Random Matrix Theory – TP 2

Kayané Robach[1] Théo Dumont[1,2]
[1]ENS Paris-Saclay [2]Mines Paris

2021–2022

Observations préliminaires

1. On écrit $\frac{1}{\sqrt{n}}A$ comme la somme de son espérance conditionnelle par rapport aux q_i, notée $\bar{A} := \mathbb{E}(A \mid q)$, et d'un terme de bruit:

$$\frac{1}{\sqrt{n}}A = \frac{1}{\sqrt{n}}\bar{A} + \frac{1}{\sqrt{n}}\underbrace{(A - \bar{A})}_{:=X}.$$

Notons $J \in \mathbb{R}^{n\times K}$ la matrice des vecteurs canoniques j_i des classes $\mathcal{C}_i$, et $C = (C_{ab})_{a,b=1}^K \in \mathbb{R}^{K\times K}$. On peut alors exprimer facilement l'espérance de A, dont chaque coefficient suit une loi binomiale :

$$\frac{1}{\sqrt{n}}\bar{A} = \frac{1}{\sqrt{n}}\bar{A} = \left[\frac{1}{\sqrt{n}}q_iq_jC_{a_ia_j}\right]_{i,j=1}^n = \frac{1}{\sqrt{n}}\text{diag}\,(q)\,J\underbrace{C}_{K\times K}J^*\text{diag}\,(q)^*,$$

ce qui prouve que $\frac{1}{\sqrt{n}}\bar{A}$ est de rang au plus K. De plus, on a bien que $\frac{1}{\sqrt{n}}X$ est de moyenne nulle et à entrées indépendantes car par hypothèse les q_i sont i.i.d. issus d'une loi indépendante des M_{ab} (donc des C_{ab}). Evaluons maintenant la variance de $\frac{1}{\sqrt{n}}X$:

$$\begin{aligned}
\mathbb{V}_q\frac{1}{\sqrt{n}}X &= \frac{1}{n}\mathbb{V}_qA \\
&= \left[\frac{1}{n}q_iq_jC_{a_ia_j}(1 - q_iq_jC_{a_ia_j})\right]_{i,j=1}^n \qquad \text{car } A_{ij} \sim \mathcal{B}(q_iq_jC_{a_ia_j}) \\
\left[\mathbb{V}_q\frac{1}{\sqrt{n}}X\right]_{i,j} &= \frac{1}{n}q_iq_j\left(1 + \frac{M_{a(i)a(j)}}{\sqrt{n}}\right)\left(1 - q_iq_j\left(1 + \frac{M_{a(i)a(j)}}{\sqrt{n}}\right)\right) \\
&= \frac{q_iq_j(1-q_iq_j)}{n} + \frac{q_iq_j(1-2q_iq_j)M_{a(i)a(j)}}{n^{3/2}} - \frac{q_i^2q_j^2M_{a(i)a(j)}^2}{n^2}
\end{aligned}$$

2. On peut écrire la matrice $\frac{1}{\sqrt{n}}B$ comme

$$\frac{1}{\sqrt{n}}B = \frac{1}{\sqrt{n}}(A - qq^*) = \frac{1}{\sqrt{n}}\underbrace{(\bar{A} - qq^*)}_{:=P} + \frac{1}{\sqrt{n}}X,$$

et on peut remarquer que

$$\begin{aligned}
P &= \text{diag}\,(q)\,(JCJ^* - I_n)\text{diag}\,(q)^* \\
&= \text{diag}\,(q)\,J(C - I_K)J^*\text{diag}\,(q)^* \\
&= \frac{1}{\sqrt{n}}\text{diag}\,(q)\,JMJ^*\text{diag}\,(q)^*
\end{aligned}$$

ce qui montre que P est de rang au plus K. Par ailleurs, les résultats sur le terme de bruit de la question **1.** restent valable, ce terme étant le même pour $\frac{1}{\sqrt{n}}B$ que pour $\frac{1}{\sqrt{n}}A$.

3. On écrit le code suivant, dont les résultats sont reproduits en figure 1 :

[1]:
```python
import numpy as np
import matplotlib.pyplot as plt
```

[2]:
```python
def get_eigenvalues(M0, q_method, q0, sigma, q1, q2, n, K):
    """Get eigenvalues of one matrix B/sqrt(n)"""
    # variables ==================================
    # M = M0 * np.ones((K,K))
    M = M0 * np.eye(K)
    C = 1 + M/np.sqrt(n)
    classes = np.random.randint(low=0, high=K, size=n)
    J = np.eye(K)[classes]
    # expression of q
    if q_method == "constant":
        q = q0 * np.ones(n)
    elif q_method == "uniforme":
        assert sigma is not None
        q = np.random.uniform(low=q0-sigma, high=q0+sigma, size=n)
    elif q_method == "choisi":
        assert q1 and q2 is not None
        q = np.random.choice([q1,q2], size=n)

    p_bernoulli = np.diag(q) @ J @ C @ J.T @ np.diag(q)
    A = np.random.binomial(n=1, p=p_bernoulli)
    # symmetrize A
    for i in range(A.shape[0]):
        for j in range(i):
            A[j,i] = A[i,j]
    B = A - np.outer(q,q)
    # compute eigenvalues ==================================
    eig_vals, eig_vec = np.linalg.eig(B/np.sqrt(n))
    eig_vec = eig_vec.T
    sorted_idx = np.argsort(eig_vals)
    return eig_vals, eig_vec[sorted_idx][-3:], classes


def experiment(M0_list, q_method, q0=None, sigma=None, q1=None, q2=None, n=1000, K=3,␣
↪show_suptitle=True, show_title=True, color="tab:blue"):
    """Perform an experiment for each value of M0 in M0_list"""

    plt.figure(figsize=(13,1.7+.2*int(show_title)+.3*int(show_suptitle)), tight_layout=True)
    eig_vec_list = []
    # plot eigenvalues
    for idx, M0 in enumerate(M0_list):
        plt.subplot(1, len(M0_list), idx+1)
        eig_vals, eig_vec, classes = get_eigenvalues(M0=M0, q_method=q_method, q0=q0,␣
↪sigma=sigma, q1=q1, q2=q2, n=n, K=K)
        if q_method == "constant":
            sigma = np.sqrt(q0**2 * (1-q0**2))
            plt.axvspan(-2*sigma, 2*sigma, alpha=0.1, color='red')
            nb_isolated = np.sum(eig_vals > 2*sigma+1e-2)
        else:
            nb_isolated = np.sum(eig_vals > 1)
        plt.hist(eig_vals, color=color, bins=100, alpha=.7)
        plt.title(show_title * f"$M_0={M0}$\n" + f"{nb_isolated} vp isolées")
        plt.grid(ls=':')
        plt.gca().spines['right'].set_visible(False)
        plt.gca().spines['left'].set_visible(False)
        plt.gca().spines['top'].set_visible(False)
```

```python
        plt.yticks([])
        plt.xlim((-1, 3))
        eig_vec_list.append(eig_vec)

    if show_suptitle:
        if q_method == "constant":
            suptitle = f"$q_i=q_0={q0}$ {q_method}"
        elif q_method == "uniforme":
            suptitle = f"$q_i$ {q_method} dans $[q_0-\sigma,q_0+\sigma]=[{q0-sigma},{q0+sigma}]$"
        elif q_method == "choisi":
            suptitle = r"$q_i$ choisi uniformément dans $\{q^{(1)},q^{(2)}\}=\{$" +␣
↪f"${q1},{q2}$" + r"$\}$"
        plt.suptitle(suptitle)
    plt.show()

    # vecteurs propres
    plt.figure(figsize=(18,8))
    for idx in range(1,4):
        plt.subplot(3, 1, idx)
        sorted_eigenvector = []
        for k in range(K):
            for i in range(len(np.array(eig_vec_list)[-1][-idx])):
                if classes[i] == k:
                    sorted_eigenvector.append(np.array(eig_vec_list)[-1][-idx][i])
        sorted_eigenvector = np.array(sorted_eigenvector)
        plt.plot(range(len(np.array(eig_vec_list)[-1][-idx])), sorted_eigenvector, alpha=.7,␣
↪color=color)
        plt.grid(ls=':')
        plt.gca().spines['right'].set_visible(False)
        plt.gca().spines['top'].set_visible(False)
    plt.show()
```

[3]:
```python
M0_list = [5, 10, 15, 20]
experiment(M0_list, q_method="constant", q0=.4)
experiment(M0_list, q_method="constant", q0=.6)
experiment(M0_list, q_method="uniforme", q0=.5, sigma=.1)
experiment(M0_list, q_method="uniforme", q0=.5, sigma=.25)
experiment(M0_list, q_method="choisi", q1=.2, q2=.6)
experiment(M0_list, q_method="choisi", q1=.4, q2=.6)
```

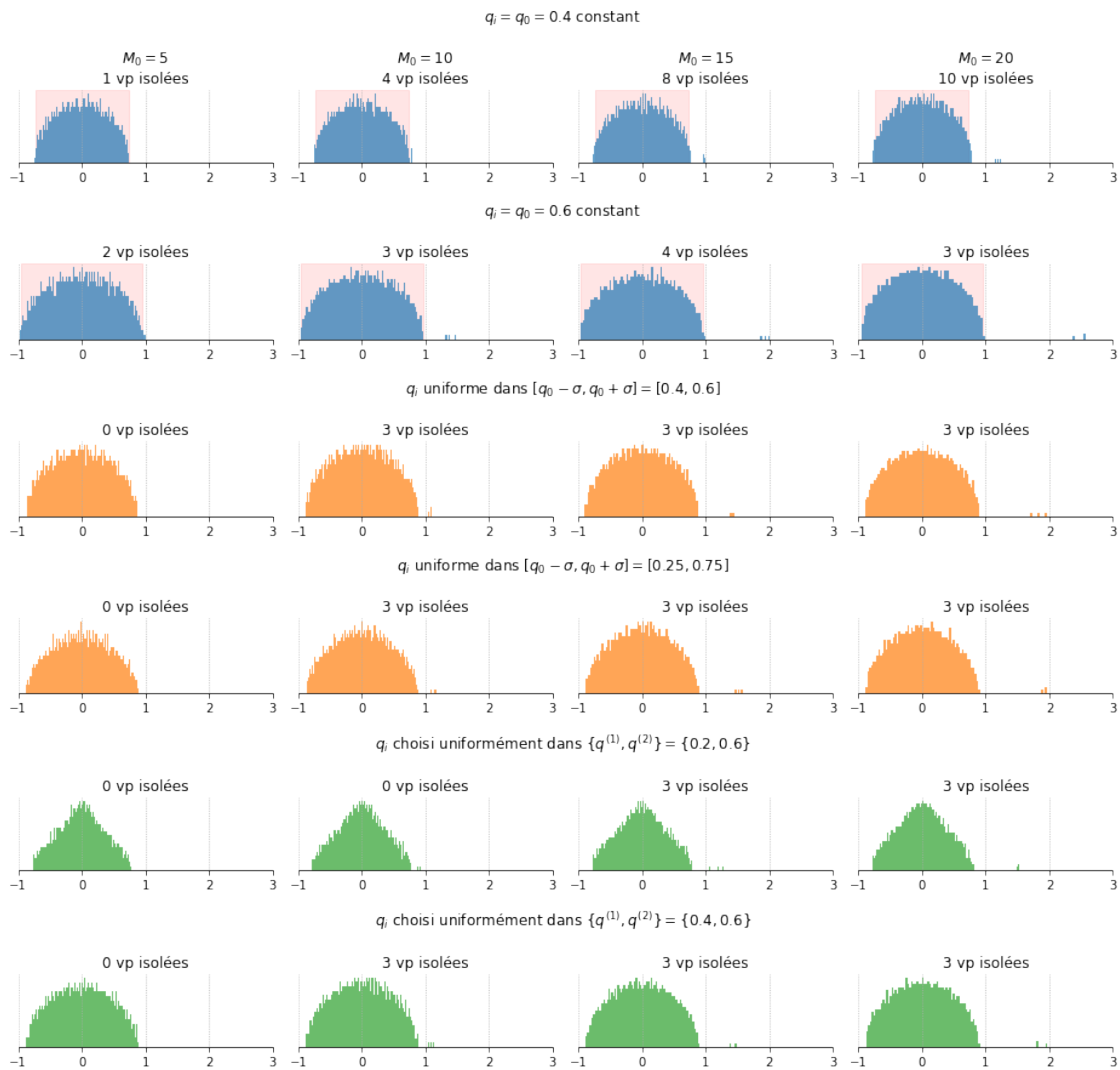
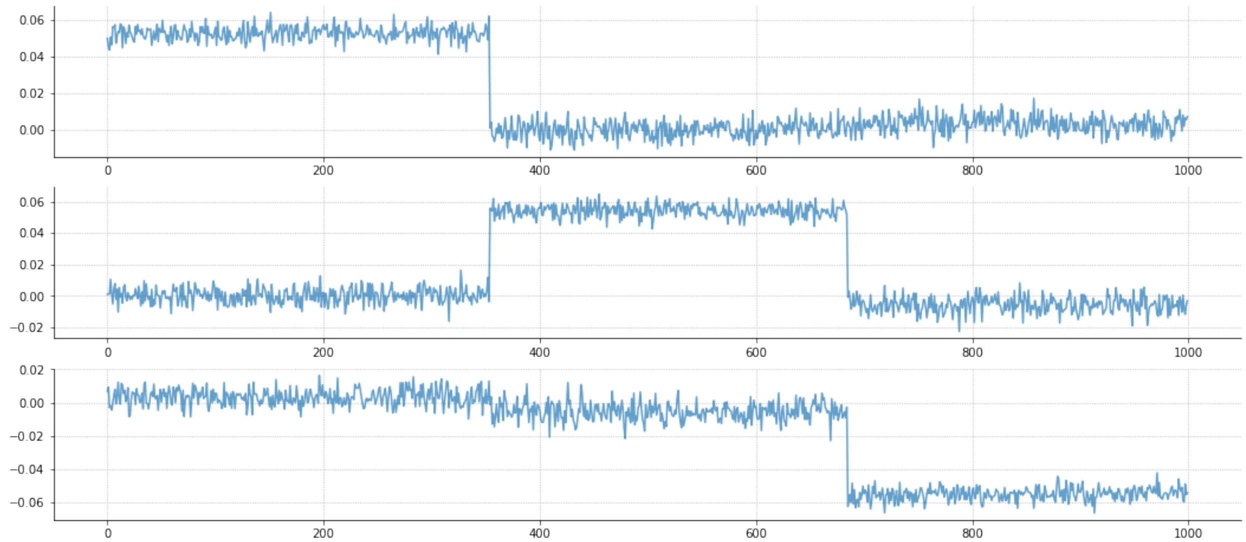


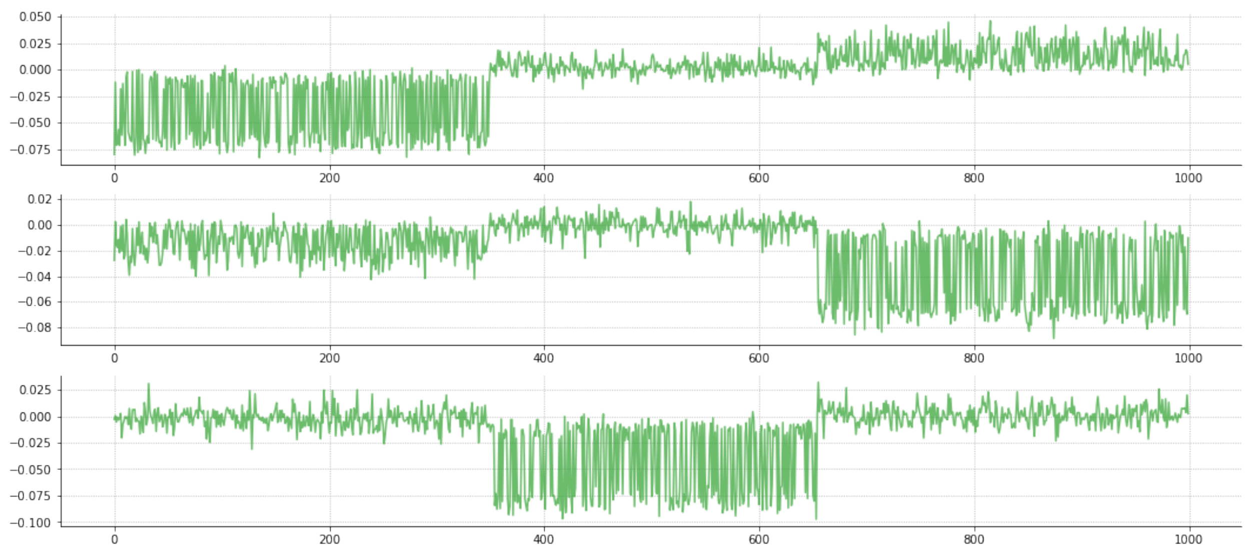
Figure 1: **Représentations graphiques du spectre de $\frac{1}{\sqrt{n}}B$** dans le cas $K = 3$ et $n = 1000$ pour plusieurs valeurs de $M = M_0 I_K$ (avec $M_0 = 5$, 10, 15, 20), et différents choix pour les q_i: une valeur constante égale à q_0 (bleu), une répartition uniforme dans un intervalle $[q_0 - \sigma, q_0 + \sigma]$ (orange), et un choix uniforme dans $\left\{q^{(1)}, q^{(2)}\right\}$ (vert). Le décompte des valeurs propres isolées se fait en comptant le nombre de valeurs propres $\lambda_i \geq 2\sigma + \varepsilon$ pour q_i constants (l'intervalle $[-2\sigma, 2\sigma]$ est représenté en rouge pâle), et $\lambda_i \geq 1$ pour les autres cas (c'est une approximation, qui se justifie par le fait que dans nos expériences les spikes seront dans tous les cas plus grands que 1 pour M_0 suffisamment grand).

Nous remarquons que dans le cas où q_0 est constant, le spectre est distribué selon la loi du demi-cercle de paramètre $q_0^2(1 - q_0^2)$ (voir cas homogène). Dans la plupart des autres cas, on observe également une distribution des valeurs propres en loi du demi-cercle. Pour des petites valeurs de M_0, il arrive que les valeurs propres extrêmes ne sortent pas du bulk ; plus la valeur de M_0 augmente, plus ces valeurs propres isolées s'éloignent du bulk, et plus il est facile de détecter les communautés du graphe. Lorsque les q_i sont uniformément choisis dans un ensemble de 2 valeurs éloignées (comme par exemple l'ensemble $\{0.2, 0.6\}$ de la Figure 1), on observe que les valeurs propres ne sont plus forcément distribuées selon la loi du demi-cercle.

4. Observons maintenant les vecteurs propres associés aux spikes :



(a) q_i constants égaux à $q_0 = 0.7$.



(b) q_i choisis uniformément dans $\left\{q^{(1)}, q^{(2)}\right\} = \{0.1, 0.7\}$.

Figure 2: **Trois plus grands vecteurs propres de** $\frac{1}{\sqrt{n}}B$ dans le cas $K = 3$ et $n = 1000$ pour $M = M_0 I_K$ (avec $M_0 = 20$) et différents choix pour les q_i: une valeur constante égale à q_0 (bleu), et un choix uniforme dans $\left\{q^{(1)}, q^{(2)}\right\}$ (vert).

- Pour détecter des communautés au sein d'un graphe nous regardons habituellement les plus petites valeurs propres du Laplacien, qui correspondent aux plus grandes valeurs propres de la matrice d'adjacence[1]. D'après la théorie sur les petites perturbations, on s'attend à ce que le spectre du Laplacien d'un graphe perturbé soit similaire à celui du graphe initial. Puisque les modèles "spike" présentent généralement une transition de phase au delà de laquelle nous pouvons détecter des grandes valeurs propres isolées, on peut extraire (par l'étude de leurs vecteurs propres associés) des informations concernant le nombre de communautés dans le graphe. Dans notre cas, nous intéresser aux vecteurs

[1] selon la manière dont est défini le Laplacien : si $L = I - A$, alors les valeurs propres de L sont les $1 - \lambda$, et les plus grandes valeurs propres de A sont donc les plus petites valeurs propres de L.

propres associés aux k valeurs propres extrêmes nous permet de détecter les k communautés présentes dans le graphe perturbé grace à un clustering du type k-means appliqué à l'espace engendré par les k plus grands vecteurs propres (de dimension k). L'algorithme basé sur le spectral clustering présenté par Rohe, Chatterjee, et Yu [Rohe et al., 2011] est comme suit. A partir du Laplacien,

1. on calcule les k vecteurs propres associés aux k valeurs propres extrêmes
2. on construit la matrice dont les colonnes sont ces vecteurs propres
3. on procède à un k-means clustering sur les lignes de cette nouvelle matrice

- Lorsque les q_i varient beaucoup, et tout particulièrement si l'on a des valeurs de q_i très éloignées, les noeuds à fort degré vont se connecter plus facilement entre eux – et ce peu importe la classe à laquelle ils appartiennent. Cela pourrait faire apparaitre des nouvelles communautés, ou altérer la distance des valeurs propres des spikes au bulk. Si celle-ci devient trop faible, cela rendrait la détection des communautés plus difficile.

Cas Homogène

1. Remarquons déjà que dans ce cas particulier, on peut simplifier l'expression de B:

$$\frac{1}{\sqrt{n}}B = \frac{1}{n}q_0^2 JMJ^* + \frac{1}{\sqrt{n}}X.$$

De plus, on peut remarquer que les X_{ij} sont i.i.d. centrées (car $X = A - \bar{A}$), et que leur variance est :

$$\begin{aligned}
[\mathbb{V}_q X]_{i,j} &= q_0^2(1-q_0^2) + \frac{q_0^2(1-2q_0^2)M_{a(i)a(j)}}{n^{1/2}} - \frac{q_0^4 M_{a(i)a(j)}^2}{n} \quad \text{car } q_i = q_0 \text{ pour tout } i \\
&= \begin{cases} q_0^2(1-q_0^2) + \frac{q_0^2(1-2q_0^2)M_{ii}}{n^{1/2}} - \frac{q_0^4 M_{ii}^2}{n} \underset{n\to\infty}{\longrightarrow} q_0^2(1-q_0^2) & \text{si } i=j \\ q_0^2(1-q_0^2) & \text{sinon,} \end{cases}
\end{aligned}$$

car M est diagonale. En posant $\sigma^2 = q_0^2(1-q_0^2)$, le moment d'ordre 4 de X est par ailleurs fini asymptotiquement, car

$$\begin{aligned}
\mathbb{E}_q(X_{ij}^4) &= \mathbb{V}_q(X_{ij}^2) + (\mathbb{E}_q X_{ij}^2)^2 \\
&= \mathbb{V}_q(X_{ij}^2) + (\mathbb{V}_q X_{ij})^2 \\
&= \mathbb{V}_q(A_{ij}^2 + \bar{A}_{ij}^2 - 2A_{ij}\bar{A}_{ij}) + (\mathbb{V}_q X_{ij})^2 \\
&\underset{n\to\infty}{\longrightarrow} \mathbb{V}_q(A_{ij}^2 + \bar{A}_{ij}^2 - 2A_{ij}\bar{A}_{ij}) + \sigma^4 \\
&< \infty \quad \text{car } A \text{ est une Bernouilli}
\end{aligned}$$

On a donc que la matrice $\frac{1}{\sqrt{n}}X$ vérifie les conditions du théorème de l'énoncé (Wigner "isotrope")[2], i.e. que pour tous vecteurs $u, v \in \mathbb{R}^n$ déterministes, on a avec probabilité 1

$$u^* \left(\frac{1}{\sqrt{n}}X - \lambda I_n\right)^{-1} v \quad \underset{n\to\infty}{\longrightarrow} \quad g_{\mathrm{sc},\sigma^2}(\lambda) u^* v = \frac{1}{\sigma}\, g_{\mathrm{sc}}\left(\frac{\lambda}{\sigma}\right) u^* v \tag{1}$$

[2]en admettant que ce théorème reste valable si le moment d'ordre 4 dépend de n, comme ici.

En gardant cela en tête, appliquons la condition du déterminant afin de trouver les valeurs propres isolées. On souhaite avoir :

$$
\begin{aligned}
0 &= \det\left(\frac{1}{\sqrt{n}}B - \lambda I_n\right) \\
&= \underbrace{\det(W-\lambda I_n)}_{\neq 0}\det\left(I_n + \frac{1}{\sqrt{n}}P(W-\lambda I_n)^{-1}\right) && \text{car } \lambda \text{ n'est pas vp du modèle de base} \\
&= \det\left(I_n + \frac{1}{n}q_0^2 JMJ^*\left(\frac{1}{\sqrt{n}}X - \lambda I_n\right)^{-1}\right) \\
&= \det\left(I_n + \frac{1}{n}q_0^2 M^{1/2}J^*\left(\frac{1}{\sqrt{n}}X - \lambda I_n\right)^{-1} JM^{1/2}\right) && \text{par le théorème de Sylvester} \\
&\underset{n\to\infty}{\overset{\text{p.s.}}{\sim}} \det\left(I_n + \frac{1}{n}q_0^2 M^{1/2}\frac{1}{\sigma}g_{\text{sc}}\left(\frac{\lambda}{\sigma}\right)J^*JM^{1/2}\right) && \text{d'après (1)} \\
&= \det\left(I_n + \frac{1}{\sigma n}q_0^2 g_{\text{sc}}\left(\frac{\lambda}{\sigma}\right)\operatorname{diag}\left(M_{ii}|\mathcal{C}_i|\right)\right) && \text{car } J^*J = \operatorname{diag}\left(|\mathcal{C}_i|\right) \\
&= \prod_{i=1}^{n}\left(1 + \frac{q_0^2 M_{ii}c_i}{\sigma}g_{\text{sc}}\left(\frac{\lambda}{\sigma}\right)\right) && \text{car } c_i \sim |\mathcal{C}_i|/n
\end{aligned}
$$

Ce qui est équivalent à l'existence d'un i tel que $g_{\text{sc}}\left(\frac{\lambda}{\sigma}\right) = -\frac{\sigma}{q_0^2 M_{ii}c_i}$. Remarquons maintenant que l'image de $]2,+\infty[$ par $g_{\text{sc}}(x) = \frac{-x+\sqrt{x^2-4}}{2}$ est $]-1,0[$ (on obtient cela avec un tableau de variation et en étudiant la limite de g_{sc} en l'infini). Il faut donc que $-\frac{\sigma}{q_0^2 M_{ii}c_i} \in]-1,0[$, ce qui est équivalent à

$$\boxed{M_{ii}c_i > \frac{\sigma}{q_0^2}.}$$

2. On sait que

$$g_{\text{sc}}\left(\frac{\lambda}{\sigma}\right) = \frac{-1}{\frac{\lambda}{\sigma} + g_{\text{sc}}\left(\frac{\lambda}{\sigma}\right)},$$

et en injectant l'expression de $g_{\text{sc}}\left(\frac{\lambda}{\sigma}\right)$ trouvée juste avant, on obtient que cette condition est équivalente à l'existence d'un i tel que

$$\boxed{\lambda_i = q_0^2 M_{ii}c_i + \frac{\sigma^2}{q_0^2 M_{ii}c_i}.}$$

Vérifions cela empiriquement :

```
[4]: def experiment_asympt(n_list, q_method, q0=None, sigma=None, M0=10, q1=None, q2=None, K=3,␣
     ↪show_suptitle=True, show_title=True, color="tab:blue"):
         """Perform an experiment for each value of n in n_list"""

         plt.figure(figsize=(13,5), tight_layout=True)
         eig_vals_list = []
         # plot eigenvalues
         plt.hlines(y=q0**2*M0/3 + (1-q0)**2/(M0/3), xmin=np.min(n_list), xmax=np.max(n_list),␣
     ↪linestyles="dashed", colors="tab:red", alpha=.7)
         for n in n_list:
             eig_vals, _, _ = get_eigenvalues(M0=M0, q_method=q_method, q0=q0, sigma=sigma, q1=q1,␣
     ↪q2=q2, n=n, K=K)

             sorted_idx = np.argsort(eig_vals)
```

```python
        eig_vals_list.append(eig_vals[sorted_idx][-3:])
    eig_vals_list = np.array(eig_vals_list)
    plt.plot(n_list, eig_vals_list[:,0], "o-", color=color, alpha=.5)
    plt.plot(n_list, eig_vals_list[:,1], "o-", color=color, alpha=.5)
    plt.plot(n_list, eig_vals_list[:,2], "o-", color=color, alpha=.5)
    plt.grid(ls=':')
    plt.gca().spines['right'].set_visible(False)
    plt.gca().spines['top'].set_visible(False)
    plt.semilogx()
    plt.show()
    return n_list, eig_vals_list, color
```

[5]:
```python
n_list = np.logspace(2,3.6).astype(int)
n_list, eig_vals_list, color = experiment_asympt(n_list, q_method="constant", q0=.7)
```

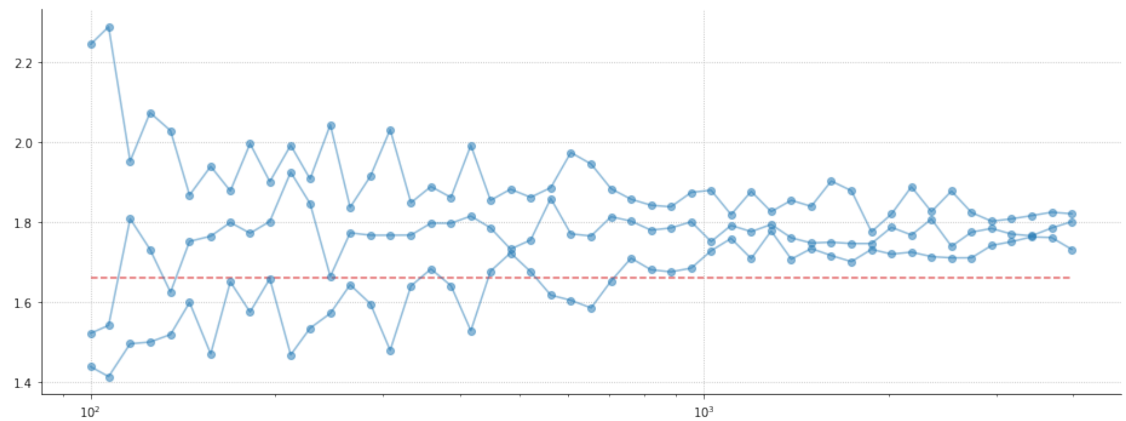


Figure 3: **Convergence des trois plus grandes valeurs propres** dans le cas $K = 3$ et pour des q_i constants égaux à $q_0 = 0.7$ pour différentes valeurs de n.

3. Donnons-nous un vecteur propre isolé u_m et le vecteur canonique j_a de la classe a. Soit Γ_m un contour fermé orienté positivement qui entoure la limite presque sûre de λ_m et laisse toutes les autres valeurs propres en dehors. Puisque $\left(\frac{1}{\sqrt{n}}X - zI\right)^{-1} = \sum_{i=1}^{p} \frac{u_i u_i^\top}{\lambda_i - z}$, on a que

$$
\begin{aligned}
\frac{1}{n_a} j_a^\top u_m u_m^\top j_a &= -\frac{1}{2i\pi} \oint_{\Gamma_m} \frac{1}{n_a} j_a^\top \left(\frac{1}{\sqrt{n}} B - zI\right)^{-1} j_a \mathrm{d}z \\
&= -\frac{1}{2i\pi} \oint_{\Gamma_m} \frac{1}{n_a} j_a^\top \left(\frac{1}{\sqrt{n}} X - zI + \frac{1}{n} q_0^2 J M J^*\right)^{-1} j_a \mathrm{d}z \\
&= \underbrace{-\frac{1}{2i\pi} \oint_{\Gamma_m} \frac{1}{n_a} j_a^\top Q(z) j_a \mathrm{d}z}_{=0 \text{ car holomorphe}} + \frac{1}{2i\pi} \oint_{\Gamma_m} \frac{1}{n_a} j_a^\top \left[Q(z) J \left(\frac{n}{q_0^2} M^{-1} + J^* Q(z) J\right)^{-1} J^* Q(z)\right] j_a \mathrm{d}z \quad \text{(Woodbury)} \\
&\underset{n\to\infty}{\overset{\text{p.s.}}{\sim}} \frac{1}{2i\pi} \oint_{\Gamma_m} \frac{1}{n_a} j_a^\top \left[g_{\mathrm{sc},\sigma^2}(z)^2 J \left(\frac{n}{q_0^2} M^{-1} + g_{\mathrm{sc},\sigma^2}(z) I\right)^{-1} J^*\right] j_a \mathrm{d}z \\
&= \frac{1}{2i\pi} \oint_{\Gamma_m} n_a e_a^\top g_{\mathrm{sc},\sigma^2}(z)^2 \left(\frac{n}{q_0^2} M^{-1} + g_{\mathrm{sc},\sigma^2}(z) I\right)^{-1} e_a \mathrm{d}z \quad \text{car } J^* j_a = n_a e_a, \text{ vecteur de la base canonique} \\
&= \frac{1}{2i\pi} \oint_{\Gamma_m} \frac{g_{\mathrm{sc},\sigma^2}(z)^2}{g_{\mathrm{sc},\sigma^2}(z) + \frac{1}{q_0^2 M_{aa} c_a}} \mathrm{d}z \quad \text{car } n_a/n \sim c_a \\
&= \frac{1}{2i\pi} \oint_{\Gamma_m} \underbrace{\frac{g_{\mathrm{sc},\sigma^2}(z)^2}{g_{\mathrm{sc},\sigma^2}(z) - g_{\mathrm{sc},\sigma^2}(\lambda_a)}}_{:=G(z)} \mathrm{d}z
\end{aligned}
$$

Appliquons maintenant le théorème des résidus. On peut déjà remarquer que si l'on considère un vecteur propre u_m qui ne correspond pas à la classe a, G n'a pas de singularité dans Γ_m et on a donc $\oint_{\Gamma_m} G(z)\mathrm{d}z = 0$. Sinon, la seule singularité de l'intégrande G est en λ_a, et on a que

$$
\begin{aligned}
\mathrm{Res}(G, \lambda_a) &= \lim_{z\to\lambda_a} (z - \lambda_a) G(z) \\
&= \frac{g_{\mathrm{sc},\sigma^2}(\lambda_a)^2}{g'_{\mathrm{sc},\sigma^2}(\lambda_a)} = 1 - \sigma^2 g_{\mathrm{sc},\sigma^2}(\lambda_a)^2,
\end{aligned}
$$

ce qui montre que

$$
\boxed{\frac{1}{n_a} j_a^\top u_m u_m^\top j_a = \begin{cases} 1 - \left(\dfrac{\sigma}{q_0^2 M_{aa} c_a}\right)^2 & \text{si } m = a \\ 0 & \text{sinon.} \end{cases}}
$$

4. Non fait par manque de temps.

5. Dans le cas où l'on peut détecter des communautés (lorsque la condition de transition de phase trouvée en **2.** est vérifiée), on connait les alignements des vecteurs propres (cf. question **3.**) et l'on peut ainsi procéder comme expliqué en **4.** des observations préliminaires.

Nous ne savons pas ce qu'il manque pour évaluer les performances de l'algorithme. Nous pouvous toujours, une fois le clustering effectué, utiliser une métrique classique de mesure de performance de clustering.

Pour être assuré que la méthode proposée plus haut peut fonctionner, il faudrait s'assurer que les valeurs propres extrêmes sont assez éloignées du bulk. Dans le cas d'un Laplacien "classique", les vecteurs propres des classes sont relativement bien alignés avec leurs vecteurs canoniques, on peut considérer qu'ils sont composés de valeurs dans $\{0, 1\}$ et ainsi, un k-means appliqué aux lignes de la matrice des vecteurs propres des spikes du Laplacien n'aura donc pas de difficulté à détecter des communautés. Néanmoins, si le Laplacien considéré est normalisé par la matrice des degrés, et que les degrés sont très différents, on se retrouve dans le cas de la question **4.** des observations préliminaires où détecter les vecteurs propres associés aux valeurs propres extrêmes pourrait s'avérer compliqué. Nous pouvons nous référer à la normalisation du Laplacien proposée par Ng, Jordan et Weiss [Ng et al., 2002] afin de contourner cette difficulté.

Cas Hétérogène

1. Non fait par manque de temps.

2. Non fait par manque de temps.

References

[Ng et al., 2002] Ng, A., Jordan, M., and Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. *Adv. Neural Inf. Process. Syst*, 14. 9

[Rohe et al., 2011] Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878 – 1915. 6