

L'épuration professionnelle du monde du spectacle à la Libération

Karine Le Bail¹ & Julien Randon-Furling²

¹ CRAL CNRS / EHESS

² Chaire *Modélisations mathématiques & Sciences humaines et sociales*, Centre Borelli, ENS Paris-Saclay



école —————
normale —————
supérieure —————
paris-saclay ————

université
PARIS-SACLAY

Cours Master Humanités Numériques

ENC-PSL - Jeudi 12 octobre 2023

Plan

Introduction : pour une conversation interdisciplinaire

Projet CNRS Défi INterFaces Interdisciplinaires Numérique et Théorique :
L'épuration des arts de la scène en France après la Libération

Pour conclure : influences croisées

Introduction : pour une conversation interdisciplinaire

Projet CNRS Défi INterFaces Interdisciplinaires Numérique et Théorique :
L'épuration des arts de la scène en France après la Libération

Pour conclure : influences croisées

Les mathématiques à l'épreuve des sciences humaines et sociales

[Les sciences de l'Homme] semblent, au premier abord, les plus éloignées de toute notion de rigueur et de mesure [...] en raison du caractère essentiellement qualitatif de leur objet.

Claude Lévi-Strauss, *Les mathématiques de l'homme*, Esprit, 243-10, 1956

Les faits humains sont, par essence, des phénomènes très délicats, dont beaucoup échappent à la mesure mathématique. Pour bien les traduire, par suite pour bien les pénétrer (car comprend-on jamais parfaitement ce qu'on ne sait dire ?), une grande finesse de langage, une juste couleur dans le ton verbal sont nécessaires.

Là où calculer est impossible, suggérer s'impose.

Marc Bloch, *Apologie pour l'histoire* (1941)

The reprint started, as usual, with the Gaussian distribution and the statistician explained to [their] former classmate the meaning of the symbols (...). [The] classmate was a bit incredulous and was not quite sure whether the statistician was pulling [their] leg.(...) "And what is this symbol here?" "Oh," said the statistician, "this is pi. (...) The ratio of the circumference of the circle to its diameter." "Well, now you are pushing your joke too far," said the classmate, "surely the population has nothing to do with the circumference of the circle."

Eugene Wigner, *The Unreasonable Effectiveness of Mathematics in the Natural Sciences*, Courant Lecture, NYU (1959)

Pour une conversation interdisciplinaire

- ▷ Sortir des clichés :
 - ◇ sciences exactes ou dures vs sciences ... ?
 - ◇ approches quantitatives vs qualitatives i.e. non math.
 - non (cf. Lévi-Strauss & Weil)
- ▷ Certes coûteux (cf ex. plus loin : 2018-2023) mais indispensable
- ▷ Circulation des méthodes et des concepts (dans leur complexité)

Une démarche commune ?

- ▷ Là où calculer est *impossible*, suggérer et calculer s'imposent.
- ▷ Régularités statistiques (\rightarrow Lois – mais : bruit ? *null model* ?)
- ▷ Modélisation (risques : théorie *per se*, déconnexion du réel etc.
 - mais : c'est une école de l'intuition)

Pour une conversation interdisciplinaire

- ▷ Sortir des clichés :
 - ◇ sciences exactes ou dures vs sciences ... ?
 - ◇ approches quantitatives vs qualitatives i.e. non math.
 - non (cf. Lévi-Strauss & Weil)
- ▷ Certes coûteux (cf ex. plus loin : 2018-2023) mais indispensable
- ▷ Circulation des méthodes et des concepts (dans leur complexité)

Une démarche commune ?

- ▷ Là où calculer est *impossible*, suggérer et calculer s'imposent.
- ▷ Régularités statistiques (\rightarrow Lois – mais : bruit ? *null model* ?)
- ▷ Modélisation (risques : théorie *per se*, déconnexion du réel etc.
 - mais : c'est une école de l'intuition)

Pour une conversation interdisciplinaire

- ▷ Sortir des clichés :
 - ◇ sciences exactes ou dures vs sciences ... ?
 - ◇ approches quantitatives vs qualitatives i.e. non math.
 - non (cf. Lévi-Strauss & Weil)
- ▷ Certes coûteux (cf ex. plus loin : 2018-2023) mais indispensable
- ▷ Circulation des méthodes et des concepts (dans leur complexité)

Une démarche commune ?

- ▷ Là où calculer est *impossible*, suggérer et calculer s'imposent.
- ▷ Régularités statistiques (\Rightarrow Lois – mais : bruit ? *null model* ?)
- ▷ Modélisation (risques : théorie *per se*, déconnexion du réel etc.
 - mais : c'est une école de l'intuition)

Pour une conversation interdisciplinaire

- ▷ Sortir des clichés :
 - ◊ sciences exactes ou dures vs sciences ... ?
 - ◊ approches quantitatives vs qualitatives i.e. non math.
 - non (cf. Lévi-Strauss & Weil)
- ▷ Certes coûteux (cf ex. plus loin : 2018-2023) mais indispensable
- ▷ Circulation des méthodes et des concepts (dans leur complexité)

Une démarche commune ?

- ▷ *Là où calculer est impossible, suggérer et calculer s'imposent.*
- ▷ Régularités statistiques (\forall Lois – mais : bruit ? *null model* ?)
- ▷ Modélisation (risques : théorie *per se*, déconnexion du réel etc.
 - mais : c'est une école de l'intuition)

Un exemple de travail interdisciplinaire sur les discriminations dans l'Épuration

Introduction : pour une conversation interdisciplinaire

Projet CNRS Défi INterFaces Interdisciplinaires Numérique et Théorique :
L'épuration des arts de la scène en France après la Libération

Pour conclure : influences croisées

Étude de cas : l'épuration des arts de la scène en France après la Libération



Mireille Balin, une des vedettes de cinéma les plus en vues dans la France des années 1930
arrêtée en septembre 1944 avec son amant Birl Deissböck, officier de la Wehrmacht.

Tino Rossi

chantait aux galas collectifs 21^e Infanterie
chantait au club de jazz 308-309, av. de la
maison de la ville av. de la ville av. de la ville

gala L. F. 1^{er} mai 1942 - Infanterie

gala de la ville - 21^e mai 1942 - Infanterie
gala de la ville - 6 mai 1942 - Infanterie

11.43 - 5^e mai - 37^e mai

Cu de la ville - 11.43 - 5^e mai - 37^e mai
PTF de la ville - 11.43 - 5^e mai - 37^e mai
à Compiègne - 11.43 - 5^e mai - 37^e mai

102nd Dimanche à Compiègne après un jour de repos
lancement Tino Rossi (affaire des deux PTB, club de
Donsommes)



Compiègne - 11.43 - 5^e mai - 37^e mai
10: 11.43 - 5^e mai - 37^e mai

mardi 22 à 30 1/2

4/2

mercredi 24.11.45

à 15h

jeudi 25.11.45

11h

Modélisation des sources en données relationnelles

↷ individus - instances : faits reprochés, date de passage, décision/sanction

↷ individus - incriminations

03 BLANCHE	1	0	0	RPA21		0	EDS	12/1/1945
04 BLAREAU	2	0	0	RPA, GAL		0	BLA	11/7/1945
05 BLONDEAU	2	0	5	EDS		0	0m	2/6/1945
05 BLONDEAU	2	0	0	FPR		0	PDS	7/31/1945
05 BLONDEAU	2	0	0	FPR		0	EDS	4/12/1945
06 BLOT	0	0	5	ATT, RPA148		0	0m	10/29/1944
06 BLOT	0	0	5	RPA157, ATT		0	4m	1/19/1945
06 BLOT	0	0	13	RPA187, ATT		0	4m	1/22/1945

~ 600 individus, 11 instances → ~ 850 observations

Processus, traces, représentations : des archives aux données

Modélisation des sources en données relationnelles

↔ individus - instances : faits reprochés, date de passage, décision/sanction

↔ individus - incriminations

...visualisables sous forme de graphes bipartites



Épuration professionnelle (1944-1949) : 358 individus, 4 instances, 24 incriminations

▷ Radio-Paris : 165 indiv.; Galas : 54 (dont 34 ont une autre incrim., dont 20 RP,...)

La question des discriminations : complexité et nécessaire formalisation

- ▷ multiples "procès en injustice" intentés au processus de l'épuration des artistes
- ▷ mais : jamais jusqu'ici d'instruction rigoureuse, d'examen systématique des dossiers individuels, des instances, des jugements et des peines :
 - Un même ensemble d'incriminations a-t-il toujours conduit à une même peine ?
 - Constate-t-on des peines plus sévères à l'encontre de certains groupes ?
 - Le cas échéant, les différences de peines apparaissent-elles justifiées ou, au contraire, semblent-elles correspondre à des biais dans les jugements ?

La question des discriminations : complexité et nécessaire formalisation

- ▷ Question sous-jacente, plus générale et fondamentale, de la discrimination d'individus ou de groupes d'individus dans un processus de justice ou de sélection
- ▷ Objet d'une abondante littérature dans plusieurs disciplines – en sciences sociales, en économie, en droit ou encore en sciences politiques
- ▷ D'aucuns considèrent comme trivial voire inutile le recours à des statistiques dès lors que leur conviction est forgée :
 - Que diraient-ils d'une même position épistémologique en sciences de la nature ? N'exigeraient-ils pas de solides modes d'administration de la preuve ? Par exemple le recours à des essais randomisés pour établir l'efficacité d'une molécule dans le traitement d'une maladie ?
 - Si les essais randomisés sont peu envisageables en histoire, il n'en est pas de même de l'analyse statistique : bien plus qu'une science des données au sens numérique, elle est justement la science des inférences rigoureuses à partir des observations, que celles-ci soient qualitatives ou quantitatives.

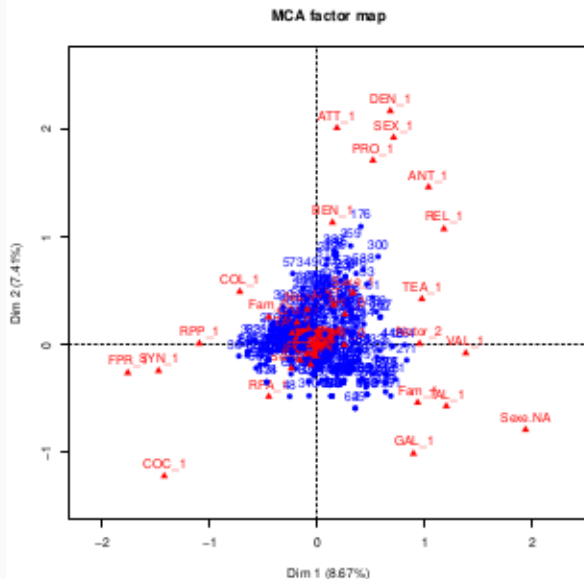
- ▷ Comparaison de sous-échantillons – Tests
- ▷ Régression généralisée – avec critères d'information et inférence multi-modèles
- ▷ Réduction de dimensionnalité – Analyse des correspondances multiples
- ▷ Réseaux complexes, réseau bayésien & inférence algorithmique

De la significativité : artefacts, valeur p et valeur s

- ▶ **Artefacts aléatoires et significativité** : les différences observées entre les peines moyennes peuvent relever du **hasard de l'échantillonnage**.
Penser par exemple : à pile ou face, 3 "pile" d'affilée peuvent se produire par hasard alors que la pièce n'est pas faussée (cela se produit 1 fois sur 8 en moyenne).
- ▶ Calculer la **probabilité d'une série d'observation sous l'hypothèse neutre** : cette probabilité est la valeur p
- ▶ Débat autour des **mésusages de la valeur p** :
R. Nuzzo, *Scientific Method : Statistical Errors*, Nature 506, 2014;
R.L. Wasserstein, N.A. Lazar, *ASA Statement on Statistical Significance and p -Values*, The American Statistician 70, 2016;
S. Greenland, *Valid p -values behave exactly as they should : some misleading criticisms of p -values and their resolution with s -values*, The American Statistician 73, 2019.
- ▶ **Dispositif cognitif simple** : valeur s = équivalent en nombre de pile ou face

Analyse des correspondances multiples

- ▶ Deux premières composantes principales du nuage ($\sim 15\%$ de sa dispersion) :

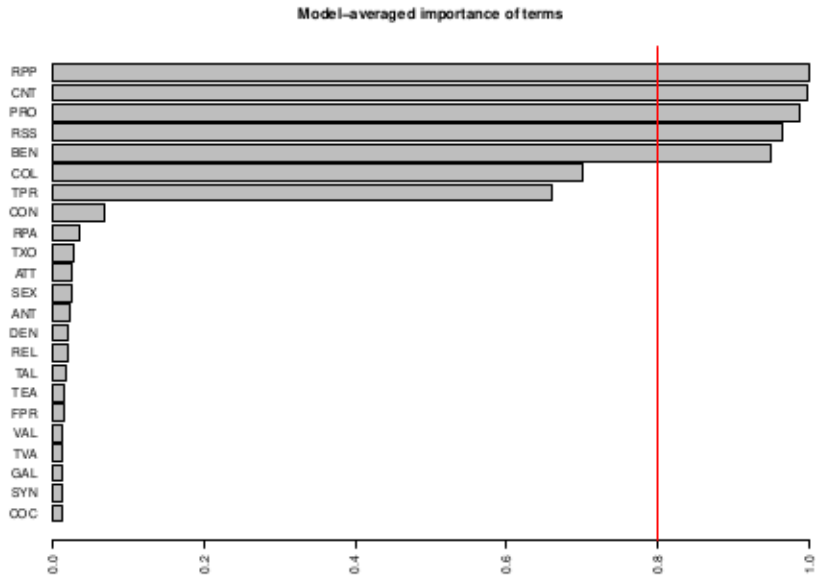


- ▷ Régression généralisée – avec critère d'information (AIC/BIC)

1 seul modèle = risque...car potentiellement hautement biaisé à cause d'artefacts

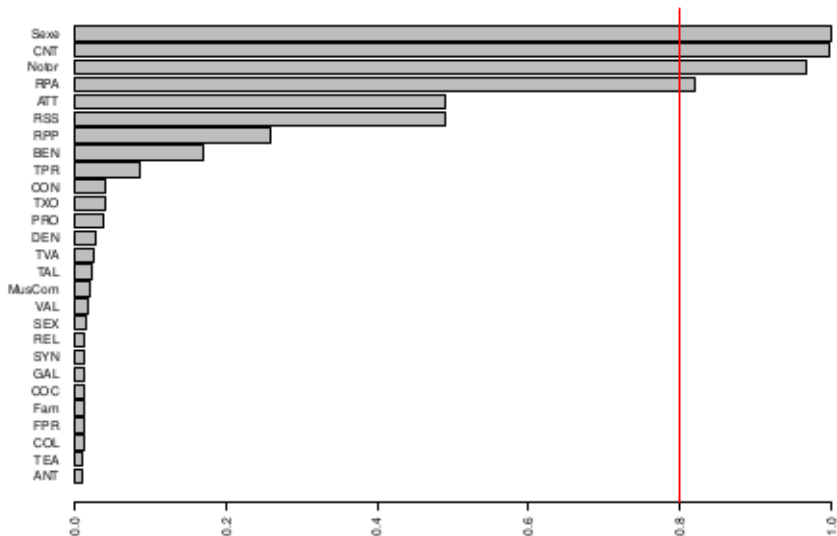
↪ Inférence multi-modèles pour sélectionner & hiérarchiser des facteurs à partir des
5, 10 ou 100 modèles qui ont le plus de probabilité d'être le meilleur modèle.

Sélection de facteurs & inférence multi-modèles



Sélection de facteurs & inférence multi-modèles

Model-averaged importance of terms



Sélection de facteurs & inférence multi-modèles

▷ Régression généralisée – avec critère d'information (AIC/BIC)

1 seul modèle = risque, car potentiellement hautement biaisé à cause d'artefacts

↪ Inférence multi-modèles pour sélectionner & hiérarchiser des facteurs à partir des
5, 10 ou 100 modèles qui ont le plus de probabilité d'être le meilleur modèle.

↪ Le facteur "Notoriété" : comparés aux artistes anonymes (notoriété 0),
interdictions plus courtes de :

1.5 mois pour les artistes de notoriété 1,

3 mois pour les vedettes (artistes de notoriété 2)

(valeur $s = 9$)

Sélection de facteurs & inférence multi-modèles

▷ Régression généralisée – avec critère d'information (AIC/BIC)

1 seul modèle = risque, car potentiellement hautement biaisé à cause d'artefacts

↪ Inférence multi-modèles pour sélectionner & hiérarchiser des facteurs à partir des
5, 10 ou 100 modèles qui ont le plus de probabilité d'être le meilleur modèle.

↪ Le facteur "Notoriété" : comparés aux artistes anonymes (notoriété 0),
interdictions plus courtes de :

1.5 mois pour les artistes de notoriété 1,

3 mois pour les vedettes (artistes de notoriété 2)

(valeur $s = 9$)

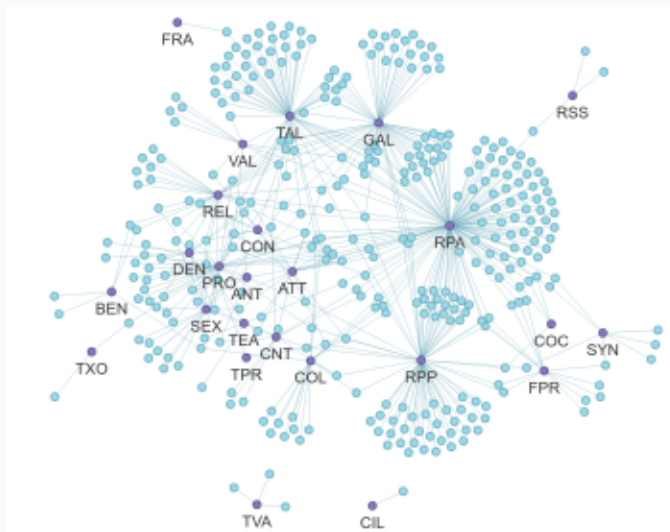
↪ le facteur "Genre" : apparaît dans tous les meilleurs modèles

mais parfois avec un coeff. positif,
parfois avec un coeff. négatif ...

et (valeur $s < 1$)...

Analyse de réseau complexe

Réseau biparti : Individus ↔ Charges



- ▷ “Ségrégation” des groupes sur le réseau

J. R-F, M. Ohtsuna, A. Lucquinard, *From urban segregation to spatial structure detection*, EPB 47(4) (2017)

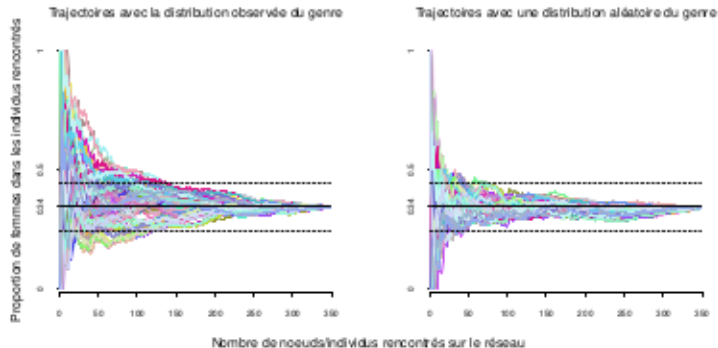
M. Ohtsuna, J. R-F, W. Clark, *Segregation through the multiscalar lens*, PNAS 116(25) (2019)

A. Bassolas, V. Nicchia, (...) *Structural correlations and heterogeneity in complex systems*, Communications Physics, 4(1) (2021)

- ▷ Idée simple :
temps moyen mis par une marche aléatoire
pour converger vers la distribution globale, selon son groupe de départ

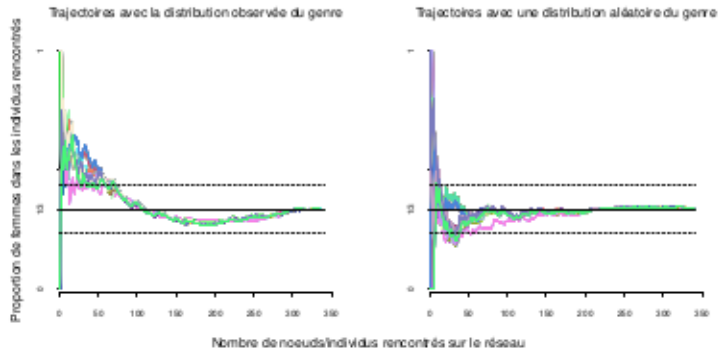
Mesure de différenciation par convergence de marches aléatoires

- ▶ Comparaison avec réseau aléatoire (permutation des individus \rightarrow *null model*)



Mesure de différenciation par convergence de marches aléatoires

- ▶ Comparaison avec réseau aléatoire (permutation des individus \rightarrow *null model*)



Mesure d'hétérogénéité par marches aléatoires

- ▷ $X(s)$ = fraction des nœuds d'un groupes visités sur le graphe *d'origine*, après $2s$ pas

Mesure d'hétérogénéité par marches aléatoires

- ▷ $X(s)$ = fraction des nœuds d'un groupes visités sur le graphe *d'origine*, après $2s$ pas
- ▷ Temps de convergence au niveau $\delta > 0$, sachant une proportion globale μ :

$$\tau_\delta = \inf \{ 1 \leq t \leq N \text{ t.q. } |X(s) - \mu| \leq \delta \text{ pour tout } s \geq t \}$$

Mesure d'hétérogénéité par marches aléatoires

- ▷ $X(s)$ = fraction des nœuds d'un groupes visités sur le graphe *d'origine*, après $2s$ pas
- ▷ Temps de convergence au niveau $\delta > 0$, sachant une proportion globale μ :

$$\tau_\delta = \inf \{ 1 \leq t \leq N \text{ t.q. } |X(s) - \mu| \leq \delta \text{ pour tout } s \geq t \}$$

- ▷ Temps de convergence moyen à $\delta > 0$ en partant du gpe G_i :

$$r_i = \mathbb{E}[\tau_\delta | X(0) \in G_i]$$

Mesure d'hétérogénéité par marches aléatoires

- ▷ $X(s)$ = fraction des nœuds d'un groupes visités sur le graphe *d'origine*, après $2s$ pas
- ▷ Temps de convergence au niveau $\delta > 0$, sachant une proportion globale μ :

$$\tau_\delta = \inf \{ 1 \leq t \leq N \text{ t.q. } |X(s) - \mu| \leq \delta \text{ pour tout } s \geq t \}$$

- ▷ Temps de convergence moyen à $\delta > 0$ en partant du gpe G_i :

$$r_i = \mathbb{E}[\tau_\delta | X(0) \in G_i]$$

- ▷ $Z(s)$ une marche aléatoire sur le réseau randomisé

Mesure d'hétérogénéité par marches aléatoires

- ▷ $X(s)$ = fraction des nœuds d'un groupes visités sur le graphe *d'origine*, après $2s$ pas
- ▷ Temps de convergence au niveau $\delta > 0$, sachant une proportion globale μ :

$$\tau_\delta = \inf \{ 1 \leq t \leq N \text{ t.q. } |X(s) - \mu| \leq \delta \text{ pour tout } s \geq t \}$$

- ▷ Temps de convergence moyen à $\delta > 0$ en partant du gpe G_i :

$$r_i = \mathbb{E}[\tau_\delta | X(0) \in G_i]$$

- ▷ $Z(s)$ une marche aléatoire sur le réseau randomisé
- ▷ Distance au *null model* :

$$d(s) = \mathbb{P}(|X(s) - \mu| > |Z(s) - \mu|) = \mathbb{E}[\mathbb{I}_{\{|X(s) - \mu| > |Z(s) - \mu|\}}]$$

- ▷ Distance moyenne au *null model* :

$$g(s) = \mathbb{E}[d(s)]$$

Mesure d'hétérogénéité par marches aléatoires

- ▷ $X(s)$ = fraction des nœuds d'un groupes visités sur le graphe *d'origine*, après $2s$ pas
- ▷ Temps de convergence au niveau $\delta > 0$, sachant une proportion globale μ :

$$\tau_\delta = \inf \{ 1 \leq t \leq N \text{ t.q. } |X(s) - \mu| \leq \delta \text{ pour tout } s \geq t \}$$

- ▷ Temps de convergence moyen à $\delta > 0$ en partant du gpe G_i :

$$r_i = \mathbb{E}[\tau_\delta | X(0) \in G_i]$$

- ▷ $Z(s)$ une marche aléatoire sur le réseau randomisé
- ▷ Distance au *null model* :

$$d(s) = \mathbb{P}(|X(s) - \mu| > |Z(s) - \mu|) = \mathbb{E}[\mathbb{I}_{\{|X(s) - \mu| > |Z(s) - \mu|\}}]$$

- ▷ Distance moyenne au *null model* :

$$g(s) = \mathbb{E}[d(s)]$$

- ▷ **Version multi-groupes** : remplacer $|X(s) - \mu|$ par entropie relative / divergence K-L :

$$d(p|q) = \sum_i p_i \ln \frac{p_i}{q_i}$$

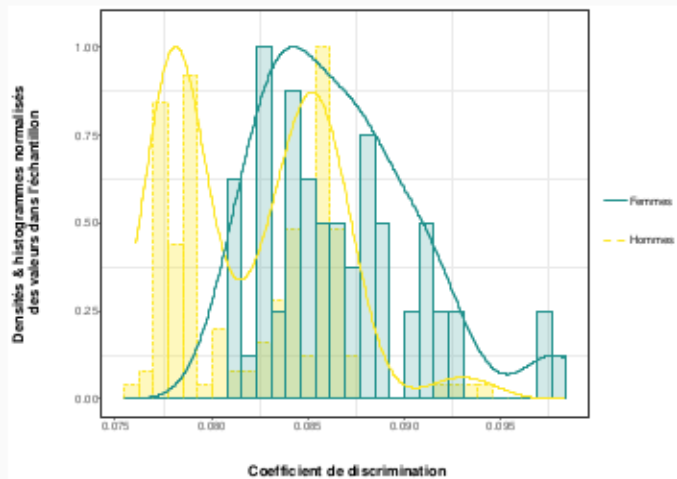
- ▷ Point de départ : graphe complet orienté entre la totalité des modalités des variables
- ▷ Décimation des arêtes selon 3 règles :
 - ▷ **temporalité des variables** :
eg un genre peut conduire à une peine, mais pas l'inverse;
 - ▷ **probabilité croissante** :
A conduit à B si la probabilité de B dans le sous-échantillon défini par A est plus grande que la probabilité de B dans le sous-échantillon restant;
 - ▷ **maximisation de la vraisemblance** du graphe étant donné les observations recueillies dans notre corpus;
- ▷ Résultat = **inférence de structure "causale"**

Modèle graphique : réseau de Suppes-Bayes appris sur les données



Scores et coefficients de discrimination calculés via PageRank (Google)

- ▷ Coefficient de discrimination (scores des noeuds des peines) en fonction du genre



- ▷ L'inférence multi-modèles révèle la *clémence vis-à-vis des vedettes*
- ▷ La "ségrégation" dans le réseau des incriminations révèle
le recours à des combinaisons spécifiques d'incriminations contre les femmes
- ▷ L'apprentissage d'un réseau causal de Suppes-Bayes révèle
la discrimination contre les femmes dans l'ensemble du processus

- ▷ Il s'agit des premiers résultats rigoureusement établis à partir de données sur cette question

Introduction : pour une conversation interdisciplinaire

Projet CNRS Défi INterFaces Interdisciplinaires Numérique et Théorique :
L'épuration des arts de la scène en France après la Libération

Pour conclure : influences croisées

Pour conclure : influences croisées

Il ne s'agit pas seulement, ni même surtout, d'emprunter en bloc aux mathématiques des méthodes et des résultats achevés. Les besoins propres aux sciences sociales, les caractères originaux de leurs objets imposent aux mathématiciens un effort spécial d'adaptation.

La collaboration ne saurait être à sens unique. D'un côté, les mathématiques contribueront au progrès des sciences sociales, mais, de l'autre, les exigences propres à ces dernières ouvriront aux mathématiques des perspectives supplémentaires.

En ce sens, il s'agit donc de mathématiques nouvelles à créer.

Claude Lévi-Strauss, *Les mathématiques de l'homme*, Esprit, 243-10, 1956

Pour conclure : influences croisées

Il ne s'agit pas seulement, ni même surtout, d'emprunter en bloc aux mathématiques des méthodes et des résultats achevés. Les besoins propres aux sciences sociales, les caractères originaux de leurs objets imposent aux mathématiciens un effort spécial d'adaptation.

La collaboration ne saurait être à sens unique. D'un côté, les mathématiques contribueront au progrès des sciences sociales, mais, de l'autre, les exigences propres à ces dernières ouvriront aux mathématiques des perspectives supplémentaires.

En ce sens, il s'agit donc de mathématiques nouvelles à créer.

Claude Lévi-Strauss, *Les mathématiques de l'homme*, Esprit, 243-10, 1956

Merci pour votre attention.