they are the main actor and unique referent for all the users, while in the latter case the company’s goal is to facilitate the exchange between the users.

In the second quadrant it can be seen how the C2C value is positively associated with the advertising form of income: advertisers are thus attracted by a large user base, with some common social features, towards which they can address a more targeted campaign. On the other hand, C2C is negatively associated with the income
form Intermediation. This is the reason for the difference, clearly visible in the graph, with respect to the market modalities B2(B)2C and C2(B)2C.

In the third quadrant it is evident the relationship between the market forms B2(B)2C and C2(B)2C, and the revenue form intermediation. Revenues for such sites derive from the amount of each individual transaction, on which a percentage is levied for the mediation and security services, and from the total number of transactions. Revenues for C2C sites, on the other hand, depend on the dimensions and classificability of the total audience, as that is what makes it a suitable advertising target. The actual C2C exchanges are thus usually free in this context.

In the fourth quadrant it can be seen that the B2B market form is positively associated with the single payment revenue form, and with subscription. From the figure 11 it also emerges a negative relation between B2B and advertising, which is as expected since B2B sites typically have a smaller, more functional audience.

It is possible to interpret the graph along the x axis according to the origin of revenues: on the extreme left there are the sites which get their revenues only from their users, on the right the sites which get their revenues from outside entities, mainly advertising. On the y axis, from the bottom up, there are sites which rely on intermediation (mediated or direct), while on the top there are those which have a unidirectional offer to their users, be them companies or individual customers.

The multivariated analysis applied to Market and Income also shows how the business models characterised by a unidirectional, direct offer are those that most resemble offline businesses. Conversely, business models characterised by direct interaction between users are related to new answers for new needs.

Finally, the whole analysis has been repeated for the DBA118, containing only sites with community. The only meaningful change is an even stronger relation between C2C market mode and advertising, reinforcing the view that a community is a crucial element in attracting advertising.

The comparison between the graphs in Figure 11 and 12 also shows a relation between absence of community and copy of offline business model.

The multivariate analysis clearly shows the existence of three separate profiles for successful online businesses:

- **Profile 1**: direct offer (B2C and B2B), Single payment and subscription, copy of offline business, no web community;
- **Profile 2**: intermediation site (C2(B)2C and B2(B)2C), intermediation revenue model, commercial or complex communities with specific controls;
- **Profile 3**: direct interaction among users (C2C), advertising, information community, light controls.

![Fig. 11. Market and Income Biplot (Need vs Technology as illustrative variable, DBA200)]
The existence of three separate profiles is an interesting result, as often there is a tendency to define a single recipe for online success. A more detailed discussion of the implications of this finding can be found in the conclusions section.

**Market and Needs vs technology.** The analysis shows a meaningful relationship between Market and Needs vs. technology. In Figure 13 the x axis can be interpreted as representing the development of content: from an active to a passive role for the user; the y axis can be interpreted as representing the level of mediation and control over interaction. The numbers next to the website names are associated to the relative position of the site in some homogenous clusters introduced to facilitate the interpretation of the display (see also Figure 14 and Figure 17).

**Market and Community.** The MCA graph for Market and Income (Figure 14) suggests an interpretation of the x axis as direct intermediation and simple community versus frontal offer and lack of community; while y axis can be interpreted in terms of intermediation and complex community.
In the DBA200 three main clusters appear:

1) Sites characterised by frontal offer (B2C and B2B) are generally devoid of community.
2) Sites with direct interaction C2C always have community (except one case), usually of information type, with minimum or light control.
3) Sites with intermediated market have a positive relation with commercial and complex communities and high level of control.

**Income and Needs vs technology.** From DBA200:

1) Single payment appears associated to copy of offline business
2) Advertising is associated to new answer for new needs.

From DBA118 there is a relation between copy of offline business and interemediation.

**Income and Community.** Income and Community does not seem to have many relations, but it can be observed an association between advertising and information community with light control on one side, and intermediation with commercial and complex community with high control on the other.

### 4.5 Results: external data

In this part, we use external data about the relative success of the sites, and superimpose it on the data of the BM*Web framework, looking for statistically significant relationships.

The external data we use are the Rank, Page View and Reach from Alexa (explained in section 4.1). These data relate to 180 (subportals are excluded), and are based on 3-months average. They have been normalised in 3 intervals as follows (quartiles):

- 25% low level
- 50% middle level
- 25% high level.

Associations have then been investigated for all variables. Only those which have proved significant are detailed here.

**Market and illustrative external data.** The relationship between number of market forms and illustrative data is given in Figure 15.
Here there is an association only with the sites which have 3 market modes: i.e., those that do are most successful in the Alexa ranking.

**Income and external data.** The same situation arises for Income: the only meaningful relationship is with the number of income channels (Figure 16).

Again, the most successful sites have multiple income channels, although those with only two (usually a main one and a secondary one) do well too, as they are placed in the middle section.

**Needs vs technology and external data.** Here we find a clear relationship: the businesses which are copy of offline activities have the lowest external values; those which offer new answers to existing needs are in the middle, and those which provide new answers to new needs are at the top (Figure 17).
Figure 18 gives the same map based only on the 115 sites with community. Here even the sites which are copies of offline businesses reach a medium external value, which shows that the community by itself has a positive relationship with those external values.

Community and external values. Figure 19 gives a map of the Community values with the external data superimposed and it shows a positive association between them, and also, a relation between information modality and high values (although this might be due to the problems in Alexa of recording correctly the interaction in connected mode, see above).
The study has produced the following general conclusions:

- The framework BM*Web is shown to be an effective tool to analyse online businesses, as it is able to identify statistically meaningful features, which in turn lend themselves to interesting interpretations.
- The 'community effect' is confirmed, as it is shown that a substantial group of successful online businesses is based around a community or is proceeding to integrate one in their model. However, the study has also shown that other types of online businesses can be just as successful, and that the key feature is the organic relationship between the various parameters in the online business structure.

5. Conclusions and future work

The study has produced the following general conclusions:
As a result, three different profiles have been determined for successful online businesses:

**Profile 1:** direct offer of goods or services; B2C or B2B, Single Payment or Subscription, Copy of Offline Business and lack of community;

**Profile 2:** goods or services offered through intermediation; C2(B)2C or B2(B)2C, Intermediation as source of revenues, commercial or complex communities with specific controls.

**Profile 3:** direct interaction between users with a marginal role for the site; C2C, Advertising, Information Community with minimal or light controls.

It thus emerges that sites of profile 1 can still be very successful, although they may appear ‘old’: in fact, they exploit very well their natural niche, and use the Internet only for the internal technological features (access from home 24/7, low fix costs, low inventory, very large virtual stock etc.).

Sites of profile 2 and 3, on the other hand, exploit the web technology in order to create new services (especially profile 3), or to expand existing ones well beyond their off-line counterparts (especially profile 2). For both profile 2 and 3, the most important feature is the ability to connect users among themselves. The main difference is that in profile 2 the site offers services related to legal, financial, privacy guarantees, allowing a user-to-user transaction which is more complex and of higher financial value, but also more controlled and fixed. Sites of profile 3 allow much more creativity and direct user-generated content (web 2.0), but usually the increased freedom is balanced by shallower, less secure transactions. This also causes a shift of main revenue form from Intermediation for profile 2 to Advertising for profile 3.

More detailed conclusions emerge from the frequency analysis (first level analysis) and the multivariate analysis with illustrative data (second level analysis).

- Community presence significantly influences the three variables Market, Income and Needs vs. technology. In fact, in the DBA200 the most common market forms are the direct ones (i.e. B2B and B2C, about 59%), while the interaction between users, direct or intermediated, represents about 40% of the total. For DBA118 (sites with community), the result is inverted: interaction between consumers grow to 52%, and direct offer sites diminish to 39%.
- The same pattern emerges for Income: on DBA118 advertising grows while single payment decreases.
- Community presence has a very large influence on the Needs vs. technology variable. In the DBA200 there is a bimodal distribution (mainly copy of offline model and new answer to new needs), while in DBA118 a unimodal distribution emerges, dominated by the latter modality. This shows a strong relationship between community and innovation on the web.

These results confirm the community effect and justify its introduction as a variable in the BM-Web framework. Unlike in Timmers’ classification scheme (Timmers 1998), the Community variable is not considered a separate business model, but an element which interacts with the other: it is felt that the results support this choice.

Analysis against ranking levels has shown the following:

- Profile 1 sites show in general a medium-low level of reach and page view.
- Profile 2 sites show a high level of reach and page view.
- Profile 3 sites show a variation (probably due to the peculiarity of Alexa counting with respect to connected protocols): commercial community sites tend to have high reach and page view, while those with complex community (e.g. virtual reality or multi-players role games) tend to have low values.

Finally, the most successful sites in terms of reach and page view tend to have three market channels and three or four income streams. This seems to show that the ability to articulate the offer to different subsections with different modalities increase the chance of success. As an example of the way in which these different profiles can interact and compete, it is instructive to look at two recent cases:

- Lack of resilience of pure-play online retailers in absence of community barriers (off-line retailers with online presence taking more than 50% of online retail market in UK from Oct. 2007): Hitwise data for the online UK market show that pure-play retailers tend to lose their first-mover advantage in absence of a community and are liable to be overtaken by brick-and-mortar competitors adding an online presence to their offline convenience (www.hitwise.co.uk/press-center/hitwiseHS2004/hotshop.php). This shows that in absence of community, offline players have an advantage in traditional markets, as predicted by our schema.
- Lack of resilience of dominant players when attacked by new community based ones (more than 50% of email in UK from Oct. 2007 originated from social networks rather than Hotmail, Yahoo, Gmail, etc. according to Hitwise): even free services (such as email) offered by dominant players have lost more than 50% of their traffic to social network, showing that most of the email use is within circles of friends which are best reproduced inside social networks (weblogs.hitwise.com/robin-goad/2007/11/social_networks_overtake_webma.html). This shows that the community factors, in innovative context, is totally dominant with respect to other elements, e.g. market dominance, financial power etc.

It should be noted that, since the data was collected, there has been a major upheaval in the online market, due to the world wide financial crisis (originated in the real-estate financial section, but quickly spread to the rest of the economy). In particular, the crisis has brought a significant contraction of the advertising expenditure, and a consolidation of the remaining part in fewer, bigger sites. It could thus be expected that some community-
based models, relying mainly on advertising, might have suffered as a result. On the other hand, other community sites, especially those related to online auctions, gambling, discounts, travel advice etc. seem to have prospered, as those activities tend to flourish offline in times of financial crisis, and the same apply in the online world. We briefly mention some possible future lines of research: a new ‘after-crisis’ snapshot with the same methodology, to check on changes and the ability of the model to identify them; more in-depth analysis of some variables, e.g. the relationship between various forms of advertising and the other variables; more area-specific sectors, e.g. the tourism area.

Acknowledgments

Roberto Garigliano’s contribution was partially supported by the University of Trento under their visiting professorships programme. The authors are very thankful to Fulvio Vanzo for his statistical data analysis; and to Federica Agostini and Stefania Marchesoni for helping to collect data with the BM*Web framework. We also thank all the eTourism research group members at Trento University that worked to the definition and validation of the different versions of BM*Web.

References


Business models for the Web: an analysis of top successful web sites

R. Garigliano, L. Mich, P.L. Novi Inverardi