Structures et Dynamique des Réseaux

Propriétés dynamiques

Clémence Magnien, Lionel Tabourier, Fabien Tarissan

LIP6 – CNRS and Université Pierre et Marie Curie

prenom.nom@lip6.fr

Plan

Outline

Motivations

Les graphes qu'on étudie sont dynamiques :

Apparitions et disparitions

- de nœuds
- de liens

Question

Comment décrire la dynamique ?

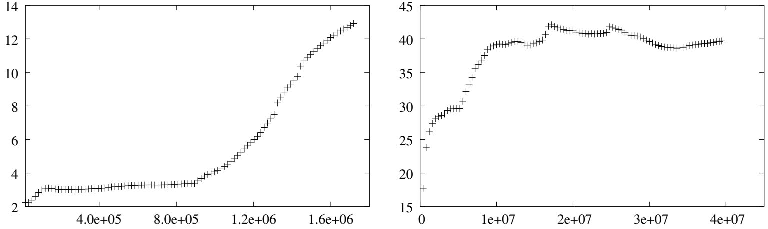
Approche naturelle

Utiliser les propriétés vues pour la description d'un réseau classique

- degrés
- clustering
- communautés
- ...

Exemple

Évolution du degré moyen pour deux graphes différents



$\longrightarrow$ Donne des informations importantes

Inconvénients

Manque les propriétés intrinsèquement dynamiques

- Durées de vie
- Chemins temporels
- ...

Inconvénients

Manque les propriétés intrinsèquement dynamiques

- Durées de vie
- Chemins temporels
- ...

Quelques propriétés déjà vues

On a abordé la question dans le cours précédent

- Nombre de nœuds distincts observés
- Présences
- Blocs de présence

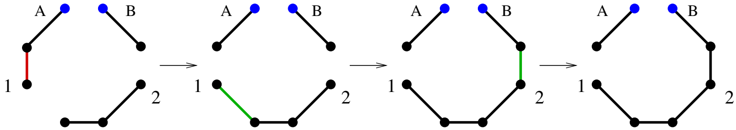
Types de données

Deux grandes familles de jeux de données

- Mesures périodiques (ex : radar)
- Suite d'événements (ex : échanges d'e-mail)

Plan

Chemin dynamique (1)



Chemin de A vers B
Pas de chemin de B vers A

Grosse différence avec la distance statique

Chemin temporel – définition

Définition intuitive

Suite de nœuds $u_1, \ldots, u_k$ telle que :

- il existe des liens (u_1, u_2) au temps t_1, (u_2, u_3) au temps t_2, ...
- $t_1 < t_2 < \ldots$

Variantes

Plusieurs variantes selon les auteurs :

- Parcours d'un lien instantané
  (nombre de parcours de liens illimité à un instant donné)
- Temps de parcours δ donné en paramètre

Plus ou moins réaliste selon les contextes
Plus ou moins facile à calculer

Distance

Trois définitions :

- plus petit nombre de sauts
- plus petit temps d'arrivée
- plus petit temps de transfert

Toutes les notions utiles selon les contextes
On se concentrera sur le plus petit temps d'arrivée

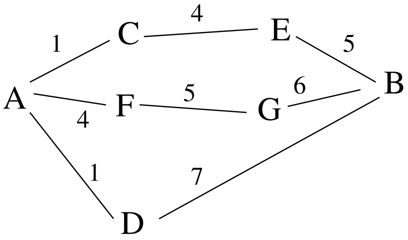
Distance – définition

Définition

i, j des nœuds, t un temps.
Soit t_a le plus petit temps d'arrivée d'un message de i vers j (s'il existe)

La distance de i vers j au temps t est : $t_a - t$

Distance – exemples



Distance de A à B

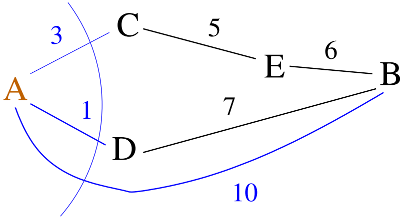
- Plus petit temps d'arrivée : A – C – E – B
- Plus petit nombre de sauts : A – D – B
- Plus petit temps de trajet : A – F – G – B

La distance dépend du temps de départ

Distance – calcul

Algorithme naïf : transmission d'un nœud vers tous les autres

Tant que tous les sommets ne sont pas visités :
prendre le premier lien de la frontière (u, v, t)
Si v non visité :
marquer v comme visité au temps t

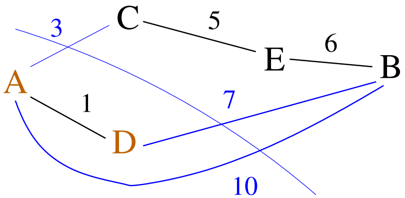


A 1

Distance – calcul

Algorithme naïf : transmission d'un nœud vers tous les autres

Tant que tous les sommets ne sont pas visités :
  prendre le premier lien de la frontière (u, v, t)
  Si v non visité :
    marquer v comme visité au temps t

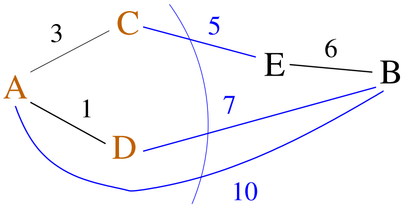


A 1
D 1

Distance – calcul

Algorithme naïf : transmission d'un nœud vers tous les autres

Tant que tous les sommets ne sont pas visités :
  prendre le premier lien de la frontière (u, v, t)
  Si v non visité :
    marquer v comme visité au temps t

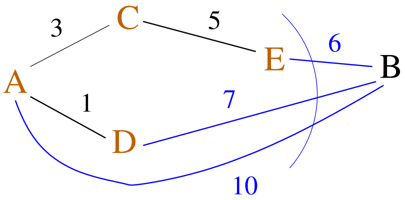


A 1
D 1
C 3

Distance – calcul

Algorithme naïf : transmission d'un nœud vers tous les autres

Tant que tous les sommets ne sont pas visités :
  prendre le premier lien de la frontière (u, v, t)
  Si v non visité :
    marquer v comme visité au temps t

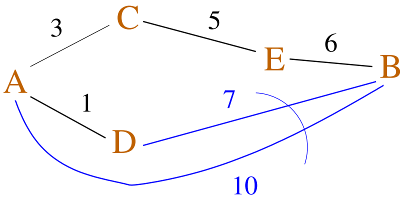


A 1
D 1
C 3
E 5

Distance – calcul

Algorithme naïf : transmission d'un nœud vers tous les autres

Tant que tous les sommets ne sont pas visités :
  prendre le premier lien de la frontière (u, v, t)
  Si v non visité :
    marquer v comme visité au temps t



A 1
D 1
C 3
E 5
B 6

Complexité de l'algorithme naïf

Algorithme naïf : transmission d'un nœud vers tous les autres

Tant que tous les sommets ne sont pas visités :
  prendre le premier lien de la frontière (u, v, t)
  Si v non visité :
    marquer v comme visité au temps t

Complexité

Tant que : regarder (presque) tous les liens : m

premier lien de la frontière : $\sim \log m$

$\sim m \log m$ étapes

Meilleur algorithme

Calcul en une passe des distances :

- pour toutes les paires de nœuds
- pour tous les temps de départ

On stocke le temps d'arrivée d'un message (plutôt que la distance)

Idée

- On a calculé tous les temps d'arrivée pour tous les temps de départ $> t$
- On traite un lien (u, v) au temps $t - 1$
  - u et v peuvent s'attendre en un temps nul à $t - 1$
  - pour chaque sommet $x \neq u, v$
    - *dux* : distance de u à x, *dvx* : distance de v à x
    - si $dux < dvx$ alors u peut utiliser v pour atteindre x plus tôt
    - et inversement

Algorithme

On trie les liens par temps décroissant

On stocke deux matrices $n \times n$: `dist` et `prev_dist`
(initialisées à -1 : pas possible d'envoyer un message)

`t_cur` = dernier temps dans les données

`dist[x][x] = t_cur, prev_dist[x][x] = t_cur` pour tout `x`

Algorithme (2)

Pour chaque lien `(u,v)` au temps `t`

```
Si t != t_cur
    Copier dist dans prev_dist
    cur_t = t
    dist[x][x] = t_cur pour tout x
dist[u][v] = cur_t, dist[v][u] = cur_t
Pour tout x != u,v
    Si prev_dist[u][x] != -1 et prev_dist[v][x] != -1
        Si dist[u][x] > prev_dist[v][x]
         dist[u][x] = prev_dist[v][x]
        Sinon, si dist[v][x] > prev_dist[u][x]
         dist[v][x] = prev_dist[u][x]
    Sinon, si prev_dist[u][x] != -1
        dist[v][x] = prev_dist[u][x]
    Sinon, si prev_dist[v][x] != -1
        dist[u][x] = prev_dist[v][x]
```

Exemple

Au tableau

Sortie de la distance

Pour utiliser les résultats :
la distance de i à j est égale à `dist[i][j] - t_cur`

Note : distance moyenne : attention aux débordement de capacité quand on fait la somme des distances
$\longrightarrow$ utilisation d'un `double`

Plan

Réseau de capteurs *Rollernet*

Tournoux et al. - *INFOCOM, 2009*

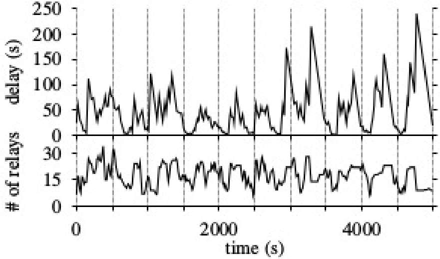
Mesure de données

- Mesurer les communications possibles entre personnes
- Chaque personne munie d'un capteur bluetooth
- Enregistrement périodique du voisinage

Randonnée de roller à Paris

- 62 nœuds
- 180 minutes

Délai moyen



Changements de délai dus aux phases d'arrêt et de redémarrage de la rando
Impact sur la qualité des communications

Plan

Graphes d'accessibilité dynamique

Whitbeck et al. - *MOBICOM, 2012*

But : mieux comprendre les graphes dynamiques

Graphe d'accessibilité : graphe dynamique orienté

Définition : $R_\delta(t)$

Étant donné δ :

$(u, v) \in R_\delta(t)$ si il existe un chemin de u vers v :

- partant au temps t
- arrivant avant $t + \delta$

Note : supposition de l'article :

suivre un lien prend un temps τ

Jeux de données

Rollernet

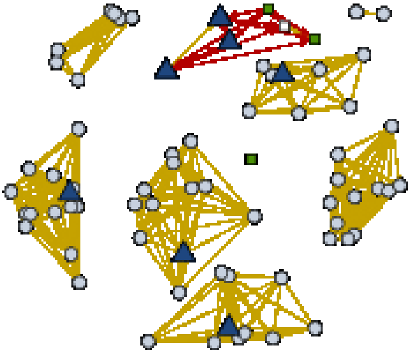
Vu plus haut

Stanford

- Une journée dans un lycée
- 782 nœuds

Observations

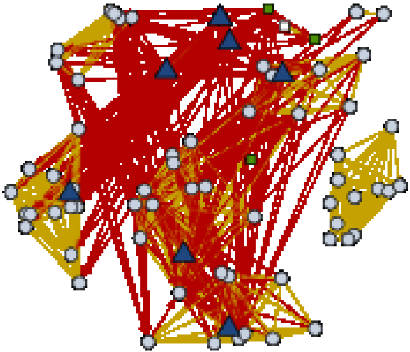
Jeu de données *Stanford* (classes) – $\delta = 20$ minutes



Triangles : professeurs
Cercles : étudiants

Observations

Jeu de données *Stanford* (classes) – $\delta = 40$ minutes



Triangles : professeurs
Cercles : étudiants

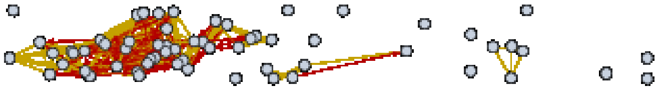
Commentaires

Observation

Étudier le graphe d'accessibilité dynamique permet de retrouver des groupes cohérents
(si δ est bien choisi)

Observations

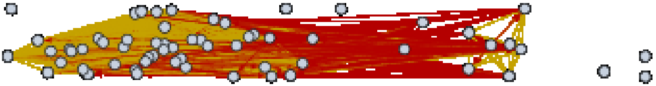
Jeu de données *Rollernet* – $\delta = 10$ secondes
Phase d'accélération



Liens rouges : asymétriques

Observations

Jeu de données *Rollernet* – $\delta = 60$ secondes
Phase d'accélération



Liens rouges : asymétriques

Commentaires

Observation

Impossible d'envoyer des messages rapidement entre la tête et la queue de la randonnée
Communication plus lente de l'avant vers l'arrière possible
Forte asymétrie

Ensemble dominant dynamique

Définition

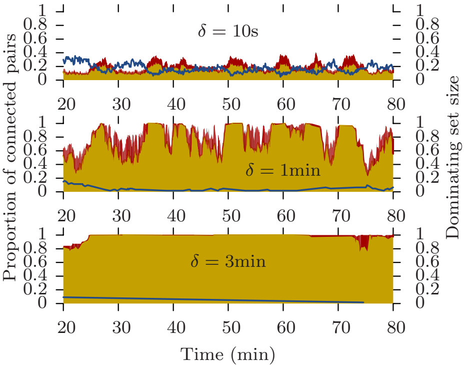
Ensemble D de sommets tel que tout sommet v est :

- soit dans D
- soit a un voisin dans D

Note : le calcul est NP-complet, besoin d'une heuristique
Étude de l'ensemble dominant pour le graphe d'accessibilité (ensemble dynamique)

Observations plus détaillées

Rollernet



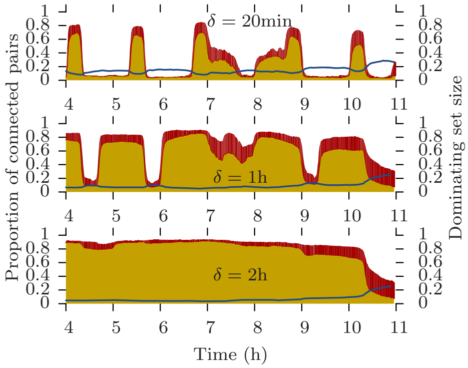
- Fraction de paires connectées (symétriques et asymétriques)
- taille de l'ensemble dominant

Remarques

- Communication avec un délai de moins d'une minute pas toujours possible
- 3 minutes suffisant pour permettre des communications efficaces
- la taille de l'ensemble dominant décroît avec δ

Observations plus détaillées

Stanford



- Fraction de paires connectées (symétriques et asymétriques)
- taille de l'ensemble dominant

Remarques

- Alternance de phases statiques (cours) et dynamiques
- Il reste de l'asymétrie
- la taille de l'ensemble dominant décroît avec δ

Plan

Centralité d'intermédiarité

But : quantifier l'importance d'un nœud
Rappel : Soit v un nœud. On appelle *centralité d'intermédiairité* la valeur :

$$BC(v) = \sum_{s,t \in V, s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

avec:

- σ_{st} : nombre de plus courts chemins entre s et t
- $\sigma_{st}(v)$: nombre de plus courts chemins entre s et t passant par v

Centralité dynamique ?

Centralité dynamique – plusieurs propositions

Évolution de la centralité statique

Calculer la centralité statique sur chaque intervalle de temps

- dépend de la taille des intervalles
- ne prend pas en compte la réalité des communications dans la plupart des cas

Extension aux plus courts chemins temporels

Pour un nœud i, calculer la fraction de plus courts chemins temporels qui passent par i.

- les chemins temporels dépendent énormément du temps de départ

Temporal Betweenness Centrality

Nicosia et al. - *in* Temporal Networks*, 2013*

Prise en compte du temps d'attente sur un nœud :

Si on a un seul plus court chemin de i vers j :

$i \longrightarrow k \longrightarrow j$

l'importance de k dépend du temps pendant lequel il stocke le message

Temporal Betweenness Centrality – définition

Nicosia et al. - *in* Temporal Networks*, 2013*

Prise en compte du temps d'attente sur un nœud :

$$C_i(t_m) = \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} \frac{U(i, t_m, j, k)}{\sigma_{jk}}$$

- $U(i, t_m, j, k)$: nombre de chemins temporels de j vers k tels que on passe par i à un temps $\leq t_m$
- σ_{jk} : nombre de plus courts chemins temporels de j vers k

Centralité moyenne : moyenne sur tous les instants

Temporal Betweenness Centrality – inconvénients

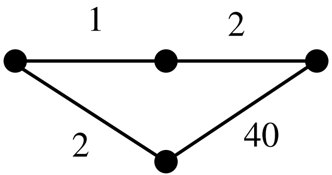
- Tous les chemins partent toujours du début des données.
- Seule la valeur moyenne est regardée

Médiation

Tang et al. - *in* Temporal Networks, *2013*

Idée

Si k est sur un plus court chemin temporel entre i et j son importance dépend du deuxième plus court chemin temporel.



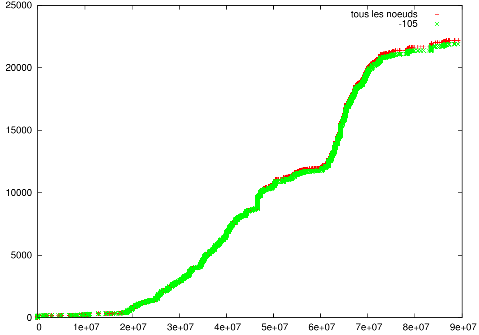
Médiation – en pratique

Principe

- Calculer les distances entre toutes les paires de nœuds
- Supprimer le nœud *i*
- Recalculer toutes les distances
  Différence : importance du nœud *i*

Médiation - exemple

Distribution cumulative des distances



Besoin de prendre en compte les chemins qui démarrent à tous les instants

Importance d'un nœud – autres proposition

D'autres propositions existent

- Centralité de proximité et extensions
- ...

Pas de consensus

Conclusion

Les propriétés statiques ne sont pas suffisantes pour décrire un réseau dynamique

Dans ce cours :

- propriétés centrées autour des chemins temporels

Autres propriétés

- Motifs temporels
- Durée de vie des nœuds et des liens
- Persistence des liens
- Communautés dynamiques
- ...