



HAL
open science

Quand l'analyse quantitative fait parler les forums de discussion

Christophe Reffay, Luigi Lancieri

► **To cite this version:**

Christophe Reffay, Luigi Lancieri. Quand l'analyse quantitative fait parler les forums de discussion. STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation), 2006, 13, 24 p. hal-00696353

HAL Id: hal-00696353

<https://hal.science/hal-00696353>

Submitted on 11 May 2012

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

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

Quand l'analyse quantitative fait parler les forums de discussion

Christophe REFFAY [LIFC, Besançon]

Luigi LANCIERI [France Télécom R&D, Caen]

■ **RÉSUMÉ** : Cet article fédère deux points de vue concernant l'analyse quantitative des forums de discussion. D'abord, nous présentons des résultats d'études sur des forums étendus. Nous proposons des indicateurs construits à partir des traces de l'activité des utilisateurs, qui permettent d'identifier et de qualifier les usages dans les grands groupes. Dans ce contexte, même si les indicateurs sont statistiquement représentatifs, ils ont l'inconvénient d'être parfois impersonnels. Dans le second point de vue nous employons des méthodologies analogues mais cette fois appliquées à de petits groupes plus fermés où chaque membre est explicitement engagé et clairement identifié. Nous nous appuyons en particulier sur les réseaux sociaux, dont nous discutons la transposition dans un environnement de communication à distance. Au delà du constat de la convergence des deux approches, nous montrons qu'elles peuvent être utilisées de manière positive afin d'explorer d'autres contextes que les forums de discussion.

■ **MOTS CLÉS** : analyses quantitatives, forum de discussion, réseaux sociaux, dynamique de groupe.

■ **ABSTRACT** : *This article merges two viewpoints concerning the quantitative analysis of discussion forums. First, we present recent results on forums studies, implying large groups with a high activity level. We propose indicators based on the data produced by user's activity that allow to identify and to qualify the usage of the forum in such groups. In this context, even if statistically representative, these indicators stay somehow impersonal. In the second viewpoint we apply similar methodologies to smaller and closed groups where each member is explicitly engaged and clearly identified. We focus on the social networks analysis. We discuss its transposition in an environment of remote communication. Beyond the convergence of the two presented approaches, we show that they can be used in a positive way in order to explore other contexts using collaborative environments.*

■ **KEYWORDS** : quantitative analysis, discussion forum, social network analysis, group dynamic.

1. Introduction
2. Métrologie des usages à travers les forums de grands groupes
3. De la métrologie des usages aux réseaux sociaux dans les forums éducatifs
4. Indicateurs et réseaux sociaux à travers les interactions en contexte éducatif
5. Comment reprendre ces outils et transférer ces résultats dans la communication médiatisée en contexte éducatif ?
6. Deux approches à unifier
7. BIBLIOGRAPHIE
8. Références à des sites Internet

1. Introduction

Convaincus de l'importance des interactions dans l'apprentissage, nombreux sont les enseignants qui entraînent leurs apprenants dans des activités de discussion à travers les forums en ligne. Ces espaces de discussion asynchrones sont souvent vus comme des compléments peu coûteux et pratiques pour approfondir et négocier les concepts abordés dans le cours. Pratiques en effet puisqu'ils relâchent les contraintes d'espace et de temps en offrant, en tout lieu, et à tout instant, la possibilité aux acteurs de contribuer à la discussion. Ces espaces peuvent être soit intégrés à des formations en ligne où ils constituent des espaces privilégiés de rencontre et d'échanges pour les acteurs distants, soit proposés en complément d'un enseignement en présence.

Les forums sont utilisés depuis plus longtemps par des communautés de pratiques en dehors des institutions de formation, composées de passionnés, de techniciens pointus, de collectionneurs, de citoyens, de malades, de professionnels, etc., couvrant une multitude de sujets. Les groupes de discussion sont appréciés par ces types de public pour la vivacité des échanges et la fraîcheur de l'information, qui, dans l'interaction, favorisent la créativité, l'innovation et le débat d'idées.

Bien que riche dans ses contenus, et porteur d'intelligence collective, le forum est un outil où la contribution individuelle est peu structurée. Face à la quantité de messages déposés dans le forum, le formateur et l'animateur d'une communauté de pratiques se sentent souvent démunis pour construire une représentation synthétique de l'activité dans le forum. Ils doivent souvent se satisfaire de l'idée qu'ils s'en font à partir du souvenir de leur lecture. Le formateur risque de manquer d'objectivité quand il s'en sert pour évaluer l'implication et la place des apprenants dans les discussions, ce qui rend plus difficile son soutien auprès des individus ou plus contestables ses appréciations. Pour tenter de répondre à ces besoins nous proposons ici de rassembler des techniques de calcul automatique de représentations synthétiques à partir de grands volumes d'échanges textuels entre acteurs.

Même en contexte éducatif, 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 ont un rôle, 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 aidant à construire 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 les messages de protestation, de révolte ou de défoulement, ou plus fréquemment, un mélange de tout cela.

Selon la durée de l'activité et les conditions d'accès à 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 ». C'est justement cette structure communicative que nous cherchons à représenter de façon synthétique avec des outils automatiques. Nous faisons l'hypothèse que les formes positives de contribution attendues en fonction des contextes étudiés sont repérables par différentes organisations communicatives. Par exemple, dans une formation institutionnelle, si les participants ne se sont jamais rencontrés, il semble difficile de débiter 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 brainstorming la diversité des idées postées est un facteur clé de réussite de l'activité. Lorsque vient la synthèse, il faut des gens motivés, ayant le sens de l'écoute et de l'organisation pour achever et réussir une synthèse. 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.

Les modèles, outils de mesure ou de suivi des formations collaboratives en ligne, ont pour but d'aider le tuteur dans différents aspects de son rôle et particulièrement :

- Soutien méta-cognitif : aider les apprenants dans leur processus d'apprentissage,
- Aide cognitive : recentrer ou ouvrir les débats, donner de nouvelles pistes, motiver les apprenants et animer le groupe pour que chacun tire le meilleur bénéfice de la formation, ces deux activités participent aussi au soutien socio-affectif,
- Soutien socio-affectif : repérer les individus en difficulté et les soutenir pour leur éviter l'abandon,
- Rôle organisationnel : respecter/ajuster le calendrier, en coordination avec d'éventuels autres groupes,
- Rôle institutionnel : évaluer la participation des apprenants dans les tâches collaboratives (cette évaluation peut aussi servir des aspects cognitifs ou méta-cognitifs).

Dans ce contexte les questions que nous débattons peuvent s'articuler autour de trois interrogations. D'abord, de quels outils génériques dispose-t-on pour mesurer et visualiser de manière synthétique les fonctionnements des groupes médiatisés. La seconde question fortement liée est comment adapter ces outils aux spécificités du contexte à investiguer. En particulier, nous considérons le cas des grands groupes puis le cadre plus restreint des forums éducatifs. Enfin, quelles relations peut-on mettre en évidence entre l'étude des grands et des petits groupes.

Dans cet article, nous commençons par présenter une série d'outils statistiques en montrant comment ils peuvent être utilisés pour mieux comprendre la dynamique et les structures d'interactions dans les forums très ouverts de l'Internet (i.e. : listsev, usenet). Puis, nous abordons l'étude des forums à vocation éducative au travers de la théorie des réseaux sociaux. Cette méthodologie initialement définie pour des rapports en face à face est discutée et transposée dans des formes de communication à distance. Nous mettons en évidence certaines propriétés de « self-similarité » dans la structure de groupes à différentes échelles, et montrons que des techniques similaires permettent de servir des objectifs différents dans ces deux contextes.

2. Métrologie des usages à travers les forums de grands groupes

Une des questions importantes que soulève la notion de métrologie est relative à la qualité des conclusions qu'il est possible d'obtenir à partir de données de base purement quantitatives. Dans cette partie, nous développons ce thème à la lumière de notre propre réflexion, appuyée par une synthèse de travaux analogues.

2.1. Métrologie et modélisation

En offrant des outils de mesure des caractéristiques et de l'évolution des forums, la métrologie des groupes de discussion a deux objectifs principaux : d'abord permettre d'uniformiser et d'automatiser l'analyse et ensuite traiter une grande quantité de forums aussi souvent que nécessaire. Ces informations permettent de mieux comprendre le mode de fonctionnement par nature complexe de ces environnements. En favorisant une « cartographie » des comportements et des interactions, ces données facilitent la prise en compte de modèles cognitifs et sociaux.

En plus de fournir une meilleure visibilité sur les interactions dans les forums, ces informations peuvent être utilisées pour faire de la veille technologique dans une stratégie d'intelligence économique. Les indicateurs quantitatifs d'activités peuvent être déterminés, puis comparés par thème de discussion, ou par participant. Ce type de données peut être calculé de manière instantanée ou sur une période ou en considérant les évolutions sur plusieurs périodes. En terme d'application, ces données peuvent être utiles pour faire une sélection automatique des groupes actifs d'une manière générale mais aussi des pointes d'activités ponctuelles qui sont le signe d'un événement peu ordinaire ou de sujets intéressants faisant l'objet d'un débat particulièrement actif.

L'évaluation quantitative de l'activité (e.g. nombre de messages par jour) est utile mais a ses limites. Par exemple, 10 messages par jour sur un groupe n'ont pas le même sens s'il s'agit de posts relatifs à un sujet nouveau (10 départs de fil) ou s'il s'agit de réponses à un même message. Pour prendre en compte cet aspect plus qualitatif de l'activité du forum, il est possible de calculer un indice de nouveauté qui reflétera le niveau de jeunesse ou de maturité des discussions dans un forum. En effet, un forum qui ne produit que de nouveaux posts restant sans réponse a peu d'intérêt du point de vue de l'activité collective. A l'autre extrême, un forum qui ne produit que des réponses au même post finira par tourner en rond et produira peu d'innovation. Il existe sans doute un équilibre entre ces 2 extrêmes mais il est clair que l'intérêt d'un forum dépendra de l'individu qui selon le cas préférera la profondeur des discussions nourries ou l'attrait de la nouveauté. Le calcul du taux de posts nouveaux ne faisant pas référence à un post ancien (hors fil) permet d'évaluer ce phénomène.

Cette grandeur a aussi l'intérêt d'évaluer la profondeur de la mémoire collective du forum. En effet, comme nous le disions, les usagers ont tendance à consulter les posts récents, considérant, parfois à tort, que les plus anciens sont périmés. En fait cette impression n'est pas toujours justifiée car l'expérience de la recherche d'information dans les posts anciens grâce aux moteurs de recherche, notamment, réserve quelquefois de bonnes surprises. Quoi qu'il en soit le fait de privilégier les posts récents a pour effet de jeter rapidement aux oubliettes ceux qui ne font pas l'objet de réponse. Une réponse en effet peut rendre actuel un post parfois ancien. La mémoire du forum, c'est à dire l'ensemble des posts qui ont une bonne probabilité d'être consultés, (i.e. plutôt récents) sera donc fortement influencée par le taux de réponses.

Il est aussi possible d'envisager les groupes sous l'angle de l'interaction. Par exemple en déterminant le taux d'activité d'un usager habitué d'un groupe dans un autre groupe (*cross-posting*). Ce taux permet d'établir une cartographie des groupes en relation, basée sur l'usage effectif de leurs membres. Ces interactions réelles peuvent être rapportées à la proximité thématique de chaque groupe. Ces 2 paramètres peuvent être associés au sein d'un indicateur de synergie potentielle. Dans ce contexte, nous avons proposé ([Lancieri, 2000](#)) une méthode pour évaluer le niveau d'interaction dans les groupes de discussion. La métrique associée basée sur la puissance itérée (formulation issue de la théorie des graphes) permet avantageusement de prendre en compte les contributions indirectes dans les faisceaux d'interactions. La valeur globale obtenue synthétise les diverses contributions dans les groupes et en particulier dans des situations de *cross-posting*. Cette méthode peut aussi être employée à l'intérieur des groupes de grande dimension où une segmentation artificielle a été opérée (algorithmes de catégorisation). La contribution croisée des usagers aux différents segments peut aussi être capturée et formalisée par cette méthode. Nous avons aussi proposé ([Lancieri, 2004](#)) une méthode de mesure du niveau de coopération potentielle entre individus basée sur le niveau de recouvrement de leur « profil thématique » (mots les plus fréquents rencontrés dans les posts d'un usager sur une période). Ces grandeurs sont intéressantes pour suivre l'activité des forums (i.e. évolution dans le temps du niveau global de synergie potentielle) mais aussi pour favoriser des mises en relation. Par ailleurs, les modèles statistiques liés à certaines grandeurs caractéristiques des forums mettent en évidence une propriété self similaire. Le modèle de loi de probabilité sous-exponentielle associé (Pareto, Weibull, log normal, etc.) dit à mémoire par opposition à des modèles de type loi de Poisson qui ignore « le passé » a, dans de nombreux travaux, été mis en avant pour refléter le comportement humain.

2.2. Evaluation de la structure d'interaction

L'analyse de la structure d'interaction est basée sur l'observation que l'activité humaine est bien décrite par des lois de probabilités sous exponentielles. Ces lois ont le grand intérêt de pouvoir s'exprimer de manière synthétique.

Dans cette approche, l'étude de l'influence interne à un groupe passe par la quantification de la participation de chaque individu et par l'analyse de cette participation relativement à celle des autres membres du groupe. Notre étude est basée sur l'examen des entêtes des messages postés dans les forums Usenet. Notre hypothèse est que la structure d'influence dans un groupe peut être approchée par l'analyse de la distribution du nombre de messages émis par membre du groupe (personne ayant émis au moins un message). En particulier, nous mettons en évidence des caractéristiques de similarité interne. La principale limite de cette approche est qu'elle repose sur l'hypothèse forte que ceux qui émettent des messages lisent ceux des autres. Le bon sens nous fait supposer que c'est bien le cas mais rien ne l'affirme. Cependant, il est difficilement concevable que ceux qui envoient beaucoup de messages ne prennent pas le temps de lire ceux des autres. Traditionnellement ce genre d'attitude (parler sans écouter) mène à l'exclusion du groupe (comme le *trolling* par exemple). Il est ainsi très courant de voir des messages de la part des habitués du groupe demandant aux novices de lire les messages anciens avant de poser des questions déjà traitées auparavant. Nous postulons donc que cette hypothèse est réaliste dans une certaine mesure.

Les données de départ sont le nombre de messages postés par un usager en fonction de son rang d'activité. Cette fonction de type sous-exponentielle (la population des classes de rang élevé est sur-représentée par rapport à une loi exponentielle) exhibe après transformation logarithmique un aspect self-similaire matérialisé par une quasi droite dont la pente (coefficient de la loi de Zipf généralisée) reflète une des caractéristiques de l'interaction. Cette caractéristique, vue comme la signature du groupe, traduit la dissymétrie entre les usagers très actifs et peu actifs. Cette structure peut être différente en fonction des groupes. Nous définissons le coefficient de complexité de l'interaction (CCI) par la valeur de cette pente. Ce coefficient traduit la « courbure » de la distribution du nombre de messages par membre du groupe. Ce coefficient va plus loin que la représentation des écarts d'activité qui aurait tout aussi bien pu être résumée par l'écart type de la distribution. Le coefficient CCI traduit la complexité et modélise de manière fine la manière dont l'activité de chacun est liée à celle des autres (voir [\(Lancieri, 2005\)](#) pour plus de détails).

Comme nous l'avons évoqué précédemment, nous associons ce coefficient à la structure des influences au sein du groupe. Les coefficients élevés, en valeur absolue (courbure importante, pente forte) traduisent une structure d'influence très polarisée. Dans ce cas, le leadership est concentré et l'influence dominante émane plutôt d'une minorité. Comme dans le cas d'une situation d'enseignement, la plupart écoutent et peu d'usagers prennent la parole. Les coefficients plus faibles (courbure faible, pente faible) représentent le cas contraire, la plupart des membres interviennent : l'influence est plus largement partagée. Nous voyons que dans les différents cas, le niveau d'inflexion de la courbe modélisée par le coefficient CCI (i.e. la pente de la droite) traduit bien cette structure d'influence. Nous avons reproduit dans la figure suivante, dans un repère log-log, le niveau d'activité correspondant aux groupes « alt.os.linux » et « intel.inbusiness » de usenet. Les deux pentes sont respectivement de $-0,7$ et -2 , avec des coefficients de corrélation respectivement de $0,92$ et $0,96$. Ces résultats ne sont pas surprenants dans la mesure où on aurait pu prévoir que le groupe « intel.inbusiness » lié à une activité commerciale serait beaucoup plus orienté (polarisation des influences) que le groupe Linux plus libre.

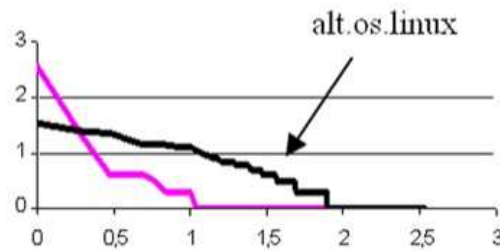


Figure 1 : CCI pour les groupes « intel.inbusiness » et « alt.os.linux »

Un aspect que nous n'avons pas traité ici mais qui mériterait d'être développé concerne l'évolution temporelle du CCI. La variation de cette évolution permettrait de rendre compte de l'évolution des influences au sein d'un groupe (voir (Lancieri, 2000) et (Lancieri, 2005) pour davantage de détails sur cette étude).

2.3. Travaux de référence sur la métrologie des forums usenet

Nous présentons dans cette section un certain nombre de travaux liés à la métrologie des forums. Saito présente dans (Saito et al., 1998) une étude statistique des groupes de news sous l'angle technologique. Il présente d'abord les protocoles et les architectures de service associés aux serveurs de forums (nature des échanges, organisation des réseaux, format et contenu des traces, etc.). En dehors de résultats quantitatifs (nombre de posts par groupe, distribution des tailles d'articles, etc.), des conclusions de type comportementales sont inférées par exemple par la comparaison des groupes en fonction du nombre de messages ou de la taille des messages, montrant que certains groupes évoluent vers des échanges plus multimédia.

Rank	Newsgroup Name	Posts	Posters	Replies	Replies	Replies	UnFMSGS	AvgLineCT	XPosts	XPTgs
1	rec.music.phish	10959	1106	501	16517	960	480	20	1082	120
2	rec.music.artists.springsteen	15237	1098	479	12872	874	792	29	793	97
3	rec.music.golead	9204	1092	334	7947	943	444	30	1145	132
4	rec.music.classical.recordings	6408	809	323	5774	479	242	30	768	46
5	rec.music.opera	8104	358	159	5550	284	181	27	366	32
6	rec.music.makers.guitar.acoustic	5875	563	306	5143	487	103	27	62	34
7	rec.music.makers.guitar.jazz	5495	600	292	4900	511	134	26	330	44
8	rec.music.makers.percussion	4340	397	205	3742	324	207	30	145	41
9	rec.music.hip-hop	4250	378	134	2631	288	237	47	1100	114
10	rec.music.classical	4171	567	180	3771	483	209	41	2303	99
11	rec.music.beatles	3935	552	231	3438	445	170	35	375	81
12	rec.music.dylan	3438	663	252	2691	502	378	35	384	78
13	rec.music.classical.guitar	3210	316	153	2816	240	140	40	83	51
14	rec.music.artists.kiss	3195	208	85	2037	163	98	27	262	58
15	rec.music.makers.guitar	2787	501	159	2529	416	81	33	1239	67
16	ntu-kpi.rec.music	1787	501	222	1026	291	412	11	0	0
17	rec.music.country.western	1691	403	69	1446	324	112	36	665	128
18	rec.music.progressive	1598	315	132	1364	216	126	26	398	49
19	rec.music.rock-00s-0b-1950s	1389	141	74	1158	105	81	40	67	13

Figure 2 • Exemple de données fournies par le site du projet NetScan

Le projet NetScan ([Netscan, 2005](#)) financé par Microsoft permet d'obtenir de nombreux détails quantitatifs sur le comportement des groupes Usenet. Les chiffres sont mis à jour mensuellement sur tous les groupes (recherche des groupes par mot clé). On y trouve des grandeurs comme le nombre de messages sans réponses, la taille des messages, le nombre de personnes ayant répondu à au moins un message, etc. Au total 9 grandeurs permettent de faire des comparaisons et de quantifier les comportements des groupes. Comme le montre la figure suivante, ces données peuvent servir à visualiser les interactions entre les différents groupes impliqués dans une relation de *cross-posting*.

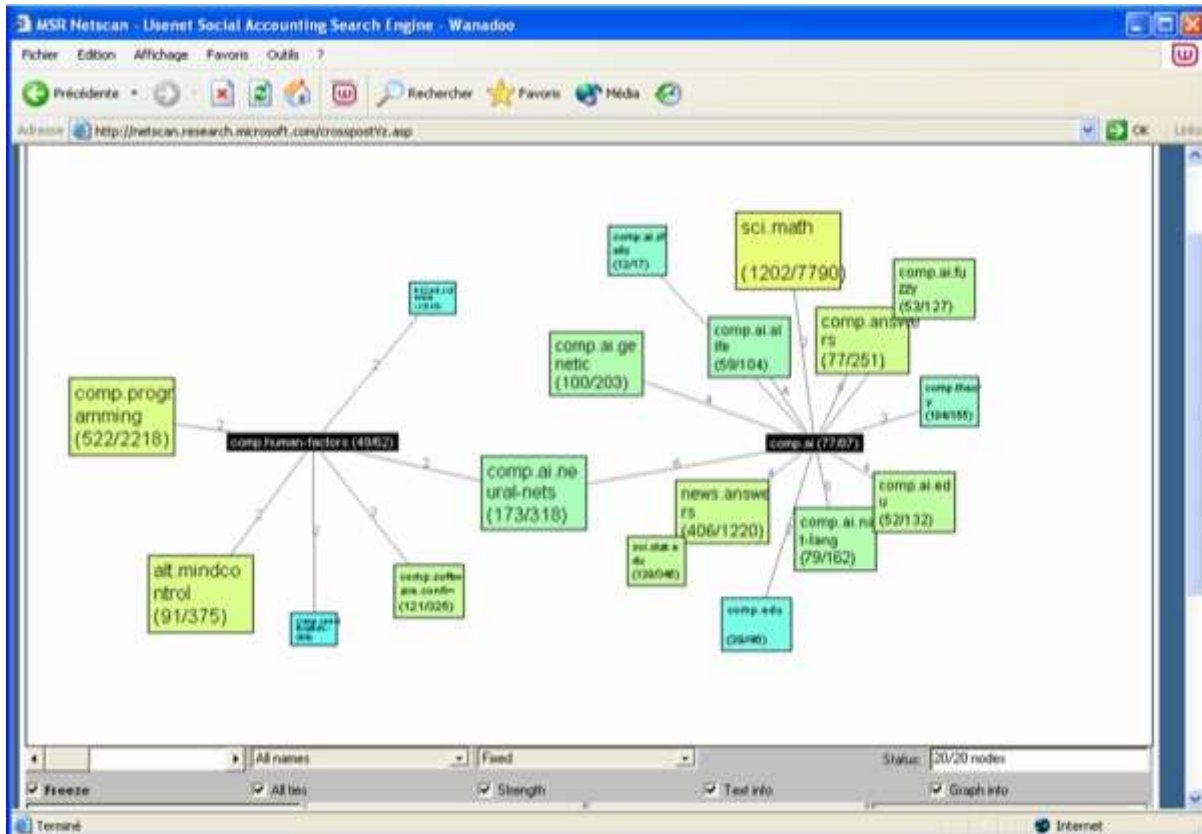


Figure 3 • Interactions entre groupes (projet Netscan)

Une des difficultés freinant l'étude fine du comportement des usagers est la nécessité d'opérationnaliser l'acquisition de traces d'activités. Pour pallier cette contrainte, certains auteurs ont modifié les outils de consultation (news readers) de manière à obtenir des traces plus détaillées. D. Maltz, par exemple, a modifié les news readers NN et XRN de manière à obtenir des traces sur les sessions de consultation ([Maltz, 1994](#)). Des données comme la durée de la session (ouverture, fermeture du news reader), le temps passé à scanner les sujets ainsi que le temps passé à lire un article (temps passé entre ouverture et fermeture de l'article) deviennent accessibles. L'étude met en évidence plusieurs résultats intéressants sur le comportement des usagers. L'auteur étudie par exemple le ratio entre le nombre de groupes consultés par session et le nombre de groupes auxquels se sont inscrits les usagers montrant que ce ratio décroît très vite (log normal) avec le nombre d'inscriptions, ce qui implique que les usagers qui s'abonnent à de nombreux groupes ont tendance à surestimer leurs capacités de lecture. La moitié des usagers s'abonne à moins de 20 groupes et arrivent dans 90 % des cas à les suivre. La fraction des usagers abonnés à plus de 100 groupes est de 8 % alors que seuls 1 % arrivent à suivre.

Dans le même esprit, Jones et al. ont étudié via une approche statistique les phénomènes de communications de masse dans les communautés électroniques (études de 578 groupes pendant 5 mois) ([Jones et al., 2002](#)). Les auteurs ont cherché à étudier l'impact du comportement des usagers dans les communications de groupes en mesurant les interactions entre utilisateurs. Ils commencent par proposer

une méthodologie pour reconstituer automatiquement les fils d'échanges sur Usenet, ce qui n'est pas simple compte tenu de la diversité des outils de consultation. En travaillant sur les indicateurs internes aux messages (Re, Reply, <, etc.) ils construisent un modèle basé sur les probabilités conditionnelles permettant d'évaluer automatiquement qu'un message est effectivement une réponse. Cette méthode permet de reconnaître automatiquement une réponse dans 99 % des cas. Ces études ont permis d'évaluer un certain nombre d'hypothèses. Elles indiquent que la complexité des messages (nombre de lignes) avait tendance à se réduire fortement (loi sous exponentielle) dans les groupes très actifs (nombreuses personnes postant des messages). Les auteurs ont aussi découvert que les messages qui démarraient un fil de discussion avaient tendance à être courts (peu complexes). Le pouvoir prédictif de ce modèle a été démontré dans 63 % des cas. Un autre point intéressant est que la participation des usagers est moins stable (usagers moins impliqués) dans les groupes très actifs où l'activité s'explique plus par une faible participation de nombreux intervenants que par des participations individuelles soutenues. D'autres travaux du même auteur (Jones, 2003) ont aussi montré que le type de technologie a un impact sur cette stabilité des contributions. L'auteur est arrivé à cette conclusion en comparant la stabilité des contributions dans le mode « groupes de news » et dans le mode « listserv ». Dans ce dernier mode les contributions sont plus stables (50 % des usagers postent des messages sur une période de 2 mois (i.e. les usagers ayant posté des messages un mois en ont posté le mois suivant) contre 11 % pour les groupes de news.

Whittaker et al, quant à eux, ont étudié l'évolution de 500 groupes de discussion pendant 6 mois (Whittaker et al., 1998). Ils ont réalisé des statistiques de manière à mettre en évidence différentes caractéristiques des groupes sous l'angle des interactions de masses. L'étude a débuté par une évaluation démographique de la population concernée (nombre de posteurs, nombre de messages par posteur, intervalle entre posts). Les auteurs font ressortir un indicateur décrivant l'évaluation du niveau de familiarité d'un posteur avec un groupe (27 % des posts proviennent de personnes n'ayant contribué qu'une fois alors que seul 2.9 % des usagers engendrent 25 % des posts). Le nombre moyen de contributions par utilisateur est de 3.1. Les auteurs n'y font pas allusion mais la courbe nombre de messages/posteur a l'allure classique d'une fonction sous exponentielle. En terme de structure d'interaction (stratégie d'échange), l'étude révèle que 34 % des messages de chaque groupe sont aussi adressés à au moins un autre groupe (3.1 groupes en moyenne). On observe aussi que sur le plan du *cross posting* chaque groupe a été en relation avec 272 groupes différents avec une cohérence faible (5.4 posts par groupe en moyenne, i.e. nombre de posts depuis l'utilisateur habituel d'un groupe vers un groupe donné). Les auteurs ont aussi mesuré le niveau d'interactivité en étudiant la profondeur des fils. Ils montrent qu'en moyenne un fil contient 1.8 messages alors que 33 % des fils contiennent plus de 2 messages (extension de conversation réussie). Les initialisations de communications manquées (messages seuls dans leur fil) correspondent à 44 % des messages. Les auteurs ont mesuré que 54 % des groupes produisaient des FAQs (Frequently Asked Questions) vues comme productions de connaissance plus structurées par le groupe. Les auteurs ont ensuite étudié l'impact de différentes variables sur l'évolution du groupe. La croissance de la production de FAQs, la décroissance du *cross posting*, la croissance de la taille des messages sont vues comme des facteurs d'augmentation de dénominateurs communs ou de la cohésion du groupe (*common ground*). L'auteur a utilisé un modèle causal construit à partir d'analyse de régression entre les variables. Il déduit par exemple que les groupes de grande taille ont tendance à engendrer du *cross-posting* et à contenir des messages de faible taille. Pareillement les groupes contenant beaucoup d'usagers familiers ont tendance à avoir moins de *cross-posting* et contenir des messages plus longs. Par contre la familiarité ne semble pas avoir d'effet sur la production de FAQs qui est surtout liée à la présence d'un modérateur. Les auteurs tirent diverses conclusions. D'abord, que même si tout le monde peut poster dans un news group, les messages sont surtout générés par une minorité très active (bavarde). Ceci contraste avec la communication plus physique (face à face ou via la vidéo ou l'audio seul) qui implique, toutes proportions gardées, une forme de communication plus égalitaire.

Smith a étudié les structures sociales émergentes ou invisibles des groupes Usenet (Smith, 1999) au

travers d'un rappel sur les racines et l'état actuel de Usenet (histoire, population, popularité par type de groupes, répartition géographique, organisation, technologie). L'auteur présente ensuite des statistiques détaillées sur les caractéristiques des messages, des usagers ou de la répartition chronologique des posts.

Viégas et Smith ont étudié la contribution individuelle des auteurs à l'effort collectif ainsi que le positionnement de cette participation individuelle dans le temps (Viégas et Smith, 2004). Différentes métriques et des représentations graphiques associées ont permis aux auteurs d'inférer un niveau de confiance à chaque contributeur en fonction de son niveau d'activité ou par croisement avec d'autres contextes où il intervient. Des études d'usages ont aussi montré que ces indicateurs étaient utilisés par les usagers pour choisir les messages à lire (modification des usages). Les auteurs ont obtenu les données issues de l'activité des groupes Usenet dans le contexte du projet NetScan (Netscan, 2005), (Smith, 1999). L'analyse visuelle des formes graphiques issues des remontées statistiques et chronologiques de l'activité a permis d'inférer des comportements de type initiateur de discussions, adepte des débats, contributeur de fonds ou perturbateur spammeur. Une approche analogue a été approfondie dans un contexte de navigation Web (Lavallard et Lancieri, 2004).

D'autres travaux n'impliquant pas seulement les forums mais aussi d'autres modes de communication ont montré l'influence de l'outil sur la coopération. Jensen et al. ont montré que la contribution des usagers à une activité en commun (un jeu en ligne) était plus importante quand le mode de communication était plus évolué (voix, synthèse vocale, texte chat, pas de communication) (Jensen et al., 2000). Les résultats montrent que la participation croît avec le niveau du mode. Les auteurs attribuent cette relation à la confiance qui s'instaure avec des formes plus évoluées de communication. On notera que la synthèse vocale engendre plus de contributions que le texte chat équivalent mais moins que la conversation humaine.

3. De la métrologie des usages aux réseaux sociaux dans les forums éducatifs

Les analyses rapportées dans la section 2 permettent de caractériser essentiellement l'activité d'un groupe dans sa globalité, ou ses liens avec les autres groupes dans l'analyse du *cross-posting*, mais permettent plus difficilement l'analyse fine de la place de chaque individu dans le groupe. La difficulté provient aussi du fait que de tels groupes, ouverts par nature, connaissent un renouvellement important des contributeurs tandis qu'un quart des contributions environ provient d'un petit noyau plus stable (Whittaker et al., 1998). Ces difficultés expliquent au moins en partie que ces études soient centrées sur la thématique des différents groupes plutôt que sur les individus.

Mais si nous nous plaçons maintenant dans un contexte plus restreint, celui des forums intégrés à des modules de formation ouverte ou à distance, nous pouvons, malgré la diversité des plateformes, compter sur des informations plus précises sur chaque individu. Nous pouvons aussi tirer avantage du fait que le public de ces forums particuliers est le plus souvent captif : i.e. les participants sont en général engagés individuellement dans un contrat d'apprentissage, le plus souvent enrôlés par leur formateur, et contraints par une consigne de participation qui peut même donner lieu à une évaluation. Cet engagement assure une certaine stabilité au groupe et donne donc à chacun des participants un cadre temporel plus clair pour tisser des liens sociaux avec les autres. Nous ne reprendrons pas ici les arguments en faveur de l'apprentissage collaboratif, mais nous prenons comme hypothèse que dans les dispositifs collaboratifs, la participation active de chacun est une condition majeure pour la réussite de l'activité collective. C'est autant le processus de construction que le résultat de cette activité qui est porteur d'apprentissages. C'est dans ce cadre que l'étude de la dynamique des groupes devient un enjeu important et les modèles issus des réseaux sociaux peuvent, si on est capable de leur donner un sens dans ce contexte, nous aider à la mesurer et à en suivre l'évolution.

Il est donc clair que l'étude des grands forums et des petits groupes ne peut pas être appréhendée de la même manière bien que composés d'individus utilisant un média analogue, les outils d'investigation peuvent dans une large mesure être identiques. En dehors de la taille du groupe une différence importante se trouve dans l'intégration du dispositif dans le contexte socioéducatif. Même si cette différence doit être prise en compte par l'observateur, il n'en reste pas moins que l'outil d'observation unit les deux contextes. Par ailleurs, d'une manière générale, les petits groupes s'inscrivent dans des groupes plus importants se recouvrant même parfois. Ceci réduit le fossé entre les notions de grands et de petits groupes qui se conçoivent plutôt dans une relation de continuité à des niveaux d'échelle différents.

Dans cette perspective et après avoir observé les forums dans une vision globale nous détaillons des formes d'investigation plus granulaires mettant en évidence jusqu'au rôle de l'individu dans le groupe.

4. Indicateurs et réseaux sociaux à travers les interactions en contexte éducatif

Dans cette partie, les dispositifs ciblés considèrent qu'un groupe (4 à 15 apprenants environ) est suivi par un tuteur, secondé par des experts de contenu dans certains cas. Nous sommes bien loin des milliers d'internautes présentés dans les forums de la partie 2. 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. Un modèle socio-constructiviste avec une organisation extensive nécessitant une durée (minimale) de deux mois caractérise le type de formation visée et confère 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. Dans le cas d'activités organisées et encadrées par un formateur, intégrant des forums de discussion, le formateur ou un tuteur en charge du suivi de l'activité peut se perdre dans le contenu et dans le volume des interactions et éprouver des difficultés à jauger l'engagement de chaque individu dans le groupe et plus globalement la dynamique du groupe. Nous pensons que ce type d'acteur en particulier a besoin d'outils de représentation synthétique des interactions pour compléter la vision perçue à travers une trajectoire de lecture des contenus du forum.

En 2000, on osait à peine parler de « groupe » dans le cas d'une formation à distance et en ligne, 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 (Chanier, 2001). Y-a-t-il une mesure objective permettant de caractériser la cohésion ? Peut-on évaluer la dynamique du groupe ?

Comme dans (Dimitracopoulou et Bruillard, dans ce volume), 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 inter-actants, la dynamique du groupe,..., toutes notions qu'un enseignant rompu aux techniques collaboratives en face-à-face évalue directement 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.

4.1. Pertinence des indicateurs quantitatifs issus d'une plateforme de téléformation

Les indicateurs calculables dans les plateformes sont à classer en plusieurs niveaux de complexité qui les rend plus ou moins coûteux (en temps de calcul) à extraire (individuel/groupe, compteurs/graphes) et apportent des informations plus ou moins pertinentes au pédagogue. Nous proposons de distinguer plusieurs indicateurs regroupés en classes 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), voir la [figure 4](#). 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 de 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.

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.

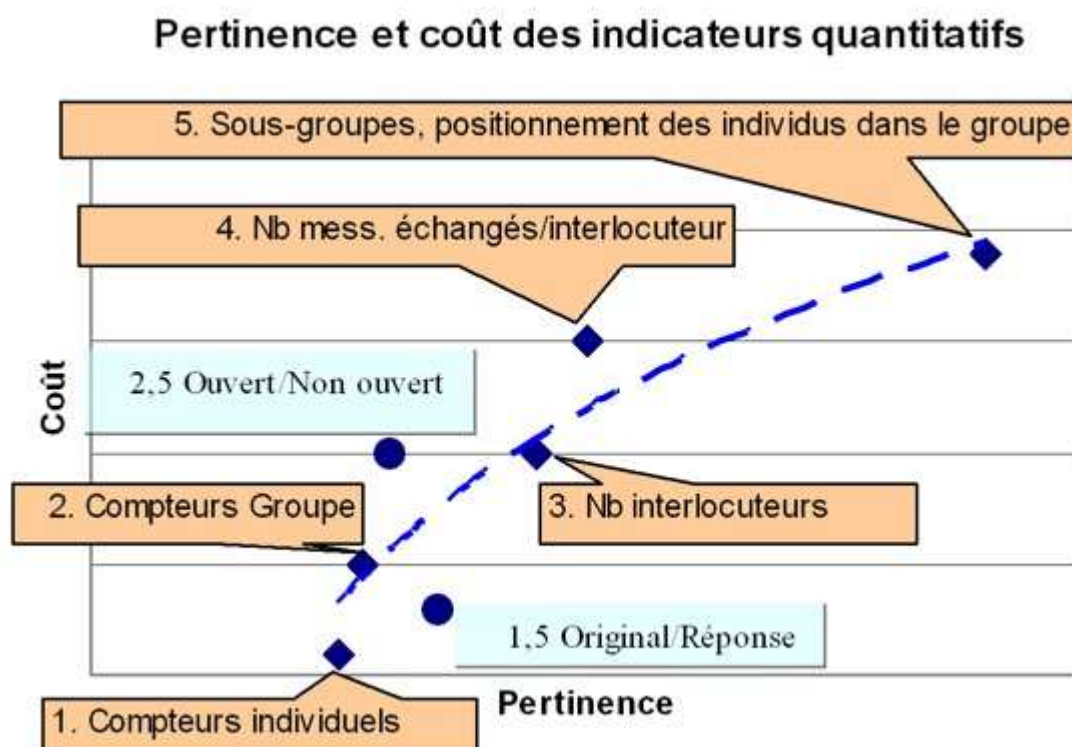


Figure 4 • Pertinence et coût des indicateurs

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. Bien qu'il s'agisse toujours de compteurs individuels, cette distinction apporte plus de pertinence à la forme de participation de l'acteur. Ces indicateurs sont donc placés dans une classe intermédiaire entre les classes 1 et 2 (1,5 sur la [figure 4](#)).

A partir de la première classe, on peut assez simplement construire dans une deuxième classe les indicateurs représentant le nombre de messages émis par 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-expert, 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 le LMS pour permettre à l'utilisateur de distinguer dans la liste des messages, ceux qui ont été déjà ouverts de ceux qui ne l'ont jamais été. Bien que cela puisse être coûteux en développement (selon le LMS), on peut utiliser cette information pour

associer à chaque message du forum, la liste des personnes l'ayant ouvert. Cette précision permet 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* » (donc « non lus ») par les destinataires.

Les classes précédentes d'indicateurs donnent quantitativement ce que chaque acteur poste au groupe ou reçoit du reste du groupe sans distinguer ses interlocuteurs. L'analyse des réseaux sociaux permet de préciser les relations en différenciant les interlocuteurs. 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'ego. Plus simplement, il s'agit du nombre d'interlocuteurs de cet acteur. De nombreuses études basées sur des relations sociales en présence et rapportées dans (Degenne et Forsé, 1994) montrent qu'une personne capable de maintenir un vaste réseau de contacts/partenaires est un acteur important dans la communauté. Cette importance, appelée « capital social », lui donne certaines facilités pour trouver un nouveau travail ou un nouveau logement par exemple. Cet indicateur (degré intérieur ou extérieur) dans le réseau d'ego permet donc d'informer sur le capital social de l'individu. Dans le cadre des communications d'un LMS, il est plus difficile à construire car il correspond à un nombre de personnes (émettrices ou réceptrices) et non à un nombre de messages. Pour un même nombre de messages échangés globalement, il permet de distinguer un bavard marginal (échangeant de nombreux messages avec très peu de personnes) d'un véritable animateur du groupe (échangeant quelques messages mais 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 leur répartition entre les différents 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 sélection peuvent restreindre l'ensemble des unités prises en compte : période (communication ayant été émise entre deux dates précises), ouverture (seules les unités contenues dans un message ayant été ouvert sont comptabilisées), etc.

Pour compléter la classification proposée par la [figure 4](#), 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 et Forsé, 1994), (Wassermann et Faust, 1994), (Scott, 2000). Nous présentons 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 invitons le lecteur intéressé à choisir ses outils parmi ceux que l'on peut trouver en particulier sur le site du « réseau international pour les réseaux sociaux » (INSNA, 2005) et qui permettent de calculer des indices ou des représentations issues des réseaux sociaux.

4.2. La cohésion proposée par les réseaux sociaux

Issus de la dynamique des groupes, de la sociologie et de la théorie des graphes, les réseaux sociaux

proposent des outils méthodologiques 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.

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. Une forme d'opérationnalisation est proposée dans (Wassermann et Faust, 1994) pour qui 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 (affinité) 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 inférieure ou égale à n,
- des **cliques de niveau s** où tous les nœuds sont directement liés par des relations d'intensité supérieure ou égale à 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 englobant intègre le $n+1^{\text{ème}}$ nœud lié avec la plus grande intensité aux n 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).

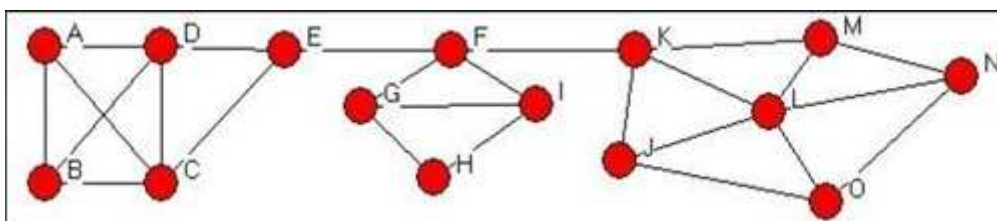


Figure 5 • (ABCD) est une clique, (ABCDE), (FGHI) et (KLMNO) sont des 2-cliques

Les liens de la figure 5 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.

4.3. Centralité, intermédiarité, influence et pouvoir

Il est indubitable que les individus les plus centraux dans un graphe occupent des positions privilégiées dans les échanges, notamment par rapport à ceux qui sont rejetés à la périphérie. De nombreuses études rapportées dans (Degenne et Forsé, 1994) établissent que la centralité est liée au pouvoir, à l'influence ou

au leadership, 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. Cette même notion est utilisée pour le calcul par Google de la popularité d'une page web en prenant en compte le nombre de liens (d'autres pages) qui pointent vers elle (Brin et Page, 1998). Mais un nœud lié à n nœuds périphériques n'est pas aussi central que celui qui est lié à n nœuds, eux-mêmes centraux. Un nœud à la périphérie dépend d'intermédiaires (plus centraux) pour sa relation avec les autres.

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 normé de centralité de proximité pour le nœud i d'un graphe comportant n sommets, est donné par la formule :

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

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 et Forsé, 1994), 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 (de longueur minimale) reliant le nœud j au nœud k , et $g_{jk}(i)$ le nombre de ces chemins passant par le nœud i , on définit C_{Ai} l'indice de centralité absolu d'intermédiarité du sommet i par :

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

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

$$C_{Nii} = \frac{2C_{Ai}}{n^2 - 3n + 2}$$

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 mesurent 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 et Fulkerson, 1956)) 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 :

$$C_{AFi} = \sum_j^n \sum_k^n f_{jk}(i), \text{ avec : } j \neq k \neq i, \text{ et } j < k$$

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

$$C_{NF_i} = \frac{\sum_j^n \sum_k^n f_{jk}(i)}{\sum_j^n \sum_k^n f_{jk}}, \text{ avec : } j \neq k \neq i, \text{ et } j < k$$

Les quatre définitions de centralité proposées ci-dessus ne tiennent pas compte de la valeur de la centralité des sommets auquel un sommet 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.

5. 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 soit 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, pérennes et donc 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 écrit est plus particulièrement destiné à une personne (en réponse à sa question par exemple), toute autre personne accède dans les mêmes conditions 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'intermédiation pour mesurer le pouvoir, l'influence ou le degré de contrôle des communications n'a plus de sens dans le contexte des échanges par forum ou liste 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 de 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.

5.1. 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 dans cet article à la proposition de deux modèles que nous illustrerons à l'aide des données issues de la formation Simuligne. Le premier mesure le volume des échanges entre individus, tandis que le second représente le volume des échanges par catégorie.

5.1.1. La formation Simuligne

Simuligne est le nom donné à une formation qui a eu lieu dans le cadre du projet ICOGAD (Interaction Cognitives dans les Groupes en formation A Distance : projet (2001-2003) du programme « Cognitique 2000 », regroupant 3 laboratoires, piloté par T. Chanier.). Il s'agit d'une *simulation globale* (Yaiche, 1996) proposée en ligne pour la pratique des langues en situation réelle de communication, textuelle et asynchrone. Un extrait des traces issues de cette formation est disponible sur (Mulce, 2007). Le scénario invite les apprenants à une production collaborative dans la langue cible. Cette collaboration implique de nombreuses interactions pour organiser, négocier, décider et finalement produire ensemble (Reffay et al., 2002). Le corpus concerne 40 adultes anglophones en formation continue, 10 natifs francophones et 4 tuteurs, répartis en 4 groupes : Aquitania, Gallia, Lugdunensis et Narbonensis). La formation s'est déroulée sur 10 semaines et a produit plus de 12 000 échanges dont 2686 messages de forum, 4062 courriels et 5680 tours de chat.

5.1.2. Volume des échanges entre individus

Pour mesurer l'importance (ici quantitative) des échanges entre les individus, nous considérons dans cette partie un graphe orienté et valué « de type 1 » qui est construit de la manière suivante : les sommets sont les membres du groupe, et chaque arc (i,j) entre deux sommets i et j est valué par le nombre d'unités (messages, actes de langage, mots, caractères, ...) postées par i et ouvertes par j.

Sur la figure 6, chaque nœud représente un membre du groupe (Aquitania). Son rôle est précisé par la seconde lettre. Ainsi At, An* et Al* représentent respectivement un tuteur, des natifs ou des apprenants (*learners*). Un suffixe numérique permet de distinguer un acteur de ses pairs. On peut constater sur cette figure que le sommet Al1 (apprenant n°1 de Aquitania, en haut à gauche sur la figure) n'est connecté à aucun autre ; ce qui traduit son absence totale d'échange avec les autres dans ce forum. Les graphes qui suivent ont été produits avec GraphViz issu de (Gansner et North, 2000)

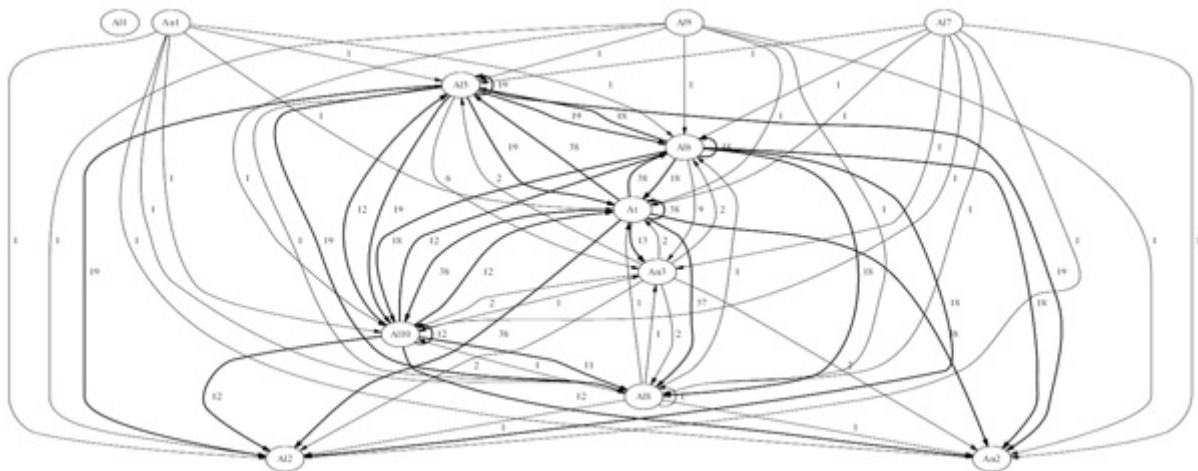


Figure 6 • Représentation graphique d'un graphe orienté et valué : (type 1) nombre de messages échangés dans le forum de socialisation au cours de la première quinzaine de la formation *Simuligne*.

Les acteurs Al2 et An2 (en bas à gauche et à droite) se distinguent par le fait qu'ils n'ont émis aucun message, mais ils ont ouvert ceux des autres. Les trois sommets An1, Al9 et Al7 (en haut de la figure) représentent des individus ayant posté un seul message et n'en ayant lu aucun. La représentation matricielle donnée à la figure 7 permet une meilleure lisibilité et une analyse plus fine des échanges.

Sur le tableau donné à la figure 7, la valeur 13 située sur la dernière ligne (At) et dans l'avant-dernière colonne (An3) signifie que 13 des messages postés par le tuteur (At) ont été lus par le natif (An3). On y voit en particulier que ceux qui ont posté de nombreux messages (Al5, Al6, Al10 et At) sont ceux qui ont lu quasiment tous les messages. Le cas de Al2 est typiquement ce qui est appelé un *learner* qui lit tout sans laisser de traces visibles aux autres, donc ne poste aucun message.

	A11	A12	A15	A16	A17	A18	A19	A110	An1	An2	An3	At
A11	0	0	0	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	0	0	0	0	0	0	0
A15	0	19	19	19	0	19	0	19	0	19	6	19
A16	0	18	18	18	0	18	0	18	0	18	9	18
A17	0	1	1	1	0	1	0	1	0	1	1	1
A18	0	1	1	1	0	1	0	1	0	1	1	1
A19	0	1	1	1	0	1	0	1	0	1	0	1
A110	0	12	12	12	0	11	0	12	0	12	1	12
An1	0	1	1	1	0	1	0	1	0	1	0	1
An2	0	0	0	0	0	0	0	0	0	0	0	0
An3	0	2	2	2	0	2	0	2	0	2	0	2
At	0	38	38	38	0	37	0	38	0	38	13	38

Figure 7 • Représentation matricielle du graphe de type 1 de la figure 6.

Après avoir présenté ce premier type de graphe mettant en exergue les relations interindividuelles, nous proposons une représentation plus proche des actions de postage et lecture des messages dans les forums en suggérant un type de graphe incluant deux sortes de sommets : les acteurs d'une part, et les classes de messages (pouvant être des forums par exemple) d'autre part.

5.1.3. Contribution des individus dans chaque catégorie

Dans cette partie, on s'intéresse davantage à la contribution de chaque participant à chacun des thèmes/sujets/forums ou à chaque catégorie définie par le chercheur.

Deux graphes représentent respectivement le postage (fig. 8) et la lecture des messages par les acteurs (fig. 9) pour s'intéresser aux relations que les individus entretiennent avec des classes de messages (définies par type, par thème, par activité, par période ou par fil de discussion ou par profondeur dans le fil par exemple) plutôt qu'aux relations interindividuelles.

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 C_j qui ont été postées par l'acteur A_i .

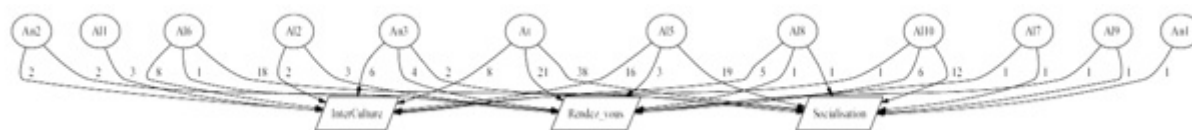


Figure 8 • représentation graphique du nombre de messages postés par chaque acteur dans les différents forums au cours de la première quinzaine de la formation Simuligne dans le groupe Aquitania.

postés	A11	A12	A15	A16	A17	A18	A19	A110	An1	An2	An3	At
Socialisation	0	0	19	18	1	1	1	12	1	0	2	38
Rendez-vous	0	3	3	1	1	1	1	6	0	2	4	21
InterCulture	3	2	16	8	0	5	0	1	0	2	6	8

Figure 9 • représentation matricielle des contributions des acteurs dans les différents forums du groupe Aquitania au cours de la première quinzaine de la formation *Simuligne*.

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 C_j qui ont été lues par l'acteur A_i .

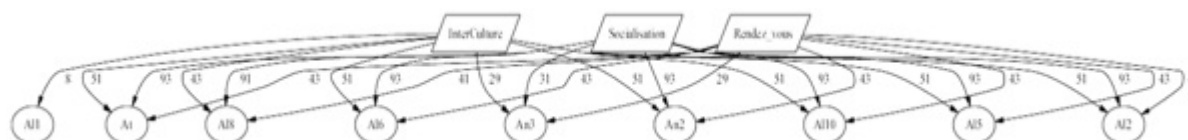


Figure 10 • représentation graphique du nombre de messages lus (ayant été déposés dans la première quinzaine).

lus	A11	A12	A15	A16	A17	A18	A19	A110	An1	An2	An3	At
Socialisation	0	93	93	93	0	91	0	93	0	93	31	93
Rendez-vous	0	43	43	43	0	41	0	43	0	43	29	43
InterCulture	8	51	51	51	0	43	0	51	0	51	29	51

Figure 11 • représentation matricielle du nombre de messages lus (ayant été déposés dans la première quinzaine).

5.1.4. Comment modéliser l'engagement ?

Puisque le premier modèle (fig. 6 et 7) constitue 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 tous moments. 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. S'il est jugé trop long, le délai de lecture d'un message (s'il est connu) peut être utilisé pour réduire l'évaluation de l'engagement d'un acteur. Le fait de répondre à un message peut s'entendre 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 donc, en pondérant les différents types de 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 (fig. 8 à 11) rend compte plus finement des contributions (en écriture ou en lecture) pour les différentes classes de messages, pour chacun des acteurs, mais ne rend pas compte 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 hypothèse 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 le thème, le type du message lu ou posté ou plus généralement la classe à laquelle ce message appartient, est prioritaire dans l'engagement de l'apprenant à le lire, à l'écrire, ou à y répondre.

Tout message étant daté, on peut bien sûr décider de ne prendre en compte que les messages ayant été postés 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.

Il convient de rappeler ici que le graphe est un indicateur des classes 3 ou 4 (en fonction des informations qu'il contient). Dès lors qu'il est disponible, il peut servir à calculer des indicateurs de la classe 5 tels que : la cohésion d'un groupe, la centralité d'un individu ou son intermédiation dans les communications du groupe.

5.2. Espoirs et limites des premières propositions

Pour les avoir étudiés en détail dans ([Reffay et Chanier, 2003a](#)) et ([Reffay et Chanier, 2003b](#)) 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 la dynamique d'un groupe en FAD. Nous avons aussi montré dans ([Reffay et Chanier, 2002](#)) qu'il était difficile de définir une mesure susceptible de représenter la cohésion du groupe. L'analyse de la cohésion dans un groupe de téléformation collaborative n'est pas une fin en soi mais sert à caractériser la position des membres par rapport au reste du groupe. Une simple valeur réelle n'apporte pas ce type d'information. En revanche, les calculs basés sur les graphes de communications tels que ceux présentés plus haut, offrent des représentations globales (sur l'ensemble des interactions considérées), simples (dépourvues des nombreux éléments quantitatifs utiles à leur construction : *data mining*) et pertinentes (permettant un repérage objectif et systématique des positions occupées par les différents acteurs). Pour faire un bon usage des modèles proposés par les réseaux sociaux, le tuteur ou le chercheur n'ont pas besoin de maîtriser la théorie des graphes, ni bien sûr les algorithmes permettant d'en extraire les indicateurs tels que la centralité ou la cohésion, mais nous ne pouvons qu'insister sur le fait que :

Ce qui est fondamental dans le bon usage de ces indicateurs, ce sont les choix de construction du graphe des relations (considérées dans leur contexte) pour pouvoir donner un sens pédagogique aux représentations qui en seront extraites automatiquement.

6. Deux approches à unifier

Les modèles quantitatifs proposés ici pour l'analyse des interactions sont issus de recherches sur des forums de natures assez différentes et concernant des groupes de tailles très variables. D'un côté, on s'intéresse à des forums ouverts, souvent de grande taille, centrés sur des thèmes assez précis et concernant un public en perpétuel renouvellement. De l'autre, on considère des forums en contexte éducatif avec des groupes de petite taille, le plus souvent captifs, encadrés par un scénario de formation qui établit la durée de vie du forum. De plus, les analyses sont motivées par des enjeux très différents.

Dans les grands forums ouverts, on tente de mesurer l'usage moyen ou de dégager des profils récurrents (ex : initiateur de discussions, adepte des débats, contributeur de fonds ou perturbateur spammeur) à partir d'un très grand nombre de groupes (au moins 500) de contributeurs et de messages sur des durées importantes (de l'ordre de 6 mois). Les usages moyens et profils récurrents qui se dégagent des analyses statistiques sur ces grands nombres nous informent sur le contexte sociétal. Dans les petits forums fermés en contexte éducatif, les analyses plus fines mais aussi plus hétérogènes cherchent plutôt à inférer des informations pédagogiques (motivation, participation, engagement, sociabilité, etc.) à partir de modèles statistiques basés sur les interactions pour situer plus précisément la position de chaque membre dans le groupe. Le contexte précis et particulier de ces interactions permet une meilleure interprétation des modèles en les liant au scénario pédagogique. Elle promet donc le développement d'outils de suivi de la dynamique des groupes plus pertinents pour les acteurs de la téléformation.

Mais au delà de ces différences de contextes et d'objectifs, il nous semble extrêmement intéressant et important de constater que les indicateurs et les outils d'analyses sont presque identiques. Mieux encore, certaines caractérisations comme la signature d'un groupe (représentant la polarité d'influence par une courbure explicitant l'écart entre les contributeurs stables et les *passants*) trouvent un écho dans le calcul de la centralité et la cohésion d'un groupe de discussion en contexte éducatif. Le positionnement des individus par rapport au reste du groupe dans la structure de communication interne utilise le même type de représentation que l'analyse des interactions entre groupes thématiques basée sur le *cross-posting*. Nous avons aussi mis en évidence certains phénomènes de « self-similarité » à différentes échelles. Cette analyse permettrait par exemple de mettre en perspective le comportement des membres du groupe vis à vis de leurs contacts.

Finalement ces outils se rejoignent sur le plan technologique ou méthodologique. L'intérêt de l'analyse des grands groupes réside dans le fait que les résultats sont représentatifs compte tenu de la loi des grands nombres. Les contextes socioéducatifs considérés dans les deux approches sont bien sûr très différents. Ce dernier point est structurant car il va conditionner l'orientation interprétative des résultats. Mais cela vaudrait la peine d'appliquer certains outils (jusque-là réservés à la métrologie des grands groupes) à des contextes éducatifs plus restreints pour tirer partie du contexte et ainsi proposer des interprétations plus riches de ces résultats. Nous proposons que « notre outil » soit adapté, en particulier avec l'aide des praticiens, au contexte dans lequel il est utilisé. Puisqu'il apporte une vision synthétique des échanges, il serait particulièrement utile pour un coordinateur en charge de nombreux groupes en formation à distance. Détecter rapidement que le comportement d'un groupe n'est pas celui attendu ou qu'un individu a une position singulière permettrait d'attirer l'attention du coordinateur sur le groupe déviant ou celle du tuteur sur l'acteur singulier.

Malgré les différences de contextes, nous défendons l'idée que certains indicateurs, méthodes et outils employés dans ces deux types de recherche peuvent traverser les échelles faisant ainsi émerger un continuum entre les groupes de quelques personnes parfaitement identifiées ayant une tâche parfaitement encadrée, et les communautés de plusieurs milliers d'internautes anonymes (au premier abord) mettant à jour leurs connaissances à travers leurs discussions publiques. Ce continuum s'ancre aussi dans certaines motivations des individus comme l'acquisition de connaissance ou le besoin de reconnaissance qui sont communs à ces contextes collectifs. Si les paramètres de ces outils sont encore à ajuster selon le contexte, pour permettre de distinguer les observables, les capacités de synthèse et de calcul automatique qui caractérisent ces outils devraient nous aider à mieux comprendre les usages dans différents contextes. Nous pensons qu'ils peuvent également offrir des services aux acteurs de ces forums dans leur recherche de contact ou d'information, ou tout simplement dans leur perception de l'activité.

BIBLIOGRAPHIE

- BRIN S., PAGE L. (1998). "The Anatomy of a Large-Scale Hypertextual Web Search Engine". Technical Report of the Stanford University, 20p. Available online. <http://infolab.stanford.edu/pub/papers/google.pdf> (dernière consultation le 10 February 2007).
- CHANIER, T. (2001). « Créer des communautés d'apprentissage à distance », *Les dossiers de l'Ingénierie Educative*, n° 36, Centre National de Documentation Pédagogique, Montrouge, France, p. 56-59. Disponible sur Internet : <http://edutice.archives-ouvertes.fr/edutice-00000119> (dernière consultation le 10 février 2007).
- DEGENNE A., FORSÉ M. (1994). *Les réseaux sociaux*. Collection U, Série Sociologie, Armand Colin, Paris.
- FORD L. R., FULKERSON D. R. (1956). "Maximal Flow Through a Network", *Canadian Journal of Mathematics*, 8, p. 399-404.
- GANSNER E., NORTH S. (2000). "An open graph visualization system and its applications to software engineering", *Software - Practice and Experience* 30(11): p. 1203-1233.
- JENSEN C., FARNHAM S., DRUCKER S., KOLLOCK P. (2000). "The Effect of Communication Modality on Cooperation in Online Environments". In *Proceedings of CHI 2000*, The Hague, Netherlands March. Available online. <http://research.microsoft.com/scg/papers/dilemmachi2000.pdf> (dernière consultation le 10 February 2007).
- JONES Q., RAVID G., RAFAELI S. (2002). "An Empirical Exploration of Mass Interaction System Dynamics: Individual Information Overload and Usenet Discourse". In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, IEEE, Big Island, Hawaii. Available online. <http://modiin.njit.edu/docs/hicss%20on%20system%20dynamics.PDF> (dernière consultation le 10 février 2007).
- JONES Q. (2003). "Applying Cyber-Archaeology". In *Proceedings of the Eighth European Conference on Computer Supported Cooperative Work*, 14–18 September 2003, Helsinki, Finland. Kluwer Academic Publishers, Dordrecht Hardbound, ISBN 1-4020-1573-9. Available online. http://modiin.njit.edu/docs/jones_ecscw03.pdf (dernière consultation le 10 février 2007).
- LANCIERI L. (1999). Description of Internet user behaviour. *IEEE International Joint Conference on Neural Network 99*. Washington, USA, 6 p.
- LANCIERI L. (2000). "A connectionist approach for evaluating the complexity of interactions in the World Wide Web. The case of News Groups". *IEEE International Joint Conference on Neural Network 2000 IJCNN 2000- (COMO - Italy)*, 5 p. Available online. http://www.ecole.ensicaen.fr/~lancieri/public_html_fichiers/ijcnn00.pdf (dernière consultation le 10 février 2007).
- LANCIERI L. (2004). "Reusing Implicit Cooperation, A novel approach to knowledge management", In *tripleC (Cognition, Cooperation, Communication) International Journal*, 2004 p. 28-46. Available online. http://www.ecole.ensicaen.fr/~lancieri/public_html_fichiers/tripleC_Lancieri.pdf (dernière consultation le 10 février 2007).
- LANCIERI L. (2005). *Interactions humaines dans les réseaux*. Edition Hermes ISBN : 2746211084
- LAVALLARD A., LANCIERI L. (2004). « Observation de l'évolution des communautés d'intérêts ». *IC 2004*, 15^{ème} journées francophones d'ingénierie des connaissances, Lyon mai. Disponible sur Internet. http://www.ecole.ensicaen.fr/~lancieri/public_html_fichiers/IC2004.pdf (dernière consultation le 10

février 2007)

MALTZ D. A. (1994). "Distributing Information for Collaborative Filtering on Usenet Net News". SM Thesis, Massachusetts Institute of Technology, Cambridge, MA. Available as MIT/LCS/TR-603 and Xerox PARC CSL-94-5. Available online. <http://www.lcs.mit.edu/publications/pubs/pdf/MIT-LCS-TR-603.pdf> (dernière consultation le 10 février 2007).

REFFAY C., CHANIER T., NICOLET J. (2002). « Produire ensemble pour apprendre : expérience d'une simulation globale en ligne ». Actes du Colloque national Apprendre avec l'Ordinateur, Bordeaux, France, Janvier 2002, p. 24.

REFFAY C., CHANIER T. (2002). "Social Network Analysis Used for Modelling Collaboration in Distance Learning Groups". Intelligent Tutoring System Conference, Biarritz, France, 2002, p. 31-40. Available online. <http://edutice.archives-ouvertes.fr/edutice-00000056> (dernière consultation le 10 février 2007).

REFFAY C., CHANIER T. (2003). « Mesurer la cohésion d'un groupe d'apprentissage en formation à distance ». Environnement Informatique pour l'Apprentissage Humain. (EIAH'03), Strasbourg, Avril 2003, p. 367-378. Disponible sur Internet. <http://edutice.archives-ouvertes.fr/edutice-00000152> (dernière consultation le 10 février 2007).

REFFAY C., CHANIER T. (2003). "How social network analysis can help to measure cohesion in collaborative distance learning?" Proceeding of the Computer Supported Collaborative Learning Conference (CSCL'03), Bergen, Kluwer Academic Publishers, p. 343-352. Available online. <http://edutice.archives-ouvertes.fr/edutice-00000422> (dernière consultation le 10 février 2007).

SAITO Y., MOGUL J., VERGHESE B. (1998). A Usenet Performance Study. Sep 1998. Project report; HP Labs. Available online. <http://citeseer.ist.psu.edu/saito98usenet.html> (dernière consultation le 10 février 2007).

SCOTT J. (2000). *Social Network Analysis : a handbook*. 2nd edition, SAGE, London.

SMITH, M. (1999). "Invisible Crowds in Cyberspace: Measuring and Mapping the Social Structure of USENET", In *Communities in Cyberspace*, edited by Marc Smith and Peter Kollock. London, Routledge Press. Available online. <http://research.microsoft.com/~masmith/Invisible%20Crowds%20in%20Cyberspace.doc> (dernière consultation le 10 février 2007).

VIÉGAS F. B., SMITH M. (2004). "Newsgroup Crowds and AuthorLines: Visualizing the Activity of Individuals in Conversational Cyberspaces". Proceedings of the 37th Hawaii International Conference on System Sciences. Available online. <http://research.microsoft.com/~masmith/Newsgroup%20Crowds%20and%20Author%20Lines.pdf> (dernière consultation le 10 février 2007).

WASSERMANN S., FAUST K. (1994). *Social Network Analysis : Methods and Applications*. Cambridge University Press, New York

WHITTAKER S., TERVEEN L., HILL W., CHERNY L. (1998). "The dynamics of mass interaction". In Proceedings of Conference on Computer Supported Cooperative Work, p. 257-264. New York, ACM Press. Available online. <http://dis.shef.ac.uk/stevewhittaker/cscw98-published.pdf> (dernière consultation le 10 février 2007).

YAICHE, F. (1996). *Les simulations globales mode d'emploi*. Hachette, Paris, 1996, 129 p.

Références à des sites Internet

International Network for Social Network Analysis (INSNA): <http://www.insna.org/> (dernière consultation le 11 février 2007).

([Mulce, 2007](#))

Multimodal contextualized Learning Corpus Exchange (MULCE) : <http://mulce.univ-fcomte.fr/> (consulté le 11 février 2007).

Usenet social accounting search engine, Microsoft: <http://netscan.research.microsoft.com/> (dernière consultation le 11 février 2007).

■ A propos des auteurs

Christophe REFFAY est maître de conférence en informatique et chercheur au Laboratoire d'Informatique de l'université de Franche-Comté. Il s'intéresse à l'analyse des interactions dans les situations collaboratives de formations/apprentissage à distance et en ligne et en particulier aux modèles issus des réseaux sociaux. Plus récemment, il s'est aussi préoccupé du recueil, de l'indexation et de la structuration des données issues de telles expérimentations, en vue d'en constituer des corpus échangeables avec des chercheurs extérieurs à l'expérimentation.

Adresse : Laboratoire de l'Informatique de Franche-Comté (LIFC), Équipe SICAH, 16, route de Gray, F-25030 BESANCON CEDEX

Courriel : Christophe.Reffay@univ-fcomte.fr

Toile : <http://lifc.univ-fcomte.fr/~reffay/>

Luigi LANCIERI est chercheur à France Telecom R&D et professeur associé à l'ensi-université de Caen. Une part importante de son activité est liée à la compréhension du facteur humain dans les réseaux informatiques à partir de traces d'activités. Il considère les interactions selon différents points de vue tels que les aspects informationnels (e.g. modélisation statistique de l'intelligence collective) ou les aspects physiques (e.g. modélisation de la mobilité ou de la localité).

Adresse : Orange-France Telecom research labs (FT R&D), 42 rue des Coutures, 14000 Caen, France

Courriel : luigi.lancieri@orange-ftgroup.com

Toile : <http://www.ecole.ensicaen.fr/~lancieri>

Référence de l'article :

Christophe REFFAY, Luigi LANCIERI, Quand l'analyse quantitative fait parler les forums de discussion, *Revue STICEF*, Volume 13, 2006, ISSN : 1764-7223, mis en ligne le 02/03/2007, <http://sticef.org>

© Revue Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation, 2006