



HAL
open science

Coévolution d'organisations sociales et spatiales dans les systèmes multi-agents : application aux systèmes de tagging collaboratifs

Maya Samaha Rupert

► To cite this version:

Maya Samaha Rupert. Coévolution d'organisations sociales et spatiales dans les systèmes multi-agents : application aux systèmes de tagging collaboratifs. Autre [cs.OH]. Université Claude Bernard - Lyon I, 2009. Français. NNT : 2009LYO10126 . tel-00688065

HAL Id: tel-00688065

<https://theses.hal.science/tel-00688065v1>

Submitted on 16 Apr 2012

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.

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE LYON

(Spécialité informatique)

Délivrée par L'UNIVERSITE CLAUDE BERNARD LYON 1

ECOLE DOCTORALE INFORMATIQUE ET MATHEMATIQUES

Présentée et soutenue publiquement par

Maya Rupert

Le 2 Septembre 2009

*Coévolution d'organisations sociales et spatiales dans les systèmes
multi-agents: application aux systèmes de tagging collaboratifs*

Jury

Mr. Olivier Boissier	Professeur ENS Mines Saint-Etienne	Rapporteur
Mr. Vincent Chevrier	Professeur Université de Nancy	Rapporteur
M ^{me} Salima Hassas	Professeur Université Lyon1	Directrice de thèse
Mr. Bernard Espinasse	Professeur Université Aix-Marseille	Examineur
Mr. Hamamache Kheddouci	Professeur Université Lyon1	Examineur

Unité de Recherche : Laboratoire d'informatique pour l'entreprise et les systèmes de production - LIESP

A ma famille, et en particulier à mon mari.

Remerciements

Je remercie tout d'abord ma directrice de thèse, Mme Salima Hassas, Professeur à l'Université Lyon 1 pour sa patience durant mes années de thèse et pour toutes nos discussions par communication électronique qui ont abouti à mettre en valeur les idées présentées dans cette thèse.

Je remercie aussi les membres du jury qui m'ont fait l'honneur de participer à ma soutenance : Mr. Olivier Boissier, professeur ENS Mines Saint-Etienne, Mr. Vincent Chevrier, professeur Université de Nancy, M^{me} Salima Hassas, professeur Université Lyon1, Mr. Bernard Espinasse, professeur Université Aix-Marseille, et Mr. Hamamache Kheddouci, professeur Université Lyon1.

Enfin, je remercie le support de mon mari Roger, mes enfants James et Patrick, ma mère et mes frères et ma sœur pour leur patience et compréhension durant mes années de thèse.

Résumé

L'évolution du Web et de ses applications subit depuis quelques années une mutation vers les technologies qui incluent la dimension sociale comme entité de première classe. Nous témoignons dans le passage du Web 1.0 au Web 2.0 puis au Web 3.0, 4.0 etc.. que les utilisateurs et les réseaux sociaux qui se forment sont au centre de cette évolution. Le web exhibe aussi toutes les caractéristiques d'un système complexe. Ces propriétés systèmes complexes et cette dimension sociale doivent être prises en considération lors de la conception et le développement des applications web. Considérons le cas des systèmes de tagging ou d'étiquetage collaboratifs. Ces systèmes sont un exemple de systèmes complexes, auto-organisés et socialement conscients. Le paradigme des systèmes multi-agents coordonné par les mécanismes d'auto-organisations a été utilisé d'une façon effective pour la conception et modélisation des systèmes complexes. Les systèmes de tagging collaboratifs actuels ne prennent pas l'avantage complet de leurs caractéristiques systèmes complexes, surtout dans l'adaptation à leur environnement et l'émergence de nouvelles fonctionnalités. Dans ce travail de thèse, nous proposons un modèle pour la conception et développement d'un nouveau système d'étiquetage collaboratif MySURF (My Similar Users, Resources, Folksonomies), utilisant une approche multi-agents gouvernée par la coévolution des organisations sociales et spatiales des agents. Nous montrons comment ce système proposé offre plusieurs nouvelles fonctionnalités qui peuvent améliorer les systèmes d'étiquetage collaboratifs actuels.

Mots-clés: Organisations dans les systèmes multi-agents, systèmes de tagging collaboratif.

Abstract

The evolution of the Web and its applications has undergone in the last few years a mutation towards technologies that include the social dimension as a first class entity. We are witnessing in the evolution of the web from the web 1.0 to web 2.0 to web 3.0 and eventually web 4.0 that the users, their interactions and the emerging social networks are in the center of this evolution. The web also exhibits all the characteristics of a complex system. These complex systems properties and this social dimension must be taken into consideration in the design and the development of new web applications. Let us consider the case of collaborative tagging systems. These systems are an example of complex, self-organized and socially aware systems. The multi-agent systems paradigm coordinated by self-organizations mechanisms was used in an effective way for the design and modeling of the complex systems. Current collaborative tagging systems do not take full advantage of the characteristics of complex systems, especially in adapting to their environment and the emergence of new features. In this thesis, we propose a model for the design and development of a new collaborative tagging system MySURF (My Similar Users, Resources, Folksonomies), using a multi-agent system approach governed by the coevolution of the social and spatial organization of the agents. We show how the proposed system offers several new features that can improve current collaborative tagging systems.

Keywords: Organizations in multi-agent systems, collaborative tagging systems.

Table des matières

Chapitre 1 Introduction.....1

1.1	Évolution du web.....	3
1.1.1	Le web 1.0	4
1.1.2	Le web 2.0	4
1.1.3	Le web 3.0, le web 4.0,.....	5
1.2	Les approches multi-agents pour le web comme système complexe	6
1.3	Problématique de recherche	7
1.4	Contributions et apports de la thèse.....	8
1.5	Plan du mémoire.....	9

Chapitre 2 Travaux Connexes et Positionnement.....11

2.1	Les systèmes complexes adaptatifs (Exemple de l'évolution du web)	12
2.1.1	Émergence et auto-organisation	13
2.1.2	Adaptation, coévolution et dynamiques	14
2.1.3	Les propriétés de Holland.....	15
2.1.4	Le web comme réseau complexe.....	17
2.1.4.1	La topologie du graphe du web	17
2.1.4.2	La propriété des réseaux invariants d'échelle ou « scale free »	17
2.1.4.3	Le phénomène « small world »	18
2.1.4.4	La structure “bowtie” et l'autosimilarité du web.....	19
2.2	Les systèmes multi-agents.....	20
2.2.1	Les systèmes multi-agents situés et l'environnement.....	20
2.2.2	La stigmergie	20
2.3	Une vue d'ensemble des approches systèmes multi-agents pour le web	21
2.3.1	Systèmes multi-agents pour la Recherche d'Information sur le Web	22
2.3.2	Le Web « Mining » et les systèmes multi-agents.....	24
2.3.3	Le Web Sémantique et les systèmes multi-agents.....	25
2.4	Synthèse et discussion	26

Chapitre 3 Cadre pour les organisations sociales et spatiales dans les systèmes multi-agents.....28

3.1	Les organisations dans les systèmes multi-agents.....	29
3.2	Modèle proposé	30
3.2.1	Organisation sociale	31
3.2.2	Organisation spatiale	35
3.2.3	Co-évolution de l'organisation sociale et l'organisation spatiale.....	37
3.2.4	Représentation par les transformations de graphe.....	39
3.2.5	Phase d'analyse	41
3.3	Synthèse.....	41

Chapitre 4 Application à la simulation d'une communauté virtuelle.....43

4.1	Éléments de base	44
4.1.1	Les communautés virtuelles	44
4.1.2	La simulation à base d'agents.....	45
4.1.3	Description de la communauté virtuelle considérée.....	46
4.2	Description du système.....	48
4.3	L'organisation spatiale dans la communauté virtuelle	49
4.3.1	Les places	49
4.3.2	Les chemins.....	50
4.3.3	Les régions.....	50
4.3.4	Les phéromones.....	50
4.4	L'organisation sociale dans la communauté virtuelle	51
4.4.1	Rôle des agents et leurs contraintes.....	51
4.4.2	Actions des agents	54
4.5	Simulation de l'évolution de la communauté virtuelle du support de la santé.....	56
4.5.1	Implémentation.....	56
4.5.2	Environnement de simulation.....	57
4.6	Les résultats	59

4.6.1	Nombre d'assignations de rôles.....	59
4.6.2	Dépendance entre la position spatiale et la position sociale de l'agent.....	60
4.6.3	Créations de nouvelles places et régions.....	61
4.6.4	Groupement par région des places similaires.....	61
4.7	Conclusion.....	62

Chapitre 5 Proposition du système d'étiquetage collaboratif MySURF ..64

5.1	Parcours des systèmes d'étiquetage existants et de leurs limites	65
5.1.1	Les réseaux sociaux en ligne	65
5.1.2	Folksonomie et systèmes d'étiquetage collaboratifs	66
5.1.3	Quelques exemples de systèmes d'étiquetage collaboratifs	67
5.1.4	Principes fondateurs des systèmes d'étiquetage collaboratifs.....	69
5.2	Étude des systèmes d'étiquetage collaboratifs comme réseaux complexes	70
5.2.1	La propriété « petit monde » des réseaux de folksonomie	71
5.2.2	La Propriété des Réseaux Invariant d'échelle dans les réseaux de folksonomie	72
5.2.3	Les limites dans les systèmes d'étiquetage actuels	72
5.3	Proposition d'un système qui répond à certaines limites dans les systèmes existants	75
5.3.1	Fonctionnalités du système MySURF	76
5.3.2	Les organisations sociales et spatiales des agents dans le système d'étiquetage proposé	79
5.4	Description détaillée du système d'étiquetage MySURF.....	79
5.4.1	Émergence des sous-communautés de ressources.....	84
5.4.1.1	Description de l'algorithme de clustering des ressources	84
5.4.1.2	Le clustering spectral.....	87
5.4.1.3	Partitionnement du graphe à partir des vecteurs propres de la matrice Q	90
5.4.2	Les étiquettes-systèmes	92
5.4.3	Les groupes virtuels d'utilisateurs.....	93
5.4.3.1	<i>Personomies</i> pour chaque utilisateur.....	93

5.4.3.2	Analyse du degré de similarité entre 2 utilisateurs.....	94
5.4.3.3	Raffinement des groupes virtuels d'utilisateurs et des sous- communautés de ressources.....	94
5.4.3.4	Algorithme de l'émergence des étiquettes-systèmes.....	95
5.5	Expériences et Résultats	97
5.5.1	Implémentation.....	97
5.5.2	Données expérimentales de Del.icio.us.....	99
5.5.3	Résultats	101
5.5.3.1	Visualisation du clustering spectral.....	102
5.5.4	Données expérimentales dans MySURF et résultats de l'échantillon créé manuellement	105
5.5.5	Exemple d'un cycle d'étapes dans MySURF pour l'étiquette « Design »	106
5.6	Synthèse.....	110
Chapitre 6 Conclusion et travail futur		112
6.1	Contributions	113
6.2	Travail futur.....	114
Annexe A		116
Bibliographie.....		119

Liste des figures

Figure 1.1 Cycle de vie du «Web Science ».....	3
Figure 2.1 La structure «bowtie» du graphe web (Broder, Kumar et al. 1999).....	19
Figure 3.1 Organisation Sociale.....	37
Figure 3.2 Organisation Spatiale.....	37
Figure 3.3 Coévolution des organisations spatiales et sociales.....	37
Figure 3.4 Transformation entre le graphe social et le graphe spatial.....	40
Figure 4.1 La page d'accueil captée le 10 septembre 2008.....	47
Figure 4.2 Les flèches sur le diagramme représentent une séquence d'actions et montrent comment l'application sera exécutée.	48
Figure 4.3 Diagramme d'un scénario possible où les questions et réponses forment les places dans l'organisation spatiale séquences pour les interactions des agents	55
Figure 5.1 Le nombre des ressources étiquetées par le tag « design » est aux alentours de 1.5 millions de ressources.	74
Figure 5.2 montre les spécifications du système en termes des différentes tâches effectuées par les utilisateurs	83
Figure 5.3 Le réseau utilisé comme exemple, avec $S = 19$ et d'une valeur aléatoire entre 1 et 10 attribuée aux liens. Trois clusters clairs apparaissent, composés par les nœuds 0 – 6, 7 -12 et 13 – 19 (Capocci, Servedio et al. 2004).	89
Figure 5.4 Les valeurs du 2e vecteur propre relatives au graphe représenté sur la figure 5.3.....	89
Figure 5.5 Le vecteur Fiedler de la matrice Q calculée pour 92 ressources liées à l'étiquette « design » et extraites de del.icio.us	91
Figure 5.6 Le vecteur Fiedler de la matrice Q calculée pour 200 ressources liées à l'étiquette « programming » et extraites de del.icio.us	91
Figure 5.7 Cycle d'évolution du système.....	96
Figure 5.8 Modèle conceptuel simplifié de la base de données	98
Figure 5.9 Résultats de recherche de l'étiquette « toread » dans Del.icio.us.....	100
Figure 5.10 Résultats de la recherche de l'étiquette «toread» dans notre système basé sur les 99 ressources « toread » extraites de Del.icio.us. Les resultats sont groupés en sous-communautés « toread programming », « toread articles ».....	101
Figure 5.11. Valeurs propres triées de la Matrice Q. La présence d'au moins 4 valeurs propres bien séparées de valeur non zéro indique la présence d'au moins 4 différents « clusters ».....	102
Figure 5.12. Représentation de cluster de ressources pour l'étiquette "toread", montrant la présence de quatre différents «clusters» de ressources.....	103
Figure 5.13 Quatre sous-communautés de ressources sont générées quand un clustering est effectué sur l'étiquette « Design »	107
Figure 5.14 <i>User1</i> appartient à 2 groupes virtuels : « Design Web » et « Design Home ».....	108
Figure 5.15 Le tableau de similarité de <i>User1</i> en comparaison avec tous les autres membres des groupes auquel il appartient. Notons que <i>User1</i> est fortement similaire à <i>User2</i> et <i>User3</i> car tous les trois partagent un intérêt commun en « Vancouver ». Ils ont étiquetées des ressources communes avec l'étiquette « Vancouver ».....	109

Figure 5.16 La recommandation personnalisée de *User1* pour le sujet de « web design » est la page web étiquetée par *User2* qui lui est fortement similaire. Notons que cette page n'apparaît pas dans la sous-communauté « web design » car il n'existe pas assez d'étiquetage de cette page pour la qualifier à la sous-communauté..... 110

Liste des tableaux

Tableau 4.1 Liste des rôles des agents qui seront utilisés dans l'application.....	52
Tableau 4.2 Liste des rôles des agents et de leurs contraintes correspondantes.....	52
Tableau 4.3 Phéromones cherchées et déposées selon les rôles des agents.....	54
Tableau 4.4 1 Fondateur 20 Fournisseurs d'Information.....	58
Tableau 4.5 20 Fournisseurs d'Opinion.....	58
Tableau 4.6 83 Chercheurs d'Information.....	58
Tableau 4.7 27 Chercheurs d'Opinion.....	59
Tableau 4.8 Type de place favorisé selon le rôle de l'agent.....	60
Tableau 4.9 Calcul du pourcentage de places dans la région où le sujet principal de la place est le même que le sujet de la région.....	62
Tableau 4.10 Pourcentage moyen des places voisines ayant le même sujet principal est 99.32% ..	62
Tableau 5.1 représentant les valeurs des éléments des dix premiers vecteurs propres de la matrice Q (colonnes 1 à 10) pour 40 ressources (lignes 1 à 40).	104
Tableau 5.2 Matrice C de corrélation calculée à partir des valeurs des éléments des vecteurs propres de la matrice Q. Ce tableau montre les valeurs pour les 20 premières ressources.....	105

Chapitre 1

Introduction

Le web continue à croître et à évoluer à un taux très rapide devenant une source d'informations primaire et indispensable. Au cours des dix dernières années, nous avons assisté à une révolution dans le contenu, l'usage et la structure du web et de ses différentes applications. Cette évolution est en grande partie motivée par les besoins des utilisateurs. Le passage du web d'une source d'informations statique (web 1.0) vers un support d'activités (communautés virtuelles), puis vers un système collaboratif (web 2.0) prend en compte l'activité de l'utilisateur comme composante principale de cette évolution.

Un domaine interdisciplinaire en a émergé : le « Web Science ». Les chercheurs ou (Web scientists) qui tentent de définir et d'étudier cette nouvelle « science » affrontent des problèmes émergents qui représentent plusieurs défis pour le futur du web. Au cours de ce travail de thèse, qui a débuté bien avant que ces concepts soient introduits, nous avons essayé de répondre en partie à une des questions posées dans (Hendler, Shadbolt et al. 2008) :

“What underlying architectural principles are needed to guide the design and efficient engineering of new Web infrastructure components for this social software?”

Partant de l'idée que le Web a dépassé sa description d'être juste une source d'information, pour devenir un mécanisme de contenu dynamique soutenant les interactions sociales entre plusieurs utilisateurs, ces auteurs pionniers du Web tels que Tim Berners-Lee, James Handler et Nigel Shadbolt ont introduit la notion de machines sociales. Les applications Web les plus populaires d'aujourd'hui tel que Facebook, MySpace, Flickr, Del.cio.us, YouTube et la blogosphere sont considérées comme « les toutes premières machines sociales, limitées par le fait qu'elles sont largement isolées les unes des autres ». La question posée ci-dessus est basée sur l'hypothèse suivante (Hendler, Shadbolt et al. 2008):

- Ces machines vont évoluer pour devenir beaucoup plus efficaces dans le futur
- Les différents processus sociaux qui sont reliés dans la société doivent être reliés dans le Web
- La conception de ces machines est un travail collectif dépendant de la technologie qui va permettre aux communautés d'utilisateurs de construire, partager et adapter ces machines sociales.

Dans le cycle de vie du « Web Science » (Fig.1.1), un des attributs de ce cycle est le « Social ». Nous argumentons que la dimension sociale est le cœur qui assure la vie à ce cycle, et qu'il affecte directement et indirectement tous les autres attributs. De ces interactions sociales au niveau de chaque individu au niveau microscopique, émerge un comportement global au niveau macroscopique non prédictible, qui influence l'évolution et le futur du web. Le web dans cette évolution exhibe toutes les propriétés d'un système complexe.

Dans notre travail, nous avons intégré cette dimension sociale et sa dynamique dans la conception d'un système informatique possédant les propriétés d'un système complexe. Ceci est exprimé à travers les couplages structures-comportements et leurs co-évolutions rétro-actives.

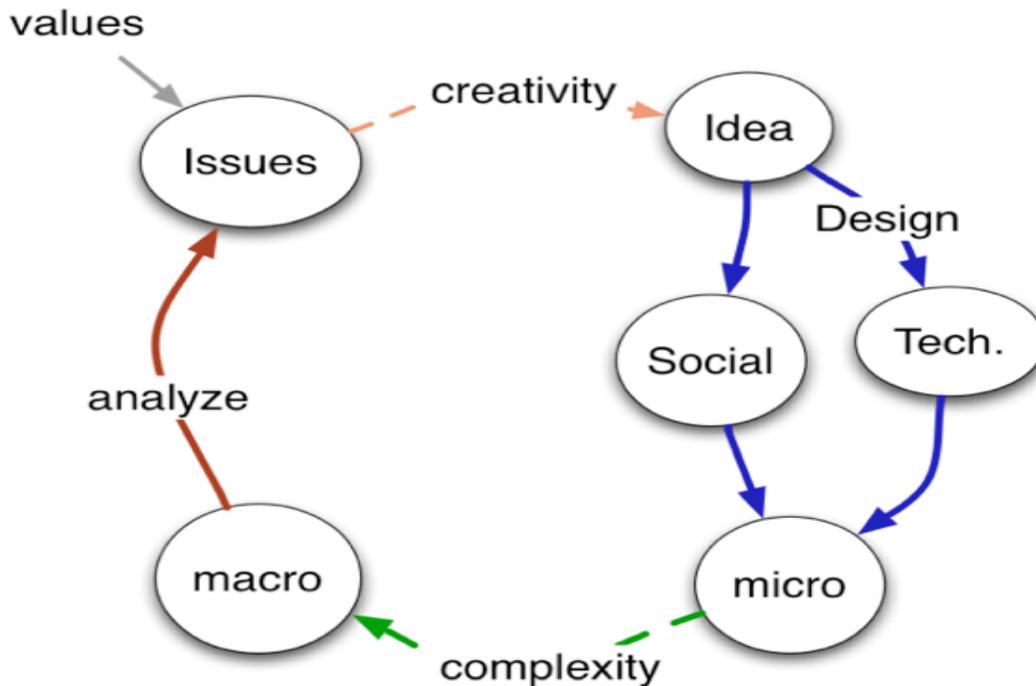


Figure 1.1 Cycle de vie du «Web Science »

1.1 Évolution du web

Nous considérerons l'évolution du web selon la vision des systèmes complexes. Le passage du web 1.0 au web 2.0 au web 3.0 est le résultat de l'agrégation des contributions des centaines de millions d'utilisateurs. De ces interactions locales des contributions des utilisateurs complétées par des agents logiciels, un comportement macro global en a émergé. Le web et son évolution ne sont contrôlés par aucune autorité, et en dépit du chaos généré par ce volume intensif de contenu et d'usage, l'ordre en a émergé et le système s'est auto-organisé.

Cette transition évolutionnaire du web s'est achevée par étapes directionnelles (Ding and Xu 2007). Auparavant, on croyait que l'évolution du web était contrôlée uniquement par l'évolution de la technologie et de ses « designs » intelligents. Mais le progrès phénoménal du web 2.0 contredit cette croyance, comme ce sont les utilisateurs et leurs interactions locales qui dictent cette évolution, le design intelligent est devenu une question secondaire dans cette

évolution. C'est ce qui nous a motivé d'accorder une importance particulière à la dimension sociale et son rôle dans l'évolution du web.

1.1.1 Le web 1.0

Le web 1.0 était relativement statique au début, où des auteurs créent des pages web et les utilisateurs reçoivent passivement le contenu de ces pages. Au cours du temps, les pratiques des utilisateurs ont évolué vers plus d'interactivité avec le web, ce qui a abouti à l'émergence des communautés en ligne ou communautés virtuelles.

Le Web 1.0, a fait l'objet de nombreuses études, que ce soit au niveau de sa structure comme les travaux sur le graphe du web, au niveau de la caractérisation/compréhension de son contenu comme dans les domaines de la recherche et l'extraction d'information (Information Retrieval) ou encore les domaine du « Web Mining » et du web sémantique et enfin au niveau des usages du web qui sont de plus en plus interactifs, collectifs et intégrant de plus en plus la dimension sociale.

Ces différentes études du web 1.0 et de ses propriétés ont montré que le web possède toutes les propriétés d'un réseau complexe. Nous voyons l'évolution des réseaux complexes et de leur usage comme un processus autopoétique selon la vision de Varela (Varela, Thompson et al. 1991). La structure de ces réseaux a la particularité d'être guidée dans son évolution par les usages. Ces derniers évoluent, à leur tour, en réaction à l'évolution la structure. La topologie du réseau est alors le résultat d'une boucle de rétroaction entre l'usage du réseau et sa structure. La base de notre travail est la prise en compte dynamique de ces trois caractéristiques du web (contenu, structure, usage) en s'appuyant sur la vision des systèmes complexes.

1.1.2 Le web 2.0

Les besoins des utilisateurs ont continué à évoluer, entraînant l'évolution de la structure du web avec eux. Le web 2.0 a émergé, en première étape, avec la création de larges réseaux sociaux comme Facebook, MySpace et YouTube, des RSSs et des blogs. En deuxième étape,

des concepts tels que la folksonomie (VanderWal 2005) ou les systèmes d'étiquetages collaboratifs, et les personomies (Hotho, Jaschke et al. 2006) sont apparus après les réseaux sociaux comme un besoin pour mieux exploiter cette évolution du web. La définition ci-dessous prise de Wikipedia décrit cette évolution au web 2..0 :

*“Le **Web 2.0** est un terme vivant décrivant les tendances changeantes dans l'utilisation de la technologie du [World Wide Web](#) et du [web design](#) qui vise à mettre en valeur la créativité, le partage de l'information, la collaboration et la fonctionnalité du web.”*

Notre vision du web 2.0 selon les systèmes complexes nous ramène à la notion d'environnement dans les systèmes multi-agents. Le web représente un environnement actif et dynamique où les utilisateurs (ou agents) laissent des traces par leurs interactions et contributions. C'est le phénomène de stigmergie (Grassé 1959). Ces interactions aboutissent à un changement dans l'environnement et ce changement affectera les actions futures des agents.

1.1.3 Le web 3.0, le web 4.0,...

Notons que pour l'évolution au web 3.0, les chercheurs n'ont pas, jusqu'à présent (lors de la rédaction de la thèse), une vision claire du futur du web. Certains chercheurs du web pensent que l'innovation majeure que le web 3.0 va avoir, c'est l'intégration du web sémantique qui est tellement anticipée. Mais en observant l'évolution du web, les prédictions que le web 2.0 va être le web sémantique, n'étaient pas correctes alors nous restons sceptiques sur la forme que le web 3.0 va avoir.

Quant au web 4.0, un mot est persistant dans les différents blogs qui abordent des prédictions du web 4.0 et c'est le mot « connections » (Godin, 2007). Donc la dimension sociale va croître de plus en plus.

Le web exhibe dans son évolution des propriétés similaires à celles d'un système complexe adaptatif. L'étude du web et de son évolution, sa complexité et son ouverture, etc. fait ressortir

la nécessité de disposer d'un support de modélisation permettant la mise en évidence de ces caractéristiques afin de faciliter et favoriser leur exploitation.

1.2 Les approches multi-agents pour le web comme système complexe

L'évolution du web comme système et son ouverture à son environnement dynamique requièrent des approches qui traitent d'une façon effective sa complexité croissante. Ce système exhibe des comportements adaptatifs, auto-organisés similaires à ceux des systèmes complexes adaptatifs: des systèmes qui sont composés de plusieurs éléments en interaction continue, créant des comportements émergents. Le comportement est dit émergent parce qu'au niveau macroscopique, le système génère de nouvelles propriétés complexes qui ne peuvent pas être observées au niveau local des différentes composantes. Ce comportement global est le fruit d'une agrégation non linéaire des comportements locaux. Les systèmes complexes adaptatifs s'auto-organisent et s'adaptent aux changements de leur environnement en l'absence d'un contrôle central gouvernant leurs comportements. Les systèmes complexes adaptatifs sont des systèmes non linéaires dont "le tout est plus que la somme de ses parties" (Kauffman 1993; Holland 1995). Dans ces systèmes, l'ordre émerge à travers des processus d'auto-organisation.

Le système informatique est alors considéré comme un ensemble de composants en interactions interrelationnelles et rétroactives, évoluant dans un environnement partagé dynamique et incertain. Le paradigme systèmes multi-agents couplé à des mécanismes d'auto-organisation et d'émergence a été utilisé pour le développement et la modélisation de ces systèmes (Serugendo, Gleizes et al. 2005). Le rôle de l'environnement est pris en considération comme une entité de première classe en construisant les systèmes multi-agents (Weyns, Schumacher et al. 2005).

Pour le développement des systèmes informatiques capables de s'adapter adéquatement à l'évolution de leur environnement, nous avons adopté le cadre proposé dans (Hassas 2003; Hassas 2005) qui est basé sur le couplage entre le système et son environnement :

- Un couplage structurel exprimé par la coévolution de la structure du système et son environnement

- Un couplage comportemental exprimé par la coévolution du comportement du système et de son environnement.

L'effet rétroactif d'un couplage sur l'autre : les changements qui se produisent dans la structure et le comportement du système durant l'évolution de l'environnement sont corrélés.

1.3 Problématique de recherche

Le travail de recherche présenté dans ce mémoire de thèse s'intéresse à la question de développement de systèmes informatiques capables d'évoluer et de s'adapter à leur environnement, tout en assurant l'émergence de nouvelles pratiques. Cette émergence est due à l'usage qui affecte l'environnement du système qui, à son tour, affecte l'usage. Ces systèmes se distinguent par les caractéristiques suivantes:

Leur Complexité exprimée par:

- leur grand volume de ressources et de données
- leur représentation par des réseaux complexes
- leur large distribution
- leur dynamique

Leur Ouverture: ces systèmes réels sont ouverts sur leur environnement et sur l'utilisateur.

Leur Auto-Organisation : ces systèmes n'ont aucun contrôle global guidant leur organisation. Ils s'auto-organisent vers de nouvelles structures émergentes mises en œuvre par l'usage.

Cette problématique se situe dans le domaine des systèmes multi-agents situés où les agents sont des entités réelles ou virtuelles, évoluant dans un environnement, capables de le percevoir et d'agir dessus.

Le système multi-agent est le système formé par l'ensemble de ces agents autonomes, situés dans un environnement qu'ils partagent, interagissant les uns avec les autres, en communication directe ou indirecte. Les agents autonomes et les systèmes multi-agents sont

de plus en plus utilisés pour la simulation, la modélisation et le développement des systèmes complexes.

Dans ces systèmes, le rôle de l'environnement est fondamental. Un des différents rôles que doit jouer l'environnement comme il a été défini dans (Weyns, Parunak et al. 2005) est :

“Structuration. L'environnement est, premièrement, un « espace » commun partagé pour les agents, qui structure le système entier. Une telle structuration peut être spatiale ou organisationnelle. Des propriétés spécifiques différentes peuvent être définies séparément pour chaque espace, telles que les positions, la région, les groupes ou les rôles.”

Nous avons essayé de résoudre ce problème par la proposition selon le cadre défini par (Hassas 2005) d'un modèle de système multi-agents organisationnel pour le développement de ces systèmes complexes ouverts, permettant leur évolution dynamique. L'utilisation des systèmes multi-agents situés permet de construire des systèmes informatiques incorporés dans leur environnement. La conception du système multi-agents doit considérer les différents couplages entre le système et son environnement :

- Un couplage spatial représenté par l'organisation spatiale du système multi-agents.
- Un couplage social représenté par l'organisation spatiale du système multi-agents.
- La coévolution des deux organisations à travers la dynamique du système multi-agents.

1.4 Contributions et apports de la thèse

Nous avons proposé un modèle pour le développement des systèmes complexes à base des systèmes multi-agents. Le modèle prend en considération la coévolution de l'organisation sociale et de l'organisation spatiale du système multi-agent et l'effet rétroactif d'une organisation sur l'autre.

Nous avons appliqué ce modèle à un premier cas d'étude portant sur la simulation d'une application du web 1.0 qui consiste en une communauté virtuelle. Les communautés virtuelles sont un bon exemple de systèmes auto-organisés complexes, ayant une bonne capacité

d'adaptation. La simulation d'une communauté en ligne nous donne une meilleure compréhension de son évolution en nous aidant à prédire la contribution des comportements de ses membres, et ainsi, nous permet de déterminer les facteurs clés de son succès, spécialement en augmentant le taux de participation des membres.

En deuxième cas d'étude, et pour approfondir notre modèle proposé, nous nous sommes intéressés au cas du web 2.0 et nous avons appliqué notre modèle pour la conception d'un système d'étiquetage collaboratif. Les systèmes d'étiquetage forment une partie intégrale du web 2.0, ils sont relativement nouveaux et c'est un domaine de recherche émergent très actif au sein de la communauté du web et des réseaux sociaux. L'idée consiste à appliquer les concepts d'organisations sociales et spatiales pour le développement des systèmes d'étiquetage collaboratifs. L'organisation spatiale des ressources va créer des clusters de ressources sémantiquement similaires qui vont être très utiles dans la recherche de ressources. Ce qui va aboutir à la génération de groupes sociaux d'utilisateurs. En analysant les similarités entre ces utilisateurs, le système recommande des ressources plus appropriées selon le goût personnalisé de chaque utilisateur, ce qui va, à son tour raffiner les clusters spatiaux de ressources. Et le cycle de coévolution continue.

Les résultats ont montré que le système d'étiquetage proposé selon notre approche organisationnelle, offre de nouvelles fonctionnalités qui améliorent les systèmes d'étiquetage collaboratifs actuels, surtout au niveau de la recherche des étiquettes qui devient plus personnalisée et performante en tenant compte de la coévolution des aspects sociaux et spatiaux.

1.5 Plan du mémoire

Le reste de cette thèse est organisé de la manière suivante :

Dans le chapitre 2, Travaux Connexes et Positionnement, nous montrons que le web exhibe les propriétés des systèmes complexes adaptatifs. Ce chapitre présente aussi les différentes approches organisationnelles dans les systèmes multi-agents, le mécanisme de stigmergie et l'auto-organisation.

Dans le chapitre 3, nous introduisons **le cadre de développement des systèmes complexes** à partir de l'organisation spatiale et l'organisation sociale des agents.

Dans le chapitre 4, nous présentons un modèle **de simulation d'une communauté en ligne** basé sur le cadre défini au chapitre 3. Nous nous intéressons en particulier à l'effet rétroactif d'une organisation sur l'autre.

Dans le chapitre 5, nous décrivons en détail notre **système d'étiquetage collaboratif MySURF** que nous avons développé, soulignant les améliorations que ce nouveau système peut ajouter aux systèmes existants.

Le chapitre 6 résume nos contributions et concerne les extensions futures liées à ce travail.

Chapitre 2

Travaux Connexes et Positionnement

Dans notre analyse de l'historique du web et de son évolution, nous avons adopté la vision des systèmes complexes à base de multi-agents. Nous avons observé les propriétés des systèmes complexes sur le Web 1.0. Ces propriétés sont exprimées par la topologie du graphe du web, l'auto-organisation, l'émergence, et les éléments fondamentaux des systèmes complexes adaptatifs définis par Holland (Holland 1995) tels que l'agrégation, le tagging, la non-linéarité, le flux et la diversité.

Avec l'apparition du web 2.0, le web est passé d'une collection à grande échelle de documents vers un environnement « stigmergique » partagé par des millions d'agents humains et logiciels, faisant émerger des nouvelles pratiques et une nouvelle manière de diffuser le savoir, les services et les biens. Ceci a transformé le web en un support d'échange intégrant la dimension sociale au centre de son évolution. Cette dimension du web est aussi le fruit et le support du concept de réseaux sociaux qui bien qu'existant depuis les années 50 avec les travaux de Elizabeth Bott, a été complètement réinventé avec l'évolution des outils et technologies.

L'approche que nous développons dans cette thèse s'inscrit dans cet esprit, et permet le développement explicite des infrastructures du web selon cette perspective des systèmes complexes. Le paradigme des systèmes multi-agents offre un outil idéal pour ce développement, en particulier concernant les concepts liés à la dimension sociale pour les représentations collectives et individuelles des agents.

2.1 Les systèmes complexes adaptatifs (Exemple de l'évolution du web)

Il n'existe pas une définition unifiée des Systèmes Complexes Adaptatifs. La majorité des chercheurs dans ce domaine (Waldrop 1992; Kauffman 1993; Gell-Mann 1994; Holland 1995) admettent que les systèmes complexes adaptatifs sont composés de plusieurs éléments en interaction continue, créant des comportements émergents. Le comportement est dit émergent parce qu'au niveau macroscopique, le système génère de nouvelles propriétés complexes qui ne peuvent pas être observées au niveau local des différentes composantes. Ce comportement global est le fruit d'une agrégation non linéaire des comportements locaux. Les systèmes complexes adaptatifs s'auto-organisent et s'adaptent aux changements de leur environnement en l'absence d'un contrôle central gouvernant leurs comportements. Les systèmes complexes adaptatifs sont des systèmes non linéaires dont "le tout est plus que la somme de ses parties" (Holland 1995). Dans ces systèmes, l'ordre émerge à travers des processus d'auto-organisation.

L'étude des systèmes complexes adaptatifs a été appliquée à différents domaines comme l'économie (Arthur, Durlauf et al. 1997), les organisations (Mitleton-Kelly 1997), les écologies (Levin 1998), la biologie, le système immunitaire (Grilo, Cateano et al. 2000) et le cerveau (Morowitz et Singer 1995). Les systèmes complexes adaptatifs partagent plusieurs caractéristiques communes. Dans les sections suivantes, une revue des caractéristiques des systèmes complexes adaptatifs est présentée, montrant que le web exhibe des comportements similaires (Rupert, Hassas et al. 2006).

Le web est composé de plusieurs éléments, interagissant rétroactivement, s'adaptant et évoluant dans un environnement dynamique. Son comportement global et son organisation émergent comme résultats d'une agrégation non linéaire des comportements locaux de ses

éléments. L'environnement du web est un réseau complexe auto-organisé, possédant les propriétés «scale free » et « small world » (Albert, Jeong et al. 1999). Ces caractéristiques sont généralement observées dans tous les systèmes naturels ou artificiels exhibant un caractère auto-organisationnel (Barabasi and Bonabeau 2003).

2.1.1 Émergence et auto-organisation

La caractéristique principale d'un système complexe adaptatif est qu'il est formé de plusieurs éléments ou agents interagissant fortement entre eux. On peut citer en exemple : les atomes, les molécules, les neurones dans le cerveau, les fourmis dans une colonie et l'air et l'eau dans le système météorologique. L'interaction entre ces agents semi-autonomes génère le comportement du système. De la même façon, le web met en jeu plusieurs acteurs (agents) y compris les utilisateurs, les auteurs du web, les moteurs de recherche, les pages web, les hyperliens, les services web et les agents web qui interagissent réciproquement de façon non linéaire.

De l'interaction locale entre ces différents agents émerge un comportement global du système. Le comportement émergent se présentant au niveau macroscopique ne pourrait pas être prédit des niveaux microscopiques. Une approche réductionniste ne peut pas s'appliquer pour comprendre ce comportement. Dans les colonies de fourmis, chaque fourmi suit une piste de phéromone déposée par les autres fourmis pour aboutir à la source de nourriture. La fourmi renforce la piste en déposant plus de phéromone pour que d'autres fourmis puissent suivre la même piste. La phéromone s'évapore avec le temps. Ce procédé est basé sur des boucles de rétroactions positives et négatives et le comportement collectif émergent est dit auto-catalytique. Dans le comportement d'une volée d'oiseaux, chaque oiseau suit ses propres règles locales (comme se tenir à une certaine distance des autres oiseaux, voler dans la direction du centre de la volée) et la volée émerge comme un tout, composant un système ayant son propre comportement.

Dans les systèmes complexes adaptatifs, le contrôle qui dicte le comportement du système est non centralisé. Les agents gouvernent leurs propres règles de comportement au niveau local, s'adaptent à leur environnement, et au niveau macroscopique, l'ordre émerge. Dans le

web, il n'y a pas d'autorité globale qui gouverne la création des pages web. Les auteurs du web sont libres d'ajouter et d'effacer des pages et des sites web et de créer des hyperliens sur n'importe quelle page. Malgré ce procédé décentralisé, le web s'auto-organise automatiquement pour créer des communautés web. (Flake, Lawrence et al. 2002) ont défini une communauté web comme étant "une collection de pages tel que chaque page a plus d'hyperliens dans la communauté qu'à l'extérieur de la communauté". L'évolution de ces communautés web a été étudiée dans (Greco, Greco et al. 2004). Ils ont identifié des communautés web basées sur le modèle évolutionniste du graphe web. (Kumar, Raghavan et al. 1999) ont proposé une méthode pour l'extraction automatique de cyber communautés utilisant le graphe bipartite. Le graphe se divise en deux ensembles de pages web où le premier ensemble appelé les fans est considéré comme point spécialisé par le deuxième ensemble appelé autorités. Dans cette structure, les cyber-communautés qui partagent des intérêts communs émergent. (Chakrabarti, Joshi et al. 2002) intègrent l'analyse de la structure et du contenu pour soutenir les cyber-communautés.

2.1.2 Adaptation, coévolution et dynamiques

Au bord du chaos, où l'ordre commence à disparaître, les agents ont besoin de s'adapter à un environnement changeant. Ils changent leurs modèles et leurs comportements internes suivant leur organisation temporelle et spatiale. Ils co-évoluent pour assurer la survie dans le nouvel environnement. Depuis sa création, la structure, le contenu et l'usage du web ont co-évolué tout en s'adaptant l'un à l'autre. Comme le nombre d'utilisateurs augmente constamment, la personnalisation et l'adaptation des contenus des sites web a émergé, et les créateurs des sites web doivent adapter leurs sites aux différents usages et fournir un meilleur contenu (Perkowitz and Etzioni 2000; Menczer 2003).

Les systèmes complexes adaptatifs changent constamment à cause des interactions continues entre les agents et leur environnement. Les dynamiques sont non linéaires rendant le comportement global du système difficilement prédictible ou carrément imprédictible.

Le web est un graphe dynamique qui évolue en ajoutant constamment de nouvelles pages et en modifiant les pages existantes (le modèle de croissance). La probabilité de relier de

nouvelles pages à une page populaire est plus grande (le modèle d'attachement préférentiel) ou "le riche devient encore plus riche". Un bon exemple de cette situation est la page web de Google (Barabasi and Bonabeau 2003). Les dynamiques du web, son âge, sa durée de vie ont été étudiés dans (Brewington and Cybenko 2000; Cho and Garcia-Molina 2000) où le taux de changement d'à peu près 7 millions de pages web a été observé entre 1999 et 2000. Les expériences ont montré que chaque année, l'intervalle moyen de modification d'une page web est de 4 mois et que 50% de toutes les pages web changent chaque 50 jours. Ces études ont aussi montré que les changements des pages du web peuvent être modélisés mathématiquement par un processus de Poisson.

2.1.3 Les propriétés de Holland

Comme les systèmes complexes adaptatifs sont formés d'un ensemble d'agents interagissant les uns avec les autres, s'adaptant et coévoluant dans leur environnement, modéliser de tels systèmes requiert une approche ascendante qui consiste à identifier les différents agents, leurs règles de comportements et leurs interactions. John Holland a identifié sept éléments fondamentaux ou éléments de base d'un système complexe adaptatif (Holland 1995):

L'agrégation est la propriété par laquelle les agents se groupent pour former des catégories ou des méta-agents qui peuvent à leur tour se combiner à un plus haut niveau (méta-méta-agents) ce qui aboutit à un système complexe. L'émergence du méta-agent est due aux actions réciproques entre les agents du niveau inférieur. Le contenu et la structure basés sur les besoins des utilisateurs sont groupés dans une page web et des pages web dans des sites web et des sites web dans les communautés du web (les méta-agents) qui émergent et s'auto-organisent sans contrôle centralisé. Cette auto-organisation est un résultat d'une interaction rétroactive entre l'usage, le contenu et la structure.

Le tagging ou l'étiquetage est le mécanisme qui facilite la formation des groupes en allouant des attributs ou étiquettes utilisés pour l'identification des agents. Une étiquette pourrait être le sujet principal d'une communauté du web ou le "vecteur de mots clefs" d'une page web spécifique utilisée dans l'analyse de texte et l'analyse de similarité des pages web.

La non linéarité est la propriété où le comportement émergent du système est le résultat d'une réponse non proportionnelle à son stimulus. Le comportement global provenant des actions réciproques entre les agents est plus compliqué qu'une simple somme ou moyenne des agents individuels. Ainsi le système ne peut pas être prédit en comprenant simplement comment chaque composante fonctionne et se comporte. La croissance du web est un processus non linéaire.

Les flux sont les ressources physiques ou les informations qui circulent à travers les nœuds d'un réseau complexe. Dans le cas du web, ce sont les informations, les requêtes, les navigations, qui traversent les sites web, considérés comme des nœuds du réseau complexe qu'est le web.

La diversité – la diversité des expériences, stratégies, et règles des différents agents garantit le comportement adaptable dynamique d'un système complexe adaptatif. Le web a un grand nombre de composants réagissant réciproquement et cette diversité dans le web contribue à sa robustesse. Nous observons la diversité dans son usage, structure et contenu. Dans (Liu, Zhang et al. 2002), les utilisateurs ont été classés en trois catégories : les utilisateurs aléatoires, les utilisateurs rationnels et les utilisateurs périodiques.

Les modèles internes ou schémas sont les fonctions qui gouvernent l'emploi des agents pour communiquer les uns avec les autres et avec leur environnement. Ces schémas dirigent les comportements des agents.

Les blocs de base sont les parties composantes qui peuvent être combinées et réutilisées pour chaque cas d'un modèle. L'identification de ces blocs constitue un premier pas vers la modélisation d'un système complexe adaptatif. (Milo, Shen-Orr et al. 2002) ont montré que les motifs des sous-graphes forment la structure des blocs de base pour le réseau WWW. Les services web sont aussi des blocs de base pour développer des applications distribuées sur le web (Kirtland 2000).

2.1.4 Le web comme réseau complexe

2.1.4.1 La topologie du graphe du web

En analysant le web comme système complexe, nous devons examiner le modèle de son graphe. Le web a été représenté par un graphe orienté $G = \{N, E\}$ où N est l'ensemble des sommets ou pages web et E est l'ensemble des arcs ou hyperliens entre les pages web (Barabasi, Albert et al. 2000), (Kumar, Raghavan et al. 2000) . L'étude de la structure du graphe du web a plusieurs objectifs comme le développement de meilleurs algorithmes pour les moteurs de recherche, la découverte des communautés du web émergentes, l'amélioration de la classification des pages du web, etc.

2.1.4.2 La propriété des réseaux invariants d'échelle ou « scale free »

Sur le web, l'émergence est observée à plusieurs niveaux. La probabilité qu'une page web ait k pages entrantes (pointant sur elle : indegree) ou qu'une page web pointe sur k pages sortantes (outdegree), suit la loi de puissance $P(k) = k^{-\lambda}$. Le graphe du web n'est pas un réseau aléatoire, mais il possède la propriété des réseaux invariants d'échelle (scale free networks), qui est la caractéristique type des réseaux résultant d'une activité auto-organisationnelle comme stipulé par (Barabasi and Bonabeau 2003).

Le web est un graphe dynamique et il évolue en ajoutant continuellement de nouvelles pages et en enlevant quelques vieilles (la croissance évolutive). Il y a une très haute probabilité d'établir des liens entre les nouvelles pages et plus une page est connectée, plus elle aura de chances d'être connectée dans l'avenir (lien préférentiel) ou "le riche devient plus riche" et un bon exemple est la page web de Google (Barabasi 2001). Cette topologie montre que l'ordre émerge sur le web malgré le contrôle décentralisé dans l'ajout de nouveaux réseaux et ceci est une caractéristique d'un système complexe.

Une autre considération sur le graphe du web, c'est son degré élevé d'agrégation (clustering). Le coefficient de clustering C est une mesure associée à un sommet i . Cette mesure représente le ratio entre le nombre d'arêtes E_i existant entre l'ensemble composé du

sommet i et ses K_i voisins immédiats, et le nombre total d'arêtes possibles entre ces sommets, si le sommet i et ses K_i voisins sont complètement connectés les uns aux autres. Ce ratio est exprimé par:

$$C_i = 2 * E_i / (K_i * (K_i - 1))$$

Si deux sommets ou leurs voisins sont connectés au même sommet, il y a une forte probabilité que ces deux sommets soient connectés l'un à l'autre. Si deux sommets sont déjà connectés à un même sommet, ils ont un contenu en commun qui peut intéresser l'utilisateur et, conséquemment, ils seront connectés ensemble. Chaque page ou sommet lui-même évolue avec le temps basé sur sa "justesse". Une page mise à jour régulièrement avec un bon contenu, une interface conviviale et un bon marketing serait plus adéquate que d'autres pages et, comme résultat, plus de pages se connecteront à elle. Les pages rivalisent entre elles sur une base d'efficacité. Ce modèle est semblable à la condensation de Bose- Einstein (Barabasi 2001).

Cette topologie montre que l'ordre émerge sur le web malgré le contrôle décentralisé dans l'ajout de nouvelles pages. Le réseau s'est révélé robuste dans le sens où des suppressions aléatoires n'affectent pas le diamètre ou la connectivité du graphe, mais le système devient vulnérable à des attaques ciblées qui suppriment les nœuds fortement connectés.

2.1.4.3 Le phénomène « small world »

L'expérience de Milgram en 1967 a montré ce qui est connu par les 6 degrés de séparation ou que la longueur de la chaîne de connaissance entre deux Américains est de 6 personnes (Milgram 1967). C'est ce qui est aussi connu par le phénomène « petit monde » ou « small world ». Si deux américains sont choisis aléatoirement, ils sont reliés par 6 connaissances en moyenne. Des techniques d'analyse des réseaux sociaux ont été explorées dans (Wasserman and Faust 1994; Adamic, Buyukkokten et al. 2003) et elles ont trouvé que les réseaux sociaux en ligne exhibent le comportement « petit monde » avec un coefficient d'agrégation (clustering) (Mislove, Marcon et al. 2007).

Des mesures ont été faites sur le graphe du web, spécialement son diamètre ou la distance moyenne minimale entre 2 pages web choisies au hasard. (Albert, Jeong et al. 1999) montrent

que le diamètre suit la formule: $d = 0.35 + 2.06 * \log(N)$ où N est le nombre total de nœuds. Le phénomène “petit monde” est observé sur le réseau web. Watts (Watts 1999) a démontré que les réseaux petit monde sont, en quelque sorte, entre les réseaux ordonnés et les réseaux aléatoires.. Le réseau a démontré qu’il est robuste, des effacements au hasard de quelques sommets n’affectent pas le diamètre ou la connectivité du graphe, mais le système devient vulnérable avec des attaques ciblées qui effacent les pages fortement connectées (Albert, Jeong et al. 2000).

2.1.4.4 La structure “bowtie” et l’autosimilarité du web

Le graphe du web était représenté par un modèle “bowtie” (Broder, Kumar et al. 2000) (fig.2.1) avec des Composantes Fortement Connectées formant 27.5% du nombre total de pages, « IN » 21.5%, « OUT » 21.5%, Vrilles et tubes 21.5% et déconnectées 8%. N’importe quel nœud de “IN” peut passer au travers SCC vers n’importe quel sommet de “OUT”. Les quatre parties sont d’un format presque équivalent.

Le web exhibe aussi un comportement “fractal ou auto-similaire. Une petite portion du web a les mêmes caractéristiques que le web tout entier. Les différents sous-graphes peuvent partager le même sujet, peuvent être hébergées sur le même site web, ou peuvent partager le même emplacement géographique, et les pages web qui partagent le même contenu sont fortement reliées entre elles. Ces sous-graphes et les distributions “indegree” et “outdegree” sont constants avec la distribution de tout le graphe du web et ont une structure “bowtie”.

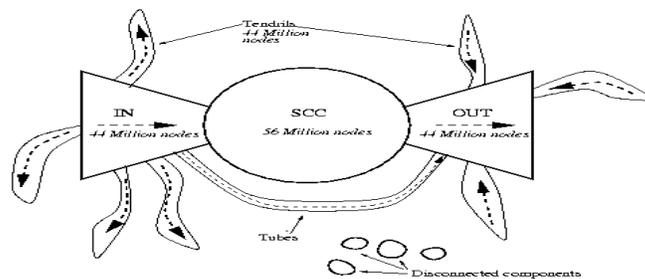


Figure 2.1 La structure “bowtie” du graphe web (Broder, Kumar et al. 1999)

2.2 Les systèmes multi-agents

Cette vision du web selon la perspective des systèmes complexes adaptatifs nous amène à considérer le système comme un ensemble d'agents situés et incarnés évoluant dans un environnement dynamique. Les systèmes multi-agents trouvent alors leur justification dans cette étude du web.

“Un système multi-agent est un ensemble d'entités (physiques ou virtuelles) appelées agents, partageant un environnement commun (physique ou virtuel), qu'ils sont capables de percevoir et sur lesquels ils peuvent agir. Ces perceptions permettent aux agents d'acquérir de l'information sur l'évolution de leur environnement, et leurs actions leur permettent de modifier l'environnement. Les agents interagissent entre eux, directement ou indirectement, et montrent des comportements corrélés créant une synergie qui leur permet de créer une coopération organisée.” (Ferber 1995)

2.2.1 Les systèmes multi-agents situés et l'environnement

Les systèmes multi-agents situés ont émergé du domaine de la robotique mobile ou de la simulation où la prise en compte de l'environnement complexe et dynamique doit être considérée pour résoudre le problème de sélection d'actions des agents en fonction de l'état de leur environnement. Dans (Weyns, Omicini et al. 2007), l'environnement est défini comme étant une abstraction de première-classe qui fournit les conditions d'existence de l'agent et fournit une abstraction de conception exploitable pour la construction des systèmes multi-agents. Les systèmes multi-agents situés utilisent l'environnement pour la communication indirecte entre les différents agents à travers le mécanisme de stigmergie décrit dans la section suivante.

2.2.2 La stigmergie

La notion de stigmergie a été introduite par le biologiste français Grassé (Grassé 1959), pour comprendre le comportement de coordination des constructions chez les insectes sociaux et surtout l'interaction avec leur environnement. Cette interaction aboutit à un changement

dans l'environnement et ce changement affectera les actions futures des agents. Grassé a donné la définition suivante : "La coordination des tâches, la régulation des constructions ne dépendent pas directement des ouvriers, mais des constructions elles-mêmes. L'ouvrier ne dirige pas son travail, il est guidé par lui. C'est à cette stimulation d'un type particulier que nous donnons le nom de STIGMERGIE ." Les fourmis, dans une colonie, utilisent le mécanisme de stigmergie pour communiquer indirectement entre elles en déposant la phéromone pour aboutir aux sources de nourriture.

Plusieurs tentatives d'associer le web avec la notion de stigmergie ont été analysées récemment. Small (Small 2004) explore la possibilité d'utiliser des stratégies stigmergiques pour créer des sites web auto-organisés. Les utilisateurs interagissent avec le site, changeant sa structure et son contenu à l'aide d'agents logiciels. Small propose aussi un mécanisme stigmergique pour localiser les sources d'information sur le web, où les utilisateurs se comporteraient de façon similaire aux fourmis dans une colonie. (Gregorio 2003) a lié le concept de stigmergie au web. Les différents utilisateurs des weblogs (les agents) communiquent indirectement les uns avec les autres, laissant des traces dans leur environnement et éventuellement, le changeant. Quand l'un modifie l'environnement, l'autre répond au nouvel environnement et le modifie à son tour et ainsi de suite. L'émergence des cybercommunautés auto-organisées est due à l'interaction indirecte entre les créateurs des pages web et leur environnement (le web). Les pages partageant le même sujet d'intérêt pointent les unes vers les autres, créant des clusters.

D'autre part, nous percevons le web 2.0 et surtout les applications d'étiquetage collaboratif comme un environnement stigmergique où les millions d'utilisateurs communiquent indirectement entre eux en laissant des traces ou leurs étiquettes comme outil puissant pour que d'autres utilisateurs découvrent l'information recherchée de façon plus efficace.

2.3 Une vue d'ensemble des approches systèmes multi-agents pour le web

Pour le web, une variété de logiciels à base d'agents ont été développés dans les domaines de la Recherche d'Information, du Web Mining et du Web Semantique. La Technologie de

l'Agent d'Information a émergé comme un nouveau champ de recherche pour solutionner le problème de la surcharge d'information disponible sur l'Internet (Klusch 2001).

“La tâche principale des agents d'information est d'exécuter des recherches pro-actives pour maintenir et médiatiser l'information pertinente pour leurs utilisateurs ou pour les autres agents. Ceci inclut des capacités telles que rapporter, analyser, manipuler, fusionner les informations hétérogènes aussi bien que d'en avoir une vue d'ensemble et guider l'utilisateur à travers l'espace disponible de l'information individuelle.”

Quelques-uns des rôles joués par l'agent du web:

- Acquisition et exploitation de l'information.
- Synthèse et présentation de l'information.
- Fournir une assistance intelligente à l'utilisateur.
- Étudier les profils des utilisateurs.
- Rechercher proactivement l'information.
- Maintenir et médiatiser l'information pour les utilisateurs.
- Extraire, analyser, filtrer et fusionner les informations hétérogènes.
- Visualiser et guider les utilisateurs à travers l'espace de l'information disponible.
- S'adapter d'une façon dynamique aux changements dans les préférences des utilisateurs, aux changements dans l'environnement et aux changements chez les autres agents.
- Suggérer des recommandations appropriées.

2.3.1 Systèmes multi-agents pour la Recherche d'Information sur le Web

La Technologie des Agents a été utilisée intensivement pour essayer de résoudre le problème de Recherche de l'Information (RI) sur le web. (Candea, Staicu et al. 2000) ont identifié quatre types de systèmes de RI à base d'agents: les systèmes “knowbot”, les systèmes adaptatifs, les systèmes collaboratifs et les systèmes proactifs. Parmi les premiers travaux, on

peut citer Letizia, un agent qui assiste l'utilisateur naviguant sur le web (Lieberman 1995). Letizia observe le comportement de l'utilisateur alors qu'il navigue sur le web et lui suggère des liens d'intérêt. Quant à Amalthea (Moukas 1997), c'est un écosystème multi-agent formé de deux groupes d'agents : des Agents de Filtrage de l'Information (AFI) qui s'adaptent aux besoins des utilisateurs en apprenant des feed-back des utilisateurs et des Agents de Découverte de l'Information qui extraient de l'information de plusieurs moteurs de recherche basés sur des demandes des AFIs. Des approches évolutionnaires, plus spécialement des algorithmes génétiques, sont utilisées pour évaluer la « justesse » des agents de façon à survivre, à se reproduire ou à mourir. L'AFI utilise une technique de recherche d'information (RI) standard: le « weighted keyword vector » pour la représentation des documents. Des expériences avec des utilisateurs virtuels ont été faites de façon à obtenir un équilibre dans l'écosystème et de faire des textes sur les utilisateurs réels pour évaluer la performance du système en trouvant des informations pertinentes. Une approche d'apprentissage multi-agent a été proposée pour récupérer des informations sur le web en utilisant un réseau de neurones (Choi and Yoo 1999). Les agents apprennent les intérêts des utilisateurs en utilisant les réseaux de neurones artificiels et en localisant les ressources appropriées de l'information et les outils de recherche pour une récupération efficace et pertinente de l'information désirée. (Müller 1999) a proposé une architecture multi-agent pour une recherche intelligente sur le web, basée sur la modélisation de l'utilisateur. CiteSeer, la “bibliothèque numérique de littérature scientifique” est basée sur une architecture multi-agent, où les agents acquièrent, font l'analyse grammaticale et naviguent à travers des documents sur le web et assistent les utilisateurs pour trouver des publications de recherches pertinentes. Un système multi-agent pour une recherche multi-cibles sur le web a été développée par (Nitto, Ghezzi et al. 2003). Une recherche multi-cibles permet aux utilisateurs de définir leurs requêtes avec des objectifs interdépendants multiples. Chaque agent qui cherche, va exécuter une sous-question correspondant à chaque objectif et les différents résultats seront combinés et offerts à l'utilisateur. Pour rechercher de l'information d'une façon sémantique sur l'Internet, (Cesarano, d'Acerno et al. 2003) ont proposé un système aussi basé sur le paradigme d'agents intelligents. Les pages Web sont catégorisées et une similarité sémantique est utilisée entre les pages rapportées sur une Base de Connaissance Sémantique. Des techniques d'apprentissage de

« machine learning » furent employées pour bâtir des agents d'auto-adaptation pour rapporter et extraire l'information (Eliassi-Rad and Shavlik 2003). Ces agents d'apprentissage sont bâtis en se basant sur les préférences des instructions des utilisateurs, en utilisant la théorie du raffinement dans les « learning machine ». Dans le secteur médical, où les médecins ont besoin d'avoir accès aux informations exactes les plus récentes sur le web, MIRRORS a été développé comme système multi-agent pour récupérer de l'information médicale qui concerne les questions intemporelles, de pertinence, ontologiques et confidentielles de l'information médicale. ARCH est un agent d'adaptation qui raffine la formulation des questions des utilisateurs pour des résultats de recherche améliorés (Parent, Mobasher. et al. 2001).

2.3.2 Le Web « Mining » et les systèmes multi-agents

L'approche à base d'agents pour le Web Mining a été classée en trois catégories (Cooley, Srivastava et al. 1997): des agents de recherche intelligente, des agents de filtrage/catégorisation de l'information et des agents personnalisés du web. (Chau, Zeng et al. 2003) ont utilisé une Araignée Collaboratrice, une approche d'un système multi-agent pour un web « mining » du contenu collaboratif. Les utilisateurs partageaient des sessions de recherche afin de bénéficier des expériences de recherche antérieures. Le problème d'adaptabilité et de passage à l'échelle (scalability) du web a été abordé avec MySpiders (Pant and Menczer 2002), où un ensemble d'agents/ « crawlers » naviguent sur les pages web au moyen d'hyperliens pour trouver l'information la plus pertinente pour répondre aux questions de l'utilisateur. Ils s'adaptent au profil de l'utilisateur et complètent le travail de l'agent de recherche. Les agents d'adaptation apprennent par les réseaux de neurones, pour évaluer la pertinence des liens sortants. Pour évaluer les « crawlers », trois mesures métriques ont été utilisées. Une approximation des mesures de rappel et de précision et une nouvelle métrique a été introduite pour évaluer la nouveauté des documents récupérés.

(Liu, Zhang et al. 2002) ont utilisé l'approche multi-agent pour étudier les régularités du comportement navigateur de l'utilisateur sur le web. Les utilisateurs, dans leur modèle, étaient les agents « fouilleurs d'information » (“information-foraging”) et leurs expériences ont montré que la fréquence de distribution des liens suit la loi de puissance. ROSA est un

système multi-agent personnalisé où les agents ont différentes tâches incluant le contenu et l'usage « mining » (Kazienko and Kiewra 2003). (Pazzani and Billsus 2002) ont développé des agents d'un site web d'adaptation qui assiste l'utilisateur navigant sur le web en faisant des recommandations sur les documents concernés. Seta2000 est un système de recommandations pour la personnalisation, basé sur l'architecture multi-agent, spécialement pour les agents de communication et de design (Ardissono, Goy et al. 2005). PII (Personal Internet Interface) est un ensemble de cinq logiciels différents, chacun ayant son propre rôle et collaborant ensemble pour aider les utilisateurs à repérer l'information pertinente. Les agents surveillent les habitudes de navigation de l'utilisateur, maintiennent l'information personnalisée de l'utilisateur, naviguent sur le web pour trouver des pages potentielles, classent les pages et suggèrent des recommandations à l'utilisateur (Byer and Depradine 2004). DIAMS est un système d'agents, fait en collaboration, qui aide l'utilisateur à recueillir de l'information et à l'organiser (Chen, Kagal et al. 2001).

2.3.3 Le Web Sémantique et les systèmes multi-agents

Les agents jouent un rôle majeur dans l'interprétation des sémantiques associées avec les données pour répondre aux questions, solutionner les problèmes et échanger de l'information avec d'autres agents par la communication et la coordination. Ils peuvent récupérer l'information, la manipuler, raisonner à son sujet, et fournir des services pour les utilisateurs. La recherche dans le domaine du système multi-agent intégré sur le web sémantique est encore à ses débuts, alors que la forme du web sémantique a encore à se définir. Van.Harmelen a défini trois secteurs dans lesquels la recherche des agents pourrait contribuer dans le web sémantique : les services web, la technologie de pair à pair et la combinaison d'ontologie (Van Harmelen 2004). Une plate-forme agent pour bâtir des agents intelligents pour le web sémantique a été proposée, basée sur les principes BDI (belief, desire, intention) pour des agents avec un raisonnement délibérant et pratique (Dickinson and Wooldridge 2003). SERSE est présenté comme un système de recherche pour le web sémantique, basé sur le paradigme du système multi-agent combiné avec les technologies ontologiques et de pair à pair. Les agents communiquent entre eux via le réseau P2P et obtiennent la connaissance des définitions ontologiques annotées avec les données (Tamma, Blacoe et al. 2004). Des problèmes de

sécurité, d'intimité et de confiance entre les agents ont été discutés dans (Finn and Joshi 2002). Comme le web sémantique est plus ouvert et dynamique pour son exploitation par les agents et pour la prise de décisions pour le compte des usagers, les langages sémantiques devraient être bien équipés avec une « politique de langage » pour définir les exigences de sécurité et de gestion en confiance. Plusieurs agents sémantiques de services du web sont proposés, comme RETSINA-RCAL (Payne, Singh et al. 2002), un agent calendrier qui assiste les utilisateurs à planifier intelligemment leurs rencontres, ITTALKS avise automatiquement les utilisateurs de conversations sur la technologie de l'information, EasyMeeting fournit des services dans un environnement d'une invitante salle de réunion, GraniteNights est un système multi-agent qui assiste l'utilisateur pour planifier une sortie en soirée dans la ville d'Aberdeen. Mais il est à mentionner qu'il existe plusieurs problèmes liés aux agents et le web sémantique qui n'ont toujours pas de solution, tels que: le partage de la connaissance, l'unification syntactique, la découverte des capacités des agents, la coordination entre agents, et les protocoles d'interaction.

2.4 Synthèse et discussion

Ce chapitre a présenté le web selon la perspective des systèmes complexes adaptatifs, en montrant que le web exhibe les propriétés et caractéristiques de ces systèmes. Il est à noter que le web, en tant que système, est façonné par les actions des utilisateurs. Les propriétés émergentes dans l'évolution du web et son caractère auto-organisationnel sont dus en grande partie à l'usage, qui, à son tour est largement affecté par cette évolution. Nous aboutissons à un phénomène auto-catalytique. Ceci peut être exprimé par le fait que les utilisateurs (que nous pouvons considérer comme agents humains) laissent des traces sur les pages web ou nœuds du graphe. C'est le phénomène de stigmergie.

Ces traces s'intensifient par l'usage donnant lieu aux nœuds fortement connectés observés dans la structure « bowtie » du graphe du web, et cela explique aussi la topologie « scale free » de ce graphe. L'attachement préférentiel exprimé par le fait que plus une page est visitée, plus elle a tendance à l'être dans l'avenir, peut être attribué au phénomène d'exploitation décrit par Holland (Holland 1992) pour les systèmes complexes adaptatifs.

Pour les nœuds faiblement connectés observés dans la structure « bowtie », ils seront accédés éventuellement avec l'évolution des usages et ceci est possible à cause de la propriété « small world » et ceci peut être attribué au phénomène d'exploration décrit par Holland.

Vu la nature du web comme réseau complexe, son évolution et son caractère auto-organisationnel, il est donc nécessaire de prendre en considération ses caractéristiques en développant ou modélisant les systèmes qui ont le web comme environnement d'évolution. Ceci en prenant en compte le rôle important que l'usage joue sur cet environnement et ce qui peut être exprimé par le mécanisme de stigmergie.

Ce chapitre a présenté aussi les différentes approches systèmes multi-agents utilisées pour le développement d'applications sur le web, pour la recherche d'informations et pour le web sémantique. Ces approches présentées illustrent la diversité d'utilisation des systèmes multi-agents dans le contexte du web. Il existe beaucoup plus d'applications ayant le web comme environnement et qui utilisent le paradigme multi-agent, mais le but de notre recherche n'est pas de présenter toutes ces applications. Comme on peut le constater, ces applications sont construites de manière ad hoc, ne permettant pas l'émergence de nouvelles propriétés façonnées par l'usage.

D'autre part, la majorité des approches organisationnelles pour les systèmes multi-agents, décrites dans le chapitre suivant, se concentrent sur l'organisation sociale et sur les rôles des agents, occultant l'influence et le rôle de l'environnement dans l'évolution de ces systèmes. En développant des systèmes qui ont les propriétés des systèmes complexes adaptatifs, comme on a pu le voir à travers l'analyse du graphe du web, les comportements des usagers et la structure du graphe du web sont fortement liés. Il semblerait alors primordial de considérer l'environnement et de le matérialiser physiquement par une représentation spatiale, corrélée à la représentation sociale des agents évoluant dans cet environnement.

Nous proposons dans le cadre de cette thèse une modélisation permettant de coupler l'organisation sociale et l'organisation spatiale au sein d'un système multi-agent, et nous l'appliquons au contexte des applications collectives sur le web.

Chapitre 3

Cadre pour les organisations sociales et spatiales dans les systèmes multi-agents

Dans les approches multi-agents proposées jusque là pour le domaine du web, à notre connaissance, la prise en compte du caractère complexe au sens de Holland (comme présenté au chapitre 2 de cette thèse), n'est pas mise en évidence (voir les différentes citations dans la section 2.3). Les modèles multi-agents proposés sont généralement ad hoc et sont destinés à apporter des réponses à des fonctionnalités précisées. Dans notre travail, nous nous distinguons de ces travaux, par la proposition d'un cadre d'intégrer les caractéristiques de Holland, comme proposé dans (Hassas 2003) et que nous instancions par la mise en évidence de la relation de couplage entre les deux organisations sociale et spatiale du système multi-agents à déployer sur le web, et leurs co-évolutions rétroactives. En tentative de réponse à la question problématique suivante posée par (Hassas 2003) :

« Comment développer des systèmes informatiques capables de travailler ”en intelligence” avec leur environnement considéré dans son sens le plus large, intégrant tant le niveau physique matériel et logiciel, que le niveau conceptuel de l’usage ? »

Nous avons proposé un modèle qui traite l’environnement comme une entité de première classe où le système est en couplage avec son environnement. En adoptant le paradigme multi-agents comme modèle de représentation, cela s’exprime par le couplage existant entre l’organisation sociale et l’organisation spatiale du collectif.

3.1 Les organisations dans les systèmes multi-agents

La notion d’organisation dans les systèmes multi-agents a débuté dans les années 80 avec les travaux de (Corkill et Lesser 1983) où ils ont exploré l’auto-conception organisationnelle et la coordination dans les systèmes multi-agents. Ces travaux ont continué (Horling and Lesser 2004) où l’organisation et les concepts reliés ont été appliqués au modèle multi-agents. Dans les travaux sur les systèmes multi-agents centrés sur l’organisation, les aspects sociaux de l’interaction entre les agents sont mis en évidence, en prenant comme exemple les organisations sociales humaines. Ferber a proposé le modèle AGR (Ferber, Gutknecht et al. 2003) où les agents, groupes, et rôles forment les concepts primitifs. Les agents jouent des rôles dans des groupes. (Odell, Nodine et al. 2005) ont utilisé UML pour proposer un méta modèle pour les agents, leurs groupes et leurs rôles basé sur les activités courantes des agents. Le cadre OMNI (Vazquez-Salceda, Dignum et al. 2005) a pris en considération les trois dimensions d’une organisation: normes et règles, relations structurales et relations contextuelles. ORA4MAS est un modèle d’infrastructure organisationnelle qui est basé sur les agents organisationnels et les artefacts organisationnels (Kitio, Boissier et al. 2008). Inspirés des organisations humaines, les agents sont munis d’outils et d’artefacts qui supportent leurs activités au sein de leur organisation. Dans l’extension du travail de Ferber (Ferber, Michel et al. 2005), les similarités entre les organisations d’aspect physique et social dans l’environnement ont été exploré où les agents existent dans des domaines appelés espace et sont capables d’entreprendre des actions dans ces domaines à travers des structures appelées modes. Les espaces géométriques correspondent au monde physique tandis que les espaces

sociaux correspondent à l'organisation sociale. MaSE (Multi-agent Systems Engineering) est une méthodologie qui inclut des concepts d'organisation tels que les buts, les rôles, et l'attribution des rôles aux agents (DeLoach 2005). D'autres concepts organisationnels ont été introduits dans les travaux sur les institutions électroniques où l'interaction sociale est structurée (Esteva, Rosell et al. 2004).

Plus récemment, MASQ (Multi-Agents Systems based on Quadrants) a été introduit comme un modèle qui intègre les 4 éléments essentiels nécessaires pour la description du processus d'interaction dans les systèmes multi-agents (Stratulat, Ferber et al. 2009). Ces éléments sont agents, environnement, organisation et institution. En rapport avec notre travail, nous nous intéressons à la notion d'espaces bruts introduite dans MASQ où les agents sont représentés par leurs corps dans ces espaces. Ces espaces sont de deux types distincts: physique ou social.

Notre travail complète ces notions en s'intéressant au **couplage** existant entre l'organisation sociale et l'organisation spatiale du système. L'environnement est essentiel dans la définition du système et de son évolution, et par conséquent l'organisation doit s'appuyer la dessus, c'est à dire elle doit prendre en compte la représentation ou la matérialisation physique de l'environnement. Ce couplage entre le système et son environnement peut être représenté au niveau du système multi-agent par la mise en œuvre des agents situés selon leur organisation spatiale dans cet environnement matérialisé physiquement. Il faut donc définir d'une part, le déploiement du système multi-agent dans l'environnement, ce qui permettra d'atteindre son organisation spatiale et d'autre part son organisation sociale qui permettra de mettre en œuvre les interactions lui permettant de maintenir son organisation (Hassas 2003). En bref, le point de départ est le couplage du système avec son environnement. Ce couplage se décline en couplage physique/conceptuel qui se décline à son tour en couplage structurel/comportemental et enfin en termes des systèmes multi-agents en couplage d'organisations spatiales/sociales.

3.2 Modèle proposé

Dans le modèle que nous proposons, comme défini dans nos travaux (Rupert, Hassas et al. 2006; Rupert, Rattrout et al. 2008), le système informatique est considéré comme un

écosystème artificiel, dans lequel des agents partagent un environnement commun. Les agents co-évoluent et interagissent les uns avec les autres via l'environnement partagé. Nous associons à l'environnement deux niveaux : un niveau physique représenté par l'organisation spatiale des agents et un niveau conceptuel représenté par l'organisation sociale.

Le niveau physique est représenté par un réseau ou graphe. Les agents sont situés sur les différents sommets du graphe appelés places. Ces places forment les positions organisationnelles que les agents peuvent occuper au niveau physique de l'environnement. Les arcs du graphe représentent les chemins entre les différentes places. Un ensemble de sommets ou places forme une région. Les régions seront définies en se basant sur une relation de voisinage entre les sommets dépendamment de la topologie du graphe. Les régions forment les unités organisationnelles de l'organisation spatiale.

Les agents peuvent percevoir les autres agents situés dans la même région et peuvent propager des signaux dans leur voisinage (à travers leurs chemins par exemple). Comme la topologie du réseau est fortement dynamique, les régions sont aussi dynamiques et changent au cours du temps. La co-évolution de l'organisation sociale et l'organisation spatiale du système est le fruit et le résultat de sa dynamique, comme dans un système « autopoétique ».

La communication indirecte entre les agents et leur coordination sont obtenues par l'utilisation du mécanisme de stigmergie et plus particulièrement par la diffusion, propagation et évaporation d'une phéromone numérique spécifique. La phéromone numérique est considérée comme la structure spatiale pour le codage de l'information de contrôle et de méta-contrôle. Cette phéromone représente de l'information selon l'application considérée. Dans le cas de la simulation des communautés virtuelles, cette phéromone a été exprimée par un type ou une étiquette qui consiste en différents sujets ou sujets discutés dans la communauté. Dans le cas des systèmes d'étiquetage collaboratifs, cette phéromone représente les étiquettes ajoutées par les utilisateurs pour des ressources données. Ces étiquettes représentent les traces laissées par les utilisateurs pour communiquer indirectement les uns avec les autres.

3.2.1 Organisation sociale

L'organisation sociale dans un système multi-agents consiste des actions suivantes :

- Diviser la société multi-agents en unités sociales ou groupes.
- Définir les rôles que les agents peuvent jouer dans chaque groupe (ou les comportements des agents dans le groupe) et les différentes relations entre ces rôles.
- Affecter les différents agents aux différents groupes

Les rôles peuvent être formés et affectés aux agents par une auto-organisation émergente dans le système, en fonction de l'état de l'environnement et de la corrélation avec l'organisation spatiale des agents. Les exemples dans les systèmes naturels incluent les guêpes Poliste qui peuvent prendre les rôles d'infirmiers, chefs, ou « fourageurs » en fonction de l'état de l'environnement (ressources alimentaires, dimension de la progéniture), et les fourmis peuvent jouer spontanément différents rôles (explorateurs vs. déposants de phéromones). Nous avons utilisé l'organisation sociale comme définie dans (Ferber, Gutknecht et al. 2003) :

Agents : Les agents sont les entités actives et mobiles qui interagissent dans le système. Ils possèdent une position spatiale et jouent au moins un rôle. Ils contrôlent leur propre exécution et peuvent changer de positions.

Rôles: Un rôle définit un ensemble de comportements qu'un agent jouant ce rôle peut accomplir. Un agent commence par un ensemble de rôles attribués et il peut dynamiquement changer de rôles au cours de l'évolution du système.

Groupes: Les groupes forment les unités organisationnelles de l'organisation sociale. Un groupe est un ensemble d'agents affectés à des rôles reliés selon un but commun ou un pattern d'interactions.

Structure des groupes: La structure des groupes décrit l'ensemble de rôles qui sont membres d'un groupe. Chaque groupe est décrit selon une structure de groupe.

Attribution de rôles: L'attribution des rôles relie un agent à un groupe. Tandis que les rôles peuvent exister dans les groupes comme part de la structure des groupes, un rôle peut n'être assigné à aucun agent. On réfère aux rôles joués par certains agents par rôles assignés.

Actions: Les actions sont les ensembles de fonctions que les agents peuvent réaliser comme partie de leurs comportements définis par les rôles.

Contrainte de multiplicité : La contrainte de multiplicité exprime le nombre maximal d'agents qui peuvent jouer un rôle particulier dans un groupe.

Contrainte de dépendance : La contrainte de dépendance décrit comment l'attribution potentielle d'un rôle dépend de l'attribution d'un rôle différent.

Contrainte de correspondance : La contrainte de correspondance décrit comment un agent jouant un rôle donné va automatiquement jouer un rôle correspondant.

Ces contraintes sont traitées comme des méta-contraintes qui aident à décrire et structurer l'organisation sociale. Elles définissent les règles structurales gouvernant l'organisation sociale. Les contraintes de multiplicité limitent le nombre d'agents qui peuvent être assignés à jouer un rôle particulier. Chaque contrainte de multiplicité correspond à une structure de groupe et à un rôle, et aide à définir les rôles tels qu'un leader dans un groupe, où la multiplicité est importante pour la structure. Les contraintes de dépendance expriment la dépendance entre l'attribution d'un rôle potentiel et l'attribution d'un rôle actuel. Elles aident à imposer des politiques et des relations sociales sur l'organisation sociale des agents. Pour aider à renforcer ces contraintes aussi bien que pour changer la structure de l'organisation sociale, des agents de groupes sont assignés à chaque groupe. Les agents de groupe sont responsables du changement de structure des rôles qui appartiennent au groupe et des structures de groupes qu'ils gèrent; ajouter et retirer des rôles des structures de groupe; mettre en place et changer les contraintes sociales (e.g. multiplicité, dépendance, correspondance); et approuver l'attribution des rôles en se basant sur la structure actuelle du groupe local.

3.2.2 Organisation spatiale

L'organisation spatiale comprend les éléments suivants (Fig. 3.2):

Places: Les places définissent les positions spatiales des agents. Comme le système évolue, les agents sont capables d'ajouter de nouvelles places et éliminer les anciennes places.

Régions: Les régions partitionnent l'organisation spatiale en unités distinctes. Une région est une collection de places qui sont spatialement proches les unes des autres. Une région forme un cluster de places basé sur une relation de voisinage. Les régions sont prédéfinies par le designer et continuent à émerger avec l'évolution du système.

Chemins: Les chemins relient les différentes places. Ils forment les arcs du graphe spatial qui relient les différents nœuds.

Les places et les régions forment une partie intégrale de l'organisation spatiale. Les places définissent les conditions sous lesquelles les agents existent. Un agent peut être situé sur une place à la fois, mais il peut appartenir à plusieurs régions quand les régions se superposent.

En plus de fournir une structure organisationnelle à l'environnement physique, les régions lui donnent de la flexibilité. Les régions sont gérées par des agents de régions (Brueckner 2000). Ces agents spécialisés fournissent des services locaux et des ressources à d'autres agents situés à chaque région. Lorsqu'un agent, à un endroit donné, requiert l'accès à des sources de données, la demande pour ce service fait son chemin et est fournie par l'agent de région de la place. Ceci procure de la flexibilité au système en permettant à différents agents de fournir différents services et ressources (Brueckner and Parunak 2002). En divisant l'environnement physique général en de plus petites unités auto-organisées, les places divisent l'ensemble des processus qui définissent et changent les états de l'environnement et distribuent la charge de l'agent dans l'environnement. Avec la division de l'environnement total en des unités plus petites, il n'est pas nécessaire d'avoir une seule entité vaste et complexe pour exécuter les processus de l'environnement. Ceci est particulièrement vrai pour les systèmes ayant un environnement physique largement diversifié et hétérogène, où une telle absence de partition englutit les possibilités de transformation de l'environnement et les paliers de performance et de croissance (Odell, Parunak et al. 2002). Grâce à cette division de

l'environnement physique et de ses constituantes hétérogènes, le comportement des agents dans le système peut varier selon leur position spatiale.

L'effet des interactions entre les agents est matérialisé dans l'environnement et peut être exprimé par l'utilisation d'une structure spatiale sous la forme d'une phéromone numérique contenant les champs suivants:

- un label : qui permet d'identifier la nature de l'information ;
- une intensité : qui exprime le degré de pertinence de l'information ;
- un taux de diffusion : qui exprime la portée d'une information, qui se traduit correctement sur le réseau par une longueur de chemin. ;
- un taux d'évaporation : qui exprime le taux de persistance d'une information. Plus une information est persistante, moins elle est amenée à disparaître facilement. Son taux d'évaporation est donc faible.

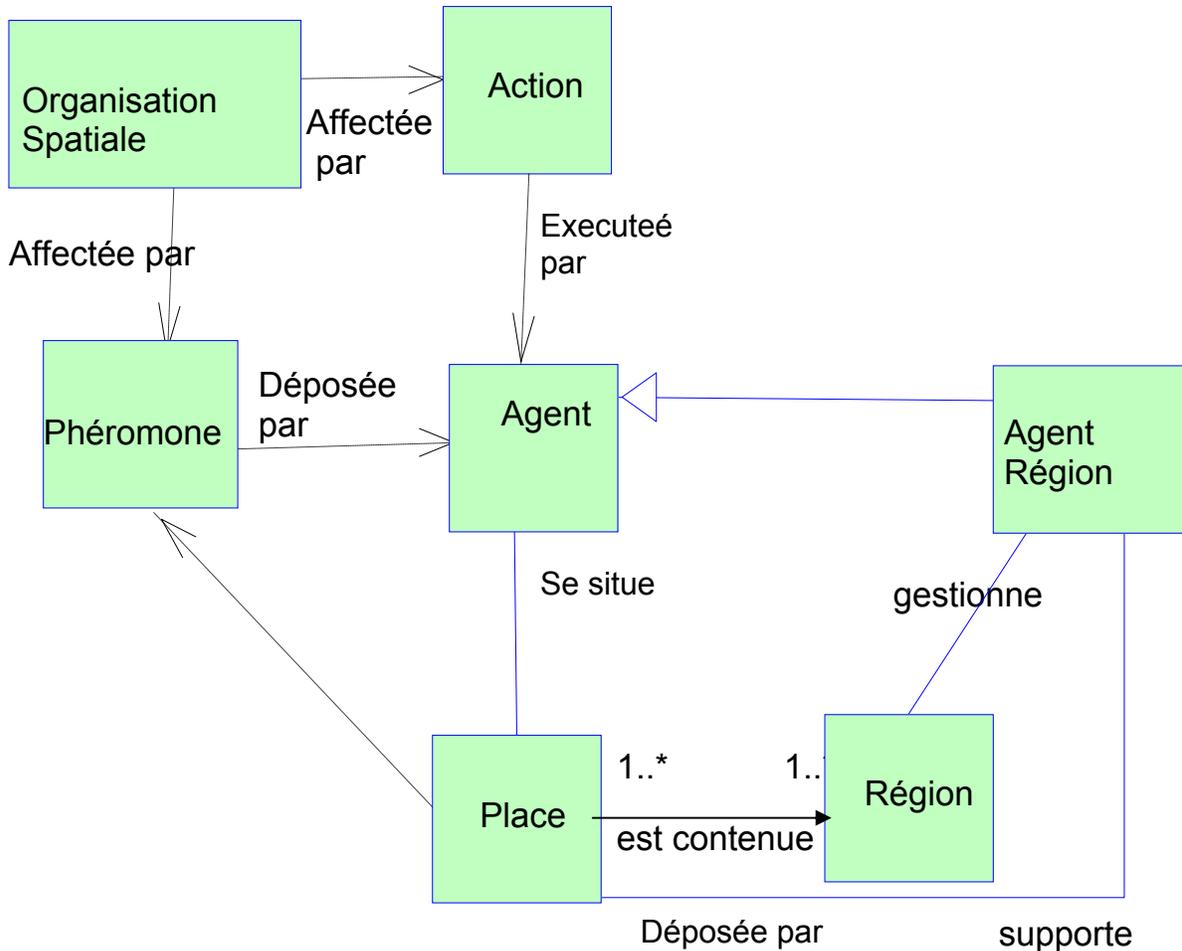


Figure 3.2 Organisation Spatiale

3.2.3 Co-évolution de l'organisation sociale et l'organisation spatiale

Le comportement d'un agent et les rôles qu'il joue sont fortement affectés par sa position dans l'organisation spatiale. Sa position spatiale est aussi affectée par les rôles qu'il joue dans l'organisation sociale. Le couplage entre l'organisation spatiale et l'organisation sociale est rétroactif et est exprimé dans la topologie du graphe. Cette coévolution est obtenue par le mécanisme de stigmergie, qui permet l'auto-structuration de l'environnement à travers les activités des agents et la diffusion de phéromone (Fig. 3.3).

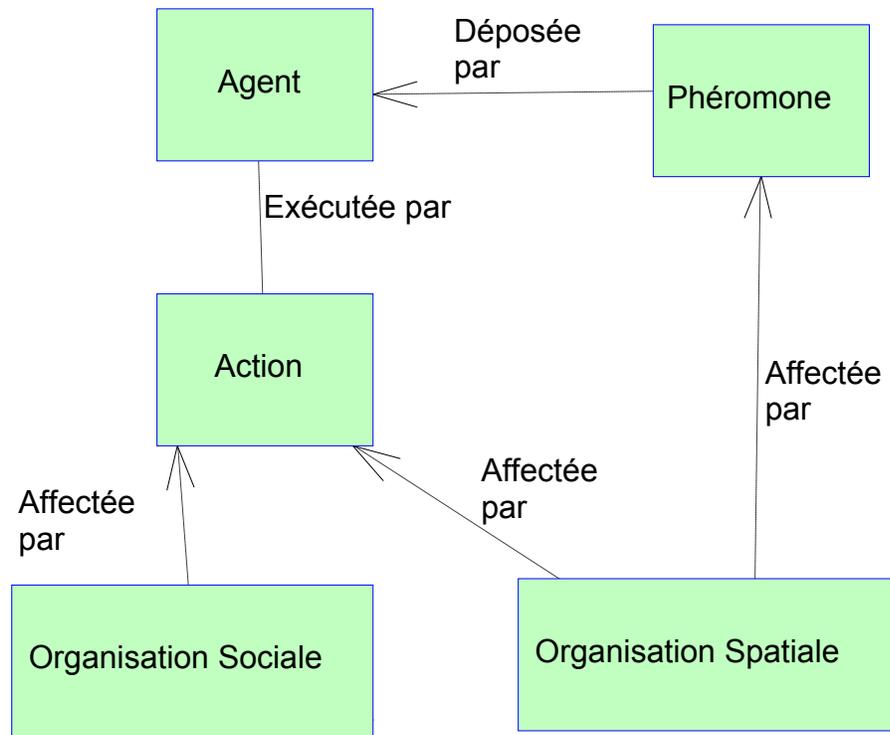


Figure 3.3 Coévolution des organisations spatiales et sociales.

Considérons quelques exemples de systèmes réels pour argumenter cette idée. On peut spécifier une organisation en terme de rôles et de relations entre rôles, qui ne sont pas définis a priori, mais qui sont identifiés et réifiés par rapport à une structure organisationnelle qui émerge, ainsi qu'à une structure spatiale. Dans le cas du comportement de fourragement chez les fourmis, nous avons les déposeurs de phéromones et les suiveurs de phéromones, et aussi les exploreurs pendant l'activité de fourragement. Au départ tout le monde est exploreur. Avec l'évolution du temps, et selon leurs positions spatiales (à proximité de nourriture, à proximité de phéromone déposée par d'autres fourmis), leurs rôles évoluent pour devenir des déposeurs ou des suiveurs. Un autre exemple est le comportement des communautés sur le web. Au départ tout le monde est navigateur, ensuite il en émerge des membres d'une communauté, créés par la proximité de leur centre d'intérêt exprimé dans leurs pages web (qui forment un

cluster). Quand un navigateur prend conscience qu'il fait partie d'une communauté ou prend l'initiative de rentrer dans une communauté, son rôle devient de renforcer la communauté ou de recruter de nouveaux membres ou même d'initier la création d'une nouvelle communauté. Enfin, dans notre vie de tous les jours, les différents rôles que nous jouons dépendent de notre position spatiale (géographique ou temporelle ou professionnelle etc..) et vice versa.

3.2.4 Représentation par les transformations de graphe

La composition des règles de transformations (Zheng-guang, Xiao-hui et al. 2006) peut être appliquée pour l'ajout ou l'effacement des nœuds et des arcs d'un graphe.

Pour les définitions ci-dessous, soit θ_{sp} l'ensemble de graphes liés à l'organisation spatiale et θ_{so} l'ensemble de graphes liés à l'organisation sociale.

Soit $M, N \in \theta_{so}$ et $S, T \in \theta_{sp}$ (Figure 3.4).

Définition 1 (transformation du graphe social)

Nous introduisons la fonction logique Sp où Sp (X) vrai si un agent a affecté X (occupé ou changé un nœud dans X) où $X \in \theta_{sp}$.

Une règle de transformation pso (Zheng-guang, Xiao-hui et al. 2006) est exprimée par un morphisme sur M et N :

$pso: M \mapsto N$. où M et N $\in \theta_{so}$.

Étant donné l'ensemble de règles Π_{pso} , et $s, f \in \Pi_{pso}$, la transformation de deux graphes dans l'organisation sociale est notée par:

Si Sp (S) vrai, $M \xrightarrow{cp} N$ où $cp \equiv (s \circ f): M \mapsto N$

où S $\in \theta_{sp}$ et un agent a pris une position sur S et M (Fig. 3.4).

Définition 2 (transformations des graphes spatiaux)

Nous introduisons la fonction logique So où $So(Y)$ vrai si un agent a affecté Y (occupé ou changé un nœud dans Y) où $Y \in \theta_{so}$

Une règle de transformation psa est exprimée par un morphisme sur S et T :

$$psa: S \mapsto T . \text{ où } S \text{ et } T \in \theta_{sp} .$$

Étant donné l'ensemble de règles Π_{psa} , et $s, f \in \Pi_{psa}$, la transformation de deux graphes dans l'organisation sociale est notée par:

$$\text{Si } So(N) \text{ vrai, } S \Rightarrow_{cp} T \text{ où } cp \equiv (s \circ f): S \mapsto T$$

où $N \in \theta_{so}$ et un agent a pris une position sur S et N (Fig. 3.4).

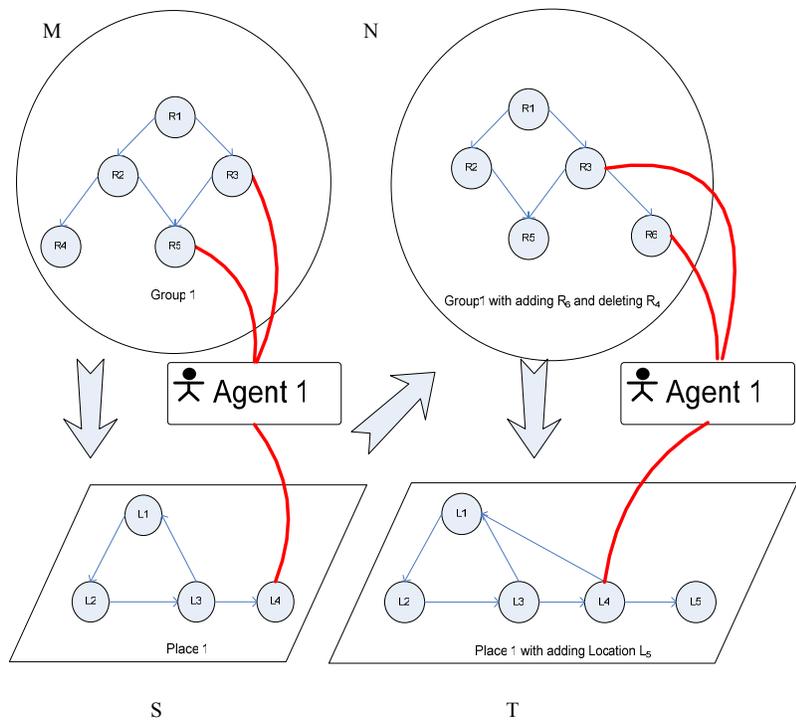


Figure 3.4 Transformation entre le graphe social et le graphe spatial

Dans la figure 3.4, un agent joue les rôles R3 et R5 et par conséquent, il se situe sur l'emplacement L3. Étant sur L3, il réalise le besoin d'ajouter et d'acquérir un nouveau rôle R6 et de quitter le rôle R3. Ceci le pousse à changer de position et se situer sur L4 et créer un nouvel emplacement L5. Cet effet rétroactif d'une organisation sur l'autre continue lors de l'évolution du système.

3.2.5 Phase d'analyse

Identification de l'organisation sociale: Dans la phase d'analyse du système et afin de définir les requis du système, nous proposons l'utilisation des « use cases » de UML pour identifier l'ensemble des groupes et rôles (Fig. 4.3). Les acteurs dans les use cases sont traités comme des agents dans les groupes. Par l'identification des use cases, nous pouvons découvrir éventuellement les rôles potentiels, et les associations entre les acteurs et les use cases permet d'assurer le mapping entre les groupes et les rôles. Après l'identification de l'ensemble initial de groupes et rôles, les relations entre les rôles comme les différentes contraintes peuvent être identifiées.

Identification de l'organisation spatiale : Concerne l'organisation de l'environnement physique dans lequel sont longés les agents. Il s'agit de définir les places (espace occupé par chaque agent) et les régions (regroupement physique de places). Cette organisation doit définir les métriques permettant d'identifier les distances et proximités entre places et entre régions, ainsi que la topologie de l'espace, conséquence de la co-évolution entre organisation sociale et organisation spatiale des agents.

Une fois les structures de base des organisations sociale et spatiale définies, l'ensemble des actions liées à chaque rôle doit être identifié ainsi que les protocoles que les agents utilisent pour interagir au sein des organisations spatiales et sociales.

3.3 Synthèse

La notion de machines sociales futures introduite dans la définition du « Web Science » trouve sa justification dans notre approche qui intègre le couplage explicite entre

l'organisation sociale et l'organisation spatiale dans l'évolution des systèmes complexes. En première application, nous avons adopté cette approche sur les communautés virtuelles qui sont un exemple de réseaux sociaux virtuels qui se forment à partir des utilisateurs qui partagent des intérêts communs. A travers notre simulation à base d'agents d'une communauté virtuelle particulière, nous avons observé l'effet de ce couplage dans l'évolution du système. Cette simulation est décrite en détail au chapitre suivant. Mais le phénomène de stigmergie n'est pas assez explicite dans ces types d'applications, et avec l'évolution du web 2.0 et la popularité croissante des systèmes d'étiquetage collaboratifs, nous avons exploré ces techniques de couplage dans un système où les actions des utilisateurs laissent des traces persistantes dans l'environnement (les étiquettes) qui affectent les actions futures. En intégrant la dimension sociale au sein de la conception du système, des nouvelles fonctionnalités ont émergé comme résultat de ce couplage entre l'organisation sociale et l'organisation spatiale, et ces fonctionnalités ont le potentiel d'améliorer les systèmes d'étiquetage actuels comme c'est décrit dans le chapitre 5.

Chapitre 4

Application à la simulation d'une communauté virtuelle

Pour illustrer l'approche proposée dans le chapitre 3, nous l'avons appliquée dans un premier temps pour l'analyse à travers la simulation de l'évolution d'une communauté virtuelle (Rupert, Hassas et al. 2007). L'objectif principal de cette application est de montrer l'effet de l'organisation spatiale et de l'organisation sociale sur les actions des agents et, par suite, sur l'évolution du système. Le rôle social qu'un agent joue et sa position spatiale déterminent son comportement et ses actions, ce qui a un effet direct sur l'environnement. En particulier, nous avons étudié lors de cette simulation, la dépendance qui existe entre la position spatiale d'un agent à un moment donné et le rôle social qu'il choisit de jouer et vice versa. Nous avons choisi comme premier exemple d'applications les communautés virtuelles car dans ce genre d'applications du web, les utilisateurs ont commencé à être impliqués directement dans l'évolution du web en participant à ces communautés. Ils ont perdu leurs statuts de lecteurs passifs du contenu du web. Leur présence sociale est imminente et doit être prise en considération dans l'étude de l'évolution du web.

Une communauté virtuelle est une communauté de personnes qui communiquent et interagissent socialement les unes avec les autres, en utilisant l'Internet comme milieu de communication (Wikipedia 2007). Les exemples de communautés virtuelles varient entre les listes de distribution de courriels, aux blogs, aux « chat rooms », aux forums et même aux messageries instantanées.

Dans cet exemple, nous nous concentrons sur les communautés virtuelles dont les membres communiquent par des articles de forum sur un site Web. Les communautés virtuelles exhibent des comportements similaires à un système complexe adaptatif auto-organisé. Les agents dans les communautés virtuelles représentent les utilisateurs appartenant à la communauté, qui naviguent à travers son contenu et participent à l'évolution de la communauté. La communauté que nous explorerons est une communauté virtuelle de support des blessures et opérations des genoux (Maloney-Krichmar and Preece 2005).

4.1 Éléments de base

4.1.1 Les communautés virtuelles

Le terme communauté en ligne ou communauté virtuelle est perçu différemment par différents chercheurs. Pour les sociologues, c'est une place de développement des interactions et relations sociales entre individus, définissant des rôles et activités liées à ces rôles et pour les informaticiens, l'important c'est la structure des logiciels utilisés formant la communauté. Preece a combiné ces deux perceptions pour donner la définition suivante : « une communauté en ligne est un groupe de personnes qui, dans un environnement virtuel, ont un intérêt ou but commun, sont supportés par la technologie et sont guidés par des normes » (Preece 2000).

Ces personnes communiquent et agissent socialement les unes avec les autres en utilisant l'Internet comme milieu de communication. Cette interaction sociale « online » entre les gens forme l'usage le plus populaire de l'Internet qui, à son tour, a laissé un grand impact sur les relations sociales humaines. Comme l'Internet continue à croître à un taux rapide, les communautés virtuelles croissent à leur tour, ce qui nécessite une étude et analyse approfondies de ce nouveau phénomène.

En dépit de la croissance et la vibration des communautés virtuelles, un grand nombre ne réussit pas, avec une participation des utilisateurs chutant à zéro. Afin de mieux comprendre comment elles se développent, évoluent et fonctionnent dans un état mûr, nous proposons de les simuler pour pouvoir prévoir leur croissance et déterminer les facteurs qui motivent la contribution des participants. Différentes études ont été faites pour analyser et comprendre les raisons des faibles participations des utilisateurs, et différentes approches ont été adoptées pour essayer d'accroître et stimuler cette participation (Nabeth, Angehrn et al. 2005; Cheng and Vassileva 2006; Du 2006; Mao, Vassileva et al. 2007).

L'aspect social des communautés virtuelles doit être pris en compte. Rheingold, l'un des pionniers dans ce domaine, a popularisé le terme communautés virtuelles dans son livre « Virtual Community page 4 » (Rheingold 1993), argumente l'importance de l'effet social de ces communautés. D'autres chercheurs ont montré que la sociabilité en ligne est un fait de la vie de chaque jour (Feenberg and Bakardjieva 2004).

4.1.2 La simulation à base d'agents

Le paradigme des systèmes multi-agents, couplé des mécanismes d'auto-organisation et d'émergence, a été utilisé pour la simulation et la modélisation des systèmes complexes (Serugendo, Gleizes et al. 2005). La simulation à base d'agents est utilisée dans différents domaines tel que les sciences sociales, la biologie, la modélisation des systèmes socio-écologiques, les systèmes économiques, etc. Dans ce type de simulations, les différentes entités sont représentées par des agents où le comportement global à l'échelle macroscopique du système émerge à partir des différents comportements à l'échelle microscopique. (Drogoul, Treuil et al. 2008) ont proposé un cadre commun théorique pour la modélisation et simulation à base d'agents tels que des modèles physiques, biologiques, comportementaux et sociaux. La simulation multi-agents est aussi utilisée pour la modélisation du comportement des utilisateurs dans les systèmes centrés-utilisateurs sur l'Internet tels que les réseaux pair-à-pair (P2P) (Siebert, Ciarletta et al. 2008). Pour la simulation de l'évolution des communautés virtuelles d'utilisateurs, la simulation à base d'agents trouve sa justification surtout pour appliquer les principes d'organisations sociales et spatiales.

4.1.3 Description de la communauté virtuelle considérée

Nous avons choisi comme exemple d'application pour la simulation, la communauté suivante qui existe déjà sur le web : le site de Bob's ACL WWWBoard (<http://factotem.org/cgi-bin/kneebbs.pl>) est une communauté virtuelle. ACL c'est le ligament croisé antérieur ou « *Anterior Cruciate Ligament* » (Fig. 4.1).

Cette communauté concerne principalement les personnes qui ont des blessures aux genoux, qui ont eu, ou vont avoir des opérations aux genoux. C'est un espace où les blessés et leur entourage partagent de l'information, partagent leur propre expérience, s'entraident et se soutiennent mutuellement, etc. Cette communauté virtuelle fournit de l'information concernant un sujet spécifique de la santé (la blessure des genoux) avec le but de servir de médium dans lequel les utilisateurs peuvent communiquer entre eux.

La communauté est composée de pages web et des utilisateurs qui sont membres de la communauté. La plupart des pages web sur le site de Bob's ACL WWWBoard sont des postes sur un forum où un utilisateur pose une question, où l'on donne de l'information, où l'on répond aux questions posées.

Cette communauté virtuelle (Bob's ACL WWWBoard) a été analysée en détail dans (Maloney-Krichmar and Preece 2005). Pendant deux ans et demi, la communauté a été étudiée à plusieurs niveaux pour examiner les différents aspects de sociabilité et de dynamique de la communauté. Dans cette analyse, une liste des différents rôles entrepris par les utilisateurs a été identifiée. Nous avons utilisé ce travail déjà existant comme base pour notre application de simulation. C'est ce qui justifie notre choix de cette communauté particulière.

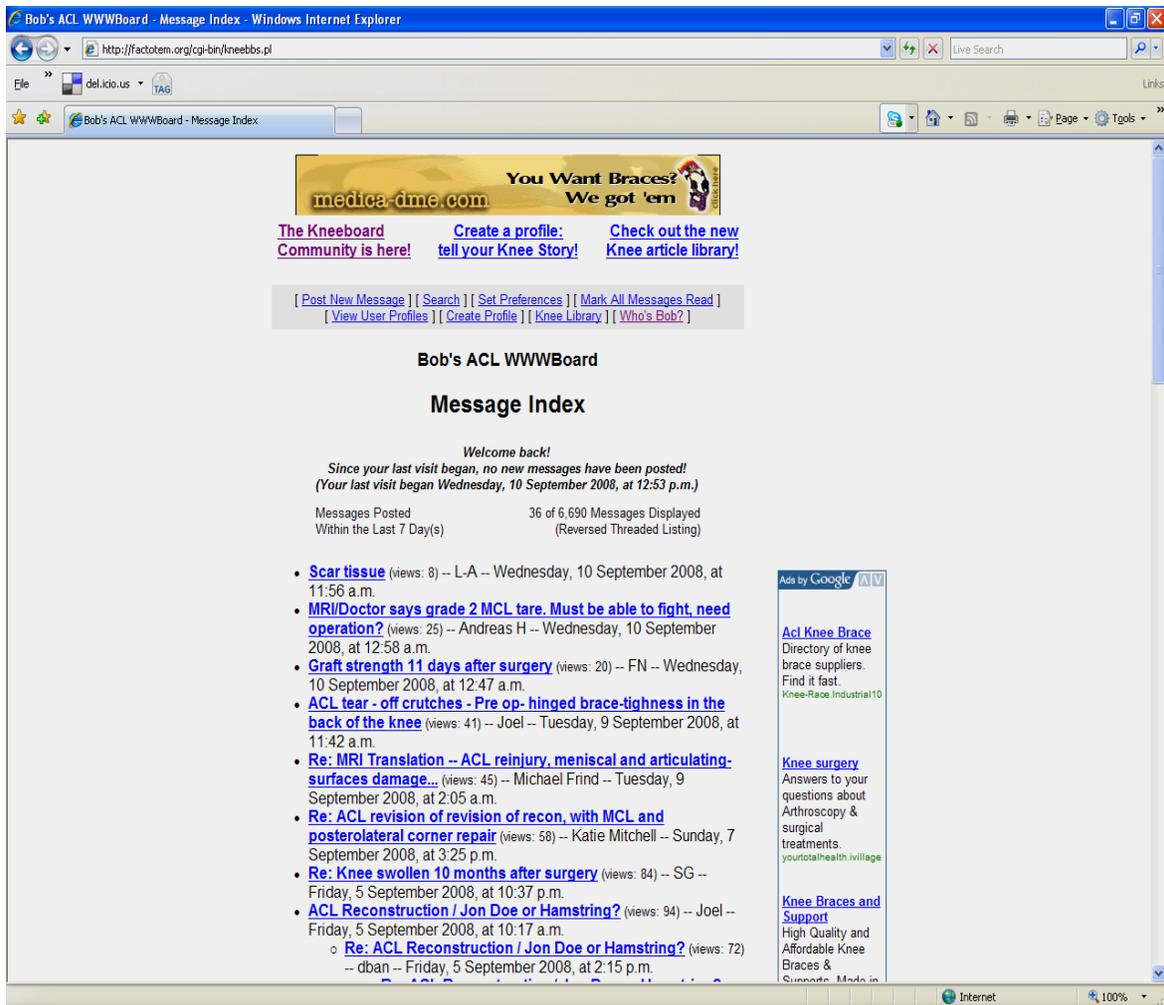


Figure 4.1 La page d'accueil captée le 10 septembre 2008

Nous avons créé une application qui simule cette communauté de santé, utilisant l'approche organisationnelle multi-agents décrite au chapitre 3. Dans notre application de simulation, les utilisateurs sont modélisés par les agents, et les postes sur le forum vont constituer les places où ces agents vont se situer, diffuser de la phéromone, et se déplacer. On attribue aux agents un ou plusieurs rôle(s). Ces rôles sont le reflet des actions que les membres de la communauté entreprennent, tel que fournir de l'information et rechercher de l'information. La communication et les interactions entre les agents sont obtenues en utilisant le mécanisme de stigmergie par la diffusion des phéromones.

Nous traitons la structure physique de cette communauté comme un environnement spatial pour nos agents. Dans cet environnement spatial, une page web est représentée par une « place » alors qu'un ensemble de pages web similaires regroupées est représenté par une « région ». Les pages web forment l'environnement physique spatial où les agents se situent, tandis que les interactions entre les agents forment l'environnement social.

4.2 Description du système

L'application multi-agents est basée sur le cadre organisationnel spatial et social. L'information pour le système multi-agents sera stockée dans une base de données et le Chargeur d'Application sera responsable de récupérer cette information et de l'utiliser pour créer les objets utilisés dans le système multi-agents. Les objets enverront de l'information sur leurs activités à l'Enregistreur des Activités qui sauvegardera l'information dans la base de données. Ceci permettra à l'information concernant le système multi-agents d'être facilement recherchée et analysée.

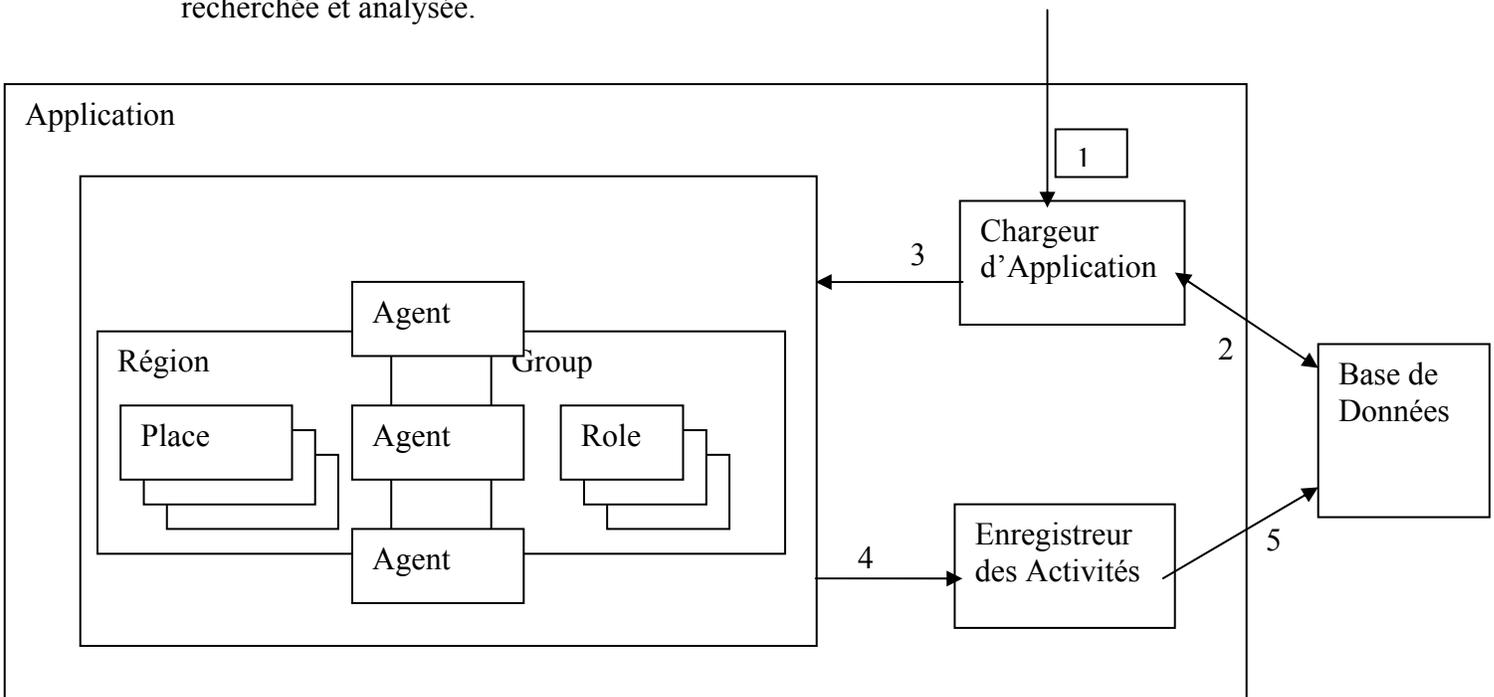


Figure 4.2 Les flèches sur le diagramme représentent une séquence d'actions et montrent comment l'application sera exécutée.

Avant de démarrer l'application, l'utilisateur nourrit la base de données avec de l'information, telle que les agents, les groupes, les rôles, les places, les régions, etc. Une fois que cette tâche est complétée, l'utilisateur démarre le système multi-agents, ce qui invoque le Chargeur d'Application (Fig. 4.2). Le Chargeur d'Application recouvre les données concernant le système de la base de données (c'est ce que l'utilisateur a déjà mis dans la base de données). Le Chargeur d'Application utilise les données pour créer les objets qui seront utilisés dans la simulation du système. Une fois que les objets seront créés, l'application démarrera. Ces objets vont générer des données concernant leurs activités. Ces données seront transmises à l'Enregistreur d'Activités utilisant des messages (l'objet d'un simple message avec de l'information sur la source d'où provient ce message, l'horodatage, le texte du message, etc.). L'Enregistreur d'Activités va traiter ces messages et stocker l'information dans la base de données.

4.3 L'organisation spatiale dans la communauté virtuelle

Les différentes pages Web de la communauté virtuelle sont le milieu physique dans lequel les agents agissent les uns avec les autres. Les pages ou places incluent une page d'accueil, une page pour les articles de base de données, une page d'inscription des membres et les pages d'articles de forum où les questions et réponses sont postées, etc (Fig. 4.3). Un agent peut être situé dans n'importe quelle place. Les liens entre les pages forment les chemins reliant les différentes places dans les rapports du voisinage. Le degré ou la force de ce rapport de voisinage est basé sur la proximité des pages Web les unes des autres. Les pages qui ont un contenu similaire ont des rapports de voisinage plus forts et ont une distance plus courte entre elles que les pages qui ont des contextes différents. L'agrégation qui en résulte est représentée dans le système par des régions, où chaque région correspond à un sujet. L'environnement spatial se compose de régions, de places, de chemins, et de phéromones.

4.3.1 Les places

Les places possèdent un contenu, qui peut être composé d'un ou de plusieurs sujets. Il existe un sujet principal, et il peut y avoir des sujets additionnels, à condition qu'ils soient liés

au sujet principal. Les sujets sont représentés par un graphe où les nœuds constituent les sujets et les arcs constituent les liens entre les sujets. Chaque sujet dans la place possède une proportion qui indique quel pourcentage de la place ce sujet occupe. En outre, chaque sujet a une valeur de pertinence qui indique combien l'information ou l'opinion sur ce sujet sont pertinentes. Plus le pourcentage est haut, plus le sujet est pertinent.

4.3.2 Les chemins

Des chemins relient les places entre elles et sont bidirectionnels. Ils fournissent aux agents des moyens de navigation parmi les places au sein de la communauté. Chaque chemin a une distance qui indique à quel point les deux places sont reliées en termes de leur contenu. Plus la distance est courte, plus les places sont liées. La distance est calculée en se basant sur le nombre de sujets qui ont deux places en commun comparé au nombre de sujets total dans les deux places. En outre, la pertinence et la proportion des sujets communs aux deux places sont également comparées.

4.3.3 Les régions

Les régions regroupent les places qui ont un contenu similaire, et chaque région porte sur un seul sujet à la fois. Les régions peuvent être considérées comme des clusters de places traitant le même sujet.

4.3.4 Les phéromones

Les phéromones font partie de l'environnement spatial et sont utilisées par les agents pour la communication indirecte. Un agent dépose une phéromone à une place après l'avoir créée pour alerter d'autres agents. Les phéromones ont un type ou une étiquette de phéromone qui consiste en sujets et une proportion pour chaque sujet qui indique à quel point la phéromone subsiste sur le sujet.

En outre, chaque phéromone a un taux d'évaporation, une intensité, et un taux de diffusion.

-Le taux d'évaporation : $e = (1 - \rho_j)$

ρ_j exprime le taux de persistance de l'information sur une place i dans l'environnement. La plus petite est sa valeur, la plus longue est la diffusion de l'information :

$$\rho_j = |N_{ij}| / |N_i| \quad [1]$$

N_{ij} est l'ensemble des voisins de la place i ayant le sujet j et N est l'ensemble de tous les voisins de la place i .

-Le taux de diffusion λ est calculé en se basant sur la distance entre les places voisines ayant le même sujet j et la place i où la phéromone a été déposée.

$$\lambda = \sum_{j \in T} p_j \left(\frac{|N_{ij}|}{N_i} \right) \quad [2]$$

où T est l'ensemble de tous les sujets, p est la proportion du sujet j dans la place.

-L'intensité $I(d,t)$: exprime la pertinence de l'information. Cette formule calcule le niveau de phéromone pour une distance donnée d de la place où la phéromone a été déposée au temps t .

$$I(d, t) = I_0 \rho_j^t (\lambda)^{-d/\lambda} \quad [3]$$

ρ_j représente le taux de persistance, $(1 - \rho_j)$ le taux d'évaporation, I_0 l'intensité initiale, λ le taux de diffusion.

4.4 L'organisation sociale dans la communauté virtuelle

4.4.1 Rôle des agents et leurs contraintes

Les agents qui appartiennent à la communauté virtuelle de la santé peuvent jouer certains des rôles définis dans (Maloney-Krichmar and Preece 2005). Ces rôles sont résumés dans le Tableau 4.1. Au début de la simulation, on rendra chaque agent conscient de tous les rôles possibles. Pour chaque rôle, l'agent recevra ou non un rôle. Un agent doit accepter au moins un rôle, mais n'est pas tenu de devoir remplir tous les rôles. Lorsqu'un agent est assigné à un

rôle, il ne quitte plus ce rôle. Si un agent est assigné à un rôle, alors il peut accomplir les actions disponibles faisant partie de ce rôle.

Rôle	Description	Actions
Fondateur / Administrateur	C'est la personne qui crée la communauté en ligne (une seule personne par communauté peut jouer ce rôle et dans chaque communauté ce rôle doit toujours être attribué à quelqu'un). Le fondateur a une connaissance significative de la matière du sujet de la communauté et est actif dans la communauté.	Crée une nouvelle communauté.
Fournisseur d'Information	Fournit de l'information sur un sujet.	Répond aux questions reçues en expédiant une réponse avec information. Envoie des phéromones pour communiquer avec les autres agents qu'une place est importante.
Fournisseur d'Opinion	Fournit des opinions sur un sujet.	Répond aux demandes des chercheurs d'opinions en leur envoyant une réponse avec une opinion. Envoie des phéromones pour communiquer avec les autres agents qu'une place est importante.
Chercheur d'Information	Cherche de l'information sur un ensemble de sujets.	Crée des places en posant des questions sur un sujet. Envoie des phéromones pour communiquer avec les autres agents qu'une place est importante.
Chercheur d'Opinion	Cherche de l'information sur un ensemble de sujets.	Crée des places en demandant des opinions sur un sujet. Envoie des phéromones pour communiquer avec les autres agents qu'une place est importante.

Tableau 4.1 Liste des rôles des agents qui seront utilisés dans l'application.

L'habileté d'un agent à s'attribuer un rôle dépend de la nature de ce rôle et des contraintes de l'organisation sociale comme définies dans le méta-modèle du chapitre 3 et sont présentées dans le Tableau 4.2 :

Contraintes Rôles	Contrainte de multiplicité	Contrainte de dépendance	Contrainte de correspondance
Fondateur / Administrateur	1	Fournisseur d'Information Fournisseur d'Opinion	Fournisseur d'Information Fournisseur d'Opinion
Fournisseur d'Information	Nombre total d'agents	Chercheur d'Information	Chercheur d'Information
Fournisseur d'Opinion	Nombre total d'agents	Chercheur d'Information Chercheur d'Opinion	Chercheur d'Information Chercheur d'Opinion
Chercheur d'Information	Nombre total d'agents	aucune	Aucune
Chercheur d'Opinion	Nombre total d'agents	aucune	aucune

Tableau 4.2 Liste des rôles et de leurs contraintes correspondantes

Pour les rôles de Chercheur d'Information et de Chercheur d'Opinion, il n'y a aucun pré-requis, la probabilité qu'un agent s'assigne l'un de ces rôles demeure uniforme. Lorsqu'un agent accède au rôle de Fondateur/Administrateur, il crée une nouvelle communauté basée sur la connaissance qu'il a acquise. Pour simplifier notre application, la nouvelle communauté ne sera pas développée, mais seulement créée. L'agent va continuer à demeurer actif dans la communauté actuelle.

Lorsqu'un agent se joint à une communauté, il peut commencer en occupant n'importe quel rôle, sauf celui de Fondateur/Administrateur (parce qu'il y en a déjà un). Si l'agent débute comme Fournisseur d'Information ou Fournisseur d'Opinion, on présume qu'il possède déjà les connaissances qui l'habilitent à remplir le rôle (comment et où l'agent a acquis ses connaissances est au-delà de la portée de l'application et ne sera pas considéré). Un agent doit s'inscrire à un rôle quand on lui assigne ce rôle. C'est nécessaire afin que l'on sache bien quels agents remplissent ce rôle. De plus, les assignations de rôles déterminent le type de phéromones qu'un agent cherche et dépose (Tableau 4.2).

Rôle de l'agent	Phéromones Cherchées	Phéromones Déposées
Chercheur d'Information	Réponse d'Information	Requête d'Information
Fournisseur d'Information	Requête d'Information	Réponse d'Information
Chercheur d'Opinion	Réponse d'Opinion	Requête d'Opinion
Fournisseur d'Opinion	Requête d'Opinion	Réponse d'Opinion

Tableau 4.3 Phéromones cherchées et déposées selon les rôles des agents

4.4.2 Actions des agents

Un agent peut être assigné à un ou plusieurs rôles. Quand un agent arrive à une place, il vérifie si le sujet qui l'intéresse est à la place. Si le sujet n'est pas à la place, l'agent se déplace à une autre place. La figure 4.3 présente un scénario où un agent peut naviguer. L'action qu'un agent exécute dépendra du rôle qu'il joue et de sa place actuelle. Les agents peuvent effectuer un certain nombre d'actions :

- Se déplacer d'une place à une autre
- Percevoir le contenu (sujet) de leur place actuelle
- Mettre à jour sa connaissance en utilisant le contenu de sa place actuelle
- Percevoir les chemins (destinations aux places voisines) de sa place actuelle
- Percevoir les phéromones de la place actuelle, y compris des phéromones diffusées de places voisines
- Déposer des phéromones
- S'assigner à un nouveau rôle.
- Changer son rôle
- Créer des places et des régions

- Créer de nouvelles communautés

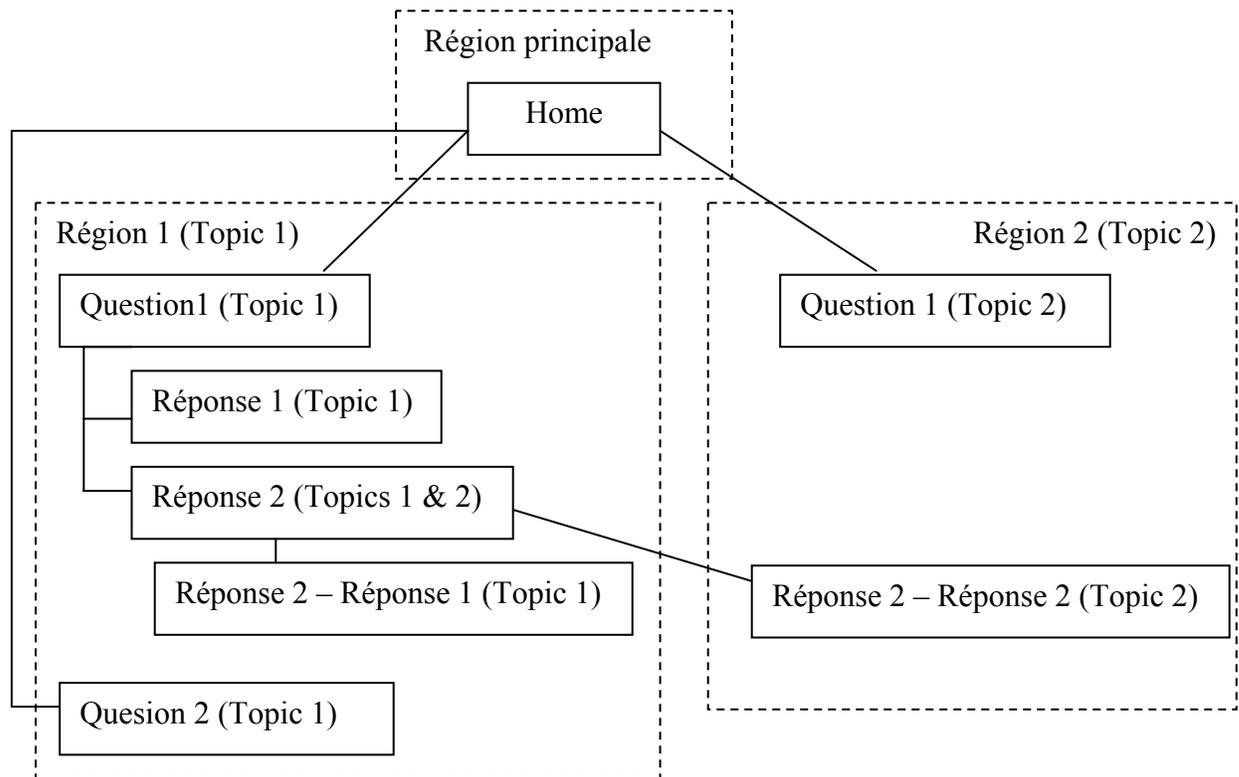


Figure 4.3 Diagramme d'un scénario possible où les questions et réponses forment les places dans l'organisation spatiale

Les actions les plus importantes effectuées par des agents sont la mise à jour de la connaissance, le déplacement à une nouvelle place, la création d'une place, et l'assignement à un rôle.

La mise à jour de la connaissance : Un agent maintient sa connaissance sur chaque sujet. L'agent a une capacité maximale de connaissance. Quand un agent met à jour sa connaissance, la quantité de connaissance qu'il gagne sur chaque sujet dépend du pourcentage du sujet à la place où il est situé et de sa pertinence.

Le déplacement à une nouvelle place : L'agent suit les phéromones pour trouver les places qui ont le sujet qui l'intéresse. S'il n'y a aucune phéromone avec le sujet qui intéresse l'agent, il choisira alors une place voisine pour s'y déplacer.

La création d'une place : Quand un agent crée une nouvelle place, il doit vérifier le type de la place, la place voisine, le sujet de la place, et la région de la place. Le type de la place sera déterminé selon l'action que l'agent prend dans le rôle qu'il joue actuellement. Le sujet principal de la place est le sujet qui intéresse l'agent. Les sujets additionnels peuvent être ajoutés, tant que chaque sujet dans la place est lié au moins à un autre sujet dans la place, et l'agent a acquis la connaissance de ce sujet. La pertinence de chaque sujet dans la place sera basée sur une combinaison de la connaissance que l'agent a sur ce sujet comme pourcentage de sa capacité de connaissance et à quel point le sujet est lié aux autres sujets dans la place. Si le sujet de la place où l'agent est actuellement est similaire au sujet principal de la nouvelle place, alors la nouvelle place sera ajoutée à cette région. Sinon, l'agent devra décider s'il crée une nouvelle région pour la place. En conclusion, l'agent se déplacera à une nouvelle place et laissera tomber une phéromone pour alerter d'autres agents de la nouvelle place.

L'assignation de rôle : Si l'agent se rend compte d'un rôle, il peut être assigné au rôle s'il répond à un ensemble d'exigences prédéfinies. Une fois que l'agent est assigné au nouveau rôle, il peut effectuer les actions liées à ce rôle.

4.5 Simulation de l'évolution de la communauté virtuelle du support de la santé

4.5.1 Implémentation

Le programme de simulation a été développé en Java. La simulation a été exécutée plusieurs fois en utilisant 151 agents. Ce nombre est dû aux limitations de nos ressources physiques. Les résultats de la simulation ont montré que le nombre d'agents assignés au rôle de fournisseur d'information augmente chaque fois. Les autres rôles n'ont pas augmenté à la même fréquence que le fournisseur d'information.

Nous avons examiné le nombre de nouvelles places et régions créées. Les résultats ont montré que le nombre de chaque type de place a augmenté. En outre, de nouvelles régions sont créées et il y a une région pour chaque sujet de la communauté virtuelle de la santé. Un examen plus précis des données, cependant, prouve que les régions qui ont un nombre restreint de places, un pourcentage très élevé de places dans la région ont le même sujet principal de la

région. Afin de déterminer à quel point les places sont groupées ensemble, nous avons examiné les voisins d'une place et nous avons calculé le pourcentage de places voisines qui ont le même sujet principal. Un pourcentage élevé indique que les places qui ont les mêmes sujets principaux sont localisées ensemble. Pour presque chaque place, toutes les places voisines ont le même sujet principal. Une autre mesure était l'étude de la dépendance entre l'organisation sociale et l'organisation spatiale et, en particulier, entre les rôles et les places afin de montrer le degré de couplage entre les deux organisations. Un test χ^2 a été effectué sur les types de places créées par des agents actifs dans un certain rôle. Une table d'éventualité a été construite pour calculer les fréquences prévues pour chaque intersection de rôle et de type de la place. Le test prouve qu'il y a une dépendance entre le type de place créé et le rôle actif de l'agent qui a créé la place.

4.5.2 Environnement de simulation

La simulation a été conduite à trois reprises, utilisant 151 agents. Les tableaux 4.1 à 4.7 montrent comment les agents furent divisés dans les rôles dans lesquels ils étaient actifs au début de la simulation (à noter que le fondateur est actif comme Fournisseur d'Information car on suppose que la communauté a déjà été créée). La capacité de connaissance pour tous les agents dans la simulation était 100.

Nombre d'agents	Sujet d'intérêt	Rôles assignés	Connaissance du sujet d'intérêt	Connaissance des autres sujets
1	Sujet principal de la communauté	Fondateur Chercheur d'Information Fournisseur d'Information Chercheur d'Opinion Fournisseur d'Opinion	90	90
5	1 par sujet	Chercheur d'Information Fournisseur d'Information	70	0
5	1 par sujet	Chercheur d'Information Fournisseur d'Information	70	10
10	2 par sujet	Chercheur d'Information Fournisseur	70	10

		d'Information Chercheur d'Opinion Fournisseur d'Opinion		
--	--	---	--	--

Tableau 4.2 1 fondateur et 20 Fournisseurs d'Information

Nombre d'agents	Sujet d'intérêt	Rôles assignés	Connaissance du sujet d'intérêt	Connaissance des autres sujets
5	1 par sujet	Chercheur d'Opinion Fournisseur d'Opinion Chercheur d'Information Fournisseur d'Information	70	0
15	3 par sujet	Chercheur d'Opinion Fournisseur d'Opinion Chercheur d'Information Fournisseur d'Information	70	10

Tableau 4.3 20 Fournisseurs d'Opinion

Nombre d'agents	Sujet d'intérêt	Rôles assignés	Connaissance du sujet d'intérêt	Connaissance des autres sujets
25	5 par sujet	Chercheur d'Information	0	0
25	5 par sujet	Chercheur d'Information	10	0
15	3 par sujet	Chercheur Information Fournisseur d'Information	0	70
15	3 par sujet	Chercheur d'Information Fournisseur d'Information	10	70
3	Communauté principale sujet (1), sous-sujet A (1), et sous-sujet B (1)	Chercheur d'Information Chercheur d'Opinion	5	0

Tableau 4.4 83 Chercheurs d'Information

Nombre d'agents	Sujet d'intérêt	Rôles assignés	Connaissance du sujet d'intérêt	Connaissance des autres sujets
15	3 par sujet	Chercheur d'Opinion	0	0
10	2 par sujet	Chercheur d'Opinion	10	0
2	Sous-sujet C (1), et sous-sujet D (1)	Chercheur d'Information Chercheur d'Opinion	5	0

Tableau 4.5 27 Chercheurs d'Opinion

4.6 Les résultats

L'objectif de cette simulation est d'étudier l'évolution des organisations sociales et spatiales des communautés virtuelles et les effets rétroactifs de l'une sur l'autre. Nous nous sommes concentrés sur les cinq aspects suivants de l'évolution de la communauté simulée :

Pour chaque rôle dans la communauté, excepté le Fondateur/Administrateur, le nombre d'agents assignés au rôle augmente (dans la simulation, il ne peut y avoir qu'un seul Fondateur/Administrateur dans la communauté).

Il y a une dépendance entre le rôle que l'agent s'assigne et le type de place où l'agent était lorsque l'assignation s'est produite.

De nouvelles places et de nouvelles régions sont ajoutées à la communauté.

Les places avec des sujets similaires sont regroupées ensemble.

La simulation a été conduite plusieurs fois pour déterminer si les résultats étaient consistants. On a trouvé que les résultats étaient relativement consistants en chaque occasion, avec des variations mineures dans les résultats. Les résultats de trois rondes de simulations sont inclus dans l'Annexe A.

4.6.1 Nombre d'assignations de rôles

Avec le temps, le nombre d'agents assignés à chaque rôle dans la communauté a augmenté, sauf pour le rôle de Fondateur, parce qu'il ne peut y avoir qu'un seul Fondateur dans une

communauté (un test séparé est fait pour créer de nouvelles communautés). C'est parce que les agents vont acquérir plus de connaissances et exécuter plus d'actions, ce qui les rendra capables de rencontrer les exigences pour remplir des rôles auxquels ils ne sont pas assignés.

Les résultats de la simulation (voir l'Annexe A) montrent que le nombre d'agents assignés au rôle de Fournisseur d'Information augmente à chaque fois. C'est consistant avec ce qui se produit dans les autres rondes de la simulation. Il n'y a pas eu d'augmentation dans le nombre d'agents assignés aux rôles de Chercheur d'Information et de Chercheur d'Opinion et ceci est attribué à notre conception du système où nous avons limité l'ensemble d'actions exécutées pour chaque combinaison de rôle actif et type de place. Le tableau 4.8 montre comment notre algorithme favorise un type de place particulier selon le rôle assigné à l'agent.

Rôle Actif	Type de place favorisé
Chercheur d'Information	Réponse d'Information
Fournisseur d'Information	Demande d'Information
Chercheur d'Opinion	Réponse d'Opinion
Fournisseur d'Opinion	Demande d'Opinion

Tableau 4.6 Type de place favorisé selon le rôle de l'agent

4.6.2 Dépendance entre la position spatiale et la position sociale de l'agent

Il y a une dépendance entre le rôle actif d'un agent et le type de nouvelles places. Le rôle actif d'un agent détermine ce qu'il veut gagner ou la connaissance qu'il veut donner. Par exemple, un Chercheur d'Information veut plus d'information sur le sujet qui l'intéresse, alors ça va créer une place de demande d'information. Un test χ^2 a été fait sur les types des places créées par les agents actifs dans un certain rôle. Le test montre qu'il y a une dépendance entre le type de places créées et le rôle actif de l'agent qui a créé la place. Les résultats de ce test peuvent être trouvés dans l'annexe A.

4.6.3 Créations de nouvelles places et régions

De nouvelles places et régions ont été créées dans la communauté. Les agents veulent donner ou acquérir la connaissance, alors ils ont créé des places pour atteindre ce but. Comme partie de la création des places, les agents ont aussi créé de nouvelles régions pour ces places.

Les résultats montrent que le nombre de chaque type de place a augmenté (voir l'Annexe A pour un compte détaillé du nombre de chaque type de place créé). De plus, de nouvelles régions sont créées et il y a une région pour chaque sujet de la communauté virtuelle. Il y a seulement un ensemble d'actions où une Demande d'Opinion est créée et c'est lorsqu'un Chercheur d'Opinion visite la place de Réponse d'Information. La place est créée seulement si le sujet qui intéresse l'agent a une large proportion et est pertinent pour cette place.

4.6.4 Groupement par région des places similaires

Les places se sont regroupées ensemble par leur sujet principal dans une région ayant le même sujet. Les agents sont responsables de l'assignation des places à une région. L'algorithme utilisé s'assure qu'une place sera ajoutée à une région si la région a un sujet identique au sujet principal de la place.

Les résultats de la simulation (Tableau 4.9) montrent que pour la plupart des régions, moins que la moitié des places dans la région ont un sujet principal qui est le même que le sujet de la région. Un examen plus approfondi des données, cependant, montre que pour les régions ayant un petit nombre de places, un très haut pourcentage de places dans la région ont un sujet principal qui est le même que celui de la région. Il apparaît que plus une région a de places, plus bas est le pourcentage des places dans la région qui ont un sujet principal qui est le même que celui de la région.

Sujet de la région	Nom du sujet	Nombre de places dans la région où le sujet principal de la place est le sujet de la région	Nombre total de places dans la région	Pourcentage de places dans la région où le sujet principal de la place est le
1	Main Community Topic/Focus	19	46	41.30%
2	Topic A	32	128	25.00%
3	Topic B	7	7	100.00%
4	Topic C	94	680	13.82%
5	Topic D	9	9	100.00%

Tableau 4.7 Calcul du pourcentage de places dans la région où le sujet principal de la place est le même que le sujet de la région

Une autre méthode pour déterminer à quel point les places sont bien groupées ensemble est d'examiner les voisines d'une place et de calculer le pourcentage des places de voisinage où le sujet principal est le même que le sujet principal de la place. Un pourcentage élevé indique que les places avec les mêmes sujets principaux sont situées les unes près des autres. Les résultats dans le Tableau 4.10 montrent que pour presque chaque place, toutes les places voisines ont le même sujet principal que celui de la place.

Pourcentage des places voisines ayant le même sujet principal	Nombre de places
0.00	5
14.29	1
85.71	1
100.00	873

Tableau 4.8 Pourcentage moyen des places voisines ayant le même sujet principal est 99.32%

4.7 Conclusion

Les communautés en ligne (ou communautés virtuelles) sont un exemple d'applications centrées-utilisateurs qui de plus en plus façonnent le web, en agissant sur son évolution. Dans ce type d'applications, nous avons voulu mettre en évidence que la structure du réseau est guidée dans son évolution par l'usage, qui à son tour évolue en réaction à l'évolution de la structure. Afin, de fournir un support à notre observation, nous avons représenté cette

coévolution par la coévolution des organisations spatiales et sociales dans les systèmes multi-agents.

Dans ce premier exemple d'application de l'approche organisationnelle décrite au chapitre 3, nous avons effectué une simulation de l'évolution de la communauté virtuelle « Bob's ACL WWWBoard ». Les résultats nous ont montré que l'intégration des organisations spatiales et sociales a un effet sur les actions des agents et, par suite, sur l'évolution du système. Le rôle social qu'un agent joue et sa position spatiale déterminent son comportement et ses actions. Nous nous sommes intéressés en particulier à l'étude des rôles des agents en relation avec leur position spatiale et nous avons constaté qu'il existe une dépendance entre les deux.

Cette dépendance nous montre qu'en effectuant des simulations sur des systèmes centrés-utilisateurs, le concept que nous proposons pour la prise en considération de la dimension spatiale en couplage avec la dimension sociale trouve sa justification dans ce genre d'applications. Cette prise en considération a permis au système d'évoluer tout en convergeant vers un état stable. Nous n'avons pas pu pousser nos expérimentations plus loin pour étudier si cette convergence va persister et continuer à observer l'évolution du système à cause des limitations des ressources physiques qui nous sont disponibles.

Par contre, et pour aller plus loin, nous avons décidé de tester l'effet de cette coévolution entre le spatial et le social dans un second exemple d'applications centrées-utilisateurs pris du web 2.0 (les systèmes de tagging ou d'étiquetage collaboratifs). Nous avons conçu notre propre système d'étiquetage collaboratif MySURF qui intègre ces notions d'organisations spatiales sous forme de sous-communautés de ressources et d'organisations sociales sous forme de groupes virtuels d'utilisateurs. Nous avons étudié les fonctionnalités de ce système proposé en les comparant aux systèmes d'étiquetage actuels, notant les améliorations que notre système présente.

Chapitre 5

Proposition du système d'étiquetage collaboratif MySURF

En étudiant et expérimentant les systèmes d'étiquetage collaboratifs existants, nous avons noté le manque d'adaptabilité et de personnalisation dans ces systèmes. A première vue ces systèmes fournissent de l'information et des ressources étiquetées qui sont ajoutées par les utilisateurs eux-mêmes. Mais, en étudiant ces systèmes de façon plus approfondie, et en tenant compte du pouvoir de l'aspect social et l'aspect spatial dans ces systèmes, nous avons réalisé que nous pouvons en extraire des connaissances bien au delà d'une liste de ressources correspondant à une étiquette particulière. Les systèmes actuels possèdent plusieurs limitations au niveau de la recherche de l'information et au niveau de l'intégration de la dimension sociale qui assure leur évolution comme systèmes complexes. Dans ce travail, nous avons profité de la coévolution qui existe entre l'organisation sociale et l'organisation spatiale dans un système complexe pour concevoir un système d'étiquetage collaboratif permettant l'émergence de nouvelles fonctionnalités. Ce système est basé sur l'effet rétroactif du social sur le spatial et vice-versa. En bref, notre contribution majeure dans ce travail de thèse est la création d'un

nouveau système **MySURF** (**My Similar Users, Resources, and Folksonomies**) où nous avons injecté la dimension sociale dans le développement du système.

L'objectif de cette recherche est de compléter les systèmes d'étiquetage existants en ajoutant de nouvelles fonctionnalités qui peuvent améliorer ces systèmes, tout en prenant en considération les caractéristiques complexes de ces systèmes ainsi que les liens forts qui existent entre les organisations spatiales et les organisations sociales, ce qui, à notre connaissance n'a pas été bien exploré jusqu'à présent dans la conception de ces systèmes.

Dans la suite de ce chapitre, en première partie, nous explorons les propriétés des systèmes d'étiquetage collaboratifs actuels, le principe fondateur de ces systèmes, les problématiques posées dans ces systèmes et les réponses apportées et leurs limites. En seconde partie, nous fournissons une description détaillée du système MySURF que nous avons développé qui propose des réponses à certaines de ces limites et nous soulignons les fonctionnalités émergentes du nouveau système dues à l'approche organisationnelle adoptée dans la conception de ce système.

5.1 Parcours des systèmes d'étiquetage existants et de leurs limites

5.1.1 Les réseaux sociaux en ligne

Les sociologues ont longtemps étudié les réseaux sociaux et leurs propriétés. Le terme a été introduit par J.Barnes en 1954 (Barnes 1954). Un réseau social est un ensemble ou une communauté d'individus qui partagent un même intérêt (même origine, même famille, même travail, même hobby etc.) et qui sont liés ensemble par des relations sociales. Le réseau ainsi formé par des nœuds (individus ou organisations) et les liens entre ces nœuds (relations sociales) est appelé réseau social.

Dès le début de l'Internet et des courriers électroniques, les réseaux sociaux en ligne ont commencé à se former. Sur une petite échelle, le graphe formé par la chaîne d'utilisateurs qui échangent des courriels forme un réseau social. Dans les années récentes, les sites de réseaux sociaux ont gagné une grande popularité. Ces sites web créent des communautés virtuelles pour les personnes intéressées par un sujet particulier ou simplement un milieu pour

communiquer ensemble et partager des photos, fichiers etc. La communication se fait par causettes « chat » ou vidéo, ou messages instantanés, ou blogs, ou vidéos conférences, etc. MySpace a été lancé en 2002, suivi de Facebook en 2004, puis de Tagged et TagWorld et d'autres comme Flickr (<http://flickr.com/>) et YouTube (<http://youtube.com/>). Ces sites ont changé la façon d'établir des relations sociales entre les individus.

Une forte similarité existe entre le web comme réseau et les réseaux sociaux. La vaste structure des liens observée sur le web a été prise en compte pour la recherche d'informations. L'algorithme HITS de Kleinberg et le fameux algorithme PageRank sont principalement basés sur l'analyse des liens sur le web. L'utilisation de la technologie des agents associée avec les réseaux sociaux a aussi été explorée pour les services web. « Referral Web » (Kautz, Selman et al. 1997) est un système multi-agents qui exploite les réseaux sociaux sur le web et fournit des chaînes de référence entre les utilisateurs et les experts dans des sujets variés. « NetExpert » est un système à base d'agents qui localise les experts dans les communautés en exploitant les réseaux de connaissance (Sangesa and Pujol 2001). « Collaboratory » est un système multi-agents qui aide les chercheurs académiques qui partagent le même intérêt à collaborer dans un projet commun (Vázquez, Barrio et al. 2001). Le système apprend le profil des utilisateurs et fournit des documents de recommandation appropriés, trouve des experts et cherche des fonctionnalités pour ses utilisateurs.

5.1.2 Folksonomie et systèmes d'étiquetage collaboratifs

Une nouvelle génération d'applications a émergé avec le « Web 2.0 » comme les wikis, les blogs, les « podcasts » et l'évolution des interactions sociales et les systèmes de partage des ressources entre les différents utilisateurs. Cette forme révolutionnaire du web permet aux utilisateurs de contribuer de façon intensive au contenu du web. Nous nous intéressons à un système particulier de partage de ressources qui est l'étiquetage collaboratif et la folksonomie ou « collaborative tagging systems ».

Thomas Vander Wal (VanderWal 2005) a introduit le terme « folksonomy » ou, en adaptation française, folksonomie. Le mot est composé des deux mots « folks » ou gens et taxonomie. Ce sont les structures conceptuelles qui sont créées par les gens eux-mêmes.

L'organisation ou la classification du contenu du web par les utilisateurs émerge de façon décentralisée. Ce concept est complété par le système d'étiquetage qui exprime le fait d'ajouter une étiquette ou mot-clé à une ressource de l'Internet. La ressource peut être une page web, un blog, une photo ou un podcast.

Les utilisateurs choisissent librement leurs propres mots pour décrire leurs ressources web favorites. Il en résulte un schéma de classification sociale émergente créé par les utilisateurs eux-mêmes. Cela est catégoriquement différent de la classification traditionnelle hiérarchique. Les termes et mots-clés existent dans un espace plat où les relations parents-enfants n'existent plus. Il faut aussi noter la présence sociale des utilisateurs qui savent que leurs étiquettes sont accessibles à partager avec les autres utilisateurs, ce qui crée une plus grande motivation pour contribuer à ces systèmes.

Golder et Huberman ont effectué une étude formelle très significative des systèmes d'étiquetage collaboratifs (Golder and Huberman 2006). Ils ont analysé la structure de ces systèmes ainsi que les dynamiques de l'information. Ils ont montré comment les étiquettes changent avec le temps mais surtout comment elles se stabilisent pour une ressource donnée. Ils ont aussi montré que l'étiquetage est fondamentalement basé sur le « faire sens ». Le « faire sens » est le processus dans lequel l'information est catégorisée et étiquetée et, à partir de ce processus, le sens et les sémantiques émergent (Weick, Sutcliffe et al. 2005). Le « faire sens » est aussi influencé par les facteurs sociaux. Les sociétés sont capables d'organiser collectivement leurs connaissances et de coordonner leurs actions. Les systèmes d'étiquetage collaboratifs peuvent résoudre les problèmes associés au flou des frontières linguistiques et cognitives en partageant et organisant les ressources de l'information (Golder and Huberman 2006). Les systèmes d'étiquetage actuels ne traitent pas le problème de catégorisation de l'information dans toutes ses dimensions. Les informations sont juste catégorisées par les étiquettes fournies par les utilisateurs mais il n'existe pas une catégorisation en profondeur caractérisée par une classification hiérarchique.

5.1.3 Quelques exemples de systèmes d'étiquetage collaboratifs

Les systèmes d'étiquetage actuels continuent à croître à un taux très rapide. Par exemple, Del.icio.us qui est l'un des plus populaires, contient à présent une centaine de millions de ressources et d'étiquettes associées et plusieurs millions d'utilisateurs. Dans les années récentes, les systèmes d'étiquetage tels que Del.icio.us (<http://del.icio.us/>), Flickr (<http://flickr.com/>), Technorati (<http://technorati.com/>) , Citeulike (<http://www.citeulike.org/>), Bibsonomy (<http://bibsonomy.org/>) et d'autres ont acquis une grande popularité. Dans ces systèmes, les utilisateurs organisent leurs signets ou page web tout en utilisant leur propre vocabulaire ou étiquettes.

Del.icio.us: Del.icio.us est l'un des systèmes les plus populaires avec plus de trois millions d'utilisateurs inscrits. Créé en 2003 par Joshua Schachter qui décrit del.icio.us comme étant « social bookmarking manager ». Les utilisateurs peuvent étiqueter des ressources ou URL:

« *del.icio.us* is a **social bookmarking** website -- the primary use of del.icio.us is to store your bookmarks online, which allows you to access the same bookmarks from any computer and add bookmarks from anywhere, too. On del.icio.us, you can use **tags** to organize and remember your bookmarks, which is a much more flexible system than folders.

You can also use del.icio.us to see the interesting links that your friends and other people bookmark, and share links with them in return. You can even browse and search del.icio.us to discover the cool and useful bookmarks that everyone else has saved -- which is made easy with tags. [<http://del.icio.us/about/>] »

Del.icio.us est considéré social parce que chaque utilisateur peut voir toutes les étiquettes et signets de tous les autres utilisateurs. On peut accéder à la page personnelle de chaque utilisateur et la filtrer par étiquette, ce qui nous apprend ce que les autres utilisateurs trouvent intéressants.

Citeulike : CiteULike est un service qui permet aux chercheurs académiques de sauvegarder, partager, et organiser les ressources et articles scolaires. Les utilisateurs peuvent ajouter les ressources en conservant le lien URL de l'article des bibliothèques numériques telles que ACM ou IEEE etc. L'utilité première de Citeulike est d'organiser les références ; il offre, en plus, la possibilité de sauvegarder l'information Bibtex et les auteurs de l'article.

Bibsonomy : Bibsonomy est similaire en contexte à Citeulike, mais il permet en plus de construire une relation entre les étiquettes, pour renforcer une certaine hiérarchie et classification explicite des différentes étiquettes.

Connotea : Connotea est un système d'étiquetage collaboratif destiné principalement aux scientifiques avec des outils spécifiques pour exporter et gérer les références des articles scientifiques. Connotea a évolué pour devenir un outil d'usage public non réservé exclusivement aux scientifiques.

Flickr : Flickr est de nature différente des systèmes précédents, car les ressources dans Flickr sont les photos personnelles de chaque utilisateur. C'est un système de partage des photos qui permet aux utilisateurs de sauvegarder, étiqueter et partager leurs photos avec les autres utilisateurs.

Il existe une multitude d'autres systèmes d'étiquetage collaboratifs comme Dogear, MovieLens, Yahoo ! MyWeb2.0, YouTube, ESP Game, Last.fm, Yahoo ! Podcasts, Odeo, Technorati, LiveJournal, etc...

5.1.4 Principes fondateurs des systèmes d'étiquetage collaboratifs

Ces systèmes sont basés sur les principes suivants qui sont bénéfiques à l'utilisateur aux niveaux personnel et social :

- Au niveau personnel : ils fournissent à l'utilisateur une façon d'organiser et d'accéder facilement à ses propres ressources à partir de n'importe quelle machine connectée à l'Internet.
- Au niveau social : La possibilité de partager les étiquettes avec d'autres utilisateurs, ce qui permet la création de réseaux sociaux. Ces systèmes poussent leurs utilisateurs à construire des communautés sociales entre eux, recommandant des liens aux pages des autres utilisateurs qui partagent le même intérêt.

Du fait de l'évolution rapide de l'Internet, les utilisateurs ont un accès facile à de grandes masses d'informations numériques. Le besoin d'exercer un genre de contrôle sur la masse

d'information numérique qu'on accumule chaque jour et de pouvoir bien l'organiser de façon intelligente et facile à accéder a contribué à la popularité croissante des systèmes d'étiquetage. Ces systèmes permettent de décrire et d'organiser cette information numérique en utilisant des catégories et terminologies qui reflètent les besoins et vues des utilisateurs eux-mêmes, au lieu des organisations et corps externes comme c'est le cas présentement. C'est donc le vocabulaire de l'utilisateur qui décrit la ressource en question. Par conséquent, l'utilisateur va trouver une ressource donnée plus facilement en la recherchant par étiquettes qui dérivent de son propre vocabulaire et non d'un système de catégorisation/classification rigide.

En bref, l'aspect social et l'interaction sociale qui en résulte permettent de partager les ressources et les étiquettes avec d'autres utilisateurs ayant le même intérêt, ce qui offre deux avantages majeurs aux utilisateurs : la découverte de connaissances et la recherche de l'information. Ceci explique la popularité croissante des systèmes d'étiquetage collaboratifs (Furnas, Fake et al. 2006) : la découverte de connaissances dans le sens de l'utilisation d'une étiquette comme une ancre pour trouver les ressources intéressantes. Les étiquettes sont utiles pour la recherche d'information, dans le sens de trouver les ressources, ce qui devient plus facile. Un ensemble de ressources étiquetées améliore les métadonnées pour tous les utilisateurs, tout en partageant la distribution du travail parmi plusieurs contributeurs.

Ces systèmes ont le potentiel d'améliorer la Recherche d'Informations sur le web en complétant les moteurs de recherche et les systèmes de recommandation. Ils peuvent aussi être utilisés pour le Filtrage d'Informations. Ces potentiels d'améliorations sont profondément exploités dans le cadre de ce travail de thèse.

5.2 Étude des systèmes d'étiquetage collaboratifs comme réseaux complexes

Les systèmes d'étiquetage collaboratifs, tout comme le web, possèdent les propriétés des systèmes complexes ayant un grand nombre de composantes, interagissant les unes avec les autres de façon dynamique et non linéaire (Rupert, Rattrout et al. 2008). Ces systèmes exhibent des propriétés d'auto-organisation et d'émergence de nouvelles dynamiques. Il n'existe pas de contrôle global sur l'évolution de ces systèmes. Cette évolution est gouvernée par les comportements individuels des différentes composantes. Les systèmes d'étiquetage

possèdent plusieurs caractéristiques des systèmes complexes, comme le grand nombre d'utilisateurs interagissant les uns avec les autres, le manque de coordination centrale, les dynamiques non-linéaires, les cycles de feedback et l'auto-organisation.

Les systèmes d'étiquetage collaboratifs peuvent être représentés par des réseaux tripartites où les nœuds sont formés par les utilisateurs, les étiquettes et les ressources. Ces trois articles forment un triplet appelé « étiquette-application », qui constitue l'unité fondamentale d'information dans ces systèmes (Farooq, Kannampallil et al. 2007). Par conséquent, il existe trois espaces séparés, donc chacun est formé d'un ensemble de nœuds liés par des arcs (Halpin, Robu et al. 2007). L'espace des utilisateurs qui contient tous les utilisateurs, l'espace des étiquettes qui contient toutes les étiquettes, et l'espace des ressources qui contient toutes les ressources. L'arc qui lie un utilisateur à une ressource à une étiquette est appelé instance d'étiquetage.

Comme ces systèmes sont relativement récents, il n'existe dans la littérature que peu d'études sur la nature de ces réseaux. Les études suivantes (Shen and Wu 2005; Cattuto, Schmitz et al. 2007; Halpin, Robu et al. 2007; Mislove, Marcon et al. 2007) ont montré que les systèmes de folksonomie ou d'étiquetage collaboratif exhibent des comportements similaires aux systèmes complexes et possèdent la propriété d'invariance d'échelle ou « scale free » et la propriété petit monde ou « small world » relatives aux réseaux complexes.

5.2.1 La propriété « petit monde » des réseaux de folksonomie

Pour les réseaux de folksonomie, (Cattuto, Schmitz et al. 2007) ont étudié des ensembles de données extraites de Del.icio.us et de Bibsonomy. Pour ces deux réseaux, le coefficient d'agrégation (clustering) est extrêmement élevé. Pour del.icio.us, la connectivité du graphe (ou la distance moyenne entre une paire quelconque de nœuds) est de 3.5. Donc à partir d'une page de del.icio.us, on peut en moyenne par 3.7 liens ou clicks, aboutir à chaque utilisateur, étiquette ou ressource. Pour Bibsonomy, qui est un réseau relativement plus petit et plus jeune que Del.icio.us, la connectivité du graphe est de l'ordre de 3.5.

5.2.2 La Propriété des Réseaux Invariant d'échelle dans les réseaux de folksonomie

Pour les propriétés « scale free » et loi de puissance des réseaux sociaux, ces mêmes études (Mislove, Marcon et al. 2007) ont démontré que la distribution des degrés des nœuds dans les réseaux collaboratifs sociaux suit la loi de puissance $P(k) = k^{-\lambda}$, mais ce qui les différencie des autres réseaux complexes et particulièrement du web, c'est le fait que les coefficients d'agrégation « indegree » et « outdegree » sont presque similaires. (Halpin, Robu et al. 2007) ont aussi démontré que les distributions dans les systèmes d'étiquetage suivent la loi de puissance. (Shen and Wu 2005) ont calculé effectivement l'exposant λ de la loi de puissance qui était de l'ordre de 1.418. Les données pour leur test étaient extraites de Del.icio.us.

5.2.3 Les limites dans les systèmes d'étiquetage actuels

Vu l'absence de classification hiérarchique dans ces systèmes, les problèmes de vocabulaire, de sémantique, deviennent de plus en plus persistants. Il n'existe pas de classification hiérarchique dans ces systèmes et la classification d'informations dans ces systèmes souffre d'une inconsistance entre les différentes communautés (dans leur vue du sens d'une étiquette) et une inconsistance dans l'usage d'un mot (quel mot ou étiquette correct doit être utilisé lors de la recherche de ressources) (Choy and Lui 2006). Les utilisateurs n'utilisent pas les étiquettes de façon consistante, par exemple ils peuvent utiliser une étiquette aujourd'hui pour une certaine ressource et utiliser dans le futur une étiquette différente pour cette même ressource, car leurs vocabulaires et sémantiques évoluent au cours du temps. En faisant de la recherche par étiquette pour certaines ressources, le chercheur doit être en accord avec le point de vue du fournisseur de l'étiquette sur les sémantiques de la ressource.

D'autres problèmes ont émergé dans les systèmes d'étiquetage collaboratifs, Nous citons ci-dessous les plus courants :

- Le problème d'imprécision dû à l'absence de classification formelle et de taxonomie, comme une ressource peut avoir un nombre d'étiquettes associées sans aucune relation sémantique réelle entre elles.

- Ambiguïté : Les termes dans une folksonomie peuvent avoir une ambiguïté due au fait que différents utilisateurs associent des termes aux ressources de manière différente. Exemple : L'étiquette MAS réfère à « Multi-Agent System » en informatique, ou en sciences « Master of Applied Science ». C'est un acronyme utilisé pour au moins une dizaine d'autres termes dans des domaines totalement différents.
- La polysémie : L'étiquette polysémique « souris » peut se référer à l'animal souris comme il peut se référer à la souris d'un ordinateur.
- Synonymes : les termes “mac”, “macintosh”, et “apple” sont utilisés pour décrire les ordinateurs Apple Macintosh. Les termes singuliers et pluriels (photo, photos) apparaissent de façon redondante.
- Spécificité : Les termes reliés qui décrivent un objet particulier peuvent varier du plus général au plus spécifique, dépendamment des connaissances et de la culture de l'utilisateur. Les documents étiquetés “MySQL” et “ajax” peuvent être très spécifiques pour certains utilisateurs, tandis qu'une étiquette « programmation » peut être très générale pour d'autres.
- Syntaxe : Il n'existe pas de directives sur la façon d'utiliser les ponctuations, les traits d'union, les mots composés, etc. . Laquelle des deux étiquettes « database design » ou « design database » est la plus correcte à utiliser?
- Usage incorrect : Les étiquettes peuvent être utilisées incorrectement; le terme “archéologie”, par exemple, peut être utilisé pour étiqueter à la fois des ressources pertinentes aux dinosaures et aux microbes primitifs.

The screenshot shows the Delicious website interface. At the top, there is a navigation bar with 'delicious' and links for 'Home', 'Bookmarks', 'People', and 'Tags'. A search bar contains the word 'design' and a 'Search' button. Below the search bar, the text reads 'Searching Everybody's bookmarks for: design'. A secondary search bar also contains 'design' and a 'Search' button. A link says 'Sign in to search your own bookmarks'. Below that, it says 'See all bookmarks tagged design' and 'Search all of Delicious for "design"'. A blue bar indicates 'Everybody's bookmarks' with '1963237 results - show all'. The main content is a list of bookmarked items, each with a title, a 'SAVE' button, and a count in a blue box. The items are: COLOURlovers (35179), kuler (31416), Smashing Magazine (26449), A List Apart (26399), Open Source Web Design (23396), deviantART (29744), Web Developer's Handbook (29538), FFFFOUND! (20918), css Zen Garden (24492), and LogoPond (16682). Each item also shows 'First saved by' and a list of related tags. At the bottom, there is a section for 'Everyone's Related Tags' showing 'design', 'webdesign', 'css', 'inspiration', and 'web'. A footer contains links for 'delicious', 'about', 'blog', 'terms of service', 'privacy policy', 'copyright policy', 'forums', and 'support', along with a 'What's new?' button.

Figure 5.1 Le nombre des ressources étiquetées par le tag « design » est aux alentours de 2 millions de ressources.

D'autre part, et du point de vue social, il existe un problème de manque d'interaction sociale entre les utilisateurs dans ces systèmes. Ces derniers ne permettent pas l'émergence

automatique de communautés ou groupes sociaux qui peuvent bénéficier au maximum de ces réseaux sociaux.

A cause de ces problèmes, la recherche d'information devient de plus en plus difficile, surtout qu'une étiquette peut être associée à presque 2 millions de ressources. Par exemple, en recherchant l'étiquette « design », Del.icio.us retourne presque 2 millions de ressources (Fig.5.1) Ces ressources peuvent être de nature différente et sans aucune relation sémantique ayant un sens pour l'utilisateur. Il est donc nécessaire d'améliorer ces systèmes en introduisant une sorte de classification hiérarchique sémantique pour pouvoir extraire les informations nécessaires à partir des métadonnées créées par les utilisateurs eux-mêmes.

5.3 Proposition d'un système qui répond à certaines limites dans les systèmes existants

Le système que nous proposons (Rupert, Li et al. 2008) répond à certaines des limites constatées dans les systèmes d'étiquetage actuels. Notre système apporte une réponse aux questions suivantes : la classification hiérarchique, la consistance des étiquettes, la prise en compte de l'aspect social entre les utilisateurs, la personnalisation des recommandations pour la recherche de ressources et enfin l'évolution du système d'étiquetage collaboratif comme système complexe. Nous citons ci-dessous ces réponses:

- La classification hiérarchique : En faisant une recherche par étiquette par exemple « Design », notre système fait un clustering des résultats de recherche, qui affecte les ressources obtenues par cette recherche à différentes sous-communautés de ressources (« Design_Web », « Design_Home », « Design_Database », etc..
- La classification des utilisateurs en groupes basés sur leurs intérêts communs : les utilisateurs qui ont étiqueté le plus de ressources avec les étiquettes « Design_Web », vont être groupés dans le groupe des utilisateurs « Design_Web ».
- Recommandations personnalisées des ressources : En analysant les « *personomies* » des utilisateurs appartenant à un même groupe, le système va

classer ces utilisateurs suivant le degré de similarité entre eux. Cette similarité va jouer un rôle important dans l'évolution du système. Le système va recommander à un utilisateur donné U_x les ressources qui ont été étiquetées par les utilisateurs qui lui sont fortement similaires.

- Consistance dans le vocabulaire des étiquettes : Le système va aussi recommander à U_x des étiquettes-systèmes correspondant à ces ressources. Ceci va créer un renforcement de ces ressources et de leurs étiquettes. Ce qui va affecter la structure des sous-communautés originelles de ressources.
- L'évolution du système dans un cadre permettant le couplage entre l'organisation sociale et l'organisation spatiale du système et les effets rétroactifs de l'une sur l'autre. Ce comportement peut être qualifié d' « enactif » selon la théorie de l'enaction de Varela, selon laquelle, la cognition est une caractéristique fondamentale des organismes vivants dans une relation dynamique et adaptative à leur environnement. *«Perception consists in perceptually guided action...» "...we must see the organism and environment as bound together in reciprocal specification and selection..."* *"Cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided."* (Varela, Thompson et al. 1991). La définition des comportements intelligents est structurée par l'environnement qui donne une importance particulière à la boucle perception-action. Les usages façonnent la structure du système qui à son tour façonne les usages futurs qui modifient cette structure et ainsi de suite.

5.3.1 Fonctionnalités du système MySURF

Le but de notre étude des systèmes collaboratifs d'étiquetage est de suggérer des méthodes qui améliorent les systèmes actuels en prenant en compte leur nature en tant que *systèmes complexes adaptatifs*.

La coévolution des organisations spatiales et sociales des agents a permis l'émergence de nouvelles propriétés. Les utilisateurs ou agents laissent une trace dans l'environnement

physique (étiquettes et renforcement des étiquettes). Comme résultat de ce comportement d'étiquetage, des sous-communautés de ressources émergent, qui à leur tour, affectent l'évolution des groupes virtuels d'utilisateurs et vice versa. C'est le mécanisme de stigmergie, où l'influence des effets persistants dans l'environnement de comportements antérieurs affecte les comportements futurs.

Ci-dessous, nous citons les principales fonctionnalités du système MySURF qui améliorent l'expérience de l'utilisateur (a priori nous faisons la comparaison avec les systèmes actuels et surtout avec Del.icio.us):

- L'amélioration de la recherche par étiquette en ayant une sorte de classification hiérarchique causée par la formation de sous-communautés de ressources. Chercher une étiquette comme « toread » avec l'option de grouper (cluster) les résultats va retourner dans notre système des sous-communautés comme “toread programming”, “toread webdesign”, etc. Chercher l'étiquette « toread » dans Del.icio.us fait retourner plus que 490k ressources dont certaines n'ont aucune relation sémantique entre elles (figure 6).
- L'avantage d'appartenir implicitement à un groupe virtuel d'utilisateur permet à chaque membre du groupe d'accéder et voir plus rapidement toute nouvelle ressource étiquetée par le groupe. Le membre aura aussi la possibilité de réagir avec d'autres utilisateurs qui partagent avec lui le même intérêt. La plupart des utilisateurs dans les systèmes d'étiquetage actuels ne rejoignent pas explicitement des groupes sociaux
- Un membre peut prendre l'initiative de devenir un leader du groupe en créant un groupe réel et envoyer des invitations aux membres du groupe virtuel pour rejoindre le nouveau groupe (construire des communautés et former des relations sociales devient plus facile quand le dirigeant du groupe a connaissance des membres potentiels).
- L'intégration de la *personomie* (personnalisation) au sein de l'application comme outil de calcul du taux de similarité entre deux utilisateurs du même

groupe. Chaque utilisateur peut se comparer à tous les autres utilisateurs et découvrir ceux qui partagent le plus ses intérêts.

- Cette similarité construit une sorte de profil d'utilisateurs. Ce qui permet l'intégration de systèmes de recommandation de pages web, car ces systèmes auront une meilleure compréhension des intérêts des utilisateurs et pourront recommander des signets plus spécifiques aux utilisateurs.
- La suggestion d'étiquettes-systèmes améliore le vocabulaire d'étiquettes utilisées par tous les utilisateurs de point de vue de consistance des termes utilisés.

Ces fonctionnalités ont émergé comme résultat du couplage et de la coévolution entre l'organisation sociale et l'organisation spatiale. Cet aspect auto- organisationnel au niveau spatial et social et l'effet rétroactif de l'un sur l'autre peut être exprimé dans notre application comme suit :

- L'évolution de l'organisation spatiale (sous-communautés de ressources) a un effet direct sur l'évolution de l'organisation sociale (création de nos groupes virtuels).
- Le renforcement des ressources par les utilisateurs a un effet direct sur l'organisation spatiale. Plus l'utilisateur laisse des traces, plus ces traces vont être renforcées par les autres utilisateurs. C'est un genre de phéromone déposée dans l'environnement.
- La présence sociale de l'individu affecte son comportement au sein de son réseau social (Lee 2006), ce qui le pousse à augmenter le nombre d'annotations et de contributions. C'est l'effet direct sur l'organisation sociale.

Notons que notre travail n'englobe pas les aspects sémantiques et cognitifs de la classification et les problèmes reliés à la taxonomie comme la polysémie, la synonymie et la variation des niveaux de base (Golder and Huberman 2006). La sous-communauté « toread programming » n'inclut pas les ressources étiquetées avec les mots "programming" (un seul « m ») ou "programmer" ou "programmacao" ou "programmation" bien qu'on reconnaît

bien que ces ressources appartiennent sémantiquement à la sous-communauté « toread programming ». En analysant les données que nous avons collectées de Del.icio.us, nous avons noté que le nombre total de ces ressources est relativement très petit, comparé aux ressources étiquetées avec l'étiquette « programming ».

5.3.2 Representation des organisations sociales et spatiales dans MySURF

Dans l'application d'étiquetage MySURF, les utilisateurs sont représentés par des agents engagés dans la réalisation d'une tâche collective. Ces agents existent dans un environnement physique matérialisé par un réseau complexe de ressources physiques qui est le système d'étiquetage lui-même comme défini dans (Hassas 2003). Ces agents ont des rôles, ils peuvent percevoir leur environnement sur lequel ils peuvent agir, en effectuant des tâches sur le réseau des ressources considéré, ce qui affecte l'évolution du système. Cette matérialisation physique permet la mise en œuvre du mécanisme de stigmergie, menant à l'auto-organisation.

Les agents communiquent indirectement les uns avec les autres en laissant des traces dans l'environnement sous forme d'étiquettes. Ces étiquettes peuvent être considérées comme un genre de phéromone, permettant au système de s'auto-structurer à travers les actions des agents sur l'environnement.

Ceci exprime l'influence exercée par les effets persistants dans l'environnement, des comportements passés des agents, sur leurs comportements futurs. Ces effets ont été classés dans (Holland et Melhuish 1999) en trois cas :

- un effet qualitatif : exprimant l'influence sur le choix de l'action à effectuer par un agent ;
- un effet quantitatif : exprimant l'influence sur les paramètres (comme la position, la force, la fréquence, la latence, la durée, etc.) de l'action, la nature de l'action restant, elle, inchangée ;
- un effet indirect qualitatif et / ou quantitatif : exprimant l'influence sur le résultat de l'action. Cette influence agit indirectement sur la manière dont sera

menée l'action et sur son résultat, du fait des changements opérés sur son environnement.

Dans MySURF, nous avons adopté l'esprit général du méta-modèle décrit au chapitre 3. Nous présentons dans ce qui suit l'association entre les différents éléments des organisations spatiales et sociales et les éléments de MySURF :

L'agent est équivalent à l'utilisateur, la place est équivalente à la ressource web, la région est équivalente à la sous-communauté de ressources, la phéromone est représentée par les étiquettes ayant un nom (label) et une fréquence par ressource (intensité). La position sociale est représentée par un sujet d'intérêt où les groupes sont associés aux groupes virtuels des utilisateurs. Vu la nature des systèmes d'étiquetage et l'objectif de l'application, il n'était pas nécessaire pour nous de représenter les contraintes sociales.

Considérons le cas d'ajouter une ressource à la librairie propre d'un agent. Si l'agent possède sa propre étiquette provenant de son propre vocabulaire et si la ressource n'existe pas déjà dans le système, cette propre étiquette va être ajoutée à la ressource. Mais si la ressource est déjà dans le système, possédant une multitude d'étiquettes dominantes, l'utilisateur va probablement choisir une de ces étiquettes dominantes. Les actions prises par les agents précédents vont affecter les actions que va prendre l'agent présent. La ressource sera renforcée par cette étiquette dominante, ce qui affectera les actions des agents futurs. C'est bien le cas d'une stigmergie passive. Par ce mécanisme de stigmergie, l'environnement s'auto-structure à travers les activités des agents dans l'environnement.

L'organisation spatiale sous forme de sous-communautés de ressources est affectée par ce changement, ce qui aboutit à une restructuration de l'environnement. Cette restructuration affectera à son tour la position spatiale de l'agent dans l'environnement physique matérialisé par le réseau des ressources. Cette position va à son tour affecter le choix de l'action à exécuter par l'agent. Comme conséquence, l'agent peut choisir de prendre un des différents rôles comme *Ressource Tagger* ou *Ressource Searcher* ou *Knowledge Expert* ou créateur d'une nouvelle communauté d'utilisateurs, etc...

Le résultat de cette action effectuée sur l'environnement aura un effet sur l'appartenance de l'agent aux différents groupes virtuels d'utilisateurs ce qui va laisser des traces dans l'environnement qui affectent à leur tour les actions futures de l'agent. Et le cycle continue, le

système évolue en se basant sur une rétroaction entre l'organisation spatiale et l'organisation sociale des agents.

5.4 Description détaillée du système d'étiquetage MySURF

Nous avons créé un système d'étiquetage collaboratif MySURF sous forme d'une application web qui permet aux utilisateurs de sauvegarder leurs signets en ajoutant des étiquettes à des ressources ou pages web. Le but de notre application est de développer, à partir des systèmes d'étiquetage collaboratif existants, un système plus performant. Pour atteindre ce but, nous nous sommes basés sur notre méthodologie d'organisations sociales et spatiales dans les systèmes multi-agents. Dans notre système, chaque étiquette est un mot-clé ou terme associé à une pièce d'information qui est une page web, dans notre cas. Les utilisateurs peuvent créer et gérer leurs ressources dans leurs bibliothèques personnelles et peuvent accéder aux ressources ajoutées par les autres utilisateurs.

Notre système d'étiquetage est une application web qui permet aux utilisateurs d'organiser et d'accéder à leurs signets en ajoutant des mots-clés ou étiquettes à des ressources web. Spatialement, les utilisateurs sont modélisés par des agents qui sont situés dans des places ou ressources web, ces emplacements se regroupent en régions ou sous-communautés de ressources. Socialement, ces agents peuvent avoir différents rôles, comme explorateurs de ressources, étiqueteurs de ressources et sont groupés selon leurs intérêts dans différents groupes virtuels d'utilisateurs.

Le système offre une interface pour les utilisateurs pour ajouter, étiqueter, rechercher et afficher les ressources. Le système regroupe les ressources sémantiquement proches l'une de l'autre dans des sous-communautés de ressources. A partir de ces sous-communautés de ressources, des étiquettes-systèmes sont suggérées aux utilisateurs quand ceux-ci essaient d'étiqueter des ressources déjà existantes dans le système. Pour chacune de ces sous-communautés de ressources, le système liste les étiquettes associées et un lien au groupe virtuel d'utilisateurs associés à ces ressources. Pour chaque groupe virtuel, le système affiche aux utilisateurs l'ensemble des ressources étiquetées pour ce groupe virtuel, l'ensemble de tous les utilisateurs appartenant au groupe, ainsi que l'ensemble des étiquettes appartenant à la

sous-communauté de ressources associées au groupe virtuel. En résumé, les principales tâches que les utilisateurs effectuent sont les suivantes (*en italique sont citées les tâches propres à notre système*) (Fig. 5.2):

- ajouter une ressource (page web) à leur librairie, associer une ou plusieurs étiquettes à cette ressource, supprimer une ressource de leur librairie, modifier les étiquettes ajoutées à une ressource de leur librairie, explorer les ressources ajoutées par les autres utilisateurs et faire de la recherche de ressource par étiquette
- *raffiner cette recherche en créant des clusters des ressources obtenues par cette recherche où une classification hiérarchique de l'information est créée.*
- *afficher des groupes virtuels d'utilisateurs partageant un intérêt commun*
- *afficher une liste d'utilisateurs qui lui sont fortement similaires basés sur leurs personnalités*
- *personnaliser les résultats des clusters de ressources en se basant sur les degrés de similarité entre les utilisateurs. Les utilisateurs se voient recommander les ressources étiquetées par les utilisateurs qui leurs sont le plus similaires.*

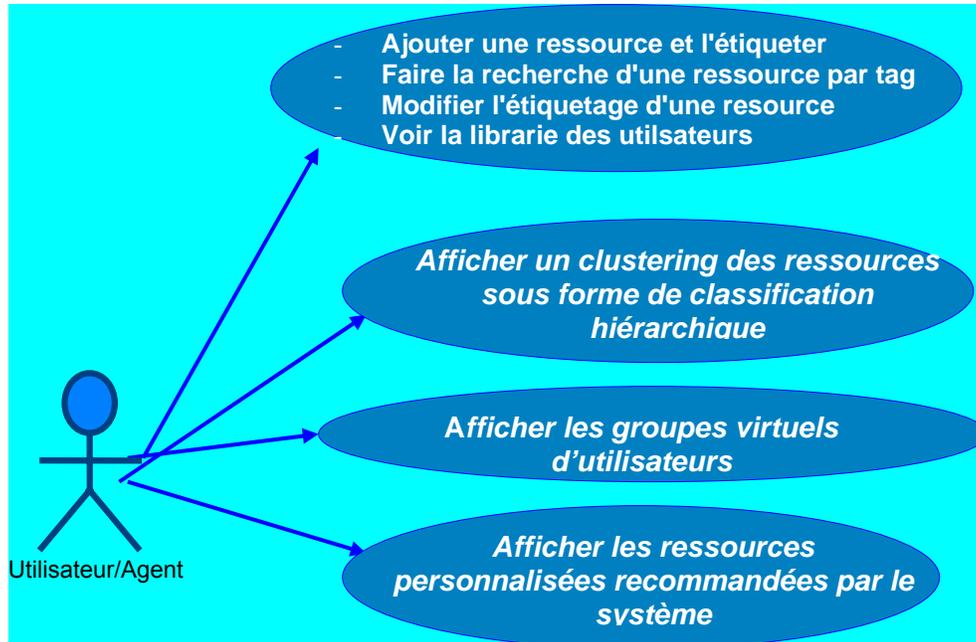


Figure 5.2 montre les spécifications du système en termes des différentes tâches effectuées par les utilisateurs

En réponse aux actions des utilisateurs, le système offre les fonctionnalités suivantes :

- Le système suggère des étiquettes aux ressources qui ont été déjà étiquetées par d'autres utilisateurs.
- Le système recommande des groupes virtuels à attribuer à certaines ressources quand les utilisateurs ajoutent des étiquettes à des ressources existantes dans le système
- Le système affiche une liste des groupes virtuels auxquels les utilisateurs appartiennent
- Le système affiche les sous-communautés de ressources après qu'une recherche s'est effectuée sur des ressources avec une étiquette spécifique.
- Le système affiche des nuages d'étiquettes associées aux sous-communautés de ressources
- Pour chaque groupe virtuel auquel appartient l'utilisateur, le système affiche une liste de ressources associées, une liste des utilisateurs qui appartiennent au

même groupe virtuel, et un nuage d'étiquettes représentant les étiquettes de la sous-communauté associée. De plus, le système montre aussi les ressources nouvellement ajoutées par d'autres utilisateurs du même groupe virtuel, même s'ils n'ont pas été classés dans la sous-communauté.

- Les ressources qui apparaissent dans un groupe virtuel, mais qui ne font pas partie de la sous-communauté associée, vont être ajoutées dans la sous-communauté, étant donné que les étiquettes suffisantes/substantielles qui sont appropriées à la sous-communauté ont été affectées par les utilisateurs du système.
- L'ensemble des groupes virtuels auxquels appartient un utilisateur évolue au cours du temps (un utilisateur peut appartenir à différents groupes virtuels à des temps différents) selon le comportement d'étiquetage de l'utilisateur.
- Le système affiche un tableau de degré de similarité entre les différents membres d'un même groupe.
- Le système affiche un raffinement des sous-communautés de ressources basé sur les degrés de similarité entre les utilisateurs calculé à partir des *personomies*.

5.4.1 Émergence des sous-communautés de ressources

Dans les systèmes d'étiquetage collaboratifs, un grand nombre d'utilisateurs autonomes ajoutent des étiquettes à des ressources sans aucun contrôle régissant leurs comportements. Cette activité non coordonnée mène à une structure de communauté émergente basée sur des relations sémantiques entre les ressources. Afin de détecter cette structure de communauté émergente, nous avons développé un algorithme basé sur les concepts d'algorithme spectral proposé dans (Ciro Cattuto 2007), pour extraire les sous-communautés de ressources.

5.4.1.1 Description de l'algorithme de clustering des ressources

Le réseau de ressources peut être représenté par un réseau valué « weighted network » introduisant la notion de similitude ou de poids w entre deux ressources R_1 et R_2 . La formule

qui définit le poids w est adoptée de la mesure “ $tf \times idf$ ” (pour le terme fréquence multiplié par la fréquence inverse d’un document), qui est un mécanisme standard utilisé dans la fouille de textes (Salton and Buckley 1987).

tf est la fréquence d’un terme dans un document ou le nombre d’occurrences de ce terme t_i dans le document considéré d_j .

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

où $n_{i,j}$ est le nombre d’occurrences du terme considéré dans un document d_j , et le dénominateur est le nombre d’occurrences de tous les termes du document d_j .

idf est la fréquence inverse d’un document qui mesure l’importance générale d’un terme obtenu, en prenant le logarithme du quotient du nombre total de documents par le nombre de documents qui contiennent le terme en question.

$$idf_i = \log \frac{|D|}{|\{d_j : t_i \in d_j\}|}$$

$|D|$: nombre total de documents dans le corpus

$$|\{d_j : t_i \in d_j\}| :$$

C’est le nombre de documents où le terme t_i apparaît

Il existe plusieurs variations de la formule « $tf \times idf$ ». Nous avons adopté celle utilisée par (Ciro Cattuto 2007). Pour deux ressources R_1 et R_2 , nous calculons le poids entre ces deux ressources comme suit:

$$w_{R_1, R_2} = \frac{\sum_{t \in T_1 \cap T_2} \frac{\min(f_t^1, f_t^2)}{f_t}}{\sum_{t \in T_1 \cap T_2} \frac{\max(f_t^1, f_t^2)}{f_t} + \sum_{t \in T_1 - T_2} \frac{f_t^1}{f_t} + \sum_{t \in T_1 - T_2} \frac{f_t^2}{f_t}}$$

Le numérateur est constitué de la somme des minimums des fréquences normalisées pour les étiquettes qui sont utilisées dans les deux ressources (intersection des ensembles T_1 et T_2).

où T_1 (respectivement T_2) est l'ensemble d'étiquettes associées avec R_1 (respectivement R_2), f_t^1 (respectivement f_t^2) est la fréquence de l'occurrence de l'étiquette t dans T_1 (respectivement T_2) et f_t est la fréquence totale de l'étiquette t ou bien le nombre total de fois où l'étiquette t a été utilisée dans toutes les ressources.

Le dénominateur contient:

la somme des fréquences normalisées des étiquettes utilisées pour la ressource R_1 seulement.

la somme des fréquences normalisées des étiquettes utilisées pour la ressource R_2 seulement pour les étiquettes qui sont utilisées dans les deux ressources (intersection des ensembles T_1 et T_2), on considère la somme des maximums des fréquences normalisées.

La matrice de similarité W formée peut être considérée comme la matrice d'adjacence d'un réseau complexe valué où il est possible d'assigner aux arcs du graphe un poids qui est proportionnel à l'intensité des connections entre les éléments du réseau.

Dans cette matrice, les ressources sont représentées aléatoirement et il n'existe pas de structure visible dans cette représentation.

Pour surmonter ce problème, nous devons faire des transformations appropriées à la matrice W .

En premier, nous compressons la matrice W en élevant tous ses éléments à une puissance arbitraire γ qui a été choisie $\gamma = 0.1$ dans l'étude (Cattuto, Schmitz et al. 2007).

$$W'_{R_1, R_2} = (W_{R_1, R_2})^{0.1}$$

Il reste à faire les permutations nécessaires de lignes et de colonnes de cette matrice afin de visualiser les différentes sous-communautés de ressources qui émergent de l'ensemble de toutes les ressources en relation avec un sujet particulier. Pour cela, nous transformons la matrice W'_{R_1, R_2} en une matrice Q comme suit :

$$Q = S - W \quad \text{où } W_{ij} = (w'_{R_1, R_2})$$

δ_{ij} est le coefficient de Kroenecker:

$$\delta_{ij} = 1 \text{ si } i = j \quad \delta_{ij} = 0 \text{ si } i \text{ est différent de } j \quad \text{et } S_{ij} = \delta_{ij} \sum_j W_{ij}$$

où S est une matrice diagonale dont chaque élément de la diagonale est égal à la somme des éléments correspondants de la ligne de W .

La matrice Q ressemble à la matrice Laplacienne de la théorie des graphes. La matrice Laplacienne d'un graphe G non-orienté de n sommets est la matrice symétrique L dont l'élément diagonal L_{ii} est le degré du sommet i , et dont l'élément diagonal L_{ij} est 1 si les sommets i et j sont connectés par un arc et zéro s'ils ne sont pas connectés.

Par définition de Q , la somme des éléments de chaque colonne de Q est nulle, ainsi que la somme des éléments de chaque ligne est nulle.

Q contient des valeurs propres positives avec au moins une valeur propre égale à zéro qui correspond au vecteur propre $1 = (1, 1, 1, \dots)$.

En étudiant les propriétés spectrales de la matrice Q , on peut déterminer le nombre de sous-communautés de ressources sémantiquement distinctes à partir du nombre des plus petites valeurs propres distinctes, bien séparées et non triviales.

La figure 3 montre les valeurs propres classées par ordre croissant de la matrice Q . La présence des 4 premières valeurs propres non nulles et bien séparées indique l'existence d'au moins 4 sous-communautés de ressources bien définies.

5.4.1.2 Le clustering spectral

Dans les années récentes, les algorithmes de clustering spectral sont devenus de plus en plus populaires comme techniques d'analyses exploratoires de données (vonLuxburg 2006).

Soit $G = (S, A)$ un graphe non-orienté avec S l'ensemble des sommets $S = \{s_1, \dots, s_n\}$. G est un graphe valué c'est-à-dire qu'il existe un poids w_{ij} non-négatif entre deux sommets s_i et s_j . La matrice d'adjacence du graphe est la matrice $W = (w_{ij}) \ i, j = 1, \dots, n$. Si $w_{ij} = 0$ alors les sommets s_i et s_j ne sont pas connectés. Comme G est non-orienté alors $w_{ij} = w_{ji}$.

Le degré d'un sommet s_i est défini comme :

$$d_i = \sum_{j=1}^n w_{ij}$$

La matrice des degrés D est définie comme la matrice diagonale avec les degrés d_1, \dots, d_n sur la diagonale.

La matrice Laplacienne L du graphe G (appelée aussi matrice d'admittance ou matrice de Kirchhoff) est définie comme étant la différence entre la matrice diagonale des degrés et la matrice d'adjacence W .

$$L = D - W$$

La matrice L possède plusieurs propriétés (Mohar 1991) :

La matrice L est symétrique, positive, semi-définie: ses valeurs propres

$$\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$$

sont des valeurs positives ou nulles.

λ_0 est toujours égale à zéro et elle correspond à un vecteur propre constant.

λ_1 est appelée la connectivité algébrique du graphe ou c'est la seconde plus petite valeur propre. Son intensité exprime la connectivité du graphe.

Les algorithmes de clustering spectral sont des techniques qui prennent en considération les valeurs propres et les vecteurs propres de la matrice de similarité pour partitionner le graphe en différentes communautés ou clusters. Ces techniques ont montré leur efficacité dans plusieurs domaines comme la recherche d'informations (Deerwester, Dumais et al. 1990), web search (Kleinberg 1998; Page, Brin et al. 1998), segmentation des images et le clustering de données (Meila and Shi 2001; Ng, Jordan et al. 2002)

Il existe plusieurs variations de ces algorithmes. Nous décrivons celui que nous avons adopté dans notre application.

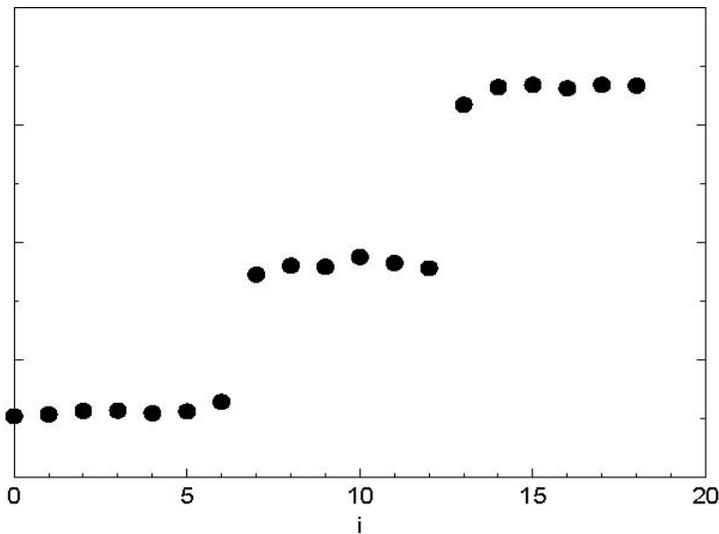


Figure 5.3 Le réseau utilisé comme exemple, avec $S = 19$ et d'une valeur aléatoire entre 1 et 10 attribuée aux liens. Trois clusters clairs apparaissent, composés par les nœuds 0 – 6, 7 -12 et 13 – 19 (Capocci, Servedio et al. 2004).

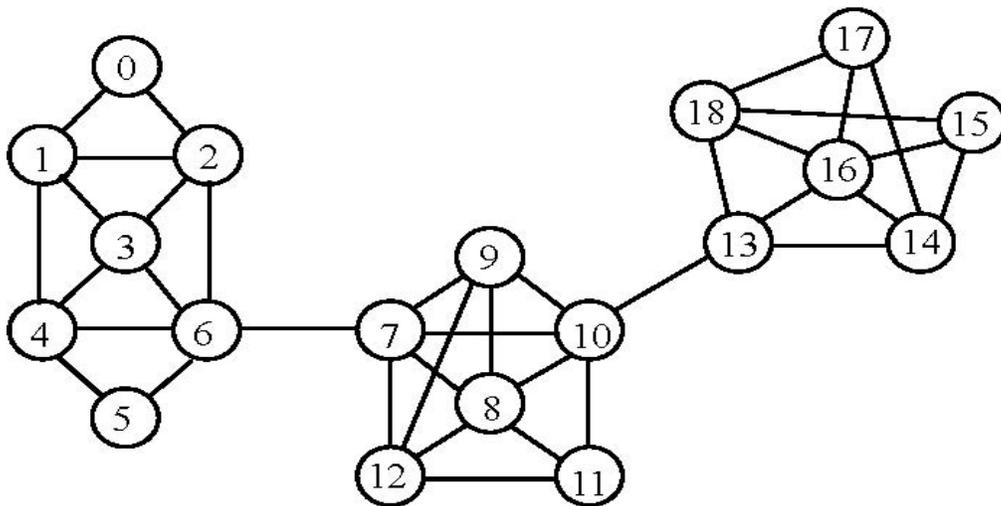


Figure 5.4 Les valeurs du 2e vecteur propre relatives au graphe représenté sur la figure 5.3.

On considère les premières plus petites valeurs propres de la matrice Laplacienne L . Le nombre de ces valeurs propres bien séparées peut nous donner une indication du nombre possible de communautés qui peuvent émerger. L'étude des premiers vecteurs propres qui correspondent à ces valeurs propres révèle la structure de ces communautés. Si on considère l'exemple suivant, pris de (Capocci, Servedio et al. 2004), qui montre le profil des composantes du second vecteur propre appelé vecteur Fiedler dans le cas d'un graphe simple $S=19$ nœuds, avec des poids aléatoires entre 1 et 10 affectés aux arcs (Fig 5.3 et Fig. 5.4). Les composantes de ce vecteur propre correspondent aux différents nœuds et le groupement de ces composantes indique les nœuds qui appartiennent à la même communauté. Il est clair d'après cette partition que le nombre de clusters émerge naturellement.

5.4.1.3 Partitionnement du graphe à partir des vecteurs propres de la matrice Q

Malheureusement, dans le cas des larges réseaux, qui n'ont pas de partitionnement clair, la valeur précise des composantes des premiers vecteurs propres n'est pas toujours efficace pour déterminer les différents partitionnements.

Si on considère le vecteur Fiedler de la matrice Q (ou son premier vecteur propre) dans le cas des données qu'on a extraites de Del.icio.us., ce vecteur ne peut pas nous donner une indication précise du nombre de sous-communautés de ressources qui peuvent émerger.

Les figures 5.5 et 5.6 sont les représentations des vecteurs Fiedler de la matrice Q correspondant, la matrice Q correspondant à 92 ressources liées à l'étiquette « design » extraites de Del.icio.us et 200 ressources liées à l'étiquette « programming ».

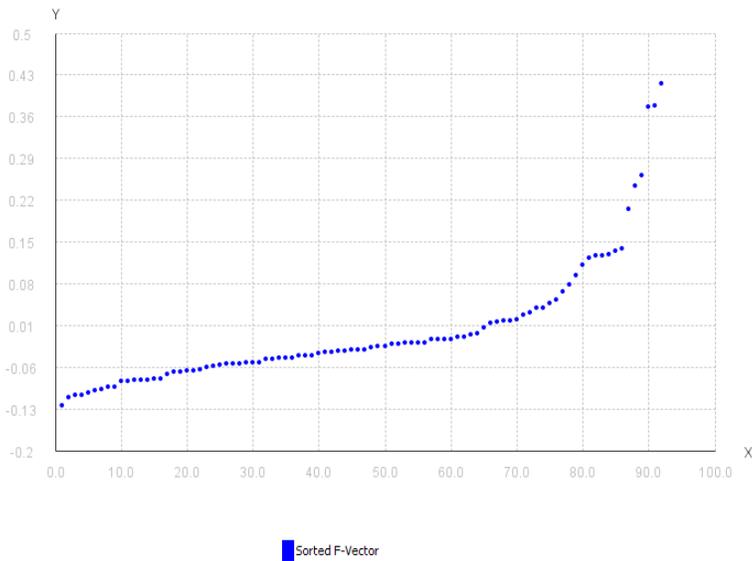


Figure 5.5 Le vecteur Fiedler de la matrice Q calculée pour 92 ressources liées à l'étiquette « design » et extraites de del.icio.us .

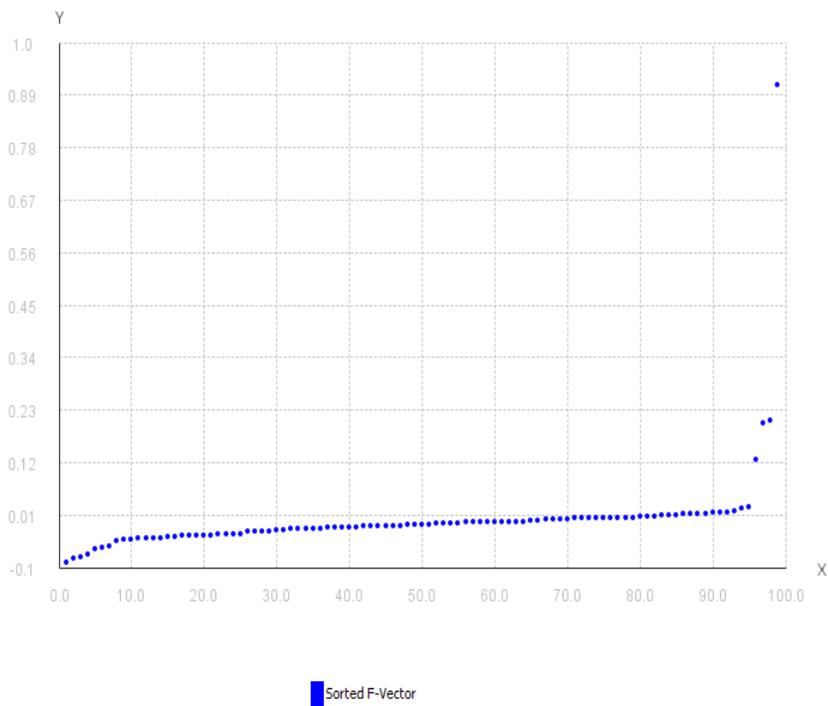


Figure 5.6 Le vecteur Fiedler de la matrice Q calculée pour 200 ressources liées à l'étiquette « programming » et extraites de del.icio.us .

Pour ces raisons, nous avons cherché d'autres méthodes pour partitionner le graphe à partir de la matrice Laplacienne obtenue. Nous avons adopté l'approche introduite dans (Capocci, Servedio et al. 2004) pour détecter le nombre de clusters à partir des vecteurs propres de la matrice Laplacienne.

L'approche consiste à étudier la corrélation entre deux nœuds, étant donné que deux ressources appartenant à la même communauté seront fortement corrélées entre elles.

Nous avons calculé la matrice de corrélation C_{ij} entre deux nœuds i et j (voir Annexe) en se basant sur la formule suivante :

$$C_{ij} = \frac{\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle}{\sqrt{[(\langle x_i^2 \rangle - \langle x_i \rangle^2)(\langle x_j^2 \rangle - \langle x_j \rangle^2)]}}$$

où x_i et x_j sont les composantes des quelques premiers vecteurs propres non triviaux, la notation $\langle \cdot \rangle$ représente la moyenne de ces composantes. Le coefficient de corrélation c_{ij} mesure la proximité entre deux nœuds i et j .

A partir de l'analyse de cette matrice, nous avons pu extraire des clusters de ressources comme montré ci-dessous dans la section 4.

En cherchant les ressources liées à une étiquette particulière, notre système fait un clustering de ces ressources en les divisant en sous-communautés. Par exemple, si l'utilisateur cherche les ressources liées à l'étiquette "programming", le système va diviser les ressources en sous-communautés de la façon suivante : "web programming", "java programming", "ajax programming", etc., tandis que tous les systèmes existants listent toutes les ressources qui ont été étiquetées par l'étiquette "programming". Ceci améliore considérablement les résultats de recherche de ressources basées sur une étiquette donnée.

5.4.2 Les étiquettes-systèmes

Une fois que le système applique l'algorithme de clustering décrit ci-dessus pour un ensemble particulier de ressources, ces ressources seront affectées à des étiquettes-systèmes composés d'une combinaison de sujet/sous-sujet (ex: "web" "design", ou "home" et "design").

Les étiquettes-systèmes sont auto- suggérées aux nouveaux utilisateurs qui sont sur le point d'étiqueter une ressource qui est déjà affectée à des étiquettes-systèmes. Cette application d'étiquette suggérée est visualisée comme une phéromone utilisée pour renforcer les traces laissées par les agents qui rencontrent une ressource spécifique.

5.4.3 Les groupes virtuels d'utilisateurs

Un groupe virtuel est un ensemble d'utilisateurs qui partagent le même intérêt dans un sujet spécifique. Les utilisateurs sont groupés dans des groupes virtuels en se basant sur leur historique d'étiquetage. Les ressources nouvellement ajoutées sont catégorisées dans des groupes virtuels basés sur le comportement pré-étiquetage de l'utilisateur. Ceci fournit un avantage aux utilisateurs qui, à l'avance, prennent conscience du réseau social et peuvent choisir de réagir avec le réseau et ajouter plus d'étiquettes et de ressources pour les partager parmi les groupes virtuels. Pour chaque sous-communauté de ressources, les utilisateurs correspondants qui ont étiqueté ces ressources, seront groupés dans des « groupes virtuels » basés sur le sujet de la sous-communauté de ressources. Exemple : tous les utilisateurs intéressés par des ressources relatives au « Web Design » appartiennent au même groupe virtuel « Web Design ».

5.4.3.1 Personomies pour chaque utilisateur

La *personomie* est définie comme étant l'ensemble des ressources et étiquettes associées à un utilisateur particulier. Formellement la *personomie* est définie comme suit (Hotho, Jaschke et al. 2006) :

Définition 1 : une folksonomie F est un tuple $F = (U, T, D, A)$ où U est l'ensemble des utilisateurs, T est l'ensemble des étiquettes, D est l'ensemble des documents web, et $A \subseteq U \times T \times D$ est l'ensemble des annotations.

Définition 2 : La *personomie* P_u d'un utilisateur u est la restriction de la folksonomie F à u : $P_u = (T_u, D_u, A_u)$, où :

A_u est l'ensemble des annotations de l'utilisateur : $A_u = \{(t, d) \mid (u, t, d) \in A\}$

T_u est l'ensemble des étiquettes de l'utilisateur : $T_u = \{t \mid (t, d) \in A_u\}$

D_u est l'ensemble des documents étiquetés par l'utilisateur : $D_u = \{d \mid (t, d) \in A_u\}$

Pour chaque groupe virtuel d'utilisateurs, nous proposons d'extraire les *personomies* de tous les utilisateurs qui appartiennent à ce groupe pour pouvoir déterminer s'il existe des similarités plus fortes entre certains membres de ce groupe.

Pour chaque utilisateur u d'un groupe virtuel, nous avons les données de la *personomie* $P_u = (T_u, D_u, A_u)$.

5.4.3.2 Analyse du degré de similarité entre 2 utilisateurs

Le but de cette analyse est de pouvoir déterminer si 2 utilisateurs donnés sont fortement similaires ou faiblement similaires.

Définition : Deux utilisateurs sont dit **fortement similaires** s'ils étiquettent de la même façon plusieurs ressources sémantiquement non similaires. Par exemple si 2 utilisateurs étiquettent de la même façon (utilisant des étiquettes similaires) des ressources liées à la programmation, il est intéressant d'analyser comment ces deux utilisateurs ont étiquetées des ressources liées à la cuisine ou aux jeux olympiques par exemple.

Si deux utilisateurs appartenant à un même groupe virtuel sont fortement similaires, les ressources ajoutées par l'un des utilisateurs seront suggérées à l'autre utilisateur et vice versa. Dans ce cas, les résultats de recherche par tag seront personnalisés selon les intérêts de chaque utilisateur.

5.4.3.3 Raffinement des groupes virtuels d'utilisateurs et des sous-communautés de ressources

Dans chaque groupe virtuel d'utilisateurs, nous allons avoir un cluster d'utilisateurs ayant un haut degré de similarité les uns avec les autres, tandis que d'autres utilisateurs dans ce même groupe auront un faible degré de similarité. Ceci aboutit à la création de plusieurs clusters dans le même groupe. Il est important que chaque utilisateur ait connaissance des utilisateurs qui lui sont fortement similaires et par conséquent il aura un accès direct aux ressources que ces utilisateurs ajoutent à leurs bibliothèques.

Ceci se manifeste dans les sous- communautés de ressources dans l'information qui va être affichée à cet utilisateur. Les sous-communautés de ressources vont à leur tour être personnalisées et raffinées selon les intérêts de l'utilisateur.

Vu cet accès direct et rapide à ces ressources qui ont été étiquetées par des utilisateurs avec un haut degré de similarité avec un utilisateur particulier, cet utilisateur particulier va avoir la tendance à étiqueter lui-même ces mêmes ressources. Cela va créer un renforcement de ces ressources dans la communauté tout à fait semblable à la phéromone.

5.4.3.4 Algorithme de l'émergence des étiquettes-systèmes

A partir de ces clusters, nous avons analysé l'histoire d'étiquetage pour chacune de ces ressources. Pour chaque cluster, nous avons extrait l'ensemble de l'intersection de toutes les étiquettes comme décrites dans l'algorithme suivant :

- On calcule T_i comme étant l'ensemble des étiquettes t tel que t appartient à chaque ressource qui appartient au cluster C_i
- Pour chaque t appartenant à T_i , on calcule la fréquence d'occurrences de l'étiquette t dans chaque ressource du cluster C_i
- La combinaison de l'étiquette principale de recherche avec l'étiquette ayant la fréquence maximale serait considérée comme étiquette-système primaire de ce cluster C_i (exemple dans notre cas d'étude : « toread articles » pour le 1er cluster, « toread programming » pour le second cluster et ainsi de suite
- La combinaison de l'étiquette principale de recherche avec l'étiquette ayant la seconde fréquence maximale serait considérée comme étiquette-système

secondaire de ce cluster C_i (exemple dans notre cas d'étude : « toread interesting » pour le 1er cluster, « toread web » pour le second cluster)

Et ainsi de suite (on peut avoir un nombre prédéterminé d'étiquettes-systèmes que l'on veut extraire). La figure 5.7 montre le cycle d'évolution du système.

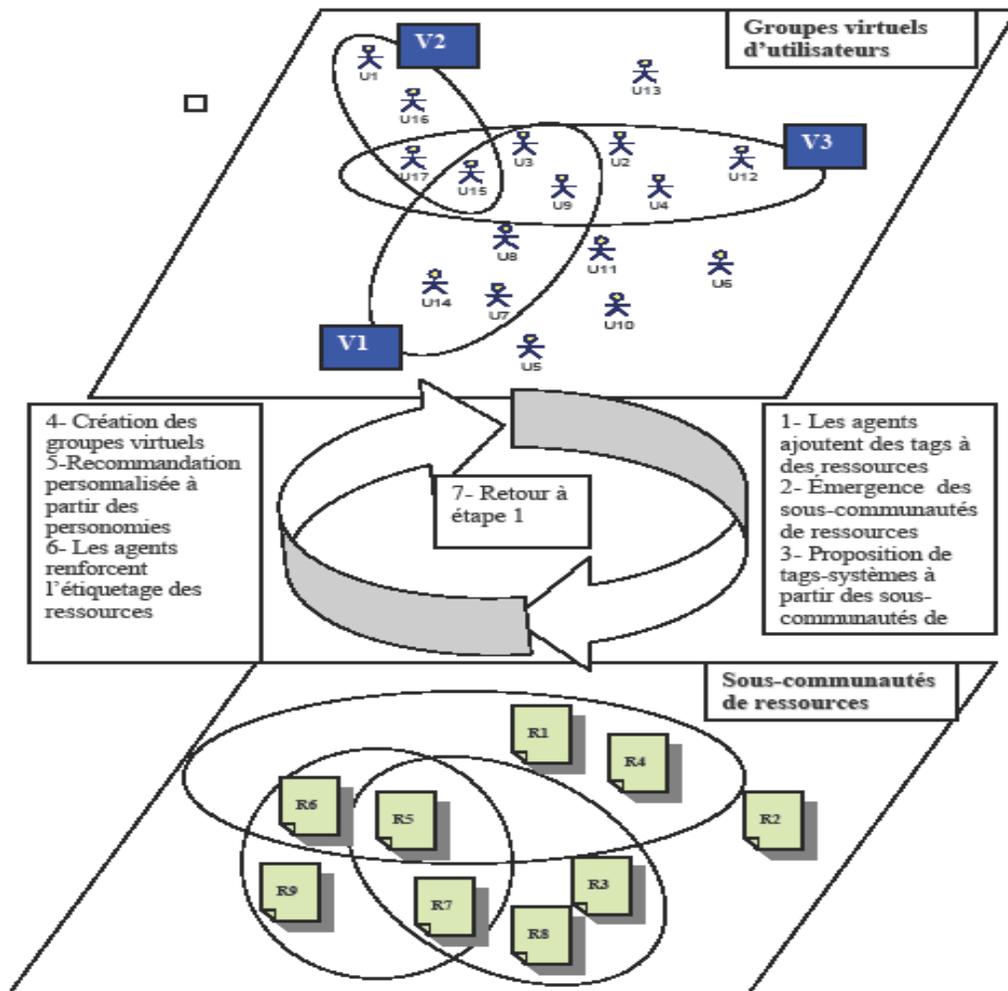


Figure 5.7 Cycle d'évolution du système.

5.5 Expériences et Résultats

5.5.1 Implémentation

L'application web a été créée sur un serveur Linux, utilisant MySQL comme base de données et le développement a été fait en Java. Le modèle conceptuel des données est relativement simple. Il est basé sur les 3 principales tables du réseau tripartite de n'importe quel système d'étiquetage (ressources, étiquettes, utilisateurs), de l'application étiquette, et de quelques tables additionnelles pour sauvegarder les sous-communautés de ressources et les groupes d'utilisateurs virtuels.

Les technologies utilisées pour le développement de l'application d'étiquetage sont les suivantes : Le serveur MySQL pour l'implémentation de la base de données et Glassfish comme serveur d'applications Open Source Java EE 5. Java était utilisé pour le calcul de la matrice des poids, le clustering spectral, les bibliothèques des Interfaces Utilisateurs et JavaScript pour le développement de l'application Web

Les trois modules principaux de l'application sont :

- Un module Client comprenant les pages web (fichiers .jsp) qui invoquent les méthodes du service web (Login.jsp, Libray.jsp, ..., Group.jsp).
- Un module Serveur (les fichiers Java) qui agit comme médiateur entre le client et la base de données. Il contient des ensembles de données pour sauvegarder les informations requises par les utilisateurs, les ressources et les étiquettes. Il requête la base de données en se basant sur les requêtes des utilisateurs.
- Un module Base de Données pour sauvegarder les informations des utilisateurs (Fig 5.8).

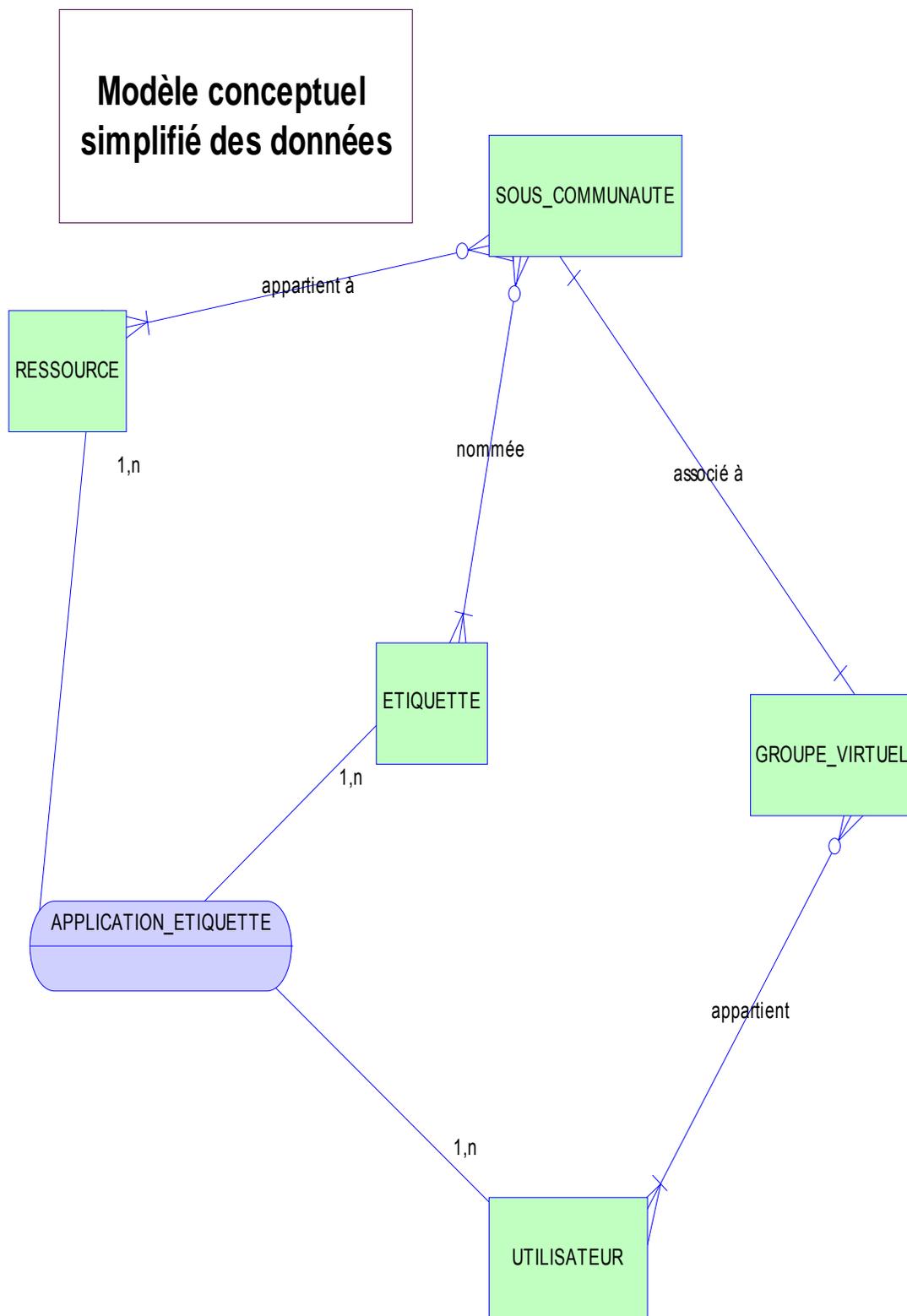


Figure 5.8 Modèle conceptuel simplifié de la base de données

5.5.2 Données expérimentales de Del.icio.us

Nous avons extrait des ensembles de données du système Del.icio.us que nous avons implémentés dans la base de données de notre système.

Plusieurs facteurs justifient notre choix de sélection des données de Del.icio.us. Del.icio.us est le plus populaire des systèmes d'étiquetage existants et le plus ancien, donc il contient la plus grande variété d'étiquettes, d'utilisateurs et de ressources.

Des données expérimentales extraites de Del.icio.us ont été utilisées dans le plus grand nombre de travaux de recherche sur la folksonomie, les réseaux sociaux en ligne et les systèmes d'étiquetage collaboratifs. Nous citons parmi les travaux qui ont basé leurs études sur des données extraites de Del.icio.us : (Shen and Wu 2005; Golder and Huberman 2006; Cattuto, Schmitz et al. 2007; Halpin, Robu et al. 2007; Yanbe, Jatowt et al. 2007; Heymann, Koutrika et al. 2008).

Pour chaque ressource extraite individuellement de Del.icio.us, il est facile d'extraire tout l'historique de l'étiquetage associée à cette ressource, c'est-à-dire tous les utilisateurs qui ont associé une étiquette à la ressource et toutes les étiquettes associées à cette ressource.

C'est une folksonomie large, contrairement à un système comme Flickr, qui offre une folksonomie étroite. Dans une folksonomie large, plusieurs personnes peuvent ajouter leurs propres étiquettes dans leur propre vocabulaire à une ressource particulière, ce qui produit une distribution suivant la loi de puissance, tandis que dans la folksonomie étroite seul le créateur de la ressource et un nombre très limité d'utilisateurs peuvent ajouter des étiquettes à la ressource en question. Les folksonomies étroites manquent de cohésion et d'échange social présents dans les folksonomies larges (VanderWal 2005).

Dans Del.icio.us, au mois de mai 2008, nous avons fait une requête pour l'étiquette "toread", et nous avons développé un programme qui a extrait les 99 premières ressources avec tout leur historique d'étiquetage. Pour chaque ressource, nous avons la liste de tous les utilisateurs qui ont étiqueté cette ressource, ainsi que la liste de toutes les étiquettes associées à cette ressource par chacun de ces utilisateurs. Nous gardons cette liste d'utilisateurs anonyme.

Au total, 20,720 étiquettes distinctes étaient associées avec les 99 ressources de “toread”, et notre nombre total d’utilisateurs est de 767,117.

Cet ensemble de données a été importé à notre base de données MySQL pour être traité dans notre application Web.

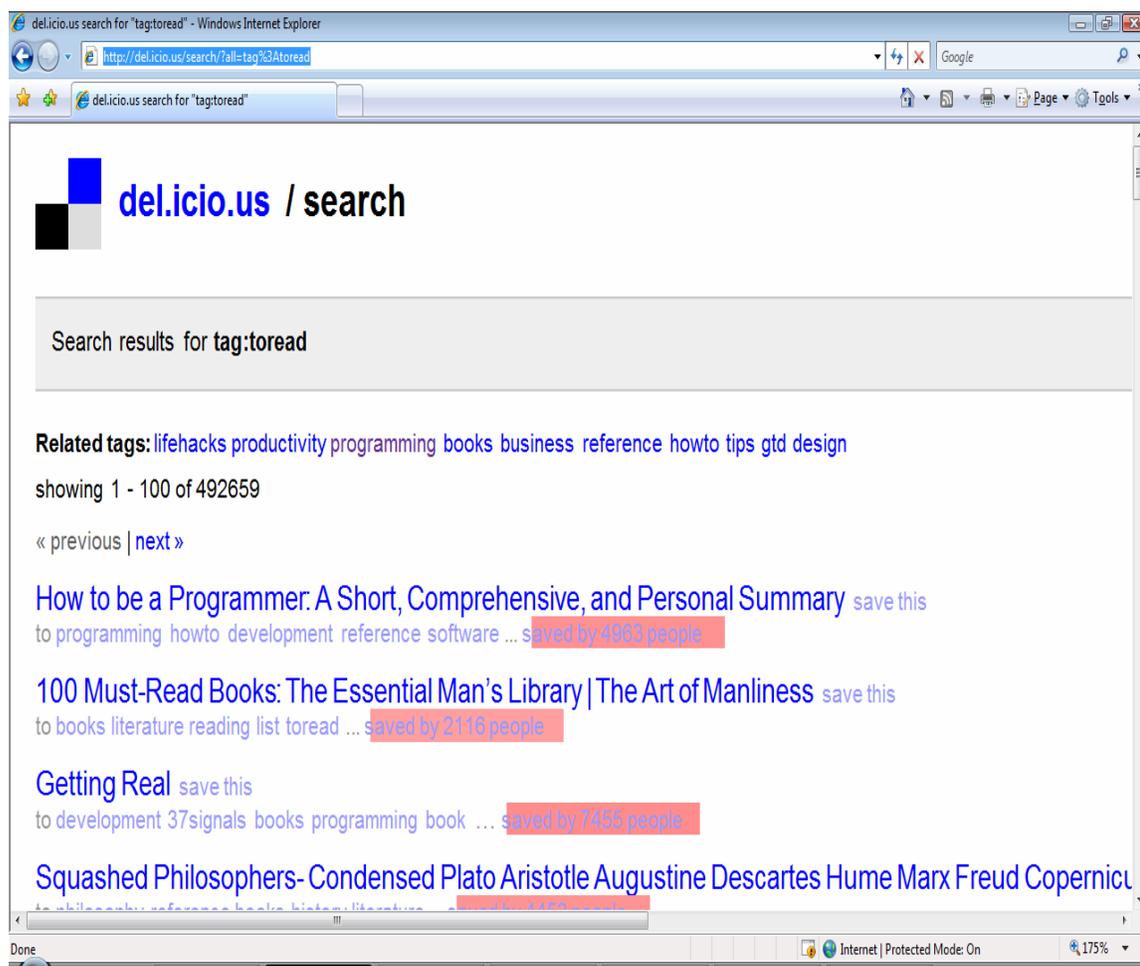


Figure 5.9 Résultats de recherche de l’étiquette « toread » dans Del.icio.us

La limitation en puissance de calcul de nos machines nous a contraint au choix de 100 ressources seulement de Del.icio.us. Le choix de l’étiquette « toread » est basé sur le fait que cette étiquette peut être sémantiquement très vaste et, a priori, l’utilisateur ne peut prévoir comment subdiviser hiérarchiquement cette étiquette en sous-catégories sémantiques.

5.5.3 Résultats

Le traitement des données extraites de Delicio.us pour les 99 ressources associées avec l'étiquette « toread » a abouti à la création de 4 sous-communautés de ressources qui appartiennent aux catégories suivantes (Fig. 5.10) (Notez que l'ancien nom de MySURF était SSOTS) :

“articles”

“programming”

“books”

”economics”

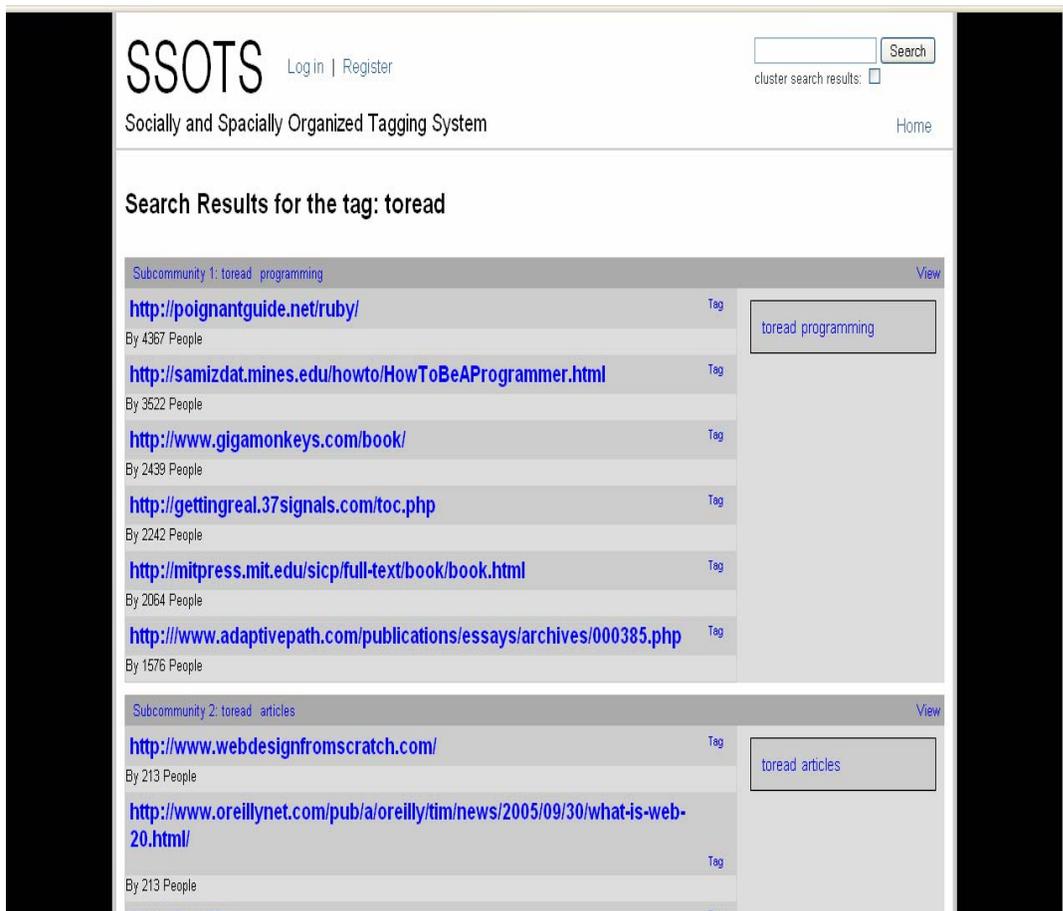


Figure 5.10 Résultats de la recherche de l'étiquette “toread” dans notre système basé sur les 99 ressources « toread » extraites de Delicio.us. Les résultats sont groupés en sous-communautés « toread programming », « toread articles »

Nous présentons dans cette section quelques graphiques et tableaux décrivant les résultats obtenus pour le calcul de la matrice Laplacienne Q , ses vecteurs propres et la matrice de corrélation pour les 99 ressources associées avec l'étiquette « toread ».

5.5.3.1 Visualisation du clustering spectral

Nous avons appliqué le clustering spectral aux données liées à l'étiquette « toread » extraites de Del.icio.us et insérées dans notre base de données. Nous avons calculé les valeurs propres de la matrice L et soit V la matrice qui contient tous les vecteurs propres VP de la matrice L .

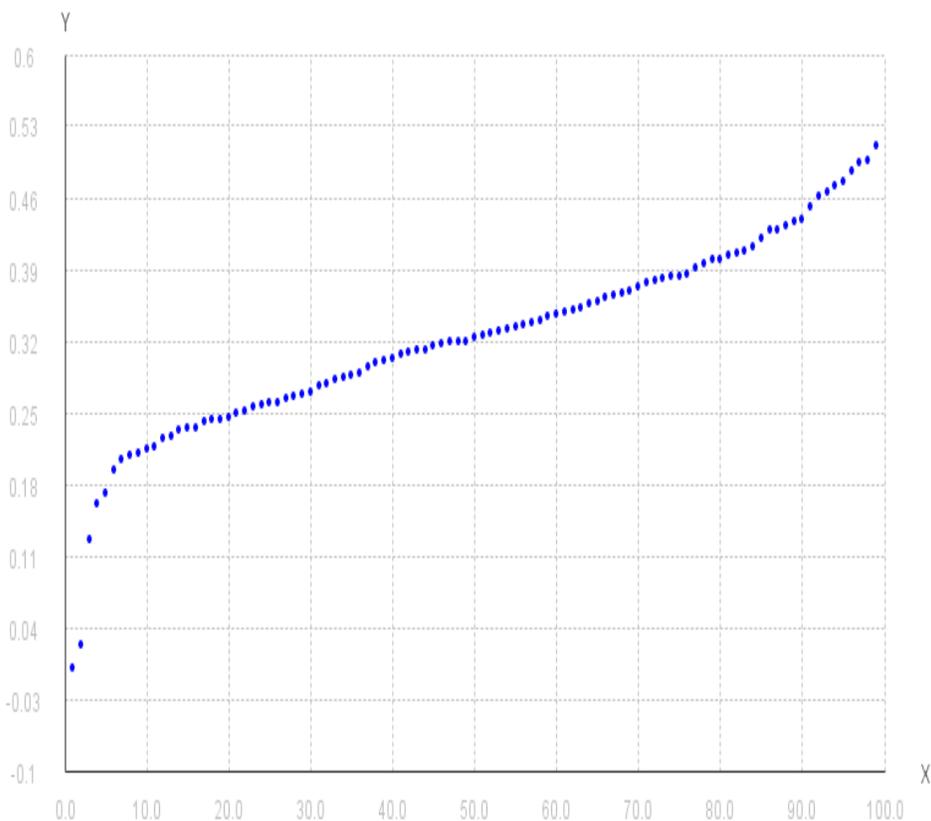


Figure 5.11. Valeurs propres triées de la Matrice Q . La présence d'au moins 4 valeurs propres bien séparées de valeur non zéro indique la présence d'au moins 4 différents « clusters ».

La figure 5.11 montre les éléments des vecteurs VP (les valeurs propres de la matrice Laplacienne L). Cette figure montre bien la présence d'une valeur propre triviale égale à zéro,

de 4 valeurs propres non triviales bien séparées (présence des 4 clusters) et le reste des éléments du vecteur VP ont des valeurs très proches les unes des autres.

Une façon de visualiser ces communautés est de représenter les éléments des vecteurs propres correspondants aux plus petites valeurs propres non triviales. Par exemple, si on considère les éléments des 3 premiers vecteurs propres V_1 , V_2 , V_3 de la matrice V où V_0 correspond au vecteur propre constant correspondant à la valeur propre nulle.

La figure 5.12 montre un graphique tridimensionnel des vecteurs propres correspondants aux 4 plus petites valeurs propres non triviales. Cette figure montre la présence des 4 clusters correspondants aux 4 sous-communautés.

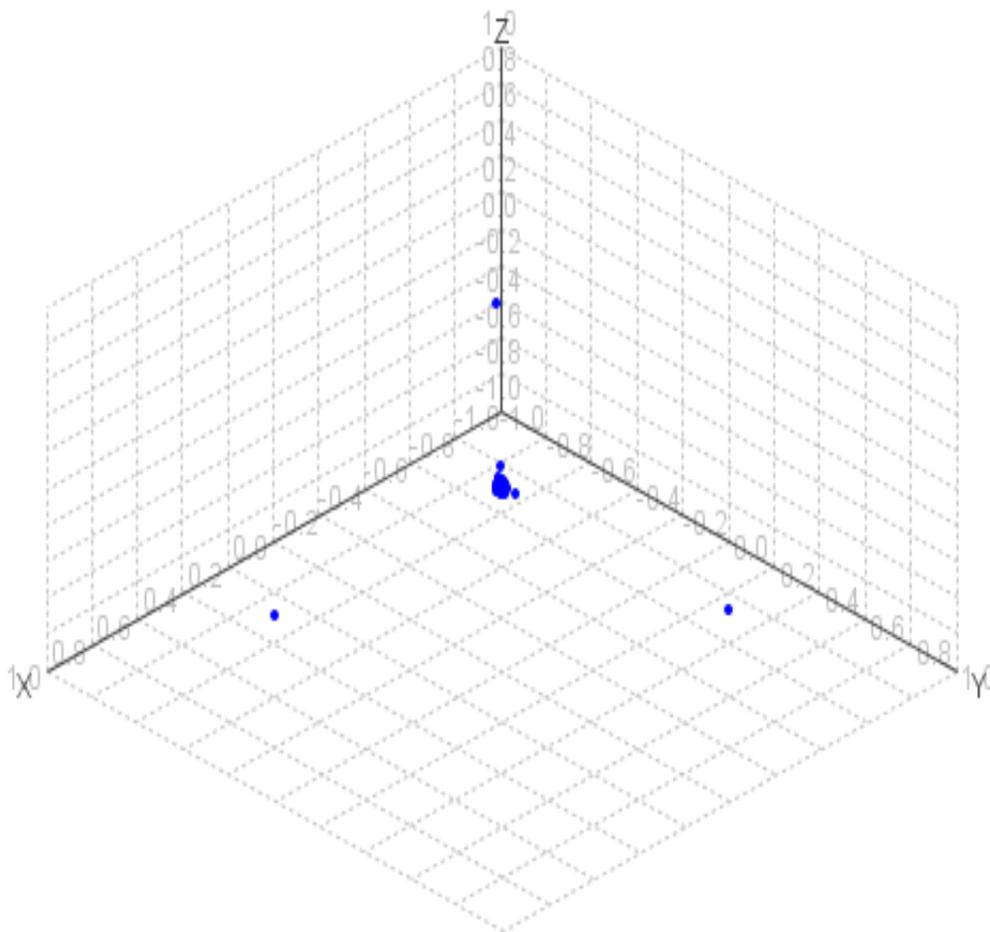


Figure 5.12 Représentation de cluster de ressources pour l'étiquette "toread", montrant la présence de quatre différents "clusters" de ressources.

Pour pouvoir extraire les ressources qui appartiennent à chaque sous-communauté ou cluster, nous avons calculé le coefficient de corrélation entre deux ressources pour les 10 premiers vecteurs propres de la matrice Laplacienne L (Tableau 5.1).

	1	2	3	4	5	6	7	8	9	10
1	-0.0007	-0.0134	0.0002	-0.0162	0.0258	-0.0086	0.2896	0.0002	-0.0060	-0.0072
2	-0.0112	0.0178	0.0285	0.0064	0.0181	-0.0305	-0.0154	-0.0003	0.0007	-0.0025
3	-0.0106	0.0072	0.0155	0.0102	0.0110	0.0031	-0.0207	0.0023	-0.0094	0.0056
4	-0.0107	0.0218	-0.0050	0.0016	0.0211	-0.0129	-0.0542	0.0022	-0.0029	0.0096
5	-0.0093	0.0112	0.0220	0.0416	0.0099	0.0056	-0.0241	0.0103	-0.0149	0.0194
6	-0.0111	-0.0024	-0.0011	0.0129	-0.0126	0.0059	-0.0051	-0.0100	0.0062	0.0100
7	-0.0063	0.0046	-0.0003	0.0022	0.0052	0.0041	0.1731	-0.0106	-0.0007	-0.0017
8	-0.0112	0.0138	-0.0094	0.0278	0.0126	0.0203	-0.0271	0.0137	-0.0042	0.0965
9	-0.0078	-0.0078	-0.0117	0.1628	0.0150	0.0979	0.0853	-0.6818	0.0009	0.1014
10	-0.0112	0.0153	0.0058	0.0268	0.0088	-0.0027	-0.0359	0.0145	0.0027	0.0811
11	-0.0110	0.0140	-0.0028	-0.0107	0.0170	-0.0105	-0.0249	0.0390	-0.1389	0.1847
12	-0.0097	0.0113	0.0030	0.0069	-0.0940	0.0022	0.1428	-0.0001	0.0098	0.0194
13	-0.0112	0.0273	0.0034	0.0125	0.0270	-0.0572	-0.1119	-0.0070	0.0047	-0.0032
14	-0.0113	0.0182	0.0271	0.0131	0.0104	-0.0640	-0.0467	-0.0154	0.0026	-0.0012
15	-0.0085	-0.0045	0.0031	-0.2029	0.0132	-0.1695	0.0700	0.6131	-0.0494	-0.1058
16	-0.0115	0.0205	-0.0036	-0.0014	0.0256	-0.0027	-0.1414	-0.0003	0.0030	0.0003
17	-0.0109	0.0171	0.1744	-0.1887	0.0157	0.9338	-0.0432	0.1282	-0.0563	-0.0038
18	-0.0121	0.0327	-0.0011	0.0009	-0.9817	0.0009	-0.0160	0.0009	-0.0010	-0.0005
19	-0.0114	-0.0014	0.0061	-0.0023	0.0111	-0.0048	-0.0381	0.0034	0.0031	0.0091
20	-0.0090	0.0165	0.0056	0.0031	0.0127	-0.0128	-0.0451	0.0153	-0.0301	0.0387
21	-0.0119	0.0246	0.0074	-0.0015	0.0222	-0.0009	-0.0909	-0.0005	-0.0003	-0.0024
22	-0.0112	0.0151	0.0227	-0.0005	-0.0063	-0.0048	-0.0391	0.0024	-0.0300	0.0228
23	-0.0113	0.0102	-0.0044	-0.0044	0.0218	-0.0108	-0.0282	-0.0014	-0.0377	0.0062
24	-0.0115	0.0191	0.0128	-0.0025	0.0128	-0.0505	-0.0551	-0.0133	0.0107	-0.0045
25	-0.0097	0.0075	0.0268	-0.0063	0.0177	-0.0182	0.0536	-0.0077	0.0018	0.0282
26	-0.0104	-0.0037	-0.0008	0.0108	0.0089	0.0063	0.1943	-0.0056	0.0046	0.0120
27	-0.0112	0.0140	0.0012	0.0160	-0.0035	0.0035	-0.0110	-0.0256	-0.0021	0.0326
28	-0.0112	0.0144	0.0211	0.0197	0.0113	-0.0060	-0.0318	0.0102	-0.0193	0.0542
29	-0.0049	-0.0017	0.0039	-0.0065	0.0124	-0.0049	0.0275	0.0353	-0.0095	0.0087
30	-0.0108	0.0118	-0.0138	0.0221	0.0116	0.0452	-0.0160	-0.0033	0.8982	-0.2999
31	-0.0114	0.0182	0.0085	0.0029	0.0047	-0.0067	-0.0446	0.0091	-0.0005	0.0022
32	-0.0106	0.0197	-0.0009	0.0023	0.0136	-0.0097	-0.0822	-0.0032	0.0138	-0.0024
33	-0.0108	0.0158	0.0274	-0.0008	0.0094	-0.0747	-0.0509	-0.0320	-0.1693	0.0808
34	-0.0113	0.0173	-0.0034	0.0149	0.0036	-0.0254	-0.0395	-0.0098	-0.0032	0.0520
35	-0.0117	0.0230	-0.0012	-0.0012	0.0287	-0.0075	-0.0666	0.0021	0.0018	0.0203
36	-0.0114	0.0149	-0.0087	0.0083	0.0097	-0.0398	-0.0496	0.0006	0.0001	-0.0049
37	-0.0101	0.0123	0.0375	-0.1124	0.0105	-0.0271	-0.0054	-0.1801	-0.3343	-0.8830
38	-0.0091	0.0080	-0.0044	-0.0035	0.0137	-0.0038	0.0183	-0.0024	-0.0059	0.0200
39	-0.0090	0.0164	0.0044	0.0017	0.0407	-0.0026	-0.1006	0.0026	0.0010	0.0149
40	-0.0114	0.0154	0.0035	-0.0085	0.0090	0.0004	-0.0329	-0.0011	-0.0424	-0.0004

Tableau 5.1 représentant les valeurs des éléments des dix premiers vecteurs propres de la matrice Q (colonnes 1 à 10) pour 40 ressources (lignes 1 à 40).

A partir des valeurs des éléments des dix premiers vecteurs propres de la matrice Q, nous avons calculé les éléments de la matrice C (Tableau 5.2) selon la formule :

$$C_{ij} = \frac{\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle}{\sqrt{[(\langle x_i^2 \rangle - \langle x_i \rangle^2)(\langle x_j^2 \rangle - \langle x_j \rangle^2)]}}$$

	1	2	3	4	5	6	7	8	9	10
1	1.0000	-0.2946	-0.6620	-0.8091	-0.5912	-0.2521	0.9885	-0.4423	0.1363	-0.5599
2	0.2946	1.0000	0.6304	0.5833	0.4692	-0.1435	-0.3125	-0.0177	-0.0824	0.2527
3	0.6620	0.6304	1.0000	0.7445	0.8815	0.1968	-0.6467	0.4451	-0.0184	0.5613
4	0.8091	0.5833	0.7445	1.0000	0.5915	0.0331	-0.8132	0.5111	-0.1423	0.6405
5	0.5912	0.4692	0.8815	0.5915	1.0000	0.4249	-0.5516	0.5530	0.0443	0.6498
6	0.2521	-0.1435	0.1968	0.0331	0.4249	1.0000	-0.1375	0.5257	0.5024	0.5011
7	0.9885	-0.3125	-0.6467	-0.8132	-0.5516	-0.1375	1.0000	-0.4135	0.2267	-0.5396
8	0.4423	-0.0177	0.4451	0.5111	0.5530	0.5257	-0.4135	1.0000	0.1125	0.9480
9	0.1363	-0.0824	-0.0184	-0.1423	0.0443	0.5024	0.2267	0.1125	1.0000	0.0311
10	0.5599	0.2527	0.5613	0.6405	0.6498	0.5011	-0.5396	0.9480	0.0311	1.0000
11	0.1305	0.0474	0.3800	0.3057	0.4478	0.0677	-0.1464	0.7897	-0.0664	0.7326
12	0.7552	-0.4082	-0.6487	-0.8132	-0.4285	0.2059	0.8072	-0.2305	0.1337	-0.3144
13	0.8049	0.7261	0.6896	0.9310	0.5806	0.0253	-0.8152	0.3210	-0.1232	0.5358
14	0.4684	0.9282	0.5766	0.6629	0.5150	0.0241	-0.4740	0.1295	0.0064	0.4142
15	0.1181	0.0832	-0.1174	-0.0167	-0.1784	-0.6144	0.0187	-0.2108	-0.9619	-0.1439
16	0.9508	0.4340	0.7286	0.9285	0.5650	0.0994	-0.9541	0.4198	-0.1450	0.5396
17	0.1627	-0.4997	0.1380	-0.1127	-0.0394	0.0784	-0.1405	0.0281	-0.0135	-0.1579
18	0.0176	-0.3200	-0.2728	-0.3623	-0.0332	0.4617	0.0577	0.0172	-0.0609	0.0329
19	0.8449	0.5362	0.7836	0.8820	0.6012	0.1515	-0.8700	0.5367	-0.1930	0.6812
20	0.6262	0.4575	0.7586	0.7979	0.7213	0.0644	-0.6413	0.7573	-0.2127	0.8392

Tableau 5.2 Matrice C de corrélation calculée à partir des valeurs des éléments des vecteurs propres de la matrice Q. Ce tableau montre les valeurs pour les 20 premières ressources

A partir de ces valeurs, nous avons extrait nos clusters en groupant ensemble toutes les ressources qui ont un coefficient de corrélation plus grand ou égal à 0.09. Nous avons obtenu 4 clusters de ressources principaux et les ressources qui n'appartenaient à aucun des clusters avaient un coefficient de corrélation faible avec les autres ressources.

5.5.4 Données expérimentales dans MySURF et résultats de l'échantillon créé manuellement

L'échantillon que nous avons créé manuellement dans la base de données pour tester la validation du cycle des fonctionnalités de notre système est construit comme suit :

- 70 utilisateurs nommés *user1* à *user70*.
- 6 étiquettes : Design, Web, Home, Database, Algorithm, Vancouver.
- 25 ressources correspondant à ces étiquettes.

5.5.5 Exemple d'un cycle d'étapes dans MySURF pour l'étiquette « Design »

Le scénario suivant basé sur l'échantillon décrit au paragraphe 5.5.4 montre bien l'effet de la coévolution des organisations spatiales et sociales sur l'évolution du système et comment cela s'exprime par des nouvelles fonctionnalités ajoutées au système. En partant de l'utilisateur *User1* qui fait la recherche de l'étiquette « Design », le scénario continue comme suit :

1. *User1* fait un Login.
2. Une recherche est faite sur l'étiquette "Design" avec l'option de générer un cluster des résultats (Fig.5.13).
3. Les ressources sont classées hiérarchiquement en 4 clusters "Web Design", "Home Design", "Database Design", "Algorithm Design" (Fig.5.13).
4. 4 groupes virtuels d'utilisateurs sont créés correspondant aux 4 clusters (Fig.5.14).
5. Comme *User1* appartient au groupe d'utilisateurs "Web Design", la *personomie* de tous les utilisateurs appartenant au groupe "Web Design" est extraite
6. Un tableau de taux de similarité basée sur les *personomies* est créé entre *User1* et tous les autres utilisateurs du groupe "Web Design" (Fig.5.15).
7. Les résultats du tableau de similarité montrent que *User1* est fortement similaire à *User2*.
8. La recommandation personnalisée de *User1* affichera les ressources étiquetées par les utilisateurs qui lui sont fortement similaires (*User2* dans ce cas). Nous considérons que ces ressources ont une importance particulière à *User1* car elles

proviennent d'un utilisateur qui partage ses mêmes intérêts. Dans notre exemple *User1* et *User2* sont tous les deux intéressés par « web design », mais ils partagent un autre intérêt commun « Vancouver » qui n'est pas du tout relié au « web design » (Fig.5.16).

Évolution du système:

9. Il existe une grande probabilité que *User1* trouve les ressources qui lui sont recommandées par le système très intéressantes pour lui. Il va lui-même les étiqueter et par ce fait, les renforcer dans le système.
10. A force d'être étiquetées, ces ressources vont apparaître dans la sous-communauté "web design", ce qui affectera la structure de l'organisation spatiale.
11. Et ainsi de suite, le système évolue, et le cycle recommence à l'étape 2 ci-dessus.

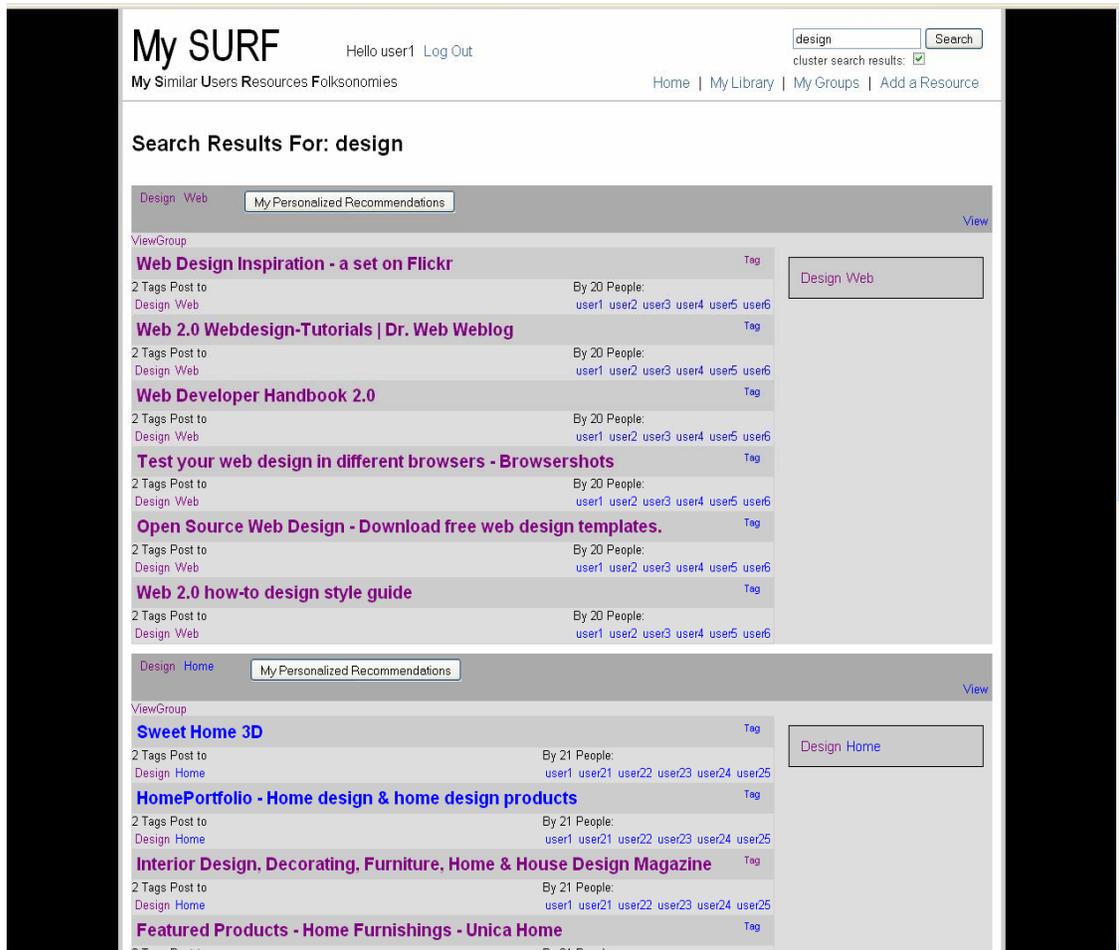


Figure 5.13 Quatre sous-communautés de ressources sont générées quand un clustering est effectué sur l'étiquette « Design »

My SURF Hello user1 Log Out design Search cluster search results: Home | My Library | My Groups | Add a Resource

My Similar Users Resources Folksonomies

My Groups: Users | Resources

user1 is in following virtual group(s)

Design	Web	View
Web Design Inspiration - a set on Flickr Tag		
2 Tags Post to	By 20 People:	
Design Web	user1 user2 user3 user4 user5 user6	Tags Users Design Web user1 user2 user3 user4 user5 user6 user7 user8 user9 user10
Web 2.0 Webdesign-Tutorials Dr. Web Weblog Tag		
2 Tags Post to	By 20 People:	
Design Web	user1 user2 user3 user4 user5 user6	
Web Developer Handbook 2.0 Tag		
2 Tags Post to	By 20 People:	
Design Web	user1 user2 user3 user4 user5 user6	
Test your web design in different browsers - Browsershots Tag		
2 Tags Post to	By 20 People:	
Design Web	user1 user2 user3 user4 user5 user6	
Open Source Web Design - Download free web design templates. Tag		
2 Tags Post to	By 20 People:	
Design Web	user1 user2 user3 user4 user5 user6	
Web 2.0 how-to design style guide Tag		
2 Tags Post to	By 20 People:	
Design Web	user1 user2 user3 user4 user5 user6	
Design	Home	View
Sweet Home 3D Tag		
2 Tags Post to	By 21 People:	
Design Home	user1 user21 user22 user23 user24 user25	Tags Users Design Home user1 user21 user22 user23 user24 user25 user26 user27 user28 user29
HomePortfolio - Home design & home design products Tag		
2 Tags Post to	By 21 People:	
Design Home	user1 user21 user22 user23 user24 user25	
Interior Design, Decorating, Furniture, Home & House Design Magazine Tag		
2 Tags Post to	By 21 People:	
Design Home	user1 user21 user22 user23 user24 user25	
Featured Products - Home Furnishings - Unica Home Tag		
2 Tags Post to	By 21 People:	
Design Home	user1 user21 user22 user23 user24 user25	
Design*Sponge Tag		
2 Tags Post to	By 21 People:	
Design Home	user1 user21 user22 user23 user24 user25	
Apartment Therapy Tag		
2 Tags Post to	By 21 People:	
Design Home	user1 user21 user22 user23 user24 user25	

Figure 5.14 User1 appartient à 2 groupes virtuels : « Design Web » et « Design Home »

My SURF Hello user1 Log Out design Search cluster search results:

My Similar Users Resources Folksonomies Home | My Library | My Groups | Add a Resource

My Groups: Users | Resources

user1 is in following virtual group(s)

Design Web View

User	Similarity Score
user1	315
user2	135
user3	135
user4	72
user5	72
user6	72
user7	72
user8	72
user9	72
user10	72
user11	72
user12	72
user13	72
user14	72
user15	72
user16	72
user17	72
user18	72
user19	72
user20	72

Design Home View

User	Similarity Score
user1	315
user21	36
user22	36
user23	36
user24	36
user25	36
user26	36
user27	36
user28	36
user29	36
user30	36
user31	36
user32	36
user33	36
user34	36
user35	36
user36	36
user37	36
user38	36
user39	36

Figure 5.15 Le tableau de similarité de *User1* en comparaison avec tous les autres membres des groupes auquel il appartient. Notons que *User1* est fortement similaire à *User2* et *User3* car tous les trois partagent un intérêt commun en « Vancouver ». Ils ont étiquetées des ressources communes avec l'étiquette « Vancouver ».

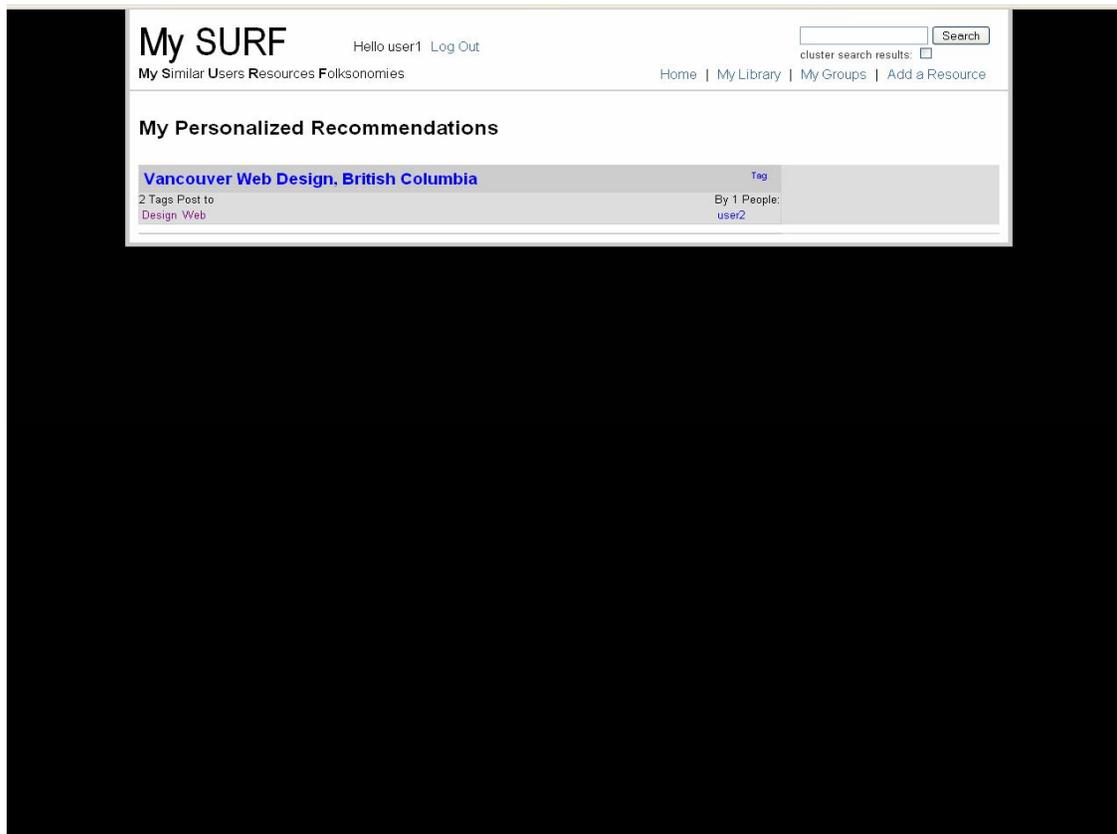


Figure 5.16 La recommandation personnalisée de *User1* pour le sujet de « web design » est la page web étiquetée par *User2* qui lui est fortement similaire. Notons que cette page n’apparaît pas dans la sous-communauté « web design » car il n’existe pas assez d’étiquetage de cette page pour la qualifier à la sous-communauté.

Ces résultats de clustering des sous-communautés de ressources, des groupes virtuels d’utilisateurs et des recommandations suggérées à *User1* sont consistants avec l’échantillon de données que nous avons planté manuellement dans la base de données.

5.6 Synthèse

Dans ce chapitre nous avons pu mettre en évidence, la mise en œuvre du couplage entre l’organisation sociale et l’organisation spatiale dans un système multi-agents, considéré selon la perspective des systèmes complexes. Nous nous sommes situés dans le cadre de l’évolution du web 2.0 et plus particulièrement dans le cadre des systèmes d’étiquetage collaboratif. Cette

vision est inscrite dans le courant enactiviste de Varela. Elle permet aussi d'approfondir le concept de la stigmergie souvent réduit au mécanisme de dépôt de phéromones dans les systèmes multi-agents.

Du point de vue de « la science du web », ce chapitre montre le potentiel apporté par une vision enactiviste et selon la perspective des systèmes complexes d'une approche multi-agents pour le futur du web. Le système d'étiquetage collaboratif que nous avons proposé, illustre ce potentiel et constitue un premier pas dans le développement des machines sociales intelligentes, en permettant de mettre en place un support pour l'expression de la dimension sociale et sa corrélation à la dimension spatiale dans l'activité des utilisateurs sur le web et leurs usages.

L'originalité de notre application MySURF se résume par la classification hiérarchique des données augmentée d'un système de recommandation des ressources basé sur le degré de similarité entre les utilisateurs. C'est ce qui distingue bien notre travail des systèmes d'étiquetage collaboratifs actuels. Grâce à l'intégration des dimensions spatiales et sociales, nous avons abouti aux nouvelles fonctionnalités citées dans la partie 5.3 de ce mémoire.

En tant que nouvelles perspectives qu'ouvre notre travail, une étude du système MYSURF comme un vrai réseau social est en cours pour intégrer les caractéristiques et propriétés des réseaux sociaux dans les systèmes d'étiquetage collaboratifs. Cela constitue nos premières démarches vers la création des futures machines sociales qui ne **sont pas** « limitées par le fait qu'elles sont largement isolées les unes des autres » comme c'est le cas aujourd'hui pour Facebook, MySpace, Flickr, Del.cio.us, YouTube et la blogosphère.

Chapitre 6

Conclusion et travail futur

Les systèmes informatiques d'aujourd'hui sont caractérisés par une complexité intrinsèque croissante, en raison de leur ouverture à un environnement répandu et incertain. Nous arguons que l'ingénierie de tels systèmes doit être déterminée par leur couplage à l'environnement, comme ça se fait dans les systèmes vivants qui sont généralement retenus comme une source d'inspiration. Lorsque nous utilisons des multi-agents situés pour réaliser un système informatique selon cette vision, le système multi-agents est lui-même sujet à ce couplage à l'environnement du système. Comme le système multi-agents est structurellement représenté par le déploiement spatial des agents dans l'environnement physique, le couplage structurel est exprimé à travers l'organisation spatiale en ce qui concerne l'environnement. Au niveau conceptuel, le couplage comportemental, qui est en relation au couplage structurel, est représenté dans le système multi-agents par l'organisation sociale et son couplage à l'organisation spatiale. Le système multi-agents doit maintenir son organisation spatiale à travers son organisation sociale et vice versa (auto-organisation).

D'autre part, l'évolution rapide du web et de ses applications, leur convergence vers les systèmes complexes, auto-organisants, le lien irréductible avec l'environnement et l'émergence continue de nouvelles pratiques requièrent des méthodologies de développement

de ces systèmes qui tiennent compte de ces aspects complexes et émergents. De même, la croissance rapide des réseaux sociaux formés à partir de l'usage intensif de certaines applications web requiert une approche qui tient compte de cette dimension sociale importante.

Nous avons utilisé le paradigme des systèmes multi-agents comme modèle de représentation de ces systèmes où l'environnement physique est représenté par l'organisation spatiale des agents et l'environnement conceptuel des usages est représenté par l'organisation sociale des agents. Le couplage entre la structure du système et son environnement est exprimé par le couplage entre l'organisation sociale et l'organisation spatiale du système multi-agent aboutissant à l'auto-organisation. Ce couplage à l'environnement s'inscrit dans une vision enactive où le système structure l'environnement tout en étant structuré par lui. Les différentes composantes du système interagissent entre elles de façon à produire et maintenir d'une manière continue la production (transformation et destruction) de ces composantes et de leurs relations.

6.1 Contributions

Nous avons étudié le cas de deux applications web de natures différentes, la simulation des communautés virtuelles et la proposition d'un système d'étiquetage collaboratif.

Dans le cas des communautés virtuelles, nous avons proposé une simulation à base d'agents qui a permis de constater que la structure du réseau est guidée dans son évolution par l'usage, qui à son tour évolue en réaction à l'évolution de la structure. Ceci a permis de constater l'influence de l'identification des rôles sociaux des acteurs dans la communauté virtuelle sur les évolutions futures. Cette observation a conforté l'idée déjà défendue dans (Hassas 2003) que le couplage entre l'organisation sociale et l'organisation spatiale du système multi-agents est un moyen de mettre en œuvre l'auto-organisation.

Le web et ses utilisateurs forment naturellement un système auto-organisé. Les systèmes déployés sur le web, comme les systèmes d'étiquetage collaboratif, gagnent à expliciter ce caractère auto-organisationnel et en profiter.

C'est ce que nous avons proposé dans le cadre du système d'étiquetage collaboratif MySURF, où les agents sont représentés par les utilisateurs eux-mêmes. Ces agents peuvent se situer spatialement sur les sous-communautés de ressources du système qui se génèrent à partir de l'étiquetage des utilisateurs et s'organisent socialement dans des groupes virtuels d'utilisateurs. Les deux organisations sont dépendantes l'une de l'autre dans leur évolution. Nous avons montré que le système ainsi construit offre de nouvelles fonctionnalités qui peuvent améliorer les systèmes d'étiquetage collaboratifs courants et ouvre ainsi une voie vers les machines sociales intelligentes.

6.2 Travail futur

Le travail présenté dans ce mémoire a permis de montrer l'intérêt de la prise en compte du couplage des organisations sociales et spatiales dans le développement des systèmes collaboratifs déployés sur le web. Les contributions présentées dans cette thèse sont un premier pas vers la réalisation de machines socialement intelligentes (Socially Intelligentes ICT).

De nombreuses pistes de recherche restent à explorer, soit à court terme, soit à long terme. Nous citons ci-dessous les directions de recherche que nous envisageons de poursuivre.

Dans le domaine des systèmes d'étiquetage collaboratifs, l'étude de leurs propriétés et de leur évolution:

Nouvelles données pour tester les résultats : Les données qui ont été testées pour obtenir les résultats de notre application d'étiquetage proposé ont été extraites de Del.icio.us. Nous espérons pouvoir tester MySURF avec des utilisateurs réels qui ajoutent des étiquettes à des ressources. En première étape, nous planifions de tester l'application dans nos laboratoires de recherche et avec nos étudiants, et en 2^{ème} étape, la mettre sur le web pour une audience plus large. Nous espérons témoigner lors de l'évolution du système, de l'émergence de nouvelles pratiques que nous n'avons pas prédites lors de la conception de notre application.

Analyse supplémentaire sur les données déjà extraites de Del.icio.us : ces données ont besoin d'une analyse approfondie pour apprendre plus sur les systèmes d'étiquetage collaboratifs. Une des observations préliminaires que nous avons remarquées, c'est que ces

systèmes, tout comme le web, sont fractals, auto-similaires. Le sous-ensemble de données relatives à une seule étiquette possède les propriétés « scale free » et « small world » observées dans le système considéré comme un tout.

Étude des profils des utilisateurs et implémentation de systèmes de recommandation:

Montrer que le système d'étiquetage collaboratif que nous avons proposé permet l'émergence de nouvelles pratiques, telles que l'identification des profils des utilisateurs. En fait, à partir des groupes virtuels auxquels appartient l'utilisateur, on peut détecter ses préférences et construire son profil. Il reste à implémenter un système de recommandation qui peut suggérer des ressources à l'utilisateur, le notifier de nouvelles ressources pertinentes ajoutées au système, etc.

Nouvelles méthodes de clustering : Nous avons choisi la méthode de clustering spectral pour regrouper les ressources dans des sous-communautés. Ce n'est pas la méthode unique, et il serait intéressant de comparer les résultats obtenus par cette méthode à des résultats obtenus par d'autres méthodes de clustering, comme le clustering par partitionnement ou le clustering hiérarchique. Si ces méthodes n'aboutissent pas à des résultats satisfaisants, aller plus loin en proposant une nouvelle méthode de clustering plus appropriée au problème de l'émergence des sous-communautés de ressources dans les systèmes d'étiquetage collaboratifs.

Du point de vue des multi-agents, il serait intéressant de considérer la proposition d'une méthodologie de conception permettant de considérer les différentes organisations (sociale, spatiales) et leur couplage rétro-actifs, en capitalisant l'expérience acquise dans le développement du système MySURF.

Enfin, notre tentative de réponse partielle à la question initiale posée dans l'introduction de ce mémoire reliée au web science « *“What underlying architectural principles are needed to guide the design and efficient engineering of new Web infrastructure components for this social software?”* » nous a amené à nous intéresser au couplage existant entre l'organisation sociale et l'organisation spatiale d'un collectif. Ceci est déterminant de notre point de vue dans ce premier pas vers la construction des machines sociales du futur. Ces notions liées au « web science » nécessitent des études plus approfondies et nous espérons continuer à orienter nos travaux de recherche dans cette direction.

Annexe A

Dépendance entre la position sociale (rôle) et la position spatiale (place) à partir du test chi-deux. 3 différentes simulations ont été exécutées.

Original Data:

Role	Location Type	Transaction Count
Information Seeker	Information Request	501
Information Giver	Information Response	372
Opinion Seeker	Opinion Request	2
Opinion Giver	Opinion Response	4

H_0 role and location type are independent

$\alpha = 0.01$

H_a role and location type dependent

Contingency Table (Actual):

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	501	0	0	0	501
Information Giver	0	372	0	0	372
Opinion Seeker	0	0	2	0	2
Opinion Giver	0	0	0	4	4
Total	501	372	2	4	879

Contingency Table (Expected):

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	285.552901	212.0273038	1.139931741	2.279863481	501
Information Giver	212.0273038	157.4334471	0.846416382	1.692832765	372
Opinion Seeker	1.139931741	0.846416382	0.004550626	0.009101251	2
Opinion Giver	2.279863481	1.692832765	0.009101251	0.018202503	4
Total	501	372	2	4	879

Chi-Square Calculation:

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	162.552901	212.0273038	1.139931741	2.279863481	378
Information Giver	212.0273038	292.4334471	0.846416382	1.692832765	507
Opinion Seeker	1.139931741	0.846416382	875.0045506	0.009101251	877
Opinion Giver	2.279863481	1.692832765	0.009101251	871.0182025	875
Total	378	507	877	875	2637

Results:

$$X^2 = 2637$$

$$\alpha = 0.01$$

$$df = 9$$

$$X_{\alpha}^2 = 21.666$$

$$X^2 = 2637 \geq 21.666 = X_{\alpha}^2 \Rightarrow \text{reject } H_0$$

Original Data:

Role	Location Type	Transaction Count
Information Seeker	Information Request	245
Information Giver	Information Response	323
Opinion Seeker	Opinion Request	5
Opinion Giver	Opinion Response	4

H_0 role and location type are independent

$\alpha = 0.01$

H_a role and location type dependent

Contingency Table (Actual):

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	245	0	0	0	245
Information Giver	0	323	0	0	323
Opinion Seeker	0	0	5	0	5
Opinion Giver	0	0	0	4	4
Total	245	323	5	4	577

Contingency Table (Expected):

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	104.0294627	137.1490468	2.12305026	1.698440208	245
Information Giver	137.1490468	180.812825	2.798960139	2.239168111	323
Opinion Seeker	2.12305026	2.798960139	0.043327556	0.034662045	5
Opinion Giver	1.698440208	2.239168111	0.034662045	0.027729636	4
Total	245	323	5	4	577

Chi-Square Calculation:

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	191.0294627	137.1490468	2.12305026	1.698440208	332
Information Giver	137.1490468	111.812825	2.798960139	2.239168111	254
Opinion Seeker	2.12305026	2.798960139	567.0433276	0.034662045	572
Opinion Giver	1.698440208	2.239168111	0.034662045	569.0277296	573
Total	332	254	572	573	1731

Results:

$$X^2 = 1731$$

$$\alpha = 0.01$$

$$df = 9$$

$$X_{\alpha}^2 = 21.666$$

$$X^2 = 1731 \geq 21.666 = X_{\alpha}^2 \Rightarrow \text{reject } H_0$$

Original Data:

Role	Location Type	Transaction Count
Information Seeker	Information Request	636
Information Giver	Information Response	453
Opinion Seeker	Opinion Request	5
Opinion Giver	Opinion Response	2

H_0 role and location type are independent

$\alpha = 0.01$

H_a role and location type dependent

Contingency Table (Actual):

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	636	0	0	0	636
Information Giver	0	453	0	0	453
Opinion Seeker	0	0	5	0	5
Opinion Giver	0	0	0	2	2
Total	636	453	5	2	1096

Contingency Table (Expected):

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	369.0656934	262.8722628	2.901459854	1.160583942	636
Information Giver	262.8722628	187.2344891	2.066605839	0.826642336	453
Opinion Seeker	2.901459854	2.066605839	0.022810219	0.009124088	5
Opinion Giver	1.160583942	0.826642336	0.009124088	0.003649635	2
Total	636	453	5	2	1096

Chi-Square Calculation:

Role	Location Type				Total
	Info Req	Info Resp	Opinion Req	Opinion Resp	
Information Seeker	193.0656934	262.8722628	2.901459854	1.160583942	460
Information Giver	262.8722628	377.2344891	2.066605839	0.826642336	643
Opinion Seeker	2.901459854	2.066605839	1086.02281	0.009124088	1091
Opinion Giver	1.160583942	0.826642336	0.009124088	1092.00365	1094
Total	460	643	1091	1094	3288

Results:

$X^2 = 3288$

$\alpha = 0.01$

$df = 9$

$X^2_{\alpha} = 21.666$

$X^2 = 3288 \geq 21.666 = X^2_{\alpha} \Rightarrow \text{reject } H_0$

Bibliographie

Adamic, L. A., O. Buyukkokten et E. Adar (2003). "A social network caught in the web." *First Monday*, vol. 8, n. 6.

Albert, R., H. Jeong et A. L. Barabasi (1999). "Internet - Diameter of the World-Wide Web." *Nature*, vol. 401, n. 6749: 130-131.

Albert, R., H. Jeong et A. L. Barabasi (2000). "Error and attack tolerance of complex networks." *Nature*, vol. 406, n. 6794: 378-382.

Ardissono, L., A. Goy, G. Petrone et M. Segnan (2005). "A Multi-Agent Infrastructure for Developing Personalized Web-based Systems." *ACM Transactions on Internet Technology (TOIT)*, vol. 5, n. 1: 47 - 69.

Arthur, W. B., S. Durlauf et D. Lane (1997). The Economy as an Evolving Complex System II, Addison Wesley Longman.

Barabasi, A. L. (2001). "The physics of the Web." *Physics World*, vol. 14, n. 7: 33-38.

Barabasi, A. L., R. Albert et H. Jeong (2000). "Scale-free characteristics of random networks: the topology of the World-Wide Web." *Physica A*, vol. 281, n. 1-4: 69-77.

Barabasi, A. L. et E. Bonabeau (2003). "Scale-free networks." *Scientific American*, vol. 288, n. 5: 60-69.

Barnes, J. (1954). "Class and Committees in a Norwegian Island Parish." *Human Relations*, vol. 7, n.: 39-58.

Brewington, B. et G. Cybenko (2000). "How dynamic is the web?" 9th International World Wide Web Conference (WWW9), Geneva

Broder, A., R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins et J. Wiener (2000). "Graph structure in the Web." *Computer Networks-the International Journal of Computer and Telecommunications Networking*, vol. 33, n. 1-6: 309-320.

Brueckner, S. (2000). Return from the Ant: Synthetic Ecosystems for Manufacturing Control. Department of Computer Science. Berlin, Humboldt University.

Brueckner, S. et H. V. D. Parunak (2002). "Swarming Agents for Distributed Pattern Detection and Classification". In Proceedings of Workshop on Ubiquitous Computing, AAMAS 2002

Byer, D. et C. Depradine (2004). "Autonomous collaborative web agents." Proceedings of the 2004 international symposium on Information and communication technologies, Las Vegas, Nevada, Trinity College Dublin

Candea, C., M. Staicu et C. B. Zamfirescu (2000). "An agent-based system for modelling the searching process on the Web." 4th IEEE International Conference on Intelligent Engineering Systems, Portoroz, Slovenia

Capocci, A., V. D. P. Servedio, G. Caldarelli et F. Colaiori (2004). Detecting communities in large networks. <http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0402499>.

Cattuto, C., C. Schmitz, A. Baldassarri, V. D. P. Servedio, V. Loreto, A. Hotho, M. Grahl et G. Stumme (2007). "Network properties of folksonomies." *AI Communications: Network Analysis in Natural Sciences and Engineering*, vol. 20, n. 4: 245-262.

Cesarano, C., A. d'Acerno et A. Picariello (2003). An intelligent search agent system for semantic information retrieval on the internet. WIDM '03: Proceedings of the 5th ACM international workshop on Web information and data management, ACM Press: 111--117.

Chau, M., D. Zeng, H. Chen, M. Huang et D. Hendriawan (2003). "Design and evaluation of a multi-agent collaborative Web mining system." *Elsevier Science Publishers B. V.*, vol. 35, n.: 167-183.

Chen, H., L. Kagal, F. Perch, Y. Zou et S. Tolia (2001). "ITTALKS: A Case Study in the Semantic Web and DAML". International Semantic Web Symposium

Cheng, R. et J. Vassileva (2006). "Design and evaluation of an adaptive incentive mechanism for sustained educational online communities." *User Modeling and User-Adapted Interaction*, vol. 16, n. 3-4: 321-348.

Cho, J. et H. Garcia-Molina (2000). "The Evolution of the Web and Implications for an Incremental Crawler". the Twenty-sixth International Conference on Very Large Databases

Choi, Y. S. et S. I. Yoo (1999). Multi-agent learning approach to WWW information retrieval using neural network. IUI '99: Proceedings of the 4th international conference on Intelligent user interfaces, ACM Press: 23--30.

Choy, S.-O. et A. Lui (2006). "Web Information Retrieval in Collaborative Tagging Systems". IEEE/WIC/ACM International Conference on Web Intelligence.352-355

Ciro Cattuto, A. B., Vito Servedio, Vittorio Loreto (2007). "Emergent Community Structure in Social Tagging Systems". European Conference on Complex Systems, Dresden

Corkill, D., V. Lesser (1983). "Use of Meta-Level Control for Coordination in a Distributed Problem-Solving Network". Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI-83, Karlsruhe, Federal Republic of Germany:748-756.

Cooley, R., J. Srivastava et B. Mobasher (1997). "Web Mining: Information and Pattern Discovery on theWorldWideWeb." 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97)

Deerwester, S., S. Dumais, T. Landauer, G. Furnas et R. Harshman (1990). "Indexing by latent semantic analysis." *Journal of the American Society of Information Science*, vol. 41, n. 6: 391-407.

DeLoach, S. A. (2005). "Multiagent systems engineering of organization-based multiagent systems". Proceedings of the fourth international workshop on Software engineering for large-scale multi-agent systems, St. Louis, Missouri, ACM Press.1-7

Dickinson, I. et M. Wooldridge (2003). "Practical reasoning agents for the semantic web." Second International Conference on Autonomous Agents and Multiagent Systems (AAMAS-03), Melbourne, Australia

Ding, Y. et L. Xu (2007). Evolution of the World Wide Web.

Drogoul, A., J. Treuil et J. Zucker (2008). Modélisation et simulation à base d'agents.

Du, Y. (2006). "Modeling the Behavior of Lurkers in Online Communities Using Intentional Agents". International Conference on Intelligent Agents Web Technologies and International Commerce, IEEE Computer Society

Eliassi-Rad, T. et J. Shavlik, Eds. (2003). Intelligent Web Agents that Learn to Retrieve and Extract Information. Intelligent Exploration of the Web, Springer-Verlag.

Esteva, M., B. Rosell, J. A. Rodriguez-Aguilar et J. L. Arcos Ameli (2004). "An agent based middleware for electronic institutions." Proceedings of the Int. Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004) 236-243.

Farooq, U., T. G. Kannampallil, Y. Song, C. H. Ganoe, J. M. Carroll et L. Giles (2007). "Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics". the international ACM conference on Supporting Group Work

Feenberg, A. et M. Bakardjieva (2004). "Virtual Community: No 'Killer Implication'." *New Media & Society*, vol. 6, n. 1: 37-43.

Ferber, J. (1995). Les systèmes multi-agents. Vers une intelligence collective., InterEditions.

Ferber, J., O. Gutknecht et F. Michel (2003). "From agents to organizations: An organizational view of multi-agent systems". AOSE 2003 : agent oriented software engineering IV, Springer, Berlin

Ferber, J., F. Michel et J. Baez (2005). Integrating Environments with Organizations. Environments for Multi-Agent Systems Lecture Notes in Computer Science, Springer Berlin / Heidelberg. **3374**: 48-56.

Finn, T. et A. Joshi (2002). "Agents, Trust, and Information Access on the Semantic Web." *ACM SIGMOD Record: Special section on semantic web and data management*, vol. 31, n. 4: 30-35.

Flake, G. W., S. Lawrence, C. L. Giles et F. M. Coetzee (2002). "Self-Organization of the Web and Identification of Communities." *IEEE Computer*, vol. 35, n.: 66--71.

Furnas, G., C. Fake, L. v. Ahn, J. Schachter, S. Golder, K. Fox, M. Davis, C. Marlow et M. Naaman (2006). "Why do tagging systems work?" Conference on Human Factors in Computing Systems (CHI), ACM Press.36-39

Gell-Mann, M. (1994). The Quark and the Jaguar: Adventures in the Simple and the Complex, W. H. Freeman and Company.

Goden S. (2007), http://sethgodin.typepad.com/seths_blog/2007/01/web4.html

Golder, S. et B. A. Huberman (2006). "Usage patterns of collaborative tagging systems." *Journal of Information Science*, vol. 32, n. 2: 198 - 208.

Grassé, P. P. (1959). "La reconstruction du nid et les coordinations inter-individuelles chez bellicoitermes natalenis et cubitermes, la théorie de la stigmergie." *Insectes Sociaux*, vol. 6, n.: 41-81.

Greco, G., S. Greco et E. Zumpano (2004). "Web Communities: Models and Algorithms." *World Wide Web*, vol. 7, n. 1: 59--82.

Gregorio, J. (2003). Stigmergy and the World-Wide Web.

Grilo, A., Caetano, A. et A. Rosa (2000). "Modelling Immune System:Principles, Models, Analysis and Perspectives." System Model, CiteSeer Scientific Literature Digital Library.

Halpin, H., V. Robu et H. Shepherd (2007). "The complex dynamics of collaborative tagging". Ithe 16th international conference on World Wide Web, Banff, Alberta, Canada.211-220

Hassas, S. (2003). *Systèmes complexes à base de multi-agents situés*, HDR, University Claude Bernard Lyon1.

Hassas, S. (2005). "Engineering Complex Adaptive Systems using Situated MAS: Some Selected Works and Contributions". Sixth International Workshop on Engineering Societies in the Agents' World (ESAW 2005), Turkey, LNCS Springer Verlag, LNAI

Hendler, J., N. Shadbolt, W. Halla, T. Berners-Lee et D. Weitzner (2008). "Web science: an interdisciplinary approach to understanding the web." *Communications of the ACM*, vol. 51, n. 7: 60-69.

Heymann, P., G. Koutrika et H. Garcia-Molina (2008). Can social bookmarking improve web search? Proceedings of the international conference on Web search and web data mining Palo Alto, California, USA ACM: 195-206

Holland, J. (1992). Adaptation in Natural and Artificial Systems. Cambridge, MA, MIT Press.

Holland, J. (1995). Hidden Order, Addison-Wesley.

Holland, O. E. et C. Melhuish (1999). "Stigmergy, self-organization, and sorting in collective robotics." *Artificial Life*, vol. 5, n.: 173-202.

Horling, B. et V. Lesser (2004). "A survey of multi-agent organizational paradigms." *Knowl. Eng. Rev.*, vol. 19, n. 4: 281-316.

Hotho, A., R. Jaschke, C. Schmitz et G. Stumme (2006). Information Retrieval in Folksonomies: Search and Ranking. The Semantic Web: Research and Applications. Y. S. a. J. Domingue, LNCS, Springer. **4011**: 411-426.

<http://bibsonomy.org/>.

<http://del.icio.us/>.

<http://flickr.com/>.

<http://technorati.com/>.

<http://www.citeulike.org/>.

<http://youtube.com/>.

Kauffman, S. (1993). The Origins of Order: Self-Organization and Selection in Evolution, Oxford University Press.

- Kautz, H., B. Selman et M. Shah (1997). "The Hidden Web." *AI Magazine*, vol. 18, n.
- Kazienko, P. et M. Kiewra (2003). "ROSA - Multi-agent System for Web Services Personalization". First Atlantic Web Intelligence Conference, Madrid, Spain, Springer Verlag.297-306
- Kirtland, M. (2000). The Programmable Web: Web Services Provides Building Blocks for the Microsoft .NET Framework. MSDN Magazine. **15**.
- Kitio, R., O. Boissier, J. F. Hübner et A. Ricci (2008). Organisational artifacts and agents for open multi-agent organisations : "giving the power back to the agents". Coordination, Organizations, Institutions, and Norms in Agent Systems III. P. N. Jaime Sichman, J. Padget, and Sascha Ossowski, LNCS. Springer. **4870**: 171-186.
- Kleinberg, J. (1998). "Authoritative Sources in a Hyperlinked Environment". ACM-SIAM Symposium on Discrete Algorithms.668-677
- Klusch, M., Ed. (2001). Information agent technology for the Internet: a survey, Elsevier Science Publishers B. V.
- Kumar, R., P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tompkins et E. Upfal (2000). "The Web as a graph". Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, ACM Press.1--10
- Kumar, R., P. Raghavan, S. Rajagopalan et A. Tomkins (1999). " Trawling the web for emerging cyber-communities". Proc. 8th International World-Wide Web Conference (WWW).403--416
- Lee, K. J. (2006). "What goes around comes around: an analysis of del.icio.us as social space". Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work
- Levin, S. (1998). "Ecosystems and the Biosphere as Complex Adaptive Systems." *Ecosystems*, vol. 1, n.: 431-436.
- Lieberman, H. (1995). "Letizia an agent that assists web browsing". Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence
- Liu, J., S. Zhang et Y. Ye (2002). "Understanding emergent web regularities with information foraging agents". Proceedings of the first international joint conference on Autonomous agents and multiagent systems, ACM Press.457--458
- Maloney-Krichmar, D. et J. Preece (2005). "A multilevel analysis of sociability, usability, and community dynamics in an online health community." *ACM Trans. Comput.-Hum. Interact.*, vol. 12, n. 2: 201-232.

Mao, Y., J. Vassileva et W. Grassmann (2007). "A System Dynamics Approach to Study Virtual Communities". 40th Annual Hawaii International Conference on System Sciences, IEEE Computer Society

Meila, M. et J. Shi (2001). "A random walks view of spectral segmentation." *AI and Statistics*, vol., n.

Menczer, F., Ed. (2003). Complementing search engines with online web mining agents, Elsevier Science Publishers B. V.

Milgram, S. (1967). "The small-world problem." *Psychology Today*, vol. 1, n.: 61-67.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii et U. Alon (2002). "Network motifs: simple building blocks of complex networks." *Science*, vol. 298, n.: 824-827.

Mislove, A., M. Marcon, K. P. Gummadi, P. Druschel et B. Bhattacharjee (2007). "Measurement and Analysis of Online Social Networks". 7th ACM SIGCOMM Internet Measurement Conference (IMC), San Diego, California, USA.29-42

Mitleton-Kelly, E. (1997). "Organisations as Co-Evolving Complex Adaptive Systems". British Academy of Management Conference

Mohar, M. (1991). "The Laplacian spectrum of graphs." *Graph theory, combinatorics, and applications*, vol. 2, n.: 871-898.

Morowitz, H et J. Singer (1995). The Mind, The Brain and Complex Adaptive Systems. Santa Fe Institute Studies in the Sciences of Complexity Proceedings.

Moukas, A. (1997). "Amalthea: Information Filtering and Discovery using a Multiagent Evolving System." *Journal of Applied Artificial Intelligence*, vol. 11, n. 5: 437-457.

Müller, M. E. (1999). An Intelligent Multi-Agent Architecture for Information Retrieval from the Internet.

Nabeth, T., A. A. Angehrn, P. K. Mitta et C. Roda (2005). "Using artificial agents to stimulate participation in virtual communities". IADIS International Conference CELDA (Cognition and Exploratory Learning in Digital Age)

Ng, A., M. Jordan et Yair Weiss (2002). "On spectral clustering: Analysis and an algorithm." In Proceedings of Advances in Neural Information Processing Systems

Nitto, E. D., C. Ghezzi et P. Selvini (2003). Using agents for multi-target search on the Web. SAC '03: Proceedings of the 2003 ACM symposium on Applied computing, ACM Press: 828--833.

Odell, J., M. Nodine et R. Levy (2005). "A Metamodel for Agents, Roles, and Groups." *Agent-Oriented Software Engineering (AOSE) Lecture Notes on Computer Science*, vol. LNCS 3382, n.

Odell, J., H. V. D. Parunak et M. Fleischer (2003). "The Role of Roles in Designing Effective Agent Organizations". *Software Engineering for Large-Scale Multi-Agent Systems, Lecture Notes on Computer Science*, Springer.27-38

Odell, J., H. V. D. Parunak, M. Fleischer et S. Breuckner (2002). "Modeling Agents and their Environment". *Agent-Oriented Software Engineering (AOSE) III, Lecture Notes on Computer Science*.16-31

Page, L., S. Brin, R. Motwani et T. Winograd (1998). *The pagerank citation ranking: Bringing order to the web*. S. D. L. T. Project.

Pant, G. et F. Menczer (2002). " MySpiders: Evolve your own intelligent Web crawlers". *Autonomous Agents and Multi-Agent Systems*.221-229

Parent, S., B. Mobasher. et S. Lytinen (2001). "An Adaptive Agent for Web Exploration Based on Concept Hierarchies". *9th International Conference on Human Computer Interaction, New Orleans*

Payne, T. R., R. Singh et K. Sycara (2002). "Calendar Agents on the Semantic Web." *IEEE Intelligent Systems*, vol. 17, n. 3: 84-86.

Pazzani, M. J. et D. Billsus (2002). "Adaptive Web Site Agents." *Autonomous Agents and Multi-Agent Systems*, vol. 5, n. 2: 205--218.

Perkowitz, M. et O. Etzioni (2000). "Towards adaptive Web sites: Conceptual framework and case study." *Artificial Intelligence*, vol. 118, n. 1-2: 245-275.

Preece, J. (2000). Online Communities: Designing Usability, Supporting Sociability, John Wiley & Sons.

Rheingold, H. (1993). The virtual community:Homesteading on the Electronic Frontier, Addison-Wesley.

Rupert, M., S. Hassas, C. Li et R. Paweska (2007). "Social and Spatial Organisations in Multi-Agent Systems". *IADIS International Conference Intelligent Systems and Agents, Lisbon, Portugal*

Rupert, M., S. Hassas, C. Li et J. Sherwood (2007). "Simulation of Online Communities Using MAS Social and Spatial Organisations." *International Journal of Computational Intelligence*, vol. 4, n. 3: 183-188.

Rupert, M., S. Hassas et A. Rattout (2006). "The Web and Complex Adaptive Systems". 20th International Conference on Advanced Information Networking and Applications (AINA 2006), Vienna, Austria, IEEE Computer Society Press ed.:200-204

Rupert, M., C. Li et S. Hassas (2008). "An Organisational Multi-Agent Systems Approach for Designing Collaborative Tagging Systems." IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Sydney, Australia, WI-IAT.114-117

Rupert, M., A. Rattout et S. Hassas (2008). "The web from a Complex Adaptive Systems Perspective." *Journal of Computer and Systems Sciences - Elsevier*, vol. 74, n. 2: 133-145.

Salton, G. et Buckley (1987). Text Weighting Approaches in Automatic Text Retrieval, Cornell University Technical Report: 870-881.

Sangesa, R. et J. M. Pujol (2001). "NetExpert: A multiagent system for expertise location". Workshop on Organizational Memories and Knowledge Management

Serugendo, G. D. M., M.-P. Gleizes et A. Karageorgos (2005). "Self-organization in multi-agent systems." *Knowl. Eng. Rev.*, vol. 20, n. 2: 165-189.

Shen, K. et L. Wu (2005). "Folksonomy as a Complex Network." *CoRR*, vol. abs/cs/0509072, n.

Siebert, J., L. Ciarletta et V. Chevrier (2008). "Impact du comportement des utilisateurs dans les réseaux pair-à-pair (P2P) : modélisation et simulation multi-agents." Journées Francophones des Systèmes Multi-Agents JFSMA08, Brest, France, Cépaduès.129-138

Small, P. (2004). Stigmergic Systems, online at <http://www.stigmergicsystems.com/index.html?>

Stratulat, T., J. Ferber et J. Tranier (2009). "MASQ: towards an integral approach to interaction." Proceedings of 8th Int. Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), Budapest, Hungary. 813-820

Tamma, V., B. Blacoe, B. Lithgow-Smith et M. Wooldridge (2004). "SERSE: Searching for Semantic Web Content." Sixteenth European Conference on Artificial Intelligence (ECAI-04), Valencia, Spain

Van Harmelen, F. (2004). "The Semantic Web:a meaningful habitat for agents?" European Multi-Agent Symposium, Barcelona, Spain

VanderWal, T. (2005). Off the Top: Folksonomy Entries <http://www.vanderwal.net/random/category.php?cat=153>. 2008.

Varela, F., E. Thompson et E. Rosch (1991). The Embodied Mind: Cognitive Science and Human Experience. Cambridge, MA, MIT Press.

Vázquez, A., I. Barrio, J. Vázquez-Salceda, J. Pujol et R. Sangüesa (2001). "An agent-based Collaboratory." CCIA

Vazquez-Salceda, J., V. Dignum et F. Dignum (2005). "Organizing Multiagent Systems." *Autonomous Agents and Multi-Agent Systems*, vol. 11, n. 3: 307-360.

vonLuxburg, U. (2006). A Tutorial on Spectral Clustering, Max Planck Institute for Biological Cybernetics.

Waldrop, M. (1992). Complexity: The Emerging Science at the Edge of Order and Chaos. New York, Simon and Schuster.

Wasserman, S. et K. Faust (1994). Social Network Analysis: Methods and Applications., Cambridge University Press.

Watts, D. (1999). Small Worlds: The Dynamics of Networks Between Order and Randomness, Princeton University Press.

Weick, K., K. Sutcliffe et D. Obstfeld (2005). "Organizing and the Process of Sensemaking." *Organization Science*, vol. 16, n. 4: 409-421.

Weyns, D., A. Omicini et J. Odell (2007). "Environment as a first class abstraction in multiagent systems." *International Journal on Autonomous Agents and Multi-Agent Systems*, vol. 14, n. 1: 5-30.

Weyns, D., V. D. Parunak, F. Michel, T. Holvoet et J. Ferber (2005). "Environments for Multiagent Systems, State-of-the-Art and Research Challenges." *Environments for Multiagent Systems Lecture Note in Artificial Intelligence LNAI Springer*, vol. 3477, n.

Weyns, D., M. Schumacher, A. Ricci, M. Viroli et T. Holvoet (2005). "Environments in multiagent systems." *Knowl. Eng. Rev.*, vol. 20, n. 2: 127-141.

Wikipedia (2007). http://wn.wikipedia.org/wiki/Virtual_community.

Yanbe, Y., A. Jatowt, S. Nakamura et K. Tanaka (2007). "Can social bookmarking enhance search in the web?" Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, Vancouver, BC, Canada, ACM.107-116

Zheng-guang, W., L. Xiao-hui et Z. Qin-ping (2006). "Adaptive mechanisms of Organizational Structures in Multi-agent Systems". Ninth Pacific Rim International Workshop on Multi-Agents, PRIMA 2006, LNAI.471-477