TD - Science des réseaux

Antoine Rey, Quentin Pointeau

1 Questions de cours

Question : Expliquer brièvement le principe d'attachement préférentiel dans un réseau complexe évolutif.

Lorsque l'on crée un réseau aléatoire, les nœuds sont reliés par des arcs ou arêtes de manière complètement aléatoire. Cependant, cette manière de procéder ne correspond pas à la réalité. En effet, les réseaux ne se sont pas créés directement de grande taille, les réseaux que l'on connaît sont plutôt le fruit d'une *"longue"* évolution.

Lorsqu'un nouveau nœud est créé dans un réseaux, il est préférable pour lui de se rattacher à ce que l'on appelle un *hub*, c'est-à-dire un nœud qui possède déjà un grand nombre de voisins. Ainsi, le nouveau nœud est assuré d'être connecté rapidement à une (très) grande partie du réseau. Ce phénomène est appelé *attachement préférentiel.*

Cette manière de procéder nous pose tout de même quelques questions concernant la robustesse et résilience d'un réseau. Si tous les nouveaux nœuds sont reliés à un hub, alors les hubs deviennent des nœuds très sensibles du réseau car sans eux, beaucoup de connexions se perdent. Des solutions comme l'ajout de connexion entre des nœuds plus petits peuvent déjà pallier, ou du moins en partie, ce problème.

Question : Comment la robustesse d'un réseau est-elle mesurée ? Proposez quelques métriques qui permettent de mesurer la robustesse.

La robustesse d'un réseau c'est sa capacité à rester fonctionnel lorsqu'un ou plusieurs liens ne fonctionnent plus.

Si on considère que les pannes dans un réseaux surviennent de manière aléatoire, alors proposer le pourcentage ou le nombre de liens à détruire de manière aléatoire pour que le réseau ne soit plus fonctionnel peut s'avérer une bonne métrique.

Cependant, comme mentionné à la question précédente, les réseaux se reposent de manière générales sur des hubs qui possèdent beaucoup de liens. Il suffirait alors de détruire que quelques hubs pour faire tomber une très grande partie des liens et ainsi mettre en péril le bon fonctionnement du réseau. Pour des réseaux construits de manière évolutives et qui se reposent beaucoup sur des hubs, proposer le nombre de liens minimum (en considérant des attaques ciblées sur les liens critiques) à faire tomber pour empêcher le bon fonctionnement du réseau semble être une meilleure métrique.

Toutefois, sur ces deux définitions, on peut toujours discuter du *bon fonctionnement* du réseau, qui est une notion assez vague et donc sujette à discussions. Ainsi, il paraît compliqué de définir une métrique universelle et correcte de la robustesse d'un réseau.

Question : Que reflète la métrique modularité ? Comment cette métrique est-elle utilisée ?

La modularité d'une partition c de la communauté globale en n_c sous-communautés est :

$$M_c = \sum_{c=1}^{n_c} \left[\frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \right]$$

où L_c est le nombre de liens au sein de la sous-communauté c, L est le nombre de liens au sein de la communauté globale et k_c est la somme des degrés des nœuds de la sous-communauté c.

La modularité reflète la qualité d'une partition d'une communauté : plus la modularité est grande, plus les liens intra-communautaires sont forts. Le but des algorithmes de division de graphe comme celui de Girvan-Newmann est donc de construire des communautés où les liens intra-communautaires sont forts, donc des communautés qui maximise la modularité.

Cette métrique est donc utile pour créer des clusters, des sous-communautés robustes au sein d'un réseau. Par exemple pour la création de clusters de nanosatellites afin de minimiser le temps des échanges au sein du réseau.

Question : Quel(s) modèle(s) peut-on utiliser pour étudier la dissémination d'une rumeur dans un réseau social ?

Afin d'étudier la dissémination d'une rumeur dans un réseau social, on peut utiliser le modèle SI (Susceptible-Infecté). Ce modèle définit deux états : Susceptible et Infecté associés aux ensembles :
— $S(t)$ le nombre d'individus susceptibles d'être infectés au temps t ;
— $I(t)$ le nombre d'individus déjà infectés au temps t.
Un taux d'infection β contrôle le taux de propagation qui est la probabilité de transmettre la maladie (ici la rumeur) entre deux individus, l'un susceptible d'être infecté et l'autre déjà infecté.

Une autre version de ce modèle existe : le modèle SIS (Susceptible-Infecté-Susceptible). Ici le fonctionnement reste le même l'exception qu'un individu infecté peut guérir et donc passer de nouveau dans l'état Susceptible en fonction du taux $\gamma = \frac{1}{D}$ de guérison où D est la durée moyenne du temps passé en étant infecté.
Ici on considérerait donc qu'après un certain temps, un individu se renseigne et ne crois plus donc à la rumeur diffusée sur le réseau social.

2 Exercice : Communautés dans un réseau social

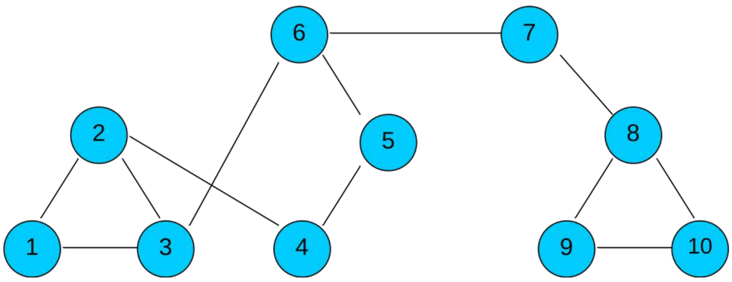

Figure 1 – Graphe G

Question : Donner le nombre chromatique du graphe précédent :

On commence par réaliser un encadrement du nombre chromatique $\gamma(G)$:

$\gamma(G)$	G
$\leq n+1-\alpha(G)$	$\alpha(G) = 4$ d'où 7
$\leq d_{max}+1$	4
$\geq \frac{n}{n-d_{min}}$	$\frac{10}{10-2} = \frac{5}{4} = 1.25$
$\geq \frac{n^2}{n^2-2m}$	$\frac{10^2}{10^2-24} = \frac{25}{19} \simeq 1.32$
$\geq \text{card}_{max}(Clique)$	3

on a donc $3 \leq \gamma(G) \leq 4$.

Avec l'algorithme de Welsh Powell on obtient la coloration du graphe suivante :

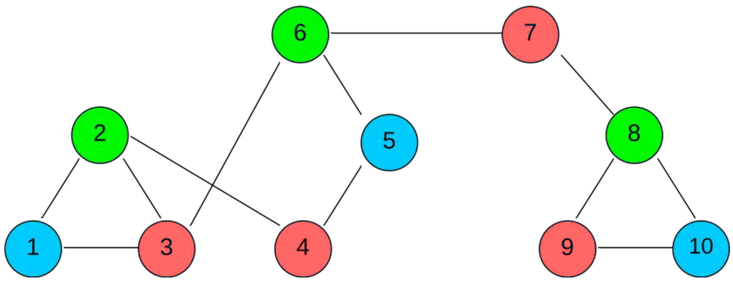

FIGURE 2 – Graphe G_c

On a donc $\gamma(G) = 3$.

Le graphe précédent modélise un réseau social composé de 10 individus. Nous voulons dans la suite déterminer les degrés de centralité des individus dans ce réseau. Trois définitions de la centralité sont données.

Centralité d'intermédiarité ($C_{\textbf{Intermédiarité}}$) : La centralité d'intermédiarité compte le nombre de fois où un nœud agit comme un point de passage le long du plus court chemin entre deux autres nœuds. Elle a été présentée comme une mesure pour quantifier le contrôle d'un humain sur la communication entre d'autres humains dans un réseau social par Linton Freeman. Dans sa conception, les sommets qui ont une forte probabilité d'apparaître sur un court chemin choisi au hasard entre deux sommets choisis au hasard ont une haute intermédiarité.
L'intermédiarité d'un sommet v dans $G = (V, E)$ avec $|V|$ sommets est calculée comme suit :
- Pour chaque couple (s, t), on calcule les plus courts chemins les reliant.
- Pour chaque couple (s, t), on détermine la proportion de plus cours chemins qui passent par le sommet en question, ici v.
- On somme cette fraction sur tous les couples (s, t) de sommets.

De façon plus concise, l'intermédiarité peut être représentée par :

$$C(v) = \sum_{s \neq v, t \neq v} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$

où $\sigma_{s,t}$ est le nombre total de plus courts chemins du sommet s au sommet t et $\sigma_{s,t}(v)$ est le nombre de tels chemins qui passent par v.
Notons que, dans le graphe étudié, il existe un seul plus court chemin entre chaque couple de sommets pris au hasard.

Centralité de proximité ($C_{\textbf{Proximité}}$) : Dans un graphe connexe il y a une mesure de distance naturelle entre paires de nœuds, définie par la longueur de leurs plus courts chemins. L'excentricité d'un nœud x est définie comme la somme des distances à tous les autres nœuds, et la proximité est définie par Bavelas comme l'inverse de l'éloignement.

$$C(x) = \frac{1}{\sum\limits_{y} d(x, y)}$$

Centralité de prestige ($C_{\textbf{Prestige}}$) : Cette métrique reflète l'importance d'un nœud dans un réseau social. Cette mesure assigne des scores relatifs à chacun des nœuds du réseau en se basant sur le principe que les connexions vers les nœuds ayant les scores les plus élevés, contribuent davantage au score du nœud en question que des connexions égales mais à de plus bas score. En d'autres termes, c'est la somme des degrés des voisins à deux sauts du nœud (directs et indirects).
La matrice suivante donne la longueur de l'ensemble des plus courts chemins de ce graphe :

	1	2	3	4	5	6	7	8	9	10
1	0	1	1	2	3	2	3	4	5	5
2	1	0	1	1	2	2	3	4	5	5
3	1	1	0	2	2	1	2	3	4	4
4	2	1	2	0	1	2	3	4	5	5
5	3	2	2	1	0	1	2	3	4	4
6	2	2	1	2	1	0	1	2	3	3
7	3	3	2	3	2	1	0	1	2	2
8	4	4	3	4	3	2	1	0	1	1
9	5	5	4	5	4	3	2	1	0	1
10	5	5	4	5	4	3	2	1	1	0

La matrice suivante présente tous les plus courts chemins :

	1	2	3	4	5	6	7	8	9	10
1	1	1.2	1.3	1.2.4	1.2.4.5	1.3.6	1.3.6.7	1.3.6.7.8	1.3.6.7.8.9	1.3.6.7.8.10
2		2	2.3	2.4	2.4.5	2.3.6	2.3.6.7	2.3.6.7.8	2.3.6.7.8.9	2.3.6.7.8.10
3			3	3.2.4	3.6.5	3.6	3.6.7	3.6.7.8	3.6.7.8.9	3.6.7.8.10
4				4	4.5	4.5.6	4.5.6.7	4.5.6.7.8	4.5.6.7.8.9	4.5.6.7.8.10
5					5	5.6	5.6.7	5.6.7.8	5.6.7.8.9	5.6.7.8.10
6						6	6.7	6.7.8	6.7.8.9	6.7.8.10
7							7	7.8	7.8.9	7.8.10
8								8	8.9	8.10
9									9	9.10
10										10

Les nombres de plus courts chemins, passant par chaque nœud, sont donnés comme suit :

	1	2	3	4	5	6	7	8	9	10
Nombre	0	3	10	2	5	21	18	14	0	0

Question : Compléter le tableau suivant et déduire les nœuds centraux du réseau :

	1	2	3	4	5	6	7	8	9	10
$C_{\text{Intermédiarité}}$	0	3	10	2	5	21	18	14	0	0
$C_{\text{Proximité}}$	1/26	1/24	1/20	1/25	1/22	1/17	1/19	1/23	1/30	1/30
C_{Prestige}	11	12	14	13	13	17	15	9	7	7

Les nombres de plus courts chemins, passant par chaque lien, sont donnés comme suit :

	1-2	1-3	2-3	2-4	3-6	4-5	5-6	6-7	7-8	8-9	8-10	9-10
Nombre chemins	3	6	6	5	16	8	10	24	21	8	8	1

Question : Divisez l'ensemble du réseau en 2, 3 et 4 sous-communautés, respectivement.

AIDE : utiliser une méthode qui divise la communauté en plusieurs sous-communautés en supprimant en priorité les liens reliant différentes sous-communautés (les liens parcourus par le plus grand nombre de plus-courts chemins).

En divisant en deux communautés, on obtient les sous-communautés suivantes : {1, 2, 3, 4, 5, 6} et {7, 8, 9, 10}.
Et on obtient le graphe suivant :

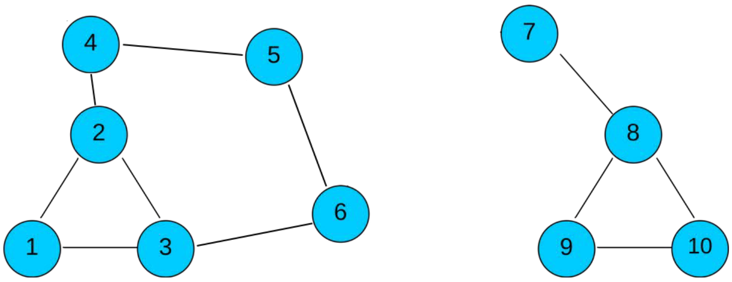

Figure 3 – Graphe G_2

Il faut donc maintenant calculer à nouveau le nombre des plus courts chemins passant par chaque sommet. Après calcul, on trouve que les arêtes 2-4 et 3-6 ont le plus grand nombre de plus courts chemins passant par eux (5 précisément). On choisit arbitrairement de supprimer le lien 2-4.

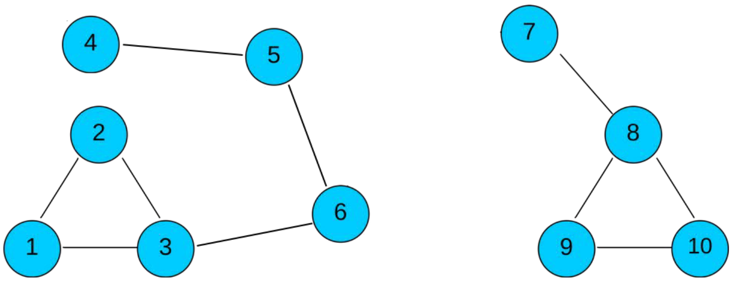

Figure 4 – Graphe $G_{2'}$

On réitère le processus et on supprime le lien 3-6. On obtient donc 3 communautés : $\{1, 2, 3\}$, $\{4, 5, 6\}$, $\{7, 8, 9, 10\}$.

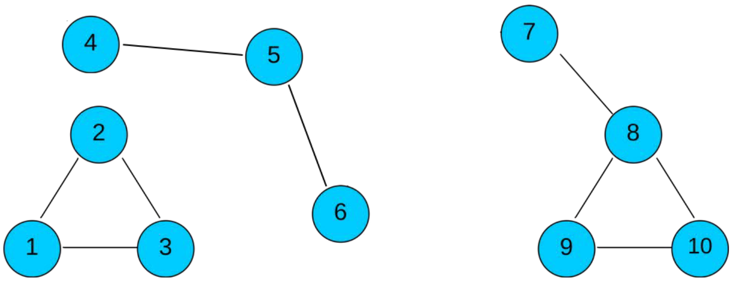

Figure 5 – Graphe G_3

Pour la dernière fois on applique l'algorithme et on supprime le lien 5-6.

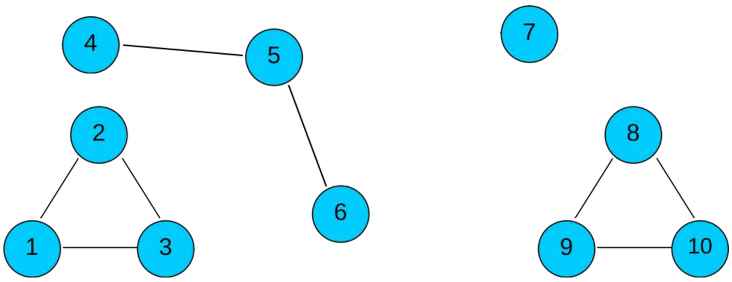

Figure 6 – Graphe G_4

On a donc découpé le graphe en 4 sous-communautés : $\{1, 2, 3\}$, $\{4, 5, 6\}$, $\{7\}$, $\{8, 9, 10\}$.

Question : Calculer la modularité pour chaque partition proposée. Quel est le meilleur partitionnement ?

RAPPEL : La modularité d'une partition c de la communauté globale en n_c sous-communautés est :

$$M_c = \sum_{c=1}^{n_c} \left[\frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \right]$$

où L_c est le nombre de liens au sein de la sous-communauté c, L est le nombre de liens au sein de la communauté globale et k_c est la somme des degrés des nœuds de la sous-communauté c.

Pour deux communautés, on a

$$M_{G_2} = \left[\frac{7}{11} - \left(\frac{14}{22} \right)^2 \right] + \left[\frac{4}{11} - \left(\frac{8}{22} \right)^2 \right] = \frac{56}{121} \simeq 0.4628099174$$

Pour trois communautés, on a

$$M_{G_3} = \left[\frac{3}{9} - \left(\frac{6}{18} \right)^2 \right] + \left[\frac{2}{9} - \left(\frac{4}{18} \right)^2 \right] + \left[\frac{4}{9} - \left(\frac{8}{18} \right)^2 \right] = \frac{52}{81} \simeq 0.6419753086$$

Pour quatre communautés, on a

$$M_{G_4} = \left[\frac{3}{8} - \left(\frac{6}{16} \right)^2 \right] + \left[\frac{2}{8} - \left(\frac{4}{16} \right)^2 \right] + \left[\frac{0}{8} - \left(\frac{0}{16} \right)^2 \right] + \left[\frac{3}{8} - \left(\frac{6}{16} \right)^2 \right] = \frac{21}{32} \simeq 0.65625$$

Ainsi, on remarque que la découpe en 4 sous-communautés est la découpe qui maximise la modularité ($M_{G_2} < M_{G_3} < M_{G_4}$). On en conclut donc que le meilleur partitionnement pour notre exemple est le partitionnement en 4 sous-communautés.