Stage LIESSE

# Promenade aléatoire dans les graphes

par Thomas BONALD
enseignant-chercheur au département InfRes de Télécom ParisTech.

mardi 23 mai 2017

Nous étions une vingtaine à nous promener aléatoirement à Télécom ParisTech, avec une écrasante majorité de matheux, enseignant en Spé.

Après un accueil autour de viennoiseries accompagnées de café/thé, nous nous sommes présentés. Thomas Bonald a été élève de l'École polytechnique, puis à Télécom Paris, et est devenu professeur au département InfRes de Télécom Paristech. Il nous propose de nous faire découvrir les aspects actuels des promenades aléatoires dans les graphes à l'aide du menu suivant.

1. Marche aléatoire.
2. Analyse spectrale (de la matrice de transition).
3. Temps d'atteinte (et non d'attente !).
4. Analogie électrique.
5. Plongement spectral.
6. Partionnement.
7. Classement.

**Introduction et motivations.** Les graphes sont bien sûr très présents dans de nombreux domaines : réseaux de neurones, réseaux sociaux (Face Book constituant un graphe non orienté, à l'inverse de Twitter), réseaux routiers/aériens, marketing (notamment le *marketing viral* qui a pour objet de cibler des clients ayant beaucoup de connaissances afin d'espérer de diffuser les produits le plus possible).

Ces graphes ont en général énormément de sommets et encore plus d'arêtes : comment alors en extraire de l'information, par exemple repérer les sommets importants, regrouper par communautés, etc.

On évoque le premier algorithme de Google, PageRank (1998), qui consiste à considérer comme populaires les pages les plus visitées par un internaute se promenant aléatoirement sur le web.

1. **Marche aléatoire.** On parle ici de chaînes de Markov (sans refaire un cours de Probabilités). Peut-être qu'un non matheux pourrait rencontrer quelques difficultés, mais rien n'a été trop technique trop longtemps et tout est expliqué, vidéo-projections à l'appui. On explique l'équivalence entre connexité du graphe et irréductibilité de la chaîne de Markov correspondante et l'on établit l'équation d'équilibre $\pi_j = \sum_i \pi_i p_{i,j}$. Les physiciens ne sont pas en reste avec le modèle d'Ehrenfest de diffusion d'un gaz qui illustre le fait qu'une chaîne de Markov peut être réversible sans pour autant rendre probable le retour à l'état initial.

2. **Analyse spectrale (de la matrice de transition).** On met en jeu le théorème de Perron-Frobenius et les matrices (à entrées) positives, le rayon spectral, des vecteurs propres gauches et droits. On introduit le laplacien (qui n'a rien à voir avec les dérivées partielles) qui est la matrice[1] $D^{-1/2}AD^{-1/2}$ qui a l'avantage d'être symétrique, contrairement à la matrice de transition $P = D^{-1}A$. Une fois ortho-diagonalisée en $\Lambda$, on démontre alors que $P = V\Lambda U^{\mathrm{T}}$ avec $U, V$ deux matrices telles que $U^{\mathrm{T}}V = I$ : c'est la décomposition spectrale de $P$.

3. **Temps d'atteinte.** On introduit le temps d'atteinte $T_i = \min\{n \geq 1 \mid X_n = i\}$ puis $h_{i,j} = \mathrm{E}_i(T_j)$. On note que $h_{i,i} = \frac{1}{\pi_i}$ n'est autre que le temps de retour en $i$. On symétrise ensuite les $h_{i,j}$ en considérant $\sigma_{i,j} = h_{i,j} + h_{j,i}$ qui peut s'interpréter comme le temps moyen d'atteinte entre $i$ et $j$ et on fait le lien avec la probabilité d'échappement $e_{i,j} = \mathrm{P}(T_j < T_i)$.

4. **Analogie électrique.** On considère des électrons parcourant un circuit électrique, et on utilise la loi de Kirchhoff et même la loi d'Ohm pour montrer, comme l'intuition nous le dicte, que le temps d'atteinte est proportionnel à la résistance. Ici les poids sur les sommets sont joués par des conductances.

*Une pause déjeuner bien agréable, dans une salle proche, où des plateaux repas nous ont été offerts.*

5. **Plongement spectral.** On construit astucieusement un produit scalaire sur $\mathbb{R}^N$ pour établir que $\sigma_{i,j}$, le temps moyen d'atteinte entre $i$ et $j$, peut se voir comme le carré d'une distance $\|v_j - v_i\|^2$. On peut alors imaginer utiliser les notions métriques (boules, etc.) pour classer la pertinences des sommets.

6. **Partionnement.** L'idée est simple : comment regrouper par paquet les sommets d'un graphe ? La réponse l'est moins, et on constate sur des exemples au vidéo-projecteur sur des graphes modestes (une 50ne de sommets) que l'appréhension visuelle est parfois trompeuse. Les exemples sont parfois très concrets : par exemple les personnages des *Misérables* de Victor Hugo, une arête entre $i$ et $j$ matérialisant la présence de $i$ et $j$ dans un même chapitre.

7. **Classement**. On évoque enfin le problème de classement, en parlant cette fois de graphes orientés. Exemple : si on tape « Jaguar », comment classer les pages parlant de voitures de celles concernant les félins, sans pour autant faire une analyse linguistique ? La méthode officielle qu'utilise Google est expliquée, mais il est fort probable que l'on ne nous dise pas tout...

**En conclusion.** Un stage de bon niveau, bien expliqué, détaillé et visuel, avec un poly en anglais, s'il vous plaît et un intervenant disponible, bon pédagogue. Il manque peut-être davantage de codes Python (le stage n'était pas prévu pour nous faire coder sur machine, ce n'était pas une surprise) pour que l'on puisse réexploiter certaines idées en TP.

1. A : la matrice d'adjacence du graphe, D : la matrice diagonale des degrés des sommets.