

---

# Stabilité globale et diversité locale dans la dynamique des commentaires de Flickr

Stéphane Raux<sup>\*,\*\*</sup> — Christophe Prieur<sup>\*\*</sup>

\* LIP6, Université Pierre et Marie Curie  
104 avenue du Président Kennedy, 75016 Paris

\*\* LIAFA, Université Paris-Diderot  
175 rue du Chevaleret, Case 7014, 75205 Paris Cedex 13  
{raux, prieur}@liafa.jussieu.fr

---

*RÉSUMÉ.* Nous analysons les dynamiques du réseau des commentaires du site Flickr à travers l'évolution de la structure globale du réseau, mais aussi celle des voisinages locaux et la manière dont ils se constituent. Nous constatons que la structure générale ne change pas de manière significative si l'on ne considère que les liens réciproques ou si on limite la durée de vie des relations.

*Les commentaires ont une forte dimension locale, avec une majorité d'échanges avec des voisins ou des voisins de voisins. Nous observons aussi une grande diversité des voisinages locaux et de leur mode de constitution. Les individus qui ont le plus de voisins choisissent ceux-ci dans leur voisinage proche, et ils se distinguent par leur activité très intensive, avec très peu de périodes d'interruption.*

*ABSTRACT.* We analyse the dynamics of the network of comments between users on Flickr website through the evolution of the overall structure of the network, but also through the local neighbourhoods and the way they are set up. We observe that there is no significant change in the overall structure of the network if we only consider the reciprocal links or if we define a lifetime limit for the relationships.

*We observe strong locality within the exchange of comments. There is also a great diversity within local neighbourhoods and the way they form and evolve. Individuals who have most neighbours are choosing them within their neighbour's neighbourhoods, and have in most cases a very intensive activity, with very few intermissions.*

*MOTS-CLÉS :* Grands réseaux, réseaux sociaux, dynamique des graphes, Flickr, réseaux égocentrés, web 2.0.

*KEYWORDS:* Complex networks, social networks, graph dynamics, Flickr, ego-centered networks, web 2.0.

## Introduction

En quelques années, l'analyse des réseaux sociaux est passée d'un courant de recherche en sciences sociales (Scott, 1992; Degenne *et al.*, 1994) à un outil central dans le développement des plus récentes plateformes du Web. Celles-ci s'appuient sur les interactions de leurs utilisateurs pour proposer des services qui couvrent aujourd'hui de nombreux domaines allant de l'organisation de l'information au commerce en ligne, en passant par la publication de photographies ou des services de rencontre ou de divertissement.

Le succès de ces plateformes, le volume très important de données qu'elles rassemblent et leur capacité à les organiser en réseau en font des sujets de recherche privilégiés (Boyd, 2004; Cardon *et al.*, 2008), tout comme l'était le World-Wide Web dans la décennie précédente, où les travaux concernant la recherche d'information sur ce vaste corpus documentaire (Efe *et al.*, 2000) ont abouti non seulement à la mise au point du *PageRank* de Google (Brin *et al.*, 1998) mais également à la création d'un nouveau champ de recherche à partir de la découverte des propriétés que partage un grand nombre de grands réseaux issus de disciplines aussi variées que la biologie, la linguistique ou l'économie (et bien entendu l'informatique et la sociologie). Ce sont ces grands réseaux qu'on appelle grands réseaux d'interactions, graphes de terrains ou *complex networks* en anglais (Barabasi *et al.*, 1999; Watts *et al.*, 1998; Newman, 2003; Bornholdt *et al.*, 2003).

L'essor des technologies de communication permet désormais de disposer de grands jeux de données horodatées concernant des réseaux ad-hoc, pair à pair ou certains réseaux de communication issus de la téléphonie ou du web participatif. La multiplication de ces données rend l'analyse statique de ces réseaux peu pertinente, et encourage les chercheurs à s'intéresser à leurs dynamiques. Ces études peuvent avoir des objectifs différents : on peut s'intéresser aux événements qui interviennent dans le réseau et voir comment la structure contraint leur diffusion, ce qui débouche sur des champs d'application comme l'épidémiologie ou le marketing viral (Watts *et al.*, 2007), mais on peut aussi étudier comment l'activité sociale crée du réseau à travers l'évolution de la structure des graphes et des groupes qui les composent (Palla *et al.*, 2007). On peut enfin s'inspirer de ces différentes approches pour modéliser la constitution des réseaux sociaux.

La plupart de ces études s'intéressent surtout aux caractéristiques globales des réseaux et à l'identification de groupes, mais on ne peut imaginer décrire avec finesse l'évolution d'ensemble d'un réseau sans porter le regard sur les innombrables dynamiques interindividuelles qui animent les acteurs de ce réseau. L'individu pris dans des chaînes d'interdépendances, tel que le décrit Norbert Elias (Elias, 1991), n'est pas seulement l'atome constitutif du réseau, il en est aussi la raison d'être. Comprendre comment l'individu peut agir malgré ou grâce aux liens qui l'entourent occupe depuis longtemps les sociologues. Tant la force des liens faibles de Granovetter que les trous structuraux de Burt montrent que le réseau peut être utilisé et ses liens mobilisés pour faciliter la recherche d'emploi (Granovetter, 1978) ou pour augmenter la produc-

tivité (Burt, 1992). Dans une perspective moins utilitariste et plus descriptive, toutes les recherches sur les réseaux égocentrés s’attachent à montrer la grande diversité des formes de sociabilité au travers des combinaisons de liens entre un individu, *ego*, et son entourage (Wellman, 1993; Gribaudo, 1998; Prieur *et al.*, 2009).

Notre travail a pour objectif de mieux comprendre les mécanismes de constitution des liens et d’évolution de la structure d’un réseau, en adoptant deux échelles d’analyse : d’une part celle du graphe dans son ensemble, en étudiant l’évolution de ses principales caractéristiques, et d’autre part celle des sommets du graphe, en s’intéressant en particulier à la structure et aux modalités de formation de leur voisinage.

L’étude présentée ici porte sur l’évolution du réseau des commentaires du site Flickr.com, un site de partage de photos et de vidéos qui compte parmi les sites phares de ce « nouveau Web » évoqué plus haut, dit « Web 2.0 » (O’Reilly, 2005). Lancé en février 2004, Flickr a rencontré un très grand succès, qui lui vaut d’être racheté par Yahoo! en mars 2005. Le réseau formé par ses utilisateurs a fait l’objet de plusieurs études, qu’il s’agisse d’analyser les usages de ses fonctionnalités (Lerman *et al.*, 2006; Beuscart *et al.*, 2009) ou d’en étudier la structure générale et son évolution (Kumar *et al.*, 2006; Mislove *et al.*, 2008). Leskovec *et al.* (2008) s’intéressent aux interactions locales pour affiner les modèles de croissance de la structure générale. Cha *et al.* s’intéressent à la diffusion de l’information à travers les liens de contact (Cha *et al.*, 2008; Cha *et al.*, 2009).

Nous consacrons la première partie de cet article à la présentation des données et des méthodes que nous avons mises en œuvre pour leur analyse. Nous étudions ensuite dans la deuxième partie la structure générale du graphe, en comparant les résultats obtenus avec différentes méthodes de construction de notre graphe. Nous constatons qu’après une courte période de mise en place, la structure est remarquablement stable, quelle que soit la méthode choisie. Nous mettons aussi en évidence l’importance des relations de courte distance au sein du graphe, ce qui nous amène à étudier dans la troisième partie la manière dont les sommets construisent leur entourage. Nous y montrons l’importance de l’origine des nouveaux voisins en distinguant le cas des voisins choisis parmi les « voisins de voisins » et en montrant les conséquences que ces choix peuvent avoir sur la structure du réseau local. Enfin, la quatrième partie est consacrée à la mesure de l’activité des sommets au fil du temps, et montre que la différenciation la plus pertinente ne concerne pas la durée totale d’activité, mais son intensité, en distinguant les sommets qui ont une activité continue de ceux qui ont une activité occasionnelle.

## 1. Données, modèles et outils

### 1.1. La base de commentaires Flickr

Nous avons travaillé à partir d’une extraction systématique réalisée en utilisant l’API publique de Flickr en août 2006 (Beuscart *et al.*, 2009). Nous disposons donc

d'un jeu de données « exhaustif » qui représente plus de 500 millions de photographies et près de 5 millions de membres.

Notre travail porte uniquement sur les commentaires de cette base de données. Un commentaire est toujours écrit par un utilisateur, que nous appellerons *émetteur*, sur la photographie d'un utilisateur que nous appellerons *destinataire*. Les messages ainsi enregistrés constituent un fil de discussion qui s'affiche en dessous de la photographie, ce qui permet aux utilisateurs d'interagir au sujet de la photographie. L'émetteur et le destinataire peuvent donc être une même personne, si l'auteur d'une photographie choisit de répondre aux commentaires déposés sur celle-ci.

Nous avons simplifié la structure de graphe biparti reliant des utilisateurs à des photographies pour nous concentrer sur les interactions entre les individus : nous n'avons conservé que l'identifiant de l'émetteur, l'identifiant du destinataire et le *timestamp* correspondant au moment où le commentaire a été écrit, à la seconde près. Nous avons par ailleurs retiré les commentaires dans lesquels l'émetteur et le destinataires sont la même personne. Un nœud est donc un internaute qui échange un commentaire ou qui a déposé une photo qui a été commentée sur le site Flickr. Nous disposons d'une base de 39 594 157 commentaires qui ont été échangés entre mars 2004 et juillet 2006 par 910 454 utilisateurs.

Cette simplification nous fait perdre une partie de l'information : on ne sait pas si plusieurs commentaires adressés à la même personne ont été déposés sur la même photographie, et il peut arriver que les commentaires prennent la forme d'un fil de discussion dans lequel les commentateurs se répondent entre eux. On pourrait alors considérer qu'il y a aussi interaction entre deux utilisateurs s'ils commentent la même photographie, mais ce serait abusif car en pratique une grande partie des commentaires restent très lacunaires et sont destinés en premier lieu au propriétaire de la photographie (e.g. : « *Nice pic !* »). Notre approche est donc réductrice, mais elle nous assure que l'émetteur du commentaire a voulu interagir avec le propriétaire de la photographie, sans avoir besoin de faire de suppositions sur l'existence d'autres destinataires.

## 1.2. Formalisation

Les informations issues de la base de commentaires se présentent sous la forme d'une liste de liens dirigés classée par ordre chronologique de la forme  $(e, d, t)$ , où  $e$  est l'identifiant de l'émetteur du commentaire,  $d$  celui du destinataire, et  $t$  le *timestamp* du commentaire.

Pour prendre en compte la dimension dynamique de nos données, on définit un intervalle de temps discret  $T$ , et pour tout  $t \in T$  le graphe non orienté  $G_t = (V, E_t)$ , où  $V$  est l'ensemble des nœuds et  $E_t$  l'ensemble des relations entre deux nœuds  $(u, v)$  qui « existent » à l'instant  $t$ . On peut adopter différentes stratégies pour déterminer si une relation existe ou non. Nous considérerons dans cet article que les commentaires sont cumulatifs : un lien entre  $u$  et  $v$  est considéré comme existant à l'instant  $t$  si les deux nœuds ont déjà échangé au moins un commentaire à un instant  $t' \leq t$ . Il est

aussi possible de limiter cette condition en considérant que les relations disparaissent si aucun commentaire n'est réémis entre deux nœuds au-delà d'un intervalle de temps donné. Nous étudierons ce cas dans la partie 2.2.

Nous définissons par ailleurs le graphe  $G = (V, E)$ , qui contient l'ensemble des commentaires. Plus précisément :  $E = \bigcup_{t \in T} E_t$

Le *voisinage*  $N_t(u)$  dans  $G_t$  (*resp.*  $N(u)$  dans  $G$ ) d'un nœud  $u$  est l'ensemble de nœuds  $v$  tels que  $(u, v) \in E_t$  (*resp.*  $(u, v) \in E$ ). Le *degré*  $\text{deg}_t(u)$  dans  $G_t$  (*resp.*  $\text{deg}(u)$  dans  $G$ ), est le nombre d'éléments dans  $N_t(u)$  (*resp.* dans  $N(u)$ ). La *distance* entre deux sommets  $u$  et  $v$ , notée  $\text{dist}_t(u, v)$  est la longueur du plus court chemin (suite d'arêtes adjacentes) entre  $u$  et  $v$  dans  $G_t$ . S'il n'existe pas de chemin, on notera  $\text{dist}_t(u, v) = \infty$ . Le *voisinage à distance 2*, noté  $N_t^2(u)$  dans  $G_t$  (*resp.*  $N^2(u)$  dans  $G$ ) d'un sommet  $u$  désignera l'ensemble des sommets qui sont des voisins de voisins de  $u$  sans être compris dans  $N_t(u)$  (*resp.*  $N(u)$ ).

Si le sommet  $v$  est un voisin de  $u$  dans  $G_t$ , on dira que c'est un *voisin proche* s'il existe un instant  $t' < t$  tel que  $\text{dist}_{t'}(u, v) = 2$ , et un *voisin lointain* sinon. Nous notons  $\bar{N}_t(u)$  l'ensemble des voisins proches de  $u$  dans  $G_t$ , que nous appelons *voisinage proche* de  $u$ . On notera  $P_t(u)$  la *proportion de voisins proches* de  $u$  dans  $G_t$  (*resp.*  $P(u)$  dans  $G$ ) :  $P_t(u) = \frac{|\bar{N}_t(u)|}{|N_t(u)|}$ .

### 1.3. Mesures

Rappelons qu'une *composante connexe* est un ensemble de sommets qui sont connectés deux à deux par au moins un chemin. Leur identification est surtout intéressante pour identifier la composante connexe principale : les graphes de terrain se caractérisent en effet par l'existence d'une très grande composante connexe qui contient une grande proportion des sommets, puis d'un grand nombre de composantes connexes contenant très peu de sommets (Albert *et al.*, 2002). En pratique, la composition des composantes connexes peut se calculer à la volée au moment du chargement du graphe au moyen d'un algorithme *union-find*.

On définit habituellement la *périphérie* d'un graphe de manière constructive en sélectionnant itérativement tous les sommets qui ne possèdent qu'un voisin et en mettant à jour les degrés de ceux-ci, ce qui revient à désigner le plus grand ensemble d'arbres induits dans le graphe. Cette notion est très utilisée en analyse de réseaux sociaux, en particulier dans le cas de la périphérie de la composante connexe principale, car elle identifie des sommets qui ont une position « marginale », mais qui restent malgré tout connectés à une grande partie du graphe (Seidman, 1983). La définition constructive de la périphérie revient à effectuer un parcours en largeur en partant des sommets de degré 1 et en ne visitant que les sommets qui n'ont qu'un seul voisin de degré 1 non encore visité.

Le nombre de *triangles* dans le graphe (*i.e.* de triplets de sommets tous trois connectés entre eux) permet de mesurer les densités locales dans  $G$ , en particulier

le *coefficient de clustering*, qui désigne la probabilité pour que deux voisins  $u$  et  $v$  d'un même sommet soient également en relation. On parcourt pour chaque sommet l'ensemble de ses voisins et on compte leurs contacts communs. Ce calcul se fait en un temps très rapide en pratique (voir Latapy (2008) pour un survol des méthodes de calcul des triangles).

#### 1.4. Méthodes

Ces mesures peuvent être effectuées directement sur le graphe  $G$ , ce qui revient à mesurer l'état de  $G_t$  à la fin de la période  $T$  si l'on considère le cas où les liens sont cumulatifs. Les mesures peuvent porter sur l'ensemble du graphe : taille et nombre de composantes connexes, taille de la périphérie ou coefficient de clustering. Elles peuvent aussi s'appliquer à chaque sommet  $u$ , par exemple la taille du voisinage ( $N(u)$  comme  $N^2(u)$ ). Les informations qui sont ainsi recueillies permettent ensuite d'effectuer des recoupements et de mieux comprendre les variations de profils entre les différents sommets.

Pour mesurer l'évolution de ces mesures au fil du temps, il faut effectuer les calculs au fur et à mesure de la lecture des commentaires. En terme d'implémentation, on charge alors l'ensemble des liens de  $G$  en mémoire sans indiquer de temps, puis on parcourt la liste des commentaires en mémorisant pour chaque lien la date de sa dernière occurrence. Dans le cas cumulatif, on teste la présence d'un lien dans  $G_t$  en vérifiant si sa dernière occurrence est inférieure ou égale à l'instant  $t$ .

On affiche ensuite l'état de  $G_t$  à intervalles réguliers : si les mesures se calculent en temps linéaire avec une faible constante et qu'on choisit des intervalles de quelques jours, on peut simplement utiliser les algorithmes des graphes statiques sur  $G_t$  lors de chaque intervalle. En revanche, on ne peut pas calculer l'évolution du coefficient de clustering avec cette méthode, car il serait trop coûteux d'effectuer la mesure à de nombreuses reprises. On utilise alors un algorithme dynamique qui fait le décompte des nouveaux triangles à chaque fois qu'une nouvelle relation apparaît dans le graphe.

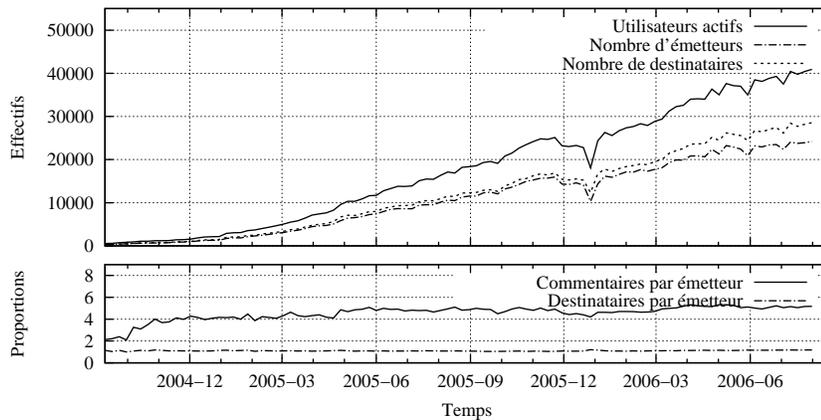
Ces mesures peuvent être calculées pour mesurer l'évolution de l'ensemble du graphe, mais on peut aussi choisir de suivre un sommet  $u$  et d'étudier l'évolution de ses caractéristiques. Cette deuxième méthode permet d'utiliser des méthodes plus coûteuses en temps car on peut alors limiter les calculs aux commentaires dans lesquels  $u$  apparaît.

Pour implémenter efficacement ces mesures dynamiques, il faut réduire au minimum les traitements de réinitialisation qui peuvent être nécessaires entre deux calculs. En fonction du type de mesure, on peut effectuer ces réinitialisations à la volée, lors de chaque étape de calcul ou bien on peut utiliser une pile pour ne réinitialiser que les valeurs qui ont été modifiées à l'étape précédente.

## 2. Dynamiques à l'échelle du graphe

### 2.1. L'évolution de l'activité

Une première approche consiste à évaluer l'évolution de l'activité des commentaires : pour chaque journée, nous avons mesuré le nombre de messages émis, le nombre d'utilisateurs différents qui ont émis un message et le nombre de destinataires différents auxquels les messages ont été adressés. Un utilisateur est considéré comme « actif » s'il a émis ou reçu au moins un message au cours d'une journée. L'évolution du nombre d'utilisateurs actifs est représentée sur la figure 1. L'activité est faible jus-



**Figure 1.** Évolution de l'activité sur l'ensemble de la période (courbe lissée).

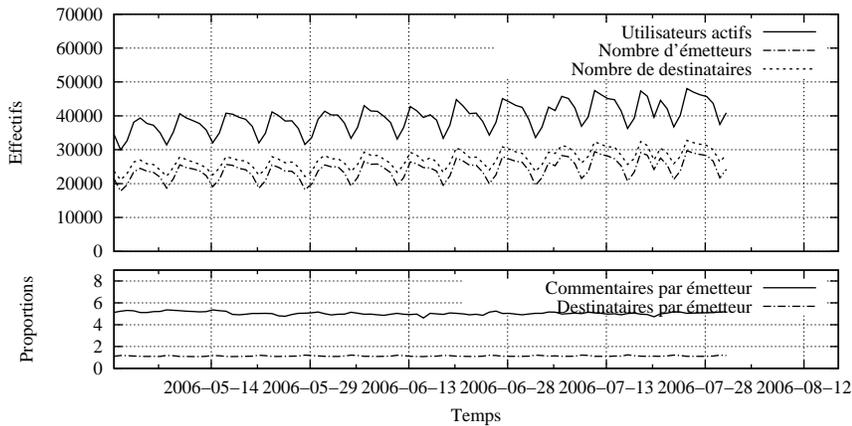
qu'en janvier 2005, ce qui correspond à une période de mise en place pendant laquelle le nombre moyen de messages par émetteur double, en passant de 2 à 4. L'activité croît ensuite de façon régulière tout au long de la période, à l'exception d'un décrochement au tournant de l'année 2005–2006. Le nombre de commentaires émis par rapport au nombre d'émetteurs augmente peu après la phase de mise en place, il se stabilise rapidement autour de 5 commentaires par émetteur et reste constant ensuite.

Le rapport entre le nombre de destinataires et le nombre d'émetteurs est lui aussi constant : il y a en moyenne 1,1 destinataires par émetteur. L'écart de cette moyenne avec celle des commentaires émis suggère que les utilisateurs ont tendance à concentrer leurs commentaires sur quelques utilisateurs. Cela peut correspondre au dépôt de commentaires sur plusieurs photographies d'un même utilisateur ou bien à l'échange de plusieurs commentaires sur une même photographie, dans le cadre d'un fil de discussion.

La constance de ces indicateurs montre que l'augmentation de l'activité des commentaires est liée uniquement à l'augmentation du volume des utilisateurs. Si l'on ne tient pas compte de la période de mise en place, ceux-ci ne sont pas plus prolifiques à

la fin de la période d'observation que lors du lancement du service : il y a un peu plus de 1 500 émetteurs par jour au début du mois de janvier 2005 et ils sont environ 20 fois plus nombreux à la fin du mois de juillet 2006, avec près de 30 000 émetteurs par jour.

Sans surprise, l'activité des utilisateurs varie en suivant des cycles d'une semaine. On observe un creux d'activité le dimanche, puis une reprise progressive de l'activité le lundi et un pic d'activité le mardi qui est suivi par une baisse progressive jusqu'à la fin de la semaine (cf. figure 2).

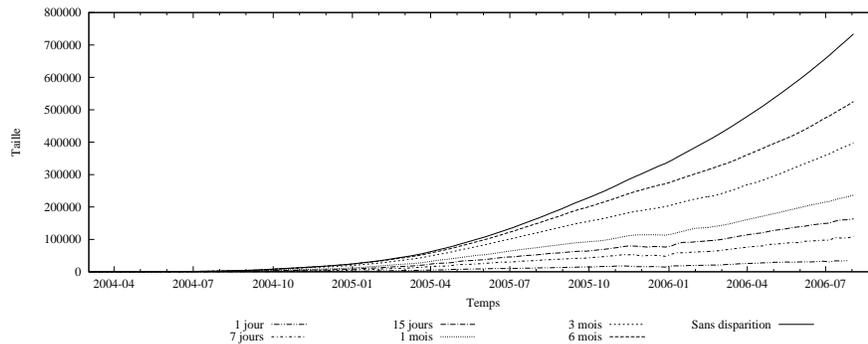


**Figure 2.** Évolution de l'activité pour les 3 derniers mois d'observation.

Ce premier survol permet d'observer deux phénomènes : l'élargissement du volume des commentaires, qui va de pair avec la popularité croissante du service, et la constance du comportement moyen des utilisateurs actifs. Pour comprendre comment cette activité se traduit en terme de relations, nous avons mesuré la structure générale du graphe et son évolution.

## 2.2. Structure générale du graphe

L'évolution de la taille de la composante connexe principale donne une bonne idée de l'accroissement de la taille du graphe car elle regroupe une grande partie de ses sommets. Nous avons d'abord construit notre graphe en adoptant le modèle cumulatif : une relation existe entre deux utilisateurs à partir du moment où ils ont échangé au moins un commentaire. On peut cependant considérer que dans un contexte dynamique, les relations perdent de leur importance si elles ne sont pas réactivées régulièrement. Le simple fait d'émettre un commentaire est un engagement peu coûteux pour un utilisateur, la relation ainsi créée doit être renouvelée régulièrement pour être significative.



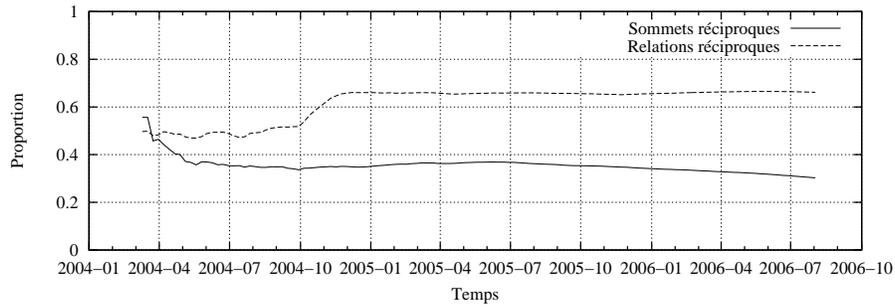
**Figure 3.** Évolution de la taille de la composante connexe principale en fonction de la méthode de construction du graphe.

On peut alors modifier le modèle de construction de  $G_t$  en fixant un délai d'activité au-delà duquel une relation est considérée comme abandonnée. Si à un moment donné la date de dernière activation d'une relation dépasse le délai fixé, on la retire du graphe. Par extension, on considère qu'un sommet  $u$  est actif dans  $G_t$  s'il est relié à au moins un autre sommet par une relation active au moment de la mesure, *i.e.* si  $\deg_t(u) \geq 1$ .

La figure 3 a été obtenue en utilisant la méthode de mesure dynamique de  $G_t$  telle que nous l'avons décrite dans la partie 1.4. La courbe la plus élevée correspond à l'accroissement de la composante connexe principale dans le cas où les liens s'ajoutent au fil des commentaires, les courbes inférieures ont été obtenues en choisissant différents délais d'activité, de six mois à une journée. La réduction du délai entraîne une réduction de la taille de la composante connexe principale, mais celle-ci augmente quel que soit le délai choisi. Ce n'est pas surprenant puisqu'on sait déjà que l'activité quotidienne augmente tout au long de la période.

Nous avons choisi de travailler sur un graphe non orienté par souci de simplification combinatoire. Les commentaires sont cependant orientés : un utilisateur peut émettre beaucoup de commentaires sans pour autant en recevoir, ou même sans avoir mis de photographie en ligne. De la même façon un utilisateur peut recevoir beaucoup de commentaires sans en émettre beaucoup en retour. L'étude de la distribution croisée des degrés entrant et sortant de chaque sommet montre qu'il existe en effet des profils dissymétriques, mais la forte corrélation entre le degré entrant et le degré sortant, en particulier pour les sommets de fort degré nous conforte dans l'idée que l'on peut avoir une vision précise de la structure générale du graphe en considérant la version non orientée. En revanche, cette simplification peut poser problème pour l'étude des dynamiques locales en raison de l'existence de profils dissymétriques.

Pour s'affranchir de ce problème on peut ajouter une contrainte de réciprocité dans les relations : on peut décider de ne conserver une relation entre deux sommets que s'il existe au moins un lien dans chaque direction pour cette relation, autrement dit si



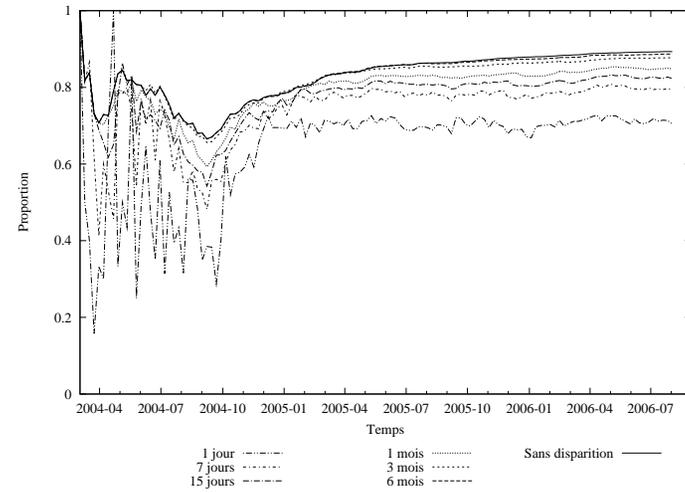
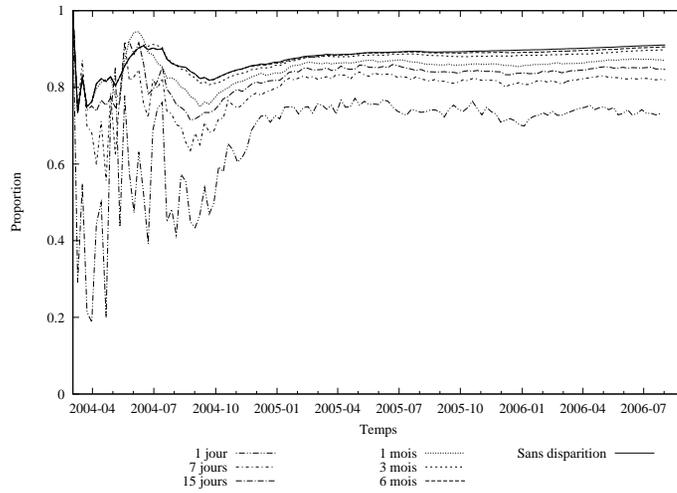
**Figure 4.** Évolution de la proportion de sommets et de relations entre le graphe de l'ensemble des liens et le graphe des liens réciproques.

chacun des sommets a été au moins une fois émetteur et récepteur. On élimine ainsi les relations unilatérales qui peuvent être considérées comme de moindre importance si l'on s'intéresse aux interactions entre utilisateurs. Le graphe des commentaires réciproques contient 259 395 sommets pour 25 818 794 commentaires. Cela signifie que les relations entretenues par moins d'un tiers (28,4%) des utilisateurs représentent près de deux tiers (65,2%) de l'ensemble des commentaires. Cette dissymétrie nous montre qu'il existe au sein du graphe une minorité d'utilisateurs dont l'activité interne constitue une part importante de l'ensemble des commentaires.

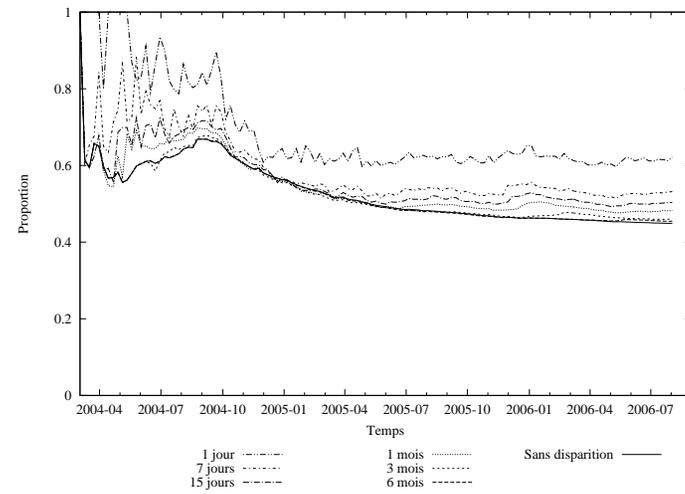
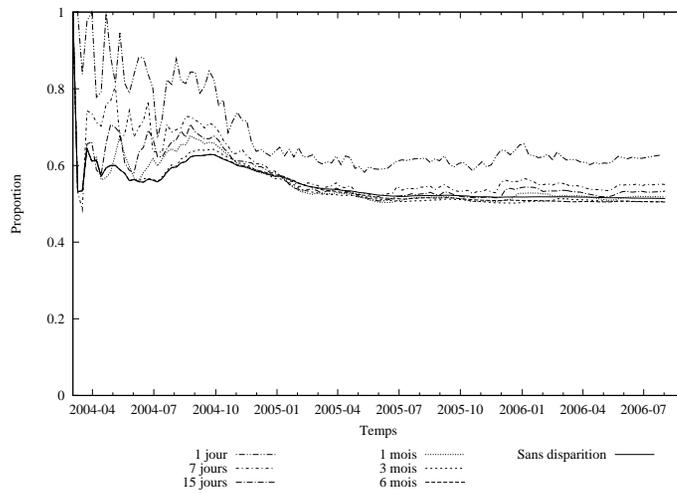
Si l'on compare l'évolution du nombre de sommets et de relations dans le graphe complet et le graphe des commentaires réciproques, sans suppression de lien (figure 4), on constate que ces proportions sont stables à partir du mois de décembre 2004, avec une légère baisse de la proportion de sommets qui entretiennent des relations réciproques.

Pour comparer les résultats obtenus avec les différentes méthodes de construction du graphe, nous rapportons les valeurs brutes obtenues pour la taille de la composante connexe principale au nombre de sommets actifs dans  $G_t$ . La figure 5 représente ainsi l'évolution de la taille relative de la composante connexe principale : les courbes de gauche concernent les mesures effectuées sur l'ensemble des commentaires, celles de droite correspondent aux commentaires réciproques. À titre de comparaison nous avons mesuré dans la figure 6 l'évolution de la taille relative de la périphérie de  $G_t$ .

On constate que les proportions sont très comparables, quels que soient le délai de suppression ou le graphe qu'on considère. Dans tous les cas, on retrouve l'évolution décrite dans Kumar *et al.* (2006) : une première période de mise en place de la structure du graphe, puis une stabilisation des proportions, toujours à partir de décembre 2004. La taille relative de la composante principale est alors en légère augmentation et reste proche de 85% des sommets actifs, la taille relative de la périphérie décroît légèrement en restant supérieure à 40%. Même en adoptant des délais de suppression très courts,



**Figure 5.** Évolution de la taille relative de la composante connexe principale pour l'ensemble des liens (à gauche) et pour les liens réciproques (à droite).



**Figure 6.** Évolution de la taille relative de la périphérie pour l'ensemble des liens (à gauche) et pour les liens réciproques (à droite).

de l'ordre de la semaine ou de la journée, ces proportions restent relativement stables, une fois passée la première période de constitution de la structure du graphe.

En effectuant la même expérience pour mesurer l'évolution du diamètre du graphe, nous constatons que celui-ci évolue lui aussi de manière très similaire quel que soit le mode de construction du graphe : le graphe de l'ensemble des liens a un diamètre qui augmente de manière régulière pour atteindre la valeur finale de 19 et celui des liens réciproques a un diamètre de 21. La distance moyenne est également légèrement plus grande dans le graphe des liens réciproques que pour le graphe de l'ensemble des liens (4, 8 contre 4, 5).

Ces résultats confortent nos premières observations sur l'activité. Ils montrent aussi que la structure relationnelle de notre graphe est remarquablement stable au fil du temps et qu'elle résiste bien aux contraintes de délai et de réciprocité des relations que l'on peut introduire dans la construction du graphe. Comme les commentaires réciproques représentent près des deux tiers de l'ensemble des commentaires, nous pouvons imaginer que l'intensité des relations entre ces sommets joue un rôle prépondérant dans l'élaboration de la structure du graphe : les commentaires non réciproques viennent grossir la taille du graphe mais ne modifient pas sensiblement sa structure.

La robustesse des caractéristiques générales du graphe permet de travailler sur des données plus « réalistes », qui tiennent compte de la durée de vie des relations sans craindre pour autant de modifier radicalement la structure des relations. Les graphes ainsi formés contiennent moins de sommets et de relations, ce qui permet d'accélérer sensiblement le temps de traitement des algorithmes de mesure. La suppression des relations ajoute cependant une difficulté dans le cas de la mesure des triangles, car il faut alors savoir à quel moment une relation disparaît pour décompter tous les triangles qui la contenaient. Cela signifie aussi que le degré des sommets peut croître et décroître, ce qui complexifie l'analyse de son évolution. Nous avons donc choisi pour des raisons de simplicité de limiter notre analyse au modèle de cumul des relations sans suppressions de liens pour le reste de l'article.

### **2.3. L'importance des phénomènes locaux**

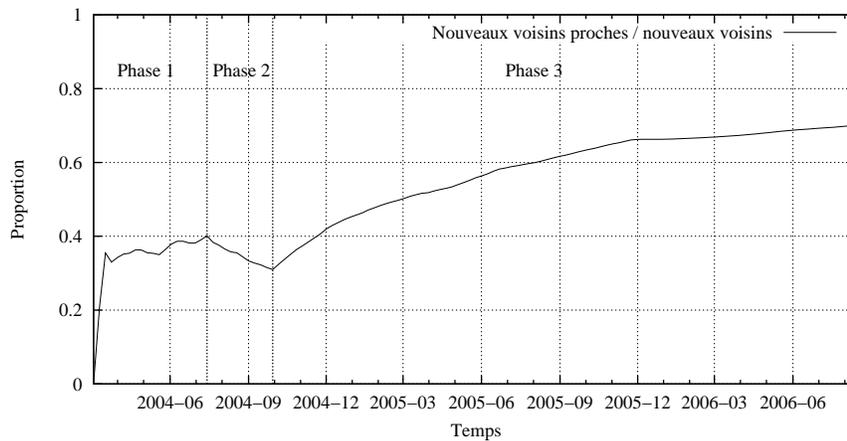
Le rôle important des relations réciproques nous montre que les utilisateurs ont tendance à échanger plus de messages avec les personnes avec qui ils sont déjà en contact. Pour évaluer l'importance de ces phénomènes locaux dans la structure du graphe, on mesure pour l'ensemble des commentaires la proportion de liens répétés et de nouvelles relations. La table 1 est obtenue en mesurant pour chaque commentaire émis à l'instant  $t$  la distance qui séparait l'émetteur du destinataire à l'instant  $t - 1$ . Si la distance est égale à 1, cela signifie que le lien est répété entre deux sommets qui étaient déjà voisins, sinon cela correspond à la création d'une nouvelle relation. Nous distinguons les nouvelles relations avec un voisin proche (distance = 2) des nouvelles relations avec un voisin lointain (distance  $\geq 3$ ). Nous regroupons les distances supérieures ou égales à 3 en raison de limitations de temps de calcul et parce que nous

considérons que les membres d'un réseau ont une vision réduite de leur entourage. Ils peuvent avoir conscience d'une partie des personnes qui se situent dans leur entourage à distance 2, mais de leur point de vue il n'y a pas de différence sensible entre des personnes situées à des distances de 3, 4 voire appartenant à une autre composante connexe : il s'agit dans tous les cas de personnes avec lesquelles ils ne partagent aucune connaissance commune.

Répétitions Distance = 1	Nouveaux voisins		Total
	Proches Distance = 2	Lointains Distance $\geq 3$	
29 946 674	6 781 686	2 865 797	39 594 157
75,6%	17,2%	7,2%	100%

**Tableau 1.** Répartition des commentaires en fonction du type de contact.

Les liens répétés représentent 75,6% de l'ensemble des commentaires et si l'on ajoute à cela les 17,2% de nouvelles relations avec des voisins proches, on constate que les commentaires sont échangés dans près de 93% des cas entre des utilisateurs qui sont déjà voisins ou qui ont au moins un voisin en commun. Une grande majorité de l'activité du réseau s'effectue donc sur de très courtes distances. Ces résultats sont cohérents avec ceux obtenus par Mislove *et al.* (2008), qui montrent que le modèle d'attachement préférentiel proposé par Barabasi *et al.* (1999) ne rend pas bien compte de la très forte proportion de voisins proches, et par Leskovec *et al.* (2008), qui soulignent l'importance de la transitivité lors de l'établissement de nouvelles relations.



**Figure 7.** Évolution de la proportion de nouveaux voisins proches par rapport à l'ensemble des nouveaux voisins.

On observe la même tendance si l'on s'intéresse seulement aux nouvelles relations : elles concernent dans 70% des cas des utilisateurs qui ont déjà au moins un

contact en commun. La figure 7 représente l'évolution de cette proportion sur l'ensemble de la période. On distingue trois étapes successives : d'abord une forte augmentation avec un pic de 40%, puis une baisse qui conduit à un creux fin septembre 2004. La proportion ne cesse ensuite de croître pour le reste de la période. On retrouve ici les trois étapes décrites dans Kumar *et al.* (2006) à propos de l'évolution de la densité. Ce n'est pas surprenant dans la mesure où un nouveau contact entre deux sommets à distance 2 entraîne la création d'un nouveau triangle, ce qui augmente la densité du réseau local. La proportion de voisins proches parmi les nouveaux liens peut donc être considérée comme un bon indicateur de la densité du graphe.

Le réseau des commentaires se constitue donc en grande partie à travers des liens courts, le plus souvent au sein du voisinage ou du voisinage à distance 2 des utilisateurs, avec une augmentation de cette tendance tout au long de la période, ce qui entraîne une densification locale du graphe. Si les fortes densités locales de réseaux « petit monde » sont un phénomène bien connu, la part très importante de relations répétées et de nouveaux voisins nous montre que les utilisateurs laissent en pratique peu de place à la rencontre de parfaits inconnus.

Cette tendance est surprenante car on s'attendrait à ce que les utilisateurs de Flickr privilégient des outils comme le moteur de recherche interne du site pour trouver de nouvelles photographies intéressantes et les commenter. On devrait alors observer des résultats qui correspondraient plus au modèle d'attachement préférentiel de Barabasi *et al.* (1999), avec une surreprésentation des thèmes les plus populaires. En pratique, la très forte redondance thématique des groupes de Flickr montre que les utilisateurs ne s'intéressent pas seulement au sujet des photographies, mais aussi à leur auteur (Beuscart *et al.*, 2009) : l'importance de la transitivité nous laisse penser que les utilisateurs privilégient plutôt les nombreux liens proposés par l'interface du site qui permettent de visiter les photographies des contacts et des personnes qui ont déjà déposé des commentaires. Cela correspond à la notion de *social browsing*, que l'on pourrait traduire par « navigation sociale » (Lerman *et al.*, 2006).

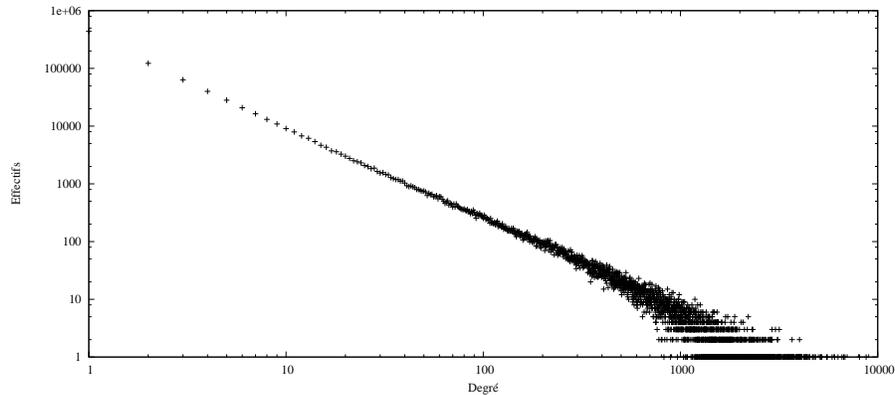
Ces informations nous permettent de mieux comprendre la structure du réseau des commentaires, mais elles ne proposent que des tendances d'ordre général : les valeurs moyennes obtenues sont en effet écrasées par le volume des commentaires, et ne peuvent pas rendre compte de la diversité des situations que l'on peut rencontrer au sein du réseau. Il est donc nécessaire de s'intéresser aux parcours individuels pour s'affranchir des effets de masse et mieux comprendre la variété des comportements.

### **3. Parcours individuels et profils de sommets**

#### **3.1. La composition de l'entourage relationnel**

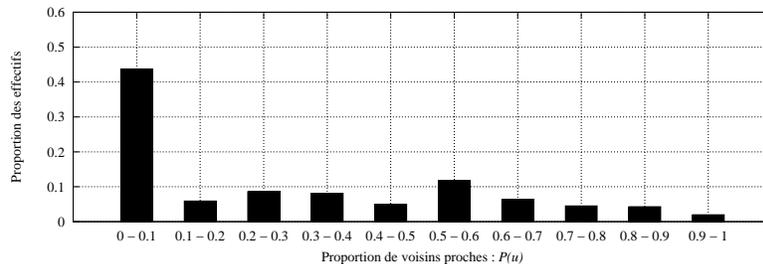
Notre objectif est de caractériser différents types de sommets et de comprendre comment ils construisent et organisent leur « réseau » de contacts. Une première étape consiste à mesurer pour chaque sommet son degré ( $\text{deg}(u)$ ) et la proportion de ses

voisins proches ( $P(u)$ ). Nous effectuons ces mesures sur le graphe de l'ensemble des commentaires, sans suppression de liens, que nous notons  $G$ .



**Figure 8.** *Distribution des degrés pour l'ensemble des commentaires.*

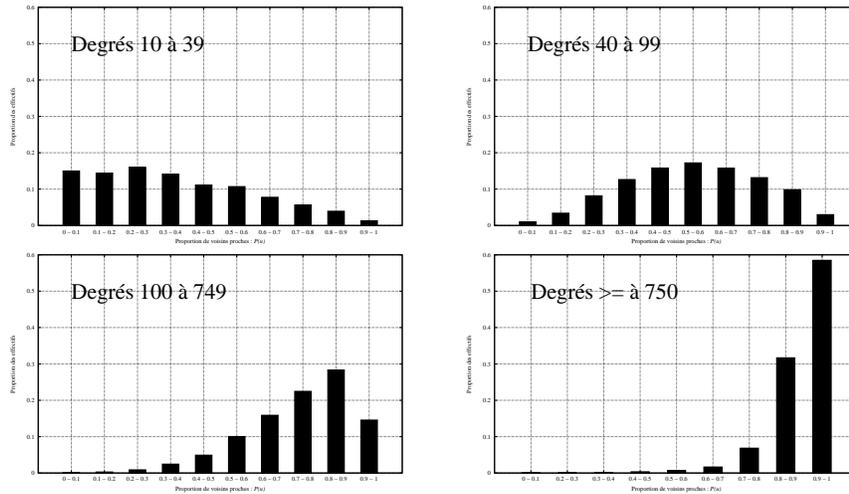
Les figures 8 et 9 montrent la distribution de ces deux indicateurs. Comme pour tous les graphes de terrain (dont les grands réseaux sociaux), la distribution des degrés est très hétérogène, il y a un très grand nombre de sommets de très faible degré (438 840 sommets de degré 1) côtoyant quelques sommets avec un degré très élevé (le degré maximum est de 8 731).



**Figure 9.** *Distribution des proportions de voisins proches pour l'ensemble des commentaires.*

Comme un sommet ne peut pas avoir de voisin proche s'il ne possède pas déjà un voisin pour servir d'intermédiaire, la distribution de la proportion de voisins proches ne concerne que les sommets qui ont au minimum 2 voisins. Cette distribution est par ailleurs difficile à apprécier car elle est en partie biaisée par la distribution des valeurs possibles. Par exemple, les sommets de degré 2 ne peuvent avoir que deux valeurs pour  $P(u)$ , qui sont 0% s'ils n'ont aucun voisin proche et 50% s'ils en ont un. Comme les sommets de faible degré sont les plus nombreux, cela conduit à une forte proportion

de sommets pour lesquels  $P(u)$  prend une valeur de 0% ou de 50%, ce qui explique en grande partie les deux pics que l'on observe sur l'histogramme pour les classes [0%–10%] et [50%–60%]. Si l'on écarte ces deux pics, on constate que l'intervalle [10%–49%] est plus peuplé que l'intervalle [60%–100%].



**Figure 10.** *Distributions des proportions de voisins proches en fonction des degrés. Les abscisses correspondent aux valeurs de  $P(u)$  et les ordonnées aux effectifs.*

Le graphe des commentaires semble donc être composé d'une majorité de sommets qui ont une faible proportion de voisins proches, ce qui semble contradictoire avec la proportion moyenne de 70% de voisins proches que nous avons calculée pour l'ensemble du graphe. Pour résoudre ce problème, il faut comparer les distributions de  $P(u)$  en regroupant les sommets en fonction de leur degré. La figure 10 représente les distributions obtenues. Plus le degré est important, plus les valeurs de  $P(u)$  sont élevées : les sommets de plus faible degré ont des proportions relativement faibles, les sommets de degré intermédiaire ont une distribution homogène, et la proportion de voisins proches devient de plus en plus élevée pour les sommets de fort degré. Près de 90% des sommets dont le degré est supérieur ou égal à 750 ont une proportion de voisins proches supérieure à 80%.

L'interprétation de ces résultats est délicate, car la taille du voisinage à distance 2 augmente très fortement en fonction du degré : on peut supposer que les sommets de fort degré ont plus de chances d'entrer en contact avec un voisin proche, même s'il est choisi de façon aléatoire. On constate en effet que les sommets de degré 20 ont en moyenne 6 700 voisins à distance 2, avec un maximum de 25 700. Ces nombres sont élevés mais ils représentent malgré tout une très faible proportion de l'ensemble des sommets. Par ailleurs, tous les sommets de degré supérieur à 6 000 ont plus de 230 000 voisins à distance 2 soit environ un quart de l'ensemble des sommets, mais

leur proportion de voisins proches dépasse les 90%. La très grande taille du voisinage à distance 2 pour les sommets de fort degré ne suffit donc pas à expliquer ce phénomène.

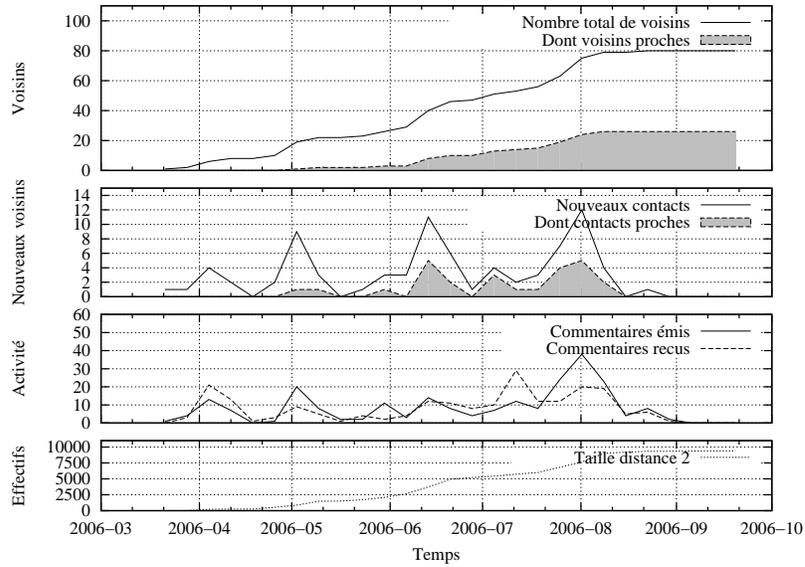
On observe des résultats semblables si l'on effectue les mêmes mesures sur le graphe des commentaires réciproques, avec des ordres de grandeur réduits en raison de sa taille plus petite. Le degré maximum est alors de 2 930, et on peut obtenir des distributions de la proportion de voisins proches presque identiques en choisissant respectivement des classes de degré 10 à 29, 30 à 79, 80 à 499 et supérieur à 500.

Ces résultats nous confirment que les sommets de fort degré jouent un rôle important dans la structuration du « petit monde » en concentrant autour d'eux l'essentiel des zones denses, mais les fortes valeurs de  $P(u)$  soulignent l'importance de la transitivité dans la construction de leur entourage. Ce caractère local de l'entourage des sommets de fort degré, qui relativise la notion de centralité de degré est aussi souligné par Cha *et al.* (2009), qui montrent que les photographies les plus populaires sont le plus souvent choisies comme favorites par des utilisateurs à très faible distance du propriétaire de la photographie.

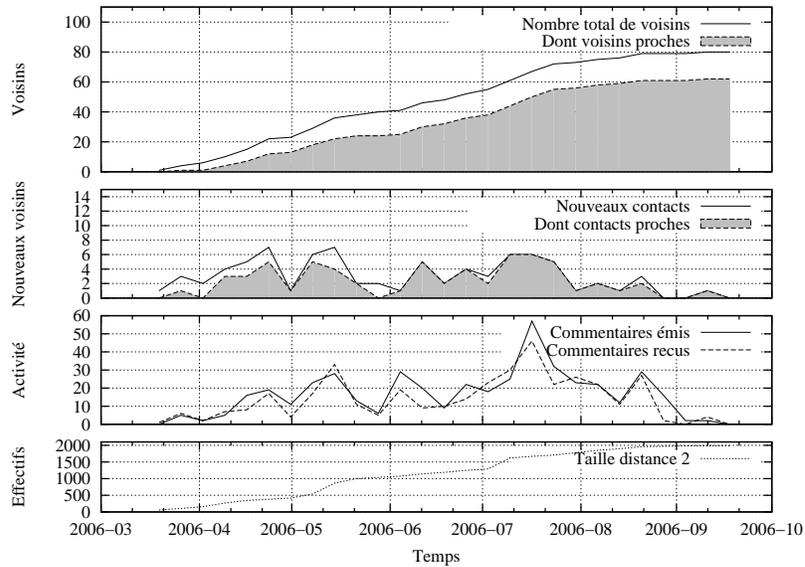
### 3.2. *Suivi individuel des sommets*

Pour mieux comprendre l'influence de l'origine des nouveaux voisins sur la constitution du voisinage, nous avons choisi d'adopter une approche égocentrée. Les figures 11 et 12 ont été obtenues en calculant l'évolution de l'entourage dans  $G_t$  de deux sommets  $A$  et  $B$  (cf. partie 1.4) : nous mesurons chaque semaine le nombre de voisins et de voisins proches, ainsi que le nombre de nouveaux voisins et de nouveaux voisins proches. Nous travaillons ici sur le graphe des commentaires réciproques, sans suppression de liens : l'évolution de la taille des voisinages n'est donc pas biaisée par l'orientation des commentaires, car deux sommets de même degré peuvent avoir des types de relations très différents si un sommet a construit son entourage en envoyant de nombreux messages sans réponse, ou si au contraire il a reçu de nombreux messages auxquels il n'a pas nécessairement répondu. Nous gardons malgré tout un indice de cette orientation en mesurant le nombre de commentaires émis et reçus chaque semaine.

Notre objectif étant d'évaluer les relations entre les proportions de voisins proches et l'évolution de la taille du voisinage, nous n'avons pas choisi de comparer des sommets de très fort degré car ils se caractérisent tous par de très fortes valeurs de  $P(u)$  et que la taille de leur voisinage augmente de manière continue tout au long de leur période d'activité. Ces sommets correspondent par ailleurs aux utilisateurs les plus investis dans le service, et ont donc un caractère très exceptionnel. Pour pouvoir observer plus de variations dans l'évolution du voisinage, nous avons choisi des sommets de degré 80 : cette valeur est suffisamment élevée pour que l'on puisse observer une réelle évolution, tout en restant dans une tranche de degrés où les valeurs de  $P(u)$  sont réparties de manière homogène. Les deux sommets sont actifs au cours de la même période, à partir de fin mars 2006.



**Figure 11.** Évolution de l’entourage du sommet A, qui privilégie les contacts lointains.



**Figure 12.** Évolution de l’entourage du sommet B, qui privilégie les contacts proches.

Les échelles sont identiques pour les deux figures, sauf pour l’évolution de la taille de la distance 2 : la figure 11 a un maximum de 10 000 tandis que la figure 12 a un maximum de 2 000 seulement.

Dans le cas du sommet  $A$ , qui a une faible proportion de voisins proches ( $P(A) = 32,5\%$ ), le degré augmente par à-coups, avec des paliers successifs. Il a une activité équilibrée entre messages émis et reçus, dans l'ensemble inférieure à une dizaine de messages par semaine en dehors des pics ponctuels d'activité. Le sommet  $B$  favorise au contraire les contacts avec son voisinage proche ( $P(B) = 77,5\%$ ) : il est plus actif, avec une vingtaine de messages émis et reçus par semaine en moyenne et un pic d'activité de plus 50 messages émis. Son voisinage augmente de façon régulière tout au long de la période. Le sommet  $A$  s'investit moins dans le service, sauf lors de pics d'activité qui peuvent être interprétés comme des périodes d'intérêt ponctuel pendant lesquelles il entre en contact avec de nouveaux voisins.

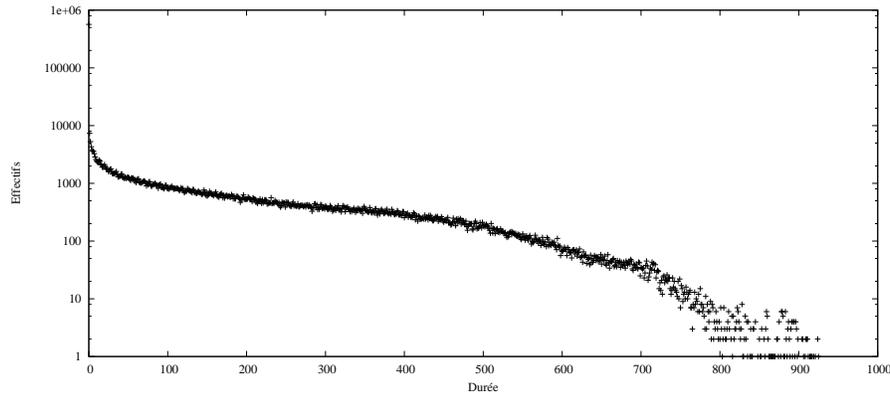
Bien sûr, ces deux exemples ne sont pas représentatifs de la variété des situations que l'on peut observer, mais la multiplication d'observations similaires nous a permis de constater que les profils d'évolution des sommets à faible  $P(u)$  sont très variés alors que ceux des sommets à  $P(u)$  élevé sont souvent réguliers, en particulier lorsqu'il s'agit de sommets de fort degré. Cha *et al.* (2009) ont adopté une démarche similaire pour étudier l'évolution du nombre d'utilisateurs qui ajoutent une photographie parmi leurs « favoris ». Ils montrent que la courbe d'évolution augmente de façon régulière, sauf en cas d'évènement extérieur (*e.g.* la photographie gagne un prix, est mise en page d'accueil du site) qui se caractérise par l'arrivée massive de nouveaux « fans » et donc par un pic de croissance. On peut donc imaginer que les profils d'évolution variés des sommets à faible  $P(u)$  sont davantage liés à des évènements extérieurs que ceux des sommets à fort  $P(u)$ , qui correspondent plutôt à une évolution caractéristique des utilisateurs les plus actifs. Pour mieux comprendre ces différents comportements, nous nous intéressons au rythme de l'activité des sommets.

#### 4. Gestion du temps par l'individu

##### 4.1. Mesure de l'activité de chaque individu

Pour mesurer les rythmes de l'activité acteurs du réseau, nous introduisons quelques mesures supplémentaires. Nous déclarons un *sommet actif* au cours d'un intervalle de temps donné s'il a émis au moins un commentaire au cours de cet intervalle. La *durée d'activité d'un lien* est l'intervalle de temps écoulé entre sa première et sa dernière occurrence. De la même manière, la *durée d'activité d'un sommet* est l'intervalle de temps écoulé entre les dates de première et dernière activité du sommet. Ces mesures ne tiennent pas compte des variations dans l'activité des sommets et en particulier des éventuelles périodes d'inactivité : nous choisissons pour cela de découper  $T$  en intervalles de temps fixe, et nous définissons *l'activité réelle d'un sommet* comptant le nombre d'intervalles au cours desquels le sommet est actif. Nous avons choisi pour notre étude d'exprimer l'activité réelle en intervalles d'une semaine.

La figure 13 montre la distribution des durées d'activité, exprimées en jours, pour l'ensemble des sommets du graphe. On dénombre 567 477 sommets avec une durée d'activité inférieure à une journée, et la durée maximale est de 925 jours. La distribu-



**Figure 13.** *Distribution des durées d'activité des sommets pour l'ensemble des commentaires.*

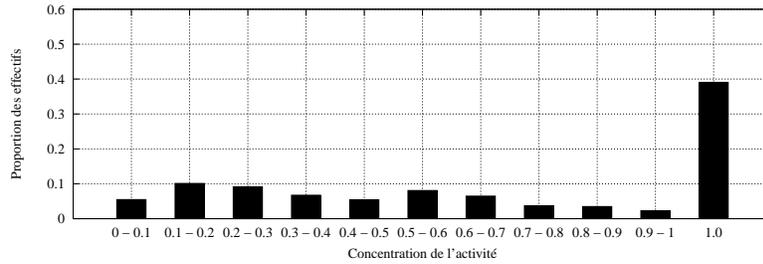
tion semble décroître faiblement<sup>1</sup> pour les sommets qui ont une durée comprise entre 100 et 500 jours. Si l'on s'intéresse à l'activité des sommets pour le graphe des commentaires réciproques, on obtient une distribution similaire, mais avec une proportion beaucoup moins importante de sommets dont la durée est inférieure à une journée, que nous désignerons comme des *sommets éphémères*. Nous ne présentons pas ici la distribution de l'activité des liens, mais elle suit une distribution tout à fait comparable, avec un grand nombre de liens de très courte durée, un intervalle de durées relativement stable, et quelques liens de longue durée. La distribution de l'activité réelle des sommets est quant à elle plus hétérogène que celle des durées d'activité, tout en restant dans les mêmes ordres de grandeur.

Durée d'activité	Tous les commentaires	Commentaires réciproques
moins d'une journée	62, 3%	26, 7%
moins d'une semaine	65, 3%	30, 8%
moins d'un mois	70, 7%	39, 3%
moins de 3 mois	78, 3%	53, 2%
moins de 6 mois	85, 5%	67, 5%
moins d'un an	94, 1%	86, 5%

**Tableau 2.** *Répartition des proportions de sommets actifs en fonction de la durée d'activité.*

Nous comparons les résultats des mesures sur les deux graphes dans le tableau 2 : les sommets actifs pendant moins d'une journée représentent 62, 3% de l'ensemble

1. L'utilisation d'une échelle logarithmique pour l'axe des coordonnées ne permet pas de voir en détail cette partie.



**Figure 14.** *Distribution de la concentration de l'activité des sommets par tranches d'une semaine.*

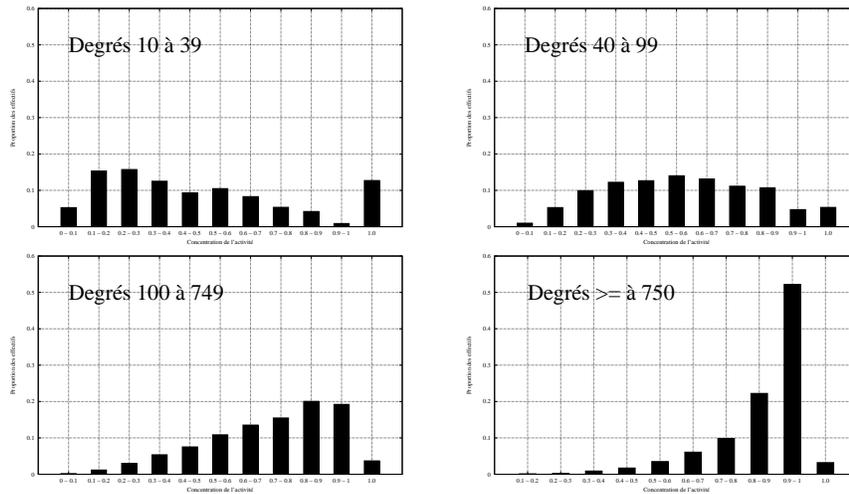
des sommets si l'on considère l'ensemble des commentaires, mais seulement 26,7% pour le graphe des commentaires réciproques. Cette différence n'est pas surprenante puisque nous savons que par construction, tout commentaire a forcément reçu une réponse dans le graphe des commentaires réciproques. Cependant en pratique, une grande partie de ces réponses arrive très rapidement après l'envoi du premier message (Kumar *et al.*, 2006). L'écart important des proportions, qui perdure pour des durées d'une semaine ou d'un mois nous confirme que le graphe des commentaires réciproques s'affranchit d'un grand nombre d'utilisateurs ponctuels, même si leur proportion reste importante.

#### 4.2. Concentration de l'activité

On peut alors mesurer la *concentration de l'activité* d'un sommet en faisant le rapport entre son activité réelle et sa durée d'activité : nous calculons donc pour chaque sommet le rapport entre le nombre de semaines pendant lesquelles il a émis au moins un commentaire et le nombre de semaines écoulées entre son premier et son dernier commentaire. La figure 14 représente la distribution de la concentration de l'activité des sommets pour l'ensemble des commentaires. On remarque la présence d'un pic de sommets qui ont une concentration de 100% : cela correspond aux utilisateurs éphémères, qui se désintéressent rapidement du service et l'abandonnent dès la première semaine.

La figure 15 représente les variations de cette distribution selon quatre classes de degré. On constate que les distributions ressemblent à celles de la proportion de voisins proches, si l'on ne tient pas compte du pic de concentration précédemment évoqué. Les sommets de faible degré ont tendance à avoir une activité plus diffuse, les sommets de degré intermédiaire ont une distribution homogène, et la concentration augmente ensuite de plus en plus pour les sommets de fort degré.

Lorsque la concentration est faible, cela signifie que le sommet a une activité diluée dans le temps, ce qui peut correspondre à des utilisateurs qui utilisent le service



**Figure 15.** Distributions de la concentration de l'activité en fonction du degré. Les abscisses correspondent à la concentration de l'activité et les ordonnées aux effectifs.

de manière occasionnelle. Les distributions obtenues pour les sommets de fort degré montrent qu'une forte proportion d'entre eux se caractérise par une activité continue, avec très peu de périodes d'interruption.

## 5. Conclusion

Nous avons expérimenté dans cet article différentes manières de modéliser un réseau de liens dynamiques à partir de nos données, et nous avons montré que la structure générale du réseau étudié reste similaire tant que l'on ne réduit pas la durée de vie des relations en dessous de la semaine. Nous avons aussi insisté sur le caractère très local des interactions au sein du réseau : on peut nuancer l'image idéalisée d'un « petit monde » où les individus sont bien connectés entre eux en lui substituant celle d'un « monde étriqué » où les contacts avec des inconnus restent exceptionnels. Le croisement des échelles nous a aussi permis d'observer la grande diversité des configurations locales, qui contrastent avec la stabilité de la structure globale. Enfin, la provenance des nouveaux voisins ou le rythme de l'activité sont autant de critères qui peuvent nous aider à caractériser les individus et à s'affranchir des moyennes liées à ces effets de masse.

Les méthodes que nous avons employées sont généralisables à de nombreux graphes de terrain, qu'ils soient issus du Web ou d'autres types de réseaux de communication (téléphonie, mail, pair à pair, ...). La compréhension des mécanismes constitutifs de la dynamique des réseaux doit bien sûr être affinée en prenant en compte

l'orientation des relations et en étudiant les répercussions des suppressions de liens sur la constitution et l'évolution des voisinages des sommets.

Ces approches nous permettent par ailleurs de mesurer le caractère essentiel de la gestion du temps par les individus : dans un contexte où le nombre d'individus et d'interactions entre eux sont foisonnants, l'élément discriminant n'est pas le nombre de photographies que l'on peut consulter ou de contacts que l'on peut solliciter, mais le temps qu'on s'accorde pour établir et entretenir ces relations. Les utilisateurs qui parviennent à entretenir un nombre élevé de relations ne sont pas ceux qui sont présents sur le service depuis le plus longtemps, mais ceux qui l'utilisent de manière la plus intensive et la plus efficace, en privilégiant les contacts avec des voisins proches, qui s'établissent plus rapidement. Dans ce contexte « d'économie de l'attention » (Simon, 1971; Lanham, 2007), les différentes configurations des voisinages sont une conséquence de la diversité des modalités de cette nécessaire gestion du temps.

La forte proportion des relations transitives nous montre que les utilisateurs de Flickr qui commentent le plus privilégient des pratiques de « navigation sociale » par rapport au moteur de recherche. On pourrait s'intéresser dans une perspective plus sociologique aux usages de la photographie comme support de la conversation en comparant le contenu des messages avec les indicateurs que nous avons établis, et les pratiques de Flickr avec celles d'autres plateformes comme les blogs (Cardon *et al.*, 2006), ou plus récemment Facebook.

## 6. Bibliographie

- Albert R., Barabási A. L., « Statistical mechanics of complex networks », *Reviews of Modern Physics*, vol. 74, n° 1, p. 47-97, Jan, 2002.
- Barabasi A.-L., Albert R., « Emergence of scaling in random networks », *Science*, 1999.
- Beuscart J.-S., Cardon D., Pissard N., Prieur C., « Pourquoi partager mes photos de vacances avec des inconnus ? Les usages de Flickr », *Réseaux*, 2009.
- Bornholdt S., Schuster H. G. (eds), *Handbook of Graphs and Networks*, Wiley-Vch, 2003.
- Boyd D., « Friendster and Publicly Articulated Social Networks », *Conference on Human Factors and Computing Systems (CHI 2004)*, 2004.
- Brin S., Page L., « The anatomy of a large-scale hypertextual Web search engine », *Computer Networks and ISDN Systems*, vol. 30, n° 1-7, p. 107-117, 1998.
- Burt R., *Structural Holes. The Social Structure of Competition*, Cambridge, Harvard University Press, 1992.
- Cardon D., Delaunay-Téterel H., « La production de soi comme technique relationnelle : un essai de typologie des blogs par leurs publics », *Réseaux*, 2006.
- Cardon D., Prieur C., *Les réseaux de relations sur Internet: un objet de recherche pour l'informatique et les sciences sociales*, Hermès, chapter in Humanités numériques, C. Brossaud and B. Reber (ed.), 2008.

- Cha M., Mislove A., Adams B., Gummadi K. P., « Characterizing social cascades in flickr », *WOSP '08: Proceedings of the first workshop on Online social networks*, ACM, New York, NY, USA, p. 13-18, 2008.
- Cha M., Mislove A., Gummadi K. P., « A Measurement-driven Analysis of Information Propagation in the Flickr Social Network », *Proceedings of the 18th Annual World Wide Web Conference (WWW'09)*, April, 2009.
- Degenne A., Forse M., *Les réseaux sociaux*, Armand Colin, 1994.
- Efe K., Raghavan V., Chu C. H., Broadwater A. L., Bolelli L., Ertekin S., « The Shape of the Web and Its Implications for Searching the Web », , CiteSeerx, 2000.
- Elias N., *La société des individus*, Fayard, 1991.
- Granovetter M., « The Strength of Weak Ties », *American Journal of Sociology*, vol. 78, n° 6, p. 1360-1380, 1978.
- Gribaudo M. (ed.), *Espaces, temporalités, stratifications : Exercices sur les réseaux sociaux*, éditions de l'EHESS, 1998.
- Kumar R., Novak J., Tomkins A., « Structure and evolution of online social networks », *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, New York, NY, USA, p. 611-617, 2006.
- Lanham R. A., *The Economics of Attention: Style and Substance in the Age of Information*, University Of Chicago Press, September, 2007.
- Latapy M., « Main-memory Triangle Computations for Very Large (Sparse (Power-Law)) Graphs », *Theoretical Computer Science (TCS)*, vol. 407, p. 458-473, 2008.
- Lerman K., Jones L., « Social Browsing on Flickr », Dec, 2006.
- Leskovec J., Backstrom L., Kumar R., Tomkins A., « Microscopic evolution of social networks », *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, New York, NY, USA, p. 462-470, 2008.
- Mislove A., Koppula H. S., Gummadi K. P., Druschel P., Bhattacharjee B., « Growth of the flickr social network », *WOSP '08: Proceedings of the first workshop on Online social networks*, ACM, New York, NY, USA, p. 25-30, 2008.
- Newman M. E. J., « The structure and function of complex networks », *SIAM Review*, 2003.
- O'Reilly T., « What is Web 2.0: design patterns and business models for the next generation of software », , <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, 2005.
- Palla G., Barabasi A.-L., Vicsek T., « Quantifying social group evolution », *Nature*, vol. 446, n° 7136, p. 664-667, 2007.
- Prieur C., Stoica A., Smoreda Z., « Extraction de réseaux égocentrés dans un (très grand) réseau social », *Bull. de Méthodologie Sociol.*, 2009.
- Scott J., *Social Network Analysis*, Sage, London, 1992.
- Seidman S. B., « Network structure and minimum degree », *Social Networks*, 1983.
- Simon H. A., *Designing Organizations for an Information-Rich World*, p. 37-72, 1971.
- Watts D. J., Dodds. P. S., « Influentials, Networks, and Public Opinion Formation », *Journal of Consumer Research*, vol. 34, p. 441-458, 2007.
- Watts D., Strogatz S., « Collective dynamics of small-world networks », *Nature*, 1998.
- Wellman B., « An egocentric network tale », *Social Networks*, vol. 15, p. 423-436, 1993.