



Explorer et comprendre les réseaux sociaux

Une perspective croisée
informatique et sociologie

Cécile Bothorel, Cecile.bothorel@telecom-bretagne.eu
Karine Roudaut, Karine.roudaut@telecom-bretagne.eu
Département LUSI, Brest





1. L'analyse des réseaux sociaux

Définition Historique



L'analyse des réseaux sociaux

- **Elle ne se réduit pas à l'étude des réseaux sociaux numériques**
- **Un ensemble de méthodes, concepts, théories, modèles, enquêtes**
- **Des disciplines diverses**
 - sociologie, anthropologie, économie, psychologie sociale, informatique, mathématiques, etc.
- **Elle étudie « les relations entre des unités sociales et leurs régularités »** [P. Merklé, 2011, Sociologie des réseaux sociaux, Paris, La Découverte]



L'analyse des réseaux sociaux

- Description
 - Relations sociales
- Formation
 - De la création de relations au Réseau social
- Transformation
 - Évolution des relations, du réseau
- Effets sur les comportements individuels.

Comportements individuels ↔ Structure du réseau



Structure de réseau

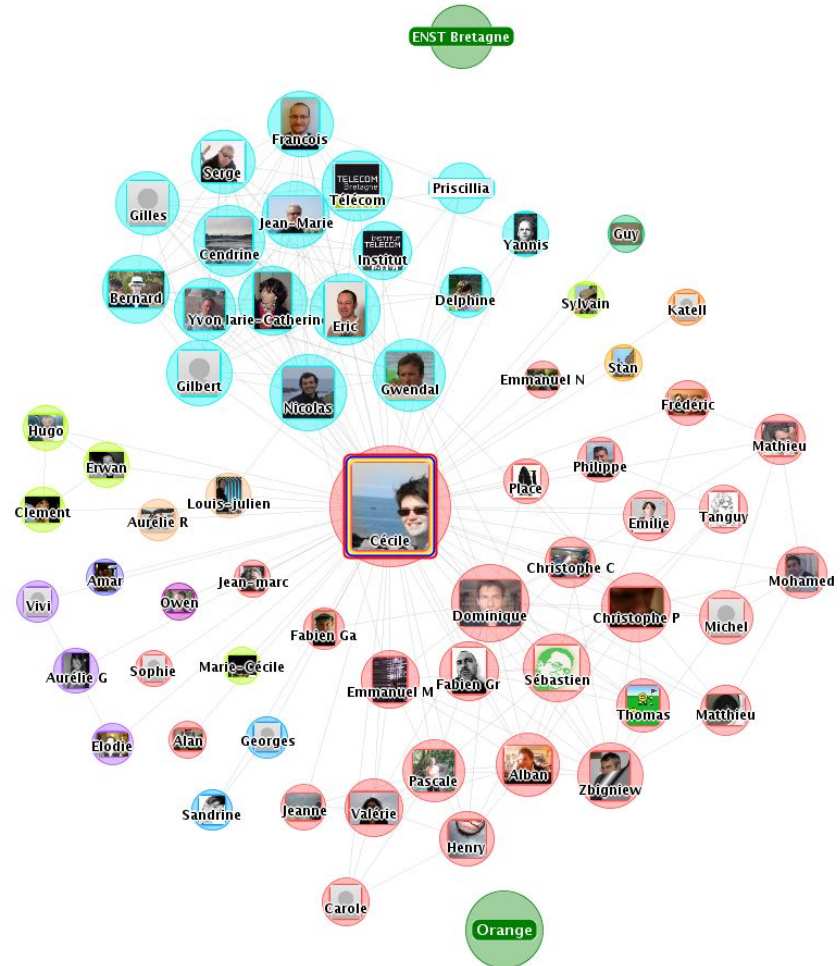
facebook Accueil Profil Amis Boîte de réception Cécile Bothorel Paramètres Déconnexion

Rechercher des amis Créer une liste Modifier la liste Supprimer la liste

Connexions Rechercher des amis Parcourir Annuaire téléphonique Ajouts récents Misses à jour récentes

Listes Amis Pages ENST Bretagne Orange Amis Colloques Famille Invitez vos amis à rejoindre Facebook

Alan Perrio	Ajouter à la liste	X
Alban Martin HEC Paris Colloques	1 liste	X
Amar Lakel	Ajouter à la liste	X
Aurélie Germain Famille	1 liste	X
Aurélie Rey Amis	1 liste	X
Bernard Gourvenec Colloques	1 liste	X
Carole Cozanet Orange Colloques	1 liste	X
Cendrine Le Locat Colloques	1 liste	X
Christophe Cariou Colloques	1 liste	X
Christophe Prieur Colloques	1 liste	X
Clement Garel Amis	1 liste	X
Delphine Le Borgne Colloques	1 liste	X
Dominique Cardon Colloques	1 liste	X
Elodie Germain	Ajouter à la liste	X
Emilie Ogez Colloques	1 liste	X
Emmanuel Mahé	Ajouter à la liste	X
Emmanuel Navarro Colloques	1 liste	X
Eric Cousin Colloques	1 liste	X
Erwan Garel Amis	1 liste	X
Fabien Gandon Colloques	1 liste	X
Fabien Granjon Colloques	1 liste	X
Francois Deltour ENST Bretagne Colloques	1 liste	X
Frédéric Amblard Colloques	1 liste	X





L'analyse des réseaux sociaux

■ Un réseau social = un ensemble d'acteurs et des relations que ces acteurs entretiennent

- Ces acteurs : des individus, des groupes, des organisations.
- « Les relations entre les éléments désignent des formes d'interactions sociales » [Merklé, p.4].
 - Très diverses :
 - ex. « transactions monétaires, échanges de biens, échanges de services, transmission d'informations, toutes sortes d'interactions verbales ou gestuelles, ou participation commune à des événements, etc. » [Merklé, p.4].
 - De natures différentes
 - liens forts (familiaux, sentimentaux) / liens faibles ou + distants (relations d'affaire, de travail, connaissances).
 - Entretien directement ou indirectement.



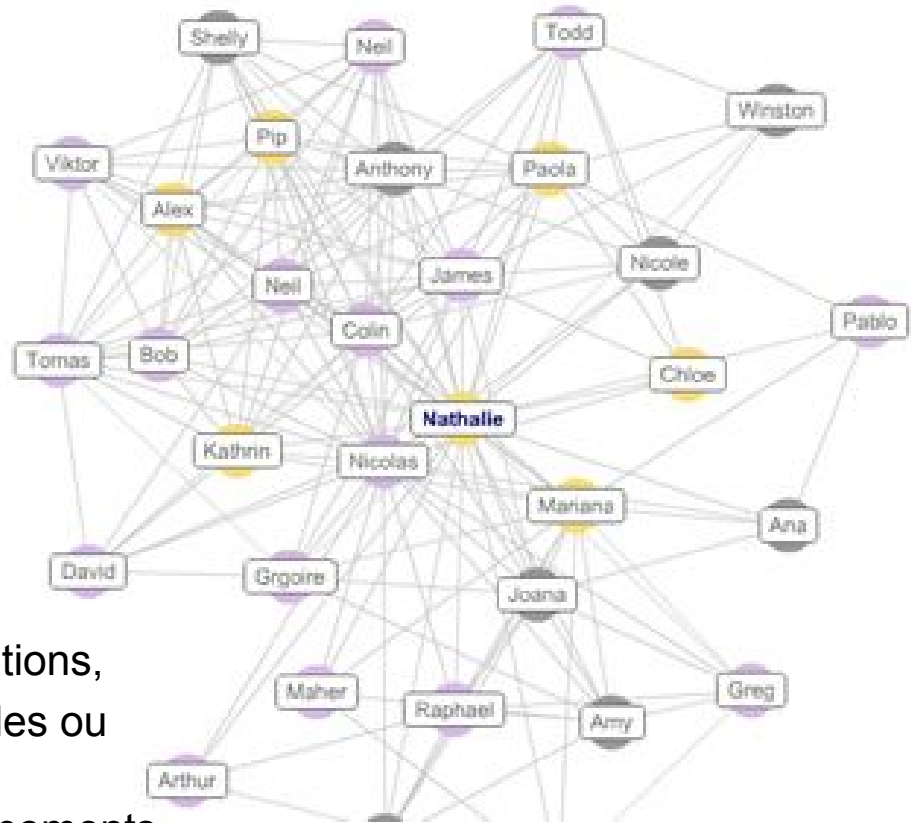
Qu'est-ce qu'un réseau social ?

■ Unités sociales

- individus,
- organisations...

■ Liens créés lors d'interactions sociales

- relation d'amitié,
- partenariat,
- intérêt commun,
- transactions monétaires,
- échanges de biens,
- échanges de services,
- Transmission maladies d'informations,
- toutes sortes d'interactions verbales ou gestuelles,
- participation commune à des événements
- transmission de maladie...





L'analyse des réseaux sociaux

■ 2 Questions :

- Échelle
- Point de vue pertinent d'observation
 - La triade = le plus petit réseau social concevable
 - L'étendue des réseaux



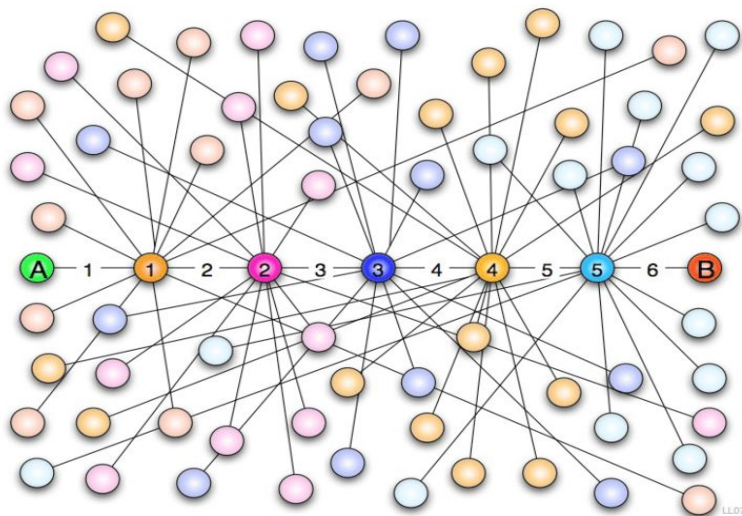
Exemples d'études empiriques fondatrices

- **1954 : John A. Barnes introduit le terme « réseau social »**
 - Une étude ethnographique de l'île de Bremmes (Norvège) sur :
 - le système des classes sociales ;
 - l'organisation de l'action sociale.
 - 3 champs sociaux distingués:
 - à base territoriale (administrations, associations);
 - le système industriel (le pêche);
 - les liens d'amitié et de connaissance entre individus.
 - Hypothèse de l'existence d'un réseau qui relie les individus par des chaînes d'interconnaissances.
 - Ce réseau n'a pas de frontière (il ne s'arrête pas avec le territoire, ni avec le système industriel).



Exemples d'études empiriques fondatrices

■ 1967 : Stanley Milgram : effet du « petit monde », six degrés de séparation

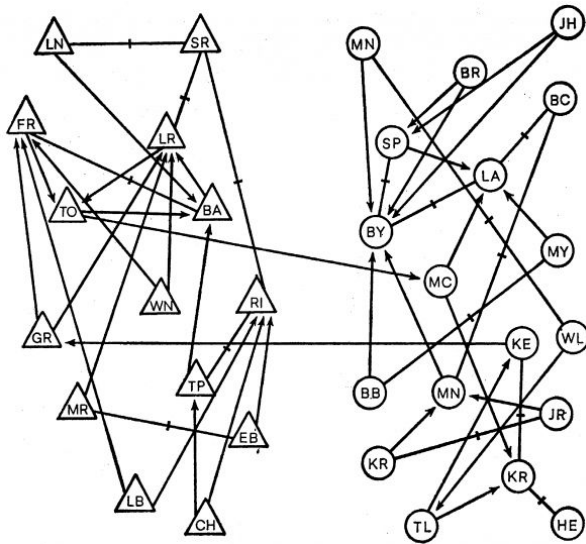


- Participants
 - un agent de change résidant à Boston comme « individu-cible »
 - 3 groupes de départ, 296 individus au total
- Mission : faire parvenir ce dossier par la poste à l'individu-cible,
 - soit directement si on le connaît
 - soit par un intermédiaire connu
- Résultats :
 - 217 ont envoyé le dossier.
 - 64 dossiers sont parvenus à l'individu-cible, des chaînes de connaissances de longueurs variables, mais dont la longueur moyenne était de 5,2 intermédiaires.

<http://fr.wikipedia.org>



Exemples d'études empiriques fondatrices



Les lignes barrées indiquent les choix réciproques
Source : Moreno [1934, annexes planche XII].

« Une représentation graphique des interrelations qui unissent les membres et les groupes d'une même communauté »

1934 : J. Moreno et les fondements de la sociométrie

- Etude des affinités entre les pensionnaires d'un Institut de Rééducation
- 506 jeunes filles, 16 pavillons d'habitation (éthnies séparées), activités communes
- Test sociométrique
 - Sympathie/antipathie à l'égard des autres membres et cohabitation
 - 2200 déclarations de choix
- Résultats
 - Structuration sociale des relations invisible et différente de l'organisation officielle
 - Un usage analytique et opératoire de la notion de réseau social

2. Les outils pour l'exploration



Analyse structurale de l'ensemble du réseau

Analyse de positions

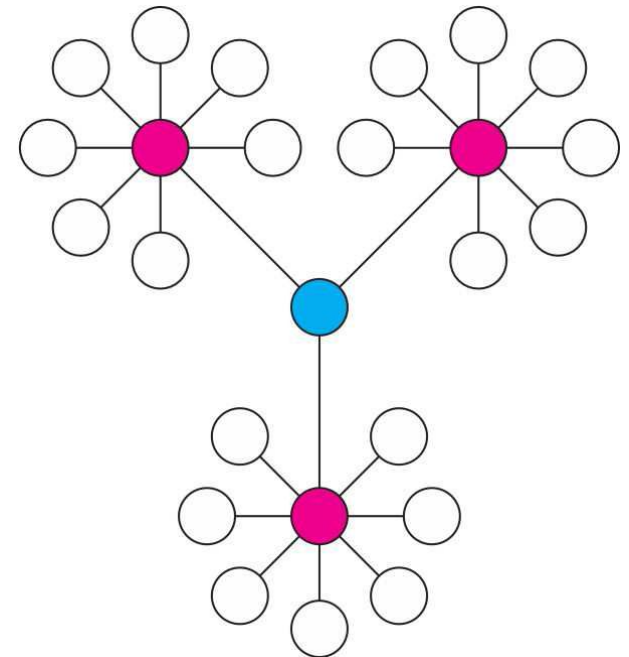
Détection de communautés



(1) Structure globale du réseau

■ Comprendre la structure globale

- Taille
- Distance moyenne, diamètre
- Connectivité, densité des relations
- Distribution des degrés, la variance des degrés
- Transitivité – coefficient de clustering
- La centralité du graphe



Réseau très central qui peut « tourner » autour d'un nœud. Si ce nœud disparaît...



(1) Structure globale du réseau

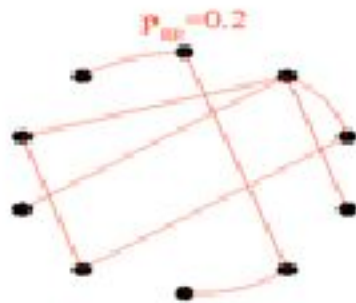
■ Formes connues de graphes

- Graphe aléatoire (Erdős et Rényi, 1959)
- Scale-free : attachement préférentiel (A.-L. Barabasi & R. Albert, 1999)
- Réseau hiérarchique : scale-free dans les clusters (A.-L. Barabasi et al., 2002)
- Petit-monde (D. Watts and S. Strogatz, 1998) différente de l'organisation officielle

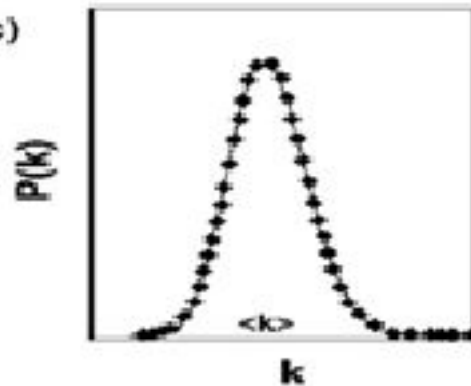


(1) Structure globale du réseau

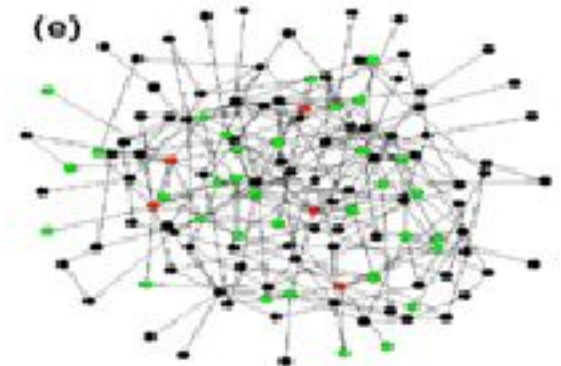
(a)



(c)



(e)



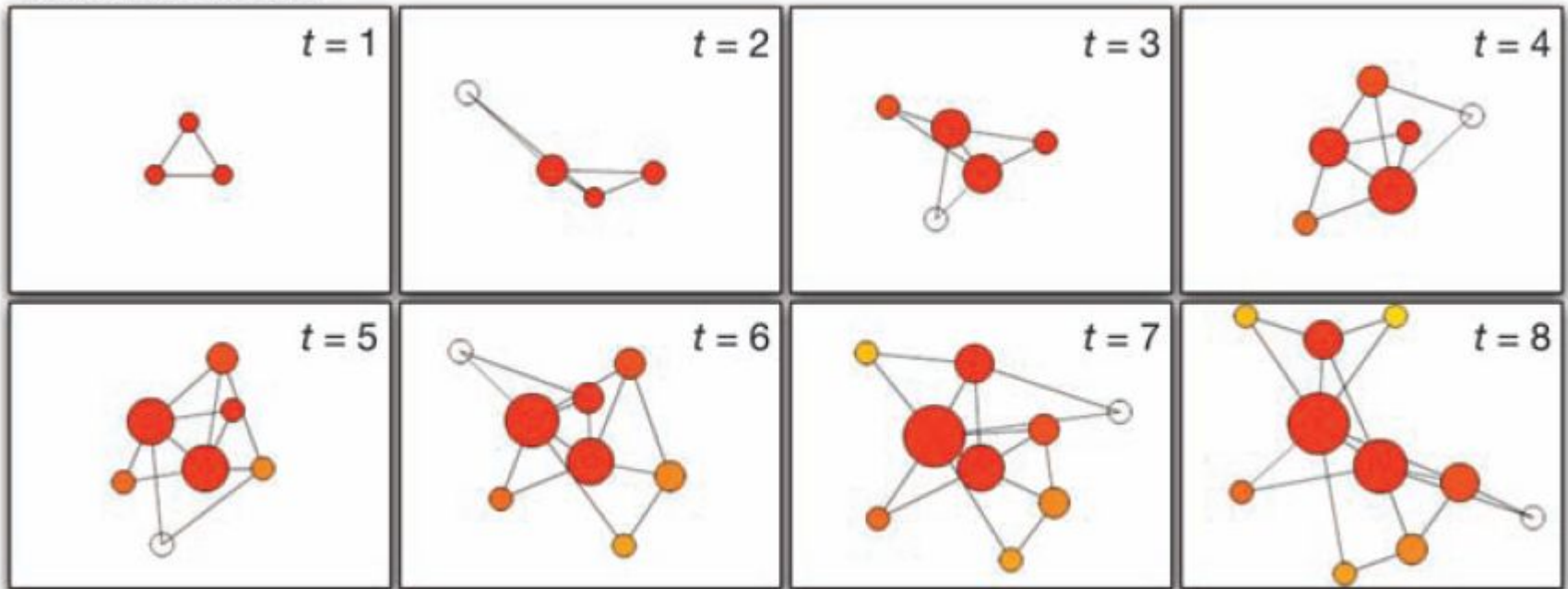
■ Graphe aléatoire (Erdős et Rényi, 1959)

- $N = 10$ et $p = 0.2$. Le réseau croît de façon homogène ($pN(N-1)/2$ liens)
- La plupart des noeuds ont le même degré (proche de la moyenne)
- Petit monde



(1) Structure globale du réseau

Scale-Free Model

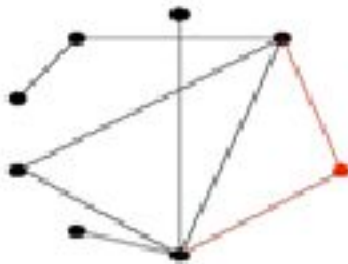


- **Scale-free : attachement préférentiel (A.-L. Barabasi & R. Albert, 1999)**
 - Décrit une dynamique et une structure

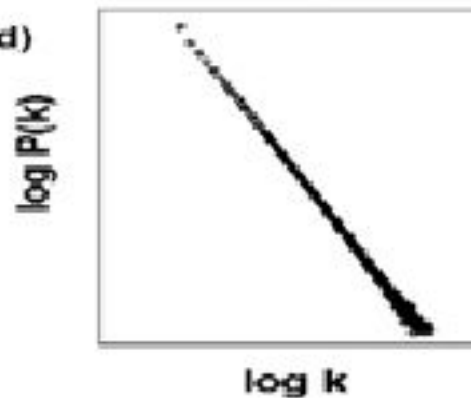


(1) Structure globale du réseau

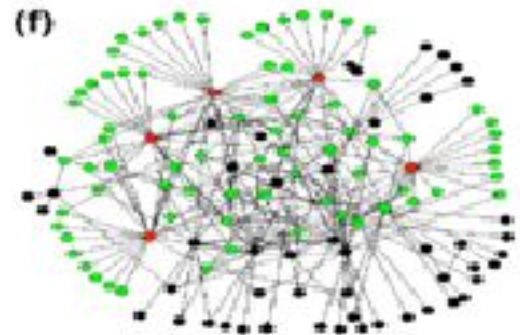
(b)



(d)



(f)



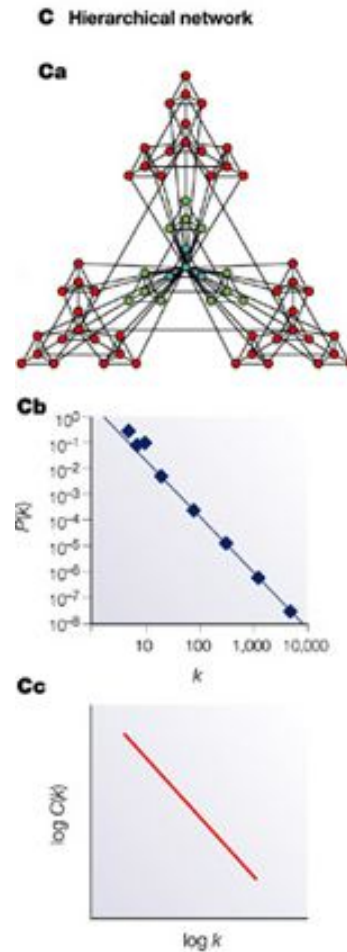
■ Scale-free : attachement préférentiel (A.-L. Barabasi & R. Albert, 1999)

- La distribution des degrés en loi de puissance
- Présence de hubs rouges qui ont attiré un grand nombre de voisins verts



(1) Structure globale du réseau

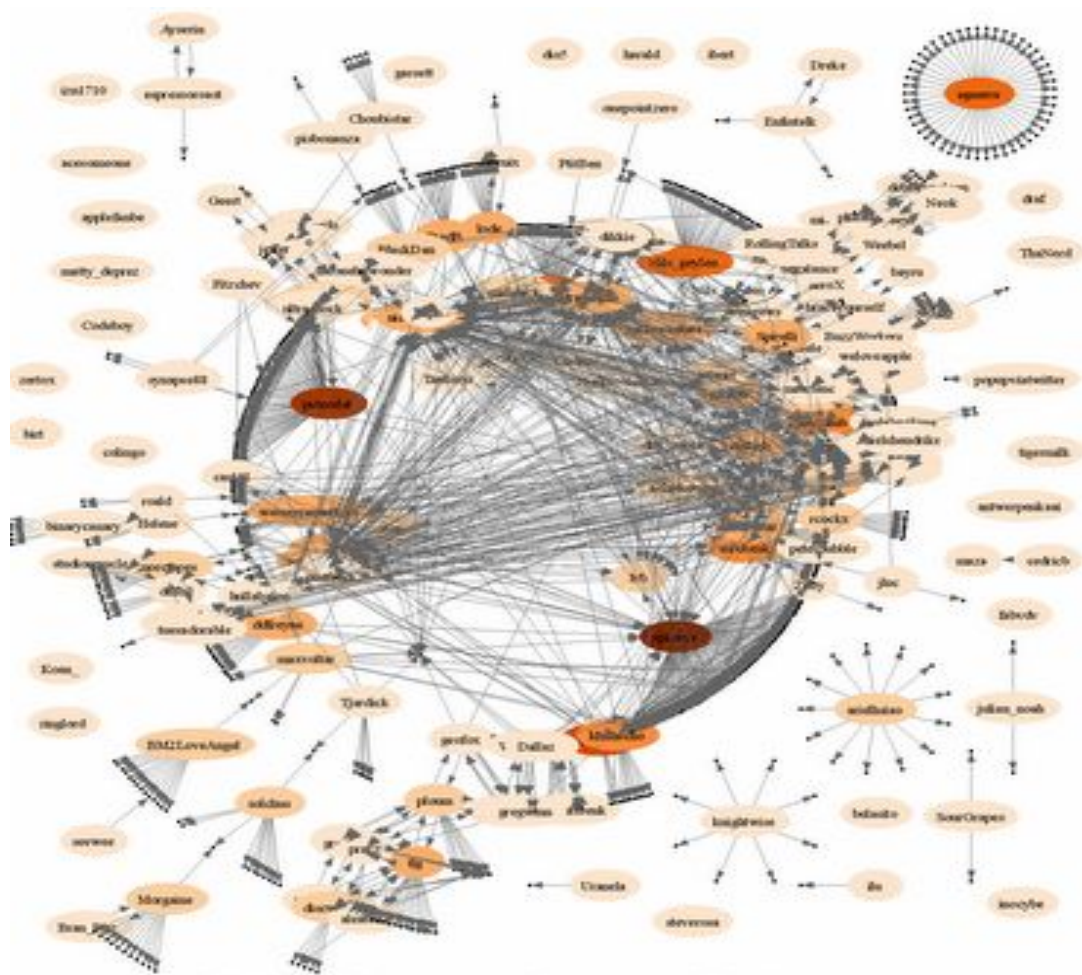
- Réseau hiérarchique : scale-free dans les clusters (A.-L. Barabasi et al., 2002)
- Petit-monde : petit diamètre de l'ordre de $\log(N)$ et fort coefficient de clustering (D. Watts and S. Strogatz, 1998)



Barabasi & Oltvai, vol.5, p101 (february 2004)



(2) Analyse de positions



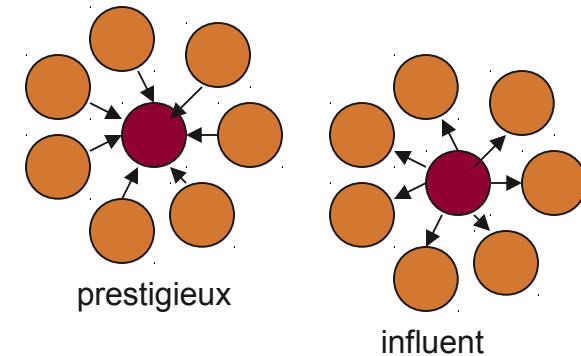
Twitter Friends van Belgische Twitteraars

(2) Analyse de positions

■ Positions clés, rôles

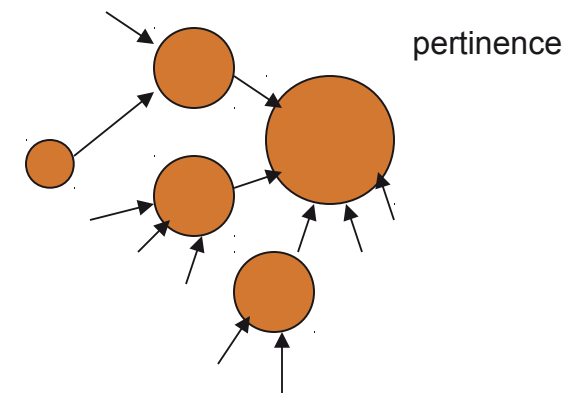
- Localement

exemple : fort degré – beaucoup de voisins immédiats



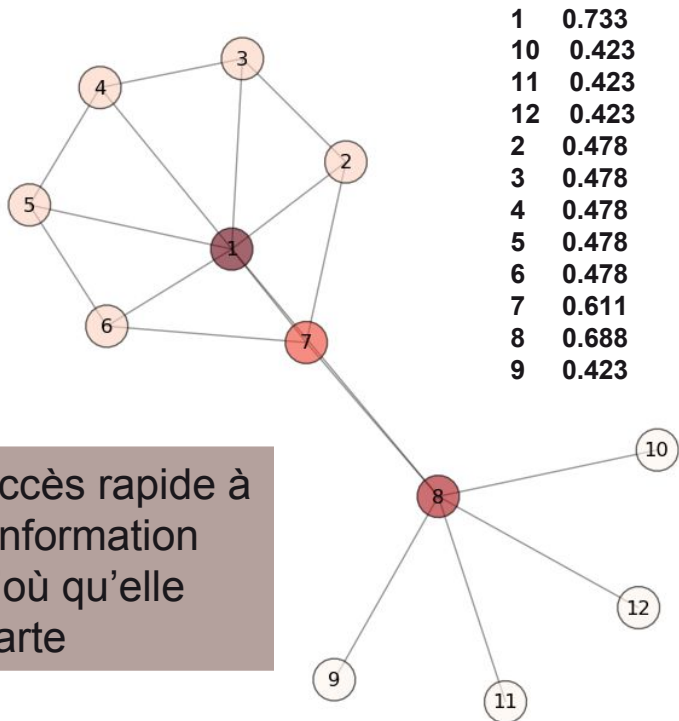
- Globalement

exemple : PageRank : être connecté à des nœuds importants me donne plus d'importance que d'être connecté à des nœuds peu importants



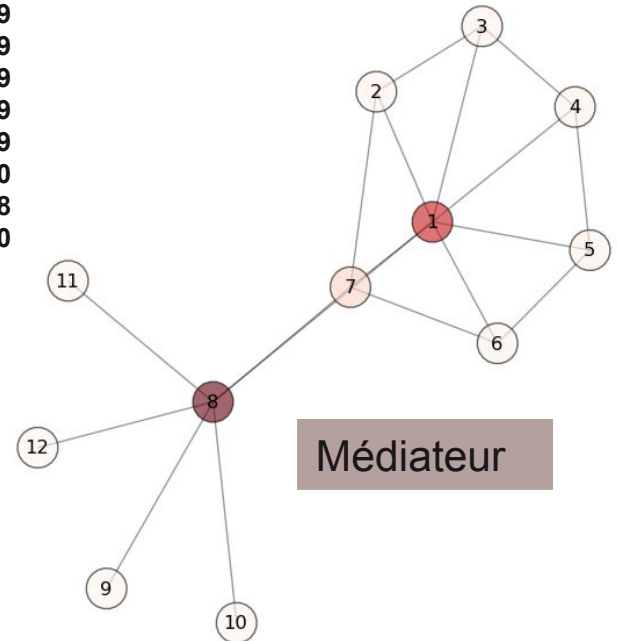


(2) Analyse de positions



Closeness centrality est l'inverse de la somme des distances entre ce nœud et chacun des autres

1	0.473
10	0.000
11	0.000
12	0.000
2	0.009
3	0.009
4	0.009
5	0.009
6	0.009
7	0.100
8	0.618
9	0.000



Betweenness centrality est le nombre de fois qu'un nœud se trouve sur le plus court chemin entre deux autres nœuds



(2) Analyse de positions

■ Centralité(s)

- Basée sur les calculs de plus courts chemins
 - Degree centrality (popularité, comportement grégaire)
 - Betweenness centrality (médiateur ou goulot d'étranglement)
 - Closeness centrality (proximité, sera informé plus rapidement)
 - Flow betweenness (contribution à augmenter le flot)
- Basée sur les marches (on peut repasser par un nœud ou un lien plusieurs fois)
 - Bonacich power = eigenvector centrality (version itérative de la centralité de degré) (importance, haut-risque d'infection)
 - Katz, Hubbel (influence, « qui a entendu ce qu'il a dit ? »)

↙ PageRank
(être pointé par des sites importants rend important)

Liste de calculs disponibles dans l'outil UCINET



(2) Analyse de positions

- Trous structuraux
 - Mesure la redondance des ego-networks (les « triangles » incluant ego) (Burt, 1992)
 - Séparation entre deux contacts non redondants
 - Zones peu denses confèrent du pouvoir au centre
 - Proche de la betwenness et de la densité des ego-network
- Equivalence structurelle
 - Nœuds de même voisinage, même centralité...
- Equivalence Régulière
 - Les voisinages sont eux-même équivalents, assurent le même rôle
 - Comme deux médecins, deux chefs d'équipe.
- ...

(2) Analyse de positions

■ Et puis de nouvelles métriques inventées tous les jours

- Calcul d'influence dans twitter : RetweetRank
- Détection de spam : SpamRank
- Calculs de confiance : TrustRank, TidalTrust
- Calculs de similarités : SimRank
- ...



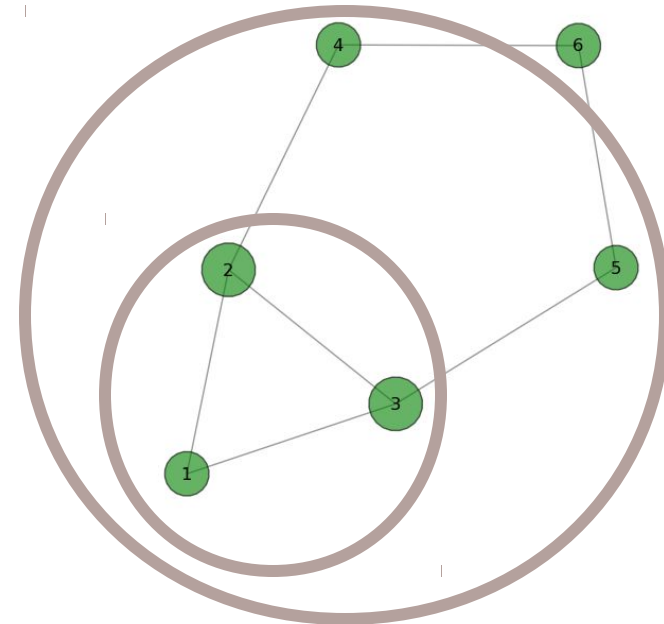
(3) Analyse de sous-groupes

- La délimitation en groupes n'est pas un *a priori*
 - On ne cherche pas qui appartient à une famille, communauté ethnique ...
 - Mettre en évidence une catégorie émergente, basée sur l'analyse de la topologie

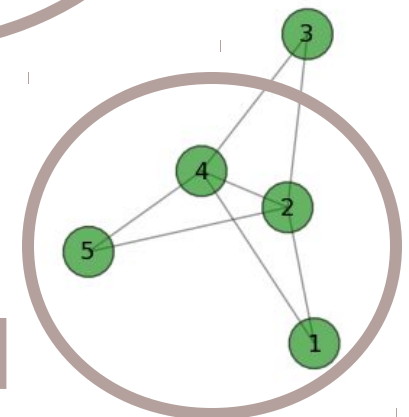
En pratique, existence de zones denses

(3) Analyse de sous-groupes

- Recherche de groupes
 - Cliques (ou 1-cliques), k-cliques : ensemble max de nœuds tous reliés à une distance k.
 - K-plex (ensemble de n nœuds où chacun a un lien avec au moins n-k nœuds)



Clique et 2-clique



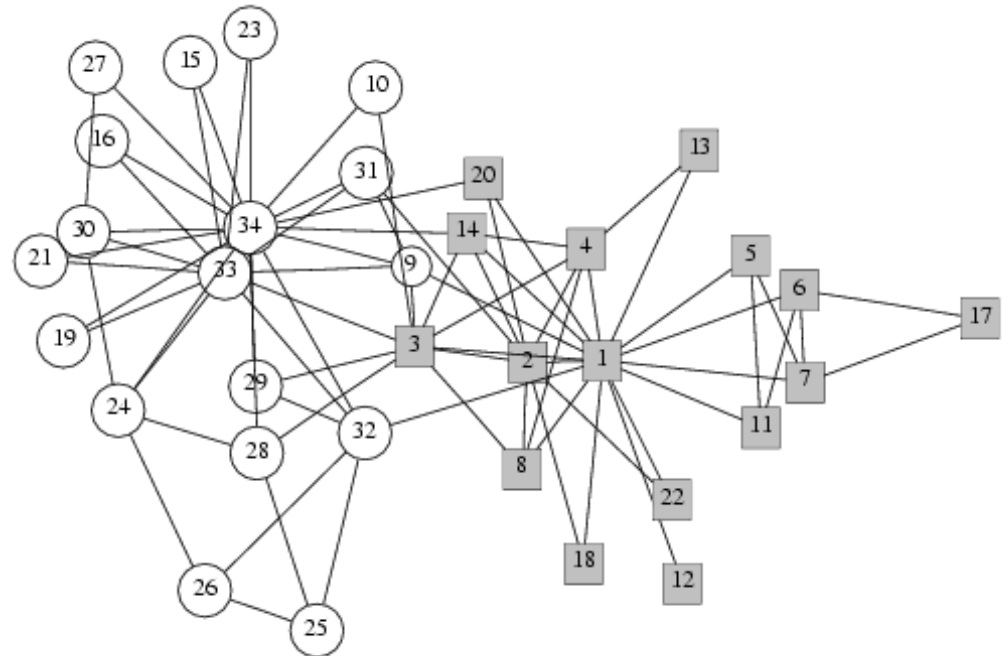
2-plex



(3) Analyse de sous-groupes

■ Détection de communautés

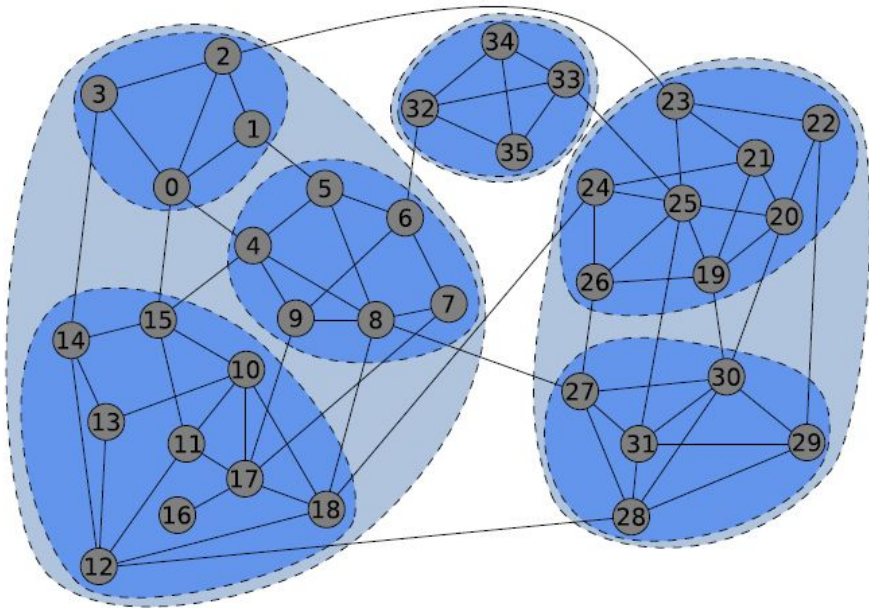
- Formes plus lâches
- Gros volumes de données
- Thématique connaît un renouveau avec l'étude de grands graphes réels



Karate Club in an American University [Zachary 1977]



(3) Analyse de sous-groupes



P. Pons, 2007

Trouver la partition
 $P = \{C_1, \dots, C_k\}$ qui optimise
la fonction de qualité $Q(P)$

■ Quatre types d'approches (thèse P. Pons, 2007)

- Classiques
- Séparatives
- Agglomératives
- Marches aléatoires



(3) Analyse de sous-groupes

■ Approches classiques

- Partitionnement de graphe en k groupes
 - Bisection spectrale (Spectral clustering)
 - théoriquement solide, basée sur le calcul de vecteur propre
 - Très coûteuse
 - Méthode de Kernighan et Lin
 - Bisection en utilisant la coupe minimisant le nombre d'arêtes tombant entre deux coupes
- Clustering hiérarchique
 - Mesure de similarité entre communauté (distance min, max ou moyenne, ou des membres au centroïd de l'autre communauté)
 - Partant avec n communautés, méthode itérative où l'on fusionne des communautés les plus proches
 - ou kmeans si on peut fixer k

(3) Analyse de sous-groupes

■ Approches séparatives

- On part d'une communauté unique puis à chaque étape, on retire une relation provoquant un découpage en sous-communautés
 - Relation la plus centrale par exemple (Girwan et Newman, 2002)
- Critère d'évaluation de la division en communautés
 - La modularité Q cherche à produire un découpage où le nombre d'arêtes reliant les groupes est plus petit que le nombre « attendu », i.e. si on avait découpé aléatoirement.



(3) Analyse de sous-groupes

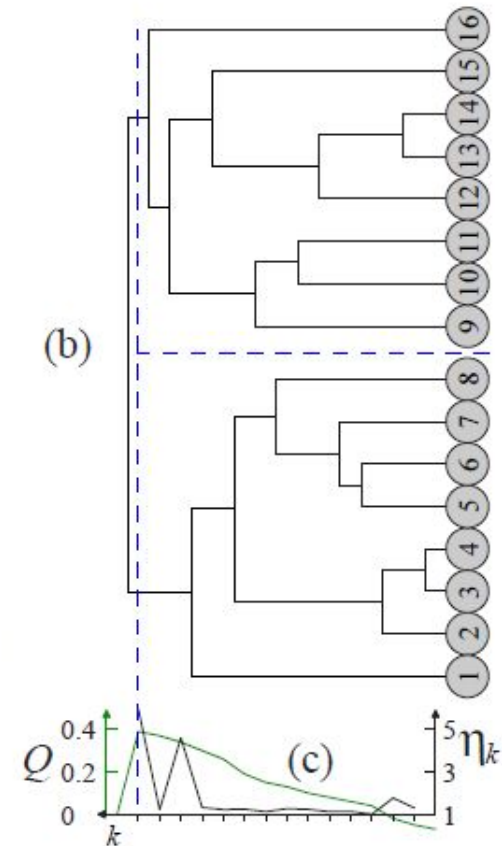
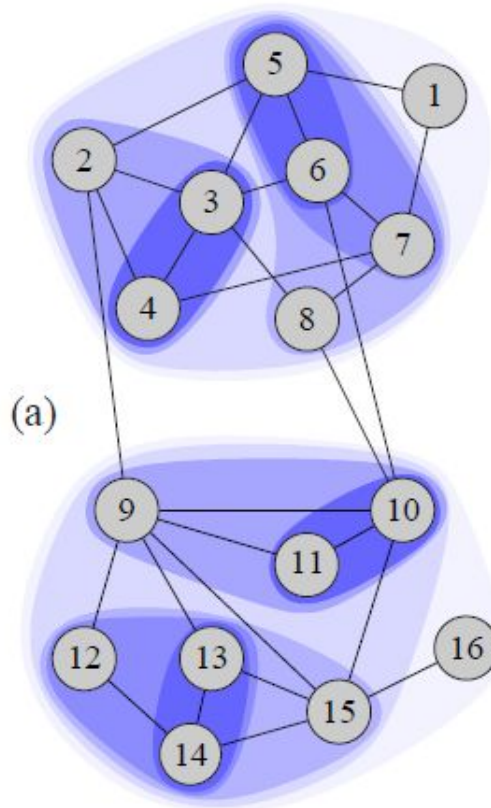
■ Approches agglomératives

- On part de n communautés composées d'un seul nœud
- Itérativement, on regroupe les nœuds en communautés, puis les communautés entre elles, jusqu'à obtenir une seule communauté.
- Critère d'évaluation de la division en communautés
 - La modularité Q cherche à produire un découpage où le nombre d'arêtes reliant les groupes est plus petit que le nombre « attendu », i.e. si on avait découpé aléatoirement.



(3) Analyse de sous-groupes

- Marches aléatoires
 - Intuition : les marcheurs vont se faire piéger dans des zones denses
 - Un marcheur va avoir tendance à rester dans sa communauté si marche courte
 - Distance : deux marcheurs proches vont voir le même voisinage
 - Clustering hiérarchique par fusions successives de communautés



Computing Communities in Large Networks Using Random Walks
Pascal Pons and Matthieu Latapy, Journal of Graph Algorithms and Applications
<http://jgaa.info/> vol. 10, no. 2, pp. 191–218 (2006)



3. Applications : compréhension des données

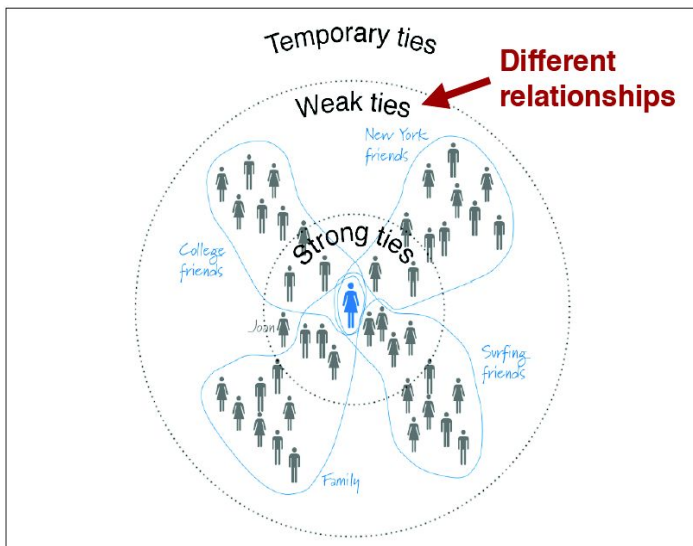
**La « force des liens faibles »
Identité numérique
eMarketing**



M. Granovetter

■ La « force des liens faibles », 1973

- Liens forts : rencontres fréquentes, échanges approfondis.
- Liens faibles : contacts brefs, occasionnels, soutien, mise en contact.
- La force d'un lien = combinaison de quantité de temps, d'intensité, d'intimité, de confiance mutuelle, et des services réciproques.





M. Granovetter, La « force des liens faibles »

■ Les liens forts ne sont jamais des « ponts »

- Ils ne permettent pas de relier entre eux des groupes d'individus disjoints.
- L'information reste à l'intérieur de cliques restreintes.

■ Les liens faibles permettent la circulation d'une information dans un réseau plus vaste

- De clique en clique.
- Ils procurent aux individus d'autres informations.
 - Ils évoluent dans des cercles différents et ont accès à une information différente de la nôtre.



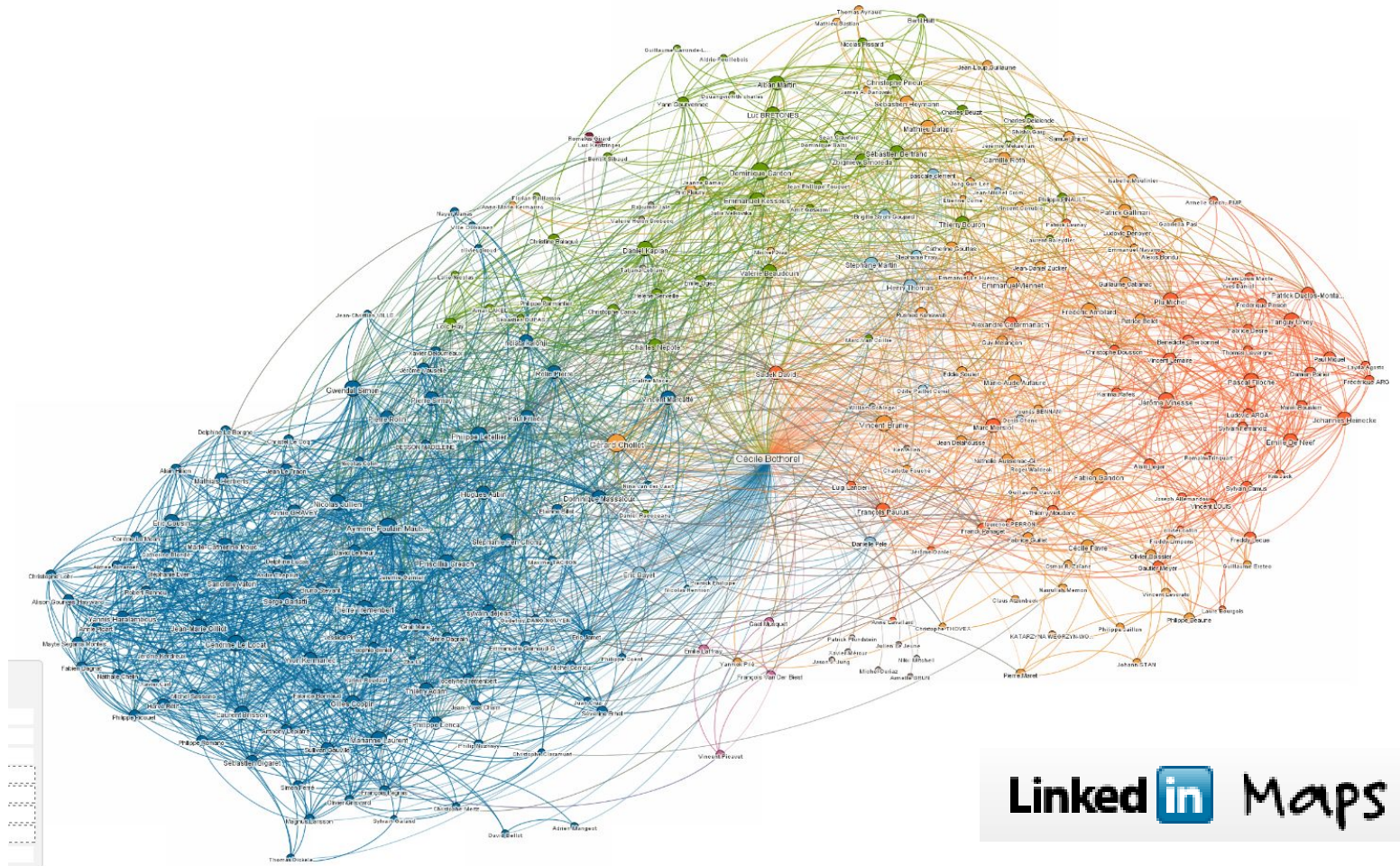
M. Granovetter, La « force des liens faibles »

■ Une application empirique de sa théorie à l'étude des processus de recherche d'emploi

- Un échantillon de 300 cadres, techniciens, gestionnaires ayant récemment changé d'emploi.
- Constats :
 - Ces salariés américains trouvent + souvent leur emploi par leurs relations personnelles que par n'importe quel autre moyen (56% des personnes dans cette enquête).
 - La fréquence de leurs contacts avec l'individu-source de l'information est faible (55,6% occasionnellement, 27,8% rarement / 16,7% souvent).



M. Granovetter, la « force des liens faibles »





M. Granovetter, la « force des liens faibles »

Your Network of Trusted Professionals



You are at the center of your network. Your connections can introduce you to 3,849,200+ professionals — here's how your network breaks down:

1 Your Connections Your trusted friends and colleagues	286
2 Two degrees away Friends of friends; each connected to one of your connections	52,600+
3 Three degrees away Reach these users through a friend and one of their friends	3,796,300+
Total users you can contact through an Introduction	3,849,200+

96,807 new people in your network since February 28

Your connections are in 44 locations but your network gives you access to **995 additional locations**, including:

- Brussels Area, Belgium
- United Kingdom
- Montreal, Canada Area

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!](#) **150,000,000+**





M. Granovetter

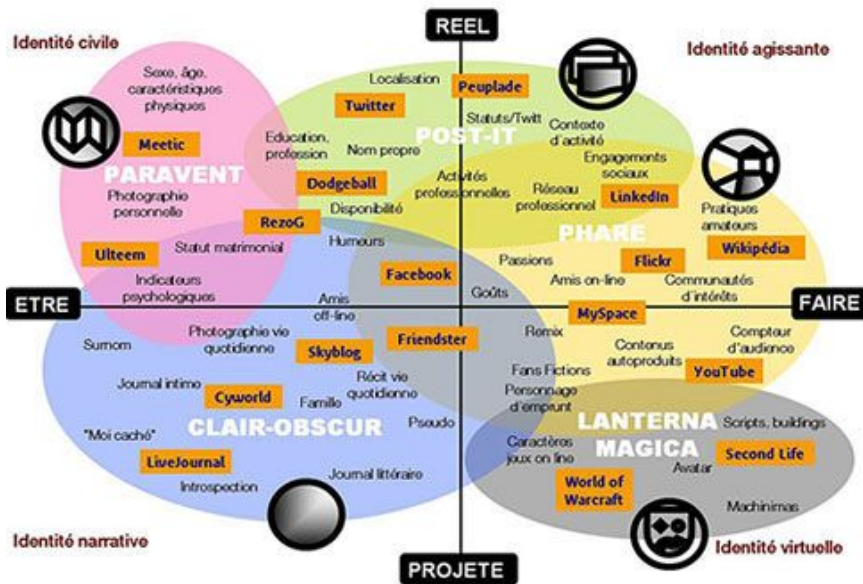
- Conclusion:
 - Pour trouver un emploi : les liens faibles sont donc plus importants
 - Ils sont + nombreux
 - Ils forment des « ponts locaux »
 - Ils créent des chemins + courts
 - Les réseaux sociaux sur Internet : des facteurs de multiplication des liens faibles, hypothèse d'un affaiblissement des hiérarchies sociales.



D. Cardon, les formats de visibilité de l'identité numérique

■ « Le design de la visibilité », 2008.

- Une cartographie structurée par deux axes:
 - Le 1er axe : les traits identitaires les + incorporés à la personne (être) 'vs' ceux qu'elle a extériorisé dans des activités, œuvres et relations (faire).
 - Le 2e axe : la distance entre identité numérique projetée 'vs' identité réelle.



[Source: D. Cardon, 2008, « Le design de la visibilité. Un essai de cartographie du web 2,0 », *Réseaux*, n°152, 93-137].



D. Cardon, la structuration de l'espace des identités numériques

■ 5 « formats de visibilité »

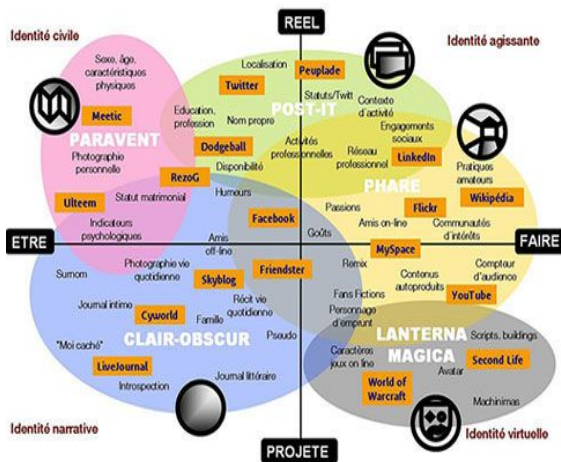
- « Coproductions des spécificités techniques des différentes plates-formes et des usages socialement contextualisés » :
 - Le « paravent » : format des sites de rencontre, les identités des individus restent cachées derrière les catégories qui les décrivent, ceux-ci ne se dévoilent qu'au cas par cas, aux personnes choisies.
 - Le « clair-obscur » : l'expression des identités intimes est beaucoup + poussée, mais seulement à destination de proches.
 - Le « phare » : la divulgation de soi à de grands réseaux relationnels.
 - Le post-it » avec le développement de Twitter et du micro-blogging.
 - La « lanterna magica » : la projection de soi dans des grands réseaux d'avatars.



D. Cardon

L'enjeu de la visibilité

Politique de visibilité spécifique à chaque plate-forme



- Diversité des formes de visibilité selon les plates-formes :

- se cacher, se voir (*le paravent*)
- se voir caché (*la lanterna magica*)
- le montrer caché (*le clair-obscur*)
- tout montrer tout voir (*le phare*)

- Des registres d'identité différents

les individus contrôlent selon les plates-formes « la distance à soi qu'ils exhibent à travers leur identité numérique »

- caractéristiques qui les décrivent le mieux dans leur vie réelle (haut) / (bas) dissimulations de traits identitaires et projection



Réseaux sociaux et eMarketing

- Evian : 55 millions de vues certifiées par Visible Measures
- T-mobile : pour le lancement du téléphone G1 (Google Phone). 13 500 personnes sont venus spontanément chanter « Hey Jude » pour un Karaoke géant
- Samsung : « Sheep LED Art ».
- Volkswagen : « FunTheory »
- ...
- Orangina : des faux comptes sur Facebook ?
- Mac Donald's « fail » en fev. 2012
- La Redoute : un modèle de gestion de crise digitale
- ...



« Fail » de La Redoute en janv. 2012, récupération par les 3 suisses

VISIBLEMENT, TOUT LE MONDE
NE SAIT PAS QUE NOUS AVONS
DES MAILLOTS DE BAIN
à partir de 9€99
3SUISSES.fr



Réseaux sociaux et eMarketing

- Bad buzz ou Stratégie bien ficelée ?
- Utiliser les médias sociaux est tentant mais
 - difficilement maîtrisable car fonctionne par bouche à oreille, le message peut être déformé
 - durée de vie limitée, un buzz en chasse un autre
- Les recettes pour faire du bu\$\$: un marché à part entière

GILBERT DELAHAYE - MARCEL MARLIER

martine apprend la recette du buzz





Réseaux sociaux et eMarketing

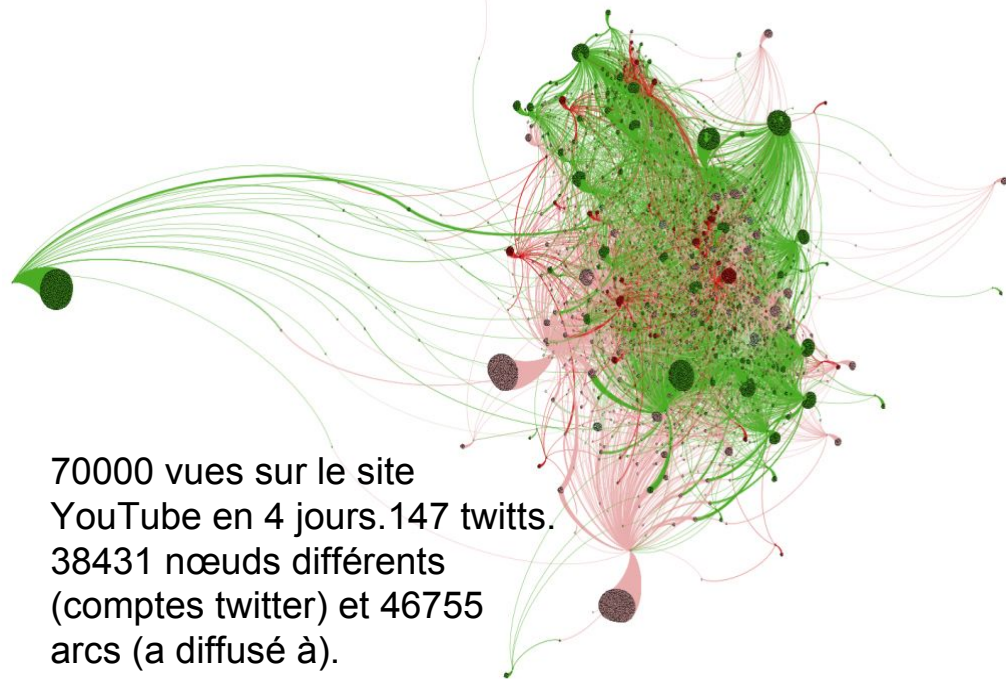
- Enjeu : Maîtriser le Bouche à oreille
- Comprendre les mécanismes de propagation et la structure des réseaux...
- ... pour mieux gérer les campagnes

**Comment choisir le blogueur qui pourra mieux parler d'un produit/service ?
Par quel utilisateur twitter se faire suivre ?
A quels clients proposer un produit en beta-test ?**



Réseaux sociaux et eMarketing

- **Comprendre le cycle de vie des buzz**
- **Analyse de données**
 - Collecte des messages et du réseau de diffusion
 - Étiquetage opinion
 - Identification des personnes clés
 - Analyser la dynamique



TelecomBretagne Télécom Bretagne
Découvrez l'École grâce à la plaquette des élèves <http://bit.ly/kai3dJ>
ou la vidéo des élèves <http://www.youtube.com/watch?v=0qYV-mrQzUA>
31 Mai



Réseaux sociaux et eMarketing

■ Déterminer les individus par qui la campagne doit démarrer/passer

- Les leaders d'opinions (calcul de popularité – centralité de degré, pageRank)
- Les médiateurs (calcul de centralité « stress », « betweenness ») : les acteurs qui contrôlent la propagation de l'information en tant que passages obligés
- Les points de référence (calcul de centralité « closeness ») : individus les plus proches de tous





Réseaux sociaux et eMarketing

- Dynamique des réseaux sociaux
 - Prédire la vitesse, la couverture de la diffusion
 - Calcul de plus courts chemins, de flots maximum
 - Social contagion, Simulation sociale
 - Une rumeur sera plus efficace si elle provient de plusieurs sources différentes

- Avec une analyse sémantique, on peut délimiter des communautés thématiques
 - Détecter les tendances
 - Cibler les campagnes
 - Les membres situés à la périphérie des communautés peuvent servir de passerelle inter-communautés



Conclusion

**Un nouveau concept, une vieille
histoire**

**Il reste des défis scientifiques et
techniques**



Conclusion

■ Outils : des challenges techniques

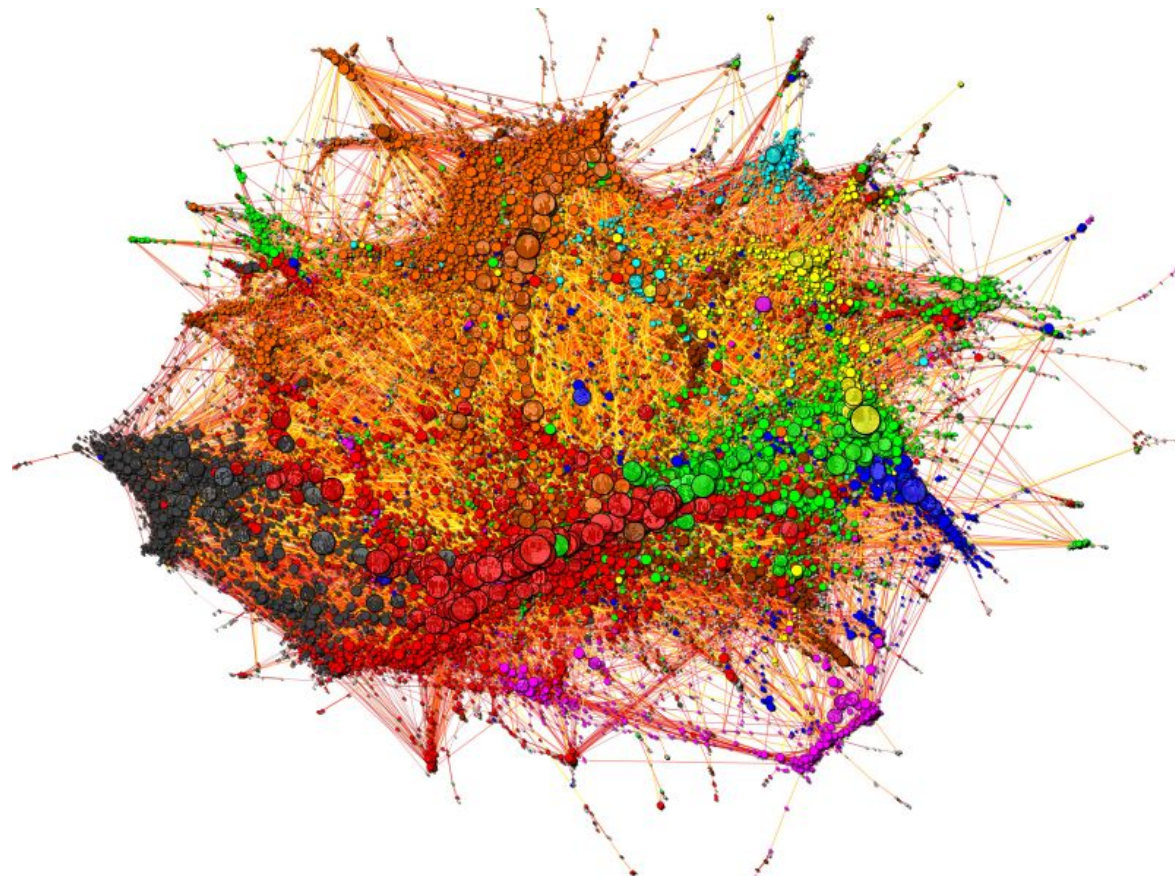
- Collecte données
 - Quelles données, biais de la collecte
 - Maintenance dans le temps de données fortement dynamique
 - Privacy : l'anonymisation n'est pas suffisante...
 - « Big Data » : volume au-delà du gérable avec les BDDs classiques
- Modélisation
 - Relations complexes : liens multiples, valuées, dynamiques
 - Données additionnelles : profil, opinion, émotion
- Analyse – métriques
 - Quelles mesures pour quelle analyse ?
 - Données incomplètes : comportements probablement influencés par « l'extérieur »
 - Valeur d'un fan Facebook, d'un RT Twitter ?
 - Evaluation des analyses : clustering, métriques
 - Evolution du graphe, prédiction de liens



Conclusion

■ Outils : des challenges techniques

- Visualisation
 - Volume
 - Nouvelles métriques
 - Interactivité



Artists in the last.fm database, by Tamas Nepusz

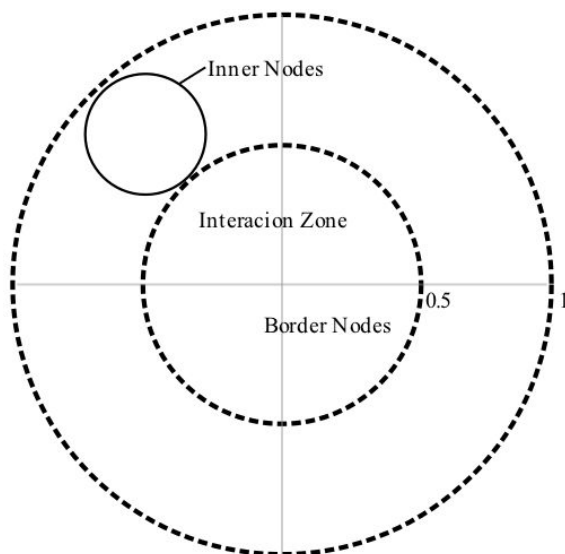


Conclusion

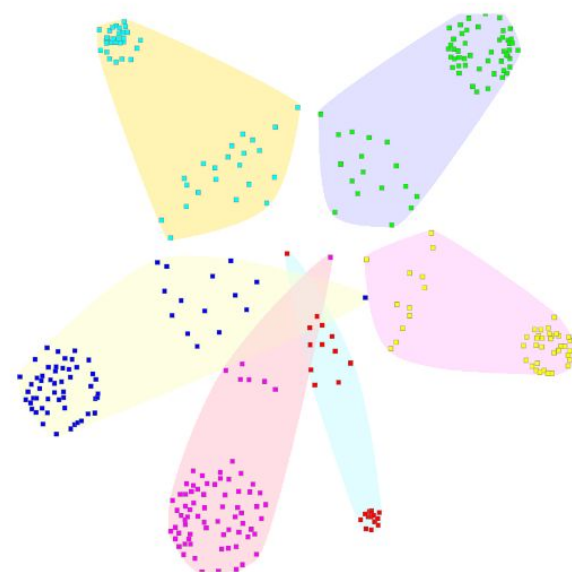
■ Outils : des challenges techniques

- Visualisation
 - Nouvelles métriques
 - Nouveaux layouts

Présentation centrée sur l'interaction entre communautés
Cruz Gomez, Bothorel, Poulet, 2012



(a) Placement of the elements of the layout model



(b) Location of the inner nodes according to the placed border nodes.



Conclusion

- **Des projets de recherche pluridisciplinaires.**

- **Des pistes de questions sociologiques à creuser :**
 - variations entre formes d'expression/exposition de soi et accroissement du capital social
 - l'émergence de réseaux ou expliciter comment un acteur choisit d'établir telle relation.
 - L'engagement ou l'implication dans un réseau.
 - la forme et le contenu des relations observées.
etc.