Algebraic topology of random simplicial complexes and applications to sensor networks

Eduardo Ferraz

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2012-ENST-006

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Informatique et Réseaux »

présentée et soutenue publiquement par

Eduardo FERRAZ

le 22 février 2012

Topologie algébrique de complexes simpliciaux aléatoires et des applications aux réseaux de capteurs

Directeur de thèse : **Laurent DECREUSEFOND**
Co-encadrement de la thèse : **Hugues RANDRIAMBOLOLONA**

Jury

M. François BACCELLI, Directeur de Recherche, CS Department, INRIA-ENS	Président du jury
M. Vin DE SILVA, Assistant Professor, Department of Mathematics, Pomona College	Rapporteur
M. Sergei ZUYEV, Professeur, Matematiska vetenskapper, Chalmers tekniska högskola	Raporteur
M. Pierre CALKA, Professeur, Lab. de Math. Raphaël Salem, Université de Rouen	Membre du jury
M. Süleyman ÜSTÜNEL, Professeur, INFRES, Télécom Paristech	Membre du jury
M. Frédéric CHAZAL, Directeur de Recherche, GEOMETRICA, INRIA-Saclay	Membre du jury
M. Laurent DECREUSEFOND, Professeur, INFRES, Télécom Paristech	Membre du jury
M. Hugues RANDRIAMBOLOLONA, Titre, INFRES, Télécom Paristech	Membre du jury

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

THÈSE

Acknoledgements

Je tiens à remercier tout d'abord la personne qui a été la plus importante pour ma thèse, professeur Laurent Decreusefond. Ses qualités en tant que directeur de thèse m'ont permis de progresser tout au long de ma thèse et de la terminer. Pour ce qui est de ses vertus de professeur et de personne, il aura toujours mon admiration et mon respect.

Je remercie Hugues Randriambololona, mon co-directeur de thèse, et également professeur Philippe Martins pour toutes les discussions et les encouragements qui ont beaucoup contribué au développement de mon travail à Télécom.

Je remercie aussi beaucoup les rapporteurs de thèse, Sergei Zuyev et Vin de Silva, leurs commentaires techniques et leurs suggestions m'ont particulièrement aidé à améliorer la thèse. Merci aussi aux membres du jury, François Baccelli, Pierre Calka, Üstünel et Frédéric Chazal de m'avoir accordé leur sollicitude.

Je suis très reconnaissant à ma collègue et amie Anaïs Vergne, qui m'a donné le plaisir de travailler avec elle et qui a réalisé une contribution brillante dans mon travail. Je remercie également Ian Flint et Gutemberg Junior pour chaque goutte de sueur qu'ils ont dépensée à m'aider et pour leurs innombrables actes de gentillesse.

Je remercie tous mes amis qui ne m'ont pas directement aidé, mais qui ont eu une grande importance pendant mon doctorat: Julie Gaudin, Dimitri Edouard, Marcia Costa, Rafael Galvão, Erica Benvindo, Davide Tammaro, Sharon Cinquegrana, Rodrigo Soulé, Sidonie Garnier, Triston Lavergne, Roberta Previtera et Lorena Marchant.

Enfin, je remercie ma famille qui m'a donné toutes les conditions pour parvenir au point de soutenir une thèse.

Résumé de la thèse

Chapitre 1 : Introduction

Une baisse des prix et de la consommation énergétique ainsi que une miniaturisation des composants ont entrainé une forte augmentation de l'utilisation de réseaux de capteurs. De nouveaux outils mathématiques ont permis une nouvelle modélisation des réseaux de capteurs : deux de ces outils sont l'analyse stochastique et la topologie algébrique. En analyse stochastique, introduite dans les années 90 par Baccelli, le déploiement des réseaux de mobiles et de capteurs est considéré comme un processus ponctuel de Poisson, à la place du comportement déterministe utilisé auparavant. Dix ans plus tard, Ghrist a choisi de modéliser la couverture des réseaux de capteurs comme un complexe simplicial. Par conséquent, on pouvait appliquer des résultats de topologie algébrique à de tels réseaux. Ces résultats ont permis le calcul explicite de l'homologie du complexe simplicial. Donc, il a été possible de répondre à des questions pertinentes et non-triviales qui se posent dans les réseaux de capteurs, chaque acteur n'ayant aucune connaissance du réseau autre que son identifiant et ceux de ses voisins. De telles questions sont : "le réseau est-il connecté ?", "la région cible est-elle couverte ?", et "combien de composantes connexes et de trous y a-t-il dans le réseau ?".

Cette thèse est composée de deux parties. La première partie utilise l'analyse stochastique pour fournir des bornes pour la probabilité de surcharge de différents systèmes grâce aux inégalités de concentration. Bien qu'ils soient généraux, nous appliquons ces résultats à des réseaux sans-fil réels tels que le WiMax et le trafic utilisateur multi-classe dans un système OFDMA. Dans la seconde partie, nous trouvons des liens entre la topologie de la couverture dans un réseau de capteur et celle du complexe simplicial correspondant. Cette analogie met en valeur de nouvelles facettes des certains objets mathématiques comme les nombres de Betti, le nombre de k-simplexes, et la caractéristique d'Euler. Puis, nous utilisons conjointement la topologie algébrique et l'analyse stochastique, en considérant que les positions des capteurs sont une réalisation d'un processus ponctuel de Poisson. Nous en déduisons les statistiques du nombre de k-simplexe et de la caractéristique d'Euler, ainsi que des bornes supérieures pour la distribution des nombres de Betti, le tout en d dimensions. Nous démontrons aussi que le nombre de k-simplexes converge vers une distribution Gaussienne quand la densité de capteurs tend vers l'infini à une vitesse de convergence connue. Enfin, nous nous limitons au cas unidimensionnel. Dans ce cas, le problème devient équivalent à résoudre une file $M/M/1/1$ préemptive. Nous obtenons ainsi des résultats analytiques pour des quantités telles que la distribution du nombre de composantes connexes et la probabilité de couverture totale.

Les capteurs et les réseaux de capteurs

Un capteur est un outil capable de balayer un domaine spatial et de transmettre un signal permettant d'obtenir des informations. Les capteurs varient selon à leurs fonctions de telle sorte que nous pouvons les utiliser pour récuperer des donnés sismiques, thermiques, magnétiques, visuelles, acoustiques, etc. Un réseau de capteurs permet l'interpolation d'informations ponctuelles d'un phénomène, obtenues par les capteurs vers des dimensions plus grandes. La définition de la couverture d'un capteur dépend de sa fonction, mais elle est essentiellement donnée par tous les points qui sont à une distance (pour une norme quelconque) inférieure à une borne ϵ.

Les outils mathématiques

Les deux outils mathématiques principaux utilisés dans la thèse sont la topologie algébrique et le processus ponctuel de Poisson. La topologie algébrique permet d'évaluer la transition locale-globale : quelques informations globales telles quelles la connectivité, la couverture et la caractéristique d'Euler peuvent être obtenues tout simplement par l'échange d'informations parmi des capteurs proches les uns des autres, sans avoir besoin de connaître leurs positions, leurs orientations ou que l'information enregistrée par un capteur soit transmise plus loin qu'à ses voisins les plus proches.

Le processus ponctuel de Poisson, depuis les années 1990, est la base des modèles stochastiques pour les réseaux. Dans cette approche, la représentation physique du réseau est préservée, mais les positions géographiques des capteurs ne sont plus modélées comme s'elles étaient fixes, donc aucun déployment du réseau non-aléatoire ne nous est utile. Au lieu de cela, cette méthode nous permet de prendre quelques caractéristiques importantes pour le résau étant donné la densité de ces points. En outre, les processus ponctuels de Poisson possèdent le plus grand nombre de résultats parmi tous les processus ponctuels, ce qui motive son usage pour la modélisation des noeuds des réseaux de capteurs.

Chapitre 2 : Le modèle stochastique

Le principal modèle stochastique utilisé dans cette thèse est celui du processus ponctuel de Poisson. A une filtration juste, ce processus peut être vu comme une Martingale et ainsi tous les outils du calcul de Malliavin peuvent être directement utilisés. Les définitions et les résultats les plus importants en déduits sont décrits à la suite.

Le processus ponctuel de poisson

Pour caractériser l'aléa du système, nous considérons que l'ensemble de points est représenté par un processus de Poisson ω à intensité λ dans un espace Y. L'espace de configurations sur Y, est l'ensemble des mesures de points simples et localement finis :

$$\Omega^Y = \left\{ \omega = \sum_{k=0}^{n} \delta(x_k) \ : \ (x_k)_{k=0}^{k=n} \subset Y, \ n \in \mathbb{N} \cup \{\infty\} \right\},$$

où $\delta(x)$ denote la mesure de Dirac en $x \in Y$. Localement fini veut dire que $\omega(K) < \infty$ quelque soit le compact K dans Y. Normalement, il convient d'identifier un élément ω dans Ω^Y avec l'ensemble qui correspond à son support, c'est-à-dire que $\sum_{k=0}^{n} \delta(x_k)$ est identifié avec l'ensemble non-ordonné $\{x_1, \cdots, x_n\}$. Si $A \in \mathcal{B}(Y)$, nous avons $\delta(x_k)(A) = \mathbf{1}_{[x_k \in A]}$, alors:

$$\omega(A) = \sum_{x_k \in \omega} \mathbf{1}_{[x_k \in A]} = \int_A \mathrm{d}\omega(x),$$

compte le nombre d'atomes contenus dans A.

Nous définissons $\mathbf{E}_\lambda [F(\omega)]$ comme l'espérance d'une fonction F qui ne dépend que de ω étant donné que l'intensité de ce processus est λ et $\mathbf{P}_\lambda[\omega \in Y] = \mathbf{E}_\lambda \left[\mathbf{1}_{[\omega \in Y]}\right]$. Par analogie, nous définissons $\mathbf{V}_\lambda [F(\omega)]$ et $\mathrm{Cov}_\lambda [F(\omega), G(\omega)]$, de même que $\Delta_n = \{(x_1, \cdots, x_n) \in Y^n \mid x_i \neq x_j, \forall i \neq j\}$. Soit $f(x_1, \cdots, x_n)$ une fonction mesurable $F(\omega)$ est une variable aléatoire donnée par :

$$F(\omega) = \sum_{\substack{x_i \in \omega \cap A, 1 \leq i \leq n \\ x_i \neq x_j \ if \ i \neq j}} f(x_1, \cdots, x_n) = \int_{A \cap \Delta_n} f(x_1, \cdots, x_n) \ \mathrm{d}\omega(x_1) \cdots \ \mathrm{d}\omega(x_n).$$

Une proprieté bien connue du processus ponctuel de Poisson, la formule de Campbell, établit que

$$\mathbf{E}_\lambda [F(\omega)] = \int_A f(x_1, \cdots, x_n) \ \mathrm{d}\lambda(x_1) \cdots \ \mathrm{d}\lambda(x_n).$$

Un processus ponctuel de Poisson marqué ayant les marque dépendantes de la position tel quel la loi de probabilité de la marque Y_n ne dépend que de la position X_n à travers d'un noyau K:

$$\mathbf{P}(Y_n \in B \,|\, \omega) = K(X_n, B), \text{ for any } B \subset X.$$

Si K est un noyau de probabilité, c'est-à-dire que $K(x, X) = 1$ pour quelque soit $x \in \mathbf{R}^k$, alors ω' est un processus de Poisson d'intensité $K(x, \ \mathrm{d}y) \ \mathrm{d}\lambda(x)$ sur $\mathbf{R}^k \times \mathbf{R}^m$. Si $f : \mathbf{R}^k \times \mathbf{R}^m \to \mathbf{R}$ est une fonction non-négative mesurable, soit

$$F = \int f \ \mathrm{d}\omega' = \sum_{X_n \in \omega} f(X_n, Y_n).$$

Alors, nous pouvons utiliser un résultat équivalent à la formule de Campbell, pour les processus ponctuels marqués :

$$\mathbf{E}_{\lambda,K}\left[F\right] = \int_{\mathbf{R}^k \times \mathbf{R}^m} f(x,\, y)\, K(x,\ \mathrm{d}y)\ \mathrm{d}\lambda(x).$$

Pour un large ensemble de fonctions symetriques f, l'intégrale stochastique de Poisson $I_n(f_n)$ est définie telle que :

$$I_n(f_n)(\omega) = \int_{\Delta_n} f_n(x_1, \cdots, x_n)(\ \mathrm{d}\omega(x_1) - \ \mathrm{d}\lambda(x_1)) \cdots (\ \mathrm{d}\omega(x_n) - \ \mathrm{d}\lambda(x_n)).$$

Cette définition nous permet de décomposer une fonction $F \in L^2(\Omega^Y,\, \mathbf{P})$ en une somme d'intégrales stochastiques :

$$F = \mathbf{E}_\lambda\left[F\right] + \sum_{n=1}^{\infty} I_n(f_n).$$

Nous utilisons trois opérateurs à partir de cette décomposition : le gradient D_t, l'opérateur de Ornstein-Uhlenbeck L et son inverse L^{-1}. Les définitions de ces opérateurs, dans des cadres où ils sont définis, sont les suivantes :

$$D_t F = \sum_{n \geq 1} n I_{n-1}(f_n(*, t)),$$

$$LF = -\sum_{n=1}^{\infty} n I_n(f_n),$$

$$L^{-1}F = -\sum_{n=1}^{\infty} \frac{1}{n} I_n(f_n).$$

Deux théorèmes importants utilisés dans cette thèse découlent des propriétés du calcul de Malliavin. Le premier établit que si la fonction F est telle que $\mathbf{E}_\lambda\left[F\right] = 0$ et $\mathrm{Var}(F) = 1$, alors

$$d_W(F, \mathcal{N}(0,1)) \leq \mathbf{E}_\lambda\left[\left|1 - \int_Y \left[D_t F \times D_t L^{-1} F\right]\ \mathrm{d}\lambda(t)\right|\right] + \int_Y \mathbf{E}_\lambda\left[\left|D_t F\right|^2 \left|D_t L^{-1} F\right|\right]\ \mathrm{d}\lambda(t).$$

L'autre, que si $D_t F < K$ pour tous les t, $K \geq 0$ et $\|DF\|_{L^\infty(\Omega, L^2(Y))} < \infty$, alors

$$\mathbf{P}(F - \mathbf{E}_\lambda\left[F\right] \geq x) \leq \exp\left(-\frac{x}{2K} \log\left(1 + \frac{xK}{\|DF\|_{L^\infty(\Omega, L^2(Y))}}\right)\right).$$

Ce résultat est aussi valable pour les processus ponctuels de Poisson marqués.

Chapitre 3 : Modèle de température d'interférence dans des réseaux de radio cognitives

Dans ce chapitre nous nous proposons d'appliquer le modèle du Processus ponctuel de Poisson dans une approche de la température d'interférence (interference temperature, IT) pour l'usage du spectre de fréquence pour des réseaux de communication sans fil. Dans cette approche, les utilisateurs secondaires (USs) peuvent utiliser les bandes des utilisateurs primaires (UPs) tant que cela ne cause pas de dommages à la communication de ces derniers.

Le modèle physique

Dans l'approche de l'IT, les utilisateurs secondaires traitent les autres USs, les UPs, le bruit et déautres sources d'interférence comme dess interférence. Puis ils ajustent leur puissance de transmission de telle sorte qu'il n'y ait pas de dommage pour les UPs. La température d'interférence est donnée par :

$$T_I(f_c, B) = \frac{P_I(f_c, B)}{kB},$$

où T_I est estimée en Kelvin, P_I en Watts, la bande considérée est centrée à f_c avec une largeur de B, en Hertz et k est la constante de Boltzmann.

Sont présentés deux modèles différents d'IT : le modèle idéal et le généralisé. Le premier limite l'interférence des signaux des UPs, et dès lors la connaissance de l'activité des UPs devient nécessaire. La relation suivante exprime ce modèle :

$$T_I(f_i, B_i) + \frac{M_i P}{kB_i} \leq T_L(f_i) \quad \forall\, 1 \leq i \leq n,$$

où P est la puissance d'interférence moyenne des USs qui opèrent avec la fréquence centrale f_c et avec largeur de bande B ; T_L est la température limite établie et la constante M_i représente l'atténuation entre l'émetteur secondaire et le récepteur primaire. Pour le modèle généralisé, l'activité des UPs n'est pas demandée, de telle sorte que le modèle peut être appliqué sur toute la largeur de bande, indépendamment des positions des UPs. Nous avons donc :

$$T_I(f_c, B) + \frac{MP}{kB} \leq T_L(f_c).$$

Le modèle utilisé pour l'atténuation de la puissance transmise est connu sous le nom de *path loss*, et établit une diminution polynommiale des signaux par rapport à la distance de la source. Ainsi, la puissance du signal dans une coordonnée y, compte tenu que le signal a été transmi dans la coordonnée x à une puissance μ_j est donnée par :

$$p_j(x, y) = \min\left(\mu_j, \mu_j \left(\frac{r_0}{\|x - y\|}\right)^{\alpha}\right),$$

α étant le coefficient du *path loss* et r_0 une distance de référence.

La capacité moyenne des utilisateurs sécondaire

Les expressions pour l'approche de l'IT étant bien définies, nous pouvons appliquer le processus ponctuel de Poisson pour placer les utilisateurs. Dans cette section, nous

calculons la capacité permise pour les utilisateurs secondaires lorsque la condition de ne pas dommager les UPs n'est respectée que pour des quantités moyennes. La densité moyenne des UPs est donnée par λ_1, celle des USs par λ_2 et celle des interféreurs est donnée par λ_3. Tout d'abord, nous obtenons la puissance maximale permise aux USs pour le cas idéal :

$$\mu_2^{id} \leq \frac{(\alpha-2)T_L k B_i - \lambda_3 \mu_3 \pi \alpha}{\lambda_2 \pi \alpha}.$$

Pour le cas généralisé, nous avons :

$$\mu_2^{gen} \leq \frac{(\alpha-2)T_L k B - \lambda_3 \mu_3 \pi \alpha - \frac{B_i}{B}\lambda_1 \mu_1 \pi \alpha}{\lambda_2 \pi \alpha}.$$

Avec ces expressions, nous utilisons le théorème de Shannon-Hartley, qui détermine la limite de la capacité donné par :

$$C(x,y) = B \log_2 \left(1 + \frac{p_2(x,y)}{\mathbf{E}_\lambda \left[I(x)\right]}\right),$$

pour avoir la capacité moyenne de communication des utilisateurs secondaires telle que si la puissance transmise est égale à la puissance moyenne, les USs ne causeront pas de dommage aux UPs :

$$C = \frac{2BK^{\frac{2}{\alpha}}}{R_2^2 \ln(2)} \left[\frac{\ln(1+K)}{2K^{\frac{1}{\alpha}}} + h\left(\frac{R_2}{K^{\frac{1}{\alpha}}}, \alpha\right) - h\left(\frac{2}{K^{\frac{1}{\alpha}}}, \alpha\right)\right].$$

Dans ce résultat, nous avons :

$$K \triangleq \frac{\mu_2}{\mathbf{E}_\lambda \left[I(x)\right]} = \frac{\mu_2(\alpha-2)}{\pi\alpha\left(\frac{B_i}{B}\mu_1\lambda_1 + \mu_2\lambda_2 + \mu_3\lambda_3\right)}$$

Pour le cas idéal, il suffit de remplacer μ_2 par μ_2^{id} et pour le cas généralisé, nous remplaçons μ_2 par μ_2^{gen}. La fonction h est définie ainsi :

$$h(r,t) \triangleq \int_0^r \ln\left(1 + \frac{1}{x^t}\right) x \ \ \mathrm{d}x.$$

Il est possible de trouver des expressions explicites pour h quand α est un entier positif.

La capacité total du réseau est donnée à partir de l'expression suivante :

$$C_{total} = \int\int_D C(x,y) \frac{\mathbf{1}_{[\{\|x-y\| \leq R_2\}]}}{\pi R_2^2} \ \mathrm{d}y \lambda_2 \ \mathrm{d}x.$$

Cela nous permet de trouver les capacités totales dans les deux cas, puisque

$$C_{total} = C\lambda_2 \pi R^2.$$

Une borne supérieure de la probabilité de dommage pour les UPs

La section précédente considérait que les puissances transmises par les USs étaient tout le temps égales à leurs valeurs moyennes. Pourtant, si la distribution de la somme d'interférences est approximativement symétrique, la restriction ne sera pas respectée dans environ 50% des cas. Dans cette section, nous utilisons les inégalités de concentration pour

trouver une relation explicite entre la probabilité de dommage pour les UPs et la puissance transmise par les USs.

En utilisant l'inégalité de concentration :

$$\mathbf{P}(F \geq t + m_F) \leq \exp\left(-\frac{t}{2s} g\left(1 + \frac{ts}{v_F}\right)\right).$$

Si l'interférence totale est donné par la somme des interférences, il nous faut maintenant utiliser la formule de Campbell afin de trouver m_F and v_F :

$$\begin{aligned} m_F &= \mathbf{E}_\lambda\left[\sum_{x_i \in \omega_2} p_2(x_i, 0)\right] + \mathbf{E}_\lambda\left[\sum_{x_i \in \omega_3} p_3(x_i, 0)\right], \\ v_F &= \mathbf{E}_\lambda\left[\sum_{x_i \in \omega_2} p_2^2(x_i, 0)\right] + \mathbf{E}_\lambda\left[\sum_{x_i \in \omega_3} p_3^2(x_i, 0)\right], \end{aligned}$$

où ω_2 et ω_3 sont des processus ponctuels de Poisson : ω_3 représent le déploiment des interféreurs et ω_2 le déploiement des USs. Pour le cas idéal, nous obtenons alors :

$$m_F = \frac{\alpha\pi(\mu_2\lambda_2 + \mu_3\lambda_3)}{(\alpha - 2)},$$

$$v_F = \frac{2\alpha\pi(\mu_2^2\lambda_2 + \mu_3^2\lambda_3)}{(2\alpha - 2)}$$

et

$$s = \max(\mu_2, \mu_3).$$

Pour le cas généralisé, nous avons :

$$m_F = \frac{\alpha\pi(\frac{B_i}{B}\mu_1\lambda_1 + \mu_2\lambda_2 + \mu_3\lambda_3)}{(\alpha - 2)},$$

$$v_F = \frac{2\alpha\pi(\left(\frac{B_i}{B}\mu_1\right)^2\lambda_1 + \mu_2^2\lambda_2 + \mu_3^2\lambda_3)}{(2\alpha - 2)}$$

et

$$s = \max\left(\frac{B_i}{B}\mu_1, \mu_2, \mu_3\right).$$

La rélation entre la probabilité de dommage et la puissance transmise est donnée par

$$\mathbf{P}(F \geq T_L(f_i)kB_i) \leq \exp\left(-\frac{T_L(f_i)kB_i - m_F}{2s} \ln\left(1 + \frac{(T_L(f_i)kB_i - m_F)s}{v_F}\right)\right) = P_{sup}.$$

Prenons le cas idéal quand $\mu_2 > \mu_3$. Si nous définissons μ_2^q comme la puissance de transmission telle que la probabilité de dommage pour les UPs est plus petite que q et η

est le rapport entre la puissance de transmission et la puissance de transmission pour le cas moyen qu'a été discuté auparavant, μ_2^{id}, alors nous avons, $\mu_2^q = \eta\mu_2^{id}$. Rappelons aussi que η est une fonction de q. Donc

$$q = \exp\left(\frac{\lambda_2\pi\alpha}{2(\alpha-2)}\frac{\eta(q)-1}{\eta(q)}\ln\left(1+\frac{2(\alpha-1)}{(\alpha-2)}\frac{1-\eta(q)}{\eta(q)}\right)\right).$$

Le résultat ci-dessus est le plus important de ce chapitre, car il signifie que dès que $q(\eta)$ est une bijection, nous pouvons écrire η en fonction de q, et donc que nous pouvons obtenir la puissance de transmission des USs à partir de la probabilité de dommage maximale pour les UPs.

Chapitre 4 : Une borne supérieure de la probabilité de perte dans un système OFDMA

Nous présentons dans ce chapitre une autre application du processus ponctuel de Poisson afin de modéliser un système sans fil basé sur l'OFDMA (*Orthogonal Frequency Division Multiple Access*). Tandis que l'IT permet l'usage d'une même partie du spectre aussi par plusieurs utilisateurs, l'OFDMA ne le permet pas. Dans ce dernier cas, le système distribue le spectre dynamiquement, de façon à permettre le plus grand nombre d'utilisateurs possible dans le réseau.

Le modèle physique considère que :

- La position de chaque utilisateur est indépendente de celles des autres et leurs positions sont distribuées identiquement ;
- Le temps d'arrivée entre deux demandes pour service consecutives dans le système est distribué exponentialement ;
- Le temps de service pour chaque utilisateur est exponentialement distribué avec une moyenne de $1/\nu$;
- La cellule C du réseau est circulaire, elle a une antenne placée en son centre et son rayon s'appelle R ;
- La densité superficielle de l'arrivée d'utilisateurs est constante.

Avec ces hypothèses, nous pouvons montrer que le processus ponctuel des utilisateurs actifs (ceux qui communiquent) est, en équilibre, un processus ponctuel de Poisson.

Si le système possède N_0 sous-canaux, nous pouvons dire que la probabilité de perte est donnée par :

$$P_{loss} = \mathbf{P}_\lambda \left(\int N \ \mathrm{d}\omega \geq N_0 \right),$$

de plus, N est le nombre de sous-canaux utilisés par un utilisateur placé à x, peut être exprimé ainsi :

$$N(x) = \left\lceil \frac{C_0}{W \log_2 \left(1 + \dfrac{P_t K G}{(I+\eta)\|x\|^\gamma} \right)} \right\rceil .$$

Dans cette expression, g est le gain de puissance dans le canal, C_0 est la capacité demandée par les utilisateurs, P_t la puissance des signaux à la source, I l'interférence des autres cellules, γ le coefficient de *path loss*, η le bruit et W la largeur de bande de chaque sous-canal.

Si, en plus, la perte de puissance d'un canal ne dépend que de la distance entre la source et le récepteur, nous obtenons :

$$\mathbf{P}_\lambda \left(\int N \ \mathrm{d}\omega \geq \alpha m_N \right) \leq \exp \left(-\frac{v_N}{N_{max}^2} \ g \left(\frac{(\alpha - 1) m_N N_{max}}{v_N} \right) \right),$$

où

$$m_N = \frac{\pi \rho}{\nu} \sum_{j=1}^{N_{max}} j(R_j^2 - R_{j-1}^2)$$

et

$$v_N = \frac{\pi\rho}{\nu} \sum_{j=1}^{N_{max}} j^2 (R_j^2 - R_{j-1}^2).$$

Nous pouvons aussi considérer que G est aléatoire et si nous prenons $S = 1/G$, tel que

$$p_S(y) = \frac{\xi}{\sqrt{2\pi}\sigma y} \exp\left[-\frac{(10\log_{10} y - \mu)^2}{2\sigma^2}\right],$$

où $\xi = 10/\ln 10$, nous avons le modèle d'interférence connu sous le nom de *shadowing*. Si un utilisateur ne peut communiquer que lorsque le rapport signal-interférence est supérieur à une constante β_{min}, le nombre maximal de sous-canaux est limité par

$$N_{max} = \left\lceil \frac{C_0}{W \log_2(1+\beta_{min})} \right\rceil.$$

Le nombre de sous-canaux demandé par un utilisateur placé à x qui a un gain y est donné par :

$$N(x,\, y) = \left\lceil \frac{C_0}{W \log_2\left(1 + \dfrac{P_t K}{\eta y \|x\|^\gamma}\right)} \right\rceil.$$

Par ailleurs, nous obtenons des résultats explicites en fonction des paramètres qui ont déjà été définis. Et, enfin, nous pouvons aussi considérer des différents besoins de capacité : par exemple, une classe d'utilisateurs qui souhaitent envoyer des données (et qui a besoin d'une capacité moins élevée) et une classe d'utilisateurs qui souhaitent télécharger des vidéos (et qui a ainsi besoin d'une capacité plus élevée). Dans ce cas là, une classe j a besoin de communiquer à une capacité C_j. Si $C_{max} = \max_j C_j$, alors

$$N^{max} = \left\lceil \frac{C_{max}}{W \log_2(1+\beta_{min})} \right\rceil.$$

En plus, le nombre de sous-canaux demandé par un utilisateur à x est aussi une fonction de j, et nous avons :

$$N(x,\, j,\, y) = \left\lceil \frac{C_j}{W \log_2\left(1 + \dfrac{P_t K}{\eta y \|x\|^\gamma}\right)} \right\rceil.$$

Le résultat le plus générique de ce chapitre (avec un gain aléatoire, et des classes d'utilisateurs différentes) est le suivant :

$$\mathbf{P}_\lambda \left(\int N \ \mathrm{d}\omega \geq \alpha m_N \right) \leq \exp\left(-\frac{v}{N_{max}^2} \ g\left(\frac{(\alpha-1) m N_{max}}{v} \right) \right),$$

où

$$m = \sum_{j=1}^{M} \int N(x,\, j,\, y) \lambda_j(x) p_S(y) \ \mathrm{d}x \ \mathrm{d}y,$$

et

$$v = \sum_{j=1}^{M} \int N(x,\, j,\, y)^2 \lambda_j(x) p_S(y)\ \mathrm{d}x\ \mathrm{d}y.$$

Soit $N_{0_{sup}}$ le nombre de sous-canaux utilisé pour avoir une probabilité de perte P_{sup} et $N_{0_{sim}}$ le nombre de sous-canaux utilisé pour avoir la probabilité de perte si nous simulons le réseau. Si nous utilisons le résultats de ce chapitre, nous observons qu'à partir de la table 4.3, $N_{0_{sup}}$ est approximativement 20% plus grand que $N_{0_{sim}}$, ce qui montre que l'usage de la borne supérieure pour le projet du système est réalisable.

Chapitre 5 : Topologie algébrique

La topologie algébrique est utilisée dans ce travail afin d'extraire certaines propriétés des réseaux à partir d'informations données par des éléments discrets qui créent ces réseaux. Les objects topologiques utilisés pour la modélisation des réseaux de capteurs sont les complexes simpliciaux. Tandis que les graphes représentent des relations binaires, les complexes simpliciaux représentent des relations d'ordre supérieur.

Étant donné un ensemble de points V, un k-simplexe est un sous-ensemble non-ordonné $\{v_0, v_1, \cdots, v_k\}$ où $v_i \in V$ et $v_i \neq v_j$ pour tout $i \neq j$. Les faces du k-simplexe $\{v_0, v_1, \cdots, v_k\}$ sont définies comme tous les $(k-1)$-simplexes de la forme $\{v_0, \cdots, v_{j-1}, v_{j+1}, \cdots, v_k\}$ où $0 \leq j \leq k$. Un complexe simplicial est une collection de simplexes fermée par rapport à l'inclusion de faces, i.e., si $\{v_0, v_1, \cdots, v_k\}$ est un k-simplexe, alors toutes leurs faces sont inclues dans l'ensemble de $(k-1)$-simplexes.

Étant donné $\mathcal{U} = (U_v, v \in \mathfrak{T})$ une collection d'ensembles ouverts le complexe de Čech de $\mathcal{U}$ appelé $\mathcal{C}(\mathcal{U})$ est le complexe simplicial abstrait où k-simplexes correspondent à $(k+1)$-tuples d'éléments distincts de $\mathcal{U}$ qui n'ont pas d'intersection vide, c'est-à-dire que $\{v_0, v_1, \cdots, v_k\}$ est un k-simplexe si et seulement si $\bigcap_{i=0}^{k} U_{v_k} \neq \emptyset$.

On peut définir l'orientation d'un simplexe en définissant un ordre dans leurs sommets. Un changement d'orientation correspond à un changement de signe des coefficients tel que

$$[v_0, \cdots, v_i, \cdots, v_j, \cdots, v_k] = -[v_0, \cdots, v_j, \cdots, v_i, \cdots, v_k].$$

Soit X un complexe simplicial. Pour chaque entier k, $C_k(X)$ est l'espace vectoriel engendré par l'ensemble de k-simplexes de X. L'opérateur appelé *boundary map* ∂_k est défini pour être une transformation linéaire $\partial_k : C_k \to C_{k-1}$ qui a comme domaine les éléments de la base de C_k, par exemple $[v_0, \cdots, v_k]$, via

$$\partial_k[v_0, \cdots, v_k] = \sum_{i=0}^{k} (-1)^k [v_0, \cdots, v_{i-1}, v_{i+1}, \cdots, v_k].$$

Cet opérateur permet la construction d'un complexe enchaîné : une suite d'espaces vectoriels et des transformations linéaires

$$\cdots \xrightarrow{\partial_{k+2}} C_{k+1}(X) \xrightarrow{\partial_{k+1}} C_k(X) \xrightarrow{\partial_k} C_{k-1}(X) \cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X).$$

Il est possible de montrer que pour quel que soit k entier positif, k,

$$\partial_k \circ \partial_{k+1} = 0.$$

Nous définissons $\ker \partial_k$ comme le noyau de ∂_k sur C_k (i.e., tous les cycles qui sont fermés), $\mathrm{im}\, \partial_k$ comme l'image de ∂_k (i.e., les $k-1$-simplexes qui sont déjà des faces de k-simplexes). Et si nous définissons

$$Z_k = \ker \partial_k \text{ and } B_k = \mathrm{im}\partial_{k+1},$$

cela induit que $B_k \subset Z_k$.

L'homologie k-dimensionelle de X, $H_k(X)$, est l'espace vectoriel quotient

$$H_k(X) = \frac{Z_k(X)}{B_k(X)}.$$

et le k-ème nombre de Betti de X est sa dimension

$$\beta_k = \dim H_k = \dim Z_k - \dim B_k.$$

L'invariant topologique nommé caractéristique d'Euler pour X, $\chi(X)$, est un entier défini par :

$$\chi(X) = \sum_{i=0}^{\infty} (-1)^i \beta_i.$$

Dans la thèse, s_k représente le nombre de k-simplexes dans un complexe simplicial X et il est connu que :

$$\chi(X) = \sum_{i=0}^{\infty} (-1)^i s_i.$$

Nous appelons aussi $\mathbb{T}_a^d$ le tore d-dimensionel de coté a.

Le complexe de Rips-Vietoris dans $\mathbb{T}_a^d$ est défini de la façon suivante : pour chaque $\epsilon > 0$, le complexe de Rips-Vietoris de ω, $\mathcal{R}_\epsilon(\omega)$ est le complexe simplicial abstrait où chaque k-simplexe correspond à $(k+1)$-tuples non-ordonnés de points dans ω qui sont deux à deux plus proche que ϵ. Nous démontrons que pour un tore $\mathbb{T}_a^d$ et pour la norme produit d_∞, $\mathcal{R}_\epsilon(\omega)$ a le même type d'homotopie que le complexe de Čech $\mathcal{C}_{2\epsilon}(\omega)$. Nous définissons aussi la fonction $h(v_1, \cdots, v_k)$ comme étant :

$$\begin{aligned} h(v_1, \cdots, v_k) &= h_k(v_1, \cdots, v_k) \\ &= \prod_{1 \le i < j \le k} \mathbf{1}_{[\|v_i - v_j\| < 2\epsilon]}, \end{aligned}$$

qui détermine si un ensemble de k points distincts ordonnés forment un $(k-1)$-simplexe $(h = 1)$ ou non $(h = 0)$.

Soit $\omega \in \mathbb{T}_a^d$ un ensemble de points qui forment le complexe simplicial $\mathcal{C}_\epsilon(\omega)$. Les trois propositions sont démontrées dans la thèse :

- Si $i > d$, $\beta_i(\omega) = 0$;
- Il n'y a que deux valeurs possibles pour le d-ème nombre de Betti de $\mathcal{C}_\epsilon(\omega)$:

  i) $\beta_d = 0$, or

  ii) $\beta_d = 1$.

  Si $\beta_d = 1$, on a aussi $\chi(\mathcal{C}_\epsilon(\omega)) = 0$.
- Soit X un sous-ensemble compact de $\mathbf{R}^d$ et $\tau : X \to Y$ où $x_i = k y_i$ pour $x_i \in X$, $y_i \in Y$ et k une constante réelle positive. Appelons $\tau_* \omega$ la mesure image de ω par τ, i.e., $\tau_* : \Omega^X \to \Omega^Y$ est l'application

  $$\omega = \sum_{i=1}^{\infty} \delta(x_i) \quad \text{vers} \quad \tau_* \sum_{i=1}^{\infty} \delta(k x_i).$$

  L'application $\tau_* : \Omega^X \to \Omega^Y$ établit une relation de la mesure de Poisson λ sur Ω^X vers la mesure de Poisson $\lambda_\tau = \lambda / k^d$ sur Ω^Y. En plus, si ϵ_τ est la distance en Y tel que deux points sont connectés, les homologies des deux complexes simpliciaux $\mathcal{C}_\epsilon(\omega)_{\omega \in \mathbb{T}^d_{[a]}}$ et $\mathcal{C}_{\epsilon_\tau}(\tau_* \omega)_{\tau_* \omega \in \mathbb{T}^d_{[ak]}}$ sont égales pour quelque soit le k si $\lambda_\tau = \lambda / k^d$ et $\epsilon_\tau = k\epsilon$.

Applications aux réseaux de capteurs

Le but de ce chapitre est d'arriver à faire le lien entre la topologie et les réseaux de capteurs. Le nombre de k-simplexes montre par lui-même des tendances dans les réseaux : si deux réseaux ont les mêmes paramètres, il est plus probable que celui qui a plus de 1-simplexes soit connecté. De la même façon, il est plus probable que celui qui a plus de 2-simplexes ait une couverture plus large que l'autre. D'une façon plus sophistiquée, les nombres de Betti mesurent directement le nombre de composants connexes (par β_0), le nombre de trous en deux dimensions (par β_1), le nombre de vides en trois dimensions (par β_3), etc.

Nous avons aussi une interprétation de la caracteristique d'Euler : $\chi = 0$ est une condition nécessaire pour que le tore soit couvert et $\beta_d = 1$ est une condition nécessaire et suffisante. Cela nous permet aussi d'évaluer la couverture en $[0, a]^d$ (i.e., nous ne prenons plus le tore), si nous faisons attention aux effets de bord. Par exemple, $\beta_d = 1$ est une condition suffisante pour la couverture de $[\epsilon, a - \epsilon]^d$.

Chapitre 6 : Les moments des k-simplexes et de la caractéristique d'Euler

Les résultats des chapitres 2 et 5 sont appliqués dans le chapitre 6. Dans le modèle physique des réseaux de capteurs proposé, chaque capteur représente un point et possède un rayon de couverture ϵ. Nous considérons aussi que les points sont ceux d'un processus ponctuel de Poisson et que la norme utilisée pour vérifier si deux points sont proches l'un de l'autre est la norme produit. En construisant le complexe de Čech à partir des points et de ϵ, nous avons, par conséquent, une représentation de la couverture du réseau par un complexe simpliciel. Comme démontré dans le chapitre précédent, la couverture de ce complexe est la même que celle des unions des couvertures. Ce fait permet de substituer un problème algébrique à un problème combinatoire. Finalement, nous supposons que ces points tombent dans un tore de d dimensions, $\mathbb{T}_a^d$, de coté a, ce qui nous permet d'éviter les effets de bord.

La méthode utilisée est la suivante : nous exprimons le nombre de k-simplexes (sommets, arrêtes, triangles, tétrahèdres, etc.) comme des intégrales itérées d'un processus de Poisson. Les calculs des moyennes se réduisent aux calculs des intégrales déterministes grâce à formule de Campbell. En utilisant la définition de la caractéristique d'Euler, nous pouvons également trouver son espérance. Pour les moments d'ordre supérieur, nous exprimons le nombre de simplexes par une somme de chaos et à partir de cela, nous utilisons la formule de multiplication de chaos. Nous avons aussi établi que la distribution du nombre de sous-complexes dans un complexe simpliciel tend vers la Gaussienne, avec une vitesse de convergence maximale de $\lambda^{-1/2}$

Premiers moments

Nous considérons toujours que $\epsilon \leq a/6$. Nous démontrons que l'espérance du nombre de $(k-1)$-simplexes, $N_k(\mathcal{C}_\epsilon(\omega))$ (où simplement N_k), est donnée par :

$$\mathbf{E}_\lambda [N_k] = \frac{\lambda^k (ak(2\epsilon)^{k-1})^d}{k!}.$$

Il est possible aussi de trouver des expressions fermées pour N_2 et N_3 quand nous utilisons la norme Euclidienne :

- $\mathbf{E}_\lambda [N_2]$ pour le complexe de Rips-Vietoris ou celui de Čech :

$$\mathbf{E}_\lambda [N_2] = \frac{\pi \lambda^2 \epsilon^2 a^2}{2} ;$$

- $\mathbf{E}_\lambda [N_3]$ pour le complexe de Rips-Vietoris :

$$\mathbf{E}_\lambda [N_3] = \pi \left(\pi - \frac{3\sqrt{3}}{4} \right) \frac{\lambda^3 a^2 \epsilon^4}{6} ;$$

- $\mathbf{E}_\lambda [N_3]$ pour le complexe de Čech :

$$\mathbf{E}_\lambda [N_3] = \frac{2\lambda^k a^2 (\pi\epsilon^2)^2}{3}.$$

Si nous utilisons la dépoissonisation, nous obtenons l'espérance du nombre de simplexes $\mathbf{E}_\lambda[N_k]$ étant donné un nombre fixe de points $N_1 = n$ (quand le nombre de points est fixé, nous appelons le processus ponctuel de processus ponctuel Binomial) :

$$\mathbf{E}[N_k|N_1 = n] = \binom{n}{k} k^d \left(\frac{2\epsilon}{a}\right)^{d(k-1)}.$$

Considérons le polynomme de Bell, défini par :

$$B_n(x) = \sum_{k=0}^{n} \left\{ n \atop k \right\} x^k,$$

où n est un entier positif et $\left\{ n \atop k \right\}$ est le nombre de Stirling du deuxième type. L'espérance de la caractéristique d'Euler est donnée par :

$$\mathbf{E}_\lambda[\chi] = \left(\frac{a}{2\epsilon}\right)^d e^{-\lambda(2\epsilon)^d}(-B_d(-\lambda(2\epsilon)^d)).$$

Si nous utilisons encore la dépoissonisation dans ce résultat, l'espérance de χ quand les points font partie d'un processus ponctuel Binomial est donnée par :

$$\mathbf{E}[\chi|N_1 = n] = \sum_{k=0}^{n} \binom{n}{k} k^d \left(\frac{2\epsilon}{a}\right)^{d(k-1)},$$

Ensuite, nous démontrons deux théorèmes qui nous permettent d'avoir des intuitions par rapport au comportement des β_k :

- La fonction $(\lambda \mapsto \mathbf{E}_\lambda[\chi\mathcal{C}_\epsilon(\omega)])$ a exactement d racines réelles non-négatives. De plus, entre deux racines consécutives et après la plus grande, il y a exactement un point critique.

- Les nombres de Betti de la couverture convergent en probabilité vers les nombres de Betti dans le tore quand λ tend vers l'infini :

$$\mathbf{P}_\lambda\left(\bigcap_{i=0}^{d}\left(\beta_i(\mathcal{C}_\epsilon) = \beta_i(\mathbb{T}^d_{[a]})\right)\right) \xrightarrow{\lambda\to\infty} 1.$$

A partir des propriétés obtenues pour l'espérance de la caractéristique d'Euler et basées sur des simulations, nous conjecturons que, dans un complexe simpliciel aléatoire tel qu'il est décrit dans ce chapitre, il y a toujours deux types de trous dominants en fonction de λ, ϵ et a : β_i et β_{i+1}. Les autres nombres de Betti sont très peu importants par rapport à ceux qui sont dominants.

Nous appliquons aussi les inégalités de concentration afin de trouver un comportement pour la queue de la distribution du nombre de composants connexes. Ce comportement valable pour $y > \lambda a^d$:

$$\mathbf{P}_\lambda(\beta_0 \geq y) \leq \exp\left(-\frac{y - \lambda a^d}{2}\log\left(1 + \frac{y - \lambda a^d}{(2^d - 1)^2\lambda}\right)\right).$$

Les moments de deuxième ordre

Pour les moments d'ordre plus grand qu'un, nous utilisons les formules du calcul de Malliavin. Pour cela, nous représentons N_k dans la façon suivante :

$$N_k = \frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda^{k-i}I_i\left(\int_{(\mathbb{T}_a^d)^i} h(x_1,\ldots,x_k)\ \mathrm{d}x_1\ldots\ \mathrm{d}x_{k-i}\right).$$

Si $\epsilon \leq a/6$ et si nous utilisons la formule de produit des chaos, la covariance entre le nombre de $(k-1)$-simplexes, N_k, et le nombre de $(l-1)$-simplexes, N_l, pour $l \leq k$ est donnée par :

$$\mathrm{Cov}_\lambda\left[N_k, N_l\right] = \sum_{i=0}^{l-1}\frac{1}{i!(k-l+i)!(l-i)!}(\lambda(2\epsilon)^d)^{k+i}\left(\frac{a}{2\epsilon}\right)^d\left(k+i+2\frac{i(k-l+i)}{l-i+1}\right)^d.$$

Il est possible de trouver la variance de N_2 et N_3 en prenant la norme euclidienne :

- Pour le Rips-Vietoris complexe :

$$\mathbf{V}_\lambda\left[N_2\right] = \left(\frac{a}{2\epsilon}\right)^2\left(\frac{\pi}{2}(4\lambda\epsilon^2)^2 + \pi^2(4\lambda\epsilon^2)^3\right),$$

- Pour le Rips-Vietoris complexe :

$$\mathbf{V}_\lambda\left[N_3\right] = \left(\frac{a}{2\epsilon}\right)^2\left((4\lambda\epsilon)^3\frac{\pi}{6}\left(\pi - \frac{3\sqrt{3}}{4}\right) + (4\lambda\epsilon^2)^4\pi\left(\frac{\pi^2}{2} - \frac{5}{12} - \frac{\pi\sqrt{3}}{2}\right)\right.$$
$$\left. +(4\lambda\epsilon^2)^5\frac{\pi^2}{4}\left(\pi - \frac{3\sqrt{3}}{4}\right)^2\right).$$

Rappelons aussi que nous pouvons obtenir ce résultat avec un nombre fixe de points, en faisant la dépoissonisation.

Les variances et covariances des nombres de simplexes nous permettent aussi de trouver la variance de la caractéristique d'Euler :

$$\mathbf{V}_\lambda\left[\chi\right] = \left(\frac{a}{2\epsilon}\right)^d\sum_{n=1}^{\infty}c_n^d(\lambda(2\epsilon)^d)^n,$$

où

$$c_n^d = \sum_{j=\lceil(n+1)/2\rceil}^{n}\left[2\sum_{i=n-j+1}^{j}\frac{(-1)^{i+j}}{(n-j)!(n-i)!(i+j-n)!}\left(n+\frac{2(n-i)(n-j)}{1+i+j-n}\right)^d\right.$$
$$\left. -\frac{1}{(n-j)!^2(2j-n)!}\left(n+\frac{2(n-j)^2}{1+2j-n}\right)^d\right].$$

Nous avons trouvé l'expression simplifiée de la variance de la caractéristique d'Euler après beaucoup d'algébrisme :

$$\mathbf{V}_\lambda\left[\chi\right] = a\left(\lambda e^{-2\lambda\epsilon} - 4\lambda^2\epsilon e^{-4\lambda\epsilon}\right),$$

pourtant, nous n'arrivons pas à trouver des expressions de la variance de la caractéristique d'Euler pour des dimensions plus grandes.

Enfin, nous appliquons des inégalités de concentration pour trouver une borne pour la queue de la distribution de χ :

$$\mathbf{P}(\chi - \bar{\chi} \geq x) \leq \exp\left(-\frac{x}{4}\log\left(1 + \frac{2x}{\mathbf{V}_\lambda[\chi]}\right)\right).$$

Les moments de n-ème ordre

Pour simplifier les calculs, mais sans perte en généralité, nous choisissons $k = 1/2\epsilon$, alors $\lambda_\tau = \lambda(2\epsilon)^d$, $\epsilon_\tau = 1/2$ et $ak = a/2\epsilon$. Nous calculons les moments centraux, définis par $\tilde{N}_k = N_k - \bar{N}_k$. De plus, nous disons que $\binom{i}{j} = 0$ si $i \leq 0$ où $j \leq 0$ où $i - j \leq 0$ pour i et j des entiers positifs.

Avant de trouver le n-ème moment pour le cas le plus générique, nous montrons comment utiliser la méthode pour le troisième moment. Soient $\mathcal{C}_1$, $\mathcal{C}_2$ et $\mathcal{C}_3$ trois simplexes qui ont des sommets en commun. Pour $L \in \mathcal{P}(\{1,2,3\})$, nous appélons m_L le nombre de sommets appartenant à la liste d'exactement L simplexes.

Alors $M = m_{123} + m_{12} + m_{13} + m_{23} + m_1 + m_2 + m_3$ est le nombre total de sommets et $\mathcal{J}_3$ représente l'intégrale dans ces trois simplexes :

$$\mathcal{J}_3 = \int_{\Delta_{p_1}}\int_{\Delta_{p_2}}\int_{\Delta_{p_3}} h_{p_1}h_{p_2}h_{p_3}\,\mathrm{d}x_1 \dots\,\mathrm{d}x_M.$$

p_i étant le nombre de sommets du simplexe $\mathcal{C}_i$ pour $i = 1, \cdots, 3$, par exemple, $p_1 = m_{123} + m_{12} + m_{13} + m_1$, et $x_1, \cdots, x_M$ étant les M sommets. De plus, $\mathcal{J}_3(i,j,s,t)$ est l'intégrale définie ci-dessus telle que

- $m_{123} = 2t - i - j + s \vee 0$
- $m_{12} = i + j - s - t \vee 0$
- $m_{13} = i - t \vee 0$
- $m_{23} = j - t \vee 0$
- $m_1 = k - i \vee 0$
- $m_2 = k - j \vee 0$
- $m_3 = k - s \vee 0$.

Avec ces définitions, nous pouvons obtenir le troisième moment centré du nombre de $(k-1)$-simplexes :

$$\mathbf{E}_\lambda\left[\widetilde{N_k}^3\right] = \sum_{i,j,s,t} \lambda^{3k-i-j} t! \binom{k}{i}\binom{k}{j}\binom{k}{s}\binom{i}{t}\binom{j}{t}\binom{t}{i+j-s-t}\mathcal{J}_3(i,j,s,t).$$

Pour trouver une expression pour le n-ème moment, nous procédons de façon de la même manière. Soient $\mathcal{C}_1, \cdots, \mathcal{C}_n$, n simplexes qui partagent quelques sommets. Pour $L \in \mathcal{P}(\{1, \cdots, n\})$, nous appelons m_L le nombre de sommets appartenant à la liste

d'exactement L simplexes. Donc $M = \sum_{L\in\mathcal{P}(\{1,\cdots,n\})} m_L$ est le nombre total de sommets et $\mathcal{J}_n$ représent l'intégrale dans ces n simplexes:

$$\mathcal{J}_n = \int_{\Delta_{p_1}} \cdots \int_{\Delta_{pn}} h_{p_1} \dots h_{p_n} \, \mathrm{d}x_1 \dots \, \mathrm{d}x_M.$$

p_i étant le nombre de sommets du simplexe $\mathcal{C}_i$ pour $i = 1, \cdots, n$, et $x_1, \cdots, x_M$ les M sommets.

L'expression de la n-ème puissance en chaos des $(k-1)$-simplexes est donnée par :

$$\widetilde{N}_k^n = \sum_{i_1,\cdots,i_n} \sum_{s_1,\cdots s_{n-2}} \sum_{t_1,\cdots t_{n-2}} \left(\prod_{j=1}^{n-2} t_j! \binom{m_{j,1}}{t_j} \binom{m_{j,2}}{t_j} \binom{t_j}{u_j - t_j} \right) I_a(\circ_{j\in A} f_{i_j}) I_b(\circ_{j\in \overline{A}} f_{i_j}),$$

où, pour $j \in \{1, \cdots, n-2\}$:

- $1 \le i_1, \cdots, i_n \le k$,
- $s_j \ge |m_{j,1} - m_{j,2}|$,
- $m_{j,1} = i_{2j-1}$ si $1 \le j \le \lfloor \frac{n}{2} \rfloor$ et $s_{2(j-\lfloor \frac{n}{2} \rfloor)-1}$ sinon,
- $m_{j,2} = i_{2j}$ si $1 \le j \le \lfloor \frac{n}{2} \rfloor$ et $s_{2(j-\lfloor \frac{n}{2} \rfloor)}$ sinon,
- $u_j = m_{j,1} + m_{j,2} - s_j$,
- $A \subset \{1, \cdots, n\}$,
- Si n si pair, alors $a = s_{n-3}$ et $b = s_{n-2}$,
- Si n si impair, alors $a = s_{n-2}$ et $b = i_n$.

Avec ce résultat, après l'application du calcul de Malliavin, nous trouvons l'expression du n-ème moment des $(k-1)$-simplexes :

$$\mathbf{E}_\lambda \left[\widetilde{N_k}^n \right] = \sum_{i_1,\cdots,i_n} \sum_{s_1,\cdots,s_{n-3}} \sum_{t_1,\cdots,t_{n-2}} \lambda^{nk+c} \left(\prod_{j=1}^{n} \lambda^{-i_j} \binom{k}{i_j} \right) \left(\prod_{j=1}^{n-2} t_j! \binom{m_{j,1}}{t_j} \binom{m_{j,2}}{t_j} \binom{t_j}{u_j - t_j} \right) \mathcal{J}_n(i_1, \cdots, i_n, s_1, \cdots, s_{n-3}, t_1, \cdots, t_{n-2}).$$

Pour $j \in \{1, \cdots, n-2\}$:

- si $j \le n-3$, $s_j \ge |m_{j,1} - m_{j,2}|$,
- $m_{j,1} = i_{2j-1}$ si $1 \le j \le \lfloor \frac{n}{2} \rfloor$ et $s_{2(j-\lfloor \frac{n}{2} \rfloor)-1}$ sinon,
- $m_{j,2} = i_{2j}$ si $1 \le j \le \lfloor \frac{n}{2} \rfloor$ et $s_{2(j-\lfloor \frac{n}{2} \rfloor)}$ sinon,
- $m_{j,3} = s_j$ si $1 \le j \le n-3$ et s_{n-3} sinon,
- $u_j = m_{j,1} + m_{j,2} - m_{j,3}$,
- si n est pair, alors $c = s_{n-3}$ et $s_{n-3} \ge |m_{n-2,1} - m_{n-2,2}| \vee |m_{n-3,1} - m_{n-3,2}|$,
- si n est impair, alors $c = i_n$ et $i_n \ge |m_{n-2,1} - m_{n-2,2}|$.

La convergence

Soient Γ un sous-complexe simplicial arbitraire qui contient n points et $C_\epsilon(\omega)$ le complexe simplicial aléatoire généré par le processus ponctuel de Poisson ω. Le nombre d'occurrences de Γ dans $C_\epsilon(\omega)$ est défini comme $G_\Gamma(\omega)$. Rappelons qu'avec cette construction de complexe simplicial, un complexe simplicial Γ apparaît en $C_\epsilon(\omega)$ si les arrêtes de Γ sont dans $C_\epsilon(\omega)$. L'ensemble d'arrêtes de Γ, appelé par J_Γ, est un sous-ensemble de $\{1, \cdots, n\} \times \{1, \cdots, n\}$. Soit aussi

$$\widetilde{h}(x_1, \cdots, x_n) = \frac{1}{c_\Gamma} \prod_{(i,j)\in J_\Gamma} \mathbf{1}_{[\|x_i - x_j\| \leq \epsilon]},$$

où c_Γ est le nombre de permutations de $\{x_1, \cdots, x_n\}$ tel que

$$\widetilde{h^\Gamma}(x_1, \cdots, x_n) = \widetilde{h^\Gamma}(x_{\sigma(1)}, \cdots, x_{\sigma(n)}),$$

et $f^\Gamma(x_1, \cdots, x_n)$ la symétrisation de $\widetilde{h^\Gamma}(x_1, \cdots, x_n)$. Alors, nous avons :

$$G_\Gamma = \sum_{\substack{x_1,\cdots,x_n\in\omega \\ x_i\neq x_j if i\neq j}} f^\Gamma(x_1, \cdots, x_n) = \int_{\Delta_n} f^\Gamma(x_1, \cdots, x_n)\ \mathrm{d}\omega(x_1) \cdots\ \mathrm{d}\omega(x_n).$$

Avec cette définition, nous obtenons le théorème de convergence le plus important de cette section. Ce théorème établi qu'il y a une constante c_Γ telle que, pour λ assez grand, la distance de Wasserstein entre $F = \frac{G_\Gamma - \mathbf{E}_\lambda[G_\Gamma]}{\sqrt{Var(G_\Gamma)}}$ et $\mathcal{N}(0,1)$ est donnée par :

$$d_W(F, \mathcal{N}(0,1)) \leq \frac{c}{\lambda^{1/2}}.$$

Chapitre 7 : Le cas unidimensionnel

Le chapitre précédent traite du problème des réseaux de capteurs en d dimensions. Bien que nous ayons trouvé plusieurs résultats, il y a quelques questions qui n'ont pas été résolues, par exemple, celle de la moyenne du nombre de composants connexes. Nous avons réalisé que ces questions pourraient être résolues dans un cas plus restreint, quand les points tombent sur une droite ou sur un cercle. Dans ces cas-là, nous n'avons pas besoin d'utiliser les outils de la topologie algébrique.

Formulation du problème

Soit $L > 0$, nous supposons que le processus ponctuel de Poisson N d'intensité λ est sur $[0, L]$. Les atomes de N sont donnés par $(X_i, i \geq 1)$. Donc, les variables aléatoires $\Delta X_i = X_{i+1} - X_i$ sont i.i.d. et exponnentialment distribuées. Nous fixons $\epsilon > 0$. Nous disons que deux points placés respectivement sur x et y, sont *directement connectés* chaque fois que $|x - y| \leq \epsilon$. Pour $i < j$, deux points de N, nous disons que X_i et X_j sont *indirectement connectés* si X_l et X_{l+1} sont directement connectés pour quelque soit $l = i, \cdots, j-1$. Un ensemble de points connectés, directement où indirectement, est appelé un *cluster*, et *cluster* complet est celui qui commence et finit dans $[0, L]$. La connectivité du réseau est mesurée par le nombre de *clusters*.

Le nombre de points dans l'intervalle $[0, x]$ est défini par $N_x = \sum_{n=0}^{\infty} \mathbf{1}_{[X_n \leq x]}$. La variable aléatoire A_i donnée par

$$A_i = \begin{cases} X_1 & \text{si } i = 1, \\ \inf\{X_j \,:\, X_j > A_{i-1},\, X_j - X_{j-1} > \epsilon\} & \text{si } i > 1, \end{cases}$$

représente le début du i-ème *cluster*, appelé par C_i. De la même manière, la fin d'un *cluster* est définée par

$$E_i = \inf\{X_j + \epsilon \,:\, X_j > A_i, X_{j+1} - X_j > \epsilon\}.$$

Donc le i-ème *cluster*, C_i, a un nombre de points donné par $N_{E_i} - N_{A_i}$. Nous définissons aussi la longueur B_i de C_i comme $E_i - A_i$. La longueur *intercluster*, D_i, est la distance entre la fin de C_i et le début de C_{i+1}, c'est-à-dire, $D_i = A_{i+1} - E_i$, et ΔA_i est la distance entre les deux premiers points de deux *clusters* consécutifs, C_i et C_{i+1}, donnée par $\Delta A_i = A_{i+1} - A_i = B_i + D_i$. Le nombre de *clusters* complets est donné par β_0, la distance entre le début du premier *cluster* et le début du $(i+1)$-ème *cluster* est définie comme étant $U_i = \sum_{k=1}^{i} \Delta A_k$. Pour finir, nous définissons aussi $\Delta X_0 = D_0 = X_1$. Nous avons trouvé les distributions de toutes les variables aléatoires présentées ci-dessus, tout d'abord avec les transformées de Laplace pour pouvoir après les inverser et ensuite trouver les expressions analytiques.

Les transformées de laplace

Les trois transformées de Laplace les plus importantes sont celles qui suivent :

- La longueur du i-ème cluster B_i :

$$\mathbf{E}\left[e^{-sB_i}\right] = \frac{1}{\lambda} \frac{\lambda + s}{\frac{se^{\lambda\epsilon}}{\lambda} e^{s\epsilon} + 1} \;;$$

- La distance ΔA_i entre les deux premiers points de deux *clusters* consécutifs C_i et C_{i+1} :

$$\mathbf{E}\left[e^{-s\Delta A_i}\right] = \mathbf{E}\left[e^{-sB_i}\right]\mathbf{E}\left[e^{-sD_i}\right] = \frac{1}{\frac{se^{\lambda\epsilon}}{\lambda}e^{s\epsilon}+1} ;$$

- La distance entre le début du premier *cluster* et le début du $(i+1)$-ème *cluster*

$$\mathbf{E}\left[e^{-sU_n}\right] = \frac{1}{\left(\frac{e^{\lambda\epsilon}}{\lambda}se^{s\epsilon}+1\right)^n} ;$$

Maintenant, nous définissons $p_n(L) = \mathbf{P}(\beta_0 = x)$, ce qui signifie que, $p_n(L)$ est la probabilité de l'existence d'exactement n *cluster* dans l'intervalle $[0, L]$. Compte tenu que pour tout $L \in \mathbb{R}_+$, $0 \leq p_n(L) \leq 1$, la transformée de Laplace par rapport à L

$$\mathcal{L}\{p_n(\cdot)\}(s) = \int_0^\infty e^{-st}p_n(t)dt,$$

est bien définie, nous pouvons la calculer en utilisant l'expression de $\mathbf{E}\left[e^{-sU_n}\right]$:

$$\mathcal{L}\{p_n(\cdot)\}(s) = \frac{e^{\epsilon\lambda}}{\lambda}\frac{e^{\epsilon s}}{\left(\frac{e^{\epsilon\lambda}}{\lambda}se^{\epsilon s}+1\right)^{n+1}}.$$

Nous avons aussi trouvé la transformée de Laplace du m-ème moment de β_0, $M^m_{\beta_0}(L)$, qui sera utilisée pour trouver la transformée inverse de $\mathcal{L}\{p_n(\cdot)\}(s)$:

$$\mathcal{L}\left\{M^m_{\beta_0}(\cdot)\right\}(s) = \frac{a}{s\,(a+1)}\mathrm{Li}_{-m}\left(\frac{1}{a+1}\right),$$

où $\mathrm{Li}_t(z)$, $z, t \in \mathbb{R}$, $z < 1$, est la fonction polylogarithmique qui a t comme paramètre, et qui esr définie par

$$\mathrm{Li}_t(z) \triangleq \sum_{k=1}^{\infty}\frac{z^k}{k^t}.$$

Les expressions analytiques

Dans cette section, nous inversons les tranformées de Laplace trouvées dans la dernière section. D'abord, nous obtenons l'expression pour les moments de β_0 :

$$M^m_{\beta_0}(L) = \sum_{k=1}^{m}\left\{{m \atop k}\right\}\left(\frac{L}{\epsilon}-k\right)^k\left(\lambda\epsilon e^{-\epsilon\lambda}\right)^k\mathbf{1}_{[L/\epsilon>k]}.$$

$\left\{{m \atop k}\right\}$ étant le nombre de Stirling du deuxième type. Nous utilisons donc ce résultat, donc, pour trouver l'expression de la distribution de β_0, $\mathbf{P}(\beta_0 = n)$:

$$\mathbf{P}(\beta_0 = n) = \frac{1}{n!}\sum_{i=0}^{\lfloor L/\epsilon\rfloor - n}\frac{(-1)^i}{i!}((L-(n+i)\epsilon)\lambda e^{-\lambda\epsilon})^{n+i}.$$

Cet inversion de la transformée de Laplace nous révèle un pair de transfomée entre les domaines x et s donnée par :

$$\frac{\mathbf{1}_{[x\geq 0]}}{n!}\sum_{i=0}^{\lfloor x/\epsilon\rfloor-n}\frac{(-1)^i}{i!}\left((x-(n+i)\epsilon)\frac{1}{a}\right)^{n+i}\overset{\mathcal{L}}{\Leftrightarrow}\frac{ae^{\epsilon s}}{(ase^{\epsilon s}+1)^{n+1}}.$$

Une fois que nous avons cette transformée, après l'usage de quelques proprietés fondamentales de la transformée de Laplace, nous pouvons obtenir les distributions de B_1 et de U_n, respectivement $f_{B_i}(x)$ et $f_{U_n}(x)$:

$$f_{B_i}(x)=\left[\lambda e^{-\epsilon\lambda}p_0(x-\epsilon)+e^{-\epsilon\lambda}\frac{d}{dx}p_0(x-\epsilon)\right]\mathbf{1}_{[x>\epsilon]},$$

et

$$f_{U_n}(x)=\lambda e^{-\epsilon\lambda}p_{n-1}(x-\epsilon)\mathbf{1}_{[x>\epsilon]},$$

où $p_n(x)=\mathbf{P}(\beta_0=n|L=x)$.

Nous pouvons aussi obtenir la probabilité que $[0,L]$ soit entièrement couvert par les capteurs. Soit $R_{m,n}(x)$ défini comme ci-dessous :

$$R_{m,n}(x)=\sum_{i=m}^{\lfloor x/\epsilon\rfloor-1}\left[\left(e^{-\lambda\epsilon}\right)^{i+n}\sum_{j=0}^{i+n}\frac{(\lambda[(1-i)\epsilon-x])^j}{j!}\right].$$

Alors

$$\begin{aligned}\mathbf{P}([0,L]\text{ soit couvert})=R_{0,1}(L)-e^{-\lambda\epsilon}R_{0,1}(L-\epsilon)\\ -e^{-\lambda\epsilon}R_{1,0}(L)+e^{-2\lambda\epsilon}R_{1,0}(L-\epsilon).\end{aligned}$$

Le problème de connectivité des réseaux de capteurs aléatoires est donc entièrement résolu: nous avons trouvé des expressions analytiques simples pour représenter les distributions de β_0, B_i, U_n de même que nous avon trouvé la probabilité que toute la région soit couverte, tout cela en fonction de L, ϵ, λ_0, μ et t (nous rappelons que $\lambda=\lambda_0 Le^{-t/\mu}$).

Cas particuliers

Nous avons résolu le problème de connectivité pour le cas où nous avions un nombre de *clusters* complets dans l'intervalle $[0,L]$. Nous pouvons aussi considérer deux cas légèrement différents. Dans le premier cas, nous comptons le nombre de *clusters* complets ou incomplets dans un intervalle, que nous appelons β_0'. Nous obtenons l'expression suivante pour cette quantité :

$$\mathbf{P}(\beta_0'=n)=\sum_{i=n}^{\lfloor L/\epsilon\rfloor+1}(-1)^{i+n}\binom{i}{n}(G(i-1)+G(i)),\text{ for } n\geq 0,$$

où

$$G(k)=(-1)^k\left(e^{-k\lambda\epsilon}\sum_{j=0}^{k}\frac{[\lambda(k\epsilon-L)]^j}{j!}-e^{-\lambda L}\right)\mathbf{1}_{[T>k\epsilon]}.$$

Le deuxième cas particulier est celui où les points sont distribués sur un cercle de rayon $L/2\pi$ et non plus sur une droite. Nous appelons χ le nombre de *clusters* dans ce cercle pour obtenir la distribution de χ :

$$\mathbf{P}(\chi = n) = e^{-\lambda L}\mathbf{1}_{[n=0]} + (1 - e^{-\lambda L})\frac{\lambda e^{-\epsilon\lambda}}{n!} \sum_{i=0}^{\lfloor L/\epsilon \rfloor - n} \left[\frac{(-1)^i}{i!} \right.$$
$$\left. ([L - (n+i)\epsilon]\lambda e^{-\epsilon\lambda})^{n+i-1} \left(L + (n+i)\left(\frac{1}{\lambda} - \epsilon\right)\right)\right].$$

Chapitre 8 : Conclusion

La contribution la plus importante de la thèse consiste en l'application simultanée de résultats de la topologie algébrique et de l'analyse stochastique aux réseaux de capteurs. Les outils du calcul de Malliavin ont été particulièrement utiles et la topologie a permis la création d'un lien entre le problème algébrique et le problème vu d'une façon combinatoire. Nous avons pu utiliser des outils de la mathématique moderne pour traiter quelques problèmes où l'usage des outils classiques a échoué. Nous remarquons aussi que nous avons obtenu des résultats qui ont des applications à la fin de chaque chapitre.

Abstract

Lower costs, lower power consumption, and hardware miniaturization have lead to a great increase in the use of sensor networks. New mathematical tools have enabled the conception of new models for sensor networks: two of these tools are stochastic analysis and algebraic topology. In stochastic analysis, as introduced by Baccelli in the 90s, deployment of mobiles and sensors in a nework is treated as a Poisson point process, instead of the deterministic behavior that was used before. Ten years later, Ghrist has started modelling the coverage of sensor networks as a simplicial complex. As a consequence, results of algebraic topology can be applied to such networks.These results have enabled the explicit computation of the homology of the simplicial complex. Therefore, it was possible to answer relevant and non-trivial questions related to sensor networks, each sensor having minimal information about the network (its identification number and the ones of his neighbors). Examples of such questions are: "is the network completely connected?", "is the target region covered?", and "how many connected components and holes are there in this network?".

This thesis has two main parts. Part I uses stochastic anlysis to provide bounds for the overload probability of different systems thanks to concentration inequalities. Although the results are general, we apply them to real wireless network systems such as WiMax and mutliclass user traffic in an OFDMA system. In part II, we find more connections between the topology of the coverage of a sensor network and the topology of its corresponding simplicial complex. These connections highlight new aspects of Betti numbers, the number of k-simplices, and Euler characteristic. Then, we use algebraic topology in conjunction with stochastic analysis, after assuming that the positions of the sensors are points of a Poisson point process. As a consequence we obtain, in d dimensions, the statistics of the number of k-simplices and of Euler characteristic, as well as upper bounds for the distribution of Betti numbers. We also prove that the number of k-simplices tends to a Gaussian distribution as the density of sensors grows, and we specify the convergence rate. Finally, we restrict ourselves to one dimension. In this case, the problem becomes equivalent to solving a $M/M/1/1$ preemptive queue. We obtain analytical results for quantites such as the distribution of the number of connected components and the probability of complete coverage.

Contents

Chapter 1

Introduction

1.1 Motivation: sensor networks

The importance of sensor networks grew tremendously in the last few decades in every field where observation of the physical world is useful. Those systems have already been inserted in a large range of sectors, and since the knowledge in electronics and communication technology is continuously expanding, it is expected that sensor networks will be an integral part of our lives, more than the present-day personal computers [1, 18].

The progressive miniaturization and the Moore law allow the development of low-cost, low-power, multifunctional sensor nodes. These features enable the practical usage of wireless sensor networks based on collaborative effort of a large number of nodes [1, 67, 72]. As a consequence of this development, we have today a very wide range of domains of application, such as health monitoring, intelligent agriculture, environmental control, telematics, and space exploration. Overview of these and other applications are included in references [1, 31, 17, 68, 30, 98, 32].

The conception of sensor networks requires, however, a multidisciplinary expertise that motivates the research that goes beyond only the improvement of the performances of the physical layer of the system. Dealing with sensor networks has to do with the integration between the physical dimension of transmission (signal, coding), to share properly the resources, to treat mechanisms that establish and maintain the connectivity and coverage (topology) of the network, to face the dynamics of the sensors (positioning, mobility and sensors default) and to model such systems. Moreover, these solutions must take into account the constraints of limited resources and low consummation [37].

In view of these particular features, sensor networks require new paradigms for communications: we need new tools (theories, heuristics, designs) conceived specifically for them. The topology of the networks is one important subject studied in this thesis, which includes connectivity and coverage. Loosely speaking, coverage describes how well sensors in the network can monitor a geographical region. The connectivity evaluates how much groups of sensors are disconnected among them.

Considering a variety of situations, it is possible to categorize three main scenarios: those where it is possible to choose the position of each sensor, perhaps best described as the 'computational geometry', e.g. in [74, 81, 115, 122], those where sensors are arbitrarily deployed in the target region with the control of a central station (see [41, 22, 107]), and those where the sensor locations are random in a decentralized system, which is analyzed, for example, in [69].

The problem of the first scenario is that, in many cases, placing the sensors is impossible.

Sometimes this impossibility comes from the fact that the cost of placing each sensor is too high and sometimes the network has an inherent random behavior (like in the ad hoc case, where users move). In addition, this policy cannot take into account the configuration of the network in the case of failure of some sensor. The drawback of the second scenario is a higher unit cost per sensor, since each one has to communicate with the central station. Besides, the central station itself increases the cost of the whole system. Moreover, if sensors are supposed to know their positions, an absolute positioning system has to be included in each sensor, making their hardware even more complex and as such, more expensive.

This has motivated us to investigate the third scenario: randomly located sensors, no central station. Actually, if we can predict some characteristics of the topology of a random network, the number of sensors (or, as well, the power supply of them) can be *a priori* determined such that a given network may operate well with high probability. For instance, we can choose the mean number of sensors such that, if they are randomly deployed, there is more than 99% probability, for the network to be completely connected.

1.2 Topology of random geometric networks

The first studies concerning random points in a space with random pairwise relations have been elaborated by Erdös and Rényi [28, 29] and Gilbert [43] in the late 50s. While Erdös and Rényi assign an equal probability to all graphs with a specific number of edges, in the first model of Gilbert every possible edge between nodes occurs to have an independent probability p. Today, the model of Gilbert is called *Bernoulli random graphs*, or *Erdös-Rényi graphs* due to their research on the model of Gilbert. Although the models used by those authors are not equivalent, none of them considered the distance between nodes in order to determinate whether two points are connected or not. This observation led to name those graphs independent random graphs (IRG). The two different ways of approach have in common that they regard the connectivity between two points as a key element. By all means those models can be applied in telecommunications when the connexion does not depend on the distance, for instance, optical networks. Besides this model can also be applied to networks of power lines, computer networks, social links, business relations between companies, to name a few. Those random graphs and many of their properties were intensely studied [15, 2, 63].

The first model of random graphs where edges depend on the position of points, the dependent random graphs (DRG), can be traced back to the work of Gilbert in 1961 [44]. In the model proposed by Gilbert, a pair of points is connected if this pair is separated by a distance smaller than some constant R. The resulting graph is called a random geometric graph (RGG). However, the dependence of points and edges of DRG's represents an important difficulty in comparison to IRG's: in the geometric setting, if the vertex V_i is close to V_j, and V_j is close to V_k, then V_i will be fairly close to V_k, that is, we do not have independence between edges any longer. On the other hand, the model of IRG's is too unrealistic to be used in wireless communication systems and the recent development of these systems has driven to intense researches of DRG's since the 90s.

Some examples of DRG's that are not RGG's are the geometric Euclidean graphs [109, 121] and the proximity graphs, where each point is connected to its k nearest neighbors [120, 11, 114]. However, these models do not capture basic properties of wireless systems: the conservation of energy imposes that the transmitted power decreases with the distance. Also the presence of noise imposes a maximum separation between source and receiver,

which drove to the use of RGG's in recent works.

The cases of finite domains and infinite domains are treated differently. In the infinite-space case, the study of RGG's is known as percolation, where the object under study is the formation of giant connected components or complete connectivity. Indeed, the first RGG studied by Gilbert was about percolation (even if this name was not used, the term *percolation* was used for the first time in 1957 by Simon Broadbent and John Hammersley [16]), where he proved that such giant components exist although. Percolation is an important branch of the modern probability theory and is intensively studied [51, 124, 125, 82, 80].

In finite domain, RGG's are usually studied to derive asymptotic behavior for large finite graphs. The first work dates back to Hafner [52]. Since then, there were several works on this subject. In the 90s, applications of graph theory to probabilistic and statistical aspects of these graphs in one dimension were studied in the work of Godehardt et al. [46, 47, 48, 57] and in the last decade we have [38, 39, 40, 55, 56], to cite a few. In higher dimensions, we have mathematical contributions from Appel and Russo [4, 5, 3] and McDiarmid [78, 77]. The book of Penrose [96] is probably the most complete reference about the asymptotics for large finite graphs, compiling and complementing a series of works starting in 1995 [88, 89, 91, 92, 90, 93, 94, 95, 97].

Beyond the work of Gilbert, several other properties of RGG's have been explored. One of the most important property for wireless communications is the connectivity. There are efforts made in order to find out if a graph is completely connected, but also in order to count the number of components if there is more than one. This property was explored by means of geometric tools until Silva, Ghrist and Muhammad, in [42, 41, 107], introduced the use of algebraic topology for models of geometric graphs. Although the authors do not apply their idea at random realizations of points, their tools allow the local-to-global transition, because it is possible to calculate the number of connected components, called β_0, with only on a list of edges and points. It means that, to apply algebraic topology to calculate β_0, it is sufficient to label the points of the process. While a graph can be defined by a list of points and edges, a simplicial complex, used by Ghrist *et al.*, is a list of points, edges, triangles, tetrahedrons and so on, which makes a graph to a particular case of a simplicial complex. As one can expect, this generalization gives much more information about the network. The first direct application is related to the coverage area in two or three dimensions.

Coverage is another widely studied subject of random realizations. In the beginning, it was not possible to point out the difference between the study of connectivity and coverage, because the research was done over circles. Many authors considered problems of covering a circle by a fixed number of arcs. In [116, 35, 24, 36, 108], the authors considered the covering when arcs have the same size. The works of Dvoretzki [25], Mandelbrot [75], Shepp [104] and Siegel and Holst [105] regard this problem when arcs are not equal. The last one provides an exact formula for the probability of covering the circle, thereby extending Stevens [110]. Intuitively, one can say that the models of these works in one dimension are coverage problems, but in a point of view of topology, these papers deal with connectivity (counting the number of non-covered gaps is the same than counting the number of connected components). In more than one dimension, one of the early and now classical coverage problems is the coverage of a sphere, proposed by Moran in [83] and solved by Gilbert in [45]. Extensions to k dimensions usually consider the asymptotics for models of fixed or Poisson distributed number of points like Hall in [53, 54], Janson in [62] and Stoyan [111]. In the case of a Poisson process, when the individual covered domain

is of constant form with a Lebesgue measure A, the coverage is a simple problem, since the probability p of covering an arbitrary point is given by $p = 1 - e^{-\lambda A}$, where λ is the intensity of the process. A natural generalization of this case is to consider a fixed number of points i.i.d. in the target region, which was studied by Roy *et al.* in [8, 103]. Other properties regarding the coverage were analyzed, for instance, the k-coverage [70] which studies the probability of a point being covered by k sensors, or the triangulation [73] which analyzes the probability of a point being inside some triangle formed by three points close from each other.

The use of topological algebraic on realizations seems to be an ideal tool to deal with the local-global transition, i.e. to obtain general aspects of coverage without knowing the exact positions of the points, such as the number of holes in two dimensions, called β_1 or voids in three dimensions, called β_2, which are important questions for wireless networks that works using only geometry were not able to do. The main motivation to the study of random geometric simplicial complexes was their application to wireless networks given by the works [42, 41, 23, 84, 22, 107]. Other connected works with persistent homology [27, 123, 26], which is a way of tracking how the homology of a sequence of spaces changes. At last, there are some works on random simplicial complexes that are not geometric, but the only ones that deal directly with geometric complexes are those from Kahle *et al.* [66, 65], where the asymptotic behavior of the number of holes and voids, called Betti numbers, are found in some particular cases.

1.3 Overview of the Problem

A sensor is a device that scans a domain and returns a signal from which information may be extracted. Sensors vary in scope, resolution and types such as seismic, low sampling, rate magnetic, thermal, visual, infrared and acoustic, which are able to monitor a wide variety of ambient conditions such as temperature, pressure, light, humidity, vehicular movement, etc. Sensor nodes can be used for continuous sensing, event detection, event identification, location sensing and local control of actuators. In this section we present the way the functionality of the sensors is related to the interpretation of their coverage, and then, we show some practical applications to illustrate the potential of those networks.

1.3.1 Coverage

The miniaturization of the sensors justifies the fact that they can be represented as points in almost all applications in modern sensor networks. However, the concept of coverage is not as clear and usually not well explained. Then we consider three different situations to determinate coverage: one where sensors are supposed to communicate, one where sensors are supposed to perform a remote control and the third one where sensors monitor some conditions in their exact location. Depending on the function of the network, sensors can have more than one kind of coverage at the same time.

Sensors communicating We first consider a sensor network where sensors upload/download data to/from other sensors. Sensors have a power supply allowing them to transmit a signal with a fixed power. It is well known that the transmitted power is attenuated over distance, an attenuation called path loss. Besides, power is affected by environmental conditions, a phenomenon modeled as shadowing. So the power received by

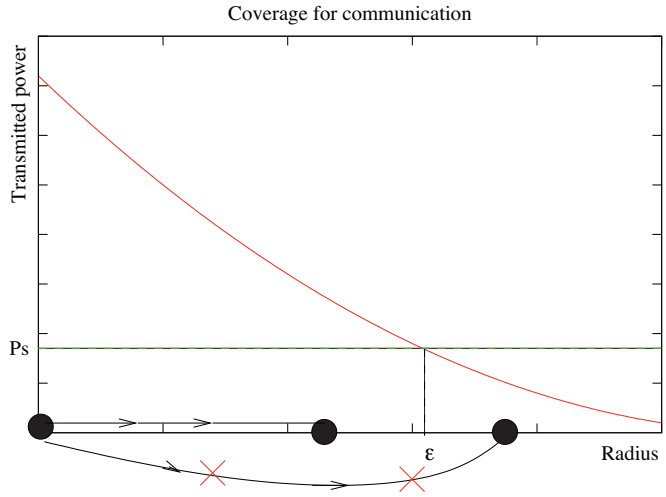
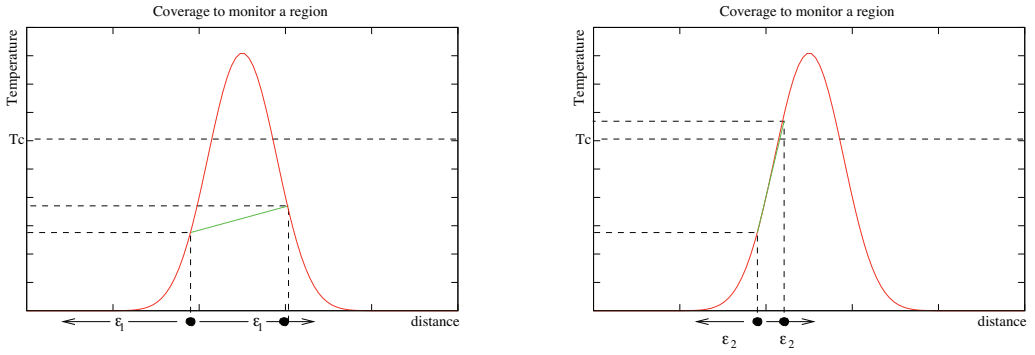


Figure 1.1: The black circles represent the positions of the sensors. The coverage radius ϵ in function of P_s under the effect of path loss only. Sensors cannot send data to their sensors further than ϵ.

a sensor is inferior to the one transmitted (except for rare events such as constructive interference due to multipath, for example). On the other hand, sensors have sensitivity, which means that they are able to identify correctly the signals with a desirable probability (with a low bit error rate) as long as the signal to noise ration is higher than some threshold, for instance, P_s. In this case, the coverage of a sensor is defined as the whole region where others sensors can receive its signal with a power higher than P_s. In Fig. 1.1 we illustrated the coverage radius ϵ in function of P_s under the effect of path loss only.

Sensors monitoring conditions in their locations In this case, the sensor network is supposed to approximate the real profile of the measured quantity based on typical gradients of this one. This means that the coverage radius have to be chosen in order to warrant that sensors separated from each other at most by this radius are still able to fit or adjust with small errors. For instance, let us suppose that sensors measure temperature and that a sensor network is supposed to detect when the temperature exceeds some temperature T_c. In Fig. 1.2, we see an example of a likely profile of the temperature gradient. In this case, we cannot choose a coverage radius ϵ_1 like the one in Fig. 1.2.a), since the sensor network could fail to detect the event of temperature higher than T_c in the monitored region. However, a smaller coverage radius ϵ_2, like in Fig. 1.2.b), would decrease considerably the probability of missing this event.

Sensors performing a remote control A large class of sensors is able to perform a remote control, such as light and movement sensors. In this case, coverage depends on the sensitivity of the receivers in the sensors, so, in order to be detected, the monitored event occurs closer than some distance.



a) b)

Figure 1.2: In both figures, the black circles represent the positions of the sensors. a) A choice of a coverage radius ϵ_1 may not capture an undesirable event; b) A choice of a smaller radius ϵ_2 increases the probability of capturing the event

1.4 Mathematical Tools

1.4.1 Algebraic topology

The tools we present in this section have been developed in the 1930s, but it was only less than 10 years ago that they were applied to sensor networks, specifically in [41, 23, 12, 22, 107]. The major contribution of algebraic topology is contemplating the local-to-global transition: global information of a sensor network such as connectivity, coverage and Euler characteristic can be obtained by using the information on the neighborhood of each node belonging to some sensor network. As a consequence, that global information is obtained without knowing the coordinates of the nodes, localization, orientation capabilities or any other information apart from their identities and the identities of close neighbors.

Roughly speaking, algebraic topology provides a way to associate to a given space X a collection of algebraic objects which gauge the global features of X. In our case, those global features are the called homology groups, $H_k(X)$, for $k = 0, 1, \ldots$, and to determinate them, we use local objects built by simple oriented pieces, called simplices.

Given a set of points V, a k-simplex is an unordered subset $\{v_0, v_1, \ldots, v_k\}$ where $v_i \in V$ and $v_i \neq v_j$ for all $i \neq j$. The faces of the k-simplex $\{v_0, v_1, \ldots, v_k\}$ are defined as all the $(k-1)$-simplices of the form $\{v_0, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k\}$ with $0 \leq j \leq k$. In terms of coverage in a network, a 0-simplex represents a single sensor and a k-simplex represents that the $k+1$ points of this simplex are covering the convex hull containing those points (see columns two and three in Fig. 1.3).

In the way that simplices were defined in this section, the Čech theorem establishes the link between a sensor network and a simplicial complex, showing that the union of the individual coverage of each sensor in the network and the simplicial complex obtained by these sensors and their coverages are topologically equivalent. This theorem is a consequence of the Nerve lemma [14, Theorem 10.7] and is seen in Chapter 5 of this thesis.

1.4.2 Poisson point processes

Since [9, 10], Poisson point processes are the basis of stochastic-geometry modeling of communication networks. This modeling consists of treating the given architecture

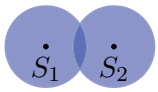
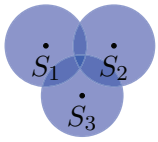
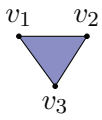
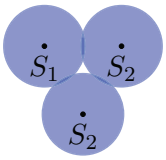
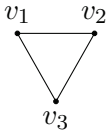
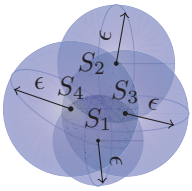
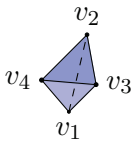
Sensor network representation	Čech complex representation	Highest order simplices
		A 0-simplex $\{v_1\}$
		An 1-simplice $\{v_1, v_2\}$
		A 2-simplex $\{v_1, v_2, v_3\}$
		Three 1-simplices $\{v_1, v_2\}$, $\{v_1, v_3\}$ and $\{v_2, v_3\}$.
		A 3-simplex $\{v_1, v_2, v_3, v_4\}$

Figure 1.3: Topological interpretations of sensor networks. Each node v_i represents a sensor S_i. We can see that the topology of the coverage is the same as for the simplicial complex

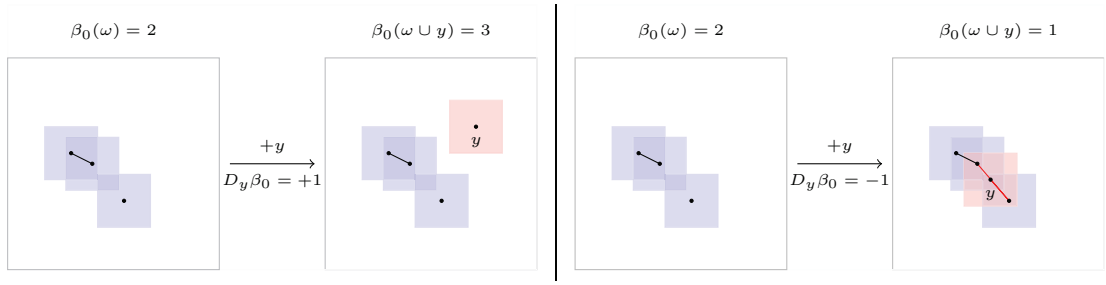


Figure 1.4: Gradient operator D_y in a Poisson process ω.

of the network as random and analyzing it in a statistical way. In this approach, the physical meaning of the network elements is preserved and reflected in the model, but their geographical locations are no longer fixed but modeled by random points. Consequently, any particular detailed pattern of locations is no longer of interest. Instead, the method allows catching the essential spatial characteristics of the network performance through the densities of these point processes.

Poisson point processes can model both static and dynamic systems. In the first case, the communicating elements do not move although each one of them can be active or inactive (a sensor can be inactive due to a flaw or, intentionally, due to some protocol for instance) and it represents a network where there is no control over the deployment of each element in the target region, so the number of elements and their positions are random. In dynamic systems, the process represents a snapshot of the network. In some cases the random number of users is overdimensioned under this hypothesis, since this system would allow an arbitrarily large number of users in a system where resources, such as power or bandwidth, are limited.

Moreover, among all studied point processes, the Poisson one has the largest arsenal of results (see [20, 111]) and by using a proper filtration, the Poisson measure can be seen as a martingale. As consequence, as showed in Chapter 6 of [99], it is possible to use Malliavin calculus on such problems. The basic tools of Malliavin calculus consist in a gradient and a divergence operator that are linked by integration by parts formula. Along this thesis, the gradient operator D_t will be used, and it can be defined by $D_y(F) = F(\omega \cup y) - F(\omega)$ of a given Poisson process ω, and an example of $D_y\beta_0$ is given in Fig. 1.4. The stochastic analysis of a Poisson process warrants also the chaos representation of a large family of random variables F depending on $\omega \in A$ as

$$F = \sum_{n=0}^{\infty} \frac{1}{n!} I_n(f_n),$$

where the multiple Poisson stochastic integral $I_n(f_n)$, is defined in Chapter 2 as

$$I_n(f_n)(\omega) = \int_{\Delta_n} f_n(x_1, \cdots, x_n)(\,\mathrm{d}\omega(x_1) - \,\mathrm{d}\lambda(x_1)) \cdots (\,\mathrm{d}\omega(x_n) - \,\mathrm{d}\lambda(x_n)).$$

The gradient operator can compute the function f_n as

$$f_n(t_1, \ldots, t_n) = \mathbf{E}_\lambda \left[D_{t_1} \ldots D_{t_n} F \right], \quad \text{a.e. } t_1, \ldots, t_n \in A.$$

1.5 Thesis Outline and Contributions

Let us turn to the contents of this thesis. Chapter 2 starts with an elementary exposition regarding the stochastic model of the positions of sensors/users. First some concepts of stochastic geometry are presented in order to define a Poisson point process and some properties such as the distribution of the number of point and the Campbell theorem. Then, we present some definitions and results of Malliavin calculus that are used in the thesis. Among the results of Malliavin calculus, we highlight the possibility of the decomposition if a large family of random variables depending on a Poisson point process as the sum of orthogonal chaos, each one being an Itô integral. If F is a random variable depending on the Poisson point process ω, the nth chaos represents the contribution of every set of n points on F. It also yields non-trivial results such as a concentration inequality used in Chapters 3, 4 and 6, allowing to find limits for the distribution of some random variables.

Chapter 3 uses the Poisson point process and concentration inequalities as analytical methods to apply to the interference temperature model. This model provides a way to reallocate spectrum bands more efficiently, giving some bandwidth at low power to opportunist users (the sensors) as long as they do not interfere the signal of licensed users (or primary users). We consider the position of the sensors in a given protocol (in this case WiMax) and we can use Campbell theorem to calculate how much power is available for those sensors such that, in the average, primary users ignore their presence in the system. However, the fact that this constraint is respected, on average, is not enough to design a system. The useful quantity is the probability of a primary user suffering an outage due to the fact that sensors use his bandwidth and it must be as low as possible. We are not able to calculate this probability but we can find an upper bound for the outage probability, so that we can design the transmitted power of the sensors such that the outage probability is lower than some threshold.

In Chapter 4 we study how an OFDMA system can be designed by using the tools presented in Chapter 2. An OFDMA system has a limited number of subchannels that distribute to the users in a cell and this system is overloaded if, at some instant, the demand of subchannels is larger than the number of available ones. Each user needs a different number of subchannels, which is a function of the channel fading and the type of service (data, voice, video, etc.). We aim to design the minimal number of subchannels a central station can provide as a function of the density of users and their needs such that the probability of overload of the system is low. This is done under reasonable assumptions, by considering that users positions at a given instant are given by the points of a Poisson process. This assumption overestimates the real number of users, since the number of available subchannels restricts the number of users in real cases. Then we use a the concentration inequality to obtain the main results.

In Chapter 5, we introduce the tools of topology. We make precise the idea of simplicial complexes, homology and Betti numbers. To determine those quantities we define a linear operator named boundary map and we use algebra. After defining these concepts, we show some other results that allow us to give even more concrete links between the model of simplicial complex in a torus and a sensor network.

It is mostly in Chapter 6 that we apply the tools of Chapters 5 and 2 and this is the first work with random geometric complexes not considering the asymptotics. We consider, as usual, that the deployment of the sensors is a Poisson point process in a d-torus. The coverage radius of the sensors is deterministic given by ϵ, which, as stated in Chapter 5, is enough to define a simplicial complex having the homotopy type and then, for our

purposes, we can only use the complex, allowing us to apply other tools from algebraic homology. Then we turn our attention to the number of simplices. We use the chaos decomposition and the Campbell theorem to find expressions for the n-th moments of k-simplices. The first three moments are explicitly found. We present a method to find the other moments, and although there are no new difficulties, the calculation is too tedious (we can have an idea comparing the complexity of the evaluation between the first and the second moments, and between the second and the third moments). This result can already tell information regarding the sensor network, but it is possible to go further and to calculate explicitly the two first moments of Euler characteristic. Moreover, all those results can be obtained for a fixed number of points by means of the Poisson transform (or de-poissonization). The results of the Euler characteristic and concentration inequalities allow us to infer some information related to the behavior of Betti numbers in d dimensions. It is also investigated the behavior of the number of k-simplices for a high density of sensors and we can show that this converges to a Gaussian distribution, which turns out to be an interesting result: since we have calculated the mean and variance of this quantity, we have a good approximation of its analytical distribution. Since we found also an upper bound of the distance between the distribution of k-simplices and the Gaussian, we can even estimate the error of this consideration.

Chapter 7 solves the problem of coverage and connectivity of sensor networks in one dimension. This can be applied to networks where sensors are deployed over a privileged dimension as well in queuing systems, since solving this problem is equivalent to evaluate a $M/D/1/1$ preemptive queue. We obtain all aforementioned results for this case: for a line segment or a circle, we find the distribution of the number of connected components (and we call each connected component a cluster), the distribution of the size of clusters, all the moments and the Laplace transform of those distributions. The mathematical tools used in this chapter are quite simple and the results could have been obtained at least fifty years ago. However, solving the problem in one dimension gives us hints and intuitions in order to solve the problem in d dimensions.

Part I

Stochastic Geometry

Chapter 2

Stochastic Model

2.1 Introduction

In this chapter, we review the main facts used in this thesis concerning the stochastic model. We give the definition of the Poisson measure on a space of configurations of a metric space X. This allows us to characterize the randomness of the system by considering that the set of points representing the sensors are modeled by a Poisson point process ω with intensity λ in a Polish space Y. This characterization is important due to the development of the theory of Poisson point process, giving us a large number of tools. From the definition of such a measure, it is possible to express a random variable depending on the random realization ω as a sum of stochastic integrals and then to obtain the probabilistic interpretation of the gradient D as a finite difference operator. Using this operator it is possible to derive other results, such as a concentration inequality on the Poisson space.

2.2 Poisson point process

To characterize the randomness of the system, we consider that the set of points is represented by a Poisson point process ω with intensity λ in a Polish space Y. The space of configurations on Y, is the set of locally finite simple point measures (cf [99]):

$$\Omega^Y = \left\{ \omega = \sum_{k=0}^{n} \delta(x_k) \ : \ (x_k)_{k=0}^{k=n} \subset Y, \ n \in \mathbb{N} \cup \{\infty\} \right\},$$

where $\delta(x)$ denotes the Dirac measure at $x \in Y$. Simple measure means that $\omega(\{x\}) \leq 1$ for any $x \in Y$. Locally finite means that $\omega(K) < \infty$ for any compact K of Y. It is often convenient to identify an element ω of Ω^Y with the set corresponding to its support, i.e., $\sum_{k=0}^{n} \delta(x_k)$ is identified with the unordered set $\{x_1, \cdots, x_n\}$. For $A \in \mathcal{B}(Y)$, we have $\delta(x_k)(A) = \mathbf{1}_{[x_k \in A]}$, so

$$\omega(A) = \sum_{x_k \in \omega} \mathbf{1}_{[x_k \in A]} = \int_A \mathrm{d}\omega(x),$$

counts the number of atoms in A. The configuration space Ω^Y is endowed with the vague topology and its associated σ-algebra denoted by $\mathcal{F}^Y$. Since ω is a Poisson point process of intensity λ:

i) For any A with Lebesgue measure $S(A)$, $\omega(A)$ is a random variable of parameter $\lambda S(A)$, i.e.,

$$\mathbf{P}(\omega(A) = k) = e^{-\lambda S(A)} \frac{(\lambda S(A))^k}{k!}.$$

ii) For $A' \in \mathcal{B}(Y)$, for any disjoints A, A', the random variables $\omega(A)$ and $\omega(A')$ are independent.

In the thesis, we refer $\mathbf{E}_\lambda[F(\omega)]$ as the mean of some function F depending on ω given that the intensity of this process is λ and $\mathbf{P}_\lambda[\omega \in Y] = \mathbf{E}_\lambda\left[\mathbf{1}_{[\omega \in Y]}\right]$. The definitions of $\mathbf{V}_\lambda[F(\omega)]$ and $\mathrm{Cov}_\lambda[F(\omega), G(\omega)]$ are straightforward. Define $\Delta_n = \{(x_1, \cdots, x_n) \in Y^n \mid x_i \neq x_j, \forall i \neq j\}$. Let $f(x_1, \cdots, x_n)$ be a measurable function and let $F(\omega)$ be a random variable given by

$$F(\omega) = \sum_{\substack{x_i \in \omega \cap A, 1 \leq i \leq n \\ x_i \neq x_j \ if \ i \neq j}} f(x_1, \cdots, x_n) = \int_{A \cap \Delta_n} f(x_1, \cdots, x_n)\, \mathrm{d}\omega(x_1) \cdots\, \mathrm{d}\omega(x_n).$$

A well known property of the Poisson point processes [20] states that

$$\mathbf{E}_\lambda[F(\omega)] = \int_A f(x_1, \cdots, x_n)\, \mathrm{d}\lambda(x_1) \cdots\, \mathrm{d}\lambda(x_n). \tag{2.1}$$

The notion of point process can be extended to configurations in $\mathbf{R}^k \times X$ where X is a subset of $\mathbf{R}^m$. A configuration is then typically of the form $\{(x_n, y_n), n \geq 1\}$ where for each $n \geq 1$, $x_n \in \mathbf{R}^k$ and $y_n \in X$. We keep writing (x_n, y_n) as a couple, though it could be thought as an element of $\mathbf{R}^{k+m}$, to stress the asymmetry between the spatial coordinate x_n and the so-called mark, y_n. For a marked point process, we denote by ω the set of locations, i.e., $\omega = \{X_n, n \geq 1\}$ and by ω' the set of both locations and marks, i.e., $\omega' = \{(X_n, Y_n), n \geq 1\}$. A marked point process with position dependent mark is one for which the law of Y_n, the mark associated to the atom located at X_n, depends only on X_n through a kernel K:

$$\mathbf{P}(Y_n \in B \,|\, \omega) = K(X_n, B), \text{ for any } B \subset X.$$

If K is a probability kernel, i.e., if $K(x, X) = 1$ for any $x \in \mathbf{R}^k$ then it is well known that ω' is a Poisson process of intensity $K(x, \mathrm{d}y)\, \mathrm{d}\lambda(x)$ on $\mathbf{R}^k \times \mathbf{R}^m$. For $f : \mathbf{R}^k \times \mathbf{R}^m \to \mathbf{R}$ a measurable non-negative function, let

$$F = \int f\ \mathrm{d}\omega' = \sum_{X_n \in \omega} f(X_n, Y_n). \tag{2.2}$$

We denote the mean of a function F of marked Poisson point process with kernel K as $\mathbf{E}_{\lambda,K}[F]$. The Laplace transform of F is given by [111]:

$$\mathbf{E}_{\lambda,K}\left[e^{-sF}\right] = \exp\left(-\int (1 - e^{-sf(x,y)}) K(x, \mathrm{d}y) d\lambda(x)\right) \tag{2.3}$$

As consequence, the Campbell formula can be extended.

Theorem 2.1 *Let ω' be a marked Poisson process on $\mathbf{R}^k \times \mathbf{R}^m$. Let λ be the intensity of the underlying Poisson process and K the kernel of the position dependent marking. Then,*

$$\mathbf{E}_{\lambda,K}[F] = \int_{\mathbf{R}^k \times \mathbf{R}^m} f(x, y)\, K(x, \ dy)\ d\lambda(x).$$

A real function $f : Y^n \to \mathbb{R}$ is called symmetric if

$$f(x_{\sigma(1)}, \cdots, x_{\sigma(n)}) = f(x_1, \cdots, x_n)$$

for all permutations σ of $\mathfrak{S}_n$.The space of symmetric square integrable random variables is denoted by $L^2(\lambda)^{\circ n}$. For $f \in L^2(\lambda)^{\circ n}$, the multiple Poisson stochastic integral $I_n(f_n)$ is then defined as

$$I_n(f_n)(\omega) = \int_{\Delta_n} f_n(x_1, \cdots, x_n)(\ \mathrm{d}\omega(x_1) - \ \mathrm{d}\lambda(x_1)) \cdots (\ \mathrm{d}\omega(x_n) - \ \mathrm{d}\lambda(x_n)).$$

If $f_n \in L^2(\lambda)^{\circ n}$ and $g_m \in L^2(\lambda)^{\circ m}$, the isometry formula

$$\mathbf{E}_\lambda[I_n(f_n) I_m(g_m)] = n! \mathbf{1}_{[m=n]} \langle f_n, g_m \rangle_{L^2(\lambda)^{\circ n}} \tag{2.4}$$

holds true (see [99]). Furthermore, we have:

Theorem 2.2 *Every random variable $F \in L^2(\Omega^Y, \mathbf{P})$ admits a (unique) Wiener-Poisson decomposition of the type*

$$F = \mathbf{E}_\lambda[F] + \sum_{n=1}^{\infty} I_n(f_n),$$

where the series converges in $L^2(\mathbf{P})$ and, for each $n \geq 1$, the kernel f_n is an element of $L^2(\lambda)^{\circ n}$. Moreover, we have the isometry

$$\| F - \mathbf{E}_\lambda[F] \|^2_{L^2(\lambda)^{\circ n}} = \sum_{n=1}^{\infty} n! \| f_n \|^2_{L^2(\mathbb{R}_+)^{\circ n}}. \tag{2.5}$$

For $f_n \in L^2(\lambda)^{\circ n}$ and $g_m \in L^2(\lambda)^{\circ m}$, we define $f_n \otimes^l_k g_m$, $0 \leq l \leq k$, to be the function:

$$(y_{l+1}, \cdots, y_n, x_{k+1}, \cdots, x_m) \longmapsto \int_{Y^l} f_n(y_1, \cdots, y_n) g_m(y_1, \cdots, y_k, x_{k+1}, \cdots, x_m)\ \mathrm{d}\lambda(y_1) \ldots\ \mathrm{d}\lambda(y_l). \tag{2.6}$$

We denote by $f_n \circ^l_k g_m$ the symmetrization in $n+m-k-l$ variables of $f_n \otimes^l_k g_m$, $0 \leq l \leq k$. This leads us to the next proposition, shown in [99]:

Proposition 2.3 *For $f_n \in L^2(\lambda)^{\circ n}$ and $g_m \in L^2(\lambda)^{\circ m}$, we have*

$$I_n(f_n) I_m(g_m) = \sum_{s=0}^{2(n \wedge m)} I_{n+m-s}(h_{n,m,s}),$$

where

$$h_{n,m,s} = \sum_{s \leq 2i \leq 2(s \wedge n \wedge m)} i! \binom{n}{i} \binom{m}{i} \binom{i}{s-i} f_n \circ^{s-i}_i g_m$$

belongs to $L^2(\lambda)^{\circ n+m-s}$, $0 \leq s \leq 2(m \wedge n)$.

In what follows, given $f \in L^2(\lambda)^{\circ q}$ $(q \geq 2)$ and $t \in Y$, we denote by $f(*,t)$ the function on Y^{q-1} given by $(x_1, \cdots, x_{q-1}) \longmapsto f(x_1, \cdots, x_{q-1}, t)$.

Definition 2.1 *Let* $\mathrm{Dom}\, D$ *be the set of random variables* $F \in L^2(P)$ *admitting a chaotic decomposition such that*

$$\sum_{n=1}^{\infty} qq! \|f_n\|^2 < \infty.$$

Let D *be defined by*

$$D \ : \ \mathrm{Dom}\, D \to L^2(\Omega^Y \times Y, P \times \lambda),$$

such that

$$F = \mathbf{E}_\lambda [F] + \sum_{n \geq 1} I_n(f_n) \longmapsto D_t F = \sum_{n \geq 1} n I_{n-1}(f_n(*,t)).$$

It is known, cf. [60], that we also have

$$D_t F(\omega) = F(\omega \cup \{t\}) - F(\omega), \ dP \times dt \ a.e..$$

Remark 2.1 It is possible to show that the expression of f_n that appears in the chaos expansion

$$F = \sum_{n=0}^{\infty} \frac{1}{n!} I_n(f_n)$$

can be expressed by using the gradient operator and we have:

$$f_n(t_1, \ldots, t_n) = \mathbf{E}_\lambda \left[D_{t_1} \ldots D_{t_n} F \right], \quad \text{a.e. } t_1, \ldots, t_n \in A.$$

This is analogous to the classical Taylor expansions for one or several variables, where we have

$$f(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!},$$

where the coefficients a_n can be found:

$$a_n = \left. \frac{\partial^n f}{\partial x^n}(x) \right|_{x=0},$$

and we have the correspondence:

calculus on $\mathbf{R}$	stochastic analysis
$f(x)$	F
$f(0)$	$\mathbf{E}_\lambda [F]$
$\dfrac{\partial^n}{\partial x^n}$	D^n
$\dfrac{\partial^n f}{\partial x^n}(0)$	$\mathbf{E}_\lambda [D^n F]$

Definition 2.2 *The Ornstein-Uhlenbeck operator L is given by*

$$LF = -\sum_{n=1}^{\infty} nI_n(f_n),$$

whenever $F \in \operatorname{Dom} L$, given by those $F \in L^2P$ such that their chaotic expansion verifies

$$\sum_{n=1}^{\infty} q^2 q! \|f_n\|^2 < \infty.$$

Note that $\mathbf{E}_\lambda [LF] = 0$, by definition and (2.4).

Definition 2.3 *For $F \in L^2(\mathbf{P})$ such that $\mathbf{E}_\lambda [F] = 0$, we may define L^{-1} by*

$$L^{-1}F = -\sum_{n=1}^{\infty} \frac{1}{n} I_n(f_n).$$

Combining Stein's method and Malliavin calculus yields the following theorem, see [87]:

Theorem 2.4 *Let $F \in \operatorname{Dom} D$ be such that $\mathbf{E}_\lambda [F] = 0$ and $Var(F) = 1$. Then,*

$$d_W(F, \mathcal{N}(0,1)) \leq \mathbf{E}_\lambda \left[\left| 1 - \int_Y [D_t F \times D_t L^{-1} F] \; d\lambda(t) \right| \right] + \int_Y \mathbf{E}_\lambda \left[|D_t F|^2 \left| D_t L^{-1} F \right| \right] \; d\lambda(t).$$

Another result from the Malliavin calculus used in this work is the following one, quoted from [99]:

Theorem 2.5 *Let $F \in \operatorname{Dom} D$ be such that $DF \leq K$, a.s., for some $K \geq 0$ and $\|DF\|_{L^\infty(\Omega, L^2(Y))} < \infty$. Then*

$$\mathbf{P}(F - \mathbf{E}_\lambda [F] \geq x) \leq \exp\left(-\frac{x}{2K} \log\left(1 + \frac{xK}{\|DF\|_{L^\infty(\Omega, L^2(Y))}} \right) \right). \tag{2.7}$$

We recall that if ω is a marked Poisson process on $\mathbf{R}^k \times \mathbf{R}^m$ of intensity λ and kernel $K(x, y)$, then

$$m_F = \mathbf{E}_{\lambda, K} [F] = \int f(x) \quad K(x, dy) \lambda(\, \mathrm{d}x)$$

and

$$v_F = \int |D_x F(\omega) \; K(x, dy)|^2 \lambda(\, \mathrm{d}x).$$

2.3 Summary

In this chapter all the stochastic tools used throughout the thesis have been introduced. First, we have presented the Poisson point process in a formal fashion in order to present some properties and extend the definition to marked Poisson point processes. After, we have introduced the notion of Malliavin calculus, which allows us to describe a large family of the Poisson point process (those square-integrable ones) as the sum of chaos. As consequence, we can apply results from [87] and concentrations inequalities from [99] to find upper bounds for the distributions of random variables depending on a Poisson point process.

Chapter 3

Interference Temperature Model in Cognitive Radio Networks

3.1 Introduction

This chapter proposes the utilization of the Poisson point process as a new analytical method to be applied in the interference temperature (IT) approach. For this purpose, we firstly develop a model for the RF environment. Afterwards, by the use of the Poisson point process, we determine essential elements for the calculation of the achievable per-link capacity and the total capacity of a secondary network following the ideal and the generalized IT models.

As long as new radio access technologies continue to appear, there are few spectrum bands to be allocated due to inefficient fixed spectrum allocations. This phenomenon obstructs the development of new wireless technology and communication services [64]. Moreover, spectrum occupancy measurements [79] evidence that fixed spectrum allocations also result in low efficiency in spectrum utilization because a large portion of the spectrum remains underutilized [33].

These observations have motivated the regulatory bodies to investigate different access methods to overcome the above problems. As result, the use of Cognitive Radio technology that allows Dynamic Spectrum Access (DSA) has emerged as a possible solution to solve the low efficiency in spectrum utilization by allowing spectrum sharing. In such an approach, Secondary Users (SUs) are allowed to dynamically access the unused spectrum in Primary Users' (PUs) bands, commonly referred as "spectrum holes".

In the last years, two different strategies of spectrum sharing have been identified. One is through opportunistic spectrum access, known as "Overlay" and the other is through the use of low power spread-spectrum, known as "Underlay" [21]. The Overlay approach is based on avoidance of PUs through the use of spectrum sensing and adaptive allocation. On the other hand in the Underlay approach, which is of interest in this chapter, the transmission of SUs is allowed in PUs bands, if the transmission power is low enough that it does not harm the PUs. As this approach imposes severe restrictions on transmitted power levels, it requires operating over "ultra" wide bandwidths. Under this framework, in November 2003, the concept of Interference Temperature (IT) was proposed by the Federal Communications Commission (FCC), as another way to dynamically manage and allocate spectrum resources [34]. The principal characteristic of the IT model, as an underlay approach, is the fact that in this model SUs attempt to coexist with PUs meanwhile, in

other proposals for DSA (i.e. overlay approaches), SUs try to avoid PU's signals [113].

After conclude the mathematical analysis enlightened by the Poisson point process model, we demonstrate the application of our model by a numerical example, in which we consider the primary user as a Universal Mobile Telecommunications Service (UMTS) network and the secondary user as an Ultra Wide Band (UWB) network. Finally, by the use of Concentration Inequalities we determine an upper bound on the outage probability of the primary network when the SUs transmit.

The rest of the chapter is organized as follows: Section 3.2 describes the ideal and the generalized interference temperature models and the physical features of the system; In Section 3.3, we present our model for the calculation of the SUs mean capacity and the numerical analysis; Bounds of the outage probability of PU using Concentration Inequalities are found in Section 3.4; Finally, in Section 3.5, we present our conclusions.

3.2 Physical Model

In the IT model, SUs equipped with cognitive radio technology must firstly sense the available spectrum band to compute the existing interference. In this approach, SUs treat PUs, other SUs, interference, and noise all as interference. Afterwards, they must adjust their transmission power to avoid raising the interference temperature above a predefined threshold, which is assumed to be established by the FCC. This threshold represents the maximum quantity of interference that a PU can tolerate. Therefore, SUs must guarantee that the existing interference temperature, added to the interference caused by their transmissions does not exceed the interference temperature limit (T_L) [113].

The Interference Temperature $T_I(f_c, B)$ is defined as:

$$T_I(f_c, B) = \frac{P_I(f_c, B)}{kB},$$

where $T_I(f_c, B)$ is estimated in Kelvin, the average interference power P_I is measured in Watts, is centered at f_c, covers a bandwidth B measured in Hertz and k is the Boltzmann's constant.

In [113], two different interference temperature models were presented: the Ideal and the Generalized. The ideal model tries to limit interference specifically to PUs signals. Therefore, the priori knowledge of PUs activity is needed. This model can be written as:

$$T_I(f_i, B_i) + \frac{M_i P}{kB_i} \leq T_L(f_i) \quad \forall\, 1 \leq i \leq n,$$

where P is the average power of SUs operating with the center frequency f_c and bandwidth B. The band [f_c−B/2, f_c+B/2] overlaps n PUs signals, with respective frequencies f_i and bandwidth B_i, T_L is the interference temperature limit. As the purpose of this model is to restrict the interference received by PUs, the constant M_i (with fractional value between 0 and 1) represents the attenuation between the primary receiver and the secondary transmitter. This constant is assumed to be fixed by a regulatory body in [113].

For the generalized model, the priori knowledge of PUs activity is not required. So this model can be applied in the entire bandwidth regardless the exact location of the PUs signals. The interference temperature limit in the generalized model can be expressed as following:

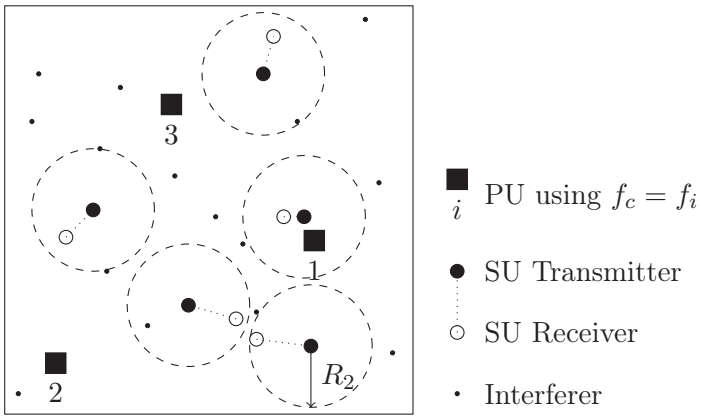


Figure 3.1: Example of a configuration with primary users, secondary users (transmitters and receivers) and interferers.

$$T_I(f_c, B) + \frac{MP}{kB} \leq T_L(f_c).$$

Here B is the entire frequency range, and not just PUs frequency band. Since the parameters of the PUs receivers are unknown, the constraint is in terms of SUs transmitter's parameters [113].

We consider a system in which the position of users are given by Poisson point processes. So the Poisson point process, ω_j on $\mathbf{R}^2$, with intensity (i.e., the mean number of users per unity of area) λ_j, represents the positions of user of kind j. Moreover, their individual transmission power is given by μ_j. We associate primary users to the index $j = 1$, secondary users that are transmitting to $j = 2$ and interferers to $j = 3$. Besides, the marked Poisson point process ω_2' on $\mathbf{R}^2 \times \mathbf{R}^2$ associates, for each point X_i from ω_2 another point Y_i uniformly distributed over a disc $D_2(X_i)$ of radius R_2 centered at X_i, in such way that Y_i is the position of the SU receiving the signal of a SU placed at X_i.

Let $p_j(x, y)$ be the received power experienced by a user located at $y \in \mathbf{R}^2$ with respect to a transmitter of kind j at $x \in \mathbf{R}^2$. We use the propagation power loss as the simplified model for path loss being a function of distance, based on Hata's model:

$$p_j(x, y) = \min\left(\mu_j, \mu_j \left(\frac{r_0}{\|x - y\|}\right)^{\alpha}\right), \tag{3.1}$$

where r_0 and α are positive constants. In this approximation α is the path loss coefficient and we take $r_0 = 1$, where r_0 is a reference distance from the antenna far field. Due the strong attenuation as a function of distance, the existence of the noise and the sensibility of the SUs, we assume that SUs are only able to communicate with some other ones closer than a distance R_2. The total interference power experienced by a point x with respect of all users of kind j is given by $Q_j(x)$. We illustrate a possible realization of the model in Fig. 3.1, representing all kinds of users. Note that, to each SU transmitting, it must exist a SU receiving, which does not cause interference. Along this work, since R_2 is considered enough small to consider that a pair of SU's (i.e. SU transmitter and SU receiver) causes the same interference in the other users, so, in terms of interference, SU's are considered as a single point.

3.3 Secondary User Mean Capacity

In this section, we consider that the constraint given by the temperature model holds for the mean of the quantities, and then, based on the physical model we analyze the mean capacity of the network. First, we proceed the calculations and then we present the numerical results.

3.3.1 Calculations of the mean capacity

The mean total network capacity of the SUs is based on the mean per-link capacity and the constraints of the IT model. Therefore, taking into account the IT model restrictions, we develop the necessary expressions to estimate the mean capacity for the ideal and the generalized IT models. In order to achieve this, we consider the following lemma.

Lemma 3.1 *Let ω_j be a Poisson point process with intensity measure λ representing the positions of active users of kind j over $\mathbf{R}^2$ transmitting with a power μ_j. If $Q_j(x)$ is the total interference power received experienced by a point $x \in \mathbf{R}^2$, then*

$$\mathbf{E}_\lambda [Q_j(x)] = \frac{\mu_j \lambda_j \pi \alpha}{(\alpha - 2)}. \tag{3.2}$$

Proof: Given the invariance under translation of a stationary Poisson point process, $\mathbf{E}_\lambda [Q_j(x)] = \mathbf{E}_\lambda [Q_j(0)]$. So, it suffices to use Eq. (2.1) for $f(x) = p_j(x, 0)$ as defined in Eq. (3.1):

$$\begin{aligned}
\mathbf{E}_\lambda [Q_j(0)] &= \mathbf{E}_\lambda \left[\sum_{X_i \in \omega} p_j(X_i, 0) \right] \\
&= \int_{\mathbf{R}^2} p_j(x, 0) \lambda_j(x) \, \mathrm{d}x \\
&= \lambda_j \int_0^{2\pi} \int_0^{\infty} \min \left(\mu_j, \mu_j \left(\frac{1}{r} \right)^{\alpha} \right) r \, \mathrm{d}r \, \mathrm{d}\theta \\
&= \frac{\mu_j \lambda_j \pi \alpha}{(\alpha - 2)},
\end{aligned}$$

concluding the proof. □

For the ideal case, we must calculate the maximum allowed SUs Transmission Power μ_2^{id}. As imposed in the physical model, the ideal interference temperature model attempt to limit interference specifically to licensed signals. This means that the objective is to guarantee that

$$\frac{P_I(f_i, B_i)}{kB_i} + \frac{M_i P}{kB_i} \leq T_L(f_i).$$

The left side of the equation represents the total temperature allowed to interferers with respect to a primary user using the center frequency f_i, placed at x_i, and we can rewrite it as

$$Q_2(x) + Q_3(x) \leq T_L(f_i) k B_i.$$

To guarantee that this inequality holds at least for the mean of interferences, we take the mean in both sides, use Lemma 3.1 and solve for μ_2^{id}, resulting in the maximum allowed transmission power to be used by secondary users:

$$\mu_2^{id} \leq \frac{(\alpha-2)T_L kB_i - \lambda_3\mu_3\pi\alpha}{\lambda_2\pi\alpha}. \tag{3.3}$$

Now we calculate the maximum mean allowed SUs transmission power μ_2^{gen} following the generalized model. The main difference between the generalized and the ideal model is that in the generalized model the priori knowledge of PUs activity is not required. Thus, the generalized model is written as:

$$\frac{P_I(f_c,B)}{kB} + \frac{MP}{kB} \leq T_L(f_c).$$

As the parameters of the PUs receivers are unknown, the constraint is in terms of the SUs transmitter's parameters. Therefore, B is the entire frequency range, and not just PUs frequency band. Again, since SUs treat PUs, other SUs, interference, and noise all as interference, we notice that for the generalized model we take into account the power from the others PUs (averaging the PUs power over the SUs bandwidth) and evaluate the T_L over the entire frequency range B. If the analyzed primary user is placed at $x \in \mathbf{R}^2$, then we rewrite this condition as a function of Q_i's:

$$\frac{B_i}{B}Q_1(x) + Q_2(x) + Q_3(x) \leq T_L(f_c)kB.$$

Assuring that, in average, this inequality holds, we take the mean, apply Lemma 3.1 and solving for μ_2^{gen}, we obtain:

$$\mu_2^{gen} \leq \frac{(\alpha-2)T_L kB - \lambda_3\mu_3\pi\alpha - \frac{B_i}{B}\lambda_1\mu_1\pi\alpha}{\lambda_2\pi\alpha}. \tag{3.4}$$

Since we are interested in the calculation of the capacity using the Shannon-Hartley theorem [61] the per-link capacity $C(x,y)$ of a user at $x \in \mathbf{R}^2$ receiving a signal from a user at $y \in \mathbf{R}^2$ such that $\|x-y\| \leq R_2$, is given by:

$$C(x,y) = B\log_2\left(1 + \frac{p_2(x,y)}{\mathbf{E}_\lambda\left[I(x)\right]}\right),$$

Where $I(x)$ is the interference power caused by the interferers, other SUs and PUs at x, given by

$$I(x) = \frac{B_i}{B}Q_1(x) + Q_2(x) + Q_3(x).$$

Since y is uniformly distributed around x, then x is uniformly distributed over y and the mean capacity per link $C(x)$ in the disc is:

$$C(x) = \int B\log_2\left(1 + \frac{p_2(x,y)}{\mathbf{E}_\lambda\left[I(x)\right]}\right)\frac{\mathbf{1}_{[\{\|x-y\|\leq R_2\}]}}{\pi R_2^2}\,\mathrm{d}y$$

By Lemma 3.1, $\mathbf{E}_\lambda\left[Q_j(x)\right]$ does not depend on y, and $p_2(x,y)$ depends only on the distance between x and y, so $C(x)=C$. Let us define

$$K \triangleq \frac{\mu_2}{\mathbf{E}_\lambda\left[I(x)\right]} = \frac{\mu_2(\alpha-2)}{\pi\alpha\left(\frac{B_i}{B}\mu_1\lambda_1+\mu_2\lambda_2+\mu_3\lambda_3\right)}.$$

Then, we can rewrite C as follows:

$$C = \int_0^{R_2}\int_0^{2\pi} \frac{B}{\pi R_2^2}\log_2\left(1+K\min(1,r^{-\alpha})\right) r \ \mathrm{d}\theta \ \mathrm{d}r. \tag{3.5}$$

Defining $h : \mathbf{R}_+ \times (2,\infty) \to \mathbf{R}_+$ as follows

$$h(r,t) \triangleq \int_0^r \ln\left(1+\frac{1}{x^t}\right) x \ \ \mathrm{d}x, \tag{3.6}$$

we rewrite Eq. (3.5) as

$$C = \frac{2BK^{\frac{2}{\alpha}}}{R_2^2\ln(2)}\left[\frac{\ln(1+K)}{2K^{\frac{1}{\alpha}}} + h\left(\frac{R_2}{K^{\frac{1}{\alpha}}},\alpha\right) - h\left(\frac{2}{K^{\frac{1}{\alpha}}},\alpha\right)\right] \tag{3.7}$$

The capacity per-link in the ideal case, C^{id}, is obtained taking $\mu_2=\mu_2^{id}$, while the one in the generalized case, C^{gen}, results of taking $\mu_2=\mu_2^{gen}$. It is possible to find analytical expressions for $h(r,t)$ when t is an integer.

Lemma 3.2 *Let $\beta_n=\pi(2n-1)$ for n integer. For t an odd integer, the expression of $h(r,t)$ is given by:*

$$\begin{aligned} h(r,t) = -\frac{1}{2}\sum_{n=1}^{\lfloor t/2\rfloor}\Bigg(&\cos\left(\frac{2\beta_n}{t}\right)\ln\left(\frac{1}{r^2}+\frac{2}{r}\cos\left(\frac{\beta_n}{t}\right)+1\right)\\ &+2\sin\left(\frac{2\beta_n}{t}\right)\arctan\left(\frac{r\sin(\frac{\beta_n}{t})}{1+r^2\cos(\frac{\beta_n}{t})}\right)\Bigg) + \frac{r^2}{2}\ln\left(1+\frac{1}{r^t}\right) - \frac{1}{2}\ln\left(1+\frac{1}{r}\right). \end{aligned}$$

If $t/2$ is odd, we obtain the following expression for $h(r,t)$:

$$\begin{aligned} h(r,t) = -\frac{1}{2}\sum_{n=1}^{\lfloor t/4\rfloor}\Bigg(&\cos\left(\frac{2\beta_n}{t}\right)\ln\left(\frac{1}{r^4}+\frac{2}{r^2}\cos\left(\frac{2\beta_n}{t}\right)+1\right)\\ &-2\sin\left(\frac{2\beta_n}{t}\right)\arctan\left(\frac{r^2\sin(\frac{2\beta_n}{t})}{1+r^4\cos(\frac{2\beta_n}{t})}\right)\Bigg) + \frac{r^2}{2}\ln\left(1+\frac{1}{r^t}\right) + \frac{1}{2}\ln\left(1+\frac{1}{r^2}\right), \end{aligned}$$

and if $t/2$ is even, then

$$\begin{aligned} h(r,t) = -\frac{1}{2}\sum_{n=1}^{\lfloor t/4\rfloor}\Bigg(&\cos\left(\frac{2\beta_n}{t}\right)\ln\left(\frac{1}{r^4}+\frac{2}{r^2}\cos\left(\frac{2\beta_n}{t}\right)+1\right)\\ &-2\sin\left(\frac{2\beta_n}{t}\right)\arctan\left(\frac{r^2\sin(\frac{2\beta_n}{t})}{1+r^4\cos(\frac{2\beta_n}{t})}\right)\Bigg) + \frac{r^2}{2}\ln\left(1+\frac{1}{r^t}\right). \end{aligned}$$

Proof: First we differentiate the right-hand terms with respect to r and after several elementary but tedious manipulations we have:

$$\frac{\partial h(r,t)}{\partial r} = r \ln\left(1 + \frac{1}{r^t}\right).$$

Then, it suffices to use the Fundamental Theorem of Calculus on the right-hand term of Eq. (3.6) to obtain that

$$\frac{\partial}{\partial r}\left(\int_0^r \ln\left(1 + \frac{1}{x^t}\right) x \ \ \mathrm{d}x\right) = r \ln\left(1 + \frac{1}{r^t}\right).$$

So both sides of the equation differ at most by a constant. Since these two functions are analytical at $r = 0$, it suffices to see that

$$h(0,t) = \int_0^0 \ln\left(1 + \frac{1}{x^t}\right) x \ \ \mathrm{d}x = 0,$$

thus the proof is concluded. □

We can use this lemma to obtain expressions for two typical values of α.

Lemma 3.3 *The expression of* $h(r,3)$ *is given by:*

$$h(r,3) = \frac{1}{4}\ln\left(\frac{r^2 - r + 1}{r^2 + 2r + 1}\right) + \frac{\sqrt{3}\pi}{12} + \frac{r^2}{2}\ln\left(1 + \frac{1}{r^3}\right) + \frac{\sqrt{3}}{2}\arctan\left(\frac{(2r-1)}{\sqrt{3}}\right)$$

and the expression of $h(r,4)$ *is the following one*

$$h(r,4) = \arctan r^2 + \frac{r^2}{2}\ln\left(1 + \frac{1}{r^4}\right).$$

Then, it is possible to calculate the mean total SUs capacity C_{total} in a secondary cell of radius, R, defining a disc D. Using theorem 2.1 on the marked Poisson point process ω_2', we obtain:

$$C_{total} = \int\int_D C(x,y) \frac{\mathbf{1}_{[\{\|x-y\| \le R_2\}]}}{\pi R_2^2} \ \mathrm{d}y\lambda_2 \ \mathrm{d}x.$$

However, Eq. 3.7 shows that the inner integral does not depend on x, so we can rewrite Eq. 3.8 as

$$C_{total} = C\int_D \lambda dx = C\lambda_2 \int_D \mathrm{d}x = C\lambda_2 \pi R^2. \tag{3.8}$$

Applying the μ_2 obtained by the temperature model, we obtain the total capacity of a network, and we denote C_{total}^{id} for the ideal case, when $\mu_2 = \mu_2^{id}$, and C_{total}^{gen} when $\mu_2 = \mu_2^{gen}$.

3.3.2 Numerical Analysis for the Mean Capacity of the Secondary Network

In this section, we demonstrate the application of the equations developed previously. We examine the achievable per-link capacity (C) of a secondary network and the total

capacity (C_{total}) of this network under some typical situations. For this analysis, we consider the primary user as a UMTS network and the secondary user as an UWB network, WiMedia. We develop this analysis following the ideal and the generalized IT models.

Concerning to UMTS or the primary network, we consider a primary user's intensity (λ_1) of 0.02 users per km^2. This corresponds to 60 active mobile stations in a macro-cell with radius equal to 30 km. According to [112], the transmission power (μ_1) of the UMTS mobile stations is equal to 250 mW or 24dBm and the PUs bandwidth (B_i) is 5 MHz.

For the secondary network or WiMedia, the bandwidth (B) is 528 MHz [118] and we consider $R_2 = 10$ m as the communication range of the secondary users (i.e. maximum distance between a secondary transmitter and a secondary receiver). This value corresponds approximately to the range of the IEEE 802.15.3a specification using UWB. Depending on the interference temperature model, the transmission power μ_2 is defined as μ_2^{id} or μ_2^{gen} and the secondary user's intensity (λ_2) is equal to 3 users per m^2.

The IT model includes not only power from primary and secondary transmitters but also the interference power of another source of interference which as been named as base interference. However, to provide an upper bound on the achievable capacity by the secondary network, we consider environments with no interference (i.e. $\mu_3 = 0$). Finally, the last parameter to complete the system is Interference Temperature Limit (T_L). This parameter was set to 50000 K, same as other studies of the IT model developed to quantify the capacity achieved by the secondary network and the interference caused to the primary network such as [113].

Considering these parameters for the ideal case we use Lemma 3.3 to obtain the SUs capacity for $\alpha = 3$ and $\alpha = 4$. In the first case we obtain an average SUs transmission power (μ_2^{id}) of -99 dBm. With this SU power, the achievable per link capacity (C^{id}) using Eq. 3.7 is 5.8 Kbps. Therefore, the total achievable capacity (C_{total}^{id}) of the secondary network present in a 100m radius secondary cell (R) is 545 Mbps. This value is obtained using Eq. 3.8. In the case of $\alpha = 4$, the average SUs transmission power is -97.37 dBm and the achievable per-link capacity (C^{id}) is 9.22 Kbps. Thus, the total achievable capacity (C_{total}^{id}) for the case $\alpha = 4$ is 869 Mbps.

For the generalized model, in this scenario, the communication is not possible if we consider the same parameters. To allow the transmission of SUs we must increase the bandwidth (B) of the secondary network, which is in fact one of the characteristics of WiMedia's medium access control (MAC) layer [117]. The WiMedia mobile station incorporates a MAC layer providing multimedia Quality of Service (QoS) and a physical layer based on multi-band orthogonal frequency-division multiplexing (MB-OFDM). This technology is well known to have robust link characteristics, meanwhile the multi-band aspect allows spectrum flexibility and support different channel modes. WiMedia's MAC layer uses a bandwidth reservation system called Distributed Reservation Protocol (DRP). The DRP provides a bandwidth reservation system that assures QoS support for multimedia traffic. This ensures that the streaming media will continue to have the bandwidth it needs once a reservation is established and without interference from other users [117]. Therefore and in order to study the performance of the generalized model, the bandwidth (B) of the secondary network was increased of 57 MHz. So, we increase the SUs bandwidth from 528 MHz (i.e. minimum channel bandwidth of WiMedia systems) to 585 MHz.

Considering the same parameters used for the ideal case, but now with a SUs bandwidth (B) of 585 MHz, applying again Lemma 3.3 with a path loss exponent $\alpha = 3$, we obtain analytically an average SUs transmission power (μ_2^{gen}) of -104 dBm. With this SU power, the achievable per link capacity (C^{gen}) using Eq. 3.7 is 2.22 Kbps. Therefore, the total

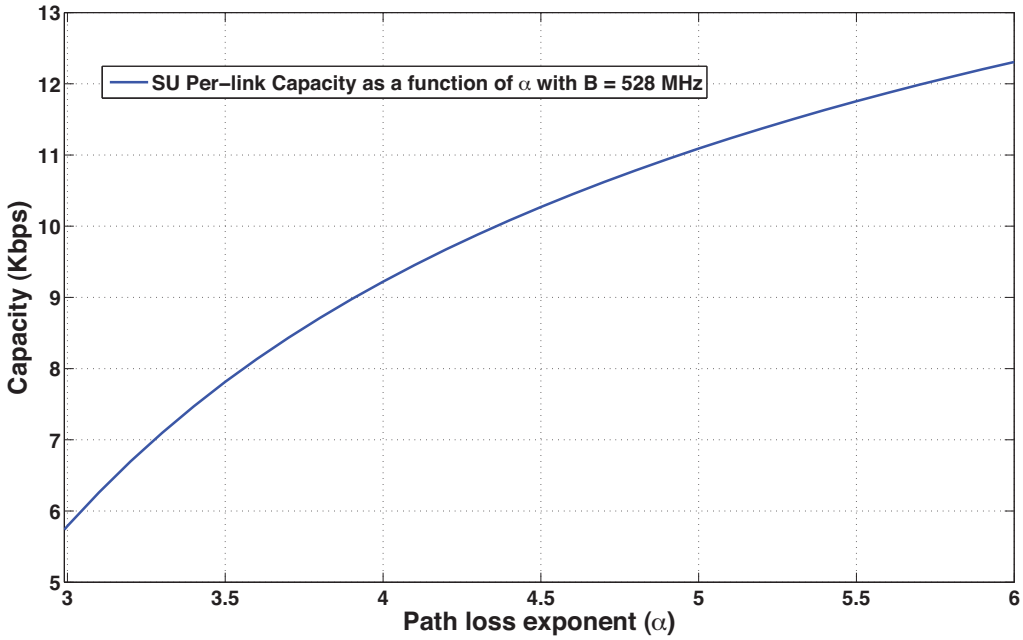


Figure 3.2: Mean SU Per-link Capacity (Kbps) as a function of the path loss exponent (α) for the Ideal case with a secondary user's intensity λ_2 equal to 3 users per m^2.

achievable capacity (C_{total}^{gen}) using Eq. 3.8 of the secondary network present in a cell with 100 m radius (R) is 209 Mbps. In the case of $\alpha = 4$, we obtain better performances compared to the case $\alpha = 3$. The average SUs transmission power is -81.44 dBm and the achievable per-link capacity (C^{gen}) is 297 Kbps. Thus, the total achievable capacity (C_{total}^{gen}) for the case $\alpha = 4$ is 2.79 Gbps.

In Figure 3.2, we analyze the performance of the achievable capacity (C^{id}) as a function of the path loss exponent (α). This figure presents the behavior of C^{id} for typical values of α. Therefore, we consider values from $\alpha = 3$ to $\alpha = 6$. These path loss exponents are used in relatively lossy environments (i.e. $\alpha = 3$) and in indoor environments (i.e. from $\alpha = 4$ to $\alpha = 6$).

To understand the behavior of C^{id} as a function of α plotted in Figure 3.2, we must take into account two different effects: the transmission effect and the reception one. In the transmission effect, with the increase of α, mobile users generate less interference and hence, the available transmission power of the SUs (μ_2^{id}) and the value of K also increase. These results are justified by Lemma 3.1, Eq. 3.3 and Eq. 3.5. On the other hand, the reception effect occurs due to the attenuation of the radio signal as it propagates through space. Here, the received signal decreases with the increase of the path loss exponent. In this case, the reception effect appears for values of α higher than 6. With the SU total per-link capacity ranging from 5.8 Kbps to 12.31 Kbps, from Figure 3.2, we can observe that the ideal IT model is robust against the variation of path loss exponent (α).

Now, using the parameters presented before, we analyze the behavior of the achievable capacity obtained by the secondary network when the secondary users' intensity (λ_2) is increased (i.e. when different load conditions are considered). Figure 3.3 presents the achieved mean SU per-link capacity as a function of the secondary users' intensity (λ_2)

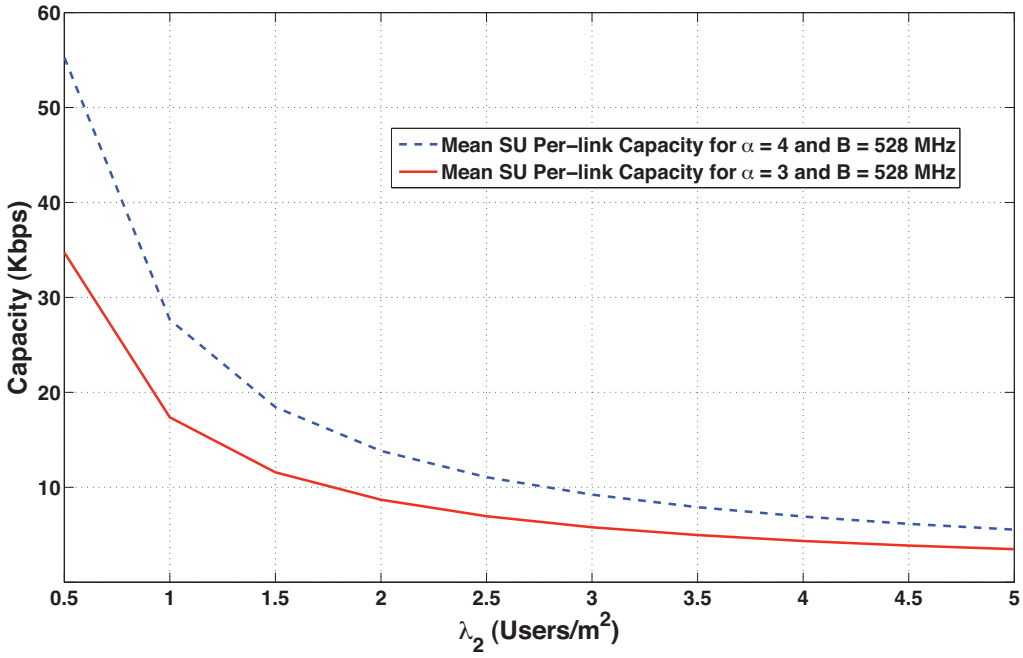


Figure 3.3: Mean SU Per-link Capacity (Kbps) as a function of the secondary users' intensity (λ_2) for the ideal IT model with B = 528 MHz for the cases α=3 and α=4.

following the ideal IT model. As we can see in this figure, with the increase of the secondary users' intensity (λ_2) the allowable SUs transmission power (μ_2^{id}) decrease and hence, the achievable SUs capacity is diminished. These results can be verified in Eq. 3.5 and Eq. 3.5. We also notice in this figure that the achievable per-link capacity in the case of $\alpha = 4$ is slightly higher compared to the case $\alpha = 3$. In the case of the achieved mean total secondary network capacity (C_{total}^{id}) as a function of the secondary users' intensity (λ_2), the total SUs capacity remains almost constant for the ideal case with $\alpha = 3$ and $\alpha = 4$. This behavior obey Eq. 3.8 presented in section 3.3.1.

Figure 3.4 and Figure 3.5 plot the achieved mean SU per-link capacity (C^{gen}) as a function of the secondary users' intensity (λ_2) for the generalized IT model. Here again, in order to analyze the performance of the generalized model the SUs bandwidth (B) was increased from 528 MHz to 585 MHz. In Figure 3.4 and Figure 3.5 we observe the same behavior occurred for the ideal case plotted in Figure 3.3. This is that the achieved mean SU per-link capacity is higher for the case $\alpha = 4$ than for the case $\alpha = 3$. These results are justified by Lemma 3.1, Eq. 3.3, Eq. 3.4 and Eq. 3.5. These expressions states that with the increase of α, MSs generate less interference and hence the available transmission power of the SUs μ_2^{id} or μ_2^{gen} and the value of K also increase. For the generalized approach, in Figure 3.5 we notice that the maximum achievable mean SU per-link capacity for $\alpha = 4$ is 1.74 Mbps, meanwhile the maximum capacity for $\alpha = 3$ in Figure 3.4 is only 13.5 Kbps. In the case of the achieved mean total secondary network capacity (C_{total}^{gen}) as a function of the secondary users' intensity (λ_2), the total SUs capacity remains almost constant for the generalized case with $\alpha = 3$ and $\alpha = 4$.

In order to directly compare the performances in terms of mean SU per-link capacity of the ideal and the generalized IT models, we have set the SUs bandwidth B to 585 MHz

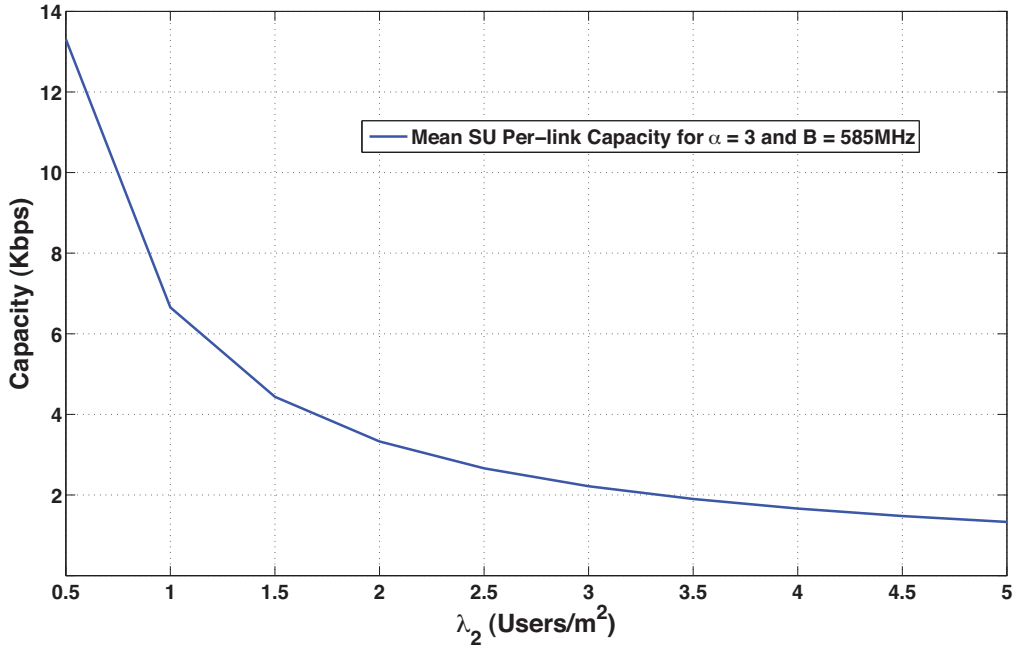


Figure 3.4: Mean SU Per-link Capacity (Kbps) as a function of the secondary users' intensity (λ_2) for the generalized IT model with $B = 585$ MHz for the case $\alpha = 3$

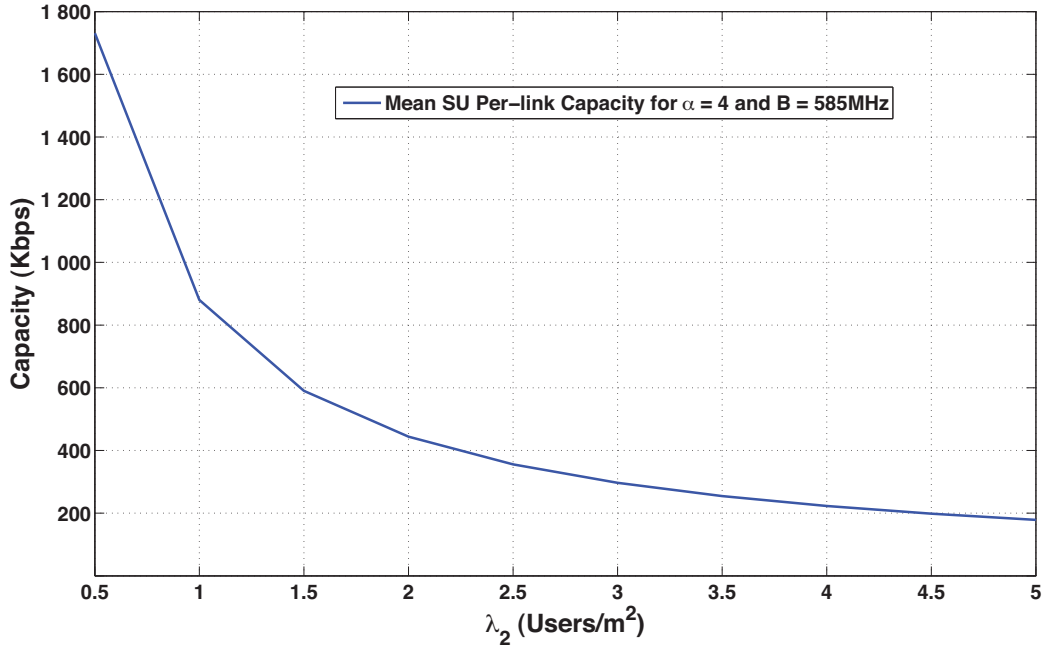


Figure 3.5: Mean SU Per-link Capacity (Kbps) as a function of the secondary users' intensity (λ_2) for the generalized IT model with $B = 585$ MHz for the case $\alpha = 4$.

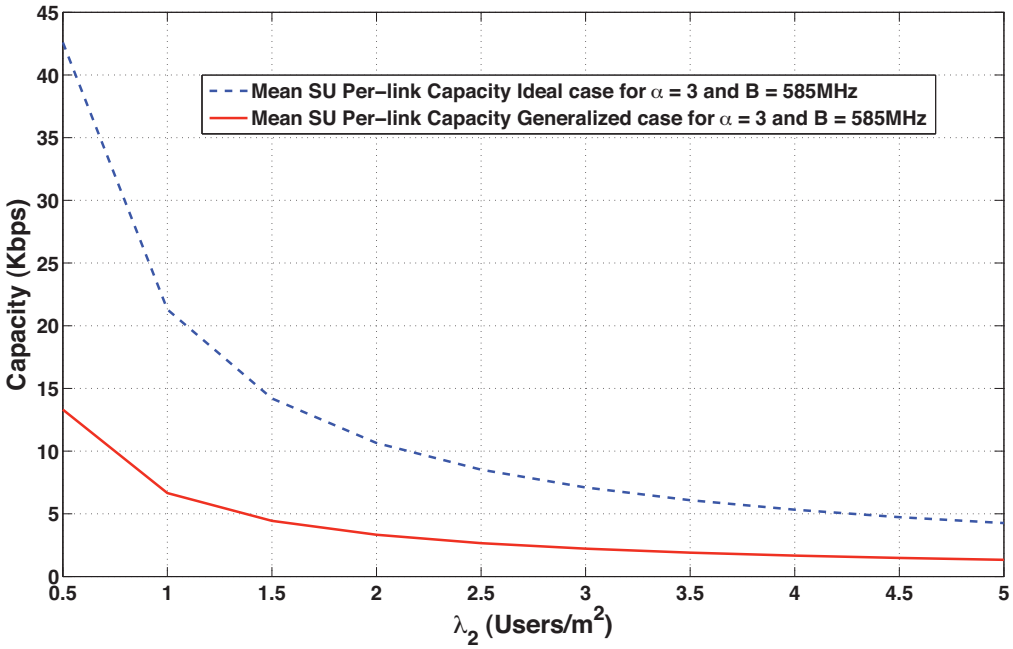


Figure 3.6: Mean SU Per-link Capacity (Kbps) as a function of the secondary users' intensity (λ_2) for the ideal and generalized IT models with B = 585 MHz for the case α = 3.

for both approaches in Figure 3.6 and Figure 3.7. Figure 3.6 plots the mean SU per-link capacity for the case $\alpha = 3$ and Figure 3.7 for the case $\alpha = 4$. As we can see in Figure 3.6 the ideal model outperforms the generalized approach for the case $\alpha = 3$. However, for $\alpha = 4$ the generalized case obtains better performance due to the higher allowable transmission power μ_2^{gen}. This behavior can be verified in Eq. 3.4.

3.4 Upper bound of PU Outage Probability

The latter section considers the constraint given by the Interference Temperature model with respect to the mean transmission power. However, if the distribution of the sum of the interferences is roughly symmetric, about 50% of PUs are not guarenteed the maximum allowed transmission power. This means that this averaged condition can only loosely gives an idea of the effect of interference, but cannot predict how probable is the occurrence of an outage of the PU network. Here, we use the concentration inequalities from Malliavin calculus to find an upper bound probability P_{sup} of outage of a PU due the interference caused by the SUs as a function of μ_2. Therefore, the system can be designed, such that the outage probability of the PUs is smaller than $q = P_{sup}$.

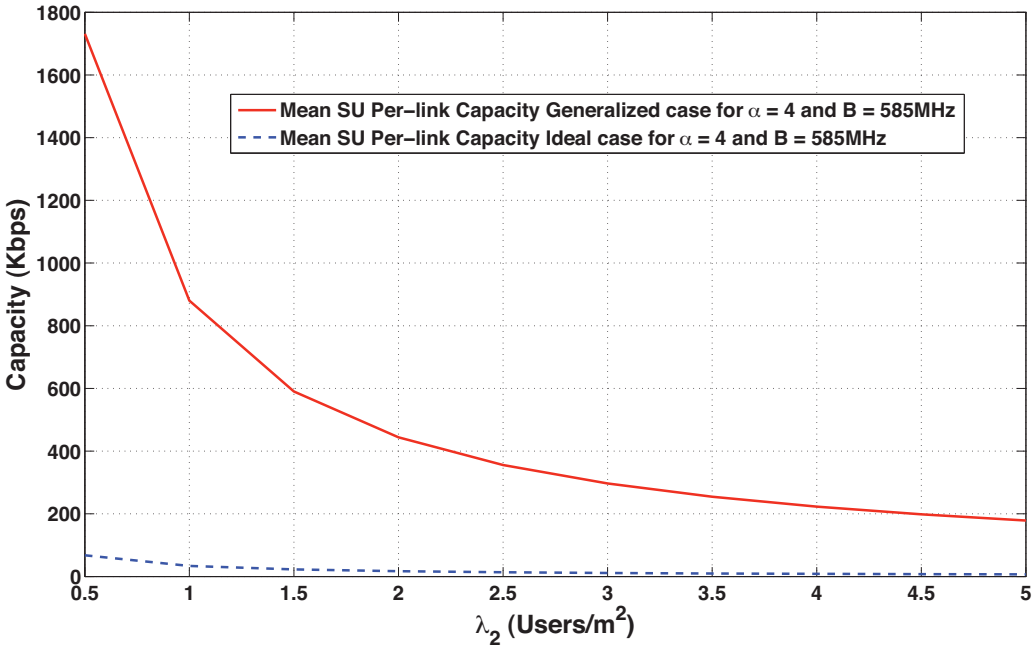


Figure 3.7: Mean SU Per-link Capacity (Kbps) as a function of the secondary users' intensity (λ_2) for the ideal and generalized IT models with $B = 585$ MHz for the case $\alpha = 4$.

3.4.1 Calculations: Upper bound of PU Outage Probability

Using the IT model and considering the ideal case, we want to find an upper bound for the probability of the following event:

$$T_L(f_i)kB_i \geq \sum_{x_i \in \omega_2} p_2(x_i, 0) + \sum_{x_i \in \omega_3} p_3(x_i, 0) \triangleq F,$$

where the primary user to be analyzed is placed at the origin.

Lemma 3.4 *Let ω_A and ω_B be two independent Poisson point processes on $\mathbf{R}^n$ with intensities λ_A and λ_B. Define G_A and G_B as follows:*

$$\begin{aligned} G_A(\omega_A) &= \sum_{X_i \in \omega_A} f_A(X_i), \\ G_B(\omega_B) &= \sum_{X_i \in \omega_B} f_B(X_i), \end{aligned}$$

for f_A and f_B two non-negative measurable real-valued functions. Then, the random variable

$$G = G_A(\omega_A) + G_B(\omega_B)$$

has the same distribution then the marked Poisson point process with intensity $\lambda = \lambda_A + \lambda_B$ and kernel

$$K(x, y) = f_A(x)\delta\left(y - \frac{\lambda_A}{\lambda_A + \lambda_B}\right) + f_B(x)\delta\left(y - \frac{\lambda_B}{\lambda_A + \lambda_B}\right).$$

Proof: We use Eq. 2.3 to obtain

$$\begin{aligned}\mathbf{E}_\lambda\left[e^{-sG}\right] &= \mathbf{E}_\lambda\left[e^{-s(G_A+G_B)}\right] = \mathbf{E}_\lambda\left[e^{-sG_A}\right]\mathbf{E}_\lambda\left[e^{-sG_B}\right] \\ &= \exp\left(-\int(1-e^{-sf_A(x)})\ \mathrm{d}\lambda_A(x)\right) \\ &\qquad \times \exp\left(-\int(1-e^{-sf_B(x)})\ \mathrm{d}\lambda_B(x)\right) \\ &= \exp\left(-\int(1-e^{-s})K(x,dy)\ \mathrm{d}(\lambda_A+\lambda_B)(x)\right),\end{aligned}$$

which concludes the proof. □

Lemma 3.5 *Let ω be a marked Poisson point processes on $\mathbf{R}^n \times \mathbf{R}^m$ with intensity λ and kernel $K(x,y)$ and define G as follows:*

$$G(\omega) = \sum_{X_i \in \omega} f(X_i, Y_i),$$

for f a non-negative measurable real-valued function. Then, for $t \in \mathbf{R}^n$

$$D_t(G) = \int f(t,y)K(t,\ dy)$$

Proof: The proof follows straightforwardly from the application of Definition 2.1 on $F(\omega)$. □

We set $m_F \triangleq \mathbf{E}_\lambda[F]$ and

$$v_F \triangleq \int |D_x F(\omega)\ K(x,dy)|^2 \lambda\ \mathrm{d}x.$$

We obtain P_{sup} via concentration inequalities, using Theorem 2.5

$$\mathbf{P}(F \geq t + m_F) \leq \exp\left(-\frac{t}{2s} g\left(1 + \frac{ts}{v_F}\right)\right).$$

where $g(x) = (1+x)\ln(1+x) - x$. Using Lemmas 3.4 and 3.5, we obtain that

$$\begin{aligned} m_F &= \mathbf{E}_\lambda\left[\sum_{x_i \in \omega_2} p_2(x_i, 0)\right] + \mathbf{E}_\lambda\left[\sum_{x_i \in \omega_3} p_3(x_i, 0)\right], \\ v_F &= \mathbf{E}_\lambda\left[\sum_{x_i \in \omega_2} p_2^2(x_i, 0)\right] + \mathbf{E}_\lambda\left[\sum_{x_i \in \omega_3} p_3^2(x_i, 0)\right], \end{aligned}$$

and we use Lemma 3.1 to find m_F:

$$m_F = \frac{\alpha\pi(\mu_2\lambda_2 + \mu_3\lambda_3)}{(\alpha - 2)}. \tag{3.9}$$

To find v_F, it suffices to use the same lemma exchanging α by 2α and μ_i by μ_i^2:

$$v_F = \frac{2\alpha\pi(\mu_2^2\lambda_2 + \mu_3^2\lambda_3)}{(2\alpha - 2)}. \tag{3.10}$$

Since the function $\max(\mu_i, \mu_i r^{-\alpha})$ is decreasing with respect to r,

$$s = \max(\mu_2, \mu_3).$$

Assuming $\mu_2 \geq \mu_3$ and taking $T_L(f_i)kB_i = t + m_F$, then

$$\mathbf{P}(F \geq T_L(f_i)kB_i) \leq \exp\left(-\frac{T_L(f_i)kB_i - m_F}{2\mu_2} \ln\left(1 + \frac{(T_L(f_i)kB_i - m_F)\mu_2}{v_F}\right)\right) = P_{sup}. \quad (3.11)$$

This inequality holds for $m_F \leq T_L(f_i)kB_i$. The generic case is similar and it suffices to define

$$F \triangleq \frac{B_i}{B} \sum_{x_i \in \omega_1} p_1(x_i, 0) + \sum_{x_i \in \omega_2} p_2(x_i, 0) + \sum_{x_i \in \omega_3} p_3(x_i, 0),$$

so

$$m_F = \frac{\alpha\pi(\frac{B_i}{B}\mu_1\lambda_1 + \mu_2\lambda_2 + \mu_3\lambda_3)}{(\alpha - 2)},$$

and

$$v_F = \frac{2\alpha\pi(\left(\frac{B_i}{B}\mu_1\right)^2\lambda_1 + \mu_2^2\lambda_2 + \mu_3^2\lambda_3)}{(2\alpha - 2)}.$$

In the generic case, s is given by

$$s = \max\left(\frac{B_i}{B}\mu_1, \mu_2, \mu_3\right),$$

and from here we apply Theorem 2.5.

3.4.2 Results: Upper bound of PU Outage Probability

In this section, we take into account a specific outage probability of PUs to design the allowed transmission power for SUs following the ideal IT model. We use the same parameters considered in Section 3.3.2 to compare this power with the results of that section such that we can evaluate the trade-off of system reliability and capacity. Besides, we set $\lambda_2 = 3$ users/m^2 and analyze the results for $\alpha = 3$ and $\alpha = 4$.

We define μ_2^q as the transmission power such that the outage probability of PUs is smaller than q in the ideal case and η as the fraction of this transmission power with respect to μ_2^{id} calculated in the previous section, i.e. $\mu_2^q = \eta(q)\mu_2^{id}$. Since $q(\eta)$ is a bijection on $(0, 1)$, there exists a function $\eta(q)$. We denote also C^q as the mean capacity per link of a SU in function of q. Setting $\lambda_3 = 0$, we can rewrite Eq. 3.11 as a function of these variables to obtain:

$$q = \exp\left(\frac{\lambda_2\pi\alpha}{2(\alpha - 2)}\frac{\eta(q) - 1}{\eta(q)}\ln\left(1 + \frac{2(\alpha - 1)}{(\alpha - 2)}\frac{1 - \eta(q)}{\eta(q)}\right)\right).$$

The function $\eta(q)$ is presented in Figures 3.8 and 3.9 for $\alpha = 3$ and $\alpha = 4$ respectively.

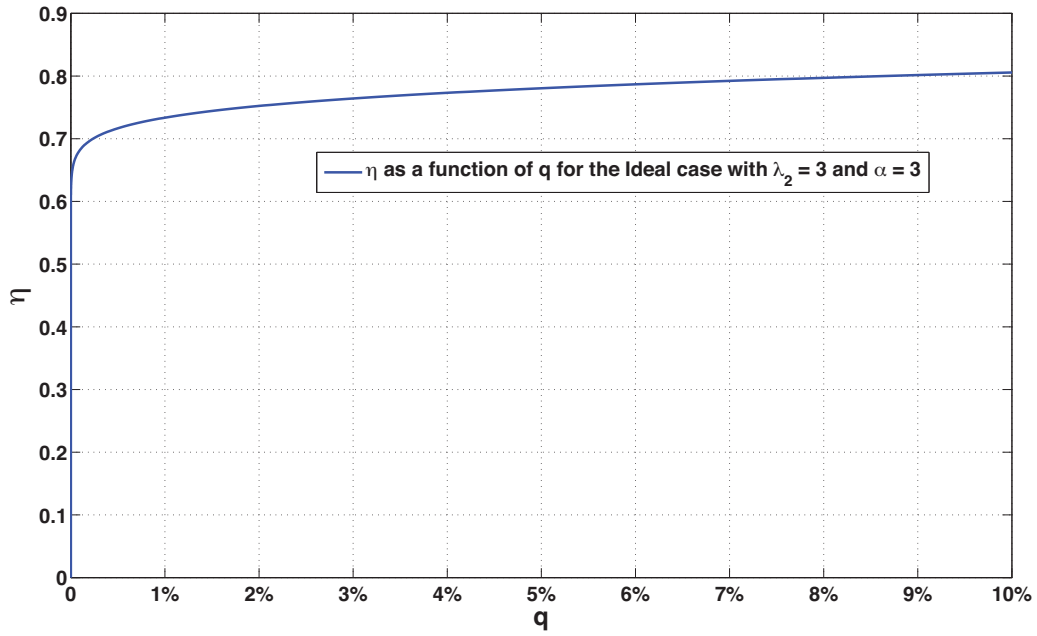


Figure 3.8: Fraction of the transmission power (η) as a function of the outage probability of the PUs (q) for the ideal case with $\alpha = 3$.

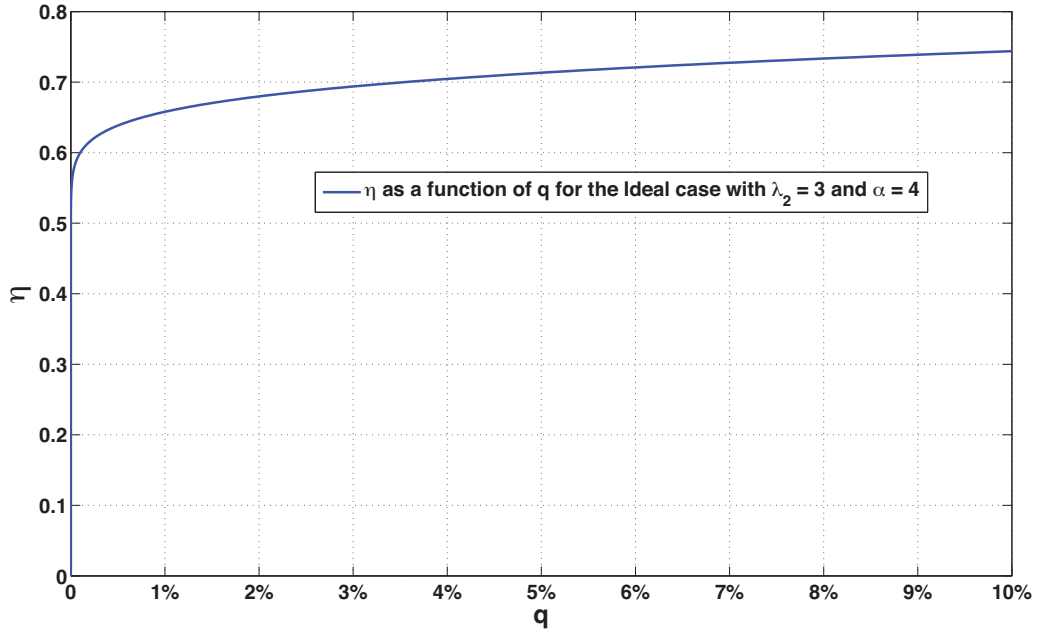


Figure 3.9: Fraction of the transmission power (η) as a function of the outage probability of the PUs (q) for the ideal case with $\alpha = 4$.

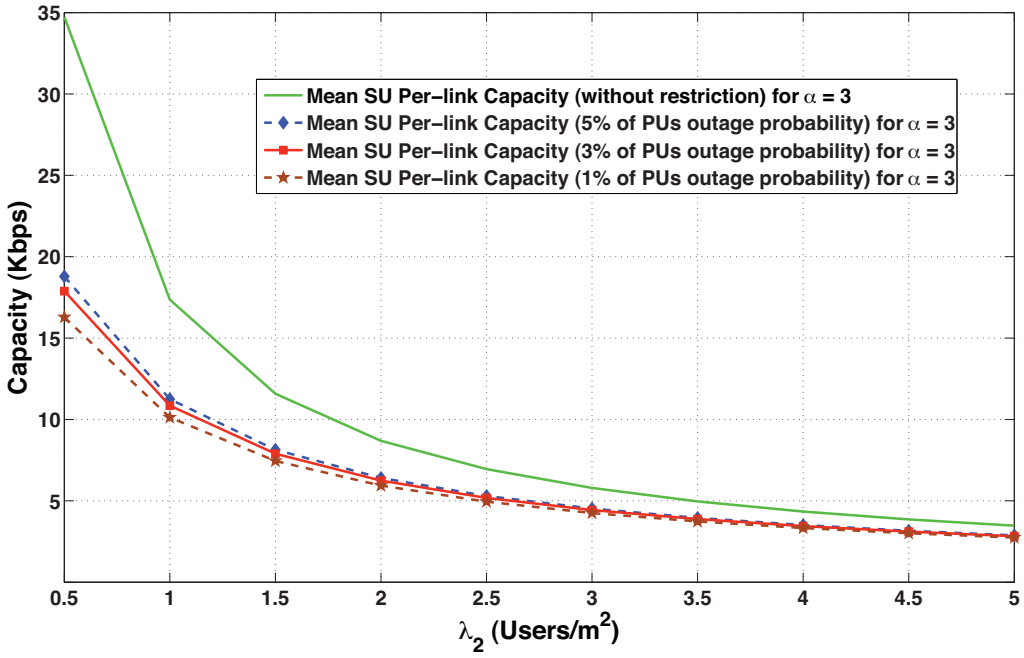


Figure 3.10: Mean SU Per-link Capacity (Kbps) as a function of the secondary users' intensity (λ_2) for different values of the PUs outage probability following the ideal IT model with $B = 528$ MHz for the case $\alpha = 3$.

We can notice from Figure 3.8 and Figure 3.9, that in order to guarantee that the outage probability of PUs remains between 1% and 5%, we must reduce the SUs transmission power μ_2^{id} between a 26% and a 22% respectively for the case $\alpha = 3$ and between a 34% and a 28% respectively for the case $\alpha = 4$.

Now, using Figure 3.10 and Figure 3.11, we evaluate the achieved performance in terms of mean SU per-link capacity for different values of the PUs outage probability following the ideal IT model. Figure 3.10 plots the mean SU per-link capacity as a function of the secondary users' intensity (λ_2) for the case $\alpha = 3$ and Figure 3.11 for the case $\alpha = 4$. Both figures show the achievable mean SU per-link capacity for 1%, 3% and 5% of the PUs outage probability and also plots the original case without restriction on the SUs transmission power μ_2^{id}. As we can see in Figure 3.10 and Figure 3.11 with the increase of λ_2, the difference between the restricted cases and the non restricted case become shorter. This means that in order to guarantee that the outage probability of PUs remains between 1% and 5%, for a scenario with a large number of SUs, the restriction of the SUs transmission power μ_2^{id} does not lead to a significant reduction with respect to the mean SU per-link capacity.

3.5 Summary

This chapter has proposed the utilization of the Poisson Point Process and concentration inequalities as new analytical methods to be applied in the Interference Temperature model. These mathematical tools help us to evaluate, in a simple fashion, the achievable capacity by a secondary network, the interference caused to the primary network and the

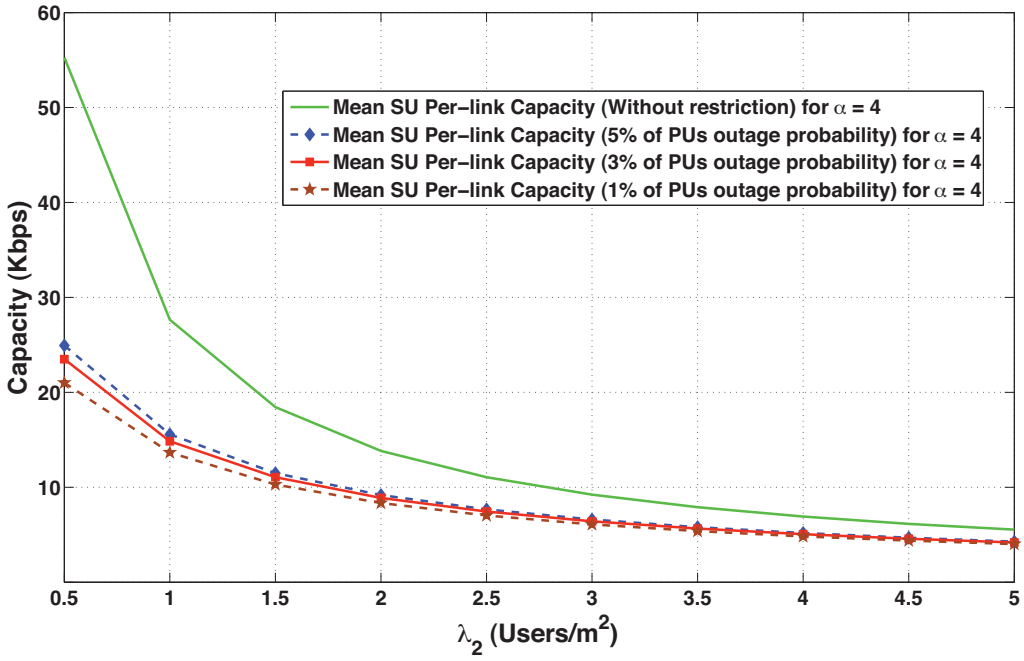


Figure 3.11: Mean SU Per-link Capacity (Kbps) as a function of the secondary users' intensity (λ_2) for different values of the PUs outage probability following the ideal IT model with $B = 528$ MHz for the case $\alpha = 4$.

outage probability of the primary network when the secondary network transmits. For this purpose, we firstly developed the necessary expressions to estimate the mean base interference, the mean interference caused by other SUs and the mean interference caused by active PUs. As we have seen through this chapter the equations developed by our model remain quite simple. Using these results, we estimated the allowed SUs transmission power to guarantee that the PUs activity will not be affected by the SUs transmission. The later analysis was performed for the ideal and the generalized IT models. Afterwards, using the Shannon-Hartley theorem, we derived the expressions of the mean SU per-link capacity and the total secondary network capacity. Finally, by the use of Concentration Inequalities we determine an upper bound on the outage probability of the primary network.

In order to obtain numerical results using our expressions in a realistic scenario, we have examined the achievable capacity of an UWB system, WiMedia, as a secondary network and a UMTS network as the primary network. Our results show that for this scenario, the secondary network achieves a limited performance in terms of capacity, compared to the real capabilities of an UWB standard (e.g. IEEE 802.15.3a). However, these performances can easily be improved if the secondary network operates with a larger channel bandwidth, which is one of the characteristics of WiMedia's MAC layer. Furthermore, we have demonstrated that SUs communication is possible causing only minor damage to primary users following the ideal and the generalized interference temperature model. Moreover, by the use of Concentration Inequalities, we have established that in order to guarantee than only 1% of the PUs is affected by the SUs transmission, it will only cost approximately 25% of the mean allowable SUs transmission power and 20% for a PUs outage probability below 5%. In addition we have demonstrated that, for a scenario with a large number of

secondary users, the restriction of the secondary users transmission power does not lead to a significant reduction of the achievable per-link capacity of the secondary network.

Chapter 4

Upper bound of loss probability of an OFDMA system

4.1 Introduction

In this chapter, we present another application of the Poisson point process modeling a wireless system based on Orthogonal Frequency Division Multiple Access (OFDMA), where future systems will widely rely. OFDMA can satisfy end user's demands in terms of throughput. It also fulfills operator's requirements in terms of capacity for high data rate services. Systems such as 802.16e and 3G-LTE (Third Generation Long Term Evolution) already use OFDMA on the downlink. For the uplink, 802.16e has also adopted OFDMA, while 3G-LTE uses SCFDMA (Single Carrier Frequency Division Multiple Access). OFDMA can also be possibly combined with multiple antenna technology to improve either quality or capacity of systems.

Dimensioning of OFDMA systems is then of the up-most importance for wireless telecommunications industry. As usual, the model introduced in this contribution takes into account the randomness of user locations and user traffic. It provides also an upper bound of loss probability in terms of sub-channels.

This chapter first provides a short introduction to OFDMA air interfaces, by providing some insights on sub-channel concepts and OFDMA jargon (see section 4.2). The dimensioning analytical model is first developed for a deterministic wireless channel, taking only into account the path-loss effect (cf. section 4.3). Section 4.4 analyses a more realistic situation, where wireless channel also encompasses shadowing effects. Section 4.5 extends the results to a multi class user traffic. The accuracy of analytical model is evaluated by comparing them with simulation.

4.2 Introduction to OFDMA air interfaces

OFDM (Orthogonal Frequency Division Multiplex) is a multi carrier technique especially designed for high data rate services. It divides the spectrum in a large number of frequency bands called sub-carriers that overlap partially in order to reduce spectrum occupation. Overlapping is made possible because the different sub-carriers are made orthogonal to each other by choosing a sub-carrier spacing multiple of the inverse of the OFDM symbol duration.

Each sub-carrier has a small bandwidth compared to the coherence bandwidth of the

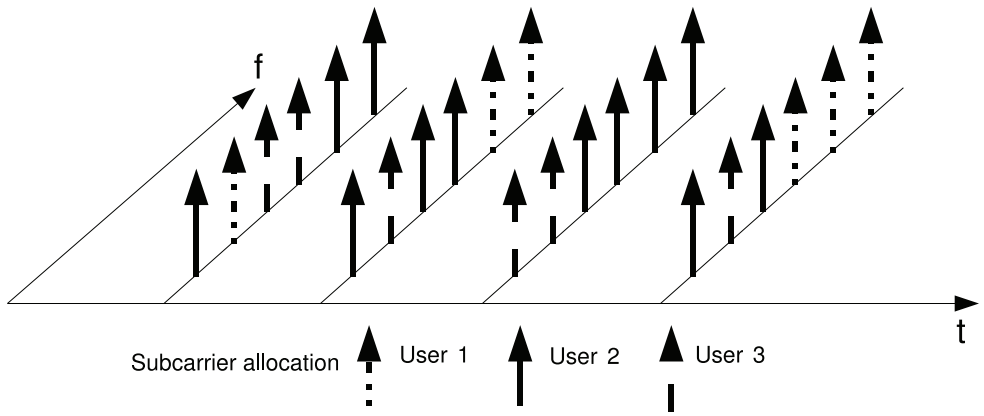


Figure 4.1: OFDMA sub-carrier allocation principle

channel in order to mitigate frequency selective fading. User data is then transmitted in parallel on each sub carrier.

Systems such as ADSL (Asymmetric Digital Subscriber Line), digital audio broadcasting (DAB) and digital video broadcasting (DVB-T) rely on OFDM modulation. Most recently, power line systems (Home Plug) and WiMedia (short range communications) have also adopted OFDM.

In OFDM systems, all available sub-carriers are affected to one user at a given time for transmission. OFDMA extends OFDM by making it possible to share dynamically the available sub-carriers between different users (see figure 4.1). In that sense, it can then be seen as multiple access technique that both combines FDMA and TDMA features.

In practical systems, such as WiMAX or 3G-LTE, the sub-carriers are not allocated individually for implementation reasons mainly inherent to the scheduler design and physical layer signaling. Several sub-carriers are then grouped in sub-channels according to different strategies specific to each system. The unit of resource allocation is the sub-channel.

For example, in WiMAX, there are three modes available for building sub-channels: FUSC (Fully Partial Usage of Sub-channels), PUSC (Partial Usage of Sub-Channels) and AMC (Adaptive modulation and coding). In FUSC, sub-channels are made of sub-carriers spread over all the frequency band. In AMC, the sub-carriers of a sub-channel are adjacent instead of being uniformly distributed over the spectrum. FUSC provides an averaging effect on quality that makes it more suitable for mobile application, while AMC is more adapted for fixed users.

The sub-channel concept makes it easier to schedule radio resources. However, it becomes more difficult to assess channel quality as it is composed by different sub-carriers that can possibly span over several timeslots. An extensive literature has addressed that problem, and we will assume in the following, that whatever the sub-channelization scheme adopted, it is possible to consider an equivalent single channel gain for all the sub-carriers making part of a sub-channel (for example the average of channel gain computed on some sub-carrier pilots). We also assume that subcarrier allocation to different sub-channels is done slot by slot.

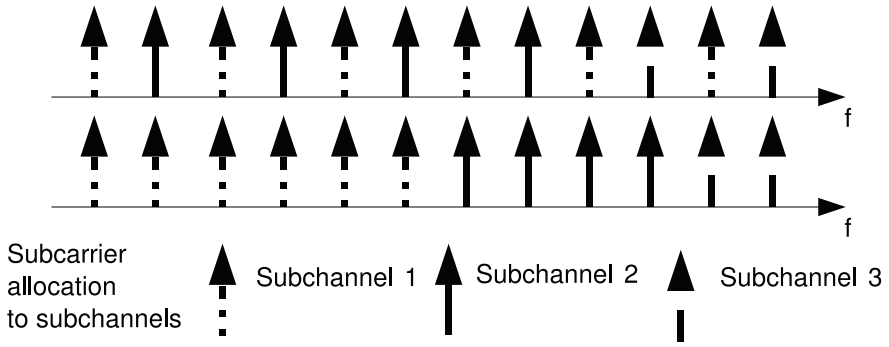


Figure 4.2: OFDMA sub-channel principle

4.3 Upper Bound under a Deterministic Gain

We state the following assumptions:

Assumption 1 *The position of each user is independent of the position of all other. The users are indistinguishable, i.e., the positions are identically distributed.*

Assumption 2 *The time between two consecutive demands of users for service in the system (or inter arrival time) is exponentially distributed.*

We define $\rho(x)$ as the surface density of inter arrival time in $\mathrm{s}^{-1}\mathrm{m}^{-2}$, constant in time. Hence, for a region $H \subseteq B$, the mean inter arrival rate is $h = \int_H \rho(x)dx$ in s^{-1}.

Assumption 3 *The service time for every user is exponentially distributed with mean* $1/\nu$.

Assumption 4 *The cell* C *is circular, with radius* R *and with the antenna in the center.*

Assumption 5 *The channel gain depends only on the distance from the transmitting antenna.*

Assumption 6 *The surface density of inter arrival time is constant.*

These assumptions are commonly done to simplify the mathematical treatment and are quite reasonable. If we show that the point process given by the location of the users is a Poisson process, then it is sufficient to have the two first moments in order to apply theorem 2.5 and then calculate an upper bound P_{sup} for the probability P_{loss} of loosing communications due to a lack of sub-channels. To do this, we consider the following lemma:

Lemma 4.1 *Considering assumptions 1, 2 and 3, the point process* ω *of the active users positions is, in equilibrium, a Poisson process with intensity* $d\lambda(x) = \rho(x)\nu^{-1}\ dx$

Proof: For a region H, in virtue of assumptions 2 and 3, the number of receiving (i.e., active) customers is the same as the number of customers in an $\mathrm{M/M/\infty}$ queue with input

rate h and mean service time ν^{-1}. It is known [100] that the distribution of the number of users U in equilibrium is then

$$\mathbf{P}(U = u) = \frac{(h/\nu)^u}{u!} e^{-h/\nu}.$$

It follows that $\lambda(H)$

$$\lambda(H) = h/\nu = \int_H \frac{\rho(x)}{\nu} \ \mathrm{d}x.$$

Using assumption 1 concludes the proof. □

Without loss of generality, we consider the cell C has its antenna located at the origin. We are looking at evaluating

$$P_{loss} = \mathbf{P}_\lambda \left(\int N \ \mathrm{d}\omega \geq N_0 \right),$$

where $N(x)$ is defined by

$$N(x) = \left\lceil \frac{C_0}{W \log_2 \left(1 + \frac{P_t K \bar{g}}{(I+\eta)\|x\|^\gamma} \right)} \right\rceil,$$

and $\bar{g}$ is the mean gain due to shadowing, C_0 is the throughput requested by users, I is the interference generated by outer cells and η the noise. We will not take into account interference generated by outer cells, so $I = 0$. Note that, with respect to x, N is increasing and piecewise constant. Let R_j, $j = 1, \cdots, N_{max}$ be the values such that $N(x) = j$ for $x \in [R_j, R_{j+1})$. We can determine them by

$$R_j = \left(\frac{P_t K \overline{g}}{\eta(2^{C_0/(jW)} - 1)} \right)^{1/\gamma}.$$

According to Eq. (2.2) for a marked process, it is then clear that

$$\mathbf{E}_\lambda \left[\int N \ \mathrm{d}\omega \right] = \int N \ \mathrm{d}\lambda(x) = \frac{\pi\rho}{\nu} \sum_{j=1}^{N_{max}} j(R_j^2 - R_{j-1}^2).$$

We denote by m_N the last quantity. Moreover,

$$\int N^2 \ \mathrm{d}\lambda(x) = \frac{\pi\rho}{\nu} \sum_{j=1}^{N_{max}} j^2(R_j^2 - R_{j-1}^2).$$

We denote by v_N the last quantity. We take N_0 of the form αm_N, so that according to Theorem 2.5:

$$\mathbf{P}_\lambda \left(\int N \ \mathrm{d}\omega \geq \alpha m_N \right) \leq P_{sup}(\alpha)$$

where

$$P_{sup}(\alpha) = \exp\left(-\frac{v_N}{N_{max}^2} \ g\left(\frac{(\alpha - 1) m_N N_{max}}{v_N} \right) \right).$$

α	1.5	1.6	1.7	1.8	1.9	2
P_{sup}	0.18	0.1	0.04	0.02	0.008	0.003
Δ	0.98	1.1	1.15	1.3	1.3	1.4

Table 4.1: Comparison between P_{sup} and P_{loss} for deterministic gain.

It is then natural to verify how far this bound is from the exact value of the loss probability in simple situations where simulation is available. We used here $\gamma = 2.8$, $C_0 = 200$ kb/s, $W = 250$ kHz and $P_t K/(\eta) = 1 \times 10^6$. For the surface density of inter arrival time we use $\rho = 0.0006$ min^{-1}m^{-2} and the service time is $1/\nu = 1$ min, so, the mean number of users in the system is $\pi R^2 \rho/\nu = 18.85$ users. If we consider the shadowing with $\sigma = \sqrt{10}$ dB and $\mu = 6$ dB, we can use the mean gain g, giving $\overline{g} = 1/12$. Thus, users in the cell boundary use 3 sub-channels, so $N_{max} = 3$. For α varying from 1 to 2, which corresponds here to loss probabilities about 2% or 0.01%, we computed $\Delta = \log_{10} P_{sup}/P_{loss}$.

Though concentration inequalities are usually thought as almost optimal, the results shown in Table 4.1 seem at first glance disappointing. Note though that the computation of the bound is immediate whereas the simulation on a fast PC took several hours to get a decent confidence interval. Note also that the error is about the same order of magnitude as the error made when using a usual trick, which consists in replacing infinite buffers by finite ones in Jackson networks (see [71]). The margin provided by the bounds may be viewed as a protection against errors in the modeling or in the estimates of the parameters.

4.4 Upper Bound under a Random Gain

Let us determine now the upper bound probability P_{sup} for P_{loss} without assumption 5 but holding all other assumptions of the preceding section. Lemma 4.1 still holds, since it is a consequence of assumptions 1, 2 and 3. We also state two other natural assumptions:

Assumption 7 *The random gain is totally described by the log-normal shadowing, with mean μ and standard deviation σ, both in dB.*

For a user at distance d from the origin, the gain is $G = 1/S$, where S follows a log-normal distribution:

$$p_S(y) = \frac{\xi}{\sqrt{2\pi}\sigma y} \exp\left[-\frac{(10\log_{10} y - \mu)^2}{2\sigma^2}\right],$$

where $\xi = 10/\ln 10$.

Assumption 8 *A user is able to receive the signal only if the signal-to-interference ratio is above some constant β_{min}.*

This means, in particular, that the number of sub-carriers needed by a transmitting user is surely bounded by

$$N_{max} = \left\lceil \frac{C_0}{W \log_2(1+\beta_{min})} \right\rceil.$$

The situation is slightly different from that of Section 4.3, since the functional depends on two random factors: positions and gains. Consider now that our configurations are

of the form (x, s) where $x \in \mathbf{R}^2$ is still a position and $s \in \mathbf{R}$ is a gain. Since gain and positions are independent, we then have a Poisson process on $\mathbf{R}^3$ of intensity measure $d\lambda(x) \otimes p_S(y)\ \mathrm{d}y$. Thus we want to evaluate an upper bound of

$$\mathbf{P}_\lambda \left(\int N\ \mathrm{d}\omega \geq N_0 \right)$$

where

$$N(x,\ y) = \left\lceil \frac{C_0}{W \log_2 \left(1 + \dfrac{P_t K}{\eta y \|x\|^\gamma} \right)} \right\rceil .$$

According to Theorem 2.5, we must compute

$$m_N = \int N(x,\ y) p_S(y)\ \mathrm{d}y\ \ \mathrm{d}\lambda(x)$$

and

$$v_N = \sup_\omega \int |D_{x,y} F(\omega)|^2 p_S(y)\ \mathrm{d}y\ \ \mathrm{d}\lambda(x) = \int N^2(x,\ y) p_S(y)\ \mathrm{d}y\ \ \mathrm{d}\lambda(x).$$

Let $\beta_0 = \infty$ and $\beta_j = 2^{C_0/(Wj)} - 1$ for $j = 1, \cdots, N_{max} - 1$. For $j = 1, \cdots, N_{max} - 1$, let

$$A_j = \int_{C \times \mathbf{R}^+} \mathbf{1}_{\{y\|x\|^\gamma \leq P_t K/\eta\beta_j\}} p_S(y)\ \mathrm{d}y\ \mathrm{d}x$$

and $A_0 = 0$.

Lemma 4.2 *For $j = 1, \cdots, N_{max} - 1$,*

$$A_j = \pi R^2 Q(\alpha_j - \zeta \ln R) + \pi e^{2/\zeta^2 + 2\alpha_j/\zeta} Q(\zeta \ln R - 2/\zeta - \alpha_j),$$

where

$$\alpha_j = \frac{1}{\sigma}(10 \log_{10}(P_t K/\eta\beta_j) - \mu) \text{ and } \zeta = \frac{10\gamma}{\sigma \ln 10}.$$

Proof: We can write

$$A_j = \int_C \mathbf{P}(S\|x\|^\gamma \leq \tilde{\beta}_j)\ \mathrm{d}x$$

where $\tilde{\beta}_j = P_t K/\eta\beta_j$. Remind that S is equal in distribution to $\exp(\mathcal{N}(\mu, \sigma^2)\xi)$ with $\xi = \ln(10)/10$. Thus after a few manipulations, we get

$$A_j = 2\pi \int_0^R r\, Q(\alpha_j - \zeta \ln r)\ \mathrm{d}r,$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{u^2}{2})\ \mathrm{d}u.$$

The final result follows by a tedious but straightforward integration by parts. □

Theorem 4.3 *For any function* $\theta : \mathbf{R} \to \mathbf{R}$*,*

$$\int \theta(N(x,\ y))p_S(y)\ dy\ \ d\lambda(x) = \sum_{j=1}^{N_{max}-1} \theta(j)(A_j - A_{j-1}) + \theta(N_{max})(\pi R^2 - A_{N_{max}-1}).$$

Proof: Since N can take only a finite number of values, we have

$$\int \theta(N(x,\ y))p_S(y)\ \mathrm{d}y\ \ \mathrm{d}\lambda(x) = \frac{\rho}{\nu} \sum_{j=1}^{N_{max}} \theta(j) \int_{C\times\mathbf{R}^+} \mathbf{1}_{\{(x,\,y),\ N(x,\,y)=j\}} p_S(y)\ \mathrm{d}y\ \ \mathrm{d}x.$$

Now we see that

$$N(x,\, y) = j \Longleftrightarrow \tilde{\beta}_{j-1} < y\|x\|^\gamma \leq \tilde{\beta}_j,$$

for $j = 1, \cdots, N_{max} - 1$ and $N(x,\, y) = N_{max}$ when $y\|x\|^\gamma > \tilde{\beta}_{N_{max}-1}$. The proof is thus complete. □

We used the same set of values as for the simulation of Section 4.3 together with assumptions 8 and 7 with $\beta_{min} = 0.2$. Results of Table 4.2 show that the theoretical bound is rather stable when gains become stochastic.

α	1.5	1.6	1.7	1.8	1.9	2
P_{sup}	0.2	0.1	0.05	0.02	0.01	0.004
Δ	1.7	1.8	2.1	2.3	2.4	2.6

Table 4.2: Comparison between P_{sup} and P_{loss} for random gain.

4.5 Multi class user traffic

4.5.1 Upper bound of loss probability

We consider in this section, M classes of users. Class j users request a throughput of C_j. The configurations associated to each class are of the form (x, y), where $x \in \mathbf{R}^2$ is a position, $y \in \mathbf{R}$ is a gain. Since gain and positions are independent, we then have for each class of users a Poisson process on $\mathbf{R}^3$ of intensity measure $\lambda_j(x)\ \mathrm{d}x \otimes p_S(y)\ \mathrm{d}y$, where $\lambda_j(x) = \rho_j(x)\nu_j^{-1}$ and j is the user class.

For the sake of computational simplicity, we assume in the following, that $\rho_j(x)$ is constant with respect to x but the theory is still valid unaltered otherwise. Furthermore we consider that the random gain is totally described by the log-normal shadowing, with mean μ and standard deviation σ, both in dB. For a user at distance d from the origin, the gain is $G = 1/S$, where S follows a log-normal distribution as in section 4.4. We also assume that a user is able to receive the signal only if the signal-to-interference ratio is above some constant β_{min}. This means, in particular, that the number of sub-channels needed by a transmitting user of class j is surely bounded by

$$N_j^{max} = \left\lceil \frac{C_j}{W \log_2(1 + \beta_{min})} \right\rceil.$$

Without loss of generality, we consider the cell C has its antenna located at the origin. We are then looking at evaluating

$$\mathbf{P}_\lambda \left(\int N\ \mathrm{d}\omega \geq \mathbf{N}_0 \right),$$

where

$$N(x,\, j,\, y) = \left\lceil \frac{C_j}{W \log_2 \left(1 + \dfrac{P_t K}{\eta y \|x\|^\gamma}\right)} \right\rceil .$$

The functional depends on two aleas: positions and gains. It has also an additional parameter that describes the class of the user.

Theorem 4.4 *With the assumptions of this Section,*

$$\mathbf{P}_\lambda \left(\int N \;\; d\omega \geq \alpha m_N \right) \leq P_{sup}(\alpha)$$

where

$$P_{sup}(\alpha) = \exp\left(-\frac{v}{N_{max}^2} \; g\left(\frac{(\alpha - 1) m N_{max}}{v} \right) \right),$$

with $N_{max} = \max_j N_j^{max}$,

$$m = \sum_{j=1}^{M} \int N(x,\, j,\, y) \lambda_j(x) p_S(y) \;\, dx \;\, dy,$$

and

$$v = \sum_{j=1}^{M} \int N(x,\, j,\, y)^2 \lambda_j(x) p_S(y) \;\, dx \;\, dy.$$

Proof: Let λ_j be the intensity of the Poisson process representing class j customers and $\lambda = \sum_{j=1}^{M} \lambda_j$. Let ω be a Poisson process on $\mathbf{R}^2$ of intensity λ. Consider the probability kernel

$$K(x, \{j\}) = \frac{\lambda_j(x)}{\lambda(x)}.$$

For a configuration $\omega = \{x_n,\, n \geq 1\}$, there is thus a sequence of marks $\{u_n,\, n \geq 1\}$, $u_n \in \{1, \cdots, M\}$ for all $n \geq 1$, corresponding to the position dependent marking according to the kernel K. According to the properties of Poisson process, the process $\omega_j = \{x_n, u_n = j\}$ is a Poisson process of intensity λ_j. Now add to each point of ω, an independent mark y_n, corresponding to the random gain, distributed according to a log-normal distribution. Denote by $\bar{\omega}$ this point process which turns to be a Poisson process since the marks are independent from the positions. From section 2, we know that the process, the atoms of which are $\bar{\omega} = (x_n, u_n, y_n)$, is a Poisson process of intensity $\sum_j K(x, \{j\}) \lambda(x) p_S(y) \; \mathrm{d}x \; \mathrm{d}y \delta_j$:

$$\begin{aligned} \mathbf{E}_\lambda \left[\sum_{n \geq 1} f(X_n, U_n, Y_n) \right] &= \sum_{j=1}^{M} \int f(x, j, y) \frac{\lambda_j(x)}{\lambda(x)} \lambda(x) p_S(y) \; \mathrm{d}x \; \mathrm{d}y \\ &= \sum_{j=1}^{M} \int f(x, j, y) \lambda_j(x) p_S(y) \; \mathrm{d}x \; \mathrm{d}y. \end{aligned}$$

We are thus in position to apply the Theorem 2.5 to the Poisson process $\bar{\omega}$. The difference operator in Definition 2.1, is here equal to

$$D_{x,j,y}F(\bar{\omega}) = F(\bar{\omega} \cup \{x, j, y\}) - F(\bar{\omega}),$$

noting that $\max(D_{x,j,y}F(\bar{\omega})) = 4$. That is to say, we look at the impact of adding a user at position x, with class j and gain y. For $F = \int N \, \mathrm{d}\omega$, we obtain

$$D_{x,j,y}F(\bar{\omega}) = N(x,\, j,\, y) \leq N_j^{max}.$$

Thus, inequality (2.7) holds with $s = \max_j N_j^{max}$,

$$m_N = \sum_{j=1}^{M} \int N(x,\, j,\, y)\lambda_j(x)p_S(y) \, \mathrm{d}x \, \mathrm{d}y,$$

and

$$v_N = \sum_{j=1}^{M} \int N(x,\, j,\, y)^2\lambda_j(x)p_S(y) \, \mathrm{d}x \, \mathrm{d}y.$$

□

Both m and v can be computed taking advantage of the fact that N is piecewise constant (see section 4.4). Let $\beta_0 = \infty$ and $\beta_{j,k} = 2^{C_j/(Wk)} - 1$ for $k = 1, \cdots, N_j^{max} - 1$. For $k = 1, \cdots, N_j^{max} - 1$, let

$$A_{j,k} = \int_{C \times \mathbf{R}^+} \mathbf{1}_{\{y\|x\|^\gamma \leq P_t K/\eta\beta_k\}} p_S(y) \;\, \mathrm{d}y \;\, \mathrm{d}x$$

and $A_0 = 0$. It can proved from results of section V that for $k = 1, \cdots, N_j^{max} - 1$,

$$A_{j,k} = \pi R^2 Q(\alpha_{j,k} - \zeta \ln R) + \pi e^{2/\zeta^2 + 2\alpha_{j,k}/\zeta} Q(\zeta \ln R - 2/\zeta - \alpha_{j,k}),$$

where

$$\alpha_{j,k} = \frac{1}{\sigma}(10 \log_{10}(P_t K/\eta\beta_{j,k}) - \mu) \text{ and } \zeta = \frac{10\gamma}{\sigma \ln 10}.$$

We finally obtain the following formula.

Theorem 4.5 *For any function* $\theta : \mathbf{R} \to \mathbf{R}$,

$$\int \theta(N(x,\, j \;\, y))p_S(y) \;\, dy \;\;\, d\lambda(x) = \sum_{l=1}^{N_j^{max}-1} \theta(l)(A_l - A_{l-1}) + \theta(N_j^{max})(\pi R^2 - A_{N_j^{max}-1}).$$

4.5.2 Numerical application

In this section we will apply the upper bound calculated previously to the dimensioning of sub-channels in a OFDMA system. We consider here a cell, where two classes of users are competing to the access of available sub-channels. More precisely we consider here $M = 2$. The capacities required by each class of user is fixed to $C_1 = 200$ kb/s and $C_2 = 100kb/s$ respectively. The path-loss exponent is fixed to $\gamma = 3.8$ and the sub-channel bandwidth

is equal to $W = 250$ kHz. We also consider $P_t K/\eta = 1 \times 10^{12}$. For the surface density of inter arrival time we use $\rho_1 = 0.0006$ min^{-1}m^{-2} and $\rho_2 = 0.0006$ min^{-1}m^{-2}. The service times are $1/\nu_1 = 1$ min and $1/\nu_2 = 0.5$ min, so the mean number of users in the system is $\pi R^2 \rho_1/\nu_1 = 18.85$ for class 1 users and $\pi R^2 \rho_2/\nu_2 = 9.425$ for class 2 users. We consider the shadowing with $\sigma = \sqrt{10}$ dB and $\mu = 6$ dB. We have also considered $\beta_{min} = 0.2$

We made α varying from 1.6 to 1.8, by steps of 0.05. This corresponds here to an upper bound of loss probability varying between 0.0068 and 0.045 As the analytical expression obtained in the previous section, is an upper bound of the real loss probability, applying it to dimension an OFDMA cell will lead to an over dimensioning in terms of sub-channels. We have computed the number of sub-channels N_0 with the analytical expression of upper bound of loss probability. We have computed by simulation the number of sub-channels required if the upper bound probability is used as the loss probability to dimension the system.

Results of table 4.3 show the over dimensioning is about 20% in terms of sub-channels. We should note that the computation of the upper bound and associated N_0 is immediate whereas the simulation on a fast PC is more tedious to get a decent confidence interval. The margin provided by the bounds may be viewed as a protection against errors in the modeling or in the estimates of the parameters.

α	1.6	1.65	1.7	1.75	1.8
P_{sup}	0.0445	0.0286	0.0180	0.0111	0.0068
$N_{0_{sup}}$	45.2	46.7	48	49.5	50.9
$N_{0_{sim}}$	38	39	40.4	41.6	42.8

Table 4.3: Difference in terms of sub-channels obtained by simulation and analytically. $N_{0_{sup}}$ is the number of sub-channels obtained with the analytical upper bound, whereas $N_{0_{sim}}$ is the one obtained by simulation for the same loss probability value as P_{sup}.

4.6 Summary

Using the concentration and deviation inequalities and the difference operator on Poisson space, we have calculated the upper bound probability of overloading the system by high demand of sub-carriers, over path loss and shadow fading. To do this we have found the first and second moment of the marked Poisson point process of users. It is possible to find an upper bound for the overloading probability, even in a relatively complex system, which is analytically computable in a very simple fashion. The method works for any functional of the configurations, possibly enriched by marks, which depends only on the positions of each user. It does not work for functionals involving relative distance between two or more users. Actually, for such a functional F, there is no bound on $D_x F(\omega)$ valid for all x and ω.

Part II

Algebraic Topology with random Simplicial complexes

Chapter 5

Algebraic Topology

5.1 Introduction

The two most important advantages of the network presented throughout this work are that we do not need to have a central station nor to know the coordinates of the sensors. The first advantage is given by the randomness of the system, and its model was presented in Chapter 2. The second one comes from the topology of the network. The recent works of Ghrist and his collaborators [41, 23] show how, in any dimension, algebraic topology can be used to compute connectivity and coverage of a given configuration of sensors by finding the homology of the network.

In this chapter we introduce some basic tools from algebraic topology. For further reading on topology, see [58, 6, 85]. The main objective of this chapter is to create a link between the physical concepts of an arbitrary sensor network and the its topological representation.

5.2 Theory

Graphs can be generalized to more generic topological objects known as simplicial complexes. While graphs model binary relations, simplicial complexes represent higher order relations. Given a set of points V, a k-simplex is an unordered subset $\{v_0, v_1, \cdots, v_k\}$ where $v_i \in V$ and $v_i \neq v_j$ for all $i \neq j$. The faces of the k-simplex $\{v_0, v_1, \cdots, v_k\}$ are defined as all the $(k-1)$-simplices of the form $\{v_0, \cdots, v_{j-1}, v_{j+1}, \cdots, v_k\}$ with $0 \leq j \leq k$. A simplicial complex is a collection of simplices which is closed with respect to the inclusion of faces, i.e., if $\{v_0, v_1, \cdots, v_k\}$ is a k-simplex then all its faces are in the set of $(k-1)$-simplices.

Given $\mathcal{U} = (U_v, v \in \mathfrak{T})$ a collection of open sets, the Čech complex of $\mathcal{U}$ denoted by $\mathcal{C}(\mathcal{U})$, is the abstract simplicial complex whose k-simplices correspond to $(k+1)$-tuples of distinct elements of $\mathcal{U}$ that have non empty intersection, so $\{v_0, v_1, \cdots, v_k\}$ is a k-simplex if and only if $\bigcap_{i=0}^{k} U_{v_k} \neq \emptyset$.

One can define an orientation for a simplicial complex by defining an order on vertices. A change in the orientation corresponds to a change in the sign of the coefficient as

$$[v_0, \cdots, v_i, \cdots, v_j, \cdots, v_k] = -[v_0, \cdots, v_j, \cdots, v_i, \cdots, v_k].$$

Let X be a simplicial complex. For each integer k, $C_k(X)$ is the vector space spanned by the set of oriented k-simplices of X. The boundary map ∂_k is defined to be the linear

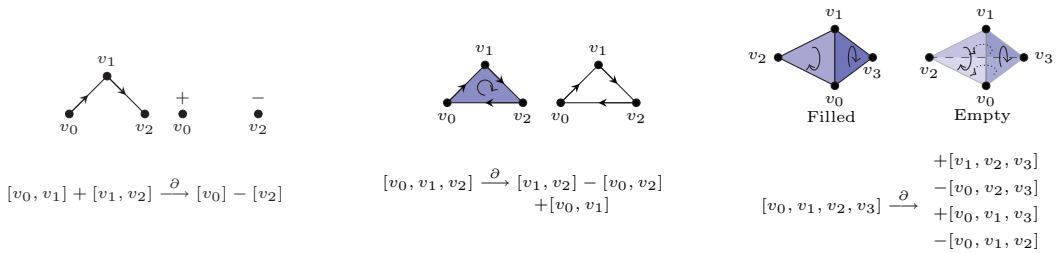


Figure 5.1: Examples of boundary maps. In a) an applications over 1-simplices, in b) we apply over a 2-simplex and in c) over a 3-simplex, turning a filled tetrahedron to an empty one

transformations $\partial_k : C_k \to C_{k-1}$ which acts on basis elements $[v_0, \cdots, v_k]$ via

$$\partial_k[v_0, \cdots, v_k] = \sum_{i=0}^{k} (-1)^k [v_0, \cdots, v_{i-1}, v_{i+1}, \cdots, v_k].$$

Examples of such operations are given in Fig. 5.1.

This map gives rise to a chain complex: a sequence of vector spaces and linear transformations

$$\cdots \xrightarrow{\partial_{k+2}} C_{k+1}(X) \xrightarrow{\partial_{k+1}} C_k(X) \xrightarrow{\partial_k} C_{k-1}(X) \cdots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X).$$

A simple lemma then asserts that for any integer k,

$$\partial_k \circ \partial_{k+1} = 0.$$

The demonstration of this lemma follows straightfowardly from the definition of ∂_k. We define $\ker \partial_k$ as the kernel of ∂_k on C_k (i.e., all closed cycles), $\text{im}\, \partial_k$ as the image of ∂_k (i.e., the $k-1$-simplices that are faces of k-simplices) and if we define

$$Z_k = \ker \partial_k \text{ and } B_k = \text{im}\partial_{k+1},$$

this induces that $B_k \subset Z_k$.

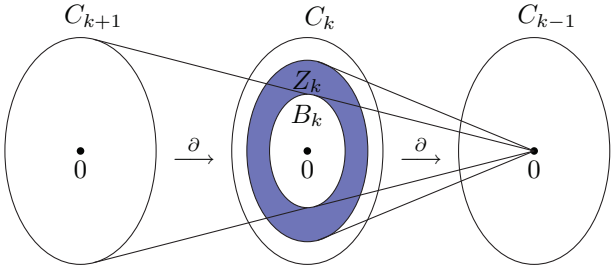


Figure 5.2: A chain complex showing the sets C_k, Z_k and B_k.

The k-dimensional homology of X, denoted $H_k(X)$ is the quotient vector space,

$$H_k(X) = \frac{Z_k(X)}{B_k(X)}.$$

and the k-th Betti number of X is its dimension:

$$\beta_k = \dim H_k = \dim Z_k - \dim B_k.$$

The well known topological invariant named Euler characteristic for X, denoted by $\chi(X)$, is an integer defined by:

$$\chi(X) = \sum_{i=0}^{\infty} (-1)^i \beta_i.$$

Denoting s_k as the number of k-simplices in a simplicial complex X, a well known theorem states that the Euler characteristic is also given by:

$$\chi(X) = \sum_{i=0}^{\infty} (-1)^i s_i.$$

The simplicial complexes we consider are of a special type. They can be considered as a generalization of geometric random graphs.

Definition 5.1 *The d-dimensional torus with with sides a is denoted by* $\mathbb{T}_a^d$.

Definition 5.2 *Given* ω *a finite set of points on the torus. For* $\epsilon > 0$*, we define* $\mathcal{U}_\epsilon(\omega) = \{B_{d_\infty}(v,\, \epsilon),\, v \in \omega\}$ *and* $\mathcal{C}_\epsilon(\omega) = \mathcal{C}(\mathcal{U}_\epsilon(\omega))$*, where* $B_{d_\infty}(x,\, r) = \{y \in \mathbb{T}_a^d,\, \|x - y\|_\infty < r\}$.

Theorem 5.1 *Suppose* $\epsilon < a/4$*. Then* $\mathcal{C}_\epsilon(\omega)$ *has the same homotopy type as* $\mathcal{U}_\epsilon(\omega)$*. In particular they have the same Betti numbers.*

Proof: This will follow from the so-called nerve lemma of Leray, as stated in [102, Theorem 7.26] or [14, Theorem 10.7]. One only needs to check that any non-empty intersection of sets $B_{d_\infty}(v, \epsilon)$ is contractible.

Consider such a non-empty intersection, and let x be a point contained in it. Then, since $\epsilon < a/4$, the ball $B_{d_\infty}(x,\, 4\epsilon)$ can be identified with a cube in the Euclidean space. Then each $B_{d_\infty}(v,\, \epsilon)$ containing x is contained in $B_{d_\infty}(x, 4\epsilon)$, hence also becomes a cube with this identification, hence convex. Then the intersection of these convex sets is convex, hence contractible. □

Definition 5.3 *Let* ω *be a finite set of points in* $\mathbb{T}_a^d$*. For any* $\epsilon > 0$*, the Rips-Vietoris complex of* ω*,* $\mathcal{R}_\epsilon(\omega)$*, is the abstract complex whose k-simplices correspond to unordered* $(k+1)$*-tuples of points in* ω *which are pairwise within distance less than* ϵ *of each other.*

Lemma 5.2 *For the torus* $\mathbb{T}_a^d$ *equipped with the product distance* d_∞*,* $\mathcal{R}_\epsilon(\omega)$ *has the homotopy type of the Čech complex* $\mathcal{C}_{2\epsilon}(\omega)$

The proof is given in [41] in a slightly different context, but it is easy to check that it works here as well. It must be pointed out that Čech and Rips-Vietoris simplicial complexes can be defined similarly for any distance on $\mathbb{T}_a^d$ but it is only for the product distance that the homotopy type of both complexes coincides.

By Lemma 5.2, k points are forming a $(k-1)$-simplex whenever they are two-by-two closer than 2ϵ from each other. We define along the thesis $h(v_1, \cdots, v_k)$ as

$$\begin{aligned} h(v_1, \cdots, v_k) &= h_k(v_1, \cdots, v_k) \\ &= \prod_{1 \le i < j \le k} \mathbf{1}_{[\|v_i - v_j\| < 2\epsilon]}, \end{aligned} \tag{5.1}$$

which determines if a set of k distinct ordered points generates a $(k-1)$-simplex.

Proposition 5.3 *Let $\omega \in \mathbb{T}^d_a$ be a set of points, generating the simplicial complex $\mathcal{C}_\epsilon(\omega)$. Then, if $i > d$, $\beta_i(\omega) = 0$.*

Proof: By Theorem 5.1, $\mathcal{C}_\epsilon(\omega)$ has the same homology as $\mathcal{U}_\epsilon(\omega)$. But $\mathcal{U}_\epsilon(\omega)$ is an open manifold of dimension d, so its Betti numbers $\beta_i(\omega)$ vanish for $i > d$, see for example [50, Theorem 22.24]. □

Proposition 5.4 *Let $\omega \in \mathbb{T}^d_a$ be a set of points, generating the simplicial complex $\mathcal{C}_\epsilon(\omega)$. There are only two possible values for the d-th Betti number of $\mathcal{C}_\epsilon(\omega)$:*

i) $\beta_d = 0$, or

ii) $\beta_d = 1$.

If the second holds, then we also have $\chi(\mathcal{C}_\epsilon(\omega)) = 0$.

Proof: By Theorem 5.1, $\mathcal{C}_\epsilon(\omega)$ has the same homology as $\mathcal{U}_\epsilon(\omega)$. Now, $\mathcal{U}_\epsilon(\omega)$ is an open submanifold of the torus, so there are only two possibilities:

i) $\mathcal{U}_\epsilon(\omega)$ is a strict open submanifold, hence non-compact

ii) $\mathcal{U}_\epsilon(\omega) = \mathbb{T}^d_a$.

In the first case, $\beta_d(\omega) = 0$ by [50, Corollary 22.25]. In the second case $\mathcal{C}_\epsilon(\omega)$ has same homology as the torus, hence $\beta_d(\omega) = 1$ and $\chi(\omega) = 0$. □

Proposition 5.5 *Let X a compact subset of $\mathbf{R}^d$ and consider the map $\tau : X \to Y$ as $x_i = ky_i$ for $x_i \in X$, $y_i \in Y$ and k a positive real constant. Denote by $\tau_*\omega$ the image measure of ω by τ, i.e., $\tau_* : \Omega^X \to \Omega^Y$ maps*

$$\omega = \sum_{i=1}^{\infty} \delta(x_i) \quad \text{to} \quad \tau_* \sum_{i=1}^{\infty} \delta(kx_i).$$

The application $\tau_* : \Omega^X \to \Omega^Y$ maps the Poisson measure λ on Ω^X to the Poisson measure $\lambda_\tau = \lambda/k^d$ on Ω^Y. Moreover, if ϵ_τ is the distance in Y such that two points will be connected, the homology of the two simplicial complexes $\mathcal{C}_\epsilon(\omega)_{\omega \in \mathbb{T}^d_{[a]}}$ and $\mathcal{C}_{\epsilon_\tau}(\tau_*\omega)_{\tau_*\omega \in \mathbb{T}^d_{[ak]}}$ are the same for any k if $\lambda_\tau = \lambda/k^d$ and $\epsilon_\tau = k\epsilon$.

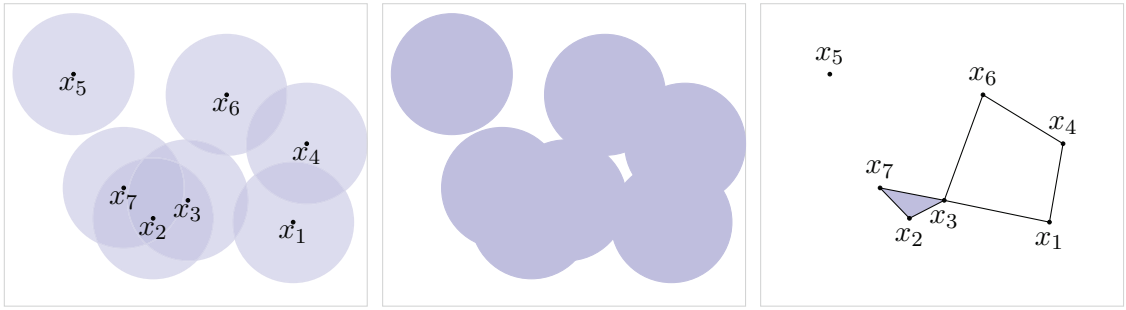
Proof: A slightly changing on Propositions 6.1.7 and 6.1.8 of [99] is enough to show that τ_* maps the Poisson measure λ on Ω^X to the Poisson measure $\lambda_\tau = \lambda/k^d$ on Ω^Y. Then, it suffices to realize that for $x_i \in X$ and for $y_i \in Y$:

$$\begin{aligned} h(x_1, \cdots, x_k) &= \prod_{1 \le i < j \le k} \mathbf{1}_{[\|x_i - x_j\| < 2\epsilon]} \\ &= \prod_{1 \le i < j \le k} \mathbf{1}_{[\|kx_i - kx_j\| < 2k\epsilon]}, \end{aligned}$$

hence

$$h(y_1, \cdots, y_k) = \prod_{1 \le i < j \le k} \mathbf{1}_{[\|y_i - y_j\| < 2\epsilon_\tau]},$$

which concludes the proof. □



a) b) c)

Figure 5.3: The relation between the coverage of a sensor network and its Čech complex. In a) the individual coverages, in b) the network coverage and in c) the correspondent simplicial complex

5.3 Applications to sensor networks

We aim to apply the definitions (k-simplexes, Euller's characteristic and Betti's numbers) and topological properties of the simplicial complexes to the connectivity and coverage problems. In a very intuitive fashion, the number of k-simplices itself shows some tendency in the network: if in two networks with identical number of sensors, one of them has more 1-simplices than the other, this first one has a tendency to be more connected; by the same reason, if a network has more 2-simplices than another one, the region on the first case tends to be well covered.

In a more sofisticated way, Theorem 5.1 formalizes that, in order to determine coverage of sensors, it suffices interpret them as Čech complexes, which is enough to applications that need to determine only connectivity and coverage. In Fig. 5.3 we see an example a sensor network seen as a simplicial complex: in a) we have the individual coverages, in b) we see the network coverage and in c) we have the correspondent simplicial complex. Nonetheless, sensors cannot provide it precisely in applications where they need to communicate with a central station or data have to be passed through them, since communication amoung them is always pairwise. The complex that represent this constraint is the Vietoris-Rips one.

An intepretation to Euller's characteristic is given by Proposition 5.4, where we see that $\chi = 0$ is a necessary condition to have a complete coverage of the torus, and $\beta_d = 1$ is a necessary and sufficient condition. This could in turn translate into conditions for coverage in $[0, a]^d$ when considered as embedded in Euclidean space (i.e. not as a torus), but then one needs to be careful about border effects. For example, one can say that $\beta_d = 1$ is a sufficient condition for coverage of $[\epsilon, a - \epsilon]^d$.

5.4 Summary

We have summarized in the chapter some concepts of algebraic topology to be used in the next two chapters. First, we have defined the concept of simplicial complex, Betti's numbers and Euler characteristic. Then, seeing the space of k-simplices as a linear space, it has been shown that the proper use of the linear operator named boundary map makes it possible to calculate any of those Betti numbers. Moreover, we have stated and proved some propositions that provide interpretations of sensor networks as a simplicial complex,

establishing a relation between the simplices, Betti numbers and Euler characteristic to sensors, connectivity and coverage.

Chapter 6

Moments of k-Simplices and Euler Characteristic

6.1 Introduction

We apply in this chapter, at the same time, the results of Chapters 2 and 5 to characterize a sensor network. The physical features are less in evidence with respect to Chapters 3 and 4 and, although these features are used to justify the mathematical model, we focus almost absolutely in the application of tools of modern mathematics.

Each sensor is represented by its location point with a coverage radius constant given by ϵ. The homology of the coverage of this sensor network, as shown in [41], can be represented by a simplicial complex. This distance ϵ, represents the distance that each sensor can control some environmental information (such as temperature, pressure, presence of an intruder, etc.) around them, but a different interpretation can be done if the sensors are communicating among them. In this case, we suppose that sensors have a power supply allowing them to transmit theirs ID's and, at the same time, sensors have receivers that can identify the transmitted ID's of other sensors above a threshold power. The sensors, knowing mutually the ID's of the close neighbors, are considered connected, creating an information network. The problem remains analogous as the previous one, except that we substitute the coverage radius ϵ by a communication one of $\epsilon/2$. We can see examples of simplicial complexes representations given by sensors communicating among them or monitoring a region in Fig. 6.1.

We consider also that sensors are points of a Poisson point processes. As usual, this assumption reflects the fact that, due the lack of control of the sensors positioning, only a random fraction of the available sensors will actually lie in the target region or some sensors may shut down by running out of energy, moreover, the position of each sensor, a priori, does not interact with the positions other sensors. Instead of using the Euclidean norm, we use the maximum norm along this chapter. We consider this for three reasons: this norm represents a superior and an inferior limits for the Euclidean norm (we can inscribe and circumscribe a circle with two squares); due to the random interactions with the environment (causing shadowing and fading), even the Euclidean norm cannot capture with precision the real behavior of this kind of sensor networks, so we choose the norm that allows us to simplify the calculations; as shown in Lemma 5.2, using the maximum norm, the Čech complex become equal to the Rips-Vietoris complexes. Finally, we assume that sensors lie over d-torus with sides a, $\mathbb{T}_a^d$. This choice was motivated by three factors:

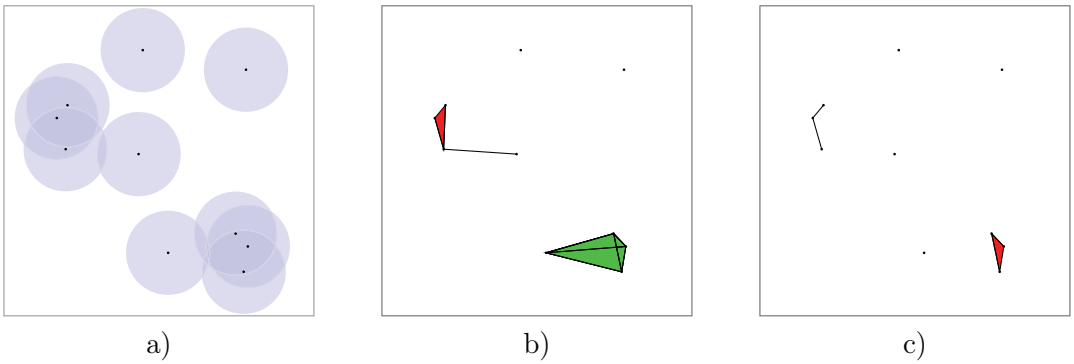


Figure 6.1: a) Sensors and their coverage; b) simplicial complex representation when sensors are monitoring the region; c) simplicial complex representation when sensors are communicating among them.

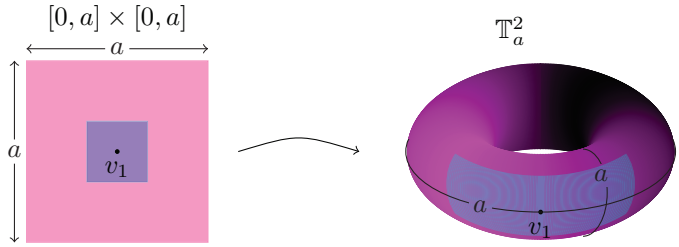


Figure 6.2: Illustration of the coverage of a point and the region where points can lie, in the 2 dimensional case

it avoids the border effects; using Proposition 5.4, it helps to determine weather or not a sensor network in the d-box is completely covered; if ϵ is small compared to a, the calculations for all parameters in the d-torus are a good approximation for the $[0,a]^d$ box. The coverage of a point and the region where points can lie in are illustrated in Fig. 6.1, representing the case where a point is deployed over a plan.

A very few papers deal with the properties of some random simplicial complexes. The most interesting one is [65] which has been followed by [66]. Though there are certain similarities between the work presented in this chapter and that of Kahle, we would like to point out the differences. In [65], the number of points is fixed and the positions are i.i.d. random variables in the plane. It differs from a Poisson point process by the randomness of the number of points. However, for our initial motivation, Poisson process fits better since due to failures or movements, we don't know at each time the number of sensors. Moreover, Kahle is concerned with asymptotic regimes of the mean value of the Betti numbers. We do give exact formulas for any moment of the number of simplices. In addition, by using Malliavin calculus, we go further since we can evaluate the speed of convergence in the CLT and give a concentration inequality to bound the distribution tail of the first Betti number. On the other hand, [66] obtains results for some ranges of ϵ, and particularly for the subcritical range asymptotic behaviors are found for β_k, including the mean, the

variance and the distribution.

Our method goes as follows: We write the numbers of k-simplices (i.e. points, edges, triangles, tetrahedron, etc.) as iterated integrals with respect to the underlying Poisson process. Then, the computation of the means simply reduces to the computation of deterministic iterated integrals thanks to Campbell formula. By using the definition of the Euler characteristic as an alternating sum of the numbers of simplices, we find its expectation. The point is that even if the summing index goes to infinity, there are so many cancellations that the expectation of χ depends only on the d-th power of the intensity of the Poisson process where d is the dimension of the underlying space. Using the multiplication formula of iterated integrals, one can reproduce the same line of thought for higher order moments to the price of an increased complexity in the computations. We obtain closed form formulas for the variance of the number of k-simplices and of the Euler characteristic and series expansions for higher order moments. Using Stein's method mixed with Malliavin calculus, we generalize the results of [96] by proving a precise (i.e. with speed of convergence) CLT for sub-complexes count. As it turns out, the speed of convergence is of the order of $\lambda^{-1/2}$.

The chapter is organized in the following way: the calculations and analytical results for the mean of simplices and Euler characteristics are presented in the Section 6.2; Section 6.3 presents the calculations for the variance and covariance of the number of simplices which leads to an expression of the Euler characteristic; next, in Section 6.4, we use the strategies calculations of the previous sections to find the third moment of k-simplices and then a method to express the n-th moment is presented; in Section 6.5, a theorem showing the convergence in law of the number of connected simplicial complex is proved.

6.2 First Moment

Consider that a Poisson point process ω generates a Čech complex $\mathcal{C}_\epsilon(\omega)$, and, even though the number of k-simplices, the Betti's number and the Euler characteristic are functions of $\mathcal{C}_\epsilon(\omega)$, we denote them, respectively, $N_{k+1}(\mathcal{C}_\epsilon(\omega)) = N_{k+1}$, $\beta_k(\mathcal{C}_\epsilon(\omega)) = \beta_k$ and $\chi(\mathcal{C}_\epsilon(\omega)) = \chi$. In this section, we evaluate the mean of the number of $k-1$-simplices, $\mathbf{E}_\lambda[N_k]$ and the mean Euler characteristic, $\mathbf{E}_\lambda[\chi]$.

Theorem 6.1 *Let $\epsilon \leq a/6$. Then, the mean number of $(k-1)$-simplices $N_k(\mathcal{C}_\epsilon(\omega))$ is given by*

$$\mathbf{E}_\lambda[N_k] = \frac{\lambda^k (ak(2\epsilon)^{k-1})^d}{k!}.$$

Proof: If $(u_{i,1}, \cdots, u_{i,d})$ represents the coordinates of a point v_i, we can separate the indicator function as follows:

$$\mathbf{1}_{[\|v_i - v_j\| < 2\epsilon]} = \prod_{l=1}^{d} \mathbf{1}_{\left[\{|u_{i,l} - u_{j,l}| < 2\epsilon\} \cup \{|u_{i,l} - u_{j,l}| > a - 2\epsilon\}\right]}.$$

According to 5.1, the number of $(k-1)$-simplices can be counted by the expression:

$$N_k = \frac{1}{k!} \sum_{\substack{v_1, \cdots, v_k \in \omega \\ v_i \neq v_j \ if \ i \neq j}} h(v_1, \cdots, v_k).$$

Since ω is a Poisson point process of intensity λ, for a Borel sets A_i, i integer we have:

$$\mathbf{E}_\lambda\left[\sum_{\substack{v_1,\cdots v_k\in\omega\\ v_i\neq v_j \ if \ i\neq j}} h(v_1,\cdots,v_k)\right] = \lambda^k \int_{A_1}\cdots\int_{A_k} h(v_1,\cdots,v_k)\ \mathrm{d}v_1\ldots\ \mathrm{d}v_k.$$

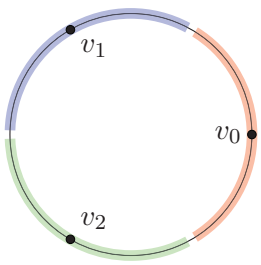
Taking $\frac{1}{k!}h = f$, $A_i = \mathbb{T}_a^d$ and defining

$$\{|x_i - x_j| < 2\epsilon\} \cup \{|x_i - x_j| > a - 2\epsilon\} = d_\epsilon(x_i, x_j),$$

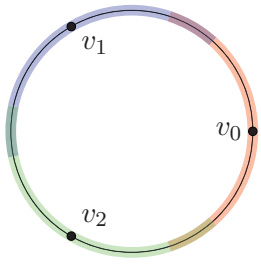
we have:

$$\begin{aligned}
\mathbf{E}_\lambda[N_k] &= \frac{\lambda^k}{k!}\int_{\mathbb{T}_a^d}\cdots\int_{\mathbb{T}_a^d} h(v_1,\cdots,v_k)\ \mathrm{d}v_1\ldots\ \mathrm{d}v_k \\
&= \frac{\lambda^k}{k!}\prod_{l=1}^{d}\int_0^a\cdots\int_0^a \prod_{1\le i<j\le k}\mathbf{1}_{[d_\epsilon(u_{i,l},u_{j,l})]}\ \mathrm{d}u_{1,l}\ldots\ \mathrm{d}u_{k,l} \\
&= \frac{\lambda^k}{k!}\left(\int_0^a\cdots\int_0^a \prod_{1\le i<j\le k}\mathbf{1}_{[d_\epsilon(x_i,x_j)]}\ \mathrm{d}x_1\ldots\ \mathrm{d}x_k\right)^d \qquad (6.1) \\
&= \frac{\lambda^k}{k!}\left(\int_0^a \underbrace{\int_0^a\mathbf{1}_{[d_\epsilon(x_k,x_{k-1})]}\cdots\int_0^a}_{m\text{ integrals}}\prod_{i=k-m+1}^{k}\mathbf{1}_{[d_\epsilon(x_i,x_{k-m})]}\cdots\right. \\
&\qquad\left.\int_0^a\prod_{i=2}^{k}\mathbf{1}_{[d_\epsilon(x_i,x_1)]}\ \mathrm{d}x_1\ldots\ \mathrm{d}x_k\right)^d. \qquad (6.2)
\end{aligned}$$

Since $6\epsilon \le a$, the integration region is convex (see Fig. 6.3). Then, we can rewrite the



a)



b)

Figure 6.3: a) Maximum cover in $\mathbb{T}_a$ and $\epsilon = a/6$. The red region shows the cover of a point v_0, the blue region is the cover of v_1 and the green region is the cover of v_2. b) Maximum cover in the same conditions of a) when $\epsilon = a/5$. In this case, we the three covers intersect each other pairwise, but there is no intersection of the three covers.

integral in Eq. (6.2) as

$$\int_0^a \int_0^a \mathbf{1}_{[d_\epsilon(x_k,x_{k-1})]} \cdots \int_0^a \prod_{i=2}^{k} \mathbf{1}_{[d_\epsilon(x_i,x_1)]} \, \mathrm{d}x_1 \ldots \, \mathrm{d}x_k = \int_0^a \int_{x_k-2\epsilon}^{x_k+2\epsilon} \int_{\max(x_k,x_{k-1})-2\epsilon}^{\min(x_k,x_{k-1})+2\epsilon} \cdots \int_{\max(x_k,\cdots,x_2)-2\epsilon}^{\min(x_k,\cdots,x_2)+2\epsilon} \mathrm{d}x_1 \ldots \, \mathrm{d}x_k. \quad (6.3)$$

Then, consider a subset of the integration region $[0,a]^d$ of Eq. (6.3), defined as $A_{1,2,\cdots,k}$, such that $x_1 \geq x_1 \geq \ldots \geq x_k$. In this case, we can write the integral over $A_{1,2,\cdots,k}$ as:

$$\int_0^a \int_{x_k-2\epsilon}^{x_k+2\epsilon} \cdots \int_{\max(x_k,\cdots,x_2)-2\epsilon}^{\min(x_k,\cdots,x_2)+2\epsilon} \mathbf{1}_{[x_i \geq x_j \text{ if } i \leq j]} \, \mathrm{d}x_1 \ldots \, \mathrm{d}x_k = \int_0^a \int_{x_k}^{x_k+2\epsilon} \int_{x_{k-1}}^{x_k+2\epsilon} \cdots \int_{x_2}^{x_k+2\epsilon} \mathrm{d}x_1 \ldots \, \mathrm{d}x_k.$$

For $\sigma \in \mathfrak{S}_k$, we denote by A_σ the set $A_{\sigma(1),\cdots,\sigma(k)}$. Then,

$$\bigcup_{\sigma \in \mathfrak{S}_k} A_\sigma = [0,a]^d.$$

Moreover, since the function $h(x_1, \cdots, x_k)$ is symmetric, we can exchange the integration variables in the integral of Eq. (6.1) without changing its result. As a consequence, if $\sigma \in \mathfrak{S}_k$,

$$\int_{A_\sigma} \cdots \int \prod_{1 \leq i < j \leq k} \mathbf{1}_{[d_\epsilon(x_i,x_j)]} \, \mathrm{d}x_1 \ldots \, \mathrm{d}x_k = \int_{A_\sigma} \cdots \int \prod_{1 \leq i < j \leq k} \mathbf{1}_{[d_\epsilon(x_i,x_j)]} \, \mathrm{d}x_{i_1} \ldots \, \mathrm{d}x_{i_k}.$$

Thus, we have

$$\begin{aligned} \int_0^a \cdots \int_0^a \prod_{1 \leq i < j \leq k} \mathbf{1}_{[d_\epsilon(x_i,x_j)]} \, \mathrm{d}x_1 \ldots \, \mathrm{d}x_k & \\ &= \sum_{\sigma \in \mathfrak{S}_k} \int_{A_\sigma} \cdots \int \prod_{1 \leq i < j \leq k} \mathbf{1}_{[d_\epsilon(x_i,x_j)]} \, \mathrm{d}x_{i_1} \ldots \, \mathrm{d}x_{i_k} \\ &= k! \int_{A_{\mathrm{Id}}} \cdots \int \prod_{1 \leq i < j \leq k} \mathbf{1}_{[d_\epsilon(x_i,x_j)]} \, \mathrm{d}x_1 \ldots \, \mathrm{d}x_k \\ &= k! \int_0^a \int_{x_k}^{x_k+2\epsilon} \int_{x_{k-1}}^{x_k+2\epsilon} \cdots \int_{x_2}^{x_k+2\epsilon} \mathrm{d}x_1 \ldots \, \mathrm{d}x_k. \end{aligned}$$

Then, by the change of variables $y_{k-i} = (x_{k-i} - x_k)/2\epsilon$ for $i = \overline{1,k}$, we get:

$$k! \int_0^a \int_{x_k}^{x_k+2\epsilon} \int_{x_{k-1}}^{x_k+2\epsilon} \cdots \int_{x_2}^{x_k+2\epsilon} \mathrm{d}x_1 \ldots \, \mathrm{d}x_k = (2\epsilon)^{k-1} k! \int_0^a \int_0^1 \int_{y_{k-1}}^1 \cdots \int_{y_2}^1 \mathrm{d}y_1 \ldots \, \mathrm{d}y_{k-1} \, \mathrm{d}x_k.$$

The integral in the right-handed term is evaluated substituting all y_i for $y_i' - 1$, so

$$(2\epsilon)^{k-1}k! \int_0^a \int_{-1}^0 \int_{y_{k-1}'}^0 \dots \int_{y_2'}^0 \mathrm{d}y_1' \dots \mathrm{d}y_{k-1}' \, \mathrm{d}x_k = (2\epsilon)^{k-1}k! a \frac{1}{(k-1)!} = a(2\epsilon)^{k-1}k. \quad (6.4)$$

Finally, plug Eq. (6.4) into Eq. (6.2) to obtain:

$$\mathbf{E}_\lambda [N_k] = \frac{\lambda^k (ak(2\epsilon)^{k-1})^d}{k!},$$

and thus the proof is complete. □

Remark 6.1 The possibility of writing $\mathbf{E}_\lambda [N_k]$ as Eq. (6.1) is due the fact that we use the maximum norm. This simplifies the calculations since we can treat each component individually. However, considering the Euclidean norm it is still possible to find a closed-form expression for $\mathbf{E}_\lambda [N_2]$ and $\mathbf{E}_\lambda [N_3]$ for the Rips-Vietoris and the Čech complexes in $\mathbb{T}_a^2$. For $\mathbf{E}_\lambda [N_2]$, we have:

$$\mathbf{E}_\lambda [N_2] = \frac{1}{2}\mathbf{E} \sum_{\substack{x_1, x_2 \in \\ \omega \\ x_1 \neq x_2}} h(x_1, x_2) = \frac{\lambda^2}{2} \int_{\mathbb{T}_a^2} \int_{\mathbb{T}_a^2} h(x_1, x_2) dx_2 dx_1.$$

Then, if $b(x, r)$ is the ball centered at x and radius r. Then

$$\mathbf{E}_\lambda [N_2] = \frac{\lambda^2}{2} \int_{\mathbb{T}_a^2} \int_{b(x_1,\epsilon)} dx_2 dx_1 = \frac{\lambda^2}{2} \int_{\mathbb{T}_a^2} \pi\epsilon^2 dx_1,$$

so

$$\mathbf{E}_\lambda [N_2] = \frac{\pi\lambda^2\epsilon^2 a^2}{2}.$$

We write $\mathbf{E}_\lambda [N_3]$, as

$$\mathbf{E}_\lambda [N_3] = \frac{1}{3!}\mathbf{E} \sum_{\substack{x_1, x_2, x_3 \in \\ \omega \\ x_1 \neq x_2, x_2 \neq x_3}} h(x_1, x_2, x_3) = \frac{\lambda^3}{6} \int_{\mathbb{T}_a^2} \int_{\mathbb{T}_a^2} \int_{\mathbb{T}_a^2} h(x_1, x_2, x_3) dx_3 dx_2 dx_1$$

Let R_3 be the region where x_3 can lie to form a two simplex. Figure 6.4 presents this region. The surface of this region $S(R_3)$ in polar coordinates (r, θ) is given by

$$S(R_3) = 2\epsilon^2 \arccos(r/(2\epsilon)) - r\sqrt{\epsilon^2 - r^2/4}$$

hence:

$$\begin{aligned} \mathbf{E}_\lambda [N_3] &= \frac{\lambda^3}{6} \int_{\mathbb{T}_a^2} \int_{b(x_1,\epsilon)} \int_{R_3} dx_1 dx_2 dx_3 \\ &= \frac{2\pi\lambda^3 S(B)}{6} \int_0^\epsilon 2r\epsilon^2 \arccos\left(\frac{r}{2\epsilon}\right) - r^2\epsilon\sqrt{1 - \left(\frac{r}{2\epsilon}\right)^2} dr \end{aligned}$$

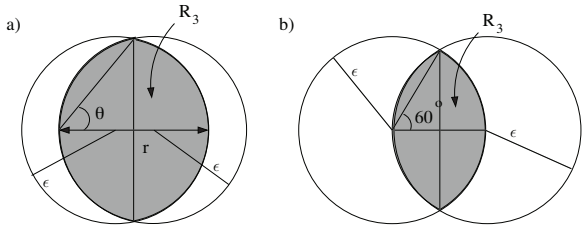


Figure 6.4: a) Region R_3 b) Limit of the region R_3

If we substitute $u = r/(2\epsilon)$, then:

$$\mathbf{E}_\lambda[N_3] = \frac{2\pi\lambda^3 S(B)}{6} 8\epsilon^4 \int_0^{\frac{1}{2}} u \arccos(u) - u^2\sqrt{1-u^2}du = \frac{2\pi\lambda^3 S(B)}{6} 8\epsilon^4 (I_1 - I_2)$$

where I_1 and I_2 are well known integrals and after solving them we find:

$$\mathbf{E}_\lambda[N_3] = \pi\left(\pi - \frac{3\sqrt{3}}{4}\right)\frac{\lambda^3 a^2 \epsilon^4}{6} \text{ for the Rips complex}\cdot$$

We can proceed do the same calculations to obtain the mean of the mean number of 2-simplices on a Čech complex, defining carefully the region of integration.

$$\mathbf{E}_\lambda[N_3] = \frac{2\lambda^k a^2 (\pi\epsilon^2)^2}{3} \text{ for the Čech complex,}$$

Corollary 6.2 *Let $\epsilon \leq a/6$. Then, the mean number of $(k-1)$-simplices N_k given that $N_1 = n$ is given by*

$$\mathbf{E}[N_k | N_1 = n] = \binom{n}{k} k^d \left(\frac{2\epsilon}{a}\right)^{d(k-1)}.$$

Proof: We use the depoissonization to obtain the mean of simplexes for a Binomial process with n points, $\mathbf{E}[N_k|N_1 = n]$, by means of the pair of Poisson transform:

$$z^k = \sum_{n=0}^{\infty} \alpha_n \frac{z^n e^{-z}}{n!} \Leftrightarrow \alpha_n = \begin{cases} \frac{n!}{(n-k)!} & , \text{ if } n \geq k, \\ 0 & , \text{ otherwise.} \end{cases}$$

□

Consider now the Bell's polynomial $B_d(x)$, defined as (see [13])

$$B_n(x) = \sum_{k=0}^{n} \left\{ {n \atop k} \right\} x^k,$$

where n is an positive integer and $\left\{ {n \atop k} \right\}$ is the Stirling number of the second kind. An equivalent definition of B_n can be:

$$B_n(x) = e^{-x} \sum_{k=0}^{\infty} \frac{x^k k^d}{k!}\cdot$$

Theorem 6.3 *Let $\epsilon \leq a/6$. The mean of the Euler characteristic mean of the simplicial complex $\mathcal{C}_\epsilon(\omega)$ is given by*

$$\mathbf{E}_\lambda[\chi] = \left(\frac{a}{2\epsilon}\right)^d e^{-\lambda(2\epsilon)^d}(-B_d(-\lambda(2\epsilon)^d)).$$

Proof: Since

$$N_k \leq \frac{1}{k!}\prod_{j=0}^{k-1}(N_1 - j) \leq \frac{N_1^k}{k!},$$

then

$$\sum_{k=1}^{\infty} N_k \leq \sum_{k=1}^{\infty}\frac{N_1^k}{k!} = e^{N_1}.$$

As $\mathbf{E}_\lambda\left[e^{N_1}\right] < \infty$, we have $\mathbf{E}_\lambda\left[-\sum_{k=1}^{\infty}(-1)^k N_k\right] = -\sum_{k=1}^{\infty}(-1)^k \mathbf{E}_\lambda[N_k]$ and

$$\begin{aligned}
\mathbf{E}_\lambda[\chi] &= \mathbf{E}_\lambda\left[-\sum_{k=1}^{\infty}(-1)^k N_k\right] \\
&= -\sum_{k=1}^{\infty}(-1)^k \mathbf{E}_\lambda[N_k] \\
&= -\sum_{k=1}^{\infty}(-1)^k \frac{\lambda^k (ak(2\epsilon)^{k-1})^d}{k!} \\
&= \frac{a^d}{-(2\epsilon)^d}\sum_{k=0}^{\infty}\frac{(-\lambda(2\epsilon)^d)^{k+1}(k+1)^d}{(k+1)!} \\
&= \frac{a^d e^{-\lambda(2\epsilon)^d}}{-(2\epsilon)^d} e^{\lambda(2\epsilon)^d}\sum_{k=0}^{\infty}\frac{(-\lambda(2\epsilon)^d)^k k^d}{k!} \\
&= \left(\frac{a}{2\epsilon}\right)^d e^{-\lambda(2\epsilon)^d}(-B_d(-\lambda(2\epsilon)^d)).
\end{aligned}$$

The proof is thus complete. □

If we take $d = 1$, $d = 2$ and $d = 3$, we obtain:

$$\begin{aligned}
\mathbf{E}_\lambda[\chi]\left(\mathcal{C}_\epsilon(\omega)\right)_{\omega\in\mathbb{T}_{[a]}} &= a\lambda e^{-\lambda 2\epsilon}, \\
\mathbf{E}_\lambda[\chi]\left(\mathcal{C}_\epsilon(\omega)\right)_{\omega\in\mathbb{T}^2_{[a^2]}} &= a^2\lambda e^{-\lambda(2\epsilon)^2}\left(1 - \lambda(2\epsilon)^2\right), \\
\mathbf{E}_\lambda[\chi]\left(\mathcal{C}_\epsilon(\omega)\right)_{\omega\in\mathbb{T}^3_{[a^3]}} &= a^3\lambda e^{-\lambda(2\epsilon)^3}\left(1 - 3\lambda(2\epsilon)^3 + (\lambda(2\epsilon)^3)^2\right).
\end{aligned}$$

Remark 6.2 For c a positive real, $\mathbf{E}_\lambda[\chi]$ is invariant under the transformation $\lambda' = \lambda/c$, $\epsilon' = c\epsilon$ and $a' = ca$. Taking $c = 1/2\epsilon$, we obtain:

$$\mathbf{E}_\lambda[\chi] = a'^d e^{-\lambda'}(-B_d(-\lambda')).$$

Hence, the mean depends actually only on a' and λ'.

Corollary 6.4 *The mean of χ in a Binomial process homogeneous with n points is given by:*

$$\mathbf{E}\left[\chi | N_1 = n\right] = \sum_{k=0}^{n} \binom{n}{k} k^d \left(\frac{2\epsilon}{a}\right)^{d(k-1)},$$

Proof: This is a consequence of Corollary 6.2. □

If we take $d = 1$, $d = 2$ and $d = 3$ and calling $\tilde{\epsilon} = \frac{2\epsilon}{a}$, we have:

$$\begin{aligned}
\mathbf{E}\left[\chi | N_1 = n, d = 1\right] &= -n\left(1 - \tilde{\epsilon}\right)^{n-1}, \\
\mathbf{E}\left[\chi | N_1 = n, d = 2\right] &= n\left(1 - \tilde{\epsilon}^2\right)^{n-2}\left(n\tilde{\epsilon}^3 - 1\right), \\
\mathbf{E}\left[\chi | N_1 = n, d = 3\right] &= -n\left(1 - \tilde{\epsilon}^3\right)^{n-3}\left(n^2\tilde{\epsilon}^6 - 3n\tilde{\epsilon}^3 + \tilde{\epsilon}^3 + 1\right).
\end{aligned}$$

The following result is well known.

Lemma 6.5 *If $B_d(x)$ is the Bell's polynomial and for $d \geq 1$, the following relations are valid:*

$$\begin{aligned}
\frac{d}{dx}B_d(x) &= \frac{B_{d+1}(x)}{x} - B_d(x), \\
\frac{d}{dx}(e^x B_d(x)) &= \frac{e^x}{x}B_{d+1}(x).
\end{aligned}$$

According to these relations, it is routine to prove the following theorem.

Theorem 6.6 *The function $(\lambda \mapsto \mathbf{E}_\lambda\left[\chi \mathcal{C}_\epsilon(\omega)\right])$ has exactly d non-negative real roots. Moreover, between each consecutive roots and after the last one, there is exactly one critical point.*

We can see by the expression of $\mathbf{E}_\lambda\left[\chi\right]$ that this quantity tends to 0 as λ tends to infinity. This convergence is due the fact that the Euler characteristic of the Čech complex of the cover tends to the Euler characteristic of the d-Torus where the points are deployed. This is shown in the following theorem.

Theorem 6.7 *The Betti numbers of $\mathcal{C}(\mathcal{U}_\epsilon)$ converge in probability to the Betti number of the torus as λ goes to infinity:*

$$\mathbf{P}_\lambda\left(\bigcap_{i=0}^{d}\left(\beta_i(\mathcal{C}_\epsilon) = \beta_i(\mathbb{T}^d_{[a]})\right)\right) \xrightarrow{\lambda\to\infty} 1.$$

Proof: Let $\eta < \epsilon/2$, by compactness of the torus, there exists a finite collection of balls $\mathfrak{B}$ of radius η covering $\mathbb{T}^d_{[a]}$. Since $\eta < \epsilon/2$, if x belongs to some ball $B \in \mathfrak{B}$ then $B \subset B(x, \epsilon)$, hence

$$\bigcap_{B\in\mathfrak{B}}\left(\omega(B) \neq 0\right) \subset \left(\mathcal{U}_\epsilon(\omega) = \mathbb{T}^d_{[a]}\right).$$

Thus,

$$\mathbf{P}_\lambda\left(\mathcal{U}_\epsilon(\omega) \neq \mathbb{T}^d_{[a]}\right) \leq \mathbf{P}_\lambda\left(\bigcup_{B\in\mathfrak{B}}\left(\omega(B) = 0\right)\right) \leq K\exp(-\lambda(2\eta)^d) \xrightarrow{\lambda\to\infty} 0.$$

Moreover, by the nerve lemma, as stated on [14, Theorem 10.7]:

$$\left(\mathcal{U}_\epsilon(\omega) = \mathbb{T}^d_{[a]}\right) \subset \bigcap_{i=0}^{d} \left(\beta_i(\mathcal{C}_\epsilon) = \beta_i(\mathbb{T}^d_{[a]})\right),$$

and the result follows. □

Remark 6.3 From the properties obtained of the Euler characteristic mean and based in some simulations, we conjuncture that, in a random simplicial complex as defined in this chapter, there is always two main kinds of holes in this complex, β_i and β_{i+1}. So, for instance, consider Fig. 6.5 where points are placed in 5 dimensions. When λ is small, in average, the components are isolated from the others, so $\beta_0 > 0$ and we do not have other kinds of holes, so β_0 is the dominating Betti number. If λ is increased (which means, in average, to increase the number of points), the components connect with each other, decreasing β_0 and some cycles appears, so β_0 and β_1 are the dominating Betti numbers. Increasing λ even more, the complex becomes completely connected and we have a large number of cycles. For λ even larger, those cycles begin to vanish and we have the first voids. Then, we follow this reasoning until all the region is completely covered, so $\beta_5 = 1$ and $\chi = 0$.

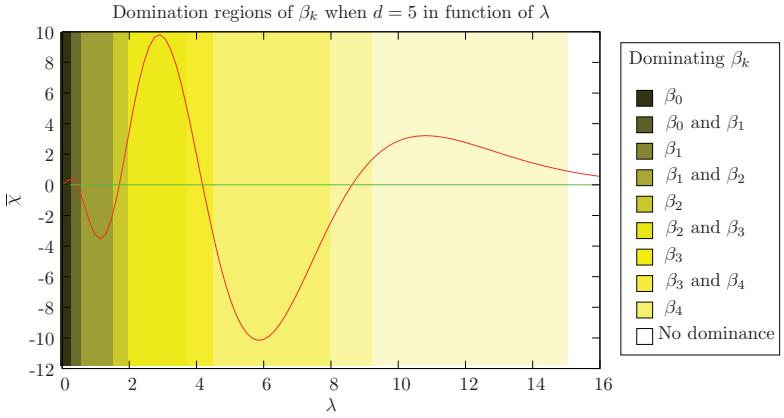


Figure 6.5: Behavior of χ related to the regions of dominance of β_i. There are at most two dominating Betti numbers

The following lemma is straightforward.

Lemma 6.8 *Let k_1, k_2 and k_3 be real positive constants and $f : \mathbf{R}^2_+ \to \mathbf{R}$ defined as*

$$f(x, y) = \exp\left(-\frac{k_1 - x}{2k_2} \log\left(1 + \frac{(k_1 - x)k_2}{k_3 y}\right)\right).$$

Then, for $k_1 - x > 0$, the function is strictly increasing with respect to x and with respect to y.

Theorem 6.9 *For $y > \lambda a^d$, we have*

$$\mathbf{P}_\lambda(\beta_0 \geq y) \leq \exp\left(-\frac{y - \lambda a^d}{2} \log\left(1 + \frac{y - \lambda a^d}{(2^d - 1)^2 \lambda}\right)\right).$$

Proof: To apply Theorem 2.5, we need to evaluate $\max(D\beta_0)$ and $\|D\beta_0\|_{L^\infty(\Omega,L^2(Y))}$. Since there are more points than connected components, $\mathbf{E}_\lambda[\beta_0] \leq \mathbf{E}_\lambda[N_0] = \lambda a^d$. According to the definition of D, $\max(D\beta_0)$ is the maximum variation of β_0 induced by the addition of an arbitrary point. If this point is at a distance smaller than ϵ from ω, then $D\beta_0 \leq 0$, otherwise, $D\beta_0 = 1$, so $\max(D\beta_0) = 1$. Besides, this added point can join at most two connected components in each dimension, so in d dimensions it can join at most 2^d connected component, which means that $D\beta_0$ ranges from $-(2^d-1)$ to 1, and then

$$\|D\beta_0\|_{L^\infty(\Omega,L^2(Y))} \leq \lambda \max|D\beta_0|^2 = \lambda(2^d-1)^2.$$

Using Lemma 6.8 and Theorem 2.5, we get:

$$\mathbf{P}_\lambda(\beta_0 \geq y) \leq \exp\left(-\frac{y-\lambda a^d}{2}\log\left(1+\frac{y-\lambda a^d}{(2^d-1)^2\lambda}\right)\right),$$

for $y > \lambda a^d \geq \mathbf{E}_\lambda[\beta_0]$. □

6.3 Second order moments

We use all the definitions of the previous section.

Lemma 6.10 *We can rewrite N_k as*

$$N_k = \frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda^{k-i} I_i\left(\int_{(\mathbb{T}_a^d)^i} h(x_1,\ldots,x_k)\ dx_1\ldots\ dx_{k-i}\right).$$

Proof: We have that

$$\int_{\Delta_k} h(x_1,\cdots,x_k)(\ \mathrm{d}\omega(x_1)-\lambda\ \mathrm{d}x_1)\ldots(\ \mathrm{d}\omega(x_i)-\lambda\ \mathrm{d}x_i)\lambda\ \mathrm{d}x_{i+1}\ldots\lambda\ \mathrm{d}x_k$$
$$= \sum_{j=0}^{i}(-1)^j\binom{i}{j}\int_{\Delta_k} h(x_1,\cdots,x_k)\ \mathrm{d}\omega(x_1)\ldots\ \mathrm{d}\omega(x_j)\lambda\ \mathrm{d}x_{j+1}\ldots\lambda\ \mathrm{d}x_k.$$

Thus, after some algebrism with the binomial factors, we have

$$\frac{1}{k!}\sum_{i=0}^{k}\binom{k}{i}\sum_{j=0}^{i}(-1)^j\binom{i}{j}\int_{\Delta_k} h(x_1,\cdots,x_k)\ \mathrm{d}\omega(x_1)\ldots\ \mathrm{d}\omega(x_j)\lambda\ \mathrm{d}x_{j+1}\ldots\lambda\ \mathrm{d}x_k$$
$$= \frac{1}{k!}\int_{\Delta_k} h(x_1,\cdots,x_k)\ \mathrm{d}\omega(x_1)\ldots\ \mathrm{d}\omega(x_k) = N_k,$$

concluding the proof. □

Definition 6.1 *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two simplices with common vertices. For $L \in \mathcal{P}(\{1,2\})$, let us denote m_L the number of vertices belonging exactly to the list L of simplices.*

Then $M = m_{12}+m_1+m_2$ is the total number of vertices and $\mathcal{J}_2$ represents the integral on these two simplices:

$$\mathcal{J}_2(m_{12},m_1,m_2) = \int_{\Delta_{m_{12}+m_1}}\int_{\Delta_{m_{12}+m_2}} h_{m_{12}+m_1}h_{m_{12}+m_2}\ dx_1\ldots\ dx_M.$$

with $x_1,\cdots,x_M$ being the M vertices.

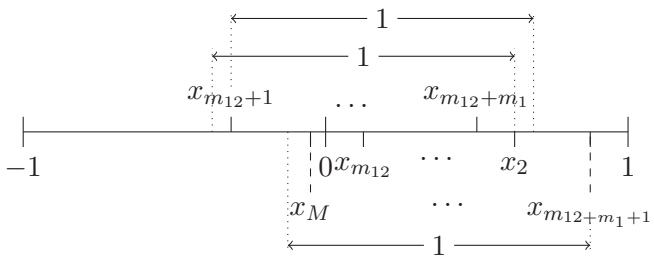


Figure 6.6: Example of relative positions of the points

Lemma 6.11 *For $d = 1$ and $\epsilon = 1/2$, we have*

$$\mathcal{J}_2(m_{12}, m_1, m_2) = m_{12} + m_1 + m_2 + \frac{2m_1m_2}{m_{12}+1}. \tag{6.5}$$

Proof: Let us split the integration domain of $\mathcal{J}_2$ in two domains S_1 and S_2 corresponding to the cases:

1. All the vertices are connected with each other, thus there is only one simplex. The integral on S_1 is simply the number of points in the simplex: $M = m_{12} + m_1 + m_2$.
2. There are at least two vertices at distance $d > 1$, which leads to two simplices. By symmetry we can choose to order the m_L vertices for each $L \in \mathcal{P}(\{1, 2\})$ from lowest to greatest or the opposite and choose which simplex is on which side of the axis. Thus we have the integral on S_2 equal to $2m_{12}!m_1!m_2!\mathcal{A}$, with $\mathcal{A}$ an integral whose calculation is detailed below.

We choose to enumerate the vertices of the simplexes such that:

- $x_1, \cdots, x_{m_{12}}$ are the m_{12} common vertices.
- $x_{m_{12}+1}, \cdots, x_{m_{12}+m_1}$ are the m_1 vertices of only $\mathcal{C}_1$.
- $x_{m_{12}+m_1+1}, \cdots, x_M$ are the m_2 vertices of only $\mathcal{C}_2$.

Without loss of generality we can choose the origin to be x_1. The vertices are now order as described in Fig. 6.6:

$$0 \leq x_{m_{12}} \leq x_{m_{12}-1} \leq \cdots \leq x_2 \leq 1,$$

$$-1 \leq x_2 - 1 \leq x_{m_{12}+1} \leq x_{m_{12}+2} \leq \cdots \leq x_{m_{12}+m_1} \leq x_{m_{12}+1} + 1 \text{ , and}$$

$$x_{m_{12}+1} \leq x_{m_{12}+m_1+1} - 1 \leq x_M \leq x_{M-1} \leq \cdots \leq x_{m_{12}+m_1+1} \leq 1.$$

Let us denote $J_a(f)(x) = \int_a^x f(u)\,\mathrm{d}u$ then we write the composition $J_a^{(2)}(f)(x) = \int_a^x \int_a^u f(v)\,\mathrm{d}v\,\mathrm{d}u$. We also denote $m = m_{12} + 1$ and $n = m_{12} + m_1 + 1$, then we have:

$$\mathcal{A} = \int_0^1 J_0^{(m_{12}-2)}(\mathbf{1})(x_2) \int_{x_2-1}^0 -J_{x_m+1}^{(m_1-1)}(\mathbf{1})(x_m) \int_{x_m+1}^1 J_{x_n-1}^{(m_2-1)}(\mathbf{1})(x_n)\,\mathrm{d}x_n\,\mathrm{d}x_m\,\mathrm{d}x_2.$$

We find that:

$$\begin{aligned} J_0^{(m_{12}-2)}(\mathbf{1})(x_2) &= \frac{x_2^{m_{12}-2}}{(m_{12}-2)!}, \\ -J_{x_m+1}^{(m_1-1)}(\mathbf{1})(x_m) &= \frac{1}{(m_1-1)!}, \\ J_{x_n-1}^{(m_2-1)}(\mathbf{1})(x_n) &= \frac{1}{(m_2-1)!}. \end{aligned}$$

Thus we have:

$$\begin{aligned} A &= \frac{1}{(m_{12}-2)!(m_1-1)!(m_2-1)!}\int_0^1 x_2^{m_{12}-2}\int_{x_2-1}^0 -x_m \,\mathrm{d}x_m \,\mathrm{d}x_2 \\ &= \frac{1}{(m_1-1)!(m_2-1)!(m_{12}+1)!}, \end{aligned}$$

concluding the proof. □

Theorem 6.12 *Let $\epsilon \leq a/6$. Then, the covariance between the number of $(k-1)$-simplices, N_k, and the number of $(l-1)$-simplices, N_l, for $l \leq k$ is given by*

$$Cov_\lambda\left[N_k, N_l\right] = \sum_{i=0}^{l-1} \frac{1}{i!(k-l+i)!(l-i)!}(\lambda(2\epsilon)^d)^{k+i}\left(\frac{a}{2\epsilon}\right)^d\left(k+i+2\frac{i(k-l+i)}{l-i+1}\right)^d. \quad (6.6)$$

Proof: We want to evaluate $\mathbf{E}_\lambda\left[(N_k - \mathbf{E}_\lambda\left[N_k\right])(N_l - \mathbf{E}_\lambda\left[N_l\right])\right]$. By Lemma 6.10, this can be written as

$$\mathbf{E}_\lambda\left[\frac{1}{k!}\sum_{i=1}^{k}\binom{k}{i}\lambda^{k-i}I_i\left(f_i^k\right)\frac{1}{l!}\sum_{i=1}^{l}\binom{l}{i}\lambda^{l-i}I_i\left(f_i^l\right)\right],$$

where

$$f_j^n = \int_{(\mathbb{T}_a^d)^j} h(v_1,\cdots,v_n)\,\mathrm{d}v_1\ldots\,\mathrm{d}v_{n-j}.$$

Using the isometry formula, given by Eq. (2.4), we have

$$\begin{aligned} \mathrm{Cov}_\lambda\left[N_k, N_l\right] &= \frac{1}{k!l!}\sum_{i=1}^{l}\binom{k}{i}\binom{l}{i}\lambda^{k+l-2i}\mathbf{E}_\lambda\left[I_i\left(f_i^k\right)I_i\left(f_i^l\right)\right] \\ &= \frac{1}{k!l!}\sum_{i=1}^{l}\binom{k}{i}\binom{l}{i}\lambda^{k+l-2i}i!\langle f_i^k f_i^l\rangle_{L^2(\lambda)^{\circ i}} \\ &= \sum_{i=0}^{l-1}\frac{1}{i!(k-l+i)!(l-i)!}\lambda^{k-l+2i}\langle f_{l-i}^k f_{l-i}^l\rangle_{L^2(\lambda)^{\circ(l-i)}}. \end{aligned} \quad (6.7)$$

Hence, we are reduced to compute

$$\langle f_j^k f_j^l\rangle_{L^2(\lambda)^{\circ(j)}} = \int_{(\mathbb{T}_a^d)^j}\left(\int_{(\mathbb{T}_a^d)^{l-j}} h(v_1,\cdots,v_l)\,\mathrm{d}v_{j+1}\ldots\,\mathrm{d}v_l \right.$$
$$\left.\int_{(\mathbb{T}_a^d)^{k-j}} h(v_1,\cdots,v_k)\,\mathrm{d}v_{j+1}\ldots\,\mathrm{d}v_k\right)\lambda\,\mathrm{d}v_1\ldots\lambda\,\mathrm{d}v_j.$$

Since $a > \epsilon/6$, we have

$$\langle f_j^k f_j^l \rangle_{L^2(\lambda)^{\circ(j)}} = \int_{[0,a]^d} \lambda \, \mathrm{d}v_1 \int_{([0,a]^d)^{k-1}} h(0, v_2, \cdots, v_k) \\ \times h(0, v_2, \cdots, v_j, v_1', \cdots, v_{l-j}') \, \mathrm{d}v_{l-j}' \ldots \, \mathrm{d}v_1' \, \mathrm{d}v_k \ldots \, \mathrm{d}v_{j+1} \lambda \, \mathrm{d}v_j \ldots \lambda \, \mathrm{d}v_2.$$

Moreover, if $v_i = (u_{i,1}, \cdots, u_{i,d})$ and $v_i' = (u_{i,1}', \cdots, u_{i,d}')$ and we proceed to the following substitutions:

$$\begin{aligned} u_{i,1} &= 2\epsilon x_i \text{ if } 2 \leq i \leq j, \\ u_{i,1} &= 2\epsilon y_{k-j} \text{ if } j+1 \leq i \leq k, \\ u_{i,1}' &= 2\epsilon z_i \text{ if } 1 \leq i \leq l-j, \end{aligned}$$

This results in a Jacobian $(2\epsilon)^{k+l-2i-1}$ and we recognize the integral to be exactly $\mathcal{J}_2(j, k-j, l-j)$ as defined in Definition 6.1. Thus, we have:

$$\langle f_j^k f_j^l \rangle_{L^2(\lambda)^{\circ(j)}} = \lambda^i a^d (2\epsilon)^{k+l-2i-1} \left(\mathcal{J}_2(j, k-j, l-j)\right)^d.$$

Finally, using Eq. (6.5) and Eq. (6.7) gives the result. □

Remark 6.4 It is possible to write $\mathrm{Var}(N_k)$ as Eq. (6.6) due the fact that we use the maximum norm. This simplifies the calculations since we can treat each component individually. However, considering the Euclidean norm it is still possible to find analytically a closed-form expression for $\mathrm{Var}(N_k)$, but its calculation involves nasty integrals and a generic term cannot be found. When we consider the Rips-Vietoris complex in $\mathbb{T}_a^2$, the variance of the number of 1-simplices and 2-simplices are given by:

$$\mathbf{V}_\lambda [N_2] = \left(\frac{a}{2\epsilon}\right)^2 \left(\frac{\pi}{2}(4\lambda\epsilon^2)^2 + \pi^2(4\lambda\epsilon^2)^3\right),$$

and

$$\mathbf{V}_\lambda [N_3] = \left(\frac{a}{2\epsilon}\right)^2 \left((4\lambda\epsilon)^3 \frac{\pi}{6}\left(\pi - \frac{3\sqrt{3}}{4}\right) + (4\lambda\epsilon^2)^4 \pi \left(\frac{\pi^2}{2} - \frac{5}{12} - \frac{\pi\sqrt{3}}{2}\right) \right. \\ \left. + (4\lambda\epsilon^2)^5 \frac{\pi^2}{4}\left(\pi - \frac{3\sqrt{3}}{4}\right)^2\right).$$

Remark 6.5 In the same way that we explicit the mean of k simplexes in corollary 6.2 for a binomial point process with n points, variances, covariances, and Nth moments can all be found since they can be written as a polynomial on λ.

Since we have an expression for the variance of the number of k-simplices, it is possible to calculate one for the Euler characteristic.

Theorem 6.13 *Let $\epsilon \leq a/6$. Then, the variance of the Euler characteristic in a d torus is:*

$$\mathbf{V}_\lambda [\chi] = \left(\frac{a}{2\epsilon}\right)^d \sum_{n=1}^{\infty} c_n^d (\lambda(2\epsilon)^d)^n,$$

where

$$c_n^d = \sum_{j=\lceil (n+1)/2 \rceil}^{n} \left[2 \sum_{i=n-j+1}^{j} \frac{(-1)^{i+j}}{(n-j)!(n-i)!(i+j-n)!} \left(n + \frac{2(n-i)(n-j)}{1+i+j-n} \right)^d \right.$$
$$\left. - \frac{1}{(n-j)!^2(2j-n)!} \left(n + \frac{2(n-j)^2}{1+2j-n} \right)^d \right].$$

Proof: The variance of χ is given by:

$$\begin{aligned} \mathbf{V}_\lambda[\chi] &= \mathbf{E}_\lambda\left[(\chi - \mathbf{E}_\lambda[\chi])^2\right] = \mathbf{E}_\lambda\left[\left(\sum_{k=1}^{\infty}(-1)^k N_k - \sum_{k=1}^{\infty}(-1)^k \mathbf{E}_\lambda[N_k]\right)^2\right] \\ &= \mathbf{E}_\lambda\left[\left(\sum_{k=1}^{\infty}(-1)^k (N_k - \mathbf{E}_\lambda[N_k]\right)^2\right] \\ &= \mathbf{E}_\lambda\left[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(-1)^{i+j}(N_i - \mathbf{E}_\lambda[N_i])(N_j - \mathbf{E}_\lambda[N_j])\right]. \end{aligned}$$

We remark that $N_i \leq \frac{N_1^i}{i!}$, so there is a constant c such that

$$\mathbf{E}_\lambda\left[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|(N_i - \mathbf{E}_\lambda[N_i])(N_j - \mathbf{E}_\lambda[N_j])|\right] \leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\left|\frac{N_1^i}{i!}\frac{N_1^j}{j!}\right| \leq c\mathbf{E}_\lambda[e^{n_1}]^2 < \infty.$$

Thus the alternating series converges absolutely allowing us to exchange the mean with the sums and we can write

$$\mathbf{V}_\lambda[\chi] = \sum_{i=1}^{\infty}(-1)^i \sum_{j=1}^{\infty}(-1)^j \mathrm{Cov}_\lambda[N_i, N_j].$$

The result follows by Eq. (6.6) and some tedious but straightforward algebra. □

Lemma 6.14 *Let n be a positive integer, then*

$$\sum_{j=1}^{n}\binom{n}{j}\left(\binom{j-1}{n-j-1} - \binom{j-1}{n-j}\right) = (-1)^n.$$

Proof: We first simplify the expression:

$$\sum_{j=1}^{n}\binom{n}{j}\left(\binom{j-1}{n-j-1} - \binom{j-1}{n-j}\right) = \sum_{j=1}^{n}\frac{2n-3j}{j}\binom{n}{j}\binom{j}{n-j},$$

Then, applying hypergeometric functions, we solve the sum:

$$\sum_{j=1}^{n}\frac{2n-3j}{j}\binom{n}{j}\binom{j}{n-j} = (-1)^n.$$

□

Theorem 6.15 *In one dimension, the expression of the variance of the Euler characteristic is:*

$$\mathbf{V}_\lambda[\chi] = a\left(\lambda e^{-2\lambda\epsilon} - 4\lambda^2\epsilon e^{-4\lambda\epsilon}\right).$$

Proof: If $d = 1$, according to Theorem 6.13:

$$\mathbf{V}_\lambda[\chi] = \frac{a}{2\epsilon}\sum_{n=1}^{\infty} c_n^1 (2\lambda\epsilon)^n, \tag{6.8}$$

and we define

$$\alpha_n = \sum_{j=\lceil\frac{n+1}{2}\rceil}^{n}\left[2\sum_{i=n-j+1}^{j}\frac{(-1)^{i+j}n}{(n-j)!(n-i)!(i+j-n)!} - \frac{n}{(n-j)!^2(2j-n)!}\right].$$

and $\beta_n = c_n^1 - \alpha$. It is well known that

$$\sum_{i=0}^{2j-n}(-1)^i\binom{j}{i} = (-1)^{2j-n-1}\binom{j-1}{2j-n},$$

using Stiffel's relation, we obtain:

$$\begin{aligned}
\alpha_n &= (-1)^n\frac{n}{n!}\sum_{j=\lceil\frac{n+1}{2}\rceil}^{n}\left[\binom{n}{j}2\sum_{i=0}^{2j-n}(-1)^i\binom{j}{i} + 2(-1)^n\binom{n}{j}\right]\\
&= \frac{1}{(n-1)!}\sum_{j=\lceil\frac{n+1}{2}\rceil}^{n}\left[2\binom{n}{j}\binom{j-1}{n-j-1} - \binom{n}{j}\binom{j}{n-j} - 2(-1)^n\binom{n}{j}\right]\\
&= \frac{1}{(n-1)!}\sum_{j=\lceil\frac{n+1}{2}\rceil}^{n}\left[\binom{n}{j}\left(\binom{j-1}{n-j} - \binom{j-1}{n-j-1}\right) - 2(-1)^n\binom{n}{j}\right]. \qquad (6.9)
\end{aligned}$$

The identity $\binom{n}{j} = \binom{n}{n-j}$ allows us to write that

$$\sum_{j=\lceil(n+1)/2\rceil}^{n}(-2(-1)^n)\binom{n}{j} = \sum_{j=0}^{n}\binom{n}{j} = 2^n, \quad n \text{ odd},$$

$$\sum_{j=\lceil(n+1)/2\rceil}^{n}(-2(-1)^n)\binom{n}{j} = \binom{n}{n/2} + \sum_{j=0}^{n}-\binom{n}{j} = -2^n + \binom{n}{n/2}, \quad n \text{ even}.$$

Since $\binom{j-1}{n-j} = 0$ for $j < \lceil\frac{n+1}{2}\rceil$, we have

$$\sum_{j=\lceil\frac{n+1}{2}\rceil}^{n}\binom{n}{j}\left(\binom{j-1}{n-j} - \binom{j-1}{n-j-1}\right) = \sum_{j=1}^{n}\binom{n}{j}\left(\binom{j-1}{n-j} - \binom{j-1}{n-j-1}\right)$$

for n odd and

$$\sum_{j=\lceil\frac{n+1}{2}\rceil}^{n}\binom{n}{j}\left(\binom{j-1}{n-j} - \binom{j-1}{n-j-1}\right) = -\binom{n}{n/2} + \sum_{j=1}^{n}\binom{n}{j}\left(\binom{j-1}{n-j} - \binom{j-1}{n-j-1}\right).$$

for n even. According to Lemma 6.14, we get:

$$\begin{aligned}\sum_{j=\lceil (n+1)/2\rceil}^{n} \binom{n}{j}\left[\binom{j-1}{n-j-1}-\binom{j-1}{n-j}\right] &= -1, \quad n \text{ odd},\\ \sum_{j=\lceil (n+1)/2\rceil}^{n} \binom{n}{j}\left[\binom{j-1}{n-j-1}-\binom{j-1}{n-j}\right] &= 1-\binom{n}{n/2}, \quad n \text{ even}.\end{aligned}$$

Then, we substitute these two last expressions in Eq. (6.9) to obtain

$$\alpha_n = (-1)^n \frac{(1-2^n)\mathbf{1}_{[n\geq 1]}}{(n-1)!},$$

and thus

$$\sum_{i=0}^{\infty} \alpha_n x^n = -xe^{-x} + 2xe^{-2x}.$$

Proceeding along the same line, β_n is given by

$$\begin{aligned}\beta_n &= \sum_{j=\lceil \frac{n+1}{2}\rceil}^{n} \left[2 \sum_{i=n-j+1}^{j} \frac{(-1)^{i+j} 2(n-i)(n-j)}{(n-j)!(n-i)!(i+j-n+1)!} \right.\\ &\qquad \left. - \frac{2(n-j)^2}{(n-j)!^2(2j-n+1)!}\right]\\ &= (-1)^n \left(\frac{(-2+2^n)\mathbf{1}_{[n\geq 1]}}{(n-1)!} - \frac{2\mathbf{1}_{[i\geq 2]}}{(i-2)!}\right),\end{aligned}$$

and again we can simplify the power series $\sum_{i=0}^{\infty} \beta_n x^n$:

$$\sum_{i=0}^{\infty} \beta_n x^n = 2xe^{-x} - 2(x+x^2)e^{-2x}.$$

Then, substituting α_n and β_n in Eq. (6.8) yields the result. □

Theorem 6.16 *We have $D\chi \leq 2$ and $\|D\chi\|_{L^\infty(\Omega, L^2(\mathbb{T}_a^d))} < \infty$ and*

$$\mathbf{P}(\chi - \bar{\chi} \geq x) \leq \exp\left(-\frac{x}{4} \log\left(1 + \frac{2x}{\mathbf{V}_\lambda[\chi]}\right)\right).$$

Proof: In two dimensions, the Euler characteristic is:

$$\chi = \beta_0 - \beta_1 + \beta_2.$$

Therefore we can bound $D\chi$ by the variation of $\beta_0 - \beta_1$ added to the variation of β_2 when we add a vertex to a simplicial complex.

If we add a vertex on the torus, either the vertex is isolated or not. In the first case, it forms a new connected component incrementing β_0 by 1, and the number of holes that is β_1 is the same. Otherwise, as there is no new connected component, β_0 is the same, but the new vertex can at most fill a hole incrementing β_1 by 1. Therefore, the variation of $\beta_0 - \beta_1$ is at most 1.

Now, let us look at the variation of β_2 when we add a vertex to a simplicial complex. According to Proposition 5.3 is at most 1, showing that $D\chi \leq 2$. Then, we use Eq. (2.7) to complete the proof. □

6.4 Nth order moments

For this section, without loss of generality, using Proposition 5.5, we can choose $k = 1/2\epsilon$, so $\lambda_\tau = \lambda(2\epsilon)^d$, $\epsilon_\tau = 1/2$ and $ak = a/2\epsilon$.

We are interested in the central moment, so we introduce the following notation for the centralized number of $(k-1)$-simplices: $\tilde{N_k} = N_k - \bar{N_k}$.

Finally, let us denote that $\binom{i}{j} = 0$ as soon as $i \leq 0$ or $j \leq 0$ or $i - j \leq 0$ for i and j integers.

We extend the Definition 6.1 used in the second order moments calculations.

Definition 6.2 *Let $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ be three simplices with common vertices. For $L \in \mathcal{P}(\{1,2,3\})$, let us denote m_L the number of vertices belonging exactly to the list L of simplices.*

Then $M = m_{123} + m_{12} + m_{13} + m_{23} + m_1 + m_2 + m_3$ is the total number of vertices and $\mathcal{J}_3$ represents the integral on these three simplices:

$$\mathcal{J}_3 = \int_{\Delta_{p_1}} \int_{\Delta_{p_2}} \int_{\Delta_{p_3}} h_{p_1} h_{p_2} h_{p_3} \; dx_1 \dots \; dx_M.$$

with p_i being the number of vertices of simplex $\mathcal{C}_i$ for $i = 1, \cdots, 3$, for instance $p_1 = m_{123} + m_{12} + m_{13} + m_1$, and $x_1, \cdots, x_M$ being the M vertices.

Definition 6.3 *We denote $\mathcal{J}_3(i,j,s,t)$ the integral defined above such that*

- $m_{123} = 2t - i - j + s \vee 0$
- $m_{12} = i + j - s - t \vee 0$
- $m_{13} = i - t \vee 0$
- $m_{23} = j - t \vee 0$
- $m_1 = k - i \vee 0$
- $m_2 = k - j \vee 0$
- $m_3 = k - s \vee 0$.

Theorem 6.17 *The third moment of the number of $(k-1)$-simplices is given by:*

$$\mathbf{E}_\lambda \left[\widetilde{N_k}^3 \right] = \sum_{i,j,s,t} \lambda^{3k-i-j} t! \binom{k}{i} \binom{k}{j} \binom{k}{s} \binom{i}{t} \binom{j}{t} \binom{t}{i+j-s-t} \mathcal{J}_3(i,j,s,t),$$

with $s \geq |i-j|$.

Proof: From Lemma 6.10 , we know that the chaos decomposition of the number of $(k-1)$-simplices is given by

$$\widetilde{N_k} = I_1(f_1) + \cdots + I_k(f_k) = \sum_{i=1}^{k} I_i(f_i),$$

with

$$f_i(x_1, \cdots, x_i) = \binom{k}{i} \int h(x_1, \cdots, x_k) \lambda^{k-i} \, \mathrm{d}x_k \dots \, \mathrm{d}x_{i+1},$$

and

$$I_i(f_i) = \int f_i(\ \mathrm{d}\omega(x_1) - \ \mathrm{d}\lambda(x_1)) \dots (\ \mathrm{d}\omega(x_i) - \ \mathrm{d}\lambda(x_i)).$$

Then, we define

$$g_{i,j,i+j-s} = \sum_{t=\lceil \frac{i+j-s}{2} \rceil}^{i+j-s\wedge i\wedge j} t! \binom{i}{t} \binom{j}{t} \binom{t}{i+j-s-t} f_i \circ_t^{u-t} f_j$$

and using the chaos expansion (cf Proposition 2.3), we get

$$\begin{aligned}
\widetilde{N_k}^3 &= (I_1(f_1) + \dots + I_k(f_k))^3 \\
&= \left(\sum_{i=1}^{k} \sum_{j=1}^{k} I_i(f_i) I_j(f_j) \right) (I_1(f_1) + \dots + I_k(f_k)) \\
&= \sum_{i,j=1}^{k} \sum_{s=|i-j|}^{i+j} I_s(g_{i,j,i+j-s}) (I_1(f_1) + \dots + I_k(f_k)) \\
&= \sum_{i,j,l=1}^{k} \sum_{s=|i-j|}^{i+j} I_s(g_{i,j,i+j-s}) I_l(f_l).
\end{aligned}$$

According to (2.4), denoting $u = i + j - s$, we obtain:

$$\begin{aligned}
\mathbf{E}_\lambda \left[\widetilde{N_k}^3 \right] &= \mathbf{E}_\lambda \left[\sum_{i,j=1}^{k} \sum_{s=|i-j|\vee 1}^{i+j\wedge k} I_s(g_{i,j,u}) I_s(f_s) \right] \\
&= \sum_{i,j=1}^{k} \sum_{s=|i-j|\vee 1}^{i+j\wedge k} \int g_{i,j,u} f_s \lambda^s \ \mathrm{d}x_1 \dots \ \mathrm{d}x_s \\
&= \sum_{i,j=1}^{k} \sum_{s=|i-j|\vee 1}^{i+j\wedge k} \sum_{t=\lceil \frac{u}{2} \rceil}^{u\wedge i\wedge j} \lambda^s t! \binom{i}{t} \binom{j}{t} \binom{t}{u-t} \int (f_i \circ_t^{u-t} f_j) f_s \ \mathrm{d}x_1 \dots \mathrm{d}x_s.
\end{aligned}$$

Then we recognize the integral defined in Definition 6.3:

$$\mathbf{E}_\lambda \left[\widetilde{N_k}^3 \right] = \sum_{i,j,=1}^{k} \sum_{s=|i-j|\vee 1}^{i+j\wedge k} \sum_{t=\lceil \frac{u}{2} \rceil}^{u\wedge i\wedge j} \lambda^{3k-i-j} t! \binom{k}{i} \binom{k}{j} \binom{k}{s} \binom{i}{t} \binom{j}{t} \binom{t}{u-t} \mathcal{J}_3(i,j,s,t).$$

Finally, relaxing the boundaries on the sums conclude the proof. □

Definition 6.4 *Let $\mathcal{C}_1, \cdots, \mathcal{C}_n$ be n simplices with some common vertices. For $L \in \mathcal{P}(\{1, \cdots, n\})$, let us denote m_L the number of vertices belonging exactly to the list L of simplices.*

Then $M = \sum_{L \in \mathcal{P}(\{1,\cdots,n\})} m_L$ is the total number of vertices and $\mathcal{J}_n$ represents the integral on these n simplices:

$$\mathcal{J}_n = \int_{\Delta_{p_1}} \cdots \int_{\Delta_{p_n}} h_{p_1} \dots h_{p_n} \ dx_1 \dots \ dx_M.$$

with p_i being the number of vertices of simplex $\mathcal{C}_i$ for $i = 1, \cdots, n$, and $x_1, \cdots, x_M$ being the M vertices.

Theorem 6.18 *The expression of the n-th power of the number of $(k-1)$-simplices is given by:*

$$\widetilde{N}_k^n = \sum_{i_1,\cdots,i_n} \sum_{s_1,\cdots s_{n-2}} \sum_{t_1,\cdots t_{n-2}} \left(\prod_{j=1}^{n-2} t_j! \binom{m_{j,1}}{t_j} \binom{m_{j,2}}{t_j} \binom{t_j}{u_j - t_j} \right) I_a(\circ_{j\in A} f_{i_j}) I_b(\circ_{j\in \overline{A}} f_{i_j}), \quad (6.10)$$

where for $j \in \{1, \cdots, n-2\}$:

- $1 \leq i_1, \cdots, i_n \leq k$,
- $s_j \geq |m_{j,1} - m_{j,2}|$,
- $m_{j,1} = i_{2j-1}$ *if* $1 \leq j \leq \lfloor \frac{n}{2} \rfloor$ *and* $s_{2(j-\lfloor \frac{n}{2} \rfloor)-1}$ *otherwise,*
- $m_{j,2} = i_{2j}$ *if* $1 \leq j \leq \lfloor \frac{n}{2} \rfloor$ *and* $s_{2(j-\lfloor \frac{n}{2} \rfloor)}$ *otherwise,*
- $u_j = m_{j,1} + m_{j,2} - s_j$,
- $A \subset \{1, \cdots, n\}$,
- *If n is even, then $a = s_{n-3}$ and $b = s_{n-2}$,*
- *If n is odd, then $a = s_{n-2}$ and $b = i_n$.*

Proof: The decomposition of the centralized number of $(k-1)$-simplices is:

$$\widetilde{N_k} = I_1(f_1) + \cdots + I_k(f_k) = \sum_{i=1}^{k} I_i(f_i).$$

Now, we raise $\widetilde{N_k}$ to the n-th power:

$$\widetilde{N_k}^n = \left(\sum_{i=1}^{k} I_i(f_i) \right)^n.$$

First, we consider the case where n is even, we can group the factors two by two:

$$\widetilde{N}_k^n = \left(\sum_{i_1=1}^{k} I_{i_1}(f_{i_1}) \sum_{i_2=1}^{k} I_{i_2}(f_{i_2}) \right) \cdots \left(\sum_{i_{n-1}=1}^{k} I_{i_{n-1}}(f_{i_{n-1}}) \sum_{i_n=1}^{k} I_{i_n}(f_{i_n}) \right).$$

We then use the chaos expansion of Proposition 2.3:

$$\begin{aligned} I_i(f_i) I_j(f_j) &= \sum_{s=0}^{2(i\wedge j)} I_{i+j-s} \left(\sum_{s \leq 2t \leq 2(s\wedge i\wedge j)} t! \binom{i}{t} \binom{j}{t} \binom{t}{s-t} f_i \circ_t^{s-t} f_j \right) \\ &= \sum_{s=|i-j|}^{i+j} I_s \left(\sum_{i+j-s \leq 2t \leq 2(i+j-s)\wedge i\wedge j)} t! \binom{i}{t} \binom{j}{t} \binom{t}{i+j-s-t} f_i \circ_t^{i+j-s-t} f_j \right). \end{aligned}$$

Let us denote

$$g_s = t!\binom{i}{t}\binom{j}{t}\binom{t}{i+j-s-t} f_i \circ_t^{i+j-s-t} f_j,$$

so we can re-write, relaxing the boundaries on the sums:

$$I_i(f_i)I_j(f_j) = \sum_{s\geq|i-j|}\sum_t I_s(g_s).$$

Thus, we have:

$$\widetilde{N}_k^n = \sum_{i_1,i_2=1}^{k}\sum_{s_1\geq|i_1-i_2|}\sum_{t_1} I_{s_1}(g_{s_1})\cdots\sum_{i_{n-1},i_n=1}^{k}\sum_{s_{n/2}\geq|i_{n-1}-i_n|}\sum_{t_{n/2}} I_{s_{n/2}}(g_{s_{n/2}}).$$

We go on grouping terms by 2 until we only have a product of 2 chaos left: First we made $n/2$ chaos expansions, leading to $n/2$ sums with indexes s_j, $j=1,\cdots,n/2$. To reduce the number of chaos to 2, we have to make other chaos expansions. For $j\geq\frac{n}{2}+1$, the sum indexed by s_j represents the expansion of the chaos indexed $s_{2(j-\frac{n}{2})-1}$ and $s_{2(j-\frac{n}{2})-1}$. We have 2 chaos remaining when $j=2(j-\frac{n}{2})+2$, i.e. when $j=n-2$.

Moreover, there are as much sums indexed with t_j as with s_j, that is $n-2$. Thus we can write:

$$\widetilde{N}_k^n = \sum_{i_1,\cdots,i_n=1}^{k}\sum_{s_1,\cdots s_{n-2}}\sum_{t_1,\cdots t_{n-2}} I_{s_{n-3}}(\phi_{s_{n-3}})I_{s_{n-2}}(\phi_{s_{n-2}}),$$

With $s_j\geq|m_{j,1}-m_{j,2}|$ for $j\in\{1,\cdots,n-2\}$ if we denote:

- $m_{j,1}=i_{2j-1}$ if $1\leq j\leq\frac{n}{2}$ and $s_{2(j-\frac{n}{2})-1}$ otherwise,
- $m_{j,2}=i_{2j}$ if $1\leq j\leq\frac{n}{2}$ and $s_{2(j-\frac{n}{2})}$ otherwise.

Then, denoting $u_j=m_{j,1}+m_{j,2}-s_j$ and A the subset of $\{1,\cdots,n\}$ such that if $j\in A$ then the chaos i_j is expanded in the chaos s_{n-3}, we have:

$$I_{s_{n-3}}(\phi_{s_{n-3}})I_{s_{n-2}}(\phi_{s_{n-2}}) = \left(\prod_{j=1}^{n-2} t_j!\binom{m_{j,1}}{t_j}\binom{m_{j,2}}{t_j}\binom{t_j}{u_j-t_j}\right) I_{s_{n-3}}(\circ_{j\in A}f_{i_j})I_{s_{n-2}}(\circ_{j\in\overline{A}}f_{i_j}).$$

The notation $\circ_{j\in A}f_{i_j}$ represents the product defined in Eq. (2.6) of the functions f_{i_j} for $j\in A$, but whom variables depend on all the $i_1,\cdots,i_n,s_1,\cdots,s_{n-2}$, and $t_1,\cdots,t_{n-2}$.

Now, if n is odd, we consider $n-1$ which is even, therefore we have:

$$\begin{aligned}\widetilde{N}_k^n &= \sum_{i_1,\cdots,i_{n-1}=1}^{k}\sum_{s_1,\cdots,s_{n-3}}\sum_{t_1,\cdots,t_{n-3}} I_{s_{n-4}}(\phi_{s_{n-4}})I_{s_{n-3}}(\phi_{s_{n-3}})\sum_{i_n=1}^{k} I_{i_n}(f_{i_n})\\ &= \sum_{i_1,\cdots,i_n=1}^{k}\sum_{s_1,\cdots,s_{n-2}}\sum_{t_1,\cdots,t_{n-2}} I_{s_{n-2}}(\phi_{s_{n-2}})I_{i_n}(f_{i_n}),\end{aligned}$$

with $s_j\geq|m_{j,1}-m_{j,2}|$ for $j\in\{1,\cdots,n-2\}$ using the same notations for $n-1$ instead of n:

- $m_{j,1} = i_{2j-1}$ if $1 \leq j \leq \frac{n-1}{2}$ and $s_{2(j-\frac{n-1}{2})-1}$ otherwise,
- $m_{j,2} = i_{2j}$ if $1 \leq j \leq \frac{n-1}{2}$ and $s_{2(j-\frac{n-1}{2})}$ otherwise.

And with $u_j = m_{j,1} + m_{j,2} - s_j$,

$$I_{s_{n-2}}(\phi_{s_{n-2}}) = \left(\prod_{j=1}^{n-2} t_j! \binom{m_{j,1}}{t_j} \binom{m_{j,2}}{t_j} \binom{t_j}{u_j - t_j} \right) I_{s_{n-2}}(\circ_{j \in \{1,\cdots,n-1\}} f_{i_j}),$$

concluding the proof. □

Theorem 6.19 *The expression of the n-th moment of the number of $(k-1)$-simplices is given by:*

$$\mathbf{E}_\lambda \left[\widetilde{N_k}^n \right] = \sum_{i_1,\cdots,i_n} \sum_{s_1,\cdots,s_{n-3}} \sum_{t_1,\cdots,t_{n-2}} \lambda^{nk+c} \left(\prod_{j=1}^{n} \lambda^{-i_j} \binom{k}{i_j} \right)$$
$$\left(\prod_{j=1}^{n-2} t_j! \binom{m_{j,1}}{t_j} \binom{m_{j,2}}{t_j} \binom{t_j}{u_j - t_j} \right) \mathcal{J}_n(i_1, \cdots, i_n, s_1, \cdots, s_{n-3}, t_1, \cdots, t_{n-2}).$$

With for $j \in \{1, \cdots, n-2\}$:

- *if $j \leq n-3$, $s_j \geq |m_{j,1} - m_{j,2}|$,*
- *$m_{j,1} = i_{2j-1}$ if $1 \leq j \leq \lfloor \frac{n}{2} \rfloor$ and $s_{2(j-\lfloor \frac{n}{2} \rfloor)-1}$ otherwise,*
- *$m_{j,2} = i_{2j}$ if $1 \leq j \leq \lfloor \frac{n}{2} \rfloor$ and $s_{2(j-\lfloor \frac{n}{2} \rfloor)}$ otherwise,*
- *$m_{j,3} = s_j$ if $1 \leq j \leq n-3$ and s_{n-3} otherwise,*
- *$u_j = m_{j,1} + m_{j,2} - m_{j,3}$,*
- *If n is even, then $c = s_{n-3}$ and $s_{n-3} \geq |m_{n-2,1} - m_{n-2,2}| \vee |m_{n-3,1} - m_{n-3,2}|$,*
- *If n is odd, then $c = i_n$ and $i_n \geq |m_{n-2,1} - m_{n-2,2}|$.*

Proof: The expression of the n-th power of the number of $(k-1)$-simplices is given in Eq. (6.10):

$$\widetilde{N}_k^n = \sum_{i_1,\cdots,i_n=1}^{k} \sum_{s_1,\cdots,s_{n-2}} \sum_{t_1,\cdots,t_{n-2}} \left(\prod_{j=1}^{n-2} t_j! \binom{m_{j,1}}{t_j} \binom{m_{j,2}}{t_j} \binom{t_j}{u_j - t_j} \right)$$
$$I_a(\circ_{j \in A} f_{i_j}) I_b(\circ_{j \in \overline{A}} f_{i_j}).$$

If n is even, we have:

$$\widetilde{N}_k^n = \sum_{i_1,\cdots,i_n=1}^{k} \sum_{s_1,\cdots,s_{n-2}} \sum_{t_1,\cdots,t_{n-2}} \left(\prod_{j=1}^{n-2} t_j! \binom{m_{j,1}}{t_j} \binom{m_{j,2}}{t_j} \binom{t_j}{u_j - t_j} \right)$$
$$I_{s_{n-3}}(\circ_{j \in A} f_{i_j}) I_{s_{n-2}}(\circ_{j \in \overline{A}} f_{i_j}).$$

So let us focus on the only part of the equation that is likely to change when we take the expected value, which we will denote:

$$K = \sum_{s_{n-3}} \sum_{s_{n-2}} I_{s_{n-3}}(\circ_{j\in A} f_{i_j}) I_{s_{n-2}}(\circ_{j\in\overline{A}} f_{i_j}).$$

We then use the property of Eq. (2.4) and recognize the integral from Definition 6.4:

$$\begin{aligned}
\mathbf{E}_\lambda[K] &= \sum_{s_{n-3}} \left(\prod_{j=1}^{n} \lambda^{k-i_j} \binom{k}{i_j} \right) \lambda^{s_{n-3}} \mathcal{J}_n(i_1, \cdots, i_n, s_1, \cdots, s_{n-3}, t_1, \cdots, t_{n-2}) \\
&= \sum_{s_{n-3}} \lambda^{nk+s_{n-3}} \left(\prod_{j=1}^{n} \lambda^{-i_j} \binom{k}{i_j} \right) \mathcal{J}_n(i_1, \cdots, i_n, s_1, \cdots, s_{n-3}, t_1, \cdots, t_{n-2}),
\end{aligned}$$

with $s_{n-3} \geq |m_{n-2,1} - m_{n-2,2}| \vee |m_{n-3,1} - m_{n-3,2}|$.

Then for n odd we directly write:

$$\begin{aligned}
K' &= \sum_{i_n} \sum_{s_{n-2}} I_{i_n}(\circ_{j\in I} f_{i_j}) I_{s_{n-2}}(\circ_{j\in\overline{I}} f_{i_j}), \\
\mathbf{E}_\lambda\left[K'\right] &= \sum_{i_n} \lambda^{nk+i_n} \left(\prod_{j=1}^{n} \lambda^{-i_j} \binom{k}{i_j} \right) \mathcal{J}_n(i_1, \cdots, i_n, s_1, \cdots, s_{n-3}, t_1, \cdots, t_{n-2}),
\end{aligned}$$

with $i_n \in \{|m_{n-2,1} - m_{n-2,2}| \vee 1, k\}$.

The binomials with the i_j allow us to relax the boundaries on the sums on i_j, concluding the proof. □

6.5 Convergence

Let Γ be an arbitrary connected simplicial complex containing n points and $C_\epsilon(\omega)$ be the random simplicial complex by the Poisson point process ω. The number of occurrences of Γ in $C_\epsilon(\omega)$ is denoted as $G_\Gamma(\omega)$. It must be noted that with our construction of the simplicial complex, a complex Γ appears in $C_\epsilon(\omega)$ as soon as its edges are in $C_\epsilon(\omega)$. The set of edges of Γ, denoted by J_Γ is a subset of $\{1, \cdots, n\} \times \{1, \cdots, n\}$. Let

$$\widetilde{h}(x_1, \cdots, x_n) = \frac{1}{c_\Gamma} \prod_{(i,j)\in J_\Gamma} \mathbf{1}_{[\|x_i - x_j\| \leq \epsilon]},$$

where c_Γ is the number of permutations of $\{x_1, \cdots, x_n\}$ such that

$$\widetilde{h^\Gamma}(x_1, \cdots, x_n) = \widetilde{h^\Gamma}(x_{\sigma(1)}, \cdots, x_{\sigma(n)}),$$

and let $f^\Gamma(x_1, \cdots, x_n)$ be the symmetrization of $\widetilde{h^\Gamma}(x_1, \cdots, x_n)$. Then, we have:

$$G_\Gamma = \sum_{\substack{x_1,\cdots,x_n\in\omega \\ x_i\neq x_j if i\neq j}} f^\Gamma(x_1, \cdots, x_n) = \int_{\Delta_n} f^\Gamma(x_1, \cdots, x_n)\, \mathrm{d}\omega(x_1) \cdots\, \mathrm{d}\omega(x_n). \tag{6.11}$$

Lemma 6.20 *The random variable G_Γ has a chaos representation given by:*

$$G_\Gamma = \sum_{i=0}^{n} I_i(f_i^\Gamma),$$

where f_i^Γ is a bounded symmetric function given by

$$f_i^\Gamma(x_{i+1}, \cdots, x_n) = \binom{n}{i} \lambda^{n-i} \int\limits_{B^{n-i}} f^\Gamma(x_1, \cdots, x_n)\ dx_1 \dots\ dx_{n-i}, \tag{6.12}$$

for any $i \in \{1, \cdots, n\}$.

Proof: From (6.11), using the binomial expansion and some algebra, we obtain

$$G_\Gamma = \sum_{i=0}^{n} \int\limits_{\Delta_i} \left(\binom{n}{i} \int\limits_{\Delta_{n-i}} f^\Gamma(x_1, \cdots, x_n) \lambda\ \mathrm{d}x_1 \dots \lambda\ \mathrm{d}x_{n-i} \right)$$
$$(\ \mathrm{d}\omega(x_{n-i+1}) - \lambda\ \mathrm{d}x_{n-i+1}) \dots (\ \mathrm{d}\omega(x_n) - \lambda\ \mathrm{d}x_n).$$

To conclude the proof, we note that, since the torus is a compact set and h^Γ is bounded, f_i^Γ is bounded. □

Lemma 6.21 *For any Γ connected simplicial complex containing n points, for λ large enough,*

$$\mathbf{E}_\lambda [G_\Gamma] \leq c\lambda^n \text{ and } \mathbf{V}_\lambda [G_\Gamma] \leq P_\Gamma^{2n-1}(\lambda).$$

where $P_\Gamma^{2n-1}(\lambda)$ is a polynomial on λ of degree $2n-1$ depending on Γ.

Proof: Using Lemma 6.20 and the chaos properties, we obtain

$$\mathbf{E}_\lambda [G_\Gamma] = \lambda^n \int_{\Delta_n} f^\Gamma(x_1, \cdots, x_n)\ \mathrm{d}x_1 \dots\ \mathrm{d}x_n \leq c\lambda^n,$$

since f^Γ is bounded. Furthermore,

$$\begin{aligned}
\mathbf{V}_\lambda [G_\Gamma] &= \sum_{i=1}^{n} i! \|f_i^\Gamma\|^2_{L^2(B,\lambda)} \\
&= \sum_{i=1}^{n} i! \int\limits_{\Delta_i} \left(\lambda^{n-i} \binom{n}{i} \int\limits_{B^{n-i}} f^\Gamma(x_1, \cdots, x_n)\ \mathrm{d}x_1 \dots\ \mathrm{d}x_{n-i} \right)^2 \lambda\ \mathrm{d}x_1 \dots \lambda\ \mathrm{d}x_i \\
&= \sum_{i=1}^{n} i!\, \lambda^{2n-i} \int\limits_{\Delta_i} \left(\binom{n}{i} \int\limits_{B^{n-i}} f^\Gamma(x_1, \cdots, x_n)\ \mathrm{d}x_1 \dots\ \mathrm{d}x_{n-i} \right)^2 \mathrm{d}x_1 \dots\ \mathrm{d}x_i.
\end{aligned}$$

and since f^Γ is bounded, $\mathbf{V}_\lambda [G_\Gamma]$ is a polynomial of degree $2n-1$. □

Lemma 6.22 *For λ large enough, if $k > 1$,*

$$\mathbf{E}_\lambda \left[(I_k(1))^2 \right] < c\lambda^{2k-1},$$

and $\mathbf{E}_\lambda \left[(I_k(1))^2 \right]$ is constant if $k = 0$.

Proof: The proof is trivial for the case $k = 0$. If $k \geq 1$, for $i \leq k$ we have

$$\int_{\Delta_i} \mathrm{d}\omega(x_1) \dots \mathrm{d}\omega(x_i) = \prod_{j=0}^{i-1} (\omega(B) - j),$$

so we can rewrite $I_k(1)$ as follows:

$$I_k(1) = \sum_{i=0}^{k} \left[\binom{k}{i} (-\lambda S(B))^i \prod_{j=0}^{k-i-1} (\omega(B) - j) \right].$$

Thus, $\mathbf{E}_\lambda \left[(I_k(1))^2 \right]$ can be written as

$$\mathbf{E}_\lambda \left[(I_k(1))^2 \right] = \mathbf{E}_\lambda \left[(\omega(B) - \lambda S(B))^{2k} + \sum_{2 \leq i+j \leq 2k-1} c_{i,j} \omega(B)^i (\lambda S(B))^j \right],$$

where the $c_{i,j}$ are integer constants.

If we differentiate the k-th central moment

$$\mathbf{E}[(N - \lambda')^k] = \sum_{i=0}^{\infty} (r - \lambda')^k e^{-\lambda'} \frac{(\lambda')^i}{i!}$$

of a random variable N distributed as Poisson with mean λ', with respect to λ' we find the following recurrence:

$$\mathbf{E}[(N - \lambda')^{k+1}] = \lambda' \left(\frac{d\mathbf{E}[(N - \lambda')^k]}{d\lambda'} + k\mathbf{E}[(N - \lambda')^{k-1}] \right).$$

Hence, using induction we can show that $\mathbf{E}[(N - \lambda')^k]$ is a polynomial on λ with maximum degree $\lfloor k/2 \rfloor$, for $k > 1$. Since $\mathbf{E}_\lambda \left[\omega(B)^i \right]$ is the Bell Polynomial of degree i on λ, it follows straightforwardly that the polynomial

$$\mathbf{E}_\lambda \left[\sum_{2 \leq i+j \leq 2k-1} c_{i,j} \omega(B)^i (\lambda S(B))^j \right]$$

has degree at most $2k - 1$, and the proof is thus complete. □

Definition 6.5 *Let f_i, g_j and h_k be, respectively, functions of i-th, j-th and k-th chaos of the Wiener-Poisson decomposition of some square integrable function of ω. For $0 \leq s \leq 2(n \wedge m)$, we define*

$$f_i \star_s g_j = \sum_{s \leq 2n \leq 2(s \wedge i \wedge j)} n! \binom{i}{n} \binom{j}{n} f_i \circ_n^{s-n} g_j.$$

For $0 \leq r \leq 2((i + j - s) \wedge k)$, we abuse of the notation to write

$$h_k \star_r (f_n \star_s g_m) = h_k \star_r f_n \star_s g_m.$$

Lemma 6.23 *If $|f_n(x_1, \cdots, x_n)|$ is bounded by a positive real c, then*

$$\mathbf{E}_\lambda \left[I_n(f_n)^2 \right] \leq c^2 \mathbf{E}_\lambda \left[I_n(1)^2 \right].$$

Proof: We use the isometry formula given by Eq. (2.5), so

$$\begin{aligned}\mathbf{E}_\lambda\left[I_n(f_n)^2\right] &= n!\|f_n\|_{L^2(\lambda)^{\circ n}}\\ &= n!\int_{B^n} f_n^2(x_1,\cdots,x_n)\lambda\ \mathrm{d}x_1\ldots\lambda\ \mathrm{d}x_n\\ &\leq n!\int_{B^n} c^2\lambda\ \mathrm{d}x_1\ldots\lambda\ \mathrm{d}x_n\\ &= c^2\mathbf{E}_\lambda\left[I_n(1)^2\right],\end{aligned}$$

and the proof is complete. □

Theorem 6.24 *Let* $F=\frac{G_\Gamma-\mathbf{E}_\lambda[G_\Gamma]}{\sqrt{\mathbf{V}_\lambda[G_\Gamma]}}$*, then, for* λ *large enough,*

$$\int_B \mathbf{E}_\lambda\left[|D_tF|^2|D_tL^{-1}F|\right]\lambda\ dt \leq \frac{c}{\lambda^{1/2}}.$$

Proof: Provided that G_Γ has n points, Lemma 6.20 shows that $G_\Gamma=\sum_{i=0}^n I_n(f_n^\Gamma)$, so

$$\begin{aligned}D_tF &= \frac{1}{\sqrt{\mathbf{V}_\lambda\left[G_\Gamma\right]}}\sum_{i=1}^n iI_{i-1}(f_i^\Gamma(*,t)),\\ D_tL^{-1}F &= \frac{1}{\sqrt{\mathbf{V}_\lambda\left[G_\Gamma\right]}}\sum_{i=1}^n I_{i-1}(f_i^\Gamma(*,t)).\end{aligned}$$

Let us define

$$g_{i-1}=\frac{f_i^\Gamma(*,t)}{\lambda^{n-i}}.$$

According to Eq. (6.12), we note that g_i does not depend on λ. Using the triangular inequality, we have

$$|D_tF|^2|D_tL^{-1}F| \leq \sum_{i,j,k=0}^{n-1}\frac{\lambda^{3n-3-i-j-k}(i+1)(j+1)}{\mathbf{V}_\lambda\left[G_\Gamma\right]^{\frac{3}{2}}}|I_i(g_i)I_j(g_j)I_k(g_k)|.$$

Then, we apply twice the chaos expansion and use again the triangular inequality to obtain:

$$|D_tF|^2|D_tL^{-1}F| \leq \sum_{i,j,k=0}^{n-1}\sum_{s=0}^{2(i\wedge j)}\sum_{r=0}^{2((i+j-s)\wedge k)}\frac{\lambda^{3n-3-i-j-k}(i+1)(j+1)}{\mathbf{V}_\lambda\left[G_\Gamma\right]^{\frac{3}{2}}}\times |I_{i+j+k-s-r}(g_i\star_r g_j\star_s g_k)|,$$

Since f_i is bounded, g_i is bounded as so $g_i\star_r g_j\star_s g_k$ for i,j,k,r,s in the range of their indexes above. We define

$$c(i,j,k,r,s)=\sup\{g_i\star_r g_j\star_s g_k\}(i+1)(j+1),$$

and we use Jensen's inequality and Lemma 6.23 to write

$$\begin{aligned}
&\mathbf{E}_\lambda\left[|D_tF|^2|D_tL^{-1}F|\right] \\
&\quad\leq \sum_{i,j,k=0}^{n-1}\sum_{s=0}^{2(i\wedge j)}\sum_{r=0}^{2((i+j-s)\wedge k)}\frac{\lambda^{3n-3-i-j-k}}{\mathbf{V}_\lambda\left[G_\Gamma\right]^{\frac{3}{2}}}\mathbf{E}_\lambda\left[\left(I_{i+j+k-s-r}(g_i\star_r g_j\star_s g_k)\right)^2\right]^{\frac{1}{2}} \\
&\quad\leq \sum_{i,j,k=0}^{n-1}\sum_{s=0}^{2(i\wedge j)}\sum_{r=0}^{2((i+j-s)\wedge k)}\frac{\lambda^{3n-3-i-j-k}c(i,j,k,r,s)}{\mathbf{V}_\lambda\left[G_\Gamma\right]^{\frac{3}{2}}}\mathbf{E}_\lambda\left[\left(I_{i+j+k-s-r}(1)\right)^2\right]^{\frac{1}{2}}.
\end{aligned}$$

Using Lemmas 6.21 and 6.22 we obtain:

$$\begin{aligned}
\int_B \mathbf{E}_\lambda\left[|D_tF|^2|D_tL^{-1}F|\right]\lambda\,\mathrm{d}t &\leq \sum_{i,j,k=0}^{n-1}\sum_{s=0}^{2(i\wedge j)}\sum_{r=0}^{2((i+j-s)\wedge k)}c(i,j,k,r,s)\times \\
&\quad\frac{\lambda^{3n-3-i-j-k}}{\mathbf{V}_\lambda\left[G_\Gamma\right]^{\frac{3}{2}}}\mathbf{E}_\lambda\left[\left(I_{i+j+k-s-r}(1)\right)^2\right]^{1/2}\int_B\lambda\,\mathrm{d}t \\
&\leq \frac{c\lambda^{3n-3}}{\mathbf{V}_\lambda\left[G_\Gamma\right]^{\frac{3}{2}}}\int_B\lambda\,\mathrm{d}t\leq\frac{c}{\lambda^{1/2}},
\end{aligned}$$

concluding the proof. □

Theorem 6.25 *Let*

$$F=\frac{G_\Gamma-\mathbf{E}_\lambda\left[G_\Gamma\right]}{\sqrt{Var(G_\Gamma)}}.$$

Then, when λ is large enough

$$\mathbf{E}_\lambda\left[\left|1-\langle DF,DL^{-1}F\rangle_{L^2(\lambda)}\right|\right]\leq\frac{c}{\lambda^{1/2}},$$

for some constant c.

Proof: The expressions of D_tF and $D_tL^{-1}F$ are given by

$$D_tF=\frac{1}{\sqrt{\mathbf{V}_\lambda\left[G_\Gamma\right]}}\sum_{i=1}^n iI_{i-1}(f_i(*,t)),$$

and

$$D_tL^{-1}Z=\frac{1}{\sqrt{\mathbf{V}_\lambda\left[G_\Gamma\right]}}\sum_{i=1}^n I_{i-1}(f_i(*,t)).$$

The inner product $\langle D_tL^{-1}F,D_tF\rangle_{L^2(\lambda)}$ is expressed by:

$$\langle D_tL^{-1}F,D_tF\rangle_{L^2(\lambda)}=\frac{1}{\mathbf{V}_\lambda\left[G_\Gamma\right]}\int_B\sum_{i,j=1}^n iI_{i-1}(f_i(*,t))I_{j-1}(f_j(*,t))\lambda\,\mathrm{d}t.$$

Then,

$$\begin{aligned}\langle D_t L^{-1} Z, D_t Z\rangle_{L^2(\lambda)} &= \frac{1}{\mathbf{V}_\lambda[G_\Gamma]} \sum_{i,j=1}^{n} i \int_B I_{i-1}(f_i(*,t)) I_{j-1}(f_j(*,t))\lambda \, \mathrm{d}t \\ &= \frac{1}{\mathbf{V}_\lambda[G_\Gamma]} \int_B I_0(f_1(t)^2 \lambda \, \mathrm{d}t \\ &+ \frac{1}{\mathbf{V}_\lambda[G_\Gamma]} \sum_{\substack{i,j=1 \\ (i,j)\neq(1,1)}}^{n} i \int_B I_{i-1}(f_i(*,t)) I_{j-1}(f_j(*,t))\lambda \, \mathrm{d}t.\end{aligned}$$

Defining g_{i-1} as in Theorem 6.24 and using the chaos expansion, we get:

$$\begin{aligned}\langle D_t L^{-1} F, D_t F\rangle_{L^2(\lambda)} &= \frac{\|f_1\|^2_{L^2(\lambda)}}{\mathbf{V}_\lambda[G_\Gamma]} + \frac{1}{\mathbf{V}_\lambda[G_\Gamma]} \sum_{i=2}^{n} i(i-1)! \int_B \|f_n(*,t)\|^2_{L^2(\lambda)} \, \lambda \, \mathrm{d}t \\ &+ \sum_{i=1}^{n-1} (i+1) \frac{\lambda^{2n-2i-2}}{\mathbf{V}_\lambda[G_\Gamma]} \int_B \sum_{s=0}^{2(i-1)} I_{2i-s}(g_i \star_s g_i)\lambda \, \mathrm{d}t \\ &+ \sum_{i,j=0}^{n-1} (i+1) \frac{\lambda^{2n-i-j-2}}{\mathbf{V}_\lambda[G_\Gamma]} \int_B \sum_{s=0}^{2(i\wedge j)} I_{i+j-s}(g_i \star_s g_j)\lambda \, \mathrm{d}t.\end{aligned}$$

Since

$$\begin{aligned}\int_B \|f_i(*,t)\|^2_{L^2(\lambda)} \, \lambda \, \mathrm{d}t &= \int_B \left(\int_{B^{i-1}} f_i^2(t_1, \cdots, t_{i-1}, t)\lambda \, \mathrm{d}t_1 \ldots \lambda \, \mathrm{d}t_{i-1} \right) \lambda \, \mathrm{d}t \\ &= \int_{B^i} f_i^2(t_1, \cdots, t_{i-1}, t)\lambda \, \mathrm{d}t_1 \ldots \lambda \, \mathrm{d}t_{i-1} \lambda \, \mathrm{d}t \\ &= \|f_i\|_{L^2(\lambda)},\end{aligned}$$

and given the isometry formula

$$\begin{aligned}\mathbf{V}_\lambda[G_\Gamma] &= \|G_\Gamma\|_{L^2\Omega} - \mathbf{E}_\lambda[G_\Gamma]^2 \\ &= \sum_{i=0} n! \|f_i\|^2_{L^2(B^i)} - \|f_0\|^2 = \sum_{i=1} n! \|f_i\|^2_{L^2(\lambda)},\end{aligned}$$

we have

$$\frac{\|f_1\|^2_{L^2(\lambda)}}{\mathbf{V}_\lambda[G_\Gamma]} + \frac{1}{\mathbf{V}_\lambda[G_\Gamma]} \sum_{i=2}^{n} i(i-1)! \int_B \|f_i(*,t)\|^2_{L^2(\lambda)} \, \lambda \, \mathrm{d}t = 1.$$

Hence

$$\langle D_t L^{-1} Z, D_t Z\rangle_{L^2(\lambda)} = 1 + \sum_{\substack{i,j=0 \\ (i,j)\neq(1,1)}}^{n-1} \sum_{\substack{s=0 \\ s\neq 2i \text{ if } i=j}}^{2(i\wedge j)} (i+1) \frac{\lambda^{2n-i-j-2}}{\mathbf{V}_\lambda[G_\Gamma]} \int_B I_{i+j-s}(g_i \star_s g_j)\lambda \, \mathrm{d}t.$$

Let $c(i,j,s)$ be defined as

$$c(i,j,s) = \sup\{g_i \star_s g_j\}(i+1).$$

Then, we use the triangular inequality, Jensen's inequality and Lemma 6.23 to obtain:

$$\begin{aligned}
&\mathbf{E}_\lambda\left[|1-\langle DF, DL^{-1}F\rangle_{L^2(\lambda)}|\right] \\
&\quad\leq \mathbf{E}_\lambda\Bigg[\sum_{\substack{i,j=0\\(i,j)\neq(1,1)}}^{n-1}\sum_{\substack{s=0\\s\neq 2i \text{ if } i=j}}^{2(i\wedge j)}(i+1)\frac{\lambda^{2n-i-j-2}}{\mathbf{V}_\lambda[G_\Gamma]}\int_B |I_{i+j-s}(g_i\star_s g_j)|\,\lambda\,\mathrm{d}t\Bigg] \\
&\quad\leq \sum_{\substack{i,j=0\\(i,j)\neq(1,1)}}^{n-1}\sum_{\substack{s=0\\s\neq 2i \text{ if } i=j}}^{2(i\wedge j)}\frac{(i+1)\lambda^{2n-i-j-2}}{\mathbf{V}_\lambda[G_\Gamma]}\int_B \mathbf{E}_\lambda\left[(I_{i+j-s}(g_i\star_s g_j))^2\right]^{\frac{1}{2}}\lambda\,\mathrm{d}t \\
&\quad\leq \sum_{\substack{i,j=0\\(i,j)\neq(1,1)}}^{n-1}\sum_{\substack{s=0\\s\neq 2i \text{ if } i=j}}^{2(i\wedge j)}c(i,j,s)\frac{\lambda^{2n-i-j-2}}{\mathbf{V}_\lambda[G_\Gamma]}\mathbf{E}_\lambda\left[(I_{i+j-s}(1))^2\right]^{1/2}\int_B \lambda\,\mathrm{d}t.
\end{aligned}$$

Finally, using Lemmas 6.22 and 6.21, there is a constant c such that:

$$\mathbf{E}_\lambda\left[|1-\langle DF, DL^{-1}F\rangle_{L^2(\lambda)}|\right]\leq\frac{c}{\lambda^{1/2}}$$

for λ large enough. □

Theorem 6.26 *There exists a constant c such that, for λ large enough, the Wasserstein distance between $F=\frac{G_\Gamma-\mathbf{E}_\lambda[G_\Gamma]}{\sqrt{Var(G_\Gamma)}}$ and $\mathcal{N}(0,1)$ is given by:*

$$d_W(F,\mathcal{N}(0,1))\leq\frac{c}{\lambda^{1/2}}.$$

Proof: The proof comes straightforwardly from Theorem 3.1 as stated in [87]:

$$d_W(F,\mathcal{N}(0,1))\leq\mathbf{E}_\lambda\left[|1-\langle DF, DL^{-1}F\rangle_{L^2(\lambda)}|\right]+\int_B\mathbf{E}_\lambda\left[|D_tF|^2|D_tL^{-1}F|\right]\lambda\,\mathrm{d}t,$$

which we can use since f_n is bounded, so $F\in\mathrm{Dom}\,D$. We use theorems 6.24 and 6.25 in the first and second terms, respectively.

□

Remark 6.6 We note that for any definition of $\tilde{h}(x_1,\cdots,x_n)$ such that $\tilde{h}(x_1,\cdots,x_n)\geq 1$, this theorem will hold. This means that, as long as the formation of simplices depends only on the positions of the points, the law of the number of connected components will converge to a Gaussian law. For example, this cover the cases on $\mathbf{R}^d$, on d-torus, for Čech complexes or Rips complexes.

6.6 Summary

In this chapter, we have obtained a way to find the n-th moment of the k-simplices in a random simplicial complex from points of a Poisson and Binomial point process as vertices in d dimensions. This is done by using the chaos expansion and evaluating some integrals. We have seen that the expressions of such moments are very complex for moments of order superior or equal to three, but the mean is expressed in a very simplified fashion and the variance is given by the sum of k terms. The same method used to find the variance can

be applied to find the covariance between the number of k-simplices and l-simplices and its expression is given by a sum of the minimum of k and l terms.

With respect to the Euler characteristic, we have calculated also a closed-form expression for its mean using the alternating sum of the mean of k-simplices, resulting in a Bell polynomial over as a function of $\lambda(2\epsilon)^d$. Moreover, we have found an infinite power series to express its variance by using the covariance result. Then, we have simplified the variance of the Euler characteristic for $d = 1$ and we have simplified two terms over $d + 1$ of the variance in d dimensions. A concentration inequality has been used to find an upper bound to the tail of the Euler characteristic distribution. We have also conjectured that at most two kinds of holes are predominant.

Although we determine a way to find the n-th moment of N_k, their expressions are too complex for $n \geq 3$ and so it is not possible to find their distributions. However, in Section 6.5, we have used Theorem 3.1 as stated in [87] to determined an upper bound for the distance between the distribution of N_k and the Gaussian one with mean $\mathbf{E}_\lambda[N_k]$ and variance $\mathrm{Var}(N_k)$. This result holds in any dimension for points deployed over any compact set and actually the result holds not just for k-simplices, but for any kind of connected graph.

Chapter 7

One-dimensional Case

7.1 Introduction

The previous chapter considers sensors represented by a random configuration on d dimensions. Although we can obtain important results concerning the moments and distributions of quantities such as the number of k-simplices and Euler characteristic, there are some unanswered questions regarding the number of connected components and coverage. Trying to answer those questions, we quickly realized that the dimension of the ambient space played a key role. We then first began by the analysis of dimension 1, which appeared to be the simplest situation. In this case, there is no need of algebraic topology so we will not go further in the description of this line of thought even if it was our first motivation.

In dimension 1, the only question of interest is that of the connexity but it can take different forms. Imagine we are given $[0, 1]$ as a domain in which n points $\{x_1, \cdots, x_n\}$ are drawn. For a radius r, one can wonder whether $[0, 1] \subset \cup_{i=1,\cdots,n}[x_i - r, x_i + r]$ or one can investigate whether $[x_i - r, x_i + r] \cap [x_{i+1} - r, x_{i+1} + r] \neq \emptyset$ for all $i = 1, \cdots, n - 1$. The second situation is less restrictive since we do not impose that the frontier of the interval to be covered. Depending on the application we have in mind, both questions are sensible. A slightly different but somehow close problem is that of the circle: Consider now that the points are dispatched along a circle of unit perimeter C_1 and ask again whether $C_1 \subset \cup_{i=1,\cdots,n}B(x_i, r)$ where $B(x, r)$ is the 2-dimensional ball of center x and radius r. Several years ago, this problem has been thoroughly analyzed [106] and references therein) for a fixed number of i.i.d. arcs over the circle. A closed form formula can be given for the probability of coverage as a function of the number and of the common law of the arcs length. Some variations of this problem have been investigated since, see for instance [59]. More recently, in [70], algorithms are devised to determine whether a domain can be protected from intrusion by a "belt" of sensors (namely a ring or the border of a rectangle). There is no performance analysis in this work which is focused on algorithmic solutions for this special problem of coverage. Still motivated by applications to sensor networks, the paper [76] considers the situation where sensors are actually placed in a plan, have a fixed radius of observation and analyses the connectivity of the trace of the covered region over a line. The works of Kahle [66, 65] are actually hardly linked to our results: the motivation is the same, studying the Betti numbers of some random simplicial complexes but the results are only asymptotic and valid in dimension greater than 2.

Our main result is the distribution of the number of connected components for a Poisson distribution of sensors in a bounded interval. We could not use the method of [106] since

the number of gaps does not determine the connectivity of the domain. For instance, one may have only one gap at the "beginning" which means that all the points are pairwise within the threshold distance and thus that the network is connected or one may have only one gap in the "middle" which means that there is a true hole of connectivity.

Actually, our method is very much related to queueing theory. Indeed, clusters, i.e. sequence of neighboring points, are the strict analogous of busy periods – see Section 7.2. As will appear below, our analysis turns down to be that of an M/D/1/1 queue with preemption: When a customer arrives during a service, it preempts the server and, since there is no buffer, the customer who was in service is removed from the queuing system. This analogy led us to use standard tools of queueing theory: Laplace transform and renewal processes – see for instance [7, 19]. This works perfectly and with a bit of calculus, we can compute all the characteristics we are interested in. It is worthwhile to note that a queueing model (namely the M/G/∞) also appears in [76].

The paper is organized as follows: Section 2 presents the model and defines the relevant quantities to be calculated. The calculations and analytical results are presented in Section 3. For our situation, we find results analogous to that of [106]. In section 4, two other scenarios are presented, considering the number of incomplete clusters and clusters placed in a circle. In Section 5, numerical examples are presented and analyzed.

7.2 Problem Formulation

Let $L > 0$, we assume that we are given a Poisson process, denoted by N, of intensity λ on $[0, L]$. Let $(X_i, i \geq 1)$ be the atoms of N. We thus know that the random variables, $\Delta X_i = X_{i+1} - X_i$ are i.i.d. and exponentially distributed. We fix $\epsilon > 0$. Two points, located respectively at x and y, are said to be *directly connected* whenever $|x - y| \leq \epsilon$. For $i < j$, two points of N, say X_i and X_j, are indirectly connected if X_l and X_{l+1} are directly connected for any $l = i, \cdots, j-1$. A set of points directly or indirectly connected is called a *cluster*, a complete cluster is a cluster which begins and ends within $[0, L]$. The connectivity of the whole network is measured by the number of clusters.

The number of points in the interval $[0, x]$ is denoted by $N_x = \sum_{n=0}^{\infty} \mathbf{1}_{[X_n \leq x]}$. The random variable A_i given by

$$A_i = \begin{cases} X_1 & \text{if } i = 1, \\ \inf\{X_j \,:\, X_j > A_{i-1},\, X_j - X_{j-1} > \epsilon\} & \text{if } i > 1, \end{cases}$$

represents the beginning of the i-th cluster, denoted by C_i. In the same way, the end of this same cluster, E_i, is defined by

$$E_i = \inf\{X_j + \epsilon \,:\, X_j > A_i, X_{j+1} - X_j > \epsilon\}.$$

So, the i-th cluster, C_i, has a number of points given by $N_{E_i} - N_{A_i}$. We define the length B_i of C_i as $E_i - A_i$. The intercluster size, D_i, is the distance between the end of C_i and the beginning of C_{i+1}, which means that $D_i = A_{i+1} - E_i$ and ΔA_i is the distance between the first points of two consecutive clusters C_i, given by $\Delta A_i = A_{i+1} - A_i = B_i + D_i$.

Remark 7.1 With this set of assumptions and definitions, we can see our problem as an $M/D/1/1$ preemptive queue, see Fig. 7.1. In this non-conservative system, the service time is deterministic and given by ϵ. When a customer arrives during a service, the served customer is removed from the system and replaced by the arriving customer. Within this

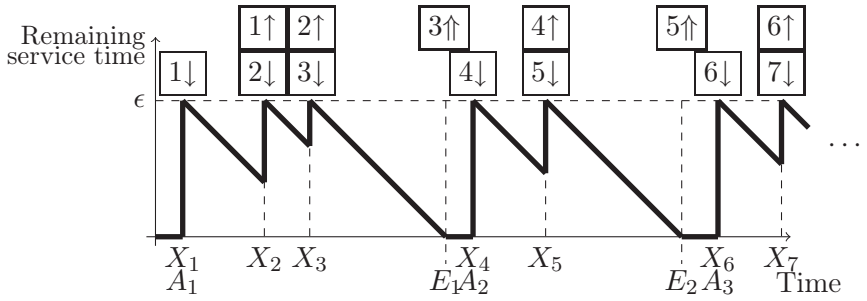


Figure 7.1: Queueing representation of the proposed problem. A down arrow denotes that user i starts to be served. An up arrow indicates that user i leaves the system without have finished the service. A double up arrow ilustrates that the service of user i finishes. It is also shown the beginning and the end of the ith busy period, respectively, A_i and E_i.

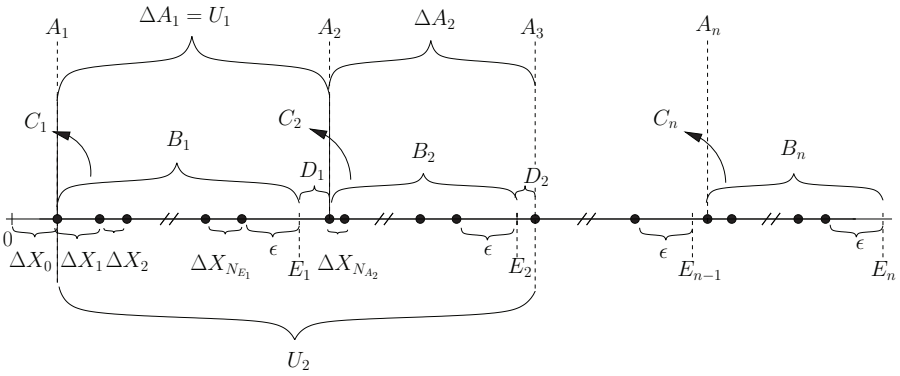


Figure 7.2: Definitions of the relevant quantities of the network: distance between points, distance between clusters, the size of clusters, the size interclusters, the beginning of clusters and the end of clusters.

framework, a cluster corresponds to what is called a busy period, the intercluster size is an idle time and $A_i + D_i$ is the length of the i-th cycle.

The number of complete clusters in $[0,\, L]$ corresponds to the number of connected components $\beta_0(L)$ (since in dimension 1, it coincides with the Euler characteristics of the union of intervals, see [42]) of the network. The distance between the beginning of the first cluster and the beginning of the $(i+1)$-th one is defined as $U_i = \sum_{k=1}^{i} \Delta A_k$. We also define $\Delta X_0 = D_0 = X_1$. Figure 7.2 illustrates these definitions.

For the sake of completeness, we recall the essentials of Markov process theory needed to go along, for further details we refer for instance to [19] and [7]. In what follows, for a process X, $(\mathcal{F}_t^X,\ t \geq 0)$ is the filtration generated by the sample-paths of X:

$$\mathcal{F}_t^X = \sigma\{X(s),\, s \geq t\}.$$

Definition 7.1 *A process $(X(t),\, t \geq 0)$ with values in a denumerable space E is said to be Markov whenever*

$$\mathbf{E}_\lambda\left[F\right](X(t+s))\mathcal{F}_t^X = \mathbf{E}_\lambda\left[F\right](X(t+s))X(t),$$

for any bounded function F from E to $\mathbf{R}$, any $t \geq 0$ and $s \geq 0$.

Equivalently, a process X is Markov if and only if given the present (i.e. given $X(t)$), the past (i.e. the sample-path of X before time t) and the future (i.e. the sample-path of X after time t) of the process are independent.

Definition 7.2 *A random variable τ with values in $\mathbf{R}^+ \cup \{+\infty\}$ is an $\mathcal{F}^X$-stopping time whenever for any $t \geq 0$, the event $\{\tau \leq t\}$ belongs to $\mathcal{F}_t^X$.*

The point is that (7.1) still holds when t is replaced by a stopping time τ: Given $X(\tau)$, the past and the future of X are independent. X is then said to be strong Markov. This property always holds for Markov processes with values in a denumerable space but is not necessarily true for Markov processes with values in an arbitrary space.

From now on, the Markov process under consideration is N, the Poisson process of intensity λ over $[0, L]$.

Lemma 7.1 *For any $i \geq 1$, A_i and E_i are stopping times.*

Proof: Let us consider the filtration $\mathcal{F}_t^N = \sigma\{N_a, a \leq t\}$. For $i = 1$, we have

$$\{A_1 \leq t\} \Leftrightarrow \{X_1 \leq t\} \Leftrightarrow \{N_t \geq 1\} \in \mathcal{F}_t^N.$$

Thus, A_1 is a stopping time. For A_2, we have

$$\{A_2 > t\} \Leftrightarrow \bigcup_{n\geq 1} \left\{ N_t = n, \bigcup_{j=1}^{n} \left\{ \Delta X_j \geq \epsilon, \bigcup_{k=j+1}^{n} \{\Delta X_k \leq \epsilon\} \right\} \right\} \in \mathcal{F}_t^N,$$

so A_2 is also a stopping time. We proceed along the same line for others A_i and as well for E_i to prove that they are stopping times. □

Since N is a (strong) Markov process, the next corollary is immediate.

Corollary 7.2 *The set $\{B_i, D_i, i \geq 1\}$ is a set of independent random variables. Moreover, D_i is distributed as an exponential random variable with mean $1/\lambda$ and the random variables $\{B_i, i \geq 1\}$ are i.i.d.*

7.3 Calculations

Throughout this section, we find first the Laplace transforms of B_i, ΔA_i, U_i and the probability that there are n clusters in the interval. Then, we find the analytical expression of those quantities by inverting their Laplace transform.

7.3.1 Laplace transforms

We find first the Laplace transform of the distribution of B_i, the size of a cluster.

Corollary 7.3 *The Laplace transform of the distribution of B_i, is given by*

$$\mathbf{E}\left[e^{-sB_i}\right] = \frac{1}{\lambda} \frac{\lambda + s}{\frac{se^{\lambda\epsilon}}{\lambda} e^{s\epsilon} + 1}. \tag{7.1}$$

Proof: Since ΔX_j is an exponentially distributed random variable,

$$\begin{aligned}\mathbf{E}\left[e^{-s\Delta X_j}\mathbf{1}_{[\Delta X_j\leq\epsilon]}\right] &= \int_0^\epsilon e^{-st}\lambda e^{-\lambda t}dt \\ &= \frac{\lambda}{s+\lambda}1-e^{-(s+\lambda)\epsilon}\end{aligned}$$

and $\mathbf{E}\left[e^{-s\Delta X_j}\mathbf{1}_{[\Delta X_j>\epsilon]}\right] = e^{-\lambda\epsilon}$. By Corollary 7.2, since the B_i's are i.i.d, it suffices to calculate $\mathbf{E}\left[e^{-sB_1}\right]$. Hence, the Laplace transform of the distribution of B_i is given by

$$\begin{aligned}\mathbf{E}\left[e^{-sB_i}\right] &= \sum_{n=1}^{\infty}\mathbf{E}\left[e^{-sB_1}, N_{E_1}=n\right] \\ &= \sum_{n=1}^{\infty}\mathbf{E}\left[e^{-s(\sum_{j=1}^{n-1}\Delta X_j+\epsilon)}\mathbf{1}_{[\Delta X_n>\epsilon]}\prod_{j=1}^{n-1}\mathbf{1}_{[\Delta X_j\leq\epsilon]}\right] \\ &= \sum_{n=1}^{\infty}\left(\mathbf{E}\left[e^{-s\Delta X_1}\mathbf{1}_{[\Delta X_1\leq\epsilon]}\right]\right)^{n-1}\mathbf{E}\left[e^{-s\Delta X_n}\mathbf{1}_{[\Delta X_n>\epsilon]}\right]e^{-s\epsilon} \\ &= \sum_{n=0}^{\infty}\left(\frac{\lambda}{s+\lambda}(1-e^{-(s+\lambda)\epsilon})\right)^n e^{-s\lambda}e^{-s\epsilon} \\ &= \frac{1}{\lambda}\frac{\lambda+s}{\frac{se^{\lambda\epsilon}}{\lambda}e^{s\epsilon}+1}, \qquad (7.2)\end{aligned}$$

which concludes the proof. □

From this result, we can immediately calculate the Laplace transform of the distribution of ΔA_i. Since $\Delta A_i = B_i + D_i$, we have $\mathbf{E}\left[e^{-s\Delta A_i}\right] = \mathbf{E}\left[e^{-s(B_i+D_i)}\right]$ and using Corollary 7.2:

$$\mathbf{E}\left[e^{-s\Delta A_i}\right] = \mathbf{E}\left[e^{-sB_i}\right]\mathbf{E}\left[e^{-sD_i}\right] = \frac{1}{\frac{se^{\lambda\epsilon}}{\lambda}e^{s\epsilon}+1}. \qquad (7.3)$$

If we turn our attention to the system seen as a queue, as a remark, this last result can lead us to calculate the probability of the server is busy, or, equivalently, the loss probability, since the interarrival time is exponentially distributed. Due to the fact that A_i and E_i are stopping times, this probability is given by $\mathbf{E}[B_i]/\mathbf{E}[\Delta A_i]$, which can be obtained by:

$$\begin{aligned}\mathbf{P}(\text{Server is busy}) &= \frac{\mathbf{E}[B_i]}{\mathbf{E}[\Delta A_i]} = \frac{-\frac{d}{ds}\mathbf{E}\left[e^{-sB_i}\right]\big|_{s=0}}{-\frac{d}{ds}\mathbf{E}\left[e^{-s\Delta A_i}\right]\big|_{s=0}} \\ &= 1-e^{-\lambda\epsilon}.\end{aligned}$$

The Laplace transform of another r.v., U_n, is found in the next corollary.

Corollary 7.4 *The Laplace transform of the distribution of U_n, for $n \geq 0$ is given by*

$$\mathbf{E}\left[e^{-sU_n}\right] = \frac{1}{\left(\frac{e^{\lambda\epsilon}}{\lambda}se^{s\epsilon}+1\right)^n}. \qquad (7.4)$$

Proof: We use Corollaries 7.2 and 7.3 to calculate the Laplace transform of the distribution of U_n, since $U_n = \sum_{i=1}^{n}(B_i + D_i)$:

$$\begin{aligned}\mathbf{E}\left[e^{-sU_n}\right] &= \mathbf{E}\left[e^{-sU_n}\right] = \mathbf{E}\left[e^{-s(\sum_{i=1}^{i=n}(B_i+D_i))}\right] \\ &= = \prod_{i=1}^{n}\mathbf{E}\left[e^{-sB_i}\right]\mathbf{E}\left[e^{-sD_i}\right] \\ &= \frac{1}{\left(\frac{e^{\lambda\epsilon}}{\lambda}se^{s\epsilon}+1\right)^n},\end{aligned}$$

for $n \geq 0$. □

Let us define $p_n(L) \triangleq \mathbf{P}(\beta_0 = x)$, i.e., $p_n(L)$ is the probability of having n clusters on the interval $[0, L]$. Since for all $L \in \mathbb{R}_+$, $0 \leq p_n(L) \leq 1$, the Laplace transform with respect to L,

$$\mathcal{L}\{p_n(\cdot)\}(s) = \int_0^\infty e^{-st}p_n(t)dt,$$

is well defined, and we are able to compute it with the expression of $\mathbf{E}\left[e^{-sU_n}\right]$:

Theorem 7.5 *For $n \geq 0$, the Laplace transform of the function $p_n(L)$ is given by*

$$\mathcal{L}\{p_n(\cdot)\}(s) = \frac{e^{\epsilon\lambda}}{\lambda}\frac{e^{\epsilon s}}{\left(\frac{e^{\epsilon\lambda}}{\lambda}se^{\epsilon s}+1\right)^{n+1}}. \tag{7.5}$$

Proof: Let $\beta_0 = \beta_0(L)$ be the number of complete clusters on $[0, L]$. To calculate the probability of $\{\beta_0(L) = n\}$, we have to note that

$$\{\beta_0 \geq n\} \Leftrightarrow \begin{cases}\{\Delta X_0 + U_{n-1} + B_n \leq L\} & \text{if } n \geq 1, \\ \{\Delta X_0 < \infty\} & \text{if } n = 0,\end{cases}$$

because the events in each side of the relation represent that there are more than β_0 clusters in the interval $[0, L]$. We can see this condition illustrated in Fig. 7.3 for $n \geq 1$. For $n = 0$, the additional case, $\{\Delta X_0 \geq \infty\}$, is trivial. Since $\{\beta_0 = n\}$ and $\{\beta_0 \geq n + 1\}$ are disjoint

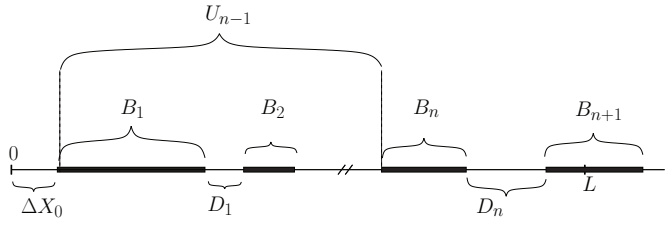


Figure 7.3: Illustration of the condition equivalent to $\beta_0 \geq n$.

events:

$$\begin{aligned}\mathbf{P}(\beta_0 \geq n) &= \mathbf{P}((\beta_0 \geq n+1) \cup (\beta_0 = n)) \\ &= \mathbf{P}(\beta_0 \geq n+1) + \mathbf{P}(\beta_0 = n).\end{aligned}$$

Thus

$$\mathbf{P}(\beta_0 = 0) = 1 - \mathbf{P}(\Delta X_0 + B_1 \leq L),$$

and

$$\mathbf{P}(\beta_0 = n) = \mathbf{P}(\Delta X_0 + U_{n-1} + B_n \leq L) - \mathbf{P}(\Delta X_0 + U_n + B_{n+1} \leq L). \tag{7.6}$$

Let

$$Y_n \triangleq \begin{cases} \Delta X_0 + U_{n-1} + B_n & \text{if } n \geq 1 \\ 0 & \text{if } n = 0 \end{cases},$$

the Laplace transform with respect to x, $\mathbf{P}(Y_n \leq x)$ can be calculated:

$$\begin{aligned} \mathcal{L}\{\mathbf{P}(Y_n \leq \cdot)\}(s) &= \int_0^\infty \mathbf{P}(Y_n \leq x) e^{-sx} dx \\ &= \int_0^\infty \int_0^x dP_{Y_n}(y) e^{-sx} dx = \int_0^\infty \left(\int_y^\infty e^{-sx} dx \right) dP_{Y_n}(y) \\ &= \frac{1}{s} \int_0^\infty e^{-sy} dP_{Y_n}(y) = \frac{1}{s} \mathbf{E}\left[e^{-sY_n}\right] \\ &= \frac{1}{s} \mathbf{E}\left[e^{-s\Delta X_0}\right] \mathbf{E}\left[e^{-sU_{n-1}}\right] \mathbf{E}\left[e^{-sB_n}\right] \\ &= \frac{1}{s} \frac{1}{\left(\frac{e^{\lambda\epsilon}}{\lambda} s e^{s\epsilon} + 1\right)^n} \end{aligned} \tag{7.7}$$

for $n \geq 1$, where we used Corollary 7.2 in the third line. For $n = 0$, the Laplace transform is trivial and given by $\mathcal{L}\{\mathbf{P}(y_0 \leq \cdot)\}(s) = 1/s$. Substituting Eq. (7.7) in the Laplace transform of both sides of Eq. (7.6) yields:

$$\begin{aligned} \mathcal{L}\{p_n(\cdot)\}(s) &= \mathcal{L}\{\mathbf{P}(Y_n \leq \cdot)\}(s) - \mathcal{L}\{\mathbf{P}(Y_{n+1} \leq \cdot)\}(s) \\ &= \frac{e^{\epsilon\lambda}}{\lambda} \frac{e^{\epsilon s}}{\left(\frac{e^{\epsilon\lambda}}{\lambda} s e^{\epsilon s} + 1\right)^{n+1}}, \quad n \geq 0. \end{aligned}$$

The proof is thus complete. □

Lemma 7.6 *Let m be an positive integer. When $\epsilon \to 0$, $\mathbf{E}[\beta_0^m] \to \mathbf{E}[N_L^m]$.*

Proof: Since $\lim_{x \to 0} \mathbf{P}(\Delta X_i < x) = 0$, for almost all N, for any non-negative integer j, if $X_j \geq \eta$, there exists η such that $\Delta X_j \geq \eta$. If $\epsilon < \eta$, then $\beta_0(\epsilon) \overset{a.s.}{=} N_L$. Besides, since $\beta_0 \leq N_L$, it is well known that all the m-th moments of a random variable distributed as Poisson converge. This means that, for any positive integer m, $\mathbf{E}[\beta_0^m]$ also exists and that when $\epsilon \to 0$, $\mathbf{E}[\beta_0^m] = \mathbf{E}[N_L^m]$. □

Let $\mathrm{Li}_t(z)$, $z, t \in \mathbb{R}$, $z < 1$, be the polylogarithm function with parameter t, defined by

$$\mathrm{Li}_t(z) \triangleq \sum_{k=1}^\infty \frac{z^k}{k^t}.$$

For m a positive integer, consider the function of x

$$M_{\beta_0}^m \left(x \mapsto \mathbf{E}[\beta_0^m(x)] = \sum_{i=0}^\infty i^m p_i(x) \right) \tag{7.8}$$

and its Laplace transform given by:

$$\mathcal{L}\left\{M_{\beta_0}^m(\cdot)\right\}(s) = \int_{0^+}^{\infty} \mathbf{E}\left[\beta_0(x)^m\right] e^{-sL} p_n(x) dx.$$

Corollary 7.7 *Let a be defined as follows:*

$$a \triangleq \frac{e^{\epsilon\lambda}}{\lambda} s e^{\epsilon s}.$$

The Laplace transform of the m-th moment of $\beta_0(L)$ is:

$$\mathcal{L}\left\{M_{\beta_0}^m(\cdot)\right\}(s) = \frac{a}{s\,(a+1)} Li_{-m}\left(\frac{1}{a+1}\right), \tag{7.9}$$

which converges, provided that $\frac{1}{a+1} < 1$.

Proof: Applying the Laplace transform of both sides of Eq. (7.8) and using its linearity:

$$\mathcal{L}\left\{M_{\beta_0}^m(\cdot)\right\}(s) = \mathcal{L}\left\{\sum_{i=1}^{\infty} i^m p_i(\cdot)\right\}(s) = \sum_{i=1}^{\infty} (i^m \mathcal{L}\{p_i(\cdot)\}(s)), \tag{7.10}$$

where we can interchange the sum and the Laplace transform due to Lemma 7.6. Thus, it is possible to find the Laplace transform of $\mathbf{E}\left[\beta_0(L)^m\right]$:

$$\begin{aligned}
\mathcal{L}\left\{M_{\beta_0}^m(\cdot)\right\}(s) &= \sum_{i=1}^{\infty} (i^m \mathcal{L}\{p_i(\cdot)\}(s)) \\
&= \frac{\frac{e^{\epsilon\lambda}}{\lambda} e^{\epsilon s}}{\left(\frac{e^{\epsilon\lambda}}{\lambda} s e^{\epsilon s} + 1\right)} \sum_{i=1}^{\infty} \frac{i^m}{\left(\frac{e^{\epsilon\lambda}}{\lambda} s e^{\epsilon s} + 1\right)^i} \\
&= \frac{a}{s\,(a+1)} \mathrm{Li}_{-m}\left(\frac{1}{a+1}\right),
\end{aligned}$$

concluding the proof. □

7.3.2 Analytical expressions

Until this point, we have all the results we are interested in, but in a Laplace transform form. In this section, we find a way to inverse every one of those transforms. We start finding the moments of β_0 and we define $\left\{ {m \atop k} \right\}$ as the Stirling number of the second kind [49].

Corollary 7.8 *The m-th moment of the number of clusters on the interval $[0, L]$ is given by:*

$$M_{\beta_0}^m(L) = \sum_{k=1}^{m} \left\{ {m \atop k} \right\} \left(\frac{L}{\epsilon} - k\right)^k \left(\lambda\epsilon e^{-\epsilon\lambda}\right)^k \mathbf{1}_{[L/\epsilon > k]}. \tag{7.11}$$

Proof: Using the following identity [119] valid for a positive integer m

$$\mathrm{Li}_{-m}(z) = \sum_{k=0}^{m} \frac{(-1)^{m+k} k! \left\{ {m+1 \atop k+1} \right\}}{(1-z)^{k+1}},$$

in the result of Corollary 7.7, we find that

$$\begin{aligned}\mathcal{L}\left\{M_{\beta_0}^m(\cdot)\right\}(s) &= \frac{a}{s}\sum_{k=0}^{m}\frac{(-1)^{m+k}k!\left\{\begin{matrix} m+1 \\ k+1 \end{matrix}\right\}(1+a)^k}{a^{k+1}(a+1)} \\ &= \frac{1}{s}\sum_{k=0}^{m}c_{k,m}\frac{1}{a^k},\end{aligned}$$

where the coeficients $c_{k,m}$ are integers given by:

$$c_{k,m}=\sum_{j=k}^{m}(-1)^j j!\left\{\begin{matrix} m+1 \\ j+1 \end{matrix}\right\}\binom{j}{k}.$$

Using the following identity of Stirling numbers [101],

$$\sum_{j=0}^{m}(-1)^j j!\left\{\begin{matrix} m+1 \\ j+1 \end{matrix}\right\}=0,$$

we find that $c_{0,m}=0$ for m a positive integer. So we can write the Laplace transform of the moments as

$$\mathcal{L}\left\{M_{\beta_0}^m(\cdot)\right\}(s)=\sum_{k=1}^{m}c_{k,m}\frac{\left(\lambda e^{-\epsilon\lambda}\right)^k}{s^{k+1}e^{ks\epsilon}} \tag{7.12}$$

and apply the inverse of the Laplace transform in both size of Eq. (7.12) to obtain:

$$\begin{aligned}M_{\beta_0}^m(L) &= \mathcal{L}^{-1}\left\{\sum_{k=1}^{m}c_{k,m}\frac{\left(\lambda e^{-\epsilon\lambda}\right)^k}{s^{k+1}e^{ks\epsilon}}\right\}(L) \\ &= \sum_{k=1}^{m}c_{k,m}\left(\lambda e^{-\epsilon\lambda}\right)^k\mathcal{L}^{-1}\left\{\frac{1}{s^{k+1}e^{ks\epsilon}}\right\}(L) \\ &= \sum_{k=1}^{m}\frac{c_{k,m}}{k!}(L-k\epsilon)^k\left(\lambda e^{-\epsilon\lambda}\right)^k\mathbf{1}_{[L>k\epsilon]} \qquad (7.13)\end{aligned}$$

According to Lemma 7.6, when $\epsilon\to 0$, we obtain

$$M_{\beta_0}^m(L)=\mathbf{E}\left[N_L^m\right]=\sum_{k=1}^{m}\frac{c_{k,m}}{k!}(L\lambda)^k\mathbf{1}_{[L>0]}.$$

Hence, for any $\lambda>0$,

$$\sum_{k=1}^{m}\frac{c_{k,m}}{k!}(L\lambda)^k\mathbf{1}_{[L>0]}=\sum_{k=1}^{m}\left\{\begin{matrix} m \\ k \end{matrix}\right\}(L\lambda)^k\mathbf{1}_{[L>0]},$$

which shows that

$$c_{k,m}=\left\{\begin{matrix} m \\ k \end{matrix}\right\}k!\ .$$

Thus, we have proved (7.11) for any positive integer m. □

Once we have an expression of the moments, we can find the Laplace transform of the distribution of β_0.

Theorem 7.9 *The expression of* $\mathbf{P}(\beta_0 = n)$ *with respect of* n, L, λ *and* ϵ *is given by:*

$$\mathbf{P}(\beta_0 = n) = \frac{1}{n!} \sum_{i=0}^{\lfloor L/\epsilon \rfloor - n} \frac{(-1)^i}{i!} ((L - (n+i)\epsilon)\lambda e^{-\lambda\epsilon})^{n+i}. \tag{7.14}$$

Proof: With the expression of the moments of β_0 found in Corollary 7.8, we can find the Laplace transform of its distribution, given by

$$\mathbf{E}\left[e^{-s\beta_0}\right] = 1 - s\mathbf{E}\left[\beta_0\right] + \frac{s^2}{2!}\mathbf{E}\left[\beta_0^2\right] - \frac{s^3}{3!}\mathbf{E}\left[\beta_0^3\right] + ...$$

Rearranging the terms of the right-side hand and substituting $M_{\beta_0}^m(L)$ by the result of Eq. (7.11), we obtain:

$$\mathbf{E}\left[e^{-s\beta_0}\right] = \sum_{k=0}^{\infty} \left((L - k\epsilon)^k \left(\lambda e^{-\lambda\epsilon}\right)^k \mathbf{1}_{[L>k\epsilon]} \sum_{j=k}^{\infty} \frac{(-s)^j}{j!} \left\{ {j \atop k} \right\} \right).$$

But we can simplify this expression, since [101] gives

$$\sum_{j=k}^{\infty} \frac{x^j}{j!} \left\{ {j \atop k} \right\} = \frac{1}{k!}(e^x - 1)^k, \tag{7.15}$$

and we have

$$\mathbf{E}\left[e^{-s\beta_0}\right] = \sum_{k=0}^{\infty} (L - k\epsilon)^k \left(\lambda e^{-\lambda\epsilon}\right)^k \mathbf{1}_{[L>k\epsilon]} \frac{(e^{-s} - 1)^k}{k!}.$$

We can then apply the inverse of Laplace to find the distribution of β_0:

$$\sum_{k=0}^{\infty} \sum_{i=k}^{\infty} \frac{(-1)^i}{i!} \binom{i}{n} \delta(k-n)(k\epsilon - L)^k \left(\lambda e^{-\lambda\epsilon}\right)^k \mathbf{1}_{[L>k\epsilon]}.$$

After some simple algebra, we find the expression of the probability of an interval containing n complete clusters:

$$\mathbf{P}(\beta_0 = n) = \frac{1}{n!} \sum_{i=0}^{\lfloor L/\epsilon \rfloor - n} \frac{(-1)^i}{i!} ([L - (n+i)\epsilon]\lambda e^{-\lambda\epsilon})^{n+i},$$

concluding the proof. □

With the explicit expression of $\mathbf{P}(\beta(x) = n) = p_n(x)$, we can show a simple lemma.

Lemma 7.10 *For* $x \geq 0$, $p_n(x)$ *has the three following properties:*

i) $p_n(x)$ *is derivable;*

ii) $\lim_{x\to\infty} p_n(x) = 0$*;*

iii) $\lim_{x\to\infty} \frac{dp_n(x)}{dx} = 0$.

Proof: Let j be a non-negative integer. The function is obviously derivable when $x/\epsilon \neq j$. Besides, we have

$$\lim_{x\to\epsilon j^+} p_n(x) - \lim_{x\to\epsilon j^-} p_n(x) = \lim_{x\to\epsilon j^+} \frac{(-1)^j}{j!}\left((x-(n+j)\epsilon)\frac{1}{a}\right)^{n+j}.$$

Since the right-hand term function of x is zero as well as its derivative for all j, the function is also derivable when $x/\epsilon = j$, which proves i). Using the Final Value theorem in the Laplace transform of $p_n(x)$ and its derivative, we show items ii) and iii). □

The expression of $p_n(x)$ gives us, indeed, a transform Laplace pair between the x and s domains:

$$\frac{\mathbf{1}_{[x\geq 0]}}{n!}\sum_{i=0}^{\lfloor x/\epsilon\rfloor - n}\frac{(-1)^i}{i!}\left((x-(n+i)\epsilon)\frac{1}{a}\right)^{n+i} \overset{\mathcal{L}}{\Leftrightarrow} \frac{ae^{\epsilon s}}{(ase^{\epsilon s}+1)^{n+1}}. \tag{7.16}$$

We can use this relation to find the distributions of B_i and U_n.

Theorem 7.11 *The distributions of B_i and U_n, respectively $f_{B_i}(x)$ and $f_{U_n}(x)$ are*

$$f_{B_i}(x) = \left[\lambda e^{-\epsilon\lambda}p_0(x-\epsilon) + e^{-\epsilon\lambda}\frac{d}{dx}p_0(x-\epsilon)\right]\mathbf{1}_{[x>\epsilon]}, \tag{7.17}$$

and

$$f_{U_n}(x) = \lambda e^{-\epsilon\lambda}p_{n-1}(x-\epsilon)\mathbf{1}_{[x>\epsilon]}, \tag{7.18}$$

where the expressions of $p_0(x-\epsilon)$ and $\frac{d}{dx}p_0(x-\epsilon)$ are straightforwardly obtained from Eq. (7.14).

Proof: According to Corollary 7.3:

$$\begin{aligned}
\mathbf{E}\left[e^{-sB_i}\right] &= \frac{1}{\lambda}\frac{(\lambda+s)}{\frac{e^{\lambda\epsilon}}{\lambda}se^{s\epsilon}+1} \\
&= \lambda e^{-\epsilon\lambda}\frac{e^{\epsilon\lambda}}{\lambda}\frac{e^{\epsilon s}}{\frac{e^{\epsilon\lambda}}{\lambda}se^{\epsilon s}+1}e^{-\epsilon s} + e^{-\epsilon\lambda}s\frac{e^{\epsilon\lambda}}{\lambda}\frac{e^{\epsilon s}}{\frac{e^{\epsilon\lambda}}{\lambda}se^{\epsilon s}+1}e^{-\epsilon s} \\
&= \lambda e^{-\epsilon\lambda}e^{-\epsilon s}\mathcal{L}\left\{p_0(\cdot)\right\}(s) + e^{-\epsilon\lambda}e^{-\epsilon s}s\mathcal{L}\left\{p_0(\cdot)\right\}(s).
\end{aligned}$$

Here, using the inverse Laplace transform established in Eq. (7.16) and remembering that $p_0(x^-) = 0$, we get an analytical expression for $f_{B_i}(x)$, proving Eq. (7.17).

Proceeding in a similar fashion, we can find the distribution of U_n by inverting its Laplace transform given by Corollary 7.4:

$$\begin{aligned}
\mathbf{E}\left[e^{-sU_n}\right] &= \frac{1}{\left(\frac{e^{\lambda\epsilon}}{\lambda}se^{s\epsilon}+1\right)^n} \\
&= \lambda e^{-\epsilon\lambda}\frac{e^{\epsilon\lambda}}{\lambda}\frac{e^{\epsilon s}}{\left(\frac{e^{\epsilon\lambda}}{\lambda}se^{\epsilon s}+1\right)^n}e^{-\epsilon s} \\
&= \lambda e^{-\epsilon\lambda}e^{-\epsilon s}\mathcal{L}\left\{p_{n-1}(\cdot)\right\}(s).
\end{aligned}$$

Inverting this Laplace transform we prove Eq. (7.18). □

We can also obtain the probability that the segment $[0, L]$ is completely covered by the sensors. To do this, we remember that the first point (if there is one) is capable to cover the interval $[X_1 - \epsilon, X_1 + \epsilon]$. This motivates the Theorem 7.12.

Theorem 7.12 *Let $R_{m,n}(x)$ be defined as follows:*

$$R_{m,n}(x) = \sum_{i=m}^{\lfloor x/\epsilon \rfloor - 1} \left[\left(e^{-\lambda\epsilon} \right)^{i+n} \sum_{j=0}^{i+n} \frac{(\lambda[(1-i)\epsilon - x])^j}{j!} \right].$$

Then,

$$\mathbf{P}([0, L] \text{ is covered}) = R_{0,1}(L) - e^{-\lambda\epsilon} R_{0,1}(L - \epsilon) - e^{-\lambda\epsilon} R_{1,0}(L) + e^{-2\lambda\epsilon} R_{1,0}(L - \epsilon). \quad (7.19)$$

Proof: The condition of total coverage is the same as

$$\{\forall x \in [0, L], \exists X_i \in [0, L] | x \in [X_1 - \epsilon, X_1 + \epsilon]\},$$

which means that:

$$\{[0, L] \text{ is covered}\} \Leftrightarrow \{B_1 \geq L - X_1\} \cap \{X_1 \leq \epsilon\}.$$

Hence,

$$\mathbf{P}([0, L] \text{ is covered}) = \int_0^\epsilon \mathbf{P}(B_1 \geq L - X_1 | X_1 = x) dP_{X_1}(x),$$

and since B_1 and X_1 are independent:

$$\mathbf{P}([0, L] \text{ is covered}) = \int_0^\epsilon \int_{L-x}^\infty f_{B_1}(u) \lambda e^{-x\lambda} du dx.$$

From this point we use Lemma 7.10 to solve some analytical integrals and to do some algebra to obtain Eq. (7.19). □

Hence, we have explicit expressions, which are quite simple, to represent the distributions of β_0, B_i, U_n and the probability of total coverage as a function of L, ϵ, λ_0, μ and t (we remember that $\lambda = \lambda_0 L e^{-t/\mu}$), so the problem of a random sensor network in one dimension is completely solved.

7.4 Other Scenarios

Although we consider the problem of finding the number of complete clusters until this point of the paper, the method can be used to calculate p_n to other definitions for the number of clusters. In this We consider particularly two other definitions: the number of incomplete clusters and the number of clusters in a circle.

7.4.1 Number of incomplete clusters

The major difference with Sec. 7.3 is that a cluster is counted as soon as one of the point of the cluster is inside the interval $[0, L]$. So, for instance, in Fig. 7.3, we count actually $n+1$ incomplete clusters. We define β_0' as the number of incomplete clusters on an interval $[0, L]$.

Theorem 7.13 *Let $G(k)$ be defined as*

$$G(k) = (-1)^k \left(e^{-k\lambda\epsilon} \sum_{j=0}^{k} \frac{[\lambda(k\epsilon - L)]^j}{j!} - e^{-\lambda L} \right) \mathbf{1}_{[T>k\epsilon]}$$

for $k \in \mathbb{N}_+$ and $G(-1) = e^{-\lambda L}$. Then

$$\mathbf{P}(\beta_0' = n) = \sum_{i=n}^{\lfloor L/\epsilon \rfloor + 1} (-1)^{i+n} \binom{i}{n} (G(i-1) + G(i)), \text{ for } n \geq 0. \tag{7.20}$$

Proof: The condition of $\beta_0' \geq n$ is slightly changed:

$$\{\beta_0' \geq n\} \Leftrightarrow \begin{cases} \{\Delta X_0 + U_{n-1} \leq L\} & \text{if } n \geq 1, \\ \{\Delta X_0 < \infty\} & \text{if } n = 0. \end{cases}$$

We define Y_n as

$$Y_n \triangleq \begin{cases} \Delta X_0 + U_{n-1} & \text{if } n \geq 1 \\ 0 & \text{if } n = 0. \end{cases}$$

Repeating the same calculations, we find the Laplace transform of $\mathbf{P}(\beta_0'(L) = n)$:

$$\mathcal{L}\{\mathbf{P}(\beta_0'(\cdot) = n)\}(s) = \begin{cases} \dfrac{\lambda}{s+\lambda} \dfrac{e^{\epsilon\lambda}}{\lambda} \dfrac{e^{\epsilon s}}{\left(\frac{e^{\epsilon\lambda}}{\lambda} s e^{\epsilon s} + 1\right)^n} & \text{if } n \geq 1, \\ \dfrac{1}{\lambda + s} & \text{if } n = 0. \end{cases} \tag{7.21}$$

With this expression, following the lines of Lemma 7.6, we obtain:

$$\mathcal{L}\left\{\mathbf{E}\left[\beta_0'(\cdot)^m\right]\right\}(s) = \sum_{k=1}^{m+1} \left\{ {m+1 \atop k} \right\} (k-1)! \frac{1}{s^k} \frac{\lambda}{\lambda + s} \left(\frac{\lambda e^{-\lambda\epsilon}}{e^{s\epsilon}} \right)^{k-1}.$$

Then, we write:

$$\frac{\lambda}{\lambda + s} \frac{1}{s^k} = \frac{(-1)^k}{\lambda^{k-1}} \frac{1}{\lambda + s} + \sum_{i=1}^{k} \frac{1}{s^i} \left(\frac{-1}{\lambda} \right)^{k-i},$$

to find an expression with a well known Laplace transform inverse, and after inverting it, we obtain:

$$\mathbf{E}\left[\beta_0'^m\right] = \sum_{k=0}^{m} \left\{ {m+1 \atop k+1} \right\} k! G(k). \tag{7.22}$$

Expanding the Laplace transform of the distribution of β_0' in a Taylor serie and rearranging terms, we get

$$\mathbf{E}\left[e^{-s\beta_0'}\right] = 1 + G(0)\sum_{j=1}^{\infty}\frac{(-s)^j}{j!}\left\{\begin{matrix} j \\ 1 \end{matrix}\right\} + \left(\sum_{k=1}^{\infty}G(k)\sum_{j=k}^{\infty}\frac{(-s)^j}{j!}\left\{\begin{matrix} j+1 \\ k+1 \end{matrix}\right\}\right).$$

Now, we use the recurrence that Stirling numbers obey [101],

$$\left\{\begin{matrix} j+1 \\ k+1 \end{matrix}\right\} = \left\{\begin{matrix} j \\ k \end{matrix}\right\} + (k+1)\left\{\begin{matrix} j \\ k+1 \end{matrix}\right\}, \tag{7.23}$$

to apply in Eq. (7.15):

$$\begin{aligned}\sum_{j=k}^{\infty}\frac{x^j}{j!}\left\{\begin{matrix} j+1 \\ k+1 \end{matrix}\right\} &= \sum_{j=k}^{\infty}\frac{x^j}{j!}\left(\left\{\begin{matrix} j \\ k \end{matrix}\right\} + (k+1)\left\{\begin{matrix} j \\ k+1 \end{matrix}\right\}\right) \\ &= \frac{1}{k!}(e^x-1)^k + \frac{1}{k!}(e^x-1)^{k+1},\end{aligned}$$

obtaining:

$$\mathbf{E}\left[e^{-s\beta_0'}\right] = 1 + \sum_{k=1}^{\infty}(G(k-1)+G(k))(e^{-s}-1)^k.$$

Inverting this expression for an non-negative integer n, we have the searched distribution. □

7.4.2 Number of clusters in a circle

We investigate now the case where the points of the process are deployed over a circumference and we want to count the number of complete clusters, which corresponds to calculate the Euler Characteristic of the total coverage, so we call this quantity χ. Due to the symmetry of the circumference , we can choose an arbitrary point to be the origin. If it is given that there is at least one point on it, we can choose the origin to be at some point of the process.

Theorem 7.14 *The distribution of the Euler Characteristic, χ, when the points are deployed over a cirumference of length L is given by*

$$\begin{aligned}\mathbf{P}(\chi = n) = e^{-\lambda L}\mathbf{1}_{[n=0]} + (1-e^{-\lambda L})\frac{\lambda e^{-\epsilon\lambda}}{n!}\sum_{i=0}^{\lfloor L/\epsilon\rfloor - n}\Bigg[\frac{(-1)^i}{i!} \\ ([L-(n+i)\epsilon]\lambda e^{-\epsilon\lambda})^{n+i-1}\left(L+(n+i)\left(\frac{1}{\lambda}-\epsilon\right)\right)\Bigg],\end{aligned} \tag{7.24}$$

for $n \geq 0$.

Proof: It is possible to establish a relation between the case in the line and the case in the circle. If there are no points in the circle, of course $\chi = 0$. Otherwise, if there is at least one point, we choose the origin at this point and we have the equivalence between the events:

$$\{\chi \geq n\} \Leftrightarrow \begin{cases} \{U_{n-1} + B_n \leq L\} \cap \{N_L > 0\} & \text{if } n \geq 1, \\ \{\Delta X_0 < \infty\} & \text{if } n = 0. \end{cases}$$

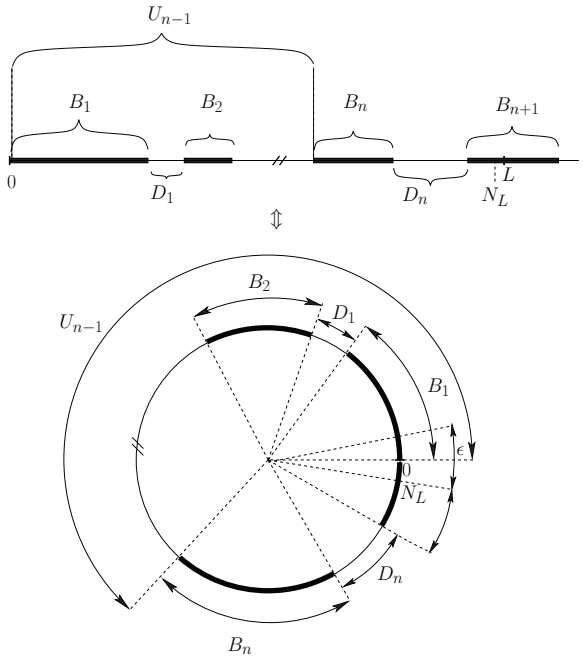


Figure 7.4: Illustration of the condition equivalent to $\chi \geq n$. Since the coverage of the last point on $[0, L]$ overlaps the cluster with a point in zero, they are actually contained in the same cluster

In Fig. 7.4 we present an example of this equivalence. We can define Y_n as

$$Y_n \triangleq \begin{cases} U_{n-1} + B_n & \text{if } n \geq 1 \\ 0 & \text{if } n = 0, \end{cases}$$

to find the Laplace transform or $\mathbf{P}(\chi(L) = n)$:

$$\mathcal{L}\{\mathbf{P}(\chi(\cdot) = n)\}(s) = (1 - e^{-\lambda L})\frac{\lambda + s}{\lambda}\frac{e^{\epsilon\lambda}}{\lambda}\frac{e^{\epsilon s}}{\left(\frac{e^{\epsilon\lambda}}{\lambda}se^{\epsilon s} + 1\right)^n}. \tag{7.25}$$

Again, we can use the fact that the number of clusters is almost surely equal to the number of points when $\epsilon \to 0$ to obtain the moments of χ:

$$\mathbf{E}\left[\chi^m\right] = (1 - e^{-\lambda L})\lambda e^{-\epsilon\lambda}\sum_{k=1}^{m}\left[\left\{ {m \atop k} \right\}([L - k\epsilon]\lambda e^{-\epsilon\lambda})^{k-1} \left(L + k\left(\frac{1}{\lambda} - \epsilon\right)\right)\mathbf{1}_{[L>k\epsilon]}\right].$$

Expanding the Laplace transform in a Taylor serie and rearranging terms, as we did previously, yields

$$\mathbf{E}\left[e^{-s\chi}\right] = (1 - e^{-\lambda L})\lambda e^{-\epsilon\lambda}\sum_{k=0}^{\infty}\left[\left([L - k\epsilon]\lambda e^{-\epsilon\lambda}\right)^{k-1}\left(L + k\left(\frac{1}{\lambda} - \epsilon\right)\right)\mathbf{1}_{[L>k\epsilon]}\sum_{j=k}^{\infty}\frac{(-s)^j}{j!}\left\{ {j \atop k} \right\}\right].$$

Since

$$\sum_{j=k}^{\infty} \frac{(-s)^j}{j!} \left\{ {j \atop k} \right\} = \frac{(e^{-s}-1)^k}{k!},$$

we can directly invert this Laplace transform, add the case where there are no points for $\chi = 0$, and the theorem is proved. □

7.5 Examples

We consider some examples to illustrate the results of the paper. Here, the behavior of the mean and the variance of β_0 as well as $Pr(\beta_0 = n)$ are presented.

From Eq. (7.11), we have that $\mathbf{E}[\beta_0]$ is given by:

$$\mathbf{E}[\beta_0] = (L-\epsilon)\lambda e^{-\epsilon\lambda}\mathbf{1}_{[L>\epsilon]}.$$

This expression agrees with the intuition that there are three typical regions given a fixed ϵ. When λ is much smaller than $1/\epsilon$, the number of clusters is approximatively the number of sensors, since the connections with few sensors will unlikely happen, which can be seen from the fact that $\overline{\beta_0} \to L\lambda$ when $\lambda \to 0$. As we increase λ, the mean number of direct connections overcomes the mean number of sensors and, at some value of λ, we expect that $\overline{\beta_0}$ decreases, when adding a point is likely to connect disconnected clusters. We remark that the maximum occurs exactely for $\epsilon = 1/\lambda$, i.e., when the mean distance between two sensors equals the threshold distance for them to be connected. In this maximum, $\overline{\beta_0}$ takes the value of $(L/\epsilon - 1)e^{-1}$. Finally, when λ is too large, all sensors tend to be connected and there is only one cluster which even goes boyond L, so there are no complete clusters into the interval $[0, L]$. This is trivial when we make $\lambda \to \infty$ in the last equation. Figure 7.5 shows this behavior when $L = 4$ and $\epsilon = 1$.

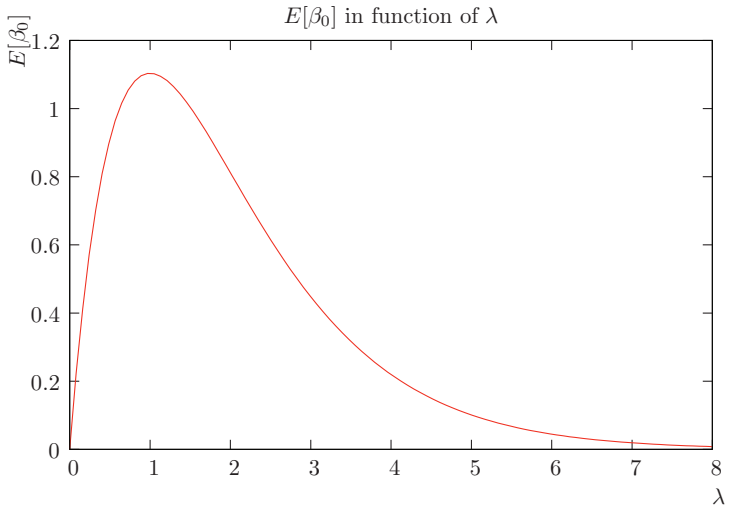


Figure 7.5: Variation of the mean number of clusters as a function of λ when $L = 4$ and $\epsilon = 1$.

The variance can be obtained also by Eq. (7.11):

$$\mathrm{Var}(\beta_0) = (L-\epsilon)\lambda e^{-\epsilon\lambda}\mathbf{1}_{[L>\epsilon]} + (L-2\epsilon)\lambda^2 e^{-2\epsilon\lambda}\mathbf{1}_{[L>2\epsilon]} - (L-\epsilon)^2\lambda^2 e^{-2\epsilon\lambda}\mathbf{1}_{[L>\epsilon]},$$

and under the condition that $L > 2\epsilon$:

$$\mathrm{Var}(\beta_0) = (L-\epsilon)\lambda e^{-\epsilon\lambda} + \epsilon(3\epsilon - 2L)\lambda^2 e^{-2\epsilon\lambda}.$$

Fig. 7.6 shows a plot of $\mathrm{Var}(\beta_0)$ as function of λ for $L = 4$ and $\epsilon = 1$. We can expect that,

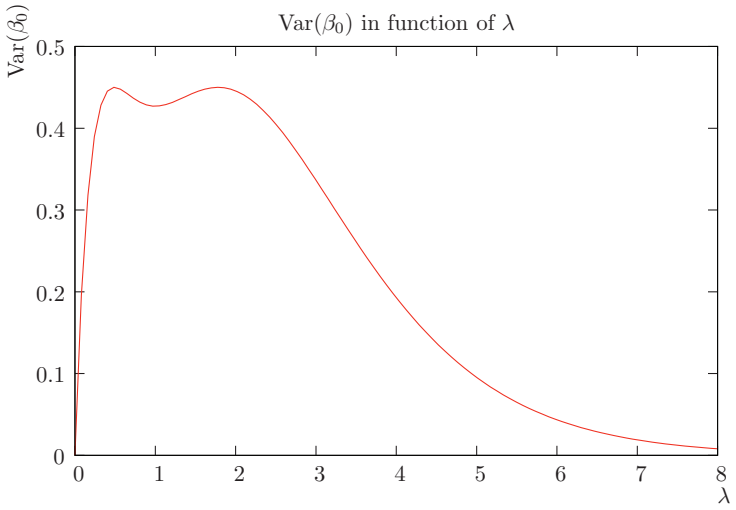


Figure 7.6: Behavior of the variance of the number of clusters as a function of λ when $L = 4$ and $\epsilon = 1$.

when λ is small compared to ϵ, the plot should be approximatively linear, since there would not be too much connections in the network and the variance of the number of clusters should be close to the variance of the number of sensors given by λL. Since β_0 tends almost surely to 0 when λ goes to infinity, $\mathrm{Var}(\beta_0)$ should also tend to 0 in this case. Those two properties are observed in the plot. Besides, we find the critical points of this function, and again, $\lambda = 1/\epsilon$ is one of them and at this value $\mathrm{Var}(\beta_0) = (L/\epsilon)e^{-1} + (3 - 2L/\epsilon)e^{-1}$. The other two are the ones satisfiy the transcendant equation:

$$\lambda e^{-\lambda\epsilon} = \frac{L-\epsilon}{2\epsilon(2L-3\epsilon)}.$$

By using the second derivative, we realize that $1/\epsilon$ is actually a minimum. Besides, if $L \leq 2\epsilon$, there is just one critical point, a maximum, at $\lambda = 1/\epsilon$.

The last example in the section is performed with the result obtained in Theorem 7.9. We consider again $L = 4$ and $\epsilon = 1$ to obtain the following distributions:

$$\begin{aligned}
\mathbf{P}(\chi = 0) &= 1 - 3\lambda e^{-\lambda} + 2\lambda^2 e^{-2\lambda} - 1/6\lambda^3 e^{-3\lambda},\\
\mathbf{P}(\chi = 1) &= 3\lambda e^{-\lambda} - 4\lambda^2 e^{-2\lambda} + 1/2\lambda^3 e^{-3\lambda},\\
\mathbf{P}(\chi = 2) &= 2\lambda^2 e^{-2\lambda} - 1/2\lambda^3 e^{-3\lambda},\\
\mathbf{P}(\chi = 3) &= 1/6\lambda^3 e^{-3\lambda},\\
\mathbf{P}(\chi > 3) &= 0.
\end{aligned}$$

Those expressions are simple and they have at most four terms, since $L = 4\epsilon$. We plot these functions in Fig. 7.7. The critical points on those plots at $\lambda = 1/\epsilon$ are confirmed for the fact that, as a function of λ, for every n, $\mathbf{P}(\chi = n)$ can be represented as a sum

$$\sum_{i=0}^{j} q_{i,j}(\lambda e^{-\lambda\epsilon})^i$$

where the coeficients $q_{i,j}$ are constant in relation to λ. However, $(\lambda e^{-\lambda\epsilon})^i$ has a critical point at $\lambda = 1/\epsilon$ for all $i > 0$, so this should be also a critical point of $\mathbf{P}(\chi = n)$. If λ is small, we should expect that $\mathbf{P}(\chi = 0)$ is close to one, since it is likely to N have no points. For this reason, in this region, $\mathbf{P}(\chi = n)$ for $n > 0$ is small. When λ is large, we expect to have very large clusters, likely to be larger than L, so it is unlikely to have a complete cluster in the interval and, again, $\mathbf{P}(\chi = 0)$ approaches to the unity, while $\mathbf{P}(\chi = n)$ for $n > 0$ become again small.

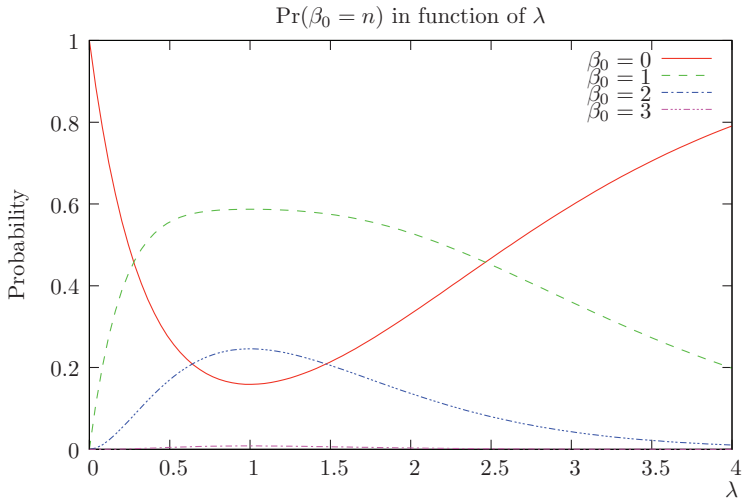


Figure 7.7: Probabilities of connectivity, $\mathbf{P}(\beta_0 = n)$, for $n = 0, 1, 2, 3$, as a function of λ when $L = 4$ and $\epsilon = 1$.

7.6 Summary

In this chapter, we have obtained expressions that model a random sensor network in a line. We have evaluated both the analytical expression and its Laplace transform of the following quantities: the length of a cluster, the length of a sequence of consecutive clusters, the distribution of the number of complete clusters, incomplete clusters and clusters on a circumference, all the moments of those distributions and the probability of complete coverage. The analytical solutions are simple since they involve only polynomial and exponential functions; they are exact; and they take into account a large number of realistic variables such as density of the network, lifetime of the sensors, power transmission and reception sensitivity, all of this with a random deployment of sensors. Moreover, these calculations solve also the queueing problem of the busy period of an $M/D/\infty$ and find many parameters of a $M/G/1/1$ when the service time is distributed as c.

Chapter 8

Concluding Discussion and Future Works

In this thesis, we explored tools of stochastic analysis and algebraic topology applied to sensor networks. We considered the usual modeling of the positions in a network as a Poisson point process. Furthermore, we used stochastic analysis under the Poisson measures to characterize random variables as the sum of stochastic integrals and we applied the gradient operator D, as well as the Ornstein-Uhlenbeck generator L and its inverse. As regards algebraic topology, we took advantage of the fact that the topology of a sensor network coverage is the same as that of its simplicial complex.

Stochastic analysis enabled us to find upper bound probabilities for various situations, and we presented two of them throughout the thesis: the probability of opportunist users in a cognitive system to damage licensed users and the probability for losing a user in an OFDMA system because all subcarriers of the central station are already in use. In both cases, the results can be used to design a system, since they are function of parameters that can be found or controlled by the operator.

In the first case we consider the power constraint to be respected in order to avoid outages of the licensed users. Then we found the maximum power allowed to opportunist users in a WiMax network such that this the mean interference caused by them respects this constraint. Then, we found out that the power allowed to these secondary users in order to have low outage probabilities for primary ones is not much weaker than the averaged one. So, for instance, the operator can control capacity of secondary users according to the outage probability, or the opposite, the intensity of users can be controlled in a region to ensure that users (primary and secondary ones) have a minimum QoS in their communication. It is worth mentioning that the secondary users capacity is improved and damage to primary ones is lessened when their intensity and bandwidth are increased, which is a tendency of the next generation wireless networks.

In the second application, we found a relationship between the probability of overloading the system, the density of active users and the number of available subchannels, thus providing a large number of possibilities to design OFDMA systems. We also compare the numerical results with simulations and note that the calculated number of subchannels leads to an overload probability overestimated by about 20% of the simulated one. The margin provided by the bounds may be viewed as a protection against errors in the modeling or in the estimations of parameters. We should remark that, in both cases, we found new analytical results for the very complex mathematical problems.

The most important results from algebraic topology came from the interpretation of a sensor network as a simplicial complex, which enables us to compute the Betti numbers of the coverage of a sensor network. We showed that Betti numbers do not allow us to determine if a region is completely covered, but we solved this problem if points lay on a torus. In this case, the Euler characteristic also can provide valuable information on coverage: if this quantity equals zero, probably the network is covered. We obtained statistics on the number of k-simplices, such as its mean, variance, covariance and third moment, and we provided a method to calculate the n-th moment. The results already show some tendencies for the network, for instance if a network has a number of 1-simplices much larger than the number of points, then probably this network is connected. We can think about similar intuitive interpretations, but these statistics lead to more interesting results: we can use them to determinate the mean and variance of the Euler characteristic and to find bounds for the distribution of connectivity. We also proved that the distribution of the number of k-simplices converges to a Gaussian distribution when the intensity of points λ tends to infinity, with a convergence rate of $O(1/\sqrt{\lambda})$. If we combine this convergence with the statistics of first and second order, we have a good approximation for the distribution of the number of k-simplices. With this result, a random Gaussian vector can represent the joint distribution of the number of different simplices, since we have the correlation between the number of any k-simplices and l-simplices.

In Chapter 7 we did not use a particular new method, but classical ones to solve the $M/M/1/1$ preemptive queue which was not solved in the literature. This result corresponds to the solution of the major problem of the thesis in one dimension: the arrival time corresponds to the beginning of a cluster of sensors, the busy period corresponds to the size of a cluster and the number of served users after a certain time corresponds to the number of clusters in a line segment. We found all relevant parameters and we believe that this example can give insights to cases in two or more dimensions.

The most important contribution of the thesis is to apply simultaneously results from algebraic topology and from stochastic analysis on sensor networks. Although tools of Malliavin calculus were more often used, topology has played a fundamental role since the random variables and their relations came from concepts of topology. One of the main purposes of this thesis was the use of modern mathematical tools in networks, which was fulfilled. By using these tools, we could obtain results for problems with very hard analytical treatment where classical tools have failed to solve. We remark that it was also important to present concrete and applicable results in each chapter.

8.1 Future Works

Although the use of topology has been essential to the work, we used but a superficial layer of it. There are a lot of works in this domain and we believe that they soon will be used in random networks (see, for instance [27, 86, 84]), especially the works of persistent homology. We remember that we were not able to find relevant results for the Betti numbers: we know how to compute them, but not how to treat them statistically. A particularly intriguing question would be to find a closed-form expression for the Euler characteristic (instead of the infinite power series representing it) in two dimensions, at least. With respect to the deployment of the points, there are various different point processes that could be used to represent the behavior of users, ranging from the randomness of the density of user (as in a Cox process) to processes where the position of each user depends on the position of other users (as in determinantal processes).

List of Figures

Bibliography

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networs: a survey. *Computer Networks*, 4(38):393–422, 2002.

[2] N. Alon, J. H. Spencer, and P. Erdös. *The Probabilistic Method*. Wiley-Interscience, New York, 1992.

[3] M. J. B. Appel and R. P. Russo. The connectivity of a graph on uniform points in $[0,1]^d$. *Statistics and Probability Letters*, 60:351–357, 1997.

[4] M. J. B. Appel and R. P. Russo. The maximum vertex degree of a graph on uniform points in $[0,1]^d$. *Advances in Applied Probability*, 29:567–581, 1997.

[5] M. J. B. Appel and R. P. Russo. The minimum vertex degree of a graph on uniform points in $[0,1]^d$. *Advances in Applied Probability*, 29:582–594, 1997.

[6] M. A. Armstrong. *Basic Topology*. Springer-Verlag, 1997.

[7] S. Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics, Stochastic Modeling and Applied Probability*. Springer-Verlag, 2003.

[8] S. Athreya, R. Roy, and A. Sarkar. On the coverage space by random sets. *Advances in Applied Probability*, 36:1–18, 2004.

[9] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev. Stochastic geometry and architecture of communication networks. *J. Telecommunication Systems*, 7:209–227, 1997.

[10] F. Baccelli and S. Zuyev. Stochastic geometry models of mobile communication networks. In *Frontiers in queueing*, pages 227–243. CRC Press, Boca Raton, FL, 1997.

[11] P. Balister, B. Bollobás, A. Sarkar, and M. Walters. Connectivity of random k-nearest-neighbour graphs. *Advances in Applied Probability*, 37(1):1–24, march 2005.

[12] Y. Baryshnikov and R. Ghrist. Target enumeration via euler characteristic integrals i: sensor fields, 2007.

[13] E. T. Bell. Exponential polynomials. *Ann, Math.*, 35:258–277, 1934.

[14] A. Björner. Topological methods. *Handbook of combinatorics (vol. 2)*, pages 1819–1872, 1995.

[15] B. Bollobás. *Random Graphs*. Academic Press, London, 1985.

[16] S. R. Broadbent and J. M. Hammersley. Percolation processes. i. crystals and mazes. In *Proceedings of the Cambridge Philosophical Society*, pages 52:629–641, 1957.

[17] E. H. Callaway. *Wireless sensor networks: architectures and protocols*. Auerbach Publications, Florida, USA, 1st edition, 2003.

[18] C.-Y. Chong and S. P. Kumar. Sensor networks: Evolution, opportunities, and challenges. In *Proceedings of IEEE*, volume 91, pages 1247–1256, august 2003.

[19] R. B. Cooper. *Introduction to queueing theory*. North-Holland Publishing Co., second edition, 1981.

[20] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Springer-Verlag, 2003.

[21] Danijela Cabric, Ian D. O'Donnell, Mike Shuo-Wei Chen and Robert W. Brodersen. Spectrum sharing radios. In *IEEE Circuits and Systems Magazine*, 2006.

[22] V. De, Silva, and R. Ghrist. Coverage in sensor networks via persistent homology.

[23] V. de Silva and R. Ghrist. Coordinate-free coverage in sensor networks with controlled boundaries via homology. *International Journal of Robotics Research*, 25, december 2006.

[24] C. Domb. The problem of random intervals on a line. *Proceedings of the Cambridge Philosophical Society*, 43:329–341, 1947.

[25] A. Dvoretzki. On covering a circle by randomly placed arcs. *Proceedings of the National Academy of Sciences U.S.A.*, 42:199–203, 1956.

[26] H. Edelsbrunner and J. Harer. Persistent homology — a survey. *Surveys on Discrete and Computational Geometry. Twenty Years Later*, 453:257–282, 2008.

[27] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 454–, Washington, DC, USA, 2000. IEEE Computer Society.

[28] P. Erdös and A. Rényi. On random graphs, i. In *Publicationes Mathematicae (Debrecen)*, volume 6, pages 290–297, 1959.

[29] P. Erdös and A. Rényi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of sciences*, pages 17–61, 1960.

[30] D. Estrin, D. Culler, K. Pister, and G. Sukhatme. Connecting the physical world with persuasive networks. In *IEEE Persuasive Computing*, volume 1, pages 59–69, 2002.

[31] D. Estrin, L. Girod, G. Pottie, and M. Srivastava. Instrumenting the world with wireless sensor networks. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing, (ICASSP 2001)*, pages 2033–2036, Salt Lake City, Utah, 2001.

[32] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. Next century challenges: scalable coordination in sensor networks. In *Proceedings of the 5th annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'99)*, pages 263–270, 1999.

[33] FCC. Report of the spectrum efficiency wg. Technical report, Spectrum Policy Task Force (SPTF), November 2002.

[34] FCC. Establishment of interference temperature metric to quantify and manage interference and to expand available unlicensed operation in certain fixed mobile and satellite frequency bands. Technical report, Notice of inquiry and notice of proposed Rulemaking, November 2003.

[35] R. A. Fisher. On the similarity of the distributions found for the test of significance in harmonic analysis, and in stevens's problem in geometrical probability. *Annals of Eugenics*, 10:14–17, 1940.

[36] L. Flatto and A. G. Konheim. The random division of an interval and the random covering of a circle. *SIAM Review*, 4:211–222, 1962.

[37] E. Fleury and D. Simplot-Ryl. *Réseaux de capteurs*. Lavoisier, 2009.

[38] C. H. Foh and B. S. Lee. A closed form network connectivity formula for one-dimensional manets. In *2004 IEEE International Conference on Communications (ICC 2004), Paris, France*, june 2004.

[39] C. H. Foh, G. Liu, B. C. Seet, K. J. Wong, and C. P. Fu. Network connectivity of one-dimensional manets with random waypoint movement. *IEEE Communications Letters*, 9:31–33, 2005.

[40] A. Ghasemi and S. Nader-Esfahani. Exact probability of connectivity in one-dimensional ad hoc wireless networks. *IEEE Communications Letters*, 10:251–253, 2006.

[41] R. Ghrist and A. Muhammad. Coverage and hole-detection in sensor networks via homology. In *Fouth International Conference on Information Processing in Sensor Networks (IPSN'05), UCLA*, pages 254–260, 2005.

[42] R. Ghrist and A. Muhammad. Coverage and hole-detection in sensor networks via homology. In *Fourth International Conference on Information Processing in Sensor Networks (IPSN'05), UCLA*, pages 254–260, 2005.

[43] E. N. Gilbert. Random graphs. *Annals of Mathematical Statistics*, 30:1141–1144, 1959.

[44] E. N. Gilbert. Random plane networks. *Journal of the Society for Industrial Applied Mathematics*, 6:533–553, 1961.

[45] E. N. Gilbert. The probability of covering a sphere with n circular caps. *Biometrika*, 52:323–330, 1965.

[46] E. Godehardt. *Graphs as Structural Models*. Wieweg, Braunschweig, 1990.

[47] E. Godehardt and J. Jaworski. On the connectivity of a random interval graph. *Random Structures and Algorithms*, 9:137–161, 1996.

[48] E. Godehardt, J. Jaworski, and D. Godehardt. The application of random coincidence graphs for testing the homogeneity of data. In *Classification, Data Analysis and Data Highways: Proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation e.C., University of Potsdam*, pages 33–45. Springer, Berlin, 1998.

[49] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics: A Foudation for Computer Science*, chapter 6.1, pages 257–267. Reading, MA: Addison-Wesley, 1994.

[50] M. J. Greenberg and J. R. Harper. *Algebraic Topology: a First Course*. Addison-Wesley, New York, NY, 1981.

[51] G. Grimmett. *Percolation*. Springer, Berlin, 1999.

[52] R. Hafner. The asymptotic distribution of random clumps. *Computing*, 10:335–351, 1972.

[53] P. Hall. On the coverage of k-dimensional space by k-dimensional spheres. *The Annals of Probability*, 13:991–1002, 1985.

[54] P. Hall. *Introduction to the Theory of Coverage Processes*. Wiley, New York, 1988.

[55] G. Han. *Connectivity Analysis of Wireless Ad-Hoc Networks*. PhD thesis, Departement of Electrical and Computer Engineering, University of Maryland, College Park, april 2007.

[56] G. Han and A. M. Makowski. One-dimensional geometric random graphs with non-vanishing densities: a very strong zero-one law for connectivity. 2011.

[57] B. Harris and E. Godehardt. Probability models and limit theorems for random interval graphs with applications to cluster analysis. In *Classification, Data Analysis and Data Highways*, pages 54–61. Springer, Berlin, 1998.

[58] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[59] L. Holst. On multiple covering of a circle with random arcs. *Journal of Applied Probability*, 17:284–290, 1980.

[60] Y. Ito. Generalized poisson functionals. *Probab. Theory Related Fields*, 77:1–28, 1988.

[61] J. G. Proakis. *Digital communications*. McGraw Hill, 1995.

[62] S. Janson. Random coverings in several dimensions. *Acta Mathematica*, 156:83–118, 1986.

[63] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Wiley, New York, 2000.

[64] Juncheng Jia, Qian Zhang. Hardware-constrained multi-channel cognitive mac. In *IEEE Global Telecommunications Conference*, November 2007.

[65] M. Kahle. Random geometric complexes. Discrete and Computational Geometry, Jannuary 2011.

[66] M. Kahle and E. Meckes. Limit theorems for betti numbers of random simplicial complexes, 2010.

[67] J. Kahn, R. Katz, and K. Pister. Mobile networking for smart dust. In *Intl. Conf. on Mobile Computing and Networking*, Seattle, WA, august 1999.

[68] H. Karl and A. Willig. *Protocols abd architectures for wireless sensors networks*. John Wiley, 2nd edition, 2003.

[69] H. Koskinen. On the coverage of a random sensor network in a bounded domain. In *in Proceedings of 16th ITC Specialist Seminar*, pages 11–18, 2004.

[70] P. Kumar. New technological vistas for systems and control: the example of wireless networks. *IEEE Control Systems Magazine*, pages 24–37, feb 2001.

[71] N. V. D. Laurent Decreusefond, H. Korezlioglu. An error bound for infinite approximations of queueing networks with large finite stations. *Queueing Networks with Finite Capacity*, pages 239–252, 1993. R.O. Onvural and I.F. Alyildiz, editors.

[72] F. Lewis. *Wireless Sensor Networks*, chapter 2. John Wiley, New York, 2004.

[73] X. Li, D. K. Hunter, and S. Zuyev. Triangulation properties of the target area in wireless sensor networks. Unpublished, 2011.

[74] X.-Y. Li, P.-J. Wan, and O. Frieder. Coverage in wireless ad hoc sensor networks. *IEEE Transactions on Computers*, 52:753–763, 2003.

[75] M. Mandelbrot. On dvoretski coverings for the circle. *Z. Warscheinlichkeitstheorie*, 22:158–160, 1972.

[76] P. Manohar, S. S. Ram, and D. Manjunath. Path coverage by a sensor field: the nonhomogeneous case. *ACM Transactions on Sensor Networks*, 5(2), March 2009.

[77] C. McDiarmid. Random channel assignment in the plane. *Random Structures and Algorithms*, 22:187–212, 2003.

[78] C. McDiarmid and B. Reed. Colouring proximity graphs in the plane. *Discrete Mathematics*, 199:123–127, 1999.

[79] M. A. McHenry. *Spectrum Occupancy Measurements*. http://www.sharedspectrum.com/, August 2005.

[80] R. Meester and R. Roy. *Continuum Percolation*. Cambridge University Press, Cambridge, 1996.

[81] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava. Coverage problems in wireless ad-hoc sensor networks. In *in IEEE INFOCOM*, pages 1380–1387, 2001.

[82] M. V. Men'shikov and S. A. Zuev. Estimation algorithms of infinite graphs percolation threshold. In *Proceedings of the International Conference on Fundamentals of Computation Theory*, pages 310–313, London, UK, 1987. Springer-Verlag.

[83] P. A. P. Moran and S. F. de St. Groth. Random circles on a sphere. *Biometrika*, 49:389–396, 1962.

[84] A. Muhammad and M. Egerstedt. Control using higher order laplacians in network topologies. In *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto, Japan*, July 2006.

[85] J. Munkres. *Elements of Algebraic Topology*. Addison Wesley, 1993.

[86] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39:419–441, March 2008.

[87] G. Peccati, J. Solé, M. Taqqu, and F. Utzet. Stein's method and normal approximation of poisson functionals. *Annals of Probability*, 38(2):443–478, 2010.

[88] M. D. Penrose. The longest edge of the random minimal spanning tree. *The Annals of Applied Probability*, 7:340–361, 1997.

[89] M. D. Penrose. Extremes for the minimal spanning tree on normally distributed points. *Advances in Applied Probability*, 30:628–639, 1998.

[90] M. D. Penrose. On k-connectivity for a geometric random graph. *Random Structures and Algorithms*, 15:145–164, 1999.

[91] M. D. Penrose. A strong law for the largest nearest-neighbour link between points. *Journal of the London Mathematical Society. Second Series*, 60:951–960, 1999.

[92] M. D. Penrose. A strong law for the longest edge of the minimal spanning tree. *The Annals of Probability*, 15:246–260, 1999.

[93] M. D. Penrose. Central limit theorem for k-nearest neighbour distances. *Stochastic Processes and their Applications*, 85:295–320, 2000.

[94] M. D. Penrose. Vertex ordering and partitioning problems for random spatial graphs. *The Annals of Applied Probability*, 10:517–538, 2000.

[95] M. D. Penrose. Focusing of the scan statistic and geometric clique number. *Advances in Applied Probability*, 34:739–753, 2001.

[96] M. D. Penrose. *Geometric Random Graphs*. Oxford University Press, 2003.

[97] M. D. Penrose and J. E. Yukich. Weak laws of large numbers in geometric probability. *The Annals of Applied Probability*, 13:277–303, 2003.

[98] G. Pottie and W. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, 43:51–58, may 2000.

[99] N. Privault. *Stochastic Analysis in Discrete and Continuous Settings*. Springer, 2009.

[100] P. Robert. *Stochastic networks and queues*, volume 52 of *Applications of Mathematics*. Springer-Verlag, Berlin, French edition, 2003. Stochastic Modelling and Applied Probability.

[101] S. Roman. *The Umbral of Calculus*. New York: Academic Press, 1984.

[102] J. J. Rotman. *An Introduction to Algebraic Topology (Graduate Texts in Mathematics)*. Springer, July 1998.

[103] R. Roy. Coverage of space in booleans models. *Dynamics and Stochastics*, 48:119–127, 2006.

[104] L. A. Shepp. Covering the circle with random arcs. *Israel Journal of Mathematics*, 22:158–160, 1972.

[105] A. F. Siegel and L. Holst. Covering the circle with random arcs of random sizes. *Journal of Applied Probability*, 19(2):373–381, 1982.

[106] A. F. Siegel and L. Holst. Covering the circle with random arcs of random sizes. *Journal of Applied Probability*, 19(2):373–381, June 1982.

[107] V. D. Silva and R. Ghrist. Homological sensor networks. *Notices of the American Mathematical Society*, 54:2007, 2007.

[108] H. Solomon. *Geometric Probability*. SIAM: Philadelphia, 1978.

[109] J. M. Steele. *Probability Theory and Combinatorial Optimization*. Society fot Industrial and Applied Mathematics, Philadelphia, 1997.

[110] W. L. Stevens. Solution to a geometrical problem in probability. *Annals of Eugenics*, 9:315–320, 1939.

[111] D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications*. John Wiley & sons, 1987.

[112] S.-F. Su. *The UMTS air-interface in RF engineering: design and operation of UMTS networksS*. McGraw-Hill, 2007.

[113] T. Charles Clancy. Achievable capacity under the interference temperature model. In *INFOCOM*, 2007.

[114] S.-H. Teng and F. Yao. k-nearest-neighbor clustering and percolation theory. *Algorithmica*, 49(11):192–211, november 2007.

[115] D. Tian and N. D. Georganas. A coverage-preserving node scheduling scheme for large wireless sensor networks. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 32–41. ACM Press, 2002.

[116] W. A. Whitworth. *DCC Exercises on Choice and Chance*. Deighton Bell and Co.: Cambridge, 1897.

[117] WiMedia. The best choice for wireless pans. Technical report, WiMedia Radio optimizes speed, range and power, August 2008.

[118] WiMedia Alliance. Multiband ofdm physical layer specification. Technical report, August 2009.

[119] D. C. Wood. Technical report 15-92. Technical report, University of Kent computing Laboratory, University of Kent, Canterburry, 1992.

[120] F. Xue and P. R. Kumar. The number of neighbors needed for connectivity of wireless networks. *Wirel. Netw.*, 10:169–181, March 2004.

[121] J. E. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*. Lecture Notes in Mathematics, vol 1675, Springer, Berlin, 1998.

[122] H. Zhang and J. Hou. Maintaining Sensing Coverage and Connectivity in Large Sensor Networks. *Ad Hoc & Sensor Wireless Networks*, 1(1-2), 2005.

[123] A. Zomorodian and G. Carlson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 2004.

[124] S. A. Zuev and A. F. Sidorenko. Continuous models of percolation theory. i. *Theoretical and Mathematical Physics*, 62:51–58, 1985.

[125] S. A. Zuev and A. F. Sidorenko. Continuous models of percolation theory. ii. *Theoretical and Mathematical Physics*, 62:171–177, 1985.