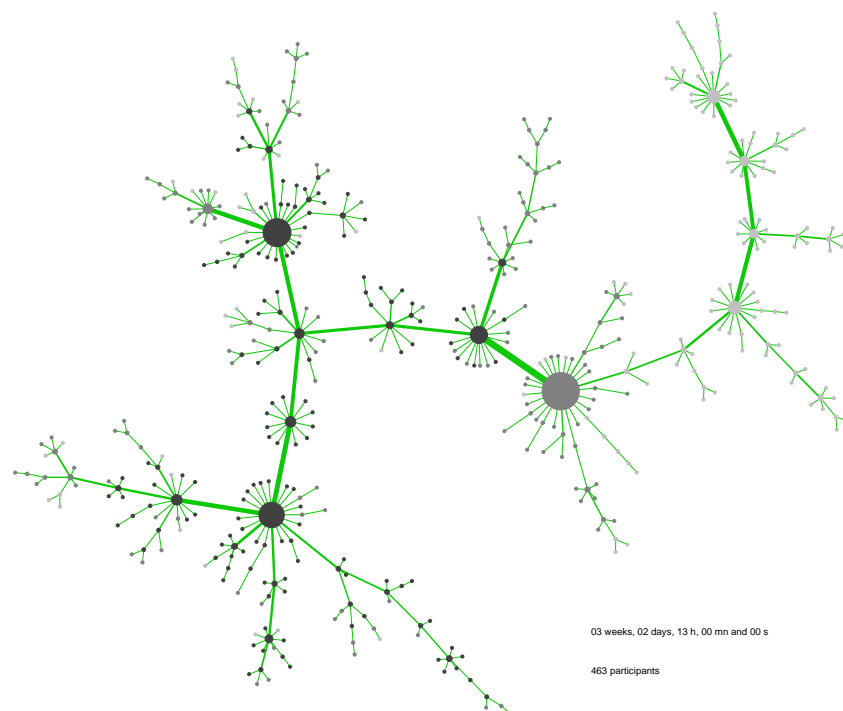


Phénomènes de diffusion sur les grands réseaux : mesure et analyse pour la modélisation

Matthieu Latapy et Bénédicte Le Grand

Prenom.Nom@lip6.fr

<http://complexnetworks.fr> – LIP6 – CNRS et UPMC – Paris



La propagation des épidémies, des rumeurs, des virus informatiques, ou la diffusion d'une information (que ce soit sur un réseau informatique ou un réseau social) sont des exemples types de *phénomènes de diffusion*. De par leur importance, ces phénomènes sont au cœur d'une intense activité de recherche.

Il est toutefois extrêmement difficile d'avoir une information précise sur *comment* des processus de diffusion **réels** se passent : il faut connaître le réseau entre les acteurs, les changements d'états et leurs causes, tout ceci au cours du temps et à une échelle suffisamment grande pour permettre l'analyse. Aujourd'hui, les données de ce type sont extrêmement rares, et souvent très partielles et biaisées.

Par conséquent, à quelques exceptions¹ près, la plupart des travaux concernant les phénomènes de diffusion reposent sur des *modèles*, qui eux-mêmes reposent sur des intuitions simples (par exemple l'idée selon laquelle une personne infectée aurait une certaine probabilité de contaminer ses contacts).

L'équipe *Complex Networks* du LIP6 est impliquée depuis plusieurs années dans la collecte de données riches et à large échelle, dont certaines permettent d'ouvrir des perspectives extrêmement prometteuses pour l'analyse de phénomènes de diffusion *réels*, et donc leur modélisation.

Le stage proposé ici vise à utiliser ces données pour obtenir, pour la première fois, des observations quantitatives sur des phénomènes de diffusion réels. L'objectif central est de confronter ces mesures aux modèles existants et ainsi d'évaluer la pertinence des hypothèses sous-jacentes. On dégagera ainsi de grands principes, confrontés à la réalité, pour la modélisation des phénomènes de diffusion.

Soulignons que les *méthodes* pour calculer les grandeurs qui nous intéressent restent à définir, ce qui est non trivial. Mener les calculs soulèvera également des *problématiques algorithmiques*, de *programmation* et d'*échantillonnage* délicates, les données en question étant complexes (graphes valués, dynamiques, etc) et de très grande taille (millions d'entités, centaines de millions de diffusions).

¹Voir en particulier http://complexnetworks.fr/videos.php?video_id=7