Projet
Web Intelligence
du Cluster ISLE

École d'été Web Intelligence 2009

sur la thématique des

« Réseaux Sociaux »

wi2009

6 au 10 JUILLET

Véranne - France

Centre G2I - Laboratoire Hubert Curien

Coordinateurs

Philippe Jaillon - Franck Thollard

http://ecole.web-intelligence-rhone-alpes.fr

Ecole Nationale Supérieure des Mines SAINT-ETIENNE

Région RhôneAlpes

Université Jean Monnet SAINT-ETIENNE

École d'été
Web Intelligence 2009

« Réseaux Sociaux »

Actes WI 2009

6-10 juillet 2009

Véranne, Loire

coordinateurs

Philippe Jaillon, ENS Mines de Saint-Étienne
Frank Thollard, Université Jean Monnet

Comités

Coordination

Ph. Jaillon, EMSE
F. Thollard, LabHC

Comité scientifique

S. Calabretto, LIRIS-INSA
M.C. Fauvet, LIG
E. Gaussier, LIG
M. Hacid, LIRIS
F. Jacquenet, LabHC
P. Maret, LabHC
M. Occello, LCIS
M.C. Rousset, LIG

Comité d'organisation

Ph. Beaune, EMSE
M. Beigbeder, EMSE
O. Boissier, EMSE
S. Calabretto, LIRIS-INSA
M.C. Fauvet, LIG
D. Genthial, LCIS
M. Géry, LabHC
M.Ph. Huget, LISTIC-Polytech'Savoie
C. Largeron, LabHC
P. Maret, LabHC
A. Mercier, LCIS
G. Picard, EMSE
L. Vercouter, EMSE

Comité d'organisation des doctoriales

L. Crépin, LIG
A. Ailt Bachir, LIG
Y. Krupa, EMSE
C. Moulin, LabHC

Programme

Lundi 6 juillet

11h30	Départ de la navette à destination de Véranne (esplanade de la gare SNCF de Saint-Étienne Châteaucreux)
12h00	Accueil des participants au château du Buisson Repas prévu sur place
14h00	Réseaux sociaux : révolution des usages sur Internet, et nouvelles problématiques – P. Maret (Univ. Saint-Étienne), A. Joly (Alcatel Lucent)
16h00	"REZO" un réseau social au service des usagers de la SNCF – C. Garnier (CRMServices)
18h00	Table ronde organisée avec Loire Numérique – Les réseaux sociaux au service de la stratégie de l'entreprise : potentialités et limites – avec P. Bizollon (LinkingBrand), G. Mathoulin (Effitel), B. Laval (RaidLight) et C. Garnier (CRMservices)

Mardi 7 juillet

9h00	Réseaux Sociaux : Structure et Sémantique – H. Hacid, J Stan (Alcatel Lucent, France)
14h00	Le Forage des Réseaux Sociaux et la découverte de communautés virtuelles – O. R. Zaïane (Univ. Alberta, Canada)
20h00	Doctoriales

Mercredi 8 juillet

9h00	Les logiciels Ucinet et Netdraw – D. Frachisse (Univ. Jean Monnet, Saint-Étienne)
11h00	Impact du web communautaire sur les missions de la DCRI – B. Sueur (Direction Centrale du Renseignement Intérieur)
Après-midi	Activités sociales – course d'orientation Visite de la Brasserie du Pilat

Jeudi 9 juillet

9h00	Web sémantique et Web social – F. Gandon (Inria, France)
14h00	Simulation multi-agent de réseaux sociaux : des modèles classiques à leur application au Web 2.0 – F. Amblard (Université de Toulouse)

Vendredi 10 Juillet

9h00	Prédiction de liens dans les réseaux sociaux – R. Kanawati, C. Rouveirol (LIPN Paris 13)
11h00	10 propriétés des coopérations faibles : analyse des réseaux sociaux en ligne – D. Cardon, M. Crépel, B. Hatt, N. Pissard, C. Prieur (Orange Labs)
13h00	Clôture de l'école d'été Repas prévu sur place
14h30	Départ de la navette à destination de la gare SNCF Saint-Étienne Châteaucreux (arrivée prévue 15h30)

Table des matières

P. JAILLON, F. THOLLARD

Avant-propos 1

Partenaires 5

Table ronde organisée avec Loire Numérique 7

USAGES (DU POINT DE VUE UTILISATEURS) **17**

A. JOLY (ALCATEL LUCENT, FRANCE), P. MARET (UNIVERSITÉ JEAN MONNET, FRANCE)

Réseaux sociaux : révolution des usages sur Internet, et nouvelles problématiques . . . 21

C. GARNIER (CRMSERVICES SNCF, FRANCE)

Le Web 2.0 moteur de la transformation SNCF 43

RÉSEAUX SOCIAUX : STRUCTURE ET SÉMANTIQUE **65**

H. HACID, J. STAN (ALCATEL LUCENT, FRANCE)

Réseaux sociaux : structure et sémantique 69

FOUILLE DE DONNÉES **117**

O. R. ZAIANE (UNIVERSITY OF ALBERTA, CANADA)

Le forage des réseaux sociaux et la découverte de communautés virtuelles 121

R. KANAWATI, C. ROUVEIROL (LIPN, PARIS 13, FRANCE)

Prédiction de liens dans les réseaux sociaux 149

ANALYSE DE RÉSEAUX **185**

D. FRACHISSE (UNIVERSITÉ JEAN MONNET, FRANCE)

Les logiciels Ucinet et Netdraw 189

SÉCURITÉ **219**

B. SUEUR (DIRECTION CENTRALE DU RENSEIGNEMENT INTÉRIEUR, FRANCE)

Impact du web communautaire sur les missions de la DCRI 221

WEB SÉMANTIQUE ET WEB SOCIAL **223**

F. GANDON (INRIA, FRANCE)

Web sémantique et Web social 227

SIMULATION ET AGENTS **263**

F. AMBLARD (UNIVERSITÉ DE TOULOUSE, FRANCE)

Simulation multi-agent de réseaux sociaux : des modèles classiques à leur application au Web 2.0 267

USAGES (DU POINT DE VUE SOCIAL) **351**

D. CARDON (ORANGE LABS, FRANCE)

Dix propriétés des coopérations faibles : analyse des réseaux sociaux en ligne 355

DOCTORIALES **397**

INDEX DES AUTEURS **419**

Avant-propos

Pour cette édition 2009, l'objectif de l'école est d'aborder les principales recherches effectuées en informatique dans le domaine du Web Intelligence en lien avec la problématique des réseaux sociaux. Les enjeux soulevés par l'essor de cette véritable "industrie de la relation sociale" sont nombreux, d'ordres à la fois scientifiques, sociaux et économiques. En effet, le développement des communautés en ligne est un phénomène qui se manifeste dans de multiples secteurs de la société et qui dépasse largement le cercle restreint de la sphère individuelle pour toucher le monde de l'entreprise, sous des formes très diverses. Il s'agit dans cette école d'été de donner aux participants une vision cohérente et actuelle des problèmes scientifiques, des outils et des services qui permettent d'instrumenter la vie des communautés et soutenir l'activité coopérative, mais aussi des multiples questions d'usages qui apparaissent, en terme d'opportunités et de menaces potentielles.

En plus d'un panorama et d'une réflexion sur les usages des réseaux sociaux tant d'un point de vue sociétal qu'économique, nous souhaitons développer au sein de l'école l'expertise technique et scientifique pour étudier la problématique des réseaux sociaux relativement à différents domaines relevant du Web Intelligence. L'objectif est de considérer l'apport de ces techniques pour les réseaux sociaux (par exemple, la découverte de nouveaux liens entre éléments du réseau par des techniques de fouilles de données), ainsi que l'apport des réseaux sociaux à ces domaines (par exemple, les nouvelles perspectives ouvertes en recherche d'information collaborative dans le contexte des réseaux sociaux).

PHILIPPE JAILLON
G2I, ENS Mines de Saint-Étienne

FRANCK THOLLARD
LabHC, Université Jean Monnet

Partenaires

Table ronde organisée avec Loire Numérique

Les réseaux sociaux au service de la stratégie de l'entreprise : potentialités et limites

La table ronde s'intéressera tout particulièrement aux relations, aux ressources humaines et à la capitalisation des connaissances. Y participeront G. Mathoulin (Effitel), P. Bizollon (LinkingBrand), B. Laval (RaidLight) et C. Garnier (CRMServices, SNCF).

Loire Numerique
Le réseau des professionnels du numérique
CREATEURS D'AVENIRS

Loire Numerique
Le réseau des professionnels du numérique
Les fondements de l'association
Contribuer à structurer la filière TIC dans la Loire par la mise en place d'une réflexion partagée associant l'ensemble des décideurs et acteurs du développement économique
Donner une lisibilité à cette filière et contribuer à sa valorisation
Favoriser l'innovation dans la filière par l'élaboration de projets communs et la mutualisation des moyens et ressources
Anticiper les évolutions à venir et proposer en conséquence toute solution permettant d'augmenter le niveau d'activité et d'efficacité des entreprises de la filière TIC

Loire Numerique
Le réseau des professionnels du numérique

L'historique

Aujourd'hui, Loire Numérique compte 138 adhérents

(4400 emplois – 400 millions d'Euros de CA)

2006 LOIRE NUMERIQUE est labellisée SPL (Système Productif Local) par la DIACT

2005 Le Club des Acteurs du Numérique devient LOIRE NUMERIQUE

2002 Naissance du Club des Acteurs du Numérique de la Loire

Loire Numerique
Le réseau des professionnels du numérique

Les permanents

Armand LULKA
Directeur

SAINT ETIENNE

ROANNE

Brigitte MARANDON
Responsable communication et développement

Roselise PEILLON
Animateur(rice) de réseau et du télépôle

Virginie VEGA
Chargée de Mission Innovation

Salem NAIT IDIR
Chargé de mission intelligence économique

Lionel GENEIX
Chargé de mission e-administration

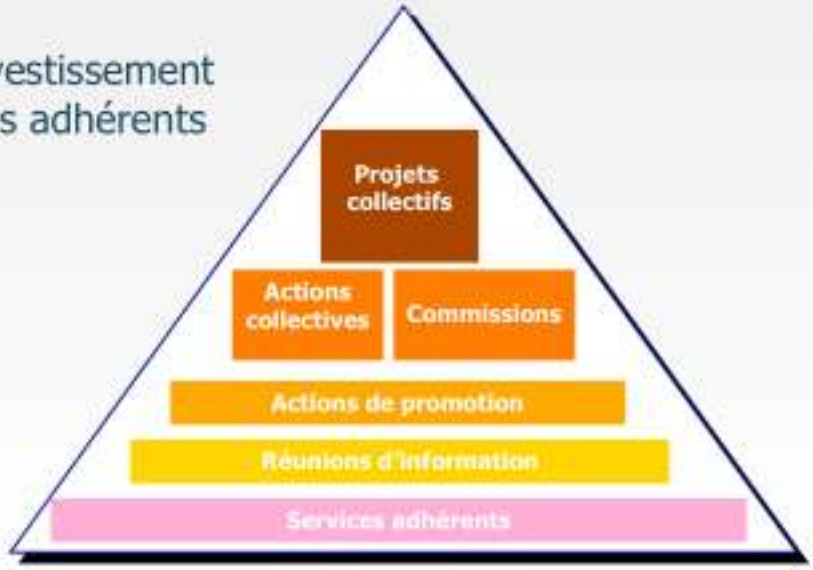
Charlène LAURENT
Chargée de recherche

Selma KANTAR
Assistante de gestion

Loire Numerique
Le réseau des professionnels du numérique

Un large panel d'actions « Réseau »

Investissement des adhérents



Loire Numerique
Le réseau des professionnels du numérique

Les services adhérents

Veille des marchés hebdomadaire

Mise à disposition de ressources matérielles

Relations presse
Campagnes presse

Mutualisation d'achats d'espaces publicitaires

Veille technologique

Loire Numerique
Le réseau des professionnels du numérique

L'animation du réseau

THEM@S : Six événements avec des ateliers à la carte.

Musclez votre développement -Travaillez la technique (Part 1 Web Part 2 Data) - Affinez votre image - Managez vos équipes - Entourez-vous d'experts

TIC EXCHANGE : Rencontres « speed-meeting business » pour favoriser les échanges.

21 avril (Roanne) 15 septembre (St-Etienne)

Loire Numerique
Le réseau des professionnels du numérique

Les actions de promotion

Network : Événement visant à valoriser des compétences spécifiques
WebdiagnosTic – Logiciels GRC – Infogérance/Externalisation

Participation à des salons : SITL, MAP, RuraliTIC, Salon des maires, Cebit, JRC, Forum du THD, Forum innovation tourisme, Journée innovation, RV de l'offre, Forum Carrières, Rencontres d'Autrans

Charte des bonnes pratiques : Déjà plus de 60 entreprises signataires

Charte de bonnes pratiques des prestataires TIC

Loire Numerique
Le réseau des professionnels du numérique
Les actions collectives
Formations inter entreprises
Gestion de projet « avancée » (28h)
Gagner des marchés publics (7h)
ITIL V3 (21h)
Contrats Informatiques (7h)
Communication Interpersonnelle (14h)
Management opérationnel (21h)
Actions Cluster :
Croissance Éditeurs Rhône-Alpes
Marchés Publics Editeurs
Loire Numérique est déclarée
Organisme de Formation
Loire Numérique est Membre fondateur
ClusterEdit
>_ Software in Rhône-Alpes

Loire Numerique
Le réseau des professionnels du numérique
Les commissions
RIO : Réseaux Infrastructures Opérateurs
ATOLL : Accompagnement TOurisme et Loisirs en Ligne
Formation: E-learning
R&D : Comité d'experts laboratoires et universités
International : Développement à l'international
Relation Clients : Logiciels GRC
Centre de contacts
Éditeurs de Logiciels : Relais du Cluster EDIT
6: Sécurité de l'Information et des Systèmes

Loire Numerique
Le réseau des professionnels du numérique

Les labels de Loire Numerique

ODYSSEE
NEOTESS
PROXITV
FRENCHOME
STN
FIDELIUM
P2P EVENTS
TWIZY
NATOO
HELIOCIS
EFICUS
GESTION DES DECHETS

Loire Numerique
Le réseau des professionnels du numérique

Les Actions « Territoire »

Centre de ressources Expert en Téléactivité

Véritable enjeu économique et social, en croissance régulière, le télétravail concerne déjà 7% des salariés en France, contre 25% dans certains pays européens, et procure des avantages tant aux entreprises qu'aux salariés et à la collectivité. Loire Numérique s'est fixée pour objectif de favoriser le télétravail sur le département en menant des actions de promotion de ces usages.

Loire Numerique
Le réseau des professionnels du numérique
Les Actions « Territoire »
E-administration
En partenariat avec le Télépôle du Grand Roanne, labellisé Centre de Ressources Expert en e-administration par le Conseil Général de la Loire, Loire Numérique a pour mission de réaliser la promotion et le déploiement d'un logiciel de dématérialisation du contrôle de légalité auprès de l'ensemble des collectivités de la Loire.
Loire Numerique
GRAND ROANNE AGGLOMERATION
CONSEIL GENERAL LOIRE

Loire Numerique
Le réseau des professionnels du numérique
Les Actions « Territoire »
Veille des appels à projets
Loire Numérique réalise une veille des appels à projets (européens, nationaux et régionaux) au bénéfice de différents pôles et filières de la Loire dans l'objectif de favoriser la coopération et l'émergence de projets entre les pôles.
Loire Numerique
ora
CIBIODE
PTI
MECA LOIRE
CONSEIL GENERAL LOIRE

Les Actions « Territoire »

GAX

Loire Numérique collabore avec l'ENISE, la CCI, le Conseil Général de la Loire, l'ISTASE et Lotim Telecom sur un projet de mutualisation de logiciels métiers pour les PME.

Les Actions « Territoire »

EUROP

Loire Numérique est partenaire de l'ARUFOG notamment dans le cadre du projet ALO (Architecture Liaison Optique), piloté par l'Institut Télécom, visant à créer une plate-forme technologique ouverte aux utilisateurs industriels et universitaires.

SESSION 1

Usages (du point de vue utilisateurs)

Réseaux sociaux : révolution des usages sur Internet, et nouvelles problématiques

A. JOLY (ALCATEL LUCENT, FRANCE),
P. MARET (UNIVERSITÉ JEAN MONNET, FRANCE)

Réseaux Sociaux

Révolution des usages sur Internet,

et nouvelles problématiques de recherche

Pierre Maret, Laboratoire Hubert Curien, St-Etienne
Adrien Joly, Alcatel-Lucent Bell Labs France, Villarceaux / Laboratoire LIRIS, Lyon

Ecole d'été Web Intelligence WI09, Juillet 2009

All Rights Reserved © Alcatel-Lucent 2009

Réseau Sociaux

Introduction de la présentation

- Qu'est-ce qu'un Réseau Social ?
- Quelles sont les caractéristiques communes des Réseaux Sociaux actuels ?
- Comment sont-ils utilisés ?
- Jusqu'où iront-ils ?
- Quelles opportunités de recherche ?

All Rights Reserved © Alcatel-Lucent 2009

Réseau Sociaux

Plan

1. Historique
2. Définition et fonctionnalités
3. Usages actuels
4. Tendances
5. Problématiques de recherche
    1. Modélisation en graphe
    2. Algorithmes
    3. Fonctions
    4. Contenus

All Rights Reserved © Alcatel-Lucent 2009

1 Historique

Evolution des BBS aux plateformes sociales sur le web

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances Problématiques

De Milgram aux Réseautage Social

"6 degrés de séparation"

- Expérience "petit monde" de Stanley Milgram, 1967
- 2 personnes choisies aléatoirement sont reliées, tout au plus, par 6 personnes intermédiaires

Communication communautaire

- Profiles utilisateurs (photos…)
- Canal "many-to-many"
- Messages en mode informel
  - Information, pas obligation
- Visibilité pour étendre le réseau

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances Problématiques

Réseaux Sociaux Informatiques

EIES: Electronic Information Exchange System
Première implémentation de logiciel social, décrit par M. Turoff, 1972

Groupware et CSCW
Réseaux employés pour améliorer la communication et collaboration dans la recherche, l'éducations et d'autres motivations professionnelles

Popularisation des modems
Communication et échanges de données sur les BBS (Bulletin Board Systems), l'IRC (Internet Relay Chat), le Minitel puis sur sites Internet

Internet social
Forums, chats, avis et recommandations, profiles utilisateurs, partage viral de contenu, réseaux sociaux

Proximité perpétuelle ?
La communication devient omniprésente, immédiate et transparente, quelle que soit la distance entre les personnes.

[redacted]

IRC

BBS

July 1980: "Groupware: The Emerging Art of Orchestrating Collective Intelligence"

Groupware

Computer-Supported Cooperative Work

Minitel

Forums

Web 2.0

Web 1.0

Internet pour tous

Social Networking Sites

Ambient Awareness ?

Informatique omniprésente

Réseaux sans-fils

Internet mobile

Téléphonie mobile

1[redacted]

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances Problématiques

Sites de réseaux sociaux incontournables 1/2

Au delà du simple échange de messages, les sites de réseaux sociaux…

- 2002: Friendster.com
  - Contrairement aux forums, les discussions ont lieu sur les profiles d'utilisateurs
  - Le but est d'afficher sa personnalité pour faire de nouvelles rencontres
- 2003: Myspace.com
  - Le profile devient une page complètement personnalisable (design, musique, vidéos)
  - Octobre 2005: 4ème site le plus consulté au monde, après Yahoo!, AOL et MSN; puis Google en 2006
  - Avril 2008: plus de 230 182 000 utilisateurs
- 2004: Thefacebook.com (renommé facebook.com en 2005)
  - Limité aux étudiants d'Harvard jusqu'en 2006, puis traduit collaborativement en 50 langues
  - Octobre 2007: Microsoft achète 1,7% de Facebook pour 240 M$ ➔ valeur totale: 15 milliards $
  - Avril 2009: plus de 200 millions de membres, dont 9 millions en France

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances Problématiques

Sites de réseaux sociaux incontournables 2/2 – Les plateformes

… puis les sites deviennent des plateformes d'applications sociales:

- Facebook.com évolue
  - Sept. 2006: Le "news-feed" résume les interactions de son réseau social ➔ révolte
  - Mai 2007: Plateforme ouverte de développement d'application sociales
- 2006: Twitter.com
  - Un "news-feed" public où les utilisateurs partagent leurs pensées en 140 caractères
  - Diffusion en temps réel, et accessible par SMS ➔ utilisation en mobilité
  - Aujourd'hui enrichi par de nombreuses applications (ex: twitpic pour les photos)
  - Vecteur inestimable pour le marketing social et la fouille de données continue
- 2009: Ning.com
  - Une plateforme de création et d'hébergement de réseaux sociaux

All Rights Reserved © Alcatel-Lucent 2009

2 Dessine moi un réseau social

Définition et fonctionnalités

All Rights Reserved © Alcatel-Lucent 2009

Définition par comparaisons 1/2

Quelle différence avec…

- *Les appels téléphoniques*

  - ➔ Moins intrusif et personnel
  - ➔ Permet de partager avec plusieurs personnes de manière asynchrone

- *Le courrier postal*

  - ➔ Moins personnel mais plus riche et interactif
  - ➔ Permet de communiquer avec des centaines de personnes à la fois

- *Le courrier électronique (e-mail)*

  - ➔ Moins formel; la liste des interactions n'a pas à être « dépilée »
  - ➔ Les destinataires sont l'intégralité du réseau.

All Rights Reserved © Alcatel-Lucent 2009

Historique **Définition** Usages Tendances Problématiques

Définition par comparaisons 2/2

Quelle différence avec…

- *La messagerie instantanée (chat) et IRC*

  - ➔ Parfois similaire, souvent à tort…
  - ➔ Comme le chat, suscitation de réponses à des messages diffusés
  - ➔ Pas forcément synchrone, on ne s'adresse pas aux personnes connectées mais à l'ensemble du réseau social
  - ➔ Chaque utilisateur ne voit que les message de ses contacts

- *Les forums et autres espaces de discussion (usenet…)*

  - ➔ Très similaire! …sauf que:
  - ➔ Discussions centrées sur les gens, et non sur un sujet (du moins, au départ)
  - ➔ Abonnement aux notifications de gens, et non aux salons/discussions

All Rights Reserved © Alcatel-Lucent 2009

Historique **Définition** Usages Tendances Problématiques

Principes de base

- Communication basée sur le partage et les commentaires
  - « Status » : dernière activité, localisation, humeur, anecdote, annonce…
  - Contenu personnel ou représentatif: photos, vidéos, liens internet
  - Les messages (publics) sont des commentaires sur le contenu et/ou le profile
- Diffusion d'interactions/événements sous forme de liste chronologique
  - On suit les nouvelles de son réseau social comme on lit un journal d'actualités
  - Les messages plus personnels peuvent être notifiés par email/sms
- Visibilité et suivi basé sur liste de contacts
  - Lien « d'amitié » mutuel (avec confirmation: Facebook) ou pas (Twitter)

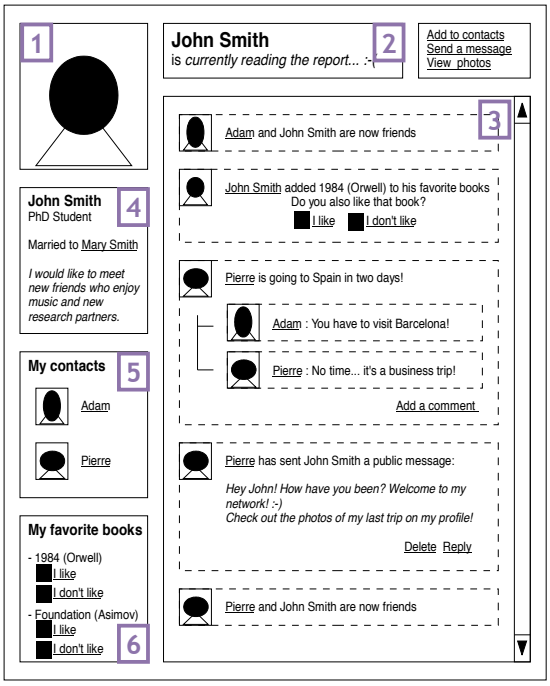
All Rights Reserved © Alcatel-Lucent 2009

Historique **Définition** Usages Tendances Problématiques

Le profile utilisateur

Le profile utilisateur est à la fois le tableau de bord personnel de son réseau social et un espace semi-public d'interaction / de discussion.

1. Photo /avatar (très important!)
2. Message d'état / "status"
3. Historique / le "mur"
4. Biographie / brève présentation
5. Liste de contacts
6. Contenu et applications



All Rights Reserved © Alcatel-Lucent 2009

3 Usages actuels

Twitter, LinkedIn, Myspace, Facebook, Couchsurfing

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition **Usages**

Les messages d'état / « status »

Le cas de Twitter:

- 140 caractères pour partager son activité, humeur, localisation, anecdote… avec son réseau social
- Messages diffusés en temps réel
- Public mais filtré par liste contacts
- Possibilité de répondre, citer, tagger
- Usages:
  - Partage viral de contenus d'intérêt
  - Actualité en temps réel (désastres…)
  - Sondages d'opinions
  - Diffusion d'intentions
  - Montrer ses gouts, ses préférences

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition **Usages** Tendances Problématiques

Valorisation « sociale » d'expertise

Le cas de LinkedIn Answers:

- Le CV est remplacé par un profile web
- Le réseau professionnel est valorisé:
  - Par les recommandations de collègues
  - Possibilité d'être mis en relation avec le réseau de 2ème/3ème degré
- Partage de plaquettes, présentations
  - Les collègues sont notifiés
- Forum où les réponses sont valorisées
  - Les réponses de bonne qualité sont promues sur le profile de leur auteur, pour prouver son expertise dans un domaine.

Une question est diffusée sur le réseau…

…les réponses sont valorisées sur le profile

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition **Usages** Tendances Problématiques

Se faire connaître et rester en contact avec ses « fans »

Le cas de Myspace:

- Artistes, marques et personnes expriment leur identité sur un profile
  - Décoration/thème et musique de fond
  - Images, vidéos… et liste d'amis / de fans
- Le profile est un espace de discussion
  - Echange de nouvelles, « coucou ! »
  - Promotion (ex: nouvel album, concert)
  - Messages de courtoisie
- Une plateforme de notifications
  - « XXX a ajouté SOMA à sa liste d'amis »
  - « SOMA joue à Paris demain soir »
  - « XXX a posté une nouvelle musique »

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition **Usages** Te

Rester en contact avec les amis / connaissances

Le cas de Facebook:

- Un réseau pour « amis »
  - « L'amitié » est mutuelle (confirmation)
  - Possibilité de modérer la visibilité du contenu (par groupes)
  - Cadeaux, « pokes »…
- News-feed: un « Twitter enrichi »
  - Diffusion de contenu en temps réel
  - Vidéos, photos, liens, résultats quizz…
  - Calendrier d'événements « sociaux »
  - Plateforme d'applications « sociales »
- Impacts sociaux
  - « Fiançailles facebook » (relation)
  - Noms déformés (anonymisation)
  - « Vous avez été taggé sur une photo! »

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition **Usages** Tendances Problématiques

Voyager, organiser des sorties, faire des rencontres

Le cas de Couchsurfing:

- L'idée
  - *Je t'héberge; un autre m'hébergera*
- Le site est un outil
  - Les voyageurs visitent le profile de leurs futurs hôtes
  - Ils laissent un commentaire public pour retour d'expérience et remerciement
- Esprit communautaire
  - Réseau social de personnes ayant des valeurs / une philosophie commune
    ➔ plus grande confiance
  - Organisation d'événements par ville
  - Groupes (forums) par thématique
  - Les membres les plus investis peuvent devenir « ambassadeurs » de leur ville

All Rights Reserved © Alcatel-Lucent 2009

4 Tendances

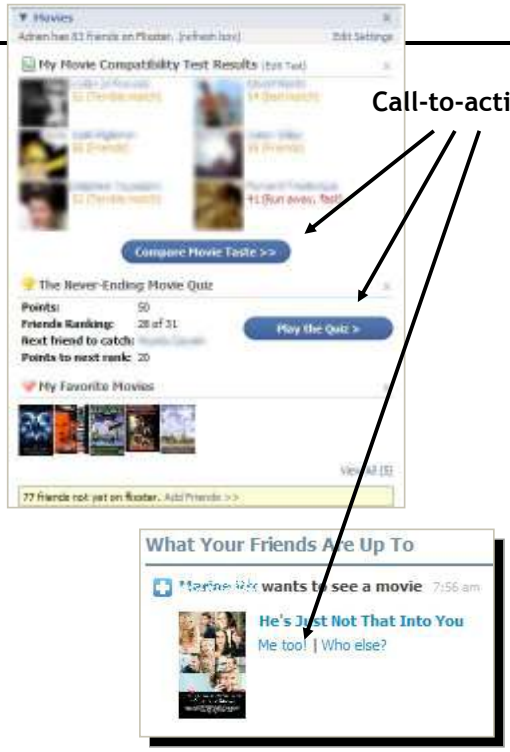
Social apps, stream mining, mobilité…

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages **Tendances** Problématiques

« Social apps » 1/2

- Application sociale ?
  - Jeux, quizz, services de partage et de recommandation, de rencontres…
  - Pour l'utilisateur: la valeur naît de son utilisation / sa visibilité par plusieurs personnes (grâce au « news-feed »)
  - Pour le développeur: la valeur est tirée de l'attention qu'elle suscite (ex: publicité)
  - Pour en bénéficier, l'utilisateur donne accès à son profile et sa liste d'amis
  - Se diffuse par recherche, invitation, et surtout « call-to-action »

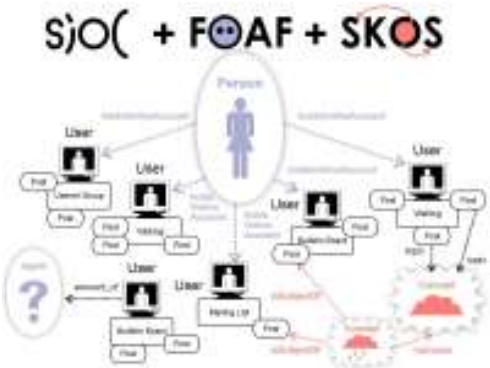


All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages **Tendances** Problématiques

« Social apps » 2/2

- Plateformes et standardisation
  - Facebook Apps: ouverture d'une API pour accéder à la plateforme et ses données
  - Open Social: réponse de Google à Facebook; construction ouverte d'un standard pour le développement d'applications sociales
  - Data portability: consortium discutant d'interopérabilité entre réseaux sociaux
- Formats et initiatives alternatives
  - FOAF / SIOC: Formats sémantiques (RDF) permettant de définir un profile social sur le web, et de fédérer les discussions intervenant sur des plateformes hétérogènes
  - DiSo project: développement de blocs pour un web social décentralisé et ouvert



All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages **Tendances** Problématiques

« Stream mining »

- Twitter, flux de données à fouiller
  - ~200 messages diffusés par seconde!
  - Twitpocalypse (2 147 483 647ème tweet) a eu lieu le 12 juin dernier!
- Exemples d'utilisations
  - Nuage de tags de sujets discutés en temps réel (mots clés)
  - Signalement de désastres (crash…)
  - Météo observée
  - Fouille d'opinion

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages **Tendances** Problématiques

Réseautage social en mobilité

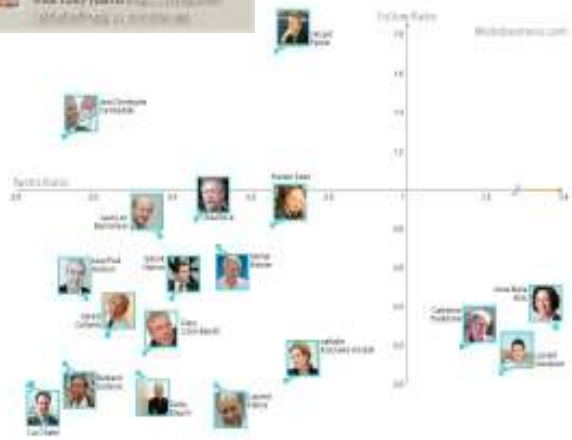
- Premières (vaines) tentatives
  - Google Dodgeball, Nokia Sensor…
  - Offre trop fragmentée
    ➔ services avec trop peu d'utilisateurs
  - Twitter (par SMS) est un peu pauvre
- Nouveaux terminaux (iPhone…)
  - Localisation d'amis
  - Planification d'activités en groupe
  - Partage d'avis (bars, restaurants…)
  - Photos/vidéos géolocalisées en temps réel
- Et demain ?
  - Rendre notre réseau social pro-actif (ou moins contraignant à utiliser) en fonction du contexte de l'utilisateur

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages **Tendances** Problématiques

Les réseaux sociaux redonnent le pouvoir au peuple ?

- Usage des réseaux sociaux pour les élections présidentielles
  - Barack Obama
  - Ségolène Royal VS Nicolas Sarkozy
- Twitter s'invite dans la politique
  - Sénateurs américains sur Twitter pour s'adresser directement aux citoyens
  - Ministres anglais sur Twitter
  - France ? Moins de 20!
  - Problème du cybersquatting…
- Facebook devient une démocratie
  - Suite à plaintes, utilisateurs invités à voter pour les termes de service
  - 60 000 votants sur 60 M. util. actifs…



All Rights Reserved © Alcatel-Lucent 2009

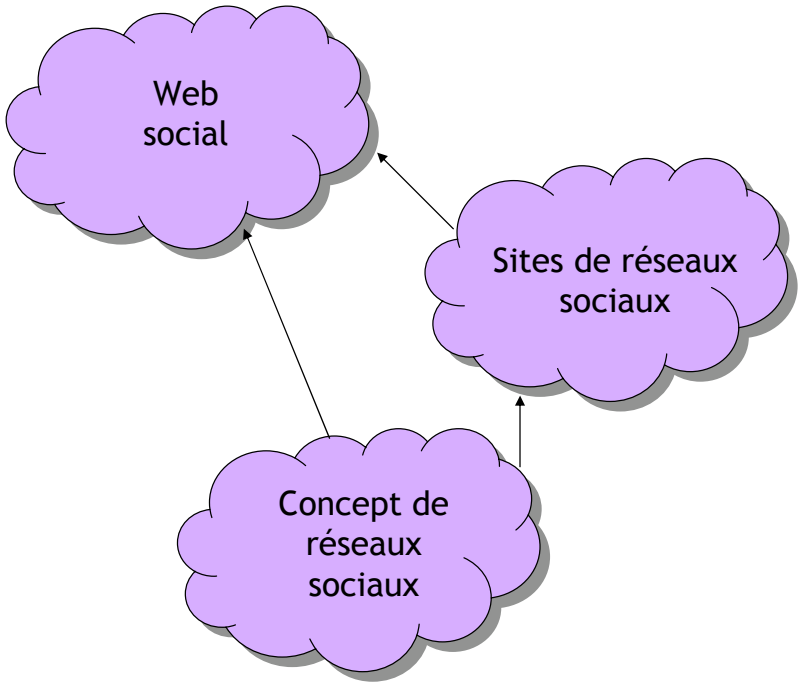
5 Problématiques de recherche

Une tentative d'organisation

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

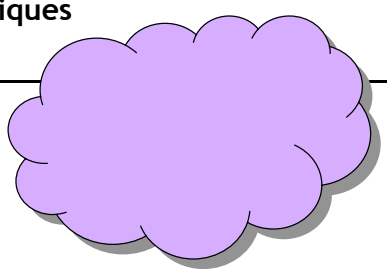
Réseaux sociaux



All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

Réseaux Sociaux



Quelles problématiques de recherche ?

Une tentative d'organisation

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

Tentative d'organisation

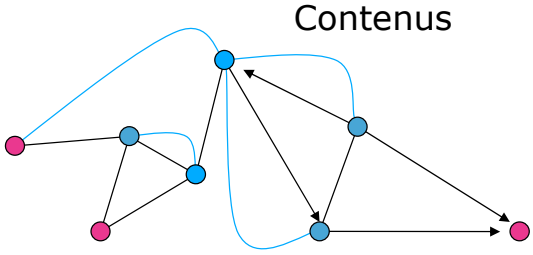
- Modélisation en graphe
- Algorithmes
  - Construction
  - Exploitation
- Fonctions
  - Usage (VE, VO, KM)
  - Contexte
  - Confidentialité
- Contenus
  - Modélisation sémantique
  - Analyse du contenu, Résumé

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

Modélisation

→ **Modèles**
Algorithmes
Fonctions
Contenus



- Modélisation en graphes
  - nœuds, liens
- Attributs
- Typages des nœuds et liens
  - Quelle maille : individu ➔ organisation

- Modèle de l'utilisateur : profil
- Standardisation des profils (FOAF, GUMO/UserML, W3C)

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**
Construire le graphe social

Modèles
→ **Algorithmes**
Fonctions
Contenus

- Analyse des interactions
  - selon approche centralisée ou décentralisée
  - ex: X appelle Y chaque jour ½ heure
- Analyse de données externes
  - ex: co-publication d'articles
- Décrit par l'utilisateur
  - "Veux-tu être mon ami ?"

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**
Exploiter le graphe social

Modèles
→ **Algorithmes**
Fonctions
Contenus

- Formuler des requêtes sur le réseau social
- Analyse du graphe/réseau social
  - Propriétés (ex: topologie du réseau, points remarquables)
  - Partages d'intérêts entre pairs
  - Élaboration de recommandations
  - Détection de communautés (analyse de la structure, analyse du contenu)
- Techniques
  - Fouille de données
  - Apprentissage (supervisé, non supervisé)

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

Utiliser le réseau social

Modèles
Algorithmes
→ **Fonctions**
Contenus

- Prise en compte de l'utilisateur
  - Interface utilisateur : interaction, représentation du RS (300 amis!), updates, notifications…
- Fonctions des sites de réseaux sociaux
  - Améliorer (ex : meilleur contrôle des données, visualisation …)
  - Étendre ➔ Web social (ex: intégration aux processus quotidiens, autres…)
- Plate-forme de communautés
  - Communautés virtuelles, collaboration, VE, VO
  - Modèles de processus (ex : communautés de soins)
- Objets communicants
  - Participation des objets (capteurs mobile or fixes)
    - Mon géranium tweete
  - Dépôt de message (post-it), etc.

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

Contexte utilisateur et réseau social

Modèles
Algorithmes
→ **Fonctions**
Contenus

- But : utiliser le contexte pour améliorer les communication de l'utilisateur (plus vs mieux)
- Sensibilité au contexte / à la situation sociale
  - Modèles de contexte
  - Algo d'apprentissage
  - Ontologies adaptés à l'utilisateur:

    37 23.516 -122 02.625 ➔ 8, Market Street ➔ Home

- Intelligence ambiante
  - L'application prend en compte l'environnement social (et les liens du réseau social)
  - Compléter le réseau social (10h dans le même compartiment!)

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**
Confidentialité, sûreté

Modèles
Algorithmes
→ **Fonctions**
Contenus

- Contrôler mes informations
- Contrôler l'information me concernant

- Où stocker l'information? (serveurs de confiance, centralisé/décentralisé)
- Comment partager, avec quel profil?
  - differentes catégories sociales (MOLI : privé/professionnel, Facebook…)

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**
Interopérabilité

Modèles
Algorithmes
→ **Fonctions**
Contenus

- Centralisation vs distribution des SRS
  - Centralisation: architecture de type client/serveur (pertinent pour la facturation)
  - Distribution: pair à pair, analogies avec les modèles agents (entités autonomes communicantes)

- Modèles de données et de processus respectifs
- Impact sur la gestion de la confidentialité, confiance

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

Les contenus

Modèles
Algorithmes
Fonctions
⟶ **Contenus**

- Prise en compte de la sémantique
  - Semantic Web and Social Networks
- Exploitation des annotations (tag) (del.icio.us)
- Noms, localisation : désambiguïsation
- Élaboration de recommandations ("Vous aimerez ce que vos amis ont aimé")

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

Quelques conférences

- W3C Workshop on the future of Social Networking (Barcelone)
- Social communication and Symposium on Social Intelligence and Networking(SocialCom, Vancouver)
- Workshop on Social Web Search and Mining (at CIKM2009, Hong Kong)
- Workshop on Confidentiality and Trust management in the Social Web (Poland)
- International Conference on User Modeling, Adaptation, and Personalization (UMAP, Trento)

All Rights Reserved © Alcatel-Lucent 2009

Historique Définition Usages Tendances **Problématiques**

Au programme de l'école d'été

- Potentialités et limites pour les entreprises
  - Relations clients (SNCF-Garnier)
  - Ressources humaines, capitalisation des connaissances (table ronde)

Fonctions

- Modèles, Sémantique, Exploitation
  - Exploitation pour nouveaux usages (Alcatel Bell Labs-Stan&Hacid)
  - Données et Web 2.0 (INRIA-Gandon)

Modèles Contenus

- Algorithmie sur les graphes
  - Fouille (Univ Alberta-Zaïane)
  - Analyse (Univ St Etienne-Frachisse)
  - Recommandation (Univ Paris 13-Kanawati)
  - Simulation (Univ Toulouse-Amblard)

Algorithmes

- Analyse des pratiques
  - Surveillance (DCRI-Sueur), Sociologie (FT-Cardon)

Fonctions

All Rights Reserved © Alcatel-Lucent 2009

Merci pour votre attention,

Avez-vous des questions ?

Prof. Pierre Maret

Laboratoire Hubert Curien, St-Etienne

pierre.maret@univ-st-etienne.fr

Adrien Joly

Alcatel-Lucent Bell Labs France, Villarceaux / Laboratoire LIRIS, Lyon

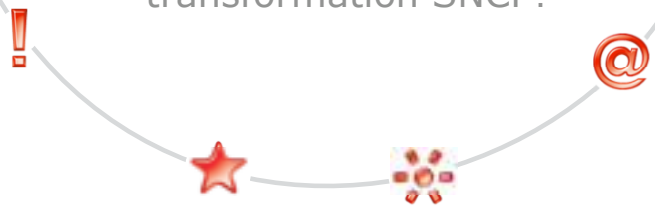
adrien.joly@alcatel-lucent.com

Le Web 2.0 moteur de la transformation SNCF

C. GARNIER
(CRMSERVICES SNCF, FRANCE)

CRM SERVICES

Le WEB 2.0 moteur de la transformation SNCF?



SNCF GROUPE

CRM SERVICES

Ordre du jour

- Qu'est ce que la SNCF ?
- Marque et Web 2.0
- Relation Client, Services et WEB 2.0
- Salariés et WEB 2.0
- La Transformation et le web 2.0
- Synthèse

2

Ordre du jour

- *Qu'est ce que la SNCF ?*
- Marque et Web 2.0
- Relation Client, Services et WEB 2.0
- Salariés et WEB 2.0
- La Transformation et le web 2.0
- Synthèse

3

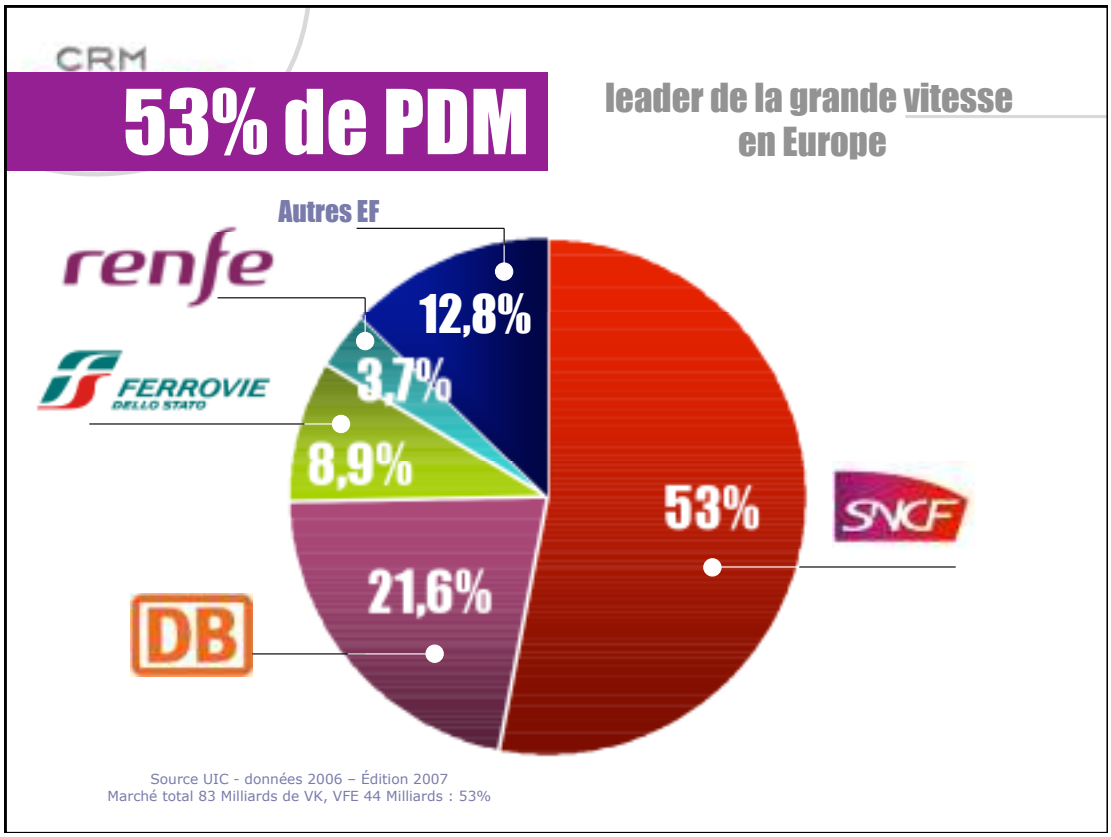
- Googlons « qu'est ce que la SNCF ?»

4

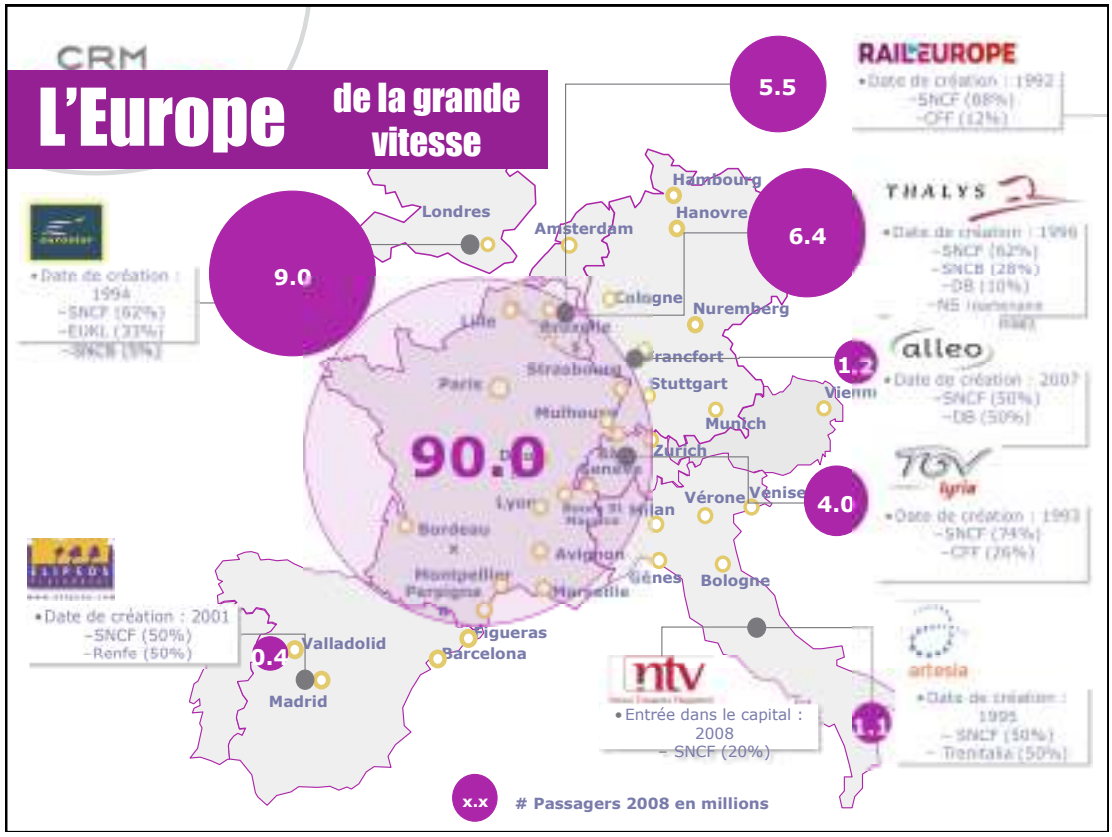
Leader mondial de la grande vitesse
du monopole à la préférence
de l'Europe au monde
SNCF VOYAGES
des idées d'avance
SNCF

CRM
A propos
de SNCF Voyag
30000
collaborateurs
130 millions
de voyageurs
15 filiales
CA 2008
7,5 M€
Une distribution
multi-canale
500 M€ = 15
investis chaque année
dans le parc TGV
rames/an
+ de 9 millions
450
rames TGV
1369
voitures
classiques

CRM
SNCF Voyages
un opérateur qui couvre
toute la chaîne du voyage
SNCF VOYAGES
Les transporteurs
Grandes Lignes
Europe
Les filiales de Distribution et de Services
Les partenariats
Les trains


CRM
53% de PDM
leader de la grande vitesse
en Europe
Autres EF
12,8%
3,7%
8,9%
53%
21,6%
Source UIC - données 2006 – Édition 2007
Marché total 83 Milliards de VK, VFE 44 Milliards : 53%

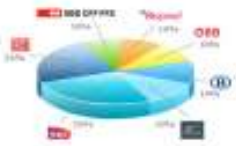
CRM
Vue d'ensemble
des lignes et partenariats
168 gares
1114 trains
Grandes Lignes quotidiens
THALYS
DB
SNCF
TGV
Iyria
artesia


CRM
L'Europe
de la grande vitesse
5.5
RAIL EUROPE
• Date de création : 1992
-SNCF (68%)
-OFF (32%)
9.0
6.4
Hambourg
Hanovre
Londres
Amsterdam
Cologne
Nuremberg
Francfort
Stuttgart
Munich
1.2
alleo
• Date de création : 2007
-SNCF (50%)
-DB (50%)
THALYS
• Date de création : 1996
-SNCF (62%)
-SNCB (28%)
-DB (10%)
-NS (entretien)
• Date de création : 1994
-SNCF (62%)
-EUKL (31%)
-SNCB (7%)
Lille
Bruxelle
Strasbourg
Paris
Mulhouse
90.0
Zurich
Genève
Lyon
Milan
Vérone
Venise
4.0
TGV
Iyria
• Date de création : 1993
-SNCF (74%)
-OFF (26%)
Bordeau
Avignon
Montpellier
Perpignan
Marseille
Gênes
Bologne
• Date de création : 2001
-SNCF (50%)
-Renfe (50%)
0.4
Valladolid
Madrid
Figueras
Barcelona
ntv
• Entrée dans le capital : 2008
- SNCF (20%)
1.1
artesia
• Date de création : 1995
- SNCF (50%)
- Trentalia (50%)
x.x
Passagers 2008 en millions

CRM

L'alliance Railteam

7 compagnies ferroviaires



Répartition du financement par les sociétés fondatrices

Faciliter les voyages à grande vitesse en Europe

25 millions de voyageurs attendus sur le réseau Railteam en **2010**

Une réelle alternative à l'avion et à la voiture

CRM

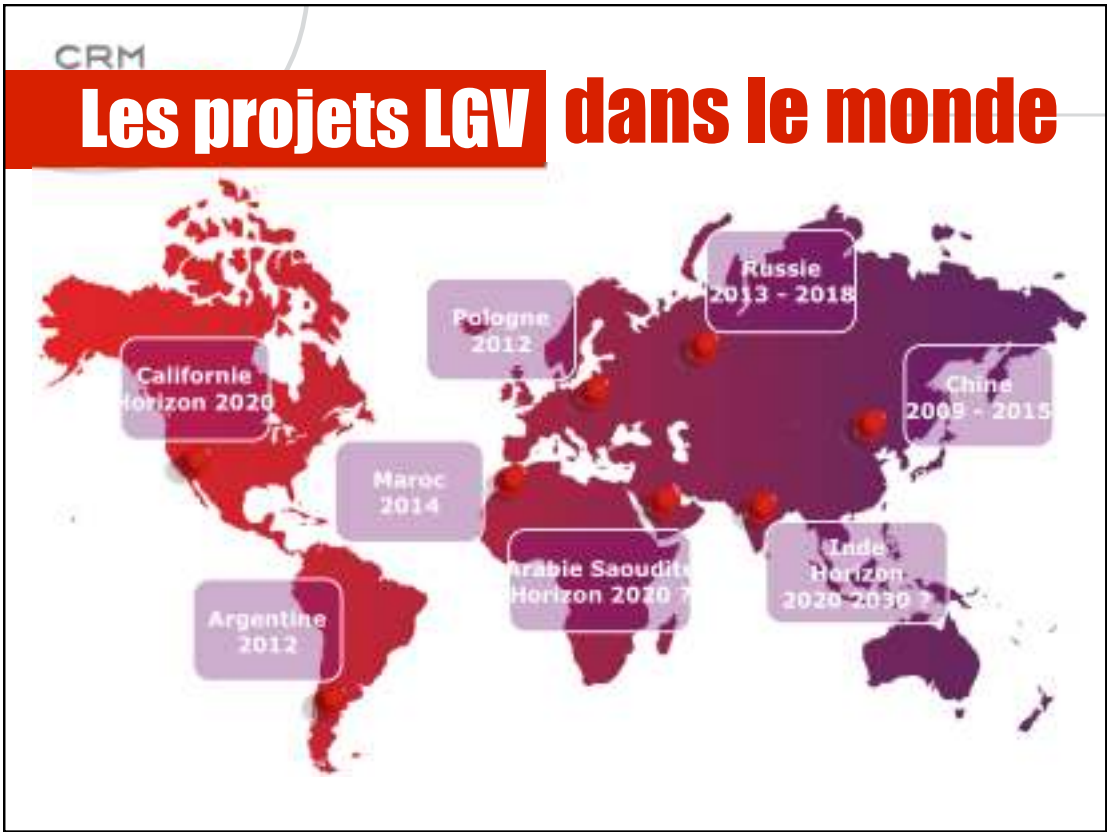
Un environnement sous le signe de la croissance

La France au cœur du trafic européen en 2020

	2008	2020
• Lyon – Francfort	6H30	4H40
• Paris – Amsterdam	4H10	3H15
• Paris – Barcelone	8H25	5H25
• Paris – Milan	7H05	4H45
• Turin – Rome	5H50	4H10
• Berlin – Munich	5H40	4H15
• Milan – Rome	4H30	3H00

Le développement du réseau rapproche les Européens

GVK – Source UIC 2006


CRM
Les projets LGV dans le monde
Californie
Horizon 2020
Pologne
2012
Russie
2013 - 2018
Chine
2009 - 2015
Maroc
2014
Arabie Saoudite
Horizon 2020
Inde
Horizon
2020-2030
Argentine
2012

CRM
SERVICES
Tous Eco mobiles
L'Eco-mobilité, c'est le développement durable by SNCF
Intermodalité
Mode doux
LE BILAN CARBONE DE MES TRAJETS
en train
32kg
CONSOMMATION À TRAJETS ÉQUIVALENTS
en avion
563kg
en voiture
718kg
TAXI
PARISIEN
Moins de CO²
pour voyager loin
LES TROPHÉES DU TOURISME responsable

CRM

A propos des Trains à Grande Vitesse

Depuis 1981,
+ de 1,5 milliard
de passagers transportés

Chaque jour,
800 TGV
parcourent les
1900 kms
de lignes à grande vitesse

En 2008,
Voyageurs France
Europe a transporté
130 millions
de passagers, dont
20 millions
en Europe

D'ici 2010,
+30% de voyageurs
attendus sur le
réseau européen
à Grande Vitesse

CRM

L'ADN de la grande vitesse française

Un TGV accessible = Prix + Volume

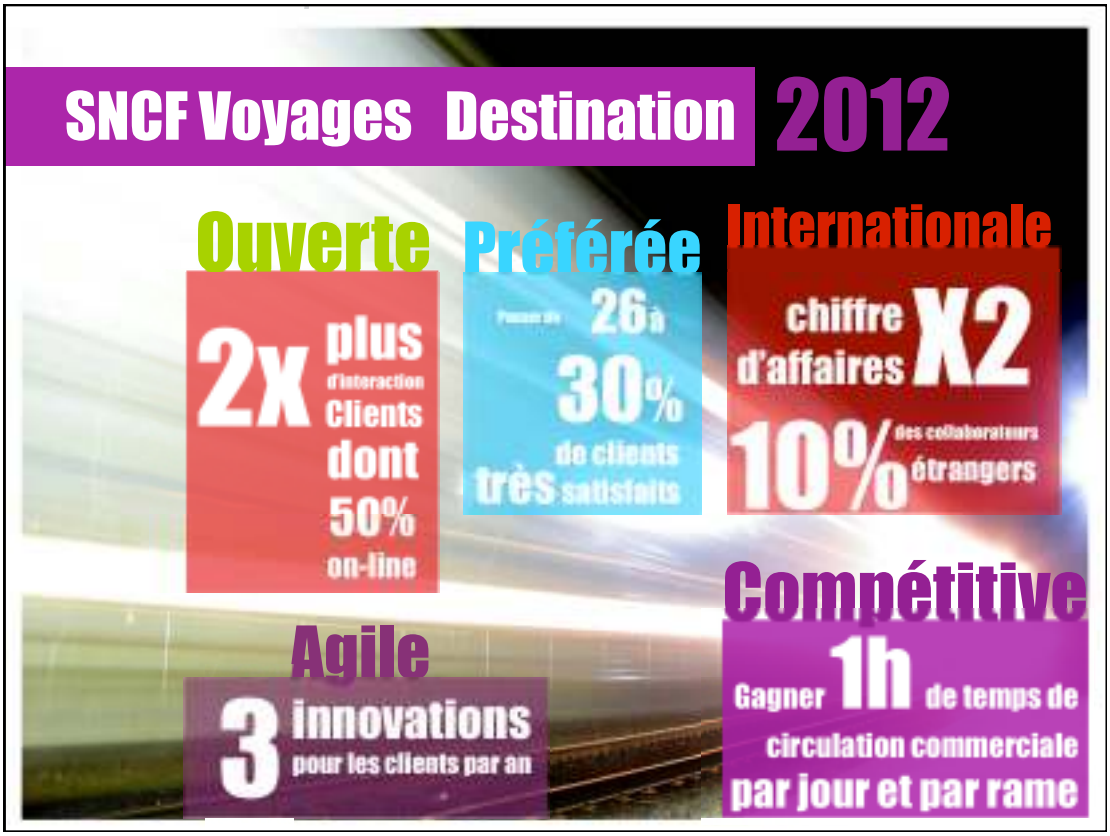
Plus de clients voyagent à petits prix

Plus de clients dans nos trains

Depuis octobre 2007

CRM
L'ADN
de la grande vitesse française
Un TGV innovant pour tous
Connectivité
Relation client
Bienvenue à l'Université du Service
TGV à la carte
TGV PRO
Faciléo s'occupe de vous
GRAND VOYAGEUR
LE CLUB
Dév. Durable
Europe
TGV-europe
Record

CRM
Une ambition de service
Nouveaux espaces de ventes
24 000 agents formés en 3 ans à l'Université du Service
Fidélisation
de plus de
3,5 millions
de clients
accès plus
Amélioration
des accès pour tous les clients
dans 350 gares

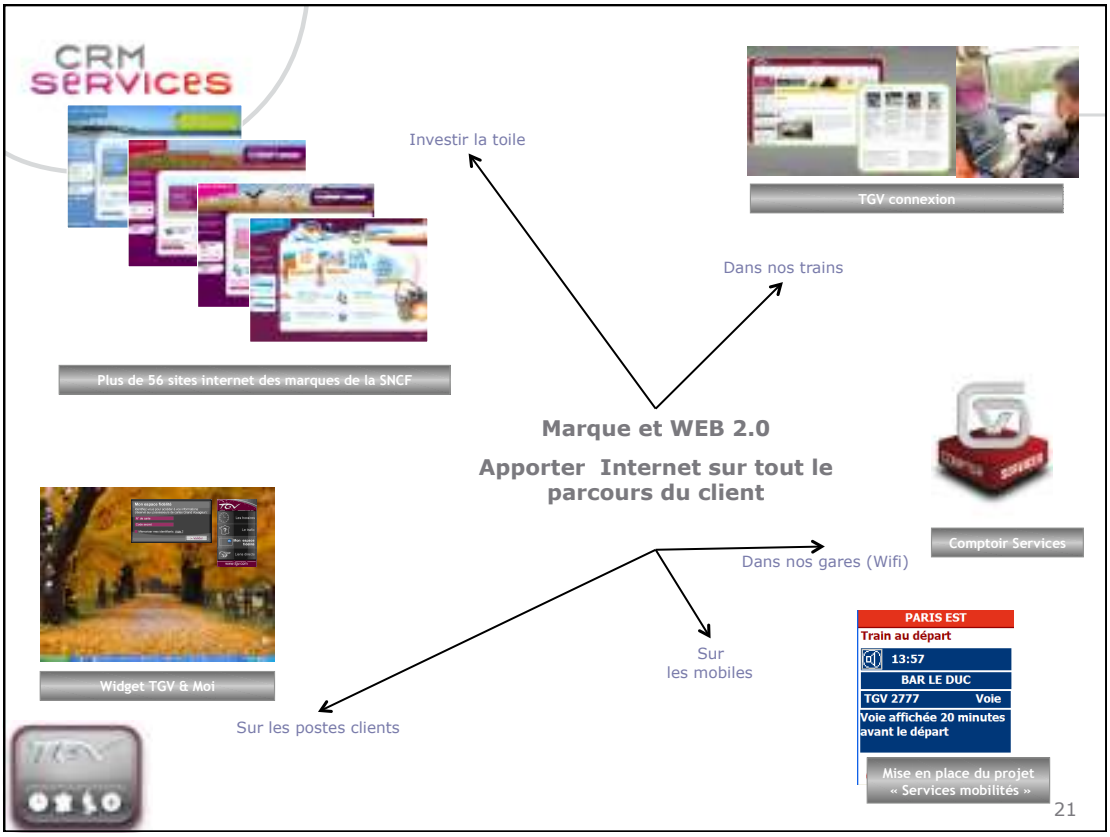


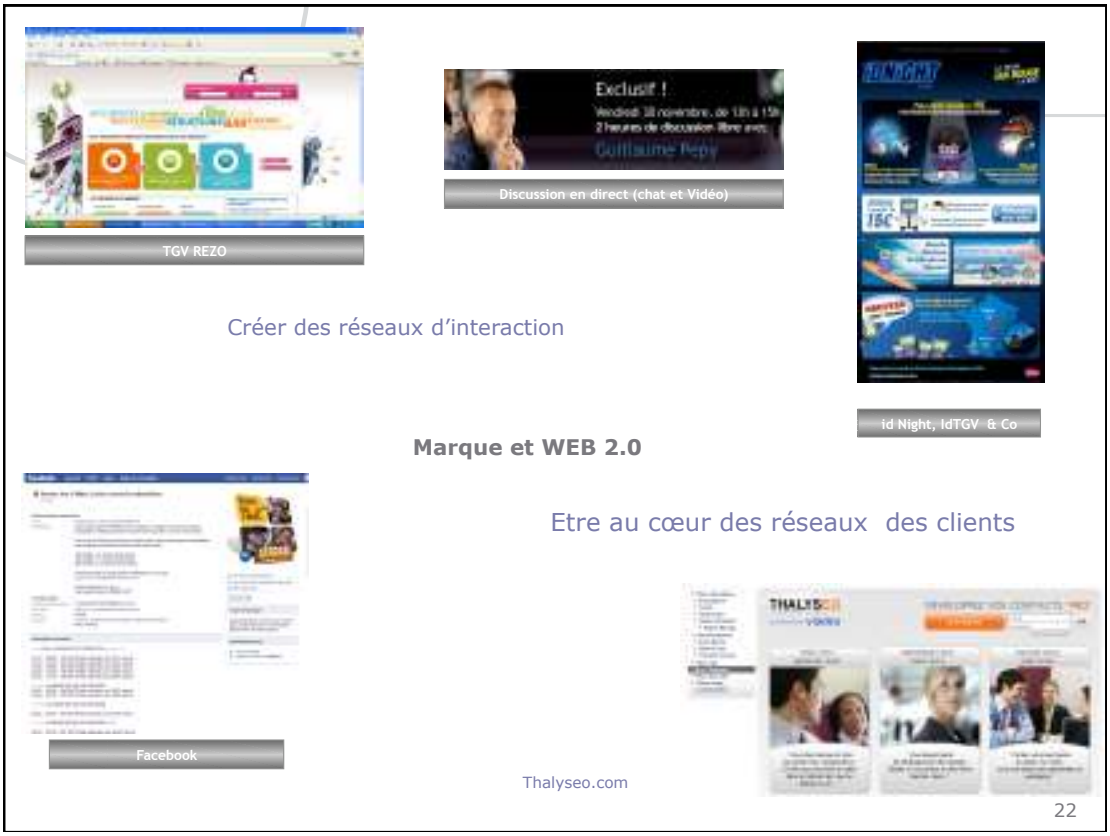
CRM SERVICES

Ordre du jour

- Qu'est ce que la SNCF ?
- *Marque et Web 2.0*
- Relation Client, Services et WEB 2.0
- Salariés et WEB 2.0
- La Transformation et le web 2.0
- Synthèse

20


CRM SERVICES
Investir la toile
TGV connexion
Dans nos trains
Plus de 56 sites internet des marques de la SNCF
Marque et WEB 2.0
Apporter Internet sur tout le parcours du client
Comptoir Services
Dans nos gares (Wifi)
Widget TGV & Moi
Sur les mobiles
PARIS EST
Train au départ
13:57
BAR LE DUC
TGV 2777 Voie
Voie affichée 20 minutes avant le départ
Sur les postes clients
Mise en place du projet « Services mobilités »
21


TGV REZO
Exclusif !
Discussion en direct (chat et Vidéo)
id Night, IdTGV & Co
Créer des réseaux d'interaction
Marque et WEB 2.0
Facebook
Etre au cœur des réseaux des clients
Thalyseo.com
22

Le Web 2.0 sur le territoire de la Marque

Contre Pouvoir

Humour

Big Brother Inversé

Traçes

Volatilité

Surveillance

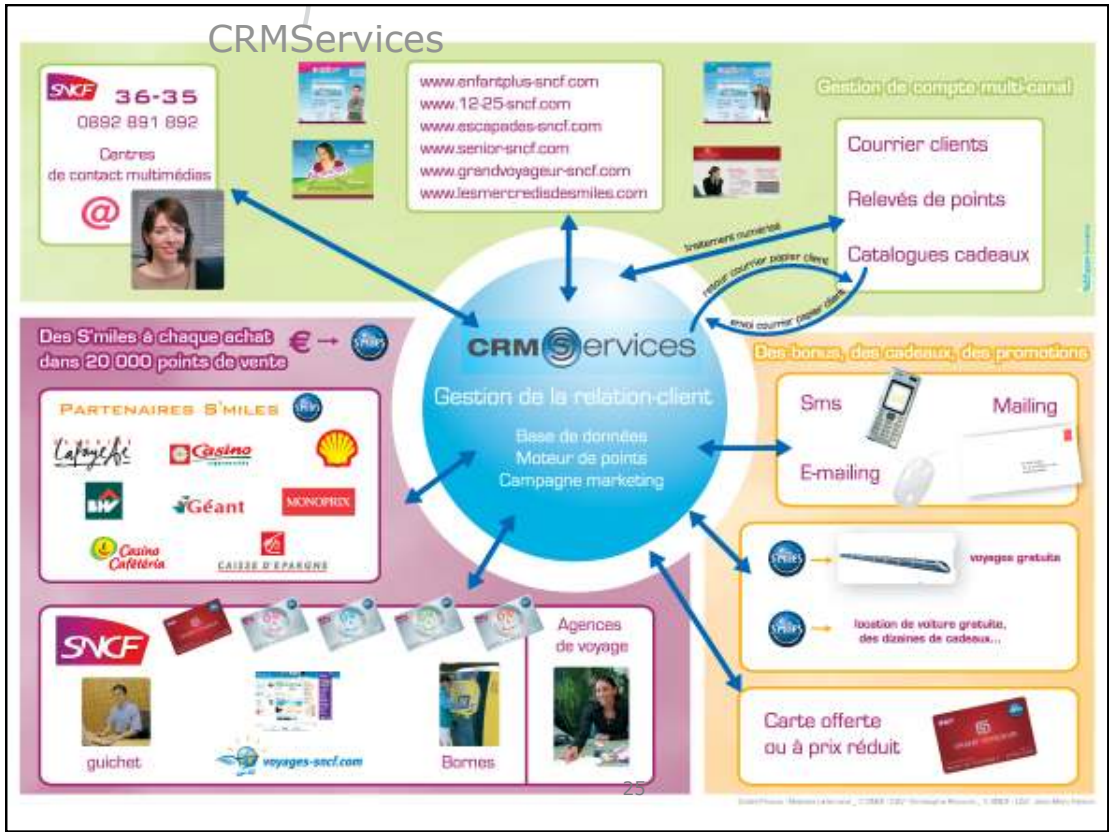
Nouveaux territoires d'expression

Les médias classiques sont les territoires de chasse des marques,
Le web 2.0 est l'eco-système où l'on traque les marques…

Ordre du jour

- Qu'est ce que la SNCF ?
- Marque et Web 2.0
- *Relation Client, Service et WEB 2.0*
- Salariés et WEB 2.0
- La Transformation et le web 2.0
- Synthèse

24


CRMServices
SNCF 36-35
0892 891 892
Centres de contact multimédias
www.enfantplus-sncf.com
www.12-25-sncf.com
www.escapades-sncf.com
www.senior-sncf.com
www.grandvoyageur-sncf.com
www.lesmercredisdesmiles.com
Gestion de compte multi-canal
Courrier clients
Relevés de points
Catalogues cadeaux
traitements numériques
retour courrier papier client
envoi courrier papier client
Des S'miles à chaque achat dans 20 000 points de vente
PARTENAIRES S'MILES
Casino
Géant
MONOPRIX
Casino Cafétéria
CAISSE D'EPARGNE
CRM Services
Gestion de la relation-client
Base de données
Moteur de points
Campagne marketing
Des bonus, des cadeaux, des promotions
Sms
Mailing
E-mailing
voyages gratuits
location de voiture gratuite, des dizaines de cadeaux...
Carte offerte ou à prix réduit
SNCF
guichet
voyages-sncf.com
Bornes
Agences de voyage
25

CRM services
WEB au cœur de la stratégie Client
Elargir la cible du programme relationnel en maitrisant les coûts contacts et en pilotant la générosité selon l'intensité de la relation
RECRUTER ET IDENTIFIER
ANALYSER ET CONVERTIR
FIDELISER ET DEVELOPPER
Statutaire
Affinitaire
ou
ADHESION ON LINE
AVIS
TGV
26

CRM SERVICES

Le lien Physique Client-Marque-Service

Carte Tarifaire

Carte Co Brandée

Carte de services

Accompagnement dans la mise en place du e-forfait

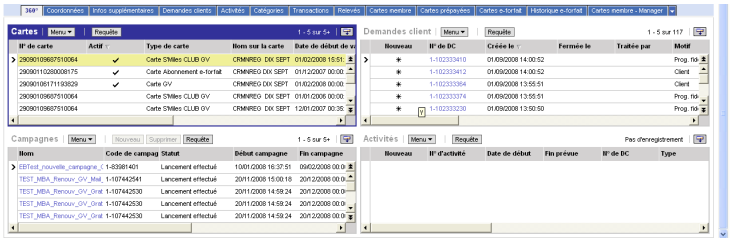
Carte de fidélité

Carte VIP

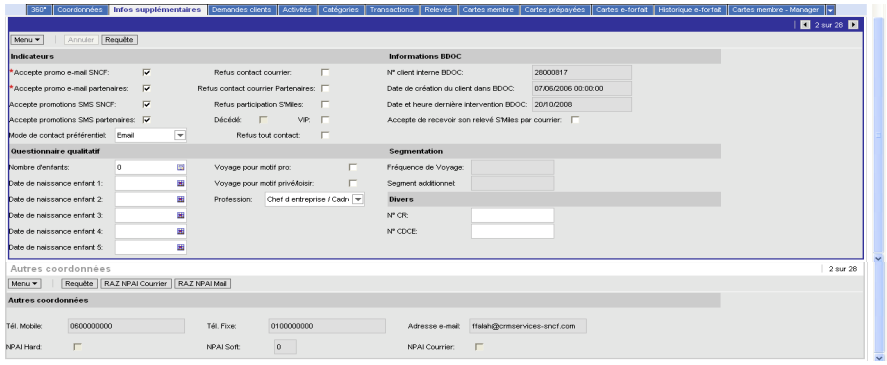
Création et lancement de GV le club

CRM SERVICES

La connaissance Client



- Le Saint-Graal de la vision 360 Client!



CRM SERVICES

Les petits gestes éco citoyens

- Le client est acteur

CRM SERVICES

Le Tchat

- Parler avec le client

CRM SERVICES

TGV REZO

Plateforme Relationnelle de TGV destinée aux 18-29 ans

Fusion de CRM et de Web 2.0

Le partage au cœur du service

CRM SERVICES

JVS

- Des services de partages , JVS sur Mesure

Le Web 2.0 nouveau levier de la relation Client

Interaction

Connaissance Client

Echange

Mise en reseau

Débat

Services Partagés

Services Client

Mais pour quelle représentativité?

Ordre du jour

- Qu'est ce que la SNCF ?
- Marque et Web 2.0
- Relation Client et WEB 2.0
- *Salariés et WEB 2.0*
- La Transformation et le web 2.0
- Synthèse

34

Site 2.0 TGV&co : www.tgvandco.com

- Une plate-forme collaborative
- Une communauté SNCF Voyages
- De l'information sur le projet

CRM SERVICES

Le Web 2.0 nouveau levier de la relation employé

Information

Interaction

Echange

Management participatif

Services Salariés

Débat

Services Partagés

Mais pour quelle efficacité?

CRM SERVICES

Ordre du jour

- Qu'est ce que la SNCF ?
- Marque et Web 2.0
- Relation Client et WEB 2.0
- Salariés et WEB 2.0
- *La Transformation et le web 2.0*
- Synthèse

37

Co-construire avec les clients

Le web 2.0 apporte l'opportunité pour les marques de s'appuyer sur l'expertise de leurs clients pour faire progresser leurs offres au travers de communautés.

Etre sur le marché au bon moment

Projets **Express** : lancer rapidement une 1ère version industrielle pour occuper le marché. Les fonctions complémentaires viendront dans un 2ème temps.

Concevoir autrement à l'intérieur des contraintes budgétaires

Diagnostics rapides : challenger la conception avec les parties prenantes pour concevoir a minima en conservant le cœur de valeur. Projets **Smart** : adresser des freins récurrents.

Ouvrir son éco-système à des innovations portées par d'autres (« Open Innovation »)

Faire appel à des startups et à des PME pour proposer et tester des solutions innovantes

Permettre aux clients de tester les nouveaux services… et de les enrichir

Google Labs, le laboratoire technologique de Google.

Google Labs présente certaines de nos idées préférées, qui ne sont pas encore prêtes pour un lancement grand public. Vos commentaires peuvent nous aider à les améliorer. N'hésitez pas à tester ces prototypes et à envoyer vos commentaires aux Googlers qui les ont développés.

- **Google Recherche de code**
  Recherche de code source public
  20/11/2006 - Commentaires (en anglais) - Discussion

Disponible en d'autres langues

- **Site-Flavored Google Search Box (en anglais) Amélioration !**
  Utilisez un champ de recherche qui personnalise les résultats en fonction de votre site Web.
  17/06/2004 - Commentaires (en anglais) - Discussion (en anglais)
- **Froogle Mobile (États-Unis) | Froogle Mobile (Royaume-Uni) Amélioration !**
  Recherchez des produits à partir de votre téléphone portable avec Froogle.
  24/02/2004 - Commentaires (en anglais) - Discussion (en anglais)
- **Google Sets (en anglais)**
  Créez automatiquement des ensembles d'éléments à partir de quelques exemples.
  20/05/2002 - Commentaires (en anglais) - Discussion (en anglais)

TGV lab
au cœur de l'innovation

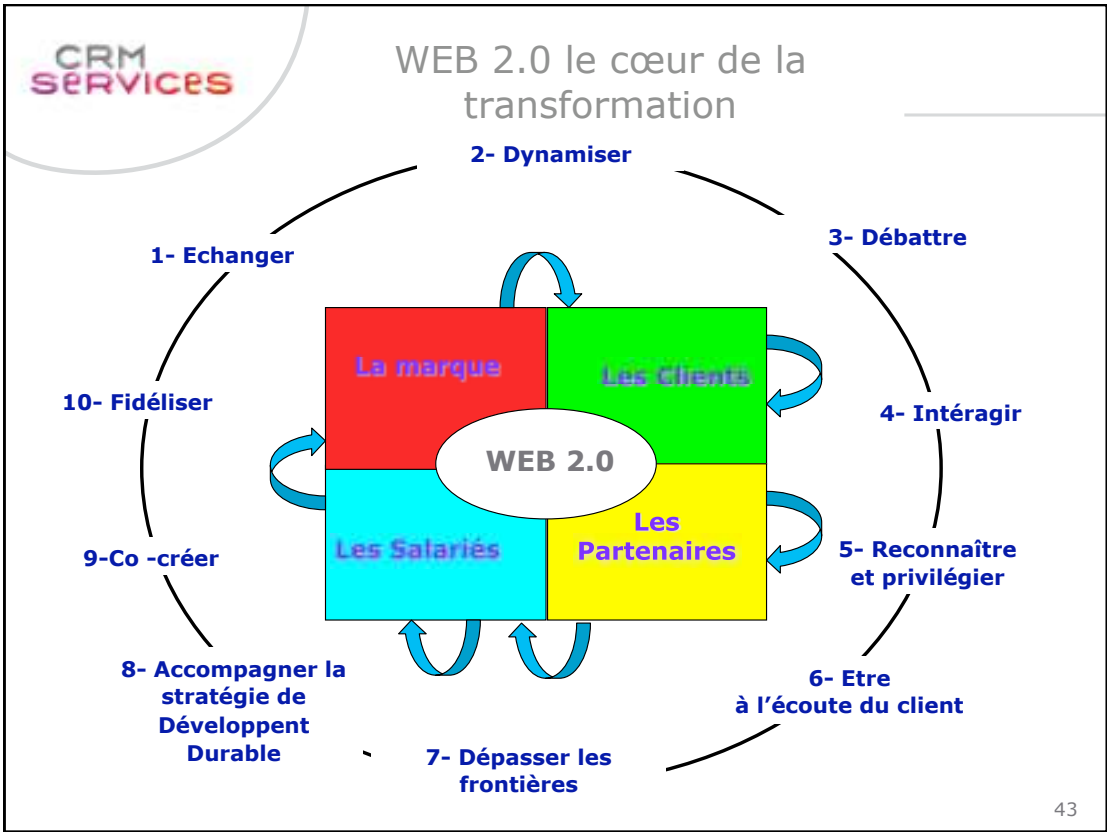
- TGV sans réservation

- TGV Familles

Ordre du jour

- Qu'est ce que la SNCF ?
- Marque et Web 2.0
- Relation Client et WEB 2.0
- Salariés et WEB 2.0
- La Transformation et le web 2.0
- *Synthèse*

42


CRM SERVICES
WEB 2.0 le cœur de la transformation
2- Dynamiser
3- Débattre
4- Intéragir
5- Reconnaître et privilégier
6- Etre à l'écoute du client
7- Dépasser les frontières
8- Accompagner la stratégie de Développent Durable
9-Co -créer
10- Fidéliser
1- Echanger
La marque
Les Clients
Les Salariés
Les Partenaires
WEB 2.0
43

SESSION 2

Réseaux sociaux : structure et sémantique

Réseaux sociaux : structure et sémantique

H. HACID, J. STAN
(ALCATEL LUCENT, FRANCE)

2009 Summer School on Social Networks
Varennes, France

July 2009

Social Networks: Structure and Semantics

Hakim Hacid[1], Johann Stan[1,2]

[1] Alcatel-Lucent Bell Labs France. Applications domain, Social Communications Department.

[2] Université de Lyon, F-42023, CNRS, UMR 5516, Laboratoire Hubert Curien, Université de Saint-Etienne, Jean-Monnet

Hakim.Hacid@alcatel-lucent.com Johann.Stan@alcatel-lucent.com

All Rights Reserved © Alcatel-Lucent 2009

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

All Rights Reserved © Alcatel-Lucent 2009

A History of Achievements

All Rights Reserved © Alcatel-Lucent 2009

Bell Labs: At a Glance

- 26,000 Active Patents Worldwide
- Active in over 100 Standards Bodies

6	Nobel Prizes in Physics
9	U.S. Medals of Science
12	U.S. Medals of Technology
2	Draper Prizes
3	John Tyndall Awards
1	GRAMMY® Award
6	Marconi Intern'l Fellowship Awards
7	C&C Prizes
28	IEEE Medal of Honor winners
1	Emmy Award
1	General Ferrié Award
1	Blondel Committee Medal

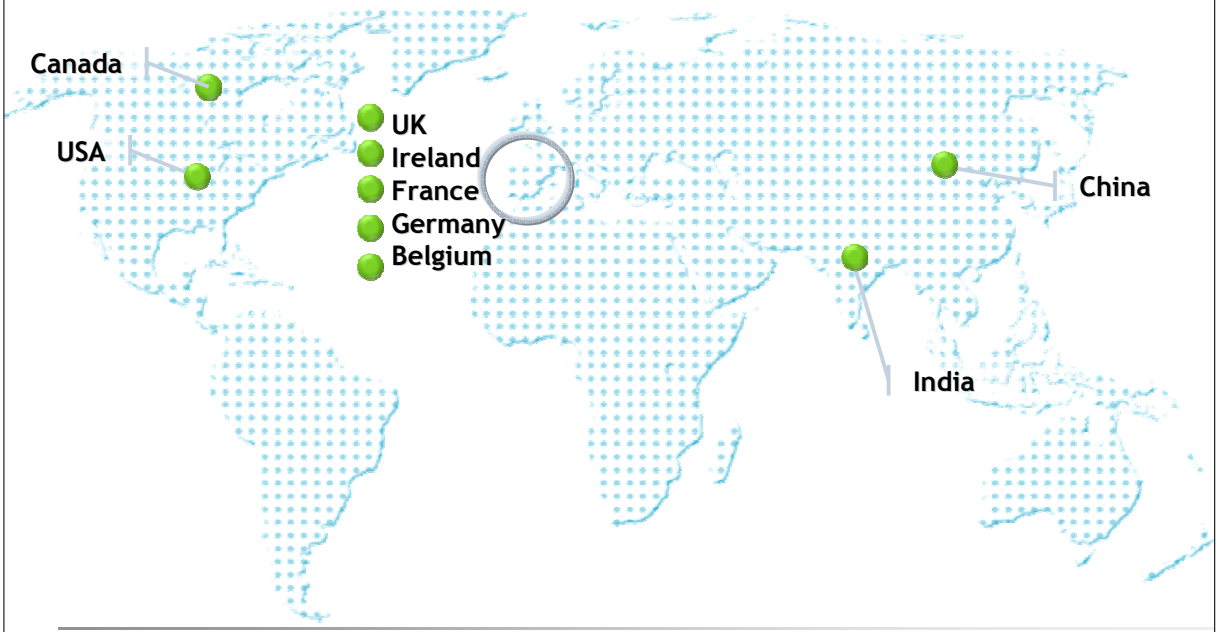
Leading non-academic "Highly Cited" institution
(Physics, Engineering, Math, Materials, and Computer Science Journals. Source: Thomson ISI)

- Collaborating with > 250 universities world-wide
- Multiple Open Innovation Initiatives

All Rights Reserved © Alcatel-Lucent 2009

Research with a Global Footprint

Major Bell Labs Locations



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

President, Bell Labs
Jeong Kim

Chief of Staff

Research | Ventures | Network Technology | Intellectual Property & Standards | Chief Scientist Office | Global Operations

Physical Technologies
- Quantum Systems
- Nano-components
- HS Electronics
- Bio-mimetics
- Optical Components
- Thermal Management

Radio Access
- Antenna
- Radio
- PHY/MAC MIMO
- Satellite

Optical Networks
- Wavelength Management
- Optical Mesh Networks
- Optical Switching
- Optical Transmission

Service Infrastructure
- Policy Management
- Messaging
- Managed Services
- IMS/SIP Infrastructure
- P2P

Computing Technology
- Algorithmic Sciences
- Math Sciences
- Computer Sciences
- Signal Processing
- Security
- Statistical Mining

Fixed Access
- DSL / Copper
- xPON / Optical
- Video
- Home Networking
- In-building

Networking & Networks
- Policy
- Autonomous Nets
- Routing Protocols
- Architecture
- Flat IP
- Softrouter

Applications
- Social Nets
- WEB 2.0
- Gaming
- Media
- Geo-based

CTOs | CF

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

All Rights Reserved © Alcatel-Lucent 2009

Web 1.0 vs Web 2.0

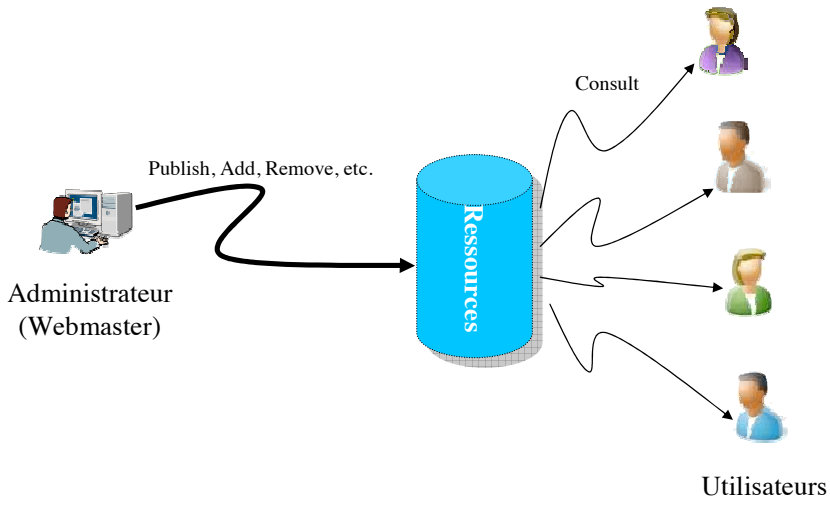


Fig. Schéma traditionnel de la publication de données sur le Web (1.0)

All Rights Reserved © Alcatel-Lucent 2009

Web 1.0 vs Web 2.0

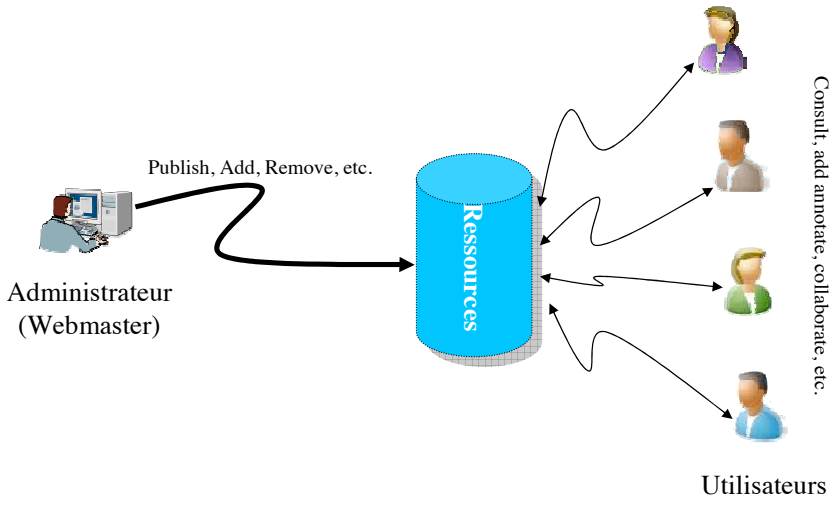


Fig. Schéma de la publication de données sur le Web 2.0

All Rights Reserved © Alcatel-Lucent 2009

Web 1.0 vs Web 2.0

Web 1.0	Web 2.0
-The Web as a tool	-The Web as a platform
-Technology centric	-User centric
-Individualistic	-Collaborative
-People-to-machine	-People-to-people
-Search and navigation	-Publish subscribe

A quick comparison between Web 1.0 and Web 2.0 technologies

All Rights Reserved © Alcatel-Lucent 2009

Social Computing and social platforms

- Any communication or interaction between two persons or a group of persons which is mediated by a machine (e.g., a computer).
  - Email, blogs, IM, collaborative filtering, tagging, etc.

- Social Platforms: Applications, generally on the Web, offring a social extension in tasks performing.
  - Social Networks
  - Social Media
  - Social bookmarking
  - Social Knowledge sharing
  - Social News

All Rights Reserved © Alcatel-Lucent 2009

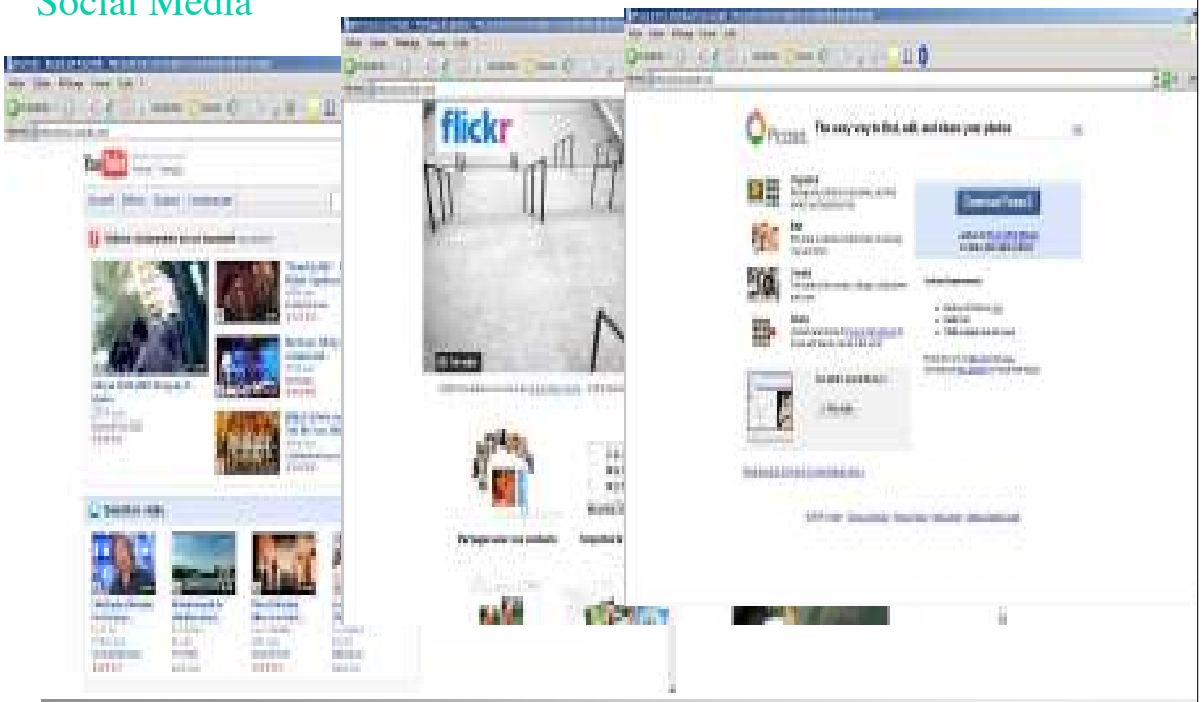
Social platforms

Social networking sites

All Rights Reserved © Alcatel-Lucent 2009

Social platforms

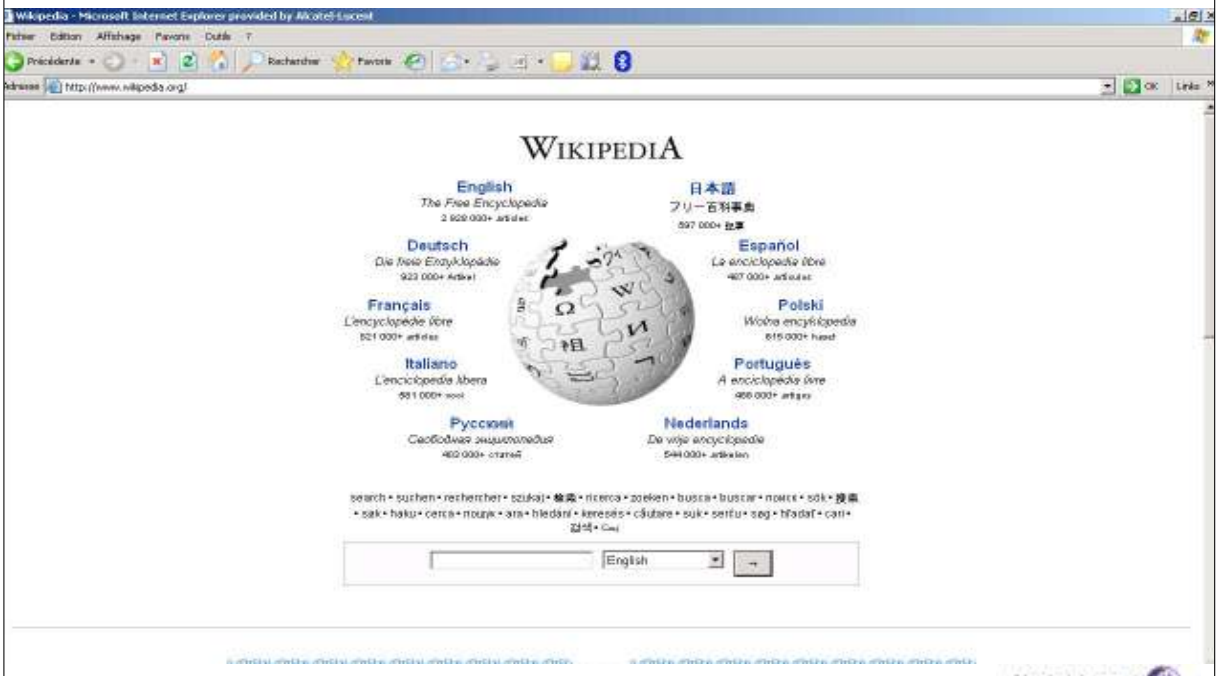
Social Media



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Social platforms

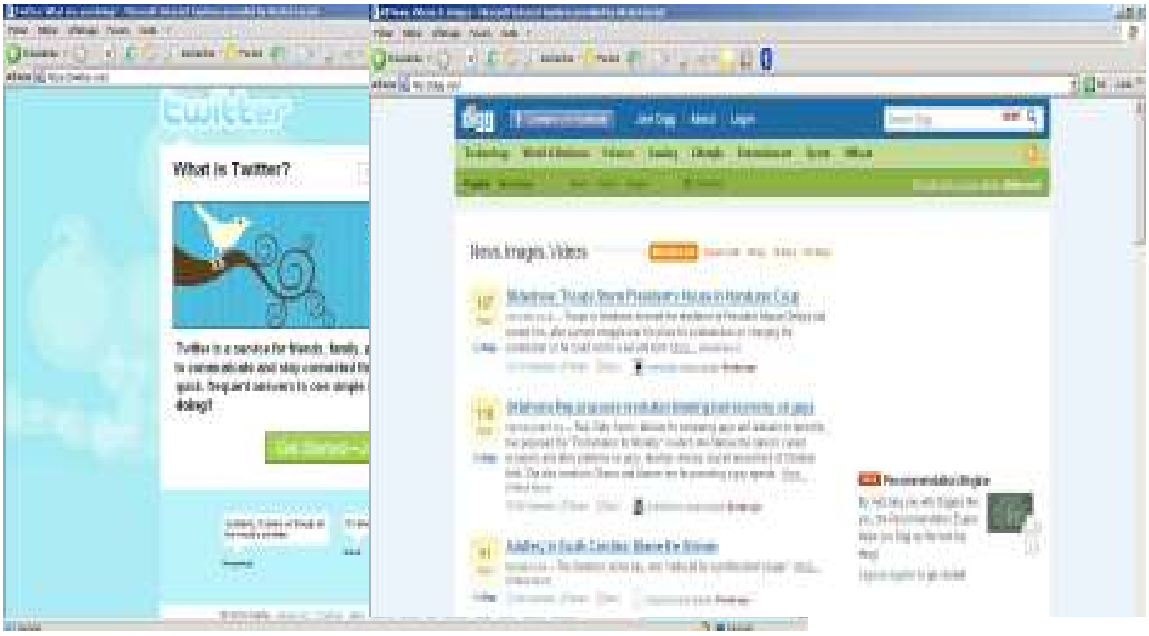
Knowledge sharing



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Social platforms

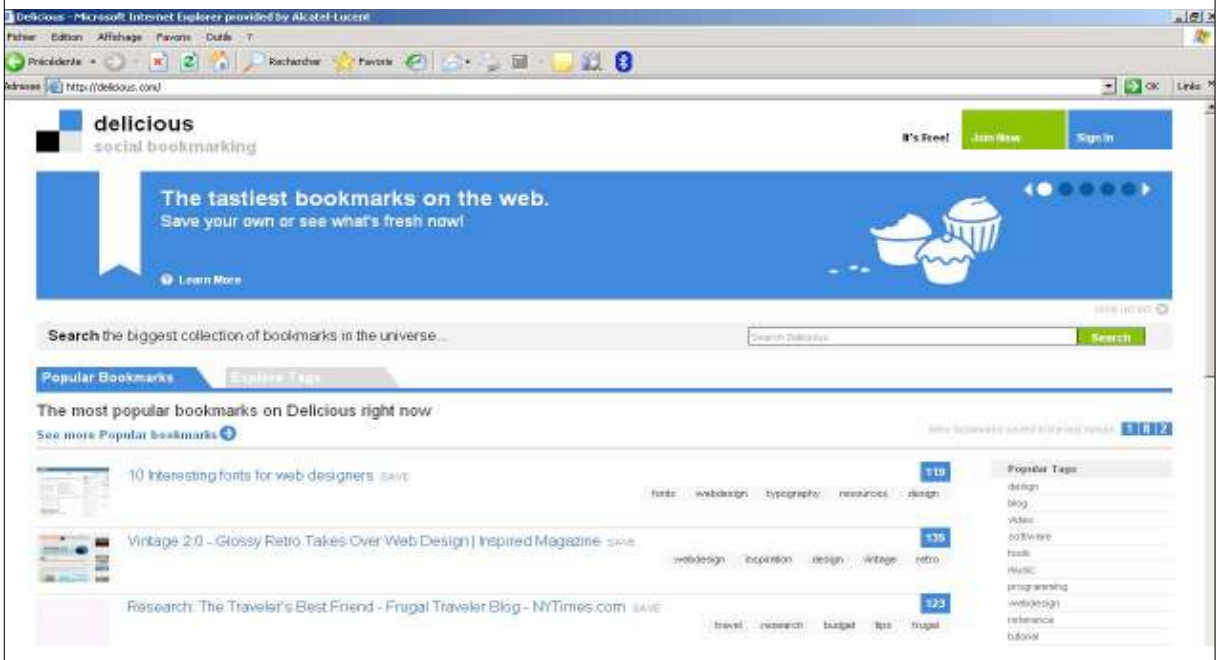
Social news



All Rights Reserved © Alcatel-Lucent 2009

Social platforms

Social bookmarking



All Rights Reserved © Alcatel-Lucent 2009

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

All Rights Reserved © Alcatel-Lucent 2009

Networks vs. Social networks?

A network : A set of nodes, points, or locations connected by means of data, voice, and video communications for the purpose of exchange.

A social network: A description of the social structure between actors, mostly individuals or organizations. It indicates the ways in which they are connected through various social familiarities ranging from casual acquaintance to close familiar bonds.

All Rights Reserved © Alcatel-Lucent 2009

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. **Part II: Semantics of social networks**
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

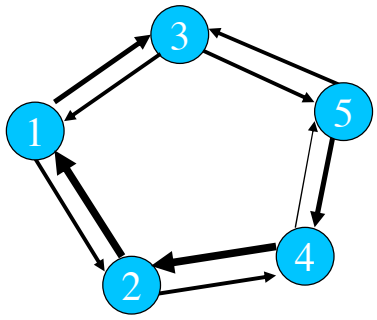
All Rights Reserved © Alcatel-Lucent 2009

Social networks: Structural properties

- **A social network = A graph of nodes and edges**

  - Nodes:
    - Individuals;
    - have properties.

  - Edges:
    - relations between nodes;
    - Have a role;
    - Generally directed and weighted.



Social Network Analysis (SNA) is the mapping and measuring of relationships and flows between people, groups, organizations, computers or other information/knowledge processing entities.

All Rights Reserved © Alcatel-Lucent 2009

Social networks in general

Social networks are considered from several angles:

- What is the size of a social network?
- What is the best representation of social networks?
- What is the behavior of a social network?
- What are the phenomena which command a social network?
- What is the shape of a social network?
- What is the role of a set of individuals in my social network?
- etc.

All Rights Reserved © Alcatel-Lucent 2009

Social networks representation

- **Matrices and graphs are compact and systematic:**
They summarize and present a lot of information quickly and easily; and they force us to be systematic and complete in describing patterns of social relations.

- **Matrices and graphs allow us to apply computers to analyzing data.**
Most of the work is dull, repetitive, and uninteresting, but requires accuracy. This is exactly the sort of thing that computers do well, and we don't.

- **Matrices and graphs have rules and conventions.**
The rules and conventions of the language of graphs and mathematics lead us to see, generally unexpected, things in our data.

All Rights Reserved © Alcatel-Lucent 2009

The most known properties of social networks

Network Diameter

- The longest path between any two edges
- Generally small (small world effect)
- Independent of the network size
- Dynamics of processes
- Erods number calculation

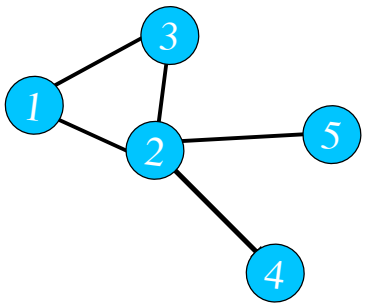
All Rights Reserved © Alcatel-Lucent 2009

The most known properties of social networks

Network Transitivity

- The friend of my friend is likely to be my friend (random graphs);
- The presence of a number of triangles in the network

$$C = \frac{\text{3x number of triangles in the network}}{\text{number of connected triples of vertices}}$$

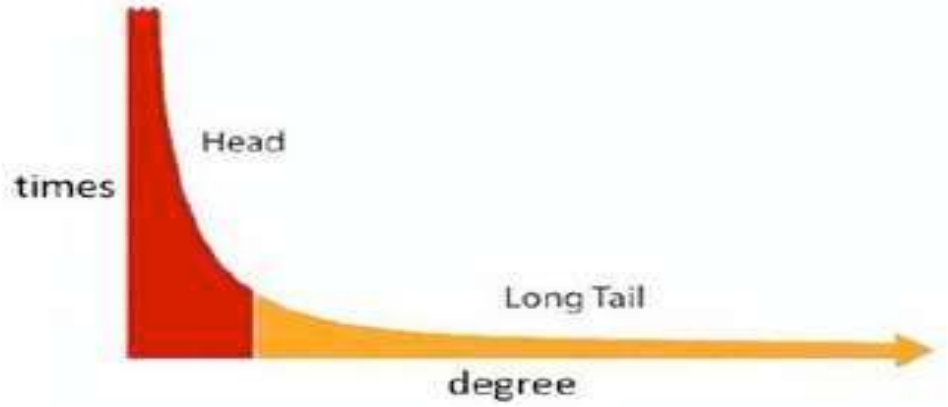


All Rights Reserved © Alcatel-Lucent 2009

The most known properties of social networks

Degree distribution

- Degree: number of nodes connected to that node;
- The fraction of nodes in the network the have a k degree;
- Long tail effect



All Rights Reserved © Alcatel-Lucent 2009

The most known properties of social networks

Network resilience

- Translates the effects of removing a node or a set of nodes from the network
- Removing a node can be:
  - Increase the length of a graph;
  - Disconnect the graph; etc.
- Two strategies to remove a node:
  - Random: doesn't affect much the graph;
  - Targeted: Remove the nodes having a high degree
    - Has an important effect on the network.

All Rights Reserved © Alcatel-Lucent 2009

The most known properties of social networks

Associative mixing

- Translates the tight relation between the nodes types
- defined the nodes which will be paired up together

		women			
		black	hispanic	white	other
men	black	506	32	69	26
	hispanic	23	308	114	38
	white	26	46	599	68
	other	10	14	47	32

All Rights Reserved © Alcatel-Lucent 2009

The most known properties of social networks

Community Structure

- Dominant property in social networks;
- Groups of nodes that have a high density of links; between them with a lower density of links between groups;
- Translates the fact that people group by their interests, age, education etc.
- Similar to a clustering problem.

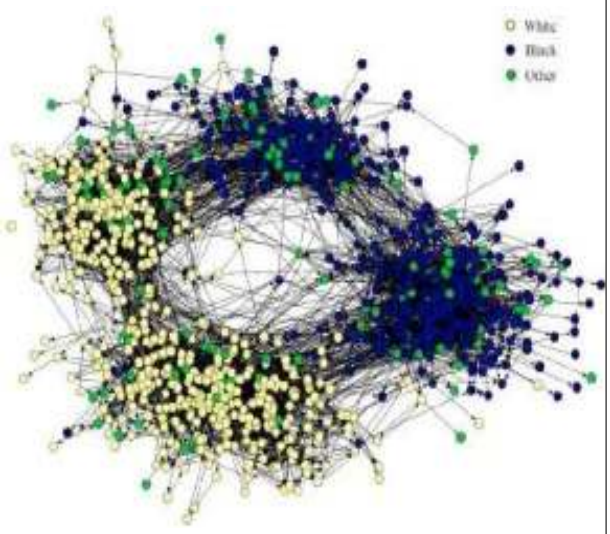


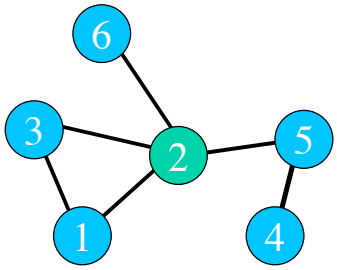
Fig. Friendship communities in a US school (Moody's study)

All Rights Reserved © Alcatel-Lucent 2009

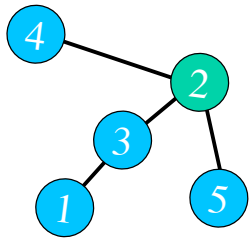
The most known properties of social networks

Centrality

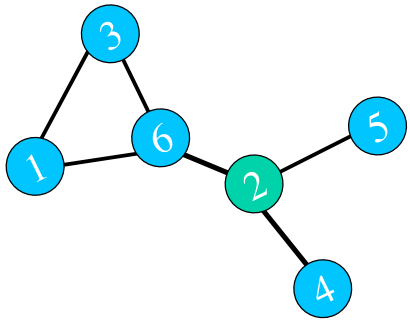
- **Degree centrality**: Nodes with high degree;
- **Closeness centrality**: nodes that have low distance to all other nodes;
- **Betweenness centrality**: nodes that occur on more shortest paths (A node with high betweenness has great influence over what flows in the network indicating important links and single point of failure).



Degree centrality



Closeness centrality



Betweenness centrality

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Other properties of social networks

- Walks;
- Reachability;
- Reciprocity;
- Transitivity;
- Size;
- Density;
- Flow;
- Cohesion;
- Influence;
- etc.

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Examples of Structural analysis of social networks

- Recommendation (Paths exploration, degrees, etc.) (Main application)
  - Find an expert in a given topic;
  - Suggest products to buy according to specific constraints;
  - Suggest a friend to connect with (most of today social networking sites);
- Privacy & trust
  - Triangles identification (i.e. Network transitivity)

All Rights Reserved © Alcatel-Lucent 2009

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

All Rights Reserved © Alcatel-Lucent 2009

Social networks analysis hot topics (structural side)

1. Understand the properties of social networks;
2. To what extent the properties of large graphs apply to social networks;
3. Reducing Complexity;
4. Visualization;
5. Privacy and trust;
6. Evolution of social networks;
7. Business valorization of social networks.

All Rights Reserved © Alcatel-Lucent 2009

References

• Stanley Wasserman, and Katherine Faust. Social Network Analysis: Methods and Applications. Stanley Wasserman. Cambridge university Press. 1994.

• Deepayan Chakrabarti, Christos Faloutsos: Graph mining: Laws, generators, and algorithms. ACM Comput. Surv. 38(1): (2006)

• Bin Zhou, Jian Pei. Preserving Privacy in Social Networks Against Neighborhood Attacks. 506-515. ICDE 2008.

• Nisheeth Shrivastava, Anirban Majumder, Rajeev Rastogi. Mining (Social) Network Graphs to Detect Random Link Attacks. 486-495. ICDE 2008.

• A. Seth and J. Zhang. A Social Network Based Approach to Personalized Recommendation of Participatory Media Content. ICWSM, 2008.

•J-P, Onnela et al. Analysis of large scale weighted network of one-to-one human communication. New journal of physics.(9) 2007.

• Hao Ma, Haixuan Yang, Michael R. Lyu, Irwin King: Mining social networks using heat diffusion processes for marketing candidates selection. CIKM 2008: 233-242.

All Rights Reserved © Alcatel-Lucent 2009

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

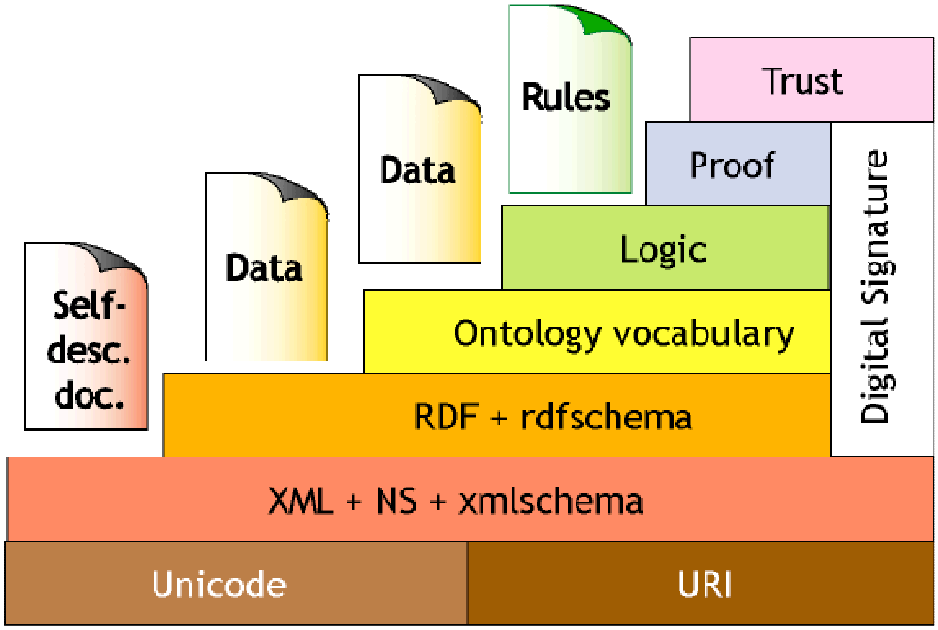
All Rights Reserved © Alcatel-Lucent 2009

The Semantic Web Meets Social Networks

The web is more a social creation than a technical one. I designed it for a social effect — to help people work together — and not as a technical toy. The ultimate goal of the Web is to support and improve our weblike existence in the world. We clump into families, associations, and companies. We develop trust across the miles and distrust around the corner.

—Tim Berners-Lee, Weaving The Web

All Rights Reserved © Alcatel-Lucent 2009

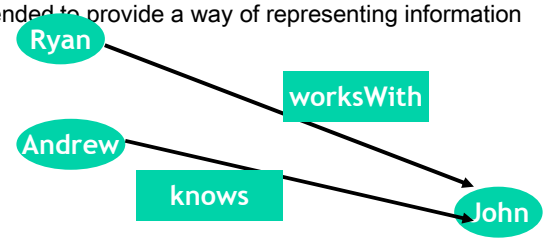


All Rights Reserved © Alcatel-Lucent 2009

Semantic Representation Tools (1/1)

From Semantic Web Challenge …

- **FOAF** ("Friend of a Friend") is an RDF format intended to provide a way of representing information about people
  - **People: Ryan ext:worksWith people:John**
  - **People: Andrew foaf:knows people:John**



- **OWL** built on top of RDF for processing information to be interpreted by computers and for reasoning
- **XFN** a lightweight method of annotating links to indicate a personal relationship

Combining XFN and FOAF to provide the deeper level of information FOAF allows, it is still possible to incorporate XFN. In this case, simply create a FOAF property that represents the relationship between two people, and use XFN values within.

All Rights Reserved © Alcatel-Lucent 2009

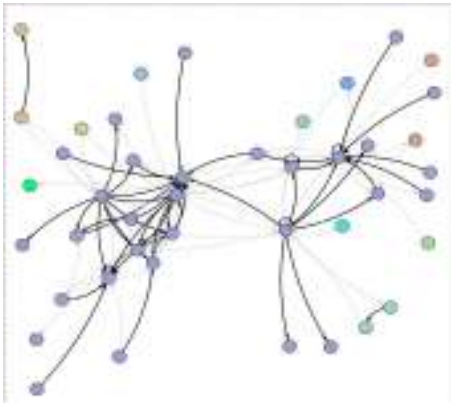
Some social networking tools (Computation & Analysis 1/2)

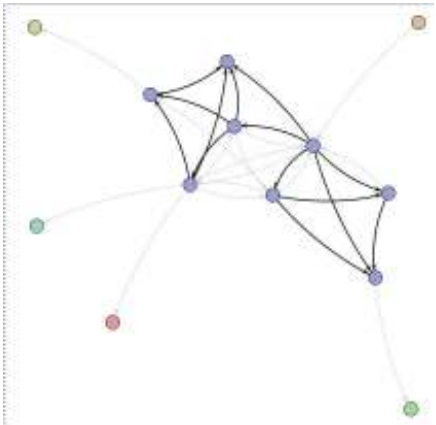
- **Jena**
  - Java framework for building Semantic Web applications.
  - Pprogrammatic environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine.
- **SPARQL**
  - RDF query language
- JUNG network data can also be exported in the formats used by the **Pajek** and **UCINET, XPlanet** and **NetDraw**.
- **Java Colt package**: Matrices are one common representation for network data taking advantage of the structure of the graph; and operations that use the Colt package for numerical matrix manipulation.

All Rights Reserved © Alcatel-Lucent 2009

Some social networking tools (Computation & Analysis 2/2)

- **JUNG** programming toolkit for (social) network analysis
- **JUNG** is a Java library that provides an object-oriented representation of networks
- includes implementations of a number of algorithms from graph theory, data mining, and social network analysis

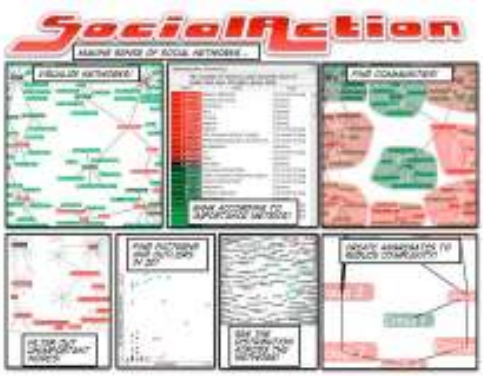




All Rights Reserved © Alcatel-Lucent 2009

Some social networking tools (Graphical analysis)

- **Prefuse** is a set of software tools for creating rich interactive data visualizations.
- **SocialAction** is a social network analysis tool (built wih Prefuse) that integrates visualization and statistics to improve the analytical process.



All Rights Reserved © Alcatel-Lucent 2009

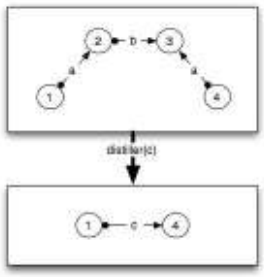
Some social networking tools (Storage and Querying)

- **Sesame**: an RDF storage and query facility allowing record of the provenance and timestamp of the social data collected as well as the weight of the relationships, when available
- **SAIL** (Storage and Inference Layer) is the RDF storage abstraction outlined by the OpenRDF APIs and implemented by Sesame.

All Rights Reserved © Alcatel-Lucent 2009

Some social networking tools (Operators and Graph transformation)

- **Banach** is a collection of operators that work on RDF graphs to infer, extend, emerge or otherwise transform a graph into another.
- **Distiller** Banach operator distills a subgraph out of a graph based on indirect properties.



Example for Social Network Extraction

Given the input graph

```
:Joe :wrote :Email1 .
:Paul :wrote :Email2 .
:Email1 :reply-to :Email2 .
```

Applying the distiller with the predicate :replied-to for the "inward" configuration generates the graph

:Joe :replied-to :Paul .

that can be interpreted as the "distilled" social network out of the email reply topology.

All Rights Reserved © Alcatel-Lucent 2009

Other stuff

- **Scutter** a robot that traverses the RDF-web by following rdfs:seeAlso links. Can be used to collect social network data from FOAF profiles.
- **Place Finder** is a web service that provides geo-location services for over three-million place names worldwide. Social network profiles can be enriched by adding the geographical locations of place names found in the FOAF profiles.
- **DiSo** (dee • zoh) is an umbrella project for a group of open source implementations of these distributed social networking concepts using the framework/platform of existing content management systems, or as Chris puts it: "to build a social network with its skin inside out".

All Rights Reserved © Alcatel-Lucent 2009

Other stuff

- **Microformat** (sometimes abbreviated µF or uF) is a web-based data formatting that seeks to re-use existing content as metadata, using only XHTML and HTML classes and attributes. This approach is intended to allow information for end-users (such as contact information, geographic coordinates, calendar events, and the like) to also be automatically processed by software. Microformats allow the encoding and extraction of events, contact information, social relationships and so on.

  http://microformats.org/

- **OpenQabal** is an open-source social-networking and collaboration platform / suite by the principles of federation, composition, and openness; with a special emphasis on enabling "distributed conversations" and the "federated social graph.«

All Rights Reserved © Alcatel-Lucent 2009

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

All Rights Reserved © Alcatel-Lucent 2009

Social Interaction Analysis
Current Limitations

Current social networking sites are based on an explicitly declared "***list of friends***" model.

- Alice is now friends with Bob.
- Jane is now friends with Amy.
- Bob is now friends with Jane.
- John is now friends with Alice.
- George is now friends with Bob.
- Grace is now friends with Alice.
- Alice is now friends with Amy.
- John is now friends with Amy.
- Grace is now friends with John.
- Grace is now friends with Bob.

The ***real*** social network behind our day-to-day social relationships is much more complex, intangible, dynamic…

I spent an hour talking to that guy… what was his name again?

I'm spending a lot more time with the dance club these weekends!

Alice doesn't know Bob is mad at her.

I look at her differently since they broke up.

George can't stand his best friend's wife.

Jane trusts Amy for financial advice, but not to keep a secret.

It's always a pleasant surprise to hear from Grace

Don't be a stranger!

It's like we've been best friends forever…

You never answer the phone when John calls?

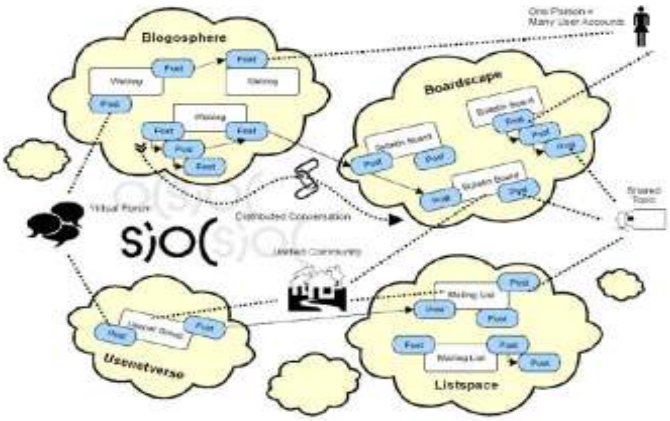
…there is no solution that builds a ***real*** social network!

All Rights Reserved © Alcatel-Lucent 2009

SIOC (Semantically Interlinked Online Communities)

Semantically-Interlinked Online Communities Project

- SIOC provides methods for interconnecting discussion methods such as blogs, forums and mailing lists to each other
- SIOC ontology, an open-standard machine readable format for expressing the information contained both explicitly and implicitly in Internet discussion methods



http://sioc-project.org/

All Rights Reserved © Alcatel-Lucent 2009

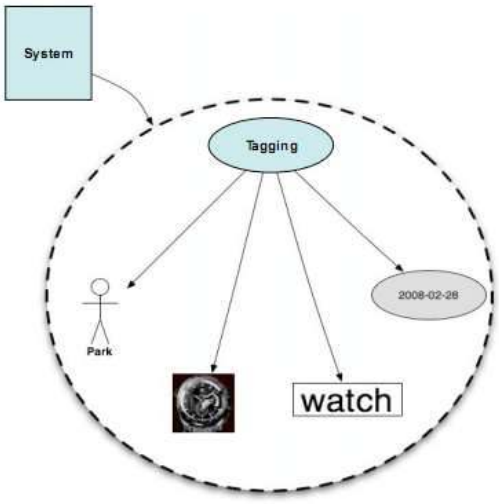
6. Peter MIKA

- Tripartite graph with hyperedges
- Three different and disjoint set of of vertices are considered by Mika:
  - the set of actors (users) -A-
  - the set of concepts (tags) -C-
  - the set of tagged objects -O-

  (e.g. photos, videos or web resources, like bookmarks, websites etc)
- Since in a social tagging system, users tag objects with concepts, ternary relations are created between the user, the concept and the object
- This tripartite hypergraph can be transformed into several bipartite graphs, each having a very specific meaning
  - AC - the graph that associates actors and concepts,
  - CO- the graph that associates concepts and objects
  - AI - the graph that associates actors and instances.

All Rights Reserved © Alcatel-Lucent 2009

CURRENTLY



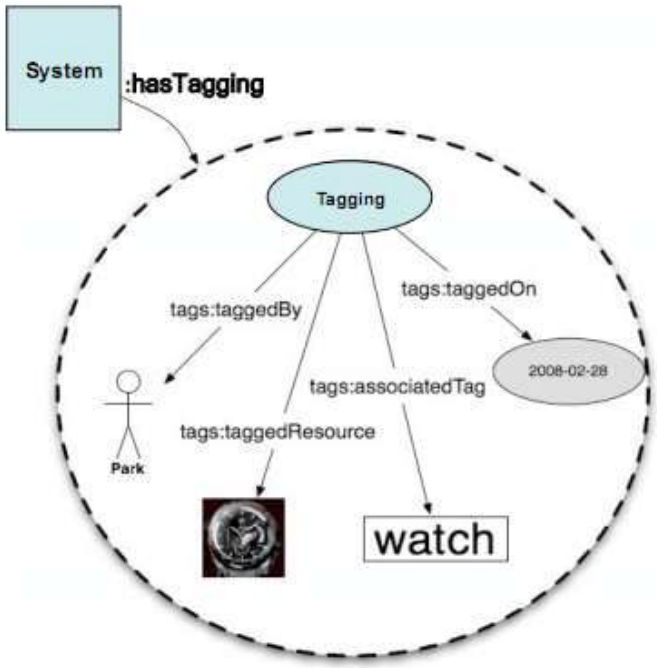
User assigns a Tag to a Resource (Object, Person) in a specific system.

We don't know the relationships.

No standard to enable reuse among different systems.

All Rights Reserved © Alcatel-Lucent 2009

TAG ONTOLOGIES : « A little semantics goes a LONG way »



All Rights Reserved © Alcatel-Lucent 2009

Social Networks: Going beyond the structure

Semantic Technologies for Interlinking Data Towards Semantic Social Networks

- Gruber's Model
- Newman
- SCOT
- MOAT
- Knerr's Model
- CommonTag Ontology
- GroupMe! Ontology

All Rights Reserved © Alcatel-Lucent 2009

1. Gruber's Model

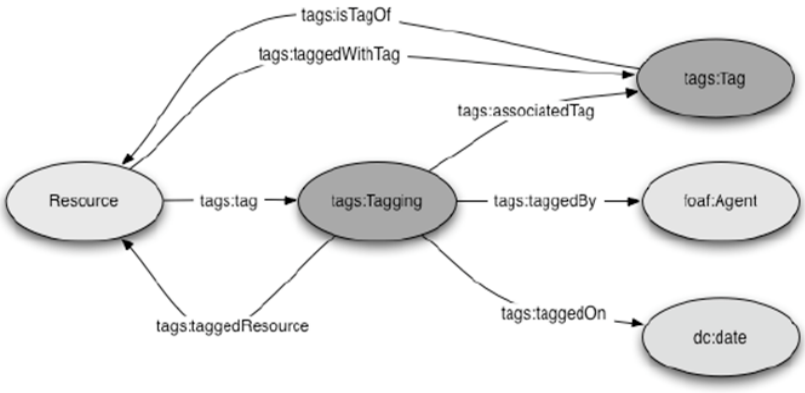
A first formal representation of tagged ressources:

- An **object**, i.e. the resource to be tagged
- A **tag**, i.e. the tag associated with the resource
- A **tagger**, i.e. the agent -more generally a person -that created the link between the tag and the object
- A **source**, i.e. the space where the tagging action has been performed, e.g.: Flickr, del.icio.us.

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

2. Newman's Model: the Tag Ontology

- Ontology of tags and tagging
- Representation of relations between tags and taggers by using existing ontologies (SKOS, Dublin Core, FOAF)
- Representation of relationships between tags
- Implementation: ***Revyu.com***



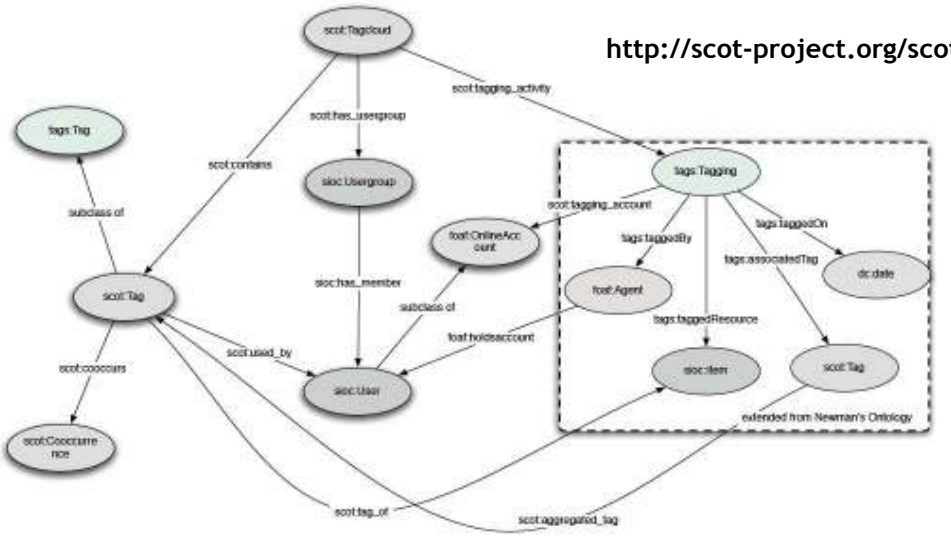
http://www.holygoat.co.uk/projects/tags/

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

3. SCOT (Semantic Cloud of Tags)

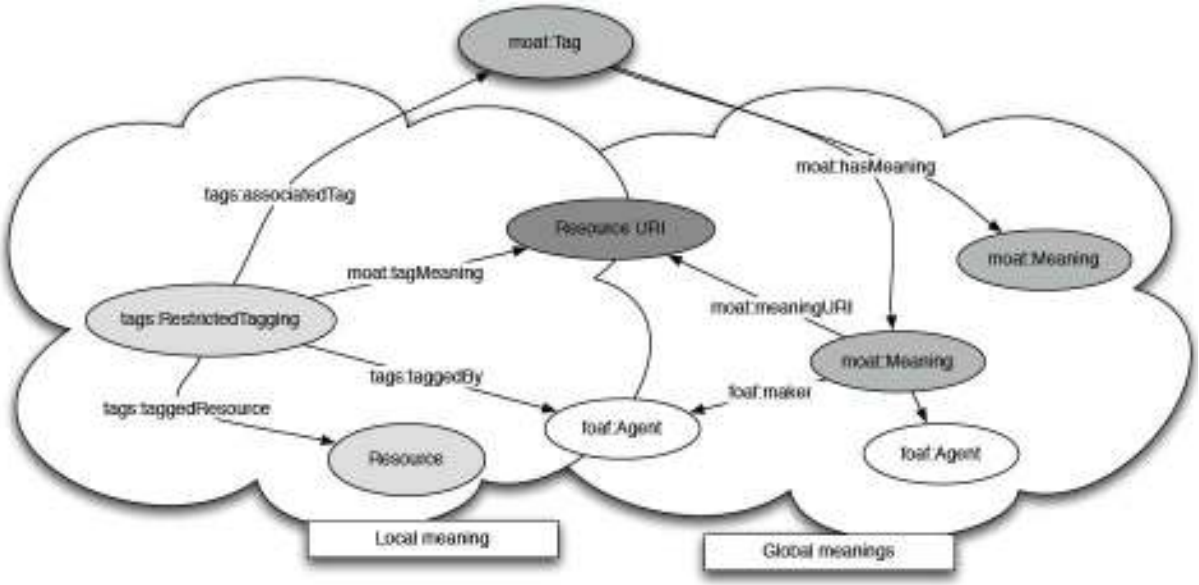
- Describes the structure and semantics of tagging data, enables social interoperability of tagging data among heterogeneous sources
- A combination of SCOT, FOAF and Dublin Core to describe tagging activity

http://scot-project.org/scot/



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

4. MOAT(Meaning of a Tag)

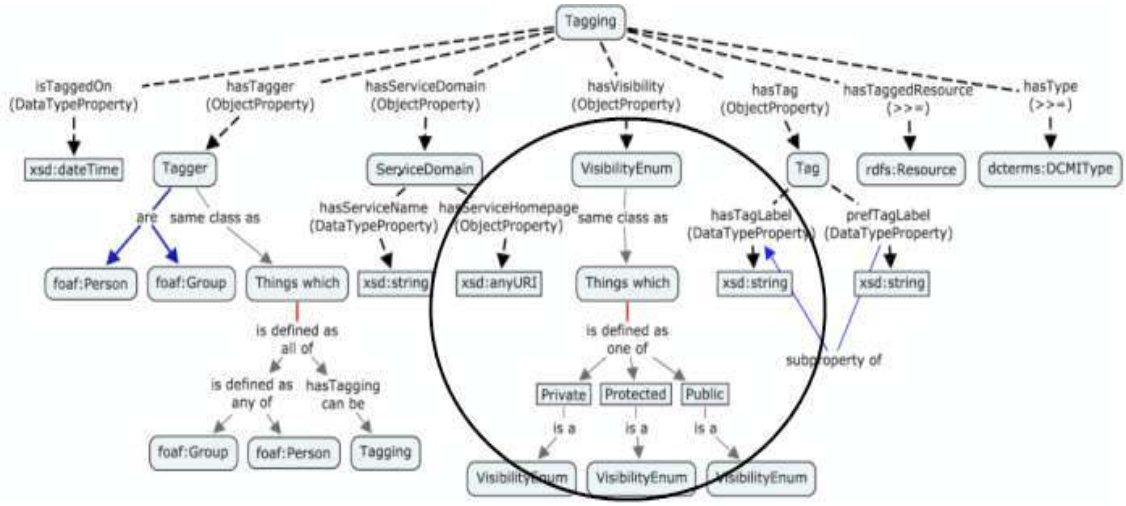


- Meaning of a Tag depends on context
- Introduction of the social aspects of tagging (community, sharing, ...)

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

5. Tagging Ontology of Knerr

- Tagging Ontology' proposed by Knerr
- The central element is the concept of Tagging, which is considered as being a tuple of (time, user, domain, visibility, tag, resource and type)

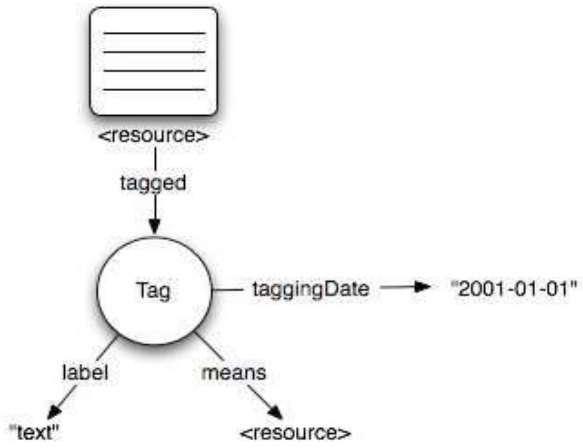


All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

6. CommonTag Ontology

- CommonTag is an open tagging format developed to make content more connected, discoverable and engaging

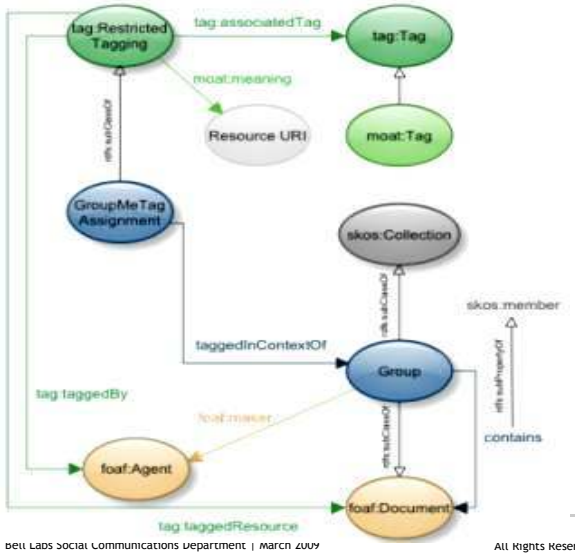
- The Common Tag structural model is very simple, states that a piece of content addressable through a URL (a resource) can be tagged with one or more Tag structures.



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

8. GroupMe! Ontology (Fabian Abel)

- Integration of additional semantics in tagging ontologies
  - User context : better description of the tagging activity
    - Timestamp, mood of the tagger
    - Background knowledge about the user: is he a reliable tagger?



All Rights Reserved © Alcatel-Lucent 2009

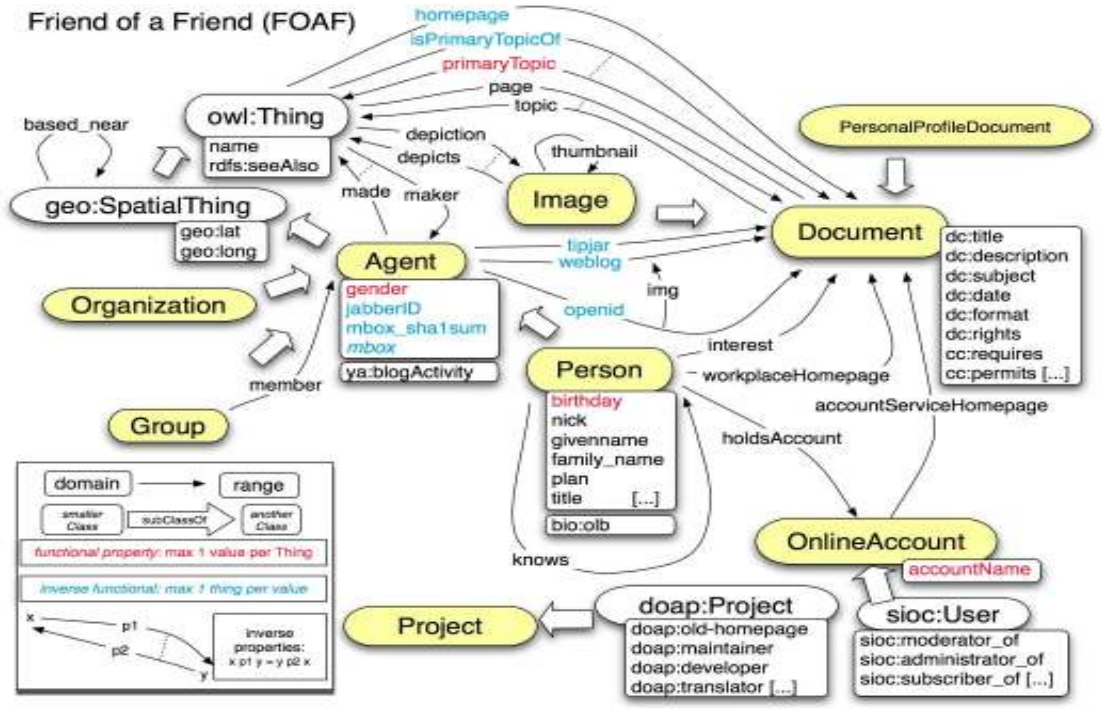
Comparison of reviewed models

Model	Namespace	Tagger Profile
Gruber	-	Tagger
Tag Ontology of Newman	tags	foaf: Agent
TagOnt	-	User
Echarte et al.	-	User
SCOT	scot	sioc: User
MOAT	moat	foat: Agent
NAO	nao	Party

Model	When	Where	Format
Gruber	TagAssertTime	TagSource	OWL
Tag Ontology of Newman	taggedOn	taggedResource	OWL
TagOnt	isTaggedOn	hasTagged	OWL
Echarte et al.	hasDateTime	Source	OWL
SCOT	updated	tagspace	OWL
MOAT	taggedOn	taggedResource	OWL
NAO	created/modified	Resource	NRL

All Rights Reserved © Alcatel-Lucent 2009

FOAF : knows



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Relationship Ontology (http://vocab.org/relationship/.html, 2004)

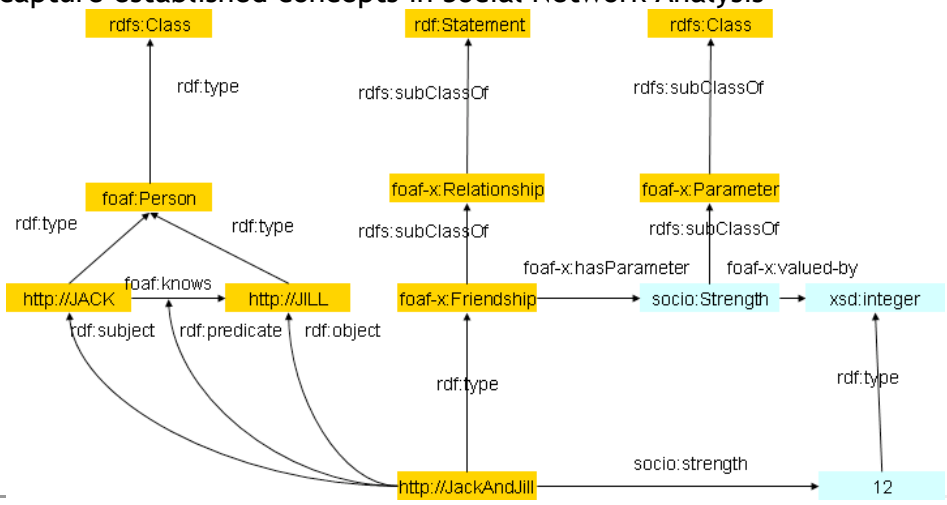
- A vocabulary for describing relationships between people
- No categories for relationships
- Extension of FOAF

acquaintance of
ambivalent of
ancestor of
antagonist of
apprentice to
child of
close friend of
collaborates with
colleague of
descendant of
employed by
life partner of
lost contact with
neighbor of
parent of
Relationship
works with
spouse of
employer of
enemy of
engaged to
friend of
grandchild of
grandparent of
has met
influenced by
knows by reputation
knows in passing
knows of
lives with
mentor of
participant
participant in
sibling of
would like to now

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

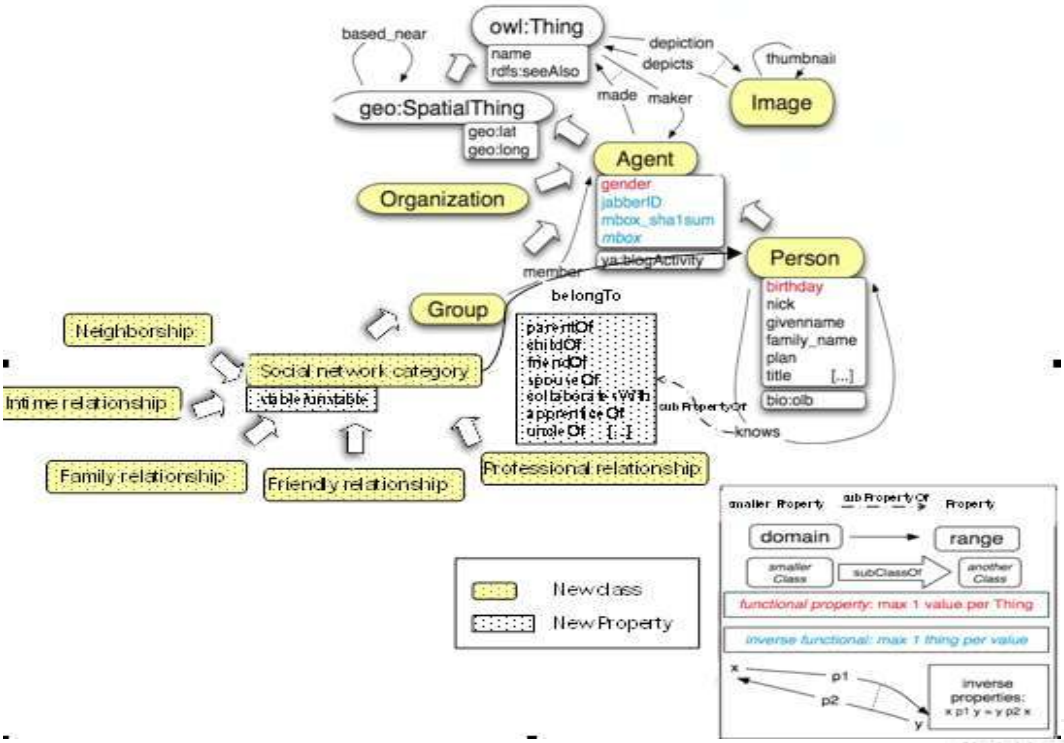
Description of Social Relations (Peter MIKA, 2006)

- ontological representations of social networks such as FOAF would need to be extended:
  - (1) to support the automated integration of social information on a semantical basis and
  - (2) to capture established concepts in Social Network Analysis



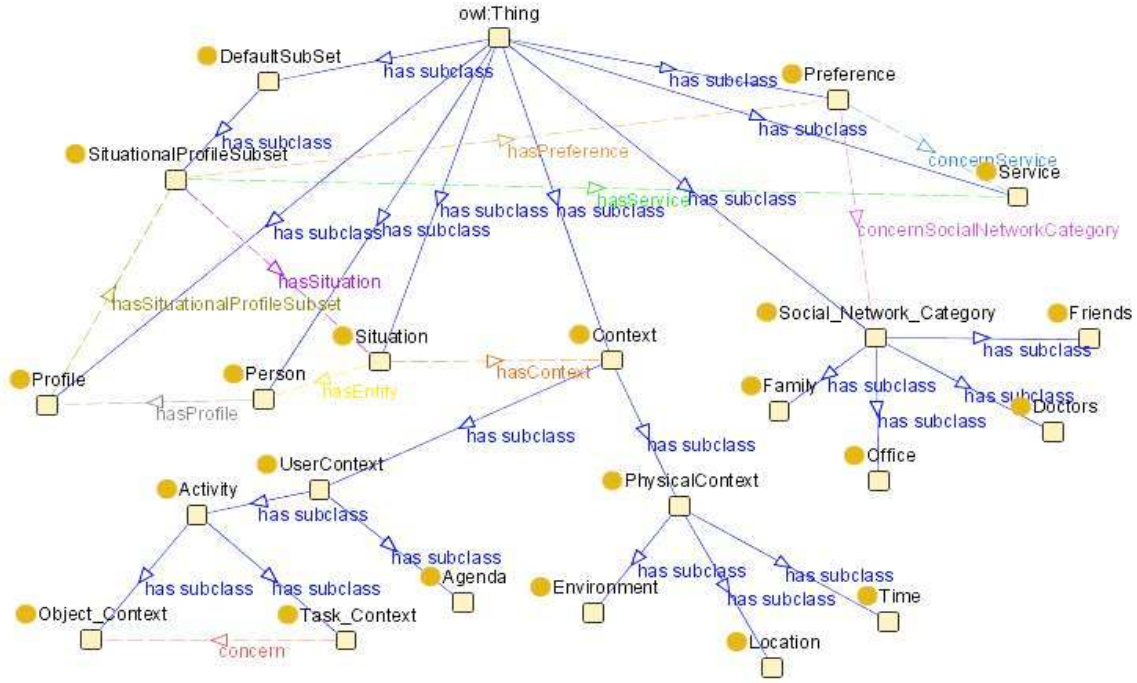
All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Relations and categories



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

SAUPO Ontology



All Rights Reserved © Alcatel-Lucent 2009

Semantic Annotation of People (1/3)

*PittCult system (Hynsuook et al., SigWeb)**

- Event recommender based on the user's trusted human network
- Explicit definition of trust network:
  - The user tags people with levels of trust and for a specific event
    - I trust Alice for thriller movies very much
    - I trust Bob for doctors pretty much
- Implicit definition of trust network:
  - Based on invitations and ratings of these invitations by the recipients

*D. Lee, Peter Buzylowski. **PITTCULT: trust-based cultural event recommender.**
Proceedings of the 2008 ACM conference on Recommender systemsLausanne, Switzerland
Pages 311-314
Year of Publication: 2008

All Rights Reserved © Alcatel-Lucent 2009

Semantic Annotation of People (2/3)

*Improving personal privacy with people tagging**

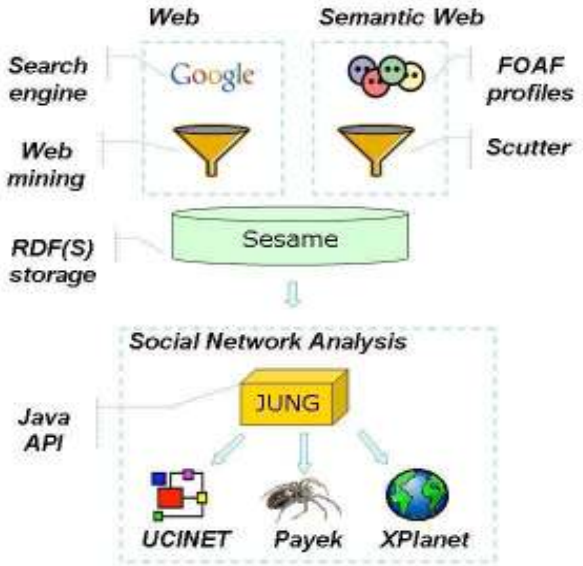
User centered mechanism for controlling access to the vast amount of user-created content that exists in social systems.

- Facilitate relationship management for personal information sharing
- Folksonomic organisation of the social networks
  - Many tags associated to one person
  - Choice of tags is free
  - For each document the owner can specify the target audience category
  - Restrict visibility of a document to a category/group

*M.Razavi. Improving personal privacy in social systems with people-tagging.
Proceedings of the ACM 2009 international conference on Supporting group work Sanibel Island, Florida, USA
SESSION: Tagging Pages 11-20
Year of Publication: 2009

All Rights Reserved © Alcatel-Lucent 2009

Case study "Flink" - Peter Mika - (1/2)



All Rights Reserved © Alcatel-Lucent 2009

Case study "Flink" - Peter Mika - (2/2)

- Flink is a website presenting the social networks and research activity of Semantic Web researchers based on a number of sources (web pages, publication databases, email archives, FOAF data). Flink has won 1st prize at the (Semantic Web Challenge) of 2004.
- RDF representation of Data from different sources

All Rights Reserved © Alcatel-Lucent 2009

Relevant Bibliography for Chapters 1, 2

- LODr -- A Linking Open Data Tagging System. Alexandre Passant. In: Proceedings of the First Social Data on the Web Workshop, Karlsruhe, Germany, October 27, 2008, CEUR Workshop Proceedings, ISSN 1613-0073, online CEUR-WS/Vol-405//paper2.pdf
- Chris Bizer, Tom Heath, Danny Ayers, and Yves Raimond. Interlinking open
- data on the web. In Poster, 4th Annual European Semantic Web Conference
- (ESWC2007), Innsbruck,Austria, 2007.
- Peter Mika. Ontologies Are Us: A Unified Model of Social Networks and Semantics. In Proc. The Semantic Web – ISWC 2005. 522--536.
- T.T. Gruber. Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human Computer Studies. Volume 43, 2005, pages 907--928.
- T. Gruber. Ontology of Folsksonomy: A Mash-Up of Apples and Oranges. Internation Journal on Semantic Web and Informaion Systems. Number 3, 2007.
- Niki Papailiou, Costas Christidis, Dimitris Apostolou, Gregoris Mentzas, Rosa Gudjonsdottir. Personal and Group Knowledge Management with the Social Semantic Desktop. Collaboration and the Knowledge Economy: issues, Applications and Case Studies, (2008)

All Rights Reserved © Alcatel-Lucent 2009

Relevant Bibliography for Chapters 1, 2

- Knerr, T.: Tagging Ontology- Towards a Common Ontology for Folksonomies (2006), Available at: http://code.google.com/p/tagont/
- A. Passant. Using Ontologies to Strenghten Folksonomies and Enrich Information Retrieval in Weblogs. In Int. Conf. on Weblogs and Social Media. 2007.
- K. Barnard, D. Forsyth. Learning the Semantics of Words and Pictures. International Conference on Computer Vision, II:408-415. 2001.
- R. Picard, T. P. Minka. Vision Texture for Annotation. MIT Media Laboratory Perceptual Computing Section. 302. 1995.
- Francisco Echarte, Jose Javier Astrain, Alberto Cordoba, Jesus Villadangos. Ontology of Folksonomy: A New Modelig Method. Proceedings of the Semantic Authoring, Annotation and Knowledge Markup Workshop (SAAKM2007) located at the 4th International Conference on Knowledge Capture (KCap 2007), Whistler, British Columbia, Canada, October 28-31, 2007.
- TagCommons Initiative. http://tagcommons.org/ -visited June 2009 -
- CommonTag Initiative. http://commontag.org -visited June 2009 -

All Rights Reserved © Alcatel-Lucent 2009

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

All Rights Reserved © Alcatel-Lucent 2009

WWW Working Group: Future of Social Networking (1/7)
What are the perspectives of Social Networks?

- 72 Positions Papers Presented
- We will survey and summarize the most interesting perspectives and research issues discussed

Source:

http://www.w3.org/2008/09/msnws/papers/

All Rights Reserved © Alcatel-Lucent 2009

What are the perspectives of Social Networks? (2/7)

Hot Topics:

- the nature of **less centralized** and more **distributed** social networks architectures, including their design and possible business and technical challenges associated with distributed social networking
- the increase of **contextual information** associated with social networking users, its use and possible abuse
- the impact of context as well as existing lack of policies within networks on **user privacy risks** and the establishment of a Web of trust, and
- the tendency for existing social networking platforms to exclude those potential users with **disabilities or constrained devices** (e.g., mobile)

All Rights Reserved © Alcatel-Lucent 2009

Less centralized social network architectures (3/7)

- Growing number and maturity of data interoperability formats and protocols
- Reduce the detrimental effects of architectural silos (walled gardens)
- A decentralized architecture for social networks would allow the user to choose how many accounts and profiles she desires, and to use any of these on any social networks which would support the proposed or user-selected features

Examples:

"Let's store our FOAF profile on a trusted server and thus have more control."

- Users define access control policies to their data

Decentralisation: The Future of Online Social Networking

(Yeung, Liccardi, Lu, Seneviratne, Berners-Lee)

All Rights Reserved © Alcatel-Lucent 2009

The role of context (4/7)

- Context is playing a large and increasingly important role in the social networking industry
- Context has many elements and sources which at this time are poorly understood and managed
- A number of social networks already take advantage of geolocation, and some even are purely based on the communities created around sharing some specific contextual data

Example(TELEFONICA, ESP):

- Generate a user social map (How much time is spent near a contact in my social graph?), automatically update weights based on real-time presence (Bluetooth)
- Create an automatic and deep user profile (applications can give deeper and more accurate recommendations)
- Analyse compatibilities between people
- Electronic post-its attached to places: they appear on the phone when a user is nearby (good food here, happy hour)

All Rights Reserved © Alcatel-Lucent 2009

Mobile Devices (5/7)

- Better adapted user experience : mobile devices
- social networking would be a priori a great opportunity for people with disabilities to increase their social interactions
- deeper involvement of mobile network operators with their user base
- heavy use of CAPTCHA, the lack of annotations brought with user generated content, and the use of javascript in an inaccessible fashion

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Privacy and trust (6/7)

- Social networking technologies need to preserve the possibility for a user to fragment its identity across various profiles
- Hide, blur or lie about the user's current context
- Differences between cultures, legislation makes this problem difficult

- Position paper from University of Reading

Problem statement:

As the number of services being provided continues to grow, it is particularly important to recognise that users will often keep accounts with systems they no longer actively use, especially if migrating data between them remains difficult.

Proposition/Motivation/Position:

We have evidence of a number of people having multiple Facebook accounts in order to maintain a separation between professional and social aspects of their lives, and in some cases multiple 'professional' accounts to provide different views of themselves to different groups of colleagues.

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Business Models for Social Networks (7/7)

Existing business models:

- subscription fees for end-users - although that option was noted as often difficult to set up
- advertising
- broker for communities sharing similar interests
- access to premium content or services

All Rights Reserved © Alcatel-Lucent 2009

Case study "Social communication analyzer" (1/2)

Based on xobni tool, this study describes what can be an in depth social communication usage analyzer tool.

- Create a profile for every person in the social network.
- Each profile can contain relationship statistics, number of incoming and outgoing communications (email, sms, SIP, IM, voice calls, files exchange, …), contact information, related peoples, threaded conversations, attachments.
- Display useful statistics about communication relationship: Day range I use IM with Tom? Is Jena a SIP user with me? Give me the time of day email conversation with my boss is very high? Give the 1st ranked person in my sms conversations. How many youtube video about Hillary Clinton I watch last week? …
- Extract automatically phone numbers, emails address, SIP numbers from communications and signatures and update profiles and my social network even if not in my address book.

All Rights Reserved © Alcatel-Lucent 2009

Case study "Social communication analyzer" (2/2)

- Exposes communication hidden social network: analyses IM conversation between Bob and me and displays a list of related peoples. Example, Bob sister appears high in his list of related people. I can click on her to see her profile.
- Offer a quickest way to find and retrieve an important communication. When I receive a mail from Bill I often want to see conversations in the past, then a conversation windows displays any conversations in the past.
- Makes finding attachment, file exchanged interesting web browsing easy, no searching.

All Rights Reserved © Alcatel-Lucent 2009

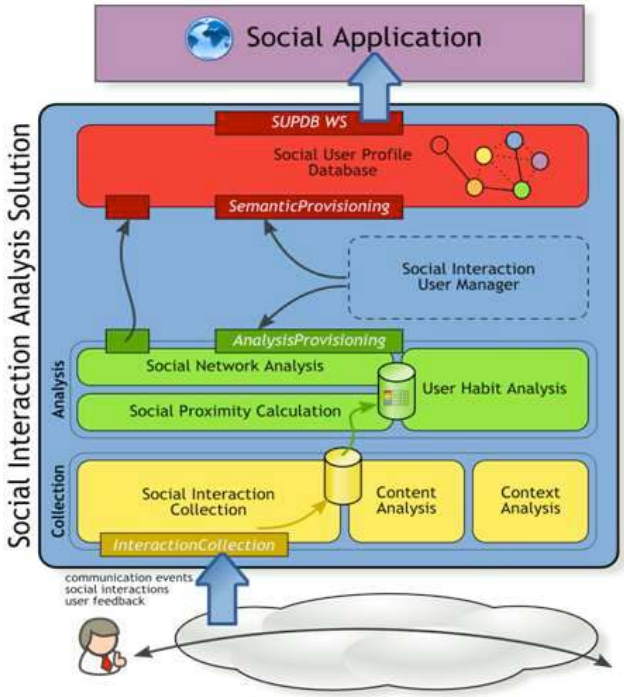
Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

All Rights Reserved © Alcatel-Lucent 2009

Social Helper: Exploiting the Structure of Social Networks

ASONAM 2009



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

ASONAM 2009

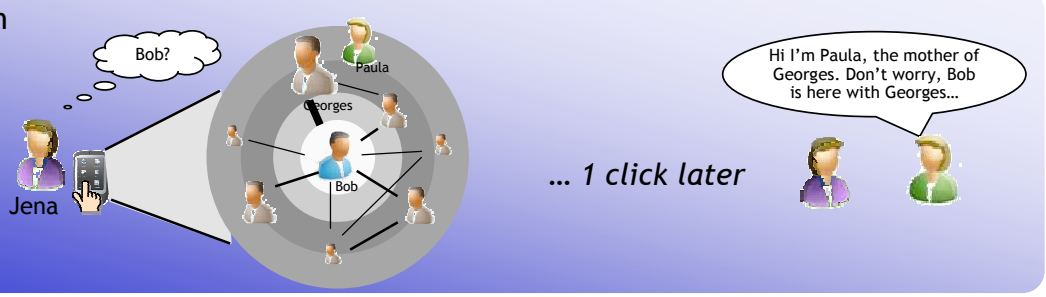
What drives us: how can applications benefit from my social interactions?

Problem

Jena, anxious mother, can not reach her kid Bob who's late…

… and after several calls, she still has no solution!

Solution

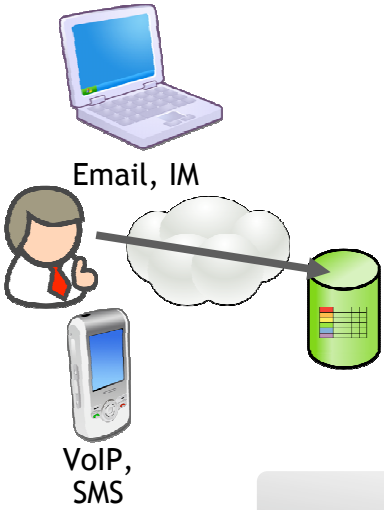


Emergency call - Exploring social graph to reach the relevant contact

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Social Helper: Social Proximity

ASONAM 2009



A calculation for social proximity can take into account

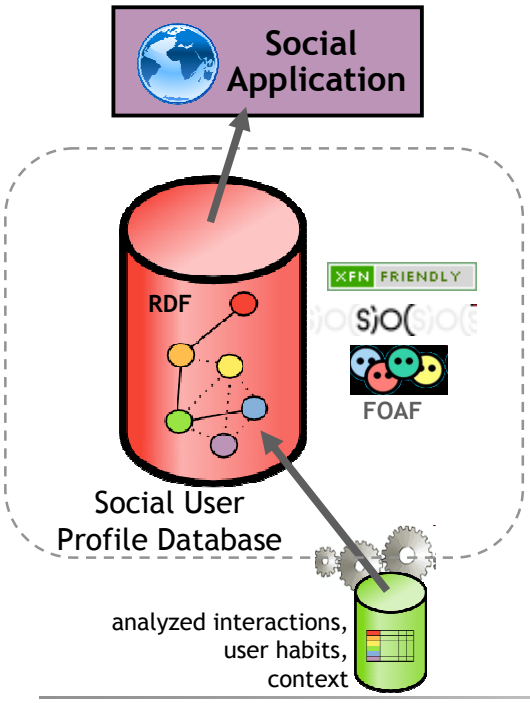
- the **type(s)** of communication,
- the **number** and **frequency** of social interactions between actors,
- the **length** or **duration** of the interaction,
- **weighted** according to the individual usage patterns and/or use cases,
- the **shape** of the conversation (new messages, reply/forward, ignored, number of participants…)

The **content** and **context** of the social interactions can be used to create useful filters for social applications.

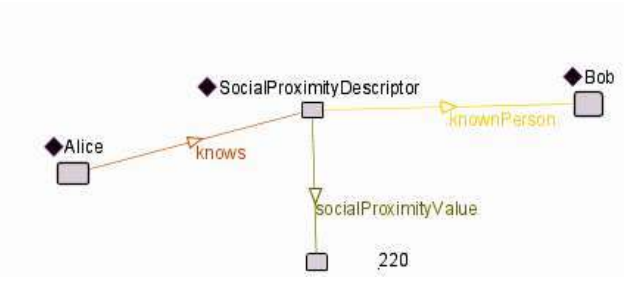
All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Social Helper: Semantic User Profile

ASONAM 2009



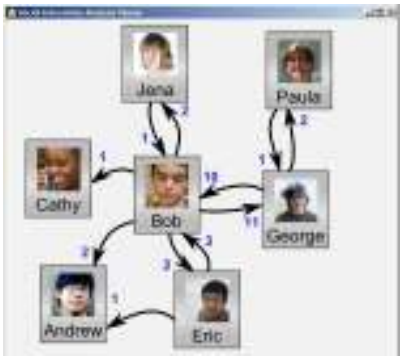
- Store the social graph in a computer-understandable vocabulary from semantic web technologies.
- Enhance the representation with standardized, rich ontologies (FOAF, XFN, SIOC, OWL)
- Current implementation:
  - Storage in RDF format, close to the FOAF standard.
  - Accessed by social applications through a well-defined web service.

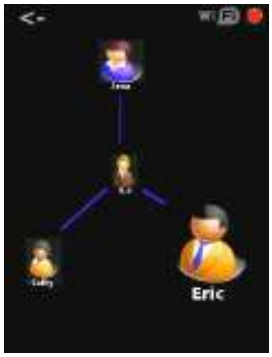


All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Social Helper: Exploiting the Structure of Social Networks

ASONAM 2009





All Rights Reserved © Alcatel-Lucent 2009

Reference

ASONAM 2009

Ryan SKRABA, Mathieu BEAUVAIS, Johann STAN, Abderrahmane MAARADJI, Johann DAIGREMONT:

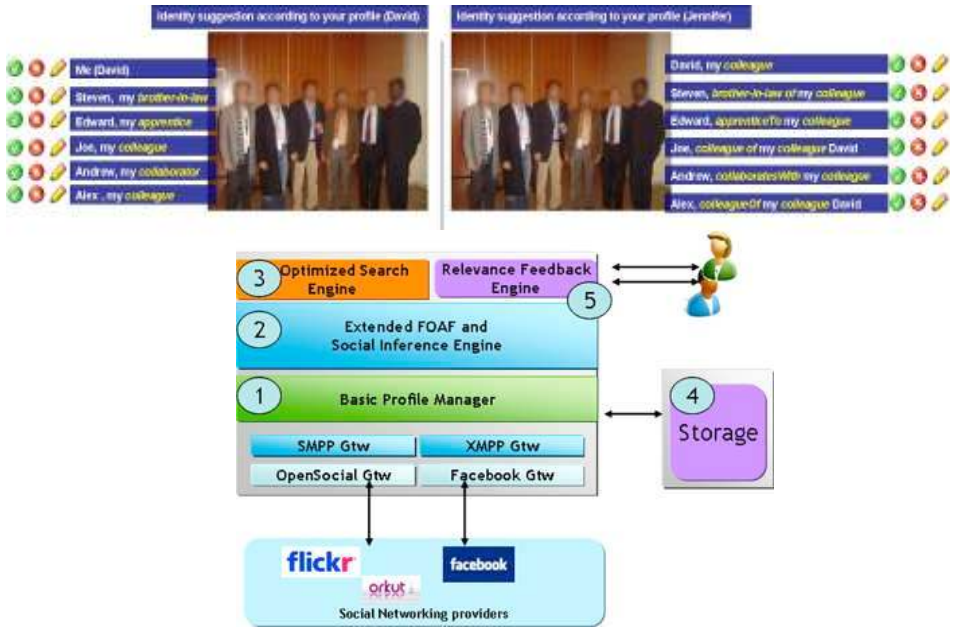
Developing Compelling Social-Enabled Applications with Context-based Social Interaction

ASONAM 2009, Athens Greece (to appear)

All Rights Reserved © Alcatel-Lucent 2009

Extended Social Tags: Identitiy Tags Meet Social Networks

Social Computing 2009



All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Reference

Social Computing 2009

Sonia LAJMI, Johann STAN, Hakim HACID, Elod EGYED, Pierre MARET:

Extended Social Tags: Identity Tags Meet Social Networks

2009 International Conference on Social Computing, Vancouver, Canada (to appear)

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Semantic Web Development Environment

- Compliling and execution tools: Java 1.6 SDK
- Eclipse
- Ontology editing tool: Protégé Ontology Editor 4.0 Alpha
- Semantic Web Programming Framework: Jena Semantic Web Framework 2.5.6
- Ontology Reasoner: Pellett 1.5.2.

http://semwebprogramming.org/

All Rights Reserved © Alcatel-Lucent 2009

Presentation outline

1. Presentation of Bell Labs
2. Part I: Introduction to the structure of social networks
   - Social computing
   - Main properties of social networks
   - Future of social networks
   - References
3. Part II: Semantics of social networks
   - Survey of Semantic Technologies for Social Information Representation
   - Communities, Tagging, Folksonomies
   - Future of Social Networking
   - Work in the Bell Labs
   - Conclusion

All Rights Reserved © Alcatel-Lucent 2009

Conclusion

- A survey of the Social Web and the Semantic Web
- Several examples of Social Networking applications
- The Semantic Web offers a compelling foundation to building innovative applications that leverage Social Data
- Social Network Data is not sufficiently exploited
- Application Enablement
- Tag Ontologies allow to represent social networks that go beyond the structural view (explicit declaration of relationships)
- Tag Ontologies provide the possibility for machine-processable representations that can be shared across social tagging systems
- The vision is of a semantic social network, where relationships are typed

All Rights Reserved © Alcatel-Lucent 2009
Alcatel·Lucent

Thank you for your attention!

www.alcatel-lucent.com

SESSION 3

Fouille de données

Le forage des réseaux sociaux et la découverte de communautés virtuelles

O. R. ZAIANE
(UNIVERSITY OF ALBERTA, CANADA)

Le Forage des Réseaux Sociaux et la découverte de communautés virtuelles

Osmar R. Zaïane
University of Alberta, Canada

École d'Été Web Intelligence 2009
Projet Web Intelligence du Cluster ISLE, Région Rhône Alpes
du 6 au 10 juillet 2009, Véranne, France

Qui est le présentateur?

Osmar R. Zaïane

Professeur en informatique à l'Université d'Alberta au Canada, et directeur sientifique du laboratoire *Alberta Ingenuity Centre for Machine Learning*. Specialiste du forage de données, il s'intéresse au applications liées à la santé.

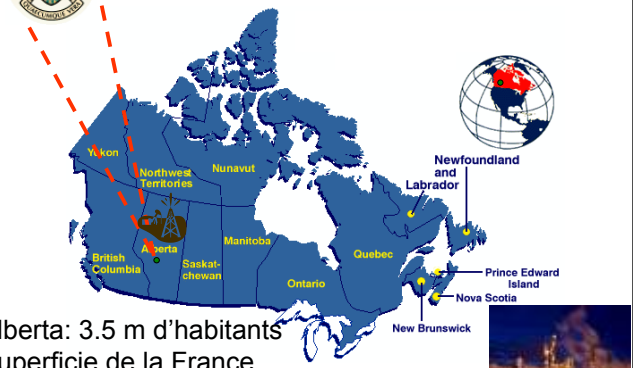
UNIVERSITY OF ALBERTA

Osmar R. Zaïane, Ph.D.
MacCalla-Killam Professor
Department of Computing Science

352 Athabasca Hall
Edmonton, Alberta
Canada T6G 2E8

Telephone: Office +1 (780) 492 2860
Fax +1 (780) 492 1071
E-mail: zaiane@cs.ualberta.ca
http://www.cs.ualberta.ca/~zaiane/

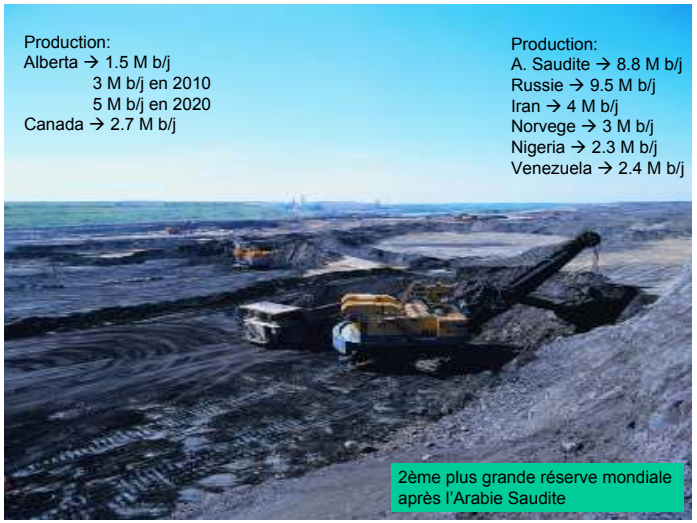
Edmonton est la 5ème plus grande ville du Canada avec un peu plus d'1 million d'habitants. C'est la capitale de l'Alberta.



- Alberta: 3.5 m d'habitants
- Superficie de la France
- Province riche en pétrole

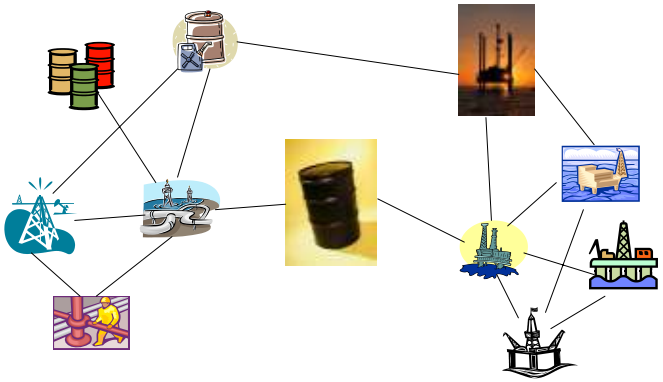
Production:
Alberta → 1.5 M b/j
3 M b/j en 2010
5 M b/j en 2020
Canada → 2.7 M b/j

Production:
A. Saudite → 8.8 M b/j
Russie → 9.5 M b/j
Iran → 4 M b/j
Norvege → 3 M b/j
Nigeria → 2.3 M b/j
Venezuela → 2.4 M b/j



2ème plus grande réserve mondiale après l'Arabie Saudite

Forage du ~~Pétrole~~ de Données (des réseaux sociaux)

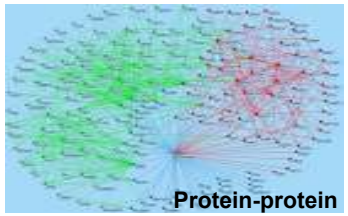


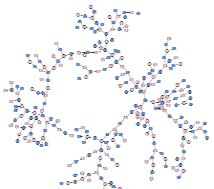
Tutorial Structure

- Analyse des Réseaux Sociaux
  - Historique, conceptes et recherche
  - Les applications et les problèmes
- Ordonnancement dans un réseau
- Découverte des communautés
  - Partition de Graphes
  - Approches hiérarchiques
  - Approches globales et approches locales
- Intersection de communautés
- Visualization des réseaux

SNA, a multidisciplinary field

Public health

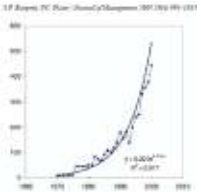

Protein-protein


Social studies http://www-personal.umich.edu/~mejn/networks

Business

A quick History

- Social network analysis is a key technique traditionally studied in sociology, anthropology, epidemiology, sociolinguistics, psychology, etc. Today it is a modern technique in marketing, economics, intelligence gathering, criminology, medicine, computer science, etc.
- J. Barnes is credited with coining the notion of social networks (theory) in 1954.
- Precursors of social network theory date from the 19th century such as Simmel, Durkheim and Tönnies.
- Massive increase in studies of social networks (in social sciences) since the 1970s.
- The increase of available data, the Internet phenomenon, Web 2.0, etc. have only catapulted the interest in SNA research



What is Social Network Analysis?

- [Wikipedia] A social network is a social structure made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of interdependency, such as values, visions, ideas, financial exchange, friendship, sexual relationships, kinship, dislike, conflict or trade.
- Social Network Analysis (SNA) is the study of social networks to understand their structure and behaviour.
- Which node is the most influential? which one is central? What are the hubs? What are the groups? Who knows who?, What are the short paths? What is perceived by who? ...

Networks in Social and Behavioral Sciences

[Monge, and Contractor, 2003]

- Social Networks
  - **Who knows who?**
- Socio-cognitive Networks
  - **Who thinks who knows who?**
- Knowledge Networks
  - **Who knows what?**
- Cognitive Knowledge Networks
  - **Who thinks who knows what?**

	Acquaintance	knowledge
Reality	Social Network	Knowledge Network
Perception	Socio-cognitive Network	Cognitive knowledge Network

- Socio-centric Analysis
  - Emerged in sociology: quantification of interaction among a group of people. Focus on Identifying global structural patterns in a network.
- Ego-centric Analysis
  - Emerged in psychology and anthropology: quantification of interaction between an individual (ego) and others (alters) directly or indirectly related to ego.

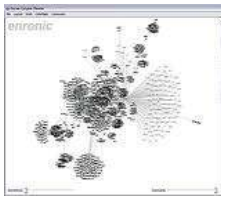
Vulgarisation

- Six degrés de séparation (Chaînes par Frigyes Kari 1929). Hypothèse: Le monde moderne est entrain de rétrécir à cause de la connectivité entre humains. A utilisé l'idée des six degrées de liberté en méchanique.

- Milgram's Paradox: effet du petit mondeb(Stanley Milgram, Sociologue 1967) Fameuse expérience en 1970 envoyant des lettres d'Omaha à Boston: 64/296 seraient arrivées, chemin moyen 5.5~6. Google
- PageRank du moteur de recherche 1998 utilise le réseaux des citations des pages Web pour estimer ... tance des pages web et les ordoner
- ... de réseautage des internautes

Le fameux cas d'Enron

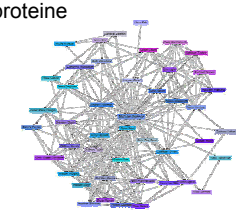
- Société de courtage pour ressources naturelles
- Faillite en 2001. Suite au scandale financier, la compagnie Arthur Andersen a été dissoute.
- Données sont publiques
  - http://www.cs.cmu.edu/~enron
  - 151 usagers
  - 200 399 messages couriel
  - Classique pour la recherche
    - Modeling a Socio-Cognitive Network
    - Quantitative Measures for Perceptual Closeness
    - Automatic Extraction of Concealed Relations
    - …



Visualisation du réseau d'email d'Enron, Jeffrey Heer, 2005

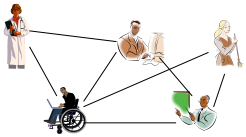
Réseaux

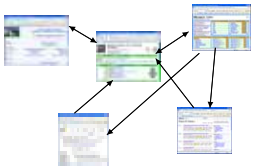
- Un réseau est représente par un graphe G(V,E):
  V = nœuds, E = liens (entre paires de nœuds)
- Exemples de réseaux:
  - V = personnes → Réseau social commun
  - V= pages Web → Toile mondiale
  - V= proteines → Interaction proteine-proteine



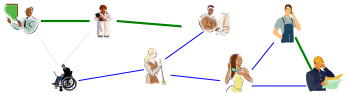
Types de relations et de réseaux (1)

- Réseaux à relation unique
  - Individus sont reliés par une même relation



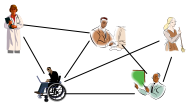


- Réseaux à multiples relations
  - Individus ont différents types de relations



Types de relations et de réseaux (2)

- Relation homogène
  - Relations entre individus du même type



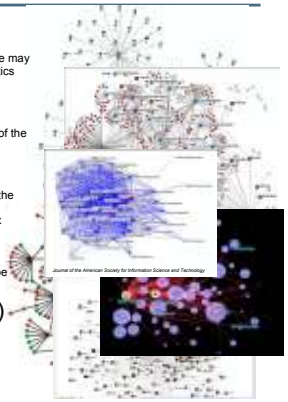
- Relation hétérogène
  - Relation entre individus de types différents

Other key concepts

- Edge Weight : interaction frequency, importance of information exchange, intimacy, emotional intensity, etc.
- Symmetric relation or not (directional)
- Centrality: determines the relative importance of a vertex (or edge) within a network.
  - Degree Centrality: Mesures the normalized number of edges incident upon a node *n;*
  - Betweeness Centrality: Measures how many times a node *n* occurs in a shortest path between any other 2 nodes in the graph;
  - Closeness Centrality: Mean shortest path distance between a node *n* and all other nodes reacheable from it;
  - Eigenvector Centrality: Measures importance of a node n by assigning a score to each node based on the principal that connections to high-scoring nodes contribute more to the score of a node in question than equal connections to low-scoring nodes (e.g. PageRank).

Applications

- Terrorism and crimes
  - Social Network analysis is an important part of a conspiracy investigation and is used as an investigative tool. Group structure may be important to investigations of racketeering enterprises, narcotics operations, illegal gambling, and business frauds.
- Medicine – epidemiology
  - valuable epidemiological tool for understanding the progression of the spread of an infectious disease.
- Marketing
  - Emarketer projected that Social Network Marketing spending in the USA will reach approximately $1.3 billion in 2009. http://www.emarketer.com/Reports/All/Emarketer_2000541.aspx
- Product Recommendation
  - Current recommendation models assume all users' opinions to be independent. Use of SNA relaxes the iid assumption.
- Bio-informatics (protein interaction)
- Relevance Ranking
- Information and Library Science



Prominent problems in SNA

- Social network extraction/construction
- Link prediction
- Approximating large social networks
- Identifying prominent/trusted/expert actors in social networks
- Search in social networks
- Discovering communities in social networks
- Knowledge discovery from social networks
- Predicting evolution

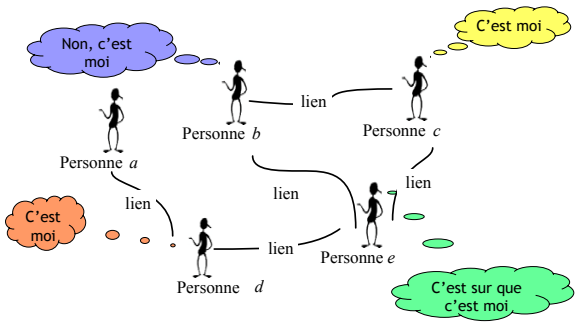
Analogy with Clustering

Tutorial Structure

- Analyse des Réseaux Sociaux
  - Historique, conceptes et recherche
  - Les applications et les problèmes
- Ordonnancement dans un réseau
- Découverte des communautés
  - Partition de Graphes
  - Approches hiérarchiques
  - Approches globales et approches locales
- Intersection de communautés
- Visualization des réseaux

Problème

Comment s'assurer de savoir qui est la persone la plus importante du réseau sociale?

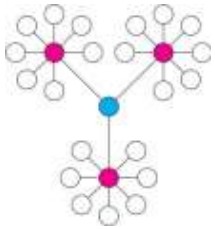


Ordonnancement d'entités

- Comprendre les relations entre entités.
- Trier les ensembles d'objets dans le réseau basé sur les relations entre ces objets et la structure générale des liens.
- Étudié surtout en sociologie (nœud influent) et en informatique

Ordonnancement Sociologique

- Centralité: mesure l'importance des individus dans un réseau sociale
- Degré de centralité
  - simple
  - mesure seulement la locale
- Centralité des vecteurs



Centralité des vecteurs propres

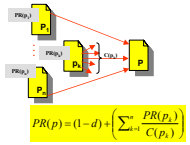
- Intuition: Les connections d'un nœud à grande valeur doivent contribuer davantage au nœud en question que les connexions venants de nœuds à faible valeur.
- Pour le nœud *i*, la centralité se définit comme étant proportionnelle à la somme des valeurs des nœuds qui connectent à *i*.

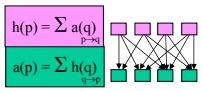
$$x_i = \frac{1}{\lambda}\sum_{j=1}^{N} A_{i,j}x_j \longrightarrow X = \frac{1}{\lambda}AX \longrightarrow AX = \lambda X$$

- La plus grande valeur propre donne la mesure de centralité désirée. (Théorème de Perron–Frobenius）

Approches en informatique

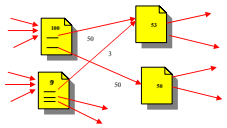
- Trier les documents Web pour un moteur de recherche
- Considérer la structure des hyperliens ainsi que le contenu des pages
- PageRank et HITS





Intuition de PageRank

- Appliquer la centralité des vecteurs propres dans le domaine du Web: un document est placé haut si d'autres documents haut placés sont connectés à celui-ci.
- PageRank peut être interprété comme une probabilité de distribution pour représenter la probabilité qu'un internaute en cliquant aléatoirement les liens d'une page à l'autre revienne au point de départ.



D'autres approches

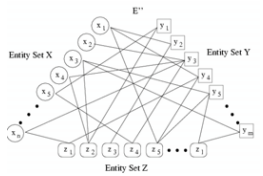
- Jeh et Widom :SimRank. Deux pages sont similaires si elles sont liées ou apparentées à des pages similaires
- Parcours aléatoires pour ordonner des tuples dans les bases de données (ObjectRank, RelationalRank, BANKS…)
- PageRank sur un ensemble de synonymes de WordNet pour trouver des propriétés sémantiques.
- Réseaux de couriel, réseaux téléphoniques, attribuer des mots clés aux images…
- Position de nœud (Node Position measure) [Bródka, Musiał, Kazienko 2009]

$$NP_{n+1}(x) = (1-\varepsilon) + \varepsilon \cdot \sum_{i=1}^{m} \left[NP_n(y_i) \cdot C(y_i, x) \right]$$

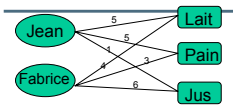
  - PageRank distribue PR d'une page equitablement a tous les fils du noed
  - NP distribue NP d'une page proportionellement au poids de chanque lien.

La solution: TupleRank

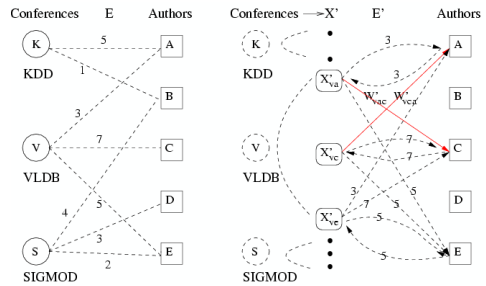
- Le réseau hétérogène est modélisé par un graphe n-aire.
- Le graphe est augmenté par des nœuds de remplacement pour modéliser des relations particulières.
- Un parcours aléatoire avec redémarrage est défini sur le graphe.
- La direction du parcours d'une partie du graphe à l'autre est contrôlée.



Extension du graphe



Les transactions de Jean ou Fabrice sont en fait perdues. La relation entre produits dans une même transaction n'est plus.

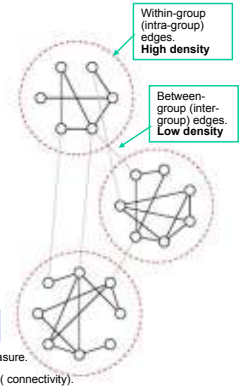


Tutorial Structure

- Analyse des Réseaux Sociaux
  - Historique, conceptes et recherche
  - Les applications et les problèmes
- Ordonnancement dans un réseau
- Découverte des communautés
  - Partition de Graphes
  - Approches hiérarchiques
  - Approches globales et approches locales
- Intersection de communautés
- Visualization des réseaux

What is Community Structure?

- *Community structure* denotes the existence of densely connected groups of nodes, with only sparser connections between groups.
- Many social networks share the property of a community structure, e.g., WWW, tele-communication networks, academic collaboration networks, friendship networks, etc.

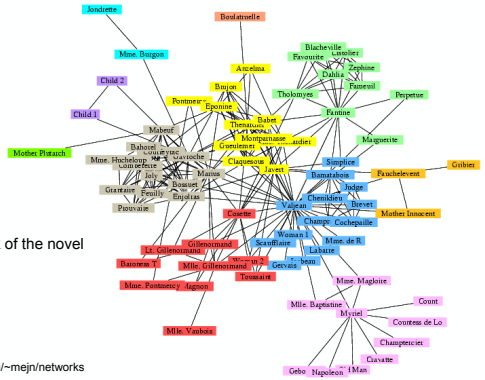


Many similarities with data **Clustering**

Clustering is dividing the data points into classes according to some similarity measure.

Community structure: dividing the network into groups according to structural info.(connectivity).

Community Structure Example

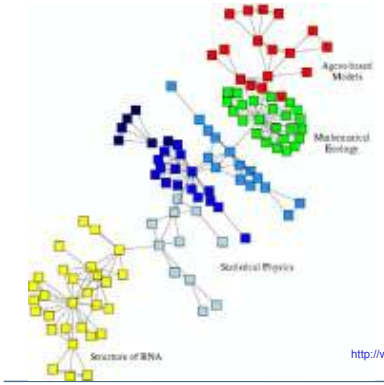


A social network of the novel *Les Misérables*

(Victor Hugo)

http://www-personal.umich.edu/~mejn/networks

Community Structure Example

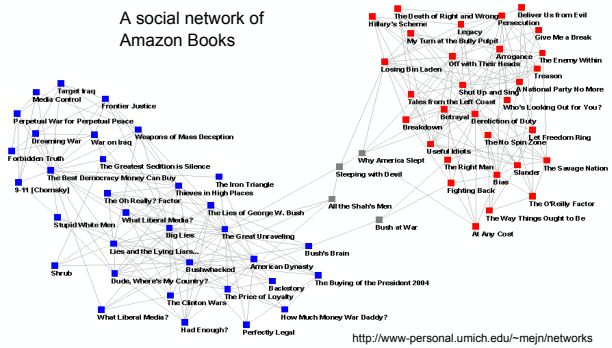


An academic (physicists) collaboration social network

http://www-personal.umich.edu/~mejn/networks

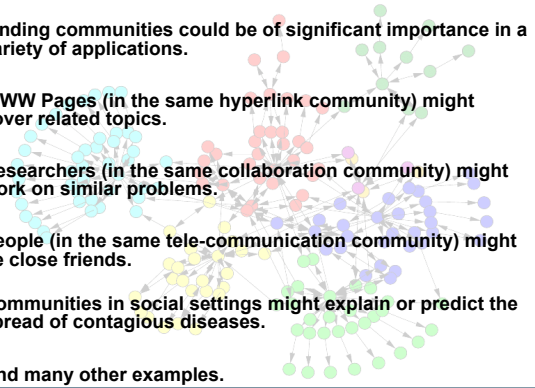
Community Structure Example

A social network of Amazon Books



http://www-personal.umich.edu/~mejn/networks

It is important!

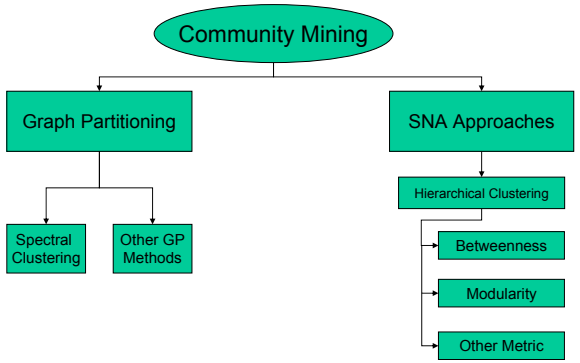


- **Finding communities could be of significant importance in a variety of applications.**
- **WWW Pages (in the same hyperlink community) might cover related topics.**
- **Researchers (in the same collaboration community) might work on similar problems.**
- **People (in the same tele-communication community) might be close friends.**
- **Communities in social settings might explain or predict the spread of contagious diseases.**
- **And many other examples.**

What is a Community

- Graph theory: Communities are those densely connected groups of vertices, with only few connections between groups.
- Sociology: Communities are social groups that entities in the same group share similar properties or connect to each other via certain relations.
- More definitions are available, however, communities are often different for different domains, even for different networks in the same domain. Thus there is no general definition.
- In community mining the community structure found is usually a byproduct of the discovery procedure.

Traditional Community Mining Taxonomy

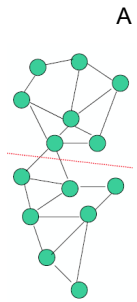


Graph Partitioning Approaches

- There is a long computer science tradition in graph partitioning: believed to be an NP-complete problem.
- Typical Solution: greedily optimize an objective function: the fraction between within-community and between-community edges.
- Iterative Bisection: find the best two-group-cut, then further sub-divide until the required community number is met.

Graph Partitioning Methods

- Graph partitioning algorithms are heavily used to find communities.
- Parameters that are difficult to decide are usually required: size of communities, number of clusters
- Spectral Clustering with benefit functions: **ratio cut** (Hagen et al. 1992), **normalized cut** (Shi et al. 1997), **min-max cut** (Ding et al. 2001)
- Unfortunately, equal-sized communities are usually favoured.



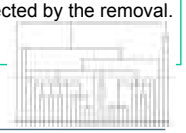
Other Problems

- Require input parameters: number of the partitions, and their sizes
- Such information would never be available for large social networks. They should be determined by the network, not the user.
- Fundamental problem: cut (sum of edge weights between communities) is simply not the right thing to optimize.

Hierarchical Clustering

- Greedily optimize a metric, which evaluates the node centrality or community quality.
- An example metric: edge betweenness, which is the number of edge occurring on the shortest path between other pair of nodes in the network.
- Up-down Algorithm: remove the edge with highest betweenness value in each step.

1. Calculate the betweenness for all edges in the network.
2. Remove the edge with the highest betweenness.
3. Recalculate betweenness measures for all edges affected by the removal.
4. Repeat from step 2 until no edges remain.
5. Cross cut the dendogram of components.



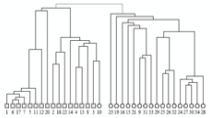
Edge Betweenness: The number of shortest paths between vertex pairs that goes along an edge.

Modularity Q

- Proposed by Newman and Girvan in 2004 as a measure of the qualify of a particular division of the network.
- Q = (*number of edges within communities*) – (*expected number of such edges*)
- Intuition: compare the division to a random network with same nodes and same degrees, but edges are placed randomly.
- Greedily maximizing Q outperformed all other methods, in most cases by an impressive margin, for community detection.

Newman's Algorithm

1. Separate each vertex solely into n community.
2. Calculate ΔQ for all possible community pairs.
3. Merge the pair of the largest increase in Q.
4. Repeat 2 & 3 until all communities merged in one community.
5. Cross cut the dendogram where Q is maximum



Notes:

$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j$

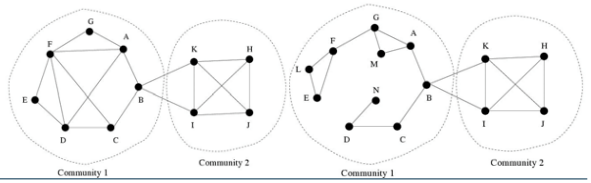
Calculate ΔQ only for pairs that are connected by an edge.

Success of the Modularity Q

- Algorithm: bottom-up agglomerative hierarchical clustering to maximize Q.
- Q has proven to be highly efficient - O((m+n)n), $O(n^2)$ for sparse graphs
- No need of prior knowledge of size of communities or number of communities.
- Q-based methods over-perform other community mining algorithms on many networks, usually with a big margin.
- FastModularity [Clauset, Newman and Moore 2004] – use of Max Heaps and binary tree to provide an efficient O(*md* log *n*) Modularity implementation where *m* is the # of edges, *n* is the number of nodes, and *d* the depth of the dendrogram.

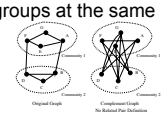
Problem Solved?

- There are three major problems for Q.
  - Q requires information of the entire network.
  - Q has a resolution limit and may fail to identify communities smaller than a certain scale.
  - Q cannot be used to compare community qualities in different networks. (Q = 0.360 for both)



Max-Min Modularity [Chen et al. SDM'09]

- Evaluation Metric: reward for connected pairs and penalty for disconnected ones.
- A "disconnection" can be "unobserved" in many social networks, e.g., biological network, dynamic Facebook.
- *Maximize* the edge number within groups and *minimize* the number of unrelated pairs defined by experts within groups at the same time.
- Use of complement graph



$$Q_{\max_\min} = Q_{\max} - Q_{\min}$$

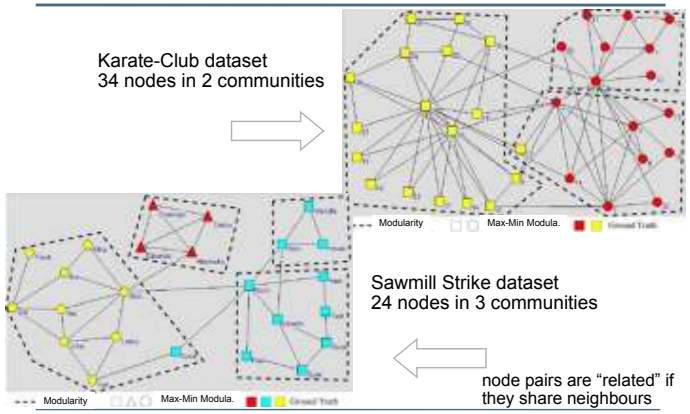
Q_{max} = Modularity Q

$$Q_{\min} = \frac{1}{n(n-1) - 2m - 2|U|} \sum_{xy} [A'_{xy} - P'_{xy}] \phi(C_x, C_y)$$

n is the node number.
U is the related but disconnected pair set defined by domain experts.

Example Results with Max-Min Modularity

Karate-Club dataset
34 nodes in 2 communities



Sawmill Strike dataset
24 nodes in 3 communities

node pairs are “related” if they share neighbours

On Real Networks?

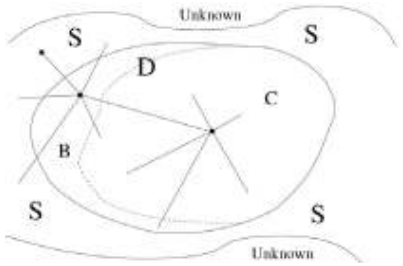
- Most of these approaches require knowledge of the entire network structure, e.g., number of nodes/edges, number of communities in the network. However, this is problematic for networks which are either too large or dynamic, e.g., the WWW.
- The size of the WWW 1 trillion unique URLs. The index size of Google is about 40 billion.
  http://www.techcrunch.com/2008/07/25/googles-misleading-blog-post-on-the-size-of-the-web/
- Facebook has more than 200 million active users
  http://www.facebook.com/press/info.php?statistics
- Vodafone has 289 million customers worldwide
  http://www.vodafone.com/start/media_relations/news/group_press_releases/2009/mobile_internet_experience.html

Local Methods

- A common assumption for the proposed methods is that the complete global network information is always available.
- For some huge networks, e.g., WWW, global information is not always accessible.
- Scenarios: Locate a friend community of a person in Facebook or Find a page cluster of a particular page in the WWW.
- The only available information are nodes that have been visited and their neighbours. All global methods fail.

Typical Problem Definition

- A local community D includes cores (C) nodes and boundary (B) nodes.
- If one new node is merged, its neighbours are added into shell nodes (S).

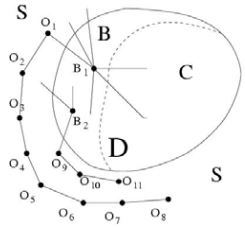


Modularity in Local Network

- Clauset proposed in 2005 the local modularity using the modularity methodology:
  - Measure R, the quality of communities
  - Greedily maximize the R measure to identify communities
- R = within edges of boundary nodes / total edges of boundary nodes
- R measures the sharpness of the boundary nodes. Identify local community by keeping merging until no merge can increase R.

Local Modularity's Problem

- Weakly linked nodes are always merged into the local community.
- In_edge / total < In_edge+1/total+1



- Outliers are merged into the local community one by one.

Measure the Local Community

- Two factors to consider in local community quality:
  - high node relations within the set
  - low relations between set and outside nodes
- R directly represent these two factors by maximizing internal degrees and minimizing external degrees
- The important missing aspect for R is the *connection density*, not the absolute number of connections, that matters in community structure evaluation.

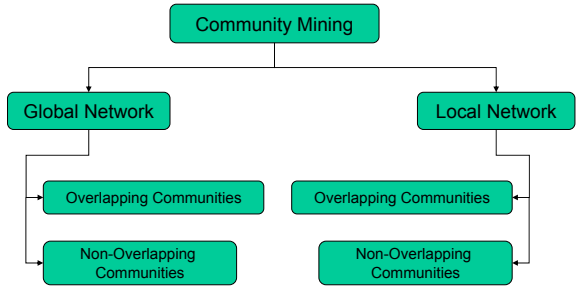
Detecting based on Local Density

- Chen et al. 2009 propose to measure the two factors by maximizing ***average*** internal degree (id) inside the whole community and minimizing ***average*** external degree (ed) of boundary nodes, by maximizing id/ed.
- The "density" idea solves the outlier problem and dramatically increases community detection accuracy:
  - F-measure 0.595 -> F-measure 0.952 (on NCAA Football dataset with ground truth)

Tutorial Structure

- Analyse des Réseaux Sociaux
  - Historique, conceptes et recherche
  - Les applications et les problèmes
- Ordonnancement dans un réseau
- Découverte des communautés
  - Partition de Graphes
  - Approches hiérarchiques
  - Approches globales et approches locales
- Intersection de communautés
- Visualization des réseaux

Community Mining Hierarchy



Global Overlapping Methods

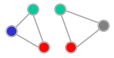
- We usually assume that one node belongs to only one community. However, in real world, it is not the case.
- One person can belong to two or more communities, thus we need to consider overlapping communities.
- Typical approach: find the cluster, then measure the relations of nodes in question to different clusters with arbitrary threshold.

CFinder

- Palla et al. proposed the CFinder system in Nature 2005, using a simple but efficient idea to detect overlaps based on cliques.
- Cliques are completely connected sub-graphs, representing strong communities.
- One node can belong to multiple cliques, which shows community overlaps.

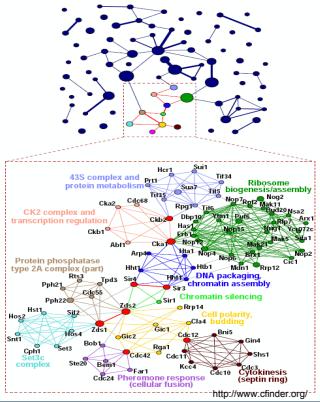
CFinder

- CFinder takes a parameter k, which is the clique size.
- Two k-cliques are adjacent if they share k-1 nodes.



- Given clique size k, merge adjacent k-cliques into one community to identify the network structure.

- Problem: also depends on parameters, k = 3,4,5 usually give reasonable results.

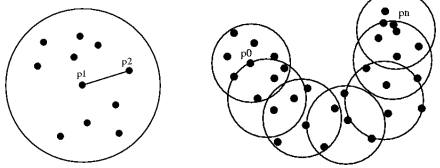


Local Overlapping Methods

- Previous works, using only local information, focus on locating the first local community given a start node.
- Iteratively applying the community identification algorithm based on local modularity may be able to find local-overlapping communities (Chen et.al CASoN 2009)
- A density based approach à la DBScan. Xu et al. proposed SCAN (Xu et al. KDD 2007)

Local Clustering: DBSCAN

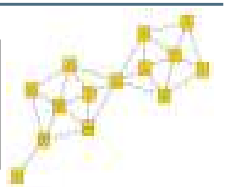
- DBSCAN:



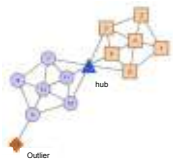
- Two parameters: epsilon and minPts: a cluster node need to have more than minPts neighbours that have a distance score smaller than epsilon

SCAN Social Network Model

- How many clusters?
- What size should they be?
- What is the best partitioning?
- Should some points be segregated?

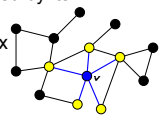


- Individuals in a tight social group, or clique, know many of the same people, regardless of the size of the group.
- Individuals who are hubs know many people in different groups but belong to no single group. Politicians, for example bridge multiple groups.
- Individuals who are outliers reside at the margins of society. Hermits, for example, know few people and belong to no group.



SCAN: Structural Clustering Algorithm for Networks

- Basic idea: the community structure of a node is defined by its neighbourhood.
- Define $\Gamma(v)$ as the immediate neighborhood of a vertex (i.e. the set of people that an individual knows).



- The community relation of two nodes can be represented by their neighbourhood **structural similarity**.
- Similarity score between connected pairs are calculated based on their shared neighbourhood and the neighbourhood sizes.

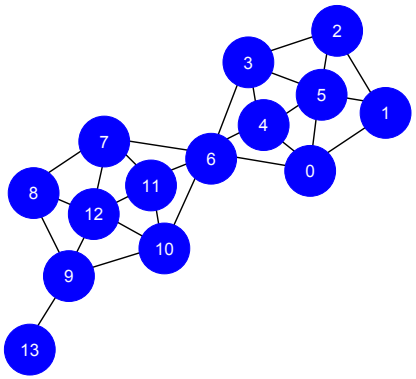
$$\sigma(v,w) = \frac{|\Gamma(v) \bigcap \Gamma(w)|}{\sqrt{|\Gamma(v)||\Gamma(w)|}}$$

SCAN: Structural Clustering Algorithm for Networks

- Give *epsilon*, build *epsilon*-neighbourhood of node A, which contains nodes that have similarity to A greater than *epsilon*.
- Give minimum cluster size *u,* a node is called a core w.r.t. *epsilon* and *u*, if its *epsilon*-neighbourhood contains at least *u* nodes.
- Node A is reachable from another core node if A can be found in the other node's *epsilon*-neighbouhood.
- Similar to DBSCAN clustering, a community is formed by a list of cores that are reachable from any other node in the community via a path.

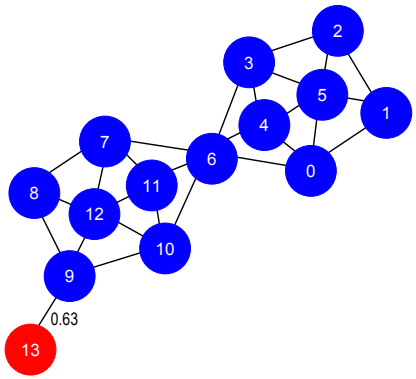
SCAN Algorithm (based on X. Xu)

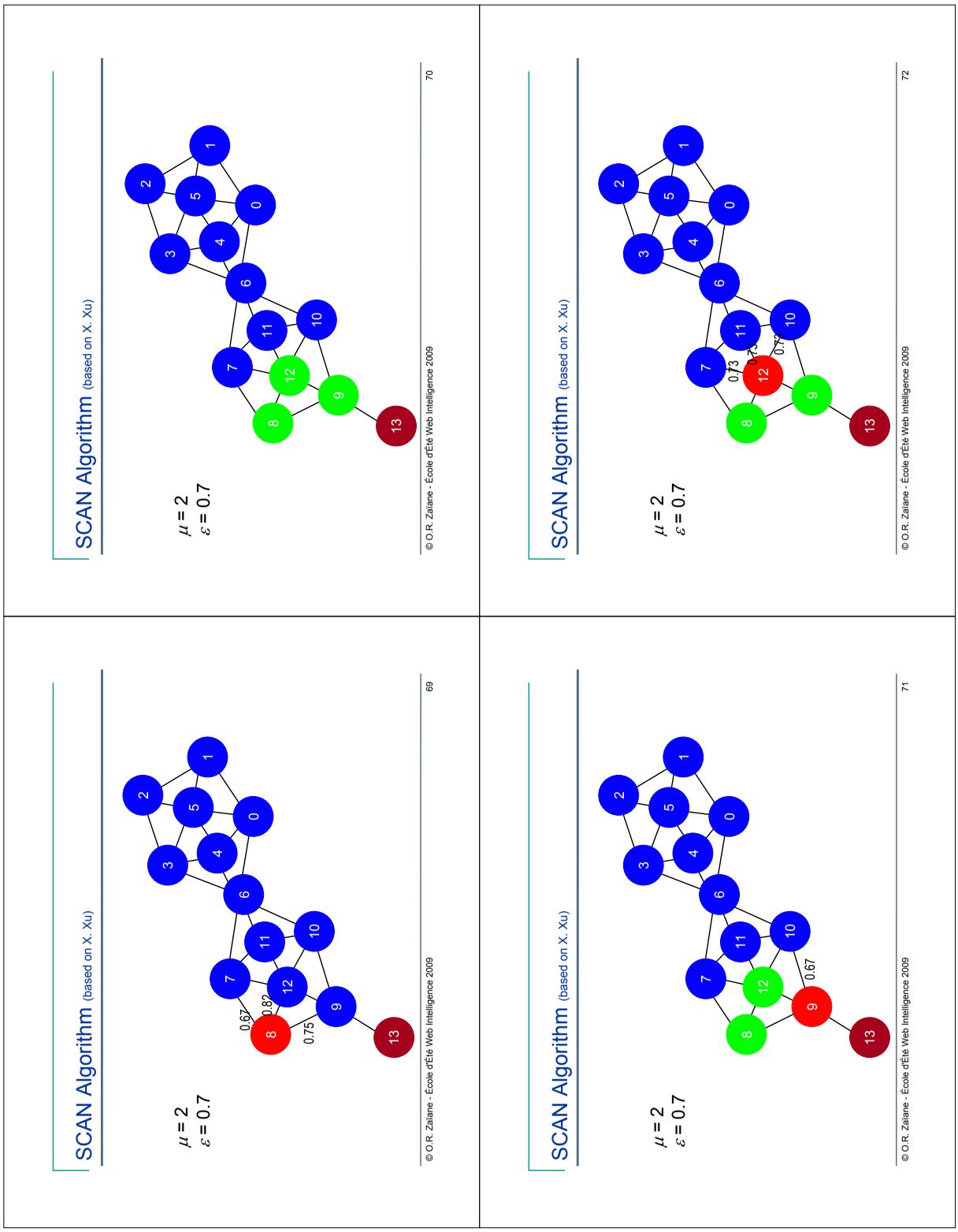
$\mu = 2$
$\varepsilon = 0.7$



SCAN Algorithm (based on X. Xu)

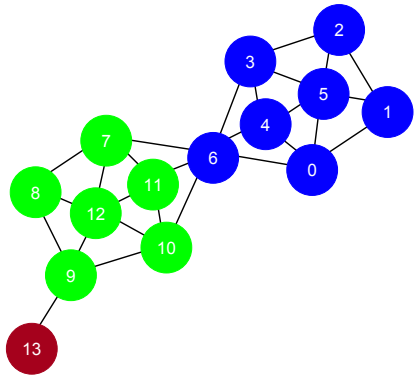
$\mu = 2$
$\varepsilon = 0.7$




SCAN Algorithm (based on X. Xu)
μ = 2
ε = 0.7
0.67
0.82
0.75
© O.R. Zaïane - École d'Été Web Intelligence 2009
69
SCAN Algorithm (based on X. Xu)
μ = 2
ε = 0.7
© O.R. Zaïane - École d'Été Web Intelligence 2009
70
SCAN Algorithm (based on X. Xu)
μ = 2
ε = 0.7
0.67
© O.R. Zaïane - École d'Été Web Intelligence 2009
71
SCAN Algorithm (based on X. Xu)
μ = 2
ε = 0.7
0.73
0.73
0.73
© O.R. Zaïane - École d'Été Web Intelligence 2009
72

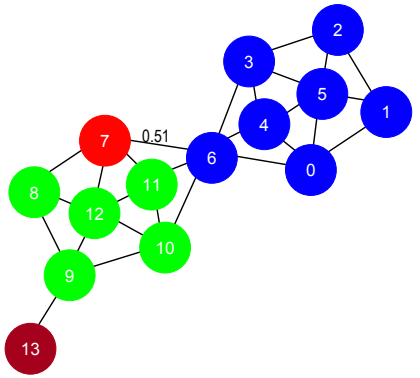
SCAN Algorithm (based on X. Xu)

$\mu = 2$
$\varepsilon = 0.7$



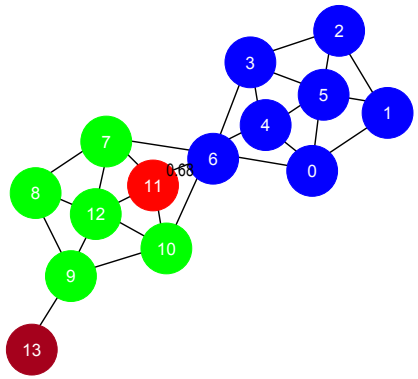
SCAN Algorithm (based on X. Xu)

$\mu = 2$
$\varepsilon = 0.7$



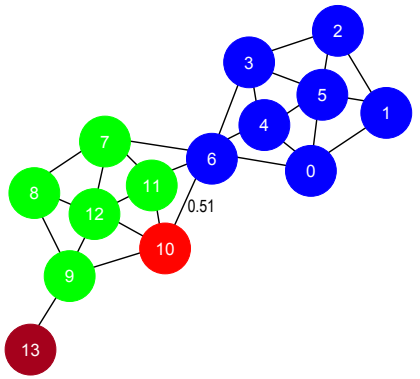
SCAN Algorithm (based on X. Xu)

$\mu = 2$
$\varepsilon = 0.7$



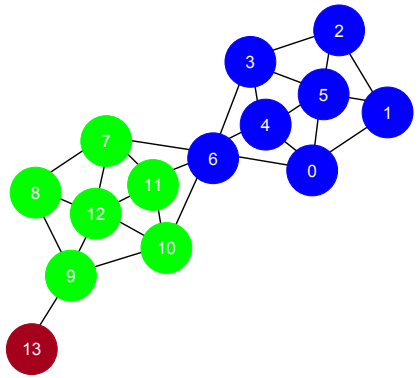
SCAN Algorithm (based on X. Xu)

$\mu = 2$
$\varepsilon = 0.7$



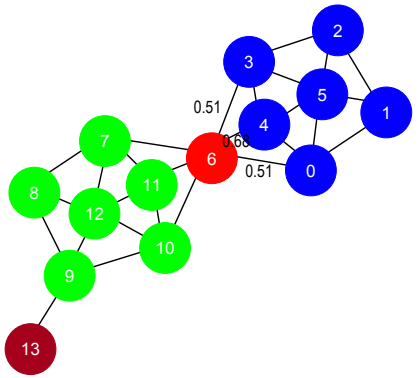
SCAN Algorithm (based on X. Xu)

$\mu = 2$
$\varepsilon = 0.7$



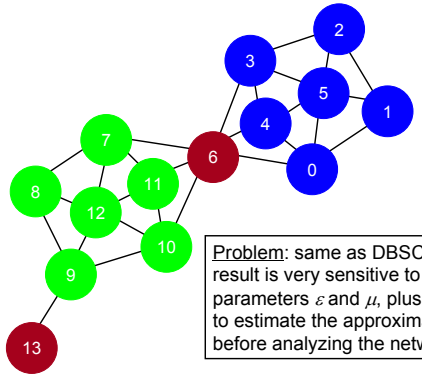
SCAN Algorithm (based on X. Xu)

$\mu = 2$
$\varepsilon = 0.7$



SCAN Algorithm (based on X. Xu)

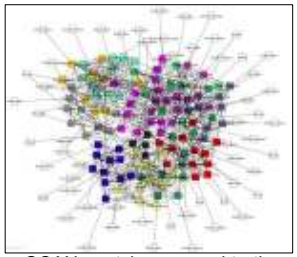
$\mu = 2$
$\varepsilon = 0.7$



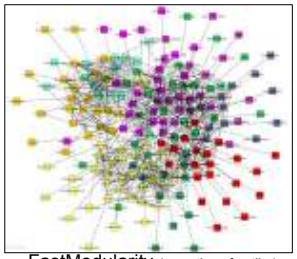
Problem: same as DBSCAN, the result is very sensitive to the two parameters ε and μ, plus it is hard to estimate the approximate value before analyzing the network.

SCAN: Example results

2006 schedule of what schools each NCAA Division 1A team met on an American football field.
Community ~ conference.



SCAN: matches ground truth



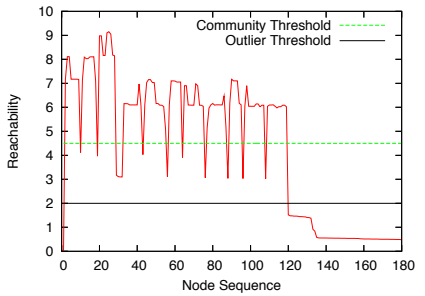
FastModularity (no notion of outlier)

Visual Community Mining

- Chen et al. 2009 proposed a visual data mining approach to detect overlapping communities.
- Given a start node, the approach first generates a sequence of nodes with their highest "*reachability score*" to former nodes in the list.
  - similar to the well-known visual data mining approach OPTICS.
- A 2D visualization is then built to show the community structure, with "mountain" and "valley" curves.

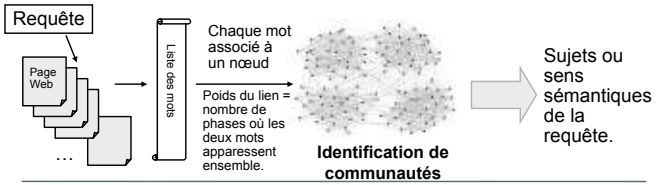
NCAA FootBall Network Visualization

- User needs to give two thresholds after observation to extract groups, hubs and outliers.
- Visual data mining helps user to get good parameters while it is hard for other approaches.



Examples pratiques pour l'extraction de communautés désambiguer les requêtes

- Jaguar: ou
- Michael Jordan: ou
- Différent groupes de mots sont utilisés pour décrire différent conceptes. Quand on cherche sur un moteur de recherche, quelle communauté est pertinente pour notre requête?



Examples pratiques pour l'extraction de communautés Segmentation d'image en vision numérique

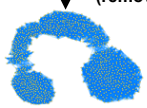
- Each pixel in the image is associated to a node in a network
- Each node has a vector of features associated (RGB, grey level,…)
- Edge weight: Euclidian distance between these feature vectors

e.g. Weight = Grey level difference between every pair of pixel

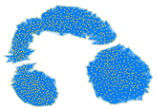
Objects with similar gray levels are identified together

Threshold (remove weak links)

Inverse transformation



Community identification

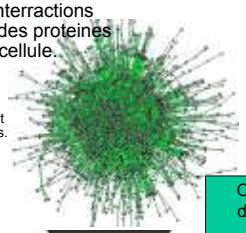


Application of community identification methods for image segmentation, F. Rodrigues, G. Travieso and L. da F. Cota.

Examples pratiques pour l'extraction de communautés proteines complexes

Reseau d'interractions obsereves des proteines dands une cellule.

Ici 6K proteine et 15K interractions.



Identification d'ensembles de proteines responsables de certaines fonctions dans une cellule.

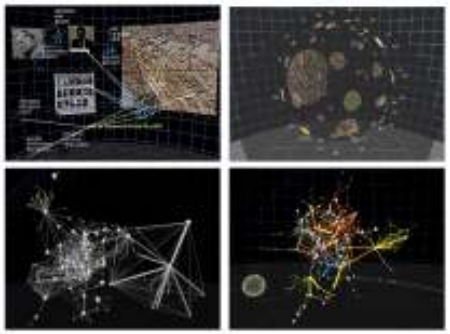
Communautés = Groupes d'ensembles de proteines

Identification de communautés

Tutorial Structure

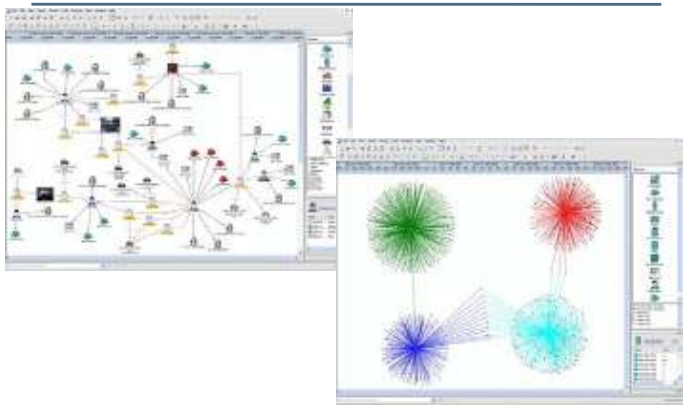
- Analyse des Réseaux Sociaux
  - Historique, conceptes et recherche
  - Les applications et les problèmes
- Ordonnancement dans un réseau
- Découverte des communautés
  - Partition de Graphes
  - Approches hiérarchiques
  - Approches globales et approches locales
- Intersection de communautés
- Visualization des réseaux

StarLight

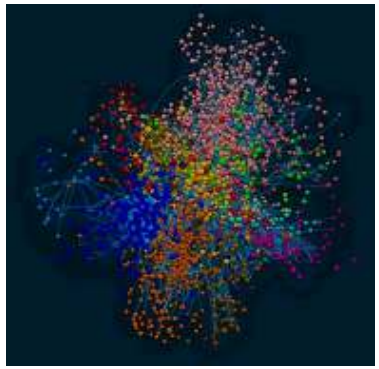


StarLight

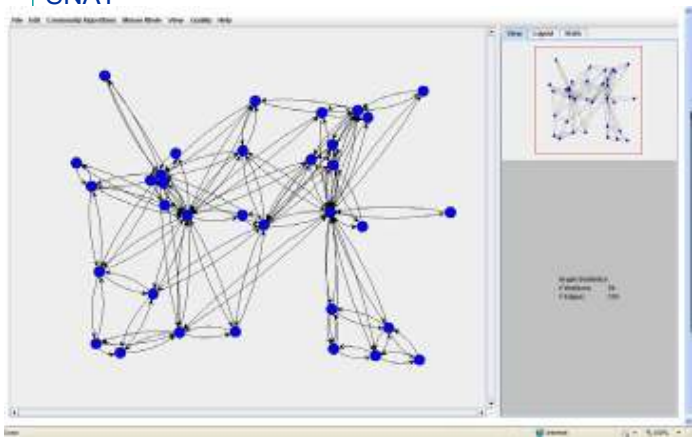
Analysts Notebook

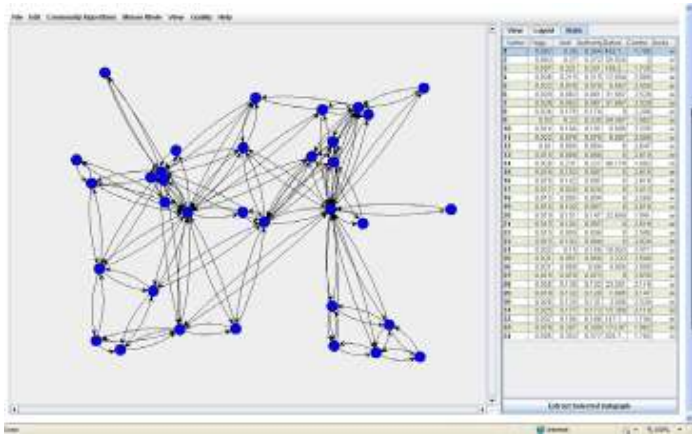


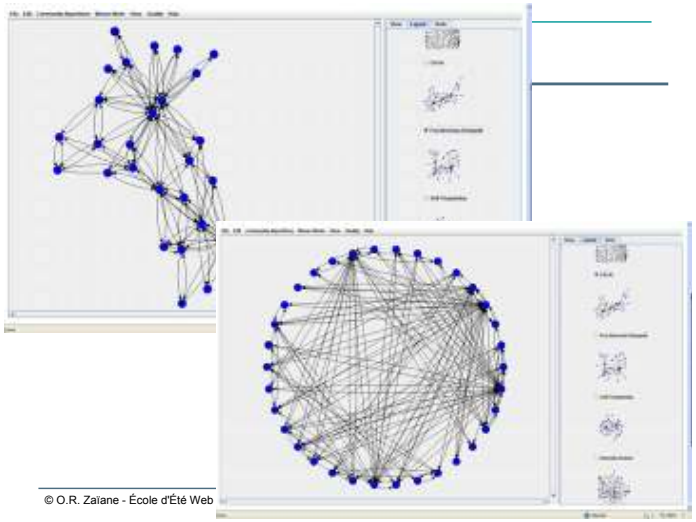
Blogopole

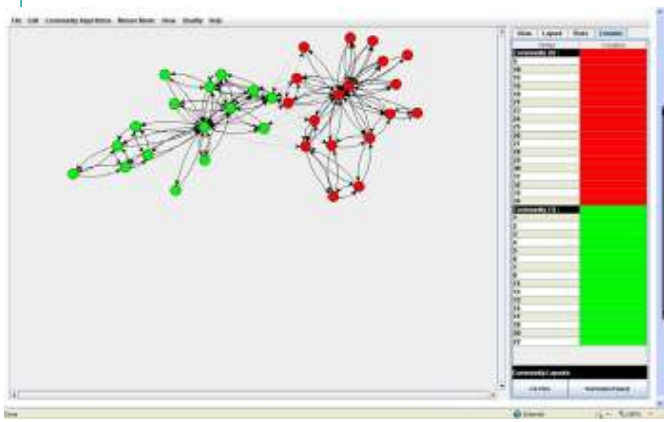


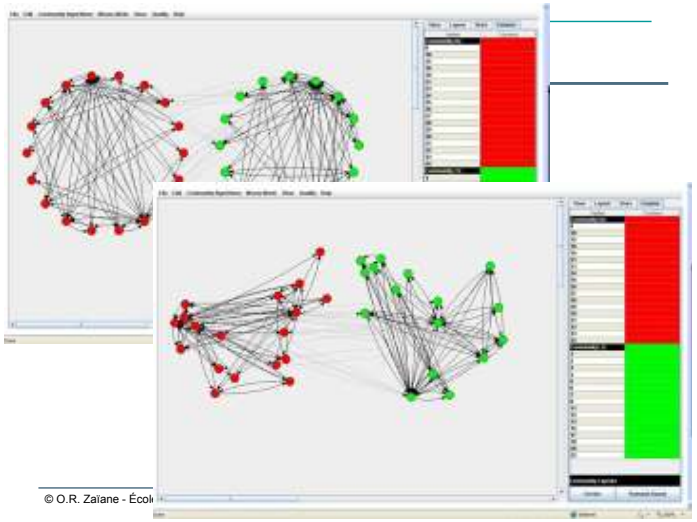
SNAT











References (Block 2) 1/5

- J. Barnes, (1954). Class and Committees in a Norwegian Island Parish. Human Relations, 7: 39-58.
- S.P. Borgatti, P.C. Foster (2003). The Network Paradigm in Organizational Research: A Review and Typology, Journal of Management, 29(6) 991–1013
- U. Brandes. A faster algorithm for betweenness centrality. Journal of Mathematical Sociology, 25:163.177, 2001
- D. Cai, Z. Shao, X. He, X. Yan, and J. Han. Community mining from multi-relational networks. In PKDD, pages 445.452, 2005.
- D. Cai, Z. Shao, X. He, X. Yan, and J. Han. Mining hidden community in heterogeneous social networks. In LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery, pages 58.65, 2005.
- P.K. Chan, M. D. F. Schlag, and J. Y. Zien. Spectral k-way ratio-cut partitioning and clustering. In Proceedings of the 30th International Conference on Design Automation, pages 749.754, 1993.
- J. Chen, O. R. Zaiane and R. Goebel, Detecting Communities in Large Networks by Iterative Local Expansion, International Conference on Computational Aspects of Social Networks (CASoN), Fontainebleau, France, June 24-27, 2009
- J. Chen, O. R. Zaiane and R. Goebel, A Visual Data Mining Approach to Find Overlapping Communities in Networks, In Proceedings of the 2009 International Conference on Advances in Social Networks Analysis and Mining, July 20-22, Athens, Greece.
- J. Chen, O. R. Zaiane and R. Goebel, Local Community Identication in Social Networks, In Proceedings of the 2009 International Conference on Advances in Social Networks Analysis and Mining, July 20-22, Athens, Greece.
- J. Chen, O. R. Zaiane and R. Goebel, Detecting Communities in Social Networks using Max-Min Modularity, In Proceedings of the 2009 SIAM International Conference on Data Mining, Apr 30 - May 2, 2009, Sparks, Nevada.

References (Block 2) 2/5

- A. Clauset. Finding local community structure in networks. Physical Review E, 72:026132, 2005.
- A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. Phys. Rev. E, 70:066111, 2004.
- L. Danon, J. Duch, A. Diaz-Guilera, and A. Arenas. Comparing community structure identification. J. Stat. Mech, page P09008, 2005.
- C. Ding, X. He, H. Zha, M. Gu, and H. D. Simon. A min-max cut algorithm for graph partitioning and data clustering. In ICDM '01: Proceedings of the 2001 IEEE International Conference on Data Mining, pages 107.114, 2001.
- J. Duch and A. Arenas. Community detection in complex networks using extreme optimization. Phys. Rev. E, 72:027104, 2005.
- S. Fortunato and M. Barthelemy. Resolution limit in community detection. PROC.NATL.ACAD.SCI.USA, 104:36, 2007.
- L. Getoor and C. P. Diehl. Link mining: a survey. SIGKDD Exploration Newsletter, 7(2):3.12, 2005.
- Michelle Girvan and M. E. J. Newman. Community structure in social and biological networks. In Proceedings of the National Academy of Science USA, 99:8271-8276, 2002.
- M. Granovetter. The strength of weak ties. American Journal of Sociology, 78:1360. 1380, 1973.
- S. Gregory. An algorithm to find overlapping community structure in networks. In PKDD, pages 91.102, 2007.
- S. Gregory. A fast algorithm to find overlapping communities in networks. In PKDD, pages 408.423, 2008.
- V. Krebs. http://www.orgnet.com/.

References (Block 2) 3/5

- X. Li, B. Liu, and P. S. Yu. Discovering overlapping communities of named entities. In PKDD, 2006.
- X. Li, B. Liu, and P. S. Yu. Mining community structure of named entities from web pages and blogs. In AAAI-CAAW, 2006.
- S. Milgram, (1967). The Small World Problem, Psychology Today, 1(1): 60-67.
- P.R. Monge, N.S. Contractor, (2003). Theories of communication networks. New York: Oxford University Press.
- T. Nepusz, A. Petro´czi, L. Ne´gyessy, and F. Bazso´ . Fuzzy communities and the concept of bridgeness in complex networks. Physical Review E, 77, 016107, 2008.
- M. E. J. Newman http://www-personal.umich.edu/~mejn/networks
- M. E. J. Newman. Detecting community structure in networks. Eur. Phys. J.B, 38:321.330, 2004.
- M. E. J. Newman. Fast algorithm for detecting community structure in networks. Physical Review E, 69, 2004.
- M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. Physical Review E, 74, 2006.
- M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. Physical Review E, 69, 2004.
- A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In NIPS, pages 849.856, 2001.

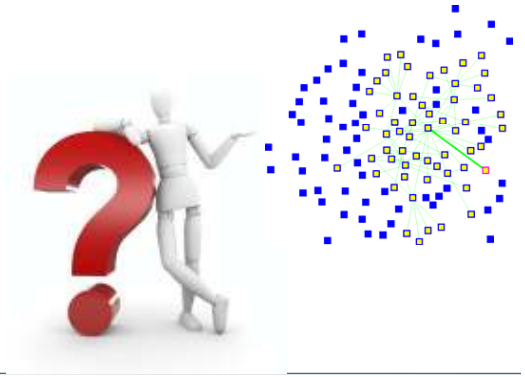
References (Block 2) 4/5

- V. Nicosia, G. Mangioni, V. Carchiolo, and M. Malgeri. Extending modularity definition for directed graphs with overlapping communities, Eprint arXiv:0801.1647 at arxiv.org. 2008.
- Pajek. http://vlado.fmf.uni-lj.si/pub/networks/pajek/.
- G. Palla, A.-L. Barabasi, and T. Vicsek. Quantifying social group evolution. Nature, 446:664.667, 2005.
- G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435:814.818, 2005.
- J. Ruan and W. Zhang. Identifying network communities with a high resolution. Physical Review E, 77:016104, 2008.
- J. Scott, (2000). Social Network Analysis: A handbook. 2nd edition. London: Sage.
- J. Scripps, P.-N. Tan, and A.-H. Esfahanian. Exploration of link structure and community-based node roles in network. In ICDM, 2007.
- J. Shi and J. Malik. Normalized cuts and image segmentation. IEEE. Trans. On Pattern Analysis and Machine Intelligence, 2000.
- J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu. Graphscope: parameter-free mining of large time-evolving graphs. In KDD, pages 687.696, 2007.
- C. Tantipathananandh, T. Berger-Wolf, and D. Kempe. A framework for community identification in dynamic social networks. In KDD, pages 717.726, 2007.

References (Block 2) 5/5

- X.Wang, N. Mohanty, and A. McCallum. Group and topic discovery from relations and their attributes. In NIPS, pages 1449.1456, 2006.
- S. Wasserman, K. Faust, (1994). Social Network Analysis: Methods and Applications, Cambridge University Press.
- D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. Nature, 393:440.442, 1998.
- S. White and P. Smyth. Algorithms for estimating relative importance in networks. In KDD, pages 266.275, 2003.
- X. Xu, N. Yuruk, Z. Feng, and T. A. J. Schweiger. Scan: a structural clustering algorithm for networks. In KDD, pages 824.833, 2007.
- O. R. Zaiane, J. Chen, and R. Goebel. Dbconnect: Mining research community on dblp data. In Joint 9th WEBKDD and 1st SNA-KDD Workshop '07 (WebKDD/SNAKDD'07), 2007.
- H. Zhang, C. L. Giles, H. C. Foley, and J. Yen. Probabilistic community discovery using hierarchical latent Gaussian mixture model. In AAAI, pages 663.668, 2007.
- S. Zhang, R. Wang, and X. Zhang. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. Physica A, 374:483.490, 2007.

Questions



Prédiction de liens dans les réseaux sociaux

R. KANAWATI, C. ROUVEIROL
(LIPN, PARIS 13, FRANCE)

,

Prédiction de liens dans les réseaux sociaux

Approches topologiques

Rushed Kanawati, Céline Rouveirol

A3 - LIPN UMR CNRS 7030
université Paris 13
prénom.nom@lipn.univ-paris13.fr

Ecole d'été Web Intelligence, 10 juillet 2009, Véranne

Plan

Introduction
Problématique de Prédiction de liens
- Définition
- Applications
- Problèmes similaires

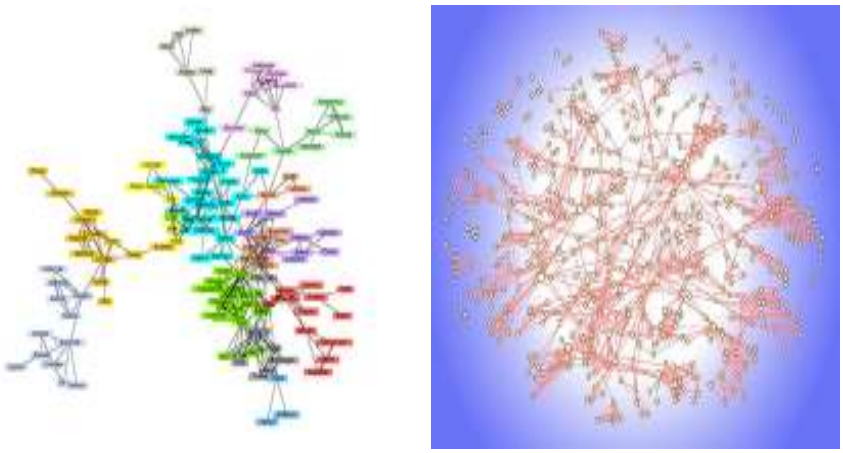
Approches de prédiction de liens
- Classification
- Approches dyadiques / topologiques
  - Apprentissage automatique
- Approches basées sur le contenus des nœuds
- Approches temporelles
- Approches structurelles

Conclusion
Bibliographie

Réseaux sociaux : Notations

- Un réseau social est un graphe $G =< V, E \subseteq V \times V >$:
  - V est l'ensemble de nœudes (i.e. acteurs sociaux)
  - E est l'ensemble de liens *sociaux*.
- Notations :
  - A_G est la matrice d'adjacence de G : $a_{ij} \neq 0$ si les nœuds $(v_i, v_j) \in E$, 0 sinon.
  - $\Gamma(v)$ est l'ensemble de voisins de v. $\Gamma(v) = \{x \in V : (x, v) \in E\}$.
  - Le degré d'un nœud $d(v) =\| \Gamma(v) \|$

Réseaux cibles



source : www.visulacomplexity.com

Réseaux cibles : Caractéristiques

- **Grand Taille** : $|V| > 10^3$.
- **Sans échelle** : La probabilité pour un nœud $v \in V$ d'avoir k voisins est : $P(k) = k^{-\gamma}$

  Beaucoup de nœuds avec peu de connexions et peu de hubs.
- **Coefficient de clustring elevé** :

$$cc(G) = \sum_{v \in V} \frac{2|E \cap (\Gamma(v) \times \Gamma(v))|}{d(v) \times (d(v) - 1)}$$

  La probabailité que deux voisins d'un nœud choisi aléatoirement soitent eux mêmes connetés.
- Réseaux à **relation unique** : tous les liens $e \in E$ sont de même type.
- Réseaux simples : La **matrice A_G est binaire et symetrique** ; $\forall i, j\ a_{ij} \in \{0, 1\}$

Réseaux cibles : Exemples

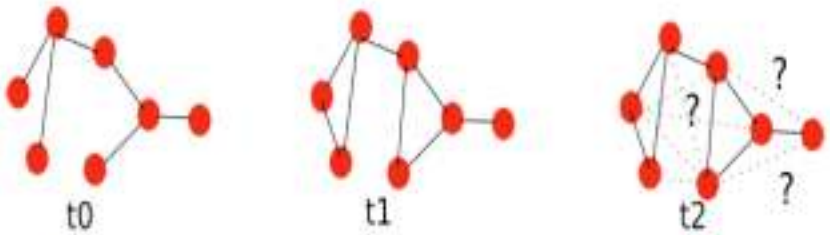
- Réseaux bibliographiques : publications, citation [New01, New04]
- Réseaux sociologiques : Interaction sur forums/blogs [Kan08, MM08], communication [SMS+08], site de rencontres [HEL04], terroriste [HZLG08a].
- Réseaux éthologiques : Comportement des Zèbres [LBW07].
- Réseaux biologiques : Interactions entre protéines [TWAK03], Métabolique [NE08], Interactions entre des gènes [OKK03].
- Réseaux technologiques : Internat, Web
- ...

Réseaux cibles : Modélisation

- Beaucoup de réseaux réels sont des graphes **bipartis** (ex. bibliographique, achats, etc.)
- Graphe biparti : $G = (\top, \bot, E \subseteq \top \times \bot)$. $\top$, $\bot$ sont deux ensembles distincts.
- Projections :
  - $G^n_\top = (V_\top \subseteq \top, E_\top = \{x, y \in \top : |\Gamma_G(x) \cap \Gamma_G(y)| \geq n\}$
  - $G^m_\bot = (V_\bot \subseteq \bot, E_\bot = \{x, y \in \bot : |\Gamma_G(x) \cap \Gamma_G(y)| \geq m\}$
  - n, m : paramètres des projections.
  - La projection augmente, artificiellement, le coefficient de clustering du graphe projeté ! !
- **Dans [IGL04] on montre que les réseaux socials cibles sont de nature bipartis**. Un algorithme est donné pour construire le graphe biparti à partir d'un graphe unimodal.

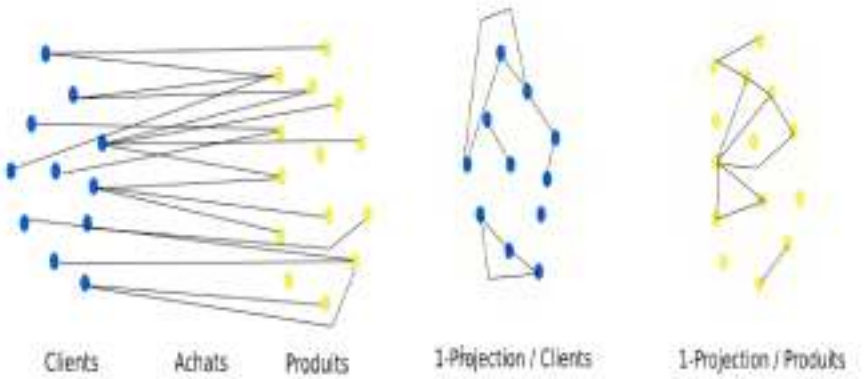
Analyse de réseaux sociaux : les tâches

Prédiction de liens : Définition



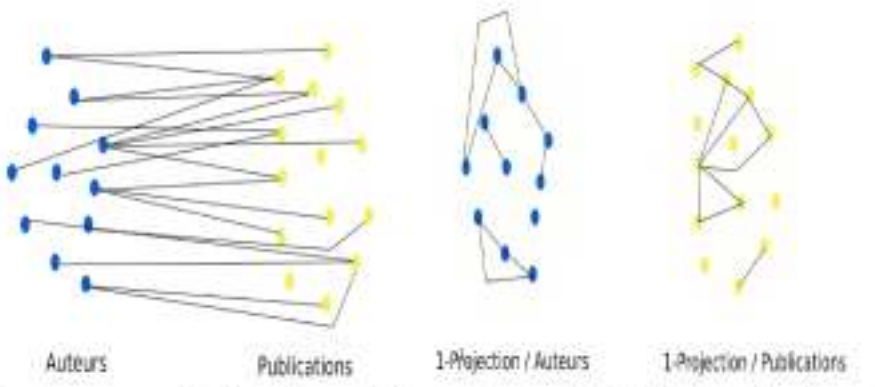
- Définition informelle : Prédire la formation d'un lien entre deux nœuds jamais connectés auparavant.
- Soit $G =< G_1, G_2, \ldots, G_T >$ un réseau temporel de réseaux sociaux, G_i est le graphe du réseau social à l'instant i. La tâche de prédiction de liens consiste à prédire pour chaque couple $v, u \in \bigcap_{i=1}^{T} V_i : \nexists E_j : (u, v) \in E_j$ si $(u, v) \in E_{T+1}$.

Application : Système de recommandation



- Recommandation de produits = Prédiction de liens dans le graphe biparti [HLC05, LHC08, bkR09].
- L'analyse des réseaux sociaux est une solution au problème posé par les matrices des transactions souvent creuses.

Application : Recommandation de collaborations académiques

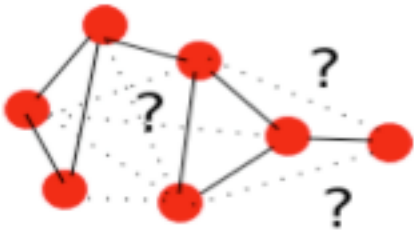


- Recommandation de collaboration = Prédiction de liens dans le graphe projeté / auteurs [PI07, LN05, KcR09].

Autres applications

- Aide à la navigation sur le Web [Zhu03]
- Aide à la réponse des questions sur forums [MM08], *help-desk* [Kan08]
- Etude de propagation des virus informatiques par e-mail [LNH03]
- . . .

Problème similaire : Liens cachés I



- Etant donné un réseau $G_t =< V, E >$ observé à l'instant t la tâche de détection de liens cachés consiste à trouver les liens manquant à l'instant t.
- Origine : information manquante, dissimulation d'information.

Problème similaire : Liens cachés II

- [Coo06] : Comparer les caractéristiques topologiques des nœuds impliqués en liens cachés (LC) / liens en formation (LF).
  - Méthodologie : éliminer aléatoirement des liens dans un réseau temporel de réseaux sociaux.
  - Pour LC le nombre de voisins communs est le double que pour LF.
  - Pour LC le produit des degrés des nœuds impliqués est la moitié que pour LF.
  - LC : distribution des degrés des nœuds plus étalée.
  - Pas de différence significative en ce qui concerene les distances !

Problème similaire : Complétion de liens

- Dans un graphe $G =< V, E >$, on considère un nœud $v \in V$ dont on connait le degré réel $d(v) > d_G(v)$. Le problème consiste à trouver les nœuds $u_i \in V$ auxquels v est susceptible d'être lié [GKK+03].
- Exemple : Un client achète 4 livres sur un site de e-commerce mais le nom d'un des livres achetés est perdu dans la transaction. Quel est ce livre ?
- Alice, Bob et une troisième personne ont fait une réunion. Trouver l'inconnue en fonction des liens connus.
- Dans [GKK+03] on se limite à examiner le cas d'un seul nœud inconnu.

Problème similaire : Liens alarmants

- Etant donné un réseau temporel $G =< G_1, G_2, \ldots, G_T >$, le problème est de classifier les nouveaux liens qui apparaissent dans G_T en deux classes : { normal, anormal}, en fonction de l'évolution du réseau [RJ05]
- Problème plus facile que la prédiction de liens.

Problème similaire : Evolution des réseaux

[HZLG08b] : définition de trois niveaux de granularité pour observer et/ou prédire l'évolution.

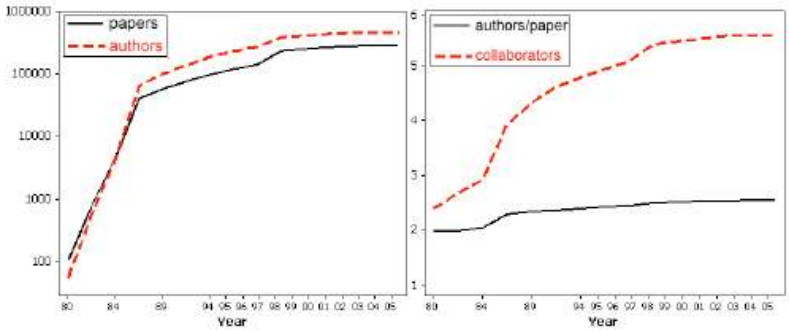
- Niveau réseau
  - Leskovec et al. [LKF05], Evolution des paramètres des réseaux : diamètre décroissant et densification des degrés dans des réseaux de co-citation.
  - Barabasi et al. [BJN+02a], réseaux de co-auteurs, domaine des maths et des neurosciences. Evolution gouvernée par l'attachement préférentiel (lien internes et externes), augmentation du degré moyen et décroissance de la séparation entre les nœuds.
- Niveau communauté : évolution des communautés par taille et des communautés de thèmes

Evolution d'un réseau de co-publications scientifiques

- Niveau "individus" :
  - Définition d'un modèle de poisson stochastique pour modéliser le nombre de collaborations dans le temps, puis d'un arbre d'optimisation pour optimiser le modèle.
  - Pour une collaboration existante au temps t $e_{i,j}(t)$, on extrait un sous-graphe $G(e_{i,j}(t))$ de voisinage de $e_{i,j}(t)$: les auteurs v_i, v_j, leurs voisins immédiats (co-auteurs) et les arcs associés.
  - Un vecteur d'attributs $a_{i,j} = (a_1, \ldots, a_p)$ est calulé à partir de ce voisinage
  - Pour un arc spécifique $e_{i,j}(t)$, le taux de collaboration $\lambda(e_{i,j}(t)) = f(a_{i,j}(t))$, apprentissage de $f(a_{i,j}(t)$ par arbre d'optimisation, technique inspirée des arbres de régression [BFOS84]

Expérimentation I

- Réseaux de co-citations en Informatique (Citeseer)
- 283 174 auteurs uniques et 451 305 papiers, publiés entre 1980 et 2005



Expérimentation II

- Tâche d'apprentissage : étant donné une paire d'auteurs v_i et v_j avec une collaboration existante $e_{i,j}$, quelle est la probabilité qu'ils collaborent k fois dans le prochain intervalle de temps Δ.

Method	Correlation Coefficient (r)	Root Mean Squared Error (RMSE)
Past Collaboration	0.227	4.74
SVR	0.673	1.53
SPOT	**0.782**	**1.42**

Approches de prédiction de liens

Trois critères de classification :

- **Approche : Dyadiques / Structurelles.**
  - Dyadique : Evaluer un *score* d'un lien entre deux nœuds v_i, v_j
  - Structurelle : Prédire l'évolution de sous-graphes (prédiction de plusieurs liens en même temps) [LBW07]
- **Type d'attributs : topologiques / caractéristiques des nœuds.**
  - Approche topologique : utiliser seulement le graphe du réseau.
  - L'emploi des approches fondées sur l'analyse du contenu des nœuds nécessite une expertise dans le domaine de l'application.
- **Prise en compte du temps : Oui / non.**

Approches dyadiques topologiques

- score d'un lien (u, v) est calculé par une fonction de *similarité* topologique.
- Deux familles de mesures de similarités topologiques :
  - Mesures basées sur le voisinage des nœuds.
  - Mesures basées sur les distances entre les nœuds.

Mesures basées sur le voisinage

- Voisins communs.
- Jaccard.
- Attachement préférentiel
- Adamic Adar
- ...

Voisins communs : VC

- Dans un graphe unimodal : $VC(x,y) = \| \Gamma(x) \cap \Gamma(y) \|$
- Dans un graphe biparti G :
  - $VC_{\top}(x,y) = \| \Gamma_{G^n_{\top}}(x) \cap \Gamma_G(y) \|$ où $x \in \top, y \in \bot$
  - Exemple : Dans un graphe d'achats où $\top$ (resp. $\bot$) est l'ensemble de clients (resp. produits) : nombre de clients similaires qui ont acheté le même produit y.
  - $VC_{\bot}(x,y) = \| \Gamma_{G^m_{\bot}}(y) \cap \Gamma_G(x) \|$ où $x \in \top, y \in \bot$
  - Le nombre de produits similaires à y achetés par le client x

Jaccard : JAC

- Dans un graphe unimodal : $JAC(x,y) = \frac{\|\Gamma(x) \cap \Gamma(y)\|}{\|\Gamma(x) \cup \Gamma(y)\|}$
- Dans un graphe biparti G :
  - $JAC_{\top}(x,y) = \frac{VC_{\top}(x,y)}{\|\Gamma^{n}_{\top}(x)\|}$
  - $JAC_{\bot}(x,y) = \frac{VC_{\bot}(x,y)}{\|\Gamma^{m}_{\bot}(y)\|}$

Attachement preférentiel : AP

- $AP(x,y) = \| \Gamma(x) \| \times \| \Gamma(y) \|$
- intuition : deux nœuds qui ont beaucoup de relations ont tendance à être liés.
- *Ex. Une fan de shopping achète aussi des produits best-seller.*
- Modèle de construction de réseaux qui génère des réseaux sans échelle [BJN+02b]

Adamic Adar : AA I

- Mesure proposée dans [AA03].
- Principe : Similarité entre nœuds = somme pondéré des *caractéristiques communes* des deux nœuds.
- Augmenter le poids des caractéristiques faibelement partagées dans le voisinage des deux nœuds.
- Exemple : caractéristique commune = nombre de voisins communs.
  - Dans un graphe unimodal : $AA(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{log(\|\Gamma(z)\|)}$
  - Un voisin commun lié à 2 nœuds a le poids $\frac{1}{log(2)} = 1.442$
  - Un voisin commun lié à 4 nœuds a le poids $\frac{1}{log(4)} = 0.721$

Adamic Adar : AA II

- Dans un graphe biparti :
- $AA_{\top}(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma_{\perp}(y)} \frac{1}{log(\|\Gamma(z)\|)}$
- $AA_{\perp}(x,y) = \sum_{z \in \Gamma(y) \cap \Gamma_{\top}(x)} \frac{1}{log(\|\Gamma(z)\|)}$

Mesures basées sur la distance I

- Chemin le plus court : $Sp(x, y)$
- Katz [Kat53] :
  - $Katz(x, y) = \sum_{l=1}^{\infty} \beta^l |paths(x, y)^l|$
  - $|paths(x, y)^l$ est le nombre de chemins de longueur l qui relient x, y.
  - Rappel : $A^l(i, j) = paths(v_i, v_j)^l$
  - Calcul : $(I - \beta A_G)^{-1} - I$ où I est la matrice identité et A_G la matrice d'adjacence.
  - Très couteux pour des grands réseaux. Appliqué à un composant connexe.
- Mesure basée sur les marches alétoires [FPRS07] :
  - Hit time.
  - temps de commutation .

Mesures basées sur la distance II

- MFA : Matrix-forest-based algorithm.
- . . .

Prédiction de lien sur la base d'attributs simples

- [LK07] : *Prédicteurs de lien* = un unique attribut *topologique*
- Tous les exemples (paires de nœuds) sont ordonnés, par ex., en fonction du plus court chemin les reliant
- Autres attributs topologiques : graph distance, voisins communs, coefficient de Jaccard, Adamic Adar, preferential attachement, Katz, hitting time, rooted pageRank, SimRank

Prédiction de lien sur la base d'attributs simples

- Réseau unimodal de co-auteurs : nœud = auteur, lien = un article co-signé. Pas de poids, pas de gestion du temps.
- Construction des exemples
  - *Graphe d'apprentissage* : les auteurs ayant co-signé pendant une période T_1
  - *Graphe d'étiquetage* : les auteurs ayant co-signa pendant une période T_2 postérieure à T_1
  - Exemples appartenant à une même **composante connexe**
- Tâche : prédire l'apparition, dans le graphe d'étiquetage, d'un lien entre deux nœuds non connectés dans le graphe d'apprentissage

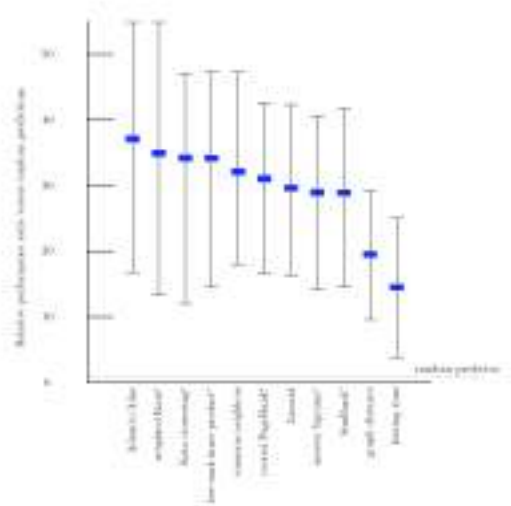
Expérimentations

- 5 réseaux de co-auteurs extraits de la base `ArXiv`

	training period			Core		
	authors	papers	edges	authors	$\|E_{old}\|$	$\|E_{new}\|$
astro-ph	5343	5816	41852	1561	6178	5751
cond-mat	5469	6700	19881	1253	1899	1150
gr-qc	2122	3287	5724	486	519	400
hep-ph	5414	10254	17806	1790	6654	3294
hep-th	5241	9498	15842	1438	2311	1576

- Core : l'ensemble des auteurs qui ont écrits au moins 3 papiers pendant la période d'apprentissage et 3 papiers pendant la période d'étiquetage
- Prédicteurs : (1) Katz pondéré, (2) Katz classification, (3) low-rank inner product, (4) rooted Pagerank, (5) bigrammes inconnus, voisins communs non pondérés (6) SimRank,

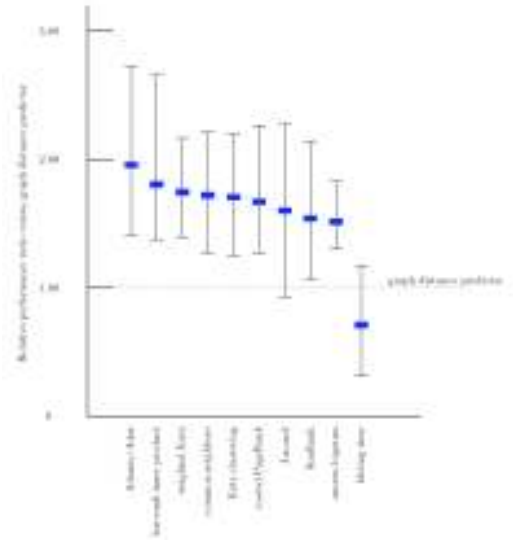
Résultats (1)



Comparaison / prédicteur aléatoire

Toutes les méthodes prédisent mieux que le prédicteur aléatoire → la topologie du réseau contient bien des informations permettant de prédire l'occurrence d'un lien

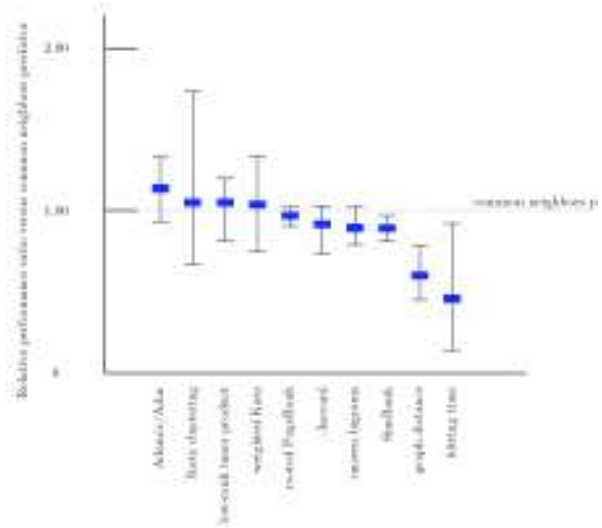
Résultats (2)



Qualité du prédicteur plus court chemin

Pas de bons résultats à cause de la propriété *petit monde* (attribut peu discriminant)

Résultats (3)



Prédicteur voisins communs

Un prédicteur simple et particulièrement efficace

Résultats (4)

	Adamic/Adar	Katz clustering	common neighbors	hitting time	Jaccard's coefficient	weighted Katz	low-rank inner product	rooted Pagerank	SimRank	unseen bigrams
Adamic/Adar	1150	638	520	193	442	1011	905	528	372	486
Katz clustering		1150	411	182	285	630	623	347	245	389
common neighbors			1150	135	506	494	467	305	332	489
hitting time				1150	87	191	192	247	130	156
Jaccard's coefficient					1150	414	382	504	845	458
weighted Katz						1150	1013	488	344	474
low-rank inner product							1150	453	320	448
rooted Pagerank								1150	678	461
SimRank									1150	423
unseen bigrams										1150

- Corrélation importante des attributs en terme de nombre de prédictions communes

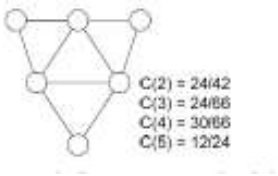
Résultats (5)

	Adamic/Adar	Katz clustering	common neighbors	hitting time	Jaccard's coefficient	weighted Katz	low-rank inner product	rooted Pagerank	SimRank	unseen bigrams
Adamic/Adar	92	65	53	22	43	87	72	44	36	49
Katz clustering		78	41	20	29	66	60	31	22	37
common neighbors			69	13	43	52	43	27	26	40
hitting time				40	8	22	19	17	9	15
Jaccard's coefficient					71	41	32	39	51	43
weighted Katz						92	75	44	32	51
low-rank inner product							79	39	26	46
rooted Pagerank								69	48	39
SimRank									66	34
unseen bigrams										68

- Corrélation importante des attributs en terme de nombre de prédictions correctes communes

Coefficient de clustering pour prédire les liens

- [Hua06] : utilisation du coefficient de clustering généralisé. La probabilité d'occurrence d'un lien donné est évaluée en fonction du nombre de cycles (de tailles différentes) qui vont être formés en ajoutant ce lien.



$$\Pr((i,j) \in E \mid |P_{ijk}| = m) = c_k^m/(c_k^m+(1-c_k)^m),\ k > 1$$

$$P_{m_2,\ldots,m_k} = \Pr((i,j) \in E \parallel P_{ij2} \mid = m_2,\ldots,\mid P_{ijk} \mid = m_k) =$$

$$c_1 {c_2}^{m_2} \ldots {c_k}^{m_k} / (c_1 {c_2}^{m_2} \ldots {c_k}^{m_k} + (1-c_1)(1-c_2)^{m_2} \ldots (1-c_k)^{m_k})$$

$$f(c_1,c_2,\ldots,c_k) = \sum_i \#(G_i) \Pr(G_i) \quad \Pr((1,k+1) \in E \mid G_i)\ .$$

Application

- Evaluation : corpus d'email Enron [1]. 40 489 emails echanges entre 151 employes d'Enron.
- Graphe d'apprentissage G
  - nœud : un employé
  - lien (e_1, e_2) : au moins un mail a été échangé entre e_1 et e_2 pendant la période $[t-3..t-1]$
- Tâche : prévoir l'occurrence d'un lien entre deux nœuds non connectés dans G au mois t

1. Données disponibles à ftp ://ftp.isi.edu/sims/philpot/data/enron-mysqldump.sql.gz

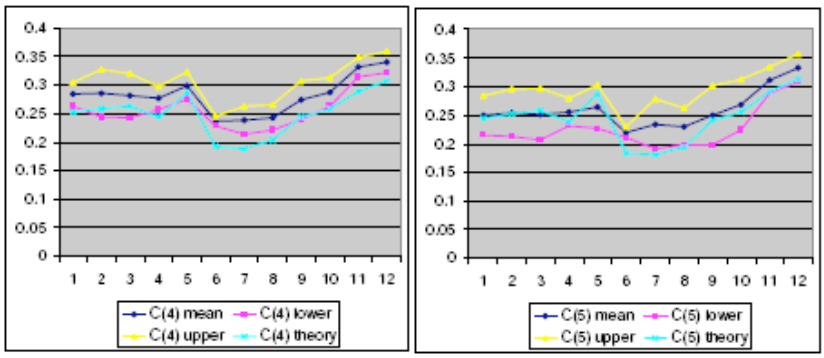
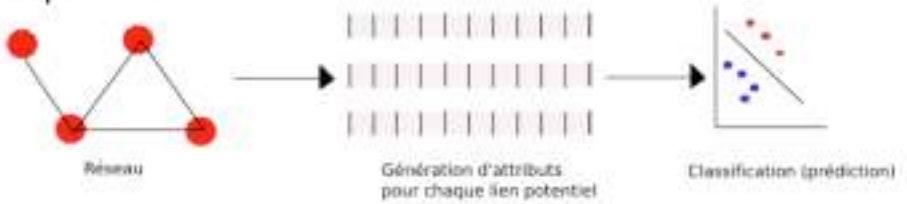


Table 4. Number of new links and AUC measures for the cycle formation and benchmark link prediction algorithms

Month (t)	New Links in G_t	$CF(3)$	PA	SA	GM
1	46	**0.76764**	0.52091	0.70977	0.67237
2	46	**0.76748**	0.48449	0.67891	0.69189
3	46	**0.81385**	0.44235	0.72598	0.73472
4	62	**0.82106**	0.45057	0.72143	0.75518
5	95	**0.73883**	0.45141	0.63052	0.69805
6	72	**0.71737**	0.41585	0.67084	0.70634
7	88	**0.73977**	0.48104	0.70304	0.69231
8	168	**0.74056**	0.60122	0.69806	0.65201
9	90	**0.72065**	0.46032	0.69501	0.70263
10	213	0.74384	0.48877	0.69840	**0.74500**
11	123	0.74860	0.47632	0.72477	**0.75116**
12	56	0.70018	0.44078	0.69573	**0.76273**

Combinaison de prédicteurs

Utilisation des techniques d'apprentissage automatique supervisé.



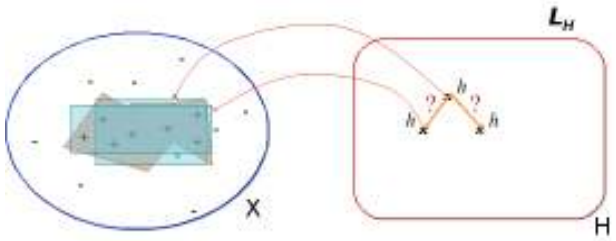
Apprentissage supervisé

A partir de :

- l'échantillon d'apprentissage $S = (x_i, y_i)_{1,m}$
- on cherche une loi de dépendance sous-jacente, par exemple une fonction h aussi proche possible de f (fonction cible) tq : $y_i = f(x_i)$
- ou bien une distribution de probabilités $P(x_i, y_i)$

afin de prédire l'avenir. Si f est une fonction binaire (booléenne), on parle d'apprentissage de concept.

Apprentissage supervisé



- Quel critère inductif ?
- Qu'est-ce qu'une hypothèse optimale étant donné l'échantillon d'apprentissage ?
- Quelle méthode d'exploration de H ?

Evaluation empirique

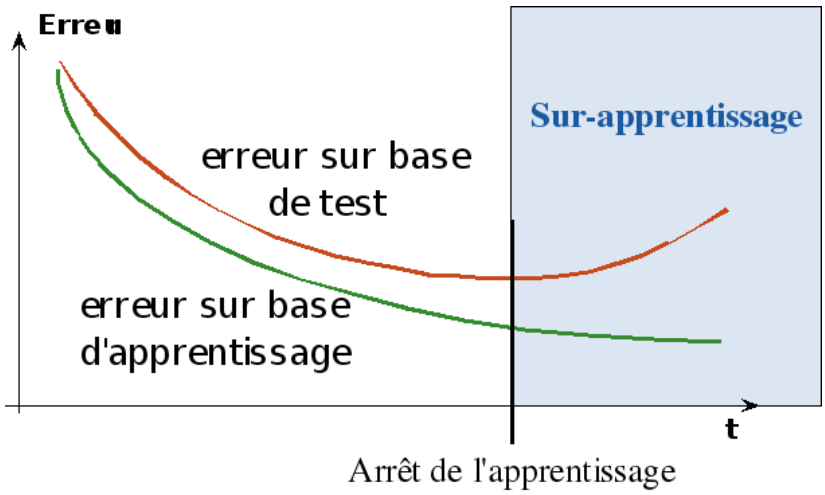
- Erreur réelle :

$$e_D = \int_D |y - f(x, \theta)| \; p(x, y) \; dx, y$$

- Erreur estimée :

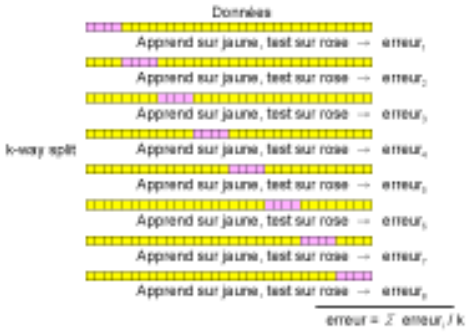
$$\hat{e}_S = \frac{1}{m} \sum_{\langle x,y \rangle \in S} |y - f(x, \theta)|$$

Problème de sur-apprentissage



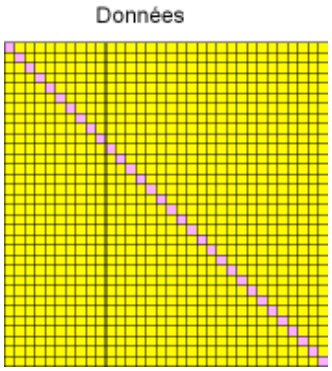
10 validation croisée

- Si peu de données disponibles
- Avantages : simplicité, généralité, temps de calcul raisonnable, robustesse, etc.
- Choix du nombre de plis : compromis entre le biais (le critère de validation croisée "n-fold" surestime d'autant moins l'erreur de prédiction que n est grand), la variabilité du critère, et le temps de calcul.



Validation par `leave one out`

- Si très peu de données disponibles
- Faible biais, haute variance
- Tendance à sous-estimer l'erreur si les données ne sont pas vraiment i.i.d.

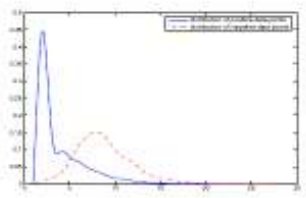


Apprentissage automatique pour la prédiction de liens

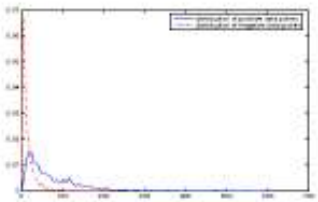
- [HCSZ06] : prédiction de liens → problème de classification binaire
  - exemple + : apparition des liens dans le graphe G_{i+1}
  - exemple - : non apparition des liens dans le graphe G_{i+1}
- Utilisation des informations topologiques dans le graphe des co-auteurs + information sur les nœuds, si disponible
- (Sous-)exploitation d'un graphe biparti auteurs - mots-clés : un lien entre un auteur et un mot-clé si mot-clé utilisé dans au moins un article, deux mots-clés qui apparaissent dans un même abstract sont connectés.
- Les autres attributs n'apportent pas d'information supplémentaires pour la discrimination

Application aux réseaux de co-publications (1)

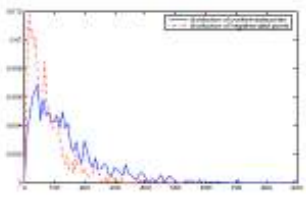
DBLP (1990-2000 : apprentissage, 2001-2004 : étiquetage) et BIOBASE (1998-2001 : apprentissage, 2002 : étiquetage)



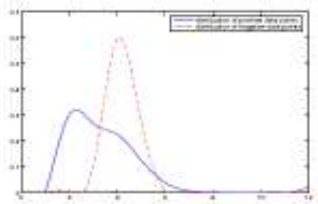
(a) Shortest Distance



(b) Sum of Neighbors count



(c) Sum of neighbors count



(d) Shortest distance

Application aux réseaux de co-publications (2)

Attribute name	Information gain	Gain Ratio	Chi-Square Attribute Eval.	SVM feature evaluator	Avg. Rank
Sum of Papers	4	4	4	2	4
Sum of Neighbors	3	3	3	4	3
Shortest distance	1	1	1	1	1
Second shortest distance	2	2	2	3	2

Attributs utilisés sur la base DBLP

Classification model	Accuracy	Precision	Recall	F-value	Squared Error
Decision Tree	82.56	87.70	79.5	83.40	0.3569
SVM(Linear Kernel)	83.04	85.88	82.92	84.37	0.1818
SVM(RBF Kernel)	83.18	87.66	80.93	84.16	0.1760
K_Nearest Neighbors	82.42	85.10	82.52	83.79	0.2354
Multilayer Perceptron	82.73	87.70	80.20	83.70	0.3481
RBF Network	78.49	78.90	83.40	81.10	0.4041
Naive Bayes	81.24	87.60	76.90	81.90	0.4073
Bagging	82.13	86.70	80.00	83.22	0.3509

Résultats sur la base DBLP

Application aux réseaux de co-publications (3)

Attribute name	Information gain	Gain Ratio	Chi-Square Attribute Eval.	SVM feature evaluator	Avg. Rank
Sum of Papers	3	4	3	4	3
Sum of Neighbors	1	3	1	2	2
Sum of KW count	6	6	6	3	5
Sum of Classification count	5	5	5	6	5
KW match count	2	1	2	1	1
Sum of log of Sec. Neighbor. count	7	7	7	8	7
Shortest distance	4	2	4	5	4
Clustering Index	9	9	9	7	8
Shortest dist. in KW-Author graph	8	8	8	9	8

Attributs utilisés sur la base BIOBASE

Classification model	Accuracy	Precision	Recall	F-value	Squared Error
Decision Tree	90.01	91.60	89.10	90.40	0.1306
SVM(Linear Kernel)	87.78	92.80	83.18	86.82	0.1221
SVM(RBF Kernel)	90.56	92.43	88.66	90.51	0.0945
K_Nearest Neighbors	88.17	92.26	83.63	87.73	0.1826
Multilayer Perceptron	89.78	93.00	87.10	90.00	0.1387
RBF Network	83.31	94.90	72.10	81.90	0.2542
Naive Bayes	83.32	95.10	71.90	81.90	0.1665
Bagging	90.87	92.5	90.00	91.23	0.1288

Résultats sur la base BIOBASE

Apprentissage supervisé : Bilan

- Avantage : Possibilité de combiner les attributs (topologiques) afin d'obtenir un meilleur taux de prédiction
- Inconvènient : problème difficile d'apprentissage
- Classes non équilibrées : Taux d'exemple positifs $\sim 0.1\%$ des exemples négatifs.
- Exemples non i.i.d.
- Solutions à envisager : échantillonnallge , apprentissage semi-supervisé ?

Approches basées sur l'analyse des attributs des nœuds

- Principe : Apprendre la dependence entre l'existence d'un lien et les attributs des nœuds reliés par le lien.
- Modèles probabilistes [TWAK03] :
  - Réseaux Bayésiens
  - Réseaux de Markov Relationels
  - Apprentissage statistique relationnel
  - . . .
- Nécessite d'avoir accès aux attributs des nœuds.
- Domaine-dependant.

Techniques hybrides I

Principe : Combiner des approches topologiques et des approches basées sur le contenu des nœuds.

Exemple : Alignement de matrices [STCE08]

- soit G un réseau décrit par la matrice d'adjacence A_G
- Soit X est la matrice (de dimension $n \times d$) des attributs des nœuds. n est le nombre de nœuds et d est le nombre d'attributs.
- XX^T est une matrice $n \times n$ de similarités entre les attributs des nœuds.
- Dans le cas idéal on cherche à avoir : $\forall i, jA_g[i,j] = XX^T[i,j]$ (après normalisation)

Techniques hybrides II

- Approche : Trouver W qui minimise : $\| A_G - XWX^T \|^2$
- Le score d'un lien est donn par $score(x,y) = xWy^T$
- Expérimentations : tester sur DBLP, TakingItGlobal.com, WebKB
- Gain en Précision de ~ 0.2 avec l'utilisation de W

Approches temporelles

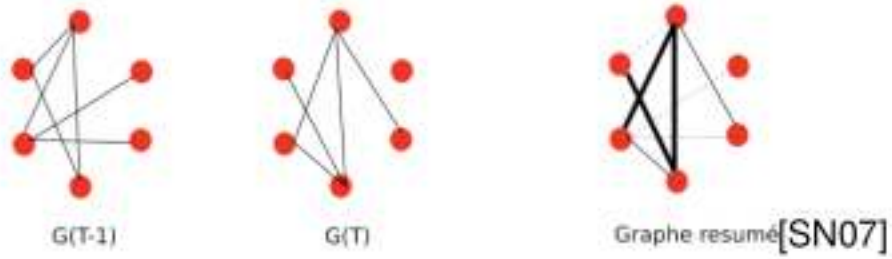
Deux grandes approches :

- Utilisation d'attributs temporels
- Résumé des graphes

Indicateurs temporels

- Indicateurs *financiers* [Coo06]
- *profit* d'un indicateur topologique sur une période de temps.
- Exemple : $\frac{D_{tf}(v)-D_{ti}(v)}{D_{tf}(v)}$
- Moyen temporel : $\sum_{t=ti}^{t=tf} \frac{D_t(v)}{tf-ti}$
- date du dernier lien
- ...

Resumé des graphes



[SN07]

Approches structurelles

- Caractérisation des liens par des sous-graphes fréquents dans le réseaux temporels.
- Apprentissage de règle de réecriture des graphes [LBW07]

Conlusion I

- Approches topologiques viables pour la prédiction de liens.
- Beaucoup de travaux sur des graphes statiques.
- Approches topologiques généralistes à combiner avec des approches basées sur l'exploration des caractéristiques des nœuds.
- Pas de benchmark pour faciliter la comparaison d'approches.
- Axes de recherche :
  - Analyse des réseaux bipartis, valués, et temporels
  - Exploration simultanée de réseaux hétérogènes, approche de graphe mining.

Conlusion II

  - Approches structurelles : Fouille parallèle des grands graphs.
- . . . .

Bibliographie I

Lada ADAMIC et Eytan ADAR :
Friends and neighbors on the web.
Social Networks, 25(3):211–230, july 2003.

L. BREIMAN, J. FRIEDMAN, R. OHLSEN et C. STONE :
Classification and regression trees.
Wadsworth International Group, Belmont, CA, 1984.

A. L. BARABASI, H. JEONG, Z. NEDA, E. RAVASZ, A. SCHUBERT et T. VICSEK :
Evolution of the social network of scientific collaborations.
PHYSICA A, 311:3, 2002.

Albert-LÃ¡szlÃ³ *Barabsi*, *H. Jeong*, *Z. N©da*, *E. Ravasz*, *A. SchubertetT. Vicsek* :
Evolution of the social network of scientific collaboration.
Physica A, 311(3-4):590–614, 2002.

Nesserine BENCHETTARA, Rushed KANAWATI et CÃ©line ROUVEIROL :
Calcul de recommendation par prÃ©diction de liens dans un graphe biparti.
In Actes de l'ateiler sur l'apprentissage et graphes pour les systÃ¨mes complexes (plate-forme AFIA 2009), Hammemt, Tunisie, Mai 2009.

Bibliographie II

Richard J.E. COOKE :
Link prediction and link detection in sequences of large social networks using temporal and local metrics.
Master thesis, University of cape Town, November 2006.

Francois FOUSS, Alain PIROTTE, Jean-Michel RENDERS et Marco SARENS :
Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation.
IEEE Transactions on knoledge and data engineering, 19(3):355–369, 2007.

A GOLDENBERG, J. KUBICA, P. KOMAREK, A. MOORE et J. SCHNEIDER :
A comparaison of statistical and machine learning algorithms on the task of linl completion.
In Proceedings of the KDD workshop on link analysis for detecting complex Behavior, August 2003.

Mohammad Al HASAN, Vineet CHAOJI, Saeed SALEM et Mohammed ZAKI :
Link prediction using supervised learning.
In SIAM Workshop on Link Analysis, Counterterrorism and Security with SIAM Data Mining Conference, Bethesda, MD, 2006.

Petter HOLME, Christofer R. EDLING et Fredrik LILJEROS :
Structure and time-evolution of an internet dating community,.
Social Networks, 26:155–174, 2004.

Bibliographie III

Zan HUANG, Xin LI et Hsinchun CHEN :
Link prediction approach to collaborative filtering.
In Mary MARLINO, Tamara SUMNER et Frank M. Shipman III, éditeurs : *JCDL*, pages 141–142. ACM, 2005.

Zan HUANG :
Link prediction based on graph topology : The predictive value of the generalized clustering coeffecient.
In Proceedings of LinkKDD'06, Philadelphia, Pennsylvania, 2006.

Jian HUANG, Ziming ZHUANG, Jia LI et C. Lee GILES :
Collaboration over time : characterizing and modeling network evolution.
In Marc NAJORK, Andrei Z. BRODER et Soumen CHAKRABARTI, éditeurs : *WSDM*, pages 107–116. ACM, 2008.

Jian HUANG, Ziming ZHUANG, Jia LI et C. Lee GILES :
Collaboration over time : characterizing and modeling network evolution.
In WSDM, pages 107–116, 2008.

Rushed KANAWATI :
On using sna techniques for enhancing performances of on-line help desks.
In Proceedings of IADIS International Conference on E-commerce, Amsterdam, july 2008. IADIS.

L. KATZ. :
A new status index derived from socimetric analysis.
Psychmetrika, 18(1), 18(1):39–43, 1953.

Bibliographie IV

Rushed KANAWATI et cÃ©line ROUVEIROL :
Lips : A sna-based system for intelligent management of academic conferences.
In Proceedings of the 4th International conference on application of Social networkk analysis, Zurich, october 2009.

Mayank LAHIRI et Tanya Y. BERGER-WOLF :
Structure prediction in temporal networks using frequent subgraphs.
In CIDM, pages 35–42. IEEE, 2007.

Jean loup GUILLAUME et Matthieur LATAPY :
Bipartite structure of all complex networks.
Information processing letters, 90:215–221, 2004.

Neal LATHIA, Stephen HAILES et Licia CAPRA :
knn cf : a temporal social network.
In Pearl PU, Derek G. BRIDGE, Bamshad MOBASHER et Francesco RICCI, éditeurs : *RecSys*, pages 227–234. ACM, 2008.

David LIBEN-NOWELL et Jon M. KLEINBERG :
The link-prediction problem for social networks.
JASIST, 58(7):1019–1031, 2007.

Bibliographie V

Jure LESKOVEC, Jon KLEINBERG et Christos FALOUTSOS :
Graphs over time : densification laws, shrinking diameters and possible explanations.
In KDD '05 : Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pages 177–187, New York, NY, USA, 2005. ACM.

David LIBEN-NOWELL :
An Algorithmic Approach tp Social networks.
Thèse de doctorat, M.I.T., june 2005.

M. LIM, M. NEGNEVITSKY et J. HARTNETT :
Artificial intelligence applications for analysis of e-mail communication activities.
In Proceedings of International Conference On Artificial Intelligence In Science And Technology, pages 109–113, 2003.

Tsuyoshi MURATA et Sakiko MORIYASU :
Link prediction based on structural proerties of online sociel networks.
new Generation Computing, 26:245–257, 2008.

Akira NINAGAWA et Koji EGUCHI :
Link prediction in metabolic networks using topology-based mixture models.
In Proceedings of the 19th International Conference on Genome Informatics, Gold Coast , Australia, December 2008.

Bibliographie VI

M. E. J. NEWMAN :
Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality.
Phys. Rev. E, 64(1):016132, Jun 2001.

M. E. J. NEWMAN :
Coauthorship networks and patterns of scientific collaboration.
Proceedings of the National Academy of Science of the United States (PNAS), 101:5200–5205, 2004.

Jung Hun OHN, Jihoon KIM et Ju Han KIM :
Social network anaysis of gene expression data.
In Proceedings of AMIA symposium : Biomediacl and health informatics, pages 958–958, Washington D.C., November 2003. AMIA.

Milen PAVLOV et Ryutaro ICHISE :
Finding experts by link prediction in co-authorship networks.
In Anna V. ZHDANOVA, Lyndon J. B. NIXON, Malgorzata MOCHOL et John G. BRESLIN, éditeurs : *FEWS*, volume 290 de *CEUR Workshop Proceedings*, pages 42–55. CEUR-WS.org, 2007.

Matthew J. RATTIGAN et David JENSEN :
The case for anomalous link discovery.
SIGKDD Explorations, 7(2):41–47, 2005.

Bibliographie VII

Mukund SESHADRI, Sridhar MACHIRAJU, Ashwin SRIDHARAN, Jean BOLOT, Christos FALOUTSOS et Jure LESKOVEC :
Mobile call graphs : beyond power-law and lognormal distributions.
In Ying LI, Bing LIU et Sunita SARAWAGI, éditeurs : *KDD*, pages 596–604. ACM, 2008.

U. SHARAN et J. NEVILLE :
Exploiting time-varying relationship in staistical relational models.
In Proceedings of the first SNA-KDD workshop, 2007.

Jerry SCRIPPS, Pang-Ning TAN, Feilong CHEN et Abdol-Hossein ESFAHANIAN :
A matrix alignment approach for link prediction.
In ICPR, pages 1–4. IEEE, 2008.

Benjamin TASKAR, Ming Fai WONG, Pieter ABBEEL et Daphne KOLLER :
Link prediction in relational data.
In Sebastian THRUN, Lawrence K. SAUL et Bernhard SCHÖLKOPF, éditeurs : *NIPS*. MIT Press, 2003.

J. ZHU :
Mining Web Site Link Structures for Adaptive Web Site Navigation and Search.
Thèse de doctorat, University of Ulster, 2003.

SESSION 4

Analyse de réseaux

Les logiciels Ucinet et Netdraw

D. FRACHISSE
(UNIVERSITÉ JEAN MONNET, FRANCE)

Les logiciels Ucinet et Netdraw

David Frachisse

École d'été Web intelligence
6 au 10 juillet 2009

Introduction
Représentation des réseaux
- Les graphes et leurs différentes formes
- Des données à la construction du graphe
- Une forme plus avancée de représentation

Propriétés générales d'un graphe
- Densité d'un graphe
- Graphe connexe et composante
- La notion de distance
- Résumé

Centralité des agents
- Le degré
- La centralité de proximité
- La centralité d'intermédiarité
- La centralité par les valeurs propres

Conclusion

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Introduction

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Qu'est ce qu'un réseau ?

- Un réseau est un ensemble de noeuds reliés entre eux par des liens.
- Le terme générique "réseau" définit un ensemble d'entités interconnectées les unes avec les autres. Un réseau permet ainsi de faire circuler des éléments matériels ou immatériels entre chacune de ces entités selon des règles bien définies.

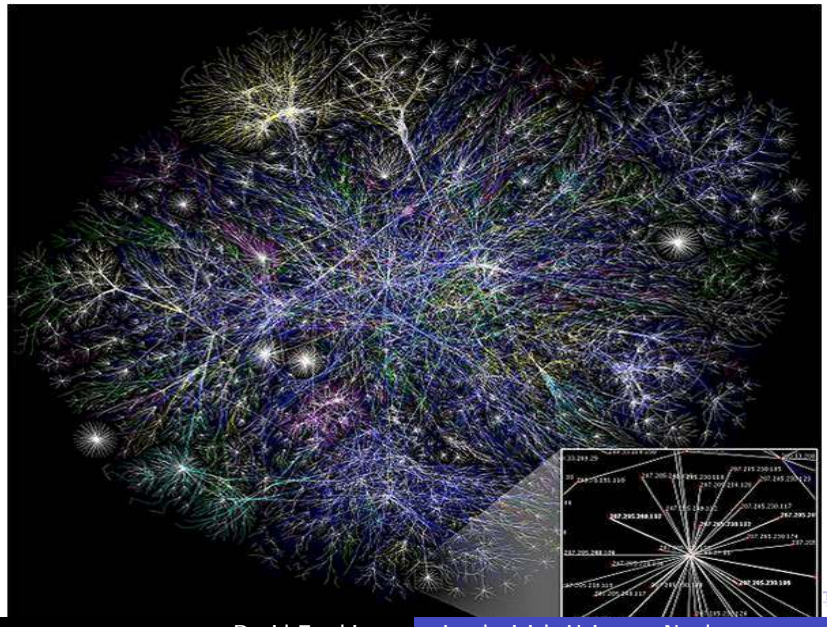
Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Les réseaux : une notion pluridisciplinaire

- La notion de réseau est utilisée dans différents domaines :
  - La médecine (réseaux neuronaux, réseaux sanguins,...) ;
  - L'informatique (Internet par exemple) ;
  - L'électrotechnique (réseau électrique) ;
  - L'économie (réseaux de collaborations) ;
  - Etc.
- Dans notre cas, on s'intéresse à la notion de réseau social utilisée à l'origine en sociologie et transposée de plus en plus à l'économie.

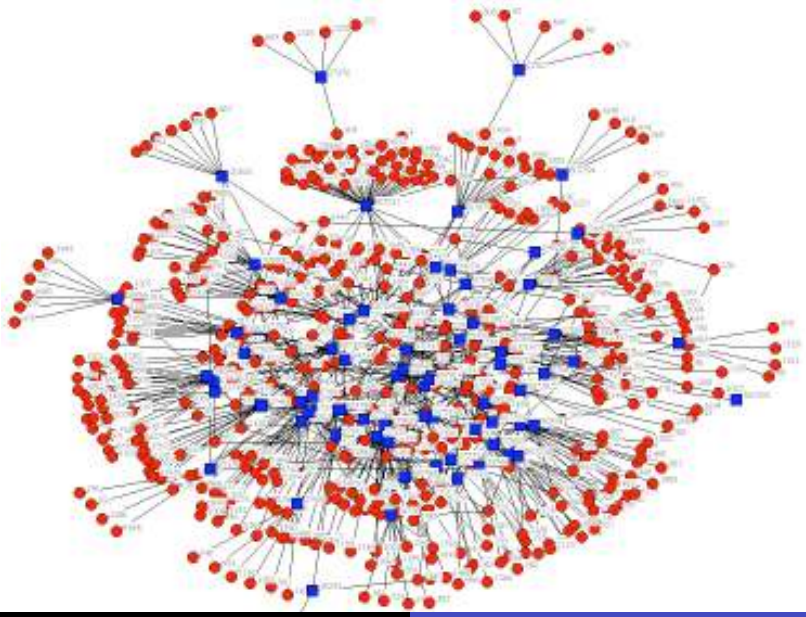
Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Exemple de réseau (1)



Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Exemple de réseau (2)



Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Exemple de réseau (3)

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Des outils pour analyser les réseaux

- Deux logiciels plus particulièrement utilisés dans l'analyse des réseaux sociaux :
  - **Ucinet** ⇒ Logiciel d'analyse des réseaux
    Borgatti, S.P., Everett, M.G. et Freeman, L.C. 2002. Ucinet for Windows : Software for Social Network Analysis. Harvard, MA : Analytic Technologies.
  - **Netdraw** ⇒ Logiciel de cartographie des réseaux
    Borgatti, S.P. 2002. NetDraw : Graph Visualization Software. Harvard : Analytic Technologies
- Ucinet et Netdraw sont complémentaires et développés par les mêmes personnes.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

D'autres logiciels

- Il existe d'autres logiciels :
  - Pajek
  - Tetralogie
  - Kinemages
  - R
  - Etc. . .
- Selon, l'objet de notre étude, certains sont plus adaptés que d'autres.

Représentation des réseaux

Le graphe

- Identification des agents qui appartiennent au réseau et des relations qui les relient entre eux.
- Seule une partie de l'information que l'on possède sur ces agents est retranscrite.
- La principale information est la relation qui relie les agents entre eux. Les caractéristiques des agents est une information secondaire.

Plusieurs types de réseau

- Le réseau peut être :
  - "One-mode network" versus "Two-modes network" ;
  - Réseaux dirigés et non dirigés ;
  - Réseaux valués et non valués ;
  - Avec ou sans liens réflexifs ;
  - Une ou plusieurs relations.

$\Rightarrow$ Un réseau est une combinaison de ces cinq possibilités.

Les données

- Les données pour représenter les réseaux se présentent la plupart du temps sous l'une de ces 3 formes :
  - Une matrice (d'adjacence ou d'affiliations) ;
  - Une liste de liens ;
  - Une table de données.

Dans le dernier cas, il faudra alors extraire les informations qui permettront de construire le réseau.

- **L'objectif est de construire une matrice qui permet de rendre compte des relations entre agents et de les cartographier.**

Exemple (1)

BDD Collaborations

ID	Projet	Domaine	Agent	Type	*Ajouter un nouveau champ*
1	Nuclea	Nucléaire	CEA	RES	
2	Nuclea	Nucléaire	EDF	IND	
3	Auto 1	Automobile	Michelin	IND	
4	Auto 1	Automobile	Renault	IND	
5	Auto 2	Automobile	Michelin	IND	
6	Auto 2	Automobile	Peugeot	IND	
7	Course	Automobile	Michelin	IND	
8	Course	Automobile	Peugeot	IND	
9	Course	Automobile	Oreca	OTH	
10	Autoelec	Automobile/E	Michelin	IND	
11	Autoelec	Automobile/E	Heuliez	IND	
12	Autoelec	Automobile/E	Orange	IND	
13	Elec	Electrique	EDF	IND	
14	Elec	Electrique	Orange	IND	
*	(Nouv.)				

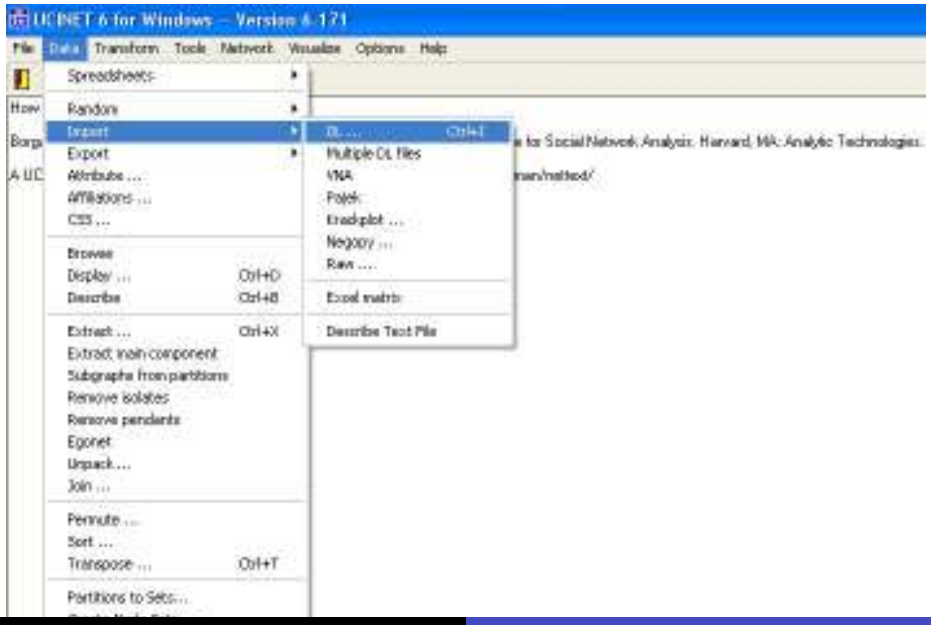
Exemple (2)

- Déclaration du fichier texte ;
- Déclaration du nombre d'agents et de projets ;
- Format du fichier texte ;
- Noms fournis en entrée ou non ;
- Données sélectionnées.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Les graphes et leurs différentes formes
Des données à la construction du graphe
Une forme plus avancée de représentation

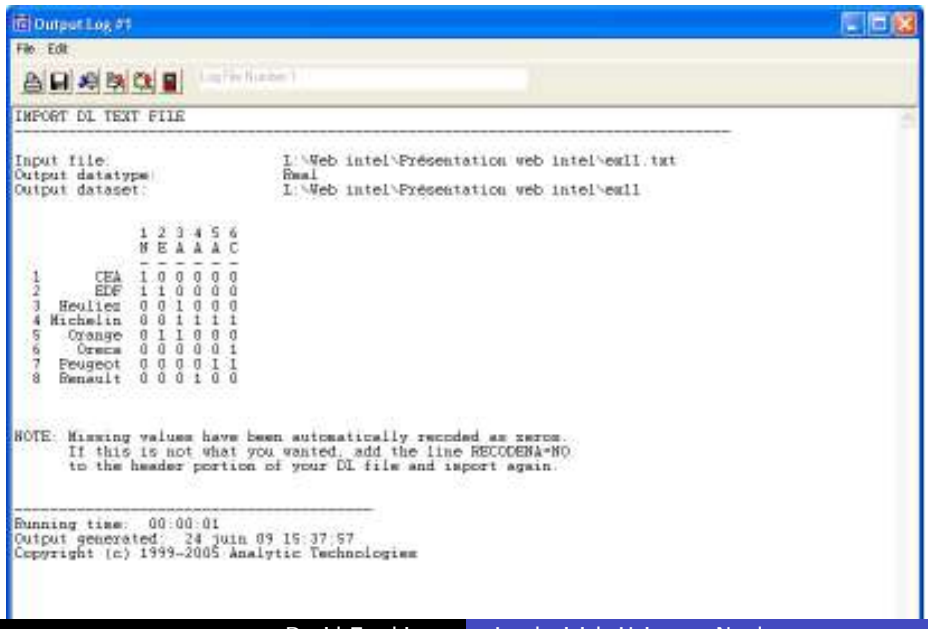
Exemple (3)



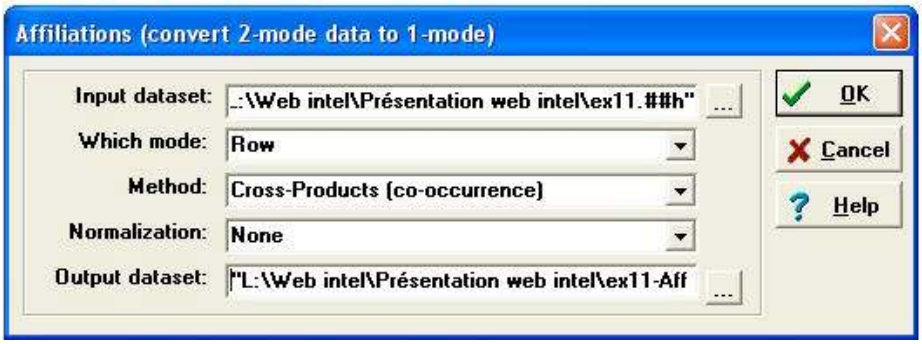
Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Les graphes et leurs différentes formes
Des données à la construction du graphe
Une forme plus avancée de représentation

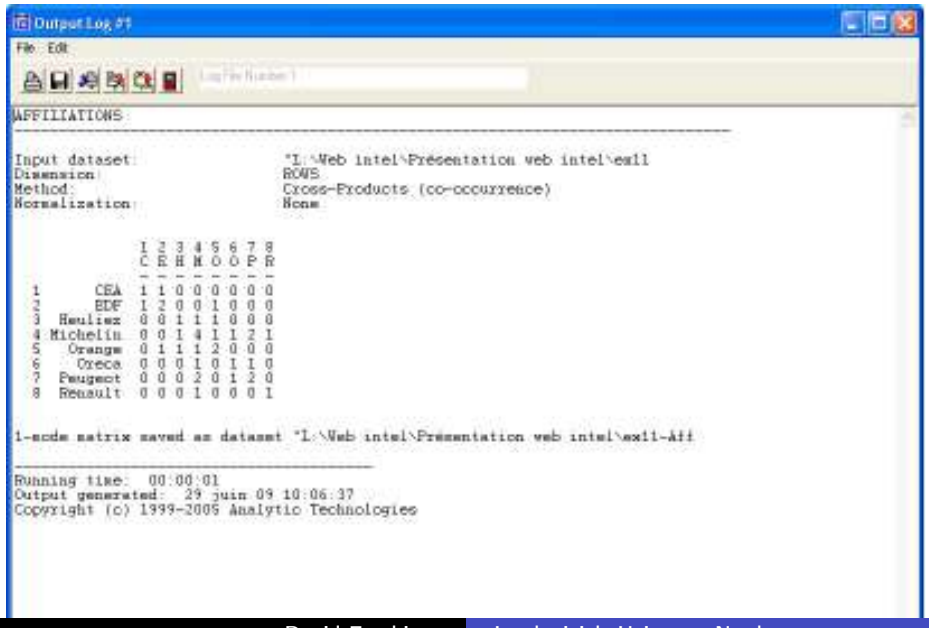
Exemple (4)



Exemple (5)



Exemple (6)

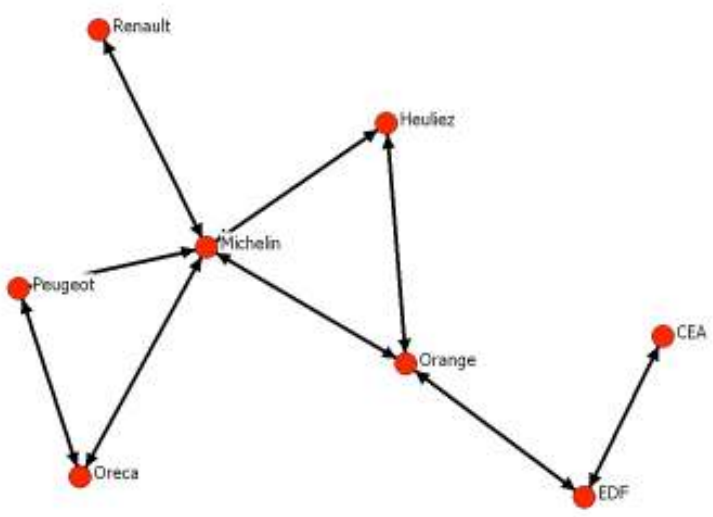


Exemple (7)

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Les graphes et leurs différentes formes
Des données à la construction du graphe
Une forme plus avancée de représentation

Exemple (8)



Quel est l'intérêt de cette méthode ?

- Améliorer la lisibilité du graphe.
- Amener de l'information supplémentaire.
- Maîtriser la construction du graphe.

L'utilisation des fichiers VNA

- Un fichier VNA est composé de 4 parties :
  - **Node data** : Les données concernant les agents, données importées de nos bases de données ;
  - **Node properties** : La manière dont sont représentés les agents sur le graphe ;
  - **Tie data** : Les données concernant les liens ;
  - **Tie properties** : La manière dont seront représentées les liens sur le graphe.

Présentation du fichier

excollaborations.vna - Bloc-notes

Fichier Edition Format Affichage ?

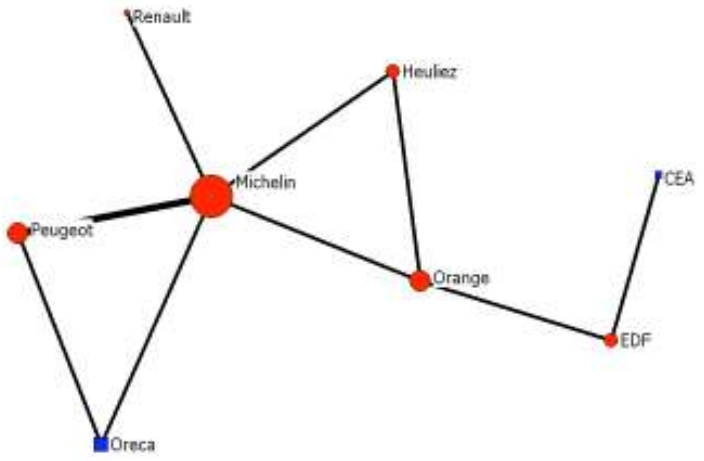
```
*Node data
"ID"
 "CEA"
 "EDF"

*Node properties
ID x y color shape size labeltext labelsize labelcolor gapx gapy active
"CEA" 940 532 255 1 16 "CEA" 18 0 3 5 TRUE
"EDF" 831 754 255 1 16 "EDF" 18 0 3 5 TRUE

*Tie data
FROM TO "ex11-Aff"
"CEA" "EDF" 1

*Tie properties
FROM TO color size headcolor headsize active
"CEA" "EDF" 0 5 0 20 TRUE
```


Exemple de représentation (1)

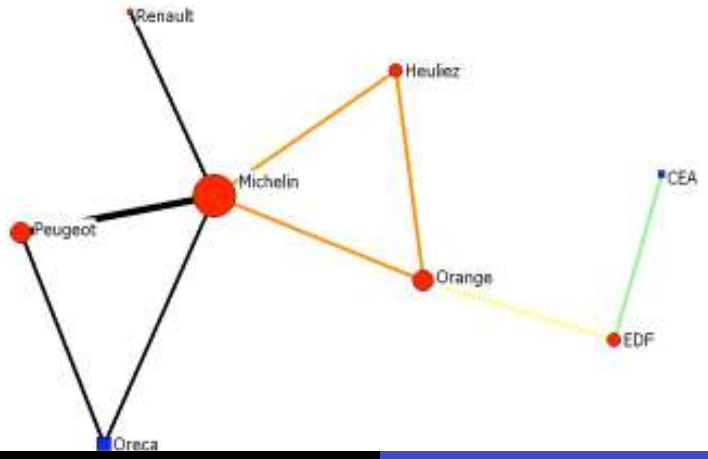


Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Les graphes et leurs différentes formes
Des données à la construction du graphe
Une forme plus avancée de représentation

Exemple de représentation (2)

Attention à ne pas surcharger le graphe et à le rendre moins lisible !



Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Densité d'un graphe
Graphe connexe et composante
La notion de distance
Résumé

Propriétés générales des réseaux

Définition formelle (1)

- La matrice d'adjacence résume les relations entre agents. Elle est notée X.

- Les graphes sont composés d'un ensemble d'agents et d'un ensemble de liens qui relie ces derniers les uns aux autres.
  - Soit $\mathcal{N}$ l'ensemble des agents avec N le nombre d'agents ;
  - Soit $\mathcal{L}$ l'ensemble des liens avec L le nombre de liens ;
  - Un graphe est noté : $g(\mathcal{N}, \mathcal{L})$.

Définition formelle (2)

- Un graphe g_s est un sous-graphe de g si l'ensemble des noeuds de g_s est un sous ensemble des noeuds de g et si l'ensemble des liens de g_s est un sous-ensemble des liens de g.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Densité d'un graphe
Graphe connexe et composante
La notion de distance
Résumé

Définition

- La densité d'un graphe est le nombre de liens potentiels qui sont actuellement présents dans le graphe.

$$D = \frac{L}{\frac{(N)(N-1)}{2}}$$

- La densité d'un graphe varie de 0 (aucun lien) à 1 (graphe complet).

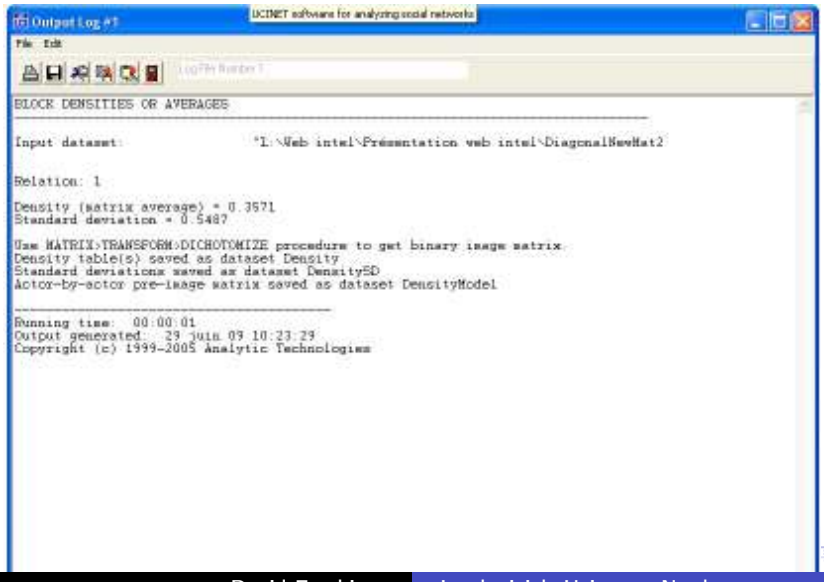
- C'est un indicateur relatif qui permet de comparer des graphes entre eux.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Densité d'un graphe
Graphe connexe et composante
La notion de distance
Résumé

Application

Network $\mapsto$ Cohesion $\mapsto$ Density



Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie
Densité d'un graphe
Graphe connexe et composante
La notion de distance
Résumé

Définition

- Un graphe est connecté s'il existe un chemin entre chaque paire de noeuds dans le graphe.

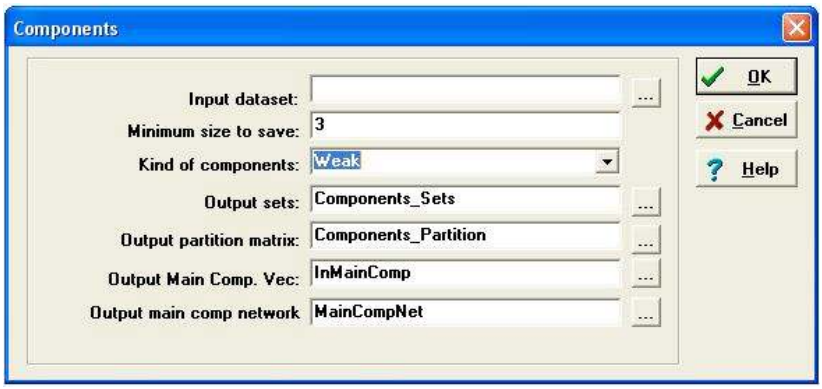
- Une composante est un sous-graphe tel qu'il existe un chemin entre tous les noeuds de la composante (lien direct ou indirect) et tel qu'il n'existe aucun chemin entre un noeud de la composante et un noeud qui n'appartient pas à la composante.

- Le singleton est un noeud isolé qui ne possède aucun lien.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie
Densité d'un graphe
Graphe connexe et composante
La notion de distance
Résumé

Application (1)

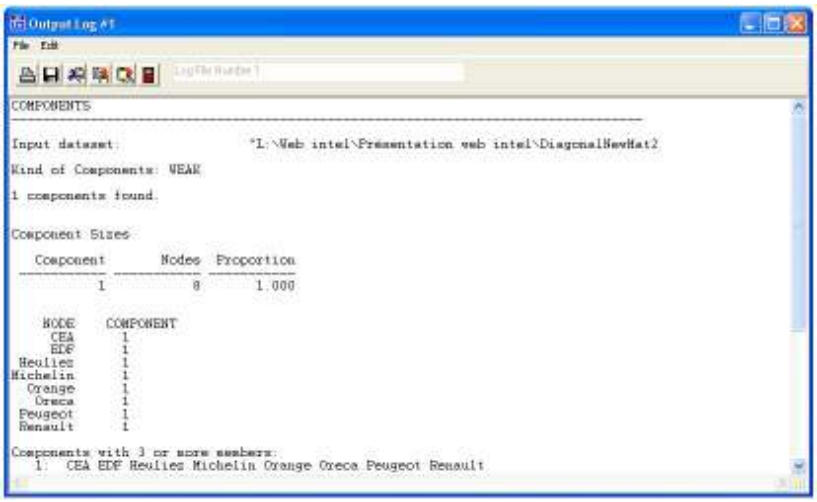
Network $\mapsto$ Regions $\mapsto$ Components



Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Densité d'un graphe
Graphe connexe et composante
La notion de distance
Résumé

Application (2)



Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

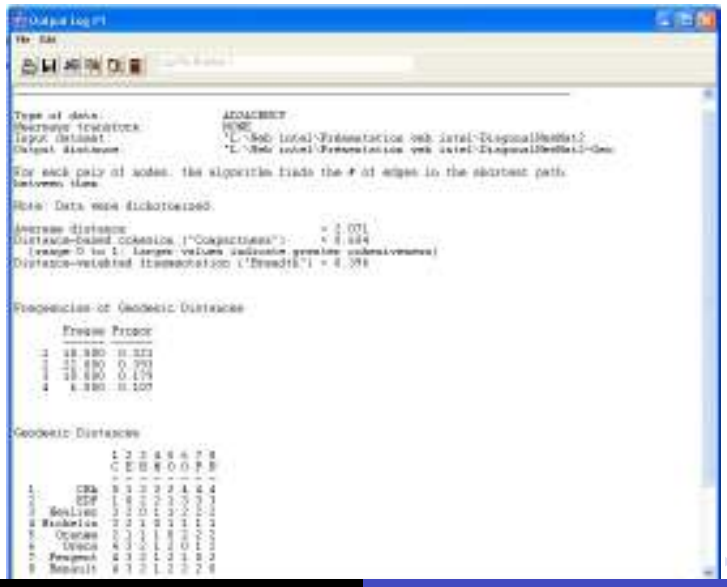
Densité d'un graphe
Graphe connexe et composante
La notion de distance
Résumé

Définition

- La distance géodésique entre deux noeuds est le plus court chemin entre ces deux noeuds.
- La distance moyenne d'un graphe connecté est égale à la moyenne des distance géodésiques entre toutes les paires d'agents.
- Le diamètre d'un graphe connecté est égal à la distance géodésique maximale au sein du graphe.

Application

Network $\mapsto$ Cohesion $\mapsto$ Distance



Caractériser le réseau

- L'ensemble de ces propriétés permettent de déterminer la forme du réseau et de confirmer le rendu visuel du graphe.

- Les propriétés générales permettent de faire des comparaisons (temporelles, sectorielles,. . .) entre graphes.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Centralité des agents

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Pourquoi mesurer la centralité d'un agent ?

Cela permet de :

- Positionner l'agent dans le réseau ;
- Établir l'importance de l'agent dans le réseau selon un critère ;
- Ordonner les agents au sein d'un graphe selon leur importance ;
- Réaliser des comparaisons pour un même agent entre deux situations (entre deux réseaux).

Une notion et plusieurs indicateurs

- Un agent sera central :
  - S'il possède un nombre important de liens ⇒ Centralité de degré ou ("Degree centrality") ;
  - S'il peut atteindre facilement les autres agents du graphe ⇒ Centralité de proximité (ou "Closeness centrality") ;
  - S'il est nécessaire à la communication au sein du graphe ⇒ Centralité d'intermédiarité (ou "Betweenness centrality") ;
  - S'il est lié ou proche d'agents qui sont eux-mêmes centraux ⇒ Centralité par les valeurs propres (ou "Eigenvector centrality").

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Quel type de mesure peut-on avoir ?

- Pour chaque indicateur, il est possible de définir trois mesures complémentaires :
  - L'indicateur brut qui donne la position d'un agent dans le graphe ;
  - L'indicateur relatif qui permet de prendre en compte la taille du graphe ;
  - L'indicateur général qui concerne le graphe et qui permet de comparer celui-ci à un graphe en forme d'étoile.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Définition du degré (1)

- Le degré d'un agent i est égal au nombre de liens de cet agent.
- Le degré d'un agent est égal à :

$$\mathbf{C_{D_i}} = \sum_{\mathbf{j}} \mathbf{x_{ij}}$$

avec $\mathbf{i} \neq \mathbf{j}$.

- Le degré standardisé ou normalisé est égal à :

$$\mathbf{C^n_{D_i}} = \frac{\mathbf{C_{D_i}}}{\mathbf{N-1}}$$

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Définition du degré (2)

- Pour chaque indicateur de centralité, il est possible d'obtenir un indicateur relatif à la centralité globale du graphe.
- Dans le cas du degré, l'indicateur est :

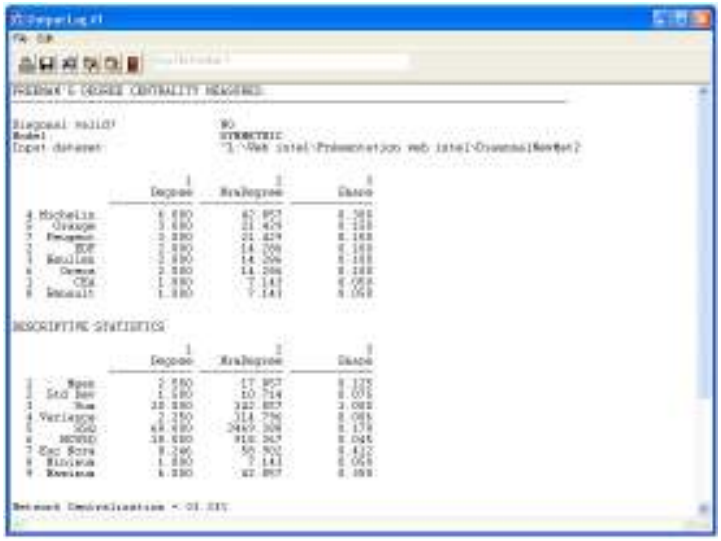
$$C_D(g) = \frac{\sum_{i=1}^{n} C_D^* - C_{D_i}}{(N-1)(N-2)}$$

avec C_D^* le degré maximal du graphe et C_{D_i} le degré de l'agent i.

- Grâce à cet indicateur, on compare le graphe étudié à un graphe en forme d'étoile.

Application

Network ↦ Centrality ↦ Degree



Avantages et inconvénients

- ⊕ Avantages :
  - Facile à calculer ;
  - Directement interprétable.
- ⊖ Inconvénients :
  - Ne prend en compte que les liens à distance 1 ;
  - Il peut entraîner rapidement des biais ;
  - Ne permet pas de discriminer entre les agents.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Définition

- Un agent est central s'il est situé à proximité de l'ensemble des autres agents du graphe.

- Intuition : un agent est central s'il peut rapidement interagir avec l'ensemble des autres agents du réseau.

- Le closeness d'un agent est égal à :

$$C_{C_i} = \left[\sum_j (dij)\right]^{-1}$$

- Le closeness normalisé est égal à :

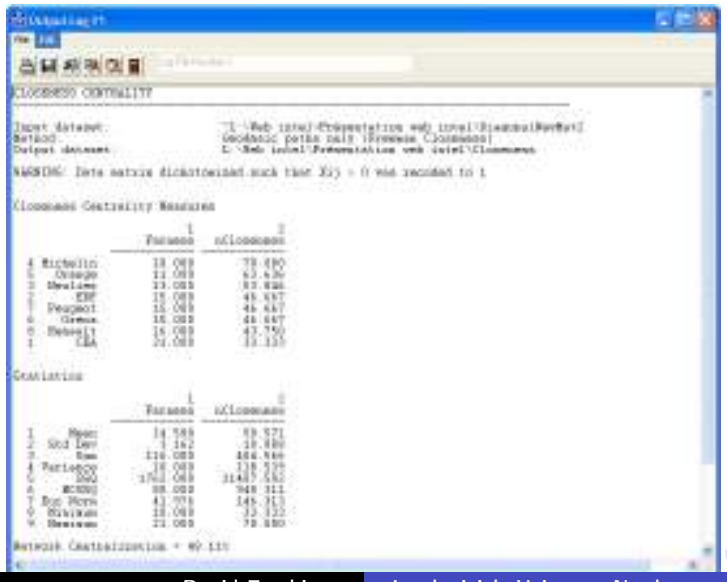
$$C^n_{C_i} = C_{C_i} \times (N - 1)$$

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Application

Network $\mapsto$ Centrality $\mapsto$ Closeness



Avantages et inconvénients

⊕ Avantages :

- Il prend en compte la position relative de l'agent par rapport à l'ensemble des autres agents du réseau.

⊖ Inconvénients :

- Il n'utilise qu'un seul chemin pour mesurer la distance entre deux agents : le chemin géodésique.
- Utilisable seulement si le graphe est connecté.

Définition

- Un agent i est d'autant plus central qu'il est situé sur le plus court chemin entre toutes les autres paires d'agents.

- Le betweenness d'un agent i est égal à :

$$C_{B_i} = \sum_{j<k} \frac{g_{jk}(i)}{g_{jk}}$$

avec g_{jk} le nombre de chemins géodésiques entre j et k et $g_{jk}(i)$ le nombre de chemins géodésiques entre j et k passant par i.

- Le betweenness normalisé d'un agent i est égal à :

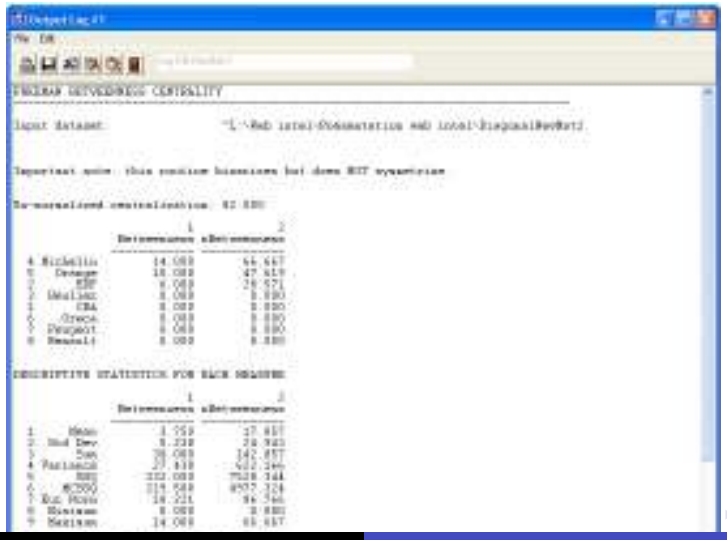
$$C^n_{B_i} = \frac{C_{B_i}}{\frac{(N-1)(N-2)}{2}}$$

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Application

Network $\mapsto$ Centrality $\mapsto$ Freeman Betweenness $\mapsto$ Node Betweenness



Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Avantages et inconvénients

⊕ Avantages :

- Un concept de centralité radicalement différent qui accorde peu d'importance au nombre de liens ;
- L'indicateur permet de localiser les agents nécessaires à la connexion du réseau, voir des cutpoints.

⊖ Inconvénients :

- L'indicateur est calculé uniquement sur les chemins géodésiques ;
- Difficile à transposer dans les cas des graphes valués et/ou dirigés.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Définition

- La centralité d'un agent dépend de la centralité des agents qui lui sont adjacents.

- Soit C_E le vecteur qui donne la centralité de l'ensemble des agents du graphe

$$\mathbf{X.CE} = \lambda \mathbf{CE}$$

avec λ la plus grande valeur propre et X la matrice d'adjacence du réseau.

- Plus la valeur est élevée plus l'agent est central.

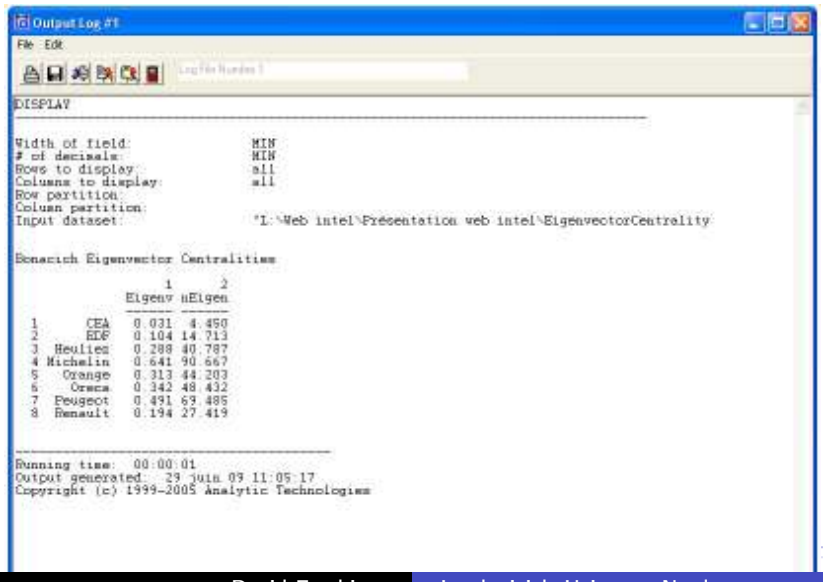
- La somme des mesures au carré fait 1.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Le degré
La centralité de proximité
La centralité d'intermédiarité
La centralité par les valeurs propres

Application

Network $\mapsto$ Centrality $\mapsto$ Eigenvector



Avantages et inconvénients

- ⊕ Avantages :
  - Il prend en compte la centralité de tous les autres agents pour calculer la centralité d'un agent ;
  - Mesure précise qui permet de discriminer entre les agents.
- ⊖ Inconvénients :
  - Difficilement calculable quand le nombre d'agent est trop important ;
  - Lorsque le graphe n'est pas connecté, seule la centralité des agents appartenant à la plus grande composante est calculée.

Conclusion

Conclusion (1)

- Les réseaux permettent de cartographier les relations qui relient les entités entre elles.
- Les propriétés générales des réseaux sont importantes afin de déterminer la forme du réseau de manière plus formelle et ainsi renforcer ou infirmer l'impression visuelle.
- Les indicateurs de centralité permettent de déterminer la position d'un agent au sein du réseau.

Conclusion (2)

- Bien déterminer quelle forme de réseau est celle qui convient le mieux à nos données.
- Adapter les définitions, formules,... au cas étudié.
- Faire attention aux interprétations.

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Bibliographie

Outline
Introduction
Représentation des réseaux
Propriétés générales d'un graphe
Centralité des agents
Conclusion
Bibliographie

Bibliographie

• Hanneman, Robert A. et Mark Riddle (2005), "Introduction to social network methods", *Riverside, CA : University of California, Riverside* (published in digital form at http ://faculty.ucr.edu/ hanneman/).

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

David Frachisse Les logiciels Ucinet et Netdraw

SESSION 5

Sécurité

Impact du web communautaire sur les missions de la DCRI

B. SUEUR
(DIRECTION CENTRALE DU RENSEIGNEMENT INTÉRIEUR, FRANCE)

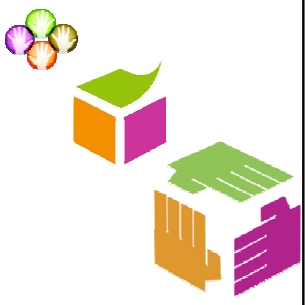
SESSION 6

Web sémantique et Web social

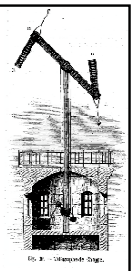
Web sémantique et Web social

F. GANDON
(INRIA, FRANCE)

Web sémantique & Web social

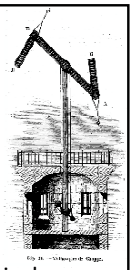


Fabien Gandon, *Inria, Sophia Antipolis, équipe Edelweiss*



préhistoire

- 1792 – frères Chappe: télégraphe optique en France
- 1835 – 1838 : télégraphe & Samuel Morse
- 1858 – Le premier câble transatlantique
- 1867 – Graham Bell invente le téléphone



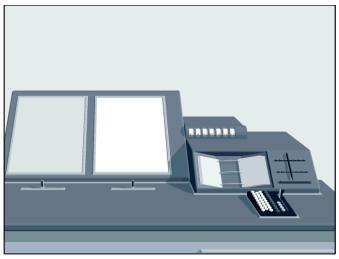
préhistoire

« est-ce un fait ou ai-je rêvé que, par le biais de l'électricité, le monde matériel est devenu un gigantesque nerf, vibrant sur des milliers de kilomètres »

Extrait traduit de: Nathaniel Hawthorne, *The House of the seven gables. The Flight of two Owls.* London, Collins' Clear-Type Press, 1851. P. 317-318.

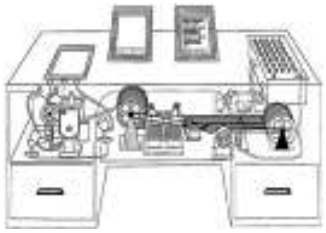
liens d'association [Vannevar Bush, 1945]

du mémex entre les informations qui nous importent

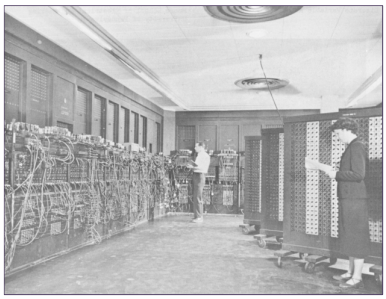


manque le virage du

digital



l'ordinateur



- 1946 – L'ENIAC le premier ordinateur électronique programmable.
- 1947 – Invention du transistor (Shocley & Brattain)
- 1958 – Invention du Circuit Intégré
- 1971 – Intel: invente et vend les microprocesseurs

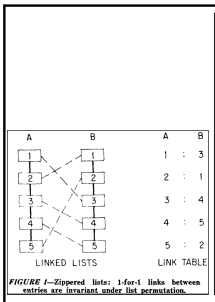


FIGURE 1—Zippered lists: 1-for-1 links between entries are invariant under list permutation.

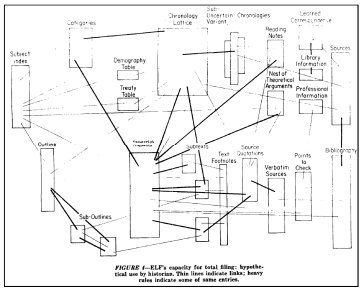


FIGURE 4—ELP's capacity for total filing; hypothetical use by historian. Thin lines indicate links; heavy rules indicate some of same entries.

hypertexte, hypermédia

[Ted Nelson, 1965]

une structure de fichier pour l'information complexe, changeante et indéterminée

manque le virage du

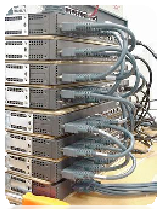
réseau

oNLine System (NLS), 1969

- 1955 – SABRE Réseau téléscripteurs réservations AA
- 1957 – ARPA: le deuxième effet Spoutnik
- 1958 – Premier bunker du réseau SAGE
  - Bell crée le premier Modem
- 1961 – Kleinrock (MIT) commutation de paquets
- 1962 – Licklider à l'ARPA & Réseau galactique
- 1963 – Code ASCII
- 1965 – Roberts & Merill: TX-2 (Massachussets) avec Q-32 (Californie) par une liaison téléphonique.
- 1967 – Plan ARPANET & NPL & Rand
- 1969 – Connexion UCLA, Stanford puis Santa Barbara et l'Utah ; 4 ordi: Arpanet

le réseau

- 1971 – ARPANET : 23 liaisons à 50 kbits/s.
- 1972 – démonstration publique ARPANET
  - email et listes de diffusion
- 1973 – protocole inter-réseaux (network) Cerf, Kahn
  - TCP, Telnet, FTP puis IP/UDP (1975)
- 1979 – Newsgroups
  - Metcalfe:Ethernet & 3Com
- 1981 – BITNET réseau universitaire (Internet)
  - IP au CERN
- 1982 – ARPANET en IP
- 1983 – Développement du DNS



les réseaux

connectés entre eux mais pas à l'humain

développement des périphériques et interfaces

- 1962 – Engelbart (Stanford) traitement de texte
- 1963 – ASCII (American Standard Code for Information Interchange)
- 1965 – TVEdit (Stanford) premiers éditeurs de texte
- 1966 – Engelbart crée la souris
- 1967 – Univ. de Brown & IBM: éditeur d'hypertexte au stylo optique pour projet Apollo.
- 1968 – NLS de Engelbart: lien et édition à la souris.

l'interaction

- 1982 – carte Hercules IBM PC 720x350 Noir & Blanc
- 1983 – Lisa Xerox micro avec interface graphique (menu, fenêtres, icônes,…) une souris
- 1987 – Hypercard livré avec le Macintosh
  – début commercialisation Internet
  – format d'image GIF de Compuserve
- 1988 – Next = Unix & GUI & Objets

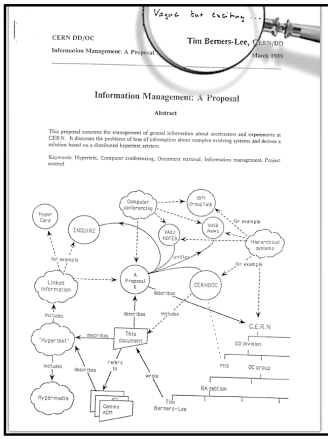
l'interaction graphique

http://www

liens utilisant des références

à travers le réseau

[Tim Berners-Lee, 1989]



le web social

a quelque chose d'un pléonasme

```
<p> je suis l'auteur de ce
 <a href="doc.html">document</a>
</p>
```

je suis l'auteur de ce document

1997, version 3.2 de **HTML**

règles d'écriture (URL Web)

https://192.54.152.134/search?name=jim#5

- minuscules & chiffres
- séparateurs : / . :
- séparateurs particuliers:
  – requêtes ?
  – fragments #

HTTP: HyperText Transfer Protocol

- HTTP: échanges de commandes et résultats

```
GET cv.html
Content-Type: text/html
If-Modified-Since: 31 Jul 1999 12:30:00
From: Fabien Gandon Fabien.Gandon@sophia.inria.fr
```

- HTTPS = HTTP+ SSL (Secure Socket Layer)
  Confidentialité, Intégrité, Authentification

HTTP

erreur ou métadonnées

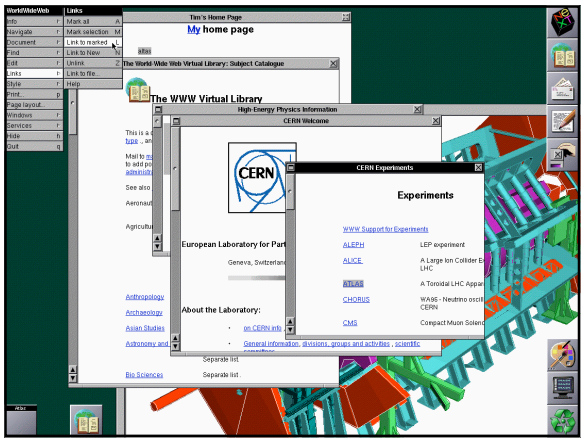
1xx: Information
100 Continue
101 Switching Protocols

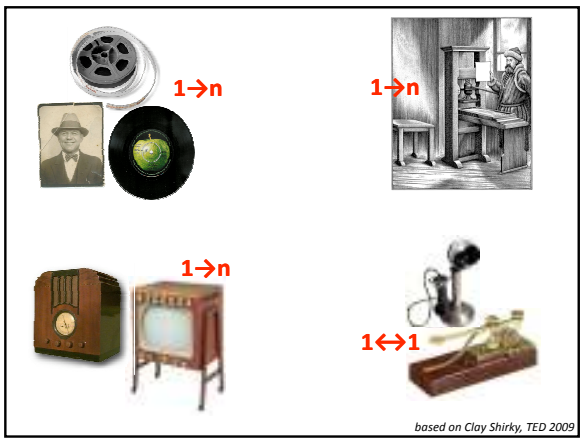
2xx: Successful
200 OK
201 Created
202 Accepted
203 Non-authoritative Information
204 No Content
205 Reset Content
206 Partial Content

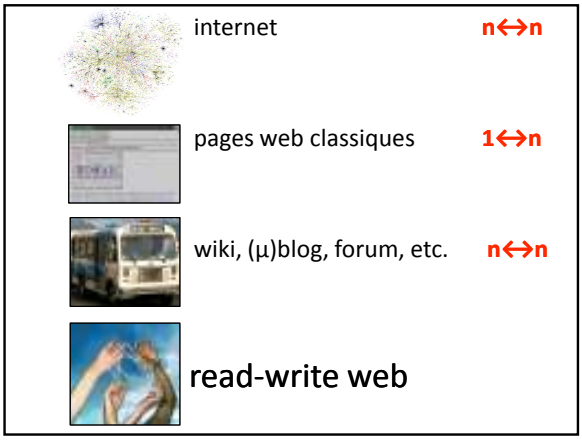
3xx: Redirection
300 Multiple Choices
301 Moved Permanently
302 Found
303 See Other
304 Not Modified
305 Use Proxy
306 Unused
307 Temporary Redirect

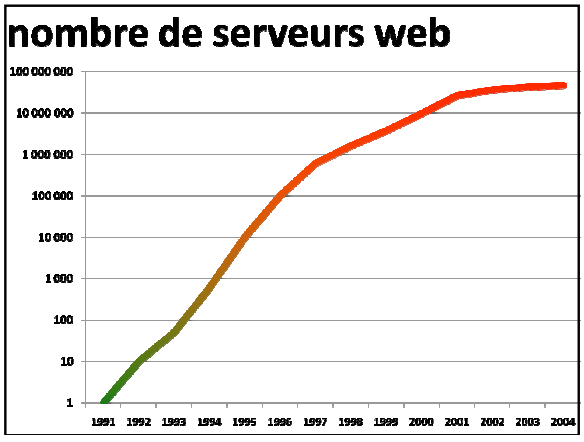
4xx: Client Error
400 Bad Request
401 Unauthorized
402 Payment Required
403 Forbidden
404 Not Found
405 Method Not Allowed
406 Not Acceptable
407 Proxy Authentication Required
408 Request Timeout
409 Conflict
410 Gone
411 Length Required
412 Precondition Failed
413 Request Entity Too Large
414 Request-URL Too Long
415 Unsupported Media Type
416 Requested Range not satisfiable
417 Expectation Failed

5xx: Server Error
500 Internal Server Error
501 Not Implemented
502 Bad Gateway
503 Service Unavailable
504 Gateway Timeout
505 HTTP Version Not Supported









1994

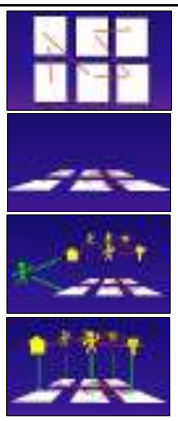
consortium pour harmoniser les évolutions du web

notes et recommandations

- recommandation
- recommandation proposée
- recommandation candidate
- dernier appel
- brouillon de travail

web sémantique

mentionné par Tim BL
en **1994** à WWW



[Tim Berners-Lee 1994, http://www.w3.org/Talks/WWW94Tim/]

séparer

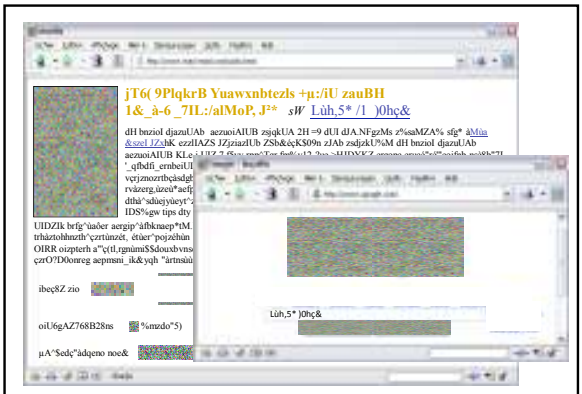
forme et contenu

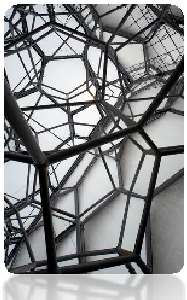
ne lisez **pas**
le panneau suivant

book victor hugo Search

Avenue Victor Hugo Bookshop -
Used and new **books** with searchable catalog and newsletter.
www.avenue**victorhugobooks**.com/ - 16k

Victor Hugo: Bibliography
Victor Hugo: A Companion to His Poetry, J.C. Ireson.
Hugo's Les Miserables and the **Novels** of the Grotesque, **...**
www.poetry-archive.com/h/**hugo_victor**_bibliography.html - 21k





vers un web
structuré

2008 XML a eu dix ans

- structurer des données
- en utilisant des balises
- dans un format textuel
- construction modulaire
- une famille de langage
- un format ouvert non propriétaire

XML en 10 points http://www.w3.org/XML/1999/XML-in-10-points.fr.html

XML
bien formé

```
<racine>...</racine> EOF
<x>...</x> ou <x/>
<a><b></a></b>
<x> ≠ <X>
<lan> <xmla> <bla bla>
<a b='c'></a> ou <a b='c'/>
```

Scalable Vector Graphics (SVG)
eXtensible Access Control Markup Language (XACML)
MusicML
OpenOffice Schemas
WSDL
MathML
XML Topic Maps (XTM)
Chemical Markup Language (CML)
Financial products Markup Language (FpML)
GPS Exchange format (GPX)
XHTML
NewsML
famille de langages
Synchronized Multimedia Integration Language (SMIL)
(...)

XML

XML

- Uniform Resource Identifiers
- sous ensemble URL
- déréférencer ?

URI

exemple:
ldap://[2001:db8::7]/c=GB?objectClass?one

namespace

- espaces de nommage
- éviter les collision de noms
- associer un préfixe de balise à un URI
- exemple

```
<ugb:note xmlns:ugb='http://www.ugb.sn/'>18</ugb:note>
<bla:note xmlns:bla='http://www.ugb.sn/'>18</bla:note>
```

- définitions héritées dans l'arbre XML
- namespace par défaut **xmlns="…"**

XPath

chemin dans un arbre xml

/livre/chapitre[position()=1]

URL+'#'+XPath=

XPointer

http://www.inria.fr/teams.xml#xpointer(edelweiss/articles[position()<=10])

XLink

des liens entre les éléments de documents XML

XQuery

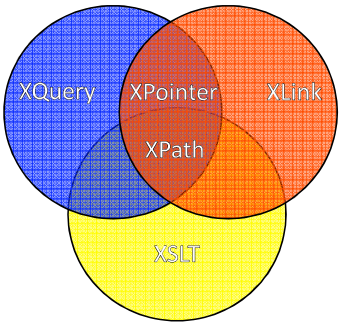
requêtes sur ces données et documents structurés

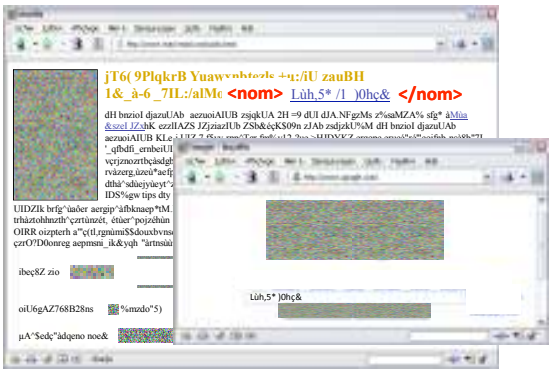
```
for $x in doc("books.xml")/bookstore/book
where $x/price>30
```

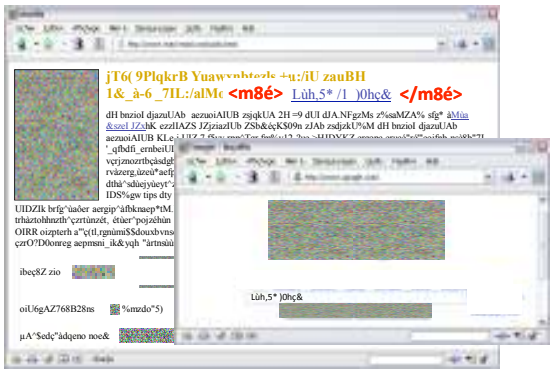
XSL

XSL-FO : style et formatage

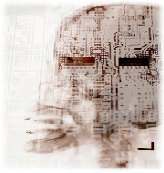
XSLT: transformation







il manque quelque chose…
une connaissance



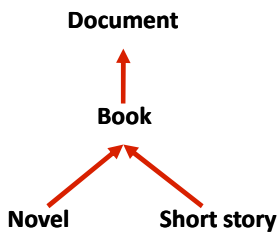
quel est le dernier
document
que vous avez lu?

documents

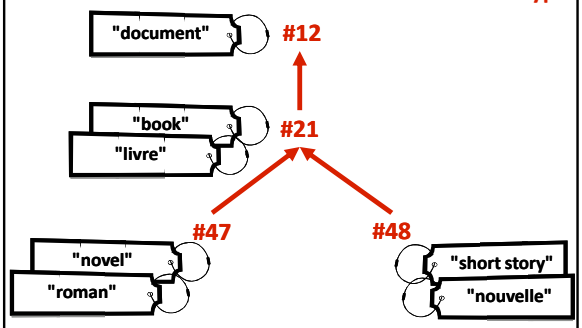
votre réponse est basée sur une
ontologie partagée

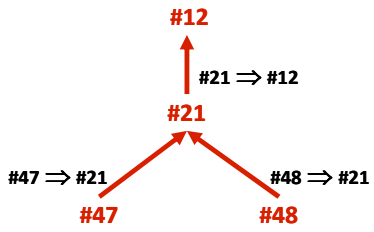
- je peux comprendre
- vous pouvez raisonner

sous type



sous type





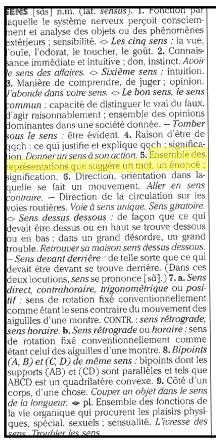
connaissance
ontologique formalisée

Ontology

a logical theory which gives an explicit, partial account of a conceptualization *i.e.* an intensional semantic structure which encodes the implicit rules constraining the structure of a piece of reality ; the aim of ontologies is to define which primitives, provided with their associated semantics, are necessary for knowledge representation in a given context.

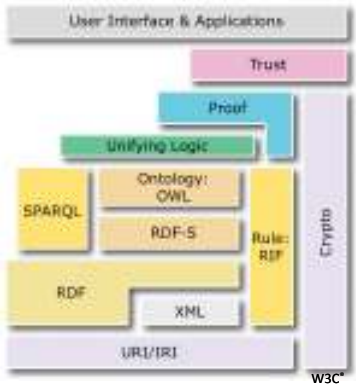
[Gruber, 1993] [Guarino & Giaretta, 1995] [Bachimont, 2000]

56



SENS [sɑ̃s] n.m. (lat. *sensus*). 1. Fonction par laquelle le système nerveux perçoit consciemment et analyse des objets ou des phénomènes extérieurs ; sensibilité. ◇ *Les cinq sens* : la vue, l'ouïe, l'odorat, le toucher, le goût. 2. Connaissance immédiate et intuitive ; don, instinct. *Avoir le sens des affaires.* ◇ *Sixième sens* : intuition. 3. Manière de comprendre, de juger ; opinion. *Il abonde dans votre sens.* ◇ *Le bon sens, le sens commun* : capacité de distinguer le vrai du faux, d'agir raisonnablement ; ensemble des opinions dominantes dans une société donnée. — *Tomber sous le sens* : être évident. 4. Raison d'être de qqch ; ce qui justifie et explique qqch ; signification. *Donner un sens à son action.* 5. Ensemble des représentations que suggère un mot, un énoncé ; signification. 6. Direction, orientation dans laquelle se fait un mouvement. *Aller en sens contraire.* — Direction de la circulation sur les voies routières. *Voie à sens unique. Sens giratoire.* ◇ *Sens dessus dessous* : de façon que ce qui devait être dessus ou en haut se trouve dessous ou en bas ; dans un grand désordre, un grand trouble. *Retrouver sa maison sens dessus dessous.* — *Sens devant derrière* : de telle sorte que ce qui devait être devant se trouve derrière. (Dans ces deux locutions, *sens* se prononce [sɑ̃].) 7. a. *Sens direct, contrahoraire, trigonométrique* ou *positif* : sens de rotation fixé conventionnellement comme étant le sens contraire du mouvement des aiguilles d'une montre. CONTR. : *sens rétrograde, sens horaire.* b. *Sens rétrograde* ou *horaire* : sens de rotation fixé conventionnellement comme étant celui des aiguilles d'une montre. 8. *Bipoints (A, B) et (C, D) de même sens* : bipoints dont les supports (AB) et (CD) sont parallèles et tels que ABCD est un quadrilatère convexe. 9. Côté d'un corps, d'une chose. *Couper un objet dans le sens de la longueur.* ◆ pl. Ensemble des fonctions de la vie organique qui procurent les plaisirs physiques, spécial. sexuels ; sensualité. *L'ivresse des sens. Troubler les sens.*

vers un web
sémantique

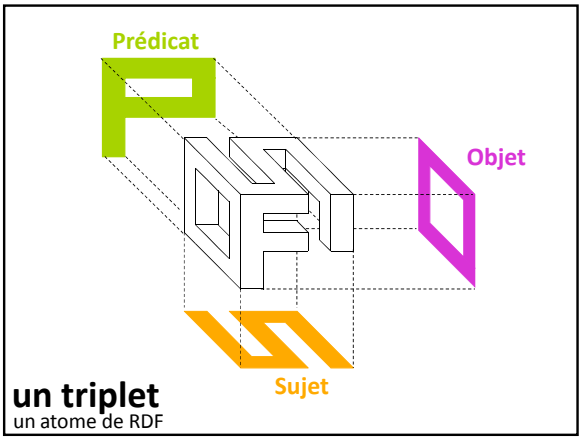


RDF est un modèle de triplets *i.e.*
toute connaissance est décomposée en

(sujet , predicat , objet)

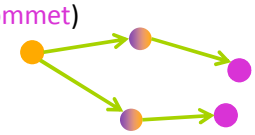
doc.html a pour auteur Fabien et a pour thème la Musique

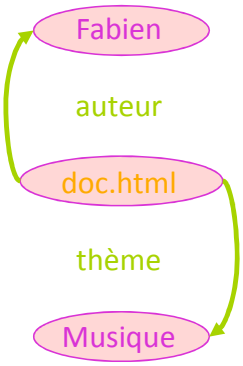
doc.html a pour auteur Fabien
doc.html a pour thème Musique



un triplet
un atome de RDF

les triplets peuvent aussi être vus comme les arcs d'un graphe RDF
(sommet, arrête, sommet)





<RDF/> une syntaxe XML pour mieux échanger ces graphes

ouvrir et relier des données dans un
graphe mondial

RDF Graph // Social network

utiliser le modèle RDF pour des graphes de réseaux sociaux

interroger avec SPARQL

SPARQL Protocol and RDF Query Language

requête SPARQL

```
SELECT ...
FROM ...
WHERE { ... }
```

résultat SPARQL

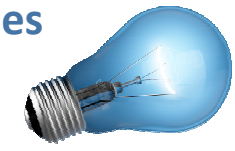
```
échec / réussite
valeurs trouvées
```

protocole SPARQL

échanger des requêtes et leurs résultats à travers le web

71

SPARQL pour accéder aux données sociales

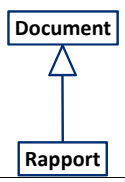


RDFS signifie RDF Schema

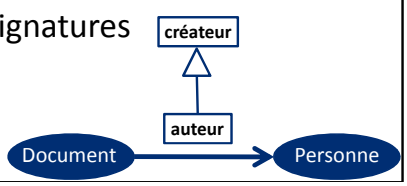
RDFS fournit les primitives pour écrire des ontologies légères

une petite goute de sémantique peut faire une grande vague à l'échelle du **Web**

RDFS pour définir les classes de ressources et organiser leur hiérarchie



RDFS pour définir les relations, leur hiérarchie et leurs signatures



OWL fournit des primitives supplémentaires pour des ontologies plus lourdes

OWL permet de définir…

{a,b,c,d,e} par énumération

par union

≠ la disjonction

↔ prop. algébriques

par intersection

par complément

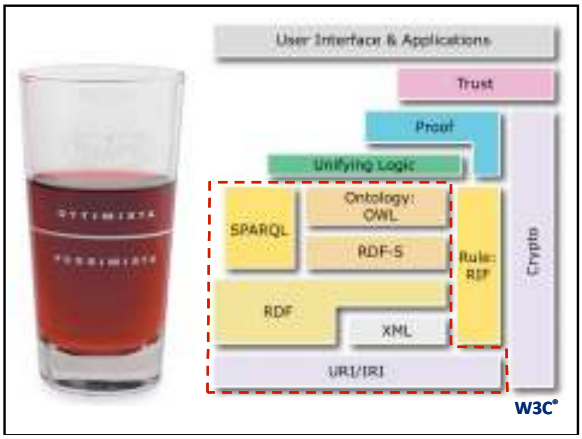
1 → par restriction

1..1 → par cardinalité

≡ par équivalence

OWL 2.0 sucre syntaxique et extensions sémantiques

RDFS/OWL pour typer les réseaux



expression naturelle pour dénoter un concept

```
inria:CorporateSemanticWeb
  skos:prefLabel "corporate semantic web"@en;
  skos:prefLabel "web sémantique d'entreprise"@fr;
  skos:altLabel "corporate SW"@en;
  skos:altLabel "CSW"@en;
  skos:hiddenLabel "web semantique d'entreprise"@fr.
```

relations

entre concepts

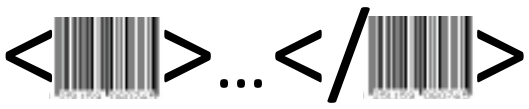
```
inria:CorporateSemanticWeb
  skos:broader w3c:SemanticWeb;
  skos:narrower inria:CorporateSemanticWiki;
  skos:related inria:KnowledgeManagement.
```

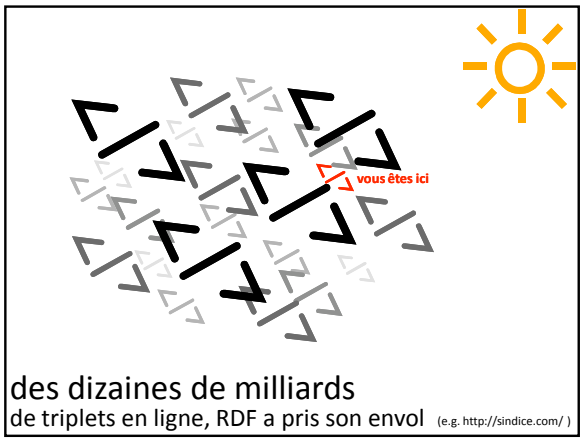
```
inria:CorporateSemanticWeb
 skos:scopeNote "only within KM community";
 skos:definition "a semantic web on an intranet";
 skos:example "Nokia's internal use of RDF gateway",
 skos:historyNote "semantic intranet until 2006";
 skos:editorialNote "keep wikipedia def. uptodate";
 skos:changeNote "acronym added by fabien".
```

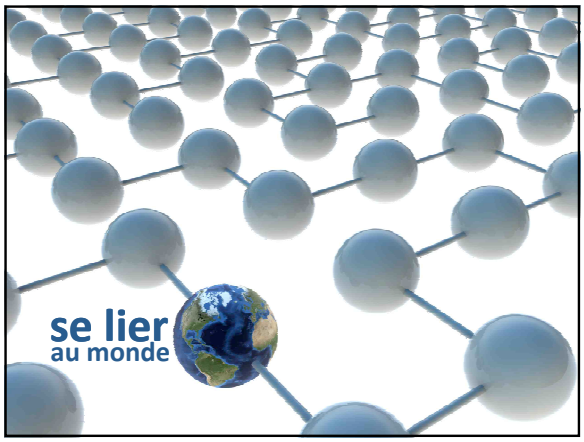
SKOS pour organiser des index thématiques, des tags, etc.

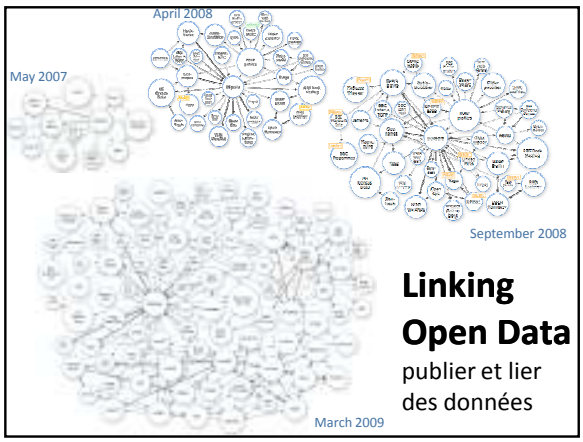
spécifier le sens

avec des identificateurs uniques




vous êtes ici
des dizaines de milliards
de triplets en ligne, RDF a pris son envol
(e.g. http://sindice.com/)


se lier
au monde


April 2008
May 2007
September 2008
March 2009
Linking
Open Data
publier et lier
des données

FOLLOW
YOUR
NOSE

silos de données
problème général sur le web

walled gardens chaque application web cultive son jardin de données de façon privative

interopérabilité

et redonner ses données à l'utilisateur (API, schémas)

données personnelles

un premier point d'accroche

BioMec

FOAF (friend of a friend)

décrire les personnes

[Brickley & Miller 2004]

Site	Example	People	links out?	OpenID Integration?
Advogato	http://www.advogato.org/person/connolly/foaf.rdf#me	11972 as of Apr 13 2009	foaf:homePage (owl:sameAs patch submitted Presbrey)	No
BCmoney	http://bcmoney-mobiletv.com/bcmoney/foaf	315 as of Jan 2009	Yes	Yes
Blogolina				
boards.ie	http://www.boards.ie/vbulletin/foaf.php?u=4	122,000	Not yet	No
Buzznet				
Crazyville				
DeadJournal				Yes
Delhometr				
Ecademy	http://www.ecademy.com/module.php?mod=network&op=foafdlBuid=1	150,000	Yes	Yes
Elgg	http://elgg.org/katelnanm/foaf.*	?	?	?
Explode	?	~6M	?	?
FilmTrust				
Fotothing		144,342		
FriendFeed	http://friendfeed.com/johnnd/subscriptions?output=foaf		Yes	
FOAFing the Music	http://foafing-the-music.iua.upf.edu/data/foaf/xcetnra.rdf	~4,000	Yes	No
GreatestJournal		1,931,283		
hi5	http://api.hi5.com/rest/profile/foaf/3525002/2 *	?	?	?
InsaneJournal		91,209		
Idea fm	http://alpha.libre.fm/user/USERNAME#me	3500 (Apr 2009)	Yes - does rdf about="WEBID" if given as external WebID	No
Linking Open Data Community	http://community.linkeddata.org/dataspace/person/kidehen2#this	stats to be gathered	Yes	Yes
LiveJournal	http://danbri.livejournal.com/data/foaf.*	● > 15M	?	Yes
LJ Provia.org				
Minilog.com				
MyBlogLog	http://www.myblogiog.com/buzz/members/jbond/foaf		No	No
MyOpera	http://my.opera.com/ujutik/xml/foaf#me	??	Yes	?
My OpenLink Service (an OpenLink Data Spaces instance)	http://my.openlinksw.com/dataspace/person/kidehen@openlinksw.com#this and http://myopenlink.net/dataspace/person/kidehen#this	> 10,097 as of 31 Oct. 2007	Yes	Yes
Orgo	http://www.orgo.at/hc.cfm/fa/Powsis	2600	Yes	Yes
Prosfile	http://prosfile.org/alfields.demo/foaf		Yes	Yes
QDOS	http://qdos.com/user/5ac/361496df09a7c29677f0d5d5f792/rdf/rdf		Yes, q:foaf	Yes
ryepeople	http://www.ryepeople.com/users/fasarison	2000000	Yes	No
Revyu	http://revyu.com/people/tom/about/rdf	in the hundreds	Yes, rdfs:seeAlso and owl:sameAs	Not yet
Twitter	http://tools.opiumfield.com/twitterfommorris/rdf		rdfs:seeAlso and owl:sameAs within same network	No
Videntity				Yes
Vox	http://path.vox.com/profile/foaf.rdf		NO	
Zopto				
Zoomeroo				

FOAF sites (ESW Wiki)

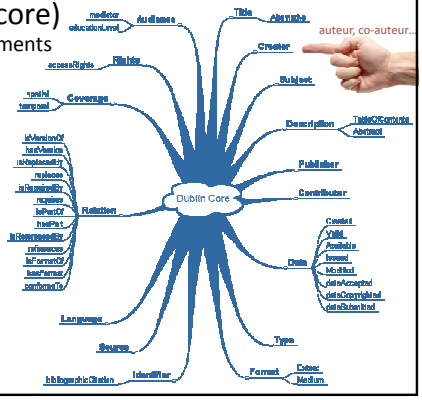
plusieurs services supportant FOAF notamment…

- Google's Social Graph API
- Yahoo's SearchMonkey

RELATIONSHIP
spécialisation des relations

http://vocab.org/relationship/

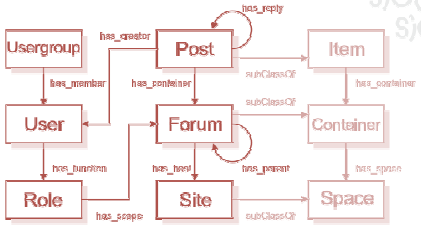
DC (dublin core)
décrire les documents



Semantically-Interlinked Online Communities
décrire des sites communautaires

SIOC

[Breslin et al., 2005]



analyse des réseaux sociaux sur le web sémantique déjà plusieurs contributions…

… [Goldbeck et al 2003] propagation de confiance

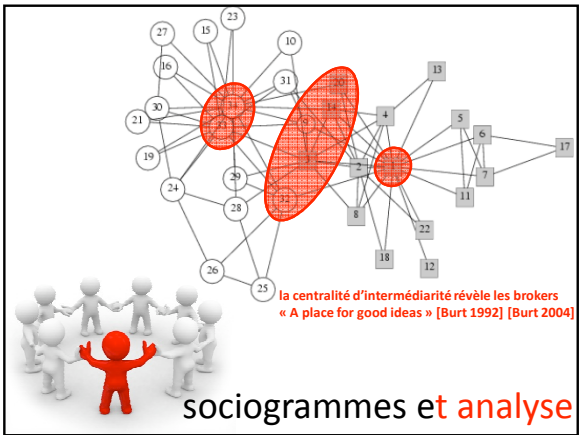
… [Finin et al 2005] loi de puissance des degrés et structure en communautés

… [Paolillo et al 2006] SNA classique sur FOAF LiveJournal

… [Goldbeck et Rothstein 2008] fusion de profils FOAF

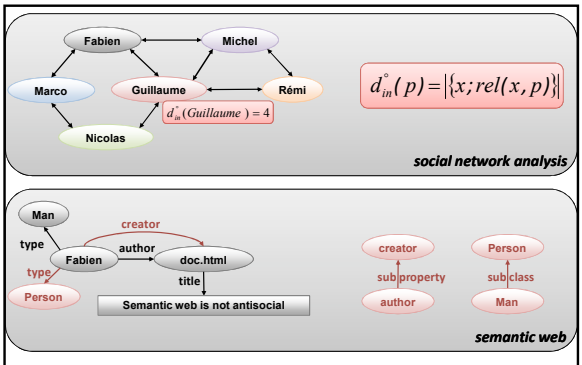
… [Anyanwu et al 2007] [Kochut et al 2007] [Corby et al 2004] [Corby 2008] extensions de SPARQL pour la manipulation des chemins

… [Ereteo et al 2009]

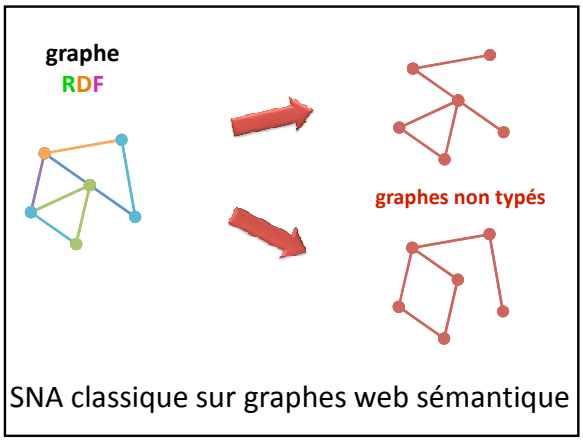


la centralité d'intermédiarité révèle les brokers « A place for good ideas » [Burt 1992] [Burt 2004]

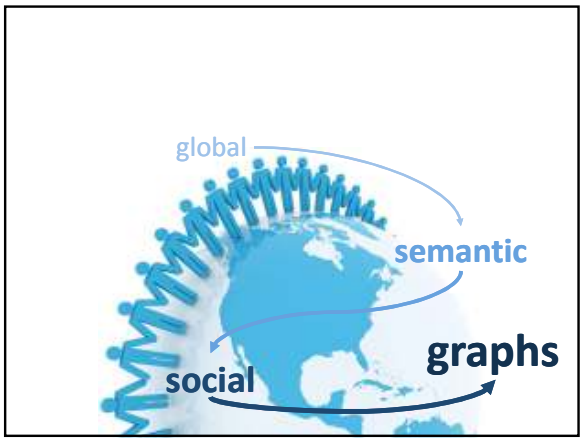
sociogrammes et analyse

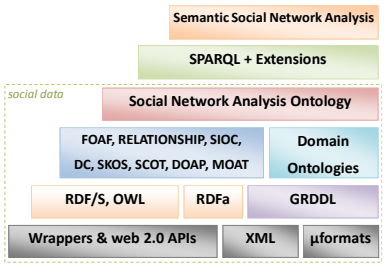


graphes, graphes, graphes, …



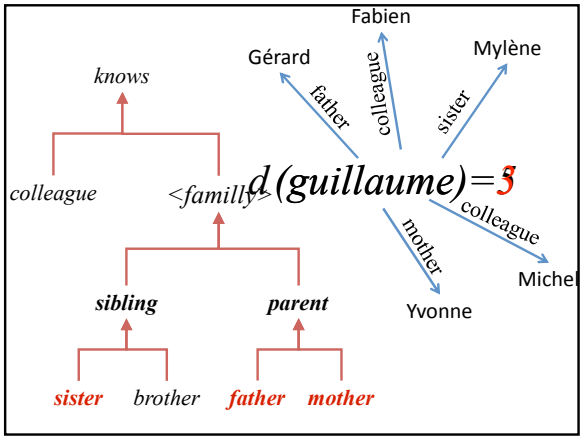
SNA classique sur graphes web sémantique





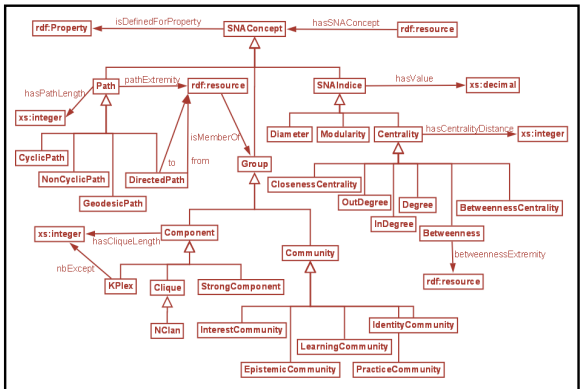
construire sur pile du web sémantique

[PhD Guillaume Erétéo]



$D_{-in<type, length>}(y)$:=

```
select ?y count(?x) as ?degreeCentrality where {
    ?x $path ?y
    filter(match($path, star(param[type])))
    filter(pathLength($path) <= param[length])
} group by ?y
```



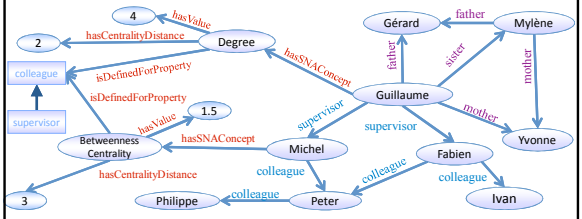
représenter les résultats

[PhD Guillaume Erétéo]

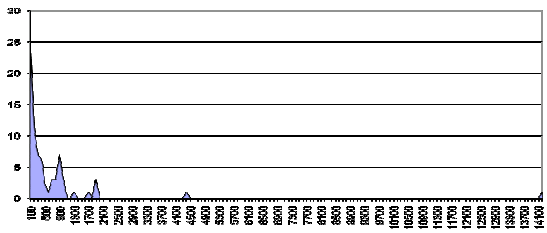
```
ADD {
  ?y semsna:hasInDegree _:b0
   _:b0 semsna:forProperty param[type]
   _:b0 rdf:value ?indegree
   _:b0 semsna:hasLength param[length]
}
SELECT ?y count(?x) as ?indegree {
  ?x $path ?y
  filter(match($path, star(param[type])))
  filter(pathLength($path)<= param[length])
} group by ?y
```

parameterized in-degree

$d^{o}_{in<type,length>}(y)$

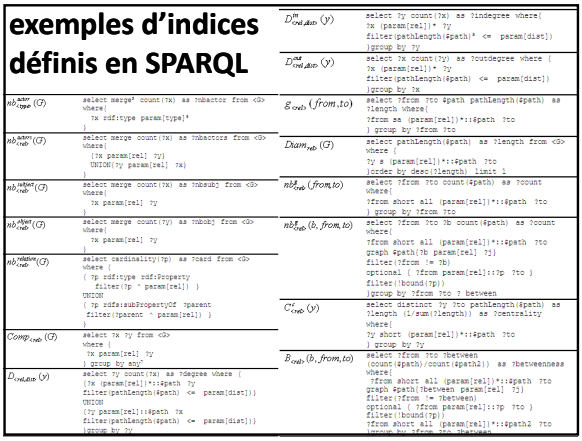


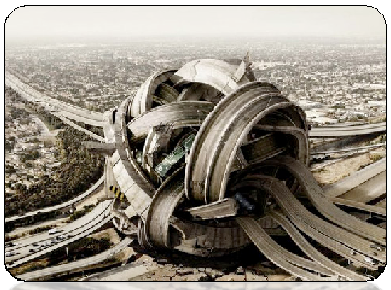
augmenter le graphe des réseaux avec
des résultats d'analyse



distribution en longue traîne de la centralité d'intermédiarité
50 000 projections sur 2020 profils FOAF de flickr.com
[Freeman, 1979]
[PhD Guillaume Erétéo]

exemples d'indices définis en SPARQL





baliser
les autoroutes de l'information
(deuxième point d'accroche, le tag)

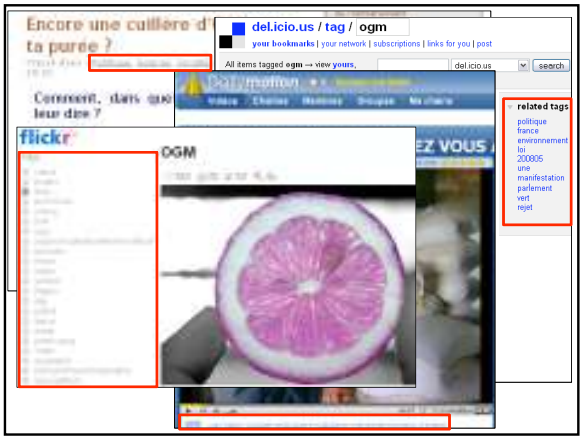
tagging
et autres pratiques web 2.0

un tag

une donnée attachée à un objet

SOCIAL TAGGING

créer et gérer collaborativement des tags pour annoter et catégoriser un contenu



une masse d'utilisateurs créant des catégorisations en masse

folksonomy

folks~taxonomy, a subject indexing systems created within internet communities. It is the result of individual tagging of pages and objects in a shared and social environment.

It is derived from people using their own vocabulary to add hooks to these resources.

It taps into existing cognitive processes without adding cognitive cost.

[Vander Wal, 2005] [Vander Wal, 2007][Rashmi Sinha, 2005]

folksonomies

ne s'opposent pas aux

ontologies

ONTOLOGIE
définition par le contenu

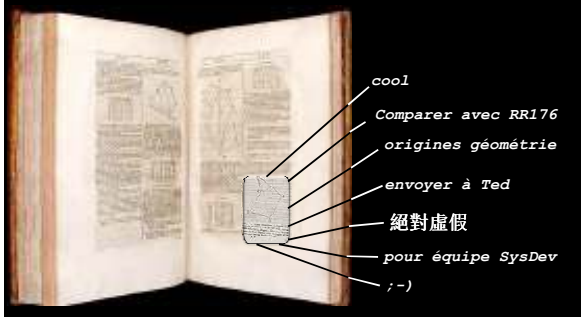
&

FOLKSONOMIE
définition par l'obtention

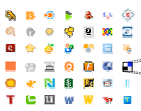
& ontologies
folksonomies

L'ACQUISITION 1.0
ce vieux démon

beaucoup de tags
pour beaucoup d'usages



beaucoup de "sociétés"



mon signet

signet partagé

signet réellement partagé

folksonomies → ontologies contributions…

… [Mika, 2005] hiérarchie par inclusion de communautés.
… [Heymann et al., 2006] hiérarchie centralité graphe Tag-Ressource.
… [Schmitz, 2006] hiérarchie probabilités conditionnelles cooccurrence .
… [Cattuto et al., 2008] [Markines et al., 2009] différentes métriques.
… [Specia et al., 2007] [Begelman et al., 2006] clustering de tags.

variations: métriques & l'espace (tag-ressources-utilisateur).

folksonomies + ontologies contributions…

… [Gruber, 2005] [Tanasescu et al., 2007] taguer les tags

… [Specia et al., 2007][Cattuto et al., 2008][Giannakidou et al., 2008] [Ronzano et al., 2008] [Tesconi et al., 2008] utilisation automatique de ressources linguistiques externes.

… [Good et al., 2007] désambigüisation manuelle par référence à un vocabulaire

… [Passant et al., 2007] désambigüisation manuelle par référence à un thésaurus

… [Huynh-Kim Bang et al. , 2008] tags structurés à la saisie

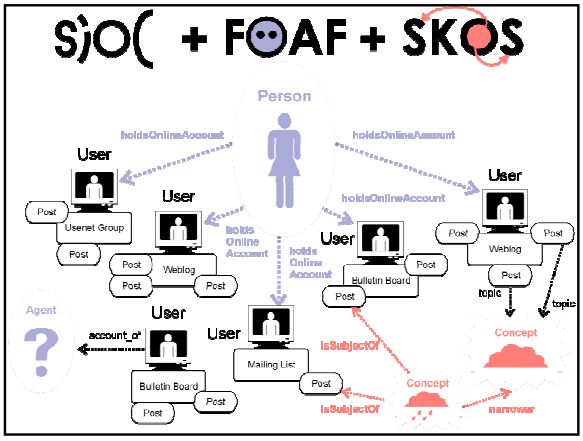
ontologies → folksonomies contributions…

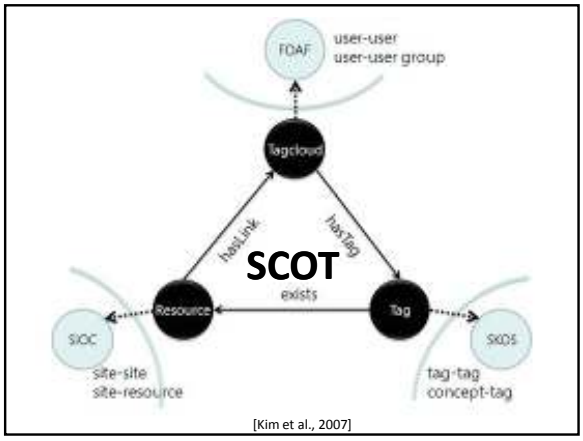
… [Gruber, 2005] [Newman et al., 2005] ontologie de l'acte de tagging

… [Breslin et al., 2005] SIOC ressources échangées dans des sites communautaires

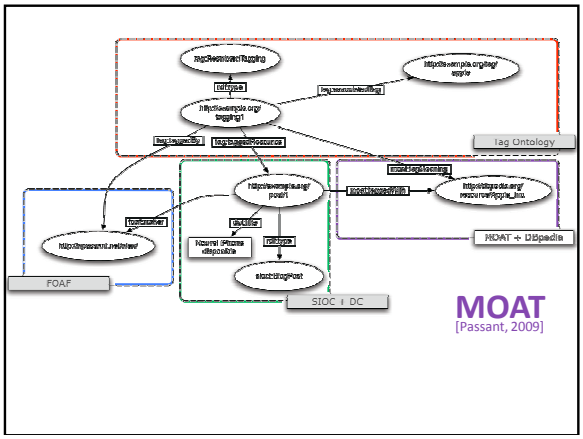
… [Kim et al., 2007] SCOT représentation des tags et de leurs nuages

… [Passant et al., 2008] MOAT, sens associé à un tag





[Kim et al., 2007]

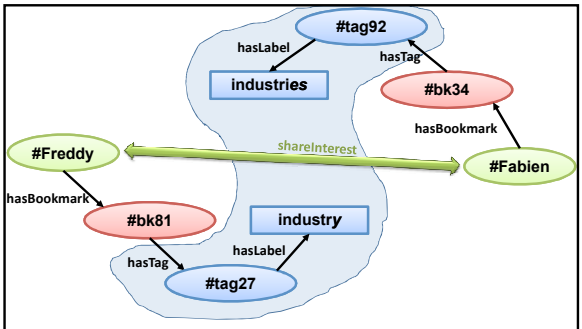


faciliter la connexion

[PhD Freddy Limpens]

ex. capturer les signets et leurs tags

co-tags extraits de del.icio.us pour “ademe”
6054 signets, 16 utilisateurs, 5153 tags, 5969 ressources



global giant graph
lier utilisateurs, actions, connaissances, groupes, etc.

signet social
maintenant, au tour des autres données…

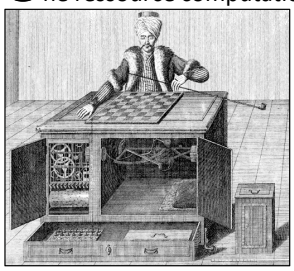
L'UTILISATEUR

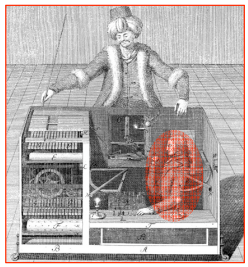
dans la boucle, certes,

mais la nôtre

UTILISATEUR 2.0

une ressource computationnelle





penser les solutions dans leur

ENSEMBLE

CAPTCHA

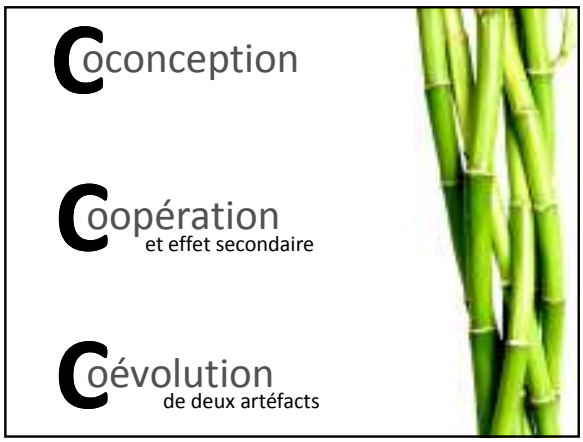
Completely Automated Public Turing test to Tell Computers and Humans Apart

RELAIS

passer la main à l'humain

How to Play
game
ESP

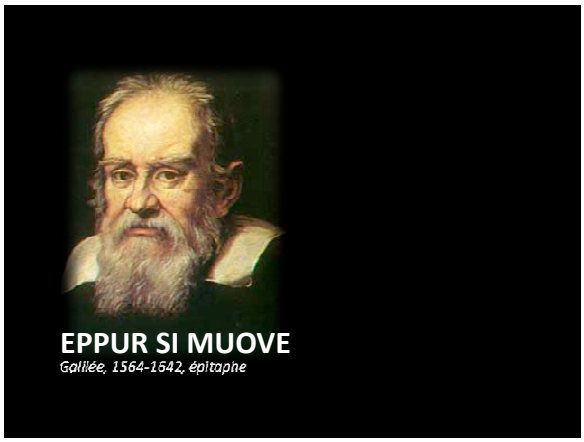
premiers essais
ONTOGAMES


Coconception
Coopération
et effet secondaire
Coévolution
de deux artéfacts

l'enjeu et l'importance de la
MOTIVATION

LE SENS
de la masse

Qu'est-ce qui gravite autour de la Terre ?
qui veut gagner des millions d'
ONTOLOGIES

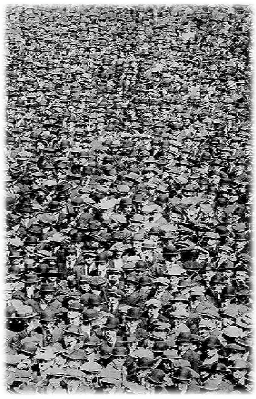


CONTRÔLE

statistiques

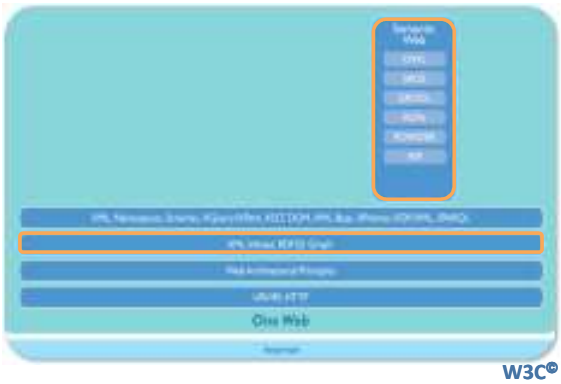
recoupements

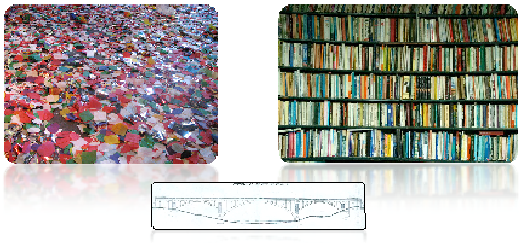
pot de miel



PASSÉ, PRÉSENT, FUTUR

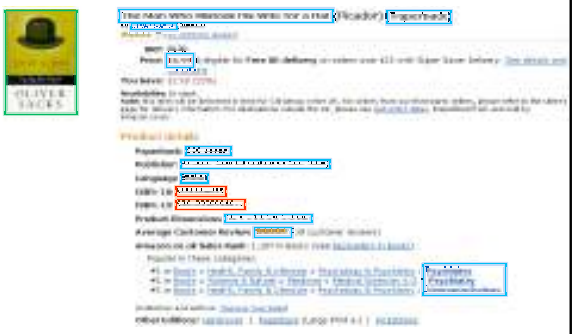
"An important aspect of the larger unit is that it contains computational elements (persons) who cannot be described entirely in computational terms." (Hutchins, 1995)





passerelles entre le web sémantique et les 'autres webs': RDFa, GRDDL

beaucoup de données
se terrent dans les pages



RDFa signifie
RDF dans des **attributs** de HTML

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/">
 <h2 property="dc:title">The trouble with Bob</h2>
 <h3 property="dc:creator">Alice</h3>
 ...
</div>
```

µFormat

des classes de style au balisage de données

```
<span class="tel">06 74 80
```

GRDDL ouvre des formats

en permettant de déclarer des mécanismes d'extraction de RDF à partir de XML

```
<head profile="http://www.w3.org/2003/g/data-view">
 <title>The man who mistook his wife for a hat</title>
 <link rel="transformation"
  href="http://www.w3.org/2000/06/ dc-extract/dc-extract.xsl" />
 <meta name="DC.Subject" content="clinical tales" />
 ...
```

µformats

web sémantique

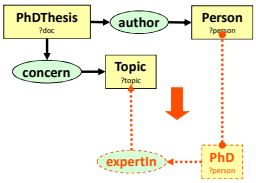
bienvenue
dans un web de données

RIF: Rules Interchange Format

– format d'échange de règles sur le web

– interopérabilité des systèmes à base de règles

– règles pour le web sémantique



POWDER... Protocol for Web Description Resources

plusieurs ressources

listes

domaines, chemins

expressions régulières

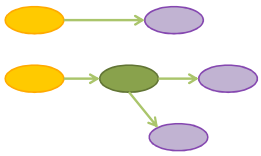
description authentifiée

exemple de description POWDER

toute ressource dont l'URI commence par http://www-sop.inria.fr/fr/ concerne le centre de recherche de l'INRIA géographiquement situé à Sophia Antipolis et est en français

recettes et bonnes pratiques

SWBPDWG



quelques
messages

impossible
de prédire tous
les usages

éviter de construire une
boite noire

rendre les conceptualisations
explicites

ouvrez vos données
à ceux qui pourraient les utiliser

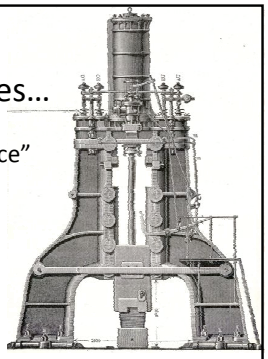
capturer l'explicite
pour ne pas avoir à extraire, fouiller, etc.

- Dublin core
- Creative Commons
- FOAF

...

à base d'ontologies…

n'est pas synonyme de
"avec moteur d'inférence"



[C. Welty, ISWC 2007]

"semantic **web**"

and not

"**semantic** web"

[J. Hendler, ISWC 2007]

"a lightweight ontology allows us to do lightweight reasoning"

le piège du domaine

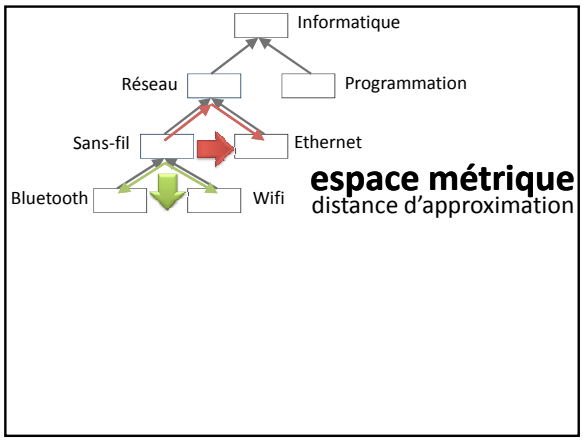
le domaine d'application n'est pas nécessairement le domaine de modélisation

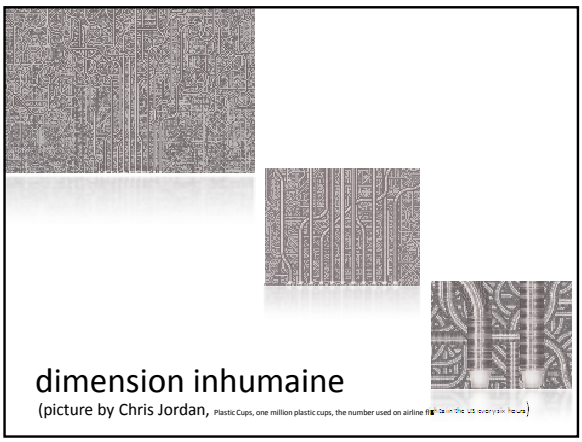
je n'ai jamais vu une

ontologie universelle

ma montre n'a qu'une aiguille
mais elle n'est pas cassée.

de l'intérêt d'un à peu près



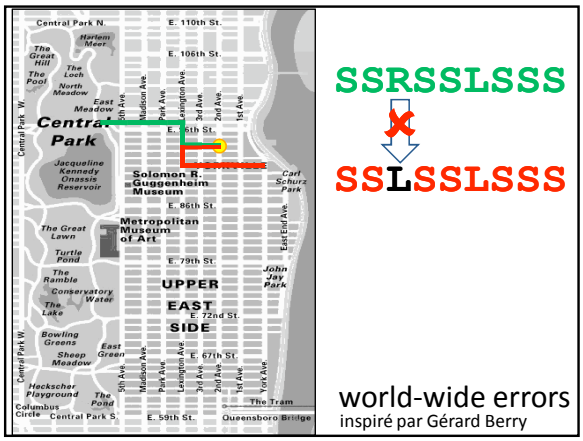


protéger les enfants
étiquettes dans les pages
émises par des autorités

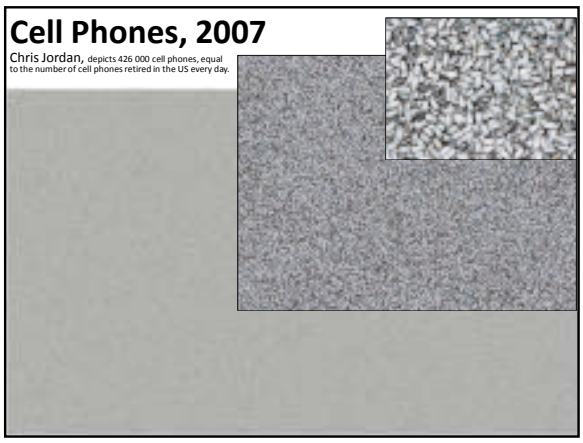
P3P Platform for Privacy Preferences
politiques des sites vs. préférences utilisateurs

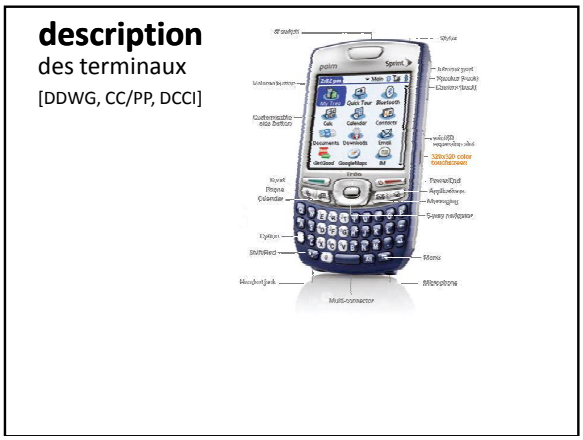
XML Security
clefs, crypto, etc.

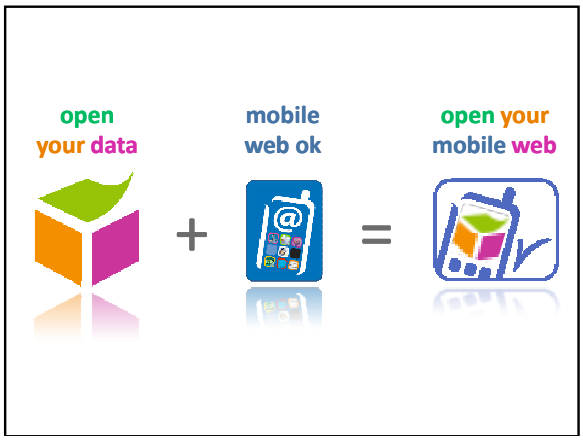
traçabilité
ex: graphes nommés.


SSRSSLSSS
SSLSSLSSS
world-wide errors
inspiré par Gérard Berry

imap-sop.inria.fr
Web Security Context
contexte de sécurité sur le Web
informer de façon ergonomique
de l'état et des changements


Cell Phones, 2007
Chris Jordan, depicts 426 000 cell phones, equal to the number of cell phones retired in the US every day.


description
des terminaux
[DDWG, CC/PP, DCCI]


open
your data
mobile
web ok
open your
mobile web

W3C en marche…

titres des 72 papiers de l'atelier

(*sauf* social, network, mobile, position, paper, web, future)

http://www.w3.org/2005/Incubator/socialweb/

Social Web Incubator Group Charter

The mission of the Social Web Incubator Group, part of the Incubator Activity, is to understand the systems and technologies that permit the description and identification of people, groups, organizations, and user-generated content in extensible and privacy-respecting ways.

The topics covered with regards to the emerging Social Web include, but are not limited to: accessibility, internationalization, portability, distributed architecture, privacy, trust, business metrics and practices, user experience, and contextual data. The scope includes issues such as widget platforms (such as OpenSocial, Facebook and W3C Widgets), as well as other user-facing technology, such as OpenID and OAuth, and mobile access to social networking services. The group is concerned also with the extensibility of Social Web descriptive schemas, so that the ability of Web users to describe themselves and their interests is not limited by the imagination of software engineers or Web site creators. Some of these technologies are independent projects, some were standardized at the IETF, W3C or elsewhere, and users of the Web shouldn't have to care. The purpose of this group is to provide a lightweight environment designed to foster and report on collaborations within the Social Web-related industry or outside which may, in due time affect the growth and usability of the Social Web, rather than to create new technology.

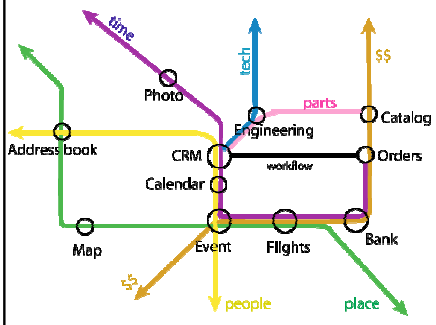
Our goal is to provide a forum through which collaborations relating to social web standards can be formed, and through which the results of practical standards-oriented collaborations can be reported and discussed. This is not a Working Group, although the members of the group are free to undertake work together, including and especially work outside the W3C, and to report it and discuss it within the group and the wider W3C.

- réseaux sociaux d'entreprise
- veille technologique et économique
- communautés d'intérêt, de pratique, d'experts

isicil.inria.fr

un web de communautés multiples

rôles, connaissances, services, contextes



Image,T. Berners-Lee, RuleML kickoff, 2005

unifier les identités

integration & consolidation données

owl:sameAs

=

Fabien

Bafien

owl:differentFrom

≠

Magic Hagrid

Semantic Hagrid

- inversement fonctionnel
- le pour vs. le contre

identité fragmentée

les $\neq^{nt}$ « moi » au cours du temps

chantiers

- sécurité & confidentialité (sémantique)
- traçabilité & qualité, confiance
- *ergonomie, conception d'interaction*
- l'égalité et légalité
- …

web à venir …

… un media modifiable, dynamique, et fluctuant
… une ressource ubiquitaire
… un web de données et d'applications connectées
… programmation orientée services et modèles
… une immense machine virtuelle
… paradigmes de programmation abstraits & déclaratifs
… un lieu de rencontre: présence, partage, collaboration

pour gérer une diversité

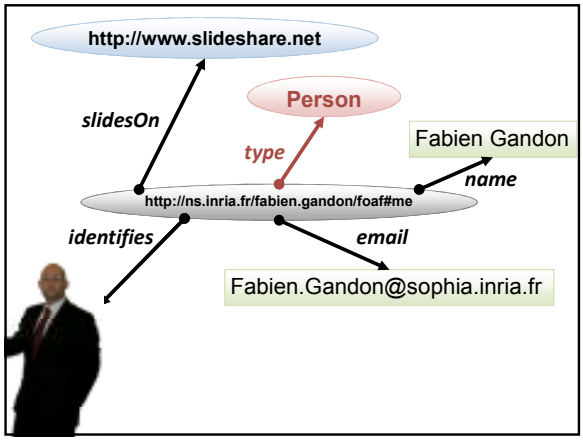
rien de tel que d'utiliser une autre diversité

la diversité des métadonnées

permet de gérer les diversités des ressources, de leurs index et de leurs utilisations

métadonnées et les standards sont des armes des passages à l'échelle

que ce soit du point de vue du nombre des ressources, de l'hétérogénéité des représentations, de la foule des utilisateurs, de la diversité des matériels ou de la multiplication des applications et des services.



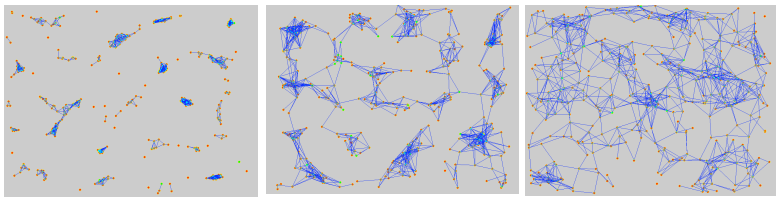
SESSION 7

Simulation et agents

Simulation multi-agent de réseaux sociaux : des modèles classiques à leur application au Web 2.0

F. AMBLARD
(UNIVERSITÉ DE TOULOUSE, FRANCE)

Simulation multi-agent de réseaux sociaux



Frédéric AMBLARD
IRIT-SMAC
Université Toulouse 1 Sciences Sociales

Plan

- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de dynamiques sur réseaux
  - Modèles de réseaux dynamiques
  - Co-évolution
- Expérimentations
- Perspectives

Plan

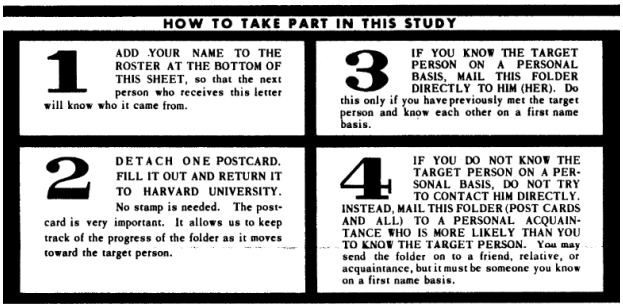
- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de dynamiques sur réseaux
  - Modèles de réseaux dynamiques
  - Co-évolution
- Expérimentations
- Perspectives

Réseaux sociaux & Internet

- Thématique interdisciplinaire
  - Sociologie
  - Sciences de l'Information
  - Informatique
  - Physique Statistique
  - Economie
- Ici, réseaux sociaux = sociabilité sur Internet
- Focalisation
  - Sur le comportement des individus
  - Sur les phénomènes sociaux émergents
  - Sur l'influence de/sur la structure sociale sur ces phénomènes

Phénomènes identifiés: l'effet Small World

- Expérience de Milgram



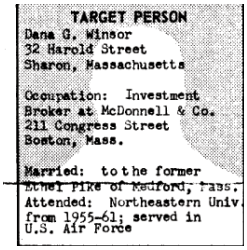
HOW TO TAKE PART IN THIS STUDY

1 ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

2 DETACH ONE POSTCARD. FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

3 IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.

4 IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST CARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder on to a friend, relative, or acquaintance, but it must be someone you know on a first name basis.



TARGET PERSON
Dana G. Winsor
32 Harold Street
Sharon, Massachusetts

Occupation: Investment Broker at McDonnell & Co.
211 Congress Street
Boston, Mass.

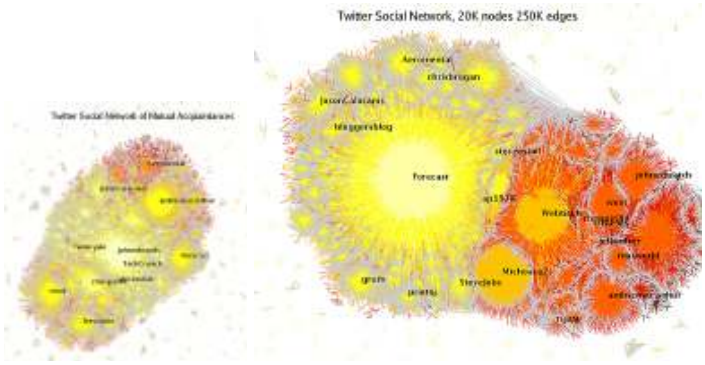
Married: to the former Ethel Pike of Medford, Mass.
Attended: Northeastern Univ. from 1955-61; served in U.S. Air Force

- Résultats : Chemins courts entre individus
  - six degrees … (quand les cartes postales reviennent…)

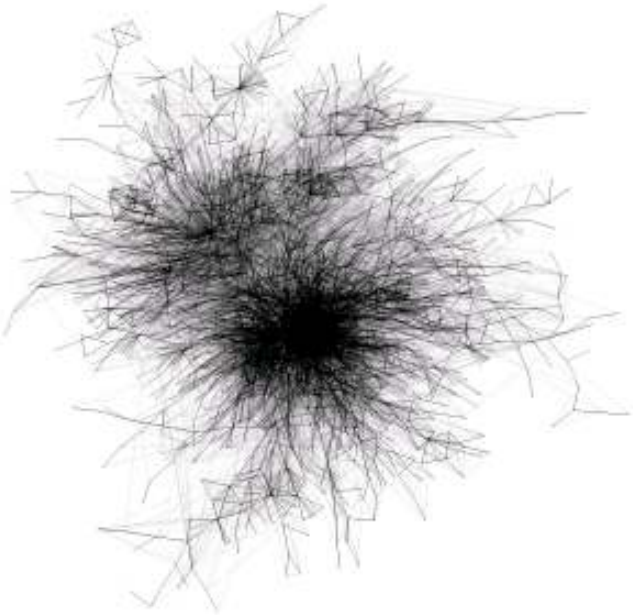
http://www.findsatoshi.com/

Rich get richer …

- Scale-free networks (Barabasi et al.)
- Plus qu'un effet d'ancienneté, phénomène de réputation
- Dynamique auto-renforçante
- … mais des basculements … phénomène Google

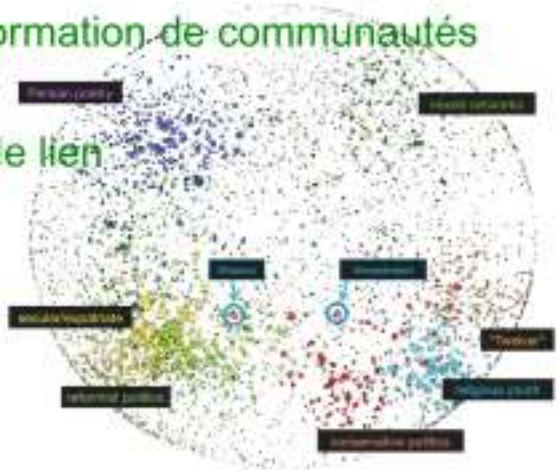


Structures Core / Periphery



Le Web communautaire ou communautariste

- Emergence / Formation de communautés
- Clustering
- Redondances de lien



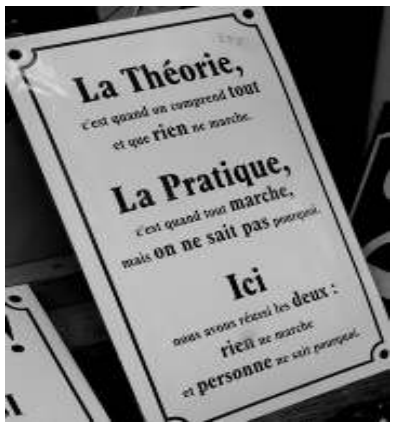
- Effets :
  - les amis de mes amis sont mes amis
  - Homophilie
  - Intérêts / informations liés

Diffusion d'information

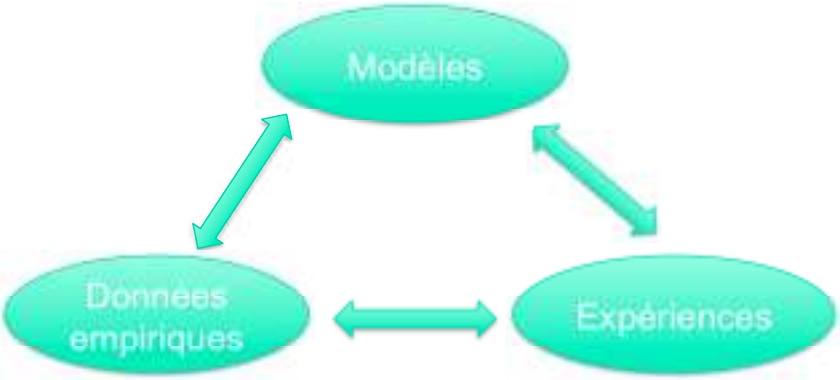
- Diffusion rapide
  - grippe porcine, Mickael Jackson
- Effets d'amplification (buzz) ou pas
- Phénomènes de régulation
- Compétition entre information (hoax Pentagone)
- Pensée unique ?

Bilan rapide…

- On sait quoi (approche descriptive)
- On ne sait pas comment ni pourquoi (approche compréhensive)
  - On ne parvient pas à anticiper
  - On ne sait pas quels sont les leviers



Organisation synthétique du programme de recherche



Plan

- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de dynamiques sur réseaux
  - Modèles de réseaux dynamiques
  - Co-évolution
- Expérimentations
- Perspectives

Données empiriques

- Large corpus de données
  - Internet
  - Email networks
  - Web
  - Blogs
  - Twitter
  - Facebook
  - P2P networks
  - …

Données empiriques limitations

- Bruitées
  - Relativement
- Signification de la relation sociale
  - Pour les utilisateurs
    - Un contact facebook est-il un ami ?
    - Un follower twitter est-il un ami ?
    - Quel contenu j'accorde aux relations via Internet en général ?
    - A chacune en particulier
  - Pour les scientifiques
- Multiplexité
- Dynamique

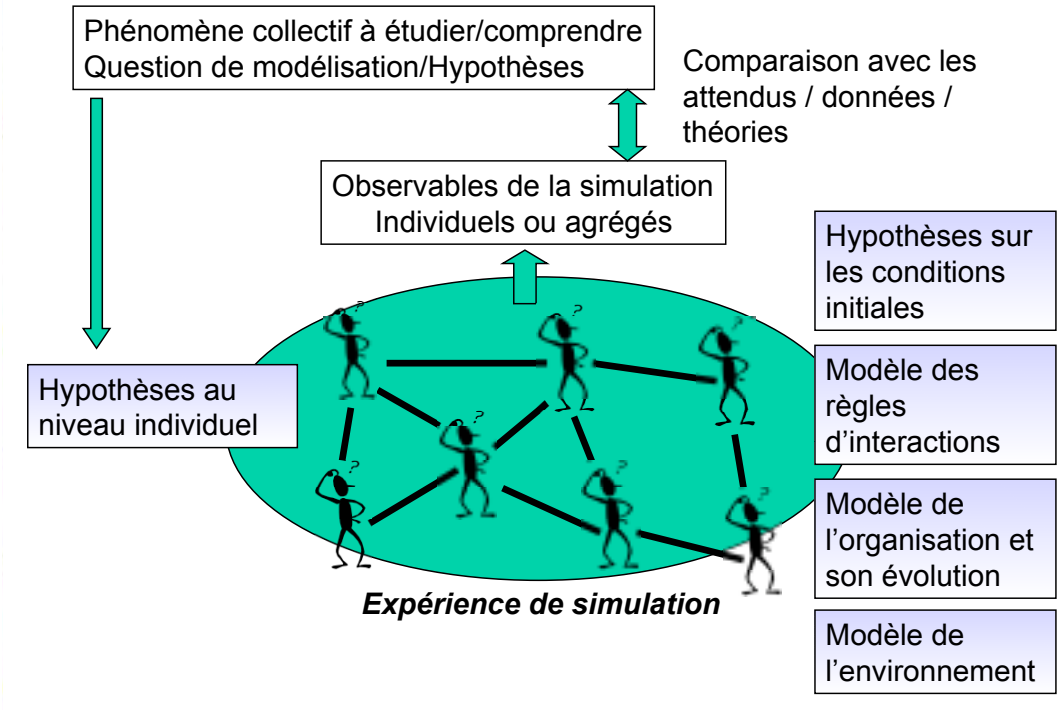
Plan

- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de dynamiques sur réseaux
  - Modèles de réseaux dynamiques
  - Co-évolution
- Expérimentations
- Perspectives

Quelle approche de modélisation ?

- Modèle statistique
  - Pas de représentation explicite des mécanismes
  - Utilisation : analyse de données, …
- Modèle macroscopique
  - Ex: équations différentielles, modèles à compartiment
- Modèle multi-agent
  - Représentation explicite des comportements individuels
  - A la recherche des mécanismes générateurs

Modèle multi-agents



Principales propriétés des simulations multi-agents

- Approche individus-centrée
  - individualisme méthodologique
- Simulation
  - le modèle plongé dans le temps
- Formalisation algorithmique souple

Intérêts de l'approche multi-agents

- Formalisation
  - Formalisme moins contrainte qu'une formalisation mathématique
  - Cependant Hypothèque Numérique
- 3eme jambe entre théorie et empirie
- Computation d'un système artificiel difficilement appréhendable sans l'ordinateur
  - Comportements individuels
  - Interactions entre agents
  - Hypothèses dynamiques

La reconstruction comme source de compréhension

- Homo Faber (Hannah Arendt)
  - Reconstruire la nature pour la comprendre
- Logo (Seymour Papert)
  - Reconstruire la géométrie
- StarLogo (Mitchell Resnick)
  - Reconstruire des systèmes distribués pour comprendre l'émergence de régularités globales
    - Fourmilière
    - Termites
    - Traffic

20

La reconstruction comme source de compréhension

- Quels comportements individuels génèrent un phénomène collectif observé ?
  - Approche par essai/erreur, visualisation de la simulation => est-ce que je suis dans la bonne direction
  - Ex: Regroupement des morts chez les fourmis en dehors de la fourmilière
- Approche spéculative : vers quel état peut tendre un système avec les règles que j'ai spécifié ?
  - Test d'hypothèses, de scénarios…

21

Applications possibles

- Artefact pour la réflexion
  - Par la formalisation
  - Par la comparaison des attendus et des résultats
  - Par la compréhension précise des mécanismes de production des résultats au niveau du modèle
- Test qualitatif d'hypothèses
  - Aménagement du territoire
  - Gestion de l'eau
- Artefact pour la négociation
- Formation/entraînement
- Prédiction (…???)

Plan

- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de dynamiques sur réseaux
  - Modèles de réseaux dynamiques
  - Co-évolution
- Expérimentations
- Perspectives

Modèles

- Modèles de réseaux
- Modèles de dynamiques sur réseau
- Modèles de réseaux dynamiques
- Co-évolution

Modèles de réseaux

- Réseaux aléatoires (Erdos-Renyi)
- Réseaux construits à partir d'une distance
- Réseaux Small-world (Watts Strogatz)

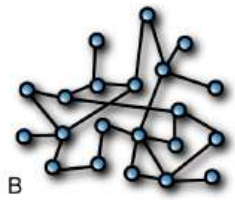
L'hypothèse nulle: le réseau aléatoire

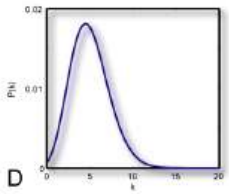
- Modèle d'Erdos-Rényi
- Une population d'individus (*N* sommets)
- Une connectivité moyenne souhaitée
  - Ex: *k*= 2.3 liens / indiv
- On tire *N.k/2* liens (a,b)
  - En tirant aléatoirement a
  - Puis b (en s assurant que b <> a)
  - On créé le lien (a,b)
    - En s'assurant qu'il n'existe pas déjà (ni (b,a))

26

Le réseau aléatoire

- Connectivité
  - Il existe un seuil (cf. percolation) en-dessous duquel le graphe obtenu n'est pas connexe et au-dessus duquel il est connexe
    - $T_c>1$ …
    - Cf Netlogo Giant Component





27

Graphes construits à partir d'un espace social

- Hypothèse d'homophilie
  - Plus on est semblable, plus on a tendance à être connecté
- La population est distribuée dans l'espace des paramètres
  - Une dimension par paramètre
  - Dimensions continues/discrètes
  - Signification de cet espace social
- Calcul d'une proximité entre individus
  - Par ex. distance euclidienne dans l'espace construit
  - Cas de la pondération des dimensions
  - Cas des dimensions complémentaires

28

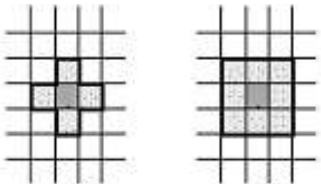
Graphe construit à partir d'un espace social

- A partir de la distance entre les individus
  - Renormalisation (pour l'assimiler à une probabilité)
- 2 possibilités
  - On active les relations en fonction de leur probabilité
    - Modèle statistique *p1* [Holland et Leinhardt]
  - On tire exactement le nombre de liens désiré en fonction de la probabilité de tirer chacune des relations

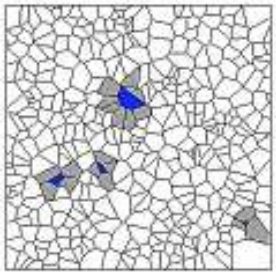
29

Graphe construit à partir d'un espace social

- Pavage de l'espace
  - Chaque individu est connecté à ses plus proches voisins
  - Si répartition régulière
    - Grille régulière



  - Sinon diagramme de Voronoi
    - Pavage irrégulier ex: 2D



30

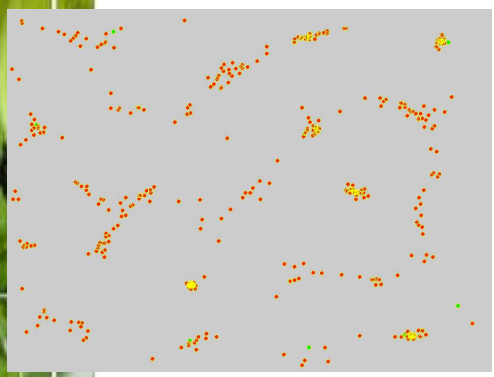
Exemple : reconstruction d'un réseau d'agriculteurs

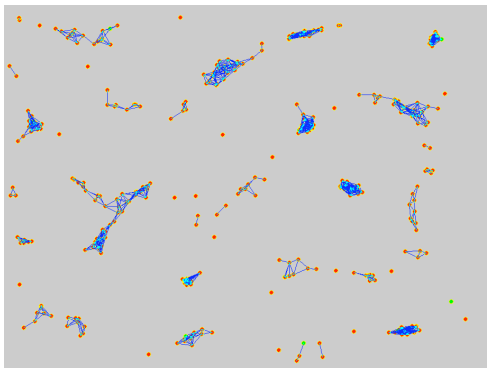
- Trois types de réseaux
  - Voisinage, professionnel, aléatoire
- Variables
  - n_l, d_l nb liens locaux, distance max locale
  - n_p, $\pi(i,j)$ nb liens professionnels, classe d'équivalence
  - n_r nb liens aléatoires

31

Liens de voisinage

- Définis à partir d'une distance max dans l'espace 2D
  - Le réseau suit la répartition spatiale

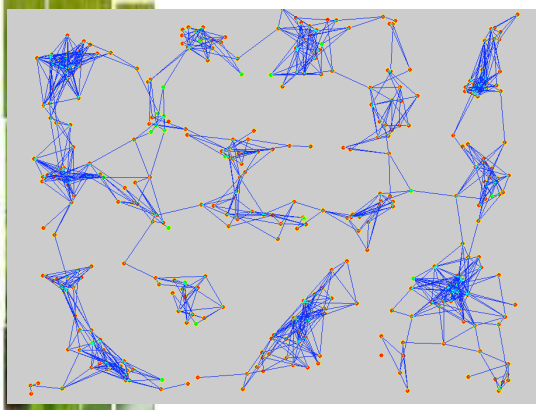


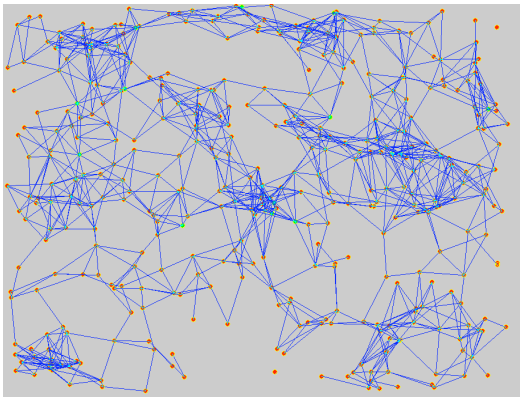


32

Réseau de voisinage

- Si on augmente la distance max …
  - Si dist max = infini ou un peu moins, réseau aléatoire





33

Breadalbane (UK)

34

Répartition spatiale initiale

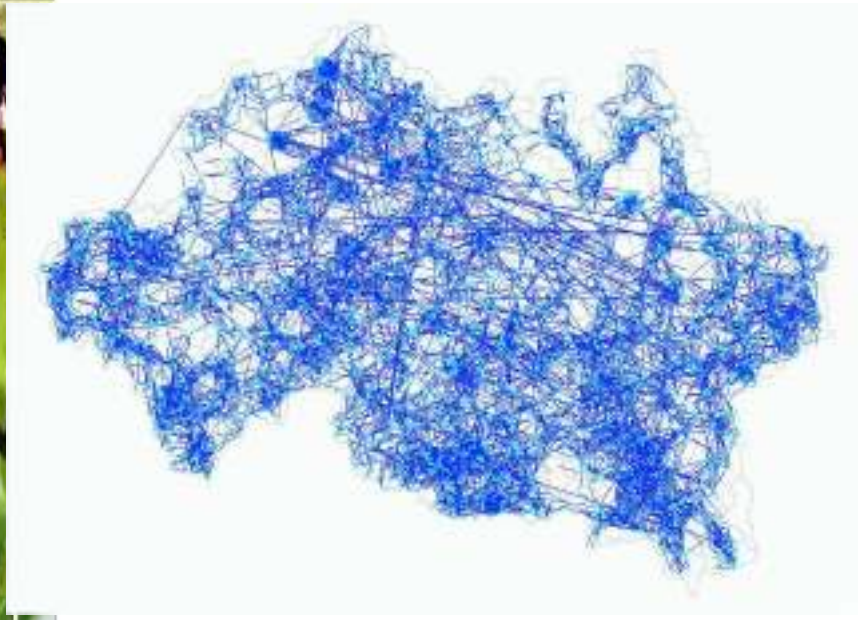
Réseau social construit sur cette distribution

On ajoute liens professionnels

…et liens aléatoires

36

Allier (Fr)



37

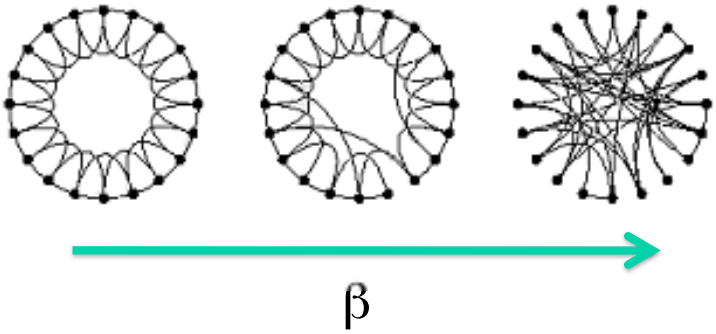
Graphes Small world (Watts)

- Empiriquement certains réseaux semblent avoir des diamètres faibles (ou des plus courts chemins courts) et un fort clustering (communautés, groupes sociaux)
- Ils ne sont ni aléatoires, ni réguliers

38

Modèle β

- β model => [ring => random graph]
- Substrat 1-lattice avec k voisins
- On rebranche les liens avec la proba β



39

Algo modèle β

- Chaque sommet *i* est choisit à tour de rôle, ainsi que la relation à son plus proche voisin dans le sens des aiguilles d'une montre
- Si random < β on enlève le lien et on tire j au hasard => création du lien (i,j)
- Cf Netlogo Small-world

40

β : propriétés length & clustering

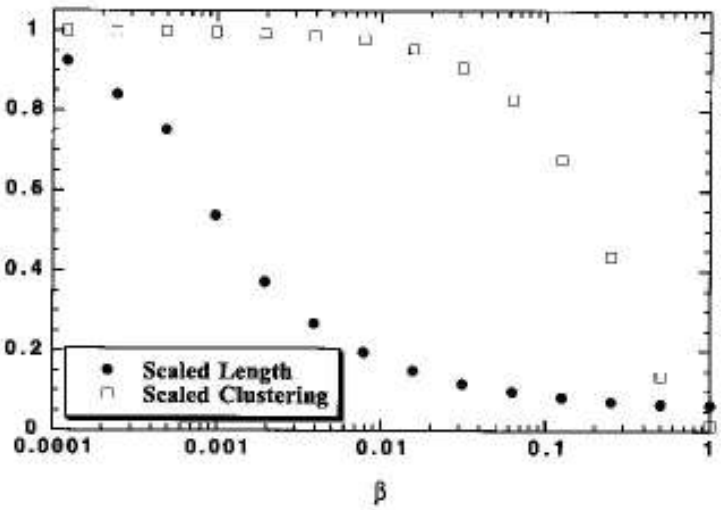


Figure 3.22 Scaled $\gamma(\beta)$ and $L(\beta)$ for β-graphs averaged over twenty realisations ($n = 1{,}000$, $k = 10$).

41

Plan

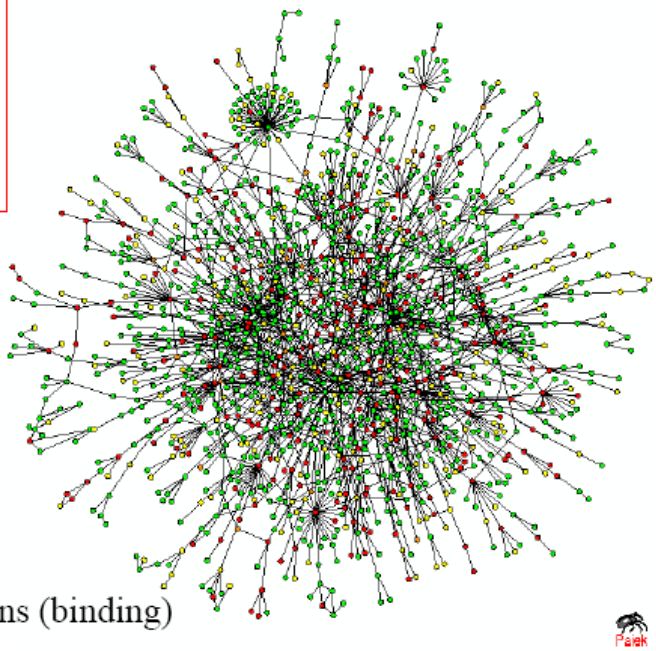
- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de réseaux dynamiques
  - Modèles de dynamiques sur réseaux
  - Co-évolution
- Expérimentations
- Perspectives

Modèles de réseaux dynamiques

- Phénomène scale-free networks
- Modèle d'attachement préférentiel

Protein Interaction network (PIN)

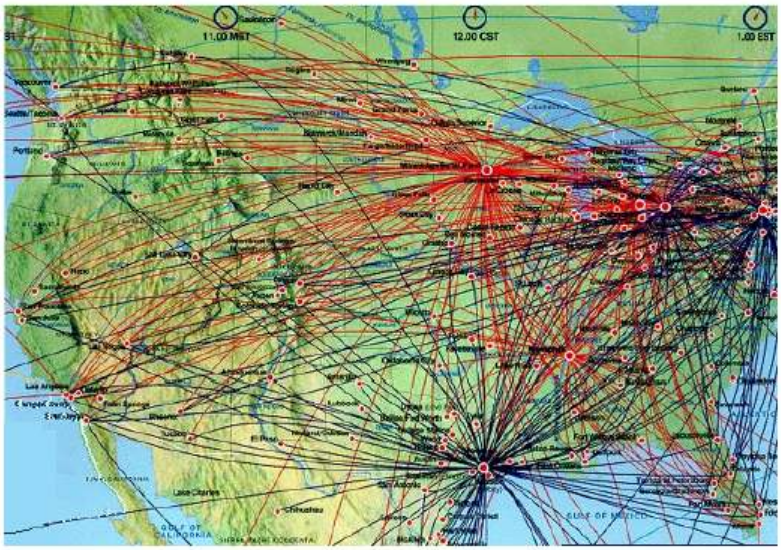
A protein interaction network is the set of binary interactions among the proteins of a given proteome



Nodes: proteins

Links: physical interactions (binding)

Airplane routes network

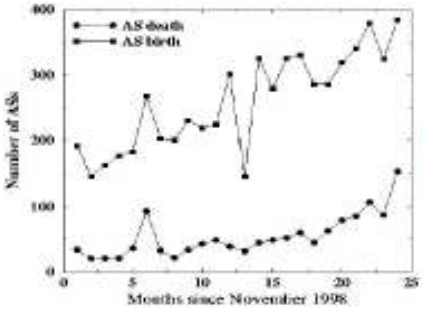


The Physical Internet

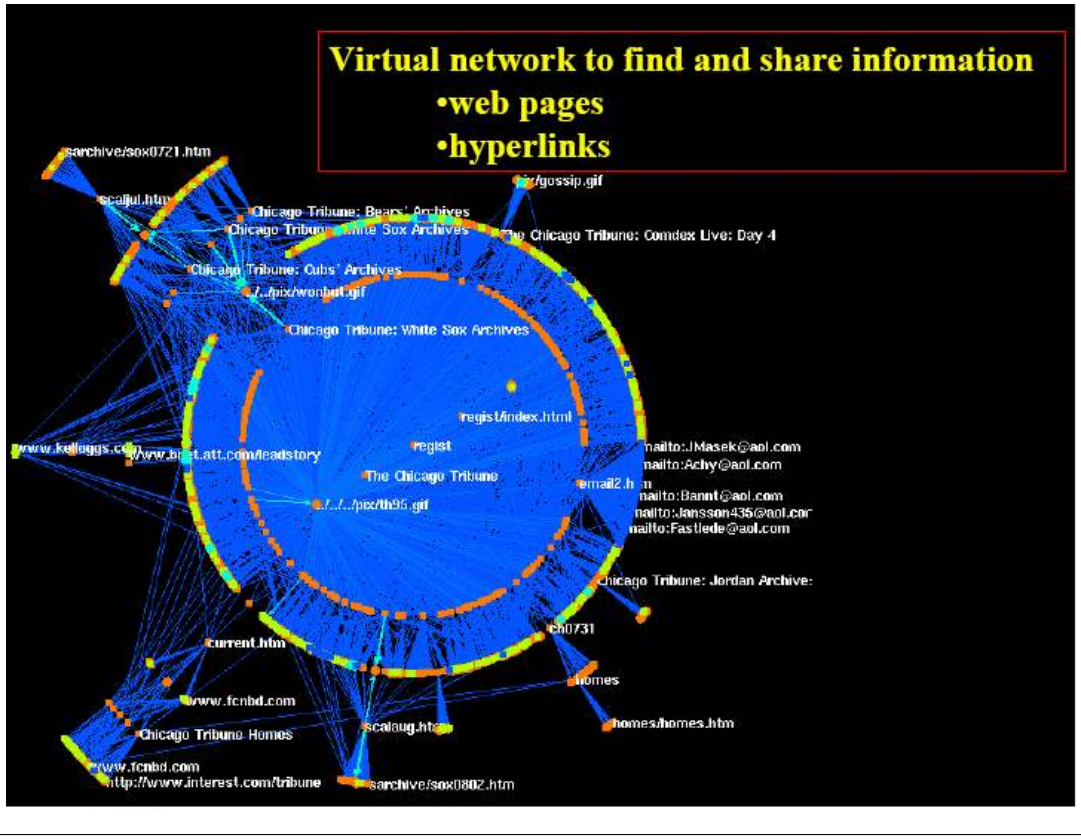
- Computers (routers)
- Satellites
- Modems (??)
- Phone cables
- Optic fibers
- EM waves

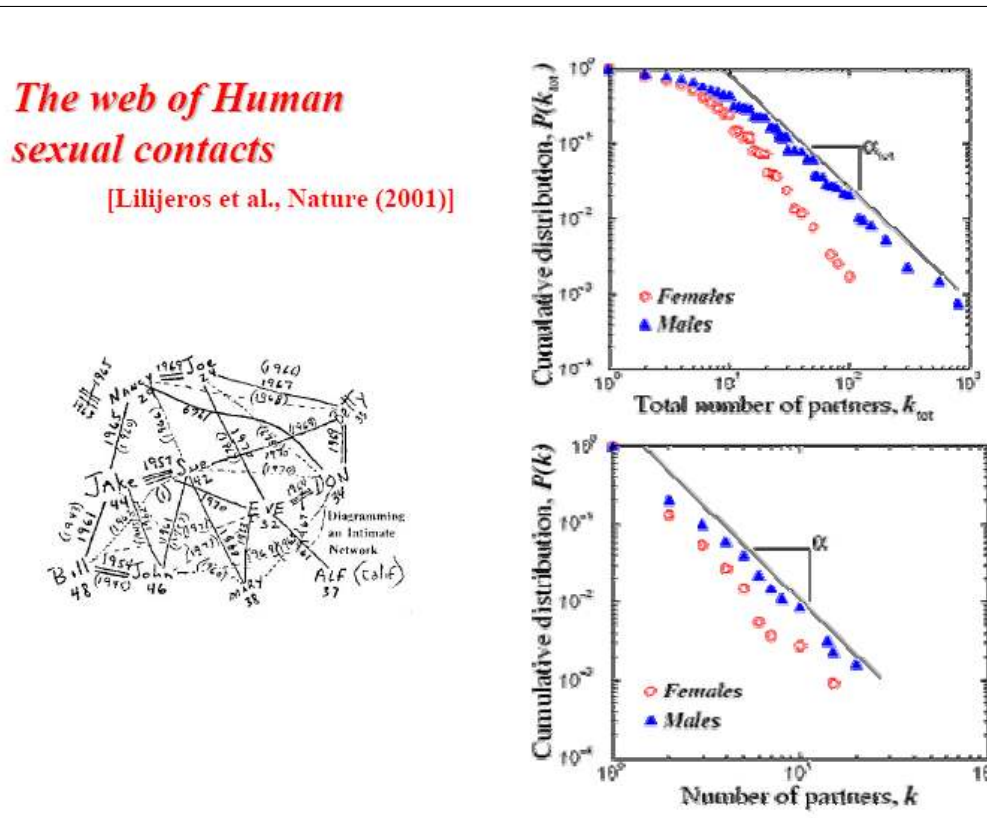
→ Technological Heterogeneity

- Système en évolution

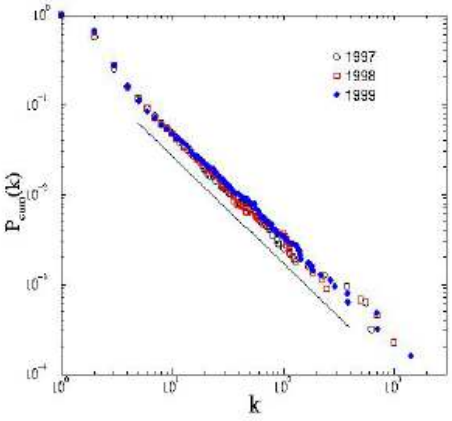


46


Virtual network to find and share information
•web pages
•hyperlinks
archive/sox0721.htm
scaljul.htm
Chicago Tribune: Bears' Archives
Chicago Tribune: White Sox Archives
The Chicago Tribune: Comdex Live: Day 4
Chicago Tribune: Cubs' Archives
Chicago Tribune: White Sox Archives
regist/index.html
regist
The Chicago Tribune
www.kelloggs.com
www.bnet.att.com/headstory
mailto:JMasek@aol.com
mailto:Achy@aol.com
mailto:Bannt@aol.com
mailto:Jansson435@aol.com
mailto:Fastlede@aol.com
Chicago Tribune: Jordan Archive:
current.htm
homes
www.fcnbd.com
Chicago Tribune Homes
scalaug.htm
homes/homes.htm
www.fcnbd.com
http://www.interest.com/tribune
archive/sox0802.htm


The web of Human sexual contacts
[Lilijeros et al., Nature (2001)]
Diagramming an Intimate Network
Cumulative distribution, $P(k_{tot})$
Females
Males
Total number of partners, k_{tot}
Cumulative distribution, $P(k)$
Females
Males
Number of partners, k

Degree distribution



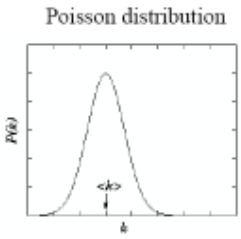
$P(k)$ = probability that a node has k links

•**Router level & AS level**

$P(k) \sim k^{-\gamma}$ $(2 < \gamma \leq 3)$

- $\langle k \rangle$ = const
- $\langle k^2 \rangle \rightarrow \infty$

Rappel: réseau aléatoire



49

Scale-free networks

- Modèle d'attachement préférentiel
  - Au départ : un noyau (3 nœuds liés)
  - On ajoute progressivement de nouveaux nœuds
  - A chaque nouveau nœud on rajoute un nombre de liens à partir de celui-ci
    - Aléatoire
    - Fixe
  - Les liens sont dirigés préférentiellement vers les nœuds les plus connectés

50

Modèle d'attachement préférentiel

- Au départ, répartition uniforme des liens
- Puis petites différences
- Le processus accroît les différences progressivement

- Cf Netlogo Preferential Attachment

- Extensions : ageing, fitness, rewiring (Dorogovtsev, Mendes…)

51

Plan

- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de réseaux dynamiques
  - Modèles de dynamiques sur réseaux
  - Co-évolution
- Expérimentations
- Perspectives

Modèles de dynamiques sur réseaux

- Modèle de percolation sociale
- Dynamique d'opinions sur réseau

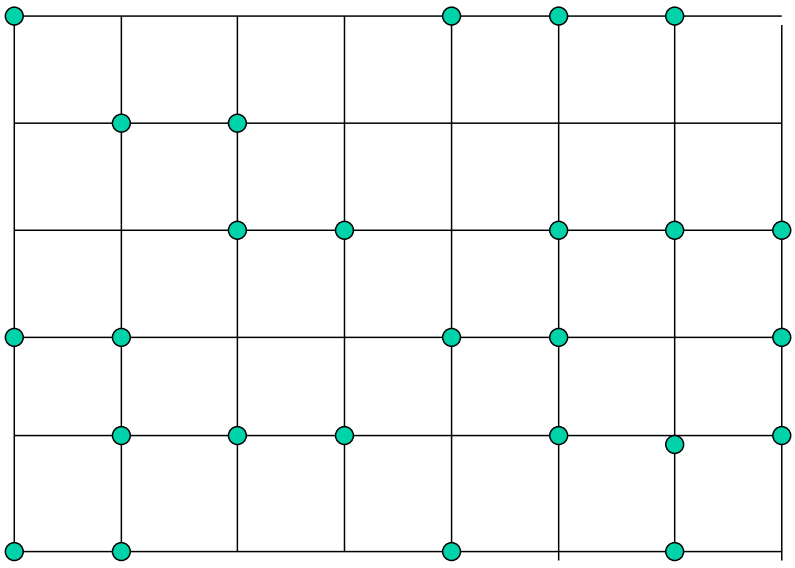
Modèle de percolation sociale

S. Solomon, G. Weisbuch, L. de Arcangellis, N. Jam, D. Stauffer, 2000. « Social Percolation Models », *Physica A*, vol.277, (1-2), pp.239-247.

Modèle de percolation sociale

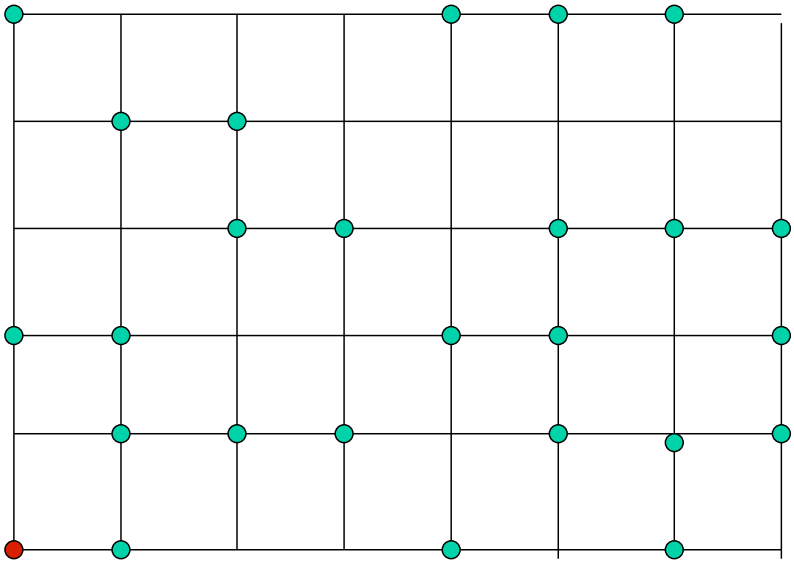
- Modélisation de diffusion d'innovations
- Sur un réseau (grille)
  - Sur chaque cellule: un agent *i* peut communiquer une information à propos de la qualité du film *q* à ses plus proches voisins
  - En fonction de l'influence de ses voisins, l'agent décide d'aller voir le film (ou pas) si $q>p_i$ (pi = préférence personnelle)
- Conditions initiales
  - Une graine (quelques agents ont vu le film)
  - P_i distribué uniformément entre 0 et 1
- Percolation en physique
  - Le film marche si $q>p_c$ seuil de percolation

55

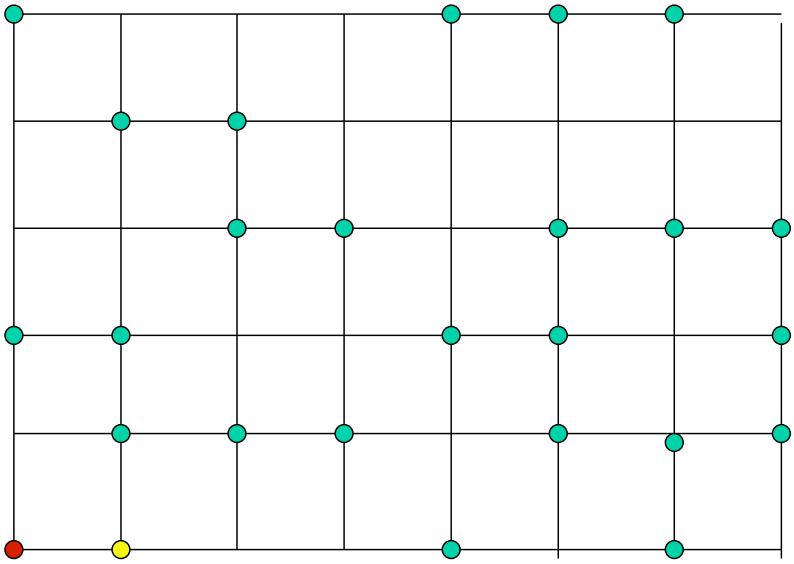


Density des adopteurs potentiels : 26 /48 > 50%

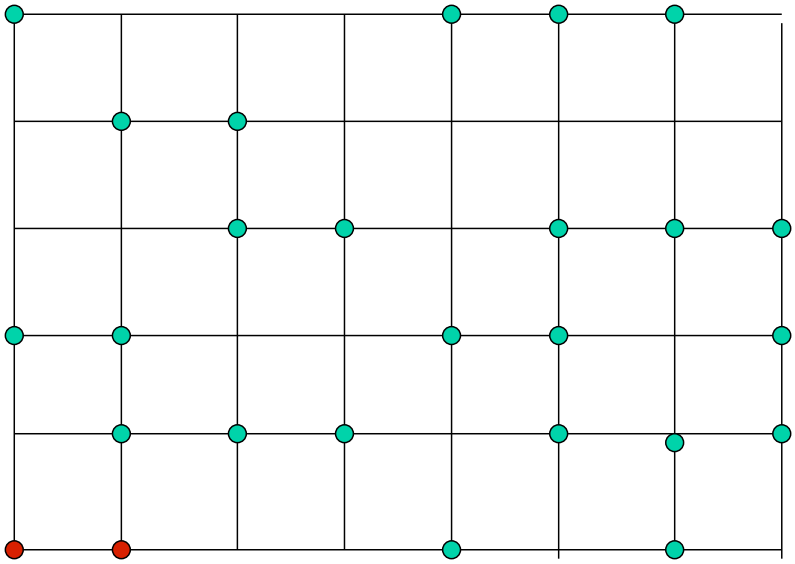
56



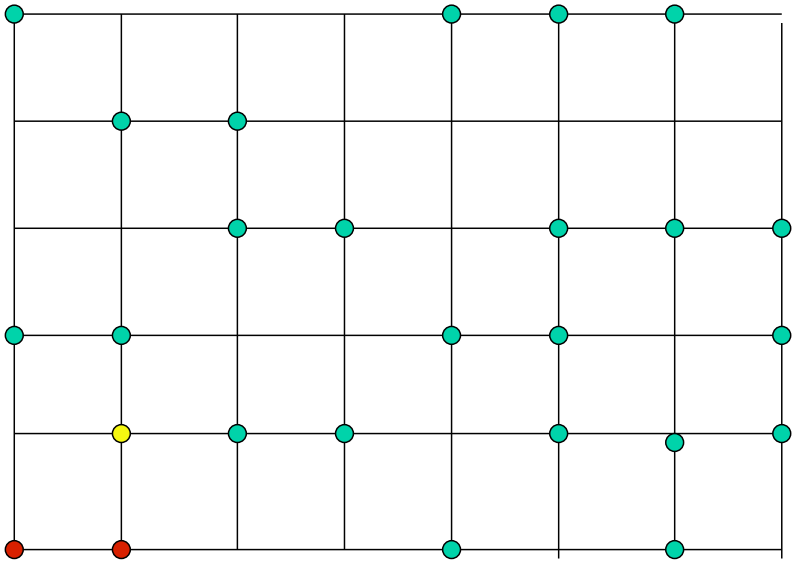
57



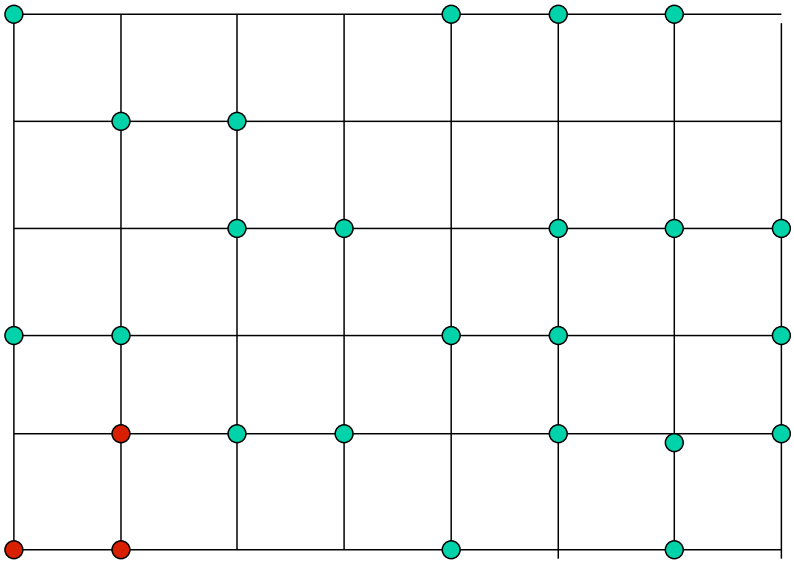
58



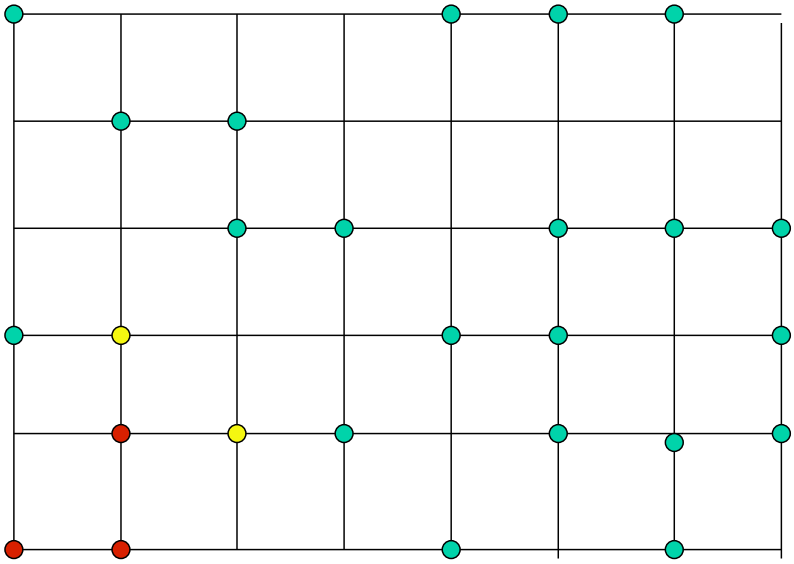
59



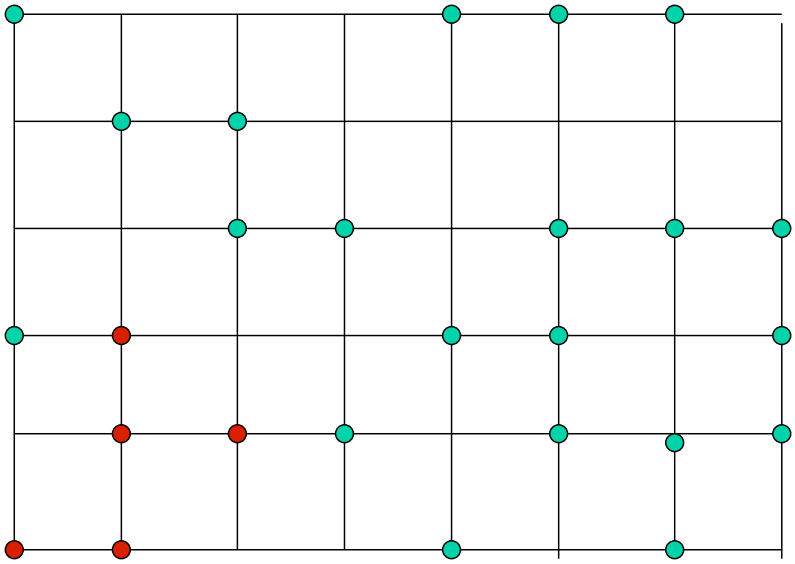
60



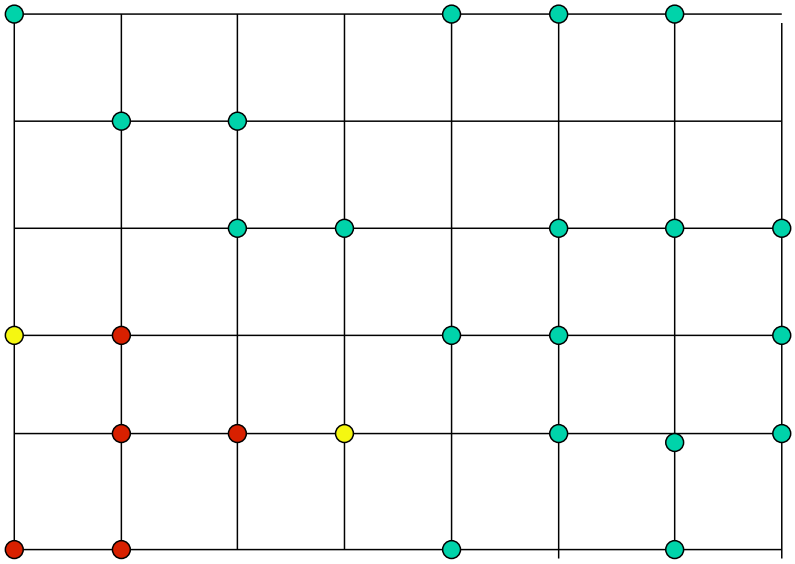
61



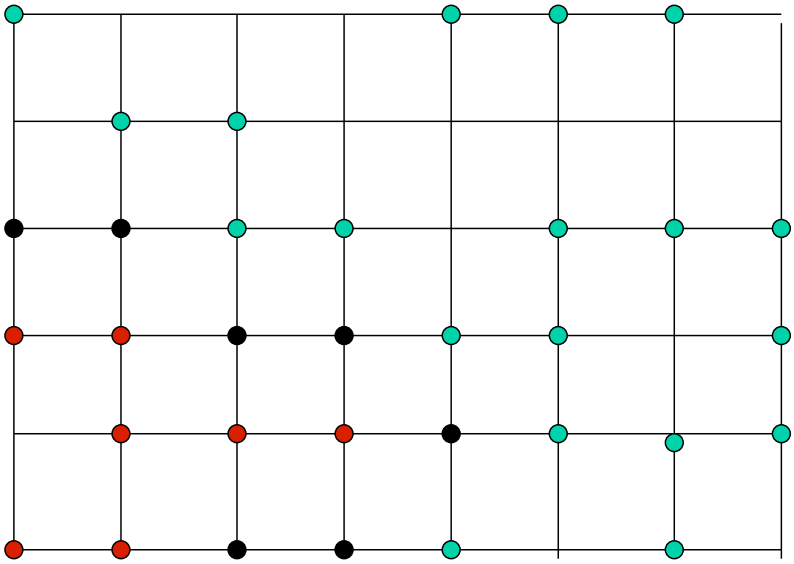
62



63



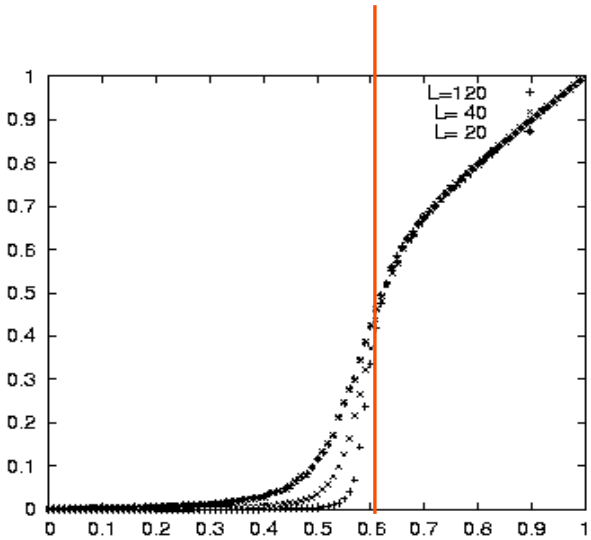
64



65

Taux d'adoptants

p_c



66

q

Variantes …

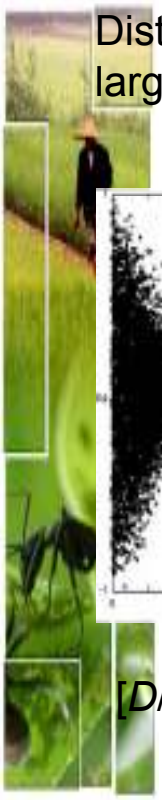
- Solomon et al.
  - Les agents qui vont voir les films ont une exigence (p_i) qui augmente
  - Ceux qui ne vont pas voir de film ont une exigence (p_i) qui diminue
  - Ceux qui ne sont pas informés ne changent pas de préférence
- Sur Réseaux Scale-free
  - Seuil critique plus bas
  - Importance du pi des hubs

67

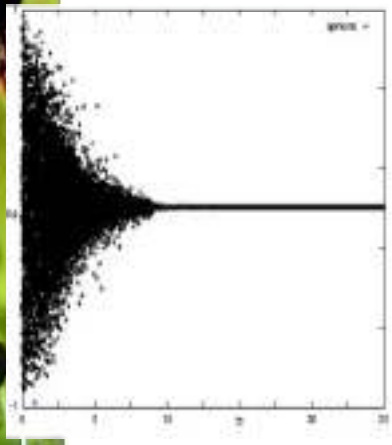
Modèles de dynamiques d'opinions

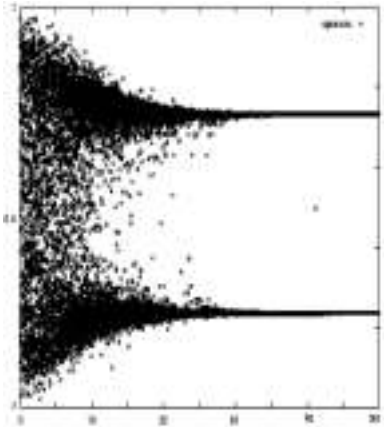
- A_0
  - X_0 : distribution uniforme des opinions
  - Même incertitude pour chaque agent
- G_0 : graphe complet
- T_0 : sélection aléatoire de relations du graphe
- f_0 : *Si* $|x - x'| < d$ alors $\delta x = \mu(x' - x)$ $\delta x' = \mu(x - x')$

68



Distribution initiale uniforme sur un segment de largeur *D*

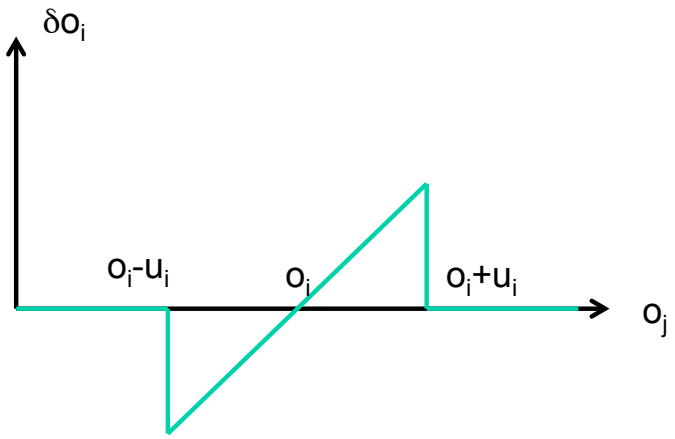




[*D*/2*d*]=1 [*D*/2*d*]=2

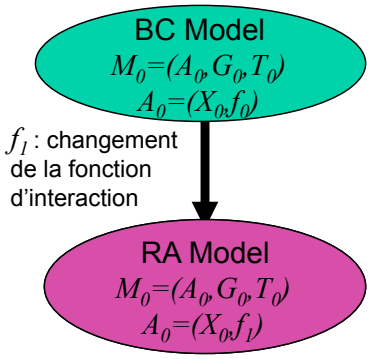
69

Principal problème du modèle BC : la fonction d'influence



Incohérence au niveau individuel
Intérêt pour des formes fonctionnelles appropriées

70



71

Modèle RA (Relative Agreement)

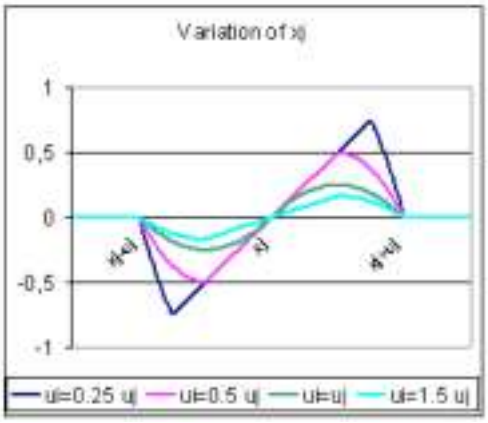
- n agents i
  - Opinion o_i (distribution uniforme $[-1 ; +1]$)
  - Incertitude u_i (initialement la même pour tous)

  ⇒Segment d'opinion $[o_i - u_i ; o_i + u_i]$
- L'influence dépend du recouvrement entre les segments d'opinion
  - Pas d'influence si ils sont trop éloignés
  - Les agents sont influencés à la fois en opinion et en incertitude
  - Plus ils sont certains, plus ils sont convaincants…

72

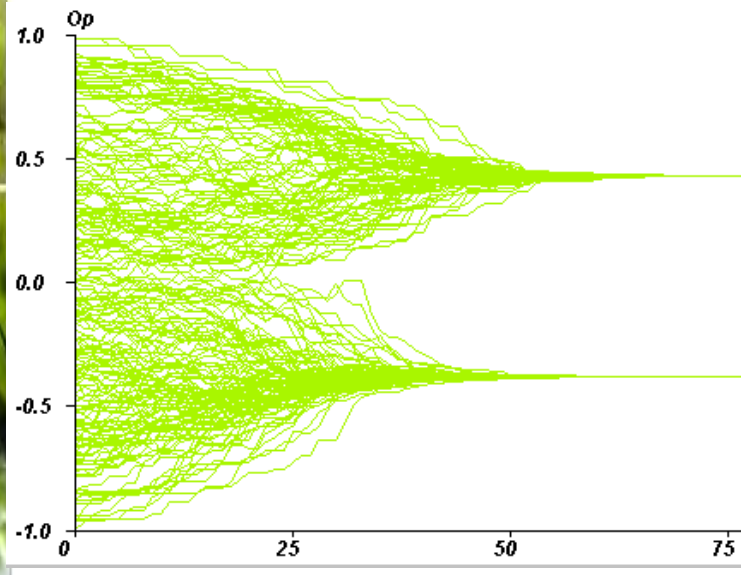
Influence continue

Plus de décroissance brutale de l'influence



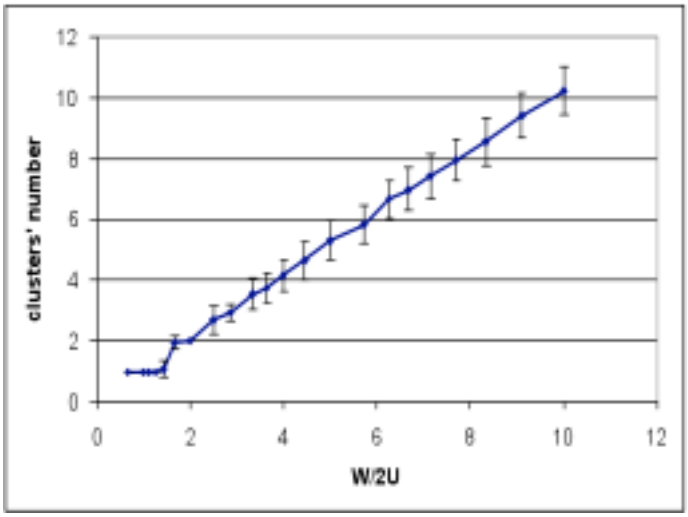
73

Résultats avec init. u = 0.5 pour chaque agent

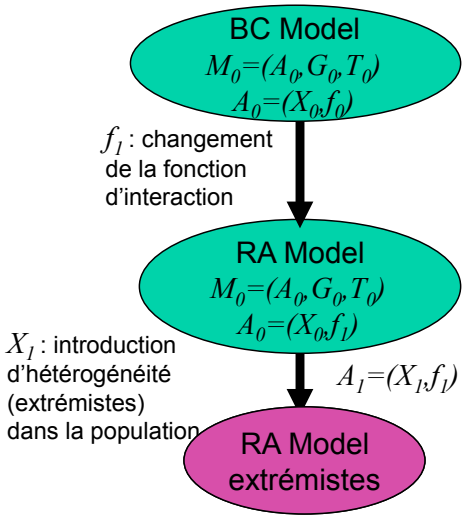


74

Variations du nombre de clusters en fonction de *u* (r^2=0.98)



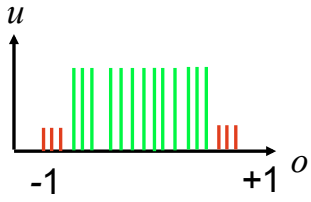
75



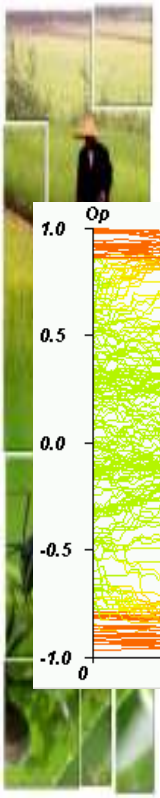
76

Introduction d'extrémistes dans le modèle RA

- X_0 : distribution uniforme des opinions, même incertitude pour tous les agents
- X_1 : distribution uniforme des opinions, hétérogénéité de l'incertitude : deux sous-populations
  - Extrémistes (aux extrémités de la distribution d'opinion)
    incertitude plus faible ue
  - Incertitude des modérés U
- Nouveaux paramètres du modèle :
  - U : incertitude initiale des agents modérés
  - ue : incertitude initiale des extrémistes
  - pe : proportion initiale des extrémistes

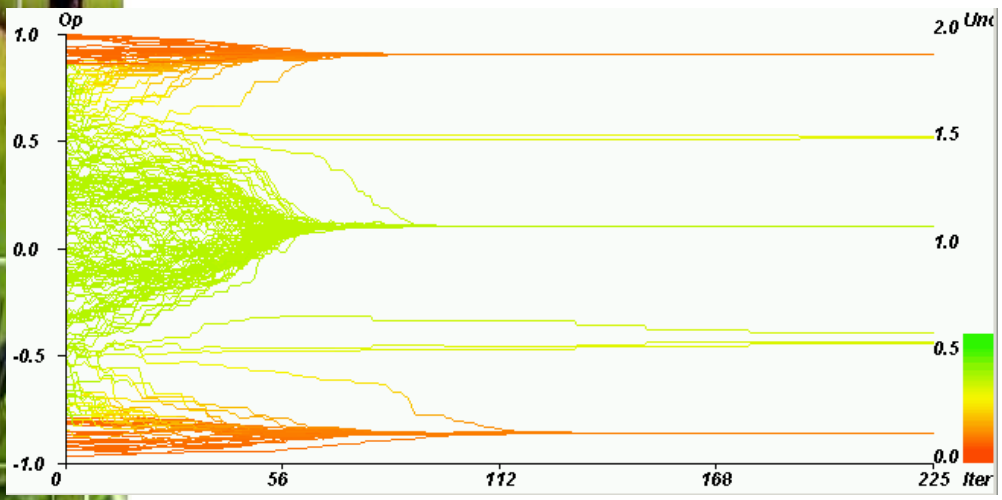


77



Convergence centrale

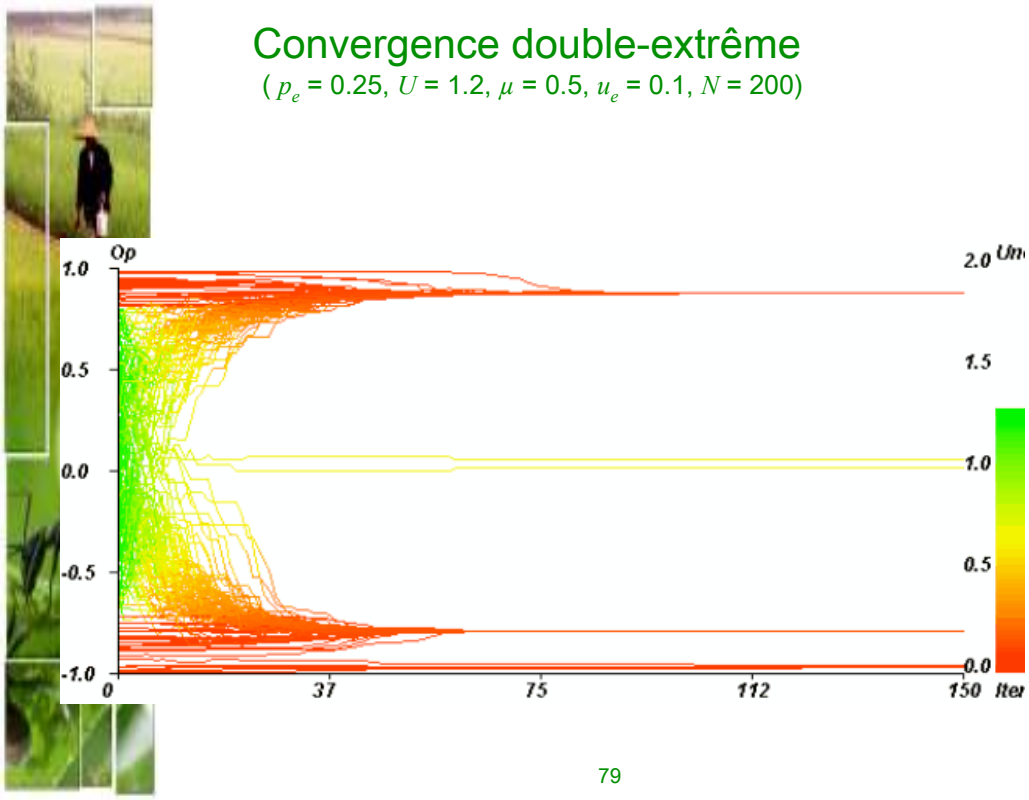
(p_e = 0.2, U = 0.4, μ = 0.5, u_e = 0.1, N = 200)



78

Convergence double-extrême

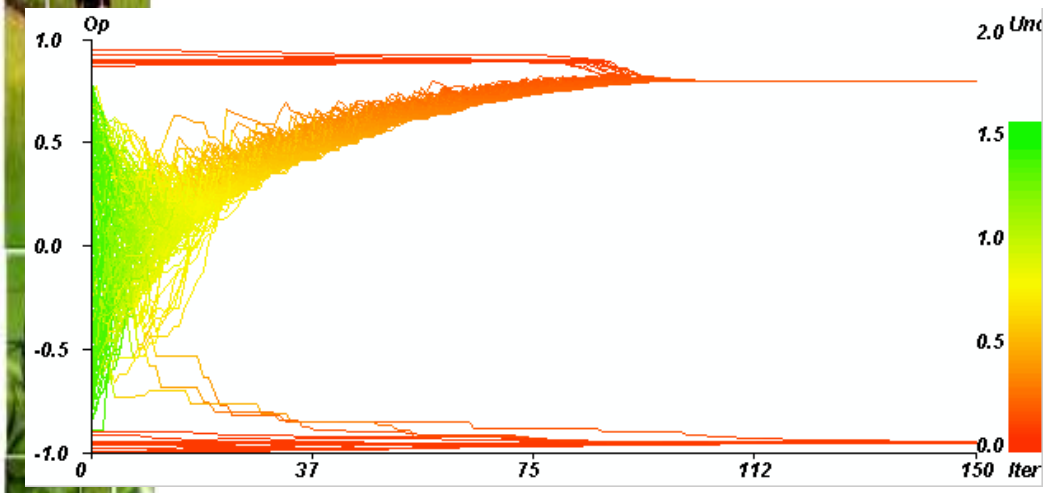
(p_e = 0.25, U = 1.2, μ = 0.5, u_e = 0.1, N = 200)



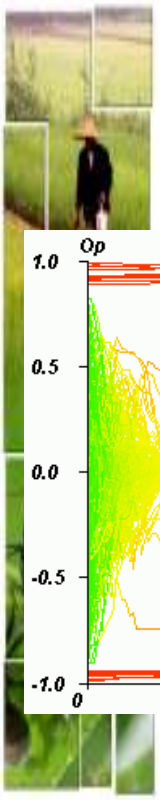
79

Convergence simple extrême

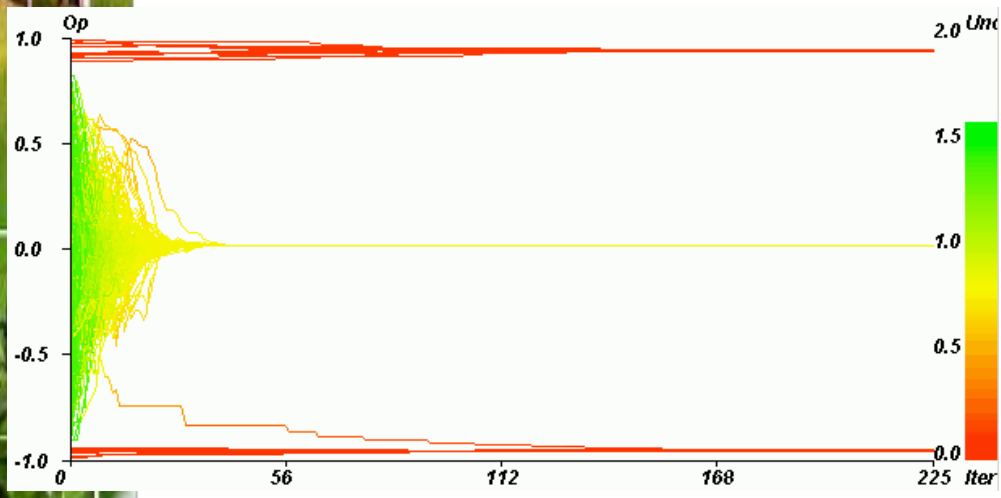
(p_e = 0.1, U = 1.4, μ = 0.5, u_e = 0.1, N = 200)



80



Sensibilité de la dynamique : pour les même paramètres, convergence centrale



81

Étude plus systématique des propriétés du modèle

- Introduction de l'indicateur $y = p'_{+}{}^{2} + p'_{-}{}^{2}$

  p'_{+} = prop. d'agents modérés qui convergent vers l'extrémité positive

  p'_{-} = prop. d'agents modérés qui convergent vers l'extrémité négative

  – Convergence centrale

  $y = 0^2 + 0^2 = 0$

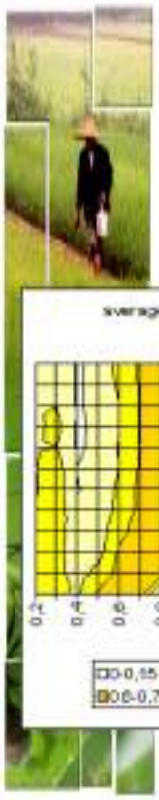
  – Convergence double extrême

  $y = 0.5^2 + 0.5^2 = 0.5$

  – Convergence simple extrême

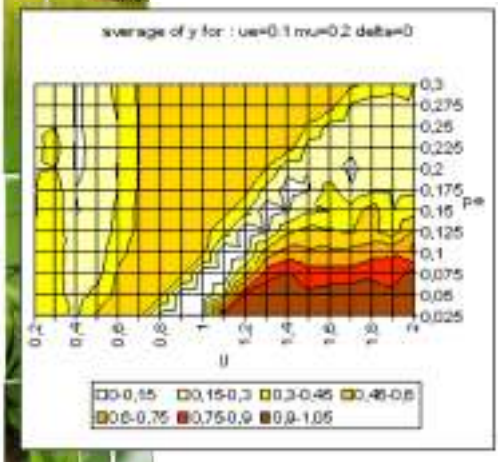
  $y = 1^2 + 0^2 = 1$

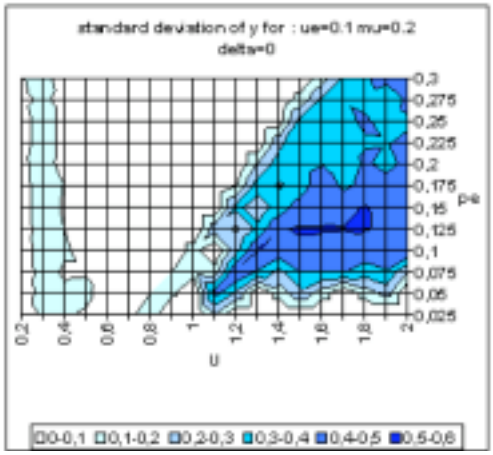
82



$\delta = 0$, $u_e = 0.1$, $\mu = 0.2$, $N=1000$ (repl.=50)

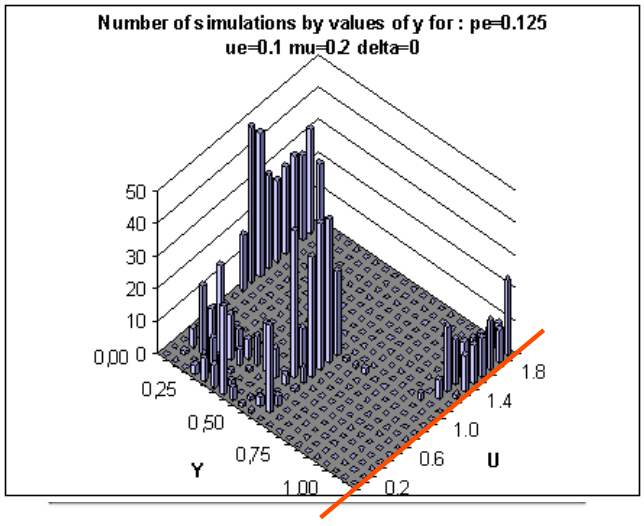
- blanc, jaune => conv. centrale
- orange => conv. double extrême
- marron => simple extrême





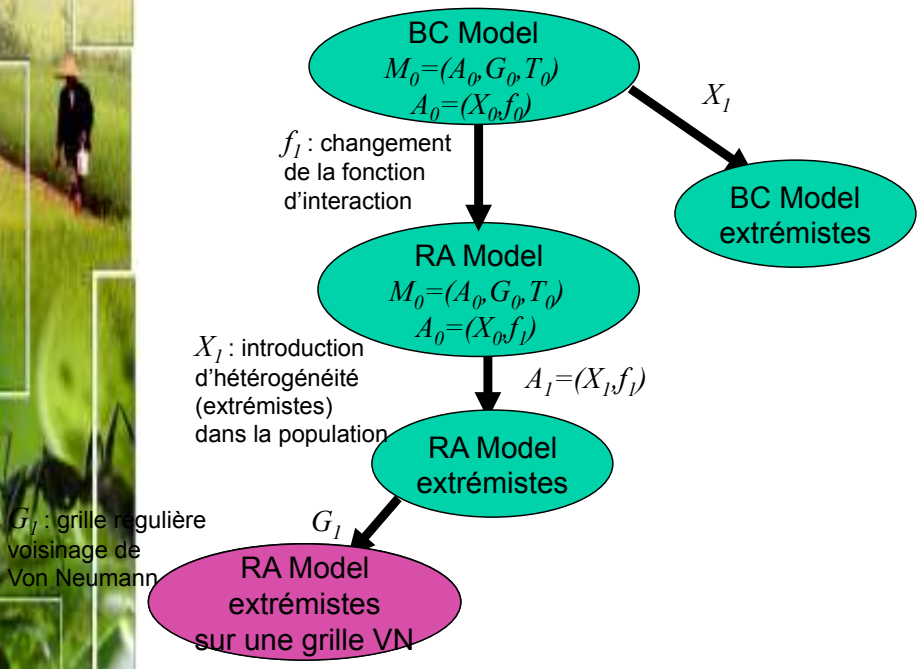
83

$p_e = 0.125$ $\delta = 0$
($U > 1$) => conv. centrale et simple extrême
($0.5 < U < 1$) => conv. double extrême
($U < 0.5$) => différentes convergences entre conv. Centrale et conv. double extrême



84

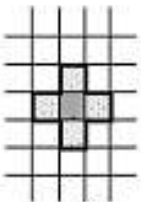
Hiérarchie de modèles de dynamiques d'opinions



85

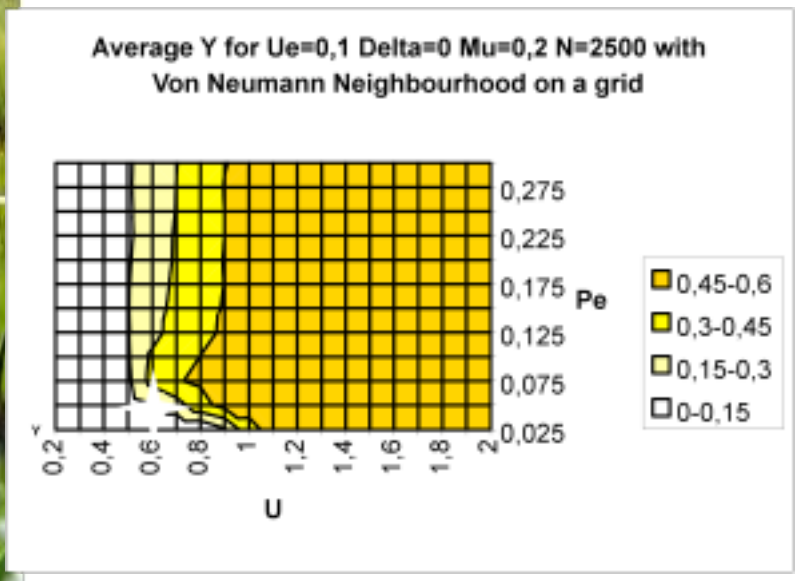
Modification du graphe d'interaction

- G_0 : graphe complet
- G_1 : grille régulière avec voisinage de Von Neumann
  - Connectivité 4 : (N,S,E,O)

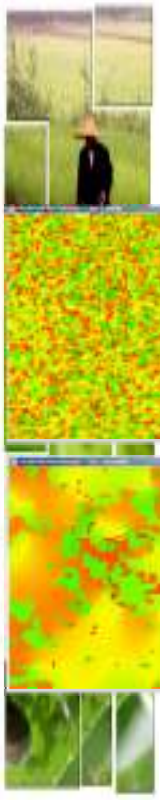


86

Sur une grille régulière voisinage de Von Neumann
Cas de convergence : moyenne de y

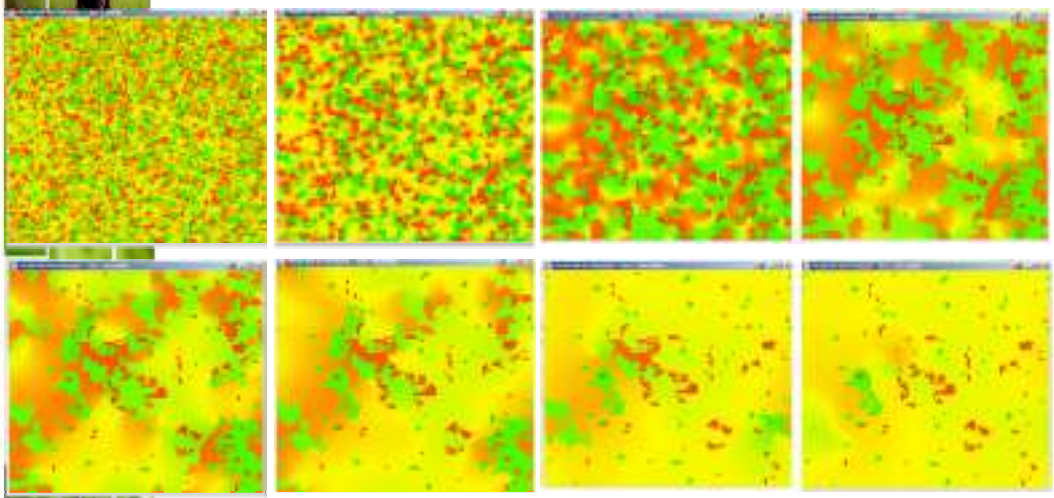


87



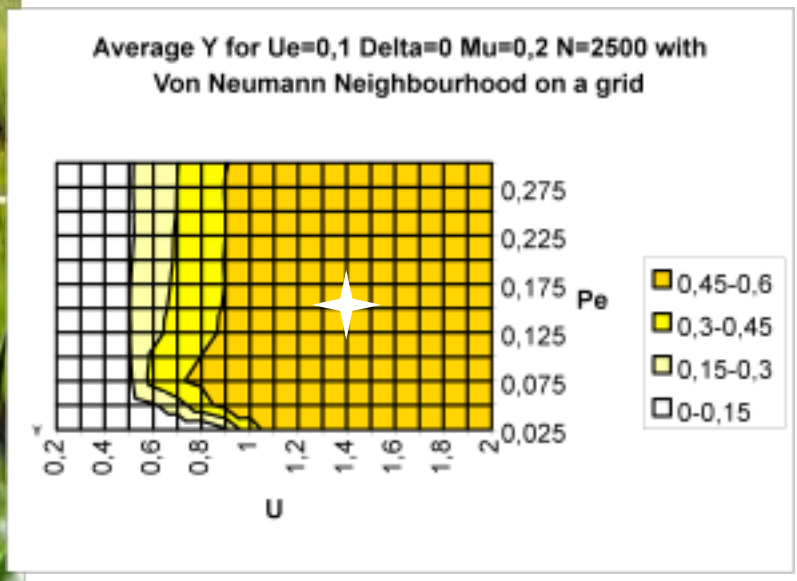
Cas de convergence centrale
(U=0.6,pe=0.05)

Extrémistes : verts et rouges



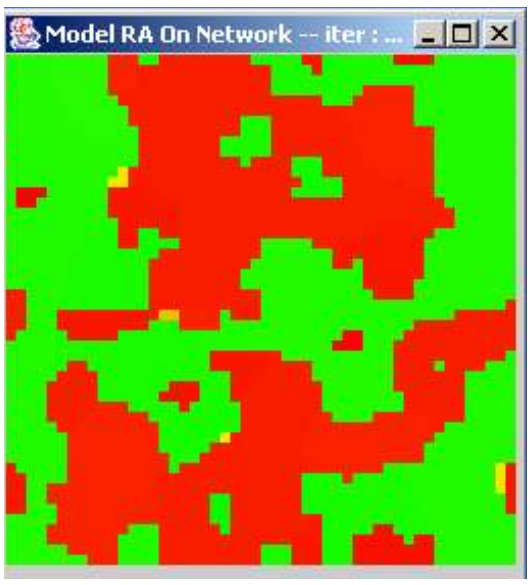
88

Sur une grille régulière voisinage de Von Neumann
Cas de convergence : moyenne de y



89

Convergence double extrême
(U=1.4 pe=0.15)



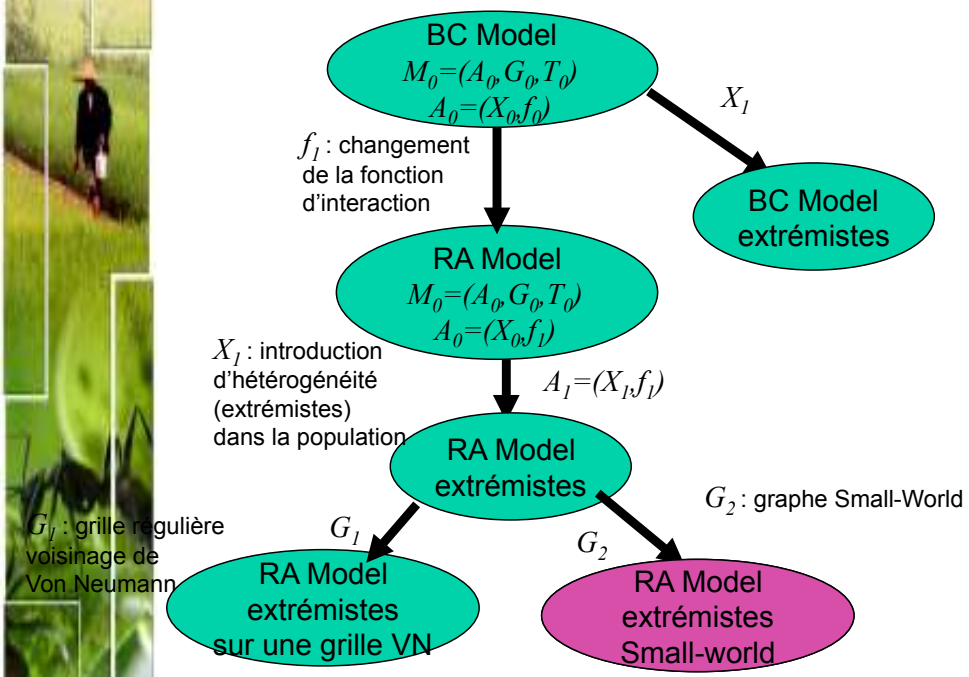
90

A la recherche du simple extrême…

- On n'observe plus de conv. simple extrême
- On l'obtenait avec un graphe complet
- On devrait l'obtenir avec un graphe de connectivité plus élevée

91

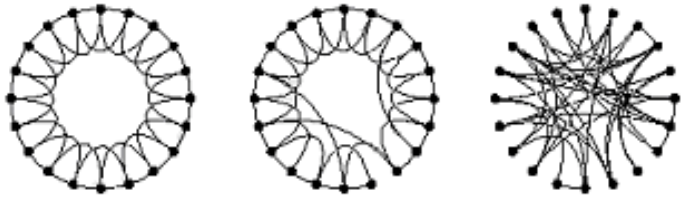
Hiérarchie de modèles de dynamiques d'opinions



92

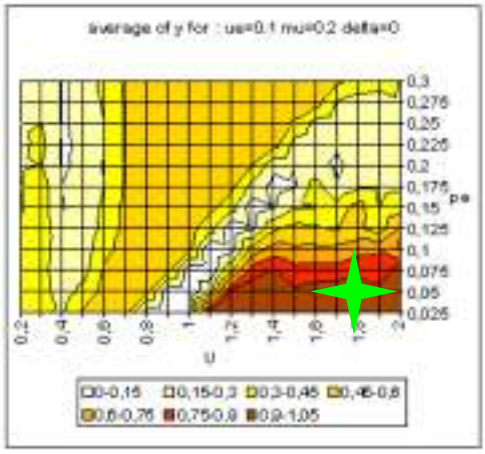
Modèle Small-World (Watts)

- G_2 : on part d'une structure régulière de connectivité k on ajoute un bruit β
- Permet de balayer un ensemble de graphe entre les graphes régulièrs (β faible à gauche) et les graphes aléatoires (β élevé à droite) en jouant également sur la connectivité

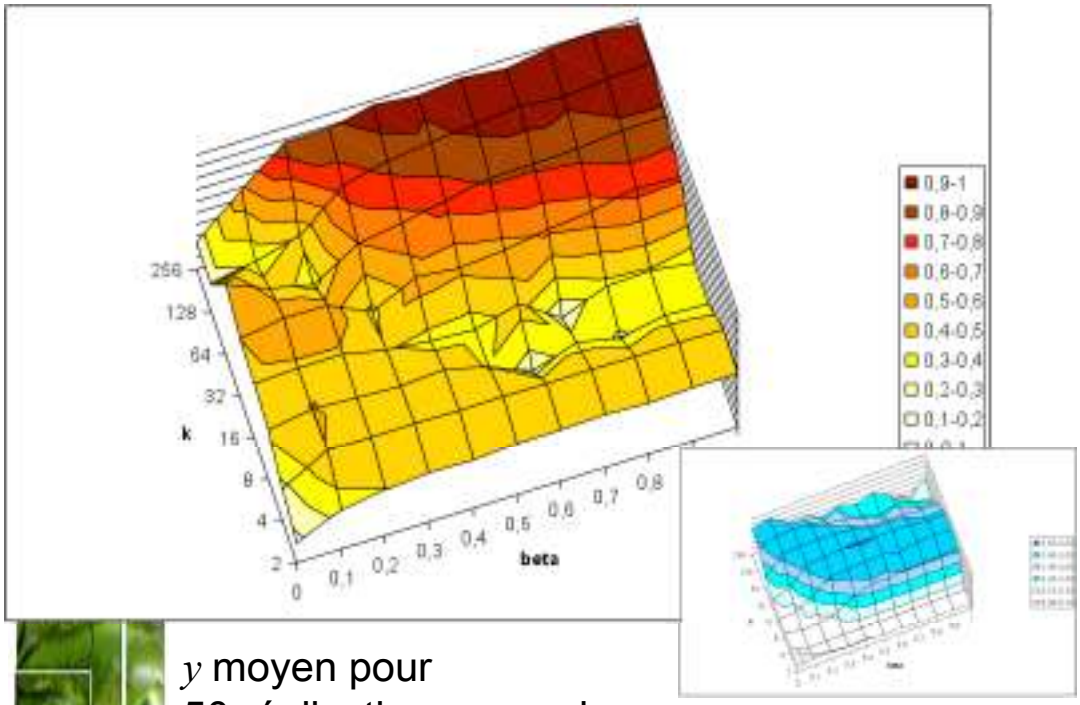


93

- Pour un point donné de l'espace (U, pe) correspondant à une convergence simple extrême (U=1.8, pe=0.05)
- On fait varier la connectivité moyenne k et β

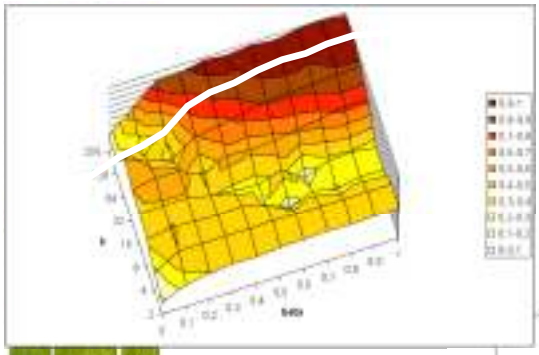


94

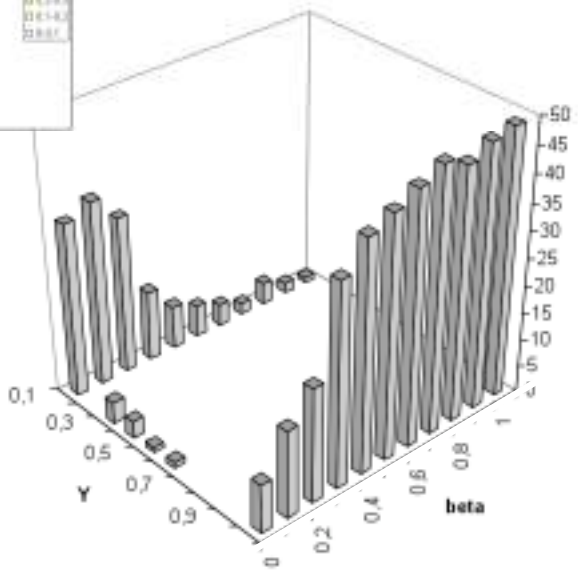


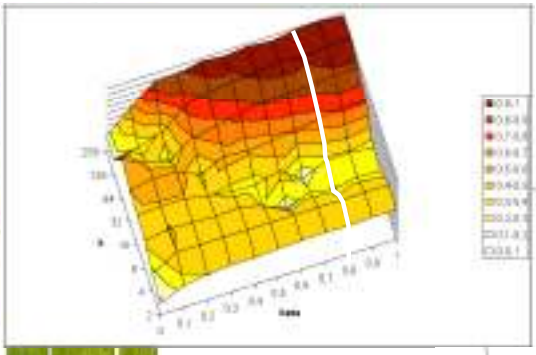
y moyen pour
50 réplications pour chaque couple (β,k)

95

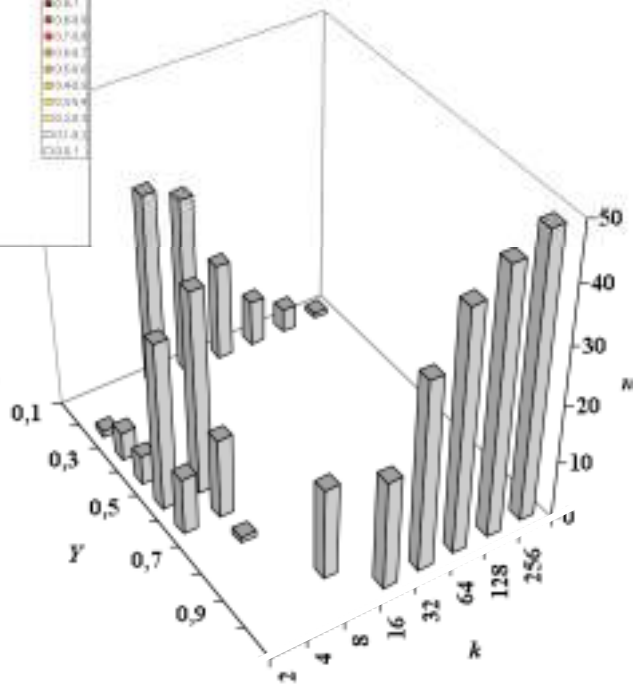


Distribution de y pour k=128





Distribution de y pour $\beta = 0.8$



Explications

- Pour des k faibles : forte influence locale des extrémistes (de chaque bord)
- Pour des k plus élevés : chaque agent a une probabilité plus forte d'interagir avec la majorité
  - Regroupement central des modérés
  - Un extrême s'impose quand la majorité est isolée de l'un des deux extrêmes (convergence centrale sinon)
- Plus le graphe est régulier (β faible), plus la transition a lieu pour des connectivités élevées
  - La régularité du graphe renforce la propagation locale de l'extrémisme, il s'en suit une convergence double extrême

98

Plan

- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de réseaux dynamiques
  - Modèles de dynamiques sur réseaux
  - Co-évolution
- Expérimentations
- Perspectives

Co-évolution des réseaux sociaux / individus

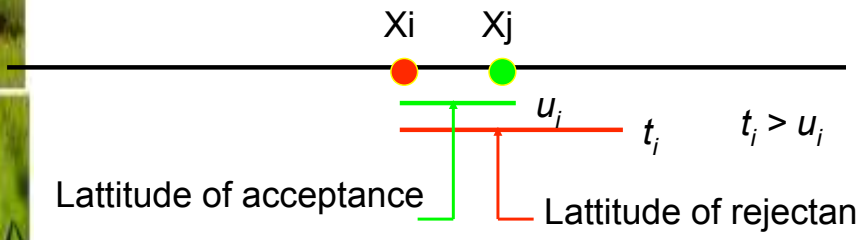
Exemple: réseaux sociaux et tabagisme

N.A.Christakis, J.H.Fowler

Formalizing Social Judgement Theory

Together with W.Jager

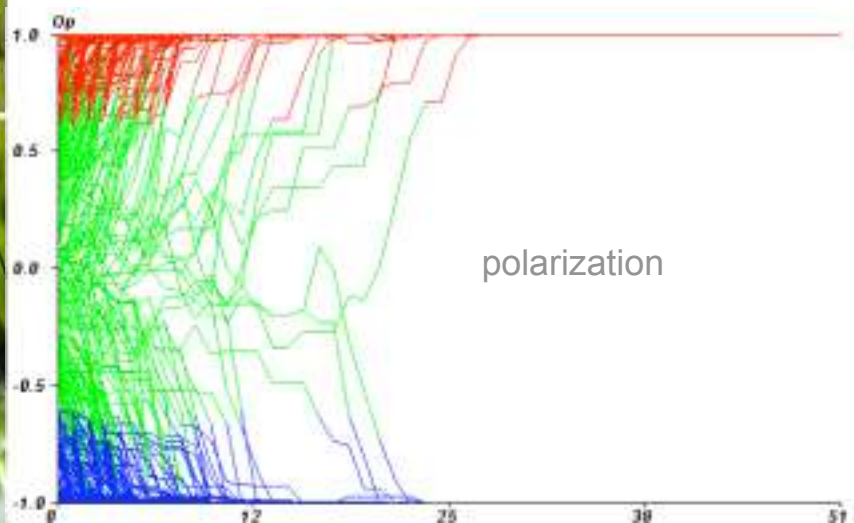
Formalising Social Judgement Theory



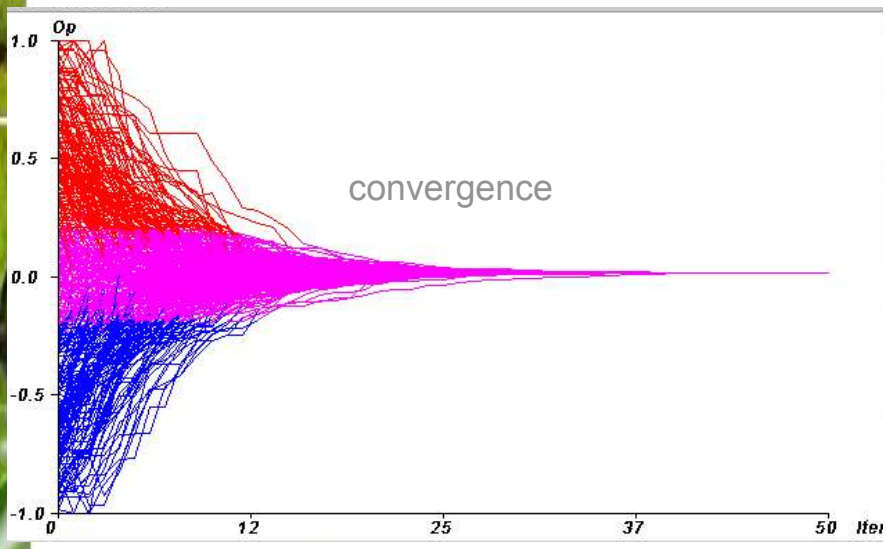
If $|x_i-x_j|<u_i$ *assimilation:* $dx_i=\mu.(x_j-x_i)$
If $|x_i-x_j|>t_i$ *contrast:* $dx_i=\mu.(x_i-x_j)$

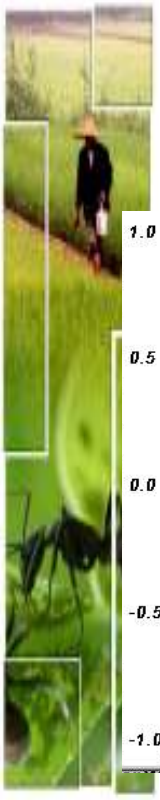
μ expresses the degree of change

Results: *high 'ego-involvement'* U = *0.4*, T = *0.6*

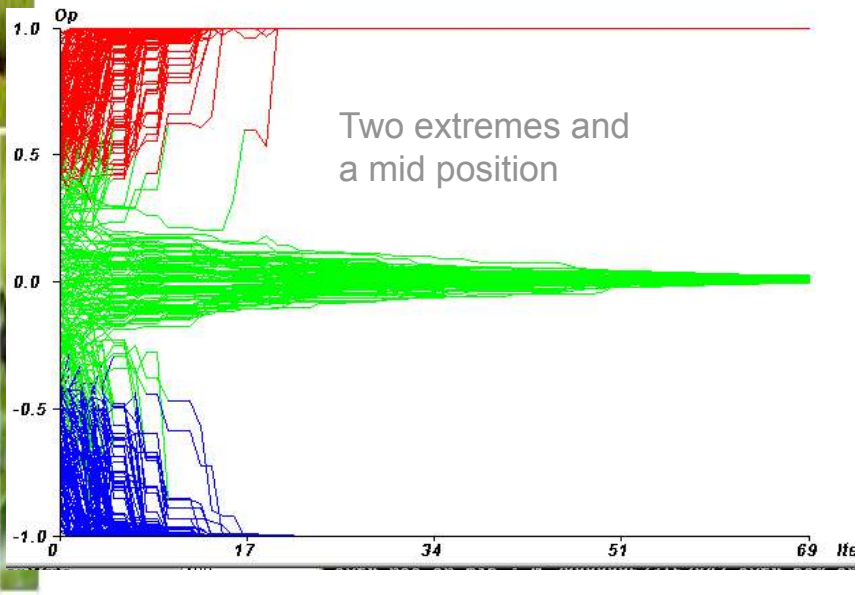


Results: *Very acceptant: U = 1.2,T = 1.6*

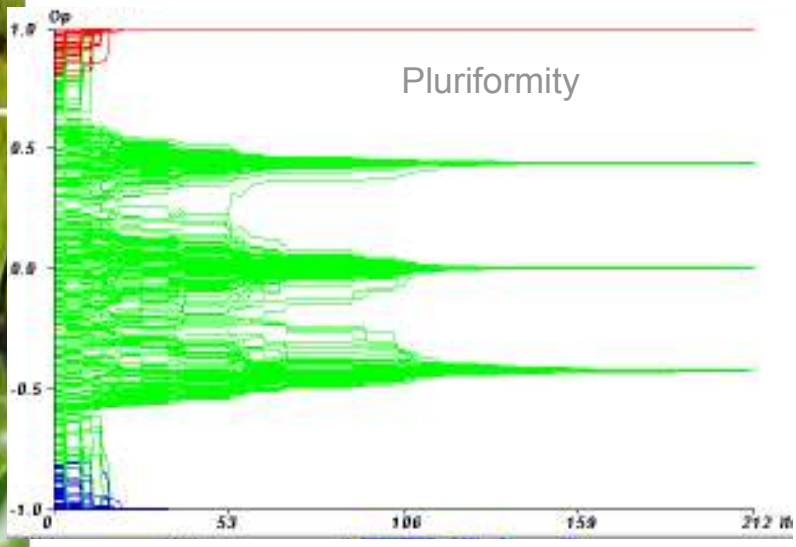




Results: small acceptance and large non-commitment (U = 0.6, *T* = 1.2).

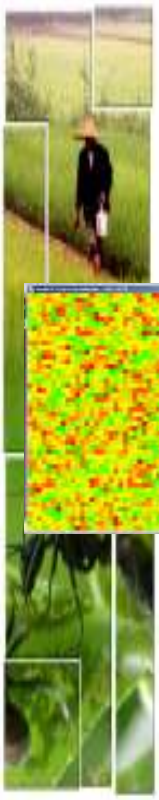


Results: Very large non-commitment area ($U = 0.2$, $T = 1.6$)

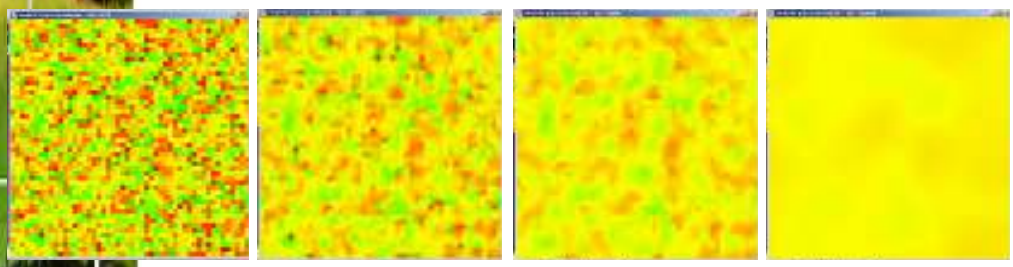


Tests on social networks

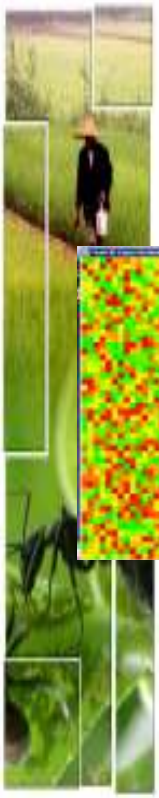
- We first test the model on regular grids (VN neighbourhood)
  - It enables a good vizualisation of what happen



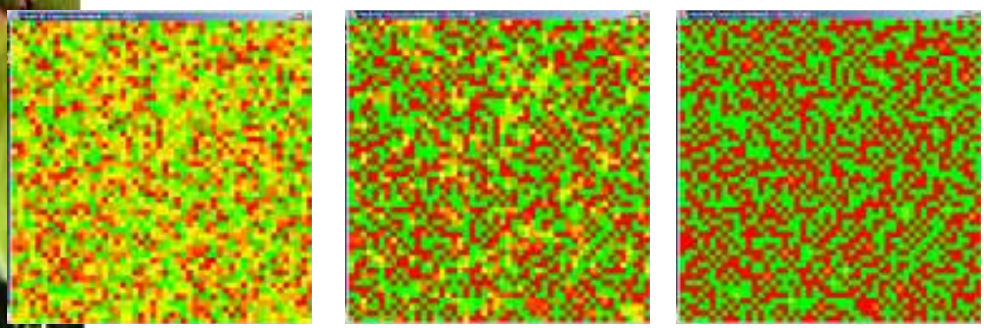
U=1.3, T=1.7 (convergence case for a totally connected population) on a grid with VN neigb.
We observe Convergence also … n=2500)

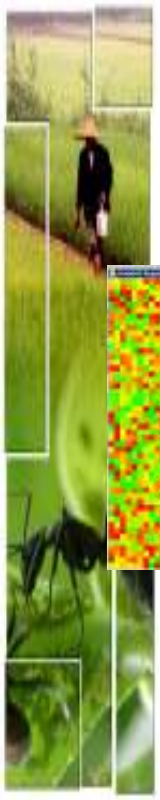


Note: the attitudes are coded between red (-1) and green (+1), yellow codes for attitudes close to zero

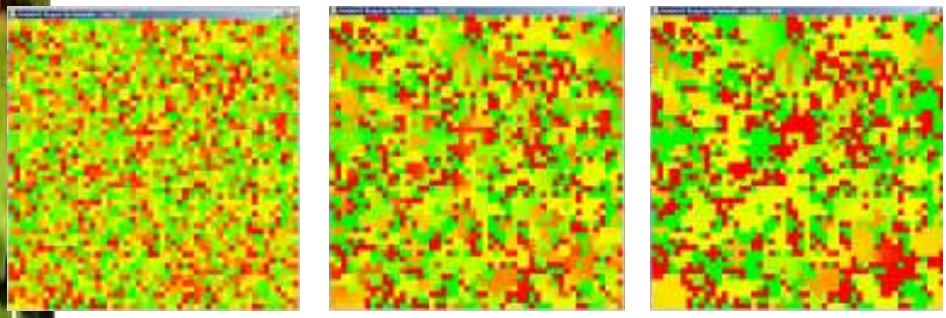


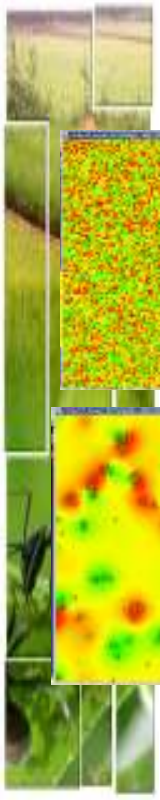
U=0.3, T=0.6 polarisation in the totally connected case …



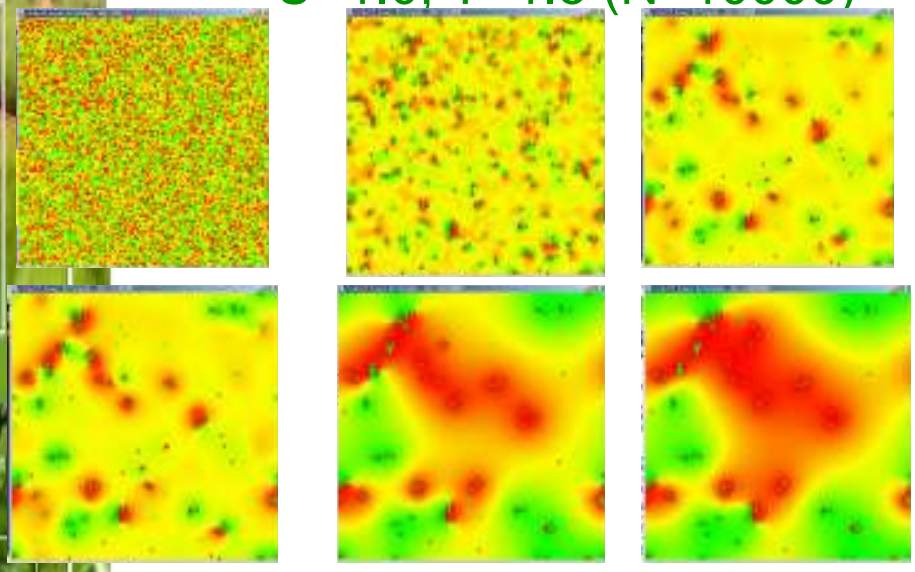


U=0.5, T=1.3 pluriformity…





Metastable states on the frontiers ?
U=1.0, T=1.5 (N=10000)



Maybe a scale effect
On the population size

Peripheral vs. Central Processing

Peripheral processing

- *2 dimensions A & B*
- *If assimilation on A, then assimilation on B*

  If $|A_i-A_j|<u_i$

  $dA_i=\mu.(A_j-A_i)$ *and* $dB_i=\mu.(B_j-B_i)$

- *If contrast on A, then contrast on B*

  If $|A_i-A_j|>t_i$

  $dA_i=\mu.(A_i-A_j)$ and $dB_i=\mu.(B_j-B_i)$

Experimental set up

1600 agents are implemented on a Von Neumann neighborhood

Attitudes A and B drawn from a uniform distribution [-1,+1]

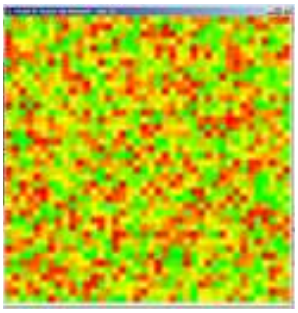
Random pairs of neighbors are selected for interacting

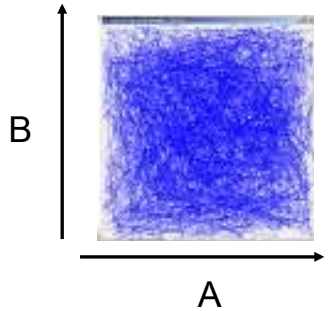
Agents first interact on issue A, and then on issue B

Latitude of acceptance and non-commitment high (U at 1.0 and T at 1.5) lead to convergence in single attitude condition (Jager & Amblard, CMOT 2005)

Representation of the results

Representation of each agent's attitude coded between red (-1) and green (+1), yellow codes for 0





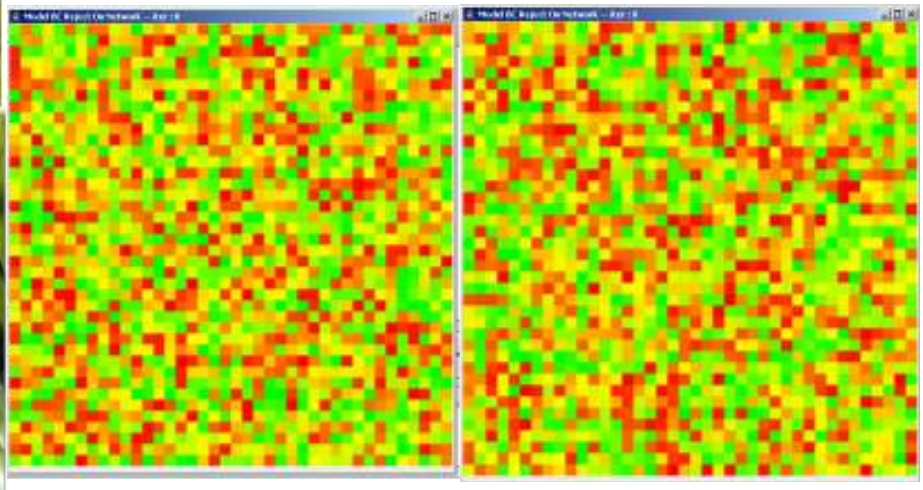
Representation of the space of states, X-axis represents the attitude A, Y-axis the attitude B, the blue lines represent the neighbourhood links between agents.

Results

Central processing on both A and B

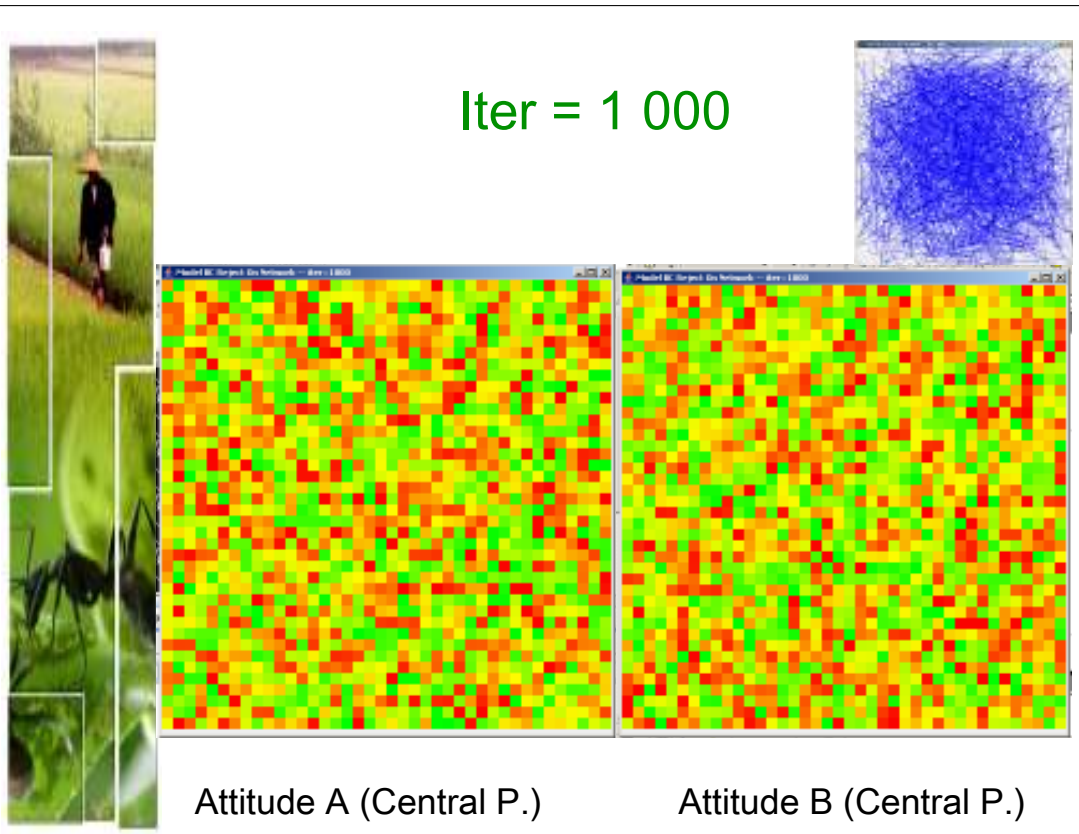
U=1.0 T=1.5

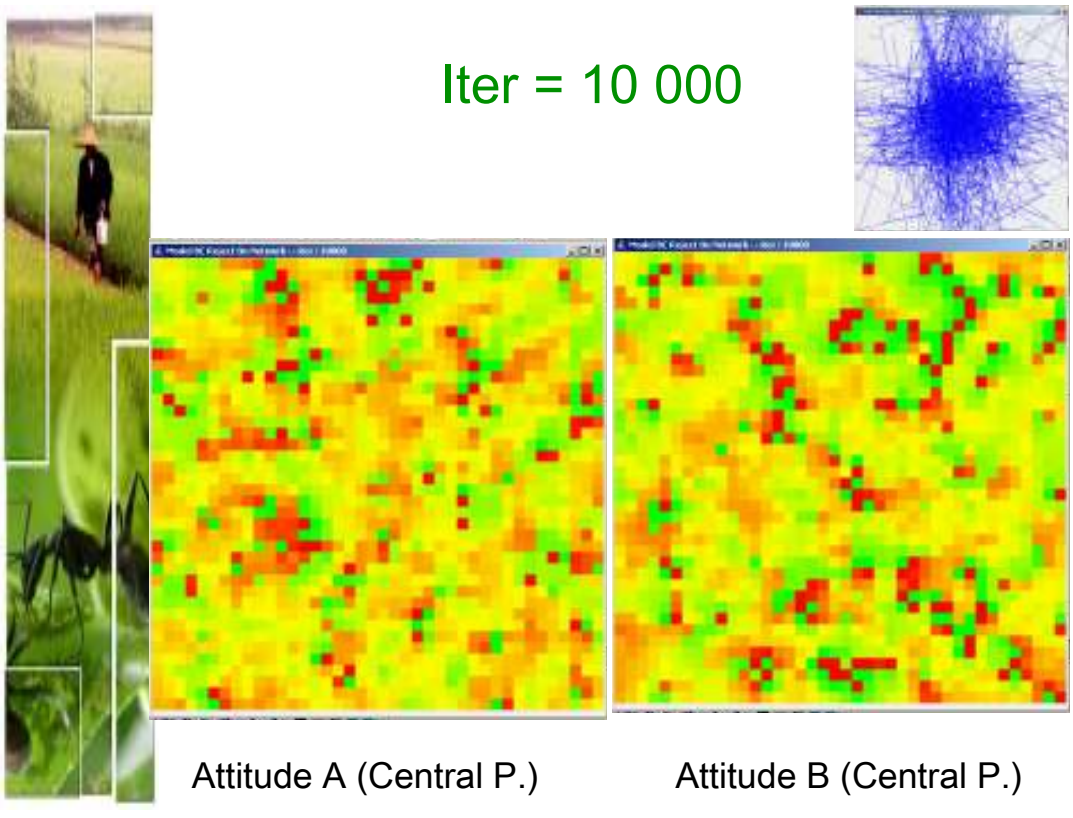
Iter = 0



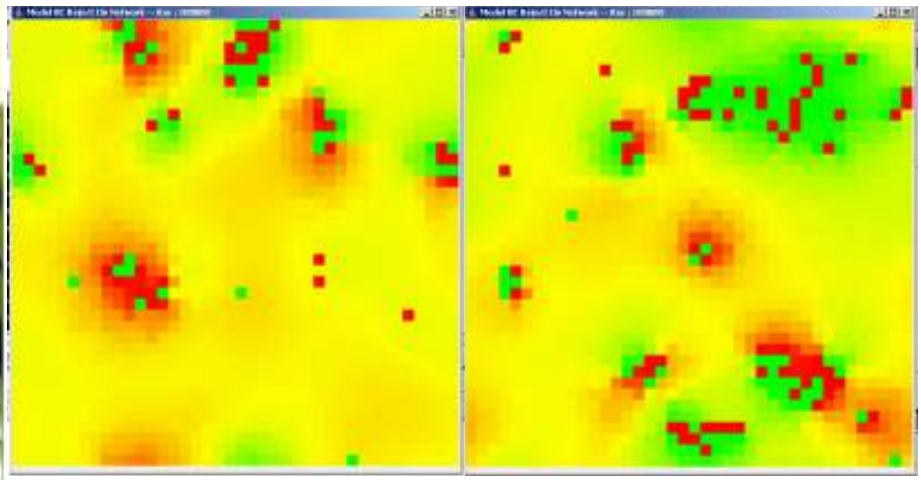
Attitude A (Central P.)

Attitude B (Central P.)


Iter = 1 000
Attitude A (Central P.)
Attitude B (Central P.)

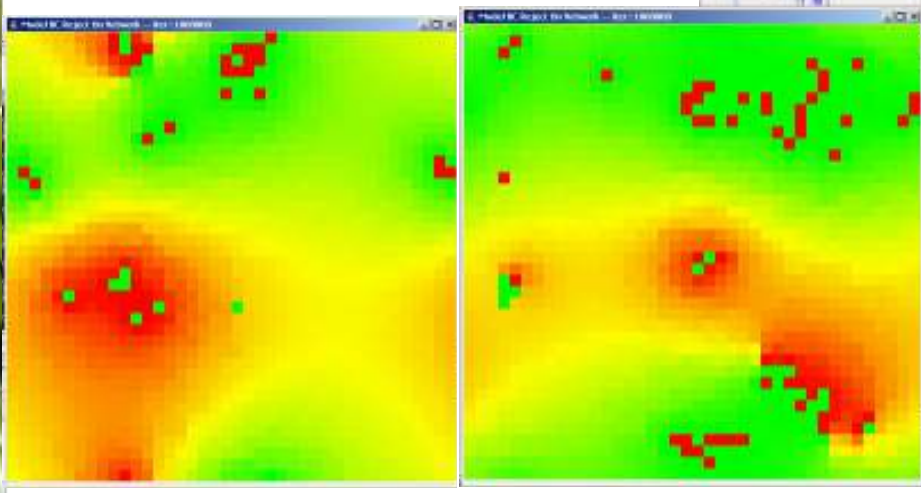

Iter = 10 000
Attitude A (Central P.)
Attitude B (Central P.)

Iter = 100 000



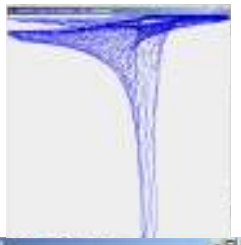
Attitude A (Central P.) Attitude B (Central P.)

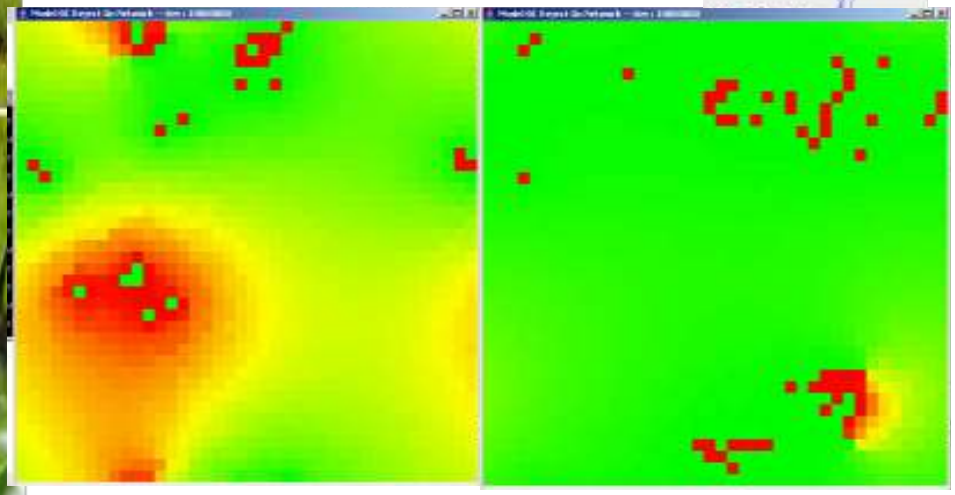
Iter = 1 000 000



Attitude A (Central P.) Attitude B (Central P.)

Iter = 10 000 000

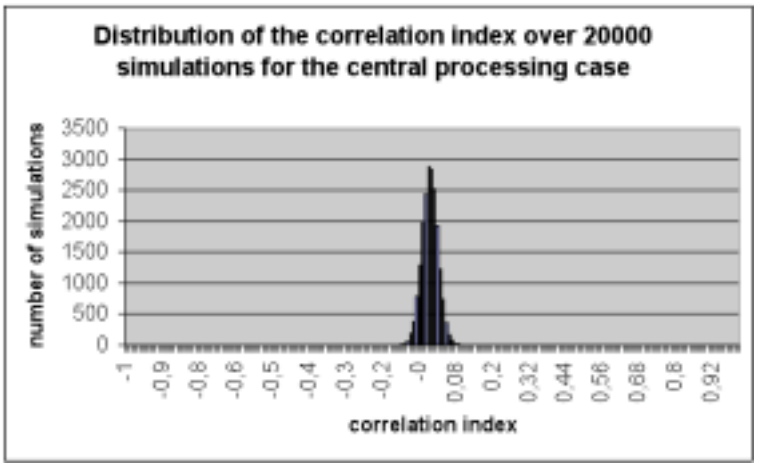




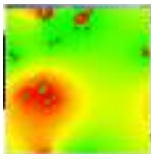
Attitude A (Central P.)

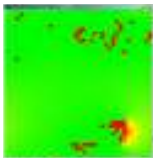
Attitude B (Central P.)

Correlation index between attitudes A and B

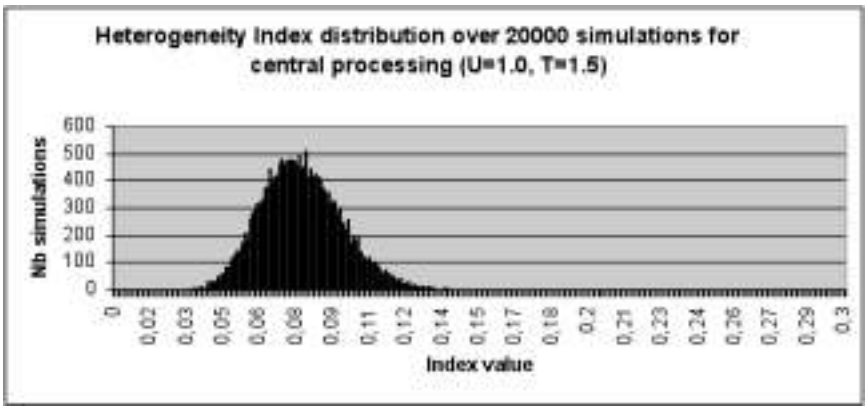


No correlation between both attitudes

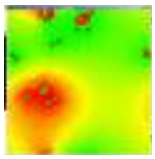


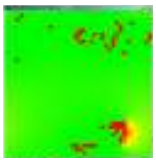


Heterogeneity index between attitudes A and B



Strong clustering on both dimensions



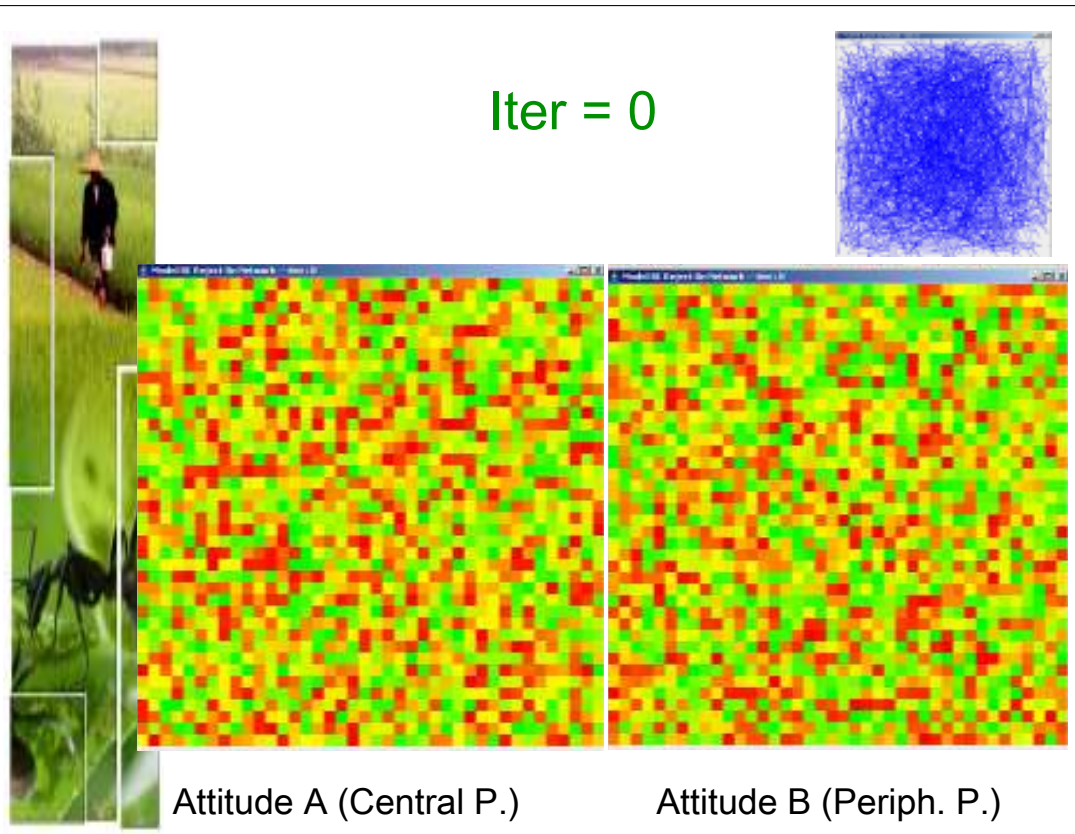


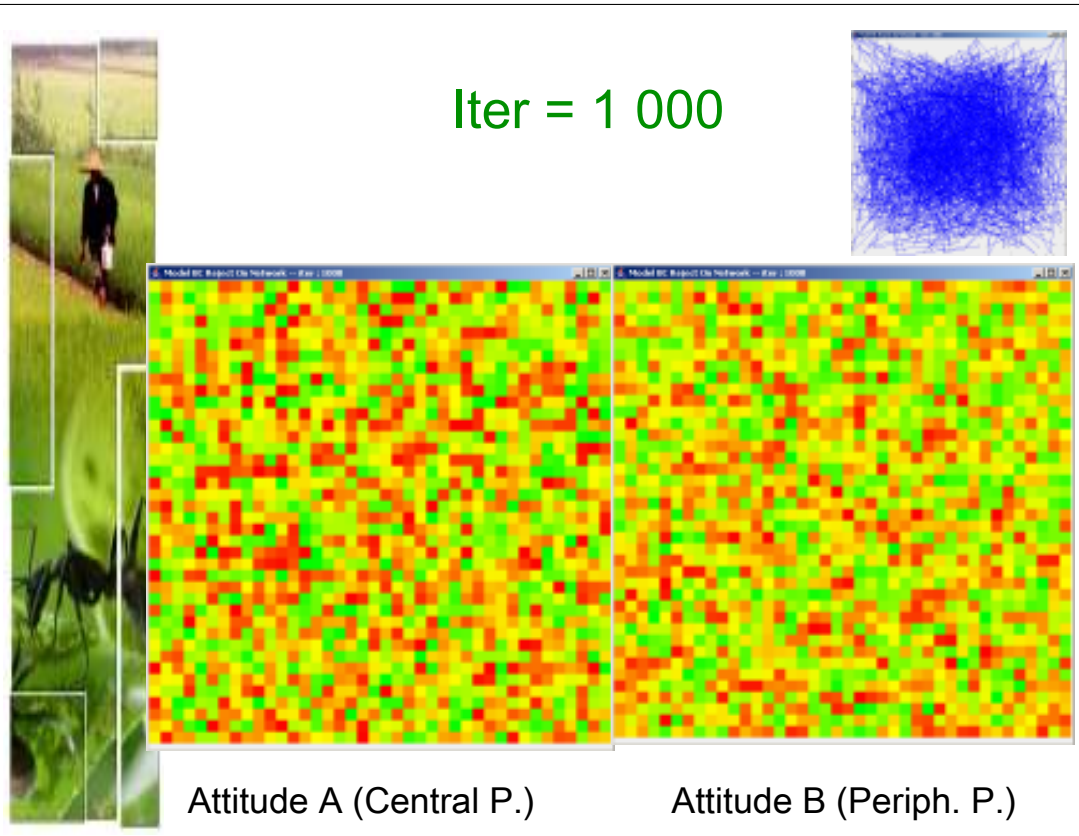
Results

Central processing on A

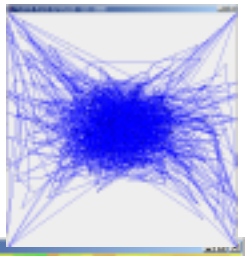
Peripheral processing on B

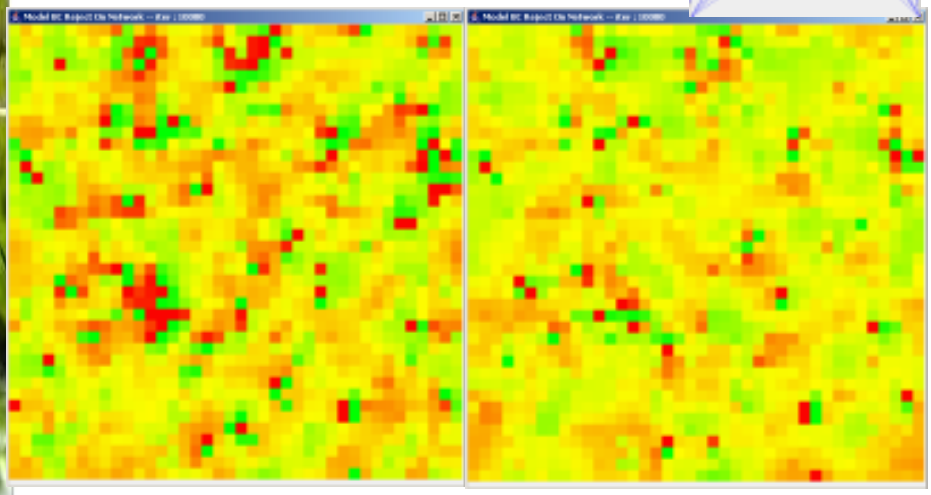
U=1.0 T=1.5


Iter = 0
Attitude A (Central P.)
Attitude B (Periph. P.)


Iter = 1 000
Attitude A (Central P.)
Attitude B (Periph. P.)

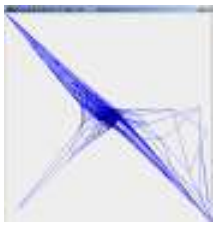
Iter = 10 000

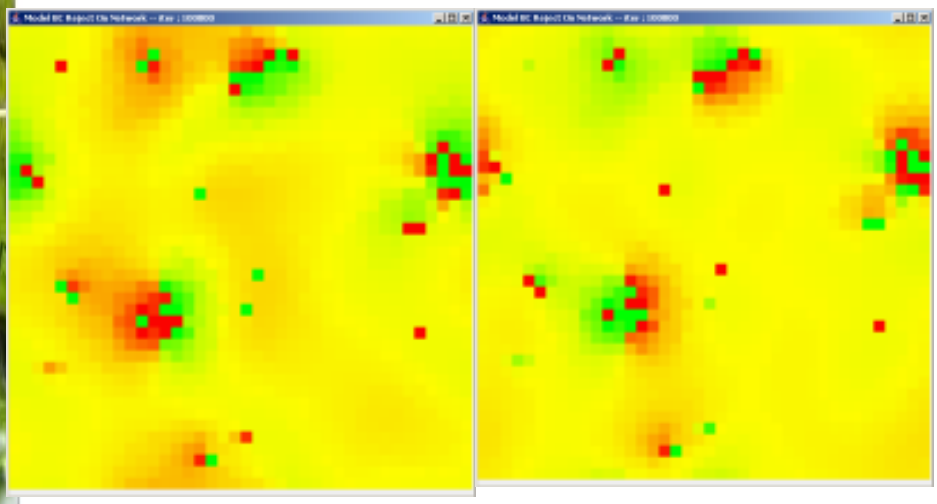




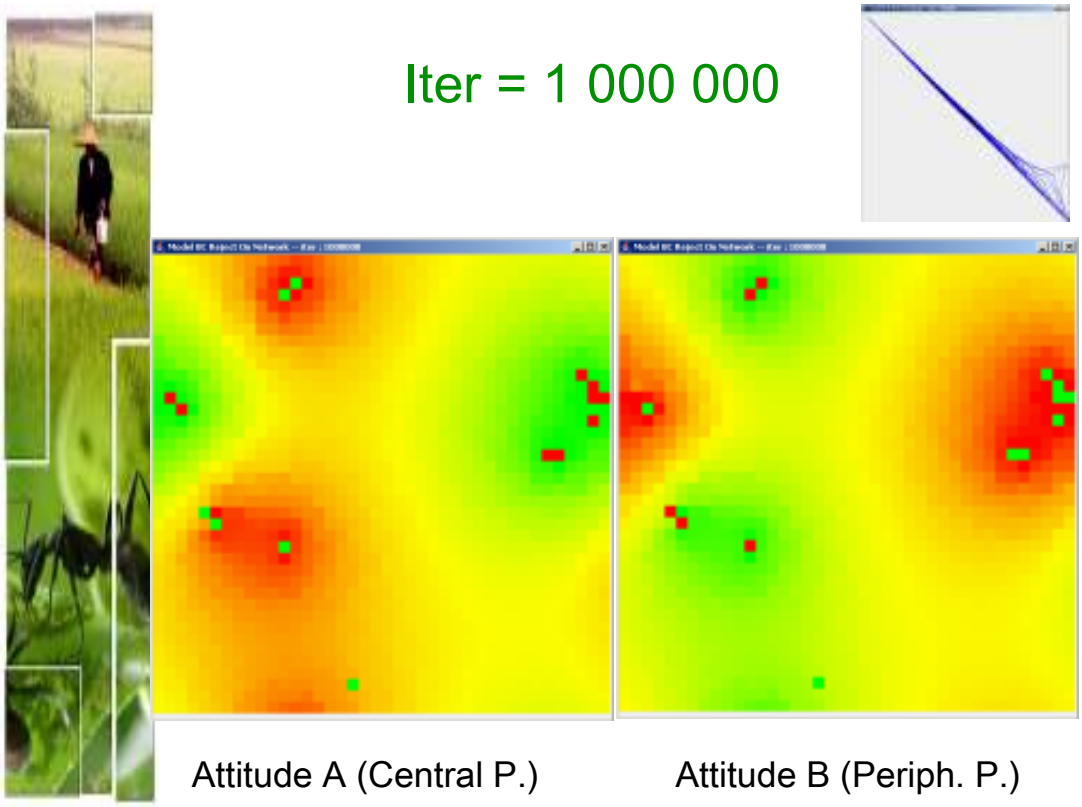
Attitude A (Central P.) Attitude B (Periph. P.)

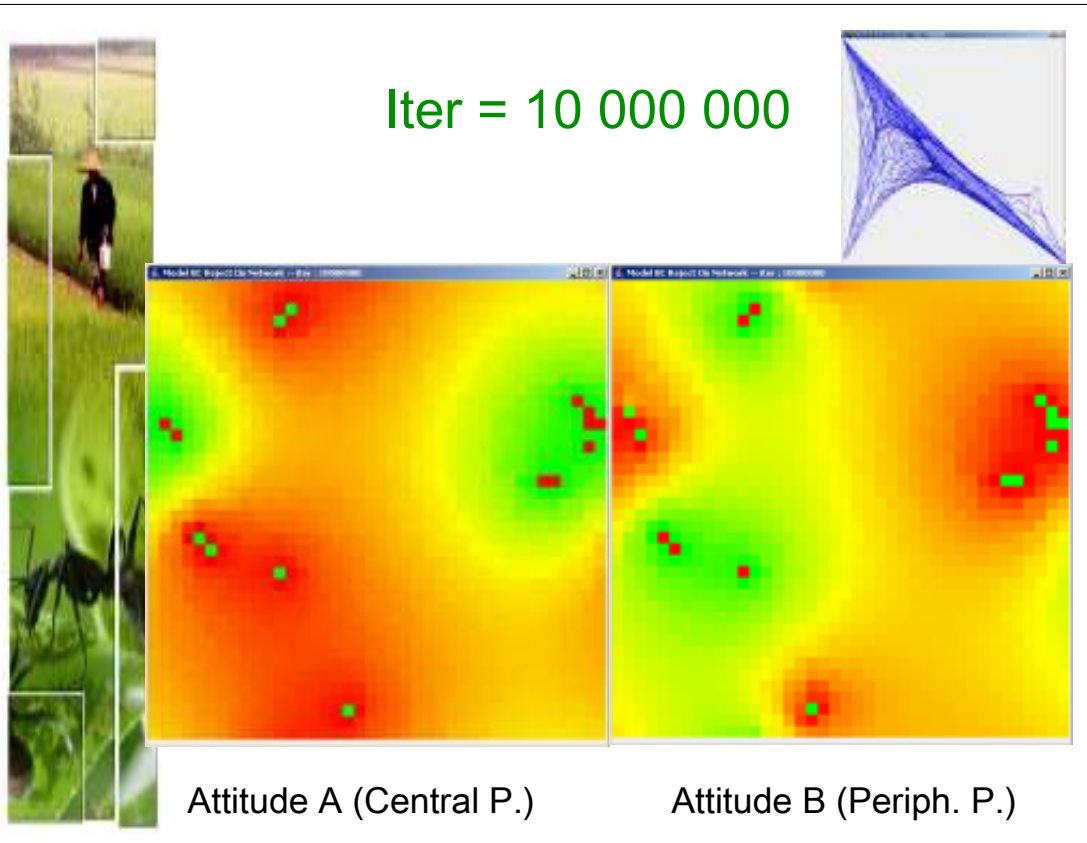
Iter = 100 000



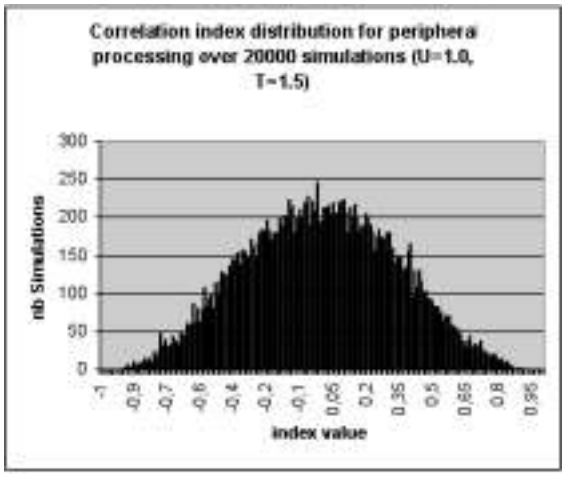


Attitude A (Central P.) Attitude B (Periph. P.)

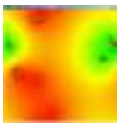

Iter = 1 000 000
Attitude A (Central P.)
Attitude B (Periph. P.)

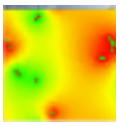

Iter = 10 000 000
Attitude A (Central P.)
Attitude B (Periph. P.)

Correlation index between attitudes A and B

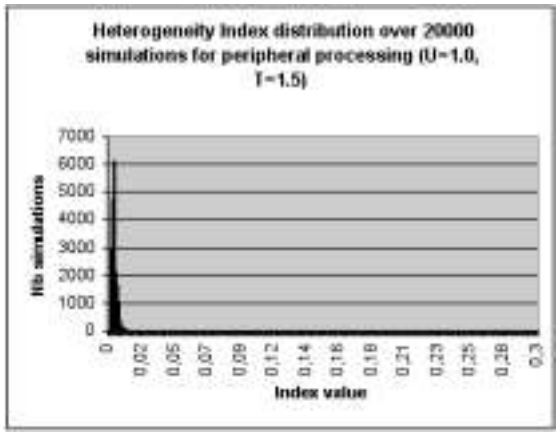


Clear correlation between both attitudes

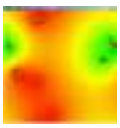


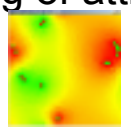


Heterogeneity index between attitudes A and B



Spatial clustering of attitudes





Social network dynamics based on assimilation and contrast effects

Model

- 2 attitudes dimensions A and B [-1,+1]
- Importance to each dimension $\beta_j * A_j$ and $(1 - \beta_j) * B_j$
- Probability of interaction

$$F_{jk} = \frac{(\beta_j + \beta_k)}{2} * |A_j - A_k| + \left(1 - \frac{(\beta_j + \beta_k)}{2}\right) * |B_j - B_k|$$

- Probability to discuss about A

$$p_{discussA} = (\beta_j + \beta_k) / 2$$

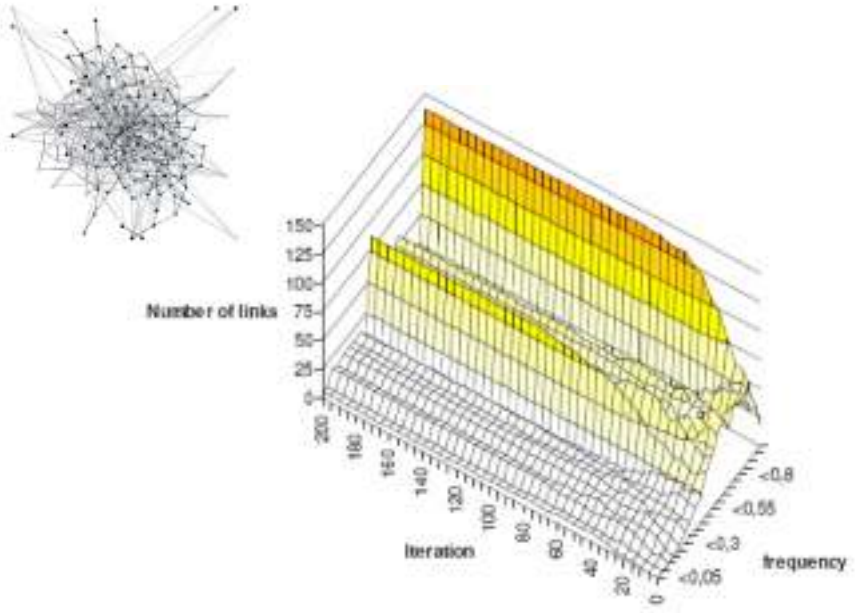
- Same rules than before for attitude dynamics

Experiments

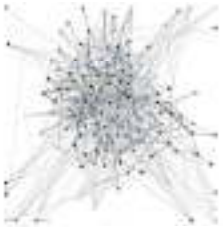
- A & B drawn at random
- β drawn at random
- Density of links 5%
- Underlying structure: random network
- 160 agents

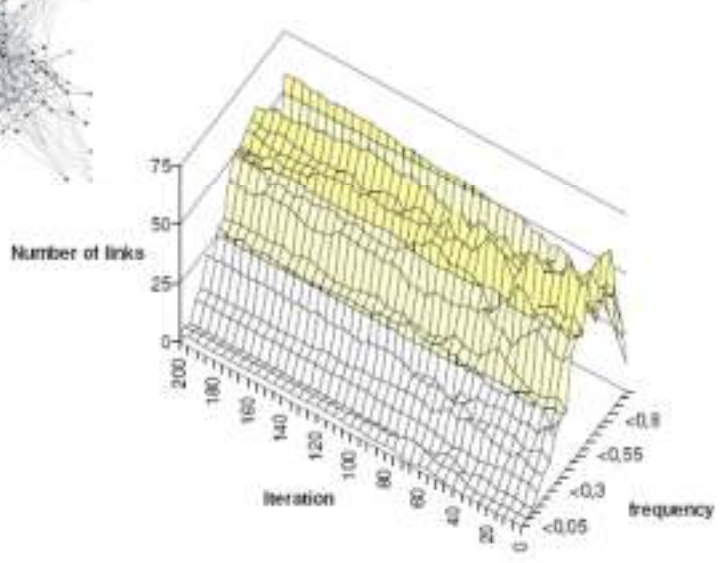
- Spring layout Netlogo
- Links from f=0 (white) to f=1 (black)
- Color of nodes = attitude A

Experiment A: Acceptance Threshold 0,5 - Rejectance Threshold 1,5 - central processing only

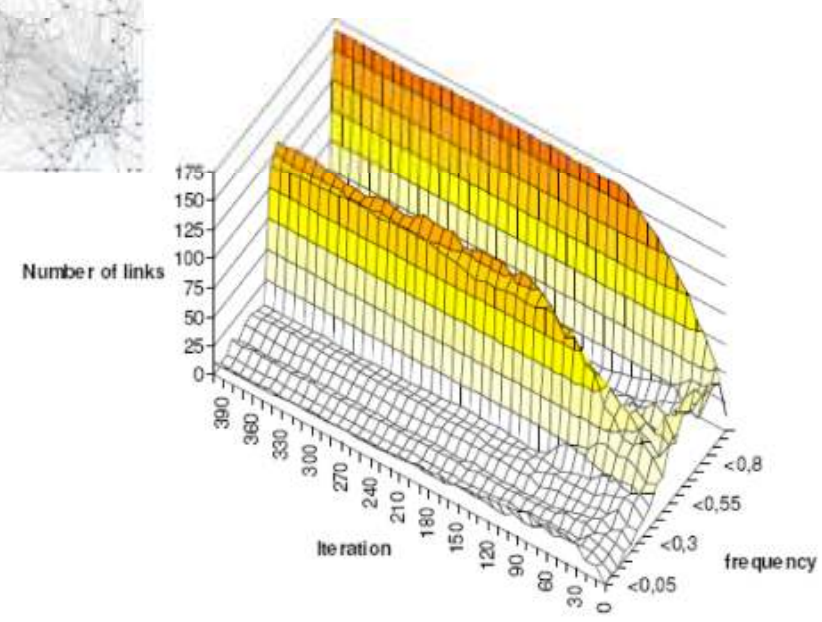


Experiment B: Acceptance threshold 0,25 – Rejectance threshold 1,75 - central processing only

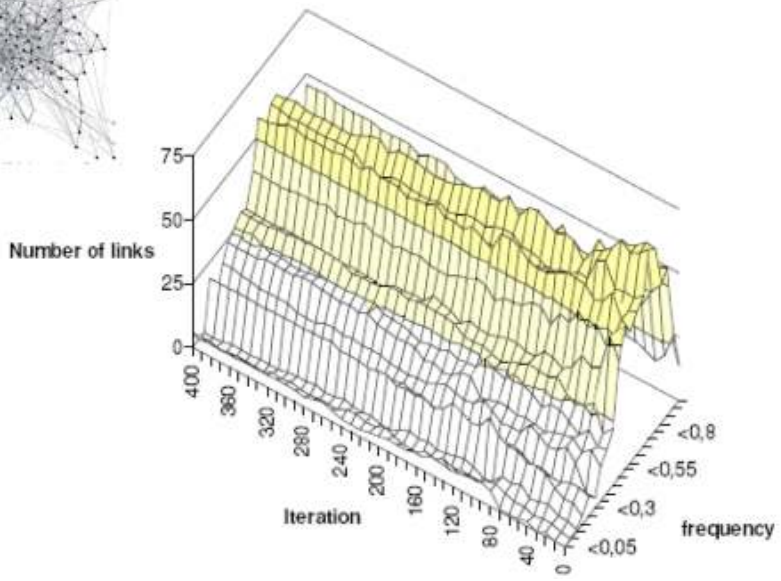


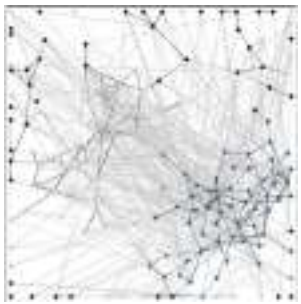
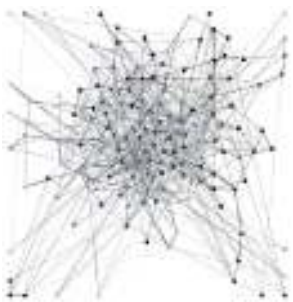


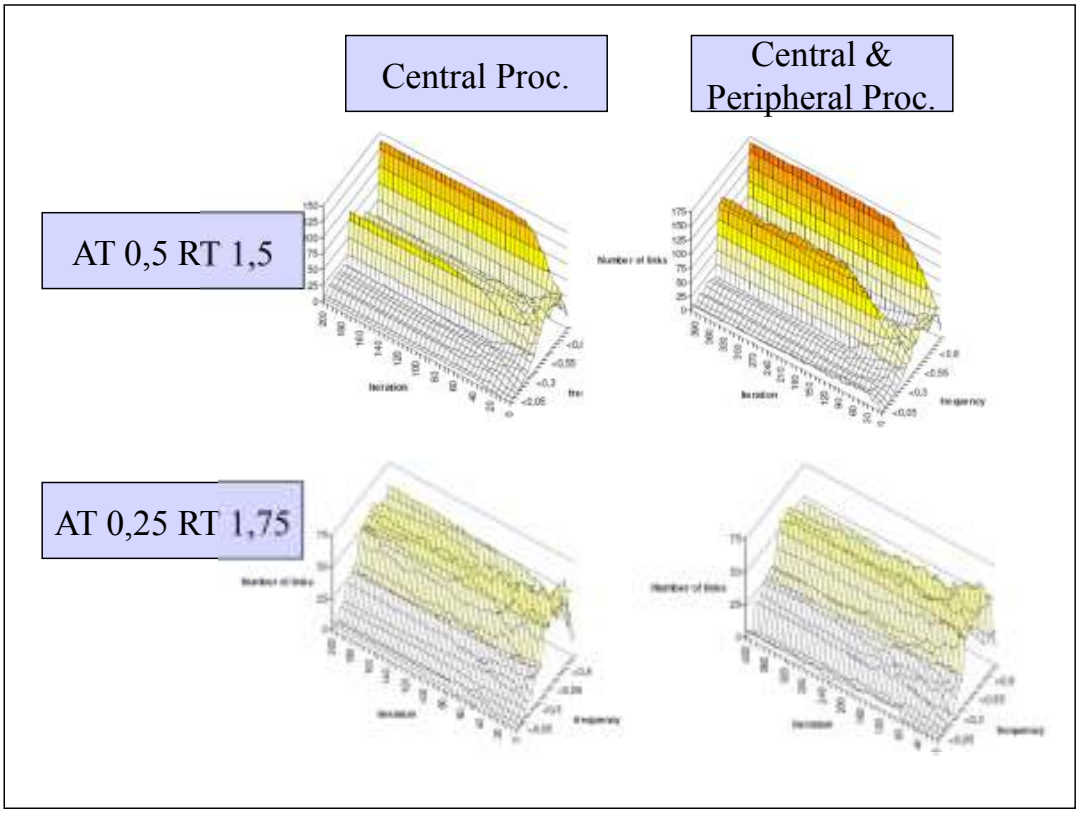
Experiment C 3: Acceptance threshold 0,5 – Rejectance threshold 1,5 – both central & peripheral processing



Experiment D: Acceptance threshold 0,25 – Rejectance threshold 1,75 – both central & peripheral processing



	Central Proc.	Central & Peripheral Proc.
AT 0,5 RT 1,5		
AT 0,25 RT 1,75		



Conclusion

- Exploration à poursuivre…
- Difficulté à élaborer des indicateurs pertinents pour suivre l'évolution d'une structure
  - Prenant en compte la valeur des liens
  - La valeur des nœuds (A & B)
  - Leur évolution
- Difficulté de visualisation

Co-évolution

- Problèmes posés
  - Indicateurs
  - Visualisation
    - Graphe dynamique
    - Dynamique sur le graphe
- Changement de perspective ?
  - Permanence des liens vs. Évènements
  - Vers une vision fréquentielle / probabiliste

Plan

- Réseaux sociaux & Internet
  - Quelques phénomènes identifiés
- Programme de recherche SMARS
- Données empiriques
- Modèles
  - Modèles de réseaux
  - Modèles de réseaux dynamiques
  - Modèles de dynamiques sur réseaux
  - Co-évolution
- Expérimentations
- Perspectives

Expériences

- Economie expérimentale version réseau
- Supports possibles
  - Facebook / plateforme web 2.0
  - En laboratoire sous conditions contrôlées
  - Ethologie (Strangers into friends)
- Connaissance récupérable
  - Données expérimentales
  - Débriefing
  - Elicitation des stratégies …

Browsing the social web

- Principe:
  - Trouver une personne qui correspond au profil recherché (Cherche quelqu'un qui…)
  - En transmettant le message de proche en proche
  - Chaque fois que vous recevez un message
    - Notez votre Prenom + initiale du nom sur le message
    - Ajoutez 1 à votre compteur de nb message
  - Si vous correspondez au profil, gardez le message

Jouez … EN SILENCE…

- A chaque étape
  - Est-ce que je correspond au profil ?
    - Si oui je garde le message
  - Sinon,
    - Identification d'un voisin à qui passer le message
    - Passer le message

Première expérience : Browsing the social web

- Inspiration
  - Small world Milgram
  - Strength of weak ties Granovetter
- Idée générale : comment les individus utilisent leur réseau au-delà du premier degré ?

Profils recherchés

- Information
  - Neutre
  - Socialement valorisante
  - Socialement dévalorisante
- Opinion
  - Neutre
  - Valorisante intragroupe
  - Dévalorisante intragroupe
- Comportement/rôle social
  - Autorité / soutien (réputation)
  - Organisation
- Délégation de tâche / compétence particulière
  - Tâche technique (simple / complexe spécifique)
  - Tâche sociale

Browsing the social web

- Debriefing
- Limitations du jeu (réseau, comm.)
- Quelles stratégies ont été mises en œuvre
- Analyse des trajets longs
- Analyse des trajets courts

Browsing the social web

- Résultats généraux attendus
- Les profils dévalorisant correspondent à des parcours plus longs
- Existence de hubs ?
  - Rôle type pages jaunes

Trust Game règles du jeu

- Chaque joueur démarre avec un stock de 5 points
- Il choisit de les répartir à ses voisins partiellement ou en intégralité ou de les garder pour lui
- (Chaque point donné est multiplié par 2)
- L'objectif: totaliser un maximum de points
- Le jeu est itéré N fois

Jouez …. EN SILENCE

- Init
  - Noter au max 4 voisins
- A chaque étape
  - Noter (discrètement) sur la feuille le nombre de points donnés (ou pas) à chaque voisin
  - Révéler les points donnés à chaque voisin
  - Noter les points que vous avez obtenus
    - En faire la somme
    - Multiplier par deux
    - Noter le nouveau solde

Deuxième jeu : trust game

- Idée générale : avoir une idée des liens mobilisés préférentiellement (vs. Broadcasting)
- Avoir une idée de la réciprocité de ces liens
- Originellement (eco expé) joué à deux joueurs en synchrone et sans visualiser le partenaire

Trust Game: debriefing

- Qui a le maximum de points
  - Quelle stratégie mise en place ?
- Est-ce qu'il y a des perdants (solde = 0 en cours de partie) ?
  - Que s'est-il passé ?
- Autres stratégies…

Trust game : résultats attendus

- Effet de la dernière itération
  - Stratégie qui intègre l'anticipation de la fin de jeu ?
- Création de partenariats privilégiés
  - Sélection parmi ses voisins d'un partenaire avec qui ça marche bien
  - Première phase d'exploration / évaluation du voisinage ?
  - Tentative d'élimination de certains voisins ? Voire du partenaire ?

Coordination game

- Chaque joueur est tagué
  - Soit de Saint-Etienne
  - Soit pas de Saint-Etienne (Web Intelligence)
- A chaque tour chacun
  - Choisit de jouer le drapeau St-Etienne ou le drapeau Web Intelligence
  - Révèle le drapeau
  - Compte le nombre de drapeaux identiques au sien (sans se compter)
    - Si le drapeau joué correspond au tag, ajouter 2 points
  - Ajouter le résultat au solde

Jouez … EN SILENCE …

- A chaque tour :
  - Je détermine (discrètement) et je pose face cachée le drapeau que je vais jouer
  - Je révèle mon drapeau
  - Calcul des points

Coordination game

- Intérêt
  - Détection de communautés ?
  - Identifier des mécanismes locaux qui permettent la synchronisation globale

Coordination game

- Débriefing
  - Stratégies utilisées

Coordination game : résultats attendus

- Effets de groupe
  - St-Etienne contre le reste du monde
  - Chacun a joué son tag au premier tour ?
- Tendance à l'homogénéisation

Conclusion expériences

- Cadre expérimental
  - Mieux spécifié
  - En particulier sur aspect réseaux
- Deux types de jeu à mener
  - En live (importance du débriefing)
  - En réseau (Appli facebook ?) à grande échelle (données)

Conclusion

- Cadre modélisation compréhensive de l'évolution / dynamique des réseaux sociaux
- Données / Modèles / Expériences
  - Pour mieux comprendre
  - Pour anticiper
  - Pour mieux designer
- Besoins méthodologiques pour l'articulation des différents aspects

SESSION 8

Usages (du point de vue social)

Dix propriétés des coopérations faibles : analyse des réseaux sociaux en ligne

D. CARDON
(ORANGE LABS, FRANCE)

10 propriétés des coopérations faibles

Dominique Cardon, Laboratoire des usages Orange Labs
dominique.cardon@gmail.com

unrestricted

avec Christophe Aguiton, Jean-Samuel Beuscard, Maxime Crepel, Bertil Hatt, Nicolas Pissard et Christophe Prieur

une innovation sans organisation ?

faire remonter l'usager dans les laboratoires

...mais l'innovation est venue d'ailleurs

craigslist

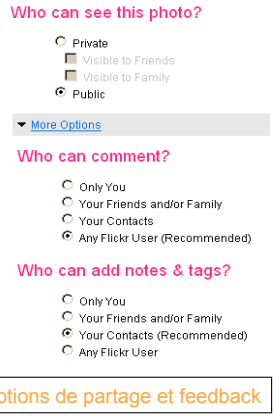
Blogosphère

Wikipedia

flickr BETA

- Partage de photos
  - avec sa famille, ses amis, et au-delà
- Feedback
  - commentaires, notes, tags, favoris
- API ouvertes
  - Publication vers la plupart des blogs du marché
  - Galaxie d'applis web qui exploitent données Flickr hors de Flickr

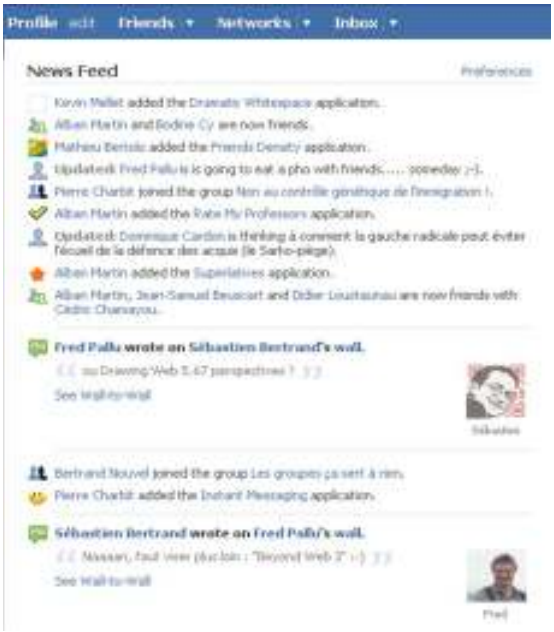
plate



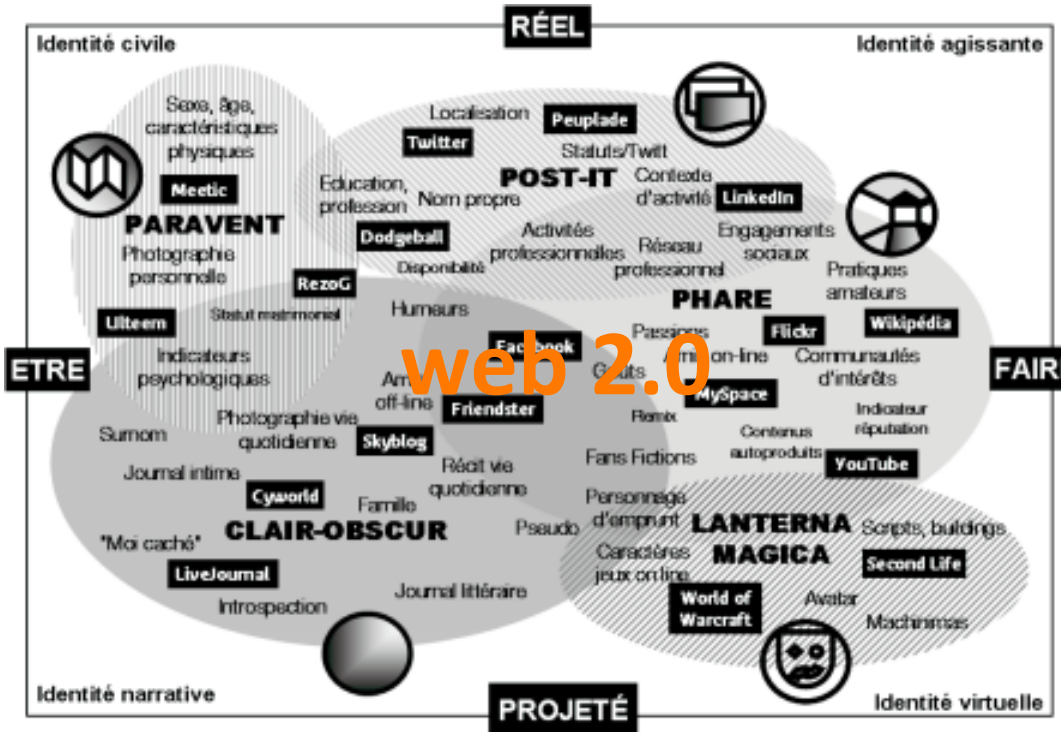
options de partage et feedback

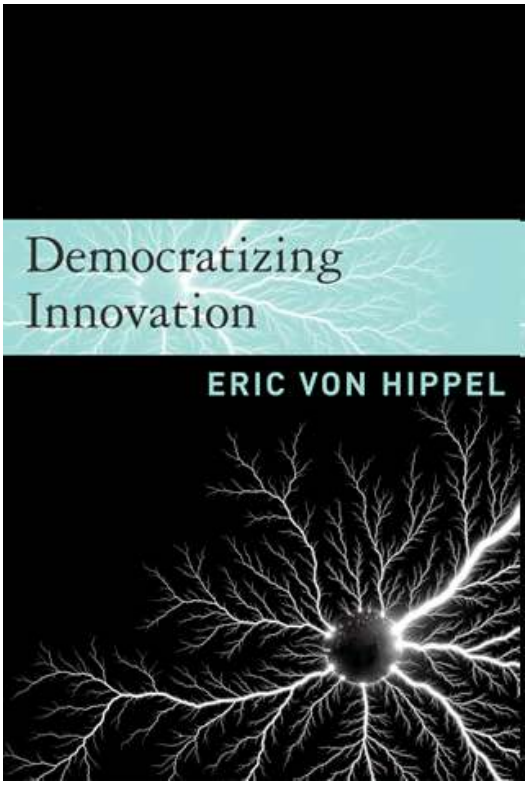
Flickrgraph : visualisation des contacts

ajout de notes sur une photo


Profile
Friends
Networks
Inbox
News Feed

twitter
Home | Your profile | Invite | Public timeline | Badges | Settings | Help | Sign out
What are you doing?
Update
Welcome back,
Dominique
Archive
Recent
What You And Your Friends Are Doing



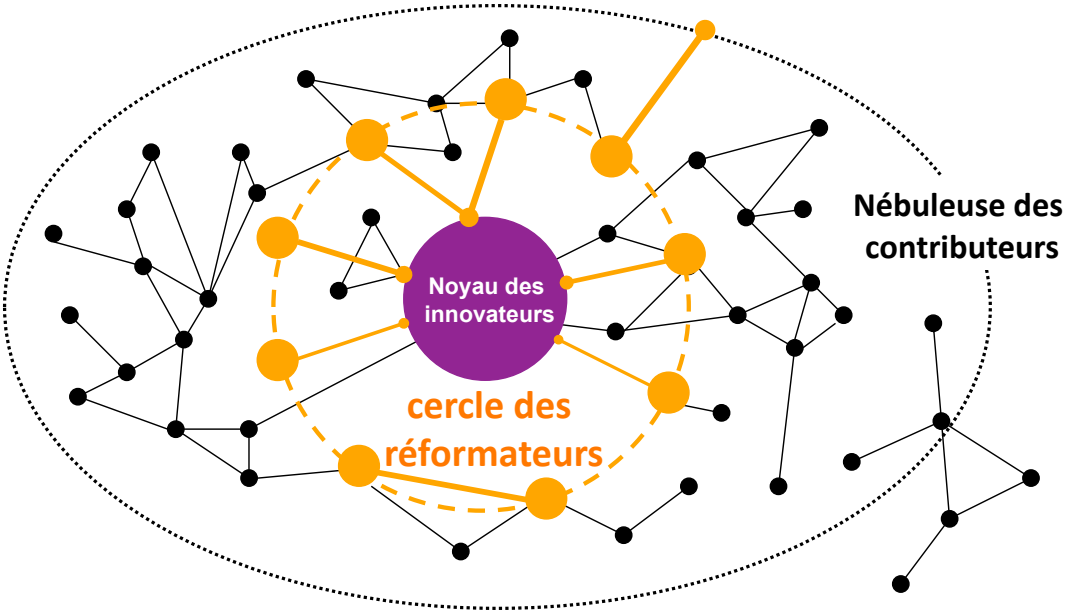


L'innovation ascendante

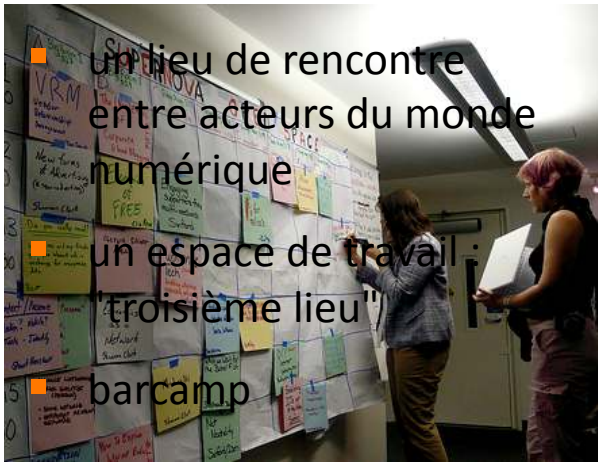
1. problème utilisateur
2. solution locale
3. révéler son innovation
4. diffusion réticulaire

- articulation avec le marché si…

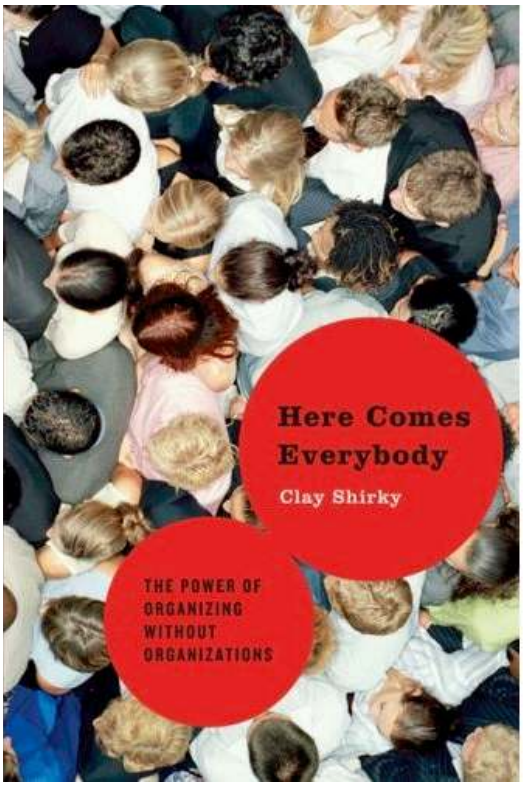
trois cercles des innovations ascendantes



la Cantine, un co-workingspace à Paris



- un lieu de rencontre entre acteurs du monde numérique
- un espace de travail : "troisième lieu"
- barcamp



Clay Shirky,
Here Comes Everybody.
The Power of Organizing
without Organizations

1.partage

2.coopération

3.esprit

les faibles ont des coopérations fortes

la force des coopérations faibles

10 propriétés des coopérations faibles

l'individu démonstratif

la visibilité comme opportunité de coopération

les "amis" ne sont pas forcément des amis

les communautés viennent après le réseau

navigation non-hiérarchique

des engagements hétérogènes

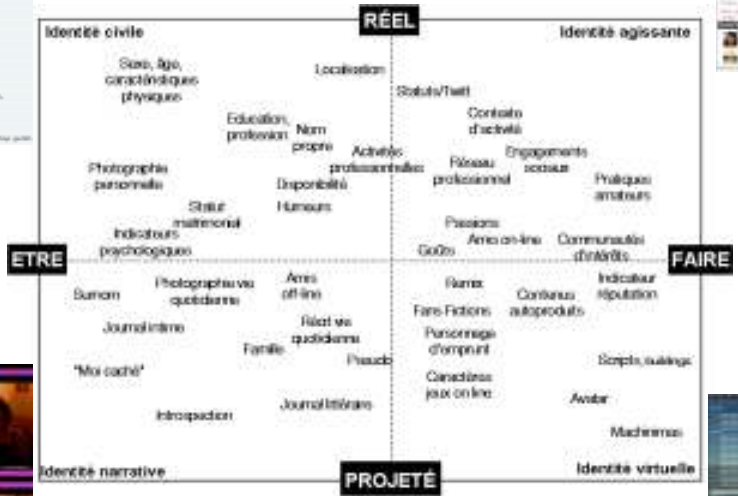
les grands ont besoin des petits

la qualité par le nombre

les hiérarchies produites par le comportement des autres

un mouvement perpétuel

l'individu démonstratif

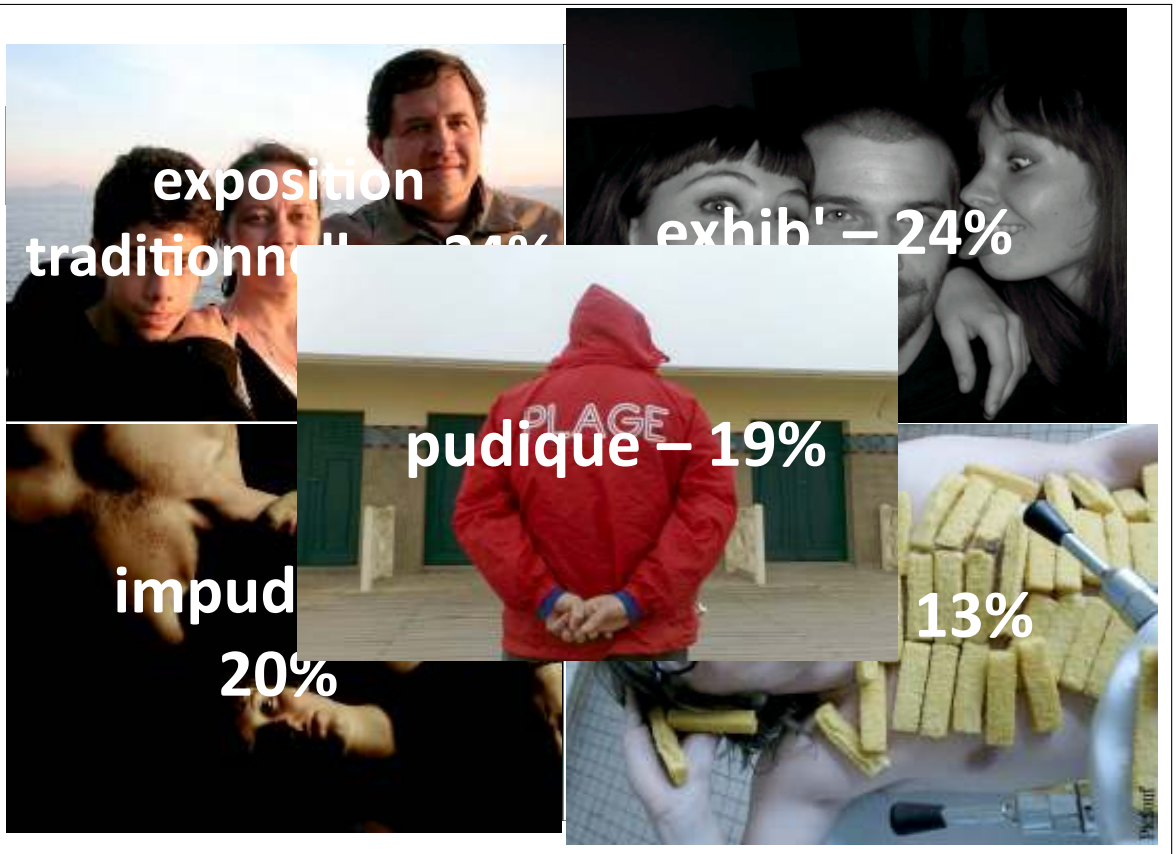


Que montre-t-on de soi sur Internet ?
jeu-enquête en ligne (15 000 répondants)
http://sociogeek.com

31% 46% 8% 15%
20% 49% 12% 19%
54% 13% 10% 23%
29% 24% 11% 36%
35% 30% 12% 23%
9% 34% 20% 37%
61% 28% 7% 5%
62% 28% 6% 4%
39% 38% 19% 4%
59% 32% 3% 6%
29% 19% 7% 45%
46% 48% 3% 3%
65% 19% 6% 9%
76% 7% 10% 7%
36% 21% 20%
17% 16% 37% 30%

une exposition mesurée

- Taux d'exposition moyen : 2,04 (2,1 : H ; 1,9 : F)
- Pas de corrélation entre la fréquence d'usage des réseaux sociaux et le niveau d'exposition
- …mais une corrélation entre le nombre d'amis et le niveau d'exposition

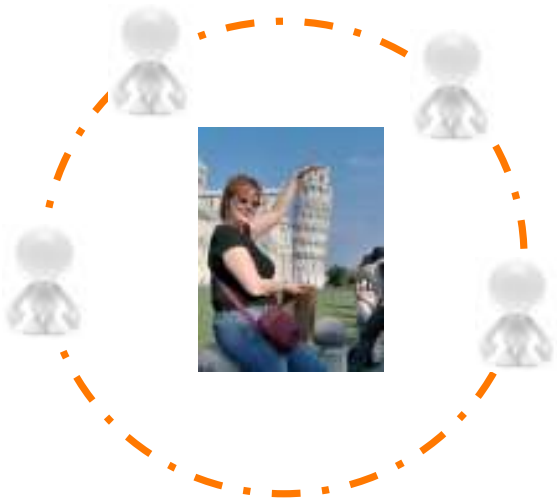


Les enjeux de protection de la vie privée

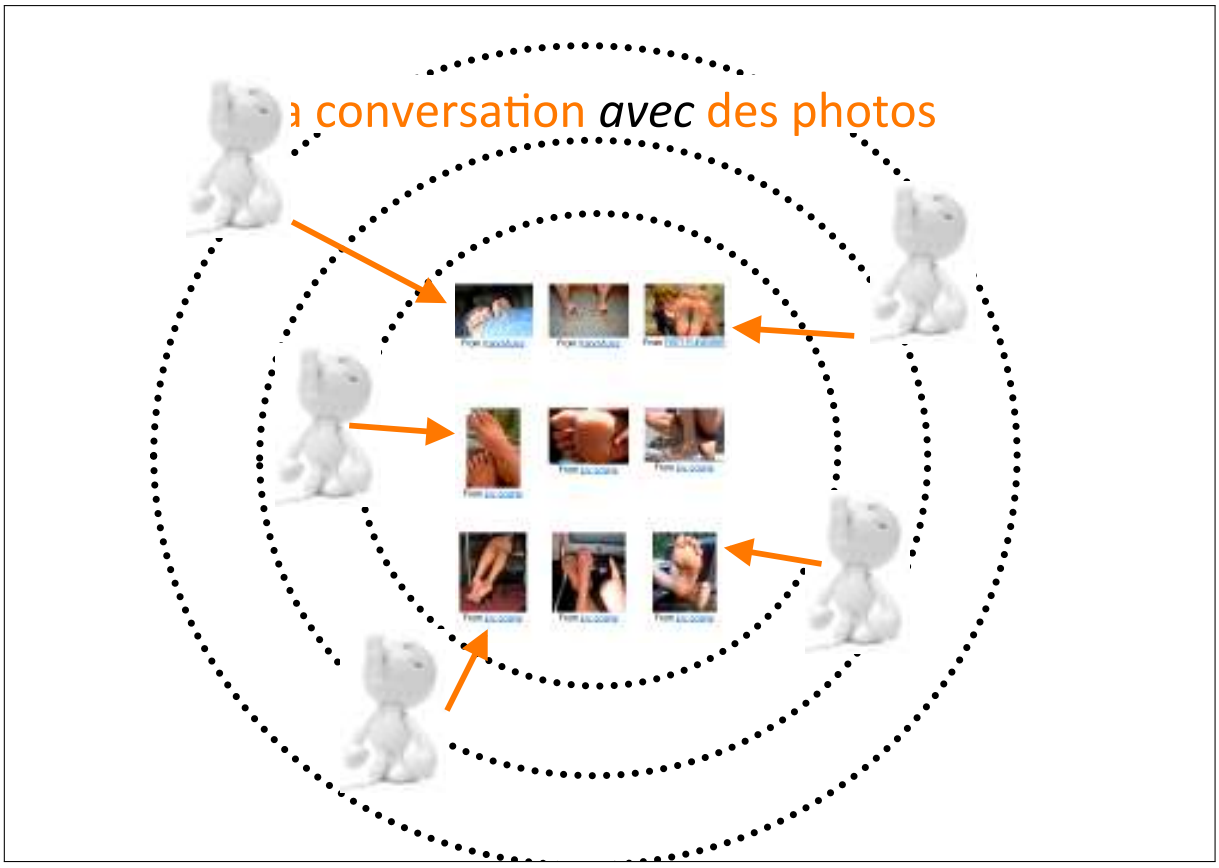
- Sculpture de soi
- Toutes les données identitaires ne sont pas des informations
- Une "fermeture" croissante des pages personnelles
- Maintenir la visibilité réduite des plateformes de réseau social de type conversationnel (Facebook)

la visibilité comme opportunité de coopération

La conversation *autour* de la photo

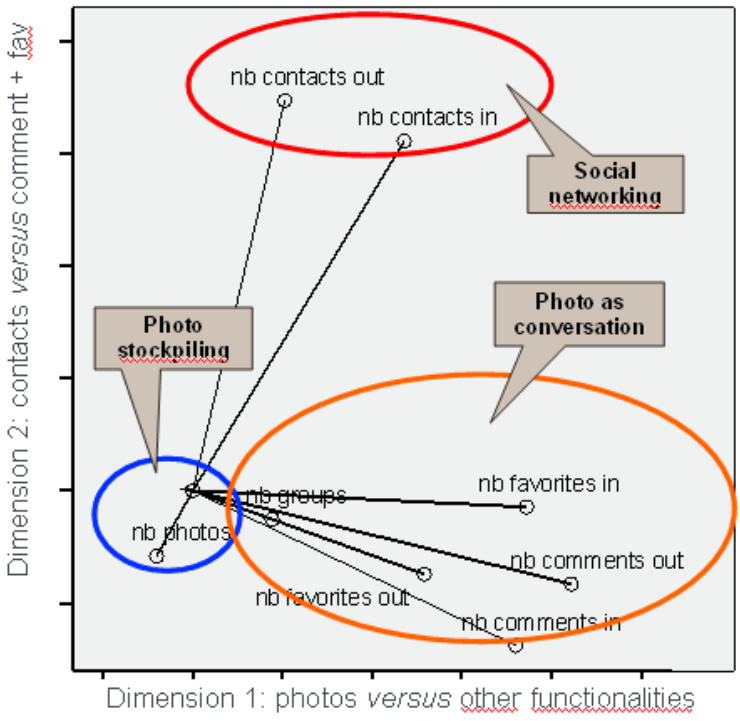


La conversation *dans* la photo



Trois usages de Flickr

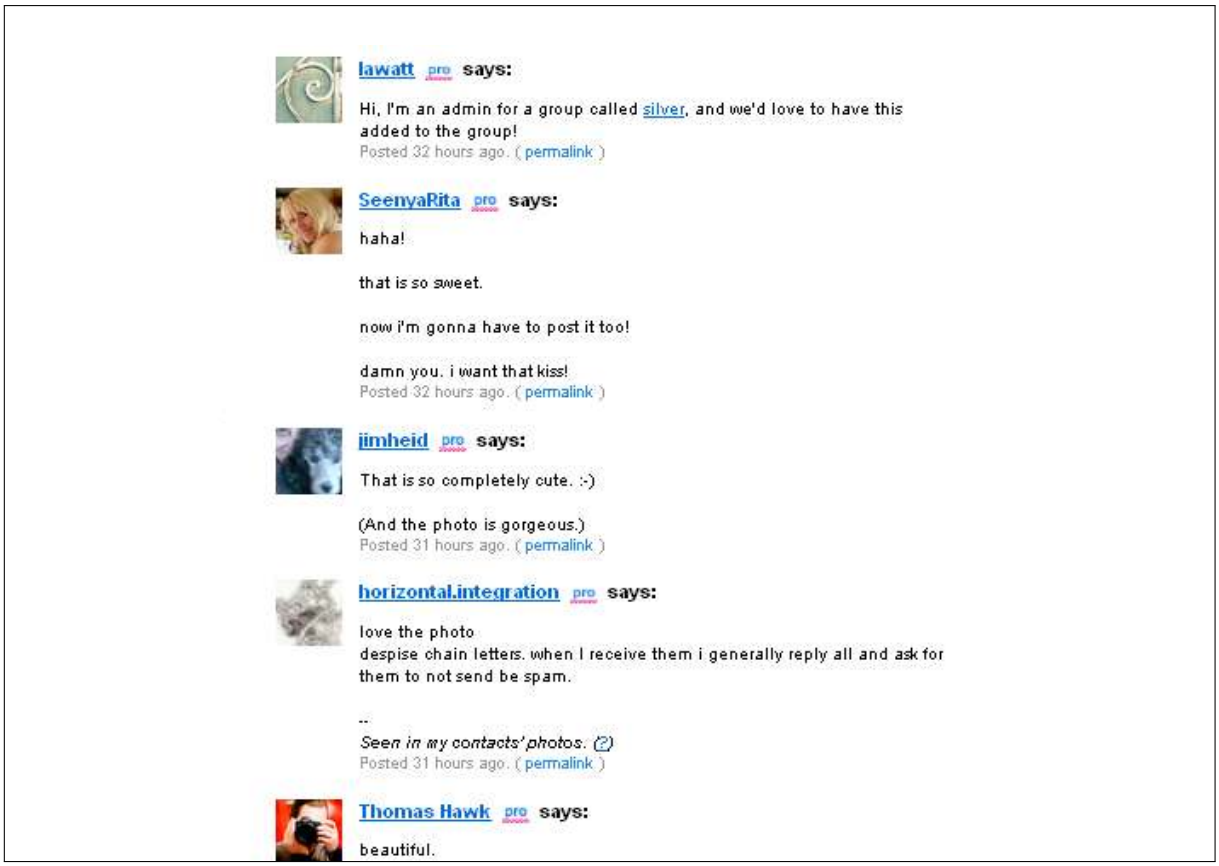
- **Stockage**
- **Relation**
- **Conversation**



repas d'avion

marabout – bout de ficelle

The Panopticon: Pictures Of Surveillance Cameras



lawatt pro **says:**

Hi, I'm an admin for a group called silver, and we'd love to have this added to the group!
Posted 32 hours ago. (permalink)

SeenyaRita pro **says:**

haha!

that is so sweet.

now i'm gonna have to post it too!

damn you. i want that kiss!
Posted 32 hours ago. (permalink)

jimheid pro **says:**

That is so completely cute. :-)

(And the photo is gorgeous.)
Posted 31 hours ago. (permalink)

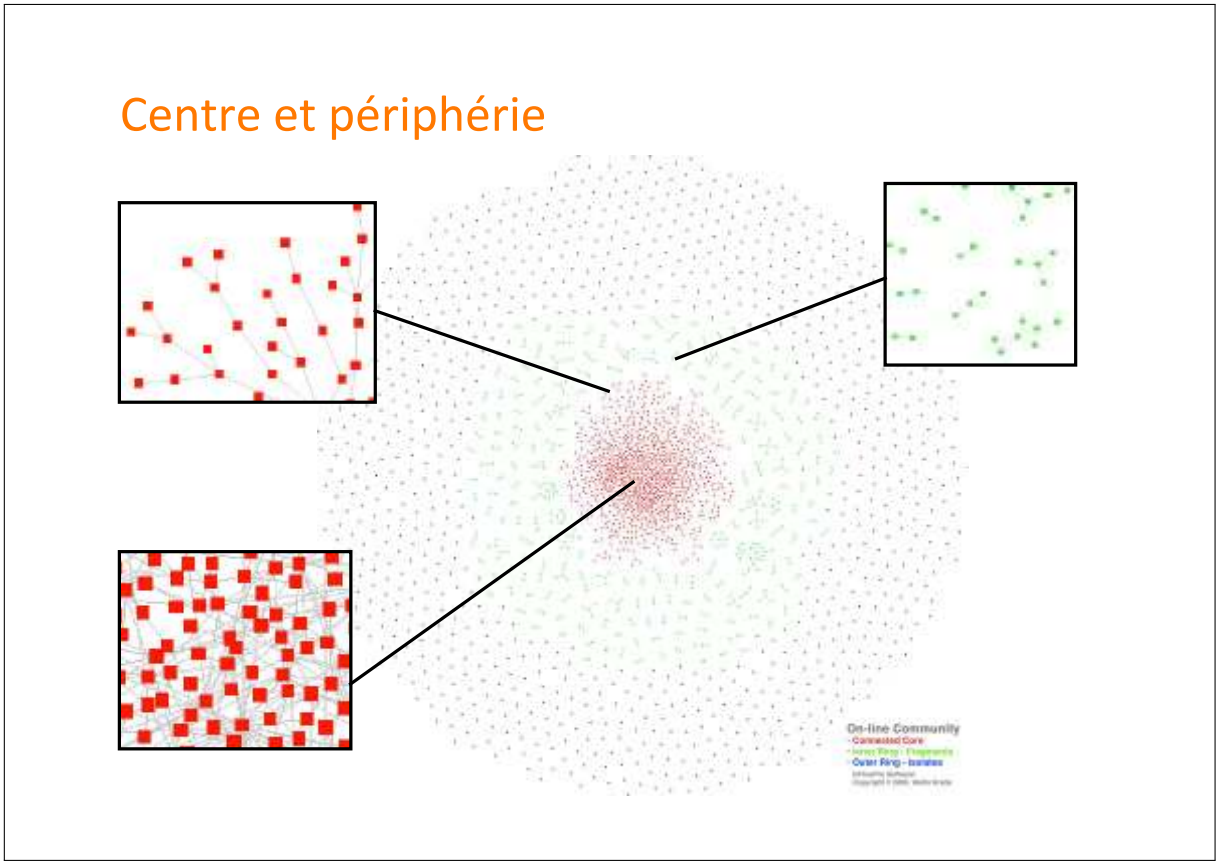
horizontal.integration pro **says:**

love the photo
despise chain letters. when I receive them i generally reply all and ask for them to not send be spam.

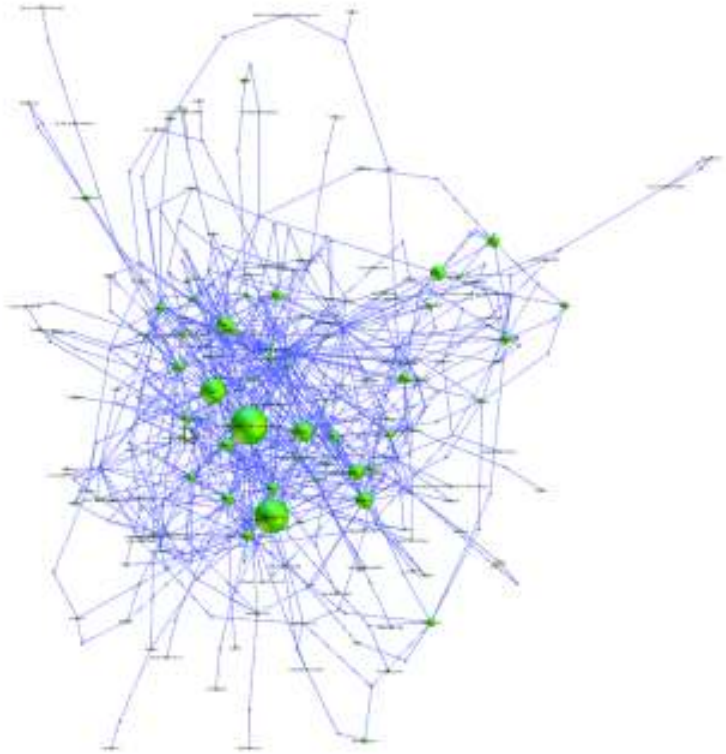
--
Seen in my contacts' photos. (?)
Posted 31 hours ago. (permalink)

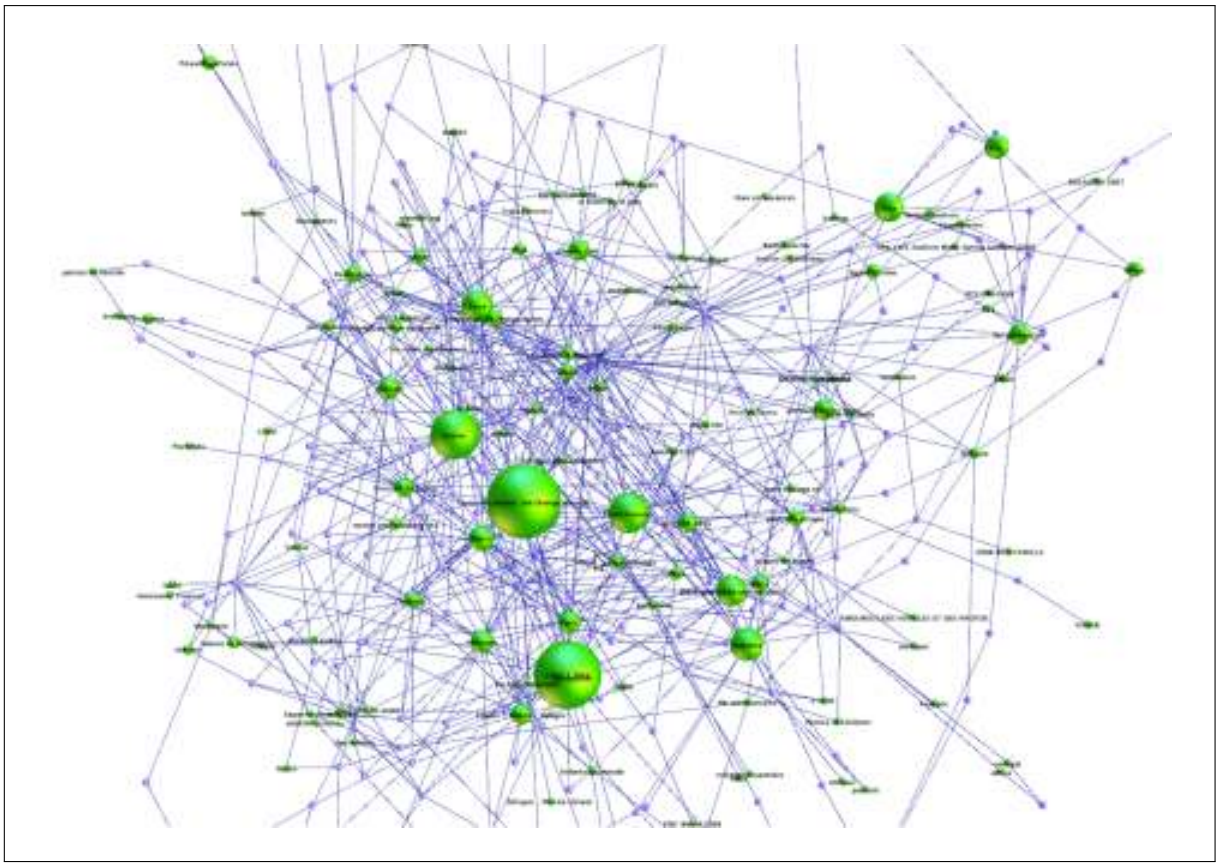
Thomas Hawk pro **says:**

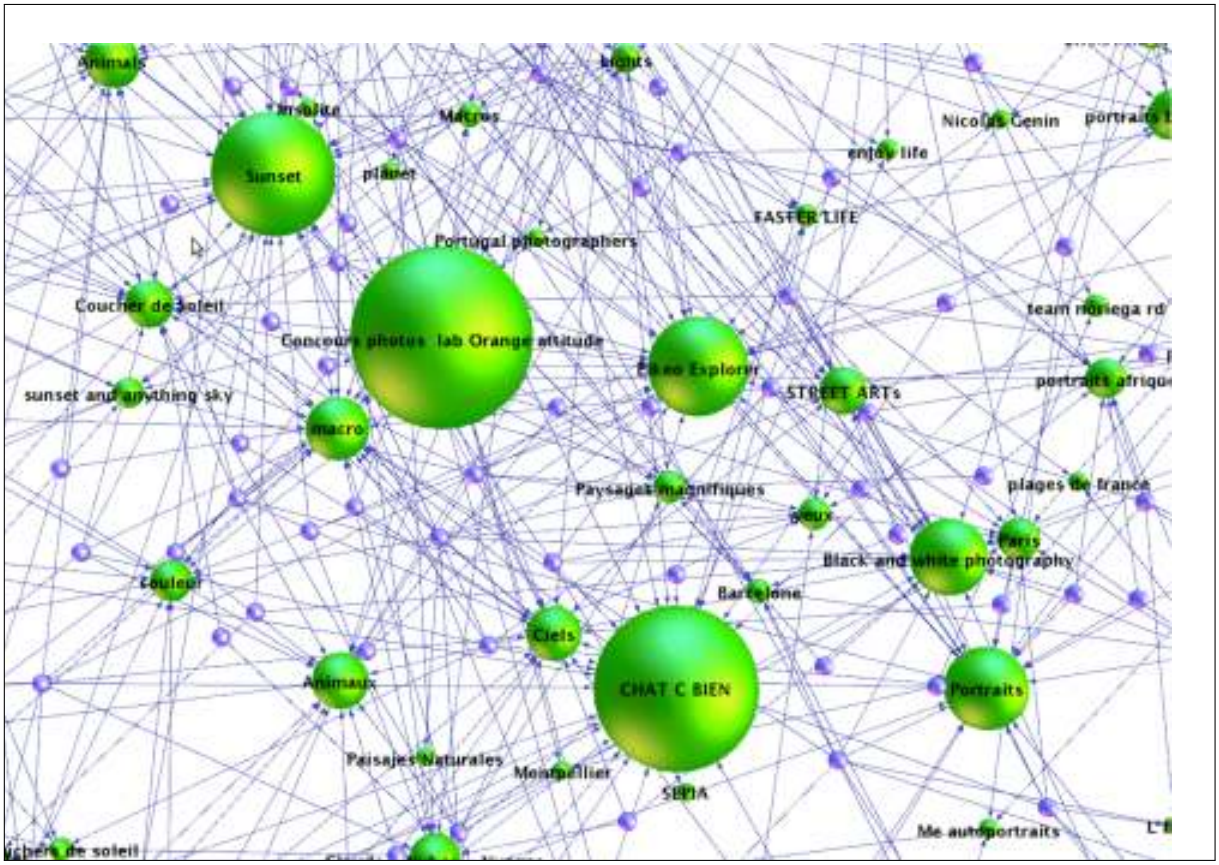
beautiful.



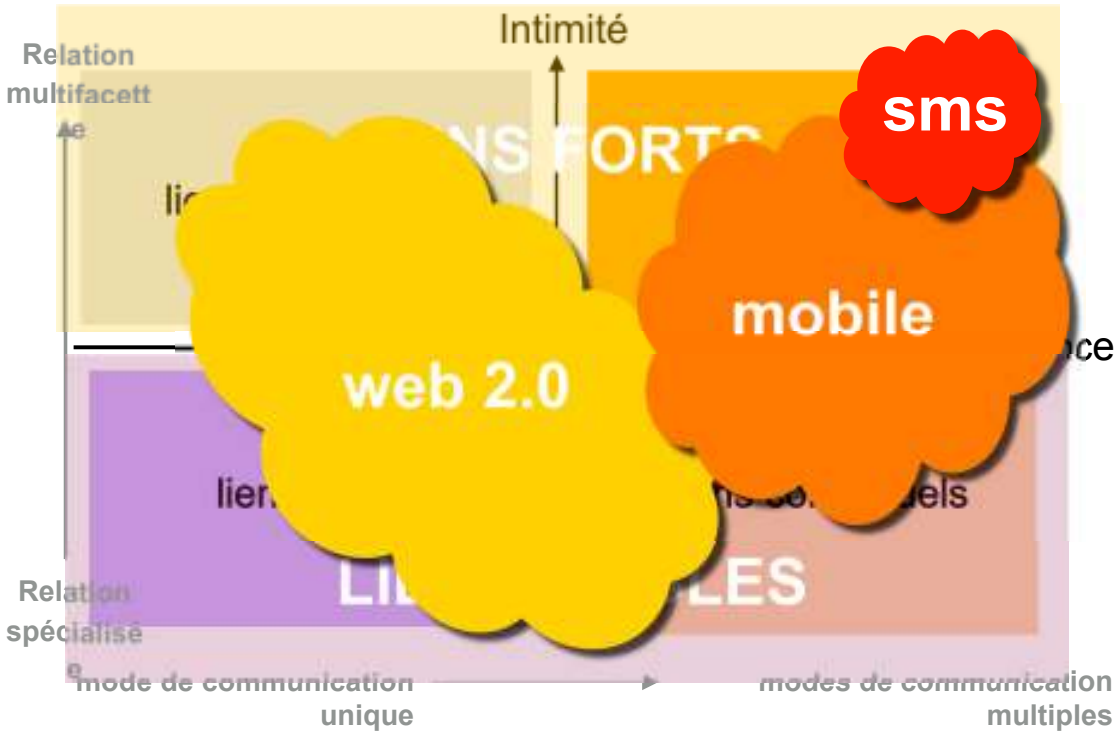
liens entre membres et groupes sur Pikeo

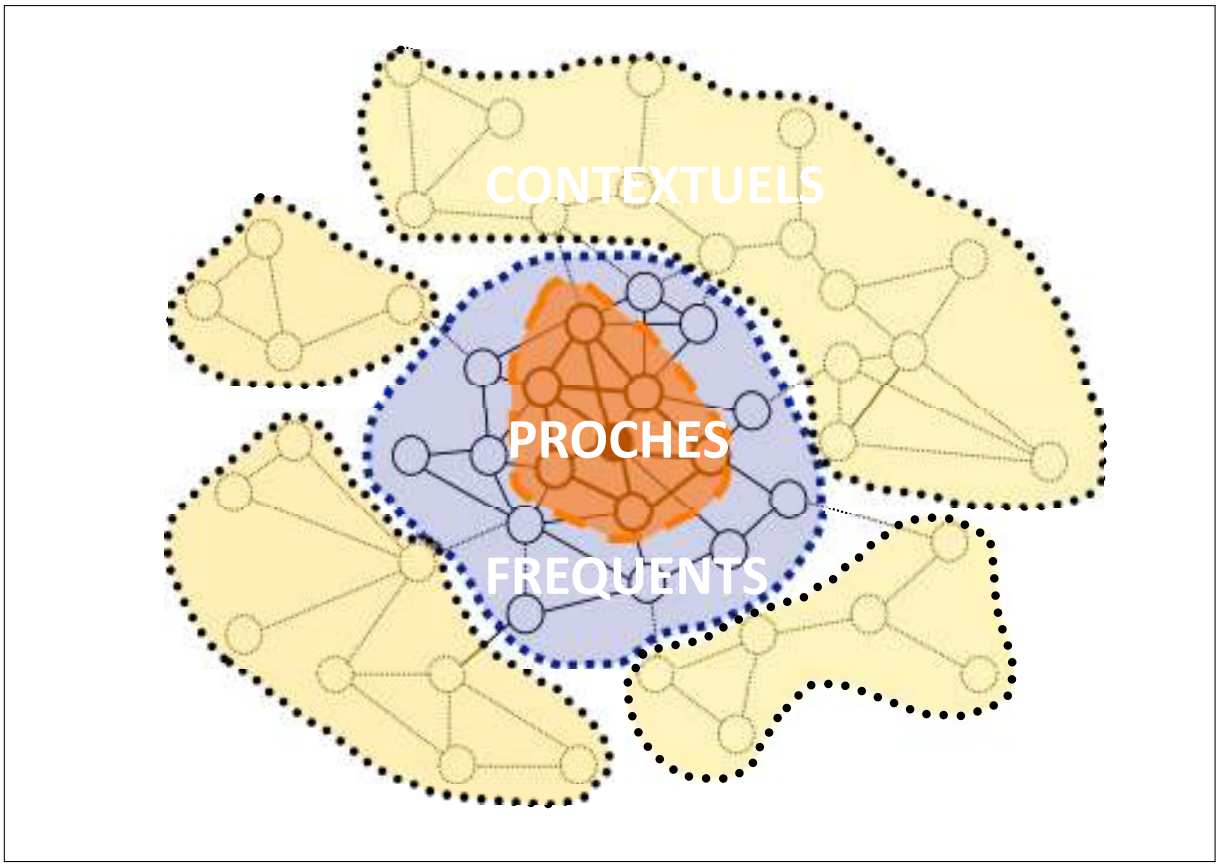




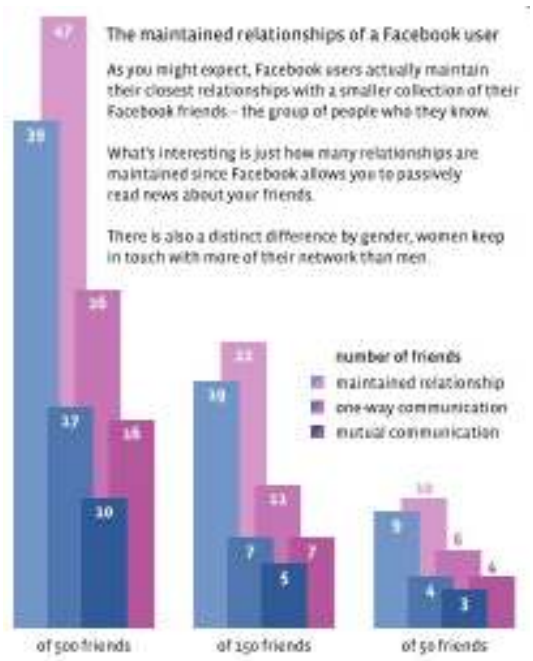

Animals
Macros
Nicolas Genin
enjoy life
Sunset
planet
FASTER LIFE
Portugal photographers
Coucher de Soleil
Concours photos lab Orange altitude
sunset and anything sky
macro
STREET ARTs
Paysages magnifiques
plages de france
Paris
Black and white photography
Couleur
Barcelone
Ciels
CHAT C BIEN
Animaux
Portraits
Paisajes Naturales
Montpellier
SEPIA
Me autoportraits

les "amis" ne sont pas forcément des amis





Lee Byron, TomLento, Camreon Marlow, Itamar Rosenn, Facebook Data



Amandine Trousseau est une petite pêche hâlée

Véronique Bisson
Bonjour mademoiselle, j'ai trouvé un pull rouge (ou ainsi que mon manteau). Quant à mes chaussures figures-toi qu'elles étaient bon marché en plus.

Amandine Trousseau
Aux Arcs ! t'ai vue hier soir arrivant à vélo bld St Marcel. Depuis le 91...

Sylvie Larame
tu te la racontes pas un peu là??

Véronique Bisson
Tu as vu ma belle, gi a tombé un peu pluvieux ici aussi le soleil ?

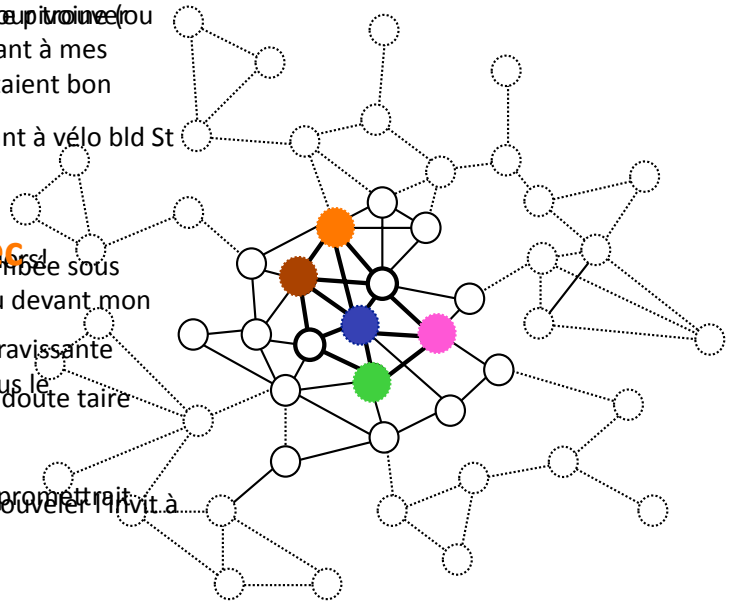
Amandine Trousseau
Toi tu nous la joueras banane flambée sous peu alors laisse-moi rêver un peu devant mon miroir !

Amandine Trousseau
Oui ! J'ai surtout vu que tu étais ravissante avec de belles pompes par dessus le marché !

Marguerite Thiel
Oui, enfin, une photo ferait sans doute taire les mauvaises langues aussi...

Amandine Trousseau
Chanceuse ! J'en profite pour renouveler l'invit à dîner, date à déterminer ! bises

Mathieu Vire
Certes...



Vendredi : La conversation avec les lieux forts : les grandes choses du quotidien

Bastien Cori : "Nous avions autrefois un mot pour désigner ceux qui souffrent de trouble déficitaire de l'attention avec hyperactivité (TDAH). Nous les appelions des garçons" (Christopher Lane, Comment la psychiatrie et l'industrie pharmaceutique ont médicalisé nos émotions)

Emmanuel Kavou,
Et les filles, c'est le syndrome de quel trouble émotionnel ?

Bertil Hamerson
Obsessionnel compulsif

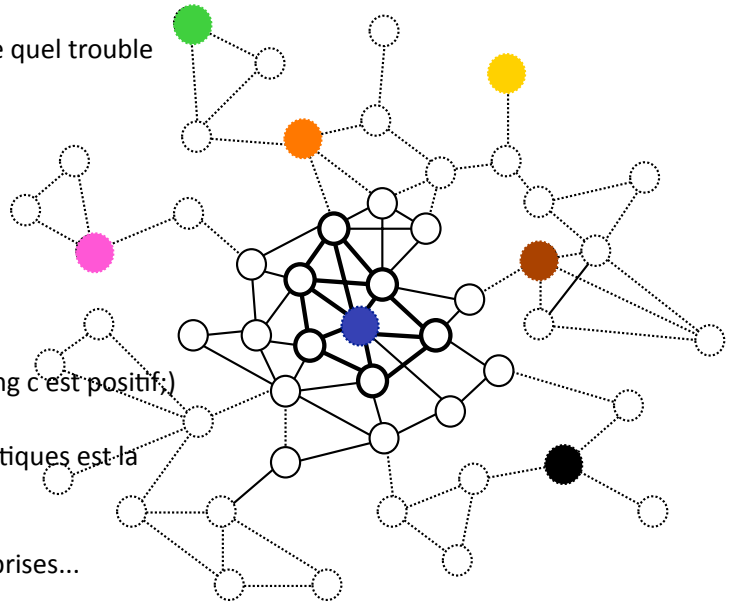
André Gavi
Et hystérique.

Emmanuel Kavou
je dirai plutôt multitasking

Stéphanie Baudricourt
@merci emmanuel, multi-tasking c'est positif;)

Olivier Malu
yeah, la pathologisation des pratiques est la vraie maladie

Cendrine Rach
surtout quand elles sont incomprises...

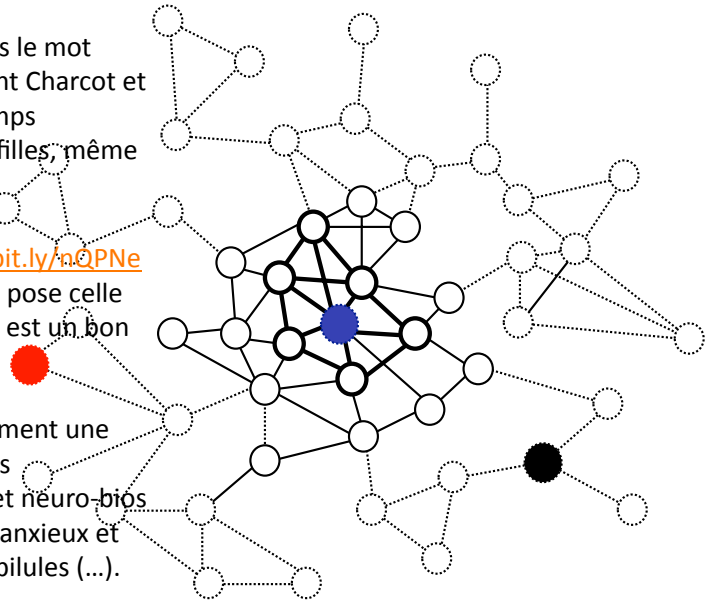


Bastien Cori : "Nous avions autrefois un mot pour désigner ceux qui souffrent de trouble déficitaire de l'attention avec hyperactivité (TDAH). Nous les appelions des garçons" (Christopher Lane, Comment la psychiatrie et l'industrie pharmaceutique ont médicalisé nos émotions)

Françoise Falloir
Les filles ? c'est le contraire : après le mot "sorcière"- si peu rationaliste -, vint Charcot et l'hystérie. Ca fait assez peu de temps finalement que les filles sont des filles, même quand elles sont un peu agitées...

Olivier Malu
sur le versant psychiatrie http://bit.ly/nQPNe
La question de la standardisation pose celle des standards choisis. Le DSM en est un bon exemple (voir Ehrenberg aussi)

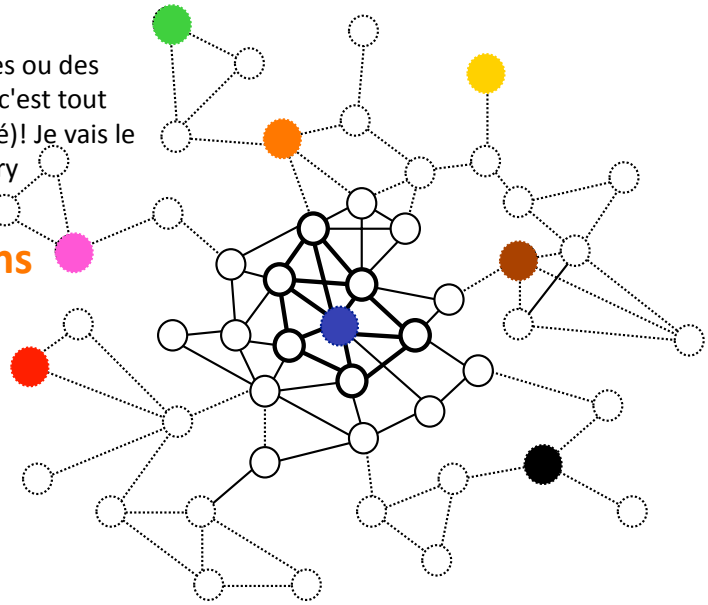
Bastien Cori
yep, le bouquin de Lane est justement une enquête très fouillée sur les petits arrangements entre psychiatres et neuro-bios pour transformer les timides, les anxieux et les agités en consommateurs de pilules (…).



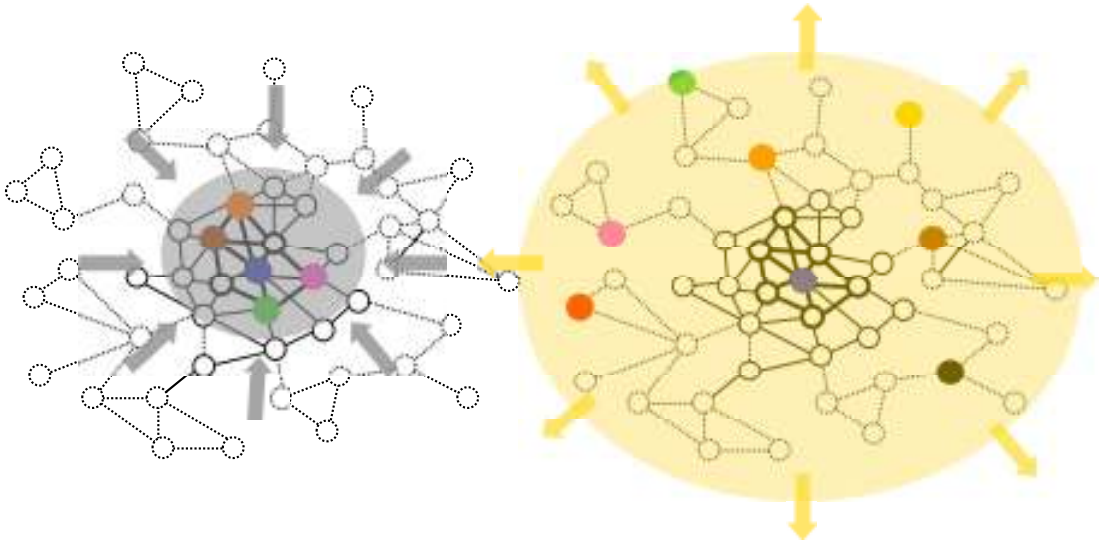
Bastien Cori : "Nous avions autrefois un mot pour désigner ceux qui souffrent de trouble déficitaire de l'attention avec hyperactivité (TDAH). Nous les appelions des garçons" (Christopher Lane, Comment la psychiatrie et l'industrie pharmaceutique ont médicalisé nos émotions)

Olivier Malu
remarque entre vendre des pilules ou des services pour la cyber-addiction, c'est tout comme (bien que moins structuré)! Je vais le lire. Il a fait un article dans (Theory Psychology, 2008, vol 19, n°1

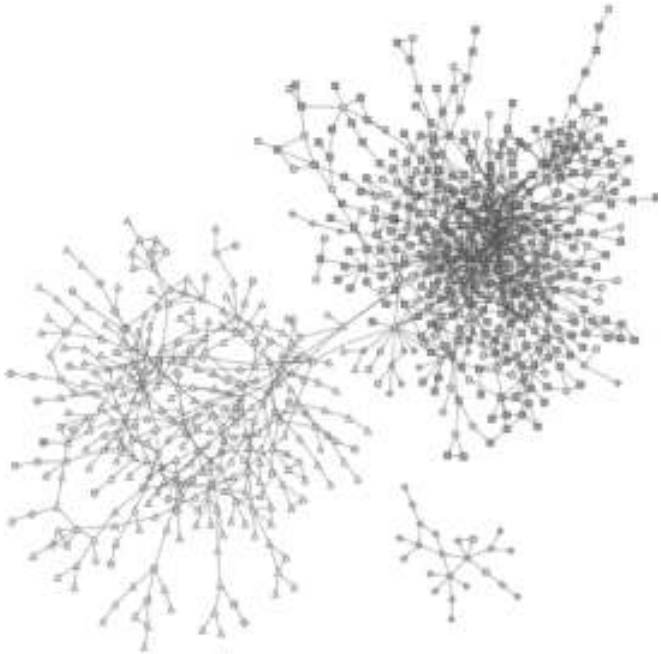
La conversation étendue vers les liens faibles permet la circulation des liens informationnels



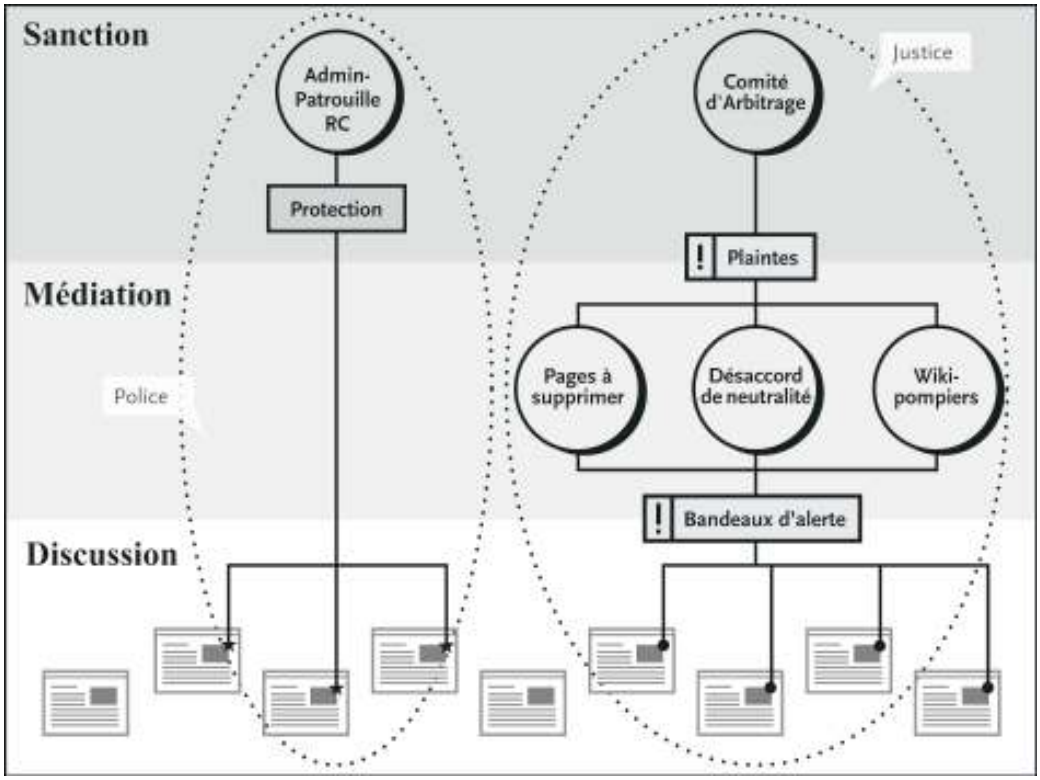
petite conversation grande conversation



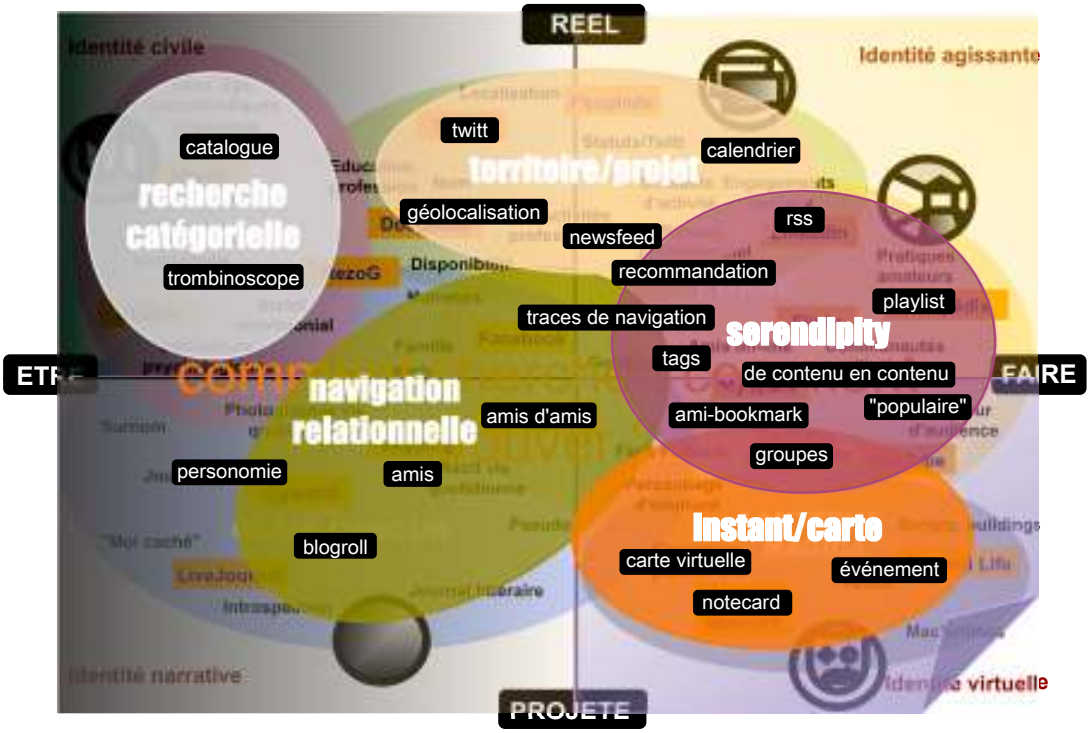
circulation de l'information entre cliques



les "communautés" viennent après le réseau

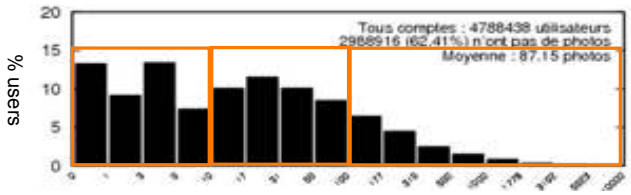


une navigation non-hiérarchique

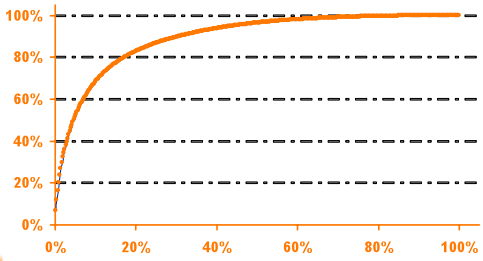


des engagements hétérogènes

6 - heterogeneous involvements: quantity



photos uploaded on Flickr



- Flickr
  - 62% no photo
  - 1 – 10
  - 11 – 100
  - 101 – 10.000
- Wiki DRS FTR&D
  - 4 users out of 9 contributed to articles
  - 63% of users have created at least a page
  - half pages are not modified after being created
  - 20% of the most active contributors have done 80% of wiki overall modifications

répartition des contributeurs de Wikipédia

	Wikipédia francophone	Wikipedia anglophone
Robots	**70** / **0,02%**	**226** / **0,006%**
IP	**286 350** / **90.22%**	**3 392 048** / **86,33%**
Inscrits	**30 985** / **9.76%**	**536 452** / **13,65%**
Top 100	**100** / **0,03%**	**100** / **0,002 %**
Total	**317 405**	**3 928 826**

57

répartition des contributions dans Wikipédia

	Wikipédia francophone	Wikipedia anglophone
Robots	**817 366** / **13,25%**	**3 075 590** / **5,04%**
IP	**998 361** / **16,18%**	**15 049 614** / **24,68%**
Inscrits	**2 693 865** / **43,65 %**	**38 702 676** / **63,48%**
Top 100	**1 661 296** / **26,92%**	**4 140 475** / **6,79 %**
Total	**6 170 888**	**60 968 355**

58

les "grands" ont besoin des "petits"

zélotes et bons samaritains (Anthony, Smith et Williamson, 2005)

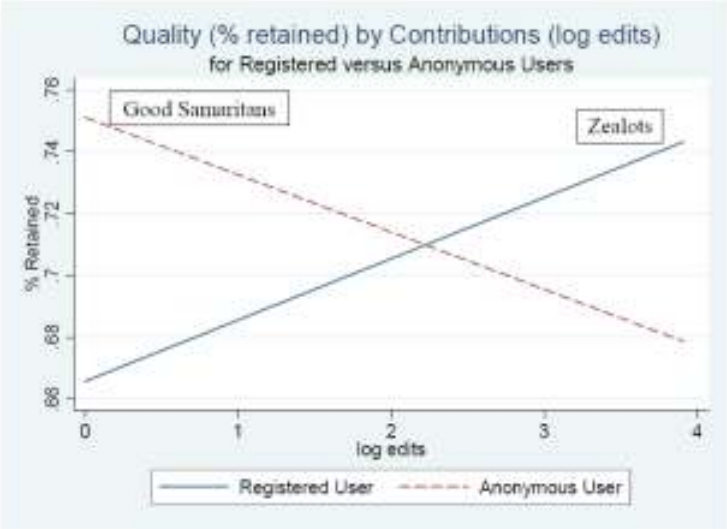
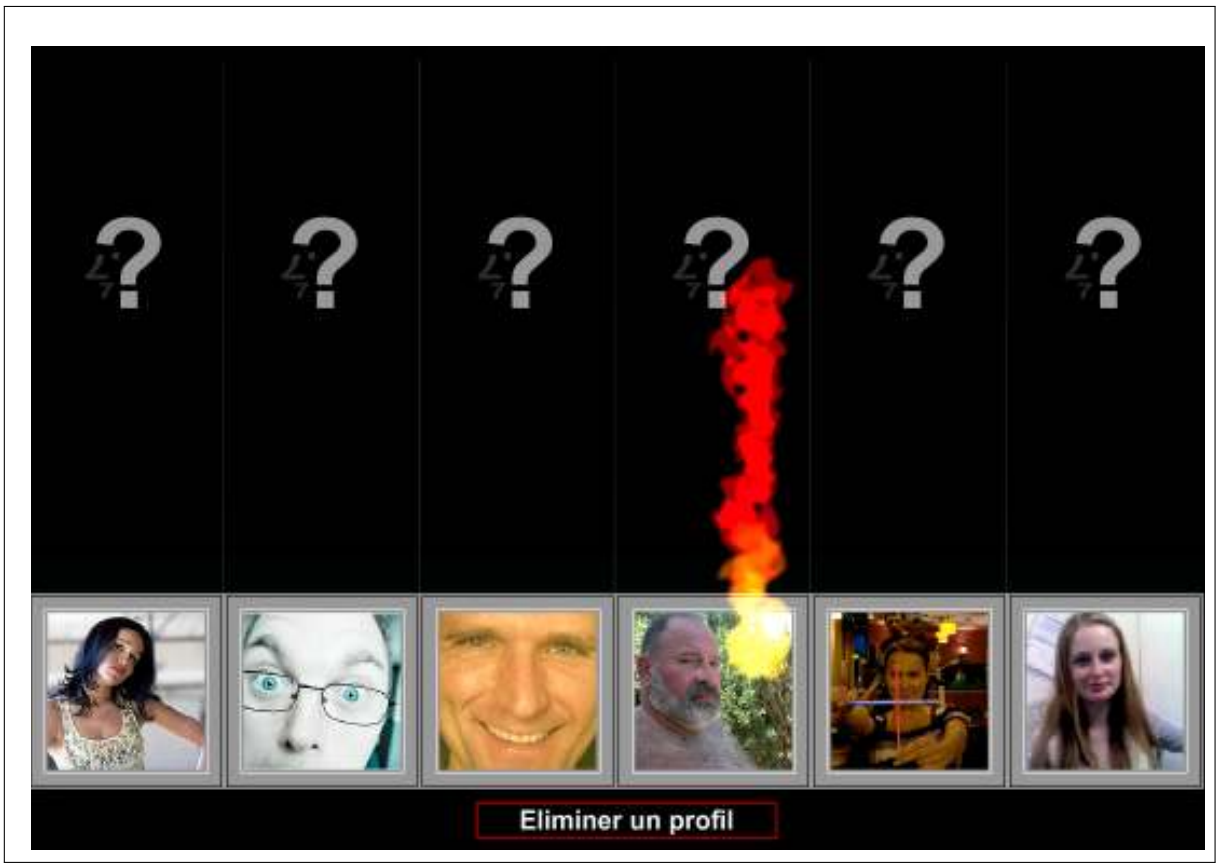
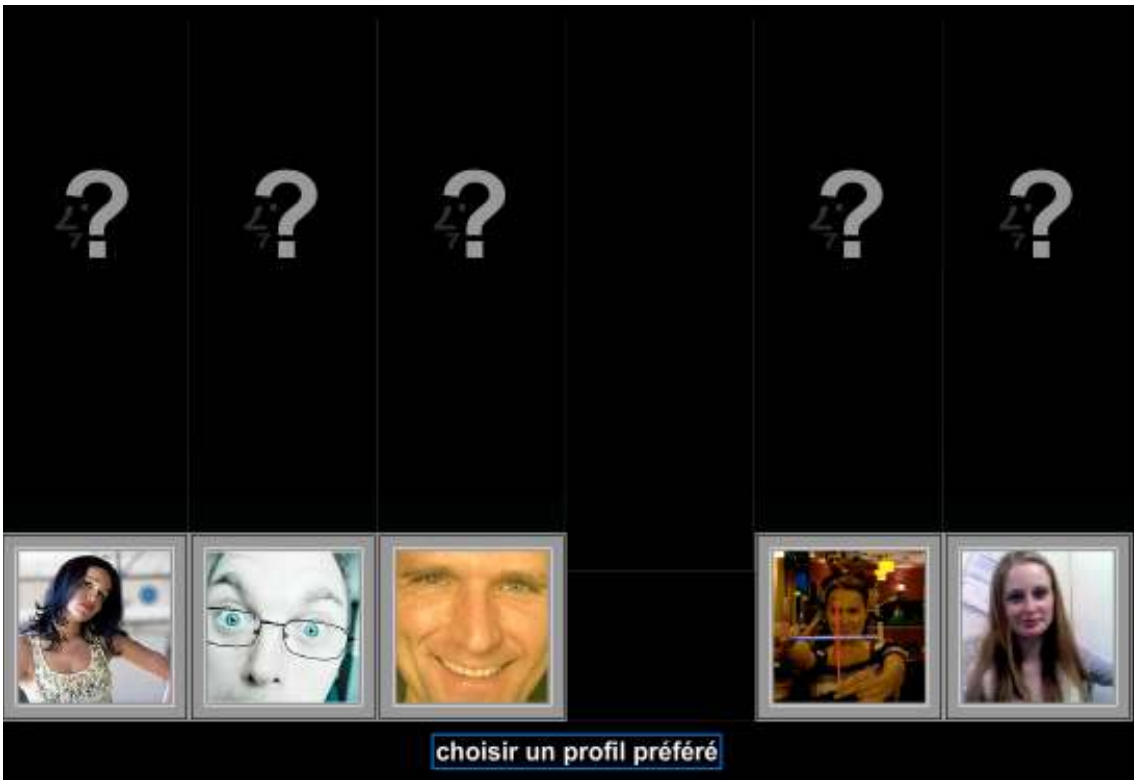


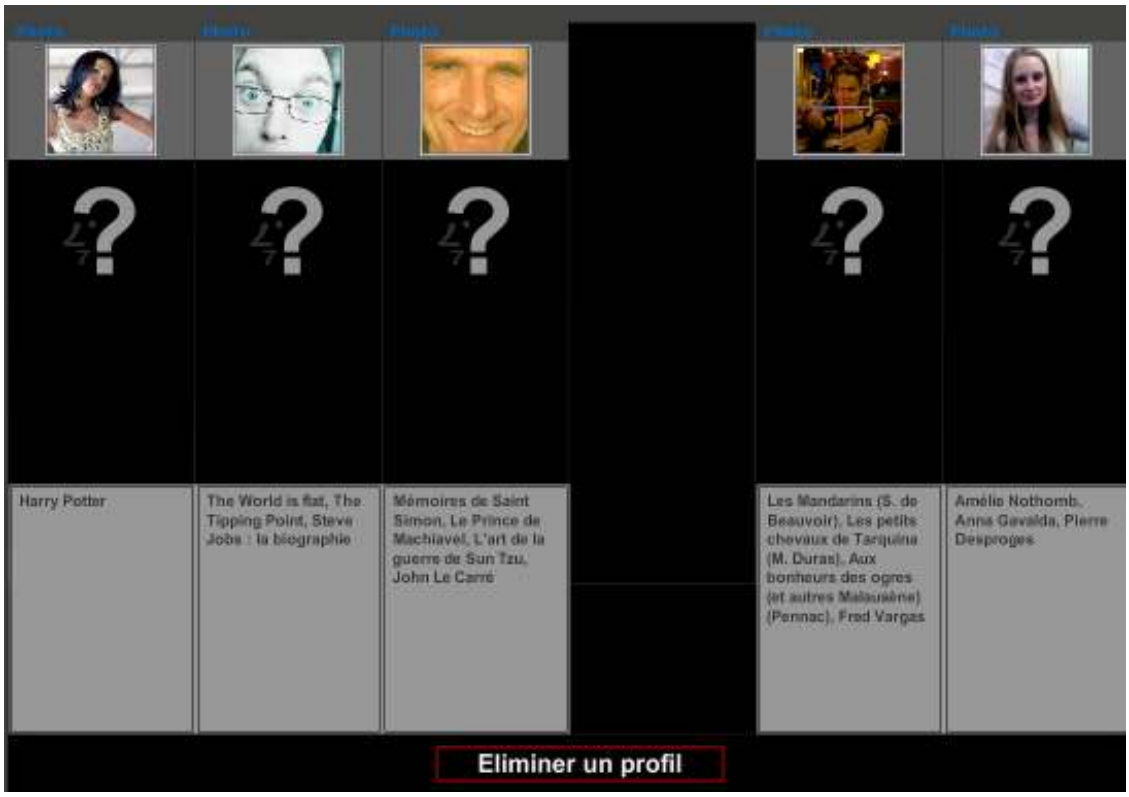
Figure 1. Quality of *Wikipedia* Contributions by Number of Contributions for Registered and Anonymous Users.

des hiérarchies produites par le comportement des autres

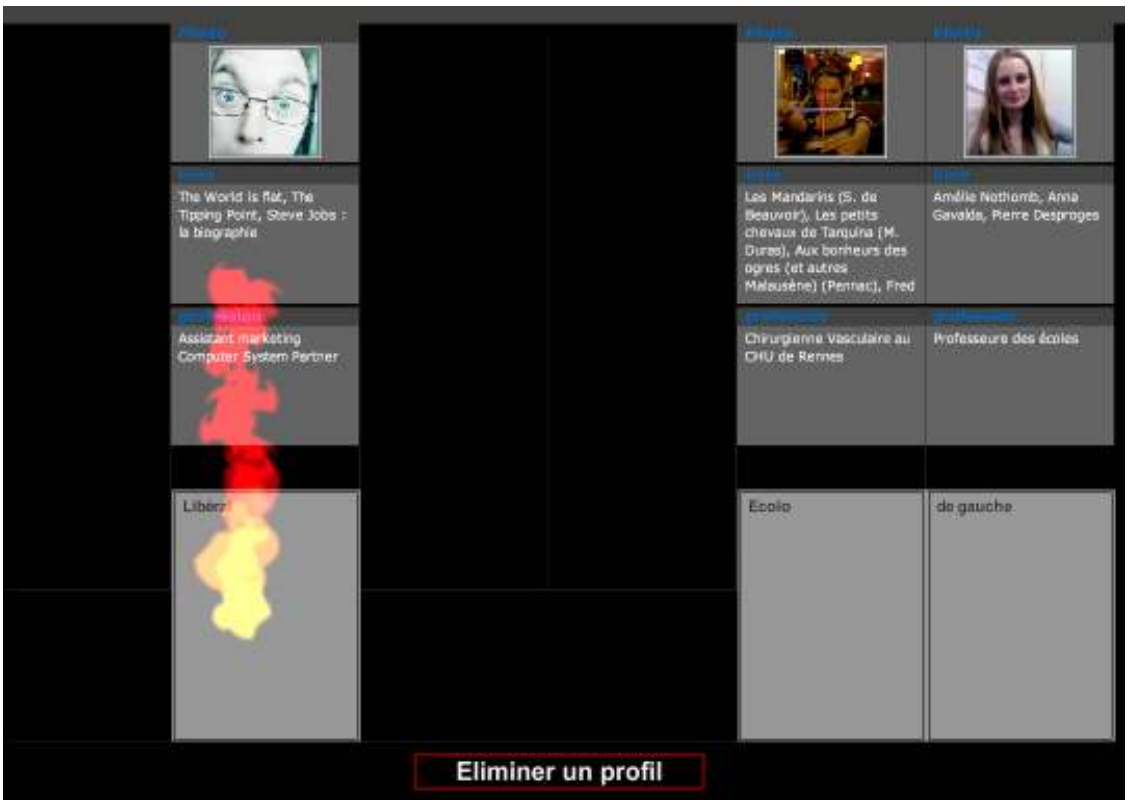
Stecher K., Counts S., 2008, "Thin Slices of Online Profile Attributes", *Proceedings of ICWSM'08*, Seattle
Cliquez sur une des variables à droite, pour faire apparaître les informations dans les 6 profils.
photo
nombre d'amis
productions personnelles
films préférés
sexe
à propos de moi
derniers messages
orientation sexuelle
livres préférés
situation matrimoniale
âge
statut
citations
religion
musiques préférées
profession
à la recherche de
groupes
opinions politiques
TV show préférés
diplômes
vu par mes amis

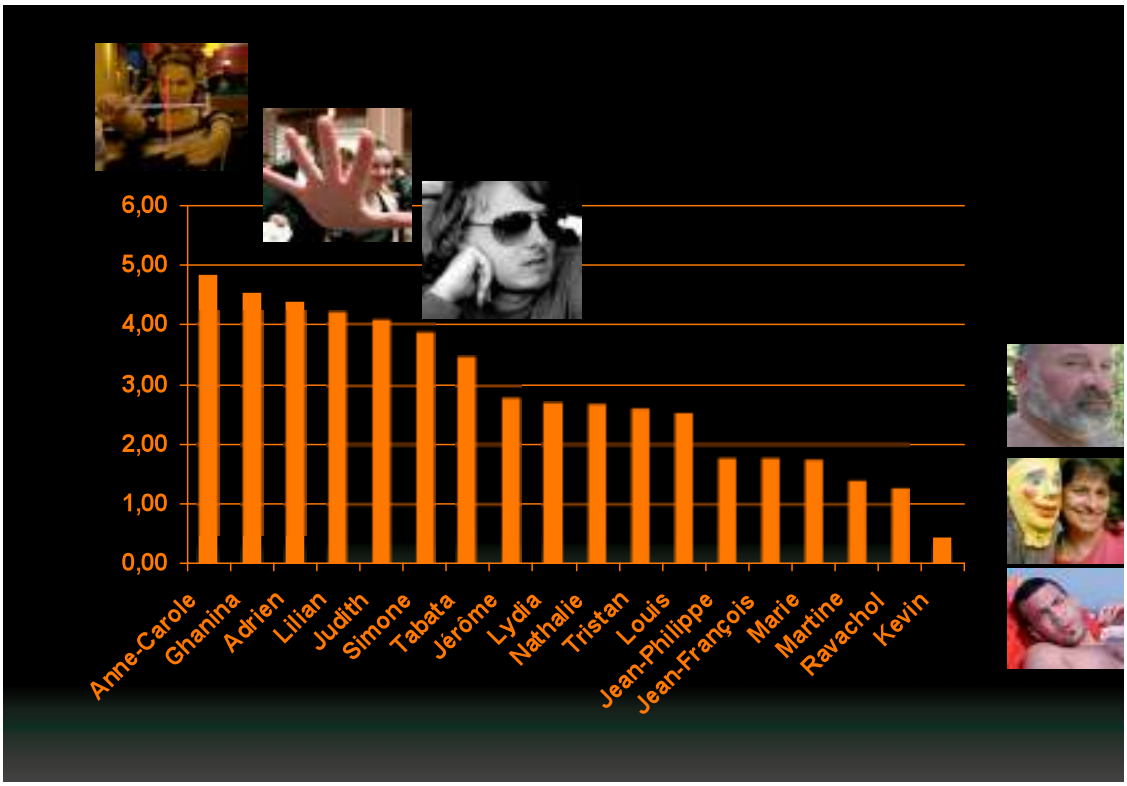

Eliminer un profil


choisir un profil préféré


Harry Potter
The World is flat, The Tipping Point, Steve Jobs : la biographie
Mémoires de Saint Simon, Le Prince de Machiavel, L'art de la guerre de Sun Tzu, John Le Carré
Les Mandarins (S. de Beauvoir), Les petits chevaux de Tarquinia (M. Duras), Aux bonheurs des ogres (et autres Malaurène) (Pennac), Fred Vargas
Amélie Nothomb, Anna Gavalda, Pierre Desproges
Eliminer un profil

The World is flat, The Tipping Point, Steve Jobs : la biographie
Mémoires de Saint Simon, Le Prince de Machiavel, L'art de la guerre de Sun Tzu, John Le Carré
Les Mandarins (S. de Beauvoir), Les petits chevaux de Tarquinia (M. Duras), Aux bonheurs des ogres (et autres Malausène) (Pennac), Fred Vargas
Amélie Nothomb, Anna Gavalda, Pierre Desproges
Eliminer un profil


The World is flat, The Tipping Point, Steve Jobs : la biographie
Assistant marketing Computer System Partner
Les Mandarins (S. de Beauvoir), Les petits chevaux de Tarquinia (M. Duras), Aux bonheurs des ogres (et autres Malausène) (Pennac), Fred
Amélie Nothomb, Anna Gavalda, Pierre Desproges
Chirurgienne vasculaire au CHU de Rennes
Professeure des écoles
Ecolo
de gauche
Eliminer un profil

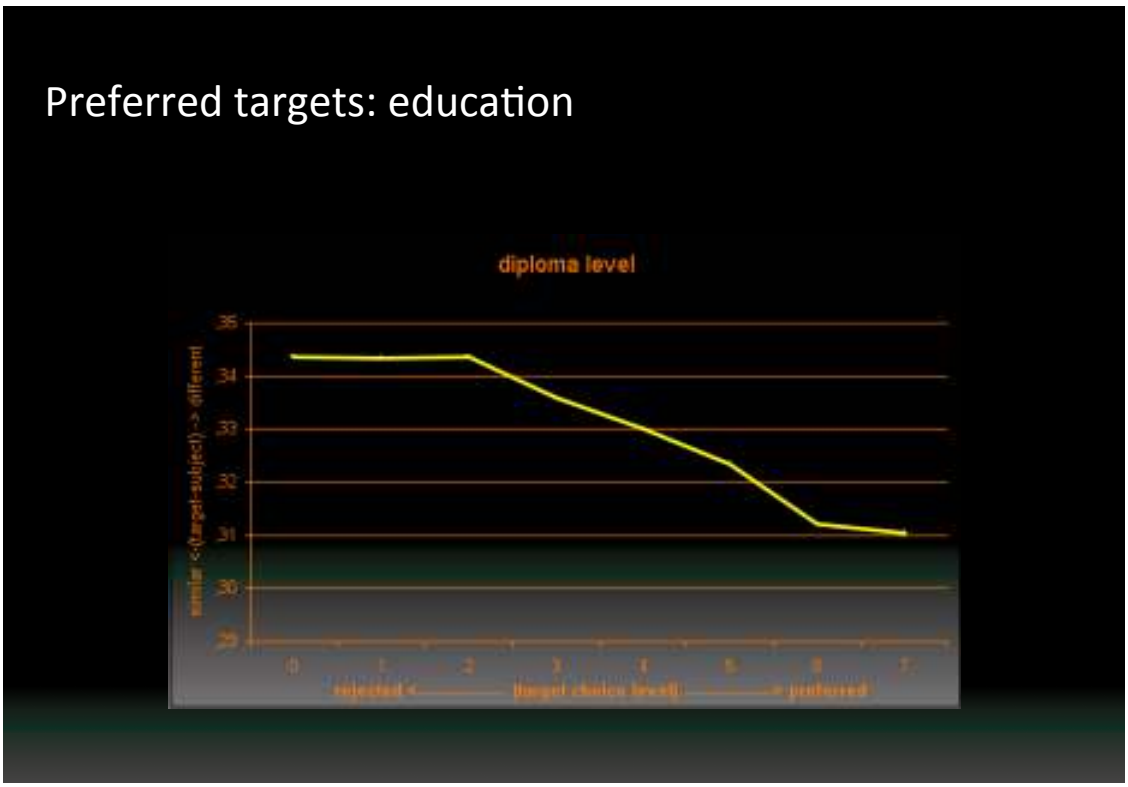


Preferred targets: homophile choices

- *Be my friend*: subject's choice is largely guided by sociological homophily, a tendency to bond with similar others

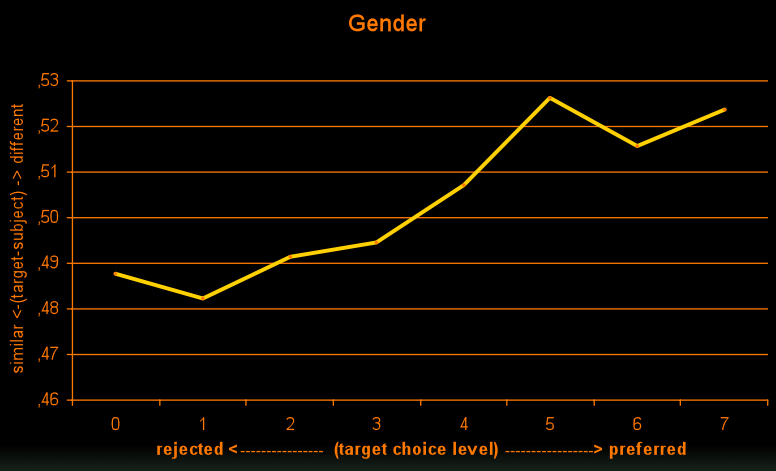
- Variables:
  - preference level (number of positive choices: 0 to 7)
  - standardized subject-target distance on a characteristic

Preferred targets: age
age
similar <-(target-subject)-> different
,21
,20
,19
,18
,17
,16
,15
0
1
2
3
4
5
6
7
rejected <
target choice level
> preferred


Preferred targets: education
diploma level
similar <-(target-subject)-> different
35
34
33
32
31
30
29
0
1
2
3
4
5
6
7
rejected <
target choice level
> preferred

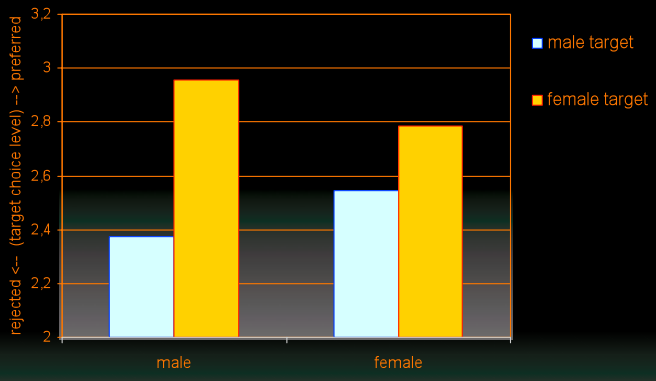
Preferred targets: socio-economic status

Preferred targets: gender



Gender heterophily: a sample artifact

- in fact, all subjects prefer female targets
- although males seem more "flirt oriented"…



Attribute	Provocative	Bodily immodesty	Show-off	Traditional exhibition	Modesty	Total
Photo	9.65	**10.43**	9.66	9.80	8.43	***9.61***
Wall	3.04	2.70	3.16	**3.36**	3.27	***3.12***
About me					**3.18**	***3.07***
Age					2.46	***3.04***
Favorite music					2.50	***2.52***
Favorite movie					**2.70**	***2.46***
Favorite book					**2.94**	***2.17***
Sexual preferences					1.70	***2.13***
Occupation					**2.52**	***2.12***
View by my friends					1.75	***1.75***
Gender					1.36	***1.73***
Personal content					**1.99**	***1.65***
Political views					**1.59**	***1.34***
Favorite TV	1.08	0.99	**1.18**	0.99	1.03	***1.06***
Groups	0.85	0.76	1.06	**1.27**	1.07	***1.02***
Number of friends	0.72	0.85	0.74	0.75	0.85	***0.78***
Diploma	0.61	0.64	0.69	**0.89**	**0.98**	***0.77***

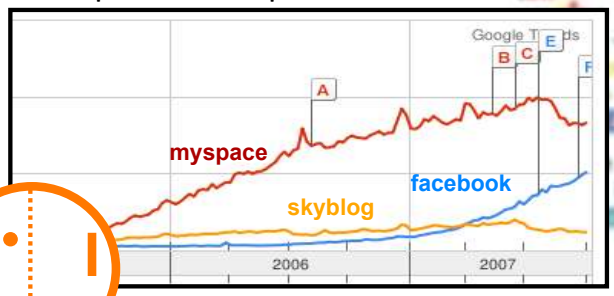
9 - hierarchies from actions

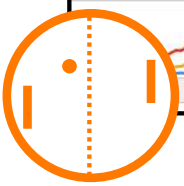
- reputation, capital of SNS
- building notoriety: strategic activity

un mouvement perpétuel ?

10 - endless motion: rise and fall of services

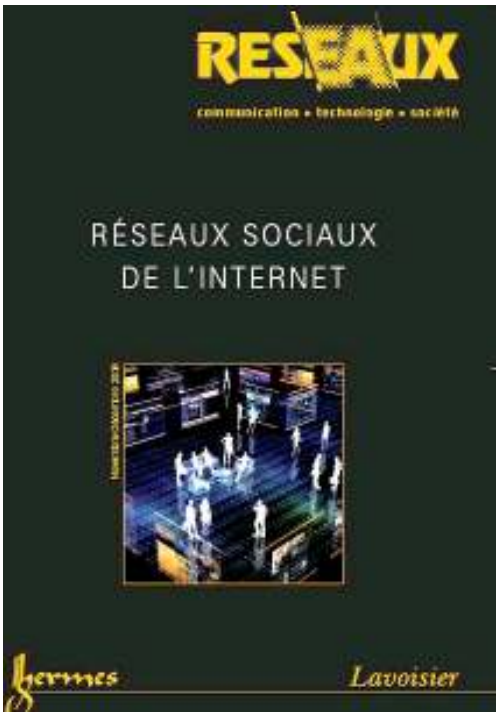
- proliferation of SNS
- 'always beta'
- hype always changes
- from project to project,
  from platform to platform





10 propriétés des coopérations faibles

- l'individualisme expressif
- la visibilité comme opportunité de coopération
- les "amis" ne sont pas forcément des amis
- les communautés viennent après le réseau
- navigation non-hiérarchique
- des engagements hétérogènes
- les grands ont besoin des petits
- la qualité par le nombre
- les hiérarchies produites par le comportement des autres
- un mouvement perpétuel



unrestricted

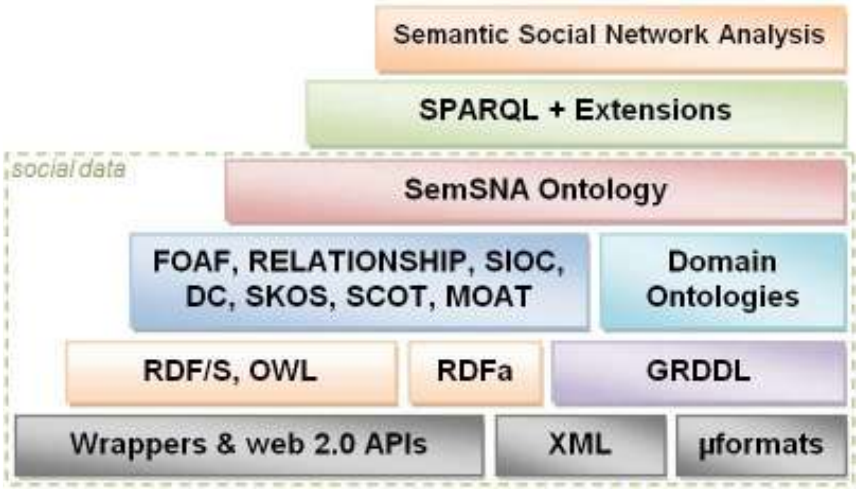
SESSION 9

Doctoriales

Doctoriales, École d'Été Web Intelligence, Projet Web Intelligence, Cluster ISLE

6 au 10 juillet 2009

Analyse des réseaux sociaux et web sémantique - *G. Ereteo*

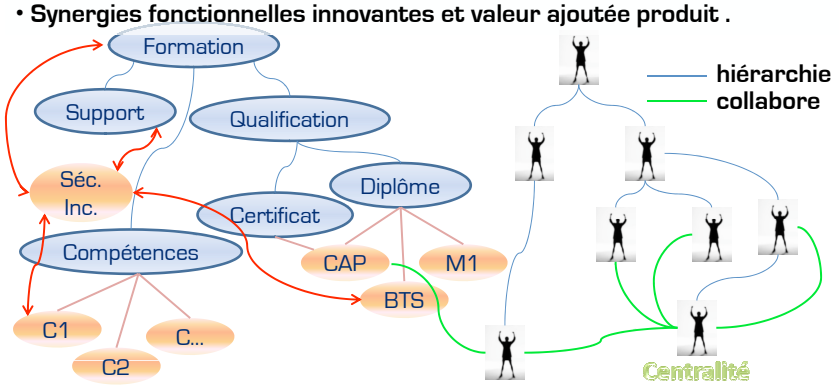


Guillaume ERETEO
Fabien Gandon, Michel Buffa

Référentiels sémantiques cross-lingues et Analyse de Réseaux sociaux en Gestion de RH. - *C. Thovex*

ONTOLOGIES ET RESEAUX SOCIAUX ORIENTES IF-GPEC

- Référentiels métiers multilingues,
- Multi-ontologies (Ingénierie de la Formation, GPEC, GRH),
- Analyse de Réseaux Sociaux,
- Synergies fonctionnelles innovantes et valeur ajoutée produit .



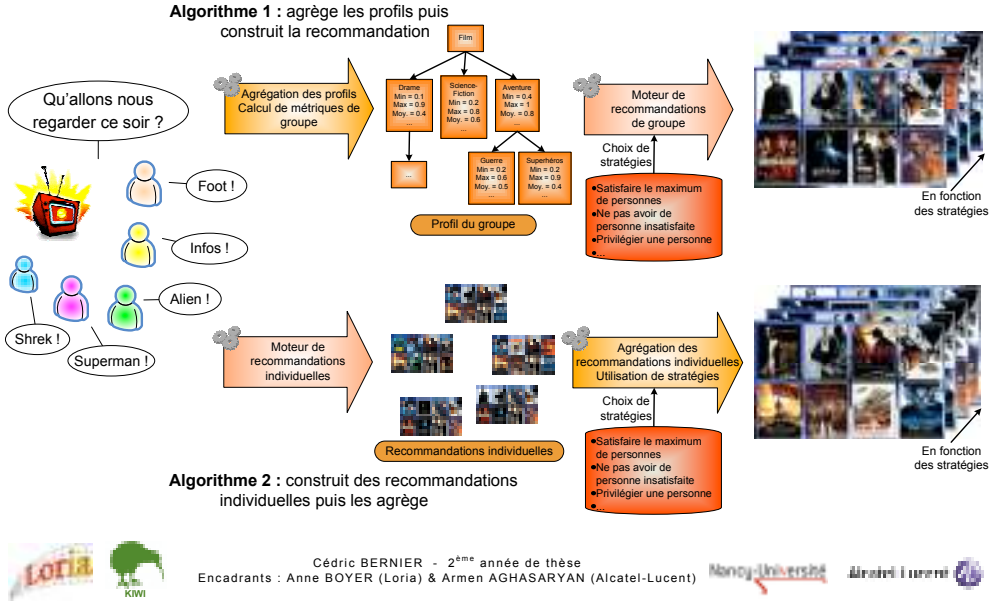
INTERMÉDIAIRE

ÉDITEUR (CONFIDENTIEL)

LABORATOIRE d'INFORMATIQUE de NANTES ATLANTIQUE UMR 6241

Recommandations de contenus pour un groupe de personnes - *C. Bernier*

Recommandations de contenus à un groupe de personnes



Cédric BERNIER - 2ème année de thèse
Encadrants : Anne BOYER (Loria) & Armen AGHASARYAN (Alcatel-Lucent)

Rapprocher ontologies et folksonomies pour le partage des connaissances - *F. Limpens*

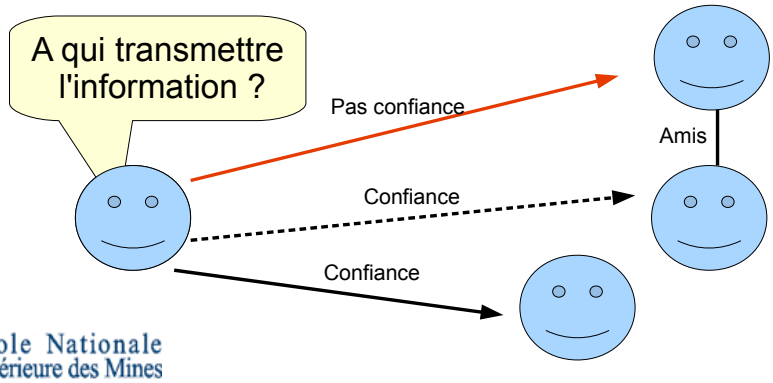
→ modèle **hybride** entre
ontologies & folksonomies

→ coupler ***social tagging*** et
co-construction d'ontologies

→ **Objectifs**:
»» folksonomies **sémantiquement enrichies**
»» émergences de **communautés d'intérêts**
»» **assistance** aux tâches de "**veille**"

Raisonnement social multidimensionnel pour la protection des informations sensibles dans les SMA - *Y. Krupa*

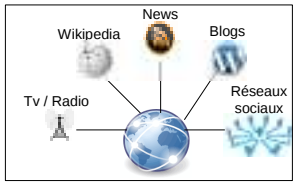
- **Contexte :** Systèmes Multi-Agents, Ouverts, Distribués.
- **Problématique :** Circulation des informations en respectant la privacy de chacun des agents.
- **Pistes :** Relations sociales pour gérer la privacy : liens d'accointances, confiance, réputation, ...
- **Applications :** Réseaux Sociaux, Transports Intelligents, ...

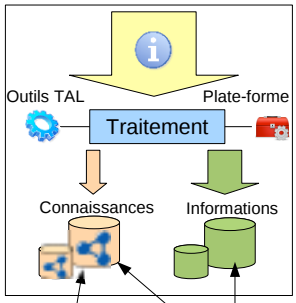


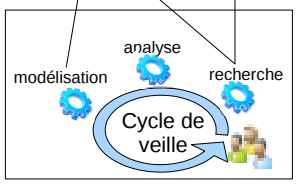
Ecole Nationale Supérieure des Mines SAINT-ETIENNE **L. Vercouter, O. Boissier**

Détection précoce de menaces pour le Renseignement d'Origine Source Ouverte - *E. Bondu*

- Contexte
  - Veille stratégique & ROSO :
    - Objectif : détection **précoce** de **menaces**
    - Sources : blog, site, flux, radio / tv, **réseaux sociaux**...
- Problématiques
  - Sélection de l'information
    - Masse d'information importante
      - L'information pertinente est **cachée**, **noyée**, **bruitée**
    - Diversité des sources et des formats de données
      - Les **méthodes**, **outils** et **objectifs** de recherche sont variés
  - Gestion de la connaissance :
    - Identifier les connaissances qui permettent d'améliorer la **collecte**, la **recherche** et l'**interprétation** des informations
    - **Modéliser**, **capitaliser** et **exploiter** ces connaissances
- Solutions
  - Utiliser un **processus de veille** stratégique
  - Capitaliser des connaissances en construisant des **scénarios**
  - Utiliser les outils du **Web sémantique** pour modéliser et exploiter les connaissances
  - Utiliser une plate-forme pour **normaliser** / **annoter** sémantiquement des documents multimédia et mettre en place des **chaînes de traitement**



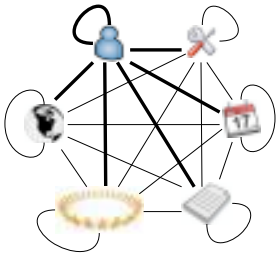


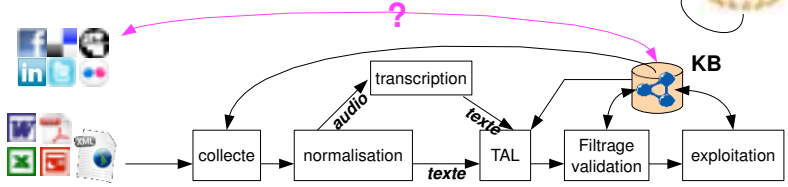


Exploitation de réseaux sociaux - Application au renseignement sur la personne - *N. Martin*

Contexte de la recherche

- Sources ouvertes
- Renseignement sur la personne
- Capacité à collecter, catégoriser l'information qui gravite autour de la personne et à capitaliser dans une base de connaissance
- Capacité à identifier les relations existantes entre ces entités
- Exploiter la base de connaissances pour valider l'information





Problématiques

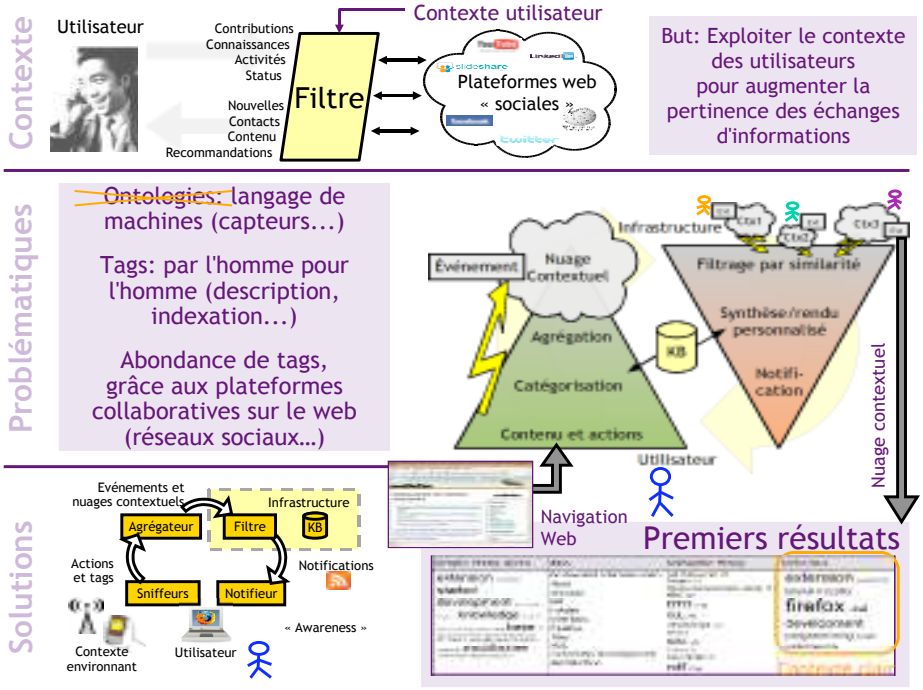
1. Quel type d'information peut-on espérer collecter sur les réseaux ?
2. Quels est la méthodologie et les outils à mettre en place pour traiter cette information ?
3. Comment exploiter les masses de données multi-modales et multi-relationnelles en KB ?
4. En quoi l'étape 3. peut nous permettre d'améliorer le processus de veille ?

Solutions

1. Catégoriser les réseaux, identifier leurs limites fonctionnelles.
2. Proposer et évaluer des chaînes de traitement.
3. i. Prétraitements sur la KB avec les raisonnements logiques du web sémantique.
   ii. Fouille de données sur les entités et relations de la KB : Link Mining.
4. Affiner la connaissance en KB, faire émerger de la nouvelle connaissance via la fouille de données.

Doctoriales, École d'Été Web Intelligence, Projet Web Intelligence, Cluster ISLE
8 / 19 - Orateur Suivant : A. Joly

Filtrage contextuel des interactions médiées - *A. Joly*



Système multi-agents pour la composition de service Web - *A. Abrougui*

École d'été Web intelligence 2009 – Réseaux sociaux

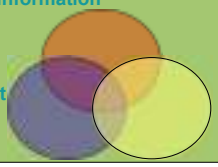
Systèmes multi-agents récursifs pour la composition de services Web

Anissa Abrougui, Pr. Michel Occello, Dr. Marc-Philippe Huget

Contexte

Technologie d'information
Internet
Service Web

Environnement
Dynamique



Besoins d'organisation
Intégration
Interopérabilité
Souplesse

Émergence
Mise en œuvre

SOA

Défit

Composition de services Web

Problématique

Services Web
Passifs
Hétérogènes
Distribués
Absence de connaissances

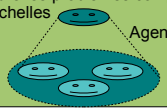
Le Web est dynamique
Apparition de nouveaux services
Disparition des services existants
Modification des services
Changement dynamique de besoins

Composition de services Web dynamique et adaptative
Besoins des utilisateurs
Aléas des services Web

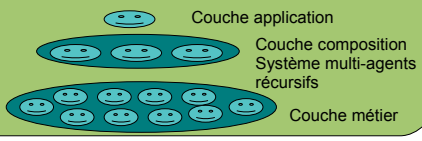
Problème complexe

Solution proposée

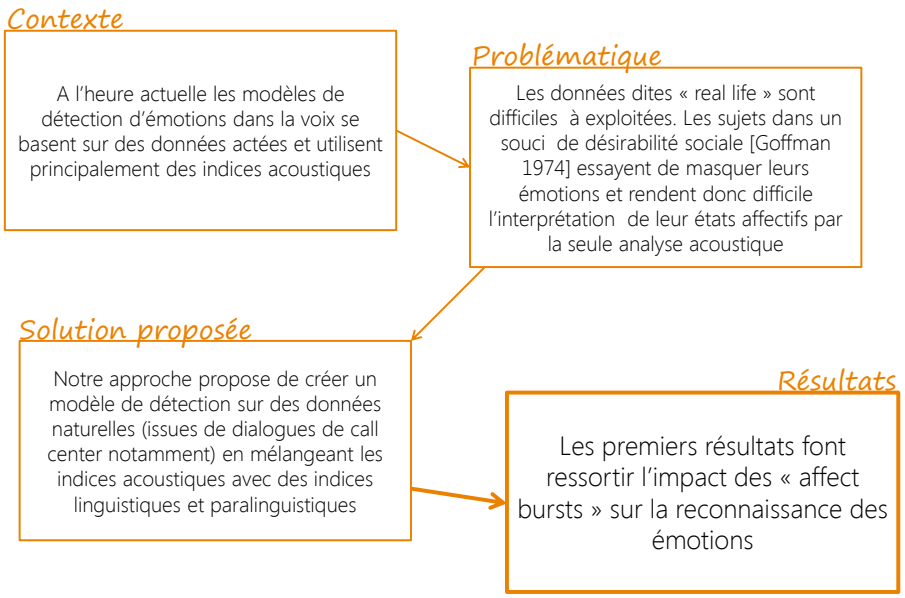
Système multi-agents récursifs
Récursivité
SMA intégrant les principes de la récursivité
Résoudre les problèmes complexes ayant des exigences multi-échelles



Un modèle de composition basé sur un systèmes multi-agents récursifs



Les marqueurs affectifs dans les intéractions orales naturelles - *C. Vaudable*



Vers un système de personnalisation automatique exploitant le Filtrage Collaboratif Navigationnel - *I. Esslimani*

Contexte de la Recherche

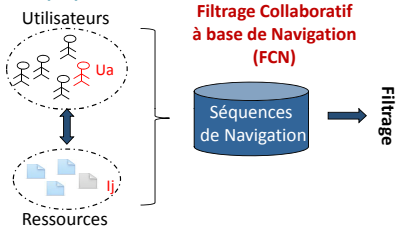
Évolution accrue du web et prolifération des ressources : Nécessité de personnalisation automatique des informations adaptées aux profils des utilisateurs

➢ Développement des systèmes de recommandation et du Filtrage Collaboratif

Problématiques

❖ Manque de données (votes explicites) requises pour le calcul des recommandations
❖ Modélisation des utilisateurs et construction automatique des profils
❖ Précision et qualité des recommandations fournies aux utilisateurs

Modèle proposé



Résultats d'expérimentation

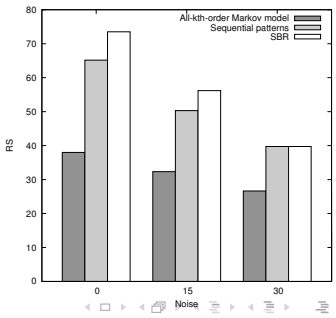
❖ DataSet : Corpus d'usage réel extrait de l'Intranet du Crédit Agricole S.A
❖ Critère d'évaluation : précision des recommandations (MAE/HMAE)
❖ Résultats : meilleure précision des prédictions calculées avec FCN par rapport aux modèles standards de FC

Doctoriales, École d'Été Web Intelligence, Projet Web Intelligence, Cluster ISLE
12 / 19 - Orateur Suivant : G. Bonnin

Vers Une Grammaire Statistique des Usages - *G. Bonnin*

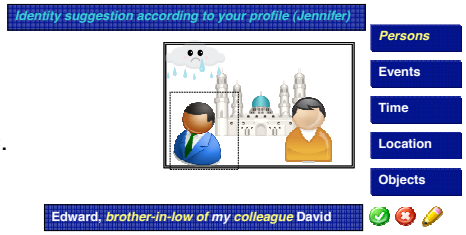
- Prise en compte de la notion d'ordre dans la recommandation d'information
- Similarités navigation web et langage naturel → exploitation des modèles statistiques de langage pour prendre en compte la notion d'ordre dans les recommandations
- Solution : version adaptée d'un modèle de n-grammes avec skipping (Skipping Based Recommender ou SBR)

Modèle	Taille	Temps
SBR	4.1 Mo	3m20s
AKO	8.3 Mo	17m02s
SP	289.6 Mo	10m50s



Annotation et recherche multimédia sémantiquement riche et adaptative d'une manière personnalisée - *S. Lajmi*

- ***Contexte de la recherche***
  - Le Web socio-sémantique.
  - Tagging des médias sociaux : pages web, photos, vidéos, etc.



- ***Problématique***
  - Adapter les tags et fournir des recommandations d'une manière personnalisée en se basant sur le contexte.
  - Optimiser l'accès à des sources de connaissances externes volumineuses.

- ***Solutions et résultats obtenus***
  - Exploitation des connaissances externes notamment les réseaux sociaux pour assister le tagging et la recherche.
  - Traçage de l'utilisateur dans ses activités de tagging/recherche/visualisation en capturant le contexte.
  - Proposition du modèle SeMAT (Semantic Model for Self-adaptative Tag) pour enrichir et adapter les tags en leur ajoutant des dimensions sémantiques.
  - Proposition d'une heuristique de recherche dans des sources de connaissances.

Technique de matching pour la réutilisation efficace des services Web : découverte et qualité - *U. Chukmol*

- Contexte de la recherche: service Web, SOC
- Problématique: découverte, réutilisation et qualité des services Web
- Proposition de solution et résultat:
  - Plateforme pour la collection, la découverte, la réutilisation et le suivi d'évolution des services Web
  - Annotation collaborative sur des services Web pour renforcer/faciliter leurs découvertes et leurs réutilisations
  - Propriétés sur la réutilisabilité des services Web

Security Patterns, their applicability, retreival and composition - *P. El Khoury*

Paul El Khoury,
Research Associate – SAP Research Security & Trust - SAP Labs France
PhD Student – LIRIS CNRS UMR 5205, University Claude Bernard Lyon I

Security Patterns,

their applicability, retreival and composition

- Why? Similarities with predecessors.
- Challenges to overcome.
- Functionalities vs New Format
- Validation of the PhD Study

Métrologie des réseaux de type ”petits mondes” - *E. Navarro*

Qui suis-je ? M2 à Toulouse (soutenance en septembre)
Thèse à partir de septembre
(encadré par Bruno Gaume & Henri Prade)

Que fais-je ? **Les grands graphes de terrain**

sparse + clusters + small-world + power-law

[Watts and Strogatz, 1998]
[Barabási and Albert, 1999]

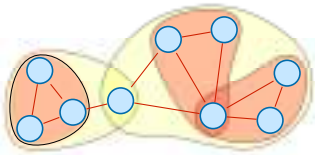
Réseaux sociaux, réseaux de protéines, etc...

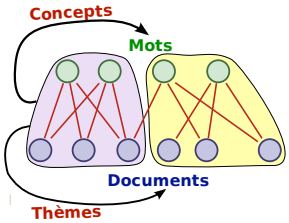
En particulier : graphes lexicaux et graphes de documents

Détection de communautés

Master 2 : Travail sur les graphes bipartis

Thèse : Hiérarchie + Chevauchement





Comment ? **Marches aléatoires** [Gaume, 2004], [Pons and Latapy, 2005]

Extraction automatique de réseaux sociaux à partir de textes et détection d'opinion - *M. Forestier*

Contexte de la recherche

Internet 2.0 a permis à ses utilisateurs d'écrire leurs opinions sous différentes formes. Que ce soit dans des forums où les individus peuvent interagir entre-eux sur des thèmes, dans des blogs etc, cette nouvelle masse de données représente une source d'information importante mais impossible à traiter à la main. Qui parle de quoi ? Qui répond à qui ? Quelles sont les personnes d'accord entre-elles ?

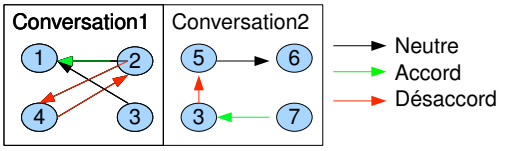
Problématiques

- Comment extraire le réseau social à partir du texte des messages ?
- Comment définir de quoi parlent les individus ?
- Comment définir l'opinion des individus ?

Solutions

- Les noeuds sont les acteurs.
- Les arêtes représente la relation "*a répondu à* " par une similarité de contenu.
- Le réseau social va être clusterisé en conversations.

Résultats attendus



Les résultats sont en cours de programmation...

Remerciements

- École d'été Web Intelligence
- Projet Web Intelligence du cluster ISLE
- Région Rhône Alpes

Index des auteurs

Abrougui, A., 399
Amblard, F., 267

Benchettara, N., 399
Bernier, C., 399
Bondu, E., 399
Bonnin, G., 399

Cardon, D., 355
Chukmol, U., 399

ElKhoury, P., 399
Ereteo, G., 399
Esslimani, I., 399

Forestier, M., 399
Frachisse, D., 189

Gandon, F., 227
Garnier, C., 43

Hacid, H., 69

Jaillon, P., 1
Joly, A., 21, 399

Kanawati, R., 149
Krupa, Y., 399

Lajmi, S., 399
Limpens, F., 399

Maret, P., 21
Martin, N., 399

Navarro, E., 399

Rouveirol, C., 149

Stan, J., 69
Sueur, B., 221

Thollard, F., 1
Thovex, C., 399

Vaudable, C., 399

Zaiane, O. R., 121