Concepts et modèles des Webs communautaires

Talel Abdessalem Pierre Senellart



21 octobre 2008, BDA 2008

http://www.infres.telecom-paristech.fr/~dbweb/

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Confiance

Conclusion

Concepts et définitions

Web 2.0 :

- « Désigne ce qui est perçu comme un renouveau du World Wide Web. Concerne aussi bien les technologies employées que les usages du »
- « Web 2.0 is a living term describing changing trends in the use of World Wide Web technology and web design that aims to enhance creativity, information sharing, collaboration and functionality of the web. »
- « Web 2.0 concepts have led to the development and evolution of web-based communities and hosted services, such as socialnetworking sites, video sharing sites, wikis, blogs, and folksonomies. »

Social Web :

- « The Social Web is currently used to describe how people socialize or interact with each other throughout the World Wide Web. »
- « The Social Web may also be used to refer to the description of web 2.0 technologies that are focused on social interaction and community before anything else. »

- · Blog ou blogue:
 - « un site web constitué par la réunion de billets agglomérés au fil du temps, et souvent, classés par ordre antéchronologique »
- Wiki :
 - « un système de gestion de contenu de site web rendant ses pages web librement modifiables par tous les visiteurs y étant autorisés »
- Bookmarking (Social Bookmarking):

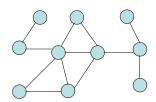
Bookmark

- → marque-page ou signet, repère par lequel on enregistre l'adresse d'une page web.
- « Social bookmarking is a method for Internet users to store, organize, search, and manage bookmarks of web pages on the Internet with the help of metadata. »

- Tag (méta-données) :
 - « a non-hierarchical keyword or term assigned to a piece of information (such as an internet bookmark, digital image, or computer file) »
 - « marqueur sémantique ou lexical utilisé sur les sites dits de réseaux sociaux Web 2.0 »
- Folksonomie (Folksonomy : fusion de Folk+Taxonomy) :
 - « Folksonomy (also known as collaborative tagging, social classification, social indexing, and social tagging) is the practice and method of collaboratively creating and managing tags to annotate and categorize content »
 - « désignant un système de classification collaborative décentralisée spontanée »

Réseau social :

« 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, kinship, dislike, conflict or trade. The resulting graph-based structures are often very complex. »



Individu/organisation

Contexte et objet de l'étude

Contexte :

- Évolution du web vers le web 2.0 et le web social
- Rôle moteur des communautés d'utilisateurs
 - · Production, indexation, partage, réseaux sociaux

Objet :

- Analyser cette évolution et les besoins qu'elle engendre
 - Analyse des outils et des modèles des web communautaires
 - Réseaux sociaux et modèles de graphes
 - Exploration des réseaux
 - Ranking, réputation et confiance

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

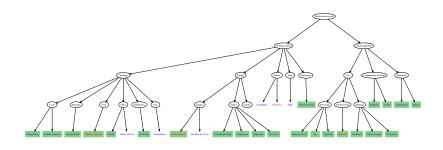
Confiance

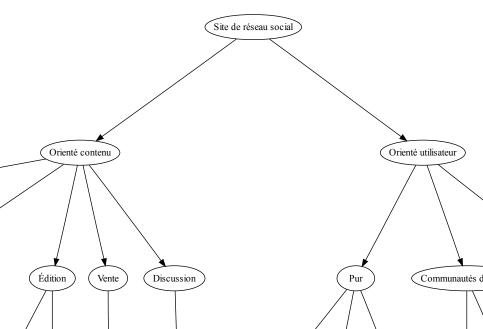
Conclusion

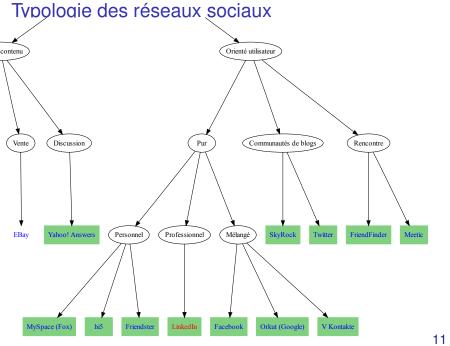
Réseaux sociaux les plus populaires

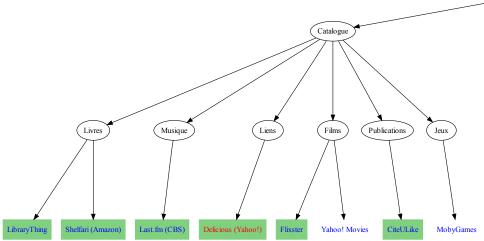
Sites de réseaux sociaux les plus populaires dans le monde et en France (rang des sites Web avec le plus de trafic, selon Alexa)

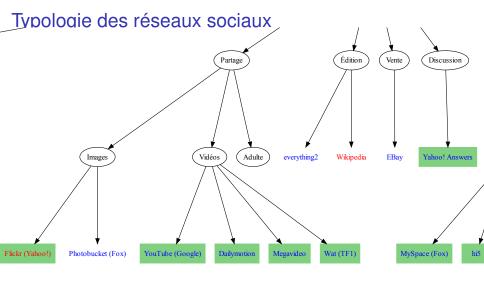
SkyRock YouTube MySpace Facebook Dailymotion EBay Wikipedia Meetic ImageShack hi5 Megavideo Adult Friendfinder Wat.tv Flickr	Monde 51 3 17 5 61 18 8 565 47 15 133 55 1568 33	France 3 4 7 8 11 12 13 27 53 59 80 82 88 94	
Orkut V Kontakte Friendster	19 28 39	>100 >100 >100	

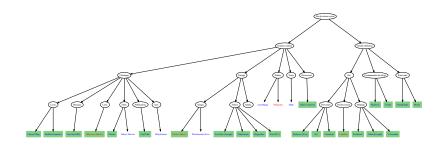








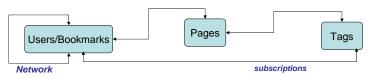




- Outil de « Social bookmarking »
- Destiné à la sauvegarde et au partage (publication) de ses favoris (bookmarks).
- Permet d'annoter et de classer ses favoris en y associant des motsclés (tags).
- Pas de partage d'informations personnelles comme MySpace ou Facebook

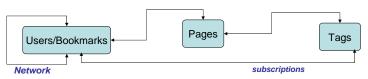
http://delicious.com http://del.icio.us

Principales entités gérées :



- Favoris (pages web)
 - adresses, titre et note associée.
- Mots-clés (données par l'utilisateur)
 - nom du tag et sa description
- Utilisateurs
 - Login, e-mail, nom et page personnelle

Confidentialité, réseaux et souscriptions :

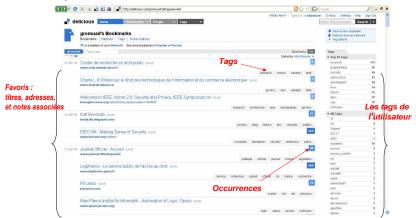


- Favoris publiques : visibles à tous (lien, annotations et tags)
- Favoris privés : visibles à leur propriétaire
- Réseau : liste d'utilisateurs
 - Donne un accès direct à la partie publique de leurs bookmarks
 - Peut être publique ou privé
- Souscription à un tag :
 - Donne un accès direct aux favoris publiques auxquelles le tag est associé

Recherche d'information :

- Par mots-clés
- · Parmi:
 - Mes favoris
 - Mon réseaux
 - Favoris publiques
- Explore:
 - les tags,
 - les titres de pages
 - et les notes

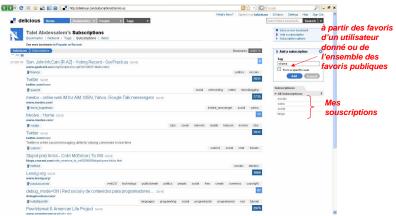
Exemple de favoris :



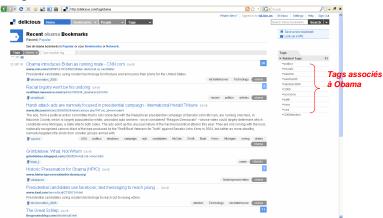
Exemple de recherche : BDA 2008



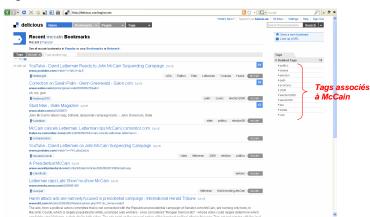
Souscriptions à un tag :



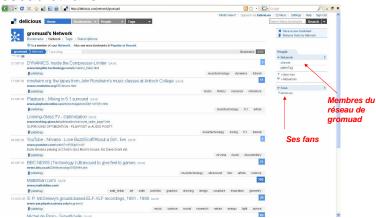
Tag associés:



Tag associés:

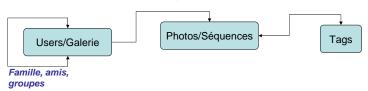


Réseau et fans :



- Gestion d'albums photos
- Permet de stocker des images, de leur associer des mots-clés et de les partager (rendre visible) avec ses contacts ou avec tous.
- Des fonctions sociales permettent aux utilisateurs de se découvrir, se retrouver, d'échanger, de se regrouper en communautés etc.
- http://flickr.com

Principales entités gérées :



- Photos
 - Titre, description, tags, dates (données EXIF, Exchangeable image file format).
- Mots-clés
 - nom du tag et sa description
- Utilisateurs
 - pseudo, e-mail, nom, page personnelle, sexe, ville pays, centres d'intérêt, Aéroport le plus proche, etc.

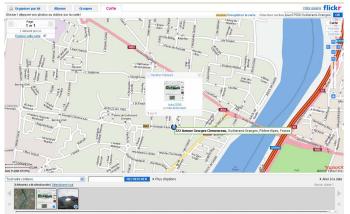
Confidentialité, réseaux et partage :

- Les contacts sont organisés en trois cercles:
 - famille, amis et public
 - laisser-passer : permettent de partager des photos avec des personnes non inscrites sur Flickr
- Groupes :
 - Objet : la mise en commun des photos
 - Groupes privés :
 - Accessibles sur invitation et masqués lors de la recherche de groupes
 - Public, sur invitation:
 - · Visible au public et contrôle sur les membres
 - Public, ouverts à tous :
 - · Accès public
 - Les discussions peuvent être masquées aux non membres.

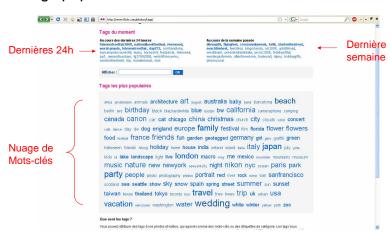
Recherche d'information :

- Photos
 - Description des photos et/ou tags
- Groupes
 - Noms et descriptions ou discussions
- Personnes
 - Nom et e-mail ou profil

GéoTagging : les photos peuvent être situées sur une carte Le droit de voir la localisation d'une photo sur la carte est géré par le propriétaire



Tags populaires:



Interestingness: les photos intéressantes.

Critères affichés : origine du cliché, commentaires, date des commentaires, qui l'a ajouté à ses favoris, tags, etc.

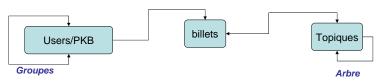


Webograph

- Un outil de « Social bookmarking »
- Sauvegarde décentralisée et partage (publication) de favoris et/ou de billets (notes personnelles).
- Permet de classer les billets selon une folksonomie propre à chaque utilisateur.
- http://webograph.enst.fr/

Webograph

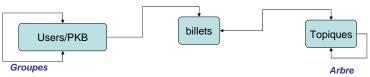
Principales entités gérées :



- Billets
 - Texte, titre du billet et URL éventuelle d'une page associée.
- Arbre de topiques (foksonomie)
 - Nom de thème et lien vers le thème parent
- Utilisateurs
 - Login, e-mail, nom et adresse de PKB

Webograph

Confidentialité, réseaux et souscriptions :



- Droit de lecture : associé à un billet ou à un topique
- Visibilité des billets et topiques : choix de la partie de l'arbre des topiques que l'on veut montrer et à qui on veut la montrer
- · Le droit de lecture est attribué aux groupes
- Groupes: listes d'utilisateurs
- Modèle Iceberg :
 - L'arbre des topiques est organisé du plus général au plus précis
 - Droit de lecture sur un billet/topique → droit de lecture sur les topiques parents
- · Liens entre arbres de topiques :
 - Permet de lier à son propre arbre de topiques tout ou partie (sous arbre) d'un autre arbre de topiques (provenant d'une autre PKB)

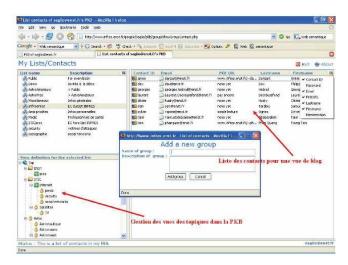
- Arbre de topiques :
 - billets indexés selon un arbre de topiques
 - Interface analogue à thunderbird folder → topique, hiérarchie de folders → hiérarchie de topique, courrier → billet



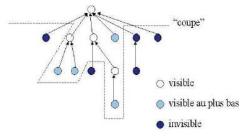






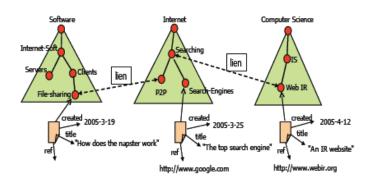


- Modèle Iceberg (modèle de confidentialité) :
 - Une vue est une coupe dans l'arbre des topiques
 - Union et intersection de coupes est une coupe
 - Associer une vue à un groupe de contacts
 - Un contact peut appartenir à plusieurs groupes



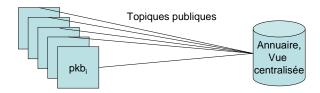
Source: Danzart et al, 2006

Liens entre PkB (arbres de topiques) :



Source: Danzart et al, 2006

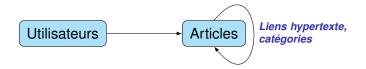
- Connexion sur invitation
- Vue centralisée des paries publiques
 - PKB
 - Topiques
 - Géré comme une pkb particulière
 - Mis à jour en fonction de l'évolution de la confidentialité



Wikipedia

- Encyclopédie collaborative
- Plus un réseau social d'articles qu'un réseau social d'utilisateurs (pas de relation d'amitié)
- Mais réseaux sociaux implicites : utilisateurs collaborant aux mêmes articles, etc.
- Historique complet disponible (permet d'étudier la dynamique du réseau)
- http://www.wikipedia.org

Wikipedia - Entités



Utilisateurs Login ou adresse IP, page de présentation Article Titre, texte, médias, catégories, liens

Aucune gestion de confidentialité (tout est public)

Wikipedia – Recherche d'information

- Recherche en texte intégral dans les articles
- Accès à la liste des modifications d'un article
- Accès à la liste des contributions d'un utilisateur
- ▶ Graphe de Wikipedia ~ graphe du Web en miniature (mais beaucoup plus propre!)

Wikipedia – Articles



The Free Encyclopedia

Wikipedia is sustained by people like you. Please donate today.

article discussion edit this page history

Log in / create account

Social network service

From Wikipedia, the free encyclopedia

Not to be confused with social network.

A **social network service** focuses on building online communities of people who share interests and activities, or who are interested in exploring the interests and activities of others. Most social network services are web based and provide a variety of ways for users to interact, such as e-mail and instant messaging services.

Social networking has created powerful new ways to communicate and share information. Social networking websites are being used regularly by millions of people, and it now seems that social networking will be an enduring part of everyday life. The main types of social networking services are those which contain directories of some categories (such as former classmates), means to connect with friends (usually with self-description pages), and recommender systems linked to trust. Popular methods now combine many of these, with MySpace and Facebook being the most widely used in North America; [1] Bebo, [2] MySpace, Facebook and Hi5 in parts of Europe; [3] Orkut and Hi5 in South America and Central America; [4] and Friendster, Orkut, and Cyworld in Asia and the Pacific Islands.

There have been some attempts to standardize these services to avoid the need to duplicate entries of friends and interests (see the FOAF standard and the Open Source Initiative), but this has led to some concerns about privacy.

Contents [hide]

- 1 History of social networking services
- 2 Research on the social impact of social networking software
- 3 Business applications

navigation

- Main page
- ContentsFeatured
- content
- Current events
- Random article

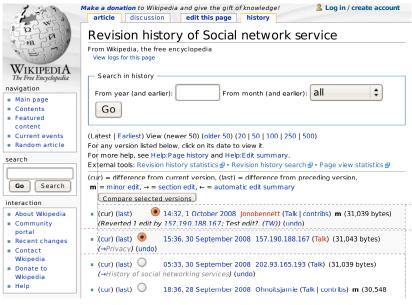
search



interaction

- About WkipediaCommunity portal
- Recent changes
- Contact Wkipedia
- Donate to Wkipedia
- Help

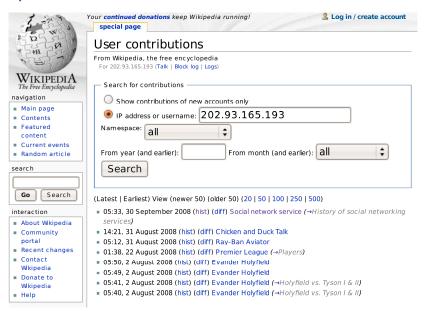
Wikipedia – Historique



Wikipedia – Utilisateur



Wikipedia - Contributions



LinkedIn

- Réseau social résolument orienté vers les professionnels
- Donc pas d'amis, mais des connexions
- But :
 - Étendre son réseau
 - Pouvoir contacter des gens par l'intermédiaire de connaissances
 - Augmenter sa visibilité dans le milieu professionnel
- Utilisé par certains employeurs pour recruter
- http://www.linkedin.com/

LinkedIn – Entités



Groupes Nom, Description, Messages, Membres Institution Nom, Membres passés et présents Utilisateur CV, connexions

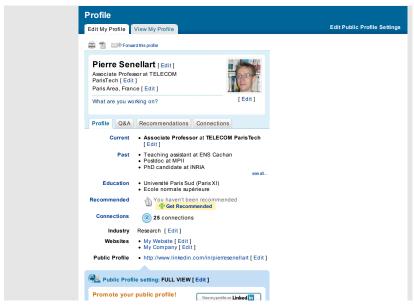
LinkedIn – Confidentialité et droits d'accès

- Inscription obligatoire
- ▶ Le profil complet d'un utilisateur n'est accessible qu'à ses connexions
- ➤ On peut savoir si on est à distance ≤ 3 d'un individu, et le chemin correspondant dans le graphe (partiellement anonymisé!)
- On peut choisir ou non de divulguer la liste de ses connections à ses connections (et à elle seulement!)
- Pour ajouter quelqu'un comme connexion, il faut connaître son email, avoir été dans la même institution, ou demander une introduction à une connexion commune
- Contacter directement un inconnu est possible, mais payant

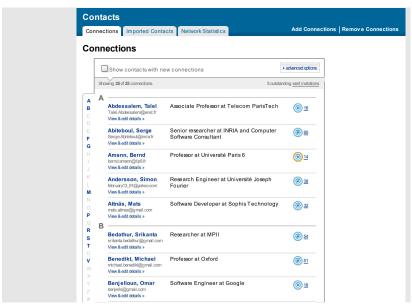
LinkedIn – Recherche d'information

- ► Recherche d'utilisateurs par mots-clés, nom, emplacement, etc.
- Affichage de la liste des utilisateurs de LinkedIn qui sont dans la même institution
- Recherches avancées possibles

LinkedIn – Profil



LinkedIn – Connexions



LinkedIn – Réseau



Here you see statistics about your network including how many users you can reach through your connections. Your network grows every time you add a connection — invite connections now.

Your Network of Trusted Professionals

You are at the center of your network. Your connections can introduce you to 82,100+ professionals—here's how your network breaks down:

1 Your Connections
25

Two degrees away
Friends of friends, each connected to one of your connections
700+

3 Rhee degrees away
Three degrees away
Total users you can contact through an introduction
61,300+
1,267 new people in your network since September 26

The LinkedIn Network

The total of all LinkedIn users, who can be contacted directly through InMail.

Total users you can contact directly — try a search now!

26.000.000+

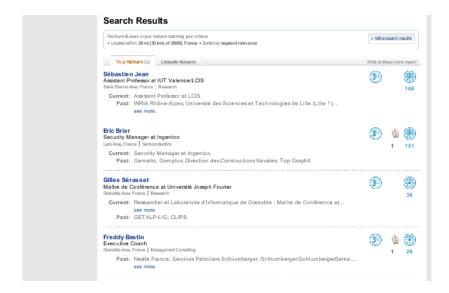
More About Your Network



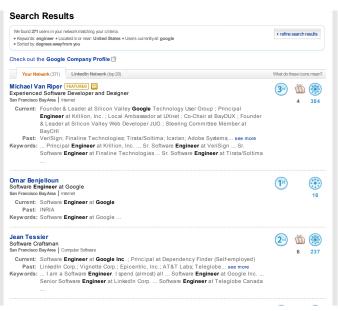


Your region: Paris Area, France: 8,100+ users

LinkedIn – Recherche



LinkedIn – Recherche avancée



Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Modèles de graphes

Recherche dans les réseaux sociaux

Recherche de communautés

Utilisation des liens sociaux pour améliorer la recherche

Confiance

Conclusion

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Modèles de graphes

Recherche dans les réseaux sociaux

Recherche de communautés

Utilisation des liens sociaux pour améliorer la recherche

Confiance

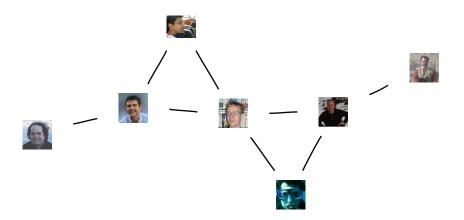
Conclusion

Graphes de réseaux sociaux

- Modélisation naturelle : réseau social = graphe
- Entités = noeuds, Relations = arêtes
- Suivant les cas :
 - graphes orientés ou non
 - bipartite, n-partites
 - arêtes annotées, pondérées

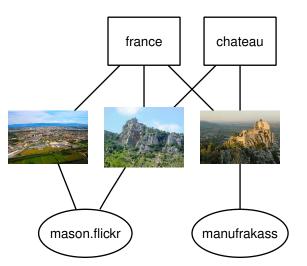
Graphe non orienté

Adapté pour les réseaux sociaux purs avec relations symétriques (p. ex., LinkedIn)



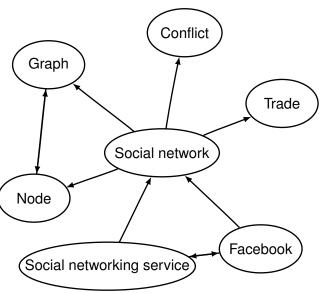
Graphe multipartite

Adapté à la plupart des réseaux sociaux de partage avec annotations, utilisateurs, contenu, etc. (p. ex., Flickr)



Graphe orienté

Rare dans le cas des réseaux sociaux, mais c'est le modèle du Web (p. ex., Wikipedia)



Réseaux implicites et explicites : aspect sociologique

- Deux types de réseaux sociaux d'individus :
 - donnés explicitement par l'application (amis, connections...)
 - dérivables implicitement des profils (mêmes intérêts, édition des mêmes articles...)
- Sociologiquement, définit quatre types de connections entre individus [Smith et al., 2007, Lin, 2001, Houard and Jacquemain, 2006], suivant la forme de capital social considéré:

		Lien implicite	
		Oui	Non
Lien explicite	Oui Non	Bonding réel Bonding potentiel	Bridging réel Bridging potentiel

- Bonding : lien « agglutinant »
- Bridging : lien « reliant »

Six degrés de séparation

- Idée que deux personnes quelconques sur Terre sont séparées par une chaîne de six personnes se connaissant deux à deux
- Mis en évidence par une expérience de Stanley
 Milgram [Travers and Milgram, 1969] (courrier à transmettre de proche en proche)
- Popularisé dans de nombreux médias
- Le chiffre 6 n'est pas à prendre au sérieux! Mais principe validé dans des expériences plus récentes
- Dans d'autres domaines :
 - Nombre d'Erdős pour les publications scientifiques
 - Kevin Bacon pour les films d'Hollywood

Caractéristique commune (de la plupart) des réseaux sociaux !

Caractéristique des graphes de réseaux sociaux

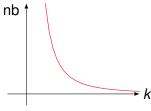
Quatre caractéristiques importantes [Newman et al., 2006] :

Graphes creux : beaucoup moins d'arêtes qu'un graphe complet

Faible distance typique: plus court chemin entre deux nœuds en général petit (logarithmique) par rapport à la taille du graphe

Haute transitivité : si *a* est relié à *b* et *b* à *c*, alors *b* a plus de chances d'être relié à *c*

Degrés suivant une loi en puissance : le nombre de sommets de degré k est de l'ordre de $k^{-\gamma}$ (γ constante)



Pas seulement les réseaux sociaux!

- Mêmes caractéristiques dans :
 - ▶ le World Wide Web
 - ▶ les systèmes nerveux
 - les graphes d'interaction de protéines
 - les graphes de citations
 - etc.
- Contre-exemples: graphes planaires, graphes de transports (plus de régularité, pas forcément de transitivité, plus grande distance typique, etc.).

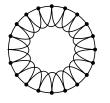
Modèles de graphes de réseaux sociaux?

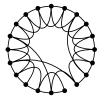
Graphes aléatoires

- Principe:
 - 1. On se fixe un nombre *n* de noeuds et un nombre *p* d'arêtes.
 - 2. On sélectionne aléatoirement (uniformément) p arêtes parmi les $\frac{n(n-1)}{2}$ possibles.
- Creux si p est bien choisi
- Distance typique (à l'intérieure d'une même composante connexe) petite!
- Pas de transitivité
- Pas de loi en puissance des degrés

Graphes petit-monde

- Principe [Watts and Strogatz, 1998, Watts, 1999] :
 - 1. On part d'un treillis régulier (une grille) périodique
 - 2. Avec probabilité *p*, on reroute chacune des arêtes aléatoirement







[Watts and Strogatz, 1998]

- Creux si le treillis initial est bien choisi
- ▶ Distance typique (à l'intérieure d'une même composante connexe) petite (vient des reroutages)
- Transitivité forte (vient du treillis initial)
- Pas de loi en puissance des degrés

Graphes sans échelle : attachement préférentiel

- Principe [Barabási and Albert, 1999, Newman et al., 2006] :
 - 1. On part d'un petit graphe de taille m_0 , on se fixe une constante $m < m_0$
 - On ajoute tour à tour n m₀ nœuds au graphe, en les connectant chacun à m nœuds existants; la probabilité de connexion à un nœud existant est proportionnelle à son degré
- Creux si m et n sont bien choisis
- Distance typique (à l'intérieure d'une même composante connexe) petite
- Transitivité forte
- ▶ Loi en puissance des degrés (en fait, exposant −3, mais des variations du modèle permettent de changer l'exposant)

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Modèles de graphes

Recherche dans les réseaux sociaux

Recherche de communautés

Utilisation des liens sociaux pour améliorer la recherche

Confiance

Conclusion

PageRank [Brin and Page, 1998]

Idée

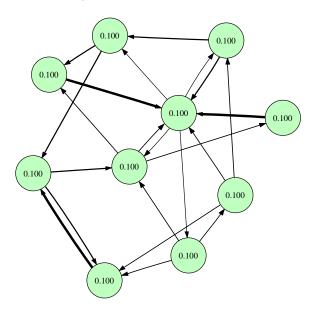
Sur le Web, les pages importantes sont les pages pointées par des pages importantes.

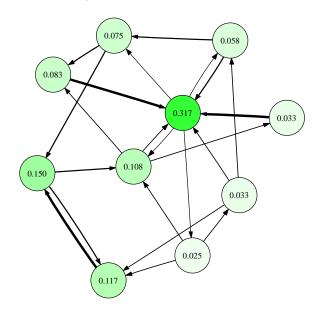
Définition (Tentative)

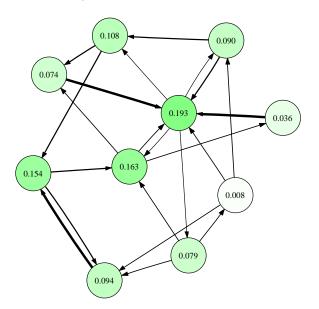
Probabilité qu'un surfeur suivant une marche aléatoire dans le graphe du Web *G* arrive à la page *i* à un point distant dans le futur.

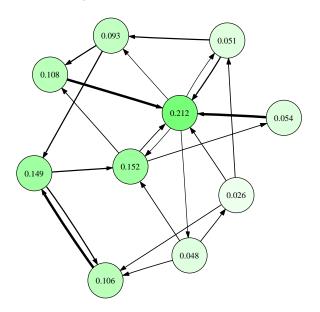
$$pr(i) = \left(\lim_{k \to +\infty} ({}^tG)^k v\right)_i$$

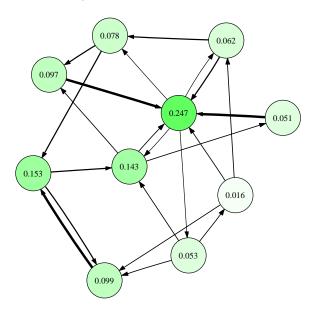
où v est un vecteur colonne initial.

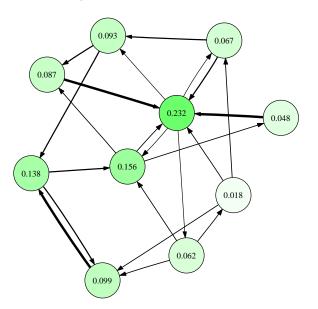


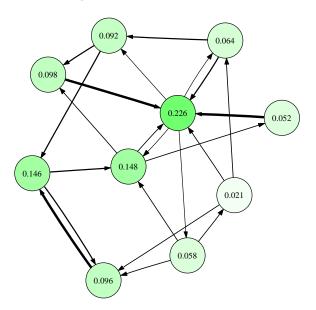


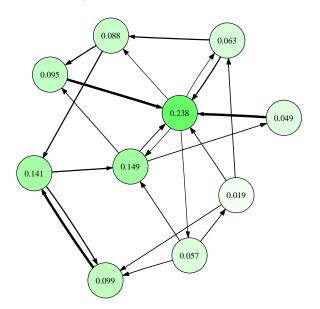


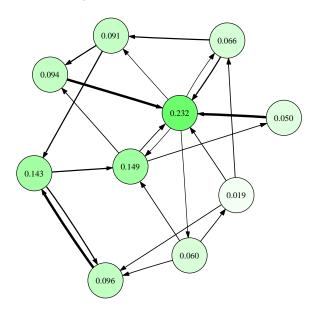


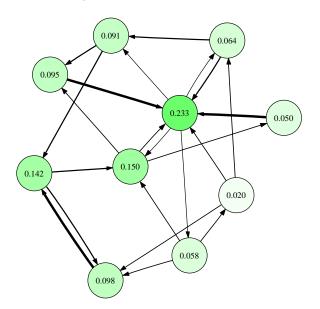


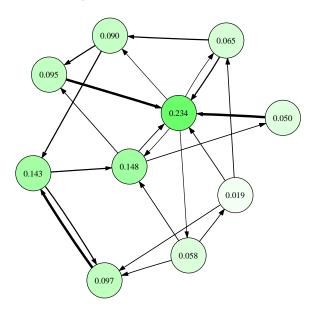


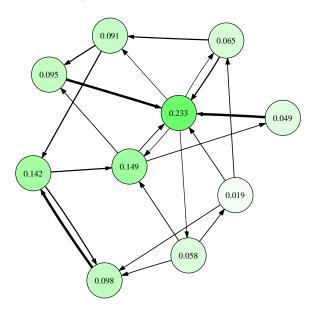


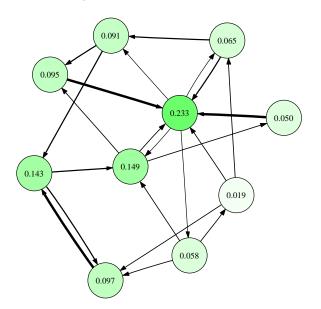


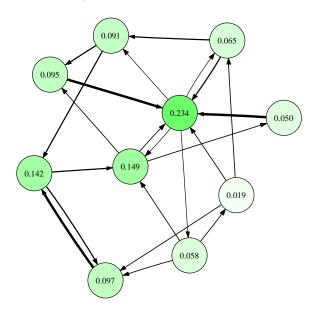












PageRank avec sauts aléatoires

Peut ne pas converger.

Pour réparer ça, le surfeur aléatoire peut, à chaque étape, sauter aléatoirement vers n'importe quelle page du Web avec probabilité d (1 -d: damping factor).

$$pr(i) = \left(\lim_{k \to +\infty} ((1 - d)^t G + dU)^k v\right)_i$$

où U est la matrice dont tous les valeurs sont à $\frac{1}{n}$ (n: nombre de nœuds).

PageRank et réseaux sociaux

- Mesure d'importance globale utilisée pour ordonner les résultats des moteurs de recherche sur le Web
- ► Très utile pour les graphes ressemblant au graphe du Web (p. ex., Wikipedia)
- Notion inappropriée pour les graphes non orientés : mesure d'équilibre de la marche aléatoire = degré du nœud
- Nombreuses variantes adaptées à certains cas particuliers :
 - ► HITS [Kleinberg, 1999] pour distinguer portails et autorités
 - Généralisation de HITS [Blondel et al., 2004] pour comparer deux graphes arbitraires
 - Mesures de Green [Ollivier and Senellart, 2007] pour biaiser le PageRank

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Modèles de graphes

Recherche dans les réseaux sociaux

Recherche de communautés

Utilisation des liens sociaux pour améliorer la recherche

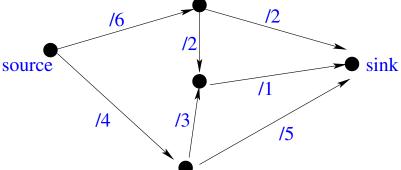
Confiance

Conclusion

Recherche de communautés

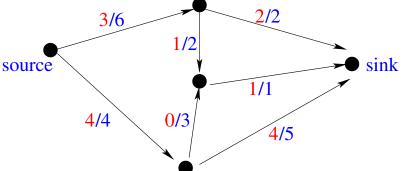
- Problème classique dans les réseaux sociaux : identifier des communautés d'utilisateurs (ou de contenu) en utilisant la structure de graphe
- Deux sous-problèmes :
 - 1. Étant donné un nœud ou ensemble de nœuds initial, trouver la communauté correspondante
 - Etant donné le graphe dans son ensemble, trouver une partition en communautés

Flot maximal / Coupe minimale



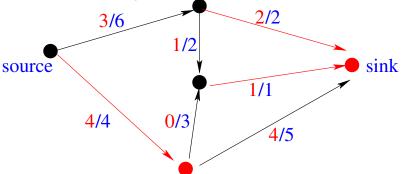
- Utilisation d'un algorithme de calcul de flot maximal [Goldberg and Tarjan, 1988] pour séparer une graine d'utilisateurs du reste du graphe
- ► Complexité : $O(n^2m)$ (n : nœuds, m : arêtes)
- Versions en ligne de ces techniques [Flake et al., 2002, Senellart, 2005] permettant de faire le calcul sans avoir à stocker le graphe localement





- Utilisation d'un algorithme de calcul de flot maximal [Goldberg and Tarjan, 1988] pour séparer une graine d'utilisateurs du reste du graphe
- ► Complexité : $O(n^2m)$ (n : nœuds, m : arêtes)
- Versions en ligne de ces techniques [Flake et al., 2002, Senellart, 2005] permettant de faire le calcul sans avoir à stocker le graphe localement

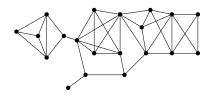




- Utilisation d'un algorithme de calcul de flot maximal [Goldberg and Tarjan, 1988] pour séparer une graine d'utilisateurs du reste du graphe
- ► Complexité : $O(n^2m)$ (n : nœuds, m : arêtes)
- Versions en ligne de ces techniques [Flake et al., 2002, Senellart, 2005] permettant de faire le calcul sans avoir à stocker le graphe localement

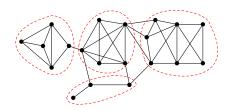
Markov Cluster Algorithm (MCL) [van Dongen, 2000]

- Algorithme de classification (clustering) de graphe
- Aussi basé sur le flot maximal, dans le graphe tout entier
- Itération d'un calcul matriciel alternant :
 - Expansion (multiplication matricielle, correspondant à la propagation de flot)
 - Inflation (opération non linéaire pour augmenter l'hétérogénéité)
- Complexité : O(n³) pour un calcul exact, O(n) pour un calcul approché



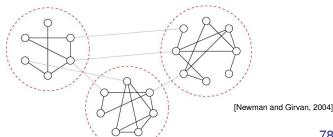
Markov Cluster Algorithm (MCL) [van Dongen, 2000]

- Algorithme de classification (clustering) de graphe
- Aussi basé sur le flot maximal, dans le graphe tout entier
- Itération d'un calcul matriciel alternant :
 - Expansion (multiplication matricielle, correspondant à la propagation de flot)
 - Inflation (opération non linéaire pour augmenter l'hétérogénéité)
- Complexité : O(n³) pour un calcul exact, O(n) pour un calcul approché



Suppression des arêtes les plus

- « au milieu » [Newman and Girvan, 2004]
 - Algorithme descendant de classification de graphe
 - Degré de « betweenness » d'une arête : nombre de chemins minimaux entre deux sommets quelconques qui passent par cette arête
 - Principe général :
 - 1. Calculer la betweenness de chaque arête du graphe
 - 2. Enlever l'arête avec la plus haute betweenness
 - 3. Recommencer, le calcul de betweenness compris
 - ► Complexité : O(n³) pour un graphe creux



Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Modèles de graphes

Recherche dans les réseaux sociaux

Recherche de communautés

Utilisation des liens sociaux pour améliorer la recherche

Confiance

Conclusion

Utilisation des liens sociaux pour améliorer la recherche

Est-il possible d'exploiter les liens entre utilisateurs et le contenu des réseaux sociaux pour améliorer la recherche d'information?

- Utilisation des tags associés à des contenus
- Contenu recommandé pour un utilisateur donné
- Recherche par mots-clefs biaisée par le réseau social d'un individu

Utilisation des marque-pages de Delicious [Heymann et al., 2008]

Tags dans Delicious : utilisables dans une recherche par mots-clefs?

Avantages

- Termes de requêtes et tags se recoupent
- Cohérents, pertinents
- Dynamiques

Inconvénients

- Faible couverture
- Manquent souvent de contexte

Systèmes de recommandation [Resnick et al., 1994]

- Filtrage collaboratif
- Ensemble de produits, ensemble d'utilisateurs
- Utilisateurs vu comme un vecteur dans l'espace des produits (pondération : plus un produit est courant, moins il a d'importance)
- Mesure de similarité entre utilisateurs (p. ex., cosinus entre les vecteurs)
- Recommandations pour un utilisateur donné :
 - 1. Calculer les utilisateurs les plus similaires
 - 2. Proposer les produits associés
- Utile pour les sites de rencontre!

Problèmes

- Passe mal à l'échelle avec un grand nombre d'utilisateurs
- Fonctionne mal pour un nouvel utilisateur

Amazon.com: recommandation produit-à-produit [Linden et al., 2003]

- ► Filtrage collaboratif produit-à-produit
- Produits : vecteurs dans l'espace des utilisateurs
- Similarité entre produits
- Précalcul des similarités entre paires de produit (coûteux, mais change peu)
- Recommandations pour un utilisateur donné :
 Proposer les objets les plus similaires aux objets de l'utilisateur

Recherche d'information classique

- ► Index en texte intégral d'une collection de documents, sous la forme de listes « terme → (document, poids) », triées par poids
- Fonction de pondération : par exemple tf-idf (term frequency-inverse document frequency)
- Requête par mots-clefs :

$$t_1$$
 AND ... AND t_n

k premiers résultats?

Notations:

```
s(t, d) pondération de t dans d (p. ex., tf-idf)
```

 $g(s_1,\ldots,s_n)$ fonction croissante calculant le score global (p. ex., addition)

Algorithme de seuil de Fagin [Fagin et al., 2001]

- 1. Soit *R* la liste vide, et $m = +\infty$.
- 2. Pour tout 1 < i < n:
 - 2.1 Récupérer le document $d^{(i)}$ contenant le terme t_i avec le meilleur score $s(t_i, d^{(i)})$ (parmi ceux non encore récupérés).
 - 2.2 Calculer le score global $g_{d^{(i)}} = g(s(t_1, d^{(i)}), \dots, s(t_n, d^{(i)}))$ en récupérant chacun des $s(t_i, d^{(i)})$ avec $i \neq i$.
 - 2.3 Si R contient moins de k documents, ou si $g_{d^{(i)}}$ est plus grand que le minimum des scores des documents dans R, ajouter $d^{(i)}$ à R (en enlevant éventuellement le document avec le score minimum).
- 3. Soit $m = g(s(t_1, d^{(1)}), s(t_2, d^{(2)}), \dots, s(t_n, d^{(n)}))$.
- 4. Si *R* contient plus de *k* documents, et le minimum des scores des documents dans *R* est supérieur ou égal à *m*, retoruner *R*.
- 5. Répéter l'étape 2.

Recherche d'information avec fonction de score sociale [Schenkel et al., 2008]

- Cadre : graphe multi-partite, p. ex., Flickr
- But : biaiser les résultats de requête en fonction de son réseau social
- Pondération sociale :
 - ► Étant donnée une relation d'amitié F(u, u') (explicite ou implicite) entre deux utilisateurs, on calcule une relation d'amitié étendue

$$\tilde{F}(u,u') = \frac{\alpha}{|U|} + (1-\alpha) \max_{\text{chemin } u = u_0 \dots u_k = u'} \prod_{i=0}^{\kappa-1} F(u_i,u_{i+1})$$

 $(0 < \alpha < 1 \text{ constante}; |U| : \text{nombre d'utilisateurs})$

► Au lieu de prendre une pondération globale

$$\mathsf{tf}\text{-}\mathsf{idf}(t,d) = \mathsf{tf}(t,d) \times \mathsf{idf}(t,d)$$

on prend une pondération sociale dépendant de u :

$$\mathsf{tf}\text{-}\mathsf{idf}_u(t,d) = \left(\sum_{t \in U} F(u,u') \cdot \mathsf{tf}_{u'}(t,d)\right) \times \mathsf{idf}(t,d)$$

Top-*k* avec score social [Benedikt et al., 2008]

- Possibilité d'adapter l'algorithme de seuil de Fagin...
- \blacktriangleright ... mais impossible de précalculer les scores tf-idf_u(t, d) pour chacun des utilisateurs
- Pour éviter une trop grande complexité :
 - Partitionner le graphe des utilisateurs en composantes d'utilisateurs fortement similaires
 - Utiliser les scores au sein de ces composantes comme estimations du seuil dans l'algorithme de Fagin
 - 3. ⇒ donne des résultats approchés, mais de bonne qualité

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Confiance

Définitions : confiance et réputation Systèmes de réputation et de confiance

Conclusion

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Confiance

Définitions : confiance et réputation

Systèmes de réputation et de confiance

Conclusion

Confiance

Dans la littérature, le terme Confiance est utilisé avec une variété de sens.

Décrite comme étant :

- Une valeur subjective (sentiment, croyance) qu'on associe à la capacité de quelqu'un ou quelque chose à réaliser une action donnée.
 Ex. Avoir confiance en votre voisin pour s'occuper de votre chat ...
- Le degré de disposition à dépendre de quelqu'un ou de quelque chose dans une situation donnée, même en cas de conséquences négatives.

Ex. Avoir confiance en votre guide de montagne ...

Mais aussi:

- 3. Un ensemble d'attentes partagées par tous ceux qui sont impliqués dans un échange (Zuker, 86). Ex. transactions financières.
- 4. ..

Définition 1 (reliability trust):

« Trust is the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends » [Jøsang et al. 07, Gambetta 88]

 Valeur subjective : Propre à chaque individu

Exprime une attente (en terme de fiabilité):
 A s'attend à ce que B réagisse ou réalise une action d'une certaine manière
 La fiabilité de B telle qu'elle est perçue par A.

Lien de dépendance :

Le bien être (la prospérité) de A dépend de l'action de B.

Est-elle uniquement liée à la fiabilité ?

Une confiance élevée (en terme de fiabilité) en quelqu'un peut ne pas suffire pour se maître en situation de dépendance envers cette personne

Lorsque:

les conséquences négatives (en cas d'échec) ne sont pas acceptables, ou ne vous permettent pas de vous décider. Indépendamment de la probabilité de l'échec

Ex. Le fait que votre voisin soit riche et très bon gestionnaire ne suffit pas pour vous décider à s'associer à lui en affaires.

Définition 2 (Decision trust):

« Trust is the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible » [Jøsang et al. 07, McKnight&Chervany 96]

• Dépendance envers le porteur de confiance :

Consentement et sentiment de sécurité Indirectement : fiabilité du porteur de confiance.

Utilité :

Positive/négative en fonction des conséquences.

Prise de risque :

Par exemple, lorsque la valeur des transactions (ex; achats) est élevée et la fiabilité <1.

Confiance et réputation

Réputation (Larousse) :

« Manière dont une personne, une chose sont considérés. Opinion favorable ou défavorable que le public a d'une personne ou d'une chose. »

Dans le Web communautaire :

Une valeur déduite d'un réseau social et visible à tous ses membres.

Lien avec la confiance :

- La confiance peut être basée sur la réputation
 Ex. : je fais confiance à mon médecin parce qu'il a une bonne réputation.
- Elle peut aussi ne donner qu'un poids marginal (ou nul) à la réputation
 Ex. : je fais confiance à mon voisin malgré sa mauvaise réputation
 Je donne plus de poids à la connaissance que j'ai de mon voisin qu'à sa réputation.

Confiance et réputation (suite)

Confiance:

Mesure personnelle et subjective qui s'appuie sur une variété de faits, dont certains peuvent avoir plus de poids que d'autres.

→ typiquement, l'expérience ou la connaissance personnelle.

Réputation :

Mesure collective de la crédibilité (fiabilité) basée sur les notations (opinions) faites par les membres d'une communauté.

Précautions :

- (1) basée uniquement sur son expérience personnelle,
- (2) ou pondérées en fonction du nombre de notations tierces prise en compte.

Confiance et réputation (suite)

Réputation d'un groupe :

- une mesure globale qui reflète comment le groupe dans son ensemble est considéré de l'extérieur,
- moyenne des réputations de ses membres,
- chaque membre hérite d'une réputation a priori sur la base de la réputation de son groupe.

Confiance et réputation (suite)

Monde réel:

le partage d'information sur le confiance et la réputation est difficile et se limite souvent à des communautés locales (bouche à oreille).

Web:

Facilite la collecte, l'échange et le partage de ces informations à une échelle beaucoup plus large.

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Confiance

Définitions : confiance et réputation

Systèmes de réputation et de confiance

Conclusion

Objet de la recherche dans ce domaine :

- 1. Trouver les bons substituts aux indicateurs traditionnels (du mode réel).
- Trouver de nouveaux indicateurs spécifiques à chaque application.
- Créer des systèmes efficaces pour collecter ces informations et en déduire des mesures de confiance et de réputation.

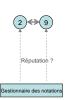
Propriétés principales [Jøsang et al. 07, Resnick et al. 00]:

- Pérennité (longévité) des acteurs :
 Pas de changement d'identité pour échapper au passé.
 Chaque échange doit pouvoir être suivi par d'autres.
- Les notations des échanges courants doivent être collectées et diffusées (notamment en cas de systèmes distribués).
- 3. Les notations des échanges passés doivent guider la décision pour les échanges actuels

Systèmes centralisés :

Les notations sont rapportées à un gestionnaire central qui se charge de du calcul de la réputation de chaque membre.





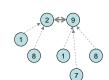
Systèmes décentralisés :

- a plusieurs gestionnaires décentralisés, pouvant être au niveau de chaque membre
- b les membres peuvent communiquer leurs scores entre eux
- c le calcul de réputation se fait au niveau de chaque membre, en fonction des notations reçues et de l'expérience personnelle
- → basé sur un sous ensemble de notations





2- Présent



Systèmes de réputation :

Produit un score public (réputation), reflète la vision de l'ensemble de la communauté de la crédibilité d'une entité.

Mécanismes : Agrégation et pondération.

Input : notations « objectives » de faits précis (Ex. vente, publication, etc.)

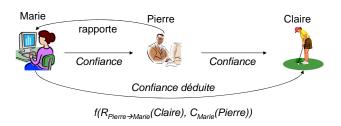
Systèmes de confiance :

Produit un score qui reflète la vision subjective des parties concernées (consultées) de la crédibilité d'une entité.

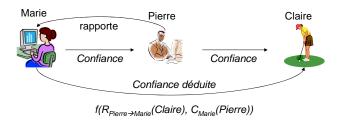
Mécanismes : Transitivité.

Input: mesures de confiance subjectives, ne s'appuyant pas sur des faits précis.

Confiance : principe de transitivité.

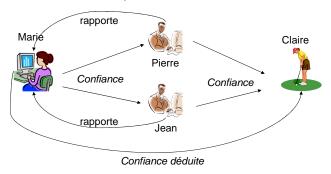


Confiance : principe de transitivité.



- 1. Confiance au sens fiabilité (Reliability trust)
- 2. Marie doit avoir confiance en Pierre pour recommander Claire pour la tache en question

Confiance: Chaînes parallèles.



 $f(R_{Pierre o Marie}(Claire), C_{Marie}(Pierre), R_{Jean o Marie}(Claire), C_{Marie}(Jean))$

Calcul de la réputation :

- Comptage des scores positifs et négatifs, ou moyenne des notations.
 - → la plus part des systèmes commerciaux (Ex. eBay et Amazon)

2. Techniques probabilistes

- Approches bayésiennes [A. Whitby et al 2005]: prise en compte du passé (réputation a priori) pour déterminer l'incertitude sur l'issue des échanges à venir et la réputation a posteriori.
 - \rightarrow Input: nbre de scores + (x) et nombre de scores (y).
 - →Incertitude : exprimée par une fonction bêta de densité de probabilité bêta(p; x, y).
 - \rightarrow E(p) = x / (x + y). La réputation et fonction de E(p).
- Modèle de convictions (belief model) [A. Jøsang 2001]
 confiance et réputation sont exprimées sous forme d'opinion (conviction, conviction du contraire, et incertitude).

Calcul de la réputation :

- Modèles de flux : La réputation augmente en fonction des flux entrants et diminue en fonction des flux sortants.
 - → Ex. PageRank: réputation d'une page augmente en fonction de la réputation des pages qui la référencent.
 - +EigenTrust, TrustRank, FolkRank, etc.
- Modèles basés sur la logique floue : modélisent la confiance et la réputation comme des concepts flous.

PageRank:

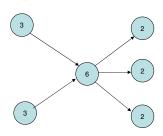
Graphe orienté $G=(V, \mathcal{E})$ Card(V)=N pages

Seuls les flux rentrants (notes positifs) comptent dans le score:

> R(p) et fonction de r(q), q étant une page qui référence p.

 $\omega(q)$ = nbre de liens sortants

$$R(p) = \alpha \cdot \sum_{\alpha:(\alpha,p) \in \varepsilon} r(\alpha)/\omega(\alpha) + (1-\alpha).1/N$$



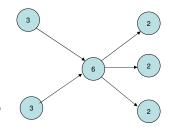
PageRank [L. Page et al. 1998]:

Graphe orienté $G=(V, \mathcal{E})$ Card(V)=N pages

Seuls les flux rentrants (notes positifs) comptent dans le score:

R(p) et fonction de r(q), q étant une page qui référence p.

 $\omega(q)$ = nbre de liens sortants



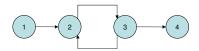
$$R(p) = \alpha \cdot \sum_{q:(q,p) \in \varepsilon} r(q)/\omega(q) + (1-\alpha).1/N$$
partie dynamique
partie statique

PageRank (forme matricielle):

$$r = \alpha . T. r + (1 - \alpha).1/N.1_N$$

$$T(p,q) = \begin{cases} 0 & \text{si } (q,p) \notin \varepsilon \\ 1/\omega(q) & \text{si } (q,p) \in \varepsilon \end{cases}$$

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

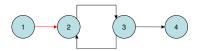


PageRank (forme matricielle):

$$r = \alpha . T. r + (1 - \alpha).1/N.1_N$$

$$T(p,q) = \begin{cases} 0 & \text{si } (q,p) \notin \varepsilon \\ 1/\omega(q) & \text{si } (q,p) \in \varepsilon \end{cases}$$

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

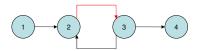


PageRank (forme matricielle):

$$r = \alpha . T. r + (1 - \alpha).1/N.1_N$$

$$T(p,q) = \begin{cases} 0 & \text{si } (q,p) \notin \varepsilon \\ 1/\omega(q) & \text{si } (q,p) \in \varepsilon \end{cases}$$

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

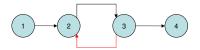


PageRank (forme matricielle):

$$r = \alpha . T. r + (1 - \alpha).1/N.1_N$$

$$T(p,q) = \begin{cases} 0 & \text{si } (q,p) \notin \varepsilon \\ 1/\omega(q) & \text{si } (q,p) \in \varepsilon \end{cases}$$

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

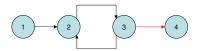


PageRank (forme matricielle):

$$r = \alpha . T. r + (1 - \alpha).1/N.1_N$$

$$T(p,q) = \begin{cases} 0 & \text{si } (q,p) \notin \varepsilon \\ 1/\omega(q) & \text{si } (q,p) \in \varepsilon \end{cases}$$

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

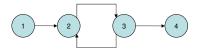


PageRank (forme matricielle):

$$r = \alpha . T. r + (1 - \alpha).1/N.1_N$$

$$T(p,q) = \begin{cases} 0 & \text{si } (q,p) \notin \varepsilon \\ 1/\omega(q) & \text{si } (q,p) \in \varepsilon \end{cases}$$

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

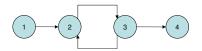


PageRank (forme matricielle):

$$r = \alpha . T. r + (1 - \alpha).1/N.1_N$$

$$T(p,q) = \begin{cases} 0 & \text{si } (q,p) \notin \varepsilon \\ 1/\omega(q) & \text{si } (q,p) \in \varepsilon \end{cases}$$

$$T = \left(\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{array} \right)$$



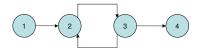
PageRank (forme matricielle):

$$r = \alpha . T. r + (1 - \alpha).1/N.1_N$$

du score statique

$$T(p,q) = \begin{cases} 0 & \text{si } (q,p) \notin \varepsilon \\ 1/\omega(q) & \text{si } (q,p) \in \varepsilon \end{cases}$$

$$T = \left(\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{array} \right)$$



TrustRank [Z. Gyöngyi et al 2004]:

- Objet → reconnaître les pages de spam sur le web.
- Idée :
 - les bonnes pages pointent vers de bonnes pages
 - détermination par un processus semi-automatique d'un ensemble de bonnes pages initiales :
 - note=0 → spam, note=1 → page normale
 - Constitution d'un vecteur de distribution statique d :

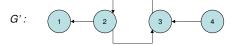
 Propagation de la réputation en utilisant un PageRank biaisé : remplacement de la distribution uniforme par d

$$r = \alpha . T. r + (1 - \alpha).d$$
d : vecteur de distribution de score statique

TrustRank [Z. Gyöngyi et al 2004]:

- Choix des pages à expertiser :
 - Préférence aux pages ayant le plus de liens sortants.
 - Classement des pages selon le nombre de liens sortants :

Utilisation du PageRank sur un graphe inversé $G'=(V, \varepsilon')$, avec $(p,q) \in \varepsilon'$ ssi $(q,p) \in \varepsilon$.



Expertise des L premières pages

FolkRank [A. Hotho et al 2006]:

les nœuds du graphe représentent des utilisateurs, des tags et des ressources

les arcs user → tag, tag → ressource et user → ressource.

La réputation de la ressource est fonction de l'importance de ses tags et de la réputation des utilisateurs qui lui ont associé des tags.

FolkRank [A. Hotho et al 2006]:

les nœuds du graphe représentent des utilisateurs, des tags et des ressources

les arcs user → tag, tag → ressource et user → ressource.

La réputation de la ressource est fonction de l'importance de ses tags et de la réputation des utilisateurs qui lui ont associé des tags.

Contrôle d'accès basé sur la confiance

[D. Hong et V. Y. Shen 2008]:

Réseaux de confiance étiquetés :

Utilisation de tags (context) associés aux liens entre individus :
 Ex. ami, collègue de travail, copain de lycée.

Propagation des droits d'accès par transitivité dans le réseau de confiance :

Permission = Trust(truster, trustee, context)

Plan de l'exposé

Web communautaires

Exemples d'outils

Recherche d'information

Confiance

Conclusion

En résumé

- Webs communautaires, réseaux sociaux : partout sur le Web!
- Modélisation naturelle avec des graphes (le plus souvent, non orientés)
- ▶ Intérêts
 - Nouveaux types de recherches (p. ex., recherches de communautés)
 - Utilisation des tags et des réseaux sociaux pour améliorer et personnaliser les résultats de recherche
- Nombreuses techniques de fouille de graphes applicables :
 - à la recherche d'information
 - à l'évaluation de la confiance

Pour aller plus loin

```
Modèles de graphes
[Watts and Strogatz, 1998, Newman et al., 2006]
[Kleinberg, 2007]

Aspects sociologiques
[Lin, 2001]

Graphe du Web, PageRank et cie
[Chakrabarti, 2003]

Web et réseaux sociaux
[Brusilovsky, 2008]
```

Perspectives de recherche



- Applications des diverses techniques à la grande diversité des réseaux sociaux :
 - un moteur de recherche intelligent dans Wikipedia
 - à qui faire confiance sur un site de rencontres ?
 - **•** . . .
- Comment se réapproprier les données des réseaux ? Architecture distribuée pour un réseau social.
- Gestion de droits d'accès plus fin que l'existant, se basant sur la confiance a priori en les utilisateurs.

MERCI.

Bibliographie I

- Marshall D. Abrams and Michael V. Joyce. Trusted system concepts. *Computers & Security*, 14(1):45–56, 1995.
- Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.
- Michael Benedikt, Sihem Amer Yahia, Laks Lakshmanan, and Julia Stoyanovich. Efficient network-aware search in collaborative tagging sites. In *Proc. VLDB*, Auckland, New Zealand, August 2008.
- Vincent D. Blondel, Anahí Gajardo, Maureen Heymans, Pierre Senellart, and Paul Van Dooren. A measure of similarity between graph vertices: applications to synonym extraction and Web searching. *SIAM Review*, 46(4):647–666, 2004.
- Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

Bibliographie II

- Peter Brusilovsky. Social information access: The other side of the social Web. *SOFSEM 2008: Theory and Practice of Computer Science*, pages 5–22, 2008.
- Soumen Chakrabarti. *Mining the Web: Discovering Knowledge from Hypertext Data*. Morgan Kaufmann, San Fransisco, USA, 2003.
- Annie Danzart, Vincent Oria, Trung Tien Phan Quang, and Jean-Marc Saglio. Iceberg: un modèle de diffusion restreinte pour un réseau de confiance entre pairs. In *Rapport de recheche, TELECOM ParisTech*, 2006.
- Ronald Fagin, Amnon Lotem, and Moni Naor. Optimal aggregation algorithms for middleware. In *Proc. PODS*, Santa Barbara, USA, May 2001.
- Gary Flake, Steve Lawrence, and C. Lee Giles. Efficient identification of Web communities. In *Proc. SIGKDD*, pages 150–160, Boston, USA, August 2000.

Bibliographie III

- Gary William Flake, Steve Lawrence, C. Lee Giles, and Frans Coetzee. Self-organization of the Web and identification of communities. *IEEE Computer*, 35(3):66–71, 2002.
- Andrew V. Goldberg and Robert E. Tarjan. A new approach to the maximum-flow problem. *Journal of the ACM*, 35(4): 921–940, October 1988.
- Tyrone Grandison and Morris Sloman. A Survey of Trust in Internet Applications. *IEEE Communications Surveys and Tutorials*, 3(4), September 2000.
- R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proc. WWW*, pages 403–412, New York, USA, 2004.
- Zoltán Gyöngyi, Hector Garcia-Molina, and Jan Pedersen. Combating web spam with trustrank. In *Proc. VLDB*, pages 576–587, Toronto, Canada, August 2004.

Bibliographie IV

- Paul Heymann, Georgia Koutrika, and Hector Garcia-Molina. Can social bookmarking improve web search? In *Proc. WSDM*, pages 195–206, Palo Alto, USA, February 2008.
- Dan Hong and Vincent Y. Shen. Setting access permission through transitive relationship in Web-based social networks. In *Proc. SWKM*, April 2008.
- Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Folkrank: A ranking algorithm for folksonomies. In *LWA*, pages 111–114, 2006.
- Jean Houard and Marc Jacquemain. *Capital social et dynamique régionale*. De Boek, Bruxelles, Belgium, 2006.
- Audun Josang. A logic for uncertain probabilities. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 9(3):279–311, 2001.

Bibliographie V

- Audun Josang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decis. Support Syst.*, 43(2):618–644, 2007.
- Jon Kleinberg. The structure of information networks. http://www.cs.cornell.edu/courses/cs685/2007fa/, 2007. Cornell University Course.
- Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.
- Nan Lin. Social Capital: A Theory of Social Structure and Action. Cambridge University Press, Cambridge, United Kingdom, 2001.
- Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.

Bibliographie VI

- M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 2004.
- Mark Newman, Albert-László Barabási, and Duncan J. Watts. The Structure and Dynamics of Networks. Princeton University Press, 2006.
- Yann Ollivier and Pierre Senellart. Finding related pages using Green measures: An illustration with Wikipedia. In *Proc. AAAI*, pages 1427–1433, Vancouver, Canada, July 2007.
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the Web. In *Tech. Report, Stanford University*, 1998.
- Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: an open architecture for collaborative filtering of netnews. In *Proc. CSCW*, Chapel Hill, United States, October 1994.

Bibliographie VII

- Ralf Schenkel, Tom Crecelius, Mouna Kacimi, Sebastian Michel, Thomas Neumann, Josiane X. Parreira, and Gerhard Weikum. Efficient top-k querying over social-tagging networks. In *Proc. SIGIR*, pages 523–530, Singapore, Singapore, July 2008.
- Pierre Senellart. Identifying Websites with flow simulation. In *Proc. ICWE*, pages 124–129, Sydney, Australia, July 2005.
- M. Smith, C. Giraud-Carrier, and B Judkins. Implicit affinity networks. In *Proc. Workshop on Information Technologies and Systems*, pages 1–7, Montreal, Canada, December 2007.
- Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, 34(4), December 1969.
- Stijn Marinus van Dongen. *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, May 2000.
- Duncan J. Watts. *Small Worlds*. Princeton University Press, 1999.

Bibliographie VIII

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

Andrew Whitby, Audun Jøsang, and Jadwiga Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proc. International Workshop on Trust in Agent Societies*, 2004.