



HAL
open science

Un système de recommandation de lieux basé sur la mesure de Katz dans les réseaux sociaux géographiques

Romain Picot-Clémente, Cécile Bothorel

► To cite this version:

Romain Picot-Clémente, Cécile Bothorel. Un système de recommandation de lieux basé sur la mesure de Katz dans les réseaux sociaux géographiques. MARAMI 2013 : 4^{ème} conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatiques, Oct 2013, Saint-Etienne, France. hal-00960139

HAL Id: hal-00960139

<https://hal.science/hal-00960139>

Submitted on 17 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Un système de recommandation de lieux basé sur la mesure de Katz dans les réseaux sociaux géographiques

Romain Picot-Clément¹, Cécile Bothorel¹

1. Lab-STICC, département LUSSI
Télécom Bretagne, Institut Mines-Télécom
UMR CNRS 3192 Lab-STICC, département LUSSI
Brest, France
{romain.picotclemente, cecile.bothorel}@telecom-bretagne.eu

RESUME. Cet article s'intéresse à la recommandation de lieux dans les réseaux sociaux géographiques. Nous proposons une méthode qui combine dans un même graphe à la fois le graphe social entre utilisateurs, le graphe bipartite de fréquentation de lieux par les utilisateurs, et un graphe géographique de distances entre les lieux. Une propagation de poids suivant la mesure de Katz est réalisée au sein du graphe global pour finalement proposer des lieux potentiellement intéressants à chaque utilisateur. Notre méthode est comparée à des méthodes de la littérature sur le jeu de données Gowalla. Nos résultats confirment le réel intérêt de considérer, en plus des fréquentations, les données sociales et les données géographiques pour de la recommandation de lieux. En général, notre méthode surpasse significativement les méthodes comparées, mais dans certaines conditions que nous analysons, nous montrons qu'elle donne parfois des résultats mitigés.

ABSTRACT. This article proposes a recommender system combining the social graph, the frequentation graph and a geographic graph in a unique graph to recommend places in LBSNs (Location-Based Social Networks). A propagation of weights by the Katz centrality method is performed on the global graph in order to induce affinity measures between users and places. Finally, potentially interesting places are deduced for each user. The proposed method is compared to popular methods from the literature on the dataset Gowalla. As in some of the few works in the domain, our results show a real interest in considering both social data and geographic data, beyond the use of frequentation data for recommending places. Generally, our method outperforms significantly the compared methods, but sometimes, it can give worse results. Then, we try to highlight the possible factors that could induce them.

MOTS-CLES : Systèmes de recommandation ; recommandation de lieux ; réseaux sociaux géographiques ; m-commerce social.

KEYWORDS : Recommender systems ; places recommendation ; location-based social networks ; Social Shopping.

4^{ème} conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatiques

1. Introduction

Avec le développement des réseaux sociaux et l'utilisation répandue des Smartphones, les utilisateurs partagent de plus en plus de contenus en mobilité. Avec des services comme Foursquare¹, le partage de lieux est même la finalité. Dans ce papier, nous nous intéressons à la recommandation de lieux dans les réseaux sociaux géographiques où les utilisateurs sont connectés aux autres utilisateurs par des relations d'amitié et où les lieux qu'ils ont visités sont connus. En posant le problème ainsi, nous nous positionnons potentiellement sur des services de m-commerce social où le mobinaute se verra conseiller des magasins ou des promotions géolocalisées. Nous proposons un nouvel algorithme considérant de manière conjointe les trois dimensions des réseaux sociaux géographiques : les relations sociales, les fréquentations (check-in), et les coordonnées géographiques des lieux.

Ce travail cherche en premier lieu à montrer que ces dimensions, si elles sont dissociées, ont un pouvoir prédictif moins fort sur la précision des recommandations que si elles sont combinées. Notre seconde contribution réside dans la méthode de combinaison elle-même qui intègre différents graphes et offre ainsi une structure pour propager une influence à travers les différentes relations reliant les différents types de noeud. Cette méthode propose en effet de générer un graphe unique qui combine le graphe social, le graphe de fréquentation et un graphe géographique. Ce dernier est construit, comme nous le montrerons, en considérant la densité de probabilités qu'un check-in soit fait en fonction de sa distance aux autres check-ins. Ensuite, l'algorithme effectue une propagation de poids par la méthode de centralité de Katz à travers le graphe fusionné. La méthode de centralité de Katz, décrite dans (Katz, 1953), était originellement utilisée dans les sciences sociales pour mesurer le degré d'influence d'un acteur dans un réseau social en propageant des poids sur chaque chemin du réseau partant de l'acteur, et en mesurant les poids qui émergent. Dans notre projet, le but de la propagation est de mettre en évidence les lieux qui seraient susceptibles d'intéresser les utilisateurs de par leur position en « zone d'influence » sur le graphe fusionné. Pour tester cet algorithme, un jeu de données issu du réseau social géographique Gowalla est utilisé.

La première partie de cet article présente des travaux qui se sont penchés sur la recommandation sociale, et plus spécifiquement la recommandation sociale géographique. La deuxième partie décrit la méthode que nous proposons pour trouver de nouveaux lieux à recommander aux utilisateurs en se basant sur un graphe fusionné et la méthode de Katz. La partie suivante détaillera quelques méthodes de la littérature que nous avons adaptées à la recommandation de lieux. Elles considèrent un ou plusieurs aspects du système : graphe social, graphe de fréquentation et/ou coordonnées géographiques. Finalement, dans la dernière partie, les résultats de ces méthodes sur le jeu de données Gowalla² seront comparés aux résultats de notre méthode et discutés.

1. <https://fr.foursquare.com/>

2. Disponible sur <http://snap.stanford.edu/data/#locnet>

2. Travaux connexes de la littérature

Dans le domaine des systèmes de recommandation, il existe deux principales familles de techniques pour réaliser des recommandations (Resnick & Varian, 1997): une famille de techniques basées sur le filtrage collaboratif qui recherche des similarités de profils utilisateurs en se basant sur des notations (un nombre d'étoiles, la liste des achats passés, les lieux visités, etc.) et une famille de techniques qui se base sur des similarités de profils de contenus sur la base de descripteurs. Le papier (Candillier, et al., 2009) compare ces méthodes et affirme que les méthodes par filtrage collaboratif donnent généralement de meilleurs résultats que le filtrage de contenu, tout du moins lorsqu'il y a suffisamment de notations disponibles. Cependant, lorsqu'un nouvel utilisateur ou un nouveau contenu survient dans le système, les méthodes par filtrage collaboratif ont des difficultés à fournir des recommandations efficaces aux nouveaux utilisateurs ou à recommander les nouveaux contenus, parce qu'aucun historique n'est disponible sur eux. Ce problème, appelé problème de démarrage à froid, est surmonté par les méthodes de filtrage de contenu. Cependant, ces techniques nécessitent des méta-données fiables : il s'avère que les profils d'internautes ou les fiches descriptives de lieux ne le sont pas, les préférences déclaratives quand elles existent sont souvent inconsistantes ; la collecte d'informations pertinentes de description de lieux pour attirer les visiteurs peut s'avérer coûteuse et complexe, voire impossible si l'on prend des commerces.

Dans cet article, nous considérons le pire cas où il n'y a pas de descripteurs sur les utilisateurs ni sur les lieux, hormis leurs coordonnées GPS. Ainsi, nous rejetons les méthodes basées sur le contenu et nous focalisons sur les méthodes à filtrage collaboratif basées sur les actions spontanées des utilisateurs, comme leur notation sur les items, et dans notre cas, comme les check-ins dans les réseaux sociaux géographiques. Parmi ces actions spontanées, encore souvent sous-exploitées en recommandation, figurent aussi les actions sociales, comme le tissage de relations ou les interactions, sur les plateformes qui le permettent.

Quelques travaux ont pris en compte le graphe social en plus des relations utilisateurs-items pour les recommandations. (Bothorel, 2011) montre que les données sociales apportent bien une amélioration des recommandations, en particulier dans un contexte de démarrage à froid, où peu d'information d'historique sont disponibles. Dans cette étude, un clustering sur le graphe social semble capturer la détection de communautés de goûts cinématographiques puisqu'il améliore la recommandation de contenus peu populaires.

(Ma, et al., 2011) proposent une méthode pour factoriser la matrice de notation utilisateurs-items par la méthode de décomposition en valeurs singulières (SVD) en minimisant une fonction objectif. Si un terme appelé régularisation sociale est ajouté à la fonction, terme basé sur les amis directs de l'utilisateur dans le réseau social, la précision des recommandations est améliorée.

Dans (Symeonidis, et al., 2011), les auteurs proposent de combiner les matrices de similarités entre utilisateurs, dérivées des réseaux sociaux implicites et explicites.

Ils calculent deux matrices de similarités, une basée sur le réseau d'amis et une basée sur le réseau bipartite (utilisateur-item). Ces deux matrices sont combinées en une seule matrice de similarité par une somme pondérée. Ils généralisent ensuite ce modèle pour incorporer plus de graphes. Leur algorithme donne de meilleures recommandations que les méthodes traditionnelles de filtrage collaboratif par voisinage. Notre méthode sera comparée à celle-ci dans la partie 4.

Dans (Vasuki, et al., 2010), les auteurs proposent un système de recommandation de groupes basé sur le réseau social d'amis, et basé sur le réseau de groupes reliant des utilisateurs à des groupes. Ils décrivent tout d'abord une manière de combiner le réseau social d'amis et le réseau de groupes en un seul graphe. Puis, ils proposent deux méthodes pour recommander des groupes : une basée sur la proximité dans le graphe en utilisant la mesure de Katz et une méthode modélisant les utilisateurs et les groupes selon des facteurs latents. Ces deux méthodes donnent de bons résultats, mais la méthode par mesure de Katz est la plus efficace en termes de temps de calcul et de qualité des recommandations. L'algorithme que nous proposons dans ce papier s'inspire en partie de leur méthode basée sur la mesure de Katz, mais nous considérons des lieux à la place des groupes et notre graphe combiné contient plus d'informations. Nous comparons notre méthode à la leur dans la section 4.

Dans notre projet, nous nous intéressons plus particulièrement à de la recommandation de lieux. Très peu de travaux se sont intéressés à la recommandation sociale géographique. Nous pouvons néanmoins citer quelques travaux récents représentatifs du domaine.

(Noulas, et al., 2012) s'intéressent à la recommandation de lieux dans les réseaux sociaux géographiques. Ils proposent un algorithme « Random Walk With Restart » dans un graphe où les nœuds sont les utilisateurs et les lieux. Les utilisateurs sont reliés aux autres utilisateurs si une relation d'amitié existe, et les utilisateurs sont reliés aux lieux en fonction des check-ins. Leur algorithme donne de meilleurs résultats que des algorithmes de recommandation traditionnels. Cependant, cet algorithme ne prend pas en compte un élément majeur des réseaux sociaux géographiques qui est la position géographique des différents lieux du système. Le papier (Ye, et al., 2011) en tient compte, il propose une méthode qui considère les aspects géographiques, sociaux et de fréquentation dans les réseaux sociaux géographiques. Il unifie trois scores de prédiction pour chaque lieu, basés sur : la similarité entre les utilisateurs selon leurs check-in ; la similarité entre les utilisateurs selon le réseau social d'amis ; les coordonnées géographiques. Pour les scores basés sur les informations géographiques, ils se basent sur la théorie naïve Bayésienne pour prédire le score de probabilité de check-in du lieu donné, en calculant le produit des probabilités de chaque distance entre le lieu considéré et les lieux visités selon la loi de distribution. Les trois scores de prédiction sont unifiés en une somme pondérée. Ils comparent leur méthode à de multiples autres méthodes et montrent que le fait de tenir compte des données sociales et des données géographiques, avec les données de fréquentation, améliore la recommandation de lieux. Cette méthode sera dénommée (F+S+G) dans la suite et sera comparée à la nôtre dans la section 4.3.

Les résultats donnés par ce dernier algorithme sont intéressants pour améliorer la recommandation dans les réseaux sociaux géographiques mais nous pensons qu'il est possible de les améliorer par la recherche de corrélations plus poussées entre les lieux et les utilisateurs. Ainsi, la méthode que nous proposons s'inspire de l'algorithme original décrit dans (Vasuki, et al., 2010) pour la propagation de poids par la méthode de Katz. Elle diverge du fait que le graphe unifié que nous considérons est composé de trois graphes : le graphe social, le graphe de fréquentation et un graphe géographique. La création de ces trois graphes est présentée dans la partie suivante. Elle insiste notamment sur le graphe géographique qui est construit suivant le comportement géographique de check-in des utilisateurs.

3. KatzFSG: Un algorithme basé sur la mesure de Katz et considérant les informations de Fréquentations, Sociales et Géographiques

Cette section décrit notre méthode KatzFSG. Elle permet de recommander des lieux aux utilisateurs selon leurs check-ins et leur réseau social, en considérant le contexte géographique des lieux visités. Elle est composée de trois parties. La première se focalise sur la définition des différents graphes. La deuxième partie insiste sur la création du graphe géographique qui est un des principaux aspects de la méthode. La troisième partie décrit la méthode de fusion de graphes et comment celui-ci est utilisé pour induire des recommandations à l'aide d'une propagation de poids utilisant la méthode de centralité de Katz.

3.1. Définitions des trois graphes

Le graphe social \mathbf{S} est graphe d'adjacence où les sommets représentent les utilisateurs et les arêtes les relations d'amitié. Dans ce travail, \mathbf{S} est représenté par une matrice S , $N \times N$ symétrique (N étant le nombre d'utilisateurs), où S_{ij} vaut 1 si une relation d'amitié existe entre l'utilisateur u_i et l'utilisateur u_j , et vaut 0 sinon.

La création du graphe de fréquentation \mathbf{F} se base sur les check-ins des utilisateurs dans les différents lieux. Dans ce graphe bipartite, les nœuds sont soit des utilisateurs, soit des lieux. Les arêtes relient les utilisateurs avec les lieux dans lesquels ils ont fait un ou plusieurs check-in(s), elles sont pondérées en fonction du nombre de visites. La matrice F , qui représente ce graphe, est une matrice $N \times M$ (M étant le nombre de lieux), où F_{ik} est le nombre de fois que l'utilisateur u_i a fréquenté le lieu l_k .

Le graphe géographique \mathbf{G} relie les lieux entre eux. La matrice G , qui représente ce graphe, est une matrice $M \times M$. Notre proposition pour la création de ce graphe (cette matrice) est décrite dans la partie suivante.

3.2. Un Graphe Géographique basé sur le comportement de check-in

Nous proposons de construire le graphe en considérant le comportement de check-in de chaque utilisateur dans l'espace géographique. Il apparaît que le comportement de check-in des utilisateurs est fortement influencé par la proximité

entre les lieux, comme vu dans (Ye, et al., 2011). Dans ce dernier, les auteurs construisent la densité de probabilité des check-ins suivant leur distance à un autre check-in et approximent les points obtenus par une courbe définie par la fonction $f(x) = ax^b$ (fonction loi de puissance). f permet d'inférer la probabilité $Pr(d(l_i, l_j))$ de check-in en un lieu pour toute distance avec un autre check-in de l'utilisateur. Ils utilisent ensuite f pour calculer la probabilité qu'un check-in soit fait en un lieu suivant ses distances avec tous les check-ins effectués, par méthode naïve bayésienne (produit des probabilités liés à chaque distance avec les check-ins). Dans notre algorithme, nous cherchons à obtenir f de la même façon mais nous ne sommes pas intéressés par le calcul de cette probabilité.

La distribution de probabilité représentée par f permet de relier chaque paire de lieux par une probabilité de check-in suivant leur distance mutuelle. Ainsi, pour chaque utilisateur, il est possible de créer un graphe géographique G_u dont les sommets sont les lieux et dont les arêtes sont pondérées par la probabilité de check-in calculée sur la distance mutuelle des lieux considérés.

Un tel graphe géographique G_u existe pour chaque utilisateur. Néanmoins, il apparaît que ces graphes ne varient que très peu d'un utilisateur à un autre. De plus, le graphe réalisé pour un utilisateur avec très peu de check-ins n'est pas vraiment pertinent car il n'y a pas assez de check-ins pour trouver la fonction de distribution réelle. Ainsi, nous proposons de générer un seul graphe géographique G reliant les lieux entre eux par des probabilités de check-ins, en se basant sur la distribution de probabilité de l'ensemble des utilisateurs, et non plus sur un seul utilisateur. La matrice géographique G est construite telle que :

$$G_{ij} = f(d(l_i, l_j), a, b) = a \times d(l_i, l_j)^b \quad (1)$$

3.3. La mesure de Katz sur le graphe fusionné

Dans notre méthode, nous proposons de fusionner les graphes S , F et G en un unique graphe C . Ainsi, dans ce graphe, les nœuds sont soit des lieux soit des utilisateurs, et les arêtes peuvent relier des utilisateurs entre eux, des lieux entre eux ou des utilisateurs avec des lieux. La matrice C représente ce graphe unifié et est construite comme suit :

$$C = \begin{pmatrix} \alpha S & \lambda F \\ \lambda F^T & \gamma G \end{pmatrix} \quad (2)$$

Les coefficients α, λ, γ sont respectivement les degrés d'influence des matrices S, F, G dans la matrice C , avec $\alpha + \lambda + \gamma = 1$. Il est aussi à noter que F et G sont normalisés avant d'intégrer C .

Notre méthode consiste ensuite à propager un poids dans le graphe C par mesure de Katz comme suit :

$$Katz(C) = \beta C + \beta^2 C^2 + \beta^3 C^3 + \dots, 0 \leq \beta \leq 1 \quad (3)$$

β est le poids qui est propagé à travers le graphe. Par cet algorithme, nous nous intéressons plus particulièrement à l'effet de la propagation de poids sur les relations

utilisateurs-lieux. Ces relations sont représentées par le bloc $Katz(C)_{12}$ dans la matrice $Katz(C)$. Etant donné le coût important de ce calcul, une matrice de Katz tronquée est calculée telle que :

$$tKatz(C, k)_{12} = \sum_{i=1}^k (\beta^i C^i)_{12}, k \text{ étant le rang maximal de calcul} \quad (4)$$

Une estimation du coût de calcul est alors $O(N \times nnz)$, où N est le nombre d'utilisateurs et nnz est le nombre de valeurs non nulles dans $(FF^T)^k$. En pratique, et comme dans (Vasuki, et al., 2010), nous choisissons d'arrêter le calcul au troisième rang $tKatz(C, 3)_{12}$.

Finalement, pour chaque utilisateur u , les n lieux non visités ayant les meilleurs poids sur la ligne de l'utilisateur dans la matrice $tKatz(C, k)_{12}$ sont sélectionnés. Ces lieux sont alors les recommandations à proposer à l'utilisateur correspondant.

4. Expérimentation

4.1. Jeu de données

Afin de tester notre algorithme, nous utilisons un jeu de données Gowalla. Il s'agit d'un jeu de données contenant 196591 utilisateurs, 950327 relations d'amitié et 6442890 check-ins de ces utilisateurs dans 1279228 lieux différents. La période de check-ins s'étend de Février 2009 à Octobre 2010. Gowalla était un site de réseautage social géolocalisé qui a fermé en Mars 2012 après avoir été racheté par Facebook en fin 2011.

Etant donné les temps de calcul importants demandés par notre algorithme sur une telle volumétrie de données, nous proposons de ne se baser que sur certaines zones géographiques restreintes pour le tester. Nous nous restreignons à des zones d'une taille maximale de 500×500 kilomètres. Arbitrairement, les zones sélectionnées sont : San Francisco, Chicago, Irlande, Sud Louisiane, Paris.

Le jeu de données est découpé en périodes de temps de 1 semaine de 12/2009 à 10/2010. Chaque semaine est utilisée comme jeu de données training et la semaine suivante est utilisée comme jeu de données test. Des mesures de rappel et de précision sont effectuées pour mesurer la qualité. Elles se basent sur les recommandations calculées sur la période training comparées aux check-ins réels de la période test. Les mesures sont moyennées sur l'ensemble des périodes pour avoir une valeur moyenne du rappel et de la précision sur une période d'une semaine. Ceci est effectué pour chaque zone géographique sélectionnée.

4.2. Méthodes de comparaison

Cette partie énonce les différentes méthodes qui seront comparées à celle que nous proposons. Elles sont tirées de la littérature que nous avons abordée précédemment :

- **Filtrage collaboratif basé sur les fréquentations (méthode F)** : filtrage collaboratif par voisinage utilisateur basé sur la similarité de Pearson sur la matrice F. La méthode est décrite dans (Breese, et al., 1998).
- **Filtrage collaboratif basé sur les relations sociales (méthode S)** : filtrage collaboratif basé sur les amis (friend-based collaborative filtering) selon (Ma, et al., 2009). Pour les paires d'amis, il s'agit de calculer un score de similarité en faisant intervenir à la fois la matrice sociale et la matrice de fréquentation.
- **Filtrage collaboratif basé sur la fusion des similarités sociales et de fréquentations (méthode FuseFS)** : décrit dans (Symeonidis, et al., 2011). La méthode est très proche de la précédente mais réalise le calcul de similarité de Pearson sur l'ensemble des utilisateurs et ne se limite pas qu'aux utilisateurs qui sont amis.
- **Algorithme de Katz sur fréquentation et social (méthode KatzFS)** : proposé dans (Vasuki, et al., 2010). Il est utilisé originellement pour de la recommandation de groupes mais nous l'adaptions ici pour de la recommandation de lieux. Le graphe social et le graphe de fréquentation sont fusionnés en un seul graphe afin d'y propager des poids pour mettre en évidence des relations qui ne sont pas évidentes au premier abord. Cet algorithme diffère de la méthode que nous proposons du fait qu'il n'intègre pas la matrice géographique.
- **Méthode géographique (méthode G)** : dénommée recommandation par influence géographique dans (Ye, et al., 2011). L'idée est de ne prendre en compte que les coordonnées géographiques des lieux visités pour en déduire pour chaque utilisateur les lieux non visités qui se retrouvent à des distances d'intérêt pour l'utilisateur. La méthode utilisée est identique à celle décrite pour trouver le graphe géographique (partie 3.2) de notre méthode à la différence que la fonction de distribution est calculée pour chaque utilisateur et qu'elle est utilisée ici pour engendrer une matrice d'intérêt utilisateurs-lieux par méthode naïve bayésienne.
- **Fusion des méthodes décrites précédemment** : par somme pondérées des matrices de prédiction des méthodes précédentes. Ainsi, nous testerons la fusion des méthodes F et S (**méthode (F+S)**), des méthodes KatzFS et G (**méthode (KatzFS+G)**), des méthodes FuseFS et G (**méthode (FuseFS+G)**) et des méthodes F, S et G (**méthode (F+S+G)**). Cette dernière méthode est la méthode dénommée recommandation collaborative unifiée de points d'intérêt (unified collaborative POI recommendation) dans (Ye, et al., 2011).

4.3. Résultats

Cette partie compare les méthodes précédentes avec celle que nous présentons (KatzFSG). De plus, comme il est fréquent, nous ajoutons une méthode « popularité » qui recommande les lieux les plus populaires de la matrice F.

Les méthodes composées (considérant plusieurs matrices pour leurs calculs) nécessitent de définir des paramètres d'importance de chaque matrice (fréquentation, social, géographique). Cependant, comme nous voulons comparer les méthodes dans les meilleures conditions, nous avons recherché les meilleures valeurs pour ces paramètres pour chaque période de temps. Ainsi, nous montrons à chaque fois les meilleures valeurs de rappel et de précision possible pour chaque méthode. Le nombre de recommandations pour chaque utilisateur prendra les valeurs 5 ou 20.

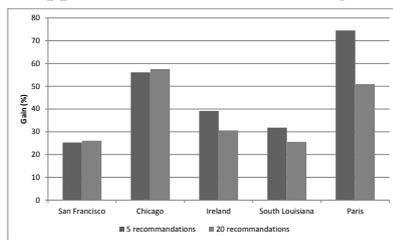
Tableau 1. Rappel moyen pour 5 et 20 recommandations dans chaque zone

Rappel moyen pour 5 recommandations					
	San Francisco	Chicago	Irlande	Sud Louisiane	Paris
Popularité	0,72	1,6	1,97	3,72	3,45
G	1,34	3,57	10,91	6,68	4,48
S	2,88	4,27	5,16	13,22	7,74
F	3,81	6,18	9,73	11,74	6,87
KatzFS	3,81	5,36	10,44	16,97	9,63
FuseFS	4,31	6,84	12,83	15,12	12,52
F+S	4,84	7,8	14,86	19,22	14,98
KatzFS+G	3,82	6,03	12,79	20,54	23,45
FuseFS+G	4,33	7,74	16,65	19,92	23,62
F+S+G	4,86	8,47	15,66	21,77	17,52
KatzFSG	6,09	13,22	21,79	28,7	30,57

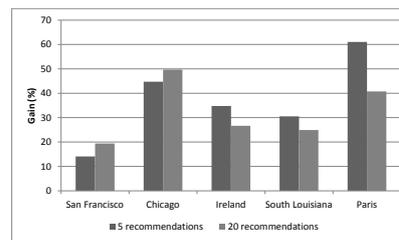
Rappel moyen pour 20 recommandations					
	San Francisco	Chicago	Irlande	Sud Louisiane	Paris
Popularité	2,94	3,62	5,81	9,11	5,13
G	3,81	8,17	12,28	13,4	16,99
S	5,13	7,64	21,02	19,02	10,34
F	9,18	12,7	25,52	30,22	23,39
KatzFS	8,23	11,49	24,25	27,02	19,05
FuseFS	10,13	13,64	30,47	35,21	32,27
F+S	10,71	15,48	32,6	37,75	29,37
KatzFS+G	8,29	12,42	26,32	30,53	33,24
FuseFS+G	10,13	14,53	34,22	38,04	41,82
F+S+G	10,75	16,38	33,48	39,61	33,97
KatzFSG	13,55	25,8	43,73	49,74	51,29

Le tableau 1 montre la valeur moyenne de rappel de chaque méthode pour 5 et 20 recommandations par utilisateur. Nous pouvons remarquer que la méthode (KatzFSG) donne les meilleures valeurs pour chaque zone testée, en moyenne. Au-delà de cette constatation, ces résultats démontrent à nouveau l'intérêt d'utiliser les informations sociales et/ou géographiques pour faire de la recommandation. En effet, nous pouvons voir que la méthode (F+S) donne de meilleurs résultats que (F), et que (F+S+G) donne de meilleurs résultats que (F+S). Cela se confirme aussi avec les autres méthodes composées.

La Figure 1(a) présente le gain en rappel de la méthode (KatzFSG) par rapport à (F+S+G), cette dernière méthode étant pour nous celle de référence pour la recommandation de lieux dans les réseaux sociaux géographiques. En moyenne, le gain en rappel est d'environ 40% sur les différentes zones testées. Pour les valeurs de précision, nous ne présentons seulement que le gain en précision de (KatzFSG) par rapport à (F+S+G) dans la Figure 1(b).



(a) Gain en rappel



(b) Gain en précision

Figure 1. Gains en rappel et précision de (KatzFSG) par rapport à (F+S+G)

Ces résultats montrent clairement l'intérêt d'utiliser notre solution pour améliorer la qualité de la recommandation dans les réseaux sociaux géographiques.

Néanmoins, avant de choisir notre solution, il est nécessaire de se demander jusqu'où il est souhaitable de sacrifier le coût de calcul pour l'amélioration en qualité. En effet, par exemple à San Francisco pour 5 recommandations, nous pouvons voir que le gain en rappel est d'environ 20% mais que la valeur est toujours basse (environ 6% dans le tableau 1). Dans ce genre de cas, est-ce que cela vaut le coût d'utiliser un algorithme coûteux sachant qu'il donne toujours de mauvais résultats ? La réponse devrait dépendre des contraintes de l'application / du système.

De plus, nous devons garder à l'esprit que ce que nous montrons ici ne sont que des valeurs moyennes. Dans certains cas, notre algorithme est moins bon que (F+S+G). Par exemple, pour Chicago, la Figure 2 montre la variation de rappel sur les périodes pour 20 recommandations, et nous pouvons voir que le rappel donné par (KatzFSG) est parfois plus petit que (F+S+G). Nous ne montrons pas les variations de précision car elles sont assez similaires. Ces variations sont difficiles à expliquer, mais en regardant les variations des densités des matrices de fréquentation et sociale dans la Figure 3, il semble que plus la densité des matrices est faible, plus la probabilité d'avoir de moins bons résultats pour (KatzFSG) augmente.

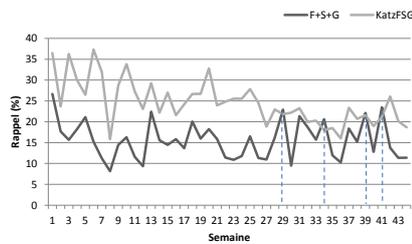


Figure 2. Rappel de (KatzFSG) et de (F+S+G) sur la zone de Chicago sur toutes les périodes, pour 20 recommandations

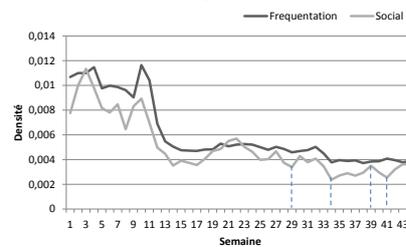


Figure 3. Variation de la densité des matrices sur les périodes, sur la zone de Chicago

Néanmoins, cela n'est pas suffisant pour expliquer ces variations. Prenons une semaine où (KatzFSG) donne un rappel moins bon que (F+S+G). Le tableau 4 montre les valeurs de rappel de chaque méthode sur la semaine 34. Nous pouvons y voir que la méthode (F+S) donne les mêmes résultats que (F+S+G). De plus, nous montrons, dans ce tableau, les résultats de (F+G) et (S+G). Nous remarquons alors que (F+G) donne les mêmes résultats que (F), et que (S+G) améliore les résultats de (S) et (G).

Tableau 4. Valeurs de rappel sur la zone de Chicago pour une semaine sélectionnée où (KatzFSG) est moins bon que (F+S+G)

	Rappel @5	Rappel @20
Popularity	0,25	0,84
S	4,56	6,84
G	4,79	9,54
F	9,63	20,96
KatzFS	7,12	12,73
FuseFS	9,63	21,93
F+S	11,57	23,48
F+G	9,63	20,96
S+G	8,27	13,71
KatzFS+G	7,12	12,73
FuseFS+G	9,63	21,93
F+S+G	11,57	23,48
KatzFSG	8,33	21,19

Etant donné ces observations, il apparaît que la partie géographique ne donne pas de nouvelles informations pour améliorer les recommandations dans cette période car elles sont déjà présentes dans la partie fréquentation. On remarque aussi ce phénomène dans les autres périodes où (KatzFSG) est moins bon que (F+S+G). Ainsi, nous avançons l'hypothèse que, dans ces cas spéciaux, la partie fréquentation et la partie géographique sont « redondantes », telle que la dimension géographique n'apporte pas de connaissance supplémentaire : les lieux corrélés en termes de fréquentation sont aussi proches en distance. Y-a-t-il des événements particuliers à ces moments-là ? Quel est le biais introduit par notre manière de générer le graphe géographique ?

Ces expériences montrent que notre algorithme surpasse généralement les autres algorithmes de manière significative. Il est aussi à noter que même dans les pires cas dans le jeu de données Gowalla, (KatzFSG), (F+S+G) et (F+S) donnent des résultats comparables avec des différences peu significatives en termes de rappel et de précision.

5. Conclusions

Cet article se penche sur la recommandation dans les réseaux sociaux géographiques. Il propose un algorithme combinant le graphe social, le graphe de fréquentation et un graphe géographique en un unique graphe, pour ensuite y réaliser un calcul de proximité entre les utilisateurs et les lieux à recommander dans le graphe à l'aide de la mesure de Katz. Il est comparé à des algorithmes connus du domaine de la recommandation et notamment l'algorithme de (Ye, et al., 2011), que nous considérons comme notre référence dans la recommandation basée sur les réseaux sociaux géographiques. Un jeu de données Gowalla est utilisé pour réaliser les comparatifs.

Tout d'abord, les tests effectués démontrent à nouveau les déclarations de la littérature, indiquant que les informations sociales et / ou géographiques ont un intérêt réel pour améliorer les recommandations. Ensuite, les résultats montrent que notre méthode améliore généralement de manière significative la qualité des recommandations en termes de rappel et de précision, quel que soit le nombre de recommandations. Néanmoins, même si notre méthode est meilleure en moyenne que les autres méthodes testées, nous avons mis en évidence le fait qu'elle ne l'était pas tout le temps. Nous avons essayé de comprendre pourquoi et nous avons remarqué que cela survenait plus souvent lorsque la densité des deux matrices sociale et de fréquentation était faible. Après quelques tests approfondis, nous avons mis en évidence le fait que cela pouvait arriver lorsque la partie géographique et la partie fréquentation sont trop corrélées que les recommandations extraites de la partie fréquentation contiennent déjà les recommandations extraites de la partie géographique. Dans le futur, nous nous intéresserons à réaliser plus d'expériences pour savoir comment identifier ces cas spéciaux.

De plus, nous sommes intéressés par les paramètres des algorithmes. En effet, dans la partie comparative, nous comparons les algorithmes dans des conditions optimales où leurs paramètres sont les meilleurs possibles pour les périodes testées.

Pour ces comparatifs, ces paramètres sont trouvés en les faisant varier et en regardant la qualité des recommandations par rapport à la période suivante. Néanmoins, cela n'est pas possible dans une utilisation réelle et ils doivent être fixés a priori. Pour la détermination de ces paramètres, une étude doit être menée pour s'approcher des paramètres optimaux suivant les caractéristiques des graphes, comme la densité.

Enfin, pour nos travaux futurs, comme nous avons inclus dans ce travail l'aspect spatial au processus de recommandation, nous aimerions aller plus loin en y intégrant l'aspect temporel afin de générer des recommandations basées sur les régularités spatio-temporelles.

Bibliographie

Bothorel, C., 2011. Analyse de réseaux sociaux et Recommandation de contenus non populaires. *Revue des nouvelles technologies de l'information (RNTI)*, 2011, vol. A.5.

Breese, J. S., Heckerman, D. & Kadie, C., 1998. *Empirical analysis of predictive algorithms for collaborative filtering*. Madison, WI.

Candillier, L., Jack, K., Fessant, F. & Meyer, F., 2009. State-of-the-art recommender systems. *Collaborative and Social Information Retrieval and Access: Techniques for Improved User Modeling*, pp. 1-22.

Katz, L., 1953. A New Status Index Derived from Sociometric Index. *Psychometrika*, pp. 39-43.

Ma, H., Lyu, M. R. & King, I., 2009. Learning to recommend with trust and distrust relationships. *s.l., s.n.*, pp. 189-196.

Ma, H. et al., 2011. *Recommender Systems with Social Regularization*. New York, s.n., pp. 287-296.

Noulas, A., Scellato, S., Lathia, N. & Mascolo, C., 2012. *A Random Walk Around the City: New Venue Recommendation in Location-Based Social Networks*. Amsterdam, The Netherlands, s.n.

Resnick, P. & Varian, H. R., 1997. Recommender systems - introduction to the special section.. *Commun. ACM* 40(3), pp. 56-58.

Symeonidis, P., Tiakas, E. & Manolopoulos, Y., 2011. *Product Recommendation and Rating Prediction based on*. New York, USA, s.n., pp. 31-68.

Vasuki, V., Natarajan, N., Lu, Z. & Dhillon, I., 2010. *Affiliation Recommendation using Auxiliary Networks*. s.l., s.n., pp. 103-110.

Ye, M., Yin, P., Lee, W.-C. & Lee, D.-L., 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. *New York, NY, USA, ACM*, pp. 325-334.