Thèse

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE

UMR 7501

Strasbourg

présentée pour obtenir le grade de docteur de
l'Université de Strasbourg
Spécialité MATHÉMATIQUES

Tsung-Hsuan Tsai

Phase transitions in random groups: free subgroups and van Kampen 2-complexes

Soutenue le 12 Juillet 2022
devant la commission d'examen

Thomas Delzant, directeur de thèse
Emmanuel Breuillard, rapporteur
Danny Calegari, rapporteur
Goulnara Arzhantseva, examinatrice
Frédérique Bassino, examinatrice
François Guéritaud, examinateur

irma.math.unistra.fr

IRMA Institut de Recherche Mathématique Avancée

cnrs

Université de Strasbourg

Phase transitions in random groups: free subgroups and van Kampen 2-complexes

Tsung-Hsuan Tsai

Institut de Recherche Mathématique Avancée

Université de Strasbourg

Thèse de doctorat sous la direction de

Thomas Delzant

12 Juillet 2022

Résumé

Dans cette thèse, nous étudions les transitions de phase dans les groupes aléatoires à densité. Un groupe aléatoire à densité d est défini par une présentation avec m générateurs et $(2m-1)^{d\ell}$ relations aléatoires, où ℓ est la longueur maximale des relations. Nous avons deux résultats principaux : un sur le problème des sous-groupes libres et l'autre sur l'existence des 2-complexes de van Kampen.

Pour tout entier r entre 1 et $m-1$, nous trouvons une transition de phase à la densité $d_r = \min\{\frac{1}{2}, 1 - \log_{2m-1}(2r-1)\}$: Si $d > d_r$, alors les r premiers générateurs engendrent le groupe entier ; si $d < d_r$, alors les r premiers générateurs engendrent un sous-groupe libre. Ce résultat donne de nouveaux exemples de présentations de groupes satisfaisant la propriété de Freiheitssatz, avec une grande variété de longueurs de relations.

Pour chaque 2-complexe d'une forme géométrique donnée, nous donnons une densité critique d_c qui caractérise l'existence d'un 2-complexe de van Kampen dont le 2-complexe sous-jacent est celui donné. Afin de prouver ce résultat, nous étudions en détail la formule d'intersection pour les sous-ensembles aléatoires et donnons une version multidimensionnelle de cette formule.

Abstract

In this thesis, we study phase transitions in random groups at density. A random group at density d is defined by a presentation with m generators and $(2m-1)^{d\ell}$ random relations, where ℓ is the maximal length of the relations. We have two main results: one on the free subgroup problem and the other on the existence of van Kampen 2-complexes.

For any integer r between 1 and $m-1$, we find a phase transition at the density $d_r = \min\{\frac{1}{2}, 1 - \log_{2m-1}(2r-1)\}$: If $d > d_r$, then the r first generators generate the whole group; if $d < d_r$, then the r first generators generate a free subgroup. This result gives new examples of group presentations satisfying the Freiheitssatz property, with a wide variety of relation lengths.

For each 2-complex of a given geometric form, we give a critical density d_c which characterizes the existence of a van Kampen 2-complex whose underlying 2-complex is the given one. In order to prove this result, we study in detail the intersection formula for random subsets and give a multidimensional version of this formula.

Remerciement

Je tiens tout d'abord, bien sûr, à remercier mille fois Thomas Delzant, mon directeur de thèse, qui m'a proposé cet excellent sujet de thèse et qui m'a guidé pendant trois ans pour réaliser ce travail. Depuis mon mémoire de M2, son enthousiasme, sa patience, sa disponibilité, sa curiosité ainsi que ses idées novatrices ont été pour moi une grande source d'inspiration et de motivation. Je n'oublierai jamais les moments de nos discussions, soit dans son bureau, soit devant le tableau au quatrième étage de l'IRMA, soit dans son agréable jardin avec son petit tableau blanc sur pattes.

Je suis très reconnaissant envers Emmanuel Breuillard et Danny Calegari d'avoir accepté d'être les rapporteurs de ma thèse, pour leur lecture détaillée et attentive de ce travail et pour leurs précieux commentaires. Je remercie également Goulnara Arzhantseva, Frédérique Bassino et François Guéritaud qui ont accepté d'être membres du jury de ma thèse.

Je tiens à remercier Rémi Coulon et Sebastien Gouezel pour leur accueil chaleureux lors de mon séjour de trois jours à Rennes, et François Dahmani de m'avoir accueilli à Grenoble. Je remercie tous les chercheurs de l'IRMA ainsi que tout le personnel pour avoir rendu mon séjour à Strasbourg très agréable. En particulier Xiaolin Zeng, Jean Berard et François Guéritaud qui ont donné des cours de M2 intéressants que j'ai suivis. Je remercie aussi Vlada Limic et Pierre Beaumann d'avoir assisté à mon comité de suivi, de m'avoir donné plusieurs conseils utiles pour compléter une thèse.

Je remercie chaleureusement mes cobureaux Nicolas, Laura, Clarence et Tom pour les moments conviviaux de discussions aux pause-cafés dans le bureau 105. Je remercie Francisco, Clarence, Basile, David, Frédérique, Laura, Raoul, Alexander, Alexandre, Suzanne, Valdo, Pierre-Alexandre, Nicolas, Martin, Coung, Kien, Hao, Lukas, Archia, Yohann, Daniel, Guillaume, Clément, Adam, Thibaut, Rorenzo, Romane, Brieuc, Khalef, Victoria et tous les autres doctorants que j'ai rencontrés durant ces trois années de thèse. Les déjeuners, les séminaires doctorants et les groupes de travail sont des moments inoubliables.

Je voudrais remercier Chen-Hao, Daniel, Basile et Clarence pour avoir relu et corrigé mes articles, m'aidant ainsi à combler mes lacunes en français et en anglais. J'aimerais remercier tous ceux qui m'ont contacté pendant les périodes de confinement sanitaire. Une discussion mathématique en ligne, une partie de jeu de société, un appel, un e-mail ou même un court message n'ont jamais été aussi précieux.

En fin, je termine en remerciant ma famille à Taïwan. Ma mère, mon père et mon frère, qui me soutiennent toujours à distance, non seulement pendant cette thèse de trois ans à Strasbourg, mais aussi pendant mon séjour de huit ans en France. Sans eux, je ne pourrais pas y arriver.

Contents

Chapter 0

Introduction en français

“I feel, random groups altogether may grow up as healthy as random graphs, for example.”

—— M. Gromov, “Spaces and Questions” [Gro00].

L’étude des groupes aléatoires est très similaire à celle des graphes aléatoires : nous nous intéressons aux *comportements asymptotiques* et aux *transitions de phase*. Dans cette thèse, nous étudions les transitions de phase dans le modèle à densité des groupes aléatoires introduit par M. Gromov dans [Gro93]. Il y a deux objectifs principaux : découvrir les propriétés génériques des présentations de groupes, et construire de nouveaux exemples de groupes hyperboliques.

Les contributions principales de cette thèse sont présentées dans les chapitres 2, 3 et 4. Dans le chapitre 2, nous étudions la *formule d’intersection* pour les sous-ensembles aléatoires à densité, qui est un outil de base pour prouver les transitions de phase dans les groupes aléatoires. Dans le chapitre 3, nous prouvons une transition de phase montrant que les r premiers générateurs d’une présentation aléatoire de groupe engendrent soit le groupe entier, soit un sous-groupe libre et quasi-convexe. Dans le chapitre 4, nous établissons une transition de phase qui caractérise l’existence des diagrammes de van Kampen. Quelques questions ouvertes sont proposées dans le chapitre 5.

0.1 Qu’est-ce qu’un groupe aléatoire ?

0.1.1 Définitions et exemples

Les groupes aléatoires sont des groupes obtenus par une construction probabiliste. En général, un groupe aléatoire peut être défini comme suit.

Définition 1. Un groupe aléatoire est une variable aléatoire à valeurs dans un ensemble donné de groupes.

Nous nous concentrons sur les groupes aléatoires construits à partir d’un ensemble fini de présentations finies de groupes. Plus précisément, en fixant un ensemble de m générateurs $X_m = \{x_1, \dots, x_m\}$, un groupe aléatoire est défini par une présentation aléatoire de groupe

$$G = \langle x_1, \dots, x_m | r_1, \dots, r_k \rangle$$

où $r_1, \dots, r_k$ sont des relateurs choisis au hasard parmi un ensemble fini de mots réduits de $X_m^{\pm}$.

Si G est un groupe aléatoire et que P est une propriété de groupe (par exemple, trivial, libre, sans torsion, etc.), nous pouvons parler de la probabilité que G satisfasse P. Nous gardons ce point de vue de "variable aléatoire" pour tout objet aléatoire. Par exemple, le rang (le nombre minimal de générateurs) d'un groupe aléatoire est une $\mathbb{N}$-variable aléatoire. Toute variable aléatoire considérée dans cette thèse est à valeurs dans un ensemble fini muni de la tribu discrète.

Voici quelques exemples de groupes aléatoires.

Exemples. Dans les exemples suivants, un groupe aléatoire est défini par une présentation $G = \langle x_1, \dots, x_m | R \rangle$ où R est un ensemble aléatoire de relateurs à spécifier.

1. (longueur fixe, modèle uniforme) Soit k, ℓ des entiers. L'ensemble des relateurs $R = \{r_1, \dots, r_k\}$ suit la loi uniforme dans l'ensemble de tous les k-uplets de mots réduits de longueur ℓ.

2. (longueurs variées, modèle d'échantillonnage) Soit k, ℓ, ℓ' des entiers avec $\ell < \ell'$. Pour construire R, choisir un relateur uniformément dans l'ensemble de tous les mots réduits de longueurs comprises entre ℓ et ℓ', répéter k fois.

   Notons qu'un mot peut être pris deux fois, mais cela se produit avec une très faible probabilité si le nombre de relateurs k est beaucoup plus petit que le nombre total de mots considérés ($(2m-1)^{\ell'+O(1)}$ dans ce cas).

3. (longueur bornée, modèle de Bernoulli) Soit ℓ un entier, soit $p \in [0, 1]$. Pour construire l'ensemble des relateurs R, chaque mot réduit r de longueur au plus ℓ est choisi indépendamment avec la même probabilité p.

   Notons que le nombre de relateurs suit la loi binomiale $B(N, p)$ avec $N = (2m-1)^{\ell+O(1)}$ le nombre total de mots de longueur au plus ℓ.

Nous nous intéressons au *comportements asymptotiques* d'un groupe aléatoire. C'est-à-dire, que se passe-t-il lorsque le nombre de générateurs ou le nombre de relateurs tend vers l'infini ? Pour décrire ce phénomène, nous utilisons la définition suivante.

Définition 2. Soit $(Q_n)_{n\geq 1}$ une suite d'événements. On dit que l'événement Q_n est satisfait *asymptotiquement presque sûrement*, noté a.p.s. Q_n, si la probabilité de Q_n converge vers 1 lorsque n tend vers l'infini.

Si (G_n) est une suite de groupes aléatoires et (P_n) est une suite de propriétés de groupes (par exemple, être trivial, être hyperbolique, de rang n etc.), on peut dire que a.p.s. G_n a la propriété P_n. L'indice n peut être le nombre de générateurs m ou la longueur des relateurs ℓ.

Exemple. Soit G_m un groupe aléatoire avec m générateurs et un relateur choisi uniformément parmi tous les mots réduits de longueur 3. Si le nombre de générateurs m est très grand, alors avec une probabilité supérieure à $1 - m^{-2/3+o(1)}$ le relateur choisi est du type $r = xyz$ avec x, y, z éléments différents dans $X_m^{\pm}$. On peut supprimer le générateur z par la relation $z = x^{-1}y^{-1}$. Ainsi, avec une probabilité convergeant vers 1 lorsque m tend vers l'infini, le groupe aléatoire G_m est un groupe libre de rang $m-1$.

On en conclut que a.p.s. le groupe aléatoire G_m est un groupe libre de rang $m-1$.

0.1.2 Modèles de groupes aléatoires et résultats classiques

Un *modèle* de groupes aléatoires est une *suite* de groupes aléatoires à spécifier. Tout au long de cette thèse, nous ferons clairement la distinction entre un *groupe aléatoire* et une *suite de groupes aléatoires*.

Voici une liste de modèles de groupes aléatoires étudiés dans la littérature.

0.1.2.a Le modèle de longueurs diverses, M. Gromov [Gro87] §0.2

Soit $m \geq 2$, $k \geq 1$. Soit $\ell_1, \dots, \ell_k$ des entiers. Un groupe aléatoire avec m générateurs et k relateurs de longueurs diverses est défini par $G(m, \ell_1, \dots, \ell_k) = \langle x_1, \dots, x_m | r_1, \dots, r_k \rangle$ où r_i est choisi uniformément parmi tous les mots réduits de $x_1^{\pm}, \dots, x_m^{\pm}$ de longueur ℓ_i.

Supposons que $\ell_1 \leq \cdots \leq \ell_k$. On étudie le comportement asymptotique lorsque la longueur minimale ℓ_1 tend vers l'infini. Nous pouvons considérer un sous-modèle dans lequel la longueur maximale ℓ_k dépend de ℓ_1. Par exemple, $\ell_k = 2\ell_1$ ou $\ell_k = e^{10\ell_1}$. Avec ce modèle, Gromov a remarqué que l'hyperbolicité est *générique* pour les groupes de présentation finie :

Théorème (M. Gromov, [Gro87] §0.2)**.** *Soit $G(m, \ell_1, \dots, \ell_k)$ un groupe aléatoire avec longueurs diverses de relateurs $\ell_1 \leq \cdots \leq \ell_k$, alors a.p.s. (lorsque la longueur minimale ℓ_1 tend vers l'infini) le groupe aléatoire $G(m, \ell_1, \dots, \ell_k)$ est un groupe hyperbolique non élémentaire.*

Une preuve détaillée de ce théorème est donnée par C. Champetier dans [Cha91].

0.1.2.b Le modèle à densité, M. Gromov [Gro93] 9.B

Soit S_ℓ l'ensemble des mots réduits d'une longueur fixée ℓ dans l'alphabet $\{x_1, \dots, x_m\}$. Notons que $|S_\ell| = (2m-1)^{\ell + O(1)}$ (lorsque $\ell \to \infty$). Une suite de groupes aléatoires $(G_\ell(m, d))$ avec $m \geq 2$ générateurs à densité $d \in [0, 1]$ est définie par $G_\ell(m, d) = \langle x_1, \dots, x_m | R_\ell \rangle$ avec R_ℓ choisi *uniformément* parmi tous les sous-ensembles de S_ℓ de cardinal $\lfloor (2m-1)^{d\ell} \rfloor$.

Par rapport au modèle précédent, la longueur des relateurs est fixée, mais le nombre de relateurs *croît exponentiellement* avec la longueur. Gromov a montré qu'il y a une *transition de phase* à la densité $d = 1/2$. Une preuve détaillée de ce théorème se trouve dans [Oll04] Section 2 par Y. Ollivier.

Théorème (M. Gromov, [Gro93] 9.B)**.** *Soit $(G_\ell(m, d))$ une suite de groupes aléatoires de densité d.*

- *Si $d < 1/2$, alors a.p.s. le groupe aléatoire $G_\ell(m, d)$ est hyperbolique non élémentaire.*
- *Si $d > 1/2$, alors a.p.s. le groupe aléatoire $G_\ell(m, d)$ est soit trivial soit isomorphe à $\mathbb{Z}/2\mathbb{Z}$.*

Dans le même paragraphe ([Gro93] 9.B), il est remarqué qu'il y a une transition de phase pour la condition de petite simplification $C'(\lambda)$ (cf. [LS77] pour une définition).

Théorème (M. Gromov, [Gro93] 9.B)**.** *Soit $(G_\ell(m, d))$ une suite de groupes aléatoires de densité d. Soit $0 < \lambda < 1$.*

- *Si $d < \lambda/2$, alors a.p.s. la présentation définissant $G_\ell(m, d)$ satisfait $C'(\lambda)$.*
- *Si $d > \lambda/2$, alors a.p.s. la présentation définissant $G_\ell(m, d)$ ne satisfait pas $C'(\lambda)$.*

Une preuve détaillée par un analogue du principe des tiroirs probabiliste est donnée dans [BNW20] Théorème 2.1. Dans cette thèse, nous donnons une preuve (Théorème 2.46) comme application de notre Théorème C, et une preuve beaucoup plus simple (Théorème 4.23) comme application de notre Théorème H.

0.1.2.c Le modèle à peu de relateurs, G. Arzhantseva et A. Ol'shanskii [AO96]

Soit $m \geq 2$ et $k \geq 1$. Une suite de groupes aléatoires (G_ℓ) avec m générateurs et k relateurs est définie par $G_\ell = \langle x_1, \dots, x_m | r_1, \dots, r_k \rangle$ où l'ensemble des relateurs $\{r_1, \dots, r_k\}$ suit la loi uniforme dans l'ensemble de tous les k-uplets de mots réduits de longueur au plus ℓ.

Notons qu'il s'agit d'une version plus simple du modèle des longueurs diverses. Bien que les relateurs de toutes les longueurs plus petit que ℓ soient pris en compte, pour tout $c < 1$ a.p.s. il n'existe aucun relateur plus court que $c\ell$. Ce modèle est un cas particulier du modèle à densité de Gromov, avec la densité $d = 0$. Nous avons les deux propriétés suivantes.

- Pour tout $\lambda > 0$, a.p.s. la présentation définissant G_ℓ satisfait la condition de petite simplification $C'(\lambda)$.
- A.p.s. G_ℓ est un groupe hyperbolique non élémentaire.

Le résultat principal dans [AO96] est le suivant.

Théorème (G. Arzhantseva et A. Ol'shanskii, [AO96]). *Soit (G_ℓ) une suite de groupes aléatoires à peu de relateurs avec $m \geq 2$ générateurs et k relateurs. A.p.s. tout sous-groupe de rang $(m-1)$ de G_ℓ est libre.*

0.1.2.d Le modèle triangulaire, A. Żuk [Żuk03]

Un relateur triangulaire par rapport à $X_m = \{x_1, \ldots, x_m\}$ est un mot cycliquement réduit de longueur 3. Notons que le nombre de relateurs triangulaires est $(2m-1)^{3+o(1)}$ (avec $m \to \infty$). Une suite de groupes triangulaires aléatoires $(G_m(d))$ à densité d est définie par $G_m(d) = \langle X_m | R_m \rangle$ avec R_m uniformément choisi parmi tous les ensembles de relateurs triangulaires de cardinal compris entre $c^{-1}(2m-1)^{3d}$ et $c(2m-1)^{3d}$ avec un $c > 1$.

Par rapport au modèle à densité, le nombre de générateurs tend vers l'infini, au lieu de la longueur des relateurs. Żuk a remarqué que la transition de phase de Gromov à la densité $d = 1/2$ est encore valable dans ce modèle, et a montré une transition de phase à la densité $d = 1/3$.

Théorème (A. Żuk, [Żuk03]). *Soit $G_m(d)$ un groupe aléatoire triangulaire avec une densité d.*

- *Si $d < 1/2$, alors a.p.s. $G_m(d)$ est hyperbolique non élémentaire.*
- *Si $d > 1/2$, alors a.p.s. $G_m(d)$ est trivial.*

Théorème (A. Żuk, [Żuk03]). *Soit $G_m(d)$ un groupe aléatoire triangulaire avec une densité d.*

- *Si $d < 1/3$, alors a.p.s. $G_m(d)$ est un groupe libre.*
- *Si $d > 1/3$, alors a.p.s. $G_m(d)$ admet la propriété (T) de Kazhdan.*

0.1.2.e Le modèle triangulaire de Bernoulli, S. Antoniuk, T. Łuczak et J. Świątkowski [AŁŚ15]

Une suite de groupes aléatoires triangulaires (de Bernoulli) $(G(m,p))_{m \geq 1}$ est définie par $G(m,p) = \langle X_m | R_m \rangle$ avec R_m construite de façon suivante : chaque relateur triangulaire est choisi indépendamment avec une probabilité $p = p(m)$.

Si $p = (2m-1)^{3d-3}$, alors l'espérance du nombre de relateurs est $(2m-1)^{3d+o(1)}$, et ce modèle est très proche du modèle triangulaire (uniforme) de Żuk. Antoniuk, Łuczak et Świątkowski ont adapté le résultat de Żuk dans leur modèle triangulaire de Bernoulli, et ont donné une transition de phase plus fine à la densité $d = 1/3$, c'est-à-dire autour de la probabilité $p \sim m^{-2}$.

Théorème (S. Antoniuk, T. Łuczak et J. Świątkowski, [AŁŚ15]). *Soit $G(m,p)$ un groupe aléatoire triangulaire de Bernoulli. Il existe des nombres réels positifs $c < c', C' < C$ tels que :*

- *Si $p < cm^{-2}$, alors a.p.s. $G_m(p)$ est un groupe libre.*
- *Si $c'm^{-2} < p < C'm^{-2}\log m$, alors a.p.s. $G_m(p)$ n'est ni libre ni ayant la propriété (T).*
- *Si $p > Cm^{-2}\log m$, alors a.p.s. $G_m(p)$ a la propriété (T).*

0.1.2.f Le modèle à densité de Bernoulli (un nouvel exemple)

L'espérance du nombre de relateurs est $|B_\ell|^d = (2m-1)^{d\ell+O(1)}$, ce modèle est très proche du modèle à densité original de [Gro93]. Les résultats suivants sont encore valables :

- La transition de phase trivialité-hyperbolicité de Gromov à la densité $d = 1/2$.
- La transition de phase $C'(\lambda)$ de Gromov à la densité $d = \lambda/2$.

0.1.3 Questions principales

En 2003, Gromov a défini la notion générale de groupes aléatoires dans [Gro03] et a proposé dans Section 1.9 le problème général suivant : "*déterminer des invariants asymptotiques et des phénomènes de transition de phase pour les groupes aléatoires.*"

0.1.3.a Sous-groupes libres

En particulier, le problème de Gromov [Gro03] 1.9 (iv) est l'"*existence/non-existence de sous-groupes non libres*". Notre objectif est de généraliser le résultat "tout sous-groupe de rang $(m-1)$ est libre" d'Arzhantseva-Ol'shanskii dans le modèle à densité de Gromov. Soit $G_\ell(m,d)$ une suite de groupes aléatoires avec m générateurs à densité d, dans le modèle à densité de Gromov. On cherche une transition de phase :

Question 1. *Existe-t-il une densité critique $d(m)$ telle que, si $d < d(m)$, alors a.p.s. tout sous-groupe de rang $(m-1)$ de $G_\ell(m,d)$ est libre ; par contre, si $d > d(m)$, alors a.p.s. il existe un sous-groupe de rang $(m-1)$ qui est non libre ?*

Plus généralement, on peut se demander :

Question 2. *Soit $1 \leq r \leq m-1$. Existe-t-il une densité critique $d(m,r)$ telle que, si $d < d(m,r)$, alors a.p.s. tout sous-groupe de rang r de $G_\ell(m,d)$ est libre ; par contre, si $d > d(m,r)$, alors a.p.s. il existe un sous-groupe de rang r qui est non libre ?*

Nous répondons partiellement à la question 2 (Théorème D et Théorème E).

0.1.3.b Les 2-complexes de van Kampen

Dans [Gro93], afin de prouver l'hyperbolicité des groupes aléatoires à densité $d < 1/2$, Gromov a montré que tout diagramme de van Kampen satisfait une certaine inégalité isopérimétrique. Plus précisément, on a le lemme suivant.

Lemme (M. Gromov [Gro93], énoncé par Y. Ollivier [Oll04])**.** *Soit $(G_\ell(m,d))$ une suite de groupes aléatoires. Si $d < 1/2$, alors pour tout entier $K \geq 1$ et tout nombre réel $s > 0$, a.p.s. tout diagramme de van Kampen réduit D de $G_\ell(m,d)$ d'au plus K faces satisfait l'inégalité isopérimétrique*

$$|\partial D| \geq (1 - 2d - s)\,\ell|D|.$$

Autrement dit, a.p.s. il n'existe pas de diagramme de van Kampen D de $G_\ell(m,d)$ d'au plus K faces satisfaisant l'inégalité $\frac{|\partial D|}{\ell |D|} < 1-2d$. On peut se demander si la réciproque est vraie :

Question 3. *Soit $c > 1-2d$ un nombre réel et soit $K \geq 1$ un nombre entier. Existe-t-il a.p.s. un diagramme de van Kampen D de $G_\ell(m,d)$ avec K faces qui satisfait l'égalité $|\partial D| = c\ell|D|$?*

Le théorème H donne une réponse à cette question dans un cas plus général, pour les 2-complexes de van Kampen.

0.1.4 Densité des sous-ensembles aléatoires

Soit S_ℓ l'ensemble de tous les mots réduits de longueur ℓ de l'alphabet $\{x_1^\pm, \dots, x_m^\pm\}$. Dans [Gro93], pour construire un groupe aléatoire $G = \langle x_1, \dots, x_m | R_\ell \rangle$, l'idée originale de Gromov est de choisir un *sous-ensemble aléatoire* R_ℓ de S_ℓ, d'un cardinal très proche de $|S_\ell|^d$. Le paramètre d est appelé la *densité* de R_ℓ dans S_ℓ.

On peut considérer la distribution uniforme dans l'ensemble des sous-ensembles de cardinal $\lfloor |S_\ell|^d \rfloor$ (comme dans les travaux d'Ollivier [Oll04; Oll05; Oll07]), ou par le modèle de Bernoulli de paramètre $|S_\ell|^{d-1}$ (comme dans [AŁŚ15], ou le modèle 0.1.2.f). Dans cette sous-section, nous introduisons un modèle de sous-ensemble aléatoire, appelé le modèle *invariant par permutation* de densité, dont l'origine se trouve dans un paragraphe de [Gro93] p.272.

Définition 3 ([Gro93] 9.A)**.** La densité d'un sous-ensemble A dans un ensemble fini E est

$$\operatorname{dens}_E A := \log_{|E|}(|A|).$$

En particulier, $\operatorname{dens}_E(A)$ est le nombre $d \in \{-\infty\} \cup [0,1]$ tel que $|A| = |E|^d$. Le cas $d = -\infty$ correspond au cas où A est un ensemble vide. S'il n'y a pas d'ambiguïté de l'ensemble d'ambiance E, nous omettons l'indice et désignons simplement la densité par $\operatorname{dens} A$.

Un *sous-ensemble aléatoire* d'un ensemble fini E est une variable aléatoire à valeurs dans l'ensemble des sous-ensembles de E.

Le *densité d'un sous-ensemble aléatoire* est alors une variable aléatoire à valeurs dans $\{-\infty\} \cup [0,1]$.

La *formule d'intersection* pour les sous-ensembles aléatoires, introduite par Gromov dans [Gro93], est l'outil principal pour prouver les transitions de phase dans le modèle à densité des groupes aléatoires. La formule s'énonce comme suit :

Métathéorem ([Gro93] p.270)**.** *Deux sous-ensembles aléatoires A et B d'un ensemble fini E satisfont*

$$\operatorname{dens}(A \cap B) = \operatorname{dens} A + \operatorname{dens} B - 1$$

avec la convention

$$\operatorname{dens}(A \cap B) < 0 \iff A \cap B = \emptyset.$$

Gromov n'a pas précisé comment prendre les sous-ensembles aléatoires à densité dans l'énoncé original de la formule d'intersection. Dans [Gro93] p.272, il affirme que la classe des sous-ensembles aléatoires qui sont *densables* et *invariants par permutations* est stable par les opérations de la théorie des ensembles et *satisfait la formule d'intersection*.

Les définitions de *densable* et *invariant par permutations* sont données ci-dessous. Comme nous nous intéressons aux comportements asymptotiques, nous fixons une suite d'ensembles finis $\boldsymbol{E} = (E_n)$ avec $|E_n| \to \infty$, et étudions une suite de sous-ensembles aléatoires $\boldsymbol{A} = (A_n)$ où A_n est un sous-ensemble aléatoire de E_n.

Définition 4. Une suite de sous-ensembles aléatoires $\boldsymbol{A} = (A_n)$ de $\boldsymbol{E} = (E_n)$ est dite *densable* de densité $d \in \{-\infty\} \cup [0, 1]$ si la suite de variables aléatoires $(\mathrm{dens}_{E_n}(A_n))$ converge en loi vers la constante d lorsque n tend vers l'infini. Notons

$$\mathrm{dens}_{\boldsymbol{E}}\,\boldsymbol{A} = d.$$

Définition 5. Une suite de sous-ensembles aléatoires $\boldsymbol{A} = (A_n)$ de $\boldsymbol{E} = (E_n)$ est dite *invariante par permutations* si la mesure de A_n est invariante sous les permutations de E_n.

C'est-à-dire que pour tout sous-ensemble $a \subset E_n$ et toute permutation $\sigma \in \mathcal{S}(E_n)$, on a $\mathbf{Pr}(A_n = a) = \mathbf{Pr}(A_n = \sigma(a))$. Un énoncé complet de la formule d'intersection pour les suites de sous-ensembles aléatoires densables et invariantes par permutations est donné dans Section 0.2, Théorème A.

0.1.5 Le modèle de groupe aléatoire considéré dans cette thèse

A l'aide des sous-ensembles aléatoires, nous pouvons définir le modèle de groupes aléatoires considéré dans cette thèse : les modèles invariants par permutation à densité de groupes aléatoires.

Fixons un alphabet $X_m = \{x_1, \dots, x_m\}$ comme générateurs de présentations de groupes. Dans le modèle à densité original de Gromov, la longueur du relateur d'un groupe aléatoire $G_\ell(m, d)$ est un nombre fixé ℓ. Dans le cas où ℓ est pair, lorsque $d > 1/2$, a.p.s. $G_\ell(m, d)$ est $\mathbb{Z}/2\mathbb{Z}$ mais pas trivial. Pour éviter cette situation, nous considérons des mots de longueur *au plus* ℓ, comme dans [AO96]. Soit B_ℓ l'ensemble des mots cycliquement réduits de longueur au plus ℓ.

Définition 6. Une suite de groupes aléatoires $(G_\ell(m, d))$ avec m générateurs de densité d est définie par

$$G_\ell(m, d) = \langle X_m | R_\ell \rangle,$$

où (R_ℓ) est une suite de sous-ensembles aléatoires densable et invariante par permutation de (B_ℓ), de densité d.

Notons que le cardinal de B_ℓ est $(2m-1)^{\ell+O(1)}$, donc le cardinal de R_ℓ est $(2m-1)^{d\ell+o(\ell)}$ avec une grande probabilité (cf. Proposition 2.6). Le modèle 0.1.2.f est inclus dans ce modèle, car un sous-ensemble aléatoire de Bernoulli est invariant par permutations. Le modèle à peu de relateurs d'Arzhantseva-Ol'shanskii est inclus dans ce modèle avec $d = 0$, car la loi uniforme est invariante par permutations, et la densité $\mathrm{dens}_{B_\ell}(R_\ell)$ converge vers 0 si le nombre de relateurs $|R_\ell| = k$ est fixé.

0.2 Résultats principaux de cette thèse

Les contributions de cette thèse sont présentées dans les chapitre 2, 3 et 4. Dans cette section, nous présentons une liste de nos résultats principaux.

0.2.1 Densité des sous-ensembles aléatoires et la formule d'intersection

Dans le chapitre 2, nous travaillons sur la formule d'intersection des sous-ensembles aléatoires, présentée dans l'article [Tsa21a], à paraître dans le *Journal of Combinatorial Algebra*.

Nous prouvons d'abord la formule d'intersection pour les suites de sous-ensembles aléatoires qui sont densables et invariantes par permutation.

Théorème A (Théorème 2.25). *Soit $\boldsymbol{A} = (A_n), \boldsymbol{B} = (B_n)$ deux suites de sous-ensembles aléatoires indépendantes, densables et invariantes par permutation d'une suite d'ensembles finis $\boldsymbol{E} = E_n$, de densités α, β. Si $\alpha + \beta \neq 1$, alors la suite de sous-ensembles aléatoires $\boldsymbol{A} \cap \boldsymbol{B} := (A_n \cap B_n)$ est encore densable et invariante par permutation. De plus :*

$$\mathrm{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{si } \alpha + \beta > 1 \\ -\infty & \text{si } \alpha + \beta < 1. \end{cases}$$

De plus, la formule d'intersection est valable entre une suite de sous-ensembles aléatoires et une suite de sous-ensembles fixés.

Théorème B (Théorème 2.28). *Soit $\boldsymbol{A}$ une suite de sous-ensembles aléatoires densable et invariante par permutations de $\boldsymbol{E}$, de densité de d. Soit $\boldsymbol{X}$ une suite de sous-ensembles fixés de $\boldsymbol{E}$, de densité α. Si $d + \alpha \neq 1$, alors la suite de sous-ensembles aléatoires $\boldsymbol{A} \cap \boldsymbol{X}$ est densable et*

$$\mathrm{dens}(\boldsymbol{A} \cap \boldsymbol{X}) = \begin{cases} d + \alpha - 1 & \text{si} d + \alpha > 1 \\ -\infty & \text{si } d + \alpha < 1. \end{cases}$$

De plus, la suite $\boldsymbol{A} \cap \boldsymbol{X}$ est une suite de sous-ensembles aléatoires densable et invariante par permutations de $\boldsymbol{X}$, de densité $\frac{d+\alpha-1}{\alpha}$.

Notons que $A_n \cap X_n$ n'est pas invariant sous les permutations de E_n si $X_n \neq E_n$.

Nous développons une forme généralisée : la *formule d'intersection multidimensionnelle*. Notons $E_n^{(k)}$ l'ensemble des k-uplets *2 à 2 distincts* de l'ensemble E_n. Soit $\boldsymbol{A}$ une suite de sous-ensembles aléatoires densable et invariante par permutations. Nous nous intéressons à l'intersection entre $\boldsymbol{A}^{(k)}$ et une suite de sous-ensembles densable $\boldsymbol{X}$ de $\boldsymbol{E}^{(k)}$.

Pour $k \geq 2$, la formule d'intersection n'est pas correcte en général (c.f. contre-exemple dans 2.3.1.a). Nous montrons que par une *condition d'auto-intersection* supplémentaire sur $\boldsymbol{X}$, nous pouvons obtenir la formule d'intersection.

Théorème C (Théorème 2.40). *Soit $\boldsymbol{A} = (A_n)$ une suite de sous-ensembles aléatoires densable et invariante par permutations de $\boldsymbol{E} = (E_n)$, de densité $0 < d < 1$. Soit $\boldsymbol{X} = (X_n)$ une suite de sous-ensembles fixés densable de $\boldsymbol{E}^{(k)}$, de densité de α.*

(i) Si $d + \alpha < 1$, alors a.p.s.

$$-A_n^{(k)} \cap X_n = \emptyset.$$

(ii) Si $d + \alpha > 1$ et $\boldsymbol{X}$ satisfait la condition d'auto-intersection d-petite (Définition 2.33), alors la suite de sous-ensembles aléatoires $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ est densable et

$$\mathrm{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = \alpha + d - 1.$$

Pour une application du théorème B, nous montrons que le résultat principal de [AO96] pour le modèle à peu de relateurs des groupes aléatoires peut être étendu au modèle à densité des groupes aléatoires avec une petite densité.

Théorème D (Théorème 2.48). *Soit $(G_\ell(m, d))$ une suite de groupes aléatoires avec m générateurs à densité*

$$0 \leq d < \frac{1}{120m^2 \ln(2m)}.$$

Alors a.p.s. tout sous-groupe de rang $(m-1)$ de $G_\ell(m, d)$ est libre.

Ceci répond partiellement à la question 1.

0.2.2 Freiheitssatz pour les groupes aléatoires et la transition de phase

Le chapitre 3 fait l'objet de la prépublication [Tsa21b]. Nous présentons l'un des résultats principaux de cette thèse, qui répond partiellement à la question 2.

Le *Freiheitssatz* (*théorème de liberté* en allemand) est un théorème fondamental en théorie combinatoires des groupes. Il a été proposé par M. Dehn et prouvé par W. Magnus dans sa thèse de doctorat [Mag30] en 1930 (cf. [LS77] II.5).

Théorème (W. Magnus, [Mag30]). *Soit $G = \langle x_1, \dots, x_m | r\rangle$ une présentation de groupe avec m générateurs et un relateur cycliquement réduit. Si le dernier générateur x_m apparaît dans l'unique relateur r, alors les $m-1$ premiers générateurs $x_1, \dots, x_{m-1}$ engendrent librement un sous-groupe libre de G.*

On dit qu'une présentation finie de groupe $G = \langle X|R\rangle$ satisfait la *propriété Freiheitssatz de Magnus* si tout sous-ensemble de X de cardinal $|X| - 1$ engendre librement un sous-groupe libre de G. En particulier, par le résultat d'Arzhantseva-Ol'shanskii [AO96], a.p.s. un groupe aléatoire à peu de relateurs G_ℓ possède cette propriété. Nous étudions la propriété de Magnus Freiheitssatz dans le modèle à densité des groupes aléatoires.

Soit $G_\ell(m, d) = \langle X|R_\ell\rangle$ un groupe aléatoire à densité d. Pour tout $1 \le r \le m-1$, nous trouvons une transition de phase à la densité

$$d_r = \min\left\{\frac{1}{2}, 1 - \log_{2m-1}(2r-1)\right\}.$$

Théorème E (Théorème 3.22). *Soit $(G_\ell(m, d))$ une suite de groupes aléatoires à densité d.*

1. *Si $d > d_r$, alors a.p.s. $x_1, \dots, x_r$ engendrent le groupe entier $G_\ell(m, d)$.*
2. *Si $d < d_r$, alors a.p.s. $x_1, \dots, x_r$ engendrent librement un sous-groupe libre de $G_\ell(m, d)$.*

Par symétrie, l'ensemble $\{x_1, \dots, x_r\}$ peut être remplacé par tout sous-ensemble X_r de X de cardinal r. En effet, dans la deuxième assertion, on peut remplacer l'ensemble $\{x_1, \dots, x_r\}$ par tout r-uplet de mots de $X^\pm$ de longueurs au plus $\frac{d_r - d}{5r}\ell$. En particulier, si $0 \le d < d_{m-1}$, alors la présentation de groupe $G_\ell(m, d) = \langle X|R_\ell\rangle$ a la propriété Freiheitssatz de Magnus.

Plus précisément pour la première assertion, nous prouvons que si $d > d_r$ alors a.p.s. tout générateur x_k avec $r < k \le m$ s'écrit en un mot réduit de $\{x_1^\pm \dots x_r^\pm\}$ de longueur $\ell - 1$ dans $G_\ell(m, d)$. Par conséquent, tout relateur $r_i\ dans R_\ell$ peut être remplacé par un mot réduit r_i' de $\{x_1^\pm \dots x_r^\pm\}$ de longueur au plus $\ell(\ell-1)$. Construisons R_ℓ' en remplaçant chaque relateur de R_ℓ de cette manière, nous avons le corollaire suivant.

Corollaire F. *Si $d_r < d < d_{r-1}$, alors a.p.s. le groupe $G_\ell(m, d) = \langle X|R_\ell\rangle$ admet une présentation avec r générateurs $\langle X_r|R_\ell'\rangle$ satisfaisant la propriété Freiheitssatz de Magnus (chaque sous-ensemble de X_r de cardinal $r-1$ engendre un sous-groupe libre).*

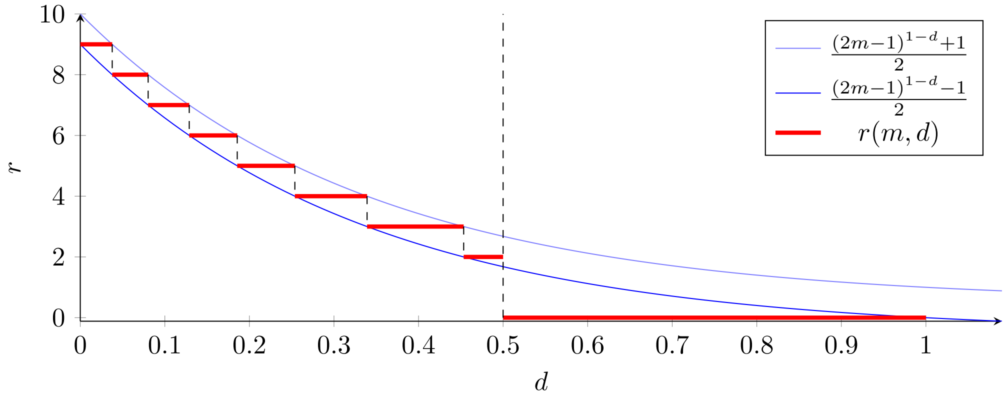
Remarque. Notons que R_ℓ' contient un ensemble de relateurs de longueurs variant de ℓ à ℓ^2. Une telle présentation ne peut pas être étudiée à l'aide des méthodes connues en théorie géométrique ou combinatoire des groupes. Néanmoins, elle nous donne de nouveaux exemples de groupes hyperboliques ayant la propriété Freiheitssatz de Magnus.

Si en plus $d_r > 1/3$ (c'est-à-dire $r < \frac{(2m-1)^{2/3}+1}{2}$), alors par le résultat de Żuk dans [Żuk03], ce sont des exemples de groupes hyperboliques ayant la propriété (T) de Kazhdan et la propriété Freiheitssatz de Magnus.

Soit $r = r(m, d)$ le nombre maximal tel qu'a.p.s. $x_1, \dots, x_r$ engendrent librement un sous-groupe libre de $G_\ell(m, d)$. Par la transition de phase à la densité $\frac{1}{2}$ [Gro93], si $d > \frac{1}{2}$, alors $r(m, d) = 0$. Si $d \leq \frac{1}{2}$, par **Théorème E**,

$$\frac{(2m-1)^{1-d} - 1}{2} \leq r(m, d) \leq \frac{(2m-1)^{1-d} + 1}{2}.$$

Comme le montre le schéma ci-dessous, puisque $r(m, d)$ est un entier, sa valeur est déterminée lorsque d n'est pas dans l'ensemble $\{d_1, \dots, d_{m-1}, 1/2\}$. Tandis que la valeur de $r(m, d)$ n'est pas déterminée lorsque $d \in \{d_1, \dots, d_{m-1}, 1/2\}$.



$r(m, d)$ with $m = 10$

0.2.3 Existence de 2-complexes de van Kampen dans les groupes aléatoires

Nous considérons un 2-complexe de van Kampen par rapport à une présentation de groupe comme un diagramme de van Kampen dans [LS77] : c'est un 2-complexe dont les arêtes sont étiquetées par des générateurs et les faces étiquetées par des relateurs. Une paire de faces est dite réductible si elles ont la même étiquette et s'il existe une arête commune sur leurs bords à la même position. Un 2-complexe de van Kampen est dit réduit s'il n'existe pas de paire de faces réductibles. C.f. Sous-section 3.1.2 et Sous-section 4.1 pour plus de détails.

Dans [Gro93], Gromov a montré que a.p.s. les diagrammes de van Kampen locaux de $G_\ell(m, d)$ satisfont une inégalité isopérimétrique (dépendant de la densité d).

Théorème (M. Gromov [Gro93] p. 274, Y. Ollivier [Oll04] chapitre 2)**.** *Pour tout $s > 0$ et $K > 0$, a.p.s. chaque diagramme de van Kampen réduit D de $G_\ell(m, d)$ avec $|D| \leq K$ satisfait l'inégalité isopérimétrique*

$$|\partial D| \leq (1 - 2d - s)\ell|D|.$$

Dans le chapitre 4, nous prouvons cette inégalité pour les 2-complexes de van Kampen, ce qui est un analogue du résultat de Gruber-Mackay [GM18] pour les groupes aléatoires triangulaires. Pour un 2-complexe Y, on note $|Y^{(1)}|$ le nombre de ses arêtes et $|Y|$ le nombre de ses faces.

Théorème G (Théorème 4.3)**.** *Soit $\varepsilon > 0$, $K > 0$. A.p.s. tout 2-complexe de van Kampen Y de complexité K de $G_\ell(m, d)$ satisfait*

$$|Y^{(1)}| + \mathrm{Red}(Y) \geq (1 - d - \varepsilon)\ell|Y|.$$

Le terme $\mathrm{Red}(Y)$ désigne le degré de réduction (Définition 4.1) du 2-complexe Y. Si le 2-complexe de van Kampen est réduit, nous pouvons omettre ce terme.

En plus, nous montrons la réciproque : si chaque sous-complexe d'un 2-complexe satisfait une inégalité donnée, alors avec une grande probabilité il est un 2-complexe sous-jacent d'un 2-complexe de van Kampen de $G_\ell(m,d)$. Nous avons en fait une transition de phase.

Théorème H (Théorème 4.15)**.** *Soit $s > 0$ et $K > 0$. Soit (Y_ℓ) une suite de 2-complexes de la même forme géométrique (Définition 4.13) de complexité K (Définition 4.2) telle que chaque face de Y_ℓ a une longueur de bord d'au plus ℓ.*

(i) Si chaque sous-2-complexe Y'_ℓ de Y_ℓ satisfait

$$|Y_\ell'^{(1)}| \geq (1-d+s)|Y'_\ell|\ell,$$

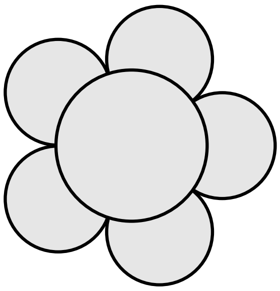
alors a.p.s. il existe un 2-complexe de van Kampen de $G_\ell(m,d)$ dont le 2-complexe sous-jacent est Y_ℓ.

(ii) S'il existe un sous 2-complexe Y'_ℓ de Y_ℓ satisfaisant

$$|Y_\ell'^{(1)}| \leq (1-d-s)|Y'_\ell|\ell,$$

alors a.p.s. il n'existe pas de 2-complexe de van Kampen de $G_\ell(m,d)$ dont le 2-complexe sous-jacent est Y_ℓ.

Rappelons qu'une présentation de groupe $G = \langle X|R\rangle$ satisfait la condition de petite simplification $C(p)$ (cf. [LS77]) si aucun relateur n'est un produit de moins de p pièces. C'est-à-dire qu'il n'existe pas de diagramme de van Kampen réduit de la forme suivante (ici $p = 5$).



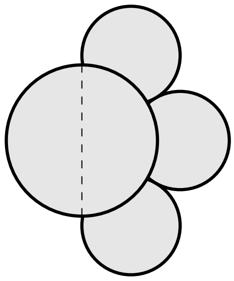
Comme application du théorème G et du théorème H, nous montrons la transition de phase pour la condition $C(p)$, mentionnée dans [OW11] Proposition 1.8 (avec seulement le cas $d < 1/(p+1)$).

Théorème I (Théorème 4.24)**.** *Soit $p \geq 2$ un entier. Il existe une transition de phase à densité $d = 1/(p+1)$:*

(i) Si $d < 1/(p+1)$, alors a.p.s. $G_\ell(m,d)$ satisfait $C(p)$.

(ii) Si $d > 1/(p+1)$, alors a.p.s. $G_\ell(m,d)$ ne satisfait pas $C(p)$.

Le même argument tient pour la condition $B(2p)$ (cf. [OW11] Définition 1.7 par Y. Ollivier et D. Wise) : la moitié d'un relateur ne peut pas être le produit de moins de p pièces ($p = 3$ dans le diagramme ci-dessous). Par des calculs similaires, on peut trouver qu'une transition de phase se réalise à densité $d = \frac{1}{2(p+1)}$.



0.3 Remarques historiques

0.3.1 Propriétés génériques des groupes de présentation finie

L'origine des groupes aléatoires provient d'un point de vue statistique : l'observation que certaines propriétés sont "génériques" pour les groupes de présentation finie. Elle est apparue pour la première fois dans les travaux de V. Guba [Gub86] et M. Gromov [Gro87]§0.2 à la fin des années 1980.

Guba [Gub86] a montré que pour "presque toute" présentation avec $m \geq 4$ générateurs et un relateur, tout sous-groupe de rang 2 est libre. Ici, "presque toute" signifie que dans l'ensemble des présentations avec $m \geq 4$ générateurs et un relateur, lorsque la longueur du relateur tend vers l'infini, le rapport entre le nombre de groupes satisfaisant la propriété demandée et le nombre de tous les groupes converge vers 1.

Le modèle de longueurs diverses construit dans [Gro87] §0.2 est en fait de ce point de vue, montrant que l'hyperbolicité est une "propriété générique" des groupes de présentation finie. Un modèle similaire est considéré par A. Ol'shanskii dans le Kourovka Notebook [MK22], question 11.75. Pour répondre à cette question, Arzhantseva et Ol'shanskii ont introduit le modèle à peu de relateurs dans [AO96], qui est un modèle plus simple mais plutôt commode.

L'étude des propriétés génériques des présentations de groupes (avec le modèle à peu de relateurs) est poursuivie par Arzhantseva dans les années suivantes : [Arz97; Arz98; Arz00]. Nous renvoyons le lecteur à [KS08] pour une revue.

0.3.2 Autres modèles de groupes aléatoires

La première construction aléatoire sur les présentations de groupes est le modèle à densité de Gromov dans [Gro93]. 9.B. De nombreux autres modèles de groupes aléatoires ont été proposés par Gromov dans [Gro93], [Gro00] et [Gro03].

Modèles de base non libres Les groupes aléatoires considérés jusqu'à présent sont des quotients aléatoires de groupes libres. Gromov a mentionné dans [Gro00] p.34 qu'on peut prendre une présentation de groupe hyperbolique comme groupe de base et ajouter des relateurs aléatoires d'une certaine densité. Ollivier [Oll04] a prouvé qu'il existe une transition de phase hyperbolicité-trivialité pour de tels groupes aléatoires, et que la densité critique est déterminée par la co-croissance du groupe hyperbolique de base. D. Gruber et J. Mackay [GM18] ont considéré des groupes triangulaires aléatoires basés sur un groupe de Burnside, et ont montré que le groupe est infini en dessous d'une certaine densité critique, mais une transition de phase n'a pas encore été trouvée.

Le modèle à température Dans [Gro00] p.34, ce modèle est la version à "présentation infinie" du modèle à densité : Un groupe aléatoire avec m générateurs et *tout* mot réduit r a la probabilité $p(r) = (2m-1)^{-\theta|r|}$

d'être dans l'ensemble des relateurs. Le paramètre θ est appelé la température. Notons F_m le groupe libre avec m générateurs, Gromov a affirmé qu'il existe une transition de phase trivialité-infinité: si la fonction $p : F_m \to \mathbb{R}$ est dans $\ell_2(F_m)$ alors le groupe est infini, sinon le groupe est trivial.

Le modèle à graphe Ce modèle est introduit dans [Gro03] p.141. Gromov a considéré le quotient d'un groupe libre F_m par le sous-groupe normal représenté par un graphe dont les arêtes sont marquées aléatoirement par les générateurs (au lieu d'un ensemble de relateurs comme dans [Gro93]). Les constructions et résultats détaillés de ce modèle sont discutés dans [AD08] par G. Arzhantseva et T. Delzant. Certaines questions ouvertes sur ce modèle sont proposées dans la section 5.3.

Pour des revues détaillées de groupes aléatoires, nous renvoyons le lecteur à (par ordre chronologique) : [Ghy04] par E. Ghys, [Oll05] par Y. Ollivier, [KS08] par I. Kapovich et P. Schupp, et [BNW20] par F. Bassino, C. Nicaud et P. Weil.

Chapter 1

Introduction

"I feel, random groups altogether may grow up as healthy as random graphs, for example."

—— M. Gromov, "Spaces and Questions" [Gro00].

The study of random groups is very similar to that of random graphs: we are interested in *asymptotic behaviors* and *phase transitions*. In this thesis, we study phase transitions in the density model of random groups introduced by M. Gromov in [Gro93]. There are two main objectives: to discover typical properties of group presentations for different densities of relators, and to construct new examples of hyperbolic groups.

Our main contributions are presented in Chapters 2, 3 and 4. In Chapter 2, we study the *intersection formula* for random subsets with density, which is a basic tool for proving phase transitions in random groups. In Chapter 3, we prove a phase transition showing that the first r generators of a random group presentation generate either the whole group or a free subgroup. In Chapter 4, we establish a phase transition that characterizes the existence of van Kampen diagrams. Some open questions are proposed in Chapter 5.

1.1 What is a random group?

1.1.1 Definitions and examples

Random groups are groups obtained by a probabilistic construction. In general, a random group can be defined as follows.

Definition 1. A random group is a random variable with values in a given set of groups.

We focus on random groups constructed from a finite set of finite group presentations. More precisely, fix a set of m generators $X_m = \{x_1, \dots, x_m\}$, a random group is defined by a random group presentation

$$G = \langle x_1, \dots, x_m | r_1, \dots, r_k \rangle$$

where $r_1, \dots, r_k$ are relators that are randomly chosen among a finite set of reduced words of $X_m^{\pm}$.

If G is a random group and P is a group property (for example, trivial, free, torsion free, etc.), we may discuss the probability that G satisfies P. We keep this "random variable" point of view for any random object. For example, the rank (the minimal number of generators) of a random group is an integer-valued random variable. All random variables considered in this thesis are with values in finite sets, endowed with the discrete σ-algebras.

Here are some examples of random groups.

Examples. In the following examples, a random group is defined by $G = \langle x_1, \dots, x_m | R\rangle$ where R is a random set of relators to be specified.

1. (fixed length, uniform model) Let k, ℓ be integers. The set of relators $R = \{r_1, \dots, r_k\}$ follows the uniform distribution in the set of all k-tuples of reduced words of length ℓ.

2. (various lengths, sampling model) Let k, ℓ, ℓ' be integers with $\ell < \ell'$. To construct R, choose one relator uniformly from the set of all reduced words of lengths between ℓ and ℓ', repeat this procedure k times.

   Note that one word may be taken twice, but it happens with a very low probability if the number of relators k is significantly smaller than the total number of considered words ($(2m-1)^{\ell'+O(1)}$ in this case).

3. (bounded length, Bernoulli model) Let ℓ be an integer, let $p \in [0,1]$. To construct the set of relators R, every reduced word r of length at most ℓ is independently chosen with probability p.

   Note that the number of relators follows the binomial distribution of $B(N,p)$ where $N = (2m-1)^{\ell+O(1)}$ is the total number of words of lengths at most ℓ.

We are interested in the *asymptotic behaviors* of a random group. That is to say, what happens when the number of generators or the number of relators goes to infinity? To describe this phenomenon briefly, we have the following definition.

Definition 2. Let $(Q_n)_{n\geq 1}$ be a sequence of probability events. We say that Q_n is satisfied *asymptotically almost surely*, denoted a.a.s. Q_n, if the probability of Q_n converges to 1 when n goes to infinity.

If (G_n) is a sequence of random groups and (P_n) is a sequence of group properties (for instance, being trivial, being hyperbolic, of rank n etc.), we may say that a.a.s. G_n has the property P_n. The index n may be the number of generators m or some relator length ℓ.

Example. let G_m be a random group with m generators and one relator uniformly chosen among all reduced words of length 3. If the number of generators m is very large, then with probability higher than $1 - m^{-2/3+o(1)}$ the chosen relator is of the type $r = xyz$ with x, y, z different elements in $X_m^{\pm}$. We can remove the generator z by the relation $z = x^{-1}y^{-1}$. So with probability converging to 1 when m tends to infinity, the random group G_m is a free group of rank $m-1$.

We conclude that a.a.s. the random group G_m is a free group of rank $m-1$.

1.1.2 Models of random groups and classic results

A *model* of random groups is a specified *sequence* of random groups. Throughout this thesis, we will clearly distinguish between a *random group* and a *sequence of random groups*.

Here is a list of random group models studied in the literature.

1.1.2.a The various lengths model, M. Gromov [Gro87] §0.2

Let $m \geq 2$, $k \geq 1$. Let $\ell_1, \dots, \ell_k$ be integers. A random group with m generators with k relators of various lengths is defined by $G(m, \ell_1, \dots, \ell_k) = \langle x_1, \dots, x_m | r_1, \dots, r_k\rangle$ where r_i is uniformly chosen among all reduced words of $x_1^{\pm}, \dots, x_m^{\pm}$ of length ℓ_i.

Suppose that $\ell_1 \leq \dots \leq \ell_k$. The asymptotic behavior is studied when the minimal length ℓ_1 goes to infinity. We may consider a sub-model that the maximal length ℓ_k depends on ℓ_1. For example, $\ell_k = 2\ell_1$ or $\ell_k = e^{10\ell_1}$. With this model, Gromov remarked that the hyperbolicity is *generic* for finitely presented groups:

Theorem (M. Gromov, [Gro87] §0.2)**.** *Let $G(m, \ell_1, \dots, \ell_k)$ be a random group with various relator lengths $\ell_1 \leq \dots \leq \ell_k$. A.a.s. (when the minimal length ℓ_1 goes to infinity) the random group $G(m, \ell_1, \dots, \ell_k)$ is a non-elementary hyperbolic group.*

A detailed proof of this theorem is given by C. Champetier in [Cha91].

1.1.2.b The density model, M. Gromov [Gro93] 9.B

Let S_ℓ be the set of reduced words of a fixed length ℓ in the alphabet $\{x_1, \dots, x_m\}$. Note that $|S_\ell| = (2m-1)^{\ell+O(1)}$ (when $\ell \to \infty$). A sequence of (uniform) random groups $(G_\ell(m, d))$ with $m \geq 2$ generators at density $d \in [0, 1]$ is defined by $G_\ell(m, d) = \langle x_1, \dots, x_m | R_\ell \rangle$ where R_ℓ is *uniformly* chosen among all subsets of S_ℓ of cardinality $\lfloor (2m-1)^{d\ell} \rfloor$.

Compare to the previous model, the length of relators is fixed, but the number of relators *grows exponentially* with the length. Gromov showed that there is a *phase transition* at density $d = 1/2$. A detailed proof of this theorem is in [Oll04] Section 2 by Y. Ollivier.

Theorem (M. Gromov, [Gro93] 9.B)**.** *Let $(G_\ell(m, d))$ be a sequence of (uniform) random groups at density d.*

- *If $d < 1/2$, then a.a.s. the random group $G_\ell(m, d)$ is non-elementary hyperbolic.*
- *If $d > 1/2$, then a.a.s. the random group $G_\ell(m, d)$ is either trivial or isomorphic to $\mathbb{Z}/2\mathbb{Z}$.*

In the same paragraph ([Gro93] 9.B), it is remarked that there is a phase transition for the $C'(\lambda)$ small cancellation condition (c.f. [LS77] for a definition).

Theorem (M. Gromov, [Gro93] 9.B)**.** *Let $(G_\ell(m, d))$ be a sequence of random groups at density d. Let $0 < \lambda < 1$.*

- *If $d < \lambda/2$, then a.a.s. the presentation defining $G_\ell(m, d)$ satisfies $C'(\lambda)$.*
- *If $d > \lambda/2$, then a.a.s. the presentation defining $G_\ell(m, d)$ does not satisfy $C'(\lambda)$.*

A detailed proof by an analog of the probabilistic pigeonhole principle is given in [BNW20] Theorem 2.1. In this thesis, we give a proof (Theorem 2.46) as an application of our Theorem C, and a much simpler one (Theorem 4.23) as an application of our Theorem H.

1.1.2.c The few relator model, G. Arzhantseva and A. Ol'shanskii [AO96]

Let $m \geq 2$ and $k \geq 1$. A sequence of random groups (G_ℓ) with m generators and k relators is defined by $G_\ell = \langle x_1, \dots, x_m | r_1, \dots, r_k \rangle$ where the set of relators $\{r_1, \dots, r_k\}$ follows the uniform distribution in the set of all k-tuples of reduced words of lengths at most ℓ.

Note that it is a simpler version of the various lengths model. Although relators of all lengths $\leq \ell$ are taken into account, for any $c < 1$ a.a.s. there is no relator shorter than $c\ell$. This model is an analog of the Gromov density model with density $d = 0$, and we have the following two properties.

- For any $\lambda > 0$, a.a.s. the presentation defining G_ℓ satisfies the $C'(\lambda)$ small cancellation condition.
- A.a.s. G_ℓ is a non-elementary hyperbolic group.

The main result in [AO96] is the following.

Theorem (G. Arzhantseva and A. Ol'shanskii, [AO96])**.** *Let (G_ℓ) be a sequence of few relator random groups with $m \geq 2$ generators and k relators. A.a.s. every $(m-1)$-generated subgroup of G_ℓ is free.*

1.1.2.d The triangular model, A. Żuk [Żuk03]

A triangular relator with respect to $X_m = \{x_1, \ldots, x_m\}$ is a cyclically reduced word of $X_m^\pm$ length 3. Note that the number of triangular relators is $(2m-1)^{3+o(1)}$ (with $m \to \infty$). A sequence of random triangular groups $(G_m(d))$ at density d is defined by $G_m(d) = \langle X_m | R_m \rangle$ where R_m is uniformly chosen among all sets of triangular relators of cardinality between $c^{-1}(2m-1)^{3d}$ and $c(2m-1)^{3d}$ with some $c > 1$.

Compare to the density model, the number of generators, instead of the relator lengths, goes to infinity. Żuk remarked that Gromov's phase transition at density $d = 1/2$ still holds in this model, and showed a phase transition at density $d = 1/3$.

Theorem (A. Żuk, [Żuk03]). *Let $G_m(d)$ be a triangular random group with density d.*

- *If $d < 1/2$ then a.a.s. $G_m(d)$ is non-elementary hyperbolic.*
- *If $d > 1/2$ then a.a.s. $G_m(d)$ is trivial.*

Theorem (A. Żuk, [Żuk03]). *Let $G_m(d)$ be a triangular random group with density d.*

- *If $d < 1/3$ then a.a.s. $G_m(d)$ is a free group.*
- *If $d > 1/3$ then a.a.s. $G_m(d)$ has Kazhdan's property (T).*

1.1.2.e The Bernoulli triangular model, S. Antoniuk, T. Łuczak and J. Świątkowski [AŁŚ15]

A sequence of (Bernoulli) random triangular groups $(G(m,p))_{m \geq 1}$ is defined by $G(m,p) = \langle X_m | R_m \rangle$ where R_m is a Bernoulli sampling in the set of triangular relators of X_m with probability $p = p(m)$ (every triangular relator is independently chosen with probability p).

If $p = (2m-1)^{3d-3}$, then the expected value of the number of relators is $(2m-1)^{3d+o(1)}$, and this model is very close to the Żuk's (uniform) triangular model. Antoniuk, Łuczak and Świątkowski adapted Żuk's result in their Bernoulli triangular model, and gave a sharper phase transition at density $d = 1/3$, i.e. with parameter $p \sim m^{-2}$.

Theorem (S. Antoniuk, T. Łuczak and J. Świątkowski, [AŁŚ15]). *Let $G(m,p)$ be a random triangular group. There exists positive real numbers $c < c', C' < C$ such that:*

- *If $p < cm^{-2}$, then a.a.s. $G_m(p)$ is a free group.*
- *If $c'm^{-2} < p < C'm^{-2} \log m$, then a.a.s. $G_m(p)$ is neither free nor having property (T).*
- *If $p > Cm^{-2} \log m$, then a.a.s. $G_m(p)$ has property (T).*

1.1.2.f The Bernoulli density model (a new example)

Let us mix up the three models in [Gro93], [AO96] and [AŁŚ15]. Let B_ℓ be the set of cyclically reduced relators of $X_m = \{x_1, \ldots, x_m\}$ of length at most ℓ. A sequence of (Bernoulli) random groups $(G_\ell(m,d))$ with m generators at density $d \in [0,1]$ is defined by $G_\ell(m,d) = \langle X_m | R_\ell \rangle$ where R_ℓ is a Bernoulli sampling of B_ℓ with parameter $|B_\ell|^{d-1}$.

The expected value of the number of relators is $|B_\ell|^d = (2m-1)^{d\ell+O(1)}$, and this model is very close to the original density model in [Gro93]. The following results are still true:

- Gromov's triviality-hyperbolicity phase transition at density $d = 1/2$.
- Gromov's $C'(\lambda)$ phase transition at density $d = \lambda/2$.

1.1.3 Main questions

In 2003, Gromov defined the general notion of random groups in [Gro03] and proposed in Section 1.9 the following general problem: "*determining asymptotic invariants and phase transition phenomena for random groups.*"

1.1.3.a Free subgroups

In particular, Gromov's problem [Gro03] 1.9 (iv) ask for "*existence/nonexistence of non-free subgroups*". Our aim is to generalize Arzhantseva-Ol'shanskii's "every $(m-1)$-generated subgroup is free" result in the Gromov density model. Let $G_\ell(m,d)$ be a sequence of random groups with m generators at density d, in the Gromov density model. We look for a phase transition:

Question 1. *Does there exist a critical density $d(m)$ such that, if $d < d(m)$ then a.a.s. every $(m-1)$-generated subgroup of $G_\ell(m,d)$ is free; while if $d > d(m)$ then a.a.s. there exists a non-free $(m-1)$-generated subgroup?*

More generally, we can ask:

Question 2. *Let $1 \leq r \leq m-1$. Does there exist a critical density $d(m,r)$ such that, if $d < d(m,r)$ then a.a.s. every r-generated subgroup of $G_\ell(m,d)$ is free; while if $d > d(m,r)$ then a.a.s. there exists a non-free r-generated subgroup?*

We will give partial answers to Question 2 (Theorem D and Theorem E).

1.1.3.b Van Kampen 2-complexes

In [Gro93], to prove the hyperbolicity of random groups at density $d < 1/2$, Gromov showed that every van Kampen diagram satisfies some isoperimetric inequality. More precisely, we have the following lemma.

Lemma (M. Gromov [Gro93], stated by Y. Ollivier [Oll04])**.** *Let $(G_\ell(m,d))$ be a sequence of random groups. If $d < 1/2$, then for any integer $K \geq 1$ and any real number $s > 0$, a.a.s. every reduced van Kampen diagram D of $G_\ell(m,d)$ with at most K faces satisfies the isoperimetric inequality*

$$|\partial D| \geq (1-2d-s)\,\ell|D|.$$

In other words, a.a.s. there is no van Kampen diagrams D of $G_\ell(m,d)$ with at most K faces satisfying the inequality $\frac{|\partial D|}{\ell|D|} < 1-2d$. We may ask the converse:

Question 3. *Let $c > 1-2d$ be a real number and let $K \geq 1$ be an integer. Does there exist a.a.s. a van Kampen diagram D of $G_\ell(m,d)$ with K faces that satisfies the equality $|\partial D| = c\ell|D|$?*

Theorem H gives an answer to this question in a more general case, for van Kampen 2-complexes.

1.1.4 Density of random subsets

Let S_ℓ be the set of all reduced words of length ℓ of the alphabet $\{x_1^\pm,\dots,x_m^\pm\}$. In [Gro93], to construct a random group $G = \langle x_1,\dots,x_m | R_\ell\rangle$, the original idea of Gromov is to choose a *random subset* R_ℓ of S_ℓ, with cardinality very close to $|S_\ell|^d$. The parameter d is called the *density* of R_ℓ in S_ℓ.

One may consider the uniform distribution in the set of subsets of cardinality $\lfloor |S_\ell|^d \rfloor$ (as in the works of Ollivier [Oll04; Oll05; Oll07]), or the Bernoulli sampling of parameter $|S_\ell|^{d-1}$ (as in [AŁŚ15], or the model 1.1.2.f). In this subsection, we introduce a random subset model, called the *permutation invariant* density model, originated in a paragraph of [Gro93] p.272.

Definition 3 ([Gro93] 9.A)**.** The density of a subset A in a finite set E is

$$\operatorname{dens}_E A := \log_{|E|}(|A|).$$

Namely, $\operatorname{dens}_E(A)$ is the number $d \in \{-\infty\} \cup [0,1]$ such that $|A| = |E|^d$. The case $d = -\infty$ corresponds to the case that A is an empty set. We omit the subscript and simply denote the density by $\operatorname{dens} A$ if there is no ambiguity of the ambiance set E.

A *random subset* of a finite set E is a random variable with values in the set of subsets of E.

The *density of a random subset* is then a random variable with values in $\{-\infty\} \cup [0,1]$.

The *intersection formula* for random subsets, introduced by Gromov in [Gro93], is the main tool for proving phase transitions in the density model of random groups. The formula is stated as follows:

Metatheorem ([Gro93] p.270)**.** *Random subsets A and B of a finite set E satisfy*

$$\operatorname{dens}(A \cap B) = \operatorname{dens} A + \operatorname{dens} B - 1$$

with the convention

$$\operatorname{dens}(A \cap B) < 0 \iff A \cap B = \emptyset.$$

Gromov did not specify how to take random subsets with densities in the original statement of the intersection formula. In [Gro93] p.272, he claimed that the class of random subsets that are *densable* and *permutation invariant* is closed under set theoretic operations and *satisfies the intersection formula*.

The definitions of *densable* and *permutation invariant* are given below. As we are interested in the asymptotic behaviors, we fix a sequence of finite sets $\boldsymbol{E} = (E_n)$ with $|E_n| \to \infty$, and study a sequence of random subsets $\boldsymbol{A} = (A_n)$ where A_n is a random subset of E_n.

Definition 4. A sequence of random subsets $\boldsymbol{A} = (A_n)$ of $\boldsymbol{E} = (E_n)$ is *densable* with density $d \in \{-\infty\} \cup [0,1]$ if the sequence of random variables $(\operatorname{dens}_{E_n}(A_n))$ converges in distribution to the constant d when n goes to infinity. We denote

$$\operatorname{dens}_{\boldsymbol{E}} \boldsymbol{A} = d.$$

Definition 5. A densable sequence of *permutation invariant* random subsets $\boldsymbol{A} = (A_n)$ of $\boldsymbol{E} = (E_n)$ is a densable sequence such that the measure of A_n is invariant under the permutations of E_n.

That is to say, for any subset $a \subset E_n$ and any permutation $\sigma \in \mathcal{S}(E_n)$, we have $\mathbf{Pr}(A_n = a) = \mathbf{Pr}(A_n = \sigma(a))$. A complete statement of the intersection formula for densable sequences of random subsets is given in Section 1.2, Theorem A.

1.1.5 The random group model considered in this thesis

Now with random subsets, we can define the model of random groups considered in this thesis: permutation invariant density models of random groups.

Fix an alphabet $X_m = \{x_1, \dots, x_m\}$ as generators of group presentations. In the original Gromov density model, the relator length of a random group $G_\ell(m,d)$ is a fixed number ℓ. In the case that ℓ is pair, when $d > 1/2$, a.a.s. $G_\ell(m,d)$ is $\mathbb{Z}/2\mathbb{Z}$ but not trivial. To avoid this situation, we consider words of length *at most* ℓ, as in [AO96]. Let B_ℓ be the set of cyclically reduced words of lengths at most ℓ.

Definition 6. A sequence of random groups $(G_\ell(m,d))$ with m generators of density d is defined by

$$G_\ell(m,d) = \langle X_m | R_\ell \rangle,$$

where (R_ℓ) is a densable sequence of permutation invariant random subsets of (B_ℓ) with density d.

Note that the cardinality of B_ℓ is $(2m-1)^{\ell+O(1)}$, so the cardinality of R_ℓ is $(2m-1)^{d\ell+o(\ell)}$ with high probability (c.f. Proposition 2.6). The model 1.1.2.f is included in this model, because a Bernoulli random subset is permutation invariant. The Arzhantseva-Ol'shanskii few relator model is included in this model with density 0, because the uniform distribution is permutation invariant, and the density $\mathrm{dens}_{B_\ell}(R_\ell)$ converges to 0 if the number of relators $|R_\ell| = k$ is fixed.

1.2 Main results of this thesis

The contributions of this thesis are presented in Chapters 2, 3 and 4. In this section, we present a list of our main results.

1.2.1 Density of random subsets and the intersection formula

In Chapter 2, we work on the intersection formula for random subsets, presented in the article [Tsa21a], to appear in the *Journal of Combinatorial Algebra*.

We first prove the intersection formula for densable sequences of permutation invariant random subsets.

Theorem A (Theorem 2.25)**.** *Let $\boldsymbol{A} = (A_n), \boldsymbol{B} = (B_n)$ be independent densable sequences of permutation invariant random subsets of a sequence of finite sets $\boldsymbol{E} = E_n$, with densities α, β. If $\alpha + \beta \neq 1$, then the sequence of random subsets $\boldsymbol{A} \cap \boldsymbol{B} := (A_n \cap B_n)$ is also densable and permutation invariant. In addition:*

$$\mathrm{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{if } \alpha + \beta > 1 \\ -\infty & \text{if } \alpha + \beta < 1. \end{cases}$$

Moreover, the intersection formula holds between a sequence of random subsets and a sequence of fixed subsets.

Theorem B (Theorem 2.28)**.** *Let $\boldsymbol{A}$ be a densable sequence of permutation invariant random subsets of $\boldsymbol{E}$ with density d. Let $\boldsymbol{X}$ be a sequence of (fixed) subsets of $\boldsymbol{E}$ with density α. If $d + \alpha \neq 1$, then the sequence of random subsets $\boldsymbol{A} \cap \boldsymbol{X}$ is densable and*

$$\mathrm{dens}(\boldsymbol{A} \cap \boldsymbol{X}) = \begin{cases} d + \alpha - 1 & \text{if } d + \alpha > 1 \\ -\infty & \text{if } d + \alpha < 1. \end{cases}$$

In addition, the sequence $\boldsymbol{A} \cap \boldsymbol{X}$ is a densable sequence of permutation invariant random subset of $\boldsymbol{X}$ with density $\frac{d+\alpha-1}{\alpha}$.

Note that $A_n \cap X_n$ is never invariant under the permutations of E_n if $X_n \neq E_n$.

We develop a generalized form: the *multi-dimensional intersection formula*. Denote $E_n^{(k)}$ the set of pairwise *distinct* k-tuples of the set E_n. Let $\boldsymbol{A}$ be a densable sequence of permutation invariant random subsets. We are interested in the intersection between $\boldsymbol{A}^{(k)}$ and a densable sequence of subsets $\boldsymbol{X}$ of $\boldsymbol{E}^{(k)}$.

For $k \geq 2$, the intersection formula is in general not correct (c.f. Counter example in 2.3.1.a). We show that by an additional *self-intersection condition* on $\boldsymbol{X}$, we can achieve the intersection formula.

Theorem C (Theorem 2.40)**.** *Let $\boldsymbol{A} = (A_n)$ be a densable sequence of permutation invariant random subsets of $\boldsymbol{E} = (E_n)$ with density $0 < d < 1$. Let $\boldsymbol{X} = (X_n)$ be a densable sequence of fixed subsets of $\boldsymbol{E}^{(k)}$ with density α.*

(i) If $d + \alpha < 1$, then a.a.s.

$$A_n^{(k)} \cap X_n = \emptyset.$$

(ii) If $d + \alpha > 1$ and $\boldsymbol{X}$ satisfies the d-small self intersection condition (Definition 2.33), then the sequence of random subsets $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and

$$\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = \alpha + d - 1.$$

As an application of Theorem B, we show that the main result of [AO96] for the few relator model of random groups can be extended to the density model of random groups with a small density.

Theorem D (Theorem 2.48)**.** *Let $(G_\ell(m, d))$ be a sequence of random groups with m generators at density*

$$0 \leq d < \frac{1}{120m^2 \ln(2m)}.$$

Then a.a.s. every $(m-1)$-generated subgroup of $G_\ell(m, d)$ is free.

This answers Question 1 partially.

1.2.2 Freiheitssatz for random groups and the phase transition

Chapter 3 is the subject of the preprint [Tsa21b]. We present one of the main results of this thesis, which answers Question 2 partially.

The *Freiheitssatz* (*freedom theorem* in German) is a fundamental theorem in combinatorial group theory. It was proposed by M. Dehn and proved by W. Magnus in his doctoral thesis [Mag30] in 1930 (c.f. [LS77] II.5).

Theorem (W. Magnus, [Mag30])**.** *Let $G = \langle x_1, \dots, x_m | r \rangle$ be a group presentation with m generators and one cyclically reduced relator. If the last generator x_m appears in the single relator r, then the first $m - 1$ generators $x_1, \dots, x_{m-1}$ freely generate a free subgroup of G.*

We say that a finite group presentation $G = \langle X | R \rangle$ satisfies the *Magnus Freiheitssatz property* if every subset of X of cardinality $|X| - 1$ freely generates a free subgroup of G. In particular, by Arzhantseva-Ol'shanskii's result [AO96], a.a.s. a few-relator random group G_ℓ has this property. We study the Magnus Freiheitssatz property in the density model of random groups.

Let $G_\ell(m, d) = \langle X | R_\ell \rangle$ be a random group at density d. For any $1 \leq r \leq m - 1$, we find a phase transition at density

$$d_r = \min \left\{ \frac{1}{2}, 1 - \log_{2m-1}(2r - 1) \right\}.$$

Theorem E (Theorem 3.22)**.** *Let $(G_\ell(m, d))$ be a sequence of random groups at density d.*

1. *If $d > d_r$, then a.a.s. $x_1, \dots, x_r$ generate the whole group $G_\ell(m, d)$.*
2. *If $d < d_r$, then a.a.s. $x_1, \dots, x_r$ freely generate a free subgroup of $G_\ell(m, d)$.*

By symmetry, the set $\{x_1, \dots, x_r\}$ can be replaced by any subset X_r of X of cardinality r. In fact, in the second assertion, we can replace the set $\{x_1, \dots, x_r\}$ by any set of r words of $X^{\pm}$ of lengths at most $\frac{d_r - d}{5r}\ell$. In particular, if $0 \leq d < d_{m-1}$, then the group presentation $G_\ell(m,d) = \langle X | R_\ell \rangle$ has the Magnus Freiheitssatz property.

More precisely for the first assertion, we prove that if $d > d_r$ then a.a.s. any generator x_k with $r < k \leq m$ equals to a reduced word of $\{x_1^{\pm} \dots x_r^{\pm}\}$ of length $\ell - 1$ in $G_\ell(m,d)$. Therefore, any relator $r_i \in R_\ell$ can be replaced by a reduced word r_i' of $\{x_1^{\pm} \dots x_r^{\pm}\}$ of length at most $\ell(\ell-1)$. Construct R_ℓ' by replacing every relator of R_ℓ in this way, we have the following corollary.

Corollary F. *If $d_r < d < d_{r-1}$, then a.a.s. the group $G_\ell(m,d) = \langle X | R_\ell \rangle$ admits a presentation with r generators $\langle X_r | R_\ell' \rangle$ satisfying the Magnus Freiheitssatz property (i.e. every subset of X_r of cardinality $r-1$ generate a free subgroup).*

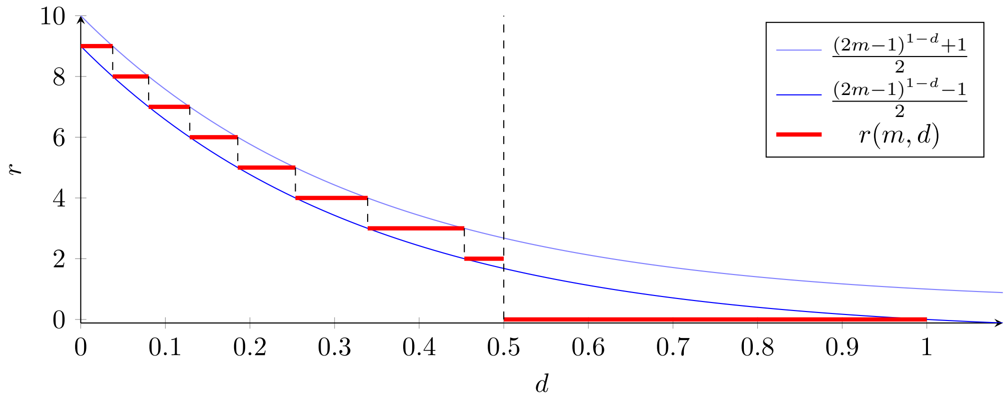
Remark. Note that R_ℓ' contains a bunch of relators of lengths varying from ℓ to ℓ^2. Such a presentation can not be studied using known methods in geometric or combinatorial group theory. Nevertheless, it gives us new examples of hyperbolic groups having the Magnus Freiheitssatz property.

If in addition $d_r > 1/3$ (i.e. $r < \frac{(2m-1)^{2/3}+1}{2}$), then by Żuk's result in [Żuk03], these are examples of hyperbolic groups with the property (T) having the Magnus Freiheitssatz property.

Let $r = r(m,d)$ be the maximal number such that a.a.s. $x_1, \dots, x_r$ freely generate a free subgroup of $G_\ell(m,d)$. By the phase transition at density $\frac{1}{2}$ [Gro93], if $d > \frac{1}{2}$, then $r(m,d) = 0$. If $d \leq \frac{1}{2}$, by **Theorem E**,

$$\frac{(2m-1)^{1-d} - 1}{2} \leq r(m,d) \leq \frac{(2m-1)^{1-d} + 1}{2}.$$

As shown in the diagram below, because $r(m,d)$ is an integer, its value is determined when d is not in the set $\{d_1, \dots, d_{m-1}, 1/2\}$. While the value of $r(m,d)$ is not determined when $d \in \{d_1, \dots, d_{m-1}, 1/2\}$.



$r(m,d)$ with $m = 10$

1.2.3 Existence of diagrams and 2-complexes in random groups

We consider van Kampen 2-complex with respect to a group presentation as van Kampen diagrams in [LS77]: it is a 2-complex with labels on edges by generators and labels on faces by relators. A pair of faces is called reducible if they have the same label and there is a common edge on their boundaries at the same position.

A van Kampen 2-complex is called reduced if there is no reducible pair of faces. C.f. Subsection 3.1.2 and Subsection 4.1 for more details.

In [Gro93], Gromov showed that a.a.s. local van Kampen diagrams of $G_\ell(m,d)$ satisfy an isoperimetric inequality (depending on the density d).

Theorem (M. Gromov [Gro93] p. 274, Y. Ollivier [Oll04] chapter 2)**.** *For any $s>0$ and $K>0$, a.a.s. every reduced van Kampen diagram D of $G_\ell(m,d)$ with $|D|\leq K$ satisfies the isoperimetric inequality*

$$|\partial D| \leq (1-2d-s)\ell|D|.$$

In Chapter 4, we prove a van Kampen 2-complex version of this inequality, which is an analog of Gruber-Mackay's [GM18] result for random triangular groups. For a 2-complex Y, denote $|Y^{(1)}|$ the number of its edges and $|Y|$ the number of its faces.

Theorem G (Theorem 4.3)**.** *Let $\varepsilon>0$, $K>0$. A.a.s. every van Kampen 2-complex Y of complexity K of $G_\ell(m,d)$ satisfies*

$$|Y^{(1)}| + \mathrm{Red}(Y) \geq (1-d-\varepsilon)\ell|Y|.$$

The term $\mathrm{Red}(Y)$ means the reduction degree (Definition 4.1) of the 2-complex Y. If the van Kampen 2-complex is reduced, we can omit this term.

More interestingly, we show the converse: if every sub-complex of a 2-complex satisfies a given inequality, then with high probability it is an underlying 2-complex of a van Kampen 2-complex of $G_\ell(m,d)$. We have in fact a phase transition.

Theorem H (Theorem 4.15)**.** *Let $s>0$ and $K>0$. Let (Y_ℓ) be a sequence of 2-complexes of the same geometrical form (Definition 4.13) of complexity K (Definition 4.2) such that every face of Y_ℓ is with boundary length at most ℓ.*

(i) If every sub-2-complex Y'_ℓ of Y_ℓ satisfies

$$|Y_\ell'^{(1)}| \geq (1-d+s)|Y'_\ell|\ell,$$

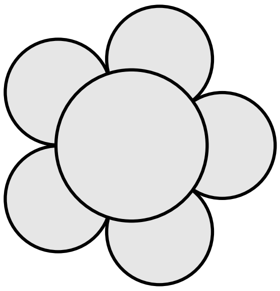
then a.a.s. there exists a van Kampen 2-complex of $G_\ell(m,d)$ whose underlying 2-complex is Y_ℓ.

(ii) If there is a sub-2-complex Y'_ℓ of Y_ℓ satisfying

$$|Y_\ell'^{(1)}| \leq (1-d-s)|Y'_\ell|\ell,$$

then a.a.s. there is no van Kampen 2-complex of $G_\ell(m,d)$ whose underlying 2-complex is Y_ℓ.

Recall that a group presentation $G=\langle X|R\rangle$ satisfies the $C(p)$ small cancellation condition (c.f. [LS77]) if no relator is a product of fewer than p pieces. That is to say, there is no reduced van Kampen diagram of the following form (here $p=5$).



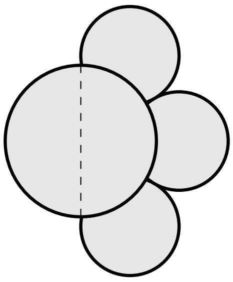
As an application of Theorem G and Theorem H, we show the phase transition for the $C(p)$ condition, mentioned in [OW11] Proposition 1.8 (with only $d < 1/(p+1)$ case).

Theorem I (Theorem 4.24). *Let $p \geq 2$ be an integer. There is a phase transition at density $d = 1/(p+1)$:*

(i) If $d < 1/(p+1)$, then a.a.s. $G_\ell(m,d)$ satisfies $C(p)$.

(ii) If $d > 1/(p+1)$, then a.a.s. $G_\ell(m,d)$ does not satisfy $C(p)$.

The same argument holds for the $B(2p)$ condition (c.f. [OW11] Definition 1.7 by Y. Ollivier and D. Wise): half of a relator can not be the product of fewer than p pieces ($p = 3$ in the diagram below). By similar computations, we can find that a phase transition occurs at density $d = \frac{1}{2(p+1)}$.



1.3 Historical remarks

1.3.1 Generic properties of finitely presented groups

The origin of random groups was from a statistical point of view: observations that some properties are "generic" for finitely presented groups. It first appeared in the works of V. Guba [Gub86] and M. Gromov [Gro87]§0.2 in the late 1980s.

Guba [Gub86] showed that “almost all” m-generated, $m \geq 4$, groups with one relator have every 2-generated subgroup free. Here "almost all" means that within the set of m-generated one relator groups, when the relator length tends to infinity, the ratio of the number of groups satisfying the demanded property to the number of all groups converges to 1.

The various lengths model constructed in [Gro87] §0.2 is actually in this point of view, showing that the hyperbolicity is a "generic property" of finitely presented groups. A similar model is considered by A. Ol’shanskii in the Kourovka Notebook [MK22], Question 11.75. To answer this question, Arzhantseva and Ol’shanskii introduced the few relator model in [AO96], which is a simpler but rather convenient model.

The study of generic properties of group presentations (with the few relator model) is continued by Arzhantseva in the following years [Arz97; Arz98; Arz00]. We refer the reader to [KS08] for a survey.

1.3.2 Other random group models

The first randomness construction on group presentations is the Gromov density model in [Gro93] 9.B. A lot more random group models were proposed by Gromov in [Gro93], [Gro00] and [Gro03].

Non-free basis models Random groups considered so far are random quotients of free groups. Gromov mentioned in [Gro00] p.34 that one can take a hyperbolic group presentation as a base group and add random relators with density. Ollivier [Oll04] proved that there is a hyperbolicity-triviality phase transition for such random groups, and the critical density is determined by the cogrowth of the base hyperbolic group. D. Gruber and J. Mackay [GM18] considered random triangular groups based on a Burnside group, and showed that the group is infinite below some critical density, but a phase transition has not been found yet.

The temperature model In [Gro00] p.34, this model is the infinite presentation version of the density model: A random group with m generators and *every* reduced word r has probability $p(r) = (2m-1)^{-\theta|r|}$ to be in the set of relators. The parameter θ is called temperature. Denoting F_m the free group with m generators, Gromov claimed that there is a triviality-infinity phase transition for $p : F_m \to \mathbb{R}$ in $\ell_2(F_m)$ or not.

The graphic model This model is introduced in [Gro03] p.141. Gromov considered the quotient of a free group F_m by the normal subgroup represented by a graph with edges randomly labeled by the generators (instead of a set of relators as in [Gro93]). Detailed constructions and results of this model are discussed in [AD08] by G. Arzhantseva and T. Delzant. Some open question about this model are proposed in Section 5.3.

For detailed surveys on random groups, we refer the reader to (in chronological order) [Ghy04] by E. Ghys, [Oll05] by Y. Ollivier, [KS08] by I. Kapovich and P. Schupp, and [BNW20] by F. Bassino, C. Nicaud and P. Weil.

Chapter 2

Random subsets with density and the intersection formula

In this chapter, we study the intersection formula for random subsets with density (Metatheorem 1.1.4) considered by M. Gromov in [Gro93] 9.A. We will mainly discuss the permutation invariant density models, with two particular cases: the Bernoulli density model and the uniform density model.

The goal is to prove the intersection formula between random subsets (Theorem A, Theorem 2.25) and the multidimensional intersection formula (Theorem C, Theorem 2.40). The random-fixed intersection formula (Theorem B) is regarded as a corollary (Theorem 2.28) of the multidimensional intersection formula.

For each theorem, we will first give a proof for the Bernoulli density model, as it is much easier to manipulate. We then adapt the proofs for the uniform density model. As any permutation invariant random subset can be decomposed into a sum of uniform random subsets (see Proposition 2.11), properties of the uniform density model are the main tools to study the permutation invariant density model.

2.1 Models of random subsets

In this section, we discuss models of random subsets in detail.

2.1.1 Examples of random subsets

Fix a finite set E. We will often use uppercase letters $A, B, C, \dots$ to denote *random* subsets and lowercase letters $a, b, c \dots$ for *fixed* subsets. Let us recall the definition of a random subset.

Definition 2.1. *A random subset A of a finite set E is a random variable with values in the set of subsets of E.*

The law of a random subset A is determined by instances $\mathbf{Pr}(A = a)$ through all subsets $a \subset E$ (or $a \in \mathcal{P}(E)$ where $\mathcal{P}(E)$ is the set of subsets of E). The cardinality $|A|$ is a usual real-valued random variable.

Here are some basic examples of random subsets.

Examples. (Examples of random subsets)

(i) (Dirac model) A fixed subset $c \subset E$ can be regarded as a constant random subset. Its law is

$$\mathbf{Pr}(A = a) = \begin{cases} 1 & \text{if } a = c \\ 0 & \text{if } a \neq c. \end{cases}$$

(ii) (Uniform random subset) Fix an integer $k \leq |E|$. A uniform random subset A of cardinality k is the uniform distribution on all subsets of E of cardinality k. Its law is

$$\mathbf{Pr}(A = a) = \begin{cases} \binom{|E|}{k}^{-1} & \text{if } |a| = k \\ 0 & \text{if } |a| \neq k. \end{cases}$$

The cardinality $|A|$ is then a constant random variable, equal to k.

(iii) (Bernoulli random subset) Let $p \in [0, 1]$. A Bernoulli random subset of parameter d is the Bernoulli sampling of parameter p on the set E: The events $\{x \in A\}$ through all $x \in E$ are independent of the same probability p. The law of A is

$$\mathbf{Pr}(A = a) = p^{|a|}(1 - p)^{|E|-|a|}.$$

In this case the random variable $|A|$ follows the binomial law $\mathcal{B}(|E|, p)$.

As usual random variables, a random subset can be constructed by other random subsets.

Examples. (Set theoretic operations) The intersection of two random subsets A, B of a finite set E is another random subset, defined by instances

$$\mathbf{Pr}(A \cap B = c) := \sum_{a,b \in \mathcal{P}(E); a \cap b = c} \mathbf{Pr}(A = a, B = b).$$

In particular, if A, B are *independent* random subsets, then

$$\mathbf{Pr}(A \cap B = c) = \sum_{a,b \in \mathcal{P}(E); a \cap b = c} \mathbf{Pr}(A = a)\mathbf{Pr}(B = b).$$

The union of two random subsets and the complement of a random subset can be similarly defined.

Definition 2.2. *Let E be a finite set.*

1. *Let a be a subset of E. The density of a in E is the number*

$$\operatorname{dens}_E a := \log_{|E|} |a| = \frac{\log |a|}{\log |E|}.$$

2. *Let A be a random subset of E. The density of A in E is the random variable*

$$\operatorname{dens}_E A := \log_{|E|} |A| = \frac{\log |A|}{\log |E|}.$$

That is to say, the density of a subset a of E is a number $d \in [0, 1] \cup \{-\infty\}$ such that $|E|^d = |a|$. We will omit the subscript if there is no ambiguity on the ambiance set E. Note that $\operatorname{dens} a = -\infty$ if and only if $a = \emptyset$.

2.1.2 Densable sequences of random subsets

We are interested in the asymptotic behavior of random subsets when the cardinality of the ambiance set $|E|$ tends to infinity. Consider a sequence of finite sets $\boldsymbol{E} = (E_n)_{n\in\mathbb{N}}$ with $|E_n| \xrightarrow[n\to\infty]{} \infty$.

Definition 2.3 (Densable sequence of random subsets)**.**

1. *A sequence of random subsets of $\boldsymbol{E}$ is a sequence $\boldsymbol{A} = (A_n)_{n\in\mathbb{N}}$ such that A_n is a random subset of E_n for all $n \in \mathbb{N}$.*
2. *Let $d \in \{-\infty\} \cup [0, 1]$. A sequence of random subsets $\boldsymbol{A}$ is **densable** with density d if the sequence of random valuables $\operatorname{dens}_{E_n}(A_n) = \log_{|E_n|} |A_n|$ converges in distribution to the constant d.*
3. *Two sequences of random subsets $\boldsymbol{A} = (A_n)$, $\boldsymbol{B} = (B_n)$ of $\boldsymbol{E}$ are independent if A_n, B_n are independent random subsets of E_n for all n.*

We are interested in asymptotic behaviors of a sequence of random groups. Recall the definition of a.a.s.

Definition 2.4. *Let $\boldsymbol{Q} = (Q_n)$ be a sequence of events. The event Q_n is **asymptotically almost surely** true if $\mathbf{Pr}(Q_n) \xrightarrow[n\to\infty]{} 1$. We denote briefly a.a.s. Q_n.*

The following proposition shows that one may study under the condition of some a.a.s. event.

Proposition 2.5. *Let $\boldsymbol{Q} = (Q_n)$, $\boldsymbol{R} = (R_n)$ be sequences of events. If a.a.s. Q_n and a.a.s. "R_n under the condition Q_n", then a.a.s. R_n.*

Proof. Denote by $\overline{Q_n}$ the complement of Q_n. By the two hypotheses, $\mathbf{Pr}(Q_n) \to 1$ and $\mathbf{Pr}(R_n \mid Q_n) \to 1$. Either $\overline{Q_n}$ is empty and $\mathbf{Pr}(R_n) = \mathbf{Pr}(R_n \mid Q_n) \to 1$, or by the formula of total probability

$$\mathbf{Pr}(R_n) = \mathbf{Pr}(Q_n)\mathbf{Pr}(R_n \mid Q_n) + \mathbf{Pr}(\overline{Q_n})\mathbf{Pr}\left(R_n \,\middle|\, \overline{Q_n}\right) \to 1.$$

□

Proposition 2.6 (Characterization of densability)**.** *Let $\boldsymbol{A}$ be a sequence of random subsets of $\boldsymbol{E}$.*

(i) The sequence $\boldsymbol{A}$ is densable with density $-\infty$ if and only if a.a.s. $A_n = \emptyset$

(ii) Let $d \geq 0$. The sequence $\boldsymbol{A}$ is densable with density d if and only if

$$\forall \varepsilon > 0 \text{ a.a.s. } |E_n|^{d-\varepsilon} \leq |A_n| \leq |E_n|^{d+\varepsilon}.$$

Proof.

(i) There is no density between $-\infty$ and 0, so $\operatorname{dens}_{E_n} A_n$ converges to $-\infty$ if and only if a.a.s. $\operatorname{dens}_{E_n}(A_n) = -\infty$, which means a.a.s. $A_n = \emptyset$.

(ii) For a sequence of random variables, the convergence in distribution to a constant is equivalent to the convergence in probability. So $\operatorname{dens}_{E_n}(A_n) = \log_{|E_n|} |A_n|$ converges in distribution to d if and only if it converges in probability to d, i.e.

$$\forall \varepsilon > 0 \quad \mathbf{Pr}(|\log_{|E_n|} |A_n| - d| \leq \varepsilon) \xrightarrow[n\to\infty]{} 1,$$

so

$$\forall \varepsilon > 0 \quad \mathbf{Pr}(|E_n|^{d-\varepsilon} \leq |A_n| \leq |E_n|^{d+\varepsilon}) \xrightarrow[n\to\infty]{} 1.$$

Which gives the assertion

$$\forall \varepsilon > 0 \text{ a.a.s. } |E_n|^{d-\varepsilon} \leq |A_n| \leq |E_n|^{d+\varepsilon}.$$

□

Remark. The terms "$\forall \varepsilon > 0$" and "a.a.s." can not be swapped. Because here "a.a.s." means "there exists n_0 such that for every $n \geq n_0$", and the number n_0 may depend on ε.

Here are some examples of densable sequences of random subsets.

Example.

(i) A sequence of fixed subsets $\boldsymbol{a} = (a_n)$ can be regarded as a sequence of random subsets (Dirac model on each term). The sequence $\boldsymbol{a}$ is densable with density $d \in \{-\infty\} \cup [0, 1]$ if the numerical sequence $\mathrm{dens}_{E_n}(a_n) = \log_{|E_n|} |a_n|$ converges to d (for example, $|a_n| = \lfloor |E_n|^d \rfloor$).

Note that $\mathrm{dens}(\boldsymbol{a}) = -\infty$ if and only if $a_n = \emptyset$ for every large enough n.

(ii) (Uniform density model) Let $d \leq 1$. A sequence of random subsets $\boldsymbol{A} = (A_n)$ of $\boldsymbol{E}$ is *a sequence of uniform random subsets with density* d if A_n is the uniform distribution on all subsets of E_n of cardinality $\lfloor |E_n|^d \rfloor$. The law of A_n is

$$\mathbf{Pr}(A_n = a) = \begin{cases} \binom{|E_n|}{\lfloor |E_n|^d \rfloor}^{-1} & \text{if } |a| = \lfloor |E_n|^d \rfloor \\ 0 & \text{if } |a| \neq \lfloor |E_n|^d \rfloor. \end{cases}$$

(iii) (Bernoulli density model) Let $d \leq 1$. A sequence of random subsets $\boldsymbol{A} = (A_n)$ of $\boldsymbol{E}$ is *a sequence of Bernoulli random subsets with density* d if A_n is a Bernoulli sampling of E_n with parameter $p = |E_n|^{d-1}$.

That is to say, every element $x \in E_n$ is taken independently with the same probability $p = |E_n|^{d-1}$. The law of A_n is

$$\mathbf{Pr}(A_n = a) = p^{|a|}(1-p)^{|E_n|-|a|}$$

with $p = |E_n|^{d-1}$.

Remark. It is not obvious that a sequence of Bernoulli random subsets is densable. In fact, a sequence of Bernoulli random subsets with density $d = 0$ is *not densable*.

Remind that the random variable $|A_n|$ follows the binomial law $\mathcal{B}(|E_n|, |E_n|^{d-1})$ and $\mathbb{E}(|A_n|) = |E_n|^d$. We have $\mathbf{Pr}(|A_n| = 0) = (1 - |E_n|^{-1})^{|E_n|} \xrightarrow[n\to\infty]{} 1/e$, so

$$\mathbf{Pr}(\mathrm{dens}\, A_n = -\infty) \xrightarrow[n\to\infty]{} 1/e.$$

Which means that the sequence of random variables $(\mathrm{dens}\, A_n)$ does not converge to any constant random variable.

Let us show that with $d \neq 0$, a sequence of Bernoulli random subsets is densable.

Proposition 2.7. *Let $\boldsymbol{A}$ be a sequence of Bernoulli random subsets with density d. If $d \neq 0$, then $\boldsymbol{A}$ is densable and:*

$$\operatorname{dens} \boldsymbol{A} = \begin{cases} d & \text{if } 0 < d \leq 1 \\ -\infty & \text{if } d < 0. \end{cases}$$

Proof. Let us separate the cases $d < 0$ and $0 < d \leq 1$.

(i) If $d < 0$, by Markov's inequality,

$$\mathbf{Pr}(|A_n| \geq 1) \leq |E_n|^d \xrightarrow[n\to\infty]{} 0.$$

Which means that $\mathbf{Pr}(|A_n| = 0) \xrightarrow[n\to\infty]{} 1$, or a.a.s. $A_n = \emptyset$. So by the characterization of density, $\boldsymbol{A}$ is $\operatorname{dens} \boldsymbol{A} = -\infty$.

(ii) If $0 < d \leq 1$, By Chebyshev's inequality,

$$\mathbf{Pr}\left(\big||A_n| - |E_n|^d\big| > \frac{1}{2}|E_n|^d\right) \leq \frac{\operatorname{Var}(|A_n|)}{\frac{1}{4}|E_n|^{2d}} \leq \frac{4|E_n|^d(1 - |E_n|^{d-1})}{|E_n|^{2d}} \xrightarrow[n\to\infty]{} 0.$$

So a.a.s.

$$\frac{1}{2}|E_n|^d \leq |A_n| \leq \frac{3}{2}|E_n|^d.$$

For every $\varepsilon > 0$, the inequality $|E_n|^{d-\varepsilon} < \frac{1}{2}|E_n|^d < \frac{3}{2}|E_n|^d < |E_n|^{d+\varepsilon}$ holds for n large enough, which gives

$$\forall \varepsilon > 0 \text{ a.a.s. } |E_n|^{d-\varepsilon} \leq |A_n| \leq |E_n|^{d+\varepsilon}.$$

By the characterization of densability (Proposition 2.6), the sequence of random subsets $\boldsymbol{A} = (A_n)$ is densable with density d

□

2.1.3 The permutation invariant density model

2.1.3.a Permutation invariant random subsets

Let E be a finite set with cardinality $|E| = n$. Denote $\mathcal{S}(E)$ as the group of permutations of E. The action of $\mathcal{S}(E)$ on E can be extended on the set of subsets $\mathcal{P}(E)$, defined by $\sigma(\{x_1, \ldots, x_k\}) := \{\sigma(x_1), \ldots, \sigma(x_k)\}$.

Note that this action has $(n+1)$ orbits of the form $\{a \in \mathcal{S}(E) \mid |a| = k\}$ for $k \in \{0, \ldots, n\}$. Moreover, the action commutes with set theoretic operations: for any permutation $\sigma \in \mathcal{S}(E)$, we have $\sigma(E\backslash a) = E\backslash\sigma(a)$, $\sigma(a \cap b) = \sigma(a) \cap \sigma(b)$ and $\sigma(a \cup b) = \sigma(a) \cup \sigma(b)$.

Definition 2.8 (Permutation invariant random subsets)**.** *Let A be a random subset of E. It is permutation invariant if its law is invariant under the permutations of E. i.e.*

$$\forall a \in \mathcal{P}(E)\ \forall \sigma \in \mathcal{S}(E) \quad \mathbf{Pr}(A = a) = \mathbf{Pr}(A = \sigma(a)).$$

Equivalently, subsets of E of the same cardinality have the same probability. That is to say, there exists real numbers $p_0, \ldots, p_n \in [0,1]$ satisfying

$$\sum_{k=0}^{n} \binom{n}{k} p_k = 1$$

such that

$$\forall a \in \mathcal{P}(E) \quad |a| = k \Rightarrow \mathbf{Pr}(A = a) = p_k.$$

Uniform random subsets and Bernoulli random subsets are permutation invariant. The class of such random subsets is closed under set theoretic operations.

Lemma 2.9 (Closed under set operations)**.** *Let E be a finite set. The class of permutation invariant random subsets of E is closed under unions, complements and intersections.*

Proof.

(i) (Complement) Let A be a permutation invariant random subset. Let $a \in \mathcal{P}(E)$ and $\sigma \in \mathcal{S}(E)$. Then

$$\begin{aligned}\mathbf{Pr}(E\backslash A = a) &= \mathbf{Pr}(A = E\backslash a) = \mathbf{Pr}(A = \sigma(E\backslash a)) \\ &= \mathbf{Pr}(A = E\backslash \sigma(a)) = \mathbf{Pr}(E\backslash A = \sigma(a)).\end{aligned}$$

(ii) (Intersection) Let A, B be independent permutation invariant random subsets. Then for $\sigma \in \mathcal{S}(E)$

$$\begin{aligned}\mathbf{Pr}(A \cap B = c) &= \sum_{a,b \in \mathcal{P}(E); a \cap b = c} \mathbf{Pr}(A = a)\mathbf{Pr}(B = b) \\ &= \sum_{a,b \in \mathcal{P}(E); \sigma(a) \cap \sigma(b) = \sigma(c)} \mathbf{Pr}(A = \sigma(a))\mathbf{Pr}(B = \sigma(b)) \\ &= \sum_{a',b' \in \mathcal{P}(E); a' \cap b' = \sigma(c)} \mathbf{Pr}(A = a')\mathbf{Pr}(B = b') \quad \text{(by substitution)} \\ &= \mathbf{Pr}(A \cap B = \sigma(c)).\end{aligned}$$

(iii) (Union) Let A, B be independent permutation invariant random subsets. We have $A \cup B = E\backslash((E\backslash A) \cap (E\backslash B))$, so $A \cup B$ is permutation invariant.

□

For a permutation invariant random subset A, let us express the expected value and the variance of the random variable $|A|$ in terms of $\mathbf{Pr}(x \in A)$ and $\mathbf{Pr}(x \in A, y \in A)$ where x, y are distinct elements in E.

Lemma 2.10. *Let E be a finite set with cardinality $|E| = n$. Let A be a permutation invariant random subset of E. Let x, y be distinct elements in E. Then*

(i) $\mathbb{E}(|A|) = n\mathbf{Pr}(x \in A)$,

(ii) $\mathrm{Var}(|A|) = \mathbb{E}(|A|) + n(n-1)\mathbf{Pr}(x \in A, y \in A) - \mathbb{E}(|A|)^2$.

Proof.

(i) By definition the probability $\mathbf{Pr}(z \in A)$ does not depend on the choice of element $z \in E$. So

$$\mathbb{E}(|A|) = \mathbb{E}\left(\sum_{z \in E} \mathbb{1}_{z \in A}\right) = \sum_{z \in E} \mathbf{Pr}(z \in A) = n\mathbf{Pr}(x \in A).$$

(ii) By the same argument, the probability $\mathbf{Pr}(z \in A, w \in A)$ does not depend on the choice of pair of distinct elements (z, w) in E. So

$$\begin{aligned}\mathbb{E}(|A|^2) &= \mathbb{E}\left[\left(\sum_{z\in E} \mathbb{1}_{z\in A}\right)^2\right] \\ &= \sum_{z\in E} \mathbf{Pr}(z \in A) + \sum_{(z,w)\in E^2; z\neq w} \mathbf{Pr}(z \in A, w \in A) \\ &= \mathbb{E}(|A|) + n(n-1)\mathbf{Pr}(x \in A, y \in A).\end{aligned}$$

□

A permutation invariant random subset can be decomposed into uniform random subsets. When proving properties for the permutation invariant model (Theorem 2.25 and Theorem 2.40), it is convenient to decompose a permutation invariant random subset into uniform random subsets and apply results of uniform random subsets.

Proposition 2.11 (Decomposition into uniform random subsets)**.** *Let E be a finite set with cardinality $|E| = n$. Let A be a permutation invariant random subset of E.*

1. *Let $k \leq n$ be an integer. If $\mathbf{Pr}(|A| = k) \neq 0$, then the random subset A under the condition $\{|A| = k\}$ is a uniform random subset of E of cardinality k.*
2. *Let Q be a probability event. Denote $\mathbb{N}_A = \{k \in \mathbb{N} \mid \mathbf{Pr}(|A| = k) \neq 0\}$. We have*

$$\mathbf{Pr}(Q) = \sum_{k\in\mathbb{N}_A} \mathbf{Pr}(Q \mid |A| = k)\mathbf{Pr}(|A| = k).$$

Proof. Let us prove the first assertion. Suppose that $\mathbf{Pr}(|A| = k) \neq 0$. Let $a \subset E$ with $|a| = k$. As A is permutation invariant, the $\binom{n}{k}$ subsets of E of cardinality k have the same probability. So

$$\mathbf{Pr}(|A| = k) = \binom{n}{k}\mathbf{Pr}(A = a).$$

Hence

$$\mathbf{Pr}\left(A = a \,|\, |A| = k\right) = \frac{\mathbf{Pr}(A = a)}{\mathbf{Pr}(|A| = k)} = \binom{n}{k}^{-1}.$$

If $|a| \neq k$, then $\mathbf{Pr}\left(A = a \,|\, |A| = k\right) = 0$. So the random subset A under the condition $\{|A| = k\}$ is a uniform random subset of cardinality k.

The second assertion is the formula of total probability. □

2.1.3.b Densable sequences of permutation invariant random subsets

Definition 2.12. *Let $\boldsymbol{A} = (A_n)$ be a sequence of random subsets of $\boldsymbol{E} = (E_n)$. It is a densable sequence of permutation invariant random subsets if it is densable and A_n is a permutation invariant random subset of E_n for all n.*

Examples.

1. A sequence of Bernoulli random subsets of $\boldsymbol{E}$ with density $d \neq 0$ is densable and permutation invariant.

2. A sequence of uniform random subsets of $\boldsymbol{E}$ with density $d \in -\infty \cup [0, 1]$ is densable and permutation invariant.

3. We will discuss another example in subsection 2.1.4, at the end of this section.

Except for some special cases (specified in the following propositions), the class of densable sequences of permutation invariant random subsets is closed under set theoretic operations. We prove here the closure under complements and unions. The proof for the closure under intersections is in the next section (Theorem 2.25, the intersection formula).

Proposition 2.13. *Let $\boldsymbol{A}, \boldsymbol{B}$ be densable sequences of permutation invariant random subsets with densities α, β. The union $\boldsymbol{A} \cup \boldsymbol{B}$ is densable and permutation invariant with density* $\operatorname{dens}(\boldsymbol{A} \cup \boldsymbol{B}) = \max(\alpha, \beta)$.

Proof. By Lemma 2.9 the sequence of random subsets $\boldsymbol{A} \cup \boldsymbol{B}$ is permutation invariant. The cases $\alpha = -\infty$ or $\beta = -\infty$ can be easily shown. Without loss of generality, assume that $\alpha \geq \beta \geq 0$.

Let $\varepsilon > 0$. By the densabilities of $\boldsymbol{A}$ and $\boldsymbol{B}$, a.a.s.

$$n^{\alpha - \varepsilon/2} \leq |A_n| \leq n^{\alpha + \varepsilon/2},$$

$$n^{\beta - \varepsilon/2} \leq |B_n| \leq n^{\beta + \varepsilon/2}.$$

Thus, a.a.s.

$$n^{\alpha - \varepsilon} \leq |A_n| \leq |A_n \cup B_n| \leq n^{\alpha + \varepsilon/2} + n^{\beta + \varepsilon/2} \leq 2n^{\alpha + \varepsilon/2} \leq n^{\alpha + \varepsilon}.$$

We conclude by the characterization of densability (Proposition 2.6). □

Proposition 2.14. *Let $\boldsymbol{A}$ be a densable sequence of permutation invariant random subsets with density $\alpha < 1$. Then the complement $\boldsymbol{E} \backslash \boldsymbol{A}$ is a densable sequence of permutation invariant random subsets and* $\operatorname{dens}(\boldsymbol{E} \backslash \boldsymbol{A}) = 1$.

Proof. Again by Lemma 2.9 the sequence of random subset $\boldsymbol{E} \backslash \boldsymbol{A}$ is permutation invariant.

Let $0 < \varepsilon < (1 - \alpha)/2$. By the densability of $\boldsymbol{A}$, a.a.s.

$$|A_n| \leq n^{\alpha + \varepsilon}.$$

As $n^{\alpha + \varepsilon} + n^{1 - \varepsilon} \leq n$ for n large enough, a.a.s.

$$|E_n \backslash A_n| \geq n - n^{\alpha + \varepsilon} \geq n^{1 - \varepsilon}.$$

We conclude by the characterization of densability (Proposition 2.6). □

Note that if $\operatorname{dens} \boldsymbol{A} = 1$, then $\boldsymbol{E} \backslash \boldsymbol{A}$ can be with any density. For example, if $\boldsymbol{A}$ is a densable sequence of random subsets with density $d < 1$, then $\boldsymbol{E} \backslash \boldsymbol{A}$ is with density 1, and $\boldsymbol{E} \backslash (\boldsymbol{E} \backslash \boldsymbol{A})$ is with density d. For another example, if $\boldsymbol{A} = (A_n)$ is a sequence of uniform random subsets defined by $|A_n| = \lfloor |E_n|/n \rfloor$, then $\operatorname{dens} \boldsymbol{A} = 1$ and $\operatorname{dens}(\boldsymbol{E} \backslash \boldsymbol{A}) = 1$.

2.1.4 Another model: random functions

We give here another natural model of random subsets: image of a random function. It can be found in [Gro93] p.270. This is also a variance of random groups considered by Ollivier in [Oll05] Lemma 59. In this subsection, we show that this model is densable and permutation invariant.

Definition 2.15. *Let E, F be finite subsets of cardinalities n, m. Denote E^F the set of functions from F to E. A random function Φ from F to E is a E^F-valued random variable.*

Let Φ be a random function from F to E. Its law is determined by the instances $\mathbf{Pr}(\Phi = \varphi)$ through all $\varphi \in E^F$.

The random function Φ can be regarded as a vector of E-valued random variables $(\Phi(y))_{y \in F}$ indexed by F. The image $\mathrm{Im}(\Phi) = \Phi(F) := \{\Phi(y) | y \in F\}$ is a random subset of E. Note that the random variables in a given vector are not necessarily independent.

Example. (Uniform random function) Let Φ be the uniform distribution on all functions from F to E. Its law is

$$\mathbf{Pr}(\Phi = \varphi) = \frac{1}{|E^F|} = \frac{1}{n^m}$$

through all $\varphi \in E^F$.

Proposition 2.16. *Let Φ be a uniform random function from F to E. Then the random elements $(\Phi(y))_{y \in F}$ are independent (identical) uniform distributions on E.*

Proof. Let $x \in E$, $y \in F$. The number of functions from F to E such that $\phi(y) = x$ is n^{m-1}. So the law of $\Phi(y)$ is

$$\mathbf{Pr}(\Phi(y) = x) = \frac{n^{m-1}}{n^m} = \frac{1}{n}.$$

Which is an uniform distribution on E.

Denote $F = \{y_1, \dots, y_m\}$. Let $(x_1, \dots, x_m)$ be a vector of m elements in E. Let $\varphi \in E^F$ such that $\varphi(y_i) = x_i$ for all $1 \leq i \leq m$. Then

$$\mathbf{Pr}\left(\bigwedge_{i=1}^{m} \Phi(y_i) = x_i\right) = \mathbf{Pr}(\Phi = \varphi) = \frac{1}{n^m} = \prod_{i=1}^{m} \mathbf{Pr}(\Phi(y_i) = x_i).$$

□

Proposition 2.17. *The image of a uniform random function is a permutation invariant random subset.*

Proof. Let Φ be a uniform random function from F to E. Let $\sigma \in \mathcal{S}(E)$, then for all $\varphi \in E^F$:

$$\mathbf{Pr}(\Phi = \varphi) = \mathbf{Pr}(\Phi = \sigma \circ \varphi) = \mathbf{Pr}(\sigma^{-1} \circ \Phi = \varphi).$$

The random function $\sigma^{-1} \circ \Phi$ has the same law of Φ. Now let $a \subset E$, we have

$$\mathbf{Pr}(\mathrm{Im}(\Phi) = a) = \mathbf{Pr}(\mathrm{Im}(\sigma^{-1} \circ \Phi) = a) = \mathbf{Pr}(\mathrm{Im}(\Phi) = \sigma(a)).$$

□

2.2 The intersection formula

In general, the intersection of two densable sequences can be not densable. The intersection formula (Metatheorem 1.1.4) is *not* satisfied by the class of densable sequences of random subsets. It is the reason that we need to introduce the permutation invariant density model of random subsets.

Here is an example.

Example. Let $\boldsymbol{E} = (E_n)$ be a sequence of sets with $|E_n| = 2n$. Let $\boldsymbol{a} = (a_n)$, $\boldsymbol{b} = (b_n)$ be sequences of subsets of $\boldsymbol{E}$ such that $b_n = E_n \backslash a_n$ and $|a_n| = |b_n| = n$. They are both densable subsets with density 1 because $\log(n)/\log(2n) \to 1$. Whereas $\operatorname{dens}(\boldsymbol{a} \cap \boldsymbol{b}) = -\infty$. They do not satisfy the intersection formula.

Define another sequence of subset $\boldsymbol{c} = (c_n)$ by $c_n := a_n$ if n is odd and $c_n := b_n$ if n is even. By its definition, $\boldsymbol{c}$ is densable with density 1. But the intersection $\boldsymbol{b} \cap \boldsymbol{c}$ is empty when n is odd and non-empty when n is even, so $\boldsymbol{b} \cap \boldsymbol{c}$ is not densable.

In this section, we prove the intersection formula for the Bernoulli density model, the uniform density model, and the permutation invariant density model. Throughout this section (and the next section), $\boldsymbol{E} = (E_n)$ is a sequence of finite sets with $|E_n| \to \infty$. To simplify, we assume that $|E_n| = n$. For an arbitrary sequence $\boldsymbol{E}$ with $|E_n| \to \infty$ we can proceed the same proofs by replacing n by $|E_n|$. Note that $|E_n|^d = n^d \sim \lfloor n^d \rfloor$ while $n \to \infty$ for $d \in [0, 1]$.

2.2.1 The Bernoulli density model

Recall that a sequence of Bernoulli random subsets with density d of $\boldsymbol{E} = (E_n)$ is a sequence of random subsets $\boldsymbol{A} = (A_n)$ such that A_n is a Bernoulli sampling of E_n with parameter $|E_n|^{d-1}$. The proof of the intersection formula for the Bernoulli density model is much easier than the proofs for the uniform density model and the permutation invariant density model.

Theorem 2.18 (The intersection formula for Bernoulli density model)**.** *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of Bernoulli random subsets of $\boldsymbol{E} = (E_n)$ with densities α, β. Then $\boldsymbol{A} \cap \boldsymbol{B}$ is a sequence of Bernoulli random subsets of $\boldsymbol{E}$ with density $\alpha + \beta - 1$, and*

$$\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{if } \alpha + \beta > 1 \\ -\infty & \text{if } \alpha + \beta < 1. \end{cases}$$

Proof. For every element $x \in E_n$, $\mathbf{Pr}(x \in A_n \cap B_n) = \mathbf{Pr}(x \in A_n)\mathbf{Pr}(x \in B_n) = |E_n|^{(\alpha+\beta-1)-1}$. In addition, for every pair of distinct elements x, y in E_n

$$\begin{aligned} \mathbf{Pr}(x, y \in A_n \cap B_n) &= \mathbf{Pr}(x, y \in A_n)\mathbf{Pr}(x, y \in B_n) \\ &= \mathbf{Pr}(x \in A_n)\mathbf{Pr}(y \in A_n)\mathbf{Pr}(x \in B_n)\mathbf{Pr}(y \in B_n) \\ &= \mathbf{Pr}(x \in A_n \cap B_n)\mathbf{Pr}(y \in A_n \cap B_n). \end{aligned}$$

So $\boldsymbol{A} \cap \boldsymbol{B}$ is a sequence of Bernoulli random subsets with density $\alpha + \beta - 1$. Proposition 2.7 gives its density. □

Note that by the remark after Proposition 2.7, if $\alpha + \beta = 1$ then the sequence $\boldsymbol{A} \cap \boldsymbol{B}$ is not densable.

As Theorem 2.18 shows, the class of Bernoulli random subsets is *closed under intersections*. Thereby the intersection formula works for multiple independent sequences of random subsets. The formula is more concise in terms of *codensities*.

Definition 2.19 (c.f. [Gro93] p.269). *Let $\boldsymbol{A}$ be a densable sequence of random subsets. The* ***codensity*** *of $\boldsymbol{A}$ is defined by*

$$\operatorname{codens} \boldsymbol{A} = 1 - \operatorname{dens} \boldsymbol{A}.$$

In particular, if $\operatorname{dens} \boldsymbol{A} = -\infty$, then $\operatorname{codens} \boldsymbol{A} = \infty$.

Theorem 2.18 can be rephrased as (see [Gro93] p.270):

Corollary 2.20 (The intersection formula by codensities). *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of Bernoulli random subsets of $\boldsymbol{E}$ with positive densities.*

(i) If $\operatorname{codens} \boldsymbol{A} + \operatorname{codens} \boldsymbol{B} < 1$, *then*

$$\operatorname{codens}(\boldsymbol{A} \cap \boldsymbol{B}) = \operatorname{codens} \boldsymbol{A} + \operatorname{codens} \boldsymbol{B}.$$

(ii) If $\operatorname{codens} \boldsymbol{A} + \operatorname{codens} \boldsymbol{B} > 1$, *then* $\operatorname{codens}(\boldsymbol{A} \cap \boldsymbol{B}) = \infty$.

□

Now we can express the codensity of the intersection of several sequences of random subsets. Note that the following corollary holds because the Bernoulli density model is closed under intersections.

Corollary 2.21 (Generalized intersection formula by codensities). *Let $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k$ be independent sequences of Bernoulli random subsets with positive densities.*

(i) If $\displaystyle\sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i < 1$, *then*

$$\operatorname{codens}\left(\bigcap_{i=1}^{k} \boldsymbol{A}_i\right) = \sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i.$$

(ii) If $\displaystyle\sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i > 1$, *then* $\operatorname{codens}\left(\bigcap_{i=1}^{k} \boldsymbol{A}_i\right) = \infty$.

□

2.2.2 The uniform density model

Recall that a sequence of uniform random subsets of $\boldsymbol{E} = (E_n)$ with density d is a sequence of random subsets $\boldsymbol{A} = (A_n)$ with the following law:

$$\mathbf{Pr}(A_n = a) = \begin{cases} \binom{n}{\lfloor n^d \rfloor}^{-1} & \text{if } |a| = \lfloor n^d \rfloor \\ 0 & \text{if } |a| \neq \lfloor n^d \rfloor. \end{cases}$$

Theorem 2.22 (The intersection formula for the uniform density model). *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of uniform random subsets of $\boldsymbol{E}$ with densities α, β. If $\alpha + \beta \neq 1$, then the sequence $\boldsymbol{A} \cap \boldsymbol{B}$ is densable and*

$$\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{if } \alpha + \beta > 1 \\ -\infty & \text{if } \alpha + \beta < 1. \end{cases}$$

Let us first estimate the expected value and the variance of the random variable $|A_n \cap B_n|$.

Lemma 2.23. *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of uniform random subsets of $\boldsymbol{E}$ with densities $\alpha, \beta \in [0, 1]$.*

(i) $n^{\alpha+\beta-1} - 2 \leq \mathbb{E}(|A_n \cap B_n|) \leq n^{\alpha+\beta-1}$.

(ii) *If $\alpha < 1$ and $\beta < 1$, then* $\mathrm{Var}(|A_n \cap B_n|) \sim n^{\alpha+\beta-1}$.

Moreover, if $n \geq 3$, then $\mathrm{Var}(|A_n \cap B_n|) \leq 3n^{\alpha+\beta-1}$.

Proof.

(i) By Lemma 2.9, as uniform random subsets are permutation invariant, $A_n \cap B_n$ is a permutation invariant random subset of E_n. Apply Lemma 2.10 (i) on $A_n \cap B_n$, A_n and B_n,

$$\begin{aligned}\mathbb{E}(|A_n \cap B_n|) &= n\mathbf{Pr}(x \in A_n \cap B_n) = n\mathbf{Pr}(x \in A_n)\mathbf{Pr}(x \in B_n) \\ &= n\frac{\mathbb{E}(|A_n|)}{n}\frac{\mathbb{E}(|B_n|)}{n} = \lfloor n^\alpha \rfloor \lfloor n^\beta \rfloor n^{-1}.\end{aligned}$$

Because $\alpha, \beta \leq 1$, we have

$$n^{\alpha+\beta-1} - 2 \leq n^{\alpha+\beta-1} - n^{\alpha-1} - n^{\beta-1} + n^{-1} \leq \lfloor n^\alpha \rfloor \lfloor n^\beta \rfloor n^{-1} \leq n^{\alpha+\beta-1}.$$

(ii) Let x, y be distinct elements in E. The number of subsets of E containing x, y of cardinality $\lfloor n^\alpha \rfloor$ is $\binom{n-2}{\lfloor n^\alpha \rfloor - 2}$, so

$$\mathbf{Pr}(x, y \in A_n) = \frac{\binom{n-2}{\lfloor n^\alpha \rfloor - 2}}{\binom{n}{\lfloor n^\alpha \rfloor}} = \frac{\lfloor n^\alpha \rfloor(\lfloor n^\alpha \rfloor - 1)}{n(n-1)}.$$

Similarly,

$$\mathbf{Pr}(x, y \in B_n) = \frac{\lfloor n^\beta \rfloor(\lfloor n^\beta \rfloor - 1)}{n(n-1)}.$$

Denote $k = \lfloor n^\alpha \rfloor$ and $l = \lfloor n^\beta \rfloor$ to simplify the notation. Note that $k = o(n)$ and $l = o(n)$ as $\alpha < 1$ and $\beta < 1$. Apply Lemma 2.10 (ii) on $|A_n \cap B_n|$. By the proof of (i), $\mathbb{E}(|A_n \cap B_n|) = kln^{-1}$. We have

$$\begin{aligned}\mathrm{Var}(|A_n \cap B_n|) &= kln^{-1} + n(n-1)\mathbf{Pr}(x, y \in A_n)\mathbf{Pr}(x, y \in B_n) - (kln^{-1})^2 \\ &= kln^{-1} + \frac{k(k-1)l(l-1)}{n(n-1)} - (kln^{-1})^2 \\ &= \frac{kl}{n^2(n-1)}(n^2 - n + nkl - nl - nk + n - nkl + kl) \\ &= \frac{kl}{n^2(n-1)}(n^2 - nl - nk + kl) \\ &\sim \frac{kl}{n^2(n-1)} \cdot n^2 \sim n^{\alpha+\beta-1}.\end{aligned}$$

Moreover, if $n \geq 3$, then

$$\mathrm{Var}(|A_n \cap B_n|) = \frac{kl}{n^2(n-1)}(n^2 - nl - nk + kl) \leq \frac{2kl}{n-1} \leq \frac{2n^{\alpha+\beta}}{n-1} \leq 3n^{\alpha+\beta-1}.$$

□

Lemma 2.24 (The concentration lemma)**.** *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of uniform random subsets of $\boldsymbol{E}$ with densities $\alpha, \beta \in [0,1]$. If $\alpha+\beta-1>0$ and $n \geq 8^{\frac{1}{\alpha+\beta-1}}$, then*

$$\mathbf{Pr}\left(\left||A_n \cap B_n| - n^{\alpha+\beta-1}\right| > \frac{1}{2}n^{\alpha+\beta-1}\right) \leq \frac{48}{n^{\alpha+\beta-1}} \xrightarrow[n\to\infty]{} 0.$$

In particular,

$$\text{a.a.s. } \left||A_n \cap B_n| - n^{\alpha+\beta-1}\right| \leq \frac{1}{2}n^{\alpha+\beta-1}.$$

Proof. By Lemma 2.23 (i) with $n \geq 8^{\frac{1}{\alpha+\beta-1}} \geq 4$, we have

$$\left|\mathbb{E}(|A_n \cap B_n|) - n^{\alpha+\beta-1}\right| \leq \frac{1}{4}n^{\alpha+\beta-1}.$$

If $\alpha = 1$ or $\beta = 1$, then the result is true as the $A_n = E_n$ or $B_n = E_n$. Otherwise, by Lemma 2.23 (ii) and Chebyshev's inequality, with $n \geq 8^{\frac{1}{\alpha+\beta-1}}$, we have

$$\begin{aligned}
&\mathbf{Pr}\left(\left||A_n \cap B_n| - n^{\alpha+\beta-1}\right| > \frac{1}{2}n^{\alpha+\beta-1}\right) \\
&\leq \mathbf{Pr}\left(||A_n \cap B_n| - \mathbb{E}(|A_n \cap B_n|)| > \frac{1}{4}n^{\alpha+\beta-1}\right) \\
&\leq \frac{16\,\mathrm{Var}(|A_n \cap B_n|)}{n^{2\alpha+2\beta-2}} \leq \frac{48}{n^{\alpha+\beta-1}}.
\end{aligned}$$

□

Remark. Some constants in this lemma ($8^{\frac{1}{\alpha+\beta-1}}$ and $\frac{48}{n^{\alpha+\beta-1}}$) are not useful for proving the intersection formula for the uniform density model, but will be used to prove the intersection formula for the permutation invariant density model (see Lemma 2.26).

Remark. The concentration lemma shows that the cardinality of $A_n \cap B_n$ is close to $n^{\alpha+\beta-1}$ with high probability, but not with probability 1. If $\alpha \neq 1$ and $\beta \neq 1$, then for n large enough $\lfloor n^\alpha \rfloor + \lfloor n^\beta \rfloor < n$, so $\mathbf{Pr}(A_n \cap B_n = \emptyset) \neq 0$. Which means that $\boldsymbol{A} \cap \boldsymbol{B}$ is not a sequence of uniform random subsets. As the class of sequences of uniform random subsets is *not* closed under intersections, Corollary 2.21 can not be applied for the uniform density model.

Now we can prove the intersection formula for the uniform density model.

Proof of Theorem 2.22.

(i) If $\alpha+\beta<1$, then by Markov's inequality and Lemma 2.23 (i),

$$\mathbf{Pr}(|A_n \cap B_n| \geq 1) \leq \mathbb{E}(|A_n \cap B_n|) \xrightarrow[n\to\infty]{} 0,$$

which implies a.a.s. $A_n \cap B_n = \emptyset$, so $\mathrm{dens}(A \cap B) = -\infty$.

(ii) If $\alpha + \beta > 1$, by Lemma 2.24,

$$\text{a.a.s. } \left||A_n \cap B_n| - n^{\alpha+\beta-1}\right| \leq \frac{1}{2} n^{\alpha+\beta-1}.$$

By the argument proving the densability for the Bernoulli density model (Proposition 2.7),

$$\forall \varepsilon > 0 \text{ a.a.s. } n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}.$$

By the characterization of densability (Proposition 2.6), the sequence of random subsets $(A_n \cap B_n)$ is densable with density $\alpha + \beta - 1$.

□

2.2.3 The permutation invariant model

Here is our Theorem A in Introduction.

Theorem 2.25 (The intersection formula)**.** *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of permutation invariant random subsets with densities α, β. If $\alpha + \beta \neq 1$, then the sequence $\boldsymbol{A} \cap \boldsymbol{B}$ is densable and permutation invariant, with density*

$$\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{if } \alpha + \beta > 1 \\ -\infty & \text{if } \alpha + \beta < 1. \end{cases}$$

Lemma 2.26. *Let $\alpha, \beta \in [0, 1]$ such that $\alpha + \beta > 1$. Let $0 < \varepsilon < \alpha + \beta - 1$. Let $\boldsymbol{A}, \boldsymbol{B}$ independent sequences of uniform random subsets of $\boldsymbol{E}$ with densities α', β' with $\alpha' \in [\alpha - \varepsilon/3, \alpha + \varepsilon/3]$ and $\beta' \in [\beta - \varepsilon/3, \beta + \varepsilon/3]$. If $n \geq \max\left\{2^{3/\varepsilon}, 8^{1/(\alpha+\beta-1-\varepsilon)}\right\}$, then*

$$\mathbf{Pr}\left(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}\right) \geq 1 - \frac{48}{n^{\alpha+\beta-1-\varepsilon}} \xrightarrow[n\to\infty]{} 1.$$

Proof. By hypothesis $\alpha' + \beta' - 1 \geq \alpha + \beta - 2\varepsilon/3 - 1 > 0$. Apply Lemma 2.24 (iii), if $n \geq 8^{1/(\alpha+\beta-1-\varepsilon)} \geq 8^{1/(\alpha'+\beta'-1)}$, then

$$\mathbf{Pr}\left(\left||A_n \cap B_n| - n^{\alpha'+\beta'-1}\right| \geq \frac{1}{2} n^{\alpha'+\beta'-1}\right) \leq \frac{48}{n^{\alpha'+\beta'-1}}.$$

This can be rewritten as

$$\mathbf{Pr}\left(\frac{1}{2} n^{\alpha'+\beta'-1} < |A_n \cap B_n| < \frac{3}{2} n^{\alpha'+\beta'-1}\right) > 1 - \frac{48}{n^{\alpha'+\beta'-1}}.$$

By hypothesis $\alpha + \beta - 1 - 2\varepsilon/3 \leq \alpha' + \beta' - 1 \leq \alpha + \beta - 1 + 2\varepsilon/3$. If $n \geq 2^{3/\varepsilon}$, then

$$n^{\alpha+\beta-1-\varepsilon} \leq \frac{1}{2} n^{\alpha+\beta-1-2\varepsilon/3} \leq \frac{3}{2} n^{\alpha+\beta-1+2\varepsilon/3} \leq n^{\alpha+\beta-1+\varepsilon}.$$

Hence, for $n \geq 2^{3/\varepsilon}$,

$$\begin{aligned} &\mathbf{Pr}\left(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}\right) \\ \geq &\mathbf{Pr}\left(\frac{1}{2} n^{\alpha+\beta-1-2\varepsilon/3} \leq |A_n \cap B_n| \leq \frac{3}{2} n^{\alpha+\beta-1+2\varepsilon/3}\right) \\ \geq &\mathbf{Pr}\left(\frac{1}{2} n^{\alpha'+\beta'-1} < |A_n \cap B_n| < \frac{3}{2} n^{\alpha'+\beta'-1}\right). \end{aligned}$$

Combine two estimations on n. If $n \geq \max\left\{2^{3/\varepsilon}, 8^{1/(\alpha+\beta-1-\varepsilon)}\right\}$, then

$$\mathbf{Pr}\left(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}\right) \geq 1 - \frac{48}{n^{\alpha'+\beta'-1}} \geq 1 - \frac{48}{n^{\alpha+\beta-1-\varepsilon}}.$$

As $\alpha + \beta - 1 - \varepsilon > 0$,

$$\frac{48}{n^{\alpha+\beta-1-\varepsilon}} \xrightarrow[n\to\infty]{} 0.$$

□

Now we can prove the intersection formula for the permutation invariant model. The idea is to decompose permutation invariant random subsets A_n and B_n into uniform random subsets, so that we can apply Lemma 2.26.

Proof of Theorem 2.25. By Lemma 2.9 the intersection $\boldsymbol{A} \cap \boldsymbol{B}$ is a sequence of permutation invariant random subsets. Denote (Q_n) the sequence of events defined by

$$Q_n = \{n^{\alpha-\varepsilon/3} \leq |A_n| \leq n^{\alpha+\varepsilon/3} \text{ and } n^{\beta-\varepsilon/3} \leq |B_n| \leq n^{\beta+\varepsilon/3}\}$$

for some small $\varepsilon > 0$ to be specified. By Proposition 2.6 and the densabilities of $\boldsymbol{A}$ and $\boldsymbol{B}$, a.a.s. Q_n is true. Note that Q_n is a union of events of type $\{|A_n| = k, |B_n| = l\}$. Denote

$$\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon} := \Big\{(k,l) \in \mathbb{N}^2 \,\Big|\, n^{\alpha-\varepsilon/3} \leq k \leq n^{\alpha+\varepsilon/3}, n^{\beta-\varepsilon/3} \leq l \leq n^{\beta+\varepsilon/3} \text{ and } \mathbf{Pr}(|A_n| = k, |B_n| = l) \neq 0\Big\}.$$

For $(k,l) \in \mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}$, by a change of variables $k = n^{\alpha'}$, $l = n^{\beta'}$, we have

$$\alpha - \varepsilon/3 \leq \alpha' \leq \alpha + \varepsilon/3 \text{ and } \beta - \varepsilon/3 \leq \beta' \leq \beta + \varepsilon/3.$$

(i) Suppose that $\alpha + \beta < 1$. Let $0 < \varepsilon < 1 - \alpha - \beta$. We shall prove that a.a.s. $A_n \cap B_n = \emptyset$.

By the formula of total probability and Markov's inequality,

$$\begin{aligned}\mathbf{Pr}(A_n \cap B_n \neq \emptyset) &\leq \mathbf{Pr}\left(|A_n \cap B_n| \geq 1 \,|\, Q_n\right)\mathbf{Pr}(Q_n) + \mathbf{Pr}(\overline{Q_n}) \\ &\leq \sum_{(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}} \Big[\mathbf{Pr}\left(|A_n \cap B_n| \geq 1 \,|\, |A_n| = k, |B_n| = l\right) \\ &\qquad \mathbf{Pr}(|A_n| = k, |B_n| = l|)\Big] + \mathbf{Pr}(\overline{Q_n}). \\ &\leq \sum_{(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}} \Big[\mathbb{E}\left(|A_n \cap B_n| \,|\, |A_n| = k, |B_n| = l\right) \\ &\qquad \mathbf{Pr}(|A_n| = k, |B_n| = l|)\Big] + \mathbf{Pr}(\overline{Q_n}).\end{aligned}$$

For any $(k,l) \in \mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}$, by Lemma 2.23 (i)

$$\begin{aligned}\mathbb{E}\left(|A_n \cap B_n| \,|\, |A_n| = k, |B_n| = l\right) &= \mathbb{E}\left(|A_n \cap B_n| \,\Big|\, |A_n| = n^{\alpha'}, |B_n| = n^{\beta'}\right) \\ &\leq n^{\alpha'+\beta'-1} \leq n^{\alpha+\beta+2\varepsilon/3-1} \leq n^{-\varepsilon/3}.\end{aligned}$$

Hence

$$\mathbf{Pr}(A_n \cap B_n \neq \emptyset) \leq n^{-\varepsilon/3}\mathbf{Pr}(Q_n) + \mathbf{Pr}(\overline{Q_n}) \xrightarrow[n\to\infty]{} 0.$$

(ii) Suppose that $\alpha+\beta>1$. Let $0<\varepsilon<\alpha+\beta-1$. We shall prove that a.a.s.

$$n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon}.$$

By the formula of total probability,

$$\begin{aligned}
&\mathbf{Pr}(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon}) \\
&\ge \mathbf{Pr}\left(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon} \,\middle|\, Q_n\right)\mathbf{Pr}(Q_n) \\
&= \sum_{(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}} \Big[\mathbf{Pr}\left(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon} \,\middle|\, |A_n|=k, |B_n|=l\right) \\
&\qquad\qquad \mathbf{Pr}(|A_n|=k, |B_n|=l|)\Big].
\end{aligned}$$

Apply Lemma 2.26 and Proposition 2.11 (decomposition into uniform random subsets). If n is large enough ($n \ge \max\left\{2^{3/\varepsilon}, 8^{1/(\alpha+\beta-1-\varepsilon)}\right\}$), then for any $(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}$,

$$\begin{aligned}
&\mathbf{Pr}\left(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon} \,\middle|\, |A_n|=k, |B_n|=l\right) \\
&= \mathbf{Pr}\left(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon} \,\middle|\, |A_n|=n^{\alpha'}, |B_n|=n^{\beta'}\right) \\
&\ge 1-\frac{48}{n^{\alpha+\beta-1+\varepsilon}} \xrightarrow[n\to\infty]{} 1.
\end{aligned}$$

Hence, for n large enough,

$$\begin{aligned}
&\mathbf{Pr}(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon}) \\
&\ge \sum_{(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}} \left(1-\frac{48}{n^{\alpha+\beta-1+\varepsilon}}\right)\mathbf{Pr}(|A_n|=k, |B_n|=l) \\
&\ge \left(1-\frac{48}{n^{\alpha+\beta-1+\varepsilon}}\right)\mathbf{Pr}(Q_n) \xrightarrow[n\to\infty]{} 1.
\end{aligned}$$

Which completes the proof by the characterization of densability (Proposition 2.6).

□

Remark that when $\alpha+\beta=1$ the density is not determined, as the remark for Proposition 2.7 showed for Bernoulli random subsets. Because the class of permutation invariant random subsets is closed under intersections, we can conclude on multiple intersections, as for the Bernoulli density model (Corollary 2.21).

Corollary 2.27. *Let $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_k$ be independent sequences of permutation invariant random subsets of* ***positive*** *densities.*

(i) If $\sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i < 1$*, then*

$$\operatorname{codens}\left(\bigcap_{i=1}^{k} \boldsymbol{A}_i\right) = \sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i.$$

(ii) If $\sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i > 1$*, then* $\operatorname{dens}\left(\bigcap_{i=1}^{k} \boldsymbol{A}_i\right) = -\infty$.

2.2.4 The random-fixed intersection formula

The intersection formula can be established between a densable sequence of random subsets and a densable sequence of *fixed* subsets (Theorem B).

Theorem 2.28 (The random-fixed intersection formula, Theorem B)**.** *Let $\boldsymbol{A}$ be a densable sequence of permutation invariant random subsets of $\boldsymbol{E}$ with density d. Let $\boldsymbol{X}$ be a sequence of (fixed) subsets of $\boldsymbol{E}$ with density α. If $d + \alpha \neq 1$, then the sequence of random subsets $\boldsymbol{A} \cap \boldsymbol{X}$ is densable and*

$$\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{X}) = \begin{cases} d + \alpha - 1 & \text{if } d + \alpha > 1 \\ -\infty & \text{if } d + \alpha < 1. \end{cases}$$

In addition, the sequence $\boldsymbol{A} \cap \boldsymbol{X}$ is a densable sequence of permutation invariant random subset of $\boldsymbol{X}$ with density $\frac{d+\alpha-1}{\alpha}$.

As this theorem can be regarded as a special case of the multidimensional intersection formula with $k = 1$ (see Remark for Theorem 2.40), we will not repeat the proof.

2.3 The multidimensional intersection formula

Let $\boldsymbol{E} = (E_n)$ be a sequence of finite sets with $|E_n| = n$. Let $k \geq 1$ be an integer. Denote the set of pairwise different k-tuples of E_n by

$$E_n^{(k)} = \{(x_1, \dots, x_k) \in E_n^k \mid x_i \neq x_j \ \forall i \neq j\}.$$

Denote $\boldsymbol{E}^{(k)} = (E_n^{(k)})_{n \in \mathbb{N}}$.

Similarly, for a sequence of random subsets $\boldsymbol{A} = (A_n)$ of $\boldsymbol{E}$,

$$A_n^{(k)} := \{(x_1, \dots, x_k) \in A_n^k \mid x_i \neq x_j \ \forall i \neq j\}.$$

The sequence $\boldsymbol{A}^{(k)} = (A_n^{(k)})$ is a sequence of random subsets of $\boldsymbol{E}^{(k)}$.

In this section, we establish an intersection formula between a sequence of random subsets of type $\boldsymbol{A}^{(k)}$ and a densable sequence of fixed subsets $\boldsymbol{X} = (X_n)$ of $\boldsymbol{E}^{(k)}$. To this end, we need an additional condition on $\boldsymbol{X}$. More precisely, $\boldsymbol{X}$ can not have too much "self-intersection". We will discuss this condition in the first subsection.

Following the path for proving the intersection formula (Theorem 2.25), we study first the case that $\boldsymbol{A}$ is a sequence of Bernoulli random subsets with density d. We then adapt the proof for the uniform density model, and prove the theorem for the permutation invariant model by decomposing permutation invariant random subsets into uniform random subsets.

2.3.1 The set of pairwise distinct k-tuples

In this subsection, we focus on the sequence of the set of pairwise distinct k-tuples $\boldsymbol{E}^{(k)}$. We will discuss the following two points.

a. Sequences of random subsets of the type $\boldsymbol{A}^{(k)}$

b. The self-intersection partition, and the d-small self-intersection condition of a sequence of subsets $\boldsymbol{X}$ of $\boldsymbol{E}^{(k)}$.

2.3.1.a Random subsets of the type $\boldsymbol{A}^{(k)}$

Proposition 2.29. *Let $\boldsymbol{A}$ be a densable sequence of random subsets of $\boldsymbol{E}$ with density $d > 0$. Then $\boldsymbol{A}^{(k)}$ is a densable sequence of random subsets of $\boldsymbol{E}^{(k)}$ with density d. Namely,*

$$\mathrm{dens}_{\boldsymbol{E}^{(k)}}(\boldsymbol{A}^{(k)}) = \mathrm{dens}_{\boldsymbol{E}}(\boldsymbol{A}).$$

Proof. Let $\varepsilon > 0$ be a small real number. Note that $n^k - k^2(n-1)^k \leq |E_n^{(k)}| \leq n^k$, so for n large enough,

$$n^{k(1-\varepsilon)} \leq |E_n^{(k)}| \leq n^{k(1+\varepsilon)}.$$

By densability of $\boldsymbol{A}$ and Proposition 2.6, a.a.s. $n^{d(1-\varepsilon)} \leq |A_n| \leq n^{d(1+\varepsilon)}$. By the same argument above, a.a.s.

$$n^{dk(1-\varepsilon)^2} \leq |A_n^{(k)}| \leq n^{dk(1+\varepsilon)^2}.$$

Hence, a.a.s.

$$|E_n^{(k)}|^{d(1-\varepsilon)^2(1+\varepsilon)^{-1}} \leq |A_n^{(k)}| \leq |E_n^{(k)}|^{d(1+\varepsilon)^2(1-\varepsilon)^{-1}}.$$

We conclude by the characterization of densability (Proposition 2.6). □

Although the densability of $\boldsymbol{A}^{(k)}$ is preserved, it is not the case for being permutation invariant. Given a permutation invariant random subset A_n of E_n, the random subset $A_n^{(k)}$ is *not* permutation invariant in $E_n^{(k)}$ for $k \geq 2$. Here is a simple example.

Counter example. Let (A_n) be a sequence of Bernoulli random subsets of (E_n) with density $0 < d < 1$. Recall that subsets of the same cardinality have the same probability. Let $x_1, \ldots, x_4$ be distinct elements in E_n.

$$\mathbf{Pr}\left(\{(x_1,x_2),(x_3,x_4)\} \subset A_n^{(2)}\right) = \mathbf{Pr}\left(\{x_1,x_2,x_3,x_4\} \subset A_n\right) = n^{4(d-1)},$$

while

$$\mathbf{Pr}\left(\{(x_1,x_2),(x_2,x_3)\} \subset A_n^{(2)}\right) = \mathbf{Pr}\left(\{x_1,x_2,x_3\} \subset A_n\right) = n^{3(d-1)}. \quad \square$$

In particular, the classical intersection formula (Theorem 2.28) can not be applied. Actually, for $k \geq 2$ the intersection formula *does not* work for some sequences of random subsets $\boldsymbol{X}$. We give here a counterexample.

Example. Let $\boldsymbol{A}$ be a densable sequence of permutation invariant random subsets with density $3/4$. Let $\boldsymbol{X} = (X_n)$ be a sequence of subsets defined by

$$X_n = \{x_n\} \times (E_n \backslash \{x_n\}) \subset E_n^{(2)}$$

with some $x_n \in E_n$. By its construction $\mathrm{dens}_{\boldsymbol{E}^{(2)}}(\boldsymbol{X}) = 1/2$, so we expected that $\mathrm{dens}(\boldsymbol{A}^{(2)} \cap \boldsymbol{X}) = 3/4 + 1/2 - 1 = 1/4$. However, we have

$$\mathrm{dens}(\boldsymbol{A}^{(2)} \cap \boldsymbol{X}) = 0$$

because a.a.s. $A_n \cap \{x_n\} = \emptyset$.

2.3.1.b The self-intersection partition of a subset

Definition 2.30 (Self-intersection). *Let $\boldsymbol{X} = (X_n)$ be a sequence of fixed subsets of $\boldsymbol{E}^{(k)}$ with density α. For $0 \le i \le k$, the i-th self-intersection of X_n is*

$$Y_{i,n} := \{(x, y) \in X_n^2 \mid |x \cap y| = i\}$$

where $|x \cap y|$ is the number of common elements between the sets $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$.

For example, with $k = 3$ and x_1, x_2, x_3, x_4 different elements of E_n, the pair $((x_1, x_2, x_3), (x_2, x_3, x_4))$ is in $Y_{2,n}$ because x_2, x_3 are repeated.

In particular, $Y_{0,n}$ is the set of pairs (x, y) in X_n^2 having no intersection; $Y_{k,n}$ is the set of pairs (x, y) in X_n^2 such that y is a permutation of x. Note that $(Y_{i,n})_{0 \le i \le k}$ is a partition of X_n^2, called the *self-intersection partition* of X_n. Namely,

$$X_n^2 = \bigsqcup_{i=0}^{k} Y_{i,n}.$$

Definition 2.31 (Self-intersection partition). *The sequence $\boldsymbol{Y}_i = (Y_{i,n})_{n \in \mathbb{N}}$ is called the i-th self intersection of $\boldsymbol{X}$. The family of sequences $(\boldsymbol{Y}_i)_{0 \le i \le k}$ is called the self-intersection partition of $\boldsymbol{X}$. Namely,*

$$\boldsymbol{X}^2 = \bigsqcup_{i=0}^{k} \boldsymbol{Y}_i.$$

The sequences $\boldsymbol{X}^2$ and $\boldsymbol{Y}_i$ are sequences of fixed subsets of $(\boldsymbol{E}^{(k)})^2 = \left((E_n^{(k)})^2\right)_{n \in \mathbb{N}}$. Note that $\mathrm{dens}_{(\boldsymbol{E}^{(k)})^2}(\boldsymbol{X}^2) = \mathrm{dens}_{\boldsymbol{E}^{(k)}}(\boldsymbol{X}) = \alpha$. To give a condition on $\boldsymbol{Y}_i$, we need the notion of *upper density*.

Definition 2.32 (Upper density). *Let $\boldsymbol{Y} = (Y_n)$ be a sequence of subsets of $\boldsymbol{E} = (E_n)$. The upper density of $\boldsymbol{Y}$ in $\boldsymbol{E}$ is*

$$\overline{\mathrm{dens}}_{\boldsymbol{E}} \boldsymbol{Y} := \overline{\lim_{n \to \infty}} \log_{|E_n|}(|Y_n|).$$

We introduce here, for a sequence of densable fixed subsets $\boldsymbol{X}$ of $\boldsymbol{E}^{(k)}$ with density α, the small self-intersection condition:

Definition 2.33 (d-small self-intersection condition). *Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α and let $(\boldsymbol{Y}_i)_{0 \le i \le k}$ be its self-intersection partition. Let $d > 1 - \alpha$. We say that $\boldsymbol{X}$ satisfies the d-small self-intersection condition if, for every $1 \le i \le k - 1$,*

$$\overline{\mathrm{dens}}_{(\boldsymbol{E}^{(k)})^2} (\boldsymbol{Y}_i) < \alpha - (1 - d) \times \frac{i}{2k}.$$

Remark (The cases $i = 0$ and $i = k$). For $i = k$, we have $|Y_{k,n}| = |\{(x, y) \in X_n^2 \mid y \text{ is a permutation of } x\}| = k!|X_n|$, so

$$\mathrm{dens}_{(\boldsymbol{E}^{(k)})^2} \boldsymbol{Y}_k = \frac{\alpha}{2} < \alpha - (1 - d)\frac{k}{2k}.$$

For $i = 0$, as the upper densities of $\boldsymbol{Y}_i$ for $1 \le i \le k$ are all smaller than α (because $0 < 1 - d < \alpha$) and $Y_{0,n} = X_n^2 \backslash \bigcup_{i=1}^{k} |Y_{i,n}|$, by Proposition 2.13 (density of unions) and Proposition 2.14 (density of complements), the sequence $\boldsymbol{Y}_0$ is densable in $(\boldsymbol{E}^{(k)})^2$ with density

$$\mathrm{dens}_{(\boldsymbol{E}^{(k)})^2} \boldsymbol{Y}_0 = \mathrm{dens}_{(\boldsymbol{E}^{(k)})^2} \boldsymbol{X}^2 = \alpha.$$

We shall represent the expected value and the variance of the random variable $|A_n^{(k)} \cap X_n|$ by probability values of the type $\mathbf{Pr}\left(\{x_1, \ldots, x_r\} \subset A_n\right)$.

Lemma 2.34. *Let $\boldsymbol{E}$, $\boldsymbol{A}$ and $\boldsymbol{X}$ given by Theorem 2.40 and let $(\boldsymbol{Y}_i)_{0 \le i \le k}$ be the self-intersection partition of $\boldsymbol{X}$. Let $x_1, \ldots, x_{2k}$ be distinct $2k$ elements of E_n.*

(i) $\mathbb{E}\left(|A_n^{(k)} \cap X_n|\right) = |X_n| \mathbf{Pr}\left(\{x_1, \ldots, x_k\} \subset A_n\right)$.

(ii) $\operatorname{Var}\left(|A_n^{(k)} \cap X_n|\right) =$

$$|X_n|^2 \Big(\mathbf{Pr}(\{x_1, \ldots, x_{2k}\} \subset A_n) - \mathbf{Pr}(\{x_1, \ldots, x_k\} \subset A_n)^2 \Big)$$
$$+ \sum_{i=1}^{k} |Y_{i,n}| \Big(\mathbf{Pr}(\{x_1, \ldots, x_{2k-i}\} \subset A_n) - \mathbf{Pr}(\{x_1, \ldots, x_{2k}\} \subset A_n) \Big).$$

Proof.

(i) As A_n is permutation invariant, the probability $\mathbf{Pr}(\{x_1, \ldots, x_k\} \subset A_n)$ does not depend on the choice of $\{x_1, \ldots, x_k\}$. So

$$\begin{aligned}\mathbb{E}(|A_n^{(k)} \cap X_n|) &= \mathbb{E}\left(\sum_{x \in X_n} \mathbb{1}_{x \in A_n^{(k)}}\right) = \sum_{x \in X_n} \mathbf{Pr}\left(x \in A_n^{(k)}\right) \\ &= |X_n| \mathbf{Pr}\left(\{x_1, \ldots, x_k\} \subset A_n\right).\end{aligned}$$

(ii) By the same reason $\mathbf{Pr}(\{x_1, \ldots, x_r\} \subset A_n)$ does not depend on the choice of $\{x_1, \ldots, x_r\}$ for all $r \in \mathbb{N}$. Note that

$$\operatorname{Var}(|A_n^{(k)} \cap X_n|) = \mathbb{E}\left(|A_n^{(k)} \cap X_n|^2\right) - \mathbb{E}\left(|A_n^{(k)} \cap X_n|\right)^2.$$

If $(x, y) \in Y_{i,n}$, then there are $2k - i$ different elements of E_n in x and y, so $\mathbf{Pr}\left(x, y \in A_n^{(k)}\right) = \mathbf{Pr}(\{x_1, \ldots, x_{2k-i}\} \subset A_n)$. Hence,

$$\begin{aligned}\mathbb{E}\left(|A_n^{(k)} \cap X_n|^2\right) &= \mathbb{E}\left(\left(\sum_{x \in X_n} \mathbb{1}_{x \in A_n^{(k)}}\right)^2\right) = \sum_{x, y \in X_n} \mathbf{Pr}\left(x, y \in A_n^{(k)}\right) \\ &= \sum_{i=0}^{k} \sum_{(x,y) \in Y_{i,n}} \mathbf{Pr}\left(x, y \in A_n^{(k)}\right) \\ &= \sum_{i=0}^{k} |Y_{i,n}| \mathbf{Pr}(\{x_1, \ldots, x_{2k-i}\} \subset A_n).\end{aligned}$$

Recall that $|Y_{0,n}| = |X_n^2| - \sum_{i=1}^k |Y_{i,n}|$. The above can be rewritten as

$$\begin{aligned}\mathbb{E}\left(|A_n^{(k)} \cap X_n|^2\right) &= \left(|X_n^2| - \sum_{i=1}^k |Y_{i,n}|\right) \mathbf{Pr}(\{x_1, \dots, x_{2k}\} \subset A_n) \\ &\quad + \sum_{i=1}^k |Y_{i,n}| \mathbf{Pr}(\{x_1, \dots, x_{2k-i}\} \subset A_n) \\ &= |X_n^2| \mathbf{Pr}(\{x_1, \dots, x_{2k}\} \subset A_n) \\ &\quad + \sum_{i=1}^k \Big(\mathbf{Pr}(\{x_1, \dots, x_{2k-i}\} \subset A_n) - \mathbf{Pr}(\{x_1, \dots, x_{2k}\} \subset A_n)\Big).\end{aligned}$$

Combined with $\mathbb{E}\left(|A_n^{(k)} \cap X_n|\right)^2 = |X_n|^2 \mathbf{Pr}\left(\{x_1, \dots, x_k\} \subset A_n\right)^2$, we have

$$\begin{aligned}\mathrm{Var}(|A_n^{(k)} \cap X_n|) &= |X_n|^2 \Big(\mathbf{Pr}(\{x_1, \dots, x_{2k}\} \subset A_n) - \mathbf{Pr}(\{x_1, \dots, x_k\} \subset A_n)^2\Big) \\ &\quad + \sum_{i=1}^k |Y_{i,n}| \Big(\mathbf{Pr}(\{x_1, \dots, x_{2k-i}\} \subset A_n) - \mathbf{Pr}(\{x_1, \dots, x_{2k}\} \subset A_n\Big).\end{aligned}$$

□

Remark. Lemma 2.10 is a special case of this lemma, with $k = 1$ and $X_n = E_n$.

2.3.2 The Bernoulli density model

In this subsection, we prove the multidimensional intersection formula for the Bernoulli density model. The Bernoulli density model is easier to manipulate because of the following proposition.

Proposition 2.35. *Let $\boldsymbol{A}$ be a sequence of Bernoulli random subsets of $\boldsymbol{E}$ with density $0 < d < 1$. For any integer $r \in \mathbb{N}$ and any distinct elements $x_1, \dots, x_r$ in E_n, we have*

$$\begin{aligned}\mathbf{Pr}\left(\{x_1, \dots, x_r\} \subset A_n\right) &= \mathbf{Pr}\left(\{x_1 \in A_n\}, \dots, \{x_r \in A_n\}\right) \\ &= \prod_{i=1}^r \mathbf{Pr}(x_i \in A_n) = n^{r(d-1)}.\end{aligned}$$

□

The proof is evident by the independence of the events $\mathbf{Pr}(x_i \in A_n)$. Because of this equality, the proof of the multidimensional intersection formula is much simpler for the Bernoulli density model. We establish later a similar proposition for the uniform density model (Proposition 2.38).

Theorem 2.36 (The multidimensional intersection formula for the Bernoulli density model)**.** *Let $\boldsymbol{A}$ be a sequence of Bernoulli random subsets of $\boldsymbol{E}$ with density $d > 0$. Let $\boldsymbol{X} = (X_n)$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α.*

(i) If $d + \alpha < 1$, then $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and

$$\mathrm{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = -\infty.$$

(ii) If $d + \alpha > 1$ and $\boldsymbol{X}$ satisfies the d-small self intersection condition (Definition 2.33), then $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and

$$\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = \alpha + d - 1.$$

Proof.

(i) Suppose that $\alpha + d < 1$. To prove that $\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = -\infty$, it is enough to prove that $\mathbf{Pr}\left(A_n^{(k)} \cap X_n \neq \varnothing\right) \xrightarrow[n\to\infty]{} 0$.

By Markov's inequality and Lemma 2.34 (i) and Proposition 2.35,

$$\begin{aligned}
\mathbf{Pr}\left(A_n^{(k)} \cap X_n \neq \varnothing\right) &= \mathbf{Pr}\left(|A_n^{(k)} \cap X_n| \geq 1\right) \\
&\leq \mathbb{E}\left(|A_n^{(k)} \cap X_n|\right) = |X_n| \mathbf{Pr}\left(\{x_1, \ldots, x_k\} \subset A_n\right) \\
&\leq n^{k\alpha + o(1)} n^{k(d-1)} \\
&\leq n^{k(\alpha+d-1)+o(1)} \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

as $\alpha + d - 1 < 0$. □

(ii) Suppose that $\alpha + d > 1$. To simplify the notation, denote $B_n = A_n^{(k)} \cap X_n$ and $\boldsymbol{B} = \boldsymbol{X} \cap \boldsymbol{A}^{(k)}$.

We shall prove that $\operatorname{dens} \boldsymbol{B} = \alpha + d - 1$. Let $\varepsilon > 0$ be a small real number. We want to prove that

$$\text{a.a.s. } n^{k(\alpha+d-1-\varepsilon)} \leq |B_n| \leq n^{k(\alpha+d-1+\varepsilon)}.$$

By Lemma 2.34 (i) and Proposition 2.35,

$$\begin{aligned}
\mathbb{E}\left(|B_n|\right) &= |X_n| \mathbf{Pr}\left(\{x_1, \ldots, x_k\} \subset A_n\right) = |X_n| n^{k(d-1)} \\
&= n^{k(\alpha+d-1)+o(1)}.
\end{aligned}$$

For n large enough,

$$n^{k(\alpha+d-1-\varepsilon)} < \frac{1}{2} n^{k(\alpha+d-1)+o(1)} < \frac{3}{2} n^{k(\alpha+d-1)+o(1)} < n^{k(\alpha+d-1+\varepsilon)}.$$

So it is enough to prove that

$$\text{a.a.s. } \frac{1}{2}\mathbb{E}\left(|B_n|\right) < |B_n| < \frac{3}{2}\mathbb{E}\left(|B_n|\right),$$

which is equivalent so

$$\text{a.a.s. } \left||B_n| - \mathbb{E}\left(|B_n|\right)\right| < \frac{1}{2}\mathbb{E}\left(|B_n|\right).$$

By Chebyshev's inequality

$$\mathbf{Pr}\left(\left||B_n| - \mathbb{E}\left(|B_n|\right)\right| \geq \frac{1}{2}\mathbb{E}\left(|B_n|\right)\right) \leq \frac{4 \operatorname{Var}\left(|B_n|\right)}{\mathbb{E}\left(|B_n|\right)^2}.$$

We shall prove that $\frac{4\operatorname{Var}(|B_n|)}{\mathbb{E}(|B_n|)^2}$ converges to 0 when n goes to infinity. By Lemma 2.34 (ii) and Proposition 2.35,

$$\begin{aligned}\operatorname{Var}(|B_n|) &= |X_n|^2\Big(\mathbf{Pr}(\{x_1,\dots,x_{2k}\}\subset A_n) - \mathbf{Pr}(\{x_1,\dots,x_k\}\subset A_n)^2\Big)\\ &\quad + \sum_{i=1}^{k}|Y_{i,n}|\Big(\mathbf{Pr}(\{x_1,\dots,x_{2k-i}\}\subset A_n) - \mathbf{Pr}(\{x_1,\dots,x_{2k}\}\subset A_n\Big)\\ &= \sum_{i=1}^{k}|Y_{i,n}|\left(n^{(2k-i)(d-1)} - n^{2k(d-1)}\right)\\ &\le \sum_{i=1}^{k}|Y_{i,n}|n^{(2k-i)(d-1)}\end{aligned}$$

Note that $n^{(2k-i)(d-1)} > n^{2k(d-1)}$ because $d < 1$. By the d-small self-intersection condition (Definition 2.33), there exists $\varepsilon > 0$ such that, for all $1 \le i \le k$,

$$|Y_{i,n}| \le n^{2k\left(\alpha+(d-1)\frac{i}{2k}\right)-\varepsilon}$$

for n large enough.
Hence, for n large enough,

$$\operatorname{Var}(|B_n|) \le kn^{2k(\alpha+d-1)-\varepsilon}.$$

Recall that $\mathbb{E}\left(|B_n|\right)^2 = n^{2k(\alpha+d-1)+o(1)}$, so

$$\frac{4\operatorname{Var}\left(|B_n|\right)}{\mathbb{E}\left(|B_n|\right)^2} \xrightarrow[n\to\infty]{} 0.$$

□

2.3.3 The uniform density model

In this subsection we prove the theorem for the uniform density model.

Theorem 2.37 (The multidimensional intersection formula for the uniform density model)**.** *Let $\boldsymbol{A}$ be a sequence of uniform random subsets of $\boldsymbol{E}$ with density d. Let $\boldsymbol{X} = (X_n)$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α.*

(i) If $d + \alpha < 1$, then $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and

$$\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = -\infty.$$

(ii) If $d + \alpha > 1$ and $\boldsymbol{X}$ satisfies the d-small self intersection condition (Definition 2.33), then $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and

$$\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = \alpha + d - 1.$$

Recall that to prove this theorem in the Bernoulli density model, we rely on the following two facts:

$$\mathbf{Pr}(\{x_1,\dots,x_r\}\subset A_n) = n^{r(d-1)},$$

and

$$\mathbf{Pr}(\{x_1,\dots,x_k\}\subset A_n)^2-\mathbf{Pr}(\{x_1,\dots,x_{2k}\}\subset A_n)=0.$$

We shall estimate the two quantities on the left-hand side for the uniform density model.

Proposition 2.38. *Let $\boldsymbol{A}$ be a sequence of uniform random subsets of $\boldsymbol{E}$ with density d. Let $0<\varepsilon<d$ be a small real number and let $k\geq 1$ be an integer. If $n\geq(1+2k)^{\frac{1}{\varepsilon}}$, then*

(i) For any integer $1\leq r\leq 2k$,

$$n^{r(d-1-\varepsilon)}\leq\mathbf{Pr}(\{x_1,\dots,x_r\}\subset A_n)\leq n^{r(d-1+\varepsilon)}.$$

(ii) $0\leq\mathbf{Pr}(\{x_1,\dots,x_k\}\subset A_n)^2-\mathbf{Pr}(\{x_1,\dots,x_{2k}\}\subset A_n)\leq n^{2k(d-1+\varepsilon)-d}$

In fact, for (ii), we will only need the left-hand side inequality.

Proof. Recall that $|E_n|=n$ and that A_n is a uniform distribution on all subsets of E_n of cardinality $\lfloor n^d\rfloor$.

(i) Because $\varepsilon<d$, we have $\lfloor n^d\rfloor\geq n^{\varepsilon}-1\geq 2k\geq r$. Among all subsets of E_n of cardinality $\lfloor n^d\rfloor$, there are $\binom{n-r}{\lfloor n^d\rfloor-r}$ subsets that include $\{x_1,\dots,x_r\}$. So

$$\mathbf{Pr}(\{x_1,\dots,x_r\}\subset A_n)=\frac{\binom{n-r}{\lfloor n^d\rfloor-r}}{\binom{n}{\lfloor n^d\rfloor}}=\frac{\lfloor n^d\rfloor\dots(\lfloor n^d\rfloor-r+1)}{n\dots(n-r-1)}.$$

We have

$$\left(\frac{n^d-r}{n}\right)^r\leq\frac{\lfloor n^d\rfloor\dots(\lfloor n^d\rfloor-r+1)}{n\dots(n-r-1)}\leq\left(\frac{n^d}{n-r}\right)^r.$$

The condition $n\geq(1+2k)^{\frac{1}{\varepsilon}}\geq(1+r)^{\frac{1}{\varepsilon}}$ implies

$$\begin{cases}n\geq n^{1-\varepsilon}(1+r)\\ n^d\geq n^{d-\varepsilon}(1+r),\end{cases}$$

so

$$\begin{cases}n^{1-\varepsilon}\leq n-r\\ n^{d-\varepsilon}\leq n^d-r.\end{cases}$$

Hence,

$$\left(n^{d-1-\varepsilon}\right)^r\leq\frac{\lfloor n^d\rfloor\dots(\lfloor n^d\rfloor-r+1)}{n\dots(n-r-1)}\leq\left(n^{d-1+\varepsilon}\right)^r.$$

Which means that

$$n^{r(d-1-\varepsilon)}\leq\mathbf{Pr}(\{x_1,\dots,x_r\}\subset A_n)\leq n^{r(d-1+\varepsilon)}.$$

□

(ii) By the same argument,

$$\begin{aligned}&\mathbf{Pr}(\{x_1,\dots,x_k\}\subset A_n)^2-\mathbf{Pr}(\{x_1,\dots,x_{2k}\}\subset A_n)\\ =&\left(\frac{\lfloor n^d\rfloor\dots(\lfloor n^d\rfloor-k+1)}{n\dots(n-k-1)}\right)^2-\frac{\lfloor n^d\rfloor\dots(\lfloor n^d\rfloor-2k+1)}{n\dots(n-2k-1)}\\ =&\left(\frac{\lfloor n^d\rfloor\dots(\lfloor n^d\rfloor-k+1)}{n\dots(n-k-1)}\right)\left(\frac{\lfloor n^d\rfloor\dots(\lfloor n^d\rfloor-k+1)}{n\dots(n-k-1)}-\frac{(\lfloor n^d\rfloor-k)\dots(\lfloor n^d\rfloor-2k+1)}{(n-k)\dots(n-2k-1)}\right).\end{aligned}$$

This number is positive because $\frac{\lfloor n^d \rfloor - i}{n-i} \geq \frac{\lfloor n^d \rfloor - i - k}{n-i-k}$ for every $0 \leq i \leq k-1$.
By a simple estimation of type $n - k + i \geq n - k$, we have

$$\begin{aligned}
&\mathbf{Pr}(\{x_1, \ldots, x_k\} \subset A_n)^2 - \mathbf{Pr}(\{x_1, \ldots, x_{2k}\} \subset A_n) \\
\leq &\left(\frac{n^d}{n-k}\right)^k \left(\frac{n^{dk}}{(n-k)^k} - \frac{(n^d - 2k)^k}{(n-k)^k}\right) \\
\leq &\frac{n^{dk}}{(n-k)^{2k}} \left(n^{dk} - \sum_{i=0}^{k} \binom{k}{i} n^{d(k-i)} (-2k)^i\right) \\
\leq &\frac{n^{dk}}{(n-k)^{2k}} (1+2k)^k n^{d(k-1)} = \left(\frac{n^d \sqrt{1+2k}}{n-k}\right)^{2k} n^{-d}.
\end{aligned}$$

As $n^\varepsilon \geq 1 + 2k$, we have

$$\begin{aligned}
n - k &\geq n^{1-\varepsilon}(1+2k) - k \\
&\geq n^{1-\varepsilon}(1+k) \\
&\geq n^{1-\varepsilon}\sqrt{1+2k},
\end{aligned}$$

so

$$\frac{n^d \sqrt{1+2k}}{n-k} \leq n^{d-1+\varepsilon}.$$

Hence,

$$\mathbf{Pr}(\{x_1, \ldots, x_k\} \subset A_n)^2 - \mathbf{Pr}(\{x_1, \ldots, x_{2k}\} \subset A_n) \leq n^{2k(d-1+\varepsilon)-d}.$$

□

Notation. Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α and let $(\boldsymbol{Y}_i)_{0 \leq i \leq k}$ be its self-intersection partition (Definition 2.31). Recall that the upper density $\overline{\text{dens}}\boldsymbol{Y}_i$ is defined as an upper limit of densities (Definition 2.32). Denote the *density difference* by

$$\varepsilon_0(d) = \min_{1 \leq i \leq k} \left\{\alpha + (d-1)\frac{i}{2k} - \overline{\text{dens}}\boldsymbol{Y}_i\right\}.$$

By Definition 2.33, the sequence of subsets $\boldsymbol{X}$ has d-small self-intersection if and only if $\varepsilon_0(d) > 0$. In addition, for every small real number $0 < \varepsilon < \frac{\varepsilon_0(d)}{10}$, there exists $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$ we have, simultaneously for all $1 \leq i \leq k$,

$$|Y_{n,i}| \leq n^{2k\left(\alpha + (d-1)\frac{i}{2k} - 10\varepsilon\right)} = n^{2k\alpha + (d-1)i - 2k \times 10\varepsilon}.$$

By the densability of $\boldsymbol{X}$, we can choose n_ε such that at the same time

$$n^{k(\alpha-\varepsilon)} \leq |X_n| \leq n^{k(\alpha+\varepsilon)}.$$

Combine with Proposition 2.38, we can now estimate the expected value and the variance of $|A_n^{(k)} \cap X_n|$ for the uniform density model.

Lemma 2.39. *Let $\boldsymbol{A}$ be a sequence of uniform random subsets of $\boldsymbol{E}$ with density d. Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α. Let $0 < \varepsilon < \min\{\frac{\varepsilon_0(d)}{10}, d\}$ be a small real number. If $n \geq \max\left\{n_\varepsilon, (1+2k)^{\frac{1}{\varepsilon}}\right\}$, then*

(i) $n^{k(\alpha+d-1-2\varepsilon)} \leq \mathbb{E}\left(|A_n^{(k)} \cap X_n|\right) \leq n^{k(\alpha+d-1+2\varepsilon)}$.

(ii) *If in addition* $\alpha + d - 1 > 2\varepsilon > 0$ *and* $\boldsymbol{X}$ *has* d*-small self-intersection, then*

$$\operatorname{Var}\left(|A_n^{(k)} \cap X_n|\right) \leq kn^{2k(\alpha+d-1-9\varepsilon)}.$$

In particular,

$$\frac{\operatorname{Var}\left(|A_n^{(k)} \cap X_n|\right)}{\mathbb{E}\left(|A_n^{(k)} \cap X_n|\right)^2} \leq kn^{-14k\varepsilon}.$$

Proof.

(i) Recall Lemma 2.34 (i),

$$\mathbb{E}\left(|A_n^{(k)} \cap X_n|\right) = |X_n|\mathbf{Pr}\left(\{x_1, \ldots, x_k\} \subset A_n\right).$$

By Proposition 2.38 (i) and the fact that $n^{k(\alpha-\varepsilon)} \leq |X_n| \leq n^{k(\alpha+\varepsilon)}$, we have

$$n^{k(\alpha-\varepsilon)}n^{k(d-1-\varepsilon)} \leq \mathbb{E}\left(|A_n^{(k)} \cap X_n|\right) \leq n^{k(\alpha+\varepsilon)}n^{k(d-1+\varepsilon)}.$$

□

(ii) By the left-hand side inequality of Proposition 2.38 (ii), $\mathbf{Pr}(\{x_1, \ldots, x_{2k}\} \subset A_n) - \mathbf{Pr}(\{x_1, \ldots, x_k\} \subset A_n)^2 \leq 0$. Let us eliminate negative parts of Lemma 2.34 (ii).

$$\begin{aligned}\operatorname{Var}\left(|A_n^{(k)} \cap X_n|\right) &= |X_n|^2\Big(\mathbf{Pr}(\{x_1, \ldots, x_{2k}\} \subset A_n) - \mathbf{Pr}(\{x_1, \ldots, x_k\} \subset A_n)^2\Big) \\ &\quad + \sum_{i=1}^{k} |Y_{i,n}|\Big(\mathbf{Pr}(\{x_1, \ldots, x_{2k-i}\} \subset A_n) - \mathbf{Pr}(\{x_1, \ldots, x_{2k}\} \subset A_n)\Big) \\ &\leq \sum_{i=1}^{k} |Y_{i,n}|\mathbf{Pr}(\{x_1, \ldots, x_{2k-i}\} \subset A_n).\end{aligned}$$

By Proposition 2.38 (i) and the fact that $|Y_{i,n}| \leq n^{2k\alpha+i(d-1)+2k\times 10\varepsilon}$,

$$\begin{aligned}\operatorname{Var}\left(|A_n^{(k)} \cap X_n|\right) &\leq \sum_{i=1}^{k} n^{2k\alpha+i(d-1)-2k\times 10\varepsilon}n^{(2k-i)(d-1+\varepsilon)} \\ &\leq kn^{2k(\alpha+d-1-9\varepsilon)}.\end{aligned}$$

□

Now we can prove the multidimensional intersection formula for the uniform density model.

Proof of Theorem 2.37.

(i) Suppose that $\alpha + d < 1$. We shall prove that $\mathbf{Pr}\left(A_n^{(k)} \cap X_n \neq \emptyset\right) \xrightarrow[n\to\infty]{} 0$.
Let $\varepsilon > 0$ such that

$$\varepsilon < \min\left\{\frac{1-d-\alpha}{2}, \frac{\varepsilon_0(d)}{10}, d\right\}.$$

By Markov's inequality and Lemma 2.39 (i). If $n \geq \max\{n_\varepsilon, (1+2k)^{\frac{1}{\varepsilon}}\}$, then

$$\begin{aligned}\mathbf{Pr}\left(A_n^{(k)} \cap X_n \neq \emptyset\right) &= \mathbf{Pr}\left(|A_n^{(k)} \cap X_n| \geq 1\right)\\ &\leq \mathbb{E}\left(|A_n^{(k)} \cap X_n|\right)\\ &\leq n^{k(\alpha+d-1+2\varepsilon)} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

□

(ii) Suppose that $\alpha + d > 1$. Denote $B_n = A_n^{(k)} \cap X_n$.
Let $\varepsilon > 0$ such that

$$\varepsilon < \min\left\{\frac{\alpha+d-1}{3}, \frac{\varepsilon_0(d)}{10}, d\right\}.$$

We shall prove that

$$\text{a.a.s.} n^{k(\alpha+d-1-3\varepsilon)} \leq |B_n| \leq n^{k(\alpha+d-1+3\varepsilon)}.$$

By Lemma 2.39 (i), if $n \geq \max\{n_\varepsilon, (1+2k)^{\frac{1}{\varepsilon}}\}$, then

$$n^{k(\alpha+d-1-2\varepsilon)} \leq \mathbb{E}\left(|B_n|\right) \leq n^{k(\alpha+d-1+2\varepsilon)}.$$

In addition, if $n \geq 2^{\frac{1}{k\varepsilon}}$, we have

$$n^{k(\alpha+d-1-3\varepsilon)} \leq \frac{1}{2} n^{k(\alpha+d-1-2\varepsilon)} \leq \frac{1}{2}\mathbb{E}\left(|B_n|\right)$$

and

$$\frac{3}{2}\mathbb{E}\left(|B_n|\right) \leq \frac{3}{2} n^{k(\alpha+d-1+2\varepsilon)} \leq n^{k(\alpha+d-1+3\varepsilon)}.$$

So it is enough to prove that

$$\text{a.a.s.} \left||B_n| - \mathbb{E}\left(|B_n|\right)\right| \leq \frac{1}{2}\mathbb{E}\left(|B_n|\right).$$

By Chebyshev's inequality,

$$\mathbf{Pr}\left(\left||B_n| - \mathbb{E}\left(|B_n|\right)\right| > \frac{1}{2}\mathbb{E}\left(|B_n|\right)\right) \leq \frac{4\,\mathrm{Var}\left(|B_n|\right)}{\mathbb{E}\left(|B_n|\right)^2}.$$

Apply Lemma 2.39 (i) and (ii), if $n \geq \max\left\{n_\varepsilon, (1+2k)^{\frac{1}{\varepsilon}}, 2^{\frac{1}{k\varepsilon}}\right\}$, then

$$\frac{4\,\mathrm{Var}\left(|B_n|\right)}{\mathbb{E}\left(|B_n|\right)^2} \leq \frac{4kn^{2k(\alpha+d-1-9\varepsilon)}}{n^{2k(\alpha+d-1-2\varepsilon)}} \leq \frac{4k}{n^{14k\varepsilon}} \xrightarrow[n\to\infty]{} 0.$$

Which completes the proof.

□

2.3.4 The permutation invariant density model

In this subsection, we prove Theorem C.

Theorem 2.40 (The multidimensional intersection formula, Theorem C)**.** *Let $\boldsymbol{A}$ be a densable sequence of permutation invariant random subsets of $\boldsymbol{E}$ with density d. Let $\boldsymbol{X} = (X_n)$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α.*

(i) If $d + \alpha < 1$, then $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and

$$\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = -\infty.$$

(ii) If $d + \alpha > 1$ and $\boldsymbol{X}$ satisfies the d-small self intersection condition (Definition 2.33), then $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and

$$\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = \alpha + d - 1.$$

Remark. When $k = 1$, we get the usual random-fixed intersection formula (Theorem 2.28). In this case, the self intersection partition of $\boldsymbol{X}$ contains only $\boldsymbol{Y}_0$ and $\boldsymbol{Y}_1$, and the self intersection condition does not make sense.

Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ having the d-small intersection condition. Recall that the density difference is $\varepsilon_0(d) = \min_{1 \le i \le k} \left\{\alpha + (d-1)\frac{i}{2k} - \overline{\operatorname{dens}}\boldsymbol{Y}_i\right\} > 0$. Note that if $d' < d$ then $\varepsilon_0(d') < \varepsilon_0(d)$. We will decompose the permutation invariant random subset A_n into uniform random subsets and apply lemma 2.39 in a small interval $[d-\varepsilon, d+\varepsilon]$. We choose $0 < \varepsilon < \min\left\{\frac{\varepsilon_0(d)}{20}, \frac{d}{2}\right\}$ so that $\varepsilon < \min\left\{\frac{\varepsilon_0(d-\varepsilon)}{10}, d-\varepsilon\right\} \le \min\left\{\frac{\varepsilon_0(d')}{10}, d'\right\}$ for every $d' \in [d-\varepsilon, d+\varepsilon]$.

By the definition of $\varepsilon_0(d)$ and the densability of $\boldsymbol{X}$, we can choose $n_\varepsilon \in \mathbb{N}$ large enough, such that for all $n \ge n_\varepsilon$ we have, for any $1 \le i \le k$,

$$|Y_{n,i}| \le n^{2k\alpha + (d-1)i - 2k \times 20\varepsilon} \le n^{2k\alpha + (d'-1)i - 2k\times 10\varepsilon}$$

and

$$n^{k(\alpha-\varepsilon)} \le |X_n| \le n^{k(\alpha+\varepsilon)}.$$

Lemma 2.41. *Let $0 < \varepsilon < \min\left\{\frac{\varepsilon_0(d)}{20}, \frac{d}{2}\right\}$ be a small real number. Let $\boldsymbol{A}$ be a sequence of uniform random subsets of $\boldsymbol{E}$ with density $d' \in [d-\varepsilon, d+\varepsilon]$. Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α. If $n \ge \max\left\{n_\varepsilon, (1+2k)^{\frac{1}{\varepsilon}}\right\}$, then*

(i) $n^{k(\alpha+d-1-3\varepsilon)} \le \mathbb{E}\left(|A_n^{(k)} \cap X_n|\right) \le n^{k(\alpha+d-1+3\varepsilon)}$.

(ii) If in addition $\alpha + d - 1 > 3\varepsilon > 0$ and $\boldsymbol{X}$ has d-small self-intersection, then $\operatorname{Var}(|A_n^{(k)} \cap X_n|) \le kn^{2k(\alpha+d-1-8\varepsilon)}$.

Proof.

(i) Recall from the above discussion that $\varepsilon < \min\left\{\frac{\varepsilon_0(d')}{10}, d'\right\}$. By Lemma 2.39 (i)

$$n^{k(\alpha+d'-1-2\varepsilon)} \le \mathbb{E}\left(|A_n^{(k)} \cap X_n|\right) \le n^{k(\alpha+d'-1+2\varepsilon)}.$$

We then have the inequality by $d - \varepsilon \le d' \le d + \varepsilon$.

(ii) Because $\varepsilon_0(d') > 0$, $\boldsymbol{X}$ has d'-small self-intersection. By Lemma 2.39 (ii) and the fact that $d' \leq d+\varepsilon$,

$$\operatorname{Var}(|A_n^{(k)} \cap X_n|) \leq kn^{2k(\alpha+d'-1-9\varepsilon)} \leq kn^{2k(\alpha+d-1-8\varepsilon)}.$$

□

Lemma 2.42. *Let $0 < \varepsilon < \min\{\frac{\varepsilon_0}{10}, \frac{d}{2}\}$. Let $\boldsymbol{A}$ be a sequence of uniform random subsets of $\boldsymbol{E}$ with density $d' \in [d-\varepsilon, d+\varepsilon]$. Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α. Suppose that $\alpha + d - 1 > 4\varepsilon > 0$ and that $\boldsymbol{X}$ satisfies the d-small self-intersection condition. If $n \geq \max\left\{n_\varepsilon, (1+2k)^{\frac{1}{\varepsilon}}\right\}$, then*

$$\mathbf{Pr}\left(n^{k(\alpha+d-1-4\varepsilon)} \leq |A_n^{(k)} \cap X_n| \leq n^{k(\alpha+d-1+4\varepsilon)}\right) > 1 - kn^{-10k\varepsilon}.$$

Proof. Denote $B_n = A_n^{(k)} \cap X_n$. By Lemma 2.41 (i) and the fact that $n^{k\varepsilon} \geq 2$,

$$n^{k(\alpha+d-1-4\varepsilon)} \leq \frac{1}{2} n^{k(\alpha+d-1-3\varepsilon)} \leq \frac{1}{2}\mathbb{E}\left(|B_n|\right)$$

and

$$\frac{3}{2}\mathbb{E}\left(|B_n|\right) \leq \frac{3}{2} n^{k(\alpha+d-1+3\varepsilon)} \leq n^{k(\alpha+d-1+4\varepsilon)}.$$

By Chebyshev's inequality,

$$\begin{aligned} &\mathbf{Pr}\left(n^{k(\alpha+d-1-4\varepsilon)} \leq |B_n| \leq n^{k(\alpha+d-1+4\varepsilon)}\right) \\ \geq &\mathbf{Pr}\left(\Big||B_n| - \mathbb{E}\left(|B_n|\right)\Big| \leq \frac{1}{2}\mathbb{E}\left(|B_n|\right)\right) \\ \geq &1 - \frac{4\operatorname{Var}(|B_n|)}{\mathbb{E}\left(|B_n|\right)^2}. \end{aligned}$$

By Lemma 2.41 (i) and (ii),

$$\begin{aligned} \frac{4\operatorname{Var}(|B_n|)}{\mathbb{E}\left(|B_n|\right)^2} \leq &\frac{kn^{2k(\alpha+d-1-8\varepsilon)}}{n^{2k(\alpha+d-1-3\varepsilon)}} \\ \leq &kn^{-10k\varepsilon}. \end{aligned}$$

□

Proof of Theorem 2.40.

Let $\varepsilon > 0$ be a small real number to be specified. Denote $Q_n = \{n^{d-\varepsilon} \leq |A_n| \leq n^{d+\varepsilon}\}$ and

$$\mathbb{N}_{\boldsymbol{A},\varepsilon,n} := \left\{\ell \in \mathbb{N} \mid n^{d-\varepsilon} \leq \ell \leq n^{d+\varepsilon} \text{ and } \mathbf{Pr}(|A_n| = \ell) > 0\right\}.$$

By the densability of $\boldsymbol{A}$, Q_n is an a.a.s. event. Denote by $\mathbf{Pr}_{Q_n} := \mathbf{Pr}\left(\cdot \mid Q_n\right)$ the probability measure under the condition Q_n. Define similarly $\mathbb{E}_{Q_n}$ and Var_{Q_n}. In order to prove that some sequence of properties (R_n) is a.a.s. true, by proposition 2.5, it is enough to prove that $\mathbf{Pr}_{Q_n}(\overline{R_n}) \xrightarrow[n\to\infty]{} 0$.

To apply Lemma 2.41 and Lemma 2.42, we assume that $\varepsilon < \min\left\{\frac{d}{2}, \frac{\varepsilon_0(d)}{20}\right\}$.

(i) Suppose that $\alpha + d < 1$. Assume in addition that $\varepsilon < \frac{1-d-\alpha}{3}$.

We shall prove that

$$\mathbf{Pr}_{Q_n}(A_n^{(k)} \cap X_n \neq \emptyset) = \mathbf{Pr}_{Q_n}(|A_n^{(k)} \cap X_n| \geq 1) \xrightarrow[n\to\infty]{} 0.$$

By the formula of total probability and Markov's inequality,

$$\begin{aligned}\mathbf{Pr}_{Q_n}\left(|A_n^{(k)} \cap X_n| \geq 1\right) &\leq \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l)\mathbf{Pr}\left(|A_n^{(k)} \cap X_n| \geq 1 \;\middle|\; |A_n| = l\right) \\ &\leq \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l)\mathbb{E}\left(|A_n^{(k)} \cap X_n| \;\middle|\; |A_n| = l\right).\end{aligned}$$

By a change of variable $l = n^{d'}$ with $d - \varepsilon \leq d' \leq d + \varepsilon$, apply Lemma 2.41 (i),

$$\begin{aligned}\mathbf{Pr}_{Q_n}\left(|A_n^{(k)} \cap X_n| \geq 1\right) &\leq \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l = n^{d'})n^{\alpha+d-1+3\varepsilon} \\ &\leq n^{\alpha+d-1+3\varepsilon} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

□

(ii) Suppose that $\alpha + d > 1$. Assume in addition that $\varepsilon < \frac{\alpha+d-1}{4}$, so that we can apply Lemma 2.42.

We shall prove that

$$\mathbf{Pr}_{Q_n}\left(n^{k(\alpha+d-1-4\varepsilon)} \leq |A_n^{(k)} \cap X_n| \leq n^{k(\alpha+d-1+4\varepsilon)}\right) \xrightarrow[n\to\infty]{} 1.$$

By Proposition 2.11 (decomposition into uniform random subsets), Lemma 2.42 and a change of variables $l = n^{d'}$,

$$\begin{aligned}&\mathbf{Pr}_{Q_n}\left(n^{k(\alpha+d-1-4\varepsilon)} \leq |A_n^{(k)} \cap X_n| \leq n^{k(\alpha+d-1+4\varepsilon)}\right) \\ =& \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l)\mathbf{Pr}\left(n^{k(\alpha+d-1-4\varepsilon)} \leq |A_n^{(k)} \cap X_n| \leq n^{k(\alpha+d-1+4\varepsilon)} \;\middle|\; |A_n| = l = n^{d'}\right) \\ \geq& \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l)\left(1 - kn^{-10k\varepsilon}\right) \\ \geq& 1 - kn^{-10k\varepsilon} \xrightarrow[n\to\infty]{} 1.\end{aligned}$$

□

2.4 Applications to random groups

Fix an alphabet $X_m = \{x_1, \ldots, x_m\}$ as generators of groups. Let B_ℓ be the set of cyclically reduced words of $X_m^{\pm}$ of length at most ℓ. Recall the definition of a random group in the permutation invariant density model.

Definition 2.43. *A sequence of random groups $(G_\ell(m,d))_{\ell\in\mathbb{N}}$ with m generators at density $0 \leq d \leq 1$ is defined by*

$$G_\ell(m,d) = \langle X_m | R_\ell \rangle$$

where $\boldsymbol{R} = (R_\ell)$ is a sequence of permutation invariant random subsets of $\boldsymbol{B} = (B_\ell)$ with density d.

By the random-fixed intersection formula (Theorem 2.28), we have the following properties.

Properties. Let $(G_\ell(m,d))$ be a sequence of random groups with m generators at density d, R_ℓ is the set of relators of $G_\ell(m,d)$.

- For any $s > 0$, a.a.s. there is no relator of length shorter than $(1-d-s)\ell$ in R_ℓ.
- For any $s > 0$, a.a.s. there are relators of length shorter than $(1-d+s)\ell$ in R_ℓ.

  The set of such relators is with density s in B_ℓ.
- Let w be a reduced word of length $\lfloor (d-s)\ell \rfloor$ with $s > 0$. Then a.a.s. there are relators in R_ℓ having w as a cyclic subword.

  The set of such relators is with density s in B_ℓ.
- If $d < 1/2$, then a.a.s. there is no relator of the form $r = w^n$ with $n \geq 2$ in R_ℓ.

Proof. Recall that the cardinality of B_ℓ is $|B_\ell| = (2m-1)^{\ell+O(1)}$, and that (R_ℓ) is a sequence of permutation invariant random subsets of density d.

- Let A_ℓ be the set of words of length shorter than $(1-d-s)\ell$. Its cardinality is $(2m-1)^{(1-d-s)\ell+O(1)}$, hence $\boldsymbol{A} = (A_\ell)$ a sequence of fixed subsets with density $1-d-s$ in $\boldsymbol{B}$. By the intersection formula a.a.s. $R_\ell \cap A_\ell = \emptyset$.
- Let A_ℓ be the set of words of length shorter than $(1-d+s)\ell$. By the same argument, $\boldsymbol{A} = A_\ell$ is a sequence of fixed subsets with density $1-d+s$ in $\boldsymbol{B}$. By the intersection formula, a.a.s. $R_\ell \cap A_\ell$ is not empty, and the sequence of with density $(1-d+s)+d-1 = s$.
- Let A_ℓ be the set of cyclically relators of length ℓ having w as a subword. There are ℓ choices for the position of w as a cyclic subword, and $(2m-1)^{\ell-\lfloor (d-s)\ell \rfloor -1}(2m-2) \quad (2m-1)^{\ell-\lfloor (d-s)\ell \rfloor}$ choices for other letters. So $\boldsymbol{A} = (A_\ell)$ is a sequence of subsets of $\boldsymbol{B}$ with density $1-d+s$. By the intersection formula, $\boldsymbol{R} \cap \boldsymbol{A}$ is with density s.
- For any integer $2 \leq n \geq \ell$, the number of words of the form w^n of length at most ℓ is smaller than $\ell(2m)(2m-1)^{\ell/n}$. So the number of words of the form w^n for some $n \geq 2$ of length at most ℓ is smaller than $\ell^2(2m)(2m-1)^{\ell}/2$. It is a set of density $1/2$. By the intersection formula, a.a.s. it does not intersect R_ℓ.

□

We have the following proposition, similar to Proposition 2.35 and Proposition 2.38. This result will be useful in the next two chapters.

Proposition 2.44. *Let $\boldsymbol{R} = (R_\ell)$ be a densable sequence of permutation invariant random subsets of $\boldsymbol{B} = (B_\ell)$ with density d. Let $0 < \varepsilon < d/2$. Denote by Q_ℓ the event $(2m-1)^{(d-\varepsilon)\ell} \leq |R_\ell| \leq (2m-1)^{(d+\varepsilon)\ell}$, we have a.a.s. Q_ℓ by the characterization of densability (Proposition 2.6). Let $r_1,\dots,r_k$ be distinct elements in R_ℓ. For ℓ large enough,*

$$(2m-1)^{k\ell(d-1-2\varepsilon)} \leq \mathbf{Pr}\left(r_1,\dots,r_k \in R_\ell \mid Q_\ell\right) \leq (2m-1)^{k\ell(d-1+2\varepsilon)}.$$

□

Proof. The key is to decompose the permutation invariant random subset R_ℓ into uniform random subsets, and apply Proposition 2.38 (i). As $|B_\ell| = (2m-1)^{\ell+O(1)}$, we choose ℓ large enough such that

$$|B_\ell| \geq (1+2k)^{\frac{4}{\varepsilon}},$$

$$|B_\ell|^{d-5\varepsilon/4} \leq (2m-1)^{(d-\varepsilon)\ell} \leq (2m-1)^{(d+\varepsilon)\ell} \leq |B_\ell|^{d+5\varepsilon/4},$$

and

$$(2m-1)^{k\ell(d-1-2\varepsilon)} \leq |B_\ell|^{k(d-1-3\varepsilon/2)} \leq |B_\ell|^{k(d-1+3\varepsilon/2)} \leq (2m-1)^{k\ell(d-1+2\varepsilon)}.$$

So under the condition Q_ℓ, we have

$$|B_\ell|^{d-5\varepsilon/4} \leq |R_\ell| \leq |B_\ell|^{d+5\varepsilon/4}.$$

Denote

$$\mathbb{N}_{\ell,\varepsilon} = \{s \in \mathbb{N} \mid |B_\ell|^{d-5\varepsilon/4} \leq s \leq |B_\ell|^{d+5\varepsilon/4}, \mathbf{Pr}(|R_\ell| = s) \neq 0\}.$$

Now we can decompose R_ℓ into uniform random subsets, with a change of variables $s = |B_\ell|^{d'}$, $d' \in \left[d - \frac{5\varepsilon}{4}, d + \frac{5\varepsilon}{4}\right]$. By Proposition 2.11,

$$\mathbf{Pr}\left(r_1, \dots, r_k \in R_\ell \mid Q_\ell\right) = \sum_{s \in \mathbb{N}_{\ell,\varepsilon}} \mathbf{Pr}\left(r_1, \dots, r_k \in R_\ell \;\middle|\; |R_\ell| = s = |B_\ell|^{d'}\right) \mathbf{Pr}\left(|R_\ell| = s \mid Q_\ell\right).$$

As $|B_\ell| \geq (1+2k)^{\frac{4}{\varepsilon}}$ and $\frac{\varepsilon}{4} < d - \frac{5\varepsilon}{4} \leq d'$ for any $d' \in \left[d - \frac{5\varepsilon}{4}, d + \frac{5\varepsilon}{4}\right]$, we can apply Proposition 2.38. For any $d' \in \left[d - \frac{5\varepsilon}{4}, d + \frac{5\varepsilon}{4}\right]$,

$$|B_\ell|^{k(d'-1-\varepsilon/4)} \leq \mathbf{Pr}\left(r_1, \dots, r_k \in R_\ell \;\middle|\; |R_\ell| = |B_\ell|^{d'}\right) \leq |B_\ell|^{k(d'-1+\varepsilon/4)},$$

so

$$|B_\ell|^{k(d-1-3\varepsilon/2)} \leq \mathbf{Pr}\left(r_1, \dots, r_k \in R_\ell \;\middle|\; |R_\ell| = |B_\ell|^{d'}\right) \leq |B_\ell|^{k(d-1+3\varepsilon/2)}.$$

Because $\sum_{s \in \mathbb{N}_{\ell,\varepsilon}} \mathbf{Pr}\left(|R_\ell| = s = |B_\ell|^{d'} \;\middle|\; Q_\ell\right) = 1$, we have

$$|B_\ell|^{k(d-1-3\varepsilon/2)} \leq \mathbf{Pr}\left(r_1, \dots, r_k \in R_\ell \mid Q_\ell\right) \leq |B_\ell|^{k(d-1+3\varepsilon/2)}.$$

Hence,

$$(2m-1)^{k\ell(d-1-2\varepsilon)\ell} \leq \mathbf{Pr}\left(r_1, \dots, r_k \in R_\ell \mid Q_\ell\right) \leq (2m-1)^{k\ell(d-1+2\varepsilon)\ell}.$$

□

2.4.1 Phase transition at density $1/2$

Theorem 2.45 (Gromov, phase transition at density 1/2)**.** *Let $(G_\ell(m,d))$ be a sequence of random groups with density d.*

- (i) *If $d > 1/2$, then a.a.s $G_\ell(m,d)$ is isomorphic to the trivial group.*
- (ii) *If $d < 1/2$, then a.a.s $G_\ell(m,d)$ is a hyperbolic group, and the Cayley graph $\mathrm{Cay}(G_\ell(m,d), X_m)$ is δ-hyperbolic with $\delta = \frac{4\ell}{1-2d}$.*

  In addition, for any $s > 0$, a.a.s. every reduced van Kampen diagram D of $G_\ell(m,d)$ satisfies the isoperimetric inequality

$$|\partial D| \geq (1-2d-s)\ell|D|.$$

In [Oll04] 2.1 (or [Oll05] I.2.b), Ollivier proved the first assertion by the probabilistic pigeon-hole principle. We give a proof here by the intersection formula (Theorem 2.25 and Theorem 2.28).

Proof of Theorem 2.45 (i). Let $x \in X_m$ be a generator. Let A_ℓ be the set of cyclically reduced words that does not start or end by x, of lengths at most $\ell - 1$ (so that $xA_\ell \subset B_\ell$). The sequences $\boldsymbol{A} = (A_\ell)$ and $x\boldsymbol{A} = (xA_\ell)$ are sequences of fixed subsets of $\boldsymbol{B} = (B_\ell)$ of density 1. By the random-fixed intersection formula (Theorem 2.28), the sequences $x(\boldsymbol{R} \cap \boldsymbol{A})$ and $\boldsymbol{R} \cap x\boldsymbol{A}$ are sequences of permutation invariant random subsets of $x\boldsymbol{A}$ of density d.

By the (random-random) intersection formula (Theorem 2.25), the intersection

$$(\boldsymbol{R} \cap x\boldsymbol{A}) \cap x(\boldsymbol{R} \cap \boldsymbol{A}) = \boldsymbol{R} \cap x\boldsymbol{R} \cap x\boldsymbol{A}$$

is a sequence of permutation invariant random subsets of $x\boldsymbol{A}$ with density $(2d - 1) > 0$. So a.a.s. (when $\ell \to \infty$) the set $R_\ell \cap xR_\ell \cap xA_\ell$ is not empty. Thus, a.a.s. there exists a word $w \in A_\ell$ such that $w \in R_\ell$ and $xw \in R_\ell$. It gives a.a.s. $x = 1$ in G_ℓ by canceling w.

This argument works for any generator $x \in X_m$. By intersecting a finite number of a.a.s. events, a.a.s. all generators $x \in X_m$ are trivial in G_ℓ. Hence, a.a.s. G_ℓ is isomorphic to the trivial group. □

The proof of Theorem 2.45 (ii) is based on the study of *van Kampen diagrams* [Kam33]. See [Gro93] 9.B. for the original idea by Gromov, [Oll04] Section 2.2 or [Oll05] Section V for a proof by Ollivier. The local case is a corollary of our Theorem 4.3 or Theorem 3.23. For a precise estimation of hyperbolicity constants, c.f. [Cha94] Lemma 3.11 by C. Champetier.

2.4.2 Phase transition at density $\lambda/2$

We prove that there is a phase transition for the λ-small cancellation condition (see [LS77] for a definition). We will give a simpler proof in Section 4.3, Theorem 4.23. using the characterization of existence of diagrams (Theorem H, Theorem 4.15).

Theorem 2.46. *Let $\boldsymbol{G}(m, d) = (G_\ell(m, d))$ be a sequence of random groups with density d. Let $\lambda \in]0, 1[$.*

1. *If $d < \lambda/2$, then a.a.s. $G_\ell(m, d)$ satisfies $C'(\lambda)$.*
2. *If $d > \lambda/2$, then a.a.s. $G_\ell(m, d)$ **does not** satisfy $C'(\lambda)$.*

Proof.

1. Recall that (Lyndon-Schupp [LS77] p.240) a *piece* with respect to a set of relators is a cyclic sub-word that appears at least twice. There are two cases to verify.

   (a) Let A_ℓ be the set of cyclically reduced words of length at most ℓ having a piece longer than λ times itself that appears twice, and these two paths do not overlap (figure 1). We shall prove that a.a.s. the intersection $A_\ell \cap R_\ell$ is empty.

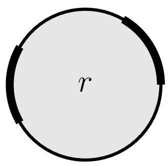


figure 1

We estimate first the number of relators of length $t \leq \ell$ with a piece of length $s \geq \lambda t$. There are $2t$ ways (including orientations) to choose the first position of the piece, and $2t - s$ ways to choose the second position (note that because r is reduced, it can not overlay the first one if they are with opposite orientations). For each way of positioning we can determine freely $t - s$ letters, each with $(2m - 1)$ choices, except for the first letter and the last letter having respectively $2m$ and $2m - 2$ or $2m - 1$ choices. So this number is $2t(2t - s)C(m)(2m - 1)^{t-s}$ where $C(m)$ is a real number that depends only on m. Hence,

$$|A_\ell| = \sum_{t=1}^{\ell} \sum_{s=\lfloor \lambda t \rfloor}^{t} 2t(2t - s)C(m)(2m - 1)^{t-s} = (2m - 1)^{(1-\lambda)\ell + o(\ell)},$$

which means that (A_ℓ) is a sequence of fixed subsets of (B_ℓ) with density $1 - \lambda$. By the intersection formula (Theorem 2.28), because $1 - \lambda + d < 1$, we have a.a.s.

$$A_\ell \cap R_\ell = \emptyset.$$

The case that the two paths on the relator are overlapping can be treated similarly. Note that if they overlap then they are in the same direction, otherwise we would either have $x = x^{-1}$ for a generator $x \in X^{\pm}$ or a sub-path xx^{-1} appearing on the reduced relator. For the set of such relators with overlapping lengths longer than $\lambda/2$, we get a subset of density at most $1-\lambda$ because the $(1 - \lambda)\ell$ letters excluding one of the pieces determine the relator; for the set of such relators with overlapping lengths shorter than $\lambda/2$, we get a subset of density at most $1-\lambda/2$; because the $(1 - \lambda/2)\ell$ letters excluding the non-overlapping part of one of the pieces determine the relator.

(b) Let X_ℓ be the set of distinct pairs of relators r_1, r_2 in B_ℓ having a piece (figure 2) longer than $\lambda \min\{|r_1|, |r_2|\}$. It is a fixed subset of $B_\ell^{(2)}$. We shall prove that a.a.s the intersection $X_\ell \cap R_\ell^{(2)}$ is empty.

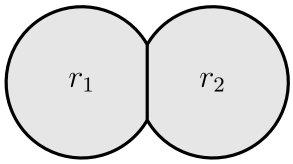


figure 2

There are $4\ell^2$ possible positions for pieces, $(2m-1)^{\ell+o(\ell)}$ choices for r_1 and $(2m-1)^{\ell-\lambda\ell+o(\ell)}$ choices for r_2. So

$$|X_\ell| = (2m - 1)^{(2-\lambda)\ell+o(\ell)},$$

which means that (X_ℓ) is a sequence of fixed subsets of $(B_\ell^{(2)})$ with density $1 - \frac{\lambda}{2}$. By the multi-dimension intersection formula (Theorem 2.40 (i)), because $1 - \frac{\lambda}{2} + d < 1$, we have a.a.s.

$$X_\ell \cap R_\ell^{(2)} = \emptyset.$$

2. Take the sequence of sets $\boldsymbol{X} = (X_\ell)$ constructed in 1(b). We shall prove that a.a.s. the intersection $X_\ell \cap R_\ell^{(2)}$ is *not* empty. We have already

$$\operatorname{dens} \boldsymbol{X} + \operatorname{dens} \boldsymbol{R}^{(2)} > 1.$$

To apply Theorem 2.40(ii), we need to calculate the size of the self-intersection

$$Y_{1,\ell} = \{(x_1, x_2) \in X_\ell^2 \,|\, |x_1 \cap x_2| = 1\}.$$

Take $x_1 = (r_1, r_2)$ and $x_2 = (r_1, r_3)$ where r_1, r_2, r_3 are three different relators in B_ℓ. There are $(2m-1)^{\ell+o(\ell)}$ choices for r_1, $(2m-1)^{\ell-\lambda\ell+o(\ell)}$ choices for r_2 and $(2m-1)^{\ell-\lambda\ell+o(\ell)}$ choices for r_3. The other three cases ($x_2 = (r_2, r_3), (r_3, r_1)$ or (r_3, r_2)) are symmetric. Multiply these numbers, we have

$$|Y_{1,\ell}| = (2m-1)^{3\ell-2\lambda\ell+o(\ell)}.$$

The density of $\boldsymbol{Y}_1 = (Y_{1,\ell})$ is $\frac{3-2\lambda}{4}$ in $(B_\ell^{(2)})^2$. As $d > 0$, we have $\frac{3-2\lambda}{4} < 1 - \frac{\lambda}{2} + \frac{1}{4}(d-1)$, which implies

$$\operatorname{dens} \boldsymbol{Y}_1 < \operatorname{dens} \boldsymbol{X} + (d-1)\frac{1}{2 \times 2}.$$

Thus we have the d-small self intersection condition (definition 2.33). By the multidimensional intersection formula, a.a.s.

$$X_\ell \cap R_\ell^{(2)} \neq \emptyset.$$

□

2.4.3 Every $(m-1)$-generated subgroup is free

Fix an integer $k \geq 1$. Recall that the few relator model of random groups is constructed by

$$G_\ell = \langle x_1, \ldots, x_m | r_1, \ldots, r_k \rangle$$

where $R_\ell = \{r_1, \ldots, r_k\}$ is a random subset of B_ℓ given by the uniform distribution on all subsets of B_ℓ with cardinality k.

It is a sequence of (permutation invariant) random groups with density $d = 0$. By Proposition 2.46, a.a.s. G_ℓ satisfies $C'(\lambda)$ for arbitrary small $\lambda > 0$. Recall the Arzhantseva-Ol'shanskii's result in [AO96].

Theorem 2.47 (Arzhantseva-Ol'shanskii, [AO96] Theorem 1)**.** *Let (G_ℓ) be a sequence of random groups with k relators. Then a.a.s. every $(m-1)$-generated subgroup of G_ℓ is free.*

Combining the (random-fixed) intersection formula and their arguments, we can extend the density from 0 to a small number depending on m.

Theorem 2.48. *Let $(G_\ell(m,d))$ be a sequence of random groups with density $0 \leq d < \frac{1}{120m^2\ln(2m)}$. Then a.a.s. every $(m-1)$-generated subgroup of $G_\ell(m,d)$ is free.*

Let us recall the definition of "μ-readable words" in [AO96] §2. Denote $X_m = \{x_1, \ldots, x_m\}$ the set of relators.

Definition 2.49 ([AO96] §2)**.** *Let $0 < \mu \leq 1$. A cyclically reduced word w of $X_m^\pm$ of length ℓ is μ-readable if there exists a graph Γ marked by $X_m^\pm$ with the following properties :*

(a) the number of edges of Γ is less than $\mu\ell$;

(b) the rank of Γ is at most $m-1$;

(c) the word w can be read along some path of Γ.

Note that the condition (b) is essential, because every word of $X_m^{\pm}$ can be read along the wedge of m circles of length 1 marked by $x_1, \dots, x_m$ respectively.

Let M_ℓ^μ be the set of words $r \in B_\ell$ having a cyclic sub-word $w < r$ such that $|w| \geq \frac{1}{2}|r|$ and w is μ-readable. The following two lemmas are from [AO96].

Lemma 2.50 ([AO96] Lemma 4)**.** *If $\mu < \log_{2m}\left(1 + \frac{1}{4m-4}\right)$, then there exists a constant $C(\mu, m)$ such that*

$$|M_\ell^\mu| \leq C(\mu, m)\ell^2 \left(2m - \frac{5}{4}\right)^\ell.$$

□

Recall that $|B_\ell| = (2m-1)^{\ell + O(1)}$, so (M_ℓ^μ) is a densable sequence of subsets of (B_ℓ) with density $\log_{2m-1}\left(2m - \frac{5}{4}\right)$.

Lemma 2.51 ([AO96] §4)**.** *Let $G = \langle X | R \rangle$ be a group presentation where $X = \{x_1, \dots, x_m\}$ and R is a subset of B_ℓ. Suppose that*

$$\mu < \log_{2m}\left(1 + \frac{1}{4m-4}\right) \quad \text{and} \quad \lambda \leq \frac{\mu}{15m + 3\mu}.$$

If R does not intersect M_ℓ^μ, has no true powers, and satisfies $C'(\lambda)$, then every $(m-1)$-generated subgroup of G is free. □

Proof of Theorem 2.48. We look for a density $d(m) \leq 1/2$ such that for any $d < d(m)$, a.a.s. the random group $G_\ell(m, d) = \langle X_m | R_\ell \rangle$ satisfies the conditions of Lemma 2.51 with $\mu = \log_{2m}\left(1 + \frac{1}{4m-4}\right) - \varepsilon$ and $\lambda = \frac{\mu}{15m+3\mu}$ with an arbitrary small $\varepsilon > 0$.

The set of true powers in B_ℓ is with density $1/2$. By the intersection formula (Theorem 2.28), because $d < 1/2$, a.a.s. R_ℓ has no true powers. By Lemma 2.50 and the intersection formula, we need $d(m) < 1 - \text{dens}(M_\ell^\mu) < 1 - \log_{2m-1}\left(2m - \frac{5}{4}\right)$ so that a.a.s. R_ℓ does not intersect M_ℓ^μ by the intersection formula.

At the end we need a.a.s. R_ℓ satisfies $C'(\lambda)$ with

$$\lambda = \frac{\log_{2m}\left(1 + \frac{1}{4m-4}\right) - \varepsilon}{15m + 3\log_{2m}\left(1 + \frac{1}{4m-4}\right) - 3\varepsilon}.$$

By Theorem 2.46, we need $d(m) < \lambda/2$. Note that this inequality implies the previous one. For ε small enough we have $\lambda > \frac{1}{60m^2 \ln(2m)}$. It is enough to take

$$d(m) = \frac{1}{120m^2 \ln(2m)}.$$

□

Chapter 3

The Freiheitssatz for random groups

Let $(G_\ell(m,d)) = \langle X_m | R_\ell \rangle$ be a sequence of random groups at density d, in the permutation invariant density model. In this chapter, we prove the phase transition at density $d_r = \min\left\{\frac{1}{2}, 1 - \log_{2m-1}(2r-1)\right\}$ (Theorem E): If $d > d_r$, then a.a.s. the first r generators $x_1, \ldots, x_r$ generate the whole group $G_\ell(m,d)$; if $d < d_r$, then a.a.s. the first r generators $x_1, \ldots, x_r$ freely generate a free subgroup of $G_\ell(m,d)$. Our main result (Theorem 3.22) is a strong version of this result, replacing "the first r generators" by a Stallings graph of rank r of bounded size.

In the first section, we recall some essential tools in combinatorial group theory (Stallings graphs [Sta83] and van Kampen diagrams [Kam33]), and introduce *distortion van Kampen diagrams* to study the distortion of subgroups of a finitely presented group. In Section 3.2, we study abstract van Kampen diagrams introduced by Y. Ollivier in [Oll04], and apply his idea to define *abstract distortion diagrams*. The main technical lemma for our main theorem (Theorem 3.22) is to estimate the *number of fillings* of a given abstract distortion diagram (Lemma 3.21). In the last section, we state a local result on distortion van Kampen diagrams (Lemma 3.23) and prove the main theorem by this lemma. The last subsection is for the proof of this lemma.

3.1 Preliminaries on group theory

In this section, we fix a finite group presentation $G = \langle X | R \rangle$ where X is the set of generators and R is the set of relators. A word u in the alphabet $X^{\pm}$ is called *reduced* if it has no sub-words of type xx^{-1} or $x^{-1}x$ for any $x \in X$. If u and v are words that represent the same element in G, we denote $u =_G v$.

3.1.1 Stallings graphs (graphs generating subgroups)

We consider oriented combinatorial graphs and 2-complexes as defined in Chapter III.2. of Lyndon and Schupp [LS77].

A *graph* is a pair $\Gamma = (V, E)$ where V is the set of *vertices* (also called points) and E is the set of (oriented) *edges*. Every edge $e \in E$ has a starting point $\alpha(e) \in V$, an ending point $\omega(e) \in V$ and an inverse edge $e^{-1} \in E$, satisfying $\alpha(e^{-1}) = \omega(e)$, $\omega(e^{-1}) = \alpha(e)$ and $(e^{-1})^{-1} = e$. The vertices $\alpha(e)$ and $\omega(e)$ are called the endpoints of the edge e. An *undirected edge* is a pair of inverse edges $\{e, e^{-1}\}$. The *size* $|\Gamma|$ of a graph is the number of its undirected edges. The *rank* $\operatorname{rk}(\Gamma)$ is the rank of its fundamental free group, which equals to $|\Gamma| - |V| + 1$ by Euler's characteristic.

A *path* on a graph Γ is a non-empty finite sequence of edges $p = e_1 \ldots e_k$ such that $\omega(e_i) = \alpha(e_{i+1})$ for $i \in \{1, \ldots k-1\}$. The starting point and the ending point of the path p are defined by $\alpha(p) = \alpha(e_1)$

and $\omega(p) = \omega(e_k)$. The inverse of p is the path $p^{-1} = e_k^{-1} \dots e_1^{-1}$. A path is called *reduced* if there is no subsequence of the form ee^{-1}. A *loop* is a path whose starting point and ending point coincide. In this case $\alpha(p) = \omega(p)$ is called the starting point of the loop. A loop $p = e_1 \dots e_k$ is *cyclically reduced* if it is a reduced path with $e_k \neq e_1^{-1}$.

An *arc* of a graph Γ is a reduced path passing only by vertices of degree 2, except possibly for its endpoints. A *maximal arc* is an arc that is not a subpath of another arc. By definition, the endpoints of a maximal arc can not be of degree 2. The following fact for finite connected graphs can be deduced by Euler's characteristic.

Lemma 3.1. *Let Γ be a finite connected graph of rank $r \geq 1$ with no vertices of degree 1.*

1. *The number of vertices of degree at least 3 is bounded by $2(r-1)$.*
2. *The number of maximal arcs of Γ is bounded by $3(r-1)$.*

Proof. If an arc of Γ can not be extended to a maximal arc, then Γ is a simple cycle and $r = 1$. In this case, both assertions are true. Otherwise, suppose that any arc can be extended to a maximal arc, so Γ can be divided into maximal arcs.

Let v be the number of vertices of degree at least 3; let a be the number of maximal arcs. By Euler's characteristic, $a - v = r - 1$. By a combinatorial fact on graphs, $3v \leq 2a$, so $a \geq \frac{3}{2}v$. Hence $v \leq 2(r-1)$ and $a \leq 3(r-1)$. □

Lemma 3.2. *The number of topological types of finite connected graphs of rank at most r with no vertices of degree 1 is bounded by $(2r)^{6r}$.*

Proof. If $r = 1$ then the only topological type is a simple cycle. If $r \geq 2$, we may draw $a \leq 3(r-1)$ arcs on a set of $v \leq 2(r-1)$ vertices. There are at most $(v^2)^a \leq (2r)^{6r}$ ways. □

A *labeled graph* (with respect to the alphabet X) is a graph $\Gamma = (V, E)$ with a labelling function on edges by generators $\varphi : E \to X^{\pm}$, satisfying $\varphi(e^{-1}) = \varphi(e)^{-1}$. We denote briefly $\Gamma = (V, E, \varphi)$. The labeling function φ extends naturally on the paths of Γ. If $p = e_1 \dots e_k$ is a path of Γ, then the word $\varphi(p) = \varphi(e_1) \dots \varphi(e_k)$ is called the *labeling word* of p. We say that a word u is *readable* on a labeled graph Γ if there exists a path p of Γ whose labeling word is u.

Labeled graphs are considered by Stallings [Sta83] to represent subgroups of a free group. Let $\Gamma = (V, E, \varphi)$ be a finite connected labeled graph. Labeling words of the loops starting at a vertex $o \in V$ form a subgroup H of $G = \langle X|R \rangle$, which is the image of the fundamental group $\pi_1(\Gamma, o)$ by the group homomorphism induced by φ. If H is a conjugate of the subgroup $\varphi(\pi_1(\Gamma, o))$ in G for some $o \in V$, we say that H is a subgroup *generated* by the labeled graph Γ.

Conversely, any finitely generated subgroup H can be generated by a labeled graph. One can choose a system of generators $h_1, \dots, h_r$ of H, and label them on the wedge of r simple cycles of lengths $|h_1|, \dots, |h_r|$. A labeled graph is *reduced* if it has no pair of edges with the same label and starting point, and, it has no vertices of degree 1. By doing reductions on the construction above, if H is a subgroup of rank r, then there is a *reduced* labeled graph of rank r that generates H. We refer the reader to [MM93] and [AO96] for details.

3.1.2 Van Kampen diagrams

A *2-complex* is a triplet $D = (V, E, F)$, where (V, E) is a graph and F is the set of (oriented) faces. Every face $f \in F$ has a boundary ∂f, which is a cyclically reduced loop of (V, E), and an inverse face $f^{-1} \in F$ satisfying $\partial(f^{-1}) = (\partial f)^{-1}$ and $(f^{-1})^{-1} = f$. An *undirected face* is a pair of inverse faces $\{f, f^{-1}\}$. The size $|D|$ is the number of undirected faces.

Note that our definition is slightly different from the definition of [LS77] chapter III.9: Every face $f \in F$ has a starting point and an orientation given by its boundary path ∂f. If $\partial f = e_1 \dots e_k$, for $1 \leq i \leq k$, we say that e_i is *attached* to f and is the i-th boundary edge of f. We say that the undirected edge $\{e_i, e_i^{-1}\}$ is attached to the undirected face $\{f, f^{-1}\}$. An edge (or an undirected edge) is called *isolated* if it is not attached to any face.

A *van Kampen diagram* with respect to a group presentation $G = \langle X|R \rangle$ is a finite, planar (embedded in $\mathbb{R}^2$) and simply connected 2-complex $D = (V, E, F)$ with two compatible labeling functions: labels on edges by generators $\varphi_1 : E \to X^{\pm}$, and labels on faces by relators $\varphi_2 : F \to R^{\pm}$. Compatible means that (V, E, φ_1) is a labeled graph, $\varphi_2(f^{-1}) = \varphi_2(f)^{-1}$ and $\varphi_1(\partial f) = \varphi_2(f)$. We denote briefly $D = (V, E, F, \varphi_1, \varphi_2)$. Note that if a van Kampen diagram D has no isolated edges (for example, a disk), then the labeling function φ_1 on edges is determined by the labeling function φ_2 on faces.

According to [CH82] p.159, a van Kampen diagram is either a disk or a concatenation of disks and segments. The boundary ∂D is the boundary of $\mathbb{R}^2 \backslash D$, which is a sub-graph of its underlying graph (V, E). A *boundary path* is a (combinatorial) path on ∂D defined as in [LS77] p.150. A *boundary word* of D is the labeling word of a boundary path, unique up to cyclic conjugations and inversions. The *boundary length* of D is the length of a boundary path, denoted by $|\partial D|$.

Let $D = (V, E, F, \varphi_1, \varphi_2)$ be a van Kampen diagram. A pair of faces $f, f' \in F$ is *reducible* if they have the same label and there is a common edge on their boundaries at the same position (see Figure **3.1**). A van Kampen diagram is called *reduced* if there is no reducible pair of faces.

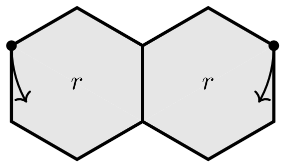


Figure 3.1: a reducible pair of faces

E. van Kampen showed in [Kam33] that a word u of $X^{\pm}$ is trivial in a finitely presented group $G = \langle X|R \rangle$ if and only if it is a boundary word of a van Kampen diagram of G. In [Ols91] §11.6, A. Ol'shanskii improved this result to *reduced* diagrams.

Lemma 3.3 (Van Kampen's lemma, Ol'shanskii's version)**.** *A word w of $X^{\pm}$ is trivial in $G = \langle X|R \rangle$ if and only if it is a boundary word of a* ***reduced*** *van Kampen diagram.*

3.1.3 Distortion van Kampen diagrams

Let $G = \langle X|R \rangle$ be a finitely presented group. For any word u of $X^{\pm}$, we denote $|u|$ its word length and $\|u\|_G$ the distance between the endpoints of its image in the Cayley graph $\mathrm{Cay}(G, X)$. Let H be a finitely generated subgroup of G, let Γ be a reduced labeled graph generating H. Its universal covering $\widetilde{\Gamma}$ is an infinite, connected and reduced labeled tree, with a natural label-preserving map $\varphi : \widetilde{\Gamma} \to \mathrm{Cay}(G, X)$.

Recall that the map between metric spaces φ is called λ-quasi isometric with some $\lambda \geq 1$ if, for any points $s, t \in \widetilde{\Gamma}$, we have $\lambda^{-1} d_{\widetilde{\Gamma}}(s, t) \leq d_G(\varphi(s), \varphi(t)) \leq \lambda d_{\widetilde{\Gamma}}(s, t)$. As the labeling on $\widetilde{\Gamma}$ is reduced, proving this inequality is equivalent to showing that for any reduced word u readable on Γ, we have $|u| \leq \lambda \|u\|_G$.

Lemma 3.4. *Let Γ be a finite connected labeled graph. If the map $\varphi : \widetilde{\Gamma} \to \mathrm{Cay}(G, X)$ is λ-quasi isometric with some $\lambda \geq 1$, then any subgroup generated by Γ is a free group.*

Proof. Let H be a subgroup of G generated by Γ. Suppose that H is *not* a free group. There exists a reduced non empty cycle p of Γ such that $p =_G 1$. But $\|p\|_G \geq \lambda^{-1}|p| > 0$, so p can not be trivial in G. □

By this lemma, to show that a subgroup generated by a labeled graph Γ is free, it is enough to show that $\widetilde{\Gamma} \to \mathrm{Cay}(G, X)$ is a quasi-isometric embedding. To study this question, we introduce *distortion van Kampen diagrams*.

Definition 3.5 (Distortion diagram)**.** *A distortion van Kampen diagram of* (G, Γ) *is a pair* (D, p) *where* D *is a van Kampen diagram of* G *and* p *is a cyclic sub-path of* ∂D *whose labeling word is readable on* Γ.

As Figure **3.2** show, a path p on ∂D is readable on a graph Γ.

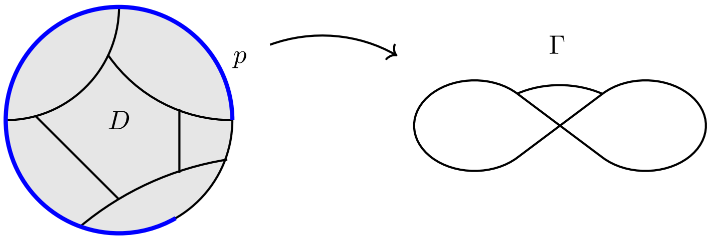


Figure 3.2: a path p on the boundary ∂D is readable on a graph Γ

Lemma 3.6. *Let* $\lambda \geq 1$. *If every* ***disk-like*** *and* ***reduced*** *distortion van Kampen diagram* (D, p) *of* (G, Γ) *satisfies*

$$|p| \leq \frac{\lambda}{1+\lambda}|\partial D|, \tag{$\star$}$$

then the map $\widetilde{\Gamma} \to \mathrm{Cay}(G, X)$ *is a* λ*-quasi isometric embedding.*

In particular, by Lemma 3.4, any subgroup generated by Γ *is free.*

Proof. Let u be a reduced word that is readable on Γ. Let v be one of the shortest word such that $uv =_G 1$ (so that $|v| = \|u\|_G$). By van Kampen's lemma (Lemma 3.3), there exists a reduced van Kampen diagram D whose boundary word is uv. If D is disk-like, then by the hypothesis $(\star)$ we have $|u| \leq \frac{\lambda}{1+\lambda}(|u| + |v|)$, which gives $|u| \leq \lambda|v|$.

Otherwise, we decompose D into disks and segments $D_1, \ldots, D_k$ (as in [CH82] p.159). The path of v does not intersect itself because it is a geodesic in G. The path of u on D does not intersect itself. If it did, as u is reduced, there would be a disk-like sub-diagram whose boundary word is readable on Γ, which is impossible because of $(\star)$.

Hence, for any $1 \leq i \leq k$, there are exactly two vertices on ∂D_i separating u and v, which are the only possible vertices of degree not equal to 2. The boundary word of D_i is written as $u_i v_i$ where u_i is a subword of u and v_i is a subword of v. If D_i is a segment, then it is read once by u and once by v with opposite directions, so $|u_i| = |v_i| \leq \lambda|v_i|$. If D_i is a disk, then $|u_i| \leq \lambda|v_i|$ by $(\star)$. We conclude that

$$|u| = \sum_{i=1}^{k} |u_i| \leq \sum_{i=1}^{k} \lambda|v_i| = \lambda|v|.$$

Now we shall prove that any subgroup generated by Γ is free. For any reduced word u readable on Γ, we have $|u| \leq \lambda\|u\|_G$. □

3.1.4 Hyperbolic groups

In this subsection, we recall several facts about hyperbolic groups introduced by M. Gromov in [Gro87]. Let $G = \langle X|R\rangle$ be a finite group presentation. A geodesic metric space is called δ-hyperbolic if each side of any geodesic triangle is δ-close to the two other sides ([CDP90] Chapter 1). A group G is called a *hyperbolic group* if there exists a finite generating set X such that the Cayley graph $\mathrm{Cay}(G, X)$ is δ-hyperbolic with some $\delta > 0$.

We start with a criterion of hyperbolicity in [Gro87] Chapter 2.3. We refer the reader to [Sho91] or [CDP90] Chapter 6 for detailed proofs. For a precise estimation of hyperbolicity constants, we refer to [Cha94] Lemma 3.11 by C. Champetier.

Theorem 3.7 (Isoperimetric inequality)**.** *Let ℓ be the longest relator length in R. The group $G = \langle X|R\rangle$ is hyperbolic if and only if there exists a real number $\beta > 0$ such that every reduced van Kampen diagram D satisfies the following isoperimetric inequality :*

$$|\partial D| \geq \beta\ell|D|.$$

In this case, the Cayley graph $\mathrm{Cay}(G, X)$ is δ-hyperbolic with

$$\delta = \frac{4\ell}{\beta}.$$

The local-global principle of hyperbolicity is due to M. Gromov in [Gro87]. For other proofs, see [Bow91] Chapter 8 by B. H. Bowditch or [Pap96] by P. Papasoglu. We state here a sharpened version by Y. Ollivier in [Oll07] Proposition 8.

Theorem 3.8. *(Local-global principal of hyperbolicity) For any $\alpha > 0$ and $\varepsilon > 0$, there exists an integer $K = K(\alpha, \varepsilon)$ such that, if every reduced disk-like diagram D with $|D| \leq K$ satisfies*

$$|\partial D| \geq \alpha\ell|D|,$$

then every reduced diagram D satisfies

$$|\partial D| \geq (\alpha - \varepsilon)\ell|D|.$$

Recall that a path p in a Cayley graph $\mathrm{Cay}(G, X)$ is a λ-quasi-geodesic if every sub-path u of p satisfies $|u| \leq \lambda\|u\|_G$. It is called a L-local λ-quasi geodesic if the inequality is satisfied by every sub-path of length at most L. Here is the local-global principle for quasi-geodesics in hyperbolic groups, stated by Gromov in [Gro87] 7.2.A and 7.2.B. See [CDP90] Chapter 3 for a proof.

Theorem 3.9. *Let $G = \langle X|R\rangle$ be a group presentation such that $\mathrm{Cay}(G, X)$ is δ-hyperbolic. Let $\lambda \geq 1$.*

1. *Every λ-quasi-geodesic is $100\delta(1 + \log \lambda)$ close to any geodesic joining its endpoints.*

2. *Every $1000\lambda\delta$-local λ-quasi-geodesic is a (global) 2λ-quasi-geodesic.*

By this theorem and the fact that random groups defined in the last chapter are a.a.s. hyperbolic, the main theorem in this chapter (Theorem 3.22) can be simplified to a local problem (Lemma 3.23).

3.2 Abstract diagrams

According to Proposition 2.44, the probability that a fixed van Kampen diagram (with respect to all possible cyclically reduced relators of length at most ℓ) appears in the random group $G_\ell(m, d)$ is determined by the number of relators used in this diagram. Two van Kampen diagrams having the same underlying 2-complex may not use the same number of relators, and should be treated separately.

For example, to check that if a group satisfies the $C'(\lambda)$ small cancellation condition (Theorem 2.46), we consider van Kampen diagrams whose underlying 2-complex consists of two faces f_1, f_2 sharing a common path of length $\lambda \min\{|\partial f_1|, |\partial f_2|\}$. We then need to consider the two types of diagrams in Figure **3.3**, one using two distinct relators and the other one using one relator.

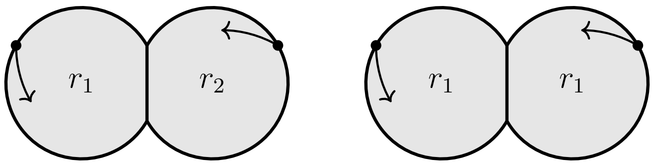


Figure 3.3: the two types of diagrams for checking if a group satisfies $C'(\lambda)$

For a 2-complex of size k, the number of diagram types is of order k^k and the problem of its existence can be very complicated. Y. Ollivier introduced *abstract van Kampen diagrams* in [Oll04] p.10 to study this problem.

3.2.1 Abstract van Kampen diagrams

Definition 3.10 (Abstract diagram, Ollivier [Oll04] p.10)**.** *An abstract van Kampen diagram $\widetilde{D}$ is a finite, planar and simply-connected 2-complex (V, E, F) with a labeling function on faces by integer numbers $\widetilde{\varphi}_2 : F \to \{1, 1^-, 2, 2^-, \ldots, k, k^-\}$ satisfying $\widetilde{\varphi}_2(f^{-1}) = \widetilde{\varphi}_2(f)^-$. We denote $\widetilde{D} = (V, E, F, \widetilde{\varphi}_2)$,*

By convention, $(i^-)^- = i$ for any $1 \le i \le k$. The numbers $\{1, \ldots, k\}$ are called *abstract relators* of $\widetilde{D}$.

Similar to a van Kampen diagram, a pair of faces $f, f' \in F$ is *reducible* if they have the same label, and they share an edge at the same position of their boundaries. An abstract diagram is called *reduced* if there is no reducible pair of faces.

Let $D = (V, E, F, \varphi_1, \varphi_2)$ be a van Kampen diagram of a group presentation $G = \langle X|R \rangle$. Let $\{r_1, \ldots, r_k\} \subset R$ be the set of relators used in D. Define $\widetilde{\varphi}_2 : F \to \{1, 1^-, \ldots, k, k^-\}$ by $\widetilde{\varphi}_2(f) = i$ if $\varphi_2(f) = r_i$. We obtain an abstract diagram $\widetilde{D} = (V, E, F, \widetilde{\varphi}_2)$ with k abstract relators, called an *underlying abstract diagram* of D.

An abstract diagram $\widetilde{D}$ is *fillable* by a group presentation $G = \langle X|R \rangle$ (or by a set of relators R) if there exists a van Kampen diagram D of G, called a *filled diagram* of $\widetilde{D}$, whose underlying abstract diagram is $\widetilde{D}$. That is to say, there exists k *different* relators $r_1, \ldots, r_k \in R$ such that the construction $\varphi_2(f) := r_{\widetilde{\varphi}_2(f)}$ gives a diagram $D = (V, E, F, \varphi_1, \varphi_2)$ of G. In Figure **3.4**, the abstract diagram has two abstract relators $1, 2$ and is filled by the relators r_1, r_2. The k-tuple $(r_1, \ldots, r_k)$ is called a *filling* of $\widetilde{D}$. As we picked different relators, $\widetilde{D}$ is reduced if and only if a filled diagram D is reduced.

We assume that faces with the same label of $\widetilde{D}$ have the same boundary length, otherwise $\widetilde{D}$ would never be fillable. Denote ℓ_i the length of the abstract relator i for $1 \le i \le k$. Let $\ell = \max\{\ell_1, \ldots, \ell_k\}$ be the maximal boundary length of faces of $\widetilde{D}$.

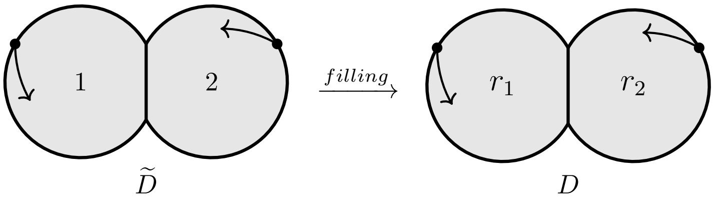


Figure 3.4: filling an abstract diagram

Notation. The pairs of integers $(i,1),\dots,(i,\ell_i)$ are called *abstract letters* of i.

The set of abstract letters of $\widetilde{D}$, denoted $\widetilde{X}$, is then a subset of $\{1,\dots,k\}\times\{1,\dots,\ell\}$, endowed with the lexicographic order.

We decorate undirected edges of $\widetilde{D}$ by abstract letters and directions. Let $f\in F$ labeled by i and let $e\in E$ at the j-th position of ∂f. The edge $\{e,e^{-1}\}$ is decorated, on the side of $\{f,f^{-1}\}$, by an arrow indicating the direction of e and the abstract letter (i,j). This decoration on $\{e,e^{-1}\}$ is called the *decoration from f at the position j*. The number of decorations on an edge $\{e,e^{-1}\}$ is the number of its adjacent faces $\{f,f^{-1}\}$ with multiplicity (0, 1 or 2 when $\widetilde{D}$ is planar).

For any filling $(r_1,\dots,r_k)$ of $\widetilde{D}$, we construct the *canonical function* $\phi:\widetilde{X}\to X^{\pm}$ such that $r_i=\phi(i,1)\dots\phi(i,\ell_i)$ for any $1\le i\le k$. If an edge $\{e,e^{-1}\}$ is decorated by two abstract letters (i,j), (i',j'), then $\phi(i',j')=\phi(i,j)$ if they have the same direction, or $\phi(i',j')=\phi(i,j)^{-1}$ if they have opposite directions. For example, in the diagram of Figure **3.5**, there is an edge decorated by two abstract letters $(1,4)$ and $(2,3)$ with opposite directions, so we have $\varphi(1,4)=\varphi(2,3)^{-1}$.

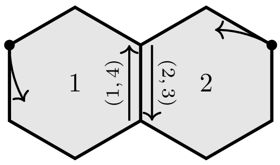


Figure 3.5: an edge decorated by two abstract letters

Note that if $\widetilde{D}$ is reduced, then by definition an abstract letter can not be decorated twice on an edge with the same direction (Figure **3.6** left-hand side). If $\widetilde{D}$ is fillable (by the set of all relators), then an abstract letter (i,j) can not be decorated twice on an undirected edge with opposite directions (Figure **3.6** right-hand side), otherwise we have $\phi(i,j)=\phi(i,j)^{-1}$ in the set of generators X.

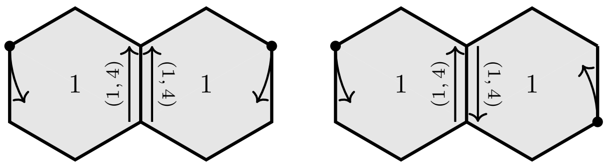


Figure 3.6: an edge decorated twice by the same abstract letter

In the following, we assume that $\widetilde{D}$ is fillable and reduced, so that the abstract letters decorated on an edge $\{e,e^{-1}\}$ are all different, and the two types of sub-diagrams in Figure **3.6** can not appear in $\widetilde{D}$. In particular, for any edge $\{e, e^{-1}\}$, there exists a unique face $\{f, f^{-1}\}$ (at a unique position) from which the decoration is (lexicographically) minimal. Whence the following two definitions.

Definition 3.11 (Preferred face of an edge)**.** *Let $\{e, e^{-1}\}$ be an edge of $\widetilde{D}$. Let $\{f, f^{-1}\}$ be the adjacent face of $\{e, e^{-1}\}$ from which the decoration is minimal. Then $\{f, f^{-1}\}$ is called the* ***preferred face*** *of $\{e, e^{-1}\}$.*

Definition 3.12 (free-to-fill)**.** *An abstract letter (i, j) of $\widetilde{D}$ is* ***free-to-fill*** *if, for any edge $\{e, e^{-1}\}$ decorated by (i, j), it is the minimal decoration on $\{e, e^{-1}\}$.*

Note that an abstract letter (i, j) is free-to-fill if and only if every face f labeled by i is the preferred face of its j-th boundary edge. In other words, if (i, j) is *not free-to-fill*, then there exists an edge $\{e, e^{-1}\}$ decorated by (i, j) that has another decoration $(i', j') < (i, j)$.

For example, in the abstract diagram of Figure **3.7**, $(1, 4)$, $(2, 1)$ and $(2, 2)$ are not free-to-fill. The other abstract letters are free-to-fill.

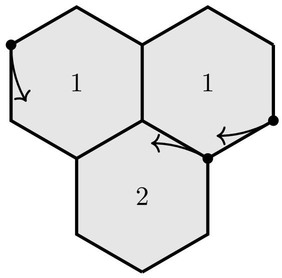


Figure 3.7: example of an abstract diagram

Denote $F^+ = \{f \in F \mid \widetilde{\varphi}_2(f) \in \{1, \ldots, k\}\}$. It gives a preferred orientation for each undirected face of $\widetilde{D} = (V, E, F, \widetilde{\varphi}_2)$. Let $\overline{E}$ be the set of undirected edges of $\widetilde{D}$.

Lemma 3.13. *Let $\widetilde{D}$ be a reduced fillable abstract diagram without isolated edges. For every face $f \in F^+$, let $\overline{E}_f$ be the set of edges $\{e, e^{-1}\}$ on the boundary of $\{f, f^{-1}\}$ such that $\{f, f^{-1}\}$ is the preferred face of $\{e, e^{-1}\}$. Then*

$$\overline{E} = \bigsqcup_{f \in F^+} \overline{E}_f.$$

Proof. For every edge $\{e, e^{-1}\}$ there exists a unique face $f \in F^+$ such that $\{e, e^{-1}\} \in \overline{E}_f$. Hence, the sets $\overline{E}_f$ with $f \in F^+$ are pairwise disjoint. Their reunion is the set of edges because every edge is adjacent to at least one face. □

3.2.2 Abstract distortion van Kampen diagrams

We generalize the idea of abstract diagrams to distortion van Kampen diagrams.

Definition 3.14 (Abstract distortion diagram)**.** *An abstract distortion van Kampen diagram is a pair $(\widetilde{D}, p)$ where $\widetilde{D}$ is an abstract diagram and p is a path on $\partial\widetilde{D}$.*

Let $G = \langle X|R \rangle$ be a group presentation and let Γ be a labeled graph. An abstract distortion diagram $(\widetilde{D}, p)$ is *fillable* by the pair (G, Γ) (or by the pair (R, Γ)) if there exists a filled diagram D of $\widetilde{D}$ such that (D, p) is a distortion diagram of (G, Γ). The distortion diagram (D, p) is called a *filled distortion diagram* of $(\widetilde{D}, p)$.

In the following, an abstract distortion diagram $(\widetilde{D}, p)$ is reduced, fillable, and without isolated edges. Recall that $\widetilde{X} \subset \{1, \dots, k\} \times \{1, \dots, \ell\}$ is the set of abstract letters. Let $\overline{p}$ be the set of undirected edges given by p. In an abstract distortion diagram, we distinguish between two types of free-to-fill abstract letters: those that decorate an edge of $\overline{p}$ and those that do not.

Definition 3.15. *Let (i, j) be an abstract letter of $(\widetilde{D}, p)$.*

*(i) (i, j) is **free-to-fill** if it is free-to-fill for the abstract diagram $\widetilde{D}$ and it does not decorate any edge of $\overline{p}$.*

*(ii) (i, j) is **semi-free-to-fill** if it is free-to-fill for the abstract diagram $\widetilde{D}$ and it decorates an edge of $\overline{p}$.*

(iii) Otherwise, (i, j) is not free-to-fill.

Notation. Let i be an abstract relator of $\widetilde{D}$. We denote α_i the number of faces labeled by i, η_i the number of free-to-fill abstract letters of i, and η'_i the number of semi-free-to-fill abstract letters of i.

Note that $\ell_i - \eta_i - \eta'_i$ is the number of non free-to-fill edges.

Lemma 3.16. *Recall that $\overline{E}_f$ is the set of edges on the boundary of f that prefers $\{f, f^{-1}\}$. Let i be an abstract relator. For any face $f \in F$ with $\widetilde{\varphi}_2(f) = i$, we have*

$$\eta'_i \leq |\overline{E}_f \cap \overline{p}| \quad \text{and} \quad \eta_i \leq |\overline{E}_f| - |\overline{E}_f \cap \overline{p}|.$$

Proof. Let $\{e, e^{-1}\}$ be the edge at the j-th position of ∂f. It is decorated by (i, j). If $\{f, f^{-1}\}$ is not preferred by $\{e, e^{-1}\}$, then (i, j) is not free-to-fill because there is a smaller decoration on $\{e, e^{-1}\}$.

Thus, if $\{e, e^{-1}\} \in \overline{E}_f \cap \overline{p}$ then (i, j) is semi-free-to-fill, which gives the first inequality. Similarly, if $\{e, e^{-1}\} \in \overline{E}_f \backslash \overline{p}$, then (i, j) is free-to-fill, so we have the second inequality. □

Lemma 3.17. *Recall that $\overline{E}$ is the set of undirected edges. The following two inequalities hold.*

$$\sum_{i=1}^{k} \alpha_i \eta'_i \leq |\overline{p}|, \quad \sum_{i=1}^{k} \alpha_i \eta_i \leq |\overline{E}| - |\overline{p}|.$$

Proof. By Lemma 3.16, for every $1 \leq i \leq k$

$$\alpha_i \eta'_i \leq \sum_{f \in F, \widetilde{\varphi}_2(f) = i} |\overline{E}_f \cap \overline{p}|.$$

Apply Lemma 3.13,

$$\sum_{i=1}^{k} \alpha_i \eta'_i \leq \sum_{f \in F^+} |\overline{E}_f \cap \overline{p}| \leq |\overline{p}|.$$

We get the second inequality by replacing η'_i by η_i and $|\overline{p}|$ by $|\overline{E} \backslash \overline{p}|$. □

3.2.3 The number of fillings of an abstract distortion diagram

Recall that B_ℓ is the set of cyclically reduced words on $X^{\pm} = \{x_1^{\pm}, \dots, x_m^{\pm}\}$ of length at most ℓ. Let Γ be a graph labeled by X with $\mathrm{rk}(\Gamma) \leq r$.

Let $(\widetilde{D}, p)$ be an abstract distortion diagram with k abstract relators. Assume that $\widetilde{D}$ is reduced, fillable and has no isolated edges. Let ℓ be the longest boundary length of faces of $\widetilde{D}$.

Denote $N_\ell(\widetilde{D}, p, \Gamma)$ the set of fillings $(r_1, \dots, r_k)$ of $(\widetilde{D}, p)$ by (B_ℓ, Γ). In this subsection, we give an upper bound of the *number of fillings* $|N_\ell(\widetilde{D}, p, \Gamma)|$ (Lemma 3.21).

Lemma 3.18. *The number of reduced words u of length L that is readable on Γ is at most* $2|\Gamma|(2r-1)^L$.

Proof. We estimate the number of paths p on Γ whose labeling word can be reduced. Take an oriented edge of Γ as the first edge of p, there are $2|\Gamma|$ choices. Every vertex is of degree at most $2r$ because $\mathrm{rk}(\Gamma) \leq r$. As p is reduced, every time we take the next edge, there are at most $(2r-1)$ choices. Hence, there are at most $2|\Gamma|(2r-1)^L$ paths. □

A vertex of $(\widetilde{D}, p)$ is called *distinguished* if it is either of degree at least 3, or the starting point of a face, or an endpoint of p. Let i be an abstract letter of $(\widetilde{D}, p)$. It can be regarded as a 2-complex (see Figure **3.8**) with two inverse faces $\{i, i^-\}$ and $2\ell_i$ edges $(i,1), \dots, (i,\ell_i)$ with their inverses, such that $\partial i = (i,1)\dots(i,\ell_i)$.

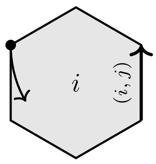


Figure 3.8: the 2-complex of the abstract letter i

A vertex of ∂i is *marked* if there exists a face f of $\widetilde{D}$ labeled by i such that the corresponding vertex is distinguished. Note that the starting point of ∂i is marked. Marked vertices divide the loop ∂i into segments, called *elementary segments*.

Consequently, an elementary segment is a sequence of abstract letters $(i,j)(i,j+1)\dots(i,j+t)$ such that, if a path $e_j \dots e_{j+t}$ on $\widetilde{D}$ is decorated by $(i,j)\dots(i,j+t)$, then it passes by no distinguished points except for its endpoints.

Lemma 3.19. *Let $(i,j)\dots(i,j+t)$ be an elementary segment. The abstract letters $(i,j), \dots, (i,j+t)$ are either all free-to-fill, or all semi-free-to-fill, or all not free-to-fill.*

Proof. We shall check that if the vertex between two consecutive abstract letters (i,j) and $(i,j+1)$ is not marked, then they are of the same type.

Recall that if an edge $\{e_1, e_1^{-1}\}$ is decorated by (i,j) from the face $\{f, f^{-1}\}$, then there is an edge $\{e_2, e_2^{-1}\}$ next to $\{e_1, e_1^{-1}\}$, decorated by $(i,j+1)$ from the same face $\{f, f^{-1}\}$. Assume that the vertex between (i,j) and $(i,j+1)$ is not marked so that the vertex between $\{e_1, e_1^{-1}\}$ and $\{e_2, e_2^{-1}\}$ is not distinguished.

We suppose by contradiction that (i,j) and $(i,j+1)$ are not of the same type. There are $3^2 - 3 = 6$ cases, grouped into three cases.

case 1. (i, j) is semi-free-to-fill and $(i, j+1)$ is free-to-fill, or inversely:

Recall that if (i, j) is semi-free-to-fill in the abstract distortion diagram $(\widetilde{D}, p)$, then it decorates an undirected edge $\{e_1, e_1^{-1}\}$ on p. As $(i, j+1)$ is free-to-fill, the edge $\{e_2, e_2^{-1}\}$ decorated by $(i, j+1)$ from the same face is not on p (see Figure **3.9**). So the vertex between $\{e_1, e_1^{-1}\}$ and $\{e_2, e_2^{-1}\}$ is distinguished, contradiction.

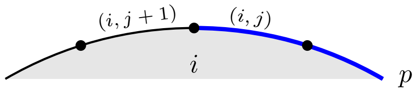


Figure 3.9: case 1 of Lemma 3.19

case 2. (i, j) is not free-to-fill, and $(i, j+1)$ is free-to-fill or semi-free-to-fill:

By definition, there is an edge $\{e_1, e_1^{-1}\}$ decorated by (i, j) having a smaller decoration $(i', j') < (i, j)$ (see Figure **3.10**). Let $\{f, f^{-1}\}$, $\{f', f'^{-1}\}$ be the faces attached by $\{e_1, e_1^{-1}\}$ such that f is labeled by i and f' is labeled by i'.

Let $\{e_2, e_2^{-1}\}$ be the edge next to $\{e_1, e_1^{-1}\}$, decorated by $(i, j+1)$ from the face $\{f, f^{-1}\}$. As the vertex between e_1 and e_2 is not distinguished, $\{e_2, e_2^{-1}\}$ is attached to the face $\{f', f'^{-1}\}$. It is then decorated by $(i', j'+1)$ or $(i', j'-1)$ from $\{f', f'^{-1}\}$. Because $(i, j+1)$ is free-to-fill, we have $(i, j+1) < (i', j'+1)$ or $(i, j+1) < (i', j'-1)$. Both are impossible because $(i, j) > (i', j')$.

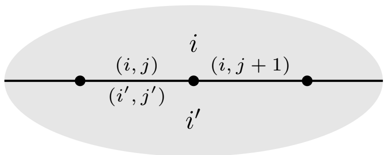


Figure 3.10: case 2 of Lemma 3.19

case 3. (i, j) is free-to-fill or semi-free-to-fill, and $(i, j+1)$ is not free-to-fill:

There is an edge $\{e_1, e_1^{-1}\}$ decorated by $(i, j+1)$ having a smaller decoration $(i', j') < (i, j+1)$ (Figure **3.11**). By the same argument of case 2, $(i, j) < (i', j'+1)$ or $(i, j) < (i', j'-1)$. The second one is obviously impossible. If the first one held, then $(i', j') < (i, j+1) < (i', j'+2)$, so $(i', j') = (i, j)$, and there was an edge decorated by (i, j) and $(i, j+1)$ with opposite directions. The canonical function $\phi : \widetilde{X} \to X$ gives $\phi(i, j+1) = \phi(i, j)^{-1}$, which is impossible because $r_i = \phi(i, 1) \ldots \phi(i, \ell_i)$ should be a reduced word.

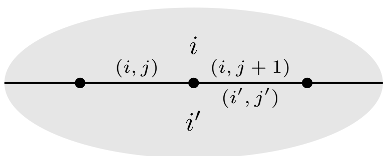


Figure 3.11: case 3 of Lemma 3.19

□

Lemma 3.20. *Let $(\widetilde{D}, p)$ be an abstract distortion diagram with no isolated edges.*

(i) *The number of distinguished vertices of $(\widetilde{D}, p)$ is at most $3|\widetilde{D}|$.*

(ii) *The number of elementary segments of an abstract letter i is at most $3|\widetilde{D}|^2$.*

Proof. The underlying 1-complex of $\widetilde{D}$ is a graph of rank $|\widetilde{D}|$ without isolated edges. By Lemma 3.1 there are at most $2(|\widetilde{D}|-1)$ vertices of degree ≥ 3. We add $k \leq |\widetilde{D}|$ starting points and 2 endpoints of p, there are at most $3|\widetilde{D}|$ distinguished vertices on $(\widetilde{D}, p)$.

The number of faces of $\widetilde{D}$ labeled by i is at most $|\widetilde{D}|$. Every face brings at most $3|\widetilde{D}|$ marked vertices to ∂i, so there are at most $3|\widetilde{D}|^2$ marked vertices on ∂i. □

Lemma 3.21. *Let $(\widetilde{D}, p)$ be a reduced abstract distortion diagram with no isolated edges and with k abstract relators. Recall that η_i is the number of free-to-fill abstract letters of i and η'_i is the number of semi-free-to-fill abstract letters of i. Let Γ be a labeled graph with $\mathrm{rk}(\Gamma) \leq r$. Recall that $N_\ell(\widetilde{D}, p, \Gamma)$ is the set of fillings $(r_1, \ldots, r_k)$ of $(\widetilde{D}, p)$ by (B_ℓ, Γ). We have*

$$|N_\ell(\widetilde{D}, p, \Gamma)| \leq \left(\frac{2m}{2m-1}\right)^k (2|\Gamma|)^{3|\widetilde{D}|^2 k} (2m-1)^{\sum_{i=1}^k \eta_i} (2r-1)^{\sum_{i=1}^k \eta'_i}.$$

Proof. We fill abstract letters of $\widetilde{D}$ in lexicographic order. We shall prove that if the abstract relators $1, \ldots, i-1$ are filled, then there are at most

$$\left(\frac{2m}{2m-1}\right) (2|\Gamma|)^{3|\widetilde{D}|^2} (2m-1)^{\eta_i} (2r-1)^{\eta'_i}$$

ways to fill the i-th abstract relator.

By Lemma 3.19, we fill elementary segments of i in order. Let u be an elementary segment of i. If u is free-to-fill, then there are at most $(2m-1)^{|u|}$ ways to fill u (or at most $2m(2m-1)^{|u|-1}$ ways if u is the first segment of i). If u is semi-free-to-fill, then there are at most $2|\Gamma|(2r-1)^{|u|}$ ways to fill u by lemma 3.18. If u is not free-to-fill, there is only one choice.

The sum of the lengths of free-to-fill segments is η_i, and the sum of the lengths of semi-free-to-fill segments is η'_i. As the number of semi-free-to-fill segments is at most $3|\widetilde{D}|^2$ (Lemma 3.20), there are at most $2m(2m-1)^{\eta_i - 1}(2|\Gamma|)^{3|\widetilde{D}|^2}(2r-1)^{\eta'_i}$ ways to fill the abstract relator i. □

3.3 The Freiheitssatz for random groups

Recall that B_ℓ is the set of cyclically reduced words of $X^\pm = \{x_1^\pm, \ldots, x_m^\pm\}$, and that $|B_\ell| = (2m-1)^{\ell+o(\ell)}$. The set of cyclically reduced words on $X_r^\pm = \{x_1^\pm, \ldots, x_r^\pm\}$ of length at most ℓ is of cardinality $(2r-1)^{\ell+o(\ell)}$. Its density in B_ℓ is

$$c_r = \log_{2m-1}(2r-1).$$

In this section, we prove that there is a phase transition at density

$$d_r = \min\left\{\frac{1}{2}, 1 - c_r\right\} = \min\left\{\frac{1}{2}, 1 - \log_{2m-1}(2r-1)\right\}.$$

3.3.1 Statement of the theorem

Theorem 3.22 (Phase transition at density d_r : The Freiheitssatz for random groups). *Let $\boldsymbol{G}(m,d) = (G_\ell(m,d))$ be a sequence of random groups at density d.*

(i) If $d > d_r$, then a.a.s., $G_\ell(m,d)$ is generated by $x_1,\dots,x_r$ (or by any subset of X of cardinality r).

(ii) If $d < d_r$, then a.a.s., for every reduced labeled graph Γ with $\mathrm{rk}(\Gamma) \le r$ and $|\Gamma| \le \frac{d_r-d}{5}\ell$, the canonical map $\widetilde{\Gamma} \to \mathrm{Cay}(G_\ell, X)$ is a $\frac{10}{d_r-d}$-quasi-isometric embedding.

In particular, a.s.s. every subgroup of $G_\ell(m,d)$ generated by a reduced labeled graph Γ with $\mathrm{rk}(\Gamma) \le r$ and $|\Gamma| \le \frac{d_r-d}{5}\ell$ is a free group of rank r.

In particular, a.a.s. $x_1,\dots,x_r$ freely generate a free subgroup of $G_\ell(m,d)$.

Proof of theorem 3.22 (i). Assume that $d < 1/2$. Otherwise, a.a.s. $G_\ell(m,d)$ is trivial by Theorem 2.45 (i). Recall that $X = \{x_1,\dots,x_m\}$ and $X_r = \{x_1,\dots,x_r\}$.

Let A_ℓ be the set of words of type $x_{r+1}w$ where w is a cyclically reduced word of $X_r^\pm$ of length $\ell-1$. The density of (A_ℓ) in (B_ℓ) is c_r. By hypothesis $c_r + d > 1$. Apply the intersection formula (Theorem 2.25), a.a.s. the intersection $R_\ell \cap A_\ell$ is not empty. Hence, a.a.s. there exists a cyclically reduced word w_{r+1} of $X_r^\pm$ such that $x_{r+1}w_{r+1} \in R_\ell$, which implies $x_{r+1} =_{G_\ell} w_{r+1}$.

Apply the same argument to the other generators $x_{r+2},\dots,x_m$. A.a.s. there are cyclically words $w_{r+1},\dots,w_m$ of $X_r^\pm$ such that $x_i =_{G_\ell} w_i$ for any $r+1 \le i \le m$. Hence, a.a.s. every word of $X^\pm$ equals to a word of $X_r^\pm$ in G_ℓ. □

The proof of the second assertion is similar to Ollivier's proof of Theorem 2.45 (ii) in [Oll04]. We work first on a local result.

Lemma 3.23 (Local diagrams). *Let $K > 0$. Let $(G_\ell) = (G_\ell(m,d))$ be a sequence of random groups with $d < d_r$. Then a.a.s. for every reduced labeled graph Γ with $\mathrm{rk}(\Gamma) \le r$ and $|\Gamma| \le \frac{d_r-d}{5}\ell$, every* ***disc-like reduced*** *distortion diagram (D,p) of (G_ℓ,Γ) with $|D| \le K$ satisfies*

$$|p| \le \left(1 - \frac{d_r - d}{5}\right)|\partial D|. \qquad (\star)$$

The proof of this lemma is in the next subsection.

Proof of Theorem 3.22 (ii) by Lemma 3.23.

By Lemma 2.5 and Theorem 2.45 (ii), we work under the condition that every diagram D of $G_\ell(m,d)$ satisfies $|\partial D| \ge (1-2d)/2|D|\ell$ and that $G_\ell(m,d)$ is δ-hyperbolic with $\delta = \frac{4\ell}{1-2d}$.

Let Γ be a reduced labeled graph with $|\Gamma| \le \frac{d_r-d}{5}\ell$ and $\mathrm{rk}(\Gamma) \le r$. Let $\lambda = \frac{5}{d_r-d}$. By the local-global principle of quasi-geodesics (Theorem 3.9), in order to prove that $\widetilde{\Gamma} \to \mathrm{Cay}(X, G_\ell)$ is a (global) 2λ-quasi-isometric embedding, we prove that every reduced word u readable on Γ is a $\frac{4000\lambda\ell}{1-2d}$-local λ-quasi-geodesic.

Let u be a reduced word that is readable on Γ with $|u| \le \frac{4000\lambda\ell}{1-2d}$. Let v be a geodesic in G_ℓ joining endpoints of the image of u in G_ℓ. We shall prove that $|u| \le \lambda|v|$. By van Kampen's lemma (Lemma 3.3) there exists a diagram D of G_ℓ whose boundary word is uv. By the isoperimetric inequality (Theorem 2.45 (ii)),

$$|D| \le \frac{2|\partial D|}{(1-2d)\ell} \le \frac{4|u|}{(1-2d)\ell} \le \frac{50000\lambda}{(1-2d)^2}.$$

Apply Lemma 3.23 with $K = \frac{50000\lambda}{(1-2d)^2} = \frac{250000}{(1-2d)^2(d_r-d)}$. If D is disk-like, then by $(\star)$, we have

$$|u| \leq \left(1 - \frac{d_r - d}{5}\right)(|u| + |v|) \leq \frac{\lambda}{1+\lambda}(|u| + |v|),$$

which implies $|u| \leq \lambda |v|$.

Otherwise, we decompose D into discs and segments. By the same argument of Lemma 3.6, because every disc-like sub-diagram is a distortion diagram satisfying $(\star)$, we still have $|u| \leq \lambda |v|$. ☐

3.3.2 Proof of Lemma 3.23

Let $(G_\ell) = (G_\ell(m,d)) = (\langle X | R_\ell \rangle)$ be a sequence of random groups with density $d < d_r$. To prove Proposition 3.23, we work first on the fillability of an abstract distortion diagram. Denote

$$\varepsilon_d = \frac{d_r - d}{5}.$$

Let Q_ℓ be the probability event $(2m-1)^{(d-\varepsilon_d)\ell} \leq |R_\ell| \leq (2m-1)^{(d+\varepsilon_d)\ell}$. We have a.a.s. Q_ℓ by the characterization of densability (Proposition 2.6).

Lemma 3.24 (Fillability of an abstract distortion diagram)**.** *Let $K > 0$. Let Γ be a reduced labeled graph with $\mathrm{rk}(\Gamma) \leq r$ and $|\Gamma| \leq \varepsilon_d \ell$. Let $(\widetilde{D}, p)$ be a disc-like abstract distortion diagram with $|\widetilde{D}| \leq K$ that satisfies*

$$|p| > (1 - \varepsilon_d)\,|\partial \widetilde{D}|.$$

Then for ℓ large enough,

$$\mathbf{P} = \mathbf{Pr}\left((\widetilde{D}, p) \text{ is fillable by } (G_\ell, \Gamma) \,\middle|\, Q_\ell\right) \leq \ell^{10K^3}(2m-1)^{-2\varepsilon_d \ell}.$$

Proof. We shall prove the lemma in four steps. We omit "for ℓ large enough" in every step. Recall that α_i is the number of faces labeled by i, η_i the number of free-to-fill abstract letters of i, and η_i' the number of semi-free-to-fill abstract letters of i.

Step 1: $\log_{2m-1} \mathbf{P} \leq \sum_{i=1}^{k} (\eta_i + c_r \eta_i' + (d - 1 + 2\varepsilon_d)\ell) + 10K^3 \log_{2m-1} \ell. \qquad (1)$

According to Proposition 2.44, if $(r_1, \dots, r_k)$ is a filling of $\widetilde{D}$ by B_ℓ, then for ℓ large enough,

$$\mathbf{Pr}\,(r_1, \dots, r_k \in R_\ell \,|\, Q_\ell) \leq (2m-1)^{k(d-1+2\varepsilon_d)\ell}.$$

Recall that $N_\ell(\widetilde{D}, p, \Gamma)$ is the set of fillings of $(\widetilde{D}, p)$ by (B_ℓ, Γ). Apply Lemma 3.21 with $|\Gamma| \leq \varepsilon_d \ell$ and $|\widetilde{D}| \leq K$,

$$\begin{aligned}
&\mathbf{Pr}\left((\widetilde{D}, p) \text{ is fillable by } (G_\ell, \Gamma) \,\middle|\, Q_\ell\right) \\
\leq &\sum_{(r_1,\dots,r_k) \in N_\ell(\widetilde{D}, p, \Gamma)} \mathbf{Pr}\,(r_1, \dots, r_k \in R_\ell \,|\, Q_\ell) \\
\leq &|N_\ell(\widetilde{D}, p, \Gamma)|(2m-1)^{k(d-1+2\varepsilon_d)\ell} \\
\leq &\ell^{10K^3}(2m-1)^{\sum_{i=1}^k \eta_i}(2r-1)^{\sum_{i=1}^k \eta_i'}(2m-1)^{k(d-1+2\varepsilon_d)\ell}.
\end{aligned}$$

Hence the inequality (1) by applying $\log_{2m-1}$.

Step 2: $|\widetilde{D}|\left(\log_{2m-1}\mathbf{P} - 10K^3\log_{2m-1}\ell\right) \leq \sum_{i=1}^{k}\alpha_i(\eta_i + c_r\eta_i' + (d-1+2\varepsilon_d)\ell)$. (2)

Let $\widetilde{D}_i$ be the sub-diagram of $\widetilde{D}$ consisting of the faces labeled by the first i abstract relators $1^{\pm},\ldots,i^{\pm}$ and the edges attached to them. Apply (1) to $\widetilde{D}_i$, and denote $\mathbf{P}_i$ the probability obtained. We have

$$\log_{2m-1}\mathbf{P} \leq \log_{2m-1}\mathbf{P}_i \leq \sum_{s=1}^{i}(\eta_s + c_r\eta_s' + (d-1+2\varepsilon_d)\ell) + 10K^3\log_{2m-1}\ell.$$

Without loss of generality, we assume $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_k$. Note that $\log_{2m-1}\mathbf{P}$ is negative and that $\alpha_1 \leq |\widetilde{D}|$. By Abel's summation formula, with convention $\alpha_{k+1} = 0$,

$$\begin{aligned}
&\sum_{i=1}^{k}\alpha_i\left(c_r\eta_i' + \eta_i + (d-1+2\varepsilon_d)\ell\right)\\
=&\sum_{i=1}^{k}\left[(\alpha_i - \alpha_{i+1})\sum_{s=1}^{i}(c_r\eta_s' + \eta_s + (d-1+2\varepsilon_d)\ell)\right]\\
\geq&\sum_{i=1}^{k}\left[(\alpha_i - \alpha_{i+1})(\log_{2m-1}\mathbf{P} - 10K^3\log_{2m-1}\ell)\right]\\
\geq&\alpha_1(\log_{2m-1}\mathbf{P} - 10K^3\log_{2m-1}\ell)\\
\geq&|\widetilde{D}|(\log_{2m-1}\mathbf{P} - 10K^3\log_{2m-1}\ell).
\end{aligned}$$

Step 3: $\log_{2m-1}\mathbf{P} \leq \left(d - \frac{1}{2} + 2\varepsilon_d\right)\ell + \left(c_r - \frac{1}{2} + \varepsilon_d\right)\frac{|\partial\widetilde{D}|}{|\widetilde{D}|} + 10K^3\log_{2m-1}\ell$. (3)

Let $\varepsilon_d' > 0$ such that $|\overline{p}| = (1-\varepsilon_d')|\partial\widetilde{D}|$. By hypothesis $\varepsilon_d' < \varepsilon_d$. Because $\widetilde{D}$ is disc-like and the boundary length of every face is $\leq \ell$, the number of undirected edges $|\overline{E}|$ is less than $\frac{|\widetilde{D}|\ell - |\partial\widetilde{D}|}{2} + |\partial\widetilde{D}|$. Apply Lemma 3.17, we get

$$\sum_{i=1}^{k}\alpha_i\eta_i' \leq |\overline{p}| = (1-\varepsilon_d')|\partial\widetilde{D}|,$$

$$\sum_{i=1}^{k}\alpha_i\eta_i \leq |\overline{E}| - |\overline{p}| \leq \frac{|\widetilde{D}|\ell}{2} + \left(\varepsilon_d' - \frac{1}{2}\right)|\partial\widetilde{D}|.$$

Note that $\sum_{i=1}^{k}\alpha_i = |\widetilde{D}|$. So we have

$$\begin{aligned}
&\sum_{i=1}^{k}\alpha_i\left(c_r\eta_i' + \eta_i + (d-1+2\varepsilon_d)\ell\right)\\
\leq&c_r(1-\varepsilon_d')|\partial\widetilde{D}| + \frac{|\widetilde{D}|\ell}{2} + \left(\varepsilon_d' - \frac{1}{2}\right)|\partial\widetilde{D}| + (d-1+2\varepsilon_d)|\widetilde{D}|\ell\\
\leq&\left(d - \frac{1}{2} + 2\varepsilon_d\right)|\widetilde{D}|\ell + \left(c_r - \frac{1}{2} + \varepsilon_d\right)|\partial\widetilde{D}|.
\end{aligned}$$

Combine this inequality with (2), we get (3)

Step 4: $\left(d - \frac{1}{2} + 2\varepsilon_d\right)\ell + \left(c_r - \frac{1}{2} + \varepsilon_d\right)\frac{|\partial\widetilde{D}|}{|\widetilde{D}|} \leq -2\varepsilon_d\ell.$ (4)

Recall that $c_r = \log_{2m-1}(2r-1)$. Note that $|\partial\widetilde{D}| \leq \ell|\widetilde{D}|$ and that $d = d_r - 5\varepsilon_d$. There are two cases:

(a) If $c_r \geq \frac{1}{2}$, then $d = 1 - c_r - 5\varepsilon_d$ and $c_r - \frac{1}{2} + \varepsilon_d \geq 0$, so

$$\begin{aligned}&\left(d - \frac{1}{2} + 2\varepsilon_d\right)|\widetilde{D}|\ell + \left(c_r - \frac{1}{2} + \varepsilon_d\right)|\partial\widetilde{D}|\\ \leq\,&(d + c_r - 1 + 3\varepsilon_d)\,|\widetilde{D}|\ell\\ \leq\,&-2\varepsilon_d|\widetilde{D}|\ell\end{aligned}$$

(b) If $c_r < \frac{1}{2}$, then $d = \frac{1}{2} - 5\varepsilon_d$ and $c_r - \frac{1}{2} + \varepsilon_d \leq \varepsilon_d$, so

$$\begin{aligned}&\left(d - \frac{1}{2} + 2\varepsilon_d\right)|\widetilde{D}|\ell + \left(c_r - \frac{1}{2} + \varepsilon_d\right)|\partial\widetilde{D}|\\ \leq\,&\left(d - \frac{1}{2} + 3\varepsilon_d\right)|\widetilde{D}|\ell\\ \leq\,&-2\varepsilon_d|\widetilde{D}|\ell\end{aligned}$$

By (3) and (4), for ℓ large enough $\log_{2m-1}(\mathbf{P}) \leq -2\varepsilon_d\ell + 10K^3\log_{2m-1}\ell$. □

By Lemma 3.1 and Lemma 3.2, we have the following two results.

Lemma 3.25. *For ℓ large enough, the number of reduced labeled connected graphs Γ (with respect to X) with $|\Gamma| \leq \varepsilon_d\ell$ and $\mathrm{rk}(\Gamma) \leq r$ is bounded by*

$$(2m-1)^{\varepsilon_d\ell}\ell^{3r}.$$

□

Lemma 3.26. *For ℓ large enough, the number of disc-like abstract distortion diagrams $(\widetilde{D}, p)$ with $|\widetilde{D}| \leq K$ and $|\partial f| \leq \ell$ for all faces $f \in F$ is bounded by*

$$\ell^{5K}.$$

□

Proof of Lemma 3.23. Recall that $\varepsilon_d = \frac{d_r - d}{5}$.

We shall prove that a.a.s. for every reduced labeled graph Γ with $\mathrm{rk}(\Gamma) \leq r$ and $|\Gamma| \leq \varepsilon_d\ell$, every reduced distortion diagram (D, p) of (G_ℓ, Γ) with $|D| \leq K$ satisfies $|p| \leq (1 - \varepsilon_d)\,|\partial D|$.

Apply Lemma 3.24, Lemma 3.25 and Lemma 3.26. The probability that there exists a reduced labeled graph Γ with $\mathrm{rk}(\Gamma) \leq r$, $|\Gamma| \leq \varepsilon_d\ell$ and there exists a disc-like reduced abstract distortion diagram $(\widetilde{D}, p)$ with $|\widetilde{D}| \leq K$, $|p| > (1 - \varepsilon_d)\,|\partial\widetilde{D}|$ such that $(\widetilde{D}, p)$ is fillable by (G_ℓ, Γ) is bounded by

$$(2m-1)^{\varepsilon_d\ell}\ell^{3r} \times \ell^{5K} \times \ell^{10K^3}(2m-1)^{-2\varepsilon_d\ell} = (2m-1)^{-\varepsilon_d\ell + O(\log\ell)}.$$

So the probability that there exists a reduced labeled graph Γ with $\mathrm{rk}(\Gamma) \leq r$, $|\Gamma| \leq \varepsilon_d \ell$ and there exists a disc-like reduced distortion diagram (D, p) of (G_ℓ, Γ) with $|D| \leq K$ that satisfies $|p| > (1 - \varepsilon_d)\, |\partial \widetilde{D}|$ is bounded by

$$(2m-1)^{-\varepsilon_d \ell + O(\log \ell)},$$

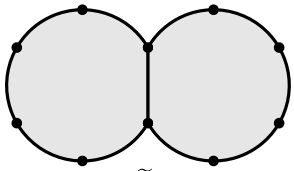
which goes to 0 when ℓ goes to infinity. □

This completes the proof of Theorem 3.22.

Chapter 4

Existence of van Kampen 2-complexes

Let $(G_\ell(m,d)) = (\langle X_m | R_\ell \rangle)$ be a sequence of random groups in the permutation invariant density model. In Subsection 2.4.2 , it is shown that there is a phase transition at density $d = \lambda/2$ for the $C'(\lambda)$ small cancellation condition. In other words, we ask if there exists van Kampen diagrams D_ℓ in $G_\ell(m,d)$ with the following "geometric form" (Definition 4.13): there are exactly 2 faces of boundary length ℓ, sharing a common path of length $\lfloor \lambda\ell \rfloor$.



$\widetilde{D}_\ell$

We consider van Kampen 2-complexes as van Kampen diagrams that are not necessarily planar. The goal of this chapter is to establish a theorem (Theorem H, Theorem 4.15) that generalize this result to any geometric form of 2-complexes. More precisely, for a given sequence of 2-complexes $\boldsymbol{Y} = (Y_\ell)$ of the same geometric form (Definition 4.13), we show that there is a critical density $d_c(Y)$ such that, if $d + d_c(Y) > 1$, then a.a.s. Y_ℓ is the underlying 2-complex of a van Kampen 2-complex of $G_\ell(m,d)$; if $d + d_c(Y) < 1$, then there is no van Kampen 2-complex of $G_\ell(m,d)$ having Y_ℓ as the underlying 2-complex. The proof is given in Section 4.2.

In [GM18] §2, D. Gruber and J. Mackay proved a 2-complex version of the isoperimetric inequality (Theorem 2.45 (ii)) for random triangular groups. In the first section of this chapter, we adapt this result to the Gromov density model, showing that a.a.s. every van Kampen 2-complex of a random group $G_\ell(m,d)$ satisfies some inequality (Theorem G, Theorem 4.3). The last section is for applications to small cancellation theory.

4.1 Isoperimetric inequality for van Kampen 2-complexes

A van Kampen 2-complex with respect to a group presentation $G = \langle X | R \rangle$ is a 2-complex $Y = (V, E, F)$ with two compatible labeling functions $\varphi_1 : E \to X^\pm$ and $\varphi_2 : F \to R^\pm$. We denote briefly $Y = (V, E, F, \varphi_1, \varphi_2)$. Notions for van Kampen diagrams in Subsection 3.1.2 can be applied for van Kampen 2-complexes.

Recall that faces $f \in F$ are oriented and $F^+ := \{f \in F | \varphi_2(f) \in R\}$ is a choice of orientation for each undirected face $\{f, f^{-1}\}$. We denote $|Y|$ the number of undirected faces and $\mathrm{Edge}(Y)$ the number of undirected edges. A pair of faces f, f' is called reducible if they have the same label and there is a common edge on their boundaries at the same position. A van Kampen 2-complex is called reduced if there is no reducible pair of faces.

In this section, we show that a.a.s. every van Kampen 2-complex Y of the random group $G_\ell(m, d)$ with a bounded *complexity* (Definition 4.2) satisfies an inequality presented by its number of edges $\mathrm{Edge}(Y)$, its number of faces $|Y|$ and its *reduction degree* $\mathrm{Red}(Y)$ (Definition 4.1).

4.1.1 Definitions and statement of the theorem

As in [GM18], we establish the inequality for non-reduced van Kampen 2-complexes. We need to count the number of edges causing reducible pair of faces, with *multiplicity*: for any edge $e \in E$, any relator $r \in R$ and any integer i, we count the number of faces $f \in F$ labeled by r and having e as the i-th boundary edge. If this number is $k \geq 2$, we add $k - 1$ to the reduction degree, otherwise we add 0. The *reduction degree* $\mathrm{Red}(Y)$ is hence defined as follows.

Definition 4.1 (Reduction degree, [GM18] Definition 2.5)**.** *Let $Y = (V, E, F, \varphi_1, \varphi_2)$ be a van Kampen 2-complex of a group presentation $G = \langle X|R\rangle$. Let ℓ be the maximal boundary length of faces of Y. The reduction degree of Y is*

$$\mathrm{Red}(Y) = \sum_{e\in E}\sum_{r\in R}\sum_{1\leq i\leq \ell}\Big(\big|\{f \in F \,|\, \varphi_2(f) = r, e \text{ is the } i\text{-th edge of } \partial f\}\big| - 1\Big)^+.$$

Remark. A van Kampen 2-complex Y is reduced if and only if, for any face f, any relator r and any integer i, the cardinality of the set $\{f \in F \,|\, \varphi_2(f) = r, e \text{ is the } i\text{-th edge of } \partial f\}$ is either 0 or 1, hence if and only if $\mathrm{Red}(Y) = 0$.

We define here the *complexity* of a 2-complex. It is not needed in [GM18] for triangular random groups, because the number of triangular 2-complexes Y of bounded size $|Y| \leq K$ is bounded by a constant depending only on K. We give later an upper bound of the number of 2-complexes with a given complexity K (Lemma 4.12). We will see that the two additional conditions (other than $|Y| \leq K$) are essential for the estimation.

Definition 4.2. *Let Y be a 2-complex. Let $K > 0$. We say that Y is of complexity K if the following three conditions hold:*

- $|Y| \leq K$.
- *The number of maximal arcs of Y is bounded by K.*
- *For any face f of Y, the boundary path ∂f is divided into at most K maximal arcs.*

Remark. If D is a disk-like 2-complex with $|D| \leq K$, then the complexity of D is $6K$. In fact, as the rank of its underlying graph is K, by Proposition 3.1, the number of maximal arcs is at most $3K$, and every boundary path is divided into at most $6K$ maximal arcs (an arc may be used twice). With this point of view, the local inequality of Theorem 2.45 (ii) is a corollary of Theorem 4.3.

We can now state Theorem G of Introduction.

Theorem 4.3 (Theorem G)**.** *Let $(G_\ell(m, d))$ be a sequence of random groups with density $d < 1/2$. Let $\varepsilon > 0$, $K > 0$. A.a.s. every van Kampen 2-complex Y of complexity K of $G_\ell(m, d)$ satisfies*

$$\mathrm{Edge}(Y) + \mathrm{Red}(Y) \geq (1 - d - \varepsilon)|Y|\ell.$$

As $\mathrm{Red}(Y) = 0$ if and only if Y is reduced. we have the following corollary.

Corollary 4.4. *Let $(G_\ell(m,d))$ be a sequence of random groups with density $d < 1/2$. Let $\varepsilon > 0$, $K > 0$. A.a.s. every* ***reduced*** *van Kampen 2-complex Y of complexity K of $G_\ell(m,d)$ satisfies*

$$\mathrm{Edge}(Y) \geq (1 - d - \varepsilon)|Y|\ell.$$

Recall that a 2-complex Y without isolated edges is called *contractible* if there is an edge of Y that is adjacent to one single face. If Y is a non contractible 2-complex, then each of its edge is adjacent to at least 2 faces, and we have $\mathrm{Edge}(Y) \leq \frac{1}{2}|Y|\ell$ where ℓ is the maximal face boundary length. As this contradicts the inequality in Corollary 4.4 for *any* density $d < 1/2$, there is a further corollary.

Corollary 4.5. *Let $(G_\ell(m,d))$ be a sequence of random groups with density $d < 1/2$. Let $K > 0$. A.a.s. every reduced van Kampen 2-complex of complexity K of $G_\ell(m,d)$ is contractible.*

Remark. Theorem 4.3 is not true if the complexity of 2-complexes is not bounded. For example, Calegari-Walker [CW15] proved that at any density $d < 1/2$, a.a.s. a random group $G_\ell(m,d)$ contains a surface subgroup of genus $O(\ell)$, so with complexity at least $O(\ell)$.

Remark. Denote $\mathrm{Cancel}(Y)$ the number of canceled undirected edges of a 2-complex Y by attaching maps. That is to say, we have

$$\mathrm{Edge}(Y) + \mathrm{Cancel}(Y) = \sum_{f \in F^+} |\partial f|.$$

If the length of every face boundary is at most ℓ, we have

$$\mathrm{Edge}(Y) + \mathrm{Cancel}(Y) = \sum_{f \in F^+} |\partial f| \leq |Y|\ell.$$

So the inequality of Theorem 4.3 implies

$$\mathrm{Cancel}(Y) - \mathrm{Red}(Y) \leq (d + \varepsilon)|Y|\ell.$$

It is an analog of Gruber-Mackay's result in [GM18] §2. for triangular random groups.

4.1.2 Abstract van Kampen 2-complexes

Abstract van Kampen 2-complexes, as abstract van Kampen diagrams and abstract distortion diagrams (Section 3.2), is a structure between 2-complexes and van Kampen 2-complexes that helps us solve random group problems.

Let $(G_\ell(m,d)) = (\langle X_m | R_\ell \rangle)$ be a sequence of random groups at density d. Let $0 < \varepsilon < 1 - d$. We shall work under the condition Q_ℓ:

$$(2m-1)^{(d - \frac{\varepsilon}{4}\ell)} \leq |R_\ell| \leq (2m-1)^{(d + \frac{\varepsilon}{4}\ell)}.$$

By the densability of R_ℓ, we have a.a.s. Q_ℓ. The following proposition is a variant of Proposition 2.44.

Proposition 4.6. *Let $r_1, \ldots, r_k$ be pairwise different relators in B_ℓ. We have*

$$\mathbf{Pr}\left(r_1, \ldots, r_k \in R_\ell \mid Q_\ell\right) \leq (2m-1)^{k(d-1+\frac{\varepsilon}{2})\ell}.$$

□

Definition 4.7 (The 2-complex version of Definition 3.10)**.** *An abstract van Kampen 2-complex $\widetilde{Y}$ is a 2-complex (V, E, F) with a labeling function on faces by integer numbers and their inverses $\widetilde{\varphi}_2 : F \to \{1, 1^-, 2, 2^-, \dots, k, k^-\}$ such that $\widetilde{\varphi}_2(f^{-1}) = \widetilde{\varphi}_2(f)^-$. We denote simply $\widetilde{Y} = (V, E, F, \widetilde{\varphi}_2)$.*

The notions of Section 3.2 for abstract van Kampen diagrams can be applied for abstract van Kampen 2-complexes. Let us recall some of them.

By convention $(i^-)^- = i$. The integers $\{1, \dots, k\}$ are called abstract relators. Denote ℓ_i the length of the abstract relator i for $1 \leq i \leq k$. Let $\ell = \max\{\ell_1, \dots, \ell_k\}$ be the maximal boundary length of faces. The pairs of integers $(i, 1), \dots, (i, \ell_i)$ are called *abstract letters* of i. The set of abstract letters of $\widetilde{Y}$ is then a subset of $\{1, \dots, k\} \times \{1, \dots, \ell\}$.

We decorate undirected edges of $\widetilde{Y}$ by abstract letters and directions. Let $f \in F$ labeled by i and let $e \in E$ at the j-th position of ∂f. The edge $\{e, e^{-1}\}$ is decorated, on the side of $\{f, f^{-1}\}$, by an arrow indicating the direction of e and the abstract letter (i, j). The number of decorations on an edge is the number of its adjacent faces with multiplicity.

Definition 4.8 (free-to-fill)**.** *An abstract letter (i, j) of $\widetilde{D}$ is **free-to-fill** if, for any edge $\{e, e^{-1}\}$ decorated by (i, j), it is the minimal decoration on $\{e, e^{-1}\}$.*

The following lemma is the 2-complex non-reduced version of Lemma 3.17.

Lemma 4.9. *Let $\widetilde{Y}$ be an abstract van Kampen 2-complex with k abstract relators. Denote α_i the number of faces labeled by the abstract relator i and η_i the number of free-to-label edges of i. Then*

$$\sum_{i=1}^{k} \alpha_i \eta_i \leq \mathrm{Edge}(\widetilde{Y}) + \mathrm{Red}(\widetilde{Y}).$$

Lemma 4.10. *Let $\widetilde{Y}$ be an abstract van Kampen 2-complex with k abstract relators.*

$$\mathbf{Pr}\left(\widetilde{Y} \text{ is fillable by } G_\ell(m, d) \,\middle|\, Q_\ell\right) \leq \left(\frac{2m}{2m-1}\right)^k (2m-1)^{\sum_{i=1}^{k}\left(\eta_i + \left(d - 1 + \frac{\varepsilon}{2}\right)\ell\right)}.$$

Proof. Let us estimate the number of fillings of $\widetilde{Y}$. For every free-to-fill abstract letter (i, j), there are at most $2m$ ways to fill a generator if $j = 1$, at most $(2m-1)$ ways to fill if $j \neq 1$ for avoiding reducible word. As there are η_i free-to-fill abstract letters on the i-th abstract relator, there are at most $2m(2m-1)^{\eta_i - 1}$ ways to fill it. So there are at most $\prod_{i=1}^{k}\left(2m(2m-1)^{\eta_i-1}\right)$ ways to fill $\widetilde{Y}$.

Let Y be a van Kampen 2-complex, which is a filling of $\widetilde{Y}$. The 2-complex Y is labeled by k different relators in B_ℓ, denoted $r_1, \dots, r_k$. By lemma 4.6,

$$\begin{aligned}\mathbf{Pr}\left(Y \text{ is a 2-complex of } G_\ell(m, d) \mid Q_\ell\right) &= \mathbf{Pr}\left(r_1, \dots, r_k \in R_\ell \mid Q_\ell\right) \\ &\leq (2m-1)^{k\left(d-1+\frac{\varepsilon}{2}\right)\ell}.\end{aligned}$$

Hence

$$\begin{aligned}\mathbf{Pr}\left(\widetilde{Y} \text{ is fillable by } G_\ell(m, d) \,\middle|\, Q_\ell\right) &\leq \sum_{Y \text{ is a filling of } \widetilde{Y}} \mathbf{Pr}\left(Y \text{ is a 2-complex of } G_\ell(m, d) \mid Q_\ell\right) \\ &\leq \prod_{i=1}^{k}\left(2m(2m-1)^{\eta_i-1}\right)(2m-1)^{k\left(d-1+\frac{\varepsilon}{2}\right)\ell} \\ &\leq \left(\frac{2m}{2m-1}\right)^k (2m-1)^{\sum_{i=1}^{k}\left(\eta_i + \left(d-1+\frac{\varepsilon}{2}\right)\ell\right)}.\end{aligned}$$

□

4.1.3 Proof of Theorem G (Theorem 4.3)

The following lemma is the key to prove Theorem 4.3. Recall that Q_ℓ is the a.a.s. event

$$(2m-1)^{(d-\frac{\varepsilon}{4}\ell)} \le |R_\ell| \le (2m-1)^{(d+\frac{\varepsilon}{4}\ell)}.$$

Lemma 4.11. *Suppose that $\widetilde{Y}$ does not satisfy the inequality given in Theorem 4.3, i.e.*

$$\mathrm{Edge}(\widetilde{Y}) + \mathrm{Red}(\widetilde{Y}) < (1-d-\varepsilon)|\widetilde{Y}|\ell,$$

then

$$\mathbf{Pr}\left(\widetilde{Y} \text{ is fillable by } G_\ell(m,d) \;\middle|\; Q_\ell\right) \le \left(\frac{2m}{2m-1}\right)(2m-1)^{-\frac{\varepsilon}{2}\ell}.$$

Proof. Let $\widetilde{Y}_i$ be the sub-2-complex of $\widetilde{Y}$ consisting of faces labeled by the i first abstract relators. Denote $P_i = \mathbf{Pr}\left(\widetilde{Y}_i \text{ is fillable by } G_\ell(m,d) \;\middle|\; Q_\ell\right)$. Apply lemma 4.10 on $\widetilde{Y}_i$, we have

$$P_i \le \left(\frac{2m}{2m-1}\right)^i (2m-1)^{\sum_{j=1}^{i}\left(\eta_j + \left(d-1+\frac{\varepsilon}{2}\right)\ell\right)}.$$

Note that if $\widetilde{Y}$ is fillable by $G_\ell(m,d)$ then its sub 2-complex $\widetilde{Y}_i$ is fillable by the same group. So for any $1 \le i \le k$,

$$\log_{2m-1}(P_k) \le \log_{2m-1}(P_i) \le \sum_{j=1}^{i}\left(\eta_j + \left(d-1+\frac{\varepsilon}{2}\right)\ell + \log_{2m-1}\left(\frac{2m}{2m-1}\right)\right).$$

Without loss of generality, suppose that $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_k$, with convention $\alpha_{i+1} = 0$. Note that $\log_{2m-1}(P_k)$ is negative and that $\alpha_1 \le |\widetilde{Y}| = \sum_{i=i}^{k}\alpha_i$, so $|\widetilde{Y}|\log_{2m-1}(P_k) \le \alpha_1 \log_{2m-1}(P_k)$. By Abel's summation formula,

$$\begin{aligned}
|\widetilde{Y}|\log_{2m-1}(P_k) \le \alpha_1 \log_{2m-1}(P_k) &= \sum_{i=1}^{k}(\alpha_i - \alpha_{i+1})\log_{2m-1}(P_k) \\
&\le \sum_{i=1}^{k}(\alpha_i - \alpha_{i+1})\sum_{j=1}^{i}\left[\eta_i + \left(d-1+\frac{\varepsilon}{2}\right)\ell + \log_{2m-1}\left(\frac{2m}{2m-1}\right)\right] \\
&= \sum_{i=1}^{k}\alpha_i\left[\eta_i + \left(d-1+\frac{\varepsilon}{2}\right)\ell + \log_{2m-1}\left(\frac{2m}{2m-1}\right)\right] \\
&= \sum_{i=1}^{k}\alpha_i\eta_i + \left(\sum_{i=1}^{k}\alpha_i\right)\left[\left(d-1+\frac{\varepsilon}{2}\right)\ell + \log_{2m-1}\left(\frac{2m}{2m-1}\right)\right].
\end{aligned}$$

Note that $\sum_{i=1}^{k} = |\widetilde{Y}|$. By Lemma 4.9 and the hypothesis of Lemma 4.3,

$$\sum_{i=1}^{k}\alpha_i\eta_i \le \mathrm{Edge}(\widetilde{Y}) + \mathrm{Red}(\widetilde{Y}) < (1-d-\varepsilon)|\widetilde{Y}|\ell.$$

Hence

$$\begin{aligned}|\widetilde{Y}|\log_{2m-1}(P_k) &\leq (1-d-\varepsilon)|\widetilde{Y}|\ell + |\widetilde{Y}|\left[\left(d-1+\frac{\varepsilon}{2}\right)\ell + \log_{2m-1}\left(\frac{2m}{2m-1}\right)\right] \\ &\leq |\widetilde{Y}|\left[-\frac{\varepsilon}{2}\ell + \log_{2m-1}\left(\frac{2m}{2m-1}\right)\right].\end{aligned}$$

□

Lemma 4.12. *Let $K > 0$. For ℓ large enough, the number of abstract van Kampen 2-complexes of complexity K with face boundary lengths at most ℓ is bounded by ℓ^{3K}.*

Proof. As there are at most K maximal arcs, the number of vertices of valency greater than 3 is at most $2K/3 \leq K$. There are at most K^2 ways to draw an arc connecting two of these vertices, every arc is with length at most ℓ. The number of underlying graphs is then at most $(K^2)^K\ell^K$.

To attach K faces on a graph, we choose K loops passing by at most K arcs, there are at most $(K^2)^{K^2}$ choices. There are at most $(2\ell)^K$ ways to choose a starting point and an orientation for every face we; and at most K^{2K} ways to label the faces by abstract relators $\{1^\pm, \ldots, K^\pm\}$. The number of such abstract 2-complex is bounded by

$$(K^2)^K\ell^K \times (K^2)^{K^2} \times (2\ell)^K \times K^{2K},$$

which is much smaller than ℓ^{3K} if ℓ is large enough. □

Proof of Theorem 4.3. Under the condition $Q_\ell = (2m-1)^{(d-\frac{\varepsilon}{4}\ell)} \leq |R_\ell| \leq (2m-1)^{(d+\frac{\varepsilon}{4}\ell)}$, the probability that there exists a van Kampen 2-complex of complexity K of $G_\ell(m,d)$ satisfying

$$\mathrm{Edge}(Y) + \mathrm{Red}(Y) < (1-d-\varepsilon)|Y|\ell \tag{$*$}$$

is smaller than

$$\sum_{\widetilde{Y} \text{ is of complexity } K, \text{ satisfying } (*)} \mathbf{Pr}\left(\widetilde{Y} \text{ is fillable by } \middle| Q_\ell\right).$$

By Lemma 4.12 and Lemma 4.11, it is smaller than

$$\ell^{3K}\left(\frac{2m}{2m-1}\right)(2m-1)^{-\frac{\varepsilon}{2}\ell}$$

which converges to 0 as $\ell \to \infty$.

By definition $\mathbf{Pr}(Q_\ell) \xrightarrow[\ell\to\infty]{} 1$, so the probability that there exists a van Kampen 2-complex of $G_\ell(m,d)$ of complexity K satisfying $(*)$ converges to 0 as ℓ goes to infinity. That is to say, a.a.s. every van Kampen diagram of $G_\ell(m,d)$ of complexity K satisfies the inequality

$$\mathrm{Edge}(Y) + \mathrm{Red}(Y) \geq (1-d-\varepsilon)|Y|\ell.$$

□

4.2 Existence of van Kampen 2-complexes

By Theorem 4.3, we may ask that, if a sequence of 2-complexes (Y_ℓ) with bounded complexity satisfies the inequality

$$\mathrm{Edge}(Y_\ell) \geq (1-d+s)|Y_\ell|\ell$$

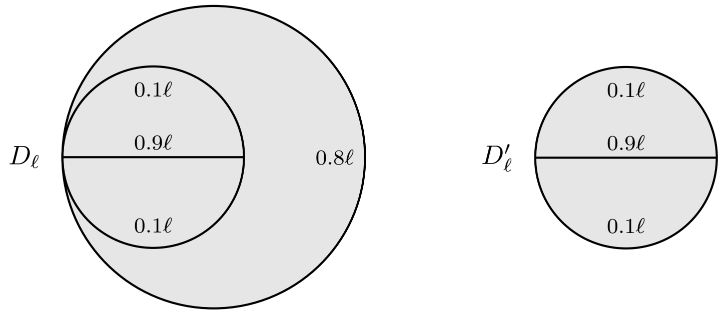
with some $s > 0$, does there exist a sequence of reduced van Kampen 2-complexes (D_ℓ) of $(G_\ell(m,d))$ such that underlying 2-complex of D_ℓ is Y_ℓ?

Although the "dimension reasoning" of Y. Ollivier in [Oll05] p.30 suggested that we have this intuition, the answer is no in general. We give here a counterexample.

4.2.1 A counterexample

Let $G_\ell(m,d)$ be a sequence of random groups at density $d = 0.4$. Let (D_ℓ) be a sequence of 2-complexes where D_ℓ is of the following form.

The given inequality is satisfied because $\mathrm{Edge}(D_\ell) = 1.9\ell > 1.8\ell = (1-d)|D_\ell|\ell$. However, the sub-diagram D'_ℓ contradicts Theorem 4.3 ($\mathrm{Edge}(D'_\ell) = 1.1\ell < 1.2\ell = (1-d)|D'_\ell|\ell$), and can not be a van Kampen diagram of $G_\ell(m,d)$.



In this section, we show that if every *sub-complex* of a 2-complex satisfies the given inequality, then with high probability it is an underlying 2-complex of a van Kampen 2-complex of $G_\ell(m,d)$. This gives a phase transition for the existence of a geometric form of van Kampen 2-complexes in a random group at density (Theorem H, Theorem 4.15). This result can be regarded as a generalization Theorem 2.46 (phase transition for the $C'(\lambda)$ condition).

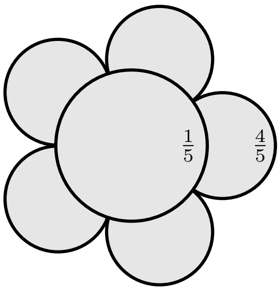
4.2.2 Definitions and statement of the theorem

Let $Y = (V,E,F)$ be a finite, connected 2-complex without isolated edges. Denote here $\mathrm{Edge}(Y) = \{\{e,e^{-1}\}|e \in E\}$ the set of undirected edges. To simplify notations, we denote e instead of $\{e,e^{-1}\}$ for undirected edges in this subsection. A *length label* on Y is a label on undirected edges of Y by real numbers $\lambda : \mathrm{Edge}(Y) \to]0,1]$. For any edge $e \in \mathrm{Edge}(Y)$, the number $\lambda_e := \lambda(e)$ is the called the *length* of e. If $p = e_1 \dots e_k$ is a path on Y, the length of p is $|p| = \sum_{i=1}^k \lambda_{e_i}$.

Definition 4.13. *A* ***marked 2-complex*** *is a couple (Y,λ) where Y is a 2-complex and λ is a length label, such that for every face f of Y, the boundary length $|\partial f|$ is bounded by* 1*.*

Let (Y_ℓ) be a sequence of 2-complexes where Y_ℓ is obtained from Y by dividing every edge e into $\lfloor \lambda_e \ell \rfloor$ edges of length 1*. The sequence (Y_ℓ) is called the* ***sequence of divided 2-complexes*** *associated to (Y,λ). We say that the sequence of 2-complexes (Y_ℓ) is with the same* ***geometric form*** *(Y,λ).*

For example, for the $C(5)$ small cancellation condition, we study the following geometric form of 2-complexes.



Note that Every face f of Y_ℓ is with boundary length at most ℓ. We may replace the sequence of lengths $(\lfloor \lambda_e \ell \rfloor)_{\ell \in \mathbb{N}}$ by any sequence with $\lambda_e \ell + o(\ell)$. If Z is a sub 2-complex of Y, we denote $Z < Y$. By convention, if (Z_ℓ) is the sequence of divided sequences of Z, we have $Z_\ell < Y_\ell$ for any integer ℓ.

Definition 4.14. *Let (Y, λ) be a marked 2-complex with coefficients $(\lambda_e)_{e \in \mathrm{Edge}(Y)}$. The density of Y is*

$$\mathrm{dens}(Y) = \frac{\sum_{e \in \mathrm{Edge}(Y)} \lambda_e}{|Y|}.$$

*The **critical density** of Y is*

$$\mathrm{dens}_c(Y) = \min_{Z < Y} \{\mathrm{dens}(Z)\}.$$

Let $(G_\ell(m, d)) = (\langle X_m | R_\ell \rangle)$ be a sequence of random groups with m generators at density d. We say that Y_ℓ is fillable by $G_\ell(m, d)$ (or by R_ℓ) if there exists a reduced van Kampen 2-complex of $G_\ell(m, d)$ whose underlying 2-complex is Y_ℓ.

Theorem 4.15 (Theorem H)**.** *Let Y be a marked 2-complex, and let (Y_ℓ) be the sequence of divided 2-complexes associated to Y.*

(i) If $\mathrm{dens}_c Y + d < 1$, then a.a.s. Y_ℓ is not fillable by $G_\ell(m, d)$.

(ii) If $\mathrm{dens}_c Y + d > 1$ and Y_ℓ is fillable by B_ℓ, then a.a.s. Y_ℓ is fillable by $G_\ell(m, d)$.

Note that in the second assertion, we need that Y_ℓ has at least one filling by the set of all possible relators B_ℓ.

4.2.3 Proof of Theorem H (Theorem 4.15)

The first assertion can be proved by Theorem 4.3.

Proof of Theorem 4.15 (i). Suppose that $\mathrm{dens}_c Y + d < 1$. There exists a sub-2-complex $Z < Y$ satisfying $\mathrm{dens}\, Z + d < 1$. Let (Z_ℓ) be the sequence of divided 2-complexes of Z. We shall prove that a.a.s. Z_ℓ is not fillable by $G_\ell(m, d)$ by Theorem 4.3.

Let $\varepsilon > 0$ such that $\mathrm{dens}\, Z = 1 - d - 3\varepsilon$. By definition,

$$\lim_{\ell \to \infty} \frac{\mathrm{Edge}(Z_\ell)}{|Z_\ell| \ell} = 1 - d - 3\varepsilon,$$

so for ℓ large enough

$$\mathrm{Edge}(Z_\ell) \leq (1-d-2\varepsilon)|Z_\ell|\ell < (1-d-\varepsilon)|Z_\ell|.$$

The complexity of Z_ℓ is $K = \max\left\{|Z|, \mathrm{Edge}(Z), \max\{\frac{1}{\lambda_e} \,|\, e \text{ edge of } Z\}\right\}$, independent of ℓ. Apply Theorem 4.3 with ε and K given above, a.a.s. every van Kampen 2-complex Z of $G_\ell(m,d)$ of complexity K satisfies

$$\mathrm{Edge}(Z) \geq (1-d-\varepsilon)\ell|Z|.$$

So a.a.s. Z_ℓ is not fillable by $G_\ell(m,d)$, which implies that a.a.s. Y_ℓ is not fillable by $G_\ell(m,d)$. □

To prove the second assertion, we need the multidimensional intersection formula (Theorem C, Theorem 2.40).

Recall that B_ℓ is the set of cyclically reduced words of length at most ℓ. For an integer k, denote $B_\ell^{(k)}$ the set of k-tuples of pairwise distinct relators $(r_1,\dots,r_k)$ in B_ℓ. Note that $|B_\ell^{(k)}| = (2m-1)^{k\ell+o(\ell)}$. Recall the following notions in Section 2.3.

Definition 4.16 (Self-intersection partition). *Let $(\mathcal{Y}_\ell)$ be a sequence of fixed subsets of the sequence $\left(B_\ell^{(k)}\right)$. For $0 \leq i \leq k$, the i-th self-intersection of $\mathcal{Y}_\ell$ is*

$$S_{i,\ell} := \{(x,y) \in \mathcal{Y}_\ell^2 \,|\, |x \cap y| = i\}$$

where $|x \cap y|$ is the number of common elements between the sets $x = (r_1,\dots,r_k)$ and $y = (r'_1,\dots,r'_k)$.

The family of subsets $\{S_{i,\ell} \,|\, 0 \leq i \leq k\}$ is a partition of $\mathcal{Y}_\ell^2$, called the *self-intersection partition* of $\mathcal{Y}_\ell$. Note that $(S_{i,\ell})$ is a sequence of subsets of the sequence $\left((B_\ell^{(k)})^2\right)_{\ell\in\mathbb{N}}$, with density smaller than $\mathrm{dens}_{\left((B_\ell^{(k)})^2\right)}\left(\mathcal{Y}_\ell^2\right) = \mathrm{dens}_{(B_\ell^{(k)})}(\mathcal{Y}_\ell)$.

Definition 4.17 (Definition 2.33). *Let $(\mathcal{Y}_\ell)$ be a sequence of fixed subsets of $\left(B_\ell^{(k)}\right)$ with density α. Let $S_{i,\ell}$ with $0 \leq i \leq k$ be its self-intersection partition. Let $d > 1-\alpha$. We say that $(\mathcal{Y}_\ell)$ satisfies the d-small self-intersection condition if, for every $1 \leq i \leq k-1$,*

$$\mathrm{dens}_{\left((B_\ell^{(k)})^2\right)}(S_{i,\ell}) < \alpha - (1-d) \times \frac{i}{2k}.$$

Theorem 4.18 (Theorem 2.40). *Let (R_ℓ) be a sequence of permutation invariant random subsets of (B_ℓ) of density d. Let $(\mathcal{Y}_\ell)$ be a sequence of subsets of $\left(B_\ell^{(k)}\right)$ of density $\alpha > 1-d$. If $(\mathcal{Y}_\ell)$ satisfies the d-small self intersection condition, then the sequence of random subsets $(\mathcal{Y}_\ell \cap R_\ell^{(k)})$ is densable with density $\alpha + d - 1$.*

In particular, a.a.s. the random set $\mathcal{Y}_\ell \cap R_\ell^{(k)}$ is not empty.

In the following k is the number of faces of Y, and $\mathcal{Y}_\ell$ is the set of fillings of Y_ℓ by pairwise distinct relators in B_ℓ. It is a subset of $B_\ell^{(k)}$. Let $(G_\ell(m,d)) = (\langle X_m | R_\ell \rangle)$ be a sequence of random groups at density d. If $\mathcal{Y}_\ell$ satisfies the hypothesis of Theorem 2.40, then a.a.s. Let $(G_\ell(m,d)) = (\langle X_m | R_\ell \rangle)$ be a sequence of random groups with m generators at density d, which means that there exists a k-tuple of pairwise distinct relators in R_ℓ that fills Y_ℓ, so Y_ℓ is fillable by $G_\ell(m,d)$.

It remains to prove that if $\operatorname{dens}_c Y > 1-d$, then the sequence $(\mathcal{Y}_\ell)$ is densable (Lemma 4.21) and it satisfies the d-small self intersection condition (Lemma 4.22).

To construct a van Kampen diagram by the set of all relators B_ℓ from the 2-complex Y_ℓ, we may start by filling edges in the neighborhoods of vertices that are originally vertices of Y (before dividing).

Definition 4.19 (Vertex labeling)**.** *Consider the set of oriented edges of Y_ℓ starting at some vertex that is originally a vertex of Y before dividing. A **vertex labeling** is a labeling on these edges by $X_m^\pm$ such that, for every pair of different edges e_1, e_2 starting at the same vertex and labeled by the same generator $x \in X_m^\pm$, the path $e_1^{-1}e_2$ is not cyclically part of a face boundary loop.*

As $m \geq 2$ and $\lfloor \lambda_e \ell \rfloor \geq 3$ for ℓ large enough, if there exists an vertex labeling, then it can be completed as a van Kampen 2-complex (with respect to B_ℓ).

Lemma 4.20. *Let $\overline{\mathcal{Y}_\ell}$ be the set of k-tuples of relators in B_ℓ filling Y_ℓ. If Y_ℓ is fillable by B_ℓ, then*

$$\operatorname{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell}) = \operatorname{dens} Y.$$

Proof. We shall estimate $|\overline{\mathcal{Y}_\ell}|$ by counting the number of labelings on edges of Y_ℓ that produce van Kampen diagrams. As the 2-complex Y_ℓ is fillable, the set of vertex labelings is not empty. Denote $C \geq 1$ the number of vertex labelings.

To label the remaining $\lfloor \lambda \ell \rfloor - 2$ edges on the arc divided from the edge $e \in Edge(Y)$, there are $2m-1$ choices for the first $\lfloor \lambda \ell \rfloor - 3$ edges, and $2m-2$ or $2m-1$ choices for the last edge. So

$$C \prod_{e \in \operatorname{Edge}(Y)} (2m-1)^{\lfloor \lambda_e \ell \rfloor - 3}(2m-2) \leq |\overline{\mathcal{Y}_\ell}| \leq C \prod_{e \in \operatorname{Edge}(Y)} (2m-1)^{\lfloor \lambda_e \ell \rfloor - 2}.$$

Recall that $k = |Y| = |Y_\ell|$, note that $|B_\ell^k| = (2m-1)^{k\ell + o(\ell)}$. We have

$$\operatorname{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell}) = \frac{\sum_{e \in \operatorname{Edge}(Y)} \lambda_e}{|Y|} = \operatorname{dens} Y.$$

□

Lemma 4.21. *If $dens_c Y > 1/2$ and Y_ℓ is fillable by B_ℓ, then $(\mathcal{Y}_\ell)$ is densable in $(B_\ell^{(k)})$ and*

$$\operatorname{dens}_{(B_\ell^{(k)})}(\mathcal{Y}_\ell) = \operatorname{dens} Y.$$

Proof. Let $Z < Y$ with two faces f_1, f_2. Because $\operatorname{dens} Z \geq \operatorname{dens}_c Y > \frac{1}{2}$, we have

$$\sum_{e \in \operatorname{Edge}(Z)} \lambda_e > \frac{1}{2}|Z| = 1 \geq |\partial f_1|.$$

Let $\overline{\mathcal{Y}_\ell^Z}$ be the set of fillings of Y_ℓ by B_ℓ such that the two faces of Z are filled by the same relator. By the same arguments of the previous lemma,

$$|\overline{\mathcal{Y}_\ell^Z}| \leq C(2m-1)^{|\partial f_1|} \prod_{e \in \operatorname{Edge}(Y) \setminus \operatorname{Edge}(Z)} (2m-1)^{\lfloor \lambda_e \ell \rfloor - 2},$$

so

$$\begin{aligned}\mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell^Z}) &\le \frac{1}{|Y|}\left[\sum_{e\in\mathrm{Edge}(Y)}\lambda_e + \left(|\partial f_1| - \sum_{e\in\mathrm{Edge}(Z)}\lambda_e\right)\right] \\ &< \frac{\sum_{e\in\mathrm{Edge}(Y)}\lambda_e}{|Y|} \\ &= \mathrm{dens}\, Y = \mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell}).\end{aligned}$$

We know that

$$\mathcal{Y}_\ell = \overline{\mathcal{Y}_\ell} \Big\backslash \bigcup_{Z<Y,|Z|=2} \overline{\mathcal{Y}_\ell^Z}\,,$$

so

$$|\overline{\mathcal{Y}_\ell}| - \sum_{Z<Y,|Z|=2} |\overline{\mathcal{Y}_\ell^Z}| \le |\mathcal{Y}_\ell| \le |\overline{\mathcal{Y}_\ell}|.$$

There are $\binom{|Y|}{2}$ terms in the union, with densities strictly smaller than $\mathrm{dens}(\overline{\mathcal{Y}_\ell})$. They do not affect the density according to Proposition 2.14 and Proposition 2.13. Together with the previous lemma,

$$\mathrm{dens}_{(B_\ell^k)}(\mathcal{Y}_\ell) = \mathrm{dens}_{(B_\ell^k)}(\overline{\mathcal{Y}_\ell}) = \mathrm{dens}\, Y.$$

Because $\mathrm{dens}_{(B_\ell^k)}(B_\ell^{(k)}) = 1$, by the random-fixed intersection formula (Theorem 2.28), we have

$$\mathrm{dens}_{(B_\ell^{(k)})}(\mathcal{Y}_\ell) = \mathrm{dens}\, Y.$$

□

Recall that $\mathcal{Y}_\ell$ is the set of fillings of Y_ℓ by pairwise distinct relators in B_ℓ. Denote $k = |Y|$.

Lemma 4.22. *Suppose that* $\mathrm{dens}_c Y > 1 - d$. *Let* $S_{i,\ell}$ *be the* i*-th self intersection of the set* $\mathcal{Y}_\ell$. *We have*

$$\mathrm{dens}_{\left((B_\ell^{(k)})^2\right)}(S_{i,\ell}) < \mathrm{dens}\, Y - (1-d) \times \frac{i}{2k}.$$

Proof. Let Z, W be sub complexes of Y with $|Z| = |W| = i < k = |Y|$. Let $(Z_\ell), (W_\ell)$ be the corresponding sequences of dividing 2-complexes. Denote $S_\ell(Z, W)$ the set of pairs of fillings $((r_1, \ldots, r_k), (r'_1 \ldots, r'_k))$ of Y_ℓ by all possible relators B_ℓ (i.e. a pair of k-tuples of relators) such that, the i relators in the first filling $(r_1, \ldots, r_k)$ corresponding to Z_ℓ are identical to the i relators in the second filling $(r'_1, \ldots, r'_k)$ corresponding to W_ℓ, and that the other $2k - 2i$ relators are pairwise different, not repeating the relators in Z_ℓ and W_ℓ.

To estimate the cardinality $|S_\ell(Z, W)|$, we first fill the first k-tuple $(r_1, \ldots, r_k)$, so the i relators in the second k-tuple $(r'_1, \ldots, r'_k)$ corresponding to the sub 2-complex W_ℓ is determined, and there are at most $i!$ choices for ordering these i relators. To fill the remaining $k - i$ relators, by the same arguments of Lemma 4.20, we get

$$|S_\ell(Z, W)| \le |\mathcal{Y}_\ell| \times i! \times C \prod_{e\in\mathrm{Edge}(Y)\setminus\mathrm{Edge}(W)} (2m-1)^{\lfloor\lambda_e \ell\rfloor - 2}.$$

Recall that the density of Y is $\mathrm{dens}\, Y = \frac{1}{|Y|}\left(\sum_{e\in Edge(Y)}\lambda_e\right)$, and that the critical density $\mathrm{dens}_c Y$ is the minimum of densities of its sub-2-complexes (Definition 4.14). Together with the hypothesis $\mathrm{dens}\, W \ge$

$\operatorname{dens}_c Y > 1 - d$, we have

$$\begin{aligned}
\operatorname{dens}_{\left((B_\ell^{(k)})^2\right)}(S_\ell(Z,W)) &\leq \frac{1}{2k}\left(\sum_{e\in Edge(Y)}\lambda_e + \sum_{e\in \mathrm{Edge}(Y)\setminus \mathrm{Edge}(W)}\lambda_e\right)\\
&= \frac{1}{2k}\left(2\sum_{e\in Edge(Y)}\lambda_e - \sum_{e\in \mathrm{Edge}(W)}\lambda_e\right)\\
&= \operatorname{dens} Y - \frac{i}{2k}\operatorname{dens} W\\
&< \operatorname{dens} Y - \frac{i}{2k}(1-d).
\end{aligned}$$

Note that

$$S_{i,\ell} = \bigcup_{Z<Y, W<Y, |Z|=|W|=i} S_\ell(Z,W).$$

It is a union of $\binom{k}{i}^2$ subsets of densities strictly smaller than $\operatorname{dens} Y - \frac{i}{2k}(1-d)$, so by Proposition 2.13

$$\operatorname{dens}_{\left((B_\ell^{(k)})^2\right)}(S_{i,\ell}) < \operatorname{dens} Y - \frac{i}{2k}(1-d).$$

□

This completes the proof of Theorem 4.15.

4.3 Applications to small cancellation theory

Let $(G_\ell(m,d))$ be a sequence of random groups in the Gromov density model. We give here two phase transitions for small cancellation theory. Theorem 4.15 allows us to construct the 2-complexes needed.

4.3.1 The $C'(\lambda)$ small cancellation condition

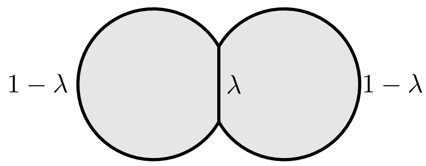
Recall that ([LS77] p.240) a piece with respect to a set of relators is a cyclic sub-word that appears at least twice, and a group presentation satisfies the $C'(\lambda)$ condition ([LS77] p.240) if the length of a piece is at most λ times the length of any relator that it appears.

The following result has been proved in [BNW20], and in Section 2.4 of this thesis using the multidimensional intersection formula (Theorem 2.40). We give here a much simpler proof using Theorem 4.15.

Theorem 4.23. *Let $0 < \lambda < 1$. There is a phase transition at density $d = \lambda/2$:*

(i) If $d < \lambda/2$, then a.a.s. $G_\ell(m,d)$ satisfies $C'(\lambda)$.

(ii) If $d > \lambda/2$, then a.a.s. $G_\ell(m,d)$ does not satisfy $C'(\lambda)$.



Proof. We refer to the proof of Theorem 2.46 (i) for the first assertion.

For the second assertion, consider a marked 2-complex Y with two faces of boundary length 1, sharing a common edge of length λ. We have $\operatorname{dens} Y = \frac{2(1-\lambda)+\lambda}{2} > 1 - d$, and every sub 2-complex is with density $1 > 1 - d$. So $\operatorname{dens}_c Y > 1 - d$.

By Theorem 4.15, a.a.s. there exists a reduced van Kampen diagram D_ℓ of $G_\ell(m, d)$ having two faces of boundary length ℓ, sharing a common side of length $\lfloor \lambda\ell \rfloor$. Hence, a.a.s. $G_\ell(m, d)$ does not satisfy $C'(\lambda)$. □

4.3.2 The $C(p)$ small cancellation condition

Recall that a group presentation satisfies the $C(p)$ small cancellation condition if no relator is a product of fewer than p pieces. To deal with this condition, we need to consider diagrams with $(p+1)$ faces. Compared to the $C'(\lambda)$ condition (with only 2 faces), it is a lot harder to prove the existence of such a diagram only by the multidimensional intersection formula. The proof of the following result shows us the convenience of Theorem 4.15 for proving the existence of certain van Kampen 2-complexes.

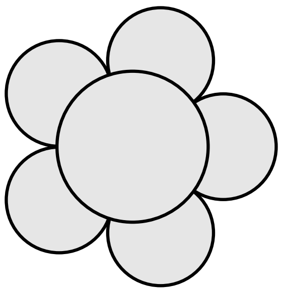
Theorem 4.24. *Let $p \geq 2$ be an integer. There is a phase transition at density $d = 1/(p+1)$:*

(i) If $d < 1/(p+1)$, then a.a.s. $G_\ell(m, d)$ satisfies $C(p)$.

(ii) If $d > 1/(p+1)$, then a.a.s. $G_\ell(m, d)$ does not satisfy $C(p)$.

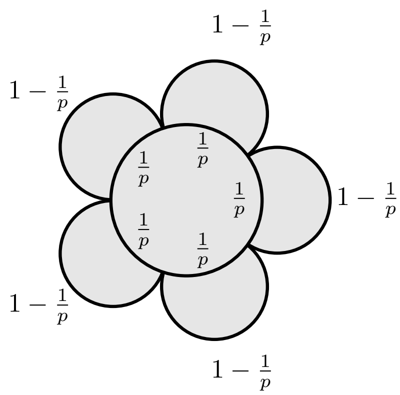
Proof.

(i) Let us prove by contradiction. Suppose that a.a.s. $G_\ell(m, d) = \langle X_m | R_\ell \rangle$ does not satisfy $C(p)$. That is to say, a.a.s. there exists a reduced van Kampen diagram D of R_ℓ with $(p+1)$ faces, one face is placed in the center, attached by the other p faces on the whole boundary, and there is no other attachments.



We have $|D| = p + 1$ and $\operatorname{Edge}(D) \leq p\ell$. By Theorem 4.3 with $\varepsilon = \left(\frac{1}{p+1} - d\right)/2$, we have $\operatorname{Edge}(D) \geq (1 - d - \varepsilon)\ell|D|$. So $d \geq 1/(p+1) - \varepsilon > d$, which gives a contradiction.

(ii) Consider a marked 2-complex with $p+1$ faces, one of the faces is placed in the center, having p edges of length $1/p$, such that every edge is attached by another faces with two edges of lengths $1/p$ and $1-1/p$. There are no other attachments.

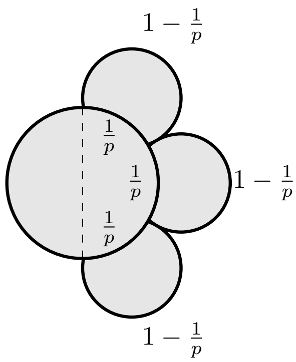


The density of Y is $\frac{1+p(1-1/p)}{p+1} = p/(p+1) > 1-d$. If Z is a sub 2-complex of Y not containing the center face, then $\operatorname{dens} Z = 1$. If Z contains the center face and $i \leq p$ other faces, then $\operatorname{dens} Z = \frac{1+i(1-1/p)}{i+1} > 1-d$. So $\operatorname{dens}_c Y > 1-d$.

By Theorem 4.15, a.a.s. there exists a reduced van Kampen diagram D_ℓ of $G_\ell(m,d)$ having $p+1$ faces of boundary length ℓ, where one of the faces is attached by p faces. Hence, a.a.s. $G_\ell(m,d)$ does not satisfy $C(p)$.

□

The same argument holds for the $B(2p)$ condition (c.f. [OW11] Definition 1.7 by Y. Ollivier and D. Wise): half of a relator can not be the product of fewer than p pieces. One can construct a marked 2-complex with p faces, one of the faces is in the center, with half of its boundary attached by the other p faces, each with length $1/p$. Its critical density is $\frac{p+1/2}{p+1}$, so a phase transition occurs at density $d = \frac{1}{2(p+1)}$.



Chapter 5

Open questions

In addition to Questions 1, 2 and 3 proposed in the Introduction, we propose here a list of open questions for random groups in different models, especially for phase transitions.

5.1 General Freiheitssatz for random groups

Let $(G_\ell(m,d))$ be a sequence of random groups with m generators at density d. Let $1 \leq r \leq m-1$. Our Question 2 ask if there is a phase transition for the property "every r-generated subgroup is free". In Section 3.3, we propose a critical density

$$d_r := \min\left\{\frac{1}{2}, 1 - \log_{2m-1}(2r-1)\right\}.$$

Theorem 3.22 (i) shows that if $d > d_r$, then a.a.s. the first r generators generate the whole group $G_\ell(m,d)$, so the "every r-generated subgroup is free" property does not hold. In the case $d < d_r$, Theorem 3.22 (ii) shows that a.a.s. every subgroup generated by a labeled graph Γ of rank r with $|\Gamma| \leq \frac{d_r - d}{5}\ell$ is free. By G. Arzhantseva and A. Ol'shanskii's method in [AO96], the "every r-generated subgroup is free" property is a.a.s. satisfied for random groups with a much smaller density $d < \frac{1}{120m^2\ln(2m)}$ (c.f. Theorem 2.48). We wish to push this density up to our critical density d_r, which fully answers Question 2.

Question 4. *Let $(G_\ell(m,d))$ be a sequence of random groups at density d. Is it true that, if $d < d_r$, then a.a.s. every r-generated subgroup of $G_\ell(m,d)$ is free?*

Our method is not enough to achieve the result for *every* r-generated subgroups, only for subgroups generated by "microscopic" graphs (of size much smaller than the hyperbolicity constant $\frac{4\ell}{1-2d}$ of $G_\ell(m,d)$).

By [Gro87] 5.3.A (c.f. [KW03] or [Arz06] for detailed proofs), as a.a.s. $G_\ell(m,d)$ is $\frac{4\ell}{1-2d}$ hyperbolic, there exists a constant $K = K(m,d)$ such that, every subgroup generated by a labeled graph Γ of rank r with every maximal arc longer than $K\ell$ is free and quasi-convex. This case is called the "macroscopic" case. It remains the "mesoscopic" case: what happens when $c\ell \leq |\Gamma| \leq C\ell$ with a given pair of constants c, C depending on m and d?

The statement of Question 4 has an interesting corollary: if $d_r < d < d_{r-1}$, then a.a.s. the rank of $G_\ell(m,d)$ is r. Effectively, by Theorem 2.45 (ii) shows that a.a.s. $G_(m,d)$ is not free, so it is not of rank $(r-1)$ because $d < d_{r-1}$, and Theorem 3.22 (i) shows that a.a.s. $G_\ell(m,d)$ is r-generated because $d > d_r$.

5.2 Parallel geodesics in Cayley graphs and the Burnside problem for random groups

Fix an alphabet $X_m = \{x_1 \ldots, x_m\}$ as generators of group presentations. Denote $F(X_m)$ the free group generated by X_m. Recall that a (free) Burnside group with m generators and exponent n, defined by $B(m,n) = \{X_m | x^n, x \in F(X_m)\}$ is *infinite* when n is large enough.

In [GM18], D. Gruber and J. Mackay studied random triangular groups based on Burnside groups. Recall that (model 1.1.2.d) a sequence of random triangular groups $(G_m(d))$ at density d is defined by $G_m(d) = \langle X_m | R_m \rangle$ where R_m is uniformly chosen among all sets of triangular relators of cardinality between $c^{-1}(2m-1)^{3d}$ and $c(2m-1)^{3d}$ with some $c > 1$.

Definition ([GM18] Definition 1.1)**.** A sequence of random triangular n-periodic groups $(G_m(d,n))$ with density d is defined by

$$G_m(d,n) = \langle X_m | R_m, x^n, x \in F(X_m) \rangle$$

where $(\langle X_m | R_m \rangle)$ is a sequence of random triangular groups at density d.

It can be regarded either as a quotient of a Burnside group by random triangular relators or as a Burnside quotient of a random triangular group. The main result of Gruber and Mackay in [GM18] is the following.

Theorem ([GM18] Theorem 1.2)**.** *Let $(G_m(d,n))$ be a sequence of random triangular n-periodic groups at density d. For any $d_0 < \frac{11-\sqrt{41}}{12} \approx 0.38307$, there exists an integer $n_0 \in \mathbb{N}$ such that for any $0 < d \leq d_0$ and any $n \geq n_0$, a.a.s. the random n-periodic triangular group $G_m(d,n)$ is infinite.*

As their method does not work for densities larger than ≈ 0.38307, Gruber and Mackay asked that does their theorem hold for random triangular groups at any density $d_0 < 1/2$ ([GM18] Question 1.6). We may ask that, is there an analog for the Gromov density model of random groups? Let us define a sequence of random n-periodic groups $(G_\ell(m,d,n))$ with density d by

$$G_\ell(m,d,n) = \langle X_m | R_\ell, x^n, x \in F(X_m) \rangle$$

where $(\langle X_m | R_\ell)$ is a sequence of random groups at density d.

Question 5. *For any $d_0 < 1/2$, does there exist $n_0 \in \mathbb{N}$ such that for any $0 < d \leq d_0$ and any $n \geq n_0$, a.a.s. the random n-periodic group $G_\ell(m,d,n)$ is infinite?*

Parallel geodesics The main technical point in [GM18] is the estimation of the number of *parallel geodesics* in a random triangular group.

Recall that random triangular groups at density $d < 1/2$ are a.a.s. hyperbolic (c.f. model 1.1.2.d). For a hyperbolic group $G = \langle X | R \rangle$, two geodesics in the Cayley graph $\mathrm{Cay}(G,X)$ are called *parallel* if they have the same pair of limit points (c.f. [CDP90] for a definition) in the hyperbolic boundary of $\mathrm{Cay}(G,X)$ and they have no intersection. A set of $k \geq 2$ geodesics are called *parallel* if they are pairwise parallel. Denote by $P(G)$ the maximal number k such that k geodesics can be parallel in the hyperbolic space $\mathrm{Cay}(G,X)$. If $\mathrm{Cay}(G,X)$ is δ-hyperbolic, as the distance between two parallel geodesics is most 2δ (c.f. [CDP90] Chapter 2), we know that $P(G)$ is bounded above by a number $C \exp a\delta$ with some universal constants C and a.

Using an analog of the isoperimetric inequality for 2-complexes (Theorem 2.45 (ii), [GM18] Theorem 2.6), Gruber and Mackay obtained an upper bound of the number of parallel geodesics in a random triangular group. In [GM18] Section 3, it is shown that for a sequence of random triangular groups $(G_m(d))$ at density

d, if $d < \frac{11-\sqrt{41}}{12}$, then a.a.s. there exists an integer $k = k(d)$ such that $P(G_m(d)) \leq k$. In particular, a.a.s. the injectivity radius of $G_m(d)$ is at least $1/k$.

By applying Theorem 4.3, we can establish an analog for the density model of random groups $(G_\ell(m, d))$, with an upper bound independent of ℓ (hence independent of the hyperbolicity constant $\frac{4\ell}{1-2d}$ of $G_\ell(m, d)$).

Proposition. *Let $(G_\ell(m, d))$ be a sequence of random groups at density d. If $d < \frac{1}{4}$, then a.a.s.*

$$P(G_\ell(m, d)) \leq 2 + \frac{2d}{1-4d}.$$

However, this proposition is not enough to answer Question 5, even for smaller densities $d < 1/4$. There is an essential difference between the triangular density model and the traditional Gromov density model: The hyperbolicity constant of (the Cayley graph of) a random triangular group $G_m(d)$ at density $d < 1/2$ is bounded by $\frac{12}{1-2d}$. This constant is independent of the sequence index m, which is an important point in Gruber-Mackay's proof. While the hyperbolicity constant for a random group $G_\ell(m, d)$ at density $d < 1/2$ is $\frac{4\ell}{1-2d}$, which depends on ℓ.

Since the maximal number of parallel geodesics in the above proposition *diverges* as the density d increases to $1/4$, we believe that when $d > 1/4$, this number is no longer uniformly bounded (i.e. increases with the relator length ℓ), and there is a phase transition at density $1/4$.

Question 6. *If $d > \frac{1}{4}$, is it true that for any $k \geq 0$ a.a.s. $P(G_\ell(m, d)) > k$?*

5.3 The graph model of random groups

As we have seen in Subsection 3.1.1, a finite connected graph Γ labeled by $X_m^\pm$ generate a subgroup of a finitely presented group $G = \langle X_m | R \rangle$. Denote G/Γ the quotient of G by the normal subgroup generated by Γ. If Γ is *randomly labeled* graph, then G/Γ defines a random group. The first appearance of this model is in [Gro03] p.141, introduced by M. Gromov, to construct a finitely presented group that can not be coarsely embedded into any Hilbert space, called *Gromov's monster group* in [AD08].

Note that this notion generalizes the uniform model of random finitely presented groups: Let $F_m = F(X_m)$ be the free group generated by X_m. Fix an integer $k \geq 1$, if Γ is a wedge of k simple cycles of length ℓ with reduced label on each cycle (not necessary on the common vertex), then F_m/Γ defines a uniform random group with k generators $G = \langle x_1, \dots, x_m | r_1, \dots, r_k \rangle$.

If (Γ_n) is a "well-chosen" interesting sequence of random labeled graphs, we may discuss asymptotic behaviors of the sequence of random groups (F_m/Γ_n). We are in particular interested in a certain family of Ramanujan graphs: let $p, q \geq 3$ be distinct prime numbers, and let $j \geq 1$ be an integer. Let $C(p, q)$ be the Cayley graph of the projective general linear group $PGL_2(q)$ over the field of q elements, for a particular set $S_{p,q}$ of $(p+1)$ generators (c.f. [LPS88], [Val97]). It is a $(p+1)$-regular graph on $q^2(q-1)$ vertices with girth (minimal simple cycle length) $\rho(p, q) \sim 4 \log_p q$. Denote $C(p, q, j)$ the graph obtained from $C(p, q)$ by dividing every edge into j edges.

Let $\Gamma(p, q, j)$ be the random labeled graph that is uniformly chosen among all labeled graphs with respect to the generators $X_m^\pm$ having $C(p, q, i)$ as the underlying graph. Let $(G_q(m, p, j))_{q \text{ prime}, q \neq p}$ be the sequence of random groups defined by $G_q(m, p, j) = F_m/\Gamma(p, q, j)$. What we are interested in is an intermediate result in Gromov's construction.

Theorem (M. Gromov, [Gro03]). *For any m, p, there exists j_0 large enough such that for any $j \geq j_0$ a.a.s. the random group $G_q(m, p, j)$ is non-elementary hyperbolic.*

By its construction, an analog of the "density" for the sequence of random groups $(G_q(m, p, j))_{q \text{ prime}, q \neq p}$ is the three parameters m, p, and j. We may ask that, is there a triviality-hyperbolicity phase transition?

Question 7. *Let $(G_q(m,p,j))_{q \text{ prime}, q \neq p}$ be a sequence of random groups defined as above. Does there exist a number $c = c(m,p,j)$ such that*

(i) if $c > 1$, then a.a.s. $G_q(m,p,j)$ is trivial,

(ii) if $c < 1$, then a.a.s. $G_q(m,p,j)$ is non-elementary hyperbolic?

We will give a speculation on the constant c at the end of this section. The same question can be proposed for any phase transition that appeared in the density model: small cancellation conditions, free subgroup problems, the existence of van Kampen diagrams, etc.

Non-reduced graphical small cancellation Let $F_m = F(X_m)$ be the free group generated by the set $X_m = \{x_1, \dots, x_m\}$.

Definition. Let Γ be a graph labeled by X_m, not necessarily reduced. A *piece* of Γ is the labeling word of a simple path of Γ that equals to, in F_m, the labeling word of another simple path of Γ.

Let ρ be the girth (minimal simple cycle length) of Γ. For any $0 < \lambda \leq 1$, we say that Γ satisfies the $G'(\lambda)$ non-reduced graphical small cancellation condition if the length of a piece is at most $\lambda\rho$.

To study the random group $G_q(m,p,j) = F_m/\Gamma(p,q,j)$ defined as above, we may ask that does the randomly labeled graph $\Gamma(p,q,j)$ satisfy the $G'(\lambda)$ condition, asymptotically when q goes to infinity. Recall that the girth of its underlying graph $C(p,q,j)$ is of order $\rho_q = 4j\log_p q + O(1)$ (c.f. [LPS88], [Val97]). Let us estimate the number of simple paths on the graph $C(p,q,j)$ of length $\lambda\rho_q$ with $\lambda < 1/2$: from each vertex of valency $(p+1)$, there are $p^{\lfloor \frac{1}{j}\lambda\rho_q \rfloor}$ such paths (because $C(p,q)$ is p-regular), the number of vertices is of order $q^3 = p^{\frac{3}{4j}\rho_q + O(1)}$. Counting paths starting at vertices of valency 2 will only multiply the number by j and will not affect the exponential term. So the number of simple paths of length $\lambda\rho_q$ is

$$C_q = p^{\frac{1}{j}\left(1+\frac{3}{4\lambda}\right)\lambda\rho_q + O(1)}.$$

The set E_q of non-reduced words of $X_m^{\pm}$ of length $\lambda\rho_q$ is with cardinality $(2m)^{\lambda\rho_q}$. The set A_q of the labeling words produced by the simple paths C_q on $\Gamma(p,q,j)$ is a *random subset* of E_q.

Let us assume that the sequence (A_q) has similar properties of a sequence of permutation invariant random subsets (although it is not). The cardinality of A_q is very close to C_q, so the density of the sequence (A_q) in (E_q) is

$$\operatorname{dens}(A_q) = \frac{1}{j}\left(1 + \frac{3}{4\lambda}\right)\log_{2m} p.$$

Studying the $G'(\lambda)$ condition is to ask if A_q has self-intersection after reduction. If fact, it happens at the same density that A_q intersects the set of words that equal to the identity in F_m, which is with density $\log_{2m}(2\sqrt{2m-1}) < 1$. Now we can ask the following question:

Question 8. *Let $(G_q(m,p,j))_{q \text{ prime}, q \neq p}$ be a sequence of random groups in the graph model. Is it true that*

(i) if $\frac{1}{j}\left(1+\frac{3}{4\lambda}\right)\log_{2m} p + \log_{2m}(2\sqrt{2m-1}) < 1$, then a.a.s. $G_q(m,p,j)$ satisfies $G'(\lambda)$,

(ii) if $\frac{1}{j}\left(1+\frac{3}{4\lambda}\right)\log_{2m} p + \log_{2m}(2\sqrt{2m-1}) > 1$, then a.a.s. $G_q(m,p,j)$ does not satisfy $G'(\lambda)$?

In the Gromov density model, the triviality-hyperbolicity phase transition happens at density $d = 1/2$, which is the critical density for the $C'(\lambda)$ condition with "$\lambda = 1$". We believe that there is an analog in the graph model, and guess that the constant $c = c(m,p,j)$ in Question 7 may be

$$c = \frac{7}{4j}\log_{2m} p + \log_{2m}(2\sqrt{2m-1}).$$

Bibliography

[AŁŚ15] Sylwia Antoniuk, Tomasz Łuczak, and Jacek Świątkowski. “Random triangular groups at density 1/3”. In: *Compositio Mathematica* 151.1 (Jan. 2015). Publisher: London Mathematical Society, pp. 167–178. DOI: 10.1112/S0010437X14007805.

[Arz97] Goulnara N. Arzhantseva. “On groups in which subgroups with a fixed number of generators are free (Russian)”. In: *"Fundamentalnaya i prikladnaya matematika (Fundamental and Applied Mathematics)* 3.3 (1997), pp. 675–683.

[Arz98] Goulnara N. Arzhantseva. “Generic properties of finitely presented groups and Howson’s theorem”. In: *Communications in Algebra* 26.11 (1998), pp. 3783–3792. DOI: 10.1080/00927879808826374.

[Arz00] Goulnara N. Arzhantseva. “A property of subgroups of infinite index in a free group”. In: *Proceedings of the American Mathematical Society* 128.11 (2000), pp. 3205–3210. DOI: 10.1080/00927879808826374.

[Arz06] Goulnara N. Arzhantseva. “A dichotomy for finitely generated subgroups of word hyperbolic groups”. In: *Topological and asymptotic aspects of group theory*. Contemporary Mathematics, 394. American Mathematical Society, Providence, RI, 2006, pp. 1–10. DOI: 10.1090/conm/394.

[AD08] Goulnara N. Arzhantseva and Thomas Delzant. “Examples of random groups”. preprint. 2008.

[AO96] Goulnara N. Arzhantseva and Alexander Yu. Ol’shanskii. “The class of groups all of whose subgroups with lesser number of generators are free is generic”. In: *Mathematical Notes* 59.4 (1996), pp. 350–355. ISSN: 1573-8876. DOI: 10.1007/BF02308683.

[BNW20] Frédérique Bassino, Cyril Nicaud, and Pascal Weil. “Random presentations and random subgroups: a survey”. In: *Complexity and Randomness in Group Theory*. GAGTA Book 1. de Gryuter, 2020. DOI: 10.1515/9783110667028.

[Bow91] Brian H. Bowditch. “Notes on Gromov’s hyperbolicity criterion for path-metric spaces”. In: *Group theory from a geometrical viewpoint*. de Gryuter, 1991. DOI: 10.1142/1235.

[CW15] Danny Calegary and Alden Walker. “Random groups contain surface subgroups”. In: *Journal of the American Mathematical Society* 28.2 (2015), pp. 383–419. DOI: 10.1090/S0894-0347-2014-00802-X.

[Cha91] Christophe Champetier. “Propriétés génériques des groupes de présentation finie”. PhD thesis. Université de Lyon 1, 1991.

[Cha94] Christophe Champetier. “Petite simplification dans les groupes hyperboliques”. In: *Annales de la Faculté des sciences de Toulouse : Mathématiques*. 6th ser. 3.2 (1994), pp. 161–221. DOI: 10.5802/afst.778.

[CH82] Donald J. Collins and Johannes Huebschmann. “Spherical diagrams and identities among relations”. In: *Mathematische Annalen* 261 (1982), pp. 155–183. DOI: 10.1007/BF01456216.

[CDP90] Michel Coornaert, Thomas Delzant, and Athanase Papadopoulos. *Géométrie et théorie des groupes : les groupes hyperboliques de Gromov*. Lecture Notes in Mathematics 1441. Springer Berlin Heidelberg, 1990. DOI: 10.1007/BFb0084913.

[Ghy04] Etienne Ghys. “Groupes aléatoires (d’après Misha Gromov, ...)” In: *Astérisque* 294 (2004), pp. viii, 173–204. DOI: 10.1090/S0894-0347-2014-00802-X.

[Gro87] Mikhael Gromov. “Hyperbolic groups”. In: *Essays in Group Theory*. Mathematical Sciences Research Institute Publications, vol 8. Springer New York, 1987, pp. 75–263. DOI: 10.1007/978-1-4613-9586-7_3.

[Gro93] Mikhael Gromov. “Finitely presented groups”. In: *Asymptotic invariants of infinite groups. Geometric Group Theory*. Vol. 2. London Math. Soc. Lecture Note Ser. 182, 1993, pp. 269–282. DOI: 10.1017/CBO9780511629273.

[Gro00] Mikhael Gromov. “Spaces and Questions”. In: *Visions in Mathematics: GAFA 2000 Special Volume, Part I*. 2000, pp. 118–161. DOI: 10.1007/978-3-0346-0422-2_5.

[Gro03] Mikhael Gromov. “Random walk in random groups”. In: *Geometric and Functional Analysis* 13.1 (2003), pp. 73–146. DOI: 10.1007/s000390300002.

[GM18] Dominik Gruber and John Mackay. “Random triangular Burnside groups”. In: *Israel Journal of Mathematics* (2018). DOI: 10.1007/s11856-021-2170-9.

[Gub86] Victor S. Guba. “Conditions under which 2-generated subgroups in small cancellation groups are free”. In: *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika* 87.7 (1986), pp. 12–19.

[Kam33] Egbert R. van Kampen. “On Some Lemmas in the Theory of Groups”. In: *American Journal of Mathematics* 55.1 (1933), pp. 268–273. DOI: 10.2307/2371129.

[KS08] Ilya Kapovich and Paul Schupp. “On group-theoretic models of randomness and genericity”. In: *Groups, Geometry, and Dynamics* 2.3 (2008), pp. 383–404. ISSN: 1661-7207. DOI: 10.4171/GGD/45.

[KW03] Ilya Kapovich and Richard Weidmann. “Nielsen Methods and Groups Acting on Hyperbolic Spaces”. In: *Geometriae Dedicata* 98 (2003), pp. 95–121. DOI: 10.1023/A:1024064029186.

[LPS88] Alexander Lubotzky, Ralph S. Phillips, and Peter Sarnak. “Ramanujan graphs”. In: *Combinatorica* 8 (1988), pp. 261–277. DOI: 10.1007/BF02126799.

[LS77] Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, 1977.

[Mag30] Wilhelm Magnus. “Ueber diskontinuierliche Gruppen mit einer definierenden Relation (der Freiheitssatz)”. In: *Journal für die reine und angewandte Mathematik* 163 (1930), pp. 141–165.

[MM93] Stuart W. Margolis and John C. Meakin. “Free inverse monoids and graph immersions”. In: *International Journal of Algebra and Computation* 3.1 (1993), pp. 97–100. DOI: 10.1142/S021819679300007X.

[MK22] V. D. Mazurov and E. I. Khukhro, eds. *Unsolved problems in group theory, The Kourovka notebook, No. 20*. 2022. DOI: 10.48550/arXiv.1401.0300.

[Ols91] Alexander Yu. Ol’shanskii. *Geometry of Defining Relations in Groups*. Springer Science+Business Media Dordrecht, 1991. DOI: 10.1007/978-94-011-3618-1.

[Oll04] Yann Ollivier. "Sharp phase transition theorems for hyperbolicity of random groups". In: *Geometric & Functional Analysis* 14.3 (2004), pp. 595–679. DOI: 10.1007/s00039-004-0470-y.

[Oll05] Yann Ollivier. "A January 2005 invitation to random groups". In: *Ensaios Matemáticos* 10 (2005), pp. 1–100. DOI: 10.21711/217504322005/em101.

[Oll07] Yann Ollivier. "Some small cancellation properties of random groups". In: *International Journal of Algebra and Computation* 17.1 (2007). Publisher: World Scientific Publishing Co., pp. 37–51. DOI: 10.1142/S021819670700338X.

[OW11] Yann Ollivier and Daniel T. Wise. "Cubulating random groups at density less than 1/6". In: *Transactions of the American Mathematical Society* 363.9 (2011), pp. 4701–4733. DOI: 10.1090/S0002-9947-2011-05197-4.

[Pap96] Panagiotis Papasoglu. "An algorithm detecting hyperbolicity". In: *Geometric and computational perspectives on infinite groups*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 25. American Mathematics Society, 1996, pp. 193–200. DOI: 10.1090/dimacs/025.

[Sho91] ed. H. Short. "Notes on word hyperbolic groups". In: *Group Theory From a Geometric Viewpoint*. World Scientific Publishing, 1991, pp. 3–63. DOI: 10.1142/1235.

[Sta83] John R. Stallings. "Topology of finite graphs". In: *Inventiones mathematicae* 71.3 (1983), pp. 551–565. DOI: 10.1007/BF02095993.

[Tsa21a] Tsung-Hsuan Tsai. "Density of random subsets and applications to group theory". arXiv: 2104.09192, to appear in the *Journal of Combinatorial Algebra*. 2021.

[Tsa21b] Tsung-Hsuan Tsai. "Freiheitssatz and phase transition for the density model of random groups". preprint arXiv: 2111.08958. 2021.

[Val97] Alain Valette. "Graphes de Ramanujan et applications". In: *Astérisque* 245 (1997). Séminaire Bourbaki, Exposés 820–834, pp. 247–276.

[Żuk03] Andrzej Żuk. "Property (T) and Kazhdan constants for discrete groups". In: *Geometric and Functional Analysis GAFA* 13 (2003), pp. 643–670. DOI: 10.1007/s00039-003-0425-8.

Dans cette thèse, nous étudions les transitions de phase dans les groupes aléatoires à densité. Un groupe aléatoire à densité d est défini par une présentation avec m générateurs et $(2m-1)^{d\ell}$ relations aléatoires, où ℓ est la longueur maximale des relations. Nous avons deux résultats principaux : un sur le problème des sous-groupes libres et l'autre sur l'existence des 2-complexes de van Kampen.

Pour tout entier r entre 1 et $m-1$, nous trouvons une transition de phase à la densité $d_r = \min\{\frac{1}{2}, 1-\log_{2m-1}(2r-1)\}$: Si $d > d_r$, alors les r premiers générateurs engendrent le groupe entier ; si $d < d_r$, alors les r premiers générateurs engendrent un sous-groupe libre. Ce résultat donne de nouveaux exemples de présentations de groupes satisfaisant la propriété de Freiheitssatz, avec une grande variété de longueurs de relations.

Pour chaque 2-complexe d'une forme géométrique donnée, nous donnons une densité critique d_c qui caractérise l'existence d'un 2-complexe de van Kampen dont le 2-complexe sous-jacent est celui donné. Afin de prouver ce résultat, nous étudions en détail la formule d'intersection pour les sous-ensembles aléatoires et donnons une version multidimensionnelle de cette formule.

In this thesis, we study phase transitions in random groups at density. A random group at density d is defined by a presentation with m generators and $(2m-1)^{d\ell}$ random relations, where ℓ is the maximal length of the relations. We have two main results : one on the free subgroup problem and the other on the existence of van Kampen 2-complexes.

For any integer r between 1 and $m-1$, we find a phase transition at the density $d_r = \min\{\frac{1}{2}, 1-\log_{2m-1}(2r-1)\}$: If $d > d_r$, then the r first generators generate the whole group ; if $d < d_r$, then the r first generators generate a free subgroup. This result gives new examples of group presentations satisfying the Freiheitssatz property, with a wide variety of relation lengths.

For each 2-complex of a given geometric form, we give a critical density d_c which characterizes the existence of a van Kampen 2-complex whose underlying 2-complex is the given one. In order to prove this result, we study in detail the intersection formula for random subsets and give a multidimensional version of this formula.

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE
UMR 7501
Université de Strasbourg
CNRS
IRMA, UMR 7501
7 rue René Descartes
F-67000 STRASBOURG

Tél. 03 68 85 01 29

irma.math.unistra.fr
irma@math.unistra.fr

IRMA 2022/001
https://tel.archives-ouvertes.fr/tel-03693833