Structure et dynamique des réseaux

Cours 4 : Modèles de graphes

Clémence Magnien, Lionel Tabourier, Fabien Tarissan

LIP6 – CNRS and Université Pierre et Marie Curie

prenom.nom@lip6.fr

Outline

Graphes aléatoires – Motivation

1: comprendre la structure observée

Propriétés observées normales ?
Réponse : comparer à un graphe aléatoire synthétique

Tirer au hasard (proba. uniforme) dans l'ensemble des graphes (d'une taille donnée) $\rightarrow$ propriétés communes à l'immense majorité des graphes $\rightarrow$ propriétés attendues

2: simuler des processus

Aussi possible avec des graphes obtenus par un modèle génératif non-aléatoire

Outline

Modèle d'Erdös-Rényi

$$G_{n,p}$$

- n sommets
- Chaque arête existe avec probablité p

Modèle d'Erdös-Rényi

$$G_{n,p}$$

- n sommets
- Chaque arête existe avec probablité p

Complexité : $\mathcal{O}(n^2)$

Modèle d'Erdös-Rényi

$$G_{n,m}$$

- n sommets
- m arêtes choisies au hasard

Modèle d'Erdös-Rényi

$$G_{n,m}$$

- n sommets
- m arêtes choisies au hasard

Complexité : $\mathcal{O}(m)$

Équivalence entre $G_{n,p}$ et $G_{n,m}$

p représente la densité

$$p = \frac{2m}{n(n-1)}$$

$G_{n,m}$ et $G_{n,p}$ sont équivalents si p et m respectent cette relation.

Arêtes en double

$G_{n,m}$: probabilité de tirer des arêtes en double

Compliqué à détecter

- structure compacte
- complexité ?

En pratique

- peu d'arêtes en double
- ne change pas les propriétés du graphe

$\rightarrow$ on les traite comme des arêtes normales
on évite les boucles

Notion de propriété attendue

Exemple : graphe aléatoire, $n = m = 4950$

Résultat : clique de 100 sommets (les autres ont degré 0)

Étonnant ?

Notion de propriété attendue

Exemple : graphe aléatoire, $n = m = 4950$

Résultat : clique de 100 sommets (les autres ont degré 0)

Étonnant ?

Probabilité d'avoir degré 0 : $q = (1-p)^{n-1} \sim 0.14$.

Nombre attendu de sommets de degré 0 :

$$nq \sim 683$$

Notion de propriété attendue

Exemple : graphe aléatoire, $n = m = 4950$

Résultat : clique de 100 sommets (les autres ont degré 0)

Étonnant ?

Probabilité d'avoir degré 0 : $q = (1-p)^{n-1} \sim 0.14$.

Nombre attendu de sommets de degré 0 :

$$nq \sim 683$$

$683 \neq\neq\neq 4850$

→ on ne peut pas obtenir ce graphe par tirage aléatoire

(autre processus en jeu)

Propriétés des graphes Erdös-Rényi

- Densité
- Connexité

- Distance moyenne, diamètre

Propriétés des graphes Erdös-Rényi

- Densité fixée
- Connexité composante géante, taille $\mathcal{O}(n)$
  (pour $m \geq \mathcal{O}(n)$)
- Distance moyenne, diamètre $\sim \log(n)$
  (pour $m \geq \mathcal{O}(n)$)

Propriétés des graphes Erdös-Rényi

- Distribution des degrés
- Coefficient de clustering

Propriétés des graphes Erdös-Rényi

- Distribution des degrés homogène
- Coefficient de clustering $\simeq$ densité

Propriétés des graphes Erdös-Rényi

	réel	aléatoire
densité	faible	faible
connexité	comp. géante	comp. géante
distances	faibles	faibles
degrés	hétérogènes	homogènes
clustering	fort	faible
communautés	oui	non

Graphes d'Erdös-Rényi – Conclusion

Les graphes de terrain sont très différents des graphes aléatoires d'Erdös-Rényi

Conséquences

- Leur ressemblance est significative
- Les graphes d'ER ne sont pas des bons modèles (simulations, preuves, . . .)

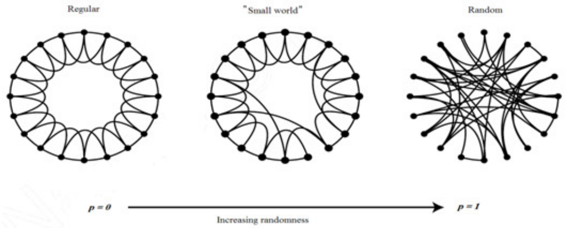
$\rightarrow$ Autres modèles ?

Outline

Modèle Small-World

Watts et Strogatz - *Nature, 1998*

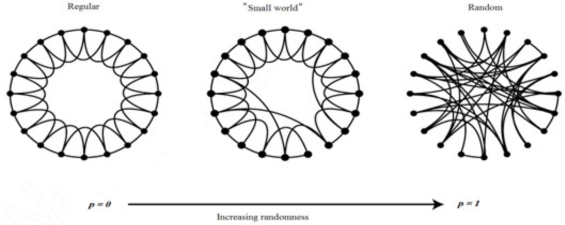
Depuis un réseau régulier, reconnexions aléatoires de liens selon une probabilité p:



Modèle Small-World

Watts et Strogatz - *Nature, 1998*

Depuis un réseau régulier, reconnexions aléatoires de liens selon une probabilité p:



Reproduit clustering élevé mais distribution non hétérogène...

Modèle *Scale-Free* (sans échelle)

Barabási et Albert - *Science, 1999*

Réseau créé selon la loi de l'attachement préférentiel: la probabilité pour un nœud i de se connecter à un nouvel arrivant est proportionnelle à son degré $\delta(i)$

Justification: ce type de processus permet d'obtenir une distribution de degré en loi de puissance (admis ici)

Autre justification (?): processus génératif des réseaux selon la règle *rich gets richer*, mais pertinence à discuter...

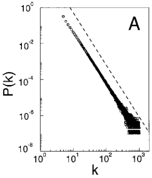
Modèle *Scale-Free* (sans échelle)

Algorithm 1: Génération d'un graphe de Barabási et Albert à n sommets et $m \simeq n.k$ liens

begin
  Paramètres: n , k
  Graphe de départ à n_0 sommets
  $i = n_0$
  while $i \leq n$ **do**
    ajouter au graphe le sommet i avec proba d'attachement à j:
    $P(i,j) = \frac{k.\delta(j)}{\sum_\alpha \delta(\alpha)}$
    $i = i + 1$
  end
end

Modèle *Scale-Free* (sans échelle)

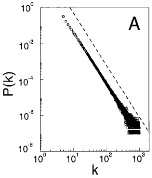
- distribution de degré hétérogène: en loi de puissance



- mais propriété cachée :

Modèle *Scale-Free* (sans échelle)

- distribution de degré hétérogène: en loi de puissance



- mais propriété cachée : le graphe a peu de cycles
  ex: pour $k = 1$*, c'est un arbre*

le graphe n'est pas un élément quelconque de l'ensemble des graphes à distribution de degré en loi de puissance...
Pas une méthode générative réaliste

Outline

Méthode générative

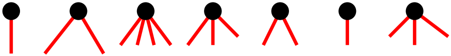
Distribution des degrés

$p_1, p_2, p_3, \ldots$

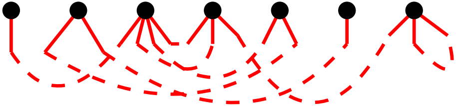
Tirer les degrés des sommets selon la distribution

1 2 4 3 2 1 3

Associer à chaque sommet des demi-arêtes



Tirer au hasard des paires de demi-arêtes



Implémentation algorithmique: méthode directe

Tableau : sommet i apparaît autant de fois que $\delta(i)$

0	1	1	2	2	2	2	3	3	3	4	4	5	6	6	6

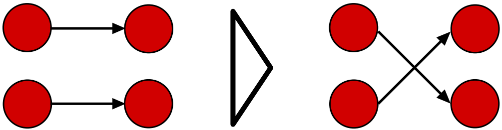
Algorithm 2: Génération d'un graphe à distribution de degré fixée

```
begin
    Choisir une paire de valeurs au hasard
    i = 2m
    while i > 0 do
        u = random (0 , i − 1)
        échanger cases u et i − 1
        v = random (0 , i − 2)
        échanger cases v et i − 2
        i = i-2
        // arête (u, v) créée
    end
end
```

Autre implémentation: méthode d'échanges

Principe

- on doit avoir un graphe respectant la distribution de degré
- on itère des permutations des extrémités d'arêtes
- après *suffisamment* d'échanges, le graphe produit est un élément quelconque de l'ensemble



Autre implémentation: méthode d'échanges

Pourquoi ça marche?

- On conserve le degré des nœuds
- Le processus est une chaîne de Markov

  on peut le voir comme une marche au hasard dans l'ensemble des graphes avec cette distribution de degré

  après un certain temps, visite tous les graphes avec la même probabilité (admis)

Comment savoir si on a effectué assez d'échanges?

mesurer des caractéristiques au cours du processus
évaluer quand celles-ci n'évoluent plus (ex: clustering)

Propriétés – Comparaison

	de terrain	densité fixée	d.d. fixée
densité	faible	faible	faible
connexité	comp géante	comp géante	comp géante
distances	faibles	faibles	faibles
degrés	hétérogènes	homogènes	hétérogènes
clustering	fort	faible	faible
communautés	oui	non	non

→ le clustering n'est pas une conséquence des degrés hétérogènes

Graphe à distribution de degrés fixé: le cas biparti

Newman, Watts et Strogatz - *PNAS, 2002*

Exemple du réseau *Internet Movie Data Base*: que signifie un lien acteur-acteur?

Représentation plus riche: réseau acteur/film

Vocabulaire

- graphe biparti: 2 sous-ensembles de sommets A et B, des liens uniquement entre sommet A et sommet B
- le réseau des acteurs est une projection de ce réseau

Graphe à distribution de degrés fixé: le cas biparti

Méthode générative directe (et à base d'échanges) s'appliquent de la même manière, avec:

- deux distributions de degré (pour sommets A et B)
- en ne connectant que des nœuds de A à ceux de B

Résultats

- justifie le clustering et le degré dans une projection pour certains graphes

  dans Newman *et al.*: *coboarding* ok, pas réseaux de collaborations
- pas de structure à grande échelle (communautés)

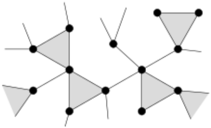
Outline

Exemple de modèle aléatoire plus élaboré

Newman - *Physical Review Letters, 2009*

Distributions de degré + clustering

- pour chaque nœud, on attribue:
  moitiés de liens simples (s_i),
  "coins" de triangles (t_i)
  selon des distributions décorrélées
- assembler les moitiés de liens et les coins séparément



clustering vérifié, structure communautaire déficiente

Contraintes quelconques et k-échanges de liens

Tabourier, Roth et Cointet - *Journal of Experimental Algorithmics, 2011*

Méthode d'échanges pour des contraintes quelconques

Problème: en général, pas de garantie d'atteindre tous les graphes de l'ensemble par la marche aléatoire

Solution

- plutôt que d'échanger les extrémités de 2 arêtes, échanger k arêtes: permet d'atteindre plus d'éléments
- il existe k tel qu'on atteint tous les éléments de l'ensemble (admis)
- mais comment trouver ce k en pratique?