



Analyse des réseaux sociaux

par

Dr. Samira LAGRINI



Année universitaire:2024/2025

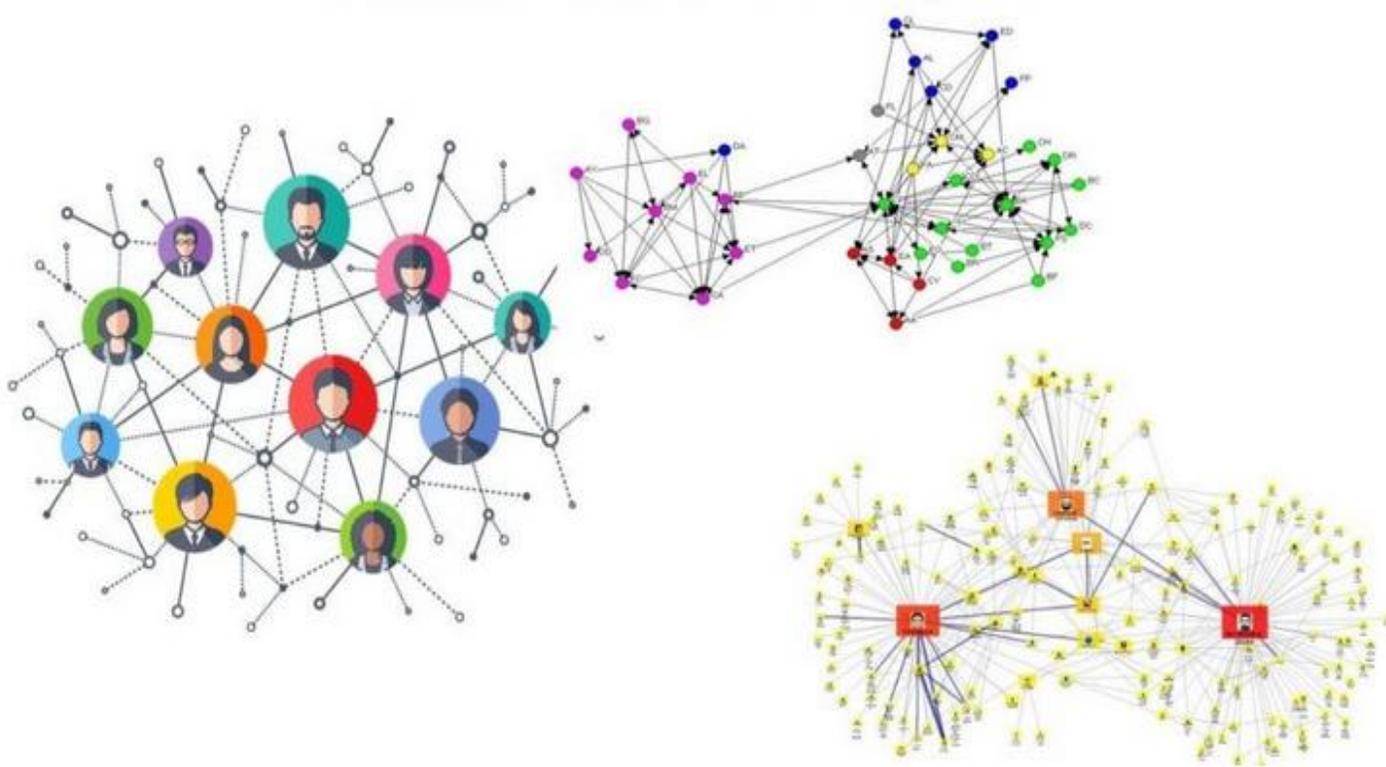
Introduction

- l'**Analyse de Réseaux Sociaux (ARS)**, ou **Social Network Analysis (SNA)** est une méthode d'analyse qui consiste à **étudier et à visualiser** les relations et les interactions entre différentes entités sur le web et comment ces entités interagissent et s'influencent mutuellement
- La SNA s'inscrit dans la cadre du **Web Structure mining** qui vise à explorer et à analyser les relations structurelles entre les entités web en se concentrant sur les hyperliens et les interactions entre les entités.

Introduction

La SNA exploite les données de la structure web pour :

- ✓ Identifier les **acteurs influents** dans un réseau (utilisateurs avec une forte centralité).
- ✓ Détecter les **communautés** au sein des interactions.
- ✓ Étudier la **dynamique des réseaux** pour comprendre les flux d'information et d'influence.



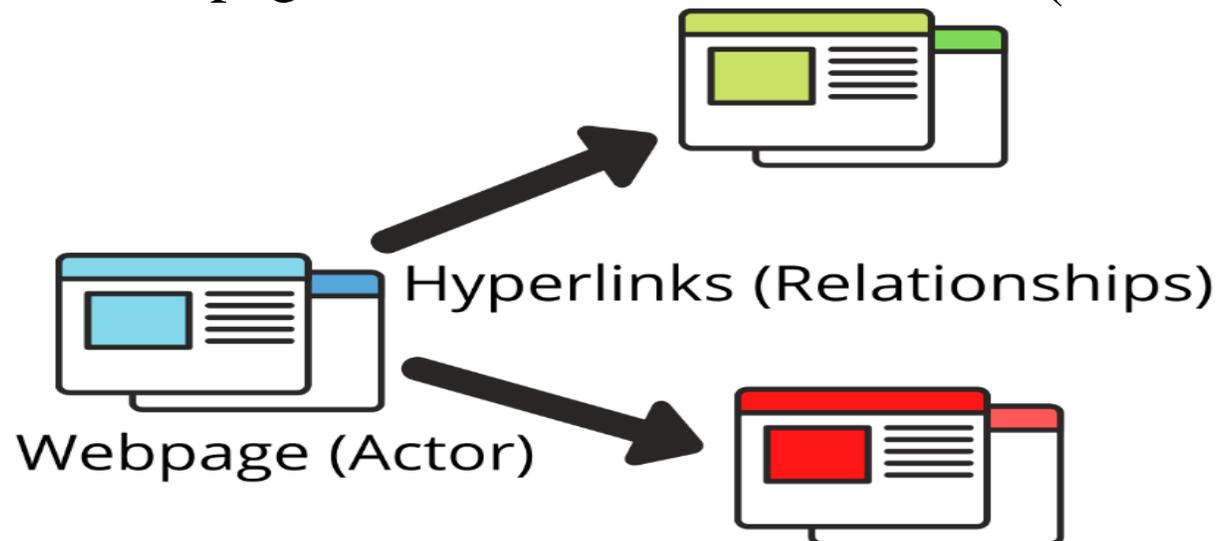
Concepts de Base

❑ Le SNA est un processus de découverte qui utilise des graphes pour représenter les réseaux, où les **nœuds** symbolisent les entités et les **liens** représentent les relations.

❑ **Les Nœuds (Nodes)** Représentent les acteurs individuels dans le réseau (ex. une personne ou une page web).

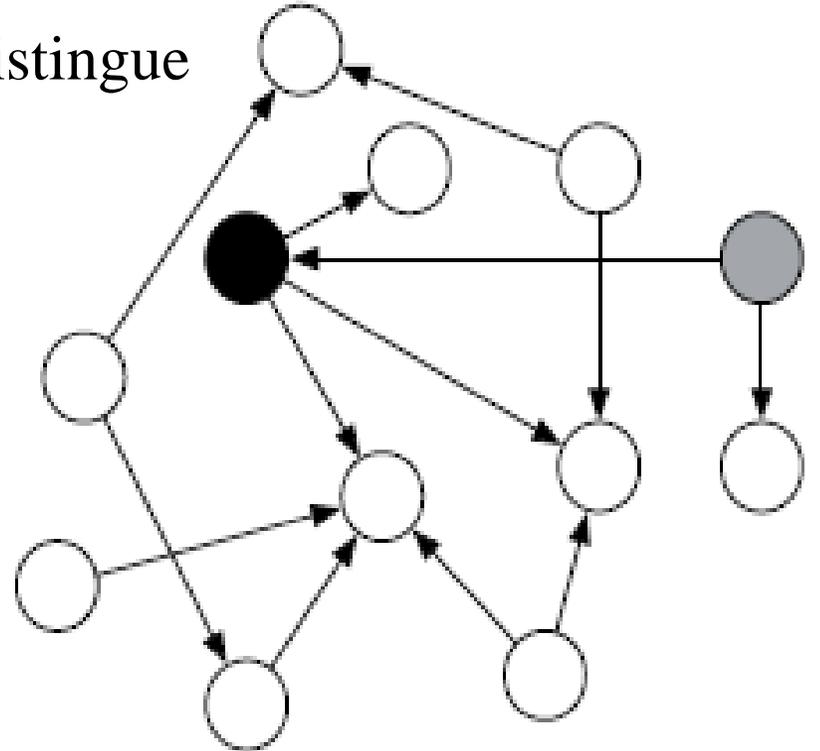
❑ **Les Liens (Edges)** Représentent les relations ou interactions entre les nœuds (ex. amitié, collaboration).

.....



Concepts de Base

- ❑ Les Graphes : des structures qui modélisent les relations entre des entités (nœuds).
- ❑ En fonction de la nature des liens entre les nœuds, on distingue les trois types de graphes :
 - Les graphes non orientés
 - Les graphes orientés
 - Les graphes pondérés

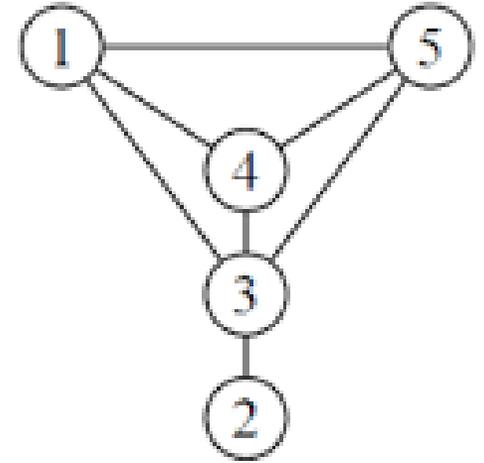


Concepts de Base -graphes non orientés-

□ Un graphe non orienté est un type de graphe où les liens entre les nœuds n'ont pas de direction (la relation est symétrique).

Exemple

La relation d'amitié entre deux personnes.



Utilisation

Modéliser des réseaux sociaux où les relations sont réciproques tel que:

- les réseaux d'amitié,
- Les réseaux de co-auteurs.



Représentation graphique

$$M = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Matrice d'adjacence

Concepts de Base -graphes orientés-

Un graphe orienté est un type de graphe où les liens entre les nœuds ont une direction. Cela signifie que la relation n'est pas forcément réciproque.

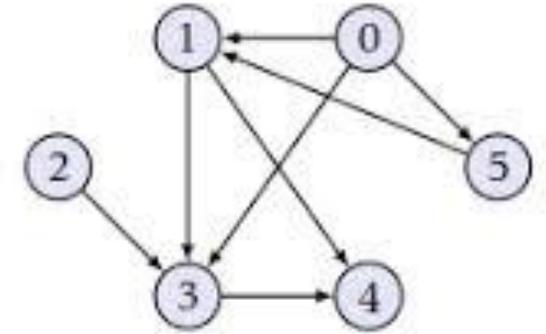
Exemple

- Sur Twitter, si A suit B, cela ne signifie pas que B suit A.
- Un e-mail envoyé d'une personne à une autre.

Utilisation

Modéliser des **relations asymétriques** tel que :

- des réseaux de communication
- des flux d'information,
- des relations hiérarchiques.



↓

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

Matrice d'adjacence



Concepts de Base -graphes pondéré-

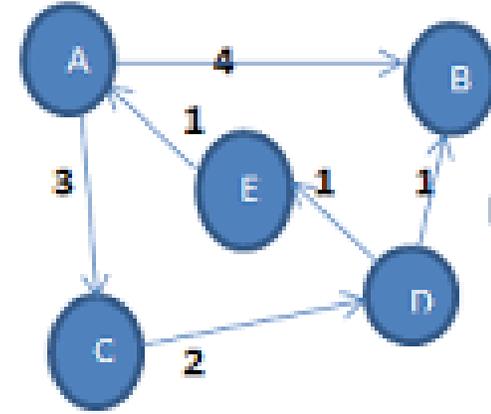
- Un graphe pondéré est un type de graphe où chaque lien est associé à un poids.
- Le poids peut représenter la force, l'intensité, la fréquence ou la durée de la relation.

Exemple

- Dans un réseau social, le poids d'un lien peut représenter le nombre de messages échangés entre deux personnes.
- Dans un réseau de transport, le poids peut indiquer la distance entre deux villes.

Utilisation

- Analyser la force des interactions dans un réseau(l'intensité de la collaboration entre des chercheurs ou la force des relations commerciales entre entreprises).



↓

	A	B	C	D	E
A	0	4	3	0	0
B	0	0	0	0	0
C	0	0	0	2	0
D	0	1	0	0	1
E	1	0	0	0	0

Matrice d'adjacence

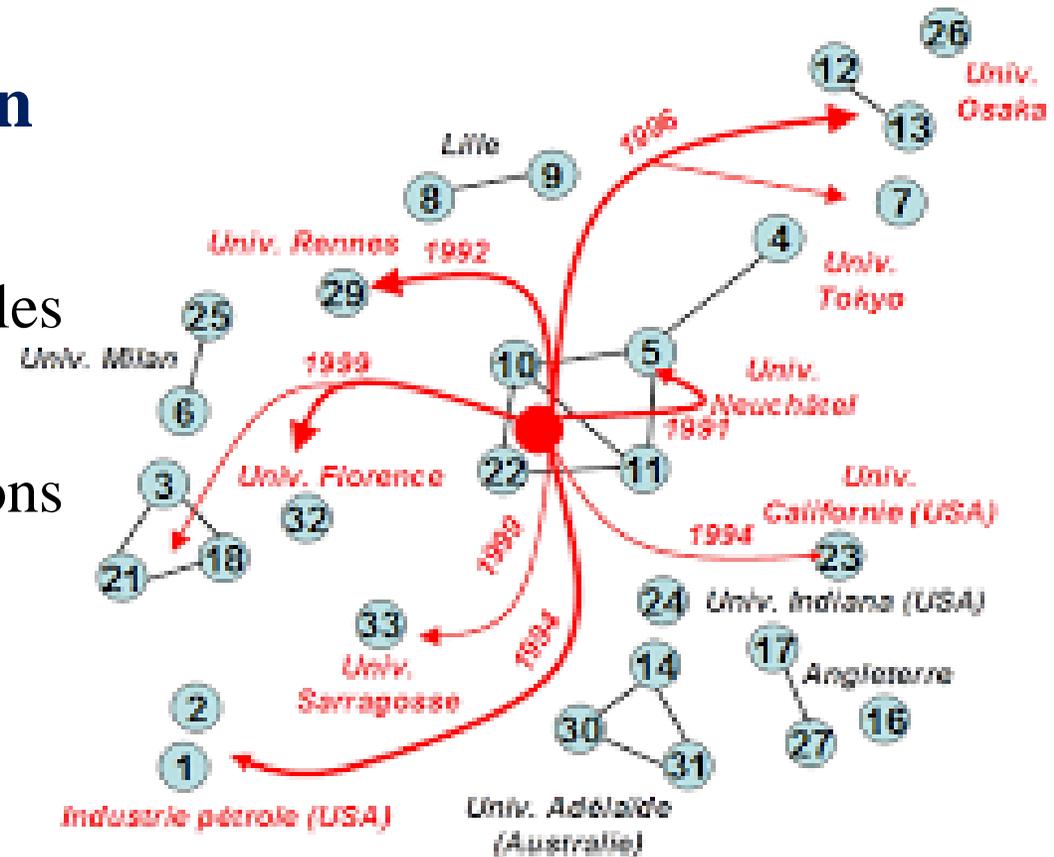
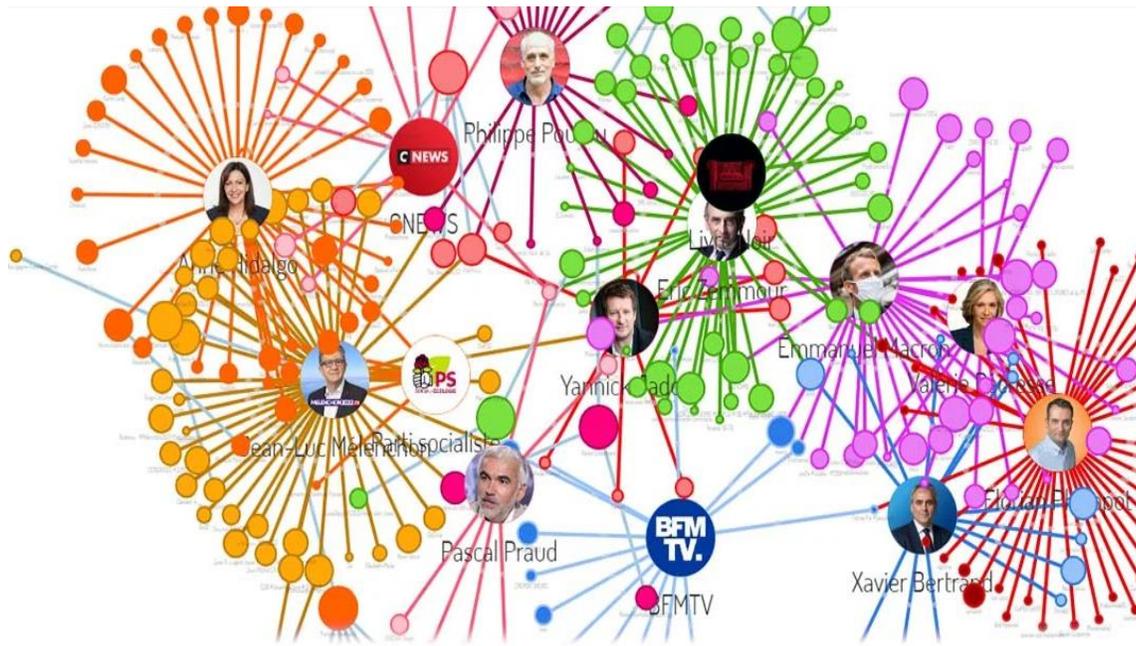
Applications de SNA

La SNA permet d'analyser non seulement les interactions humaines, mais aussi des systèmes complexes comme les réseaux informatiques, financiers et biologiques.



❑ Recherche Académique et Collaboration Scientifique

- Étudier la collaboration entre chercheurs en analysant les réseaux de co-publication.
- Identifier les leaders de domaine et les collaborations interdisciplinaires.



❑ Analyse des Médias Sociaux

- Étudier la propagation de l'information et des fake news sur les réseaux sociaux.
- Analyser les communautés d'utilisateurs pour mieux comprendre les comportements en ligne.

❑ Marketing et Détection des Influenceurs

- Identifier les influenceurs dans les réseaux sociaux pour maximiser l'impact des campagnes de marketing.



❑ Sécurité et Intelligence

- Détection des réseaux criminels

- Analyse des schémas de fraude dans le secteur financier: identifier des comportements suspects, comme des transferts d'argent anormaux entre comptes, en détectant des connexions inhabituelles qui pourraient indiquer une fraude.

❑ Santé Publique

❑ Étude de la propagation des maladies en identifiant les individus responsables de la transmission rapide des infections.

❑

Métriques en Analyse des Réseaux Sociaux

- Les **métriques** permettent d'évaluer différentes propriétés des nœuds des relations dans un réseau.
- Elles sont essentielles pour comprendre la structure du réseau, identifier les acteurs clés, et analyser les dynamiques de communication et d'influence

Centralité de degré

□ Mesure le nombre de connexions directes qu'un nœud a dans le réseau. Pour le graphe orienté, on distingue:

- ✓ Degré entrant (in-degree) : Nombre de liens entrants vers un nœud
- ✓ Degré sortant (out-degree) : Nombre de liens sortants depuis un nœud,

□ La centralité de degré d'un nœuds ($CD(i)$) est mesuré par son degré $d(i)$ normalisé par le degré maximum ($n-1$)

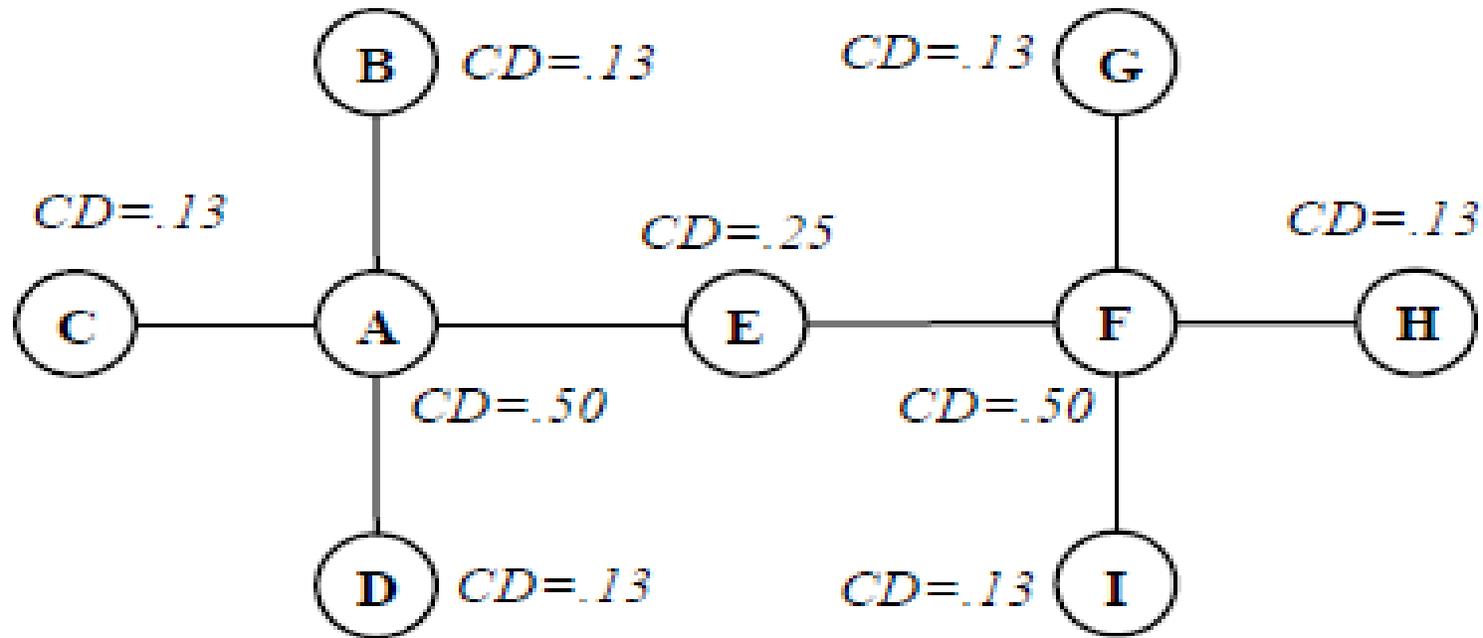
$$C_D(i) = \frac{d(i)}{n-1}.$$

▪ n : le nombre de nœud dans le graphe

Exemple : Dans un réseau social, une personne ayant un grand nombre d'amis ou de suiveurs aura une centralité de degré élevée.

Centralité de degré

Exemple:



les noeuds A et F, ayant quatre liens chacun, sont les plus importants

Centralité de proximité

- Basée sur l'idée qu'un acteur est important s'il est capable de contacter facilement un grand nombre d'acteurs avec un minimum d'effort (l'effort ici est relatif à la longueur des chemins).

Exemple : Dans une entreprise, un employé ayant une centralité de proximité élevée est capable de communiquer rapidement avec de nombreuses personnes.

- La centralité de proximité d'un nœud v_i est défini par:

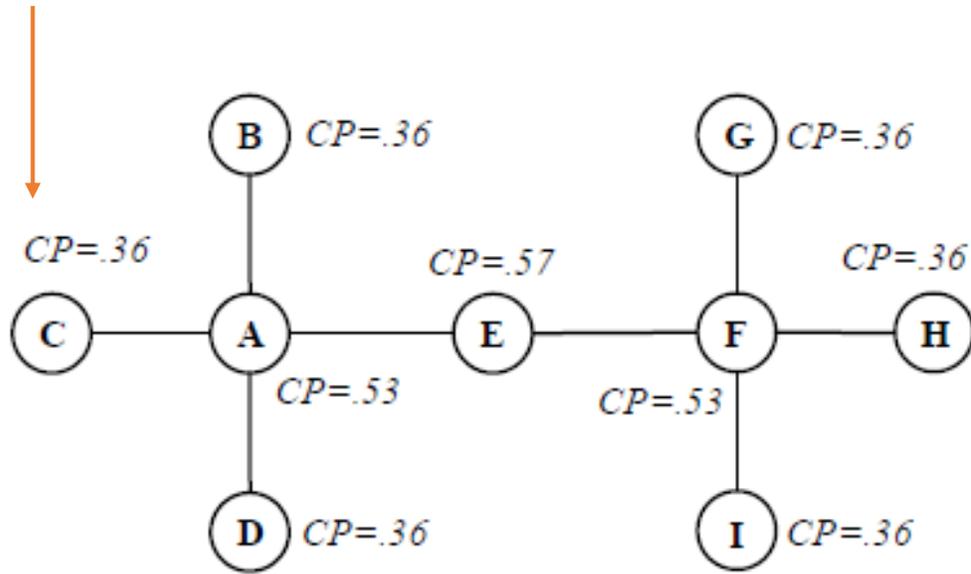
$$C^{pro}(v_i) = \frac{N-1}{\sum_{j=1}^N dist(v_i, v_j)}$$

Dist (v_i, v_j) est la distance (nombre d'arcs) entre les deux sommets v_i et v_j

Centralité de proximité

Graphe non orienté

$$CP(C) = 8 / (1 + 2 + 2 + 2 + 3 + 4 + 4 + 4) = 0,36$$

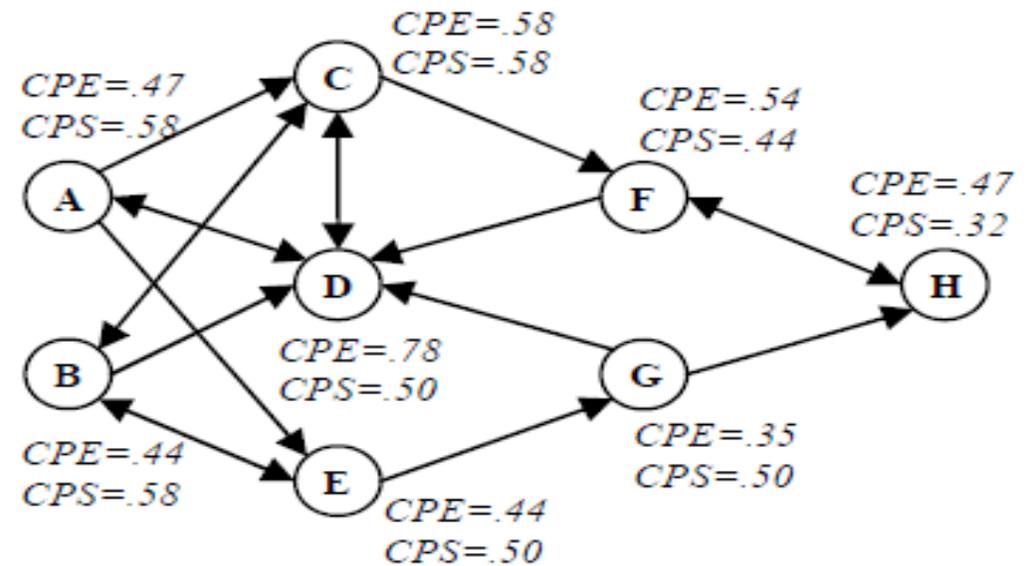


le noeud E est le plus important dans le graphe G1 alors qu'il n'a que deux liens

Graphe orienté

Identique au cas non orienté mais en prenant en compte la direction des arcs

$$C_{out}^{pro}(v_i) = \frac{N-1}{\sum_{j=1}^n dist(v_i, v_j)} ; C_{in}^{pro}(v_i) = \frac{N-1}{\sum_{j=1}^n dist(v_j, v_i)}$$



Centralités de proximité entrante (CPE) et sortante (CPS)

Centralité d'Intermédiation (Betweenness Centrality)

- Mesure le nombre de fois qu'un nœud se trouve sur le chemin le plus court reliant deux autres nœuds du réseau.
- **Interprétation** : Les nœuds avec une centralité d'intermédiation élevée jouent souvent un rôle de médiateur ou de "pont" entre différentes parties du réseau. Ils sont cruciaux pour la transmission d'informations.
- **Exemple** : Dans un réseau criminel, les acteurs ayant une centralité d'intermédiation élevée sont souvent les intermédiaires ou les coordinateurs.

Densité du Réseau (Network Density)

- La densité du réseau est le rapport entre le nombre de liens présents et le nombre maximal de liens possibles dans le réseau.
- **Interprétation** : Une densité élevée indique un réseau très connecté, tandis qu'une densité faible suggère peu de connexions entre les nœuds.
- **Exemple** : Un groupe de collègues travaillant étroitement ensemble aura une densité élevée, tandis qu'un réseau de personnes venant de différentes équipes aura une densité plus faible.

Densité du Réseau (Network Density)

- Densité = Nombre de liens existants / Nombre total de liens possibles
- Pour un **réseau non orienté** avec n nœuds, le nombre total de liens possibles est :

$$n(n-1)/2$$

- Pour un **réseau orienté**, le nombre total de liens possibles est :

$$n(n-1)$$

Modularité (Modularity)

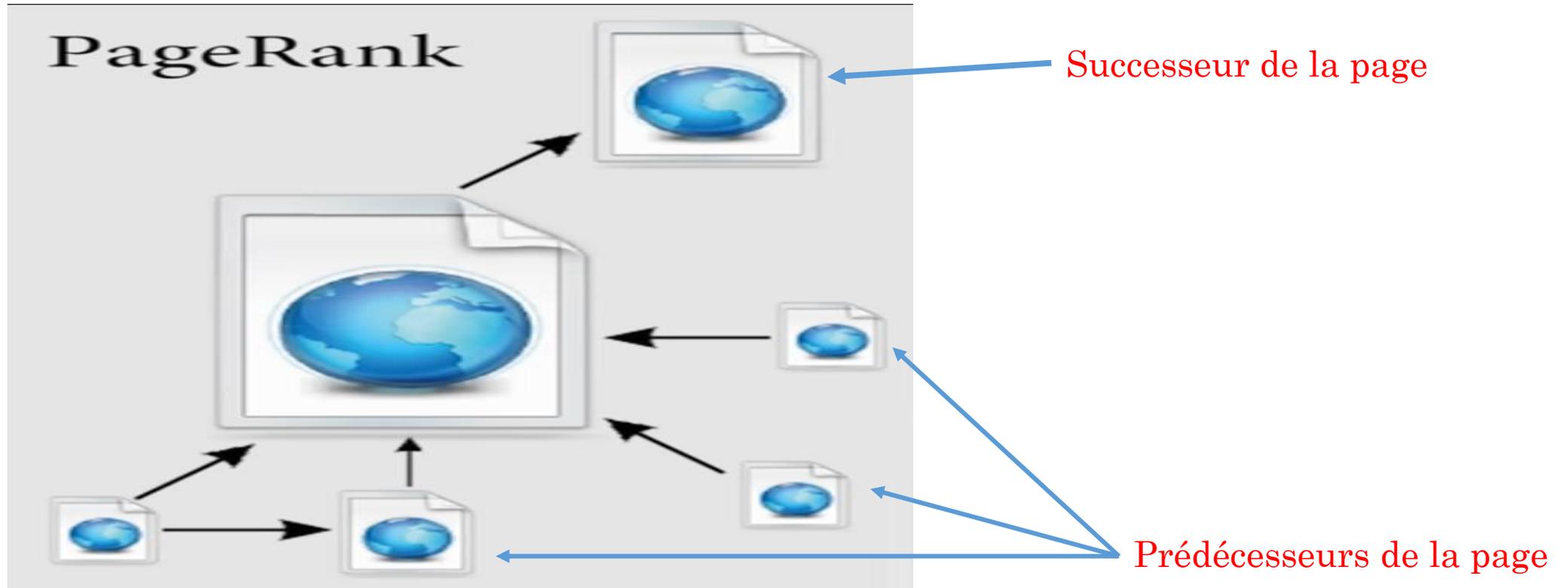
- La modularité mesure la force de la division d'un réseau en communautés ou groupes.
- Elle évalue si le réseau peut être partitionné en sous-groupes densément connectés.
- **Interprétation** : Une modularité élevée indique la présence de communautés bien définies.
- **Exemple** : Dans les réseaux sociaux, les communautés peuvent représenter des groupes d'intérêts ou des sous-cultures.

PageRank

- un algorithme d'analyse des réseaux développé par Larry Page et Sergey Brin, les fondateurs de Google.
- Il est utilisé pour mesurer l'importance relative d'un nœud dans un graphe orienté, basé sur la quantité et la qualité des liens entrants.
- L'idée principale est que si une page est pointée par de nombreuses pages importantes, elle est probablement aussi importante.

PageRank

PageRank considère que l'importance (la popularité) d'une page est fonction des popularités des pages qui la **citent**



PageRank

La popularité ou *le page rank* d'une page (p) est calculée comme suit :

$$PageRank(u) = (1 - d) / N + d \sum_{(v,u) \in E} \frac{PageRank(v)}{OutDegree(v)}$$

d: '*Damping factor*' compris entre 0 et 1, généralement d=0,85

N: Nombre de pages (nœuds) dans le graphe

OutDegree (v): le nombre de liens sortants de la page v

PageRank(v): le page rank d'une page v qui cite la page u (*prédécesseur de la page u*)

Étapes de l'algorithme Page rank

1. Initialiser tous les page avec la même importance (score=1/N). (N: nombre de page dans le graphe)
2. Recalculer les nouveaux scores des pages en utilisant la formule suivante:

$$PageRank(u) = (1 - d) / N + d \sum_{(v,u) \in E} \frac{PageRank(v)}{OutDegree(v)}$$

// A chaque itération, chaque page partage son importance avec ses successeurs (et reçoit de l'importance de ses prédécesseurs)

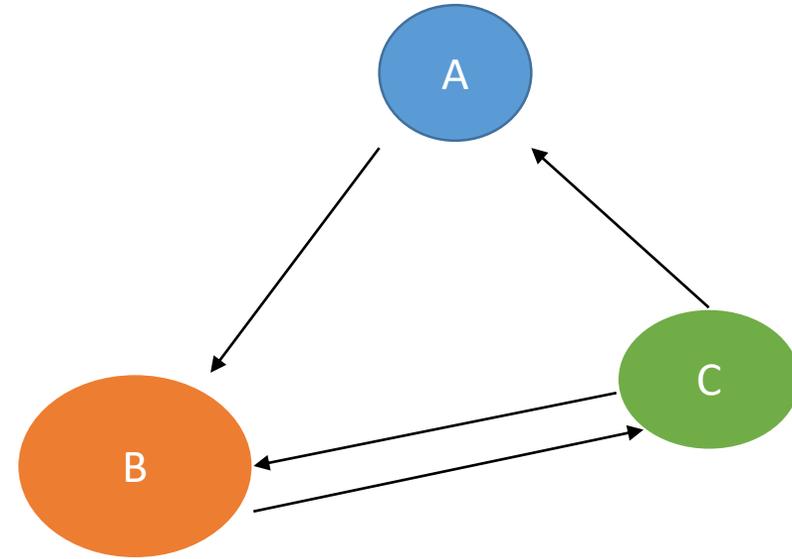
3. Répéter l'étape 2 jusqu'à convergence (les scores se stabilisent, c_a_d la différence entre deux itérations des scores est plus petit qu'une certaine marge acceptable d'erreur)

Exemple:

$$PR(A) = 0,15/3 + 0,85 \left(\frac{PR(C)}{2} \right)$$

$$PR(B) = 0,15/3 + 0,85 \left(\frac{PR(A)}{1} + \frac{PR(C)}{2} \right)$$

$$PR(C) = 0,15/3 + 0,85 * \frac{PR(B)}{1}$$



	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=11	i=12	i=13	i=14
A	0,33	0,19025	0,1904625	0,2413084	0,1984732	0,2169065	0,2170452	0,2105511	0,2162614	0,2139580	0,2140268	0,2149481	0,2142656	0,2146196
B	0,33	0,47075	0,352175	0,4032015	0,4035853	0,3856088	0,4014158	0,3950396	0,3952299	0,3977803	0,3958911	0,3968709	0,3969716	0,3967454
C	0,33	0,3305	0,4501375	0,3493487	0,3927213	0,3930475	0,3777674	0,3912034	0,3857836	0,3859454	0,3881132	0,3865075	0,3873402	0,3874258



La page la plus populaire est la page B (page rank= 0,39)

La page C est plus autoritaire que la page A parce qu'elle est citée par la page B (la plus autoritaire)

PageRank

Exemple d'application :

- **Recherche sur le Web** : PageRank est utilisé par les moteurs de recherche pour classer les pages Web selon leur importance.
- **Analyse des réseaux sociaux** : Identifier les utilisateurs influents dans des plateformes comme Twitter ou LinkedIn.

Interprétation :

- Un nœud avec un PageRank élevé est considéré comme important et influent dans le réseau, car il est pointé par des nœuds eux-mêmes importants.

Outils d'Analyse des Réseaux Sociaux

❖ Gephi

- Un logiciel open-source pour la visualisation et l'analyse des réseaux.
- Il permet de représenter graphiquement les réseaux et d'effectuer des analyses quantitatives de ces réseaux.
- Gephi est particulièrement puissant pour :
 - Visualiser les nœuds et les arêtes d'un réseau.
 - Calculer des métriques comme la centralité, ...
 - Effectuer des algorithmes de détection de communautés pour identifier des groupes d'individus dans le réseau.



Outils d'Analyse des Réseaux Sociaux



❖ **NetworkX** : Bibliothèque Python pour la manipulation et l'analyse des graphes.

❖ **GraphML**

format de fichier XML utilisé pour représenter des graphes.

Il est compatible avec **Gephi**, facilitant ainsi l'importation des données et leur analyse (détection des communautés ou calcul de métriques).

EXAMPLE



NetworkX
Network Analysis in Python

```
import networkx as nx
import matplotlib.pyplot as plt

# Création d'un graphe orienté
G = nx.DiGraph()

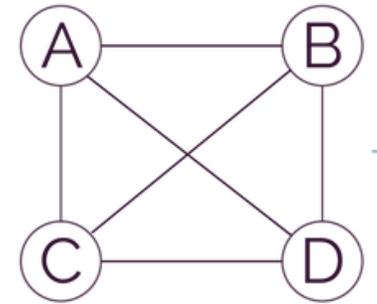
# Ajout des nœuds et des liens
G.add_edges_from([('A', 'B'), ('A', 'C'), ('B', 'C'), ('C',
'D')])

# Calcul de la centralité de degré
degree_centrality = nx.degree_centrality(G)
print("Centralité de degré :", degree_centrality)

# Visualisation du graphe
nx.draw(G, with_labels=True)
plt.show()
```

Représentation des Réseaux Sociaux sous Forme de Matrices- **matrice d'adjacence** -

- Les **matrices d'adjacence** sont une représentation matricielle des réseaux.
- Elles permettent de représenter les relations entre les nœuds d'un graphe, où chaque ligne et chaque colonne correspond à un nœud.
- Une case dans la matrice contient une valeur qui indique la présence (ou l'absence) d'une connexion entre deux nœuds.



$$\begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array} \begin{bmatrix} \text{A} & \text{B} & \text{C} & \text{D} \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Représentation des Réseaux Sociaux sous Forme de Matrices- **matrice d'adjacence** -

- ❑ Une **matrice d'adjacence binaire** pour un réseau non pondéré : Les cases prennent la valeur 1 si une relation existe, et 0 sinon.
- ❑ Une **matrice d'adjacence pondérée** pour un réseau où les relations peuvent avoir des poids (par exemple, l'intensité d'une interaction).

Calcul et Interprétation des Métriques via Gephi

1. Importation de données :

- Importez le fichier de données (GraphML ou csv) dans Gephi via le menu **Fichier > Importer**.
- Il est possible ensuite d'explorer et visualiser le réseau.

2. Calculer des métriques :

- Allez dans l'onglet **Statistiques** pour calculer les différentes métriques (degré, intermédiarité, proximité, etc.).
- Gephi permet de visualiser ces mesures sous forme de graphiques ou de cartes de couleurs sur le réseau.

3. Visualisation du réseau :

- Utilisez les **algorithmes de mise en page** dans Gephi pour organiser les nœuds du réseau
- Ces algorithmes utilisent les métriques de centralité pour positionner les nœuds de manière significative.

4. Interprétation des résultats

Travaux pratiques



TP 1: Analyse d'un réseau social à l'aide de Gephi

Etapas:

1. Importer des données de réseau sous forme de matrice d'adjacence.

- (*large_network.csv*)

2. Calculer les métriques des nœuds et du réseau

- *degré , centralité d'intermédiarité , Centralité de , modularité , densité du réseau,*

3. Visualiser les résultats et interpréter les résultats.

- *modifier la taille des nœuds en fonction du degré.*

- *Utilisez les couleurs pour représenter les communautés (modularité)*

- *Affichez les noms des nœuds*

4. Exporter les résultats

- *La visualisation du graphe en tant qu'image (PNG).*

- *Les métriques calculées en fichier CSV.*

TP 2 : Analyse d'un Réseau Social avec Gephi et GraphML

Objectif

- Apprendre à importer un fichier **GraphML** dans Gephi.
 - *Utiliser le fichier : large_manual_social_network.graphml*
Nombre de nœuds : 300, Attributs des nœuds : label et group.
Attributs des arêtes : weight et type.
- Visualiser un réseau social complexe avec des communautés et des attributs (poids, types de relations).
- Calculer et interpréter les métriques avancées.

