



# Using Social Networks to Enhance Customer Relationship Management

Ian Basaille, Nadine Cullot, Lylia Abrouk, Eric Leclercq

► **To cite this version:**

Ian Basaille, Nadine Cullot, Lylia Abrouk, Eric Leclercq. Using Social Networks to Enhance Customer Relationship Management. 5th International Conference on Management of Emergent Digital Ecosystems, Oct 2013, Luxembourg, Luxembourg. pp.169-176. hal-00904251

**HAL Id: hal-00904251**

**<https://hal-univ-bourgogne.archives-ouvertes.fr/hal-00904251>**

Submitted on 14 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Using Social Networks to Enhance Customer Relationship Management

Ian Basaille-Gahitte  
Laboratoire Électronique,  
Informatique et Image  
UMR CNRS 6306 - Université  
de Bourgogne  
9 Avenue Alain Savary, 21078  
Dijon CEDEX  
ian.basaille@u-  
bourgogne.fr

Lylia Abrouk  
Laboratoire Électronique,  
Informatique et Image  
UMR CNRS 6306 - Université  
de Bourgogne  
9 Avenue Alain Savary, 21078  
Dijon CEDEX  
lylia.abrouk@u-  
bourgogne.fr

Nadine Cullot  
Laboratoire Électronique,  
Informatique et Image  
UMR CNRS 6306 - Université  
de Bourgogne  
9 Avenue Alain Savary, 21078  
Dijon CEDEX  
nadine.cullot@u-  
bourgogne.fr

Eric Leclercq  
Laboratoire Électronique,  
Informatique et Image  
UMR CNRS 6306 - Université  
de Bourgogne  
9 Avenue Alain Savary, 21078  
Dijon CEDEX  
eric.leclercq@u-  
bourgogne.fr

## ABSTRACT

In recent years, the Web has evolved into an exchange platform. Customer Relationship Management (CRM) must follow this evolution and connect CRM tools to social networks in order to place companies in the center of all the exchanges. We propose, in this article, a community detection approach that identifies clusters of customers of a company using their explicit and implicit behaviour. Our contribution is the definition of a composite profile that integrates various informations gathered from different applications, such as the information system of the company, the existing CRM, or Twitter. We define a similarity measure, between a user and a tag, that takes into account the rating and consultation of resources, as well as actions on social networks and user contacts. We validate this approach against a test database and we discuss results and future works.

## Categories and Subject Descriptors

[Web]: Semantic Web; Web 2, Web 3; Social Networks and Communities; [Data & Knowledge Management Systems]: Knowledge Management, Sharing and Discovery

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MEDES'13 October 29-31, 2013, Neumünster Abbey, Luxembourg  
Copyright 2013 ACM 978-1-4503-2004-7 ...\$10.00.

## General Terms

Social networks, community detection, CRM, Social CRM

## Keywords

Social networks, CRM, profile, tag, Web 2

## 1. INTRODUCTION

Companies develop Web applications as a marketing tool, by creating dedicated websites, or by developing on-line advertisement tools dynamically included in other websites. Moreover, companies maintain complex applications for managing interactions with customers through retail websites which are used not only by customers to buy products, but also to gather information about products and to contact on-line support services. Avatars and customer advisers (using phone calls) are used to answer customers questions and expectations. Customers also become actors on behalf of companies or of their brands, by giving grades, opinions or comments on informations available on dedicated websites of companies, customers forums, or on social networks.

The study of customer, client or prospect behaviour is an essential task for the competitiveness of a company. A lot of CRM (Customer Relationship Management) tools exist. They provide users with rich environments dedicated to improve the management of commercial, marketing and after-sales services, and to offer analysis tools. Three main categories of CRM tools can be identified:

1. *Universal, or non specialized tools*, like *SugarCRM*<sup>1</sup> or *SalesForce*<sup>2</sup>, that offers general CRM features such

<sup>1</sup><http://www.sugarcrm.com/>

<sup>2</sup><https://www.salesforce.com/>

as commercial relationship management, marketing, analysis tools, but also some collaborative capabilities;

2. *Integrable tools*, that offer software modules that can be linked with other applications of the company's information system. They usually have more specialized features like data analysis or data mining algorithms. *Smarter Analytics from IBM*<sup>3</sup> is one of these tools, that can analyse data for business intelligence purposes;
3. *Generic tools or frameworks*, that are developed to be customized and instantiated in order to answer the specific needs of a company. These types of software are modular and can cover every aspect of the CRM. They can be easily extended to evolve towards social networks applications.

The CRM is evolving, taking into account the emergence of Web 2.0, social networks and their growing popularity. Indeed, the Web has evolved into a generic exchange platform, where every user becomes a content provider by means of tools like blogs with comments, wikis with their collaboration and contribution features, or social networks with their annotation and resource sharing features. Therefore, CRM must take into account a better integration of the *social* dimension of exchanges between a company and its clients, or among the customers writing about the company. The term *Social CRM* is associated with the use of social medias in the customer relationship context. Furthermore, *customers on the Internet* are a part of the *e-reputation* of a company, acting directly on its growth and competitiveness.

*Social CRM* applications will take advantage of the potential of social networks in terms of knowledge about the clients and customers. Platforms that connect CRM tools and social networks are emerging needs. These platforms will be able to project companies inside the conversations that are already taking place on the Web and in which they don't take any part of yet. Nevertheless, the increasing number of communication channels and the scattering of information and conversations among users on the Web remain challenging issues.

Community detection tools are essential in order to meet the needs of *Social CRM*. Users communities exist in an implicit way and can be found in general use social networks or inside the professional social network of a company. Their discovery benefits the company as well as its clients. In this context, the detection of communities of users can be based on their interests or behaviour. Other issues also concern computation of user-generated data in order to analyse the logs of the interactions of the users on the Web. These interactions can either be with a dedicated website, forums, annotations or comments. It also includes the development of methods or algorithms able to detect the users on the Web and group them into communities and to interact in a more targeted way with these communities.

Our approach, named DisCoCRM (*Discovering Communities for CRM*), takes user-generated data in order to compute user profiles. These profiles are then used to detect communities of users that will be used, for example, by a recommender system to recommend items to the users. Our approach will be integrated in a Social CRM platform.

<sup>3</sup><http://www.ibm.com/analytics/us/en/>

We propose in this paper an approach to integrate different sources of information for detecting communities of customers based on their behaviour.

This paper is organized as follows: Section 2 relates some existing works, and Section 3 gives the definition of users profile and community building process. Before concluding this paper in Section 5, we describe our experiments and discuss the results in Section 4.

## 2. RELATED WORK

A social network can be modelled as a graph. Nodes are the elements of the network, such as users or resources; and edges, or arcs if the graph is directed, are the relations between these elements. The relations can describe affinity links between users, thematic similarity between resources, etc. The term *social network* is usually associated to Web 2.0, where the user became an essential actor: he can add, annotate, interact and broadcast content. Hence, this "collaborative" Web promotes on-line social interactions.

General purpose social networks, such as **Facebook**<sup>4</sup> and **Twitter**<sup>5</sup>, are now used daily by a lot of persons; and companies develop more and more their own social network in order to share and elaborate, in-house, content; or in order to create links between employees. Several specific tools, labelled as enterprise social networks, exist, like for instance **Yammer**<sup>6</sup> that allows users to work inside a network of their colleagues, or **Bluekiwi**<sup>7</sup>, a collaboration and dialogue platform for in-house and outside of the company exchanges.

Several works are devoted to Social CRM. Ajmera et al. [4] analyze posts and messages in social platforms and identify posts relevant to enterprise. The system takes into account different parameters like the intent of the post and the nature of users behind the post. Wu and al. [19] propose a novel telecom CRM framework: Group CRM (GCRM) to transform the traditional CRM methods from individuals to social groups based on emerged methods of social networks analysis.

Since the inception of the Web, implicit communities research has greatly evolved. At first, it focused on the links between documents, and went on to study the links between users by analysing their interactions, via Web applications. Quan [15] presents a technical state of the art of social networks, their features, the platforms and the new research problematics such as distributed structure of information, interoperability of platforms and information systems, identity research, data ownership and security.

Many specific works focus on the notion of community (Kumar et al. [12]). Cohen et al. [5] present a state of the art of the measures of proximity in social networks, allowing to quantify the similarity degree between two users.

In order to regroup users into communities, defining a profile for each user is mandatory. This profile consists of a set of data about a specific user, such as his name, age, town. It also usually consists of information about his interests and the grades he gives to resources [8]. These interests can be explicitly given by the user, or implicitly by analysing his behaviour. The profile can be represented by a set of key words (tags).

<sup>4</sup><https://www.facebook.com/>

<sup>5</sup><https://twitter.com/>

<sup>6</sup>[www.yammer.com](http://www.yammer.com)

<sup>7</sup><http://www.bluekiwi-software.com>

Using data collected from social networks can improve the content of a user profile. Abel et al. [2] propose to use tweets in order to model the interests of a user. They use tweets and hashtags<sup>8</sup> from the user and the links included in these tweets in order to give context to Twitter activities. Abel et al. [1] propose a user modelling framework based on the activities of a user on several social medias such as **Flickr**<sup>9</sup>, **Twitter** and **Delicious**<sup>10</sup>, using tag-based profiles. They noticed that this method can improve the quality of recommendations in systems that have few informations about their users.

**Activity Streams**<sup>11</sup> is a standardised format allowing to associate meta-data to the actions performed by a user on the different social networks he has an account on, in order to distinguish them and add more sense to them. It consists of an actor, a verb, an object and a target, for instance *a user associates a tag to a resource, or a user is in contact with another user*.

Hung et al. in [9] define a profile as a set of tags with weights. They compute a tag-to-tag matrix to capture public preferences, *i.e.* how much percentage of people who like topic A will also like topic B. Firan et al. in [6] use Last.fm<sup>12</sup> data to compare different algorithms, based on the type of profile and technique used to compute recommendations, and divide them into three categories : collaborative filtering based on songs, based on tags, and research based on tags. They concluded that research based on tags is the most efficient one and does not suffer from the cold start problem. However, they pointed out the need for disambiguating tags and reducing the weight of non important tags.

Adomavicius and Thuzhilin in [3] describe the limitations of recommender systems. These systems need to have a better understanding of users and items and to incorporate contextual information into the recommendation process. The understanding can be done by using the user's transactional histories and with better profiling that can describe the user's interests. The contextual information can be used to compute multi-dimension recommendations, by adding time (of the year, of the week, etc.), the person with whom the item will be consumed, and knowledge about the user's task. Another limitation is the intrusiveness of the recommender systems as they usually require explicit feedback from the users.

The study of related works shows that the use of tags is a good solution. The complexity comes from the choice of tags and their weight. These two parameters have a great influence on the final result. We believe that social information can be used to have a better understanding of customers and their interests. The contextual information needed to improve the recommendation process can also be extracted from the customer's social network activities and relations.

### 3. PROFILE MODELLING AND COMMUNITY BUILDING

Enterprises have to elaborate knowledge on their customers

<sup>8</sup>Key word, meta-data tag, symbolised by a # on Twitter and Facebook, tool that helps categorising a message

<sup>9</sup><https://secure.flickr.com/>

<sup>10</sup><https://delicious.com/>

<sup>11</sup><http://activitystrea.ms/>

<sup>12</sup><http://www.last.fm/>

to improve their competitiveness, to be able to better fulfill customers needs and also to offer new services or products. In this section, we propose a solution to improve customer relationship management by extending the enterprise social network, which usually only focuses on the members of the company, to the customers. This extension is performed by taking into account their interests and behaviour as well as their own contacts in social networks such as Twitter. Our approach consists in three parts: 1) the specification of DisCoCRM; 2) the modelling of the user profiles; and 3) the detection of the communities of users by applying classification algorithms.

#### 3.1 DisCoCRM

Users of DisCoCRM, who are customers of a company, interact with the information system of the company via a dedicated Web site and also have some other exchanges on public social networks. DisCoCRM takes into account the relationships between the company, its resources and its users. Hence, it is mandatory to model resources and users interactions, as well as interactions between users themselves, specifying their particular profiles. These user profiles will then be used to build communities of users with similar needs or interests. Figure 1 depicts DisCoCRM which consists of two major parts:

1. *The information system under the Web site* dedicated to the company (right side of figure 1) contains a set of resources which are categorized using a tag base. We assume that each resource is tagged only with one category. The user behaviour is monitored via the historic of the actions he has performed on the Web site. He can consult, share and comment on existing resources and he also can post new ones. The user actions on the Web site are used to build the user profile on the Web site.
2. *The public social networks*, such as Facebook or Twitter (left side of figure 1), can help refining the user profile with informations that are not available in the information system. A user can post, share or bookmark resources. He can create tagged resources (with tools like hashtags) or post links to other Web sites. He can also have a list of contacts inside the social network. The user actions that are done on resources tagged with categories of the tag base of the information system of the company are managed and used to build the user profile on the social network.

*The CRM of the company* which contains a knowledge base of the customers or prospects of the company is also used to complete the profiles of the users who are known and identified in the information system. We assume that the knowledge base contains general informations such as the user address, his phone number, the list of products he has bought and possibly the list of problems he has encountered, and so on.

#### 3.2 Modelling of the user profile

The user profile is based on the user's interest in the resources of the system. This interest can explicitly be expressed by the user when, for example, he gives a grade to a resource or implicitly when he performs some other actions on a resource. The user profile is modelled by taking into

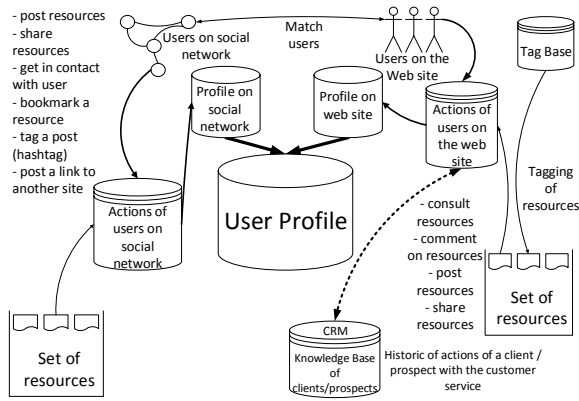


Figure 1: DisCoCRM and User Profile

account these two kinds of interests. We retain three types of information: i) the evaluation of the resources by the user, in form of grades; ii) the user's interest in a resource by the fact that the user posts or consults this resource; and iii) the network of contacts of the user.

### 3.2.1 Definitions

We consider a set of users  $U = \{u_1, \dots, u_n\}$  and a set of resources  $R$  that can be either inside the information system of the company  $R_{ins}$  or outside the system  $R_{out}$ , on social networks for instance.

We suppose that users can comment on a resource of the system  $R_{ins} \subseteq R$ . We don't qualify these comments, but we suppose that if a user comments on a resource, it means that he is interested in this resource. The user can also share a resource of  $R_{ins}$  on social networks such as Twitter or Facebook. If a user shares a resource, we consider that he is interested in this resource. The combination of comments and shares can serve as a grade from the user to a resource. That is, if the user has commented on a resource, the grade is 1. If he comments on or shares this resource on a social network, the grade is 2, and if he has commented on and shared it on two social networks the grade is 3, and so on. The grades are stored in a matrix  $M : U \times R_{ins}$  for a user  $u_i \in U$  and a resource  $r_j \in R_{ins}$ ,  $M(u_i, r_j) = n_{ij}$ . This matrix is dynamically updated when new users join the system, when new resources are added into the system, and when new actions from users modify any grade.

Resources in  $R_{ins}$  are annotated with tags from the tag base stored inside the system. We define  $T = \{t_1, \dots, t_m\}$  as the set of tags. Each resource is annotated with a subset of  $T$ . Resources in  $R_{out}$  are already tagged when they are collected, and we only consider the ones that are tagged with tags included in  $T$ . Social networks can be used to collect resources from outside the information system. For instance, users on Twitter can post tweets with links and hashtags that can be used to tag the tweet.

Tags that are associated to the resources are stored in a matrix  $MT : R \times T$ . The definition, for a resource  $r_j \in R$  and a tag  $t_k \in T$ , is:

$$MT(r_j, t_k) = \begin{cases} 1 & \text{if } t_k \text{ is associated to } r_j, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Social networks can also be used to get access to the social network of the user. On Facebook, one can use the friends list to visualize the social network of a user. Twitter offers two types of links between users:

- the *followers*, who are the people who explicitly stated that they want to be in contact with the user, without any approval from the user;
- the *following*, who are to people the user explicitly stated he wants to be in contact with.

We only use the second type, because the first one is independent from the user actions, and the second one represents the people the user is interested in.

As users of social networks are coming from Twitter and/or Facebook and/or any other general use social network, there can be users that are not inside the information system. We only consider users that are on social networks and are also part of the information system to be able to identify them on both platforms. The social network is stored in a symmetric matrix  $A : U \times U$ , defined as follows, for two users  $u_i, u_j \in U$ :

$$A(u_i, u_j) = \begin{cases} 1 & \text{if } u_i \text{ is in contact with } u_j, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

### 3.2.2 Aggregating information in the profile

We have identified five important steps to aggregate information into a profile:

1. Collecting data from the grades given by the user;
2. Collecting data from the activities performed by the user (consultations, Twitter, Facebook, etc.);
3. Collecting data from the contacts of the user;
4. Observing the activities of the users to define the weighting coefficients and the window of time;
5. Computing profiles, with different parameters.

**Step 1.** The goal is to group users in thematic communities, based on the resources they are interested in. We compute the degree of membership  $da_{ij}$  of a user  $u_i$  to a tag  $t_j$ :

$$da_{ij} = \frac{|R(u_i, t_j)|}{|R|} \times m_j \quad (3)$$

with

$$m_j = \frac{\sum n_{ij}}{n_{max} \times |n_{ij}|} \quad (4)$$

- $R(u_i, t_j)$  is the set of resources graded by user  $u_i$  where the tag  $t_j$  appears and  $|R(u_i, t_j)|$  the cardinality of  $R(u_i, t_j)$
- $m_j$  the average grade given by the user  $u_i$  to resources  $R(u_i, t_j)$  divided by  $n_{max}$  to have a value between 0 and 1
- $n_{max}$  is the maximal grade given to resources by the users

**Step 2.** As we would like to take into account the behaviour of users inside and outside the system, we refine the expression of  $da_{ij}$  with the behaviour of  $u_i$ , either with:

- his interest in a resource with either:
  - the post or the consultation of the resource
  - the tweeting, re-tweeting or bookmarking on Twitter or the post or sharing of resources on Facebook
- his social network: by taking into account the grades given by his contacts

The membership degree of  $u_i$  to a tag  $t_j$  is modified if this tag is associated to a consulted resource inside the set of resources  $R$  by  $u_i$ , and also by the membership degree of the contacts of  $u_i$ .

In order to integrate these elements, we define a membership degree  $d'_{ij}$  that is computed via the consulted resources  $R_{consult}$  annotated with the tag  $t_j$ . We use the logs of consultation of  $u_i$  on the website, and the tweets, re-tweets and bookmarks of  $u_i$  on Twitter and the posted and shared resources on Facebook. The equation is defined as follows, with the example of only Twitter:

$$d'_{ij} = a \times \frac{|R_{consult}(u_i, t_j)|}{|R_{consult}|} + b \times \frac{|R_{tweet}(u_i, t_j)|}{|R_{tweet}|} + c \times \frac{|R_{re-tweet}(u_i, t_j)|}{|R_{re-tweet}|} + d \times \frac{|R_{bookmark}(u_i, t_j)|}{|R_{bookmark}|}$$

with

- $R_{consult}$  the set of resources  $R_{ins}$  consulted by  $u_i$
- $R_{tweet}$  the set of tweets of  $u_i$
- $R_{re-tweet}$  the set of re-tweets of  $u_i$
- $R_{bookmark}$  the set of bookmarks of tweets of  $u_i$
- $a, b, c$  and  $d$  weighting coefficients with  $a+b+c+d = 1$

Depending on how the users of the system interact with the website and with social networks, one can tweak the equation into giving more weight to one or more specific parts, for instance tweets and consulted resources if users don't re-tweet or bookmark often on Twitter.

**Step 3.** In order to use the information given by the users that are in contact with  $u_i$ , we use the degree of membership from *Step 1* for every user  $u_k \in A(u_i)$ . Therefore, for a user  $u_i$  and a tag  $t_j$ , we define  $dc_{ij}$  as :

$$dc_{ij} = \frac{\sum_{k=1}^m da_{kj}}{m} \quad (5)$$

- $A(u_i) \subseteq U$  is the set of contacts of the user  $u_i$  and  $m = |A(u_i)|$
- $da_{kj}$  is the degree of membership of the user  $u_k$  to the tag  $t_j$ , with  $u_k \in A(u_i)$

**Step 4.** We define a window of time, and we will only consider the actions that have occurred during this window. We still get every action performed by the user, to save data for visualising the evolution of the user interest through

time, but the user profile at a given time only takes into account actions that have occurred during the window of time. This window can be defined by studying the activities of the users. For example, if they don't participate a lot, the window has to be wider to incorporate the same number of activities than in a website where users interact a lot. One can imagine setting the time window automatically by taking into account a minimum number of activities per user.

With the three parameters taken into account in our approach (grades, consultations/actions on social networks and contacts), the degree of membership  $d_{ij}$  of a user  $u_i$  to a tag  $t_j$  is defined as follows:

$$d_{ij} = \alpha \times da_{ij} + \beta \times dc_{ij} + \gamma \times d'_{ij} \quad (6)$$

- $\alpha$  and  $\beta$  and  $\gamma$  are weighting coefficients with  $\alpha + \beta + \gamma = 1$

**Step 5.** The weighting coefficients enable us to take into account the different business-specific aspects and to give more or less importance to some of interactions, depending on the context of the application. This also helps us to define a generic tool that can be used in various business environments. The coefficients can be adjusted according to the activities of the users, depending on how they interact between each other.

As our tool is designed to be used by Community Managers to help them in their work, their feedback will be useful to set the weighting coefficients. One can imagine to define a few templates of weighting coefficients, each designed for a specific use, for instance emphasizing an aspect of the user profile, or taking into account the different parameters the same way. These templates can be defined by the feedback of the Community Managers and they will choose which template to use according to the context instead of setting up manually the parameters.

The profile of the user  $u_i$ , noted  $X_i$ , is a vector of his degrees of membership to each tag:  $X_i = (d_{i1}, d_{i2}, \dots, d_{ij})$ .

### 3.3 Community detection algorithms

The early works on community construction use hypertext links as a computation base ([11] and [10]) by structuring usually thematic communities of Web pages. These pages were created by persons or organisations, that have a shared interest on a specified subject.

Nowadays, the concept of community primarily involves users with multiple roles such as customer, resource producer, message transmitter, or grader of products. Hence, links are replaced by the notion of user profiles. In the social Web, communities are usually a set of resources, users or tags ([14]).

Smyth and White [17] propose two spectral clustering algorithms that seek to maximise  $Q$ , the modularity function proposed by Newman and Girvan [13]. They are efficient and effective at finding good clusterings and the appropriate number of clusters. Radicchi et al. [16] propose a way to implement a quantitative definition of community in a generic divisive algorithm, based on the edge betweenness measure defined by Girvan and Newman in [7], that as proven to be faster than the one proposed by Girvan and Newman. Other clustering algorithms can be used, such as k-means for instance, or hierarchical clustering.

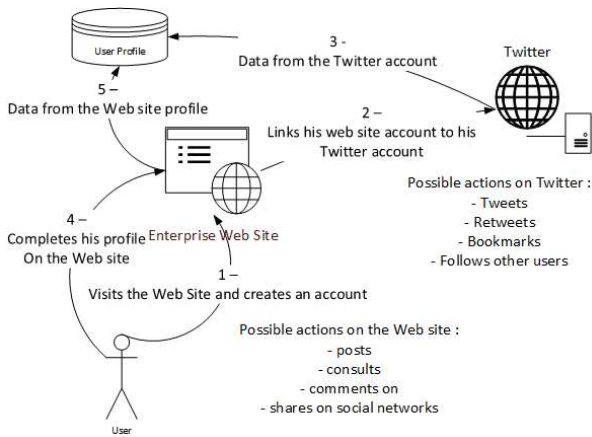


Figure 2: User actions and profile building in our experimental system

## 4. EXPERIMENTS AND RESULTS

We have tested our approach with a test sample. This sample consists of a website, used by one of our partners, that hosts a number of resources on a specific thematic, in our case *client relationship management*. These resources are divided into categories which serve as tags. Resources can be, either written papers or text interviews, or videos of interviews. Users of this website are marketing directors, specialists in client relationship management, bloggers, etc. The website is also a collaborative platform, where users can search, post, comment on, and share resources. Some users of this platform have a Twitter profile and have explicitly declared their Twitter account. Linking their website profile and their Twitter profile allows us to use both profiles in order to enhance their user profile. Moreover, users on Twitter can tweet, re-tweet, bookmark tweets and follow other people. We only consider users on Twitter that can be linked to users on the website. Regarding our time window, we only take into account the actions of users that have occurred during the past year. Figure 2 shows a diagram of our experimental setup and the various actions that can be performed by the users.

A typical use case of the platform would be:

1. A user visits the web site. He creates an account on it;
2. The user then links his Twitter account to his account on the web site;
3. Data from his Twitter account is gathered into his user profile (tweets, re-tweets, bookmarks, contacts);
4. He completes his profile on the web site if necessary;
5. Data from his web site profile his gathered into his user profile (grades, consultations).

This experiment is divided in two steps: 1) building the user profiles, by computing the membership degrees; and 2) building the communities of users.

### 4.1 Computing user profiles

We start by computing the matrix  $M$  with the set of users  $U$  and resources  $R$  with the grades of the users. We also

Table 1: Example of tagged resources

Resources	Tags
$R_1$	Cars
$R_3$	Banks and Insurances
$R_8$	Blogging
$R_{11}$	Services
$R_{16}$	Consulting
$R_{30}$	Energy

Table 2: Example of grades

	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$
$u_1$	0	1	0	0	0
$u_3$	3	3	0	0	0
$u_5$	3	3	0	0	0
$u_{19}$	0	3	0	2	0

have the list of resources with their associated tag, the table 1 shows a few examples of tags associated to resources and the table 2 of grades given by users.

Then, we compute the matrix  $M_d$  by evaluating the degrees of membership of the users to the various tags. The table 3 presents some of these degrees.

Then, we refine our profile by using the implicit behaviour of the users: we take into account the consultation of resources. For the users who have a Twitter account, we also take into account the tweets, re-tweets and bookmarks, and other users they follow on Twitter. The table 4 shows the degrees of membership that will be used as a user profile and that takes into account these new elements. We experiment with different weighting coefficients, both for the consultations, tweets, re-tweets and bookmarks and for the final formula of the degree. We also plan to pursue the experiments in order to refine the weighting coefficients values, according to the different types of communities one might want. In order to do so, we plan on defining a few templates of weighting coefficients designed, for instance, to emphasize a specific aspect of a profile, or to take into account the different parameters the same way. The Community Managers, who will be the end users of our system, will give us feedback in order to refine these templates.

Table 4 presents an example of the weighting coefficients as follows:  $a = 0.5$ ,  $b = 0.3$ ,  $c = 0.2$ ,  $d = 0.1$  and  $\alpha = 0.5$ ,  $\beta = 0.2$ ,  $\gamma = 0.3$ .

We can already notice, for instance, that the  $d_{ij}$  for user  $u_1$  and the tag *Cars* has gone up. This can be explained by the fact that  $u_1$  and  $u_3$  are in contact on Twitter and by the fact that  $u_3$  is interested in the tag *Cars*. We can also notice that  $u_1$  consults, tweets, re-tweets and bookmarks tweets that are tagged with *Cars*. Also, the  $d_{ij}$  for user  $u_3$  and the tag *Banks and insurances* has gone up. He hasn't given any grade to any resource tagged with this tag, but he consults resources tagged with *Banks and insurances* and is in contact with  $u_1$ , who is interested in this tag.

Table 3: Explicit user profiles

	Cars	Banks and insurances	Blogging	Energy
$u_1$	0.0002	0	0.0206	0.0002
$u_3$	0.0035	0	0.0003	0.0029
$u_5$	0.0059	0	0	0
$u_{19}$	0.0005	0.0009	0.0033	0.0002

**Table 4: Explicit and implicit user profile**

	Cars	Banks and insurances	Blogging	Energy
$u_1$	0.0339	0.1018	0.1271	0.0122
$u_3$	0.0581	0.1084	0.0849	0.0158
$u_5$	0.0784	0.0088	0.1806	0
$u_{19}$	0.0145	0.0148	0.0305	0.0144

## 4.2 Community detection

For the community detection phase, we used the k-means algorithm, and the open source software WEKA [18].

After studying the behaviour of the users, we choose to set the community number to 4. The *Services* category is the one with the most resources in it (19 resources), and some only have one or two resources.

Non refined communities (built using only the explicit profile) are illustrated in table 5. Refined communities (built using both explicit and implicit profiles) are illustrated in tables 6 and 7. The difference between both tables is the different weighting coefficients used for computing the membership degrees. The communities in table 7 have an highlight on consultations and actions on Twitter ( $\gamma = 0.3$ ), whereas the communities in table 6 have an highlight on the network of contacts ( $\beta = 0.3$ ). Both tables are computed with  $a = 0.4$ ,  $b = 0.3$ ,  $c = 0.2$ ,  $d = 0.2$ . This difference between the highlights on the profile has an influence on the membership degrees, hence on the communities that result from these degrees. For instance, when the highlight is placed on Twitter,  $u_2$  and  $u_5$  are in the same community, which is not the case when the highlight is placed on the network of contacts. We now describe the results by analysing the results for a few remarkable users.

When studying user behaviour, we noticed that  $u_3$ ,  $u_5$  and  $u_{13}$  have similar interests: *car*, *transportation*. Also,  $u_{15}$  likes *transportation* too.  $u_8$  and  $u_3$  also have common interests: *cars* and *consulting*. They are regrouped in the same community.  $u_3$  and  $u_5$  also tweet, re-tweet and bookmark a lot about *cars* and *transportation*.

$u_{16}$  doesn't grade a lot of resources, and ends up as the sole member of his community. However, he consults resources, and on the same topics as  $u_7$  and  $u_{20}$ .  $u_7$  and  $u_{20}$  have common interests and grade resources on the same topics: *medias* and *Internet*.  $u_7$  and  $u_{20}$  are in the same community, with other users, if we use only the explicit profile. If we use the explicit and the implicit profile,  $u_7$ ,  $u_{16}$  and  $u_{20}$  are regrouped together in the same community. They also have crossed common interests: *medias* and *phones* for  $u_{16}$  and  $u_{20}$  and *medias* and *internet* for  $u_7$  and  $u_{20}$ .

$u_4$  and  $u_9$  are always in the same community, regardless of the weighting coefficients. They have the same interests, and grade and consult resources accordingly.  $u_2$  is also often regrouped with  $u_4$  and  $u_9$ , depending on the weighting coefficients. The more relevant the bookmarks on Twitter are, the less  $u_2$  is likely to be with  $u_4$  and  $u_9$ . Indeed,  $u_2$  and  $u_4$  are on Twitter, but do not bookmark resources the same way. They share some common interest on the Web site and on Twitter, mostly *tourism*.

Even though  $u_2$  and  $u_5$  are in contact on Twitter, they don't share the same interests at all, and they rarely are in the same community. They don't grade the same resources, so when using the explicit profile of the other user to refine their own profile, the membership degrees are affected but not in a significant enough way to make the user change

**Table 5: Non refined communities**

Communities	Explicit profile
1	$U_3, U_5, U_8, U_{11}, U_{13}, U_{14}$
2	$U_{10}, U_{18}$
3	$U_1, U_2, U_4, U_6, U_7, U_9$ $U_{12}, U_{15}, U_{17}, U_{19}, U_{20}$
4	$U_{16}$

**Table 6: Refined communities 1**

Communities	$\alpha = 0.5$ et $\beta = 0.3$ et $\gamma = 0.2$
1	$U_1, U_3, U_5, U_6, U_8, U_{11}$ $U_{13}, U_{14}, U_{15}, U_{17}, U_{19}$
2	$U_{10}, U_{12}, U_{18}$
3	$U_2, U_4, U_9$
4	$U_7, U_{16}, U_{20}$

community. However, they have a common interest on Twitter, *real estate*. This common interest bring them together in the same community when the highlight is placed on the actions on Twitter.

$u_1$  is interested in most of the categories of the Web site and participates a lot, both on the Web site and on Twitter. He is also in contact with  $u_3$  on Twitter. When using the explicit profile,  $u_1$  and  $u_3$  are not in the same community, because they don't grade resources the same way. However, when using both the explicit and implicit profile, they are in the same community, regardless of the weighting coefficients. They visit resources with the same tag, and also their actions of Twitter have tags in common: *cars*, *consulting*, and *energy*.

$u_{10}$  and  $u_{18}$  share a common interest: *volume retailing*. They are the only users who are interested in this category in the system. They are regrouped into a community when using only the explicit profile.  $u_{12}$  is always in their community when using the explicit and implicit profile. He doesn't grade resources the same way  $u_{10}$  and  $u_{18}$  do, but he does consult quite a lot of resources with the tag *volume retailing*.

## 4.3 Summary of the experiments

By taking into account the user behaviour and by adding into the equation the contact network of the user, our approach is able to regroup users into communities more relevant than the communities created with only a simple profile. The different weighting coefficients of the refined profile can help discovering communities on the fly, by making them variate according to the different parameters one wants to highlight: grades given by the users, the network of contacts, or the consultation of resources and the activity on Twitter. This tweaking enables users of our tool (a Community Manager) to favour an aspect of the user behaviour according to the context and the activity of the users, and according to the type of community he wants to build. It is also possible to make the communities vary according to the number of clusters set in the k-means algorithm.

**Table 7: Refined communities 2**

Communities	$\alpha = 0.5$ et $\beta = 0.2$ et $\gamma = 0.3$
1	$U_1, U_2, U_3, U_5, U_8, U_{11}, U_{13}, U_{14}, U_{15}$
2	$U_{10}, U_{12}, U_{17}, U_{18}, U_{19}$
3	$U_4, U_6, U_9$
4	$U_7, U_{16}, U_{20}$



## 5. CONCLUSION AND FUTURE WORK

In this paper, we have presented DisCoCRM, a community detection method based on the usage, the behaviour and the relationships (social network) of users. Behavioural and social data allow us to refine the user profile and to create communities based on the interests of users, but also on the thematics that interested them at a given moment and on the interests of their contacts. This allows us to compute dynamic communities that change with the behaviour of the users in the system. It also gives opportunities in terms of recommendation of users and resources. DisCoCRM also offers the possibility to discover communities by tweaking the different weighting coefficients of the refined profiles, thus allowing the detection of communities to evolve according to the criteria one wants to see reinforced.

Our future research is directed towards the development of a recommender system, in order to recommend resources to consult, thematics that can interest users, and also to recommend users that have similar interest and / or behaviour in order to take advantage of the communities and the social network of a user. We don't plan on developing a new recommender system from scratch, *per se*. Our approach will fit between the application from where the data is derived and the recommender system that uses the data to compute recommendations. Our goal is to improve these recommendations by enhancing the data used by the recommender system.

As we said in our experiments, we plan to refine the weighting coefficients values, according to the different type of communities one might want, by defining templates and taking into account the feedback from the Community Managers. We also plan to work on bigger user, resources and tag databases. At the application level, we would like to extend the gathering of information by using other social networks on the Web such as Facebook, or forums.

## 6. ACKNOWLEDGEMENTS

This work is sponsored by a CIFRE grant, number 2012/0261, funded by the company eb-Lab in Dijon, France.

## 7. REFERENCES

- [1] F. Abel, S. Araújo, Q. Gao, and G.-J. Houben. Analyzing cross-system user modeling on the social web. In *Web Engineering*, pages 28–43. Springer, 2011.
- [2] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Semantic enrichment of twitter posts for user profile construction on the social web. In *The Semantic Web: Research and Applications*, pages 375–389. Springer, 2011.
- [3] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749.
- [4] J. Ajmera, H.-i. Ahn, M. Nagarajan, A. Verma, D. Contractor, S. Dill, and M. Denesuk. A crm system for social media: challenges and experiences. In *Proceedings of the 22nd international conference on World Wide Web*, pages 49–58. International World Wide Web Conferences Steering Committee, 2013.
- [5] S. Cohen, B. Kimelfeld, and G. Koutrika. A survey on proximity measures for social networks. *Search Computing*, 7538:191–206, 2012.
- [6] C. S. Firan, W. Nejdl, and R. Paiu. The benefit of using tag-based profiles. In *Web Conference, 2007. LA- WEB 2007. Latin American*], pages=32–41, year=2007, organization=IEEE.
- [7] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.
- [8] J. Golbeck. Trust and nuanced profile similarity in online social networks. *ACM Transactions on the Web (TWEB)*, 3(4):12, 2009.
- [9] C.-C. Hung, Y.-C. Huang, J. Y.-j. Hsu, and D. K.-C. Wu. Tag-based user profiling for social media recommendation. In *Workshop on Intelligent Techniques for Web Personalization & Recommender Systems at AAAI*, 2008.
- [10] N. Imafuji and M. Kitsuregawa. Effects of maximum flow algorithm on identifying web community. In *Proceedings of the 4th international workshop on Web information and data management*, pages 43–48. ACM, 2002.
- [11] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.
- [12] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the web for emerging cyber-communities. *Computer networks*, 31(11):1481–1493, 1999.
- [13] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
- [14] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali. Cluster-based landmark and event detection on tagged photo collections. *IEEE Multimedia*, 18(1):52–63, 2010.
- [15] H. Quan. *Online Social Networks & Social Network Services: A Technical Survey*. CRC Press, 2011.
- [16] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658–2663, 2004.
- [17] S. Smyth and S. White. A spectral clustering approach to finding communities in graphs. In *Proceedings of the fifth SIAM international conference on data mining*, volume 119, page 274, 2005.
- [18] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2 edition, 2005. (Ercument-2011-11-01).
- [19] B. Wu, Q. Ye, S. Yang, and B. Wang. Group crm: a new telecom crm framework from social network perspective. In *Proceedings of the 1st ACM international workshop on Complex networks meet information & knowledge management, CNIKM '09*, pages 3–10, New York, NY, USA, 2009. ACM.