

Etude des quartiers: défis et pistes de recherche

1

Loïc Bonneval, Fabien Duchateau, Franck Favetta, Aurélien Gentil, Mohamed Nader Jelassi, Maryvonne Miquel, Ludovic Moncla

PLAN

1. Contexte et motivations
2. Pistes de recherche
 1. Description des quartiers
 2. Délimitation des quartiers
 3. Comparaison entre quartiers
 4. Evaluation
3. Conclusion et perspectives

Principaux cas d'études autour des quartiers

- Efficacité énergétique (Perez et al., 2016)
- Gestion du patrimoine (Devernois et al., 2014)
- Étude d'impact de l'environnement sur la santé (Morland et al., 2002)
- Reconstitution numérique du patrimoine (Pardoen, 2015)

Perez, N., A. Mailhac, C. Inard, et P. Riederer (2016). Outil d'aide à la décision multicritère pour la conception de systèmes énergétiques à l'échelle du quartier.

In IBPSA France

Devernois, N., S. Muller, et G. Le Bihan (2014). *Gestion du patrimoine urbain et revitalisation des quartiers anciens : l'éclairage de l'expérience française*.

Agence française de développement (AFD).

Morland, K., S. Wing, A. D. Roux, et C. Poole (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American journal of preventive medicine*

Pardoen, M. (2015). Les oreilles à l'affût! restitution d'un paysage sonore : œuvre de l'imaginaire ou recherche d'authenticité. *Silences et bruits du Moyen Âge à nos jours : Perceptions,*

identités sonores et patrimonialisation. L'Harmattan

Contexte



Home in Love (HiL) est un ensemble de processus destinés à accompagner les personnes en mobilité géographique afin de leur trouver un logement dans une ville qu'ils ne connaissent pas

Projet pluridisciplinaire (informaticiens, sociologues,...)

Objectif de HiL: recommandation de quartiers/logements

- **pour étudier ces quartiers, il est alors important de bien les définir**

Nos pistes de recherche

- 1) **Description:** le concept de quartier est souvent flou et sujet à diverses interprétations
- 2) **Délimitation:** l'information n'est pas souvent disponible
- 3) **Comparaison:** quels algorithmes et indicateurs sont pertinents ?
- 4) **Evaluation:** quelle méthodologie/jeux de données pour valider ?



6

Pistes de recherche

Description de quartiers

Liste des données

- Quelles informations pertinentes permettent la description des quartiers ?
- Points de vue différents entre divers acteurs pour la description (informaticiens, sociologues,...)

Problème 1 : Liste des données utiles pour un domaine d'application donné

Données de cadrage		Nombre de communes	Population en 2005 ⁽¹⁾ (habitants)	Population moyenne par commune en 2005	Évolution de population 1999-2005 (%)	Densité de population en 2005 (hab./km ²)	Capacité d'accueil en 1998 ⁽²⁾ (nbre de lits)	Densité touristique en 1998 (lits/km ²)	Capacité d'accueil moyenne par commune en 1998
Manche - mer du Nord	Communes littorales	264	933 445	3 536	-0,9	347	809 245	301	3 065
	Arrière-pays	619	445 824	720	3,4	93	124 739	26	202
	Cantons littoraux	883	1 379 269	1 562	0,4	184	933 984	124	1 058
Atlantique	Communes littorales	405	1 930 209	4 766	4,7	189	3 085 170	301	7 618
	Arrière-pays	349	558 821	1 601	10,3	74	204 476	27	586
	Cantons littoraux	754	2 489 030	3 301	5,9	140	3 289 646	185	4 363
Méditerranée	Communes littorales	119	2 803 990	23 563	4,3	554	2 754 007	544	23 143
	Arrière-pays	89	302 437	3 398	9,0	162	109 737	58	1 233
	Cantons littoraux	208	3 106 427	14 935	4,7	448	2 863 744	413	13 768
Littoral (hors Corse)	Communes littorales	788	5 667 644	7 192	3,5	315	6 648 422	370	8 437
	Arrière-pays	1 057	1 307 082	1 237	7,5	92	438 952	31	415
	Cantons littoraux	1 845	6 974 726	3 780	4,3	217	7 087 374	220	3 841
Métropole		36 570	59 418 631	1 625	3,8	109	18 467 552	34	505
Corse	Communes littorales	97					353 301	97	3 642
	Arrière-pays	124					47 550	22	383
	Cantons littoraux	221					400 851	69	1 814

⁽¹⁾ Source : Insee - estimations de population supracommunales.

⁽²⁾ Source : Insee, recensement de la population (RP), 1999, Enquête tourisme 1998 (hôtellerie et hébergement de plein air), Inventaire communal 1998 (autres hébergements).

Recherche de sources de données

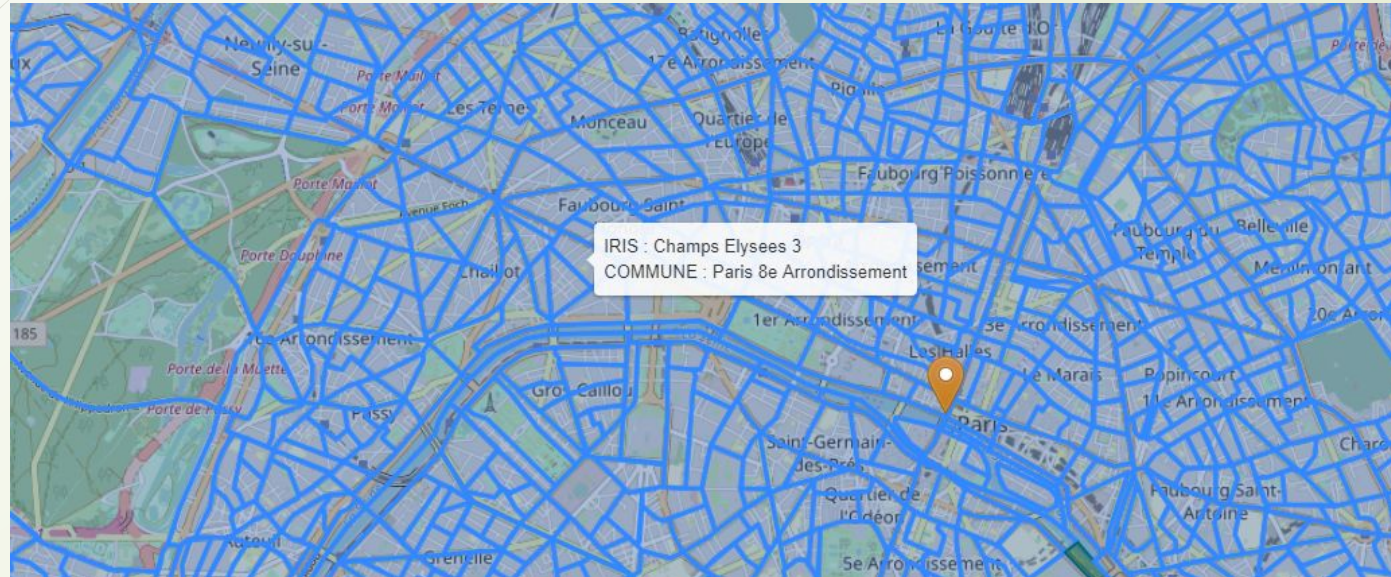
Problème 2 : Recherche de sources de données disponibles

- ❑ Recherche facilitée par l'*open data*
- ❑ Nécessite la vérification de la **qualité** des données (date de mise à jour, etc.)
- ❑ Certaines données sont **rare**s ou inexistantes (comme le prix)
- ❑ Certaines API sont **indisponibles** ou limitées en utilisation



Recherche de sources de données: INSEE

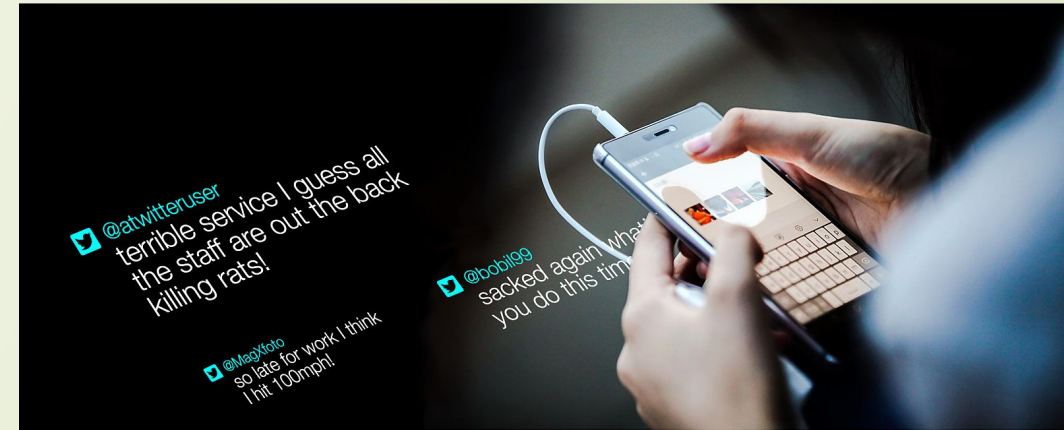
9



- ❑ L'INSEE quadrille le territoire français en 50 000 IRIS
- ❑ **IRIS** = Ilots regroupés pour l'information statistique
- ❑ + de 1000 indicateurs décrivent les iris (nombre d'épiceries, répartition par CSP,...)
- ❑ Certains indicateurs peuvent être agrégés (Actif et al., 2013)(exemple: commerces de proximité = nombre de boulangeries + nombre d'épiceries)
- ❑ Limitées à la France

Réseaux sociaux

- La collecte de données des **réseaux sociaux** qui peuvent être **localisées**:
 - Tweets
 - Likes
 - *Checkins*, ...



- Kennedy et Naaman (2008)

Intégration des données

Problème 3 : L'hétérogénéité des données



- ❑ Les indicateurs de l'INSEE sont fournis dans **plusieurs** fichiers
- ❑ Interprétation **différente** des concepts (exemple: LIBIRIS ou LIB_IRIS selon la source)
- ❑ **Hétérogénéité** des concepts
- ❑ Evolution de la granularité spatiale (exemple: regroupement ou division des IRIS)

- ❑ **Solutions possibles :**
 - ❑ Manuelle: Saisie dans un tableur
 - ❑ Automatique: Outils existants (Open Refine, Talend,...)

Pistes de recherche

Délimitation des quartiers

Délimitation des quartiers (1)

- C'est quoi ? Liste de coordonnées géographiques formant un polygone
- Comment elle se définit ?
 - Se base sur les définitions de l'administration
 - Exploite des cartes ou des systèmes d'information géographique (SIG)
 - Regroupement/Divisions de quartiers
 - Exploite des textes, des images (aériennes)
 - Exploite des likes, checkins des réseaux sociaux
 - Réalise des enquêtes auprès des populations

Définition Administrative

- L'administration met à disposition le **contour géographique** mais uniquement pour certaines zones (exemple: Nanterre)
- La méthodologie de découpage n'est pas toujours accessible
- Prise en compte uniquement du point de vue administratif

Systemes d'information géographique (SIG)

□ Google maps

- Noms de quartiers avec leur contour, méthodologie non décrite

□ Bing et Here maps

- Un seul point pour représenter un quartier (insuffisant)

□ Wikimapia

- Découpage en quartiers décrits par leurs contours
- Site collaboratif □ découpage subjectif et incomplet

Regroupements/Divisions de quartiers

□ (Barret et al., 2019)

- Prendre en compte le voisinage (IRIS adjacents à un iris) pour lisser les données
- Se rapprocher au mieux de la notion de quartier

□ (Actif et al., 2013)

- De «grands quartiers» créés à partir d'IRIS
- Validité des données ? Possibilité de biais dans les analyses



Sources textuelles



□ (Brindley et al., 2014)

- S'appuient sur des sources de données textuelles
- Exemple de source: *patrimoine-lyon.org*
- Extraction de noms de quartiers à partir d'un grand volume de documents

□ (Guérois et Madelin, 2017)

- S'appuient sur des annonces immobilières (comme airbnb)
- Descriptions détaillées sur les quartiers du point de vue expert ou habitant



Brindley, P., J. Goulding, et M. L. Wilson (2014). A data driven approach to mapping urban neighbourhoods. In *SIGSPATIAL*

Guérois, M. et M. Madelin (2017). Comment les hôtes et clients d'Airbnb parlent-ils des lieux ? Une analyse exploratoire à partir du cas parisien. In *EXCES*

Images aériennes et satellites

- **(Kang et al., 2018) et (Huang et al., 2015)**
 - S'appuient sur les images aériennes et satellites
 - Exploitent les images street view pour détecter des quartiers (Analyse de bâtiments, objets, véhicules,...)

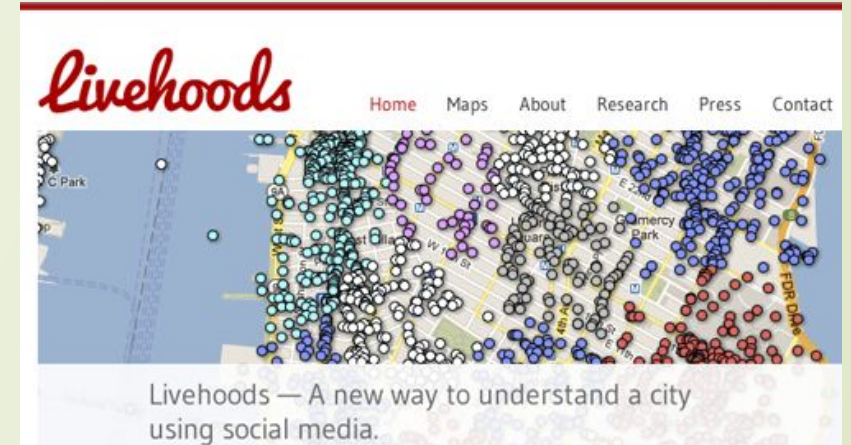


Kang, J., M. Körner, Y. Wang, H. Taubenböck, et X. X. Zhu (2018). Building instance classification using street view images. *ISPRS*

Huang, X., H. Liu, et L. Zhang (2015). Spatiotemporal detection and analysis of urban villages in mega city regions of china using high-resolution remotely sensed imagery. *IEEE*

La piste des réseaux sociaux

- Exploiter les données de géolocalisation issus des réseaux sociaux
- **LiveHoods** (*livehoods.org*)
 - Délimitation des quartiers grâce aux checkins
 - (Cranshaw et al., 2012)
- **Hoodsquare**
 - Analyse d'activités des habitants
 - (Zhang et al., 2013)



Cranshaw, J., R. Schwartz, J. Hong, et N. Sadeh (2012). The livehoods project : Utilizing social media to understand the dynamics of a city
Zhang, A. X., A. Noulas, S. Scellato, et C. Mascolo (2013). Hoodsquare : Modeling and recommending neighborhoods in location-based social networks. In *Social Computing*

Enquête auprès des populations

□ (Kumar et al., 2015)

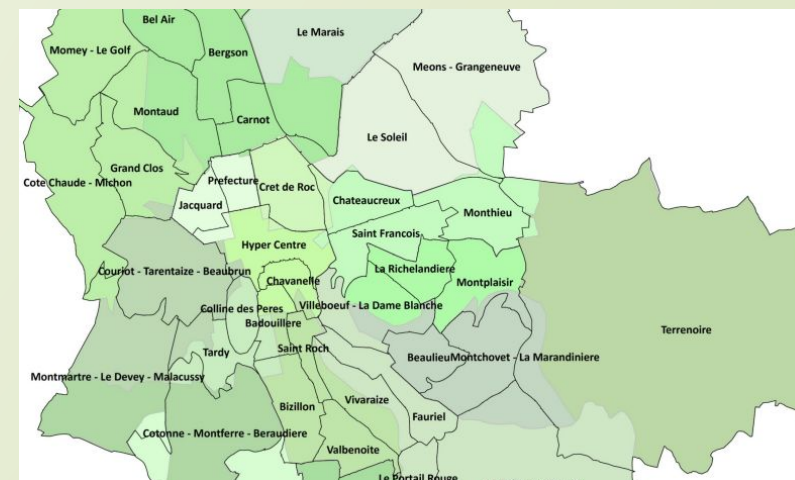
- Proposent aux utilisateurs de dessiner sur une carte interactive leur quartier

□ (Berjawi et al., 2013)

- Proposent aux utilisateurs de décrire les points d'intérêt de leur quartier afin d'en déduire les frontières

Kumar, C., W. Heuten, et S. Boll (2015). Visual overlay on openstreetmap data to support spatial exploration of urban environments. ISPRS

Berjawi, B., M. Colomb, T. Joliveau, F. Favetta, F. Duchateau, et M. Miquel (2013). Outil de repérage urbain à travers la prise de points de repère. Prototype, Laboratoires EVS et LIRIS.



Pistes de recherche

Comparaison de quartiers

Motivation

- ✓ Plusieurs cas d'application (recherche immobilière, étude sociologique des quartiers,...) ont besoin d'établir la ressemblance (similarité) entre quartiers
- ✓ En exploitant les descriptions des quartiers

Mesures de similarité

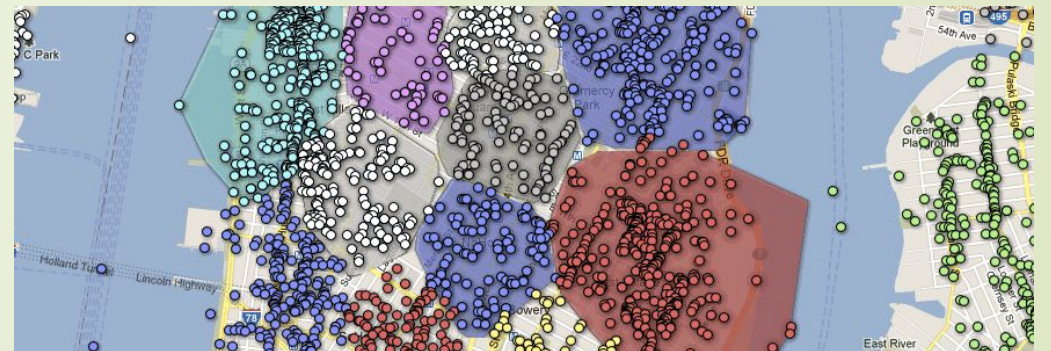
- ❑ Mesurer le degré de ressemblance entre 2 quartiers
- ❑ Un quartier = Vecteur de valeurs (le caractérisant)
- ❑ Comparaison d'objets: la mesure *cosine*, *jaccard*

Yu, M., G. Li, D. Deng, et J. Feng (2016). String similarity search and join : a survey

Zhang, A. X., A. Noulas, S. Scellato, et C. Mascolo (2013). Hoodsquare : Modeling and recommending neighborhoods in location-based social networks. In *Social Computing*,

Clustering et partitionnement

- Objectif: Regrouper les quartiers les plus similaires par:
 - **Partitionnement**: *KMeans*, *Affinity propagation*, ...
 - **Clustering**: *DBSCAN*, ...
- Exemple: *Livehoods* se base sur l'algorithme *spectral clustering* pour comparer des quartiers



Case-Based Reasoning

□ (Yuan et al., 2013)

- Détection de cas similaires pour rapprocher des quartiers

□ **Exemple:**

- Une personne (marié, 2 enfants) est à la recherche d'une résidence
- Sa situation est analysée pour lui proposer des quartiers où vivent des résidents dans une même situation



26

Pistes de recherche

Evaluation

Motivation

□ Pourquoi ?

- Certains processus comme la délimitation et la comparaison de quartiers ont besoin d'être vérifiés et validés
- Automatisation de la vérification ?

Pistes

- **Enquêtes** ou questionnaires auprès des habitants
 - **Comparaisons** entre les résultats des enquêtes et les résultats des algorithmes
- **Evaluation** automatique: Evaluer la comparaison de quartiers revient à disposer d'un jeu de données manuellement rempli qui liste les quartiers **similaires**
 - **(Kittur et al., 2008)**: Jeux de données annotés ou vérifiés par des personnes rémunérées
 - **(Smarzaro et al., 2017)**: Vérifier si les fournisseurs cartographiques (Facebook, Yelp,...) confirment les statistiques expertisées

Conclusion et perspectives

Conclusion



Objectif: Recommander le quartier idéal lors d'une recherche immobilière

- ❖ Description: à partir des données IRIS
 - ❖ Délimitation: prise en compte du voisinage d'un IRIS partageant des descriptions similaires
 - ❖ Comparaison: Recommandation de quartiers (algs déjà testés: *cosine*, *DBSCAN*,...)
 - ❖ Evaluation: peu de données disponibles (~150 utilisateurs)
- Utilisables dans le regroupement, classement et recommandation des quartiers

Limites

- Pas de définition claire d'un quartier
- Il faut caractériser l'utilisation de chaque algorithme de clustering (avec des sociologues ?)
- Impossible de vérifier ou de justifier les résultats des algorithmes

Perspectives

- Proposition d'une ontologie du domaine immobilier
- Système de prédiction des likes/dislikes des utilisateurs sur les annonces immobilières
- Définition de profils types d'utilisateurs (en mutation, premier job, alternants,...)
- Système personnalisé de recommandation de logements