

# Présentation :

## Exploration de la puissance de l'approche par graphe disjoint de l'analyse des données compositionnelles

Organisme d'accueil :  
Unité de recherche BioSTM,  
UR7537 (Laboratoire de  
biomathématique)



Encadré par :  
Emmanuel Curis

Soumis par :  
Jules Bertin

# Aperçu

Présentation des grandes sections de cette présentation

- Introduction
- Présentation de la méthode des graphes
- Étude de la puissance appliquée aux graphes
- Introduction de nouveaux indicateurs et de leurs estimations
- Étude de cas sur un graphe à 3 sommets
- Bilan

# Introduction

Qu'est-ce que sont des données compositionnelles ?

Ex : volume de matière grise dans le cervelet

Quels problèmes amènent ce type de données pour leur étude?

Une solution : l'approche par méthode des graphes

# Méthode des graphes

## **Objectif :**

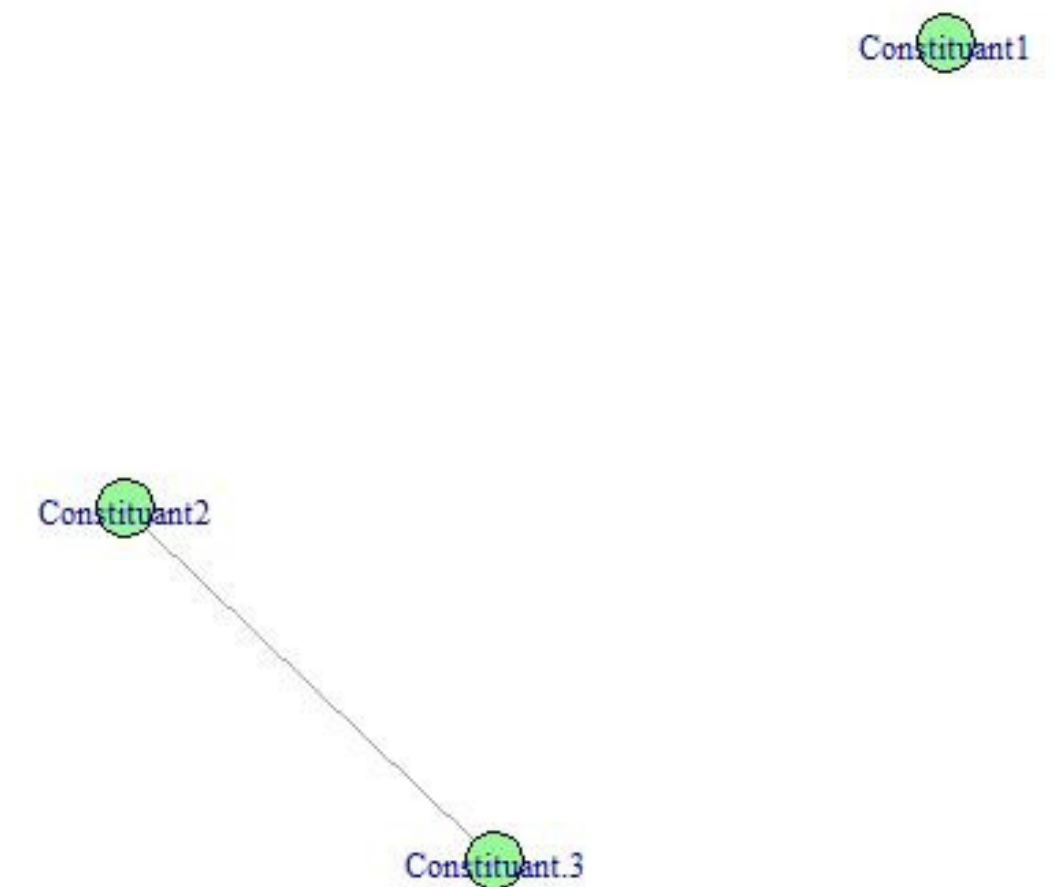
Traduire les variations de moyennes entre deux groupes pour des données compositionnelles

## **Présence d'arête :**

Non-évolution significative de la moyenne du rapport des 2 variables entre les deux groupes.

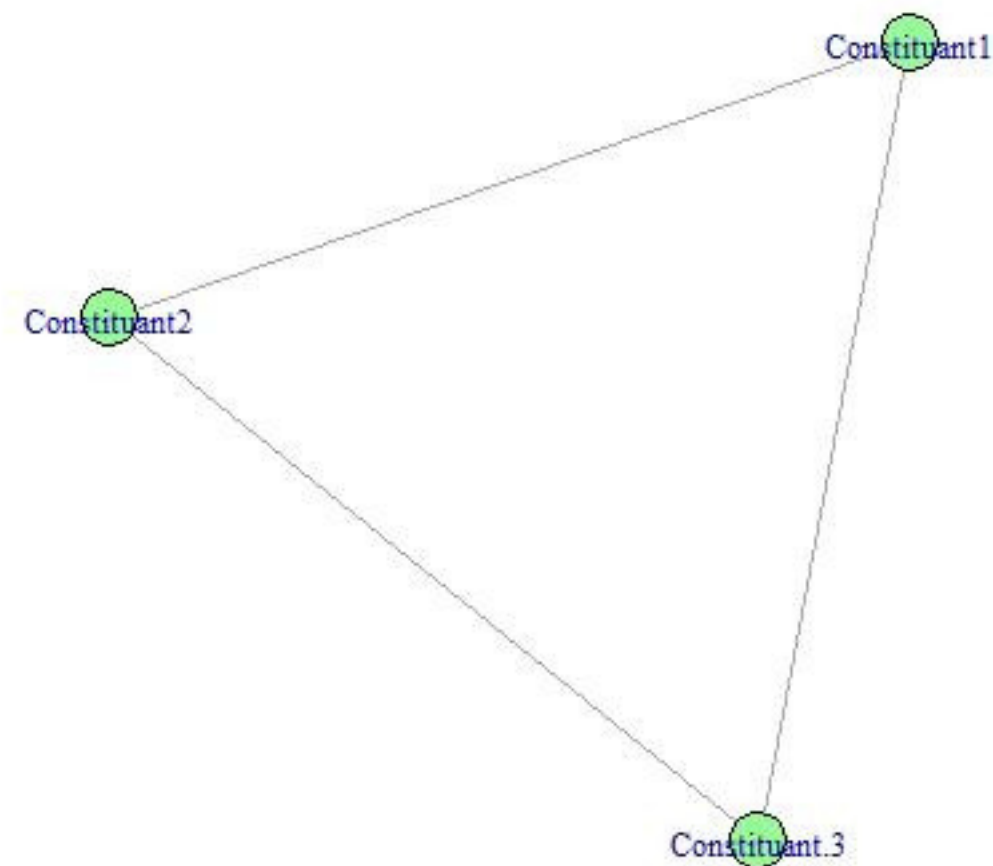
## **Critère de jugement de la significativité :**

Test de Student sur l'égalité des moyennes et seuil de coupure



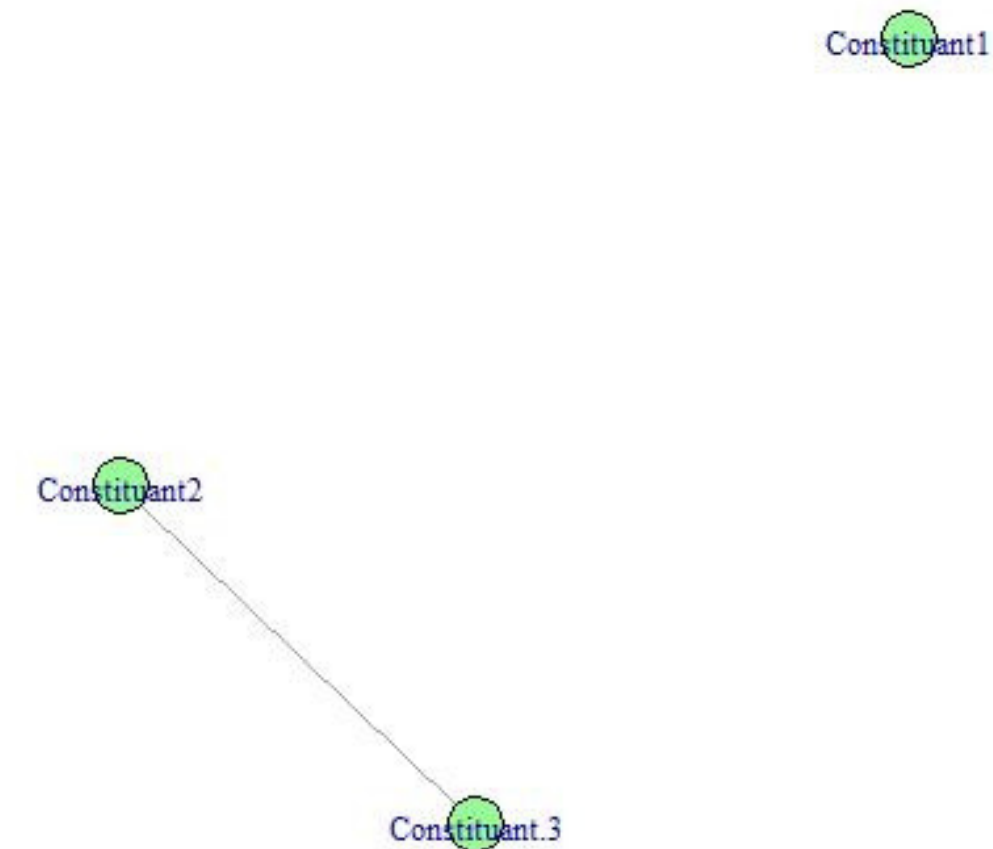
# Hypothèses sur les graphes

H0: Graphe connexe



$$\mathcal{H}_0 : \mu_{\ln(\frac{C.1}{C.3})}^A = \mu_{\ln(\frac{C.1}{C.3})}^B$$

H1: Graphe disjoint



$$\mathcal{H}_1 : \mu_{\ln(\frac{C.1}{C.3})}^A \neq \mu_{\ln(\frac{C.1}{C.3})}^B$$

$$\text{où } \mu_{\ln(\frac{C.1}{C.3})}^A = E \left[ \ln \left( \frac{\text{Constituant1}}{\text{Constituant3}} \right) \mid A \right] \text{ et } \mu_{\ln(\frac{C.1}{C.3})}^B = E \left[ \ln \left( \frac{\text{Constituant1}}{\text{Constituant3}} \right) \mid B \right]$$

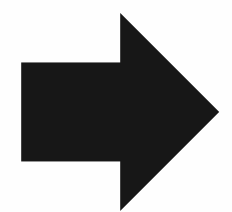
# Questionnement sur la puissance appliquée à la méthode des graphes

Exemple :

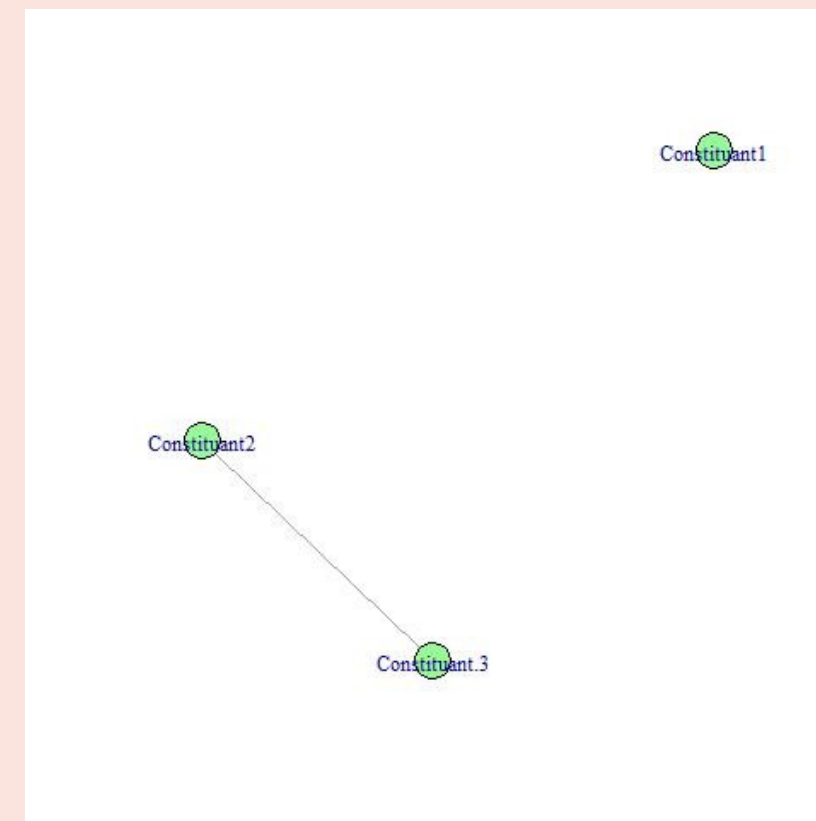
Définition de la puissance usuelle :

$$\forall \theta \in \mathcal{H}_1 \quad h(\theta) = \mathbb{P}(T \in \mathcal{R}_{\mathcal{H}_0} | \theta)$$

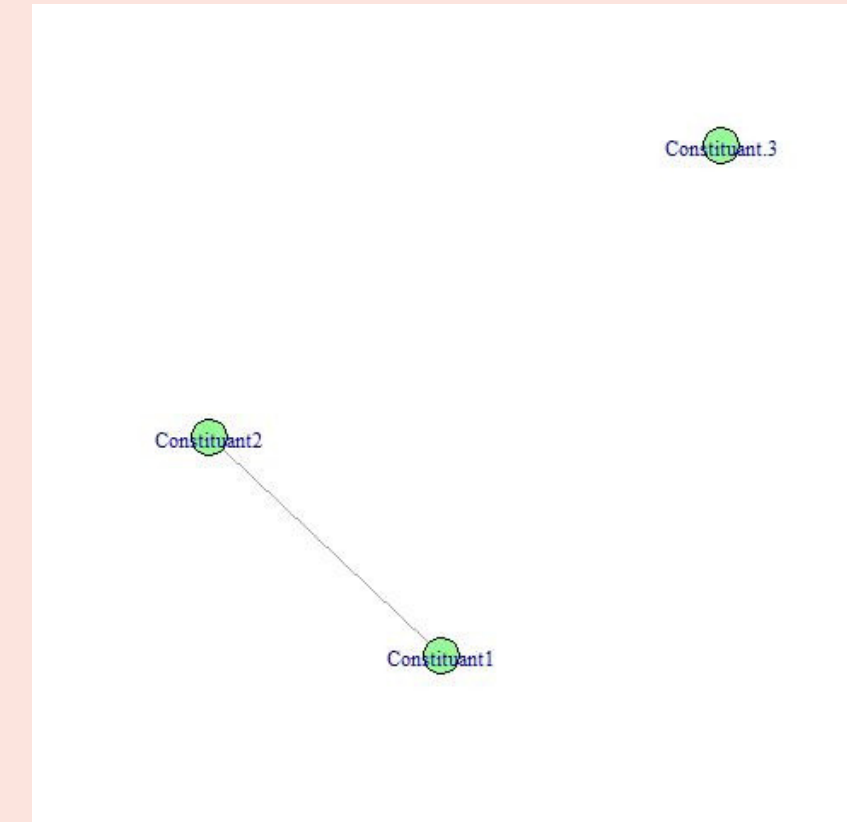
Problème de la composition de  $\mathcal{H}_1$  :



Intérêt de la puissance pour seulement un graphe de  $\mathcal{H}_1$



graphe d'intérêt



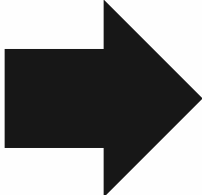
graphe appartenant à  $\mathcal{R}_{\mathcal{H}_0}$  mais qui ne nous intéresse pas pour la puissance

# 2 idées de solution

## 1<sup>re</sup> idée :

2 hypothèses :

- $\mathcal{H}_0$  : tous les graphes sauf celui d'intérêt
- $\mathcal{H}_1$  : graphe d'intérêt

 Problème : complexité pour calculer le risque de 1<sup>re</sup> espèce

## 2<sup>e</sup> idée :

Diviser  $\mathcal{H}_1$  en deux sous hypothèses :

- le nouveau  $\mathcal{H}_1$  : le graphe d'intérêt
- $\mathcal{H}_2$  : tous les graphes sauf le graphe d'intérêt et le graphe connexe

# Nouvelles hypothèses pour les graphes

## Modèle initial

2 hypothèses :

- $\mathcal{H}_0$  : graphe connexe
- $\mathcal{H}_1$  : les graphes disjoints

## Modèle amélioré

3 hypothèses :

- $\mathcal{H}_0$  : graphe connexe
- $\mathcal{H}_1$  : le graphe d'intérêt
- $\mathcal{H}_2$  : le restant des graphes



# Introduction de nouveaux indicateurs

## Puissance corrigée :

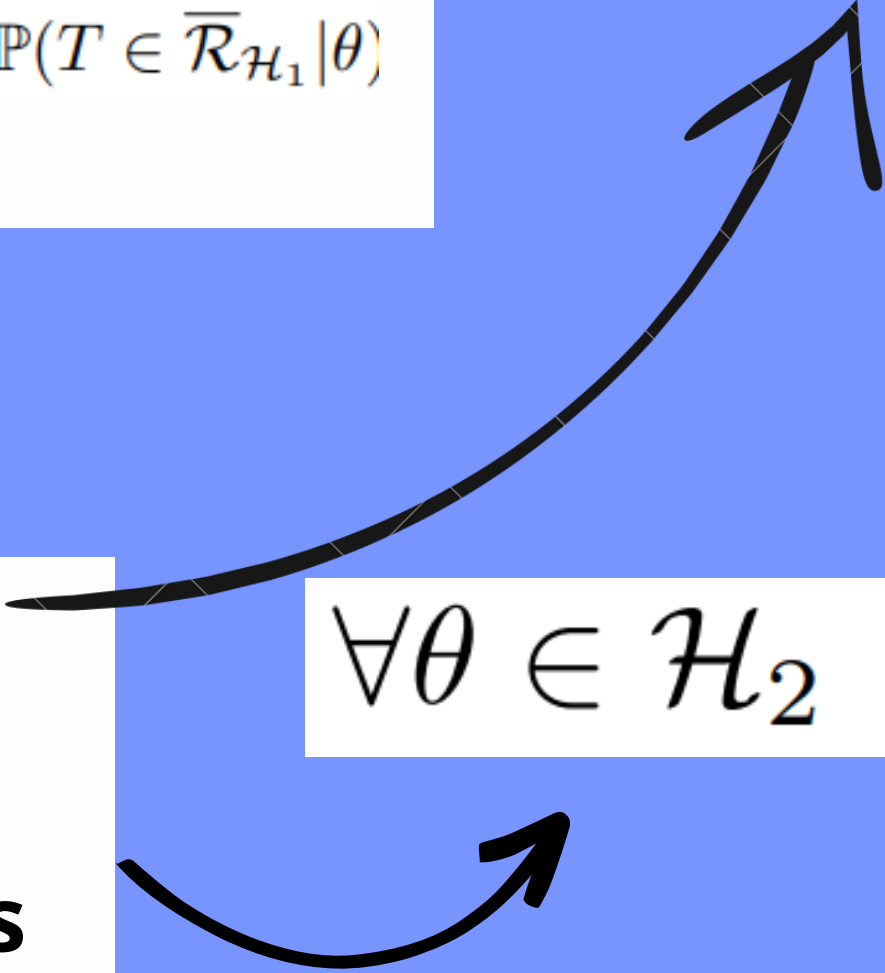
Probabilité d'accepter  $\mathcal{H}_1$  à raison

$$\forall \theta \in \mathcal{H}_1 \quad h'(\theta) = \mathbb{P}(T \in \mathcal{R}_{\mathcal{H}_0} \cap \mathcal{R}_{\mathcal{H}_2} | \theta) = \mathbb{P}(T \in \overline{\mathcal{R}}_{\mathcal{H}_1} | \theta)$$

$$\forall \theta \in \mathcal{H}_2 \quad \gamma_0 = \mathbb{P}(T \notin \mathcal{R}_{\mathcal{H}_0} | \theta)$$

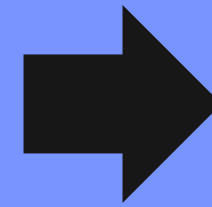
L'hypothèse  $\mathcal{H}_2$  introduit de nouveaux risques :

**les risques de 3èmes espèces**

$$\forall \theta \in \mathcal{H}_2 \quad \gamma_1 = \mathbb{P}(T \in \mathcal{R}_{\mathcal{H}_0} | \theta)$$


# Estimations des indicateurs

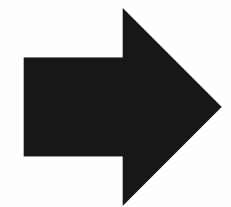
Utilisation de simulations



Être sûr que les données suivent l'hypothèse recherchée pour l'indicateur

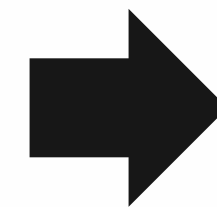
## Risque de première espèce :

Pour déterminer le seuil de coupure du graphe associé à un risque de première espèce



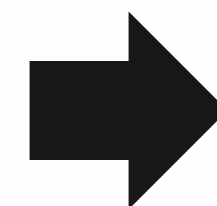
**Simulation sous des données qui donnent le graphe  $\mathcal{H}_0$  (connexe)**

## Puissance :



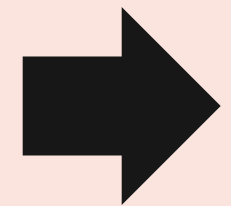
Simulation sous des données qui donnent le graphe  $\mathcal{H}_1$  (d'intérêt)

## Risques de 3<sup>e</sup> espèce :

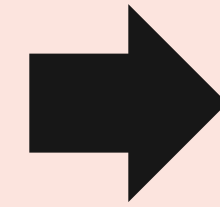


Simulation sous des données qui donnent l'un des graphes de  $\mathcal{H}_2$

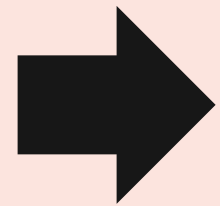
# L'exemple du graphe à 3 sommets



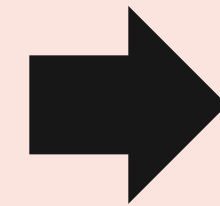
Données préliminaires sur 52 patients (réparties dans 5 catégories dont 4 pathologies neurodégénératives)



Calibration des simulations à l'aide de ces données

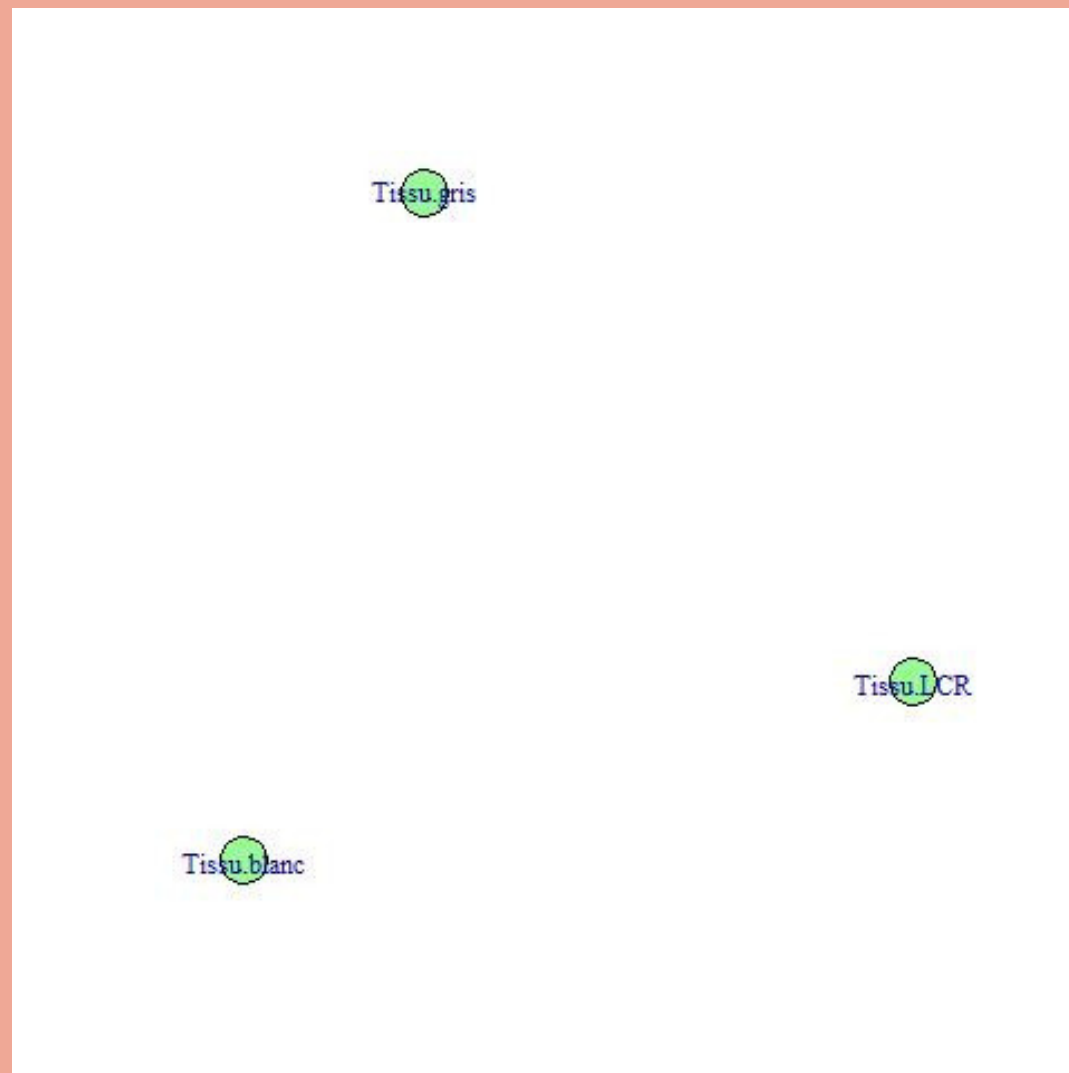


3 variables d'intérêts : Tissu gris, Tissu blanc et Tissu LCR

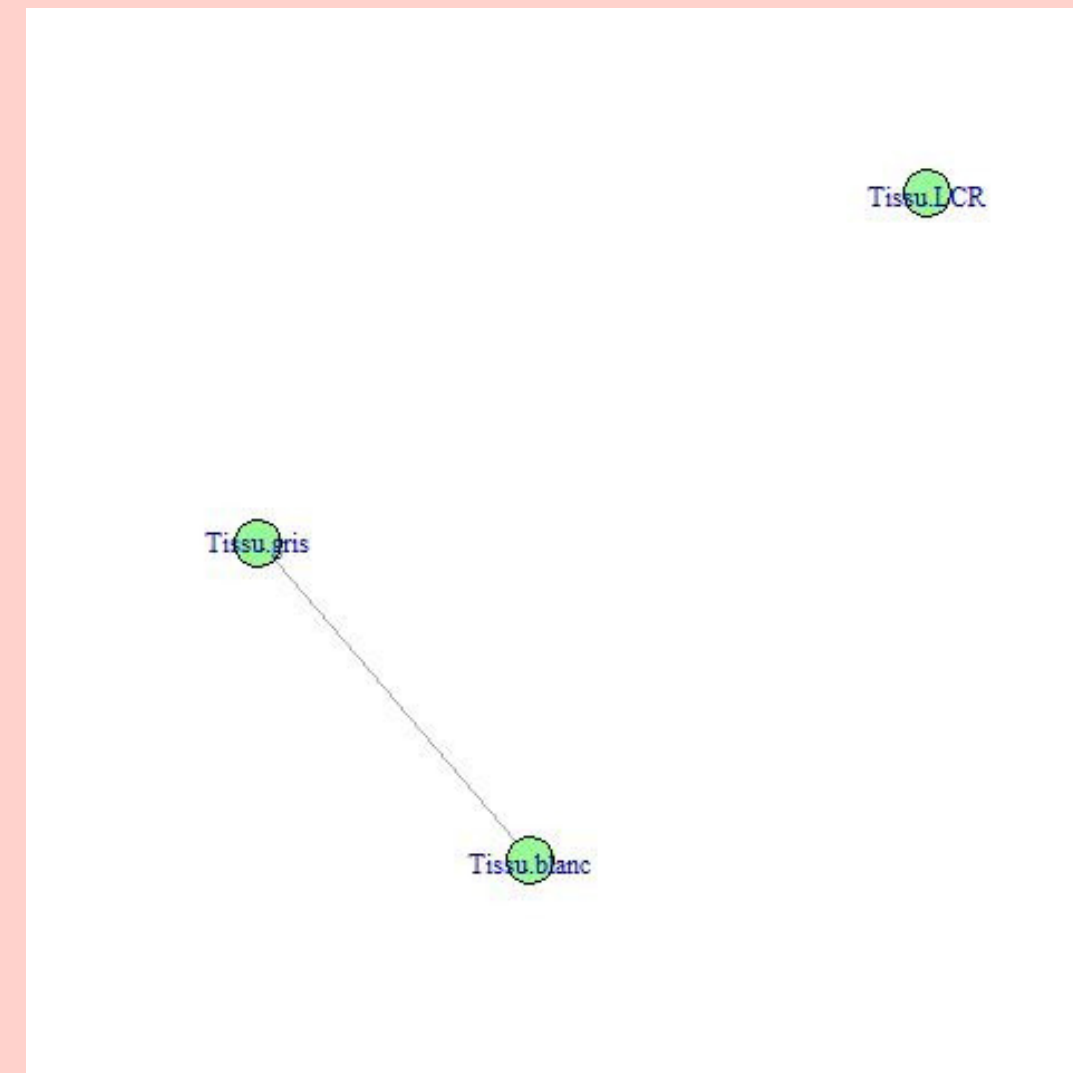


On fait suivre les volumes des Tissus des lois lognormales

# L'exemple du graphe à 3 sommets : 2 Cas à étudier



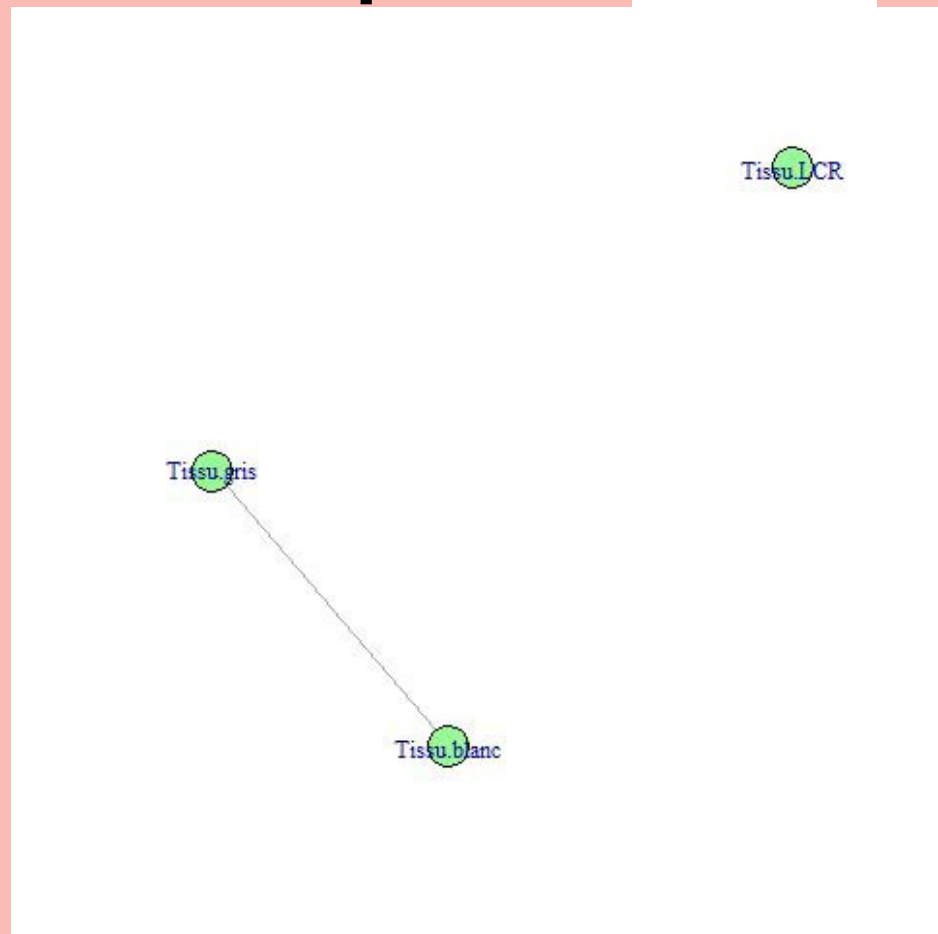
le graphe d'intérêt :  
totalement disjoint



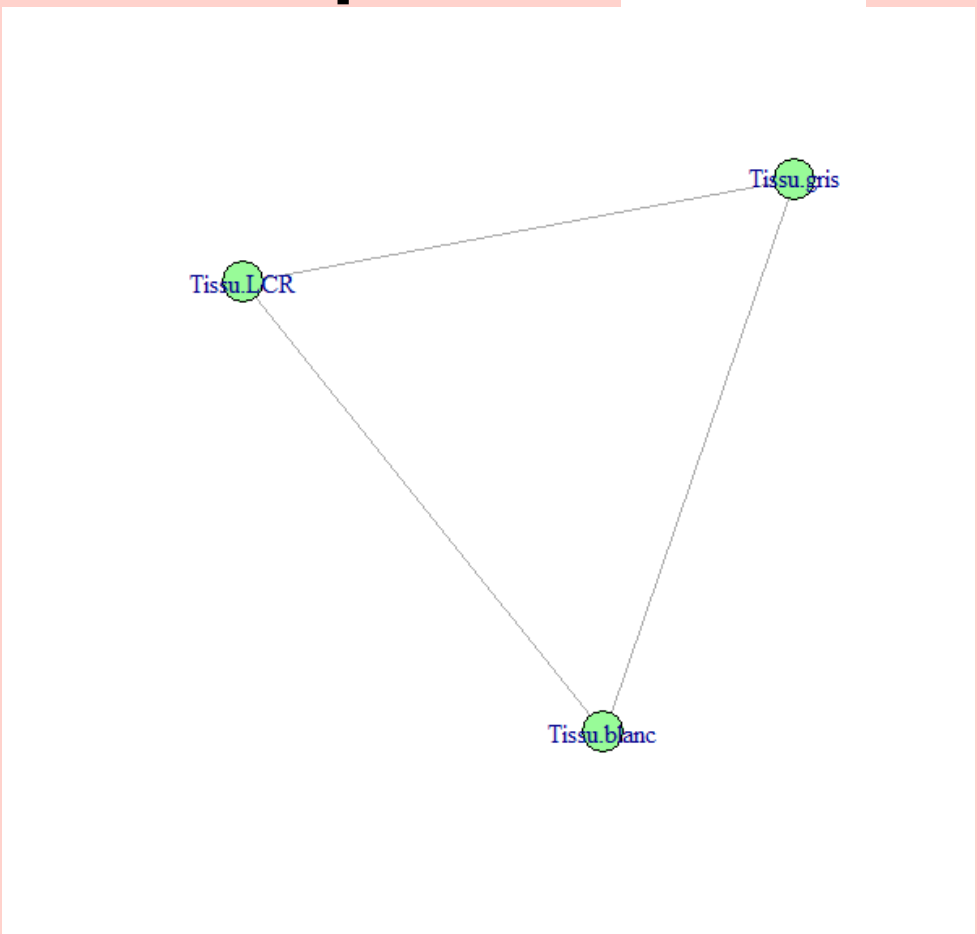
le graphe d'intérêt :  
avec un sommet isolé

# Le graphe disjoint (à 2 groupes)

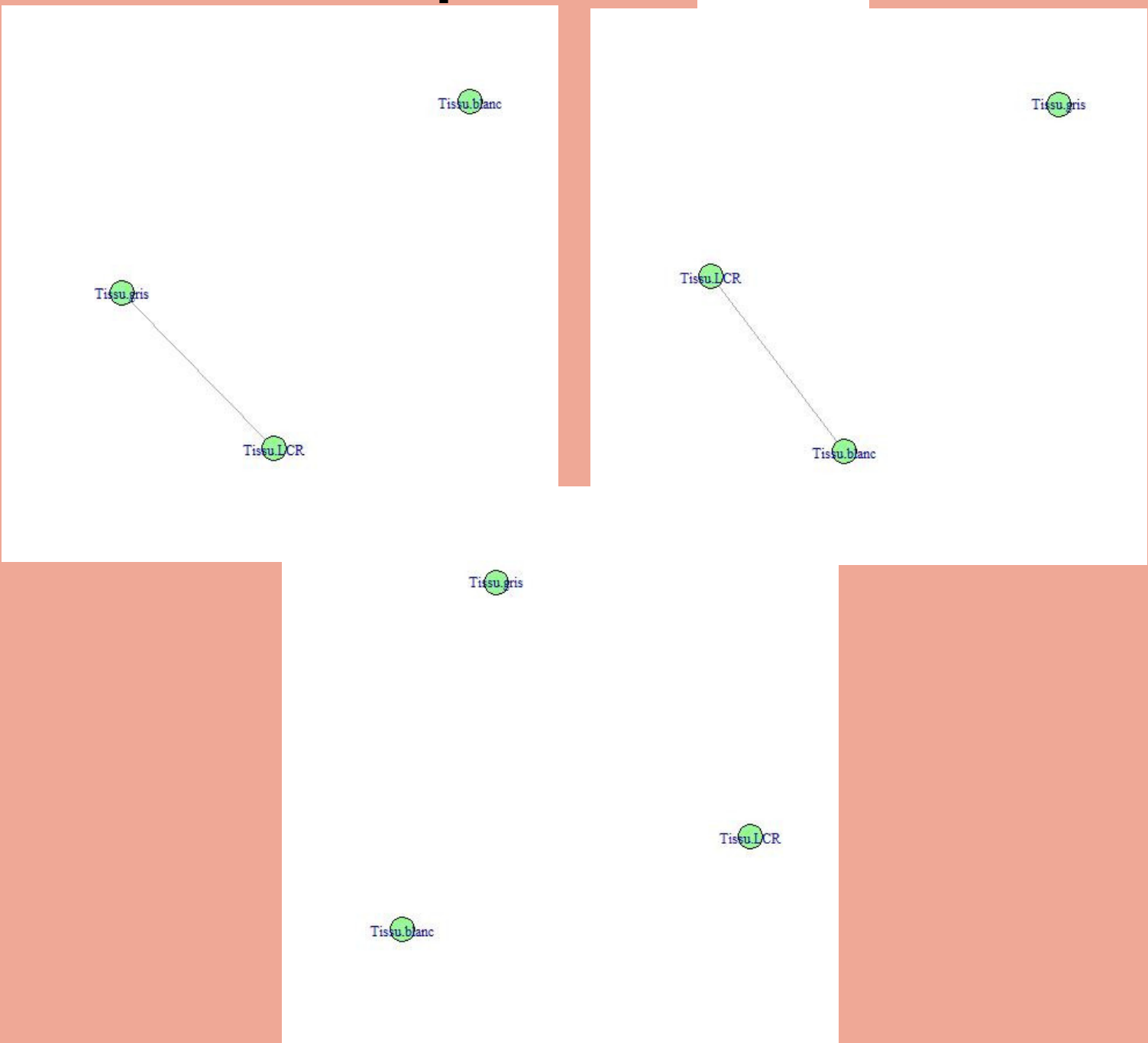
Graphe  $\mathcal{H}_1$



Graphe  $\mathcal{H}_0$

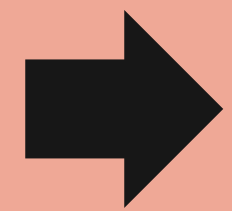


Graphes  $\mathcal{H}_2$

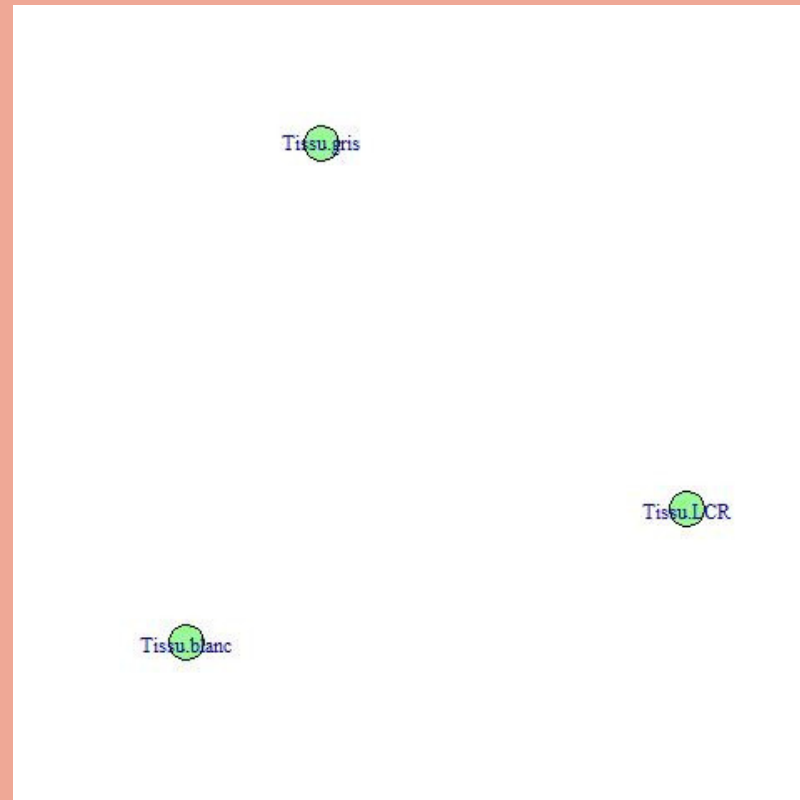


# Introduction d'erreurs

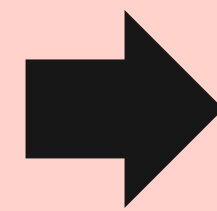
## La simple erreur :



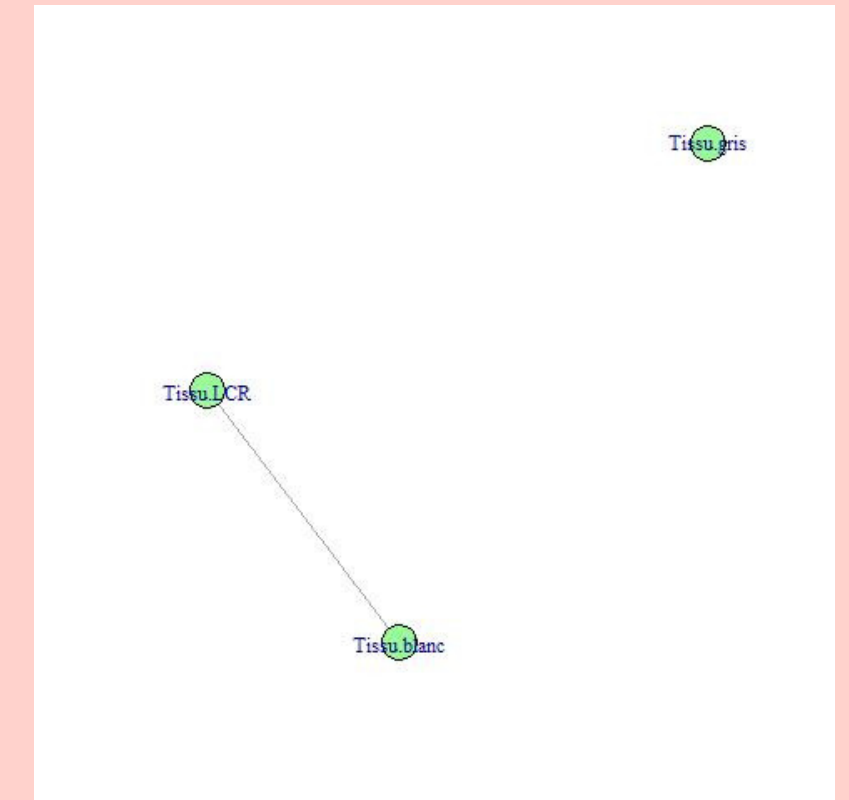
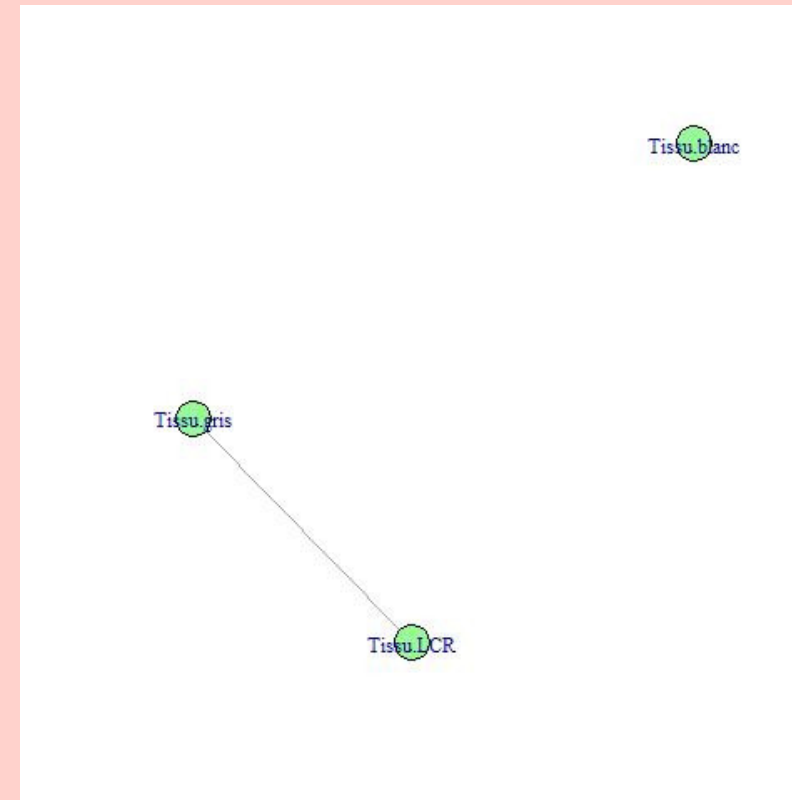
Arête entre Tissu\_gris et tissu\_blanc a été coupé



## La double erreur :



Arête entre Tissu\_gris et tissu\_blanc a été coupé + **Création d'une nouvelle arête**



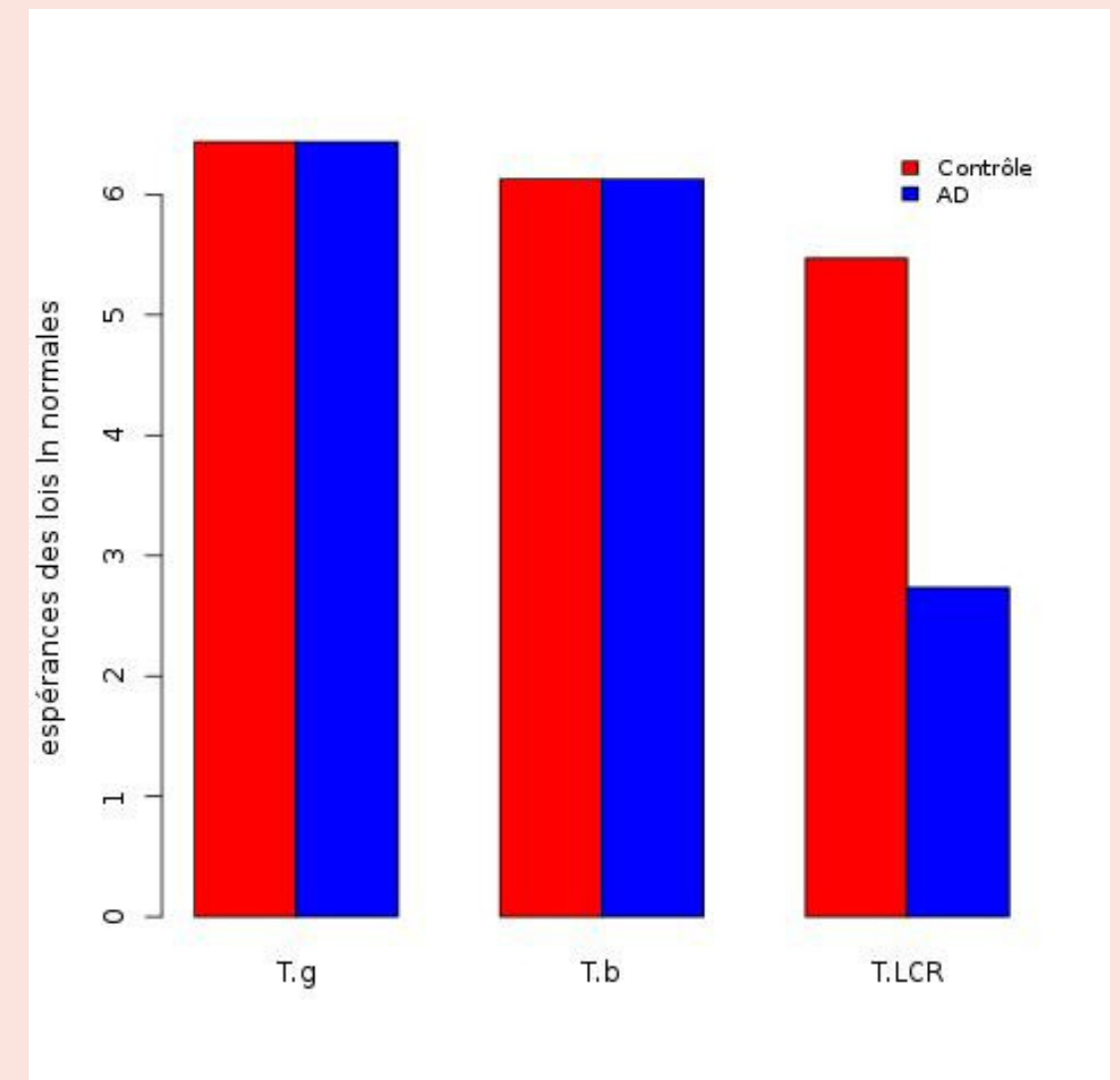
# Comparaison puissance et puissance corrigée

Puissance = puissance corrigée + simple erreur + double erreur

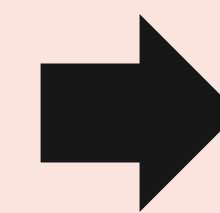
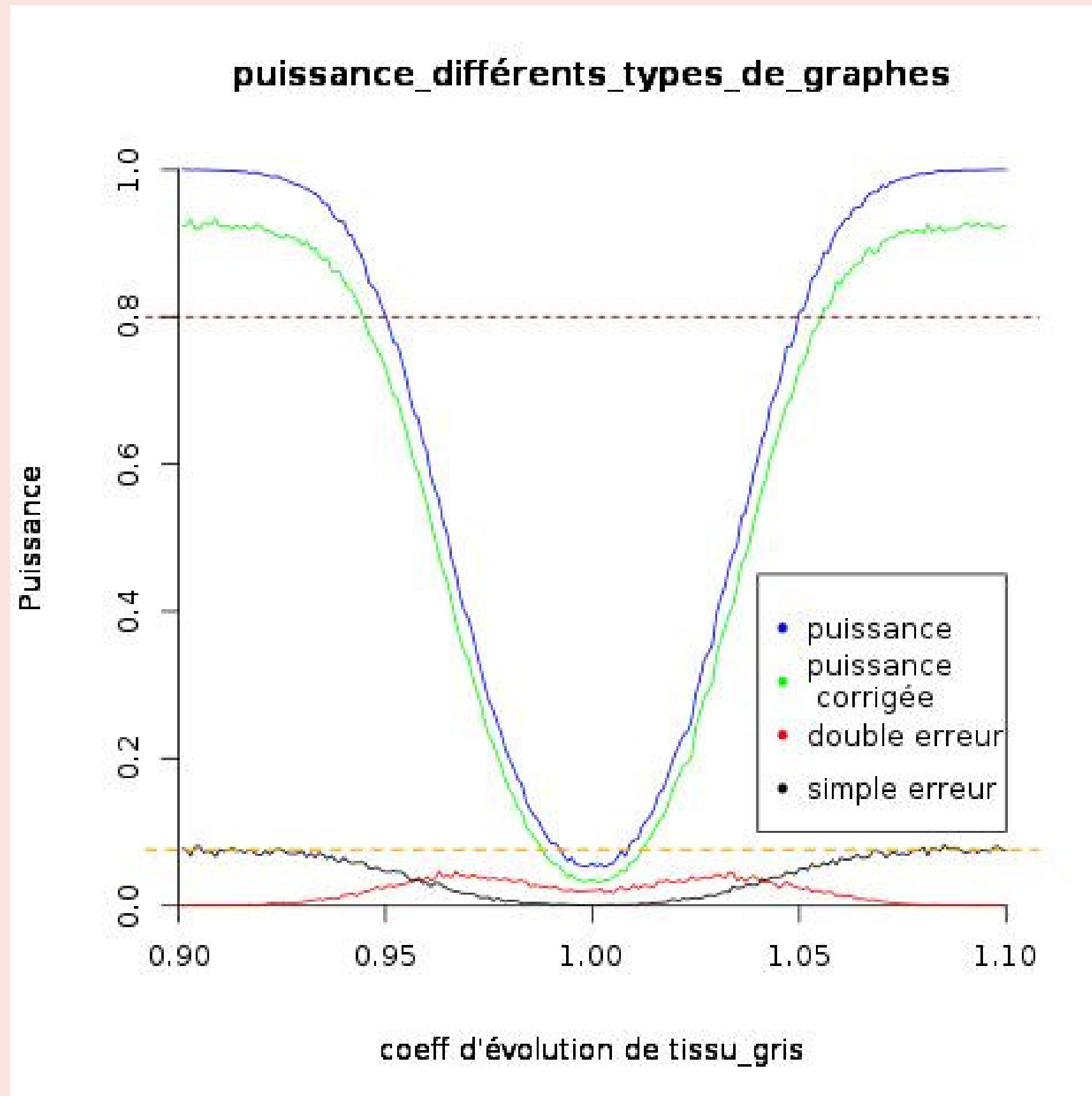
Objectif : Étudier l'évolution des puissances suivant les données de simulations

➡ Simulations avec l'espérance de la loi ln-normale du groupe AD qui va varier (à l'aide d'un coefficient d'évolution noté  $C_E$ )

➡  $E[\ln(Tissu.LCR) | AD] = C_E * E[\ln(Tissu.LCR) | Controle]$



# Graphe puissances

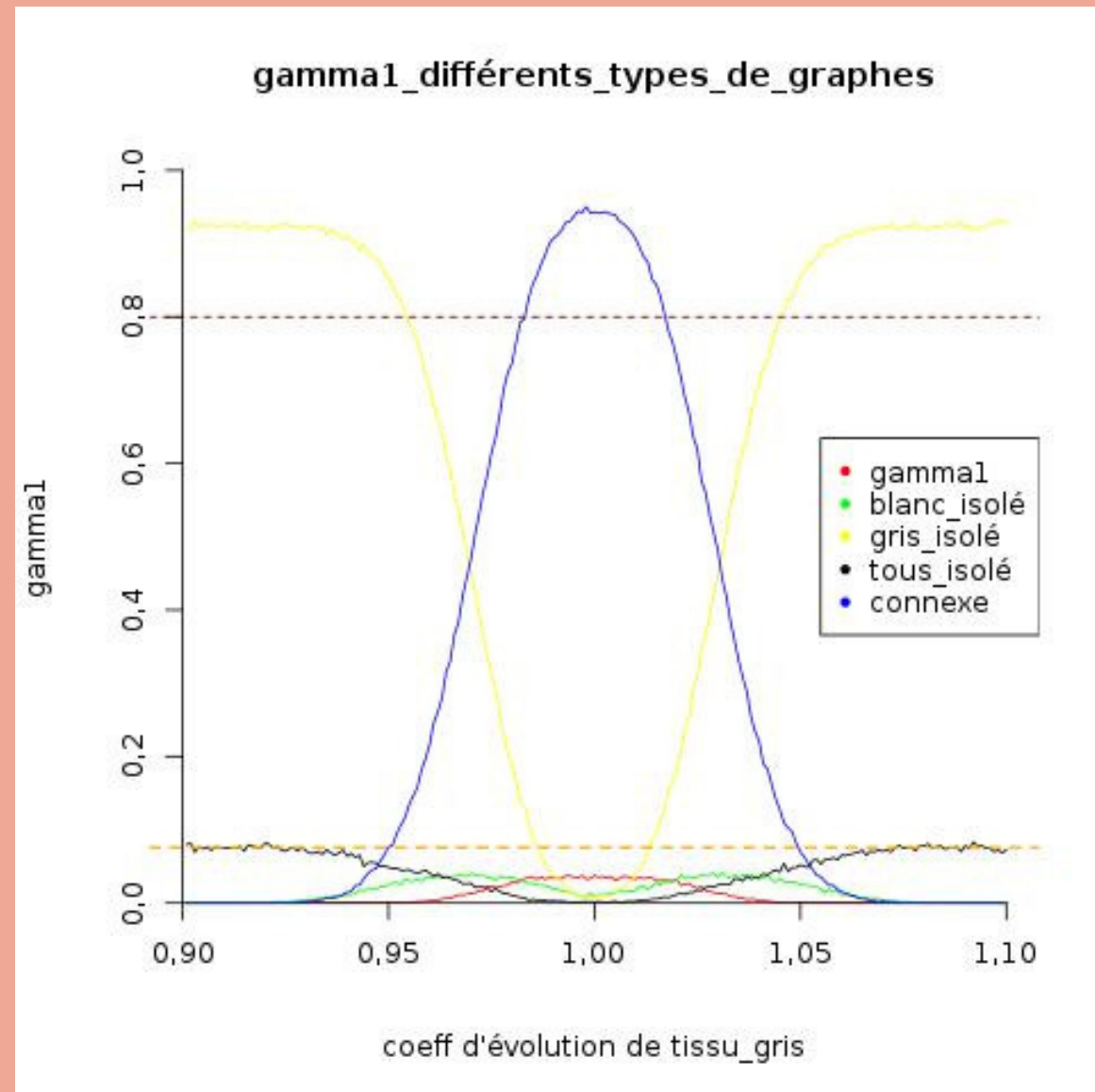


Seuil de coupure : 0.076  
->pointillés orange

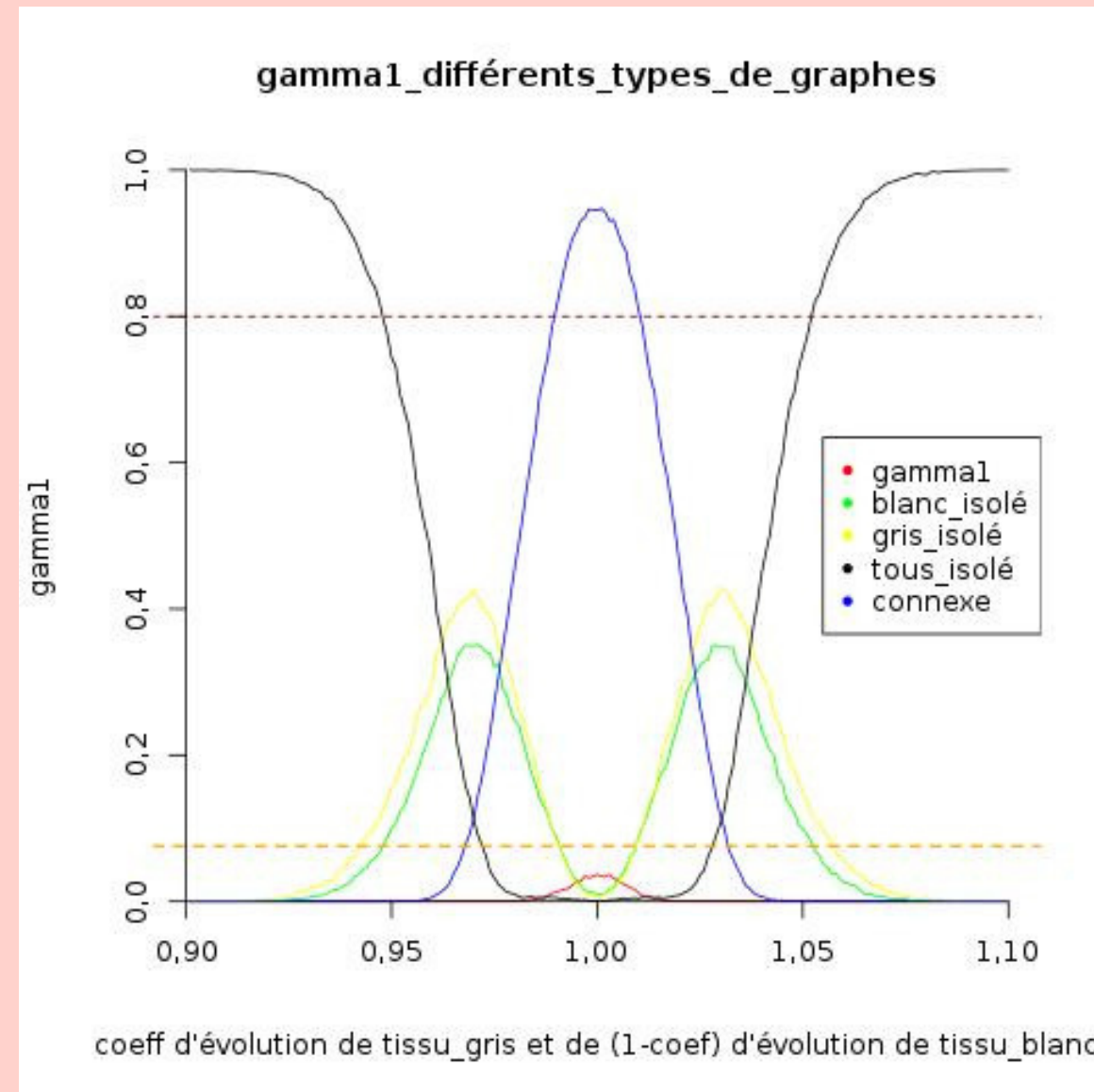


# Risque de 3ème espèce

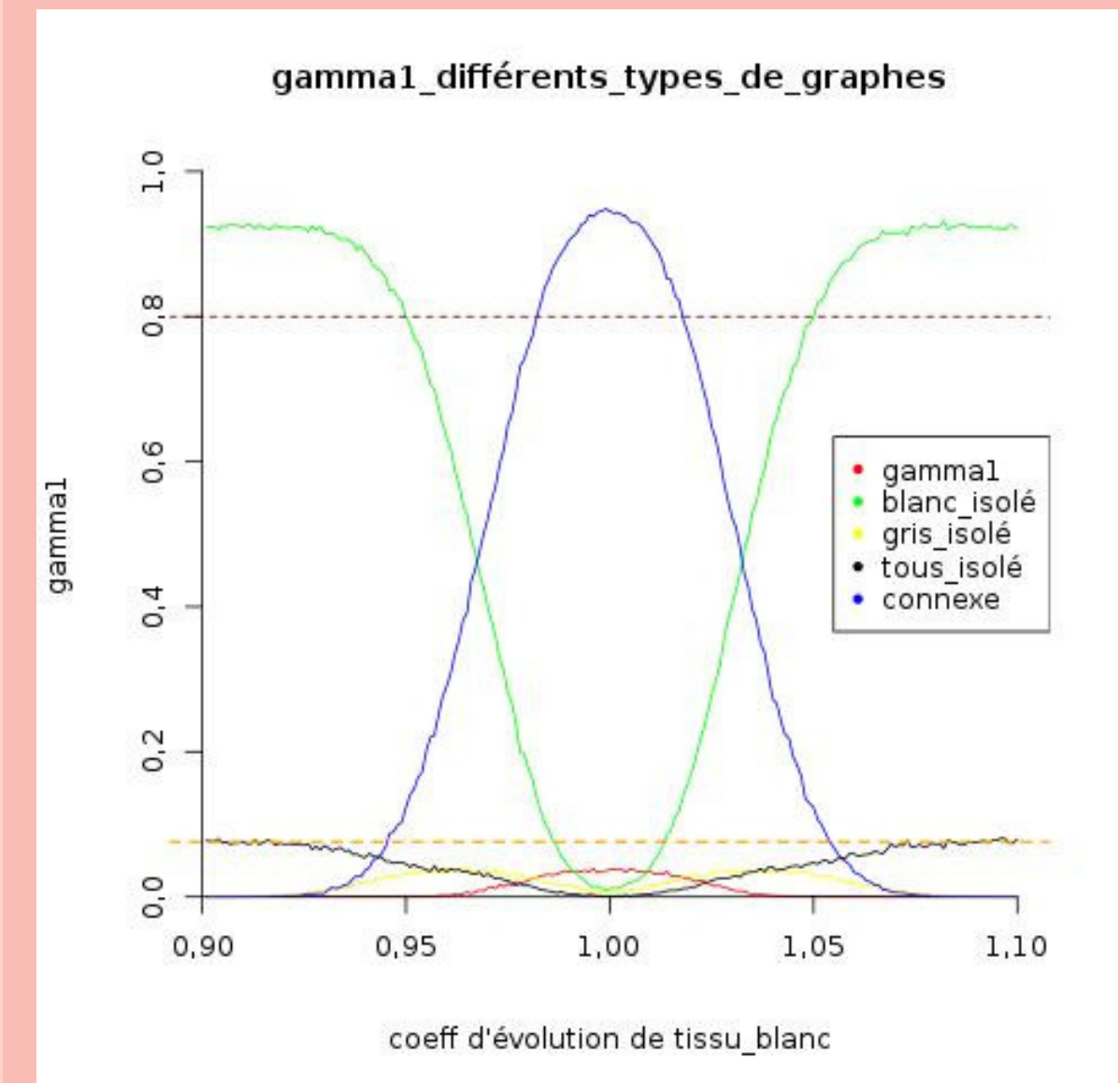
Simulation sous les 3 graphes de  $\mathcal{H}_2$



Graphe avec tissu\_gris  
isolé



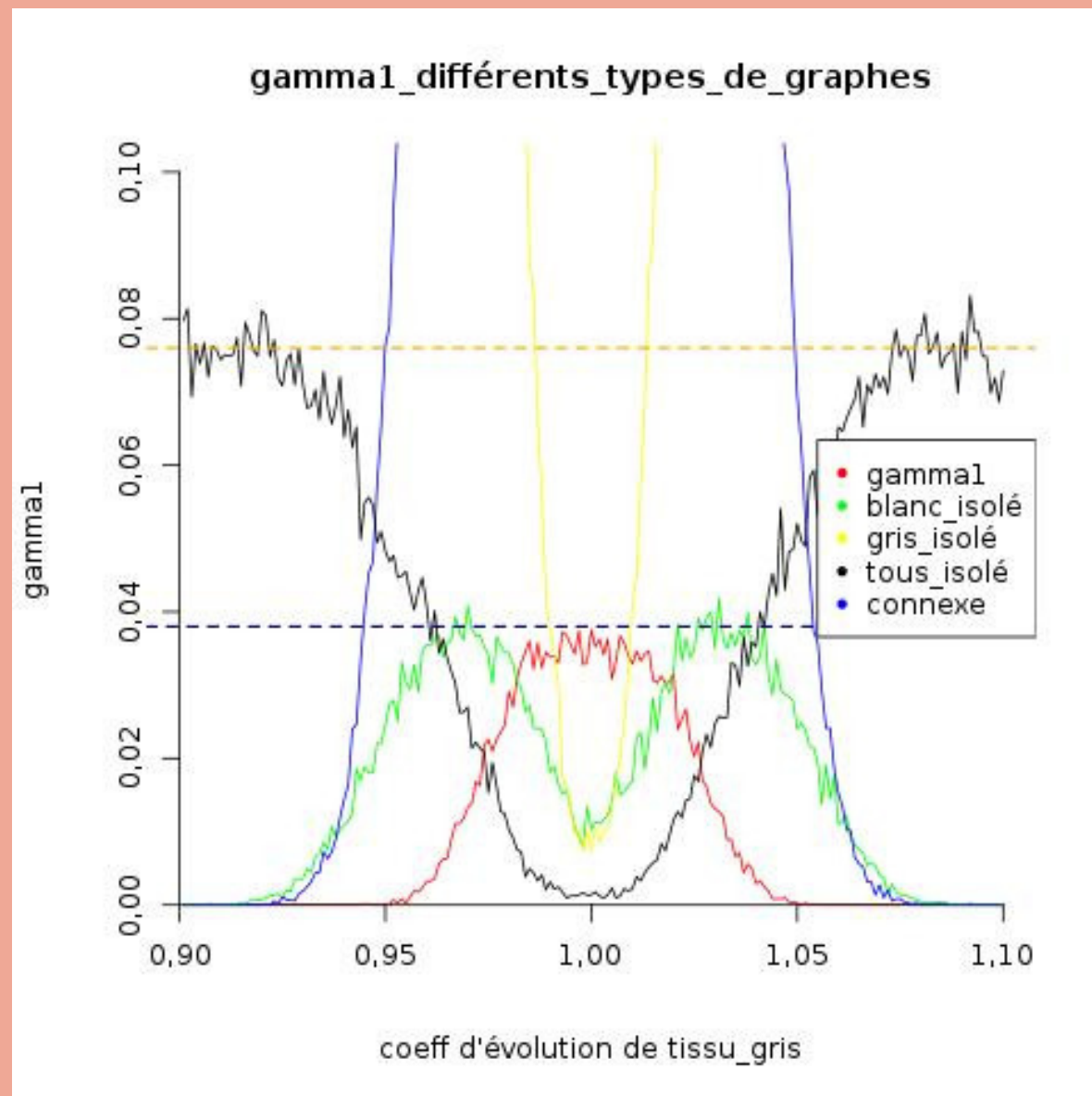
Graphe complètement  
disjoint



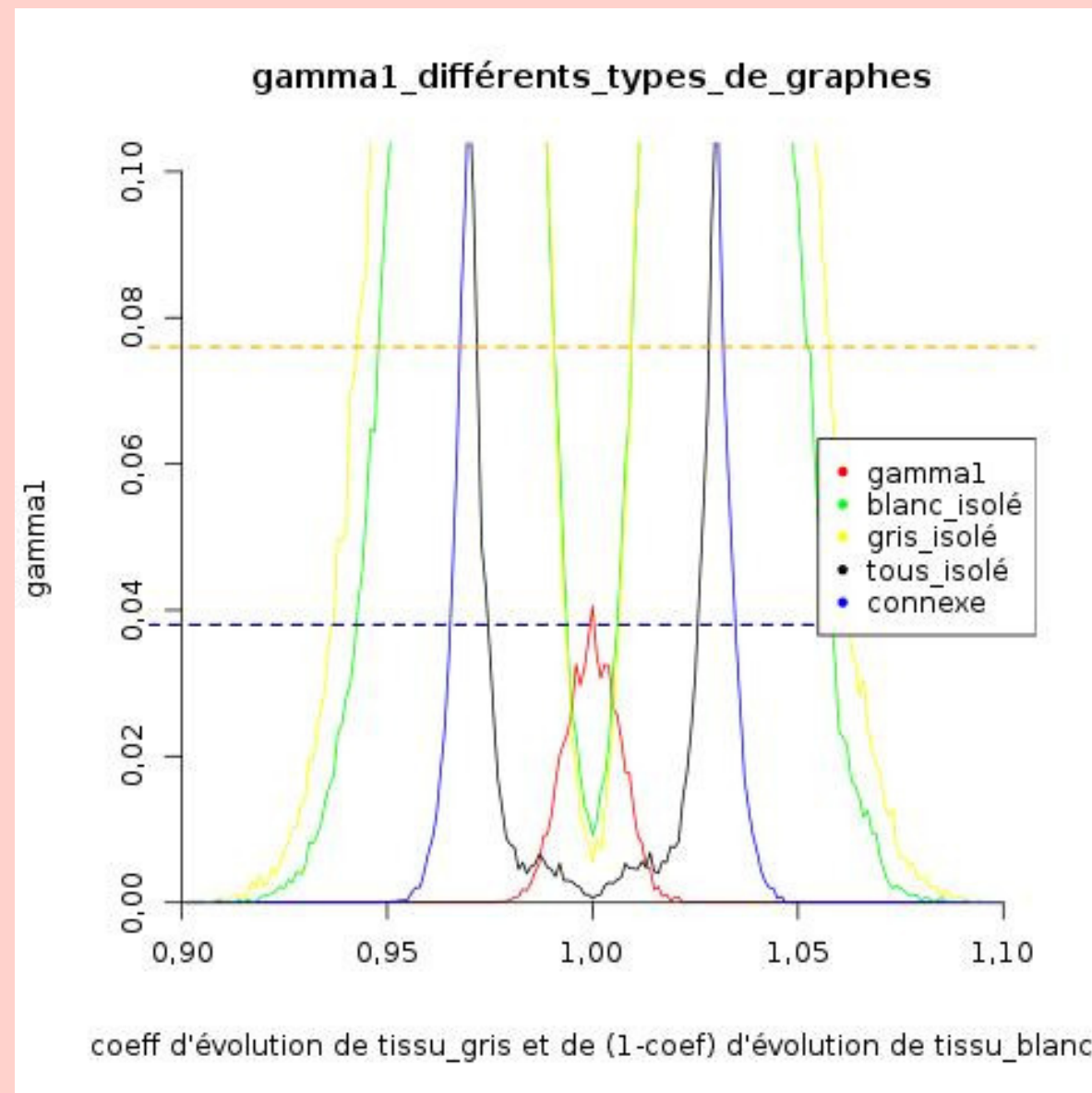
Graphe avec tissu\_blanc  
isolé

# Risque de 3ème espèce

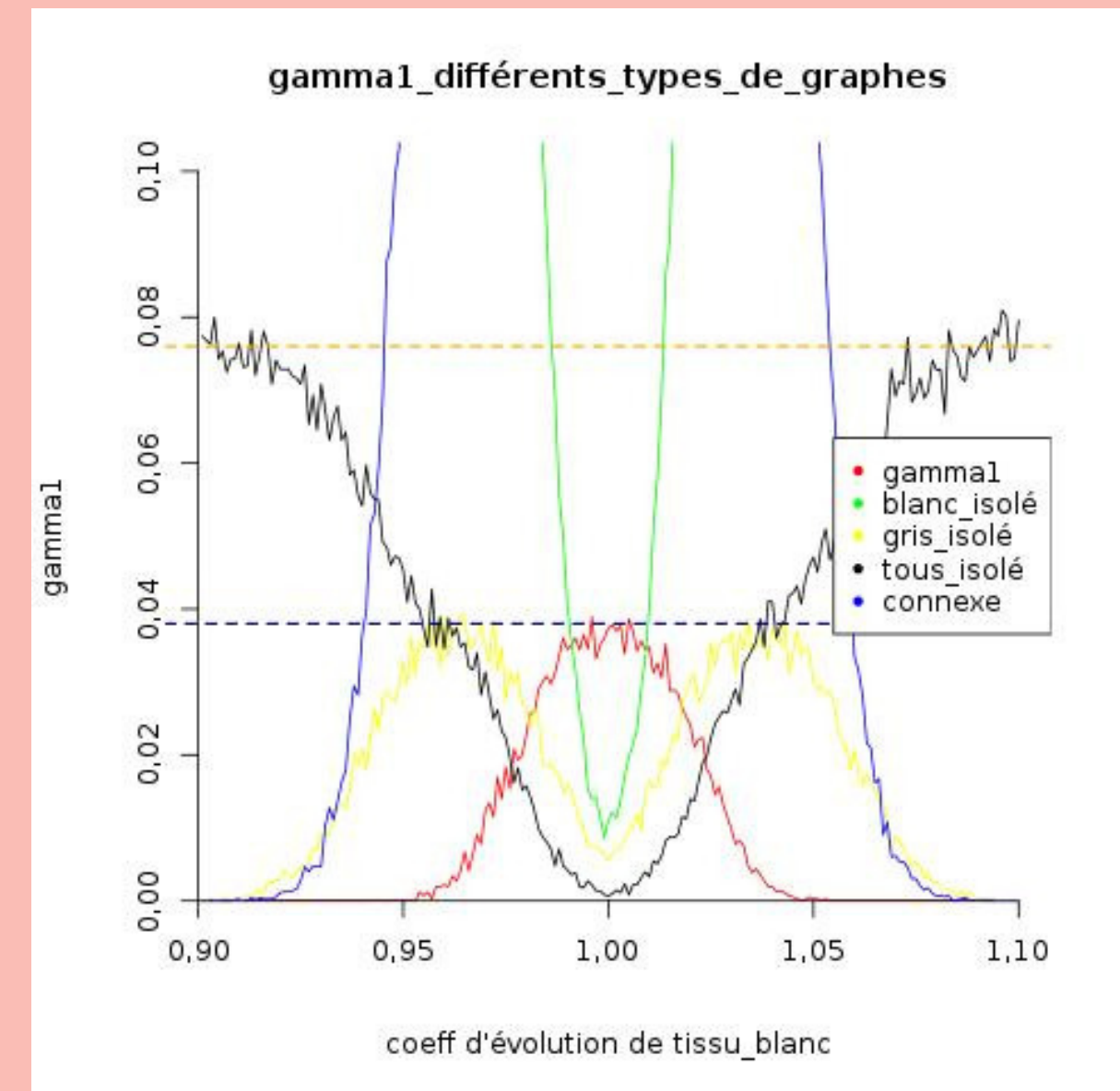
## Graphes zoomés



Graphe avec tissu\_gris  
isolé



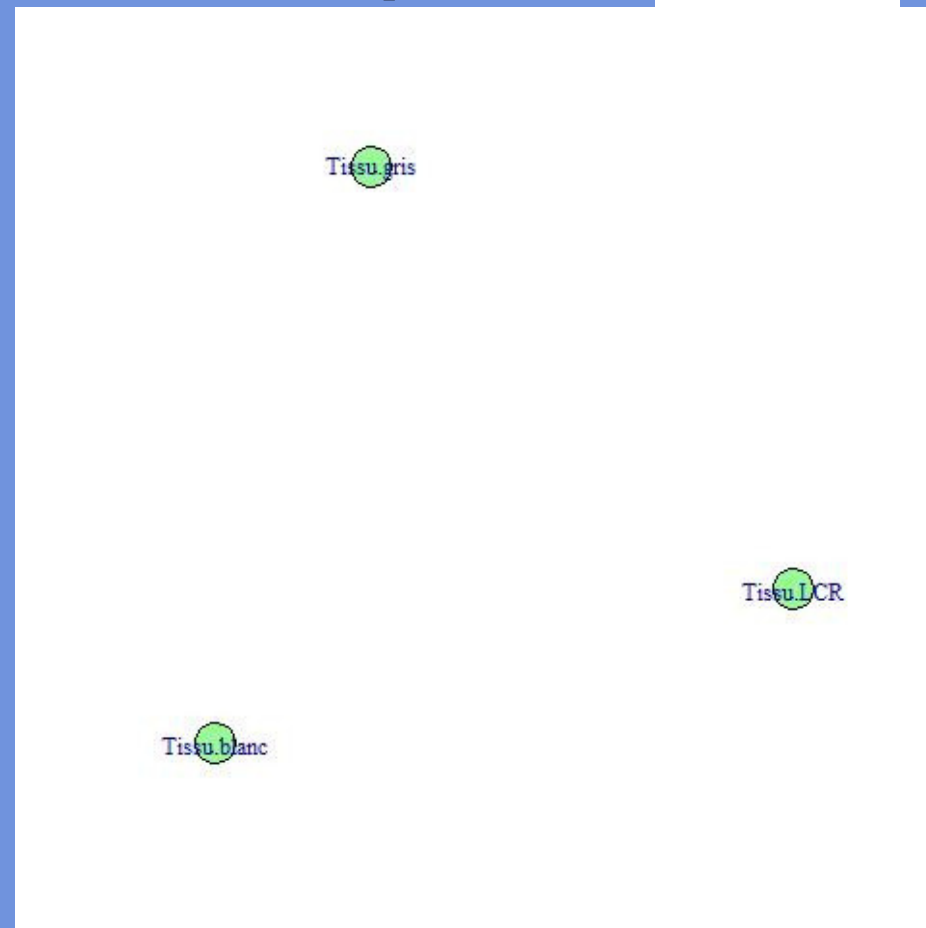
Graphe complètement  
disjoint



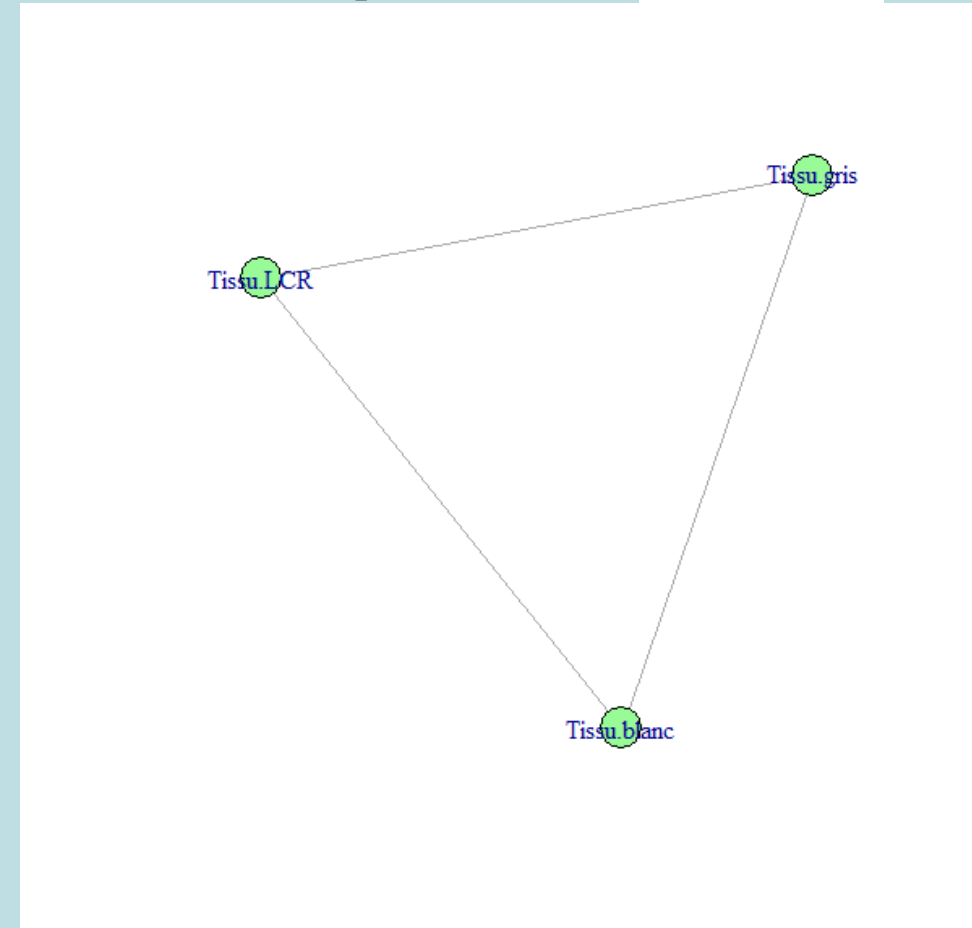
Graphe avec tissu\_blanc  
isolé

# Le graphe totalement disjoint

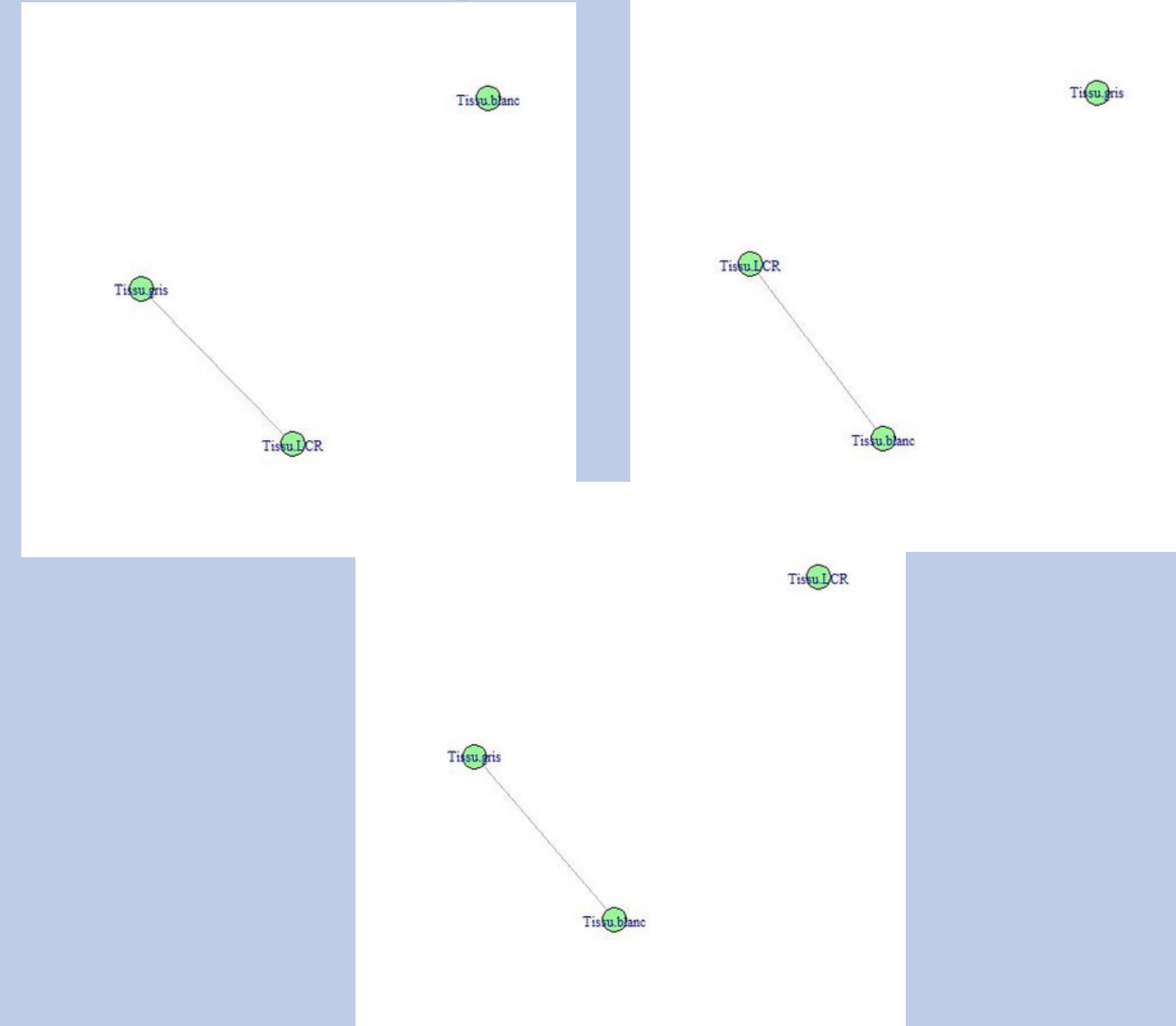
Graphe  $\mathcal{H}_1$



Graphe  $\mathcal{H}_0$



Graphes  $\mathcal{H}_2$



# Comparaison puissance et puissance corrigée

➡ Remise en question de la double erreur  
Puissance = puissance corrigée + simple erreur

**Objectif : Étudier l'évolution des puissances suivant les données de simulations**

➡ Simulations avec l'espérance de la loi ln-normale du groupe OD qui va varier (seulement sur un tissu à la fois)(à réaliser pour chacun des 3 tissus)

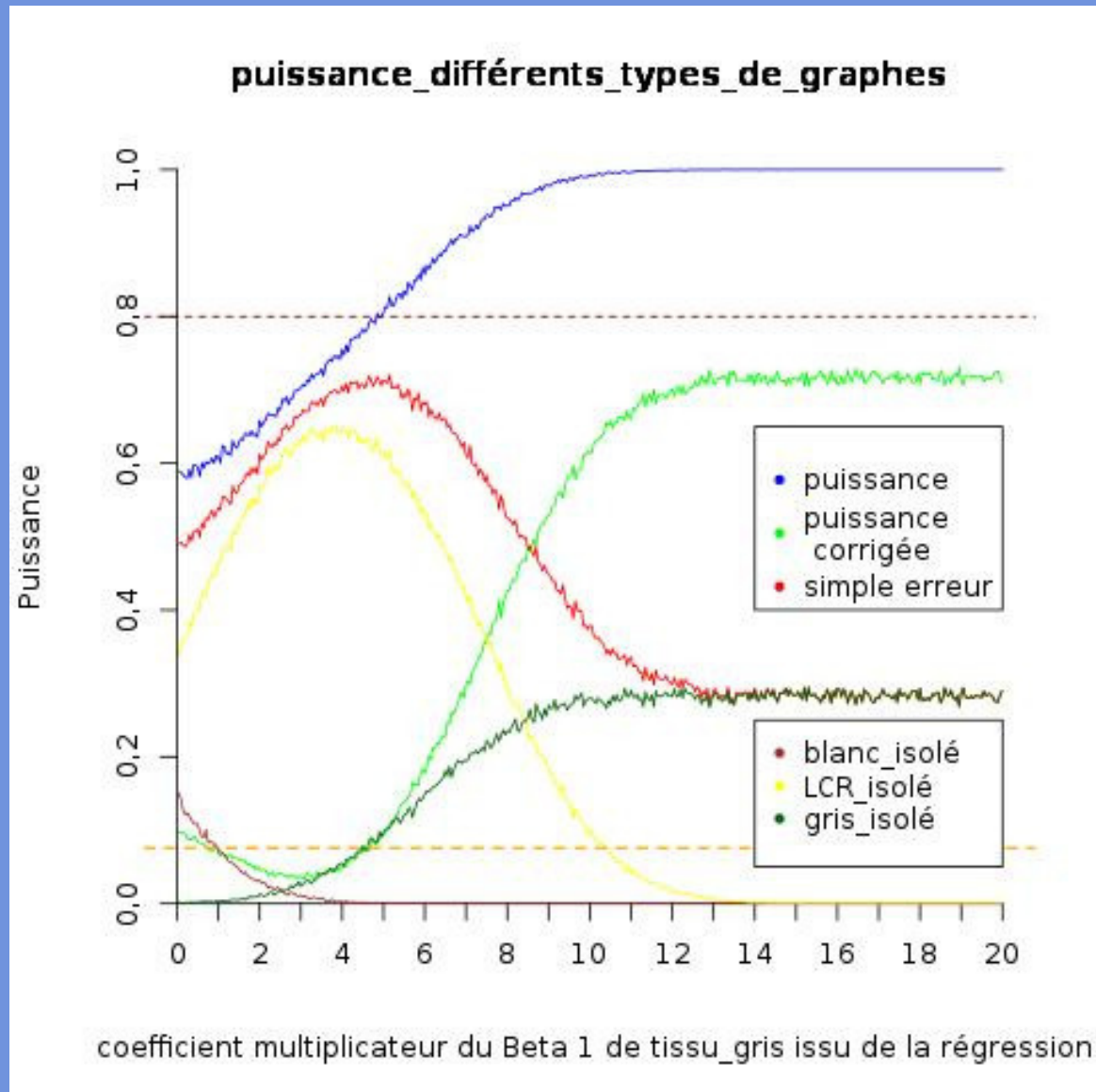
**Exemple :** l'espérance de la loi ln-normale pour Tissu\_LCR évolue pour le groupe OD (à l'aide d'un coefficient)

➡ Groupe contrôle :  
 $\hat{\beta}_0^{T.LCR}$

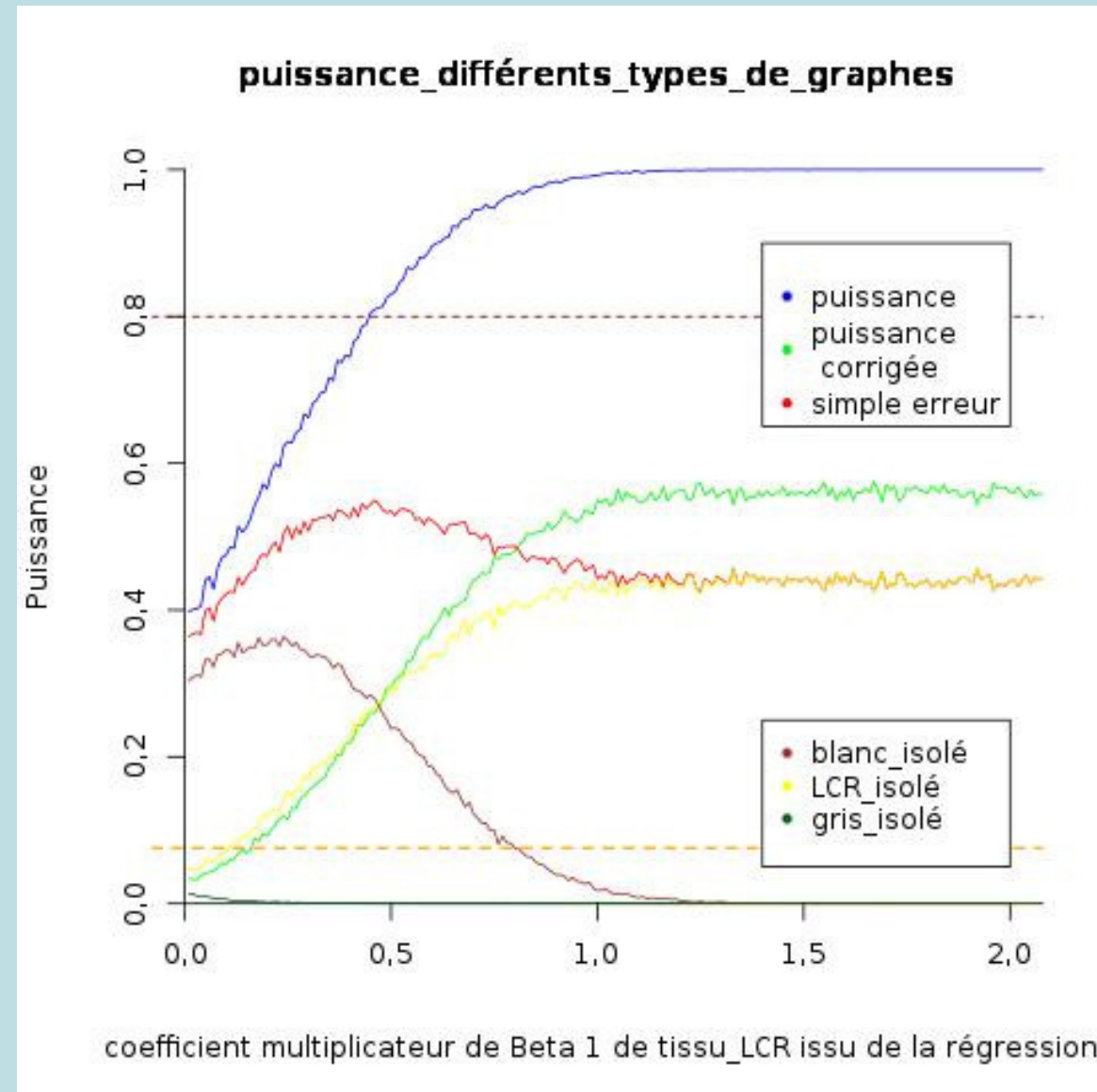
➡ Groupe OD:  
 $\hat{\beta}_0^{T.LCR} + C_{MR} * \hat{\beta}_1^{T.LCR}$



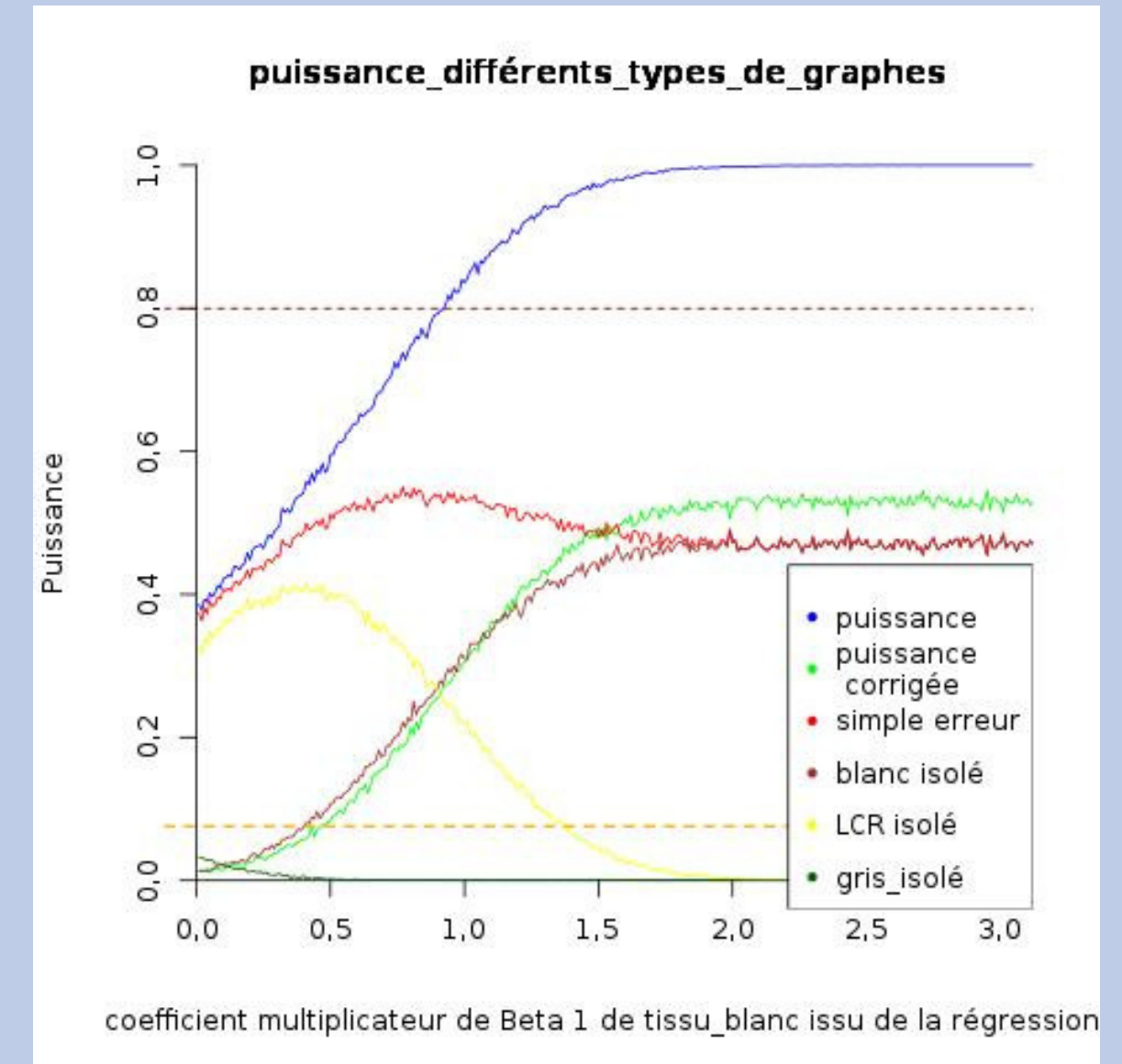
# Graphes puissances



Coefficient sur tissu\_gris  
isolé



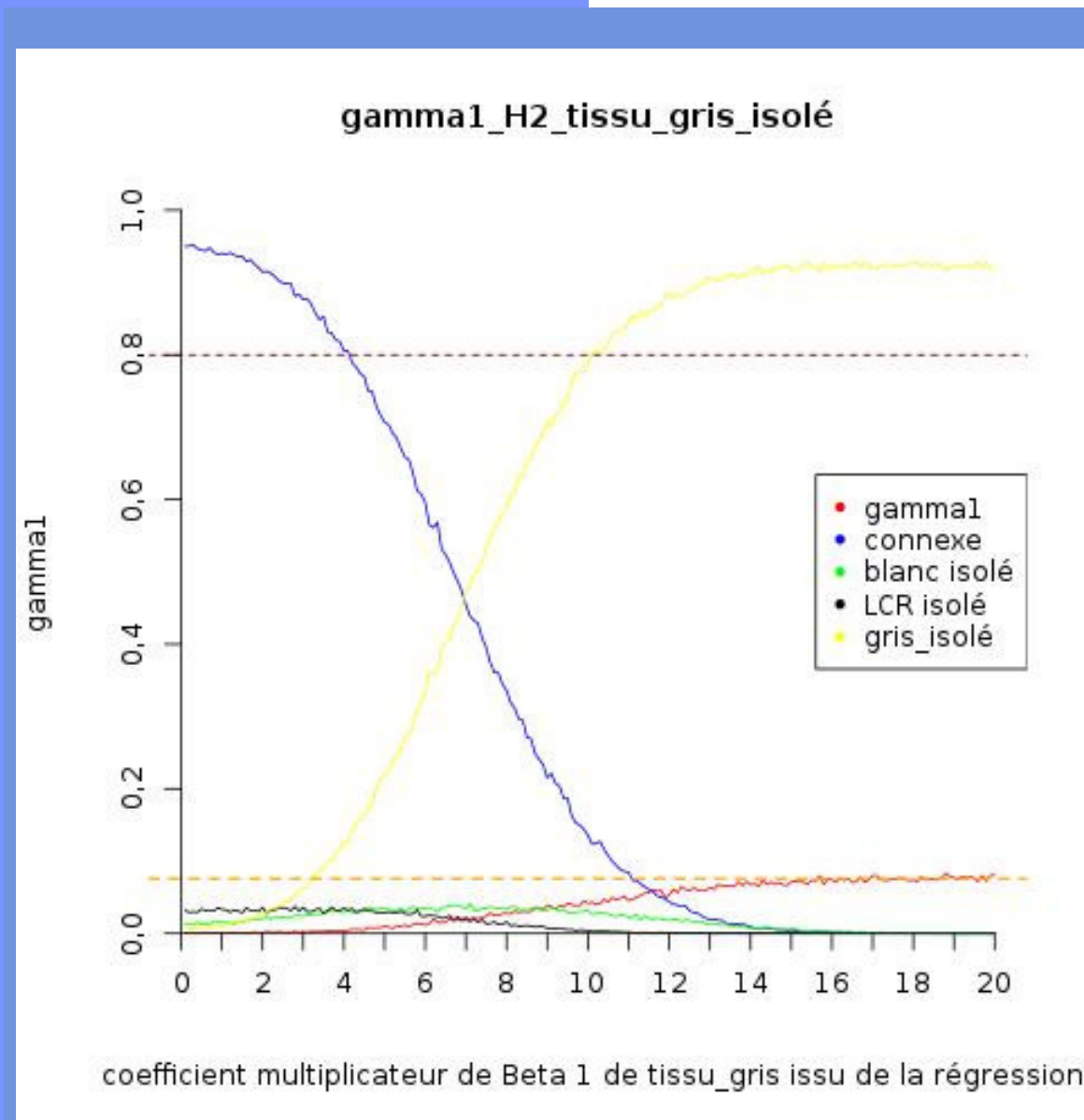
Coefficient sur tissu\_LCR  
isolé



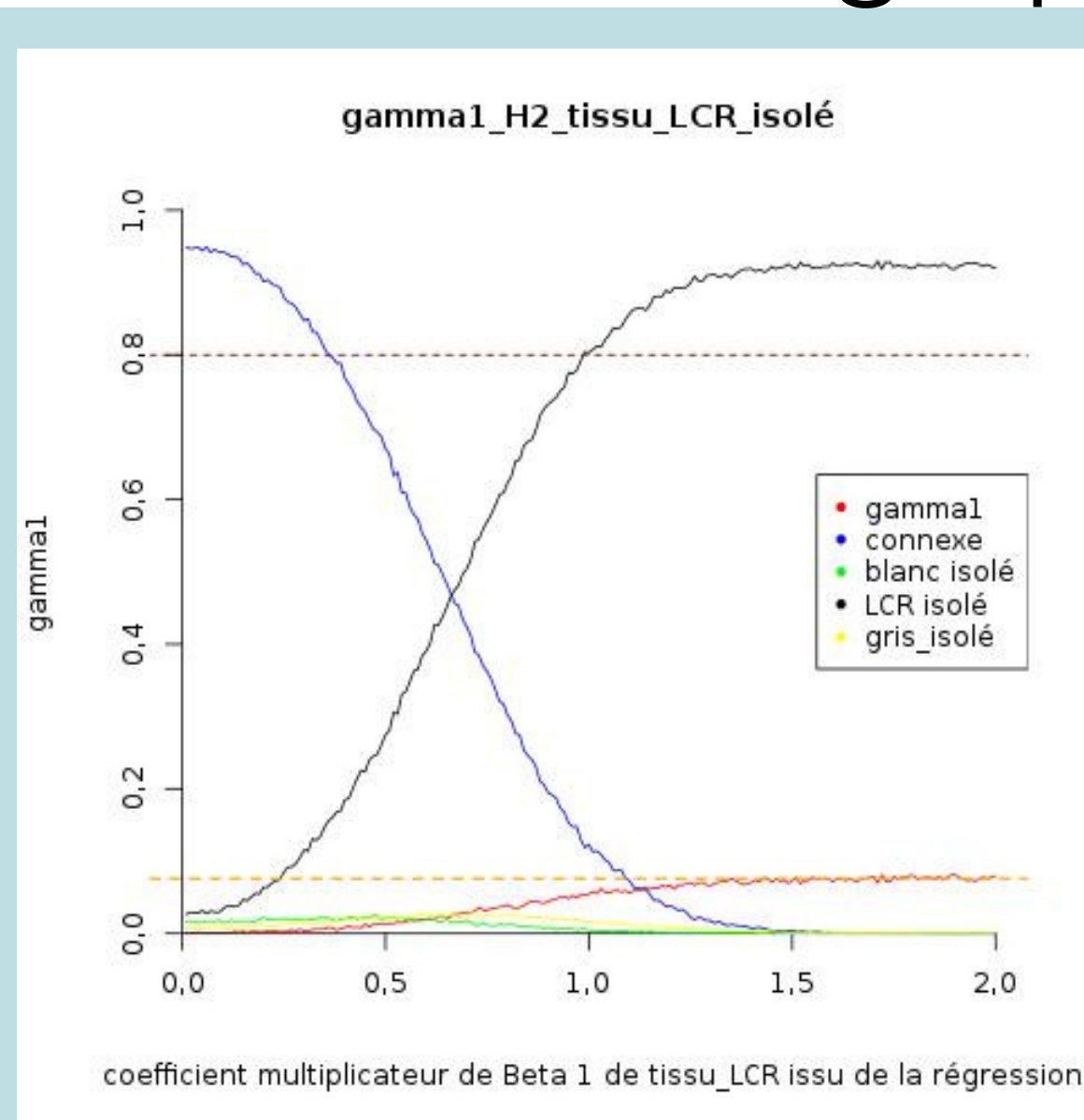
Coefficient sur tissu\_blanc  
isolé

# Risque de 3ème espèce

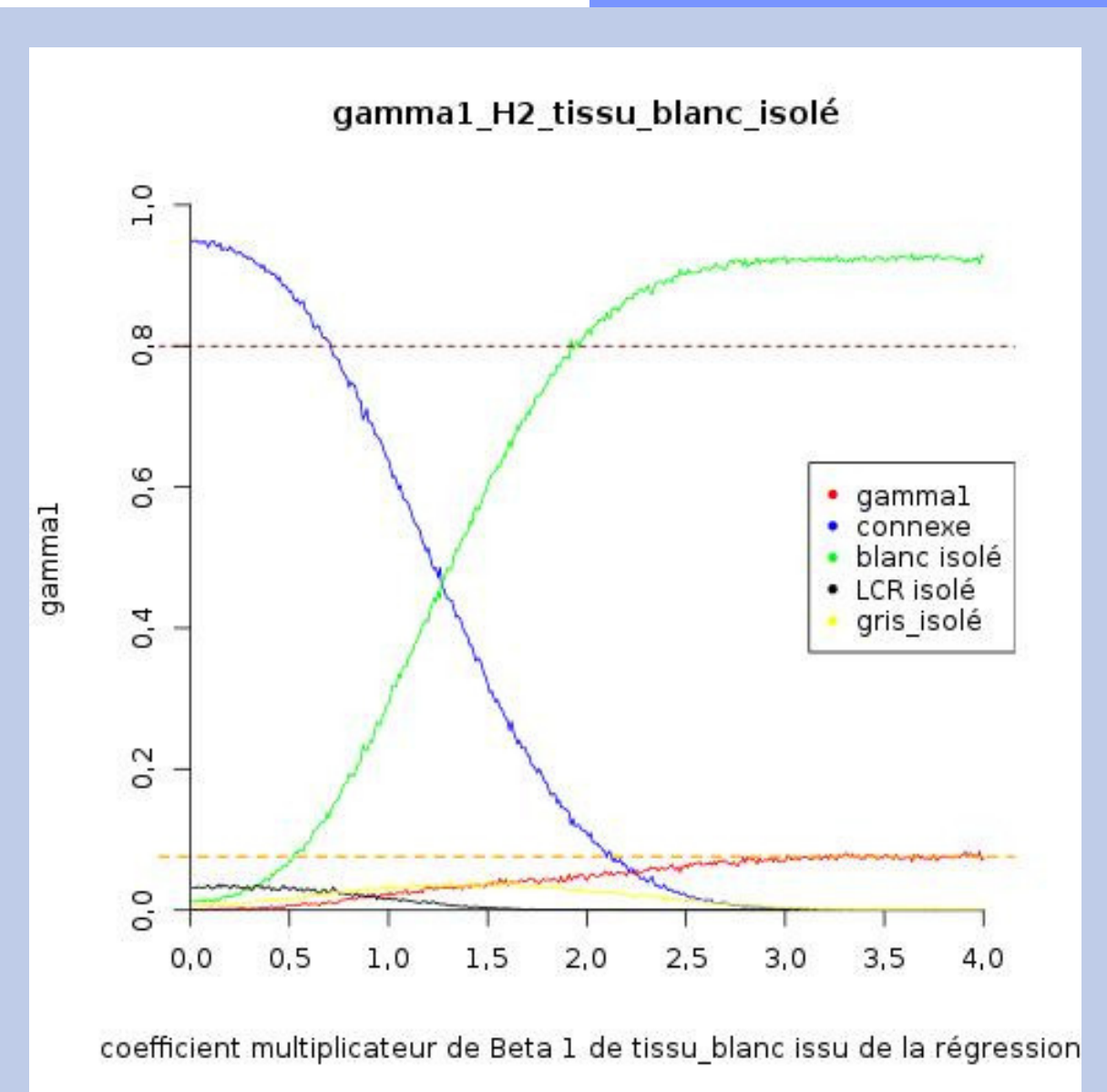
Simulation sous les 3 graphes de  $\mathcal{H}_2$



Graphe avec tissu\_gris  
isolé



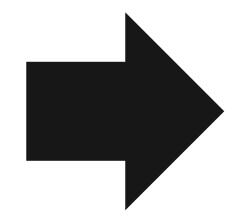
Graphe avec tissu\_LCR  
isolé



Graphe avec tissu\_blanc  
isolé

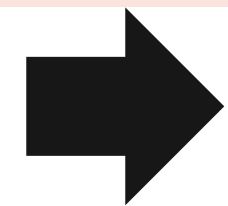
# Conclusion :

## Lien puissance/puissance corrigée :



Dépend du graphe  $\mathcal{H}_1$  et aussi des données choisies pour calibrer les simulations

## Risque de 3ème espèce :



Lié au seuil de coupure :

- borné par le seuil de coupure si le graphe de  $\mathcal{H}_1$  est totalement disjoint
- borné par la moitié du seuil de coupure si le graphe de  $\mathcal{H}_1$  où l'un des composant est isolé