
Etude d'un modèle aléatoire conservant les bicliques
maximales dans les graphes bipartis

Présenté par
Cécile Pov

le 12 octobre 2020

Jury : MM. Fabien Tarissan & Lionel Tabourier (ComplexNetworks) - M. Nabil Hassan Mustafa (ESIEE Paris)

Plan

1. Contexte du stage
2. Un domaine d'étude : les réseaux complexes
3. Problématique du stage
4. Le modèle triparti
5. Résultats
Comparaison des graphes bipartis générés
6. Conclusion

Contexte du stage

LIP6

- Laboratoire de recherche sous tutelle de Sorbonne Université et du CNRS
- \approx 200 permanents (enseignants chercheurs, etc.)
 \approx 300 non-permanents (200 doctorants, postdocs, etc.)
- 4 axes transverses : IA et datascience, architecture, sécurité, théorie et outils mathématiques pour l'informatique
- 22 équipes
 - > ComplexNetworks : 18 membres, 4 permanents
 - Répondre aux questions transverses autour des réseaux complexes

Domaine d'étude : les réseaux complexes

Les réseaux complexes nous entourent :

- Internet
- Interactions sociales
- Interactions protéines-protéines

Nombreux contextes applicatifs et tâches critiques :

- Diffusion (information, virus)
- Prédiction de cyber attaques
- Identifier des personnes “influentes” dans un réseau

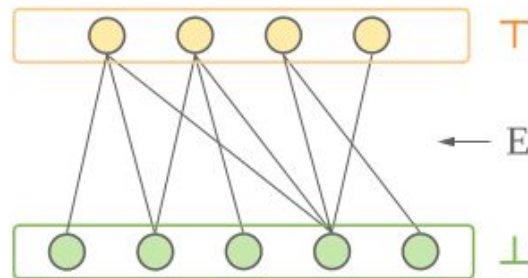


[1]

Domaine d'étude : les réseaux complexes

Les réseaux réels ont des propriétés non triviales :

- Densité globale très faible
- Densité locale forte
- Distance moyenne faible
- Distribution de degré hétérogène

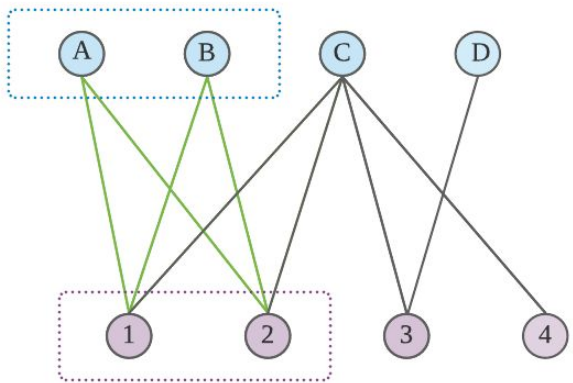


On capture ces propriétés dans des modèles de génération de graphe.

A ce jour, aucun modèle de génération de graphe ne permet d'aboutir à un graphe de terrain réaliste.

Problématique : les bicliques dans les graphes réels

Biclique (X_n, Y_n) : un sous-graphe de G où tous les noeuds X_n du haut sont reliés à tous les noeuds Y_n du bas



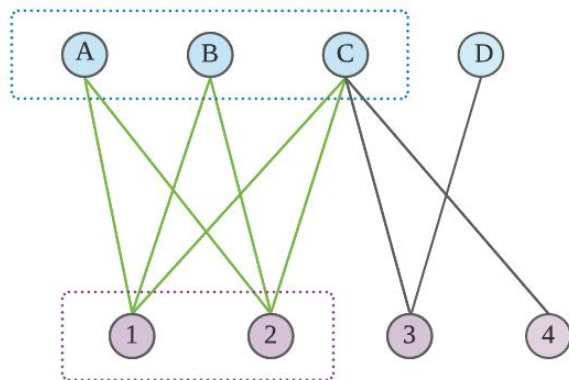
$(\{A,B\}, \{1,2\})$ est une biclique ? **OUI**

$(\{A,B\}, \{1,2\})$ est une biclique maximale ?

- Structure forte
- Nombreux contextes applicatifs

Problématique : les bicliques dans les graphes réels

Biclique (X_n, Y_n) : un sous-graphe de G où tous les noeuds X_n du haut sont reliés à tous les noeuds Y_n du bas

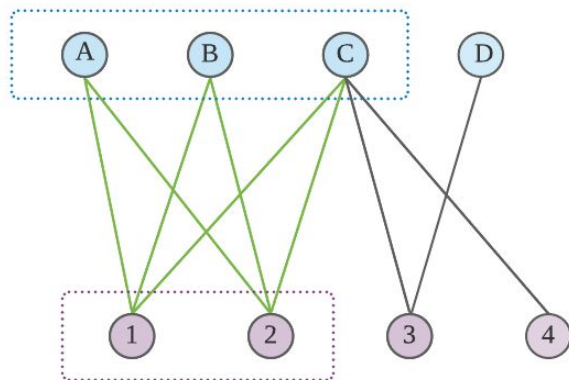


$(\{A,B\}, \{1,2\})$ est une biclique ? **OUI**

$(\{A,B\}, \{1,2\})$ est une biclique maximale ? **NON**

Problématique : les bicliques dans les graphes réels

Biclique (X_n, Y_n) : un sous-graphe de G où tous les noeuds X_n du haut sont reliés à tous les noeuds Y_n du bas

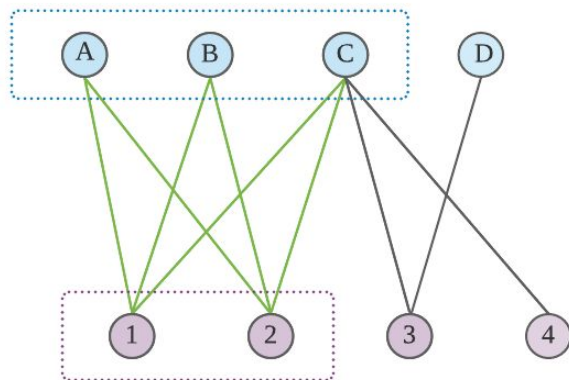


$(\{A,B,C\}, \{1,2\})$ est une biclique ? **OUI**

$(\{A,B,C\}, \{1,2\})$ est une biclique maximale ? **OUI**

Problématique : les bicliques dans les graphes réels

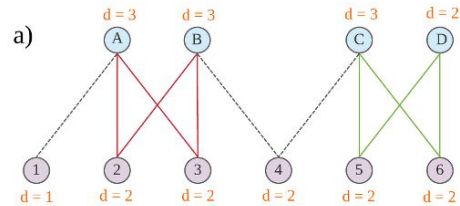
Biclique (X_n, Y_n) : un sous-graphe de G où tous les noeuds X_n du haut sont reliés à tous les noeuds Y_n du bas



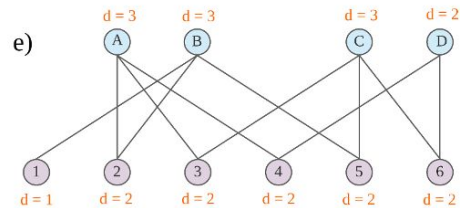
$(\{A,B,C\}, \{1,2\})$ est une biclique ? **OUI**

$(\{A,B,C\}, \{1,2\})$ est une biclique maximale ? **OUI**

Problématique : le configuration model biparti

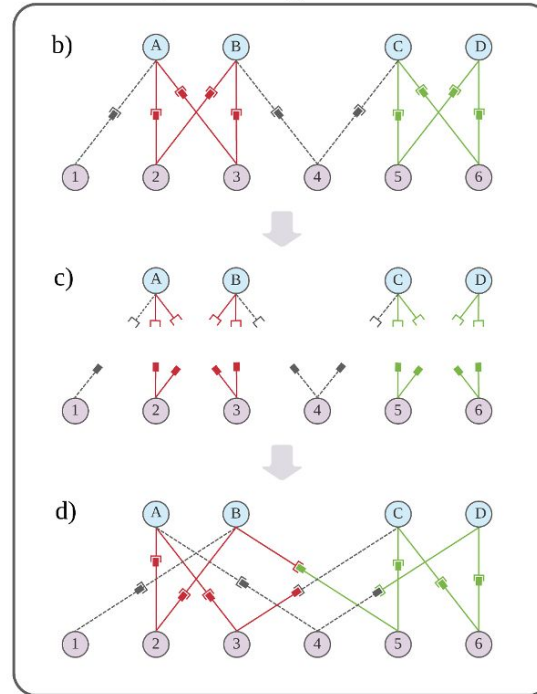


2 bicliques maximales

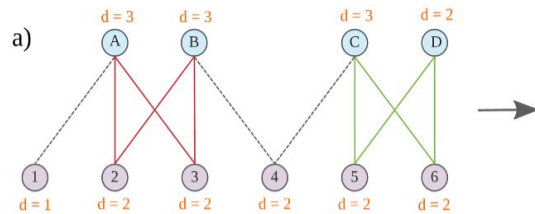


0 biclique maximale

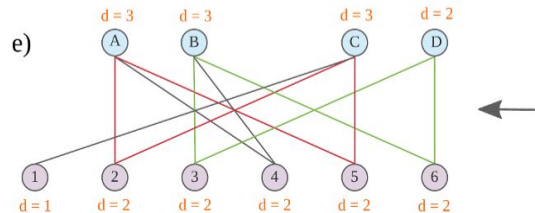
Bipartite Configuration Model



Problématique : le configuration model biparti



2 bicliques maximales



2 bicliques maximales

Modèle triparti

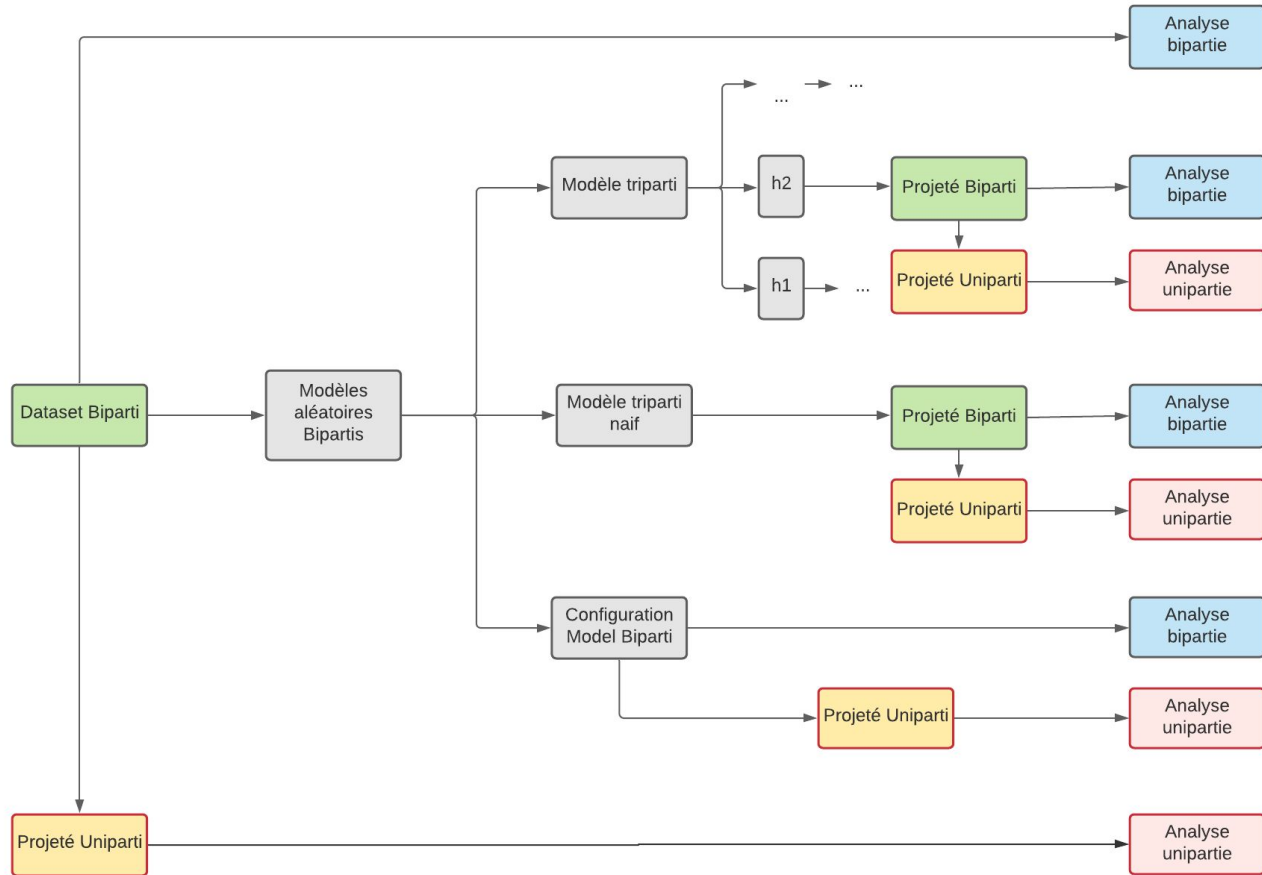
?

Objectif du stage

Etudier la pertinence du modèle triparti en tant que support pour générer des graphes réalistes :

- Etude statistique
- Etude structurelle

Objectif du stage



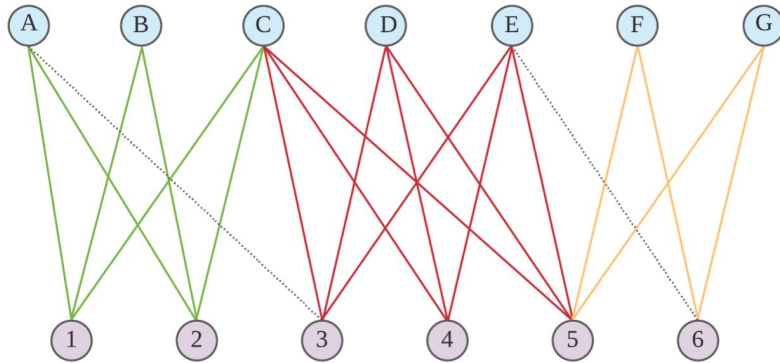
Problématique

Déroulement de l'algorithme

Modèle triparti :

1. Enumérer toutes les bicliques maximales
2. Encoder les b.m sélectionnées dans un 3ème niveau
3. Générer un graphe triparti aléatoire
4. Projeter le graphe triparti généré pour récupérer un graphe biparti aléatoire.

Problématique



3 bicliques maximales non recouvrantes dans G :

- $(\{A, B, C\}, \{1, 2\})$
- $(\{C, D, E\}, \{3, 4, 5\})$
- $(\{F, G\}, \{5, 6\})$

2 arêtes non-impliquées dans ces bicliques maximales :

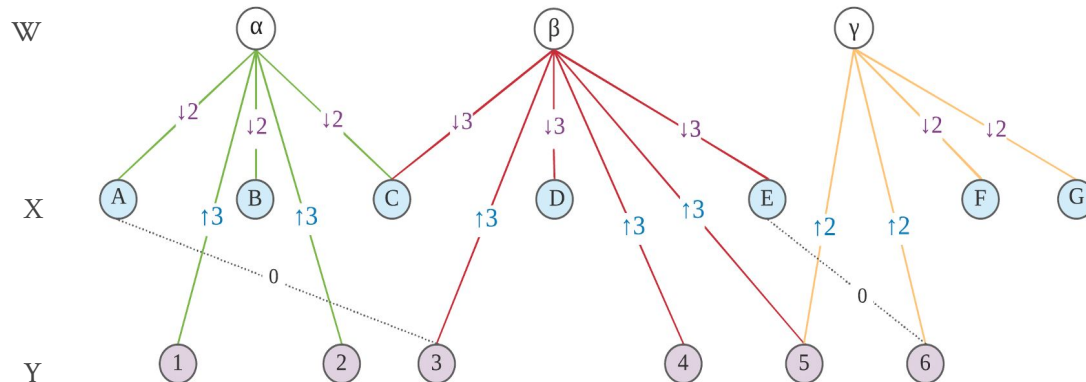
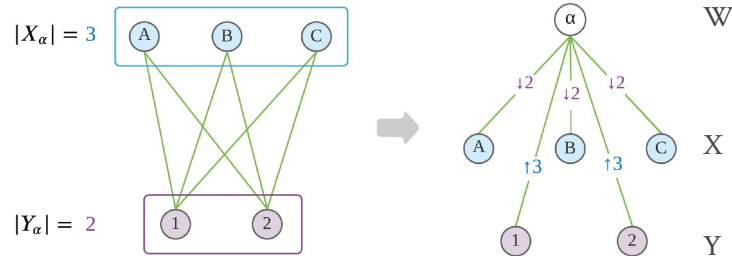
- (A,E)
- (E,6)

Modèle triparti :

1. **Enumérer toutes les bicliques maximales**
> En extraire une séquence de bicliques maximales non recouvrantes (via heuristique, etc.)
2. Encoder les b.m sélectionnées dans un 3ème niveau
3. Générer un graphe triparti aléatoire
4. Projeter le graphe triparti généré pour récupérer un graphe biparti aléatoire.

Problématique

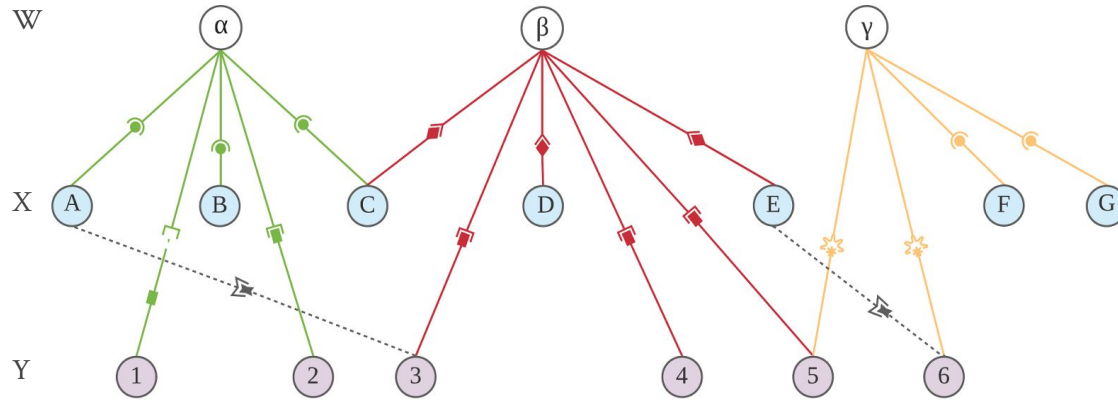
Le noeud α encode la biclique maximale (X_α, Y_α)



Modèle triparti :

1. Enumérer toutes les bicliques maximales
2. Encoder les b.m sélectionnées dans un 3ème niveau
> Chaque noeud dans le 3ème niveau encode une biclique maximale dans G.
> Les arêtes sont labellisées selon leur degré vers l'autre ensemble
3. Générer un graphe triparti aléatoire
4. Projeter le graphe triparti généré pour récupérer un graphe biparti aléatoire.

Problématique

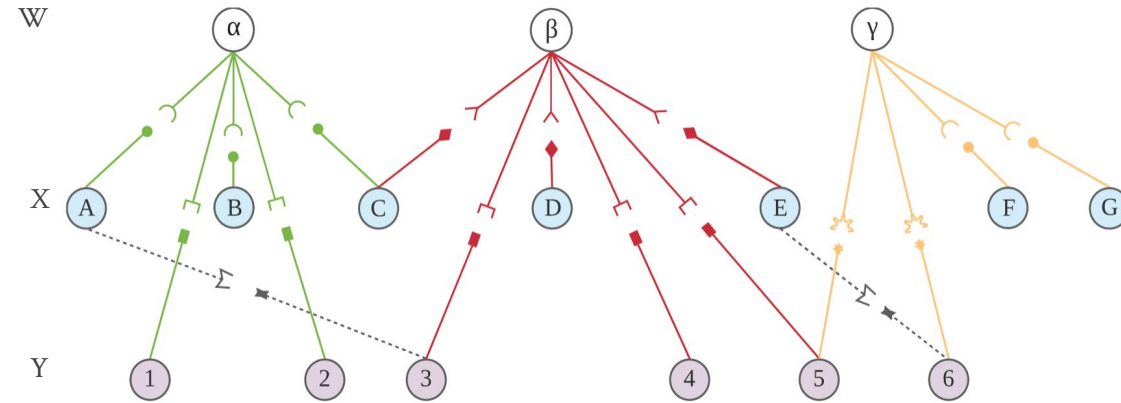


interlockable pairs	not interlockable pairs
12	
13	
13	
12	
0	
	...

Modèle triparti :

1. Enumérer toutes les bicliques maximales
2. Encoder les b.m sélectionnées dans un 3ème niveau
3. Générer un graphe triparti aléatoire
 - > Couper les arêtes
 - > Les demi-arêtes provenant d'arêtes ayant le même label sont connectés aléatoirement.
4. Projeter le graphe triparti généré pour récupérer un graphe biparti aléatoire.

Problématique

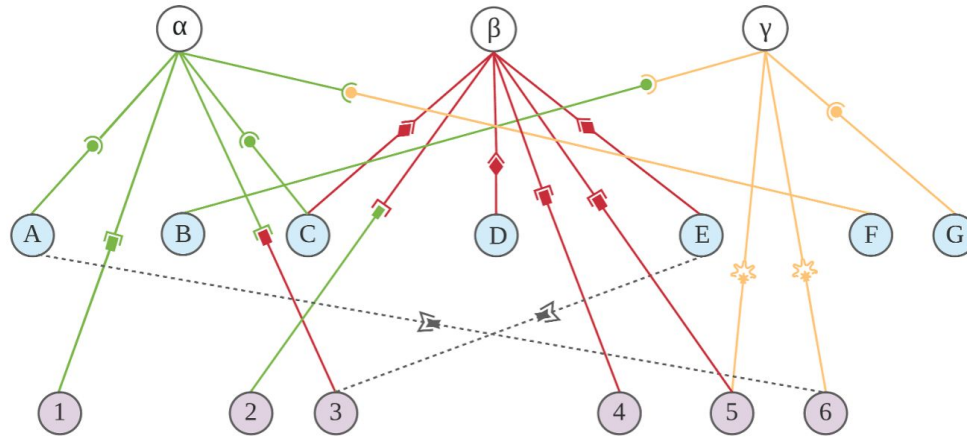


interlockable pairs	not interlockable pairs
12	
13	
13	
12	
0	
	...

Modèle triparti :

1. Enumérer toutes les bicliques maximales
2. Encoder les b.m sélectionnées dans un 3ème niveau
3. **Générer un graphe triparti aléatoire**
 > Couper les arêtes
 > Les demi-arêtes provenant d'arêtes ayant le même label sont connectés aléatoirement.
4. Projeter le graphe triparti généré pour récupérer un graphe biparti aléatoire.

Problématique

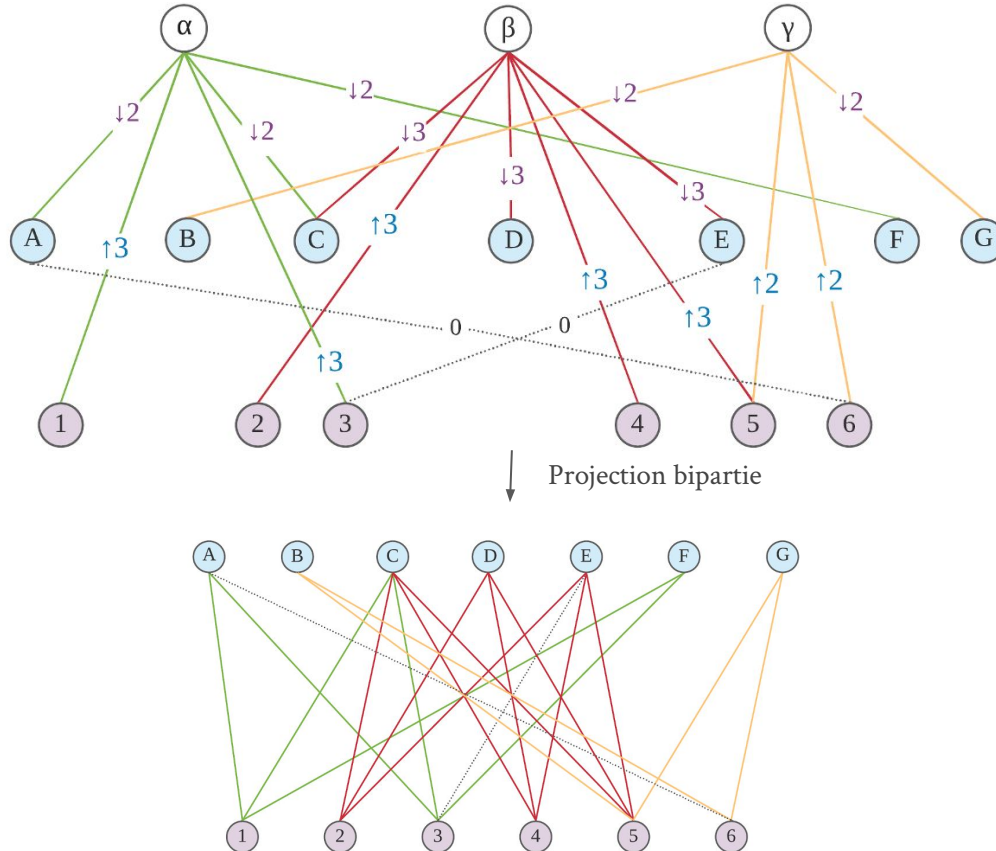


interlockable pairs	not interlockable pairs
↓2	
↓3	
↑3	
↑2	
0	...

Modèle triparti :

1. Enumérer toutes les bicliques maximales
2. Encoder les b.m sélectionnées dans un 3ème niveau
3. **Générer un graphe triparti aléatoire**
 - > Couper les arêtes
 - > Les demi-arêtes provenant d'arêtes ayant le même label sont connectés aléatoirement.
4. Projeter le graphe triparti généré pour récupérer un graphe biparti aléatoire.

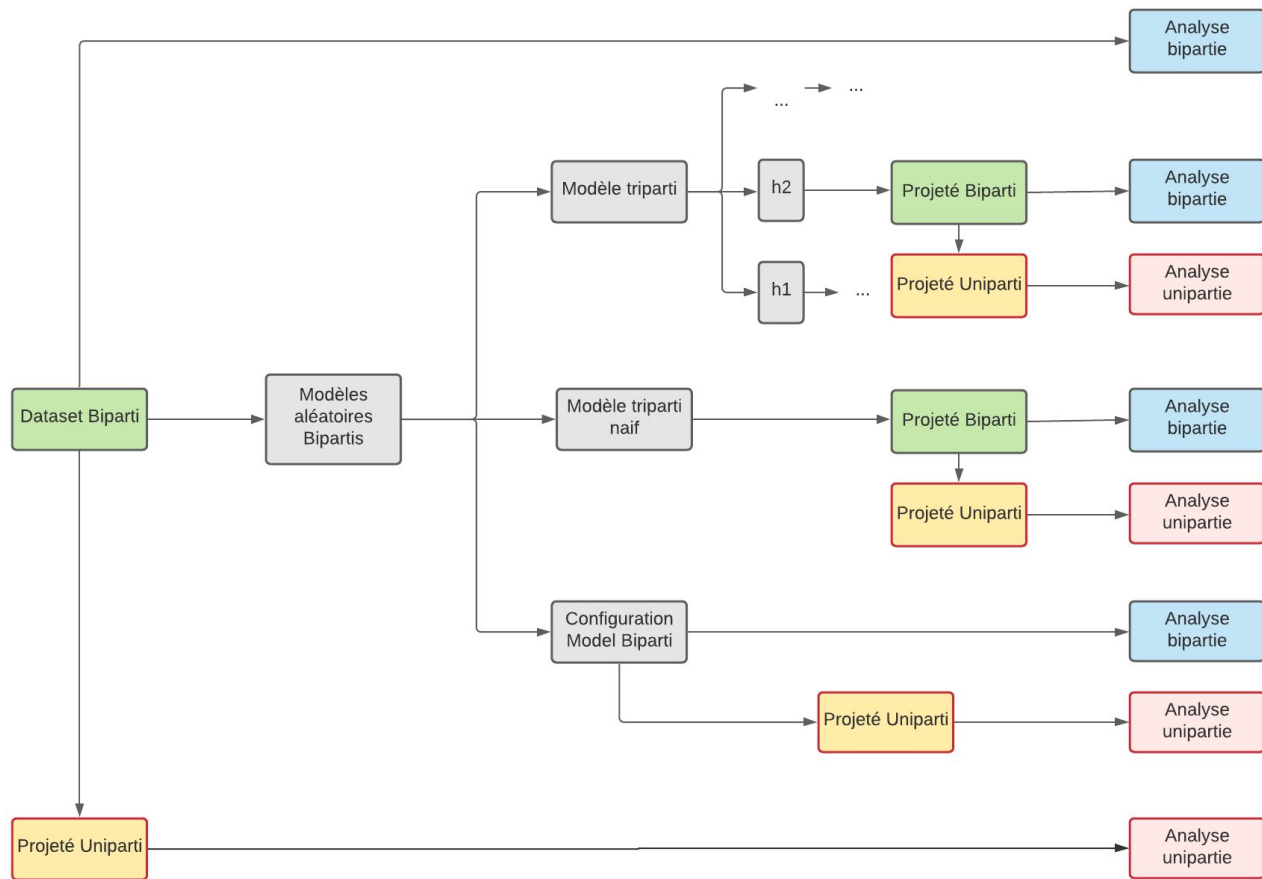
Problématique



Modèle triparti :

1. Enumérer toutes les bicliques maximales
2. Encoder les b.m sélectionnées dans un 3ème niveau
3. Générer un graphe triparti aléatoire
4. **Projeter le graphe triparti généré pour récupérer un graphe bipartite aléatoire.**
Les noeuds de X et Y reliés à un même noeud dans W forment une biclique maximale dans le graphe bipartite projeté.

Résultats



Résultats

Test sur 6 datasets

Figures présentés dans la suite :

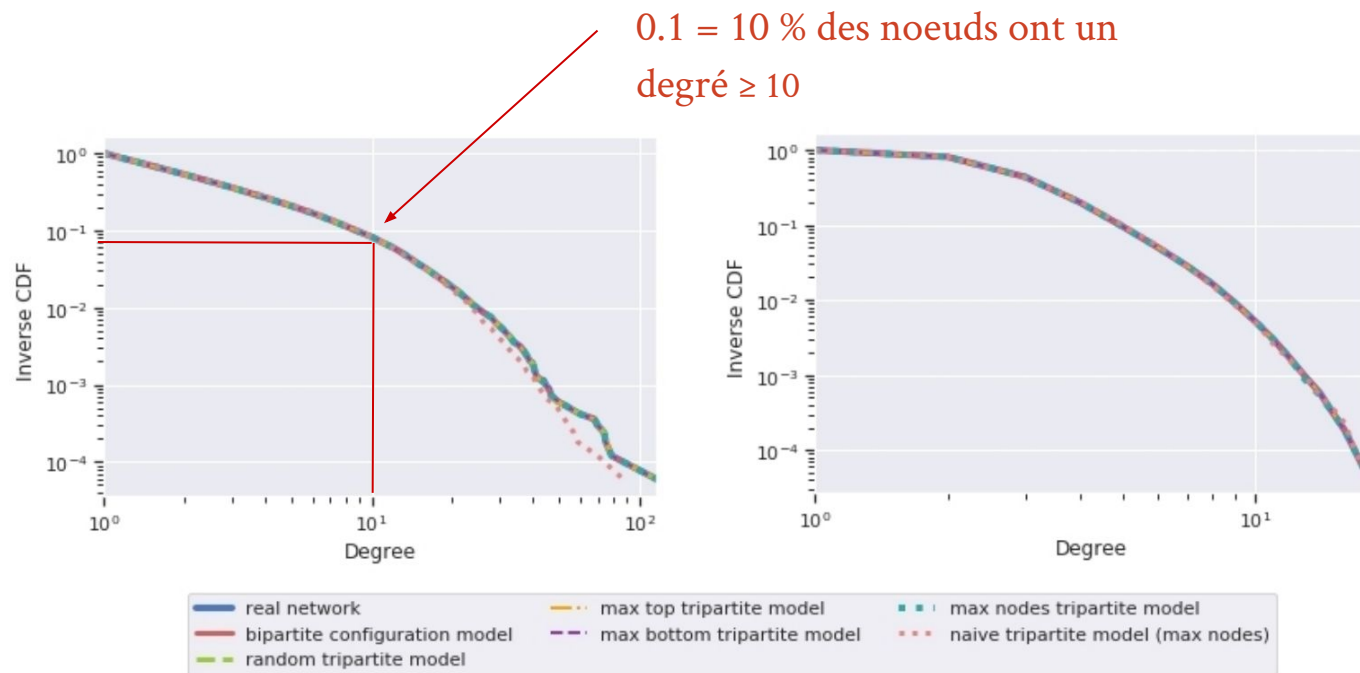
- Dataset de co-authoring (auteur - article) :
 - nombre de noeuds top : 38741
 - nombre de noeuds bottom : 16726
 - nombre d'arêtes : **58595**

Résultats - Analyse bipartie

	real	configuration model	random	maxtop	maxbottom	maxnodes	naive_tri
nb_vertices	16726	16726	16726	16726	16726	16726	16726
nb_edges	58595	58577	58587	58586	58584	58579	58586
density	0.000159129	0.00015908	0.000159107	0.000159105	0.000159099	0.000159086	0.000159105
cc_bullet_set	0.370592	0.198609	0.255805	0.265567	0.246961	0.253785	0.253337
nb_connected_components	1188	747	1437	1490	1454	1450	1441
degree_avg	3.50323	3.50215	3.50275	3.50269	3.50257	3.50227	3.50269
degree_sum	58595	58577	58587	58586	58584	58579	58586
degree_min	1	1	1	1	1	1	1
degree_max	116	116	116	115	116	116	87

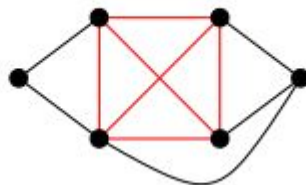
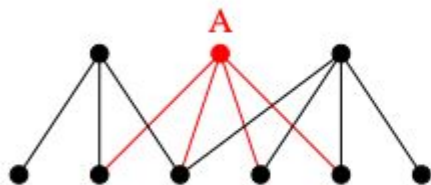
Résultats - Analyse bipartie

Degré : Distribution cumulative inverse (ICDF)

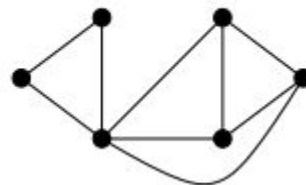


Résultats - Analyse bipartie

Redondance : Distribution cumulative inverse (ICDF)



A crée 6
liens



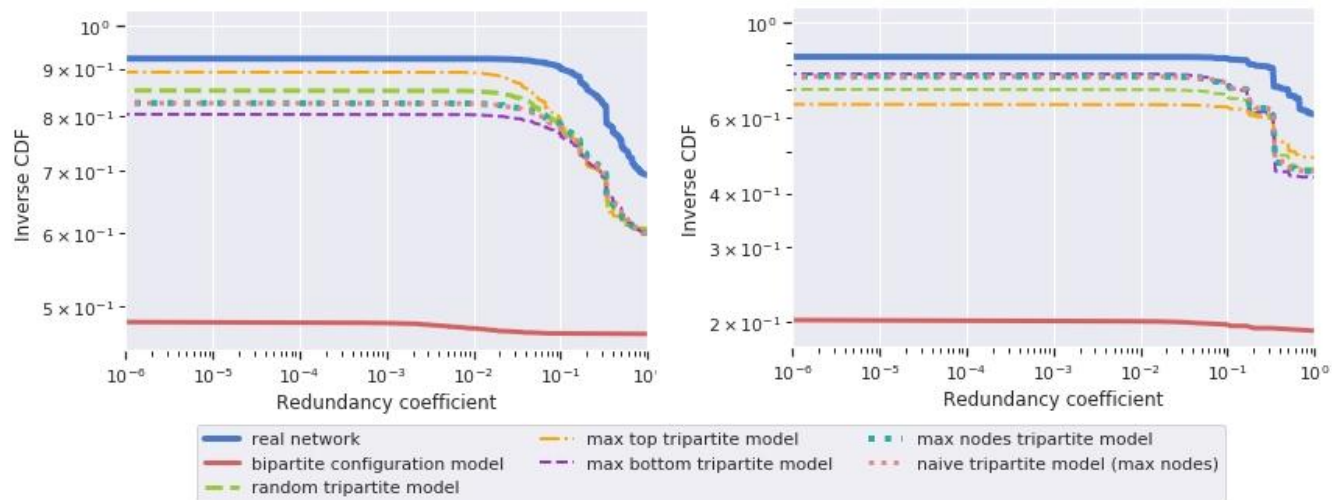
en supprimant A,
il reste 4 liens
parmi les 6

Fraction de voisins de A qui sont liés ensemble dans la projection, même si on supprime A dans le graphe biparti.

Figure : Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. Basic Notions for the Analysis of Large Two-mode Networks. volume 30, pages 31–48. Elsevier, January 2008.

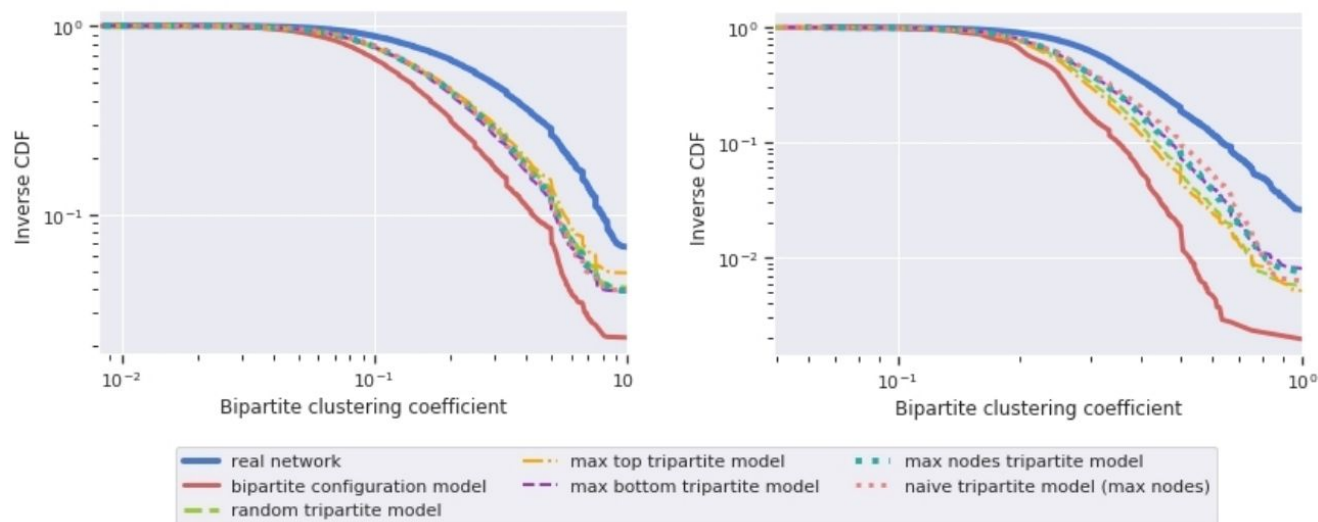
Résultats - Analyse bipartie

Redondance : Distribution cumulative inverse (ICDF)



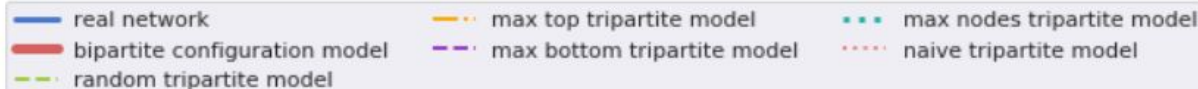
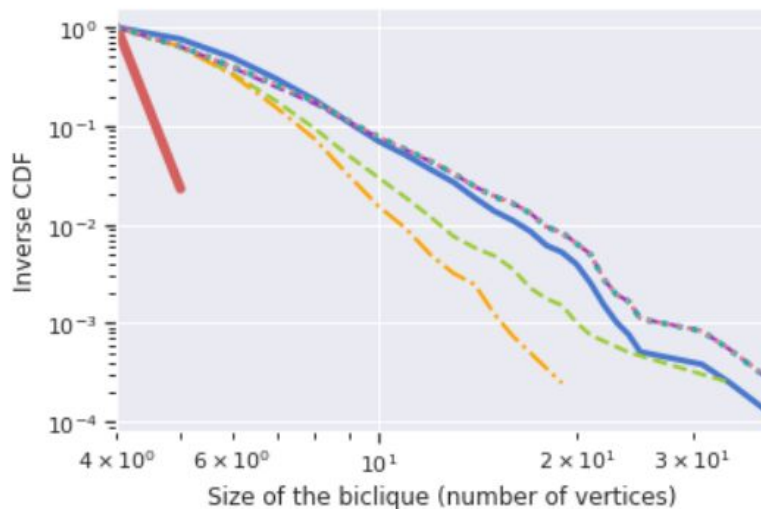
Résultats - Analyse bipartie

Clustering coefficient biparti : distribution cumulative inverse (ICDF)



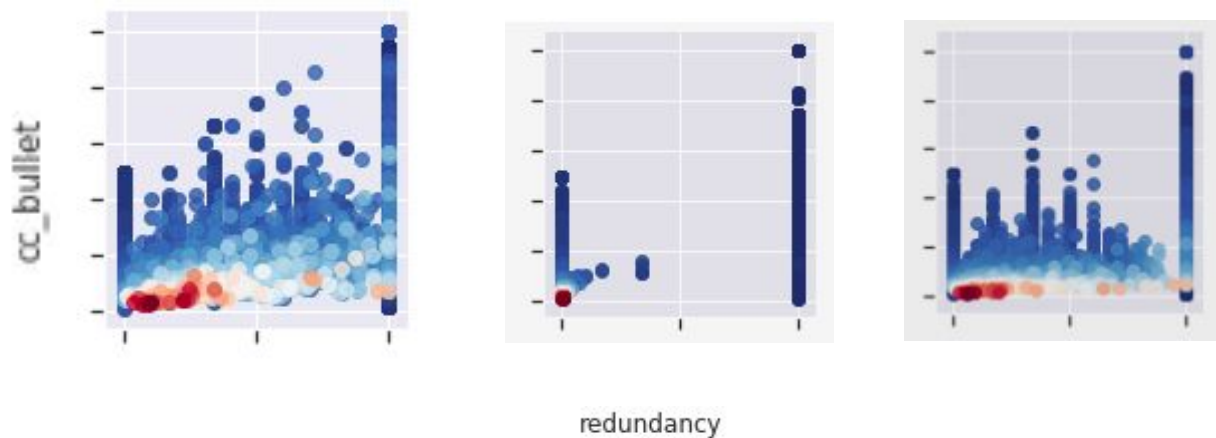
Résultats - Analyse bipartie

Taille des bicliques : Distribution cumulative inverse (ICDF)



Résultats - Analyse bipartie

Corrélation redondance -coefficient de clustering



Autres résultats : Confère code.

Conclusion

Résultats prometteurs pour les métriques mesurées en biparti.

Améliorations :

- Algorithme d'énumération de bicliques plus efficace
- Etendre l'étude à d'autres métriques (monopoly coefficient, dispersion coefficient, diffusion, etc.)
- Faire une étude qualitative des résultats obtenus

Sources principales

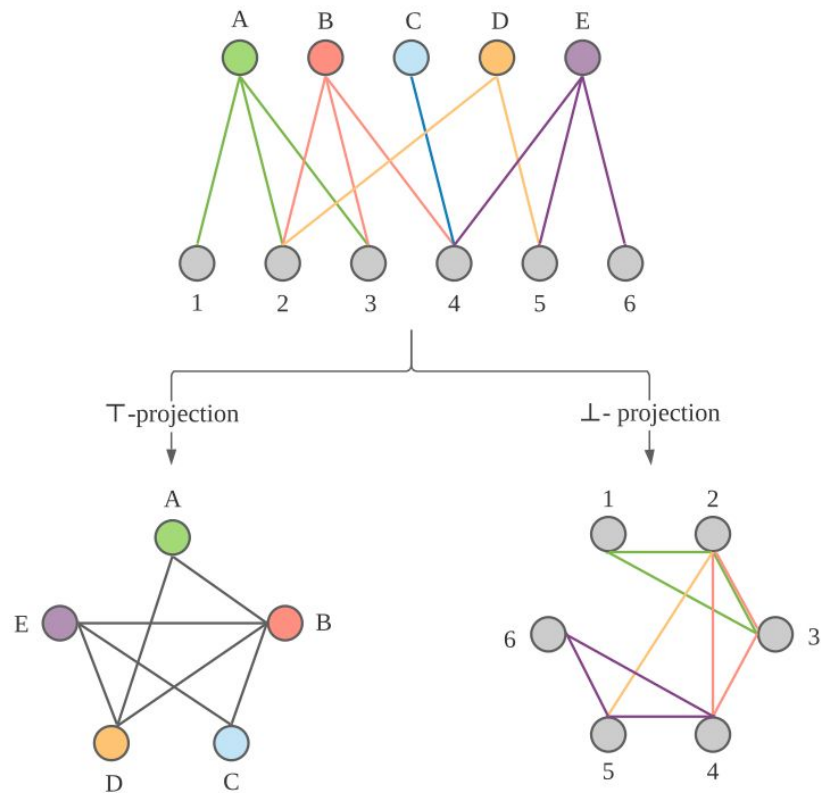
- [1] Fabien Tarissan and Lionel Tabourier. A random model that relies on maximal bicliques to preserve the overlaps in bipartite networks. In 8th International Conference on Complex Networks and their Applications, Lisbonne, Portugal, December 2019.
- [2] Émilie Coupechoux and Fabien Tarissan. Un modèle pour les graphes bipartis aléatoires avec redondance. In 4ème Journées Modèles et l'Analyse des Réseaux : Approches Mathématiques et Informatique (MARAMI'13), Saint-Etienne, France, October 2013.
- [3] Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. Basic Notions for the Analysis of Large Two-mode Networks. volume 30, pages 31–48. Elsevier, January 2008.
- [4] Enver Kayaaslan. On enumerating all maximal bicliques of bipartite graphs. pages 105–108, 01 2010.
- [5] Peter Damaschke. Enumerating maximal bicliques in bipartite graphs with favorable degree sequences. Information Processing Letters, 114:317–321, June 2014.

Questions

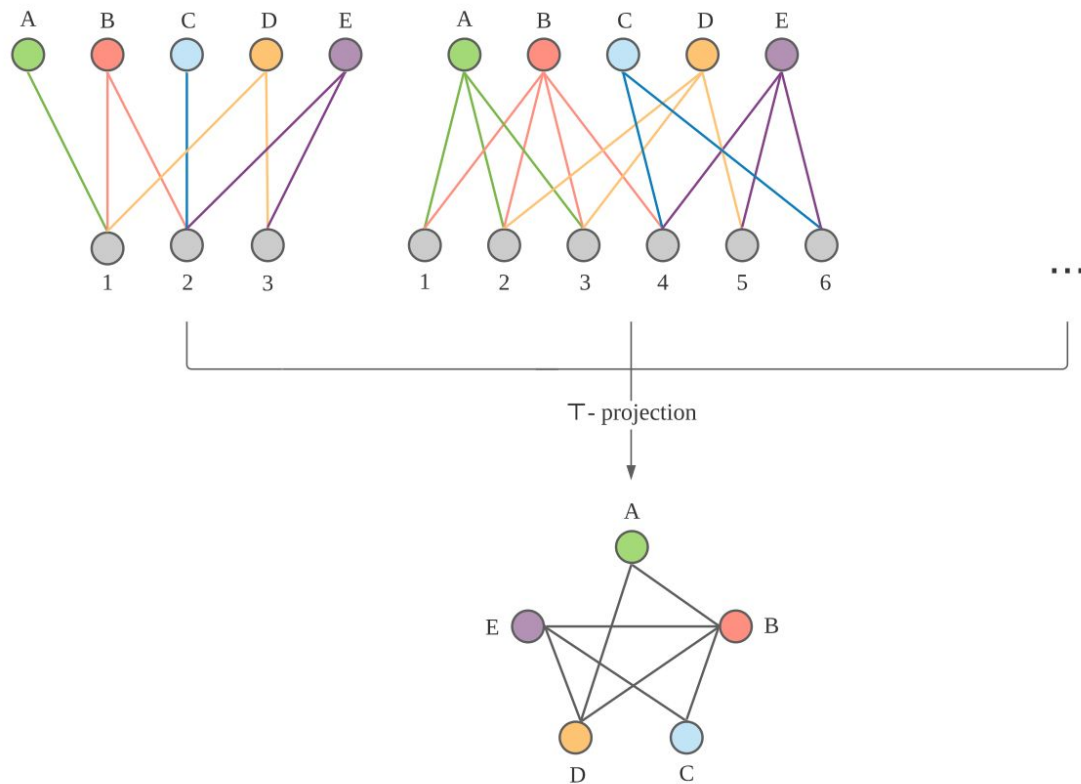
Résultats - Analyse unipartie

	real	configuration model	random	maxtop	maxbottom	maxnodes	naive_tri
nb_vertices	22015	292	292	292	292	292	292
nb_edges	226873	3129	2013	2501	1449	2381	2483
density	0.000936257	0.0736478	0.0473803	0.0588665	0.0341054	0.056042	0.0584428
clustering_coeff	0.804555	0.414234	0.506564	0.488632	0.47663	0.491209	0.496199
nb_connected_components	1188	9	13	17	23	17	22
diameter	17	5	5	6	5	5	5
Assortativity	0.70699	-0.0654871	0.0498075	-0.0399532	-0.015579	-0.0347291	-0.00481483
degree_avg	20.6108	21.4315	13.7877	17.1301	9.92466	16.3082	17.0068
degree_sum	453746	6258	4026	5002	2898	4762	4966
degree_min	0	0	0	0	0	0	0
degree_max	176	124	90	106	61	92	85

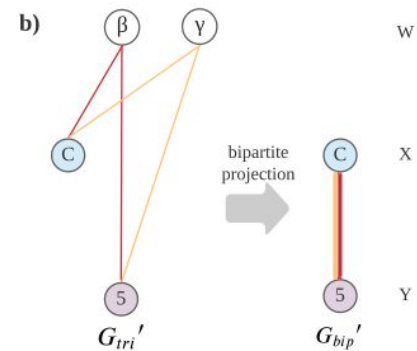
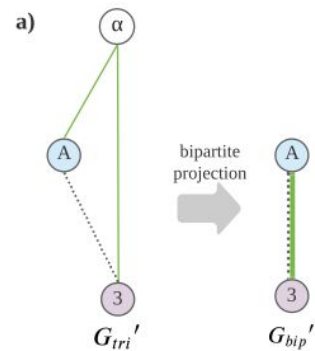
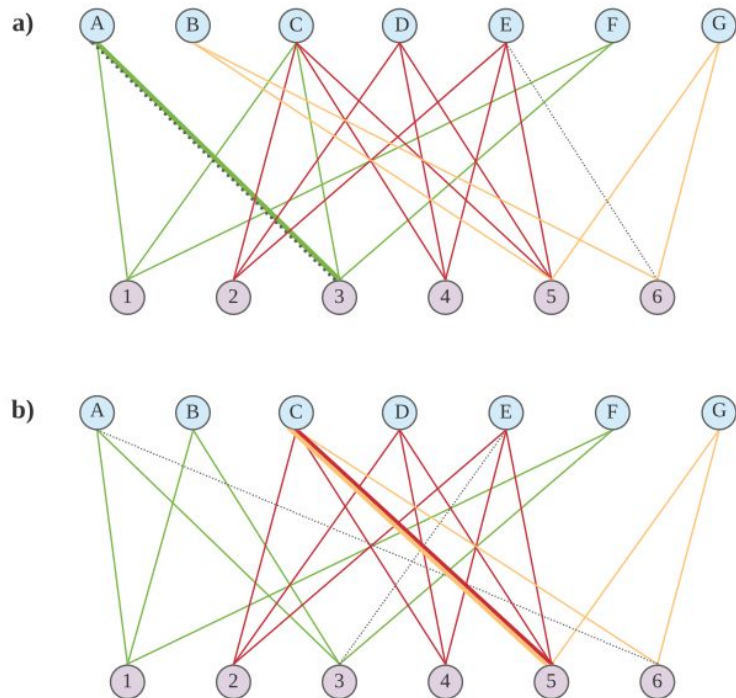
Projeté d'un graphe



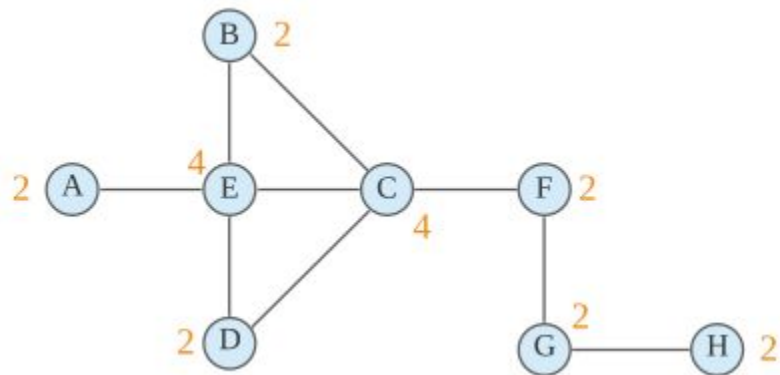
Projeté d'un graphe



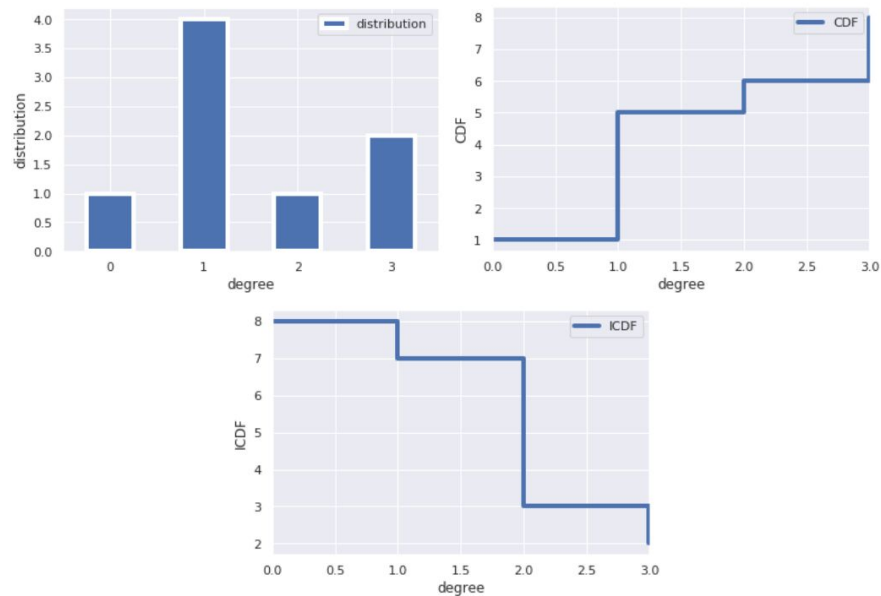
Liens multiples



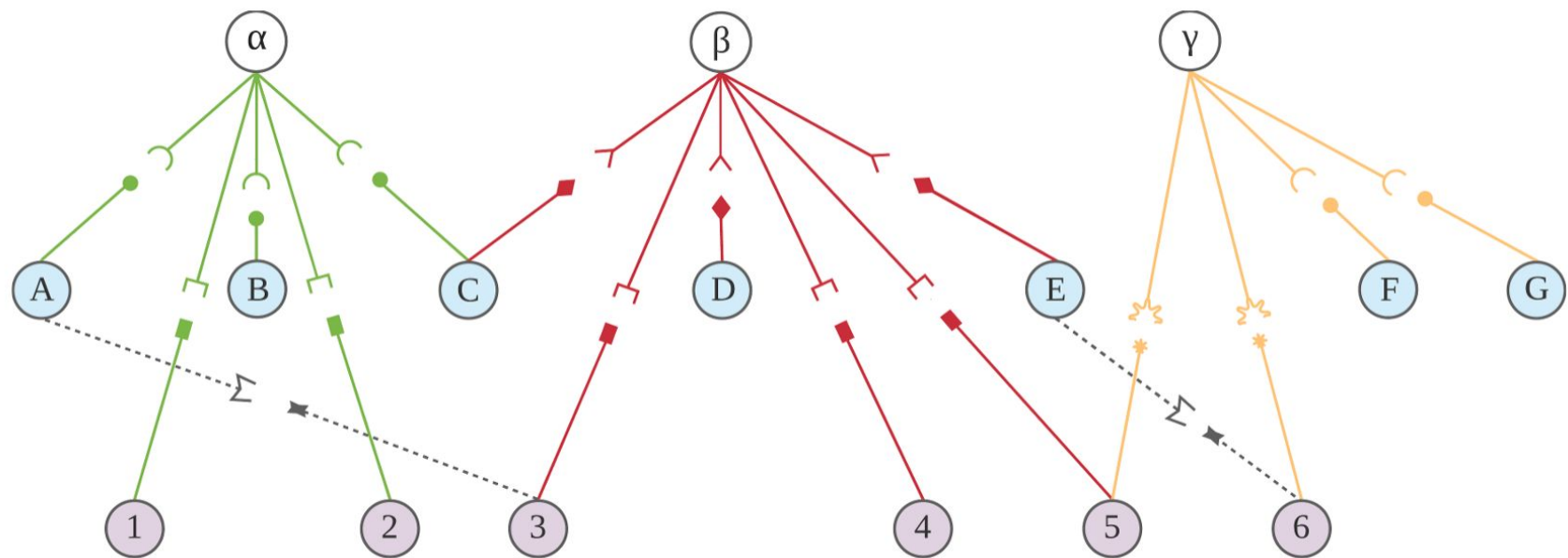
ICDF



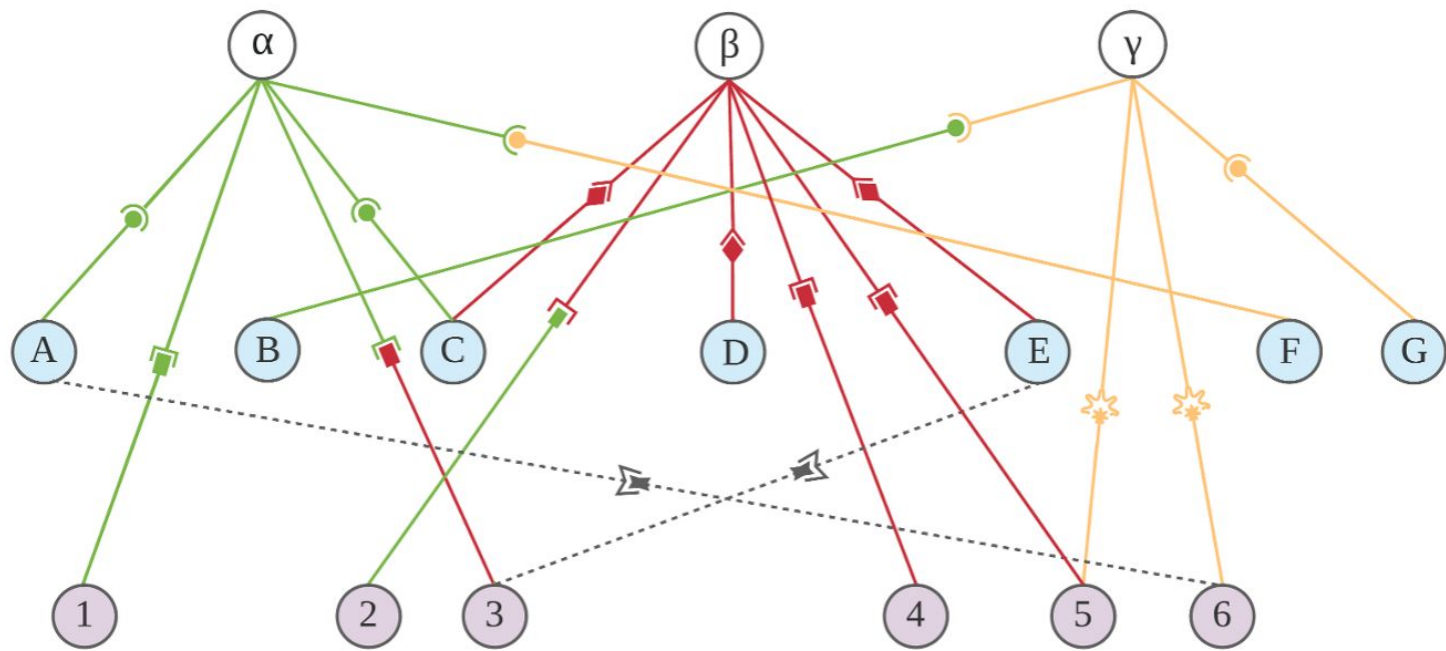
degree	distribution	CDF	ICDF
1	1	1	8
2	4	5	7
3	1	6	3
4	2	8	2



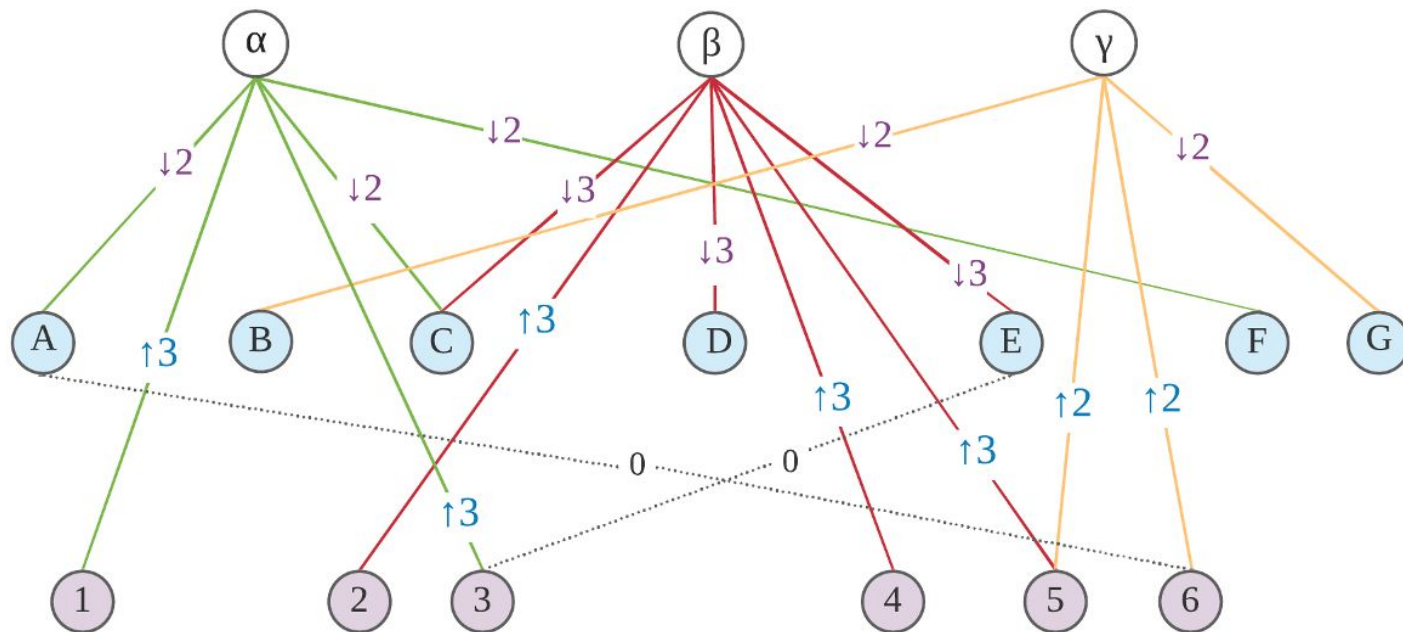
Problématique



Problématique



Problématique



Problématique

