Réseaux sociaux et analyse de traces des forums d'une communauté d'apprentissage

Christophe Reffay

▶ **To cite this version:**

Christophe Reffay. Réseaux sociaux et analyse de traces des forums d'une communauté d'apprentissage. Symposium, formation et nouveaux instruments de communication, Jan 2005, Amiens, France. edutice-00000811

HAL Id: edutice-00000811
https://edutice.hal.science/edutice-00000811

Submitted on 12 Apr 2005

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

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

Résumé

Après une mise en contexte de nos travaux, proposant une classification des indicateurs quantitatifs de suivi des interactions humaines médiatisées par les LMS (Learning Management System), cette prépublication de chapitre aborde diverses notions proposées par les réseaux sociaux (Social Network Analysis), en vue de les intégrer dans un tableau de bord du tuteur pour le suivi de la formation à distance, basée sur la production collaborative. Quelques-uns de nos résultats antérieurs concernant la mesure et la représentation de la cohésion sont repris pour discuter de la validité, de l'expressivité et de la transférabilité des modèles issus des réseaux sociaux dans le contexte des communications médiatisées. Basée par essence sur l'analyse des relations, l'analyse structurale (autre nom des réseaux sociaux) dépend avant tout d'une bonne définition des relations étudiées. Elle passe par la construction d'un graphe des relations appelé sociogramme. Des algorithmes sont appliqués sur ces graphes pour construire des représentations synthétiques mettant en relief certaines propriétés du groupe ou la place singulière qu'y ont les individus. On peut alors proposer des images de la cohésion du groupe, de l'intermédiarité ou de la centralité d'un individu pour le situer dans la structure de communication du groupe afin d'en déduire son rôle, son pouvoir ou son implication.

Mots - clés : Formation à Distance, Analyse structurale des interactions, Réseaux Sociaux.

Contexte et objectif

Avant de décrire les méthodes d'analyse utilisées sur les forums de communautés d'apprentissage, il convient de préciser le type de forum et le type de communautés d'apprentissage auxquels elles s'adressent. On trouve des études de métrologie sur de grands groupes tels que les forums de usenet [LANCIERI 05] où les internautes vont et viennent par milliers, chaque forum recelant en moyenne 200 messages postés par jour mais pouvant aller jusqu'à 8200 par jour ! On y découvre par exemple que la longueur moyenne d'un message dans un forum est inversement proportionnelle au nombre de contributeurs dans ce forum. On peut aussi s'intéresser à la profondeur des fils de discussion ou plus globalement aux usages d'un tel média. Si ces études nous intéressent pour nourrir l'ingénierie pédagogique des systèmes de formation en ligne, elles ne peuvent s'appliquer au suivi des groupes plus restreints dans le cadre spécifique que nous ciblons. Dans nos dispositifs, un groupe est composé de 4 à 10 apprenants environs, suivi par un tuteur, secondé par des experts de contenu dans certains cas. Nous sommes bien loin des milliers d'internautes présentés par L. Lancieri dans les forums usenet. Il est très important de souligner également que nos apprenants sont liés par un contrat pédagogique dès lors qu'ils se sont inscrits à la formation. Le modèle extensif et socio-constructiviste et une durée (minimale) de deux mois caractérisent le type de formations visées et confèrent à la dimension sociale une importance capitale. La qualité des relations interindividuelles entre les pairs et avec le tuteur constitue le ferment des possibles dans une pédagogie interactionniste, où apprendre, c'est d'abord écouter, confronter, négocier pour construire la connaissance. On comprend alors que le tuteur en charge du suivi de ce type de formations, soit intéressé par des instruments de mesure permettant de jauger l'engagement de chaque individu dans le groupe et plus globalement la dynamique du groupe.

En novembre 2000, quand le projet ICOGAD [CHANIER 02] a été lancé, on se demandait alors si cela avait un sens de parler de «groupe» ou, nous dirions aujourd'hui «communauté d'apprentissage» en ligne. Très vite, la question essentielle devint de savoir comment

reconnaître si des personnes, engagées dans une formation collaborative, constituaient un groupe ou non ? Y-a-t-il une mesure objective permettant de caractériser la cohésion ? Peut-on évaluer la dynamique du groupe ?

L'idée de base était de tirer partie de l'énorme quantité de données brutes stockées sur le LMS (Learning Management System) pour tenter de renseigner efficacement le tuteur sur l'intensité de l'activité, la vivacité des échanges, la position des différents interactants, la dynamique du groupe, … toutes notions qu'un enseignant rompu aux techniques collaboratives en face-à-face évalue intuitivement à partir de la perception des informations visuelles (présence, absence, mouvements, gestes, postures, regards, etc.) ou auditives (volume sonore, dialogues, bruits dus à certains déplacements, etc.) sans nécessairement entrer dans la sémantique du discours des acteurs. Il fallait donc trouver un modèle permettant de formaliser cette évaluation pour qu'elle puisse opérer sur les traces accessibles du LMS.

D'un point de vue méthodologique, nous avons souhaité réunir un corpus à partir des données disponibles sur un LMS déjà largement utilisée plutôt que de développer un prototype qui fournirait les traces ad-hoc. Un tel prototype permettrait probablement de mieux ajuster le type et le stockage des données pour définir les indicateurs souhaités, mais limiterait l'usage de ces indicateurs à ce seul prototype. Il nous avait semblé raisonnable d'ausculter les données existantes dans un LMS démocratisé et de chercher à en tirer le meilleur profit informationnel sur les aspects particuliers de la dynamique de groupe.

Nous avons rapidement distingué plusieurs indicateurs que nous pouvons positionner sur un quart de plan dont les deux dimensions représentent la pertinence (en abscisse) et le coût d'extraction (en ordonnée) (figure 1). La pertinence doit être ici comprise comme l'intérêt qu'une donnée quantitative peut représenter aux yeux du tuteur d'abord d'un point de vue pédagogique puis d'un point de vue social. Dans un contexte d'une formation collaborative, ce qui est attendu des apprenants par le tuteur, c'est de participer de façon constructive aux échanges, ce qui implique en particulier de lire ce qu'écrivent les autres et de répondre à leurs demandes, leurs questions ou de commenter leurs productions. L'objectif de ces travaux est donc de chercher les meilleurs modèles pour caractériser ces formes d'engagement.

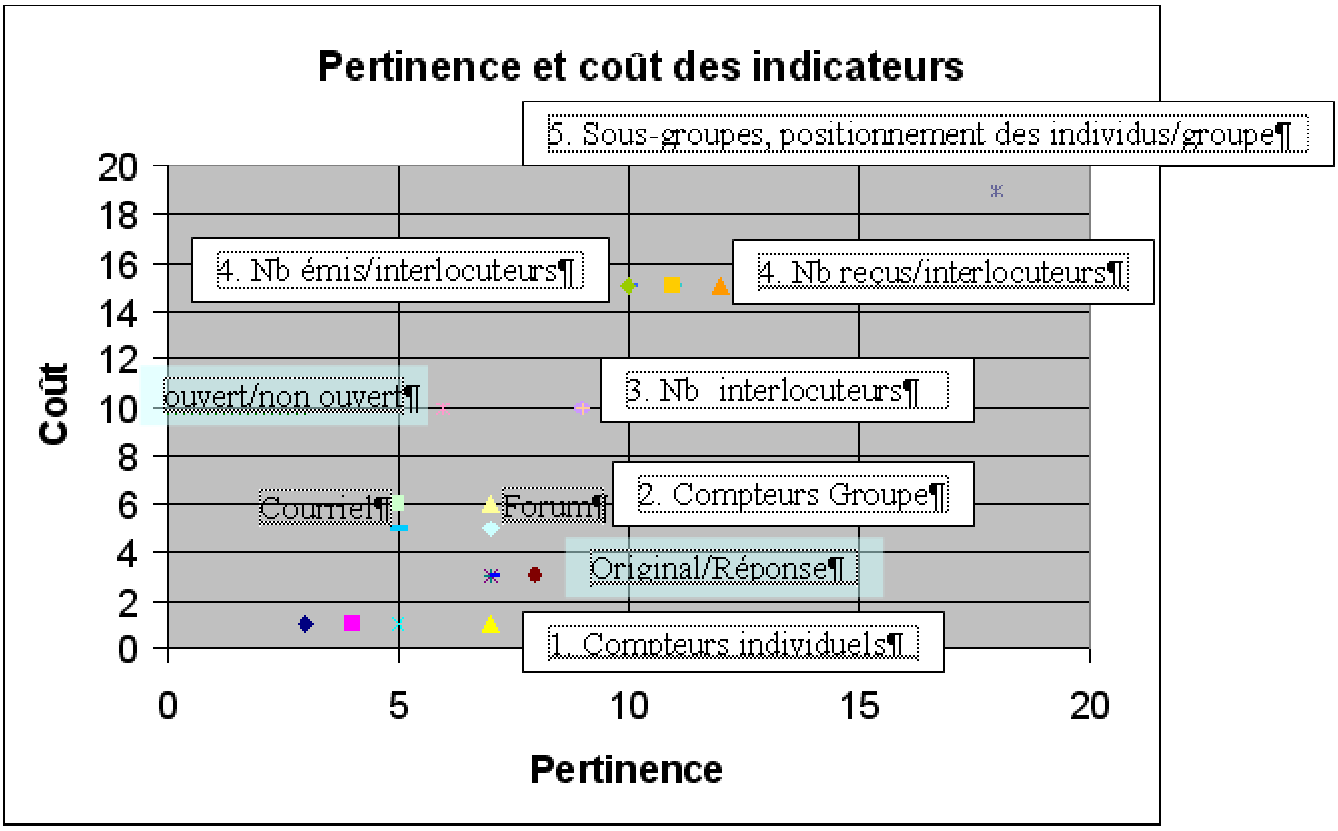

Figure 1 : Pertinence et coût des indicateurs

Le positionnement des indicateurs sur l'axe des coûts nous suggère de les regrouper par classes.

La première classe regroupe l'ensemble des indicateurs déjà disponibles sur le LMS et qui sont donc immédiatement lisibles (sans effort). Elle regroupe les compteurs individuels quantifiant, pour un acteur le nombre de connexions, le nombre de pages demandées, le nombre de messages postés dans le forum, le nombre de messages postés et reçus par courriel.

Dans certains LMS comme WebCT, on trouve des compteurs distinguant les messages originaux (initiant un fil de discussion dans un forum ou un échange par courriel) des messages postés en réponse à d'autres.

A partir de cette première classe, on peut assez simplement construire dans une deuxième classe les indicateurs représentant le nombre de messages émis dans le courriel ou postés dans le forum pour l'ensemble du groupe ou en moyenne par membre d'un groupe. On distingue en général deux niveaux de regroupement : les sous-groupes ou équipes (projets) et le groupe classe. Les acteurs sont distingués par rôle : tuteur, (assistant), apprenant.

A chaque message est associée une information permettant de préciser si le message a été lu ou non par l'utilisateur (message courriel ou forum). Elle est utilisée par la LMS pour indiquer à l'utilisateur si un message a été déjà ouvert ou non, mais rarement pour recenser les personnes ayant ouvert un message posté dans le forum. Cet indicateur est généralement jugé assez coûteux et relativement peu pertinent (car « ouvert » ne signifie pas « lu »). Nous y attachons cependant une certaine importance car un message « non ouvert » implique qu'il est « non lu ». Cette information nous permet ainsi de distinguer la « participation » (un message

émis témoigne d'une intention de participation) de la « communication » et donc d'éliminer des « communications » les messages « non ouverts » par les destinataires.

Dans les classes précédentes, on a regardé globalement ce que chaque acteur poste vers ou reçoit du reste du groupe globalement sans distinguer ses interlocuteurs. L'analyse des réseaux sociaux nous a amené à étudier ces interconnexions de façon plus fine. Nous plaçons donc dans la troisième classe ce que la théorie des graphes appelle les degrés intérieur et extérieur d'un nœud du réseau : c'est à dire respectivement le nombre de personnes ayant émis des messages vers cet acteur et le nombre de personnes ayant reçu au moins un message émis par cet acteur. Ces différentes personnes constituent le réseau d'égo. Plus simplement, il s'agit du nombre d'interlocuteurs de cet acteur. L'analyse des réseaux sociaux montre qu'une personne capable de maintenir un vaste réseau est un acteur important dans la communauté. Cet indicateur (degré intérieur ou extérieur) permet de mesurer cette importance. Il est plus difficile à construire car il compte le nombre de personnes (émettrices ou réceptrices) et non le nombre de messages. Pour un même nombre de messages échangés globalement, il permet de distinguer un couple de bavards (deux personnes échangeant de nombreux messages exclusivement entre elles) d'un véritable animateur du groupe (échangeant quelques messages avec beaucoup de personnes).

La quatrième classe d'indicateurs intègre 3 notions : le nombre de messages, le fait qu'ils soient émis ou reçus et répartit ces valeurs pour chacun des interlocuteurs d'un acteur : ce sont des graphes orientés et valués. Ils expriment donc à la fois la direction et l'intensité du flux des messages en précisant les sources et les destinataires. Ce type de graphe est sans doute applicable à différents média de communication : forum, courriel, bavardage, audio ou vidéo-conférence,… L'intensité d'une relation peut utiliser des unités variées : message, acte de langage, mot, caractère, tour de parole, temps de communication, etc. Enfin, des critères de sélections peuvent restreindre l'ensemble des unités prises en compte : période (communication ayant été émise entre deux dates précises), ouvert (seuls les unités contenues dans un message ayant été ouvert sont comptabilisée), etc.

Enfin, nous plaçons dans la cinquième classe les indicateurs qui permettent de préciser (à partir des indicateurs de la classe 4), de façon plus synthétique et au niveau du groupe, la position relative de chacun des individus dans le groupe, compte tenu des rôles qu'ils tiennent dans la communication du groupe.

C'est de cette cinquième catégorie que relèvent la cohésion du groupe, la centralité et l'intermédiarité d'un individu dans le groupe : trois notions essentielles que nous retenons de la littérature dans le champ des réseaux sociaux [DEGENNE&FORSE 94, WASSERMAN&FAUST 94, SCOTT 2000]. Nous présentons rapidement ces notions telles qu'elles sont définies dans la littérature : c'est à dire en considérant des relations en face-à-face. Nous discuterons ensuite des conditions de transfert de ces notions lorsqu'il s'agit de relations à distance, passant par le réseau et dans un contexte d'apprentissage. Enfin, nous rapporterons les premiers résultats de nos recherches pour représenter la cohésion d'un groupe de formation collaborative à distance. A la lumière de ces premières tentatives, nous proposerons quelques recommandations dans l'usage des nombreux outils que l'on peut trouver en particulier sur le site du réseau international pour les réseaux sociaux [INSNA 05] et qui permettent de calculer des indices ou des représentations issues des réseaux sociaux.

Trois notions extraites des réseaux sociaux

Issu de la dynamique des groupes, de la sociologie et de la théorie des graphes, les réseaux sociaux proposent des outils et des mesures pour analyser la structure d'un groupe en fonction des relations que les membres entretiennent. Cette analyse structurale fournit des indicateurs permettant de faire ressortir certaines propriétés du groupe ou de caractériser la place singulière qu'occupe un individu au sein du groupe. Trois de ces notions sont présentées ici dans leur contexte originel : la cohésion, la centralité et l'intermédiarité.

Cohésion

La cohésion peut être vue comme une force d'attraction entre les individus, une interdépendance positive et nécessaire entre les membres d'un groupe pour que celui-ci conserve sa raison d'être, et atteigne les objectifs fixés. Si cette première description donne une idée de l'intérêt et de l'importance de la cohésion dans un groupe, elle ne donne pas de moyen de la mesurer. Cette notion se précise dans [WASSERMAN&FAUST 94] où on peut lire qu'un sous-groupe cohésif est un sous-ensemble d'acteurs entre lesquels les relations sont plus fortes, fréquentes, directes ou intenses, que celles qui existent entre ces acteurs et les autres. Selon les relations étudiées (affinité, services de voisinage, collaborations, parenté, etc.) on peut valuer la force d'un lien par la fréquence (ex : services de voisinage, collaborations), l'intensité (affinités) ou le fait d'être direct (parenté). Dès lors, la théorie des graphes permet d'identifier différents sous-ensembles caractéristiques :

- des sous-graphes complets appelés « cliques » où chaque nœud est directement lié à chacun des autres,
- des « n-cliques » où chaque nœud peut atteindre tous les autres par un chemin de longueur<=n,
- des « cliques de niveau s » où tous les nœuds sont directement liés par des relations d'intensité>=s,
- des « composantes fortement connexes » où chaque nœud appartient à un circuit maximal,
- des « clusters hiérarchiques » construits du centre vers l'extérieur par ordre décroissant de l'intensité des liens, où le nouveau cluster intègre le nœud lié avec la plus grande intensité à tous les nœuds du cluster précédent.

Chacun de ces sous-ensembles caractéristiques ne pourra ensuite être interprété qu'en fonction de la nature des liens composant le graphe initial, mais aussi en fonction des choix de certaines transformations utilisées dans les calculs intermédiaires (symétrisation de la matrice des relations par une fonction : min, max, somme ou moyenne par exemple).

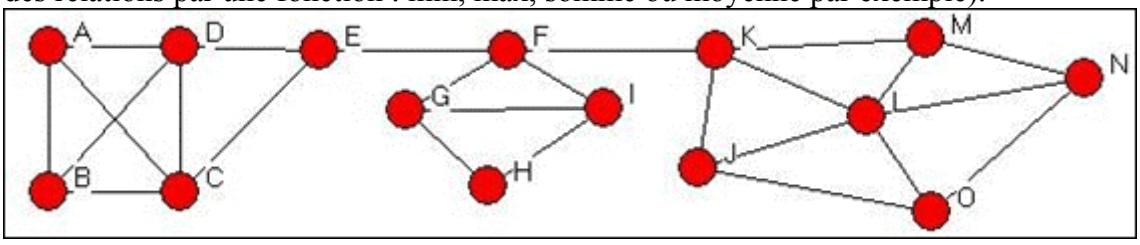

Figure 2 : (ABCD) est une clique, (ABCDE), (FGHI) et (KLMNO) sont des 2-cliques

Les liens de la figure 2 n'étant pas orientés, tous les nœuds font partie d'une seule composante fortement connexe. Et puisque ces liens ne sont pas valués (par une intensité), on ne peut parler ni de cliques de niveau s ni de clusters hiérarchiques, mais des exemples de tels sous-ensembles caractéristiques seront données plus loin pour les forums de discussion.

Centralité, intermédiarité, influence et pouvoir

Il est indubitable que les individus les plus centraux dans un graphe occupent des relations privilégiées dans les échanges, notamment par rapport à ceux qui sont rejetés plus à la périphérie. De nombreuses études rapportées dans [DEGENNE&FORSE 94] établissent que la centralité est liée au pouvoir, à l'influence ou au leardership, même si cette relation n'est ni simple ni univoque. Il existe différents points de vue qui ont conduit Linton Freeman (1979) à proposer trois définitions pour la centralité :

La centralité de degré (la plus simple) considère le degré (pour un graphe non orienté, le demi-degré intérieur ou extérieur pour un graphe orienté) d'un nœud : c'est-à-dire le nombre de connexions directes irriguant ce nœud. Un individu est d'autant plus central qu'il est directement lié à un grand nombre d'individus. La centralité de degré mesure (localement) la capacité d'un individu à communiquer, indépendamment de la centralité des individus auxquels il est directement lié. Or, intuitivement, un individu lié à n individus périphériques n'est pas aussi central que celui qui est lié à n individus, eux-même centraux. Un individu à la périphérie dépend des intermédiaires (plus centraux) pour sa communication avec les autres. Il ne contrôle pas sa communication.

La centralité de proximité donne un point de vue plus global à la centralité puisqu'elle considère la proximité d'un individu avec tous les autres. Un indice de Centralité normé de proximité pour le nœud i d'un graphe comportant n sommets, est donné par la formule :

$$C_{NPi}=\frac{n-1}{\sum_{j=1}^{n} d_{ij}}$$ avec d_{ij}=distance du plus petit chemin allant du nœud i au nœud j.

A l'aide de cette définition de la centralité, *"on a montré que plus un individu est proche des autres, plus il est susceptible d'avoir d'informations (Leavitt, 1951), de pouvoir (Coleman, 1973), de prestige (Burt, 1982), d'influence (Bavelas, 1950 ; Friedkin, 1991), et finalement d'accéder à un plus haut statut social (Katz, 1953)."*
[DEGENNE & FORSE 94, p. 157]

La centralité d'intermédiarité proposée par Freeman en 1966 défend l'idée qu'un individu peut bien être faiblement connecté aux autres et même relativement éloigné, mais servir d'intermédiaire dans bon nombre des échanges entre les autres membres du groupe. Plus il sert ou peut servir d'intermédiaire pour tous les membres, plus il est en position de contrôler la communication ou d'être indépendant des autres pour communiquer. Un tel individu peut influencer le groupe plus facilement en filtrant ou distordant les informations qui y circulent. Sa position lui permet également d'assurer la coordination du groupe. Soit n le nombre de sommets d'un graphe, g_{jk} le nombre de chemins géodésiques reliant le nœud j au nœud k, et $g_{jk}(i)$ le nombre de ces chemins passant par le nœud i, on définit C_{AIi} l'indice de centralité absolu d'intermédiarité du sommet i par :

$$C_{APi}=\sum_{j}^{n}\sum_{k=1}^{n}\frac{g_{jk}(i)}{g_{jk}}$$ avec : $\mathbf{j \neq k \neq i, et\ j < k.}$

L'indice de centralité normé d'intermédiarité qui en découle pour le sommet i est donné par :

$$C_{NIi}=\frac{2C_{AIi}}{n^2-3n+2}$$

A ce stade, les réseaux sociaux proposent 3 mesures de la centralité. La centralité de degré correspond à la capacité d'un individu à développer des communications au sein d'un réseau ; elle s'exprime par le nombre ou la proportion de liens avec son environnement local. Les

centralités selon la proximité ou l'intermédiarité mesure la capacité d'un individu à contrôler cette communication qui ne dépend pas forcément du nombre de ses liens avec ses voisins, mais de son rapport à l'ensemble des membres du réseau : rapport de proximité ou d'intermédiarité. On peut caractériser une mesure globale (pour un graphe dans son ensemble) de centralisation d'un graphe pour chacune de ces trois définitions de la centralité. Une forte centralisation de connexion est l'indice d'une communication active tandis qu'une forte centralisation de proximité ou d'intermédiarité traduit le fait qu'un petit nombre d'acteurs contrôlent cette communication.
Mais dans les trois définitions de la centralité données ci-avant, le graphe était supposé non valué. On ne tient pas compte du volume des échanges. C'est pour cette raison que Linton Freeman, Stephen Borgatti et Douglas White proposent en 1991 une généralisation de la centralité d'intermédiarité aux graphes valués :
l'intermédiarité de flot. Soit f_{jk} le flot maximum entre les sommets j et k (défini dans [FORD & FULKERSON 56]) et $f_{jk}(i)$ celui qui passe par i, le degré (absolu) auquel le flot maximum entre toute paire de sommets (j,k) dépend du sommet i considéré comme intermédiaire, appelé C_{AFi} est donné par :

$$C_{AFi}=\sum_{j}^{n}\sum_{k}^{n}f_{jk}(i)$$, avec : $\mathbf{j \neq k \neq i, et\ j < k.}$

L'indice normé d'intermédiarité de flot pour le sommet i noté C_{NFi} est donné par :

$$C_{NFi}=\frac{\sum_{j}^{n}\sum_{k}^{n}f_{jk}(i)}{\sum_{j}^{n}\sum_{k}^{n}f_{jk}}$$, avec : $\mathbf{j \neq k \neq i, et\ j < k.}$

Les quatre définitions de centralité proposées ci-dessus ne tiennent pas compte de la valeur de la centralité des sommets auquel i est lié. Or, ce n'est peut-être pas la même chose d'être le centre d'un groupe d'individus périphériques que d'être celui d'un groupe de membres eux-mêmes centraux. Bonacich propose en 1972 de calculer la solution matricielle $\lambda c=Rc$, où c est le vecteur des scores de centralité de chaque sommet, R la matrice des relations (valuées) entre les sommets et λ une constante de proportionnalité. De cette équation, on déduit que λ est une valeur propre de R et c son vecteur propre associé. Ce vecteur propre, normé par la plus grande valeur propre peut, selon Bonacich, servir de mesure pour le pouvoir de chaque acteur. Mais, comme le rapportent Degenne et Forsé, cette dernière mesure de la centralité est basée sur l'hypothèse qu'un acteur en relation avec des personnes centrales (qui peuvent contrôler leurs communications) voit sa centralité renforcée. Or, les sociologues ne sont pas tous d'accords sur le fait par exemple que le pouvoir des relations d'ego renforce le pouvoir d'ego. Dans certaines négociations, il peut être plus avantageux d'être connecté à des acteurs peu influents qu'à des acteurs puissants. De même, la théorie des coalitions de Caplow (1968) amène à penser que le pouvoir des autres réduit le pouvoir d'ego, bien plus qu'il ne l'augmente.
Comment reprendre ces outils et transférer ces résultats dans la communication médiatisée en contexte éducatif ?
Dans les travaux à l'origine des réseaux sociaux, d'où l'on puise les modèles de mesure de la cohésion et des différents types de centralité, les liens considérés supposent le plus souvent

une co-présence des individus, ce qui n'est évidemment plus le cas dans les communications médiatisées par les réseaux.
Si l'on reprend la transitivité dans les liens affinitaires qui s'exprime très simplement et assez justement par :
« Les amis de mes amis sont mes amis »,
on peut accepter que cette transitivité est due à la co-présence nécessaire des relations affinitaires. En effet, si A et C ont un ami commun B, A et C peuvent se trouver fréquemment chez B, ils ont donc de bonnes chances de s'y rencontrer. De plus, si A et C entretiennent des relations d'amitié avec B, c'est peut-être qu'ils partagent des centres d'intérêt, mais aussi des positions sociales et des comportements proches ou « compatibles ». C'est au moins en partie cet ensemble d'interactions directes et incarnées qui font que A et C ont de bonnes chances de devenir amis : ils sont assez probablement « compatibles » et ont de bonnes chances de se rencontrer. Bien sûr tout ceci ne tient que si les relations considérées sont plus basées sur la ressemblance que sur la complémentarité.
Si l'on cherche à comparer les réseaux d'acteurs communicant en face-à-face à ceux que nous connaissons aujourd'hui en ligne, nous venons de souligner la co-présence et la distance, mais nous pouvons aussi relever le fait que les échanges par forum (ou liste de diffusion ou courriel) sont presque exclusivement écrits asynchrones et pérennes et consultables à volonté tandis qu'en face-à-face, synchrone par définition, l'oral prévaut très largement et ne laisse pas systématiquement de traces pérennes. Les échanges en face-à-face ne peuvent profiter qu'à ceux qui étaient au rendez-vous tandis que les communications de groupe médiatisées et écrites peuvent être lues par toute personne ayant accès au médium, et à tout moment. En ligne, même si le message (de groupe) écrit est plus particulièrement destiné à une personne (en réponse à sa question par exemple), toutes autre personne accède dans les même condition que le destinataire, à la même information.
Mais si l'on s'intéresse plus particulièrement au cas des forums de discussion (ou listes de diffusion), le fait le plus marquant est probablement que toute personne a accès à toute information. Il ne peut y avoir rétention ou distorsion de l'information par un membre influent ou central comme c'était le cas dans les réseaux sociaux étudiés par Linton Freeman ou Ronald Burt pour ne citer que des exemples. Le calcul de la centralité de proximité ou d'intermédiarité pour mesurer le pouvoir, l'influence ou le degré de contrôle des communications n'a plus de sens dans le contexte des échanges par forums ou listes de diffusion. Cependant, dans le contexte éducatif (comme c'est probablement le cas généralement), le fait d'avoir accès à l'information ne suffit pas : il faut que l'apprenant fasse la démarche et y accède effectivement ! On peut donc reprendre les propositions Linton Freeman définissant la centralité et envisager, dans l'analyse des forums de discussion en contexte éducatif, d'en interpréter les mesures comme des représentations de *l'engagement* dans la tâche ou dans la collaboration plutôt qu'une mesure du pouvoir ou de l'influence. Mais pour ce faire, il ne suffit plus de comptabiliser ce qui a été posté et par qui, il faut aussi et surtout répertorier qui lit quels messages.
Quel graphe permet de modéliser un forum ?
Par la suite, selon l'objet de l'étude, on peut envisager différentes façons de modéliser les échanges par forum. On se limitera à la proposition de deux modèles que nous illustreront par la suite. Notons que les graphes qui suivent ont été produits avec GraphViz [GraphViz 05] issu de [GANSNER & NORTH 00] :

1. Un graphe orienté et valué dont les sommets sont les membres du groupe, et chaque arc (i,j) est valué par le nombre d'unités (messages, actes de langage, mots, caractères, …) postées par i et ouvertes par j.
2.

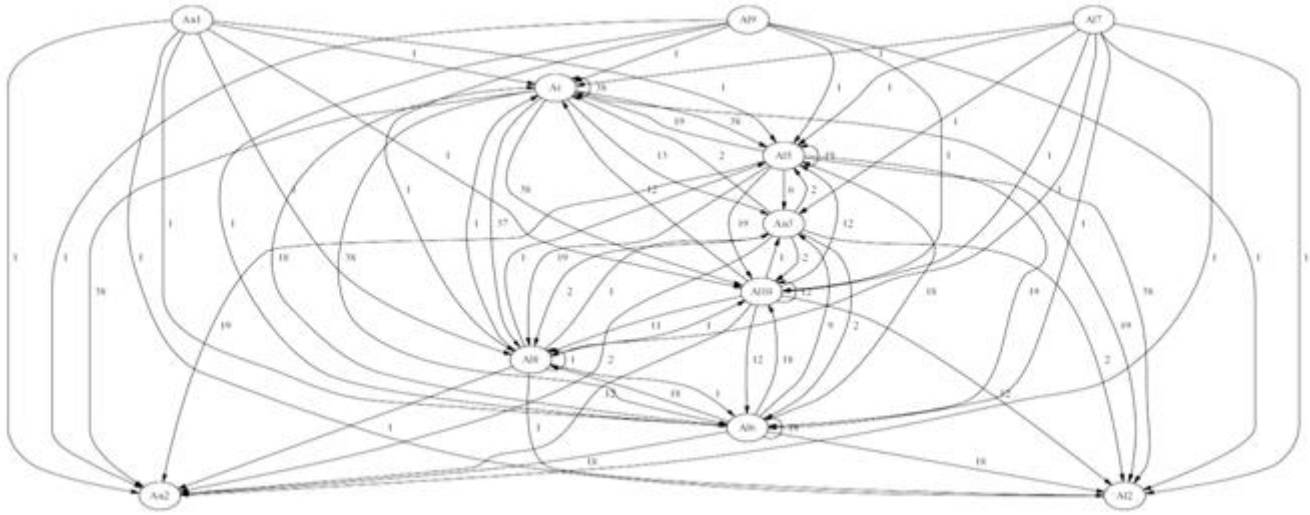

Figure 3 : Représentation graphique d'un graphe orienté et valué : nombre de messages échangés dans le forum de socialisation au cours de la première semaine de la formation « Simuligne ».

	At	Al1	Al2	Al3	Al4	Al5	Al6	Al7	Al8	Al9	Al10	An1	An2	An3
At	38		38			38	38		37		38		38	13
Al1														
Al2														
Al3														
Al4														
Al5	19		19			19	19		19		19		19	6
Al6	18		18			18	18		18		18		18	9
Al7	1		1			1	1		1		1		1	1
Al8	1		1			1	1		1		1		1	1
Al9	1		1			1	1		1		1		1	
Al10	12		12			12	12		11		12		12	1
An1	1		1			1	1		1		1		1	
An2														
An3	2		2			2	2		2		2		2	

Figure 4 : Représentation matricielle du graphe de Type 1 de la figure 2.

Sur le tableau donné à la figure 3, la valeur 13 située sur la première ligne (AT) et sur la dernière colonne (An3) signifie que 13 des messages postés par le tuteur (At) ont été lus par le natif (An3).

2. Deux graphes représentant respectivement le postage et la lecture des messages par les acteurs pour s'intéresser aux relations que les individus entretiennent avec des classes de messages (définies par types, par thèmes, par activités, par périodes ou par fils de discussion ou par profondeur dans le fil par exemple) plutôt qu'aux relations interindividuelles.

o Un graphe des postages dont les sommets sont divisés en 2 catégories : les acteurs A_i et les classes de messages C_j et les arcs orientés (A_i, C_j) sont valués par le nombre d'unités de la classe Cj qui ont été postées par l'acteur A_i.

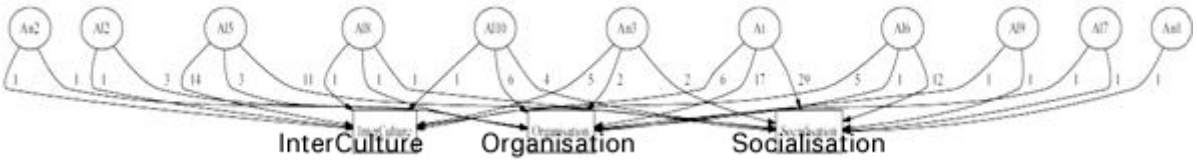

Figure 5 : représentation graphique du nombre de messages postés par chaque acteurs dans les différents forums au cours de la première semaine de la formation « Simuligne » dans le groupe Aquitania.

Auteur	Organisation	InterCulture	Socialisation
Al10	6	1	4
Al2	3	1	
Al5	3	14	11
Al6	1	5	12
Al7	1		1
Al8	1	1	1
Al9	1		1
An1			1
An2	1	1	
An3	2	5	2
At	17	6	29

Figure 6 : représentation matricielle des contributions des acteurs dans les différents forums du groupe Aquitania au cours de la première semaine de la formation Simuligne.

o Un graphe des lectures dont les sommets sont divisés en 2 catégories : les acteurs A_i et les classes de messages C_j et les arcs orientés (C_j, A_i) sont valués par le nombre d'unités de la classe Cj qui ont été lues par l'acteur A_i.

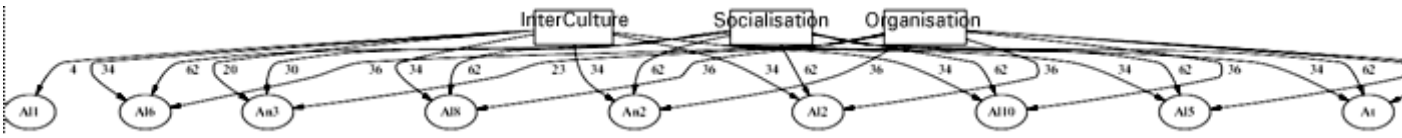

Figure 7 : représentation graphique du nombre de messages (déposés dans la première semaine) lus.

Lecteur	Organisation	InterCulture	Socialisation
Al1		4	
Al10	36	34	62
Al2	36	34	62
Al5	36	34	62
Al6	36	34	62
Al8	36	34	62
An2	36	34	62
An3	23	20	30
At	36	34	62

Figure 8 : représentation matricielle du nombre de messages (déposés dans la première semaine) lus.

Le premier modèle constituant un sociogramme basé sur les communications par forum, il permet de rester dans le cadre des réseaux sociaux, tout en prenant en compte les remarques faites précédemment sur la nature nouvelle de ces relations distantes, écrites, accessibles à tous et à tout moment. Si l'on considère que le fait de poster des messages est une marque importante de l'engagement dans la tâche, plus forte encore lorsqu'il s'agit d'un message pour initier un fil de discussion, le fait de les lire peut être considéré comme une forme légèrement plus faible de cet engagement dans la tâche. Le délai de lecture (s'il est connu) d'un message peut éventuellement être utilisé pour débiter l'engagement d'un acteur qui n'aurait pas consulté un message. Le fait de répondre à un message peut être considéré comme une forme d'engagement vis-à-vis de la personne à qui l'on répond tout autant que cette réponse dénote d'un engagement dans la tâche. On peut, en pondérant les différents types de contributions (ou non-contributions), représenter l'engagement de chacun des acteurs. Cet engagement peut être calculé globalement, mais on peut aussi différencier l'engagement dans la tâche de celui vis à vis des autres acteurs.
Le deuxième modèle rend compte plus finement des contributions (en écriture ou en lecture) pour les différentes classes de messages, pour chacun des acteurs, mais fait une totale abstraction des relations interindividuelles, comme si les acteurs discutaient entre eux indépendamment de toute forme de sociabilité. Cette deuxième approche permet en revanche des analyses plus variées. En effet, chaque chercheur pourra définir sa propre classification de messages pour étudier les profils des différents acteurs au regard de cette classification.
Chacun des deux modèles repose sur une motivation différente :

- Dans le premier cas, on suppose qu'un acteur est influencé pour une part non négligeable, par son engagement vis-à-vis des autres acteurs du groupe,
- Dans le deuxième, on suppose que seul le thème, le type du message lu ou posté ou plus généralement la classe à laquelle ce message appartient, engage l'apprenant à le lire ou à l'écrire.

Tout message étant daté, on peut bien sûr décider de ne prendre en compte que les messages ayant été posté et lus entre deux dates précises. On peut également envisager de ne sélectionner que les messages d'un certain fil de discussion ou relatifs à telle ou telle activité pour regarder plus précisément l'allure du graphe.
Quel forum ? Pour y analyser quoi ?

Comme nous le verrons par ailleurs dans ce livre, les forums ne sont pas tous de même nature de par le nombre de leurs contributeurs et ou lecteurs, mais aussi selon la façon dont ils sont intégrés ou non à un enseignement (lui-même pouvant être partiellement ou totalement à distance), si le tuteur ou l'enseignant y a un rôle ou non, s'ils servent précisément de support à une activité collaborative ou s'ils sont simplement des espaces ouverts pour les apprenants qui cherchent à rompre leur isolement ou simplement répondre à un problème pratique. Ils génèrent par conséquent ni la même activité, allant du vide au bouillonnement, ni les mêmes contenus, allant du message épistémique (questionnant ou construisant la connaissance en question) au message de salutation, de remerciement ou de félicitation en passant par les mises au point de calendrier et les rappels à l'ordre ou même les messages de protestation, de révolte ou de défoulement. Selon la durée de l'activité et les conditions de fermeture de l'espace, les temps du forum ne sont ni homogènes en intensité de l'activité, ni semblables dans la forme que peut prendre la structure communicative. Les formes positives de contribution attendues en fonction des différentes phases de la formation sont repérables par différentes organisations communicatives avec des paramètres ajustés comme nous le montrerons plus loin. Par exemple, dans un groupe où les membres ne se sont jamais rencontrés, il semble difficile de débuter une quelconque activité collaborative sans être passé par une phase de socialisation, de négociation des objectifs et de l'identité du groupe et de positionnement de chacun dans cette nouvelle structure. Dans cette première phase, les plus extravertis auront un rôle moteur dans la construction de la cohésion du groupe en répondant par exemple systématiquement au premier message d'un nouveau membre pour encourager son intégration. Dans une phase de brain storming la diversité des idées postées est un facteur clé de réussite de l'activité. Lorsque vient la synthèse, il faut des gens à la fois motivés, impartiaux, ayant le sens de l'écoute et de l'organisation pour achever et réussir une synthèse. Les questions et débats épistémiques sont rares, il est donc important pour l'enseignant de les détecter et d'en tirer tout le profit possible. Les apports d'information restent bien souvent les contributions de contenu les plus nombreuses. Certaines contraintes liées à l'organisation de la formation (vacances, examens, séances de regroupement, etc.) modifient profondément les contenus et le rythme des contributions dans le forum. Par ces quelques exemples, nous voulons souligner l'importance du contexte de l'utilisation d'un forum sur la forme communicative qui peut en être extraire et représentée à des fin d'analyse par les acteurs « dans le feu de l'action » ou par des chercheurs « hors le feu de l'action » [PERRENOUD 2001].
Rappelons ici que nous sommes motivés pour apporter à terme, des modèles et outils de suivi des formations collaboratives en ligne, pour aider le tuteur dans différents aspects de son rôle et particulièrement :

- Aider les apprenants dans leur méthode d'apprentissage : soutient méta-cognitif,
- Recentrer ou ouvrir les débats, donner de nouvelles pistes : aide cognitive,
- Repérer les individus en difficulté et les soutenir pour leur éviter l'abandon : soutient socio-affectif,
- Motiver les apprenants et animer le groupe pour que chacun tire le meilleur bénéfice de la formation,
- Respecter/ajuster le calendrier, en coordination avec d'éventuels autres groupes : rôle organisationnel
- Evaluer la participation des apprenants dans les tâches collaboratives.

Espoirs et limites des premières propositions
Pour les avoir étudiés en détails dans [REFFAY & CHANIER 03a ; REFFAY & CHANIER 03b] il nous semble que les outils de mesure de la cohésion proposés dans la

littérature des réseaux sociaux peuvent servir à la construction d'indicateurs prometteurs pour caractériser de la dynamique d'un groupe en FAD. Nous avons aussi montré dans [REFFAY & CHANIER 02] qu'il était difficile de définir une quantité (simple réel entre 0 et 1) capable de représenter la cohésion du groupe. L'analyse de la cohésion dans un groupe de télé-formation collaborative n'est pas une fin en soit mais sert à caractériser les membres du groupe qui assurent ou au contraire réduisent cette cohésion. Une simple valeur réelle n'apporte pas ce type d'information. C'est pourquoi nous développerons sur un même exemple une analyse à l'aide de cliques de niveau n d'une part et une analyse à l'aide de clusters hiérarchiques d'autre part.

Références

[CHANIER 02] CHANIER, T., « Rapport de fin de recherche », *Cognitique2000*, rapport de fin de projet, projet n°23, ministère de la recherche, 2002.

[DEGENNE & FORSE 94] DEGENNE A., FORSE M., *Les réseaux sociaux*, Collection U, Série Sociologie, Armand Colin, Paris, 1994.

[FORD & FULKERSON 56] Ford, L. R. Jr., Fulkerson, D. R., "*Maximal Flow Through a Network"*, Canadian Journal of Mathematics, 8, 399-404 (1956). http://wisl.ece.cornell.edu/ECE794/Jan29/ford_fulkerson/ff1956.pdf

[GANSNER & NORTH 00] Gansner, E., North, S., "An open graph visualization system and its applications to software engineering", Software - Practice and Experience 30(11): 1203-1233 (2000).

[PERRENOUD 01] Perrenoud, P., « Développer la pratique réflexive dans le métier d'enseignant. Professionnalisation et raison pédagogique », Paris, ESF, 2001.

[Reffay & Chanier 02] Reffay, C., Chanier, T., « Social Network Analysis Used for Modelling Collaboration in Distance Learning Groups », *Intelligent Tutoring System Conference*, Biarritz, France, 2002, p. 31-40.

[Reffay & Chanier 03a] Reffay, C., Chanier, T., « Mesurer la cohésion d'un groupe d'apprentissage en formation à distance », *Environnement Informatique pour l'Apprentissage Humain. (EIAH'03)*, *Strasbourg, Avril 2003*.

[Reffay & Chanier 03b] Reffay, C., Chanier, T., « How social network analysis can help to measure cohesion in collaborative distance learning? », *Proceeding of the Computer Supported Collaborative Learning Conference. (CSCL'03)*, Bergen, 2003.

[SCOTT 00] SCOTT, J., *Social Network Analysis : a handbook*, 2nd ed., SAGE, London, 2000.

[WASSERMANN & FAUST 94] WASSERMANN S., FAUST K., *Social Network Analysis : Methods and Applications*, Cambridge University Press, New York, 1994.

Toile

[INSNA 05] "International Network for Social Network Analysis" http://www.insna.org/ , last visited 16/02/2005.

[GraphViz 00], "Graph Visualization Software", http://www.graphviz.org/ , last visited 16/02/2005