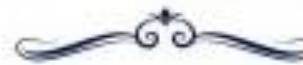




Analyse des réseaux sociaux

par

Dr. Samira LAGRINI



Année universitaire:2025/2026

Qu'est-ce que l'Analyse de Réseaux Sociaux (SNA) ?

- **L'Analyse de Réseaux Sociaux (SNA)** est une approche méthodologique et théorique qui étudie les structures sociales en modélisant les entités sociales (individus, organisations, ordinateurs) comme des **nœuds** connectés par des **liens** représentant leurs relations ou interactions.

Objectif : révéler les **motifs structurels** sous-jacents, notamment :

- ✓ La **centralité** et l'**influence** des acteurs.
- ✓ La **cohésion** des groupes.
- ✓ La **détection de communautés**.
- ✓ Les **canaux de diffusion** de l'information, des innovations ou des comportements.

SNA et le Web Mining

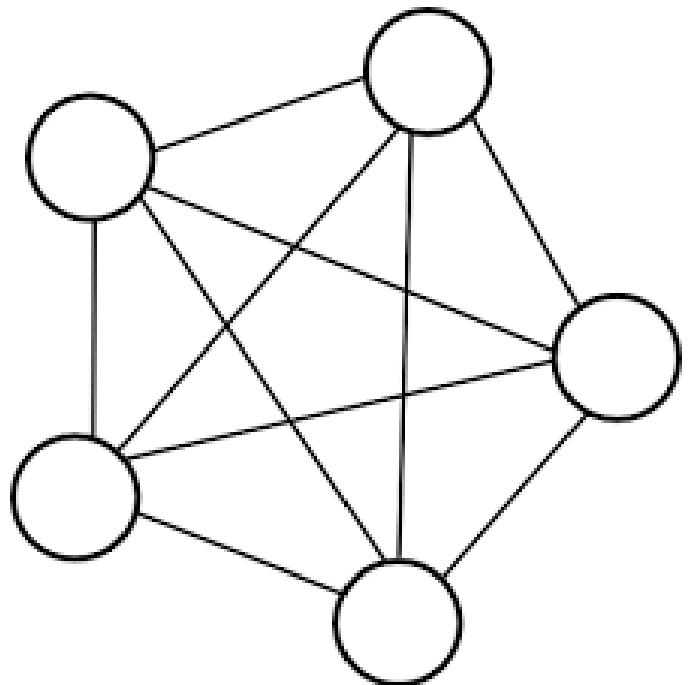
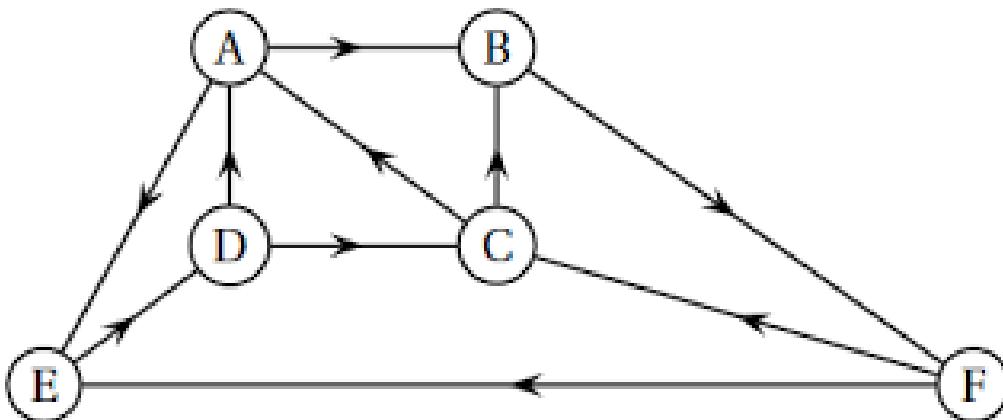
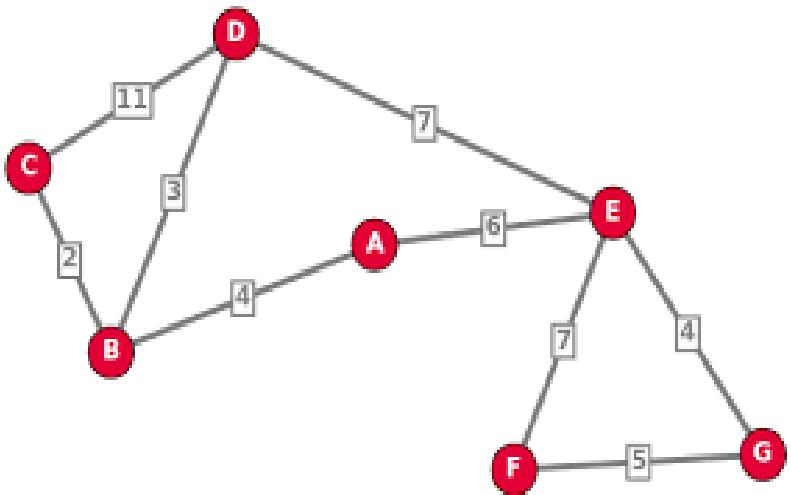
- Lorsque le SNA est appliquée au Web Mining, Elle se concentre sur :
 - **L'analyse des réseaux sociaux numériques** (Facebook, Twitter, Instagram, LinkedIn) pour identifier des influenceurs, étudier la viralité (la diffusion d'informations) et détecter des communautés en ligne.
 - **L'analyse de la structure hypertexte du Web** (web structure mining), où les pages web sont des nœuds et les liens hypertextes sont des arêtes, utilisée par des algorithmes comme **PageRank**.
 - **La modélisation des réseaux d'interaction** dérivés du Web Usage Mining (ex: "les utilisateurs qui ont acheté X ont aussi acheté Y").

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 du réseau (ex : Personnes sur un réseau social, chercheurs dans un réseau de co-auteurs).
- **Les Liens (Edges)** Représentent les relations ou interactions entre les nœuds (ex : Amitiés sur Facebook, citations d'articles dans un réseau scientifique).
- **Graphes** : Utilisés pour modéliser les réseaux. Un graphe est une structure qui relie des entités (nœuds) via des relations (arêtes).

Types de Graphes utilisés en SNA

- Graphes non orientés
- Graphes orientés
- Graphes pondéré

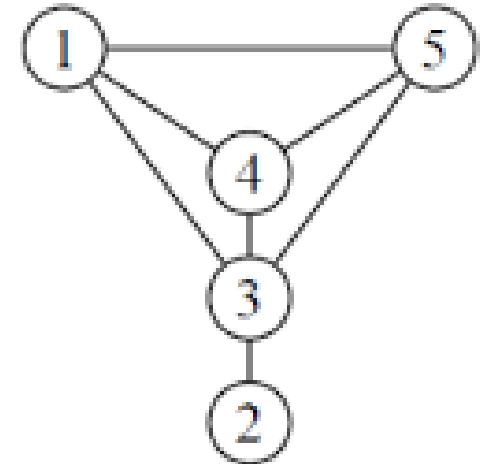


Graphes non orientés

Graphes non orientés : Les liens entre les nœuds n'ont pas de direction (relation 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

Graphes orientés

Graphes orientés : Les liens ont une direction (relation asymétrique).

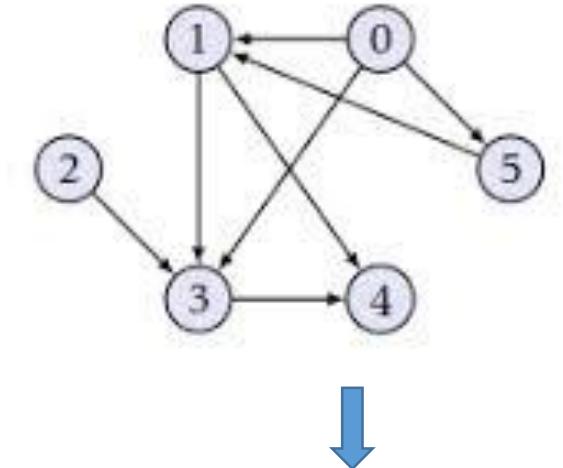
Exemple

- Suivi 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,



Matrice d'adjacence

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



Graphes pondérés

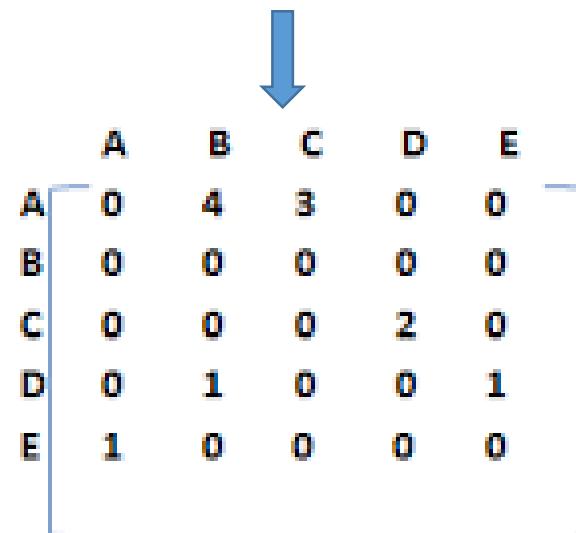
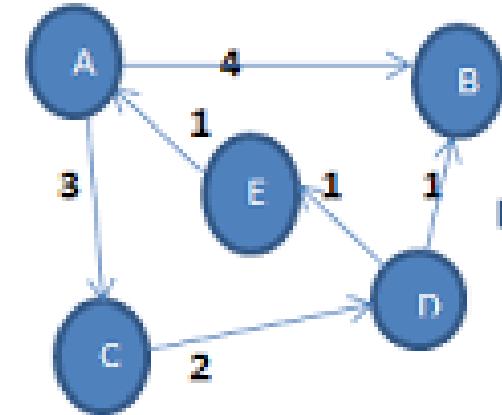
- **Graphes pondérés** : Les liens ont un poids associé (représente l'intensité de la relation, la fréquence ou la durée de la relation.).

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

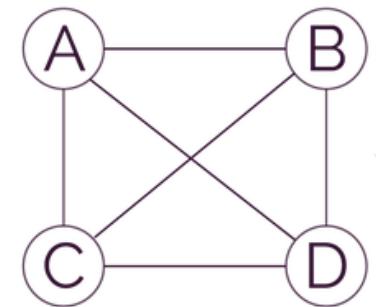
- Analyser la force des interactions dans un réseau (l'intensité de la collaboration entre des chercheurs)
- la force des relations commerciales entre entreprises.



Matrice d'adjacence

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.



	A	B	C	D
A	0	1	1	1
B	1	0	1	1
C	1	1	0	1
D	1	1	1	0

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).

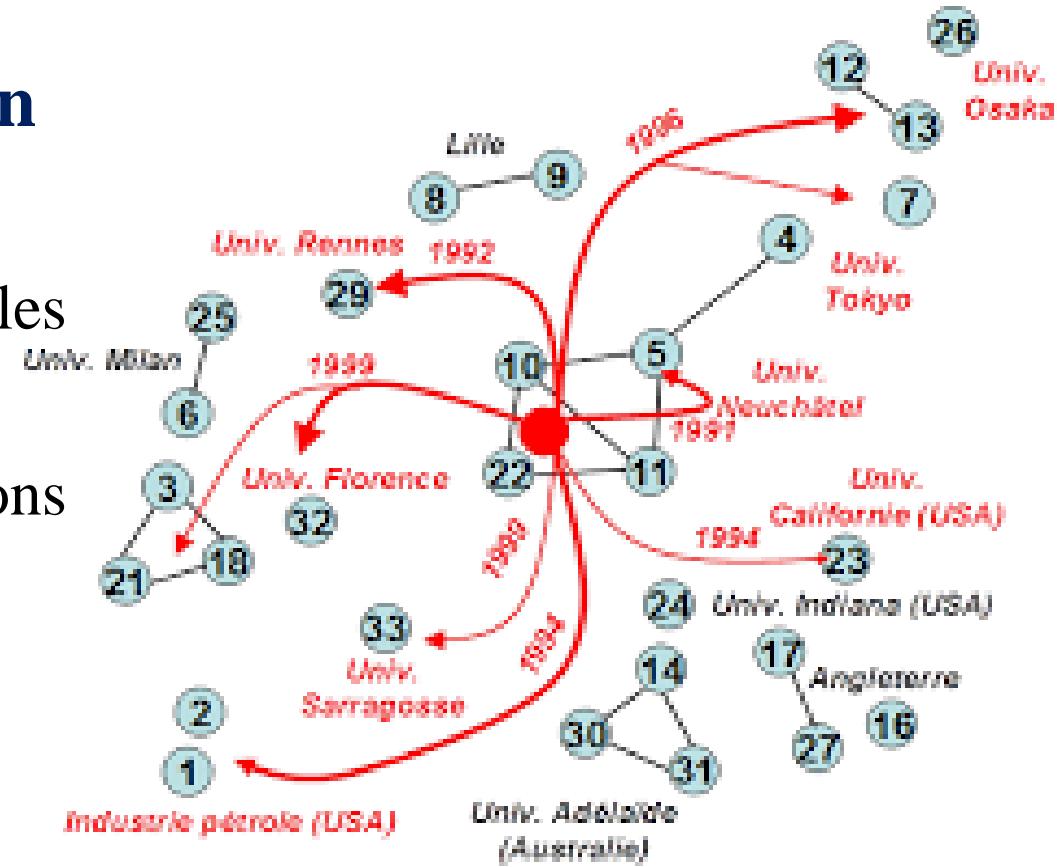
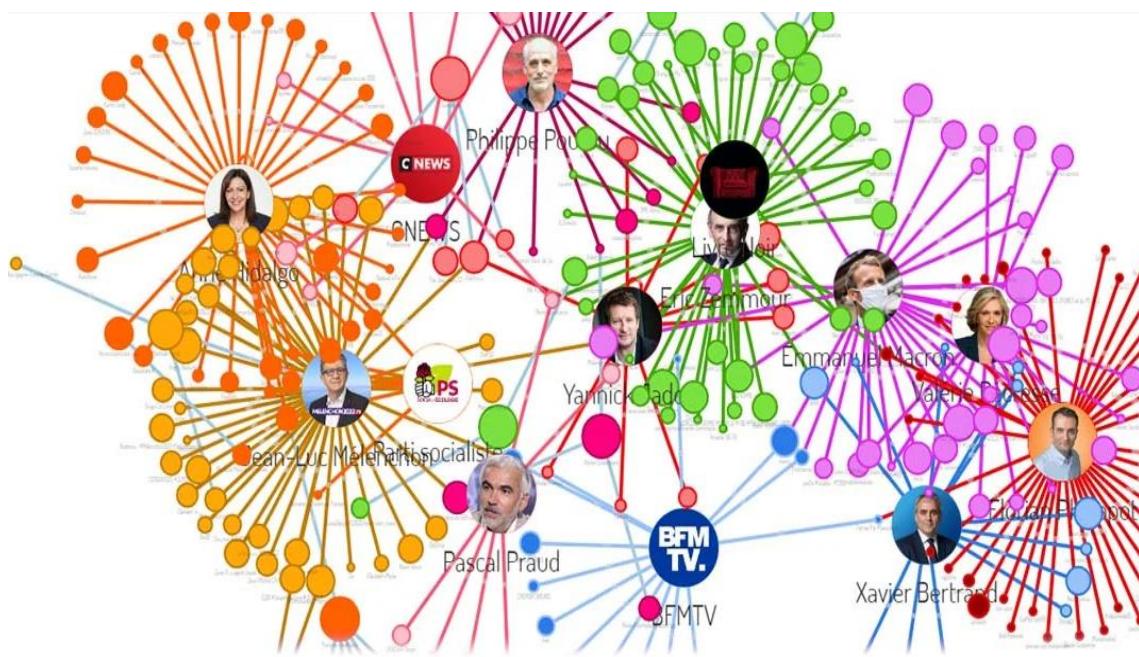
Applications de SNA

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

□ 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

- Métriques de nœuds
- Métriques du réseau

Centralité de degré

- Mesure du nombre de connexions directes d'un nœud.
- 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 dégrée** d'un nœuds ($CD(i)$) est mesuré par son degré $d(i)$ normalisé par le degré maximum ($n-1$)

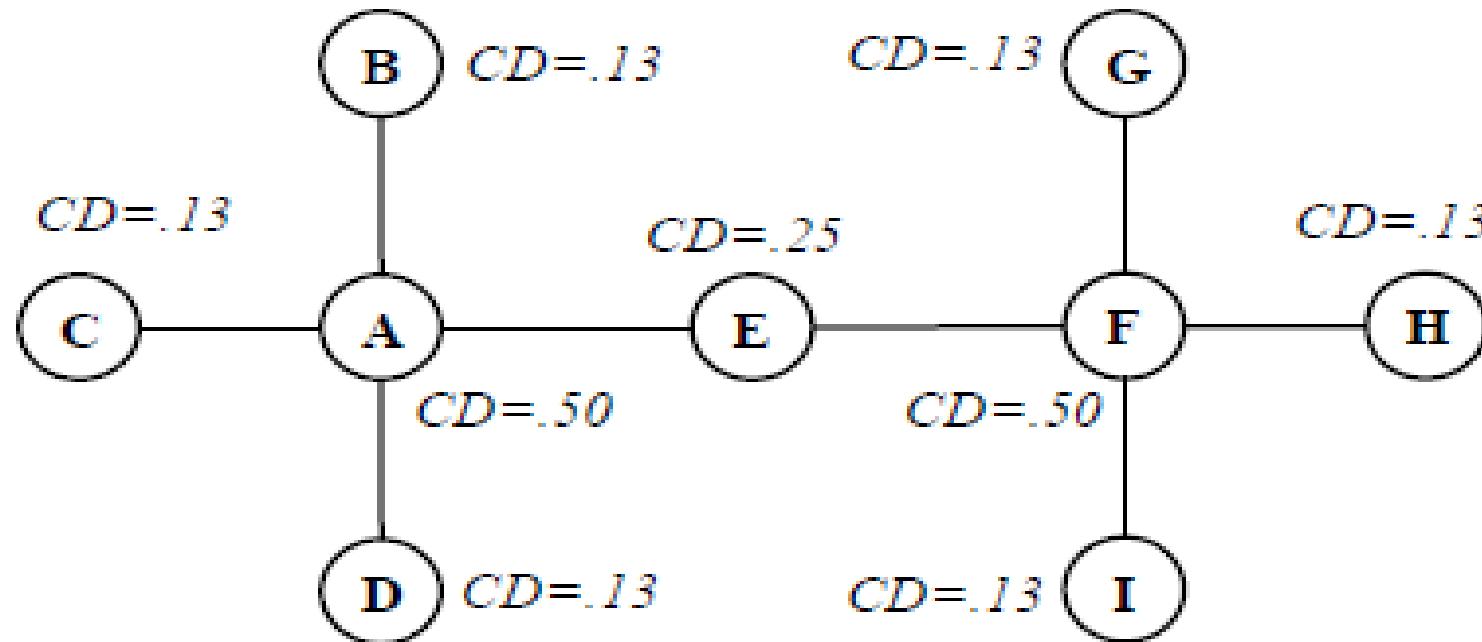
$$CD(i) = \frac{d(i)}{n-1}.$$

- 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 a 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:

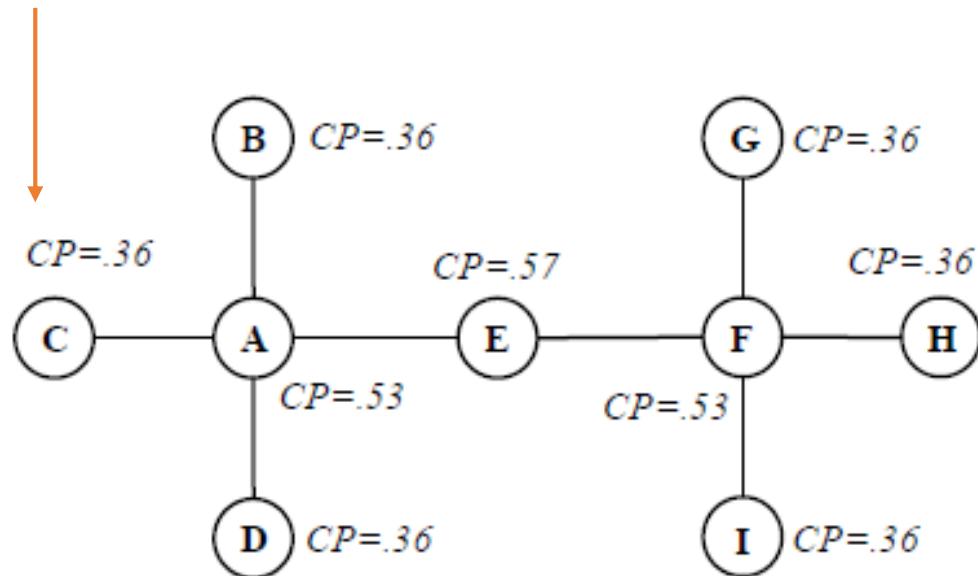
$$C^{pro}(v_i) = \frac{N-1}{\sum_{j=1}^N dist(v_i, v_j)}$$

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

Centralité de proximité

Graphe non orienté

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

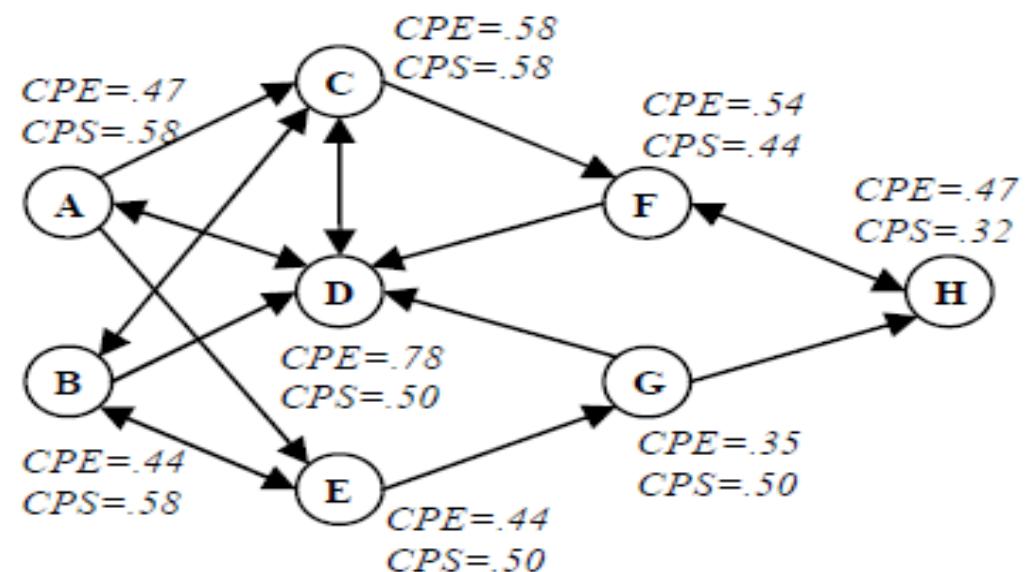


le noeud E est le plus important dans le graphe G1 alors qu'il n'a que deux liens

Graphe orienté

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

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



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

Centralité d'Intermédiairité (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'intermédiairité élevée jouent souvent un rôle de médiateur ou de "pont" entre différentes parties du réseau. Ils sont cruciaux pour la transmission d'informations.
- **Exemple** : Un individu qui agit comme médiateur dans un réseau..

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)

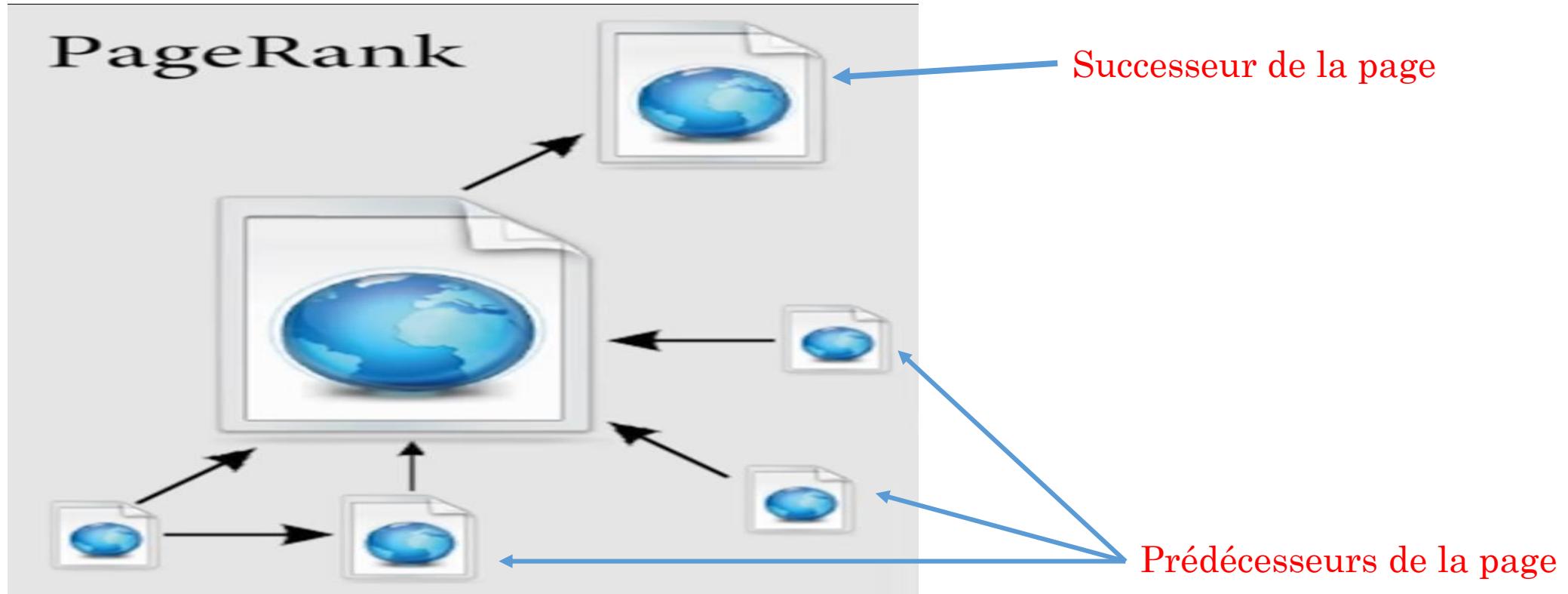
- La modularité mesure la force de la division d'un réseau en communautés ou groupes.
- Elle évalue si le réseau peut être partitionné en sous-groupes densément connectés.
- **Interprétation** : Une modularité élevée indique la présence de communautés bien définies.
- **Exemple** : Dans les réseaux sociaux,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

PageRank

- Mesure l'importance d'un nœud dans un réseau orienté en fonction des liens entrants.
- Un nœud pointé par des nœuds importants est considéré comme important et influent dans le réseau → un PageRank élevé.
- PageRank est utilisé par Google pour classer les pages web.

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 (*prédecesseur 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 prédecesseurs)

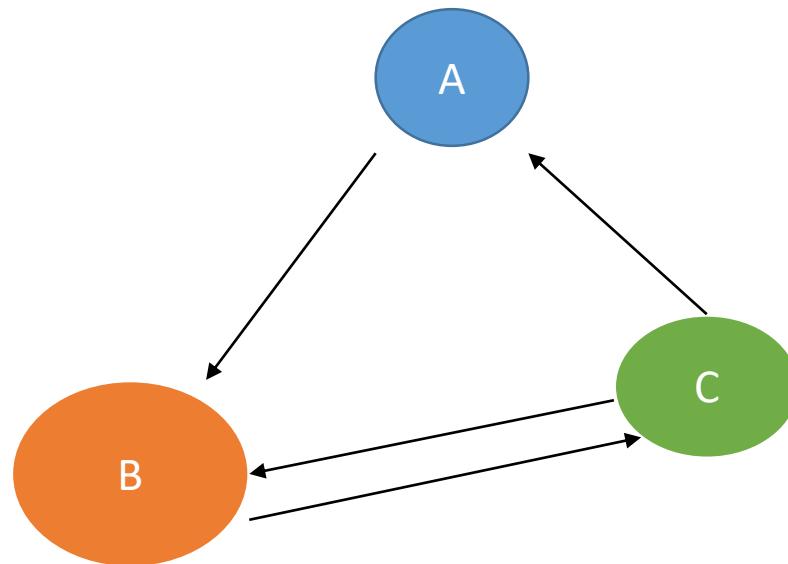
3. Répéter l'étape 2 jusqu'à convergence (les scores se stabilisent, c_a_d la différence entre deux itérations des scores est plus petit qu'une certaine marge acceptable d'erreur)

Exemple:

$$PR(A) = 0,15/3 + 0,85 (PR(C)/2)$$

$$PR(B) = 0,15/3 + 0,85 (PR(A)/1 + PR(C)/2)$$

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



	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 cité par la page B (la plus autoritaire)

Outils d'Analyse des Réseaux Sociaux

❖ Gephi

Un logiciel open-source pour visualiser et analyser les réseaux sociaux. Il permet de calculer des métriques comme la centralité et de détecter des communautés.



❖ NetworkX

Bibliothèque Python pour la manipulation et l'analyse des graphes.



NetworkX
Network Analysis in Python

Outils d'Analyse des Réseaux Sociaux

❖ GraphML

Format de fichier XML utilisé pour représenter des graphes, compatible avec Gephi.

```
<graphml xmlns="http://graphml.graphdrawing.org/xmlns">
  <graph id="G" edgedefault="undirected">
    <!-- Définition des nœuds -->
    <node id="A"/>
    <node id="B"/>
    <node id="C"/>
    <node id="D"/>
    <node id="E"/>

    <!-- Définition des arêtes -->
    <edge source="A" target="B"/>
    <edge source="B" target="C"/>
    <edge source="C" target="D"/>
    <edge source="D" target="E"/>
    <edge source="E" target="A"/>
  </graph>
</graphml>
```

EXAMPLE

```
import networkx as nx
import matplotlib.pyplot as plt
# Cr ation d'un graphe orient 
G = nx.DiGraph()
# Ajout des n euds et des liens
G.add_edges_from([('A', 'B'), ('A', 'C'), ('B', 'C'), ('C',
'D')])
# Calcul de la centralit  de degr 
degree_centrality = nx.degree_centrality(G)
print("Centralit  de degr  :", degree_centrality)
# Visualisation du graphe
nx.draw(G, with_labels=True)
plt.show()
```



Travaux pratiques



TP1: Importation et analyse d'un réseau social à partir du dataset "The Crisis"

1. Importez les données du réseau social issues du **Modernist Journals Project - The Crisis**, disponible sur le lien suivant : <https://github.com/melaniewalsh/sample-social-network-datasets/tree/master/sample-datasets/modernist-journals-project/crisis> (utiliser les deux formats csv et GraphML)
2. Calculer des métriques de nœuds (la centralité de degré, l'intermédiarité,...) pour identifier les nœuds les plus influents et du réseau (la densité, la modularité..) pour détecter les communautés
3. Visualiser le réseau: Mettez en évidence les nœuds importants en fonction des métriques calculées (ex. : taille des nœuds en fonction de leur centralité)
4. Exporter les résultats
 - Visualisation : Exportez la visualisation du réseau sous forme d'image (format PNG ou JPEG).
 - Métriques : Exportez les résultats des métriques sous forme de fichier CSV pour permettre une analyse détaillée des propriétés du réseau.

