

Le mobile, reflet des frontières françaises

TÉLÉCOMMUNICATIONS | C'est à l'échelle de la région qu'on passe le plus grand nombre d'appels, selon une étude des transmissions entre antennes-relais. Une illustration du potentiel de la toute jeune science des réseaux

DAVID LAROUSSERIE

Lyon, envoyé spécial

A l'heure des communications tous azimuts, SMS et téléphone portable, on ne s'est finalement guère affranchis des distances. 85% des communications se font au sein d'une région et seulement 15% en sortent. Bref, les Bretons parlent aux Bretons et peu aux Normands. Idem pour les Alsaciens, les Parisiens ou les Corses. Les « ondes » semblent s'arrêter aux frontières administratives !

Telle est l'une des conclusions étonnantes récemment obtenue par une équipe de chercheurs de l'université de Louvain, en Belgique, du Massachusetts Institute of Technology (MIT) aux Etats-Unis, et de l'Orange Labs, le laboratoire de recherche de l'opérateur de télécommunication Orange. Elle a été présentée, à Lyon, les 12 et 13 décembre, lors du colloque « Réseaux sociaux : des structures à la politique », organisé par le CNRS, l'Institut national de la recherche en informatique et automatique (Inria) et l'Ecole normale supérieure de Lyon. Plus de 1,5 milliard de communications (SMS et voix) entre 17 millions de clients particuliers de l'opérateur recensés pendant six mois, en 2007, ont été analysés. Le contenu en était effacé, seuls restait la

Les Bretons parlent

aux Bretons

et peu aux Normands.

Idem pour les Alsaciens,

les Parisiens

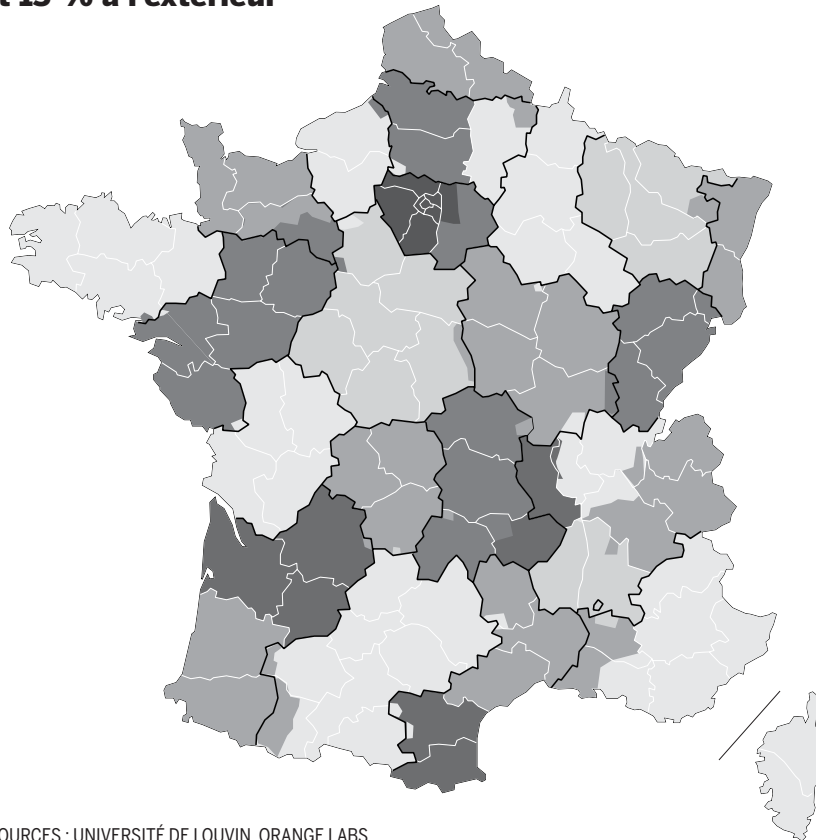
ou les Corses

durée et les lieux d'appel (identifiés grâce aux antennes-relais). Les chercheurs se sont demandés si des structures apparaissent au sein de cette gigantesque mer de données.

« Les géographes savent bien que les comportements sociaux se concentrent autour de points centraux. Nous nous attendions donc à découvrir un découpage territorial. Mais nous pensions que ce serait le département plutôt que la région », explique Zbigniew Smoreda, sociologue de l'Orange Labs. Cette séparation est assez nette car aux frontières, le ratio 85%-15% entre « extérieur » et « intérieur » ne tombe qu'à 75%-25%.

L'algorithme de découverte des communautés ou des agrégats a donc trouvé que le meilleur découpage comme les frontières plus anciennes des départements pour coller à celles, plus récentes, des régions. Dans un agrégat, les clients échangent moult communications entre eux mais peu avec les autres. Cet algorithme, dit méthode de Louvain, a été inventé en 2008, et a depuis servi à traiter d'autres grands réseaux comme Twitter ou LinkedIn et leurs millions d'utilisateurs et « amis » (ou respectivement « nœuds » et « liens » dans le jargon du monde des réseaux mathématiques). En fait, la méthode permet d'obtenir le meilleur découpage, donnant des régions les plus denses possibles en liens. « Tous les trois jours je reçois des demandes

85 % des communications se font entre régions et 15 % à l'extérieur



SOURCES : UNIVERSITÉ DE LOUVAIN, ORANGE LABS

d'information à propos de cette méthode », témoigne Vincent Blondel, de l'université de Louvain et du MIT, co-inventeur de la technique. « Il y a dix ans, de tels outils n'existaient pas. Il y en a aujourd'hui des centaines, mais le nôtre est réputé comme étant le plus rapide. » Ainsi, si le nombre de nœuds du réseau est multiplié par 100, le temps de calcul n'est multiplié que par 200, alors qu'il aurait été multiplié par 10 000 pour des méthodes moins performantes.

La France des mobiles n'a donc pas mis longtemps à livrer ses secrets. Avant elle, la Belgique était aussi passée à la moulinette de Louvain. La frontière linguistique Wallonie-Flandre était apparue assez nettement, tout comme l'îlot bruxellois plus multilingue. « Le test sur la France était intéressant car nous n'avions pas le biais

linguistique mais a priori un fort poids de l'administration », se souvient Zbigniew Smoreda. Pour le sociologue, la carte scolaire, donc les collèges et les lycées, oriente fortement les flux d'interactions entre individus. Les adolescents parlent à leurs copains de la même zone ; tout comme les parents. « C'est une sorte de formatage administratif des relations sociales. Ces liens entre espace et relations sociales intéresseront sûrement le marketing, soucieux de bien diffuser ses messages. » Il y a cependant des exceptions. Le nord Aquitaine parle peu au sud et réciproquement. Le Languedoc ne cause pas au Roussillon. Le département de la Haute-Loire est tirailé entre Auvergne et Rhône-Alpes...

Les chercheurs ont aussi confirmé une loi quasi universelle bien connue des sociologues. Il

Pendant six mois, en 2007, les communications SMS et portables de 17 millions d'utilisateurs ont été collectées anonymement et localisées grâce aux antennes-relais. Un traitement informatique a regroupé ces échanges en zones riches en interactions internes mais pauvres vers les autres. Une ségrégation sociale (zones grisées) a été mise en évidence, correspondant aux régions administratives (traits noirs) et non aux départements (traits blancs). Ou presque. Rhône-Alpes est coupé en trois. Le Languedoc-Roussillon « grignote » PACA. Tout comme l'Alsace vis-à-vis de la Lorraine.

existe comme une loi de Newton de l'attraction universelle qui impose que le volume de données échangées entre deux entités, comme des communes ou des régions, est inversement proportionnel au carré de la distance entre elles. A taille égale, deux villes distantes de 100 kilomètres échangent cent fois moins de communications que deux villes distantes de 10 kilomètres.

Cette loi fonctionne même pour des individus : la probabilité d'appeler un ami à 100 kilomètres est cent fois plus faible que celle d'appeler un ami à 10 kilomètres. « Cette loi de décroissance avait aussi été constatée auparavant sur les échanges économiques entre villes ou sur les lieux de déménagement », rappelle Vincent Blondel. En choisissant pour logo la pomme, symbole de Newton, Apple avait donc vu juste. ■

A la recherche des corrélations perdues

Le volume des données engendrées par les activités humaines dépasse notre capacité à les traiter », rappelle Vincent Blondel (Université catholique de Louvain et MIT), qui travaille à la mise au point d'algorithmes performants pour s'y retrouver dans cette masse d'information. C'est aussi ce que propose une équipe américaine du Massachusetts Institute of Technology (MIT) et de l'université d'Harvard dans la revue *Science* du 16 décembre.

Ces chercheurs ont élaboré une recette pour trouver des corrélations entre différents jeux de paires de données. Y a-t-il un lien entre passage des cigognes et naissance ? Entre pleine lune et criminalité ? Plus sérieusement, ils s'intéressent à des bases de données de gènes, de bactéries ou de médecine, mais aussi... de base-ball.

Le but est de détecter, un peu à l'aveuglette, si des données a priori indépendantes le sont vraiment. La nouvelle technique a été comparée avec une méthode traditionnelle en statistique, appelée régression linéaire.

Selon les chercheurs, elle a permis de trouver des corrélations insoupçonnées et plus complexes. Comme par exemple un lien entre obésité et revenus dans des îles du Pacifique. Des gènes ont aussi été mieux identifiés par leur méthode que par d'autres. Cependant, il reste à comparer cette méthode à des techniques plus sophistiquées.

En outre, elle ne prend que les données deux à deux sans tenir compte de critères supplémentaires, ce qui est souvent nécessaire pour les analyses plus complexes. L'invention mon-

tre que l'exploration des bases de données devient aussi un enjeu de société.

« Nous entrons dans une nouvelle période des usages de la statistique avec des collectes de plus en plus massives de données via les réseaux sociaux, par exemple, et toutes les traces que nous laissons. On ne cherche plus à faire des hypothèses pour les traiter mais on cherche seulement des corrélations, érigées en normes scientifiques. Les individus disparaissent au profit de profils statistiques », constate Thomas Berns, philosophe de l'Université libre de Bruxelles, spécialiste des relations entre politique et traitement des données. Les géants du Web comme Amazon ou Facebook sont d'ailleurs à l'avant-garde de ce qu'il appelle le « gouvernement algorithmique ». ■ D.L.

La NASA a perdu des centaines d'échantillons spatiaux

Un audit dénonce la faiblesse des contrôles de l'agence américaine, notamment lors de prêts de fragments de météorites

JULIEN JOLY

La NASA devra faire plus attention aux échantillons qu'elle prête aux chercheurs. Selon un audit de l'agence spatiale américaine publié jeudi 8 décembre, 517 de ces fragments « ont été perdus ou volés entre 1970 et juin 2010 ». Des ressources uniques, comprenant des morceaux de Lune et de météorites, de la poussière de comète et des ions issus de la couche extérieure du Soleil.

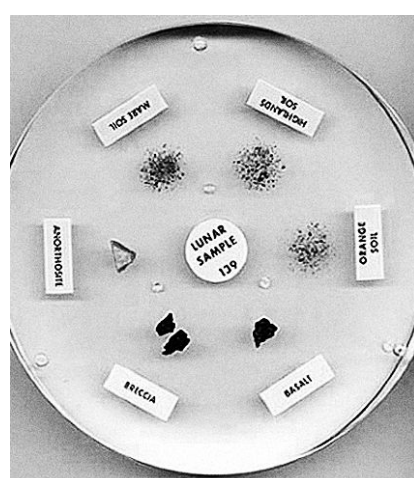
« La plupart des faiblesses que nous avons observées dans les procédures de la NASA sont associées aux prêts de météorites et de poussières cosmiques », dénonce Paul Martin, inspecteur général de l'agence spatiale. Contrairement à d'autres

types de matériaux, elles peuvent être recueillies sur Terre et détenues légalement par des particuliers. Par conséquent, une proposition de vente de météorite volée attire moins l'attention des autorités que des échantillons lunaires d'Apollo.

Sur les 517 échantillons du rapport, 218 fragments de Lune et de météorites ont été volés à un chercheur américain en 2002, mais ont été retrouvés ensuite.

« Matière rare et limitée »

Les météorites sont gardées moins précieusement que les fragments lunaires, et ne font pas l'objet de conventions de prêt. « Les gens font moins attention aux météorites », confirme Jérôme Gattacceca, chercheur au CNRS. D'autant plus qu'en cas de perte d'un échantillon, quel qu'il soit, « il n'y a pas de sanction financière ».



Echantillons de roches lunaires.

NASA/AP

En ce qui concerne les échantillons lunaires, les mesures de prêt sont plus strictes, précise Marc Chaussidon, directeur de recherche au CNRS : « Il faut envoyer notre projet scientifique à un comité de la NASA, qui nous attribue ensuite une petite quantité d'échantillons. On les garde dans un coffre-fort au centre de recherche. Chaque année, la NASA vérifie l'avancement de nos travaux. »

Malgré ces précautions, le rapport cite douze cas de chercheurs décédés, partis en retraite ou mutés sans qu'on sache ce qu'il était advenu de leurs échantillons. D'autres avaient encore en leur possession des fragments que la NASA croyait détruits lors des expériences.

Cinq cent dix-sept échantillons disparus en quarante ans, cela peut paraître dérisoire quand on sait que la collection

totale de la NASA s'élève à 140 000 morceaux de Lune, 18 000 échantillons de météorite et 5 000 échantillons de poussières cosmiques et de comète.

Dans le cas des fragments lunaires, les quantités prêtées aux laboratoires n'excèdent pas les quelques dizaines de milligrammes, sur les 350 kg rapportés sur Terre. Enfin, certains de ces « échantillons disparus » ont été retrouvés par la suite.

Leur nombre est toutefois révélateur des lacunes dans les contrôles de la NASA. Le rapport préconise que l'agence devra désormais mieux répertorier les morceaux qu'elle détient et mettre en place un inventaire annuel, car « ces échantillons constituent une matière rare et limitée et jouent un rôle important dans la recherche et l'éducation. » ■