Recommendation of shopping places based on social and geographical influences

Romain Picot-Clémente
Lab-STICC, LUSSI department
Télécom Bretagne, Institut Mines-Télécom
Brest, France
romain.picotclemente@telecom-bretagne.eu

Cécile Bothorel
Lab-STICC, LUSSI department
Télécom Bretagne, Institut Mines-Télécom
Brest, France
cecile.bothorel@telecom-bretagne.eu

ABSTRACT
The project tackled in this article is a shopping recommender system that aims at providing recommendations of new interesting shopping places to users, by considering their tastes and those of their friends, since social friends are often sharing common interests. This kind of system is a Location-Based Social Network. It considers social relationships and check-ins; i.e. the action of visiting a shopping place. In order to recommend shopping places, we are proposing a method combining three separated graphs, namely the social graph, the frequentation graph and a geographic graph into one graph. Hence, in this merged graph, nodes can represent users or places, and edges can connect users to each other (social links), users with places (frequentation relations) or places to each other (geographic relations). Given that check-in behavior of users is strongly dependent on the distances, the geographic graph is constructed considering the density of probabilities that a check-in is done according to its distance to the other check-ins. The Katz centrality is then used on the merged graph to compute the scores of candidate locations to be recommended. Finally, the top-n unvisited shopping places are recommended to the target user. The proposed method is compared to methods from the literature on a real-world datatset. The results confirm the real interest of considering both social and geographic data beyond the frequentations for recommending new places. Generally, our method outperforms significantly the compared methods, but under certain conditions that we analyze, we show it gives sometimes mixed results.

Categories and Subject Descriptors
H.3.3 [**Information storage and retrieval**]: Information Search and Retrieval – *information filtering, search process, selection process.*

General Terms
Algorithms, Performance.

Keywords
Social shopping; recommender systems; places recommendation; location-based social networks.

1. INTRODUCTION
With the development of social networks and the widespread use of smartphones, users are sharing more and more contents in mobile situations. In Location-Based Social Networks (LBSNs) like Foursquare[1], sharing places with friends is even the finality.

In this paper, we are interested in the recommendation of shopping places. The aim of this project is to improve the user shopping experience by providing potential interesting new shopping places according to her tastes and those of her friends. The friendship relations are taken into account because it is shown that social friends tend to share common interests and that it is beneficial to improve recommendations [1] [2]. In this article, we are proposing an algorithm considering jointly the three-dimensions of a LBSN: social relations, check-ins, and the geographic coordinates of places.

This work tries first to show that these dimensions, if they are separated, have a less strong predictive power on the accuracy of recommendations than if they are combined. Our second contribution lies in the combination method itself which includes various graphs and then provides a structure to spread influence through the different relationships between the different node types. Indeed, this method proposes to generate a single graph that combines the social graph, the frequentation graph and a geographic graph. [3] [4] [5] [6] show that the check-ins behavior of users in LBSNs often fits a power law distribution (i.e. the closer locations have a much higher probability of being visited). Our third contribution is a method to construct this geographic graph by linking places together with a weight which is the check-in probability following the mutual distances. Finally, we propose an algorithm realizing a propagation of weights by the Katz centrality method through the merged graph. The Katz centrality method described in [7] was originally used in social sciences to measure the degree of influence of an actor in a social network by propagating weights on every path of the network starting from the considered actor, and by measuring the weights that emerge. In our project, the goal of the propagation is to highlight shopping places that would most likely be of interest for users. To test this algorithm, like related works of the domain, we use the publicly available Gowalla[2] dataset.

The first part of this article presents works that have focused on social recommendations, and more specifically on social geographic recommendations. The second part describes our proposed method for finding new shopping places to recommend to users based on a merged graph and the Katz method. In order to evaluate it, the next section will detail some benchmarking methods from the literature. They consider one or more aspects of the system: social graph, frequentation graph and/or geographic influence. Finally, in the last part, the results of these methods on the Gowalla dataset will be compared with the results of our method.

2. RELATED WORKS
In the field of recommender system, there are two main families of techniques to make recommendations [8] : a family of techniques based on collaborative filtering to seek similarities of user profiles based on ratings (the number of stars, the list of past purchases, places visited, etc..) and a family of techniques based on similarities

[1] https://fr.foursquare.com/

[2] Available at *http://snap.stanford.edu/data/#locnet*

of profiles on the basis of content descriptors. The paper [9] compared these methods with different types of similarity measures and different evaluation methods on two reference datasets. According to these tests, they argue that the collaborative filtering methods usually give better results than content filtering methods, at least when there are sufficient available ratings. However, when a new user or a new content appears in the system, collaborative filtering methods have difficulties in providing effective recommendations for new users or to recommend new content, because no historical information are available on them. This problem called the "cold start problem" is overcome by content filtering methods. Nevertheless, these techniques need reliable metadata: it appears that the users profiles or descriptive sheets of places are not reliable, declarative preferences are often inconsistent; collecting relevant information describing places to attract visitors can be costly and complex, or even impossible when taking shops.

In this article, we assume the worst case where there are no descriptors on users and places, except their GPS coordinates. Thus, we reject the content based methods and we focus on collaborative filtering methods for the recommendations based on spontaneous actions of users like their ratings on items, and in our case, like check-ins in LBSNs. Among these spontaneous actions, still often underexploited in recommendation, there are also social activities, like creating relationships or interacting, on platforms allowing it.

Some studies have taken into account the social graph beyond the use of user-item relationships for recommendations. The paper [10] proposes a method for factorizing the rating matrix users-items based on the singular value decomposition (SVD) by minimizing an objective function. A term called "social regularization" is included into the function. This term is constructed considering the direct friends of the user in the social network. The results of this method are compared to a similar method that does not consider a social regularization. This comparison shows that the use of social data gives an improvement of the recommendations.

In [11], the authors propose to combine user similarity matrices from the implicit social network and the explicit one. In their example, they compute two user similarity matrices, one based on the friendship social network (users-users) and one based on the bipartite network (users-items). These two matrices are combined in one similarity matrix by a weighted sum, where weights define the importance of each network in the computation. Then, they generalize this model to consider more graphs. Their tests on Flixster and Epinions show that their algorithm gives better recommendations than traditional collaborative filtering methods based on the neighborhood (see *section 4.1*). We will describe this method in the context of our project in the *section 4.3*, in order to compare it with our solution.

In [12], the authors propose a recommender system of groups based on the friendship social network, and based on the users-groups relations. For this purpose, they first describe a way to combine the social graph with the users-groups graph. Then, they suggest two methods for recommending groups: one based on the proximity in the unique graph using the Katz measure and a method modeling users and groups by latent factors. These two methods give good results, but the method using the Katz measure is the most efficient in terms of computation time and recommendations quality. Their method will be described in the context of our project in the *section 4.4*, in order to be compared with our solution.

In our project, we are more specifically interested in recommendation of places. This implies that we have at our disposal the geographical coordinates of items to possibly improve the recommendations. Very few studies have focused on geosocial recommendations. Nevertheless, we can mention a few recent works that are representative of the field.

The paper [13] focuses on recommending places in LBSNs. It highlights the fact that most current algorithms of recommendation does not take into account all aspects of LBSNs. It proposes an algorithm Random Walk With Restart in a graph where the nodes are the users and the places. Users are connected to other users if a friendship relation exists, and users are connected to places according to the check-ins. The algorithm is compared to other popular algorithms of the domain on the Gowalla and Foursquare (via Twitter) datasets, and gives better results than these latters. Nevertheless, this algorithm does not take into account of a major aspect of LBSNs which is the geographic position of the places. The paper [14] takes it into account. It proposes a method considering the geographic, the social and the frequentation aspects in a LBSN. This algorithm unifies three scores of prediction for each place, based on: the similarity between users based on their check-ins; the similarity between users based on their friendship social network; the geographic information. They based the scores of the geographic information on the naïve Bayesian theory to predict the probability score of a check-in in a given place, by calculating the product of the probabilities of each distance between a given place and the visited places under the distribution law. The three scores are unified in one score by a weighted sum to give more or less importance to each criterion (social, frequentation, geographic). The method has been compared to several methods and they show that taking into account both social data and geographic data, with frequentation data improves the recommendation of places. This method will be called (F + S + G) in the remainder of this article and will be compared to our method in *section 5.2*.

The results given by this latter algorithm are interesting for improving the recommendations in LBSNs but we believe it is possible to improve recommendations by searching for more advanced correlations between places and users. Thus, the method we are proposing aggregates three graphs – the social graph, the frequentation graph and a geographic graph – into one graph that is then used to propagate weights by the Katz centrality method in order to highlight new candidate locations to be recommended.

3. KatzFSG: A KATZ-BASED ALGORITHM CONSIDERING FREQUENTATION, SOCIAL AND GEOGRAPHIC INFORMATION

This section describes the method, called **KatzFSG**, we are proposing for recommending new places to users according to their check-ins and their social network. It is composed of three parts, the first one focuses on the definition of the different graphs. The second part insists on the creation of the geographic graph which is one of the main aspects of the method. The third part describes the way the merged graph is made and how it is used to induce recommendations by a propagation of weights using the Katz centrality method.

3.1 Definitions

The social graph $\boldsymbol{S}$, also called adjacency graph, is based on the friendship relations between users. In this graph $\boldsymbol{S}$, nodes represent users, and edges connect nodes corresponding to users with a friendship relation. The matrix S representing this graph is a symmetric matrix $N \times N$ (N is the number of users), where S_{ij} is 1 when there is a friendship relation between the user u_i and the user u_j, and is 0 otherwise.

The frequentation graph $\boldsymbol{F}$ is based on the check-ins of users in the different places. In this bipartite graph, nodes are either users or places. Edges connect users with the places they have visited (in which they have made one or several check-in(s)), they are weighted according to the number of visits. The matrix F representing this graph is a matrix $N \times M$ (M is the number of places), where F_{ik} is the number of time the user u_i has visited the place l_k.

The geographic graph $\boldsymbol{G}$ connects places together. The matrix G representing this graph is a $M \times M$ matrix. Our proposition for the construction of this matrix/graph is described in detail in the following part.

3.2 A Geographic Graph based on the check-in behavior of users

As shown in several works [3] [4] [5] [6], the check-in behavior of users in a LBSN is highly influenced by the geographic proximity of places, and more precisely the check-in probability in a place according to its distance to another check-in follows a power law distribution. From this observation, we are proposing to construct a geographic graph linking places together by weights which are the probabilities of check-ins in them according to their mutual distance.

For this purpose, we first construct the density of probability of the check-ins following to their distance to another check-in. Figure 1 shows an example of a density of probability.

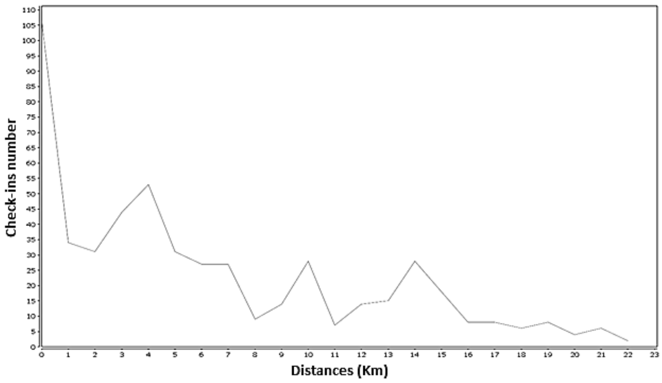

Figure 1. Example of a probability density of checking-in in a place according to its distances with the visited places

Then, in order to have the probability of checking-in every place following to their distance to the other check-ins, this density of probability is approximated by a curve defined by a power law function $f(x) = ax^b$ where x is the distance and a, b are the coefficients.

For finding the optimal coefficients a and b of this function, we pass in log-log scale to have a linear model for the distribution function:

$$\begin{gathered} f(x) = ax^b \\ Let\ f' = \log(f)\ and\ x' = \log(x), \\ f'(x') = a + \log(b)\, x' \\ w_0 = a\ and\ w_1 = \log(b) \\ f'(x', w) = w_0 + w_1 x', with\ w = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} \end{gathered} \tag{1}$$

Thus, we seek w such that $f'(x', w)$ best approaches the points of the real distribution. To do this, the method consists in minimizing a function $E(w)$ such as:

$$E(w) = \frac{1}{2} \sum_{i=0}^{N} (f'(x'_i, w) - y_i)^2 + \frac{\lambda}{2} \|w\|^2 \tag{2}$$

y_i are the real probability measures for the distances x_i.

For minimizing this function, we choose the gradient algorithm allowing us to find a value of a and b such that $f(x)$ approximates the real probability distribution of check-in according to the distances of places taken two by two.

The probability distribution represented by f allows connecting each pair of places by a check-in probability following to their mutual distance. Thus, for each user, it is possible to create a geographic graph $\boldsymbol{G_u}$ where the nodes are the places and where the edges connect places together and are weighted by the check-in probability computed on the distance between the considered places.

A geographic graph $\boldsymbol{G_u}$ exists for every user. However, it appears that these graphs do not differ a lot from one user to another. Moreover, a graph that has been realized for a user with very few check-ins is not really relevant because there is not enough check-ins to find the real distribution function. Thus, we propose to generate one unique graph $\boldsymbol{G}$ connecting places together with weights that are based on the probability distribution of all users, and no longer on a single user. The geographic matrix G is constructed such as

$$G_{ij} = f\left(d(l_i, l_j)\right) = a \times d(l_i, l_j)^b \tag{3}$$

where $d(l_i, l_j)$ is the distance between the places l_i and l_j.

3.3 Katz measure on the merged graph

In our method, we propose to merge the graphs $\boldsymbol{S}, \boldsymbol{F}, \boldsymbol{G}$ in one unique graph $\boldsymbol{C}$. So, in this graph, nodes are either places or users, and edges can connect users together, places together or users with places.

The matrix C represents this merged graph and is constructed as follows:

$$C = \begin{pmatrix} \alpha S & \lambda F \\ \lambda F^T & \gamma G \end{pmatrix} \tag{4}$$

The coefficients α, λ, γ are respectively the influence coefficients of the matrices S, F, G in the matrix C, with $\alpha + \lambda + \gamma = 1$. Note also that F and G have been normalized before integrating C.

The proposed method consists then in propagating a weight in the graph $\boldsymbol{C}$ by the Katz measure as follows:

$$Katz(C) = \beta C + \beta^2 C^2 + \beta^3 C^3 + \cdots, 0 \leq \beta \leq 1 \tag{5}$$

β is the weight that is propagated through the graph. For this algorithm, we are more specifically interested in the effect of the weight propagation on the users-places relations. These relations are represented by the block $Katz(C)_{12}$ in the matrix $Katz(C)$. Given the expensiveness of this computation, a truncated Katz matrix is considered:

$tKatz(C,k)_{12} = \sum_{i=1}^{k} \left(\beta^i C^i\right)_{12}$, k being the maximal rank of calculation

A conservative estimate of the computation cost is $O(N_u \times nnz)$, where N_u is the number of users and nnz is the number of non-zeros in $(FF^T)^k$. In practice, and like [12], we choose to stop at the third rank of calculation[3]:

$$tKatz(C,3)_{12} = \sum_{i=1}^{3} \beta^i \left(C^i\right)_{12} \tag{6}$$

Finally, for each user u, the n unvisited places that have the best weights on the line of the user in the matrix $tKatz(C,k)_{12}$ are selected. They are then the recommendations to provide to the corresponding user.

[3] Improved algorithm like [21] could be used to consider more than three ranks of calculation

4. BENCHMARKING METHODS

This part defines the different methods that will be compared to the one we are proposing. These methods are taken from the literature we have discussed previously. The most of them are not designed originally to recommend places, so we have adapted them for this purpose. They consider one or more aspects of the system: the social network, frequentations and/or geographic coordinates.

4.1 Collaborative filtering based on frequentations (method F)

The method described here is well known in the field of collaborative filtering, it realizes a collaborative filtering on the user neighborhood based on the Pearson similarity. It is described in [15].

Let F be the frequentation matrix connecting users with places, where F_{ik} represents the number of times the user i has visited the place k.

The Pearson similarity between the users i and j based on the frequentation matrix is denoted $sim_F(i,j)$. It is calculated as follows:

$$sim_F(i,j) = \left| \frac{\sum_{k=0}^{K}\left[(F_{ik}-\overline{F_i}).\left(F_{jk}-\overline{F_j}\right)\right]}{\sqrt{\sum_{k=0}^{K}(F_{ik}-\overline{F_i})^2}.\sqrt{\sum_{k=0}^{K}\left(F_{jk}-\overline{F_j}\right)^2}} \right| \quad (7)$$

The similarity matrix based on the frequentation matrix is denoted SIM_F and constructed such as $SIM_{F_{ij}} = sim_F(i,j)$.

Let F' be the matrix representing the prediction scores of the future frequentations of *users* in the places. F'_{ik}, the prediction score of frequentation of the user i in the place k, is calculated as follows:

$$F'_{ik} = \frac{\sum_{j=0}^{N} SIM_{F_{ij}}.F_{jk}}{\sum_{j=0}^{N} SF_{ij}} \quad (8)$$

4.2 Collaborative filtering based on the social relations (method S)

The idea of this algorithm is similar to the one described previously. It is described in [16] as friend-based collaborative filtering. It consists in computing similarity scores between users based on the friendship relations, involving both the social matrix and the frequentation matrix.

Let S be the social matrix connecting users with users, where $S_{ij} = S_{ji} = 1$ if the user i has a friendship relation with the user j, and $S_{ij} = S_{ji} = 0$ otherwise. S is a symmetric matrix.

The similarity between users i and j based on the social matrix is denoted $sim_S(i,j)$. It is equal to the Pearson similarity between i and j, if i and j are friends, and null otherwise:

$$sim_S(i,j) = \begin{cases} \left| \frac{\sum_{n=0}^{N}\left[(S_{in}-\overline{S_i}).\left(S_{jn}-\overline{S_j}\right)\right]}{\sqrt{\sum_{n=0}^{N}(S_{in}-\overline{S_i})^2}.\sqrt{\sum_{n=0}^{N}\left(S_{jn}-\overline{S_j}\right)^2}} \right|, & S_{ij} > 0 \\ 0, & S_{ij} = 0 \end{cases} \quad (9)$$

Then, a similarity matrix SIM_S based on the social matrix can be constructed such as $SIM_{S_{ij}} = sim_S(i,j)$.

Similarly, the similarity $sim_F(i,j)$ between i and j on the frequentation matrix is equal to the Pearson similarity if i and j are friends, and is null otherwise:

$$sim_F(i,j) = \begin{cases} \left| \frac{\sum_{k=0}^{K}\left[(F_{ik}-\overline{F_i}).\left(F_{jk}-\overline{F_j}\right)\right]}{\sqrt{\sum_{k=0}^{K}(F_{ik}-\overline{F_i})^2}.\sqrt{\sum_{k=0}^{K}\left(F_{jk}-\overline{F_j}\right)^2}} \right|, & S_{ij} > 0 \\ 0, & S_{ij} = 0 \end{cases} \quad (10)$$

SIM_F is then constructed such as $SIM_{F_{ij}} = sim_F(i,j)$.

The global similarity between users is represented by the matrix SIM such as:

$$SIM = \alpha.SIM_S + \beta.SIM_F \quad (11)$$

α and β represent respectively the influence degrees of similarity based on the social relations and based on frequentations in the global similarity score.

The prediction scores matrix F' is then constructed as follows :

$$F'_{ik} = \frac{\sum_{j=0}^{N} SIM_{ij}.F_{jk}}{\sum_{j=0}^{N} S_{ij}} \quad (12)$$

4.3 Collaborative filtering based on the fusion of social and frequentation similarities (method FuseFS)

This method is described in [11]. It is very similar to the previous one, but the difference lies in the fact that it is not restricted to users who are friends for calculating similarities.

Thus, we only show the definition of $sim_S(i,j)$ and $sim_F(i,j)$.

$$sim_S(i,j) = \left| \frac{\sum_{n=0}^{N}\left[(S_{in}-\overline{S_i}).\left(S_{jn}-\overline{S_j}\right)\right]}{\sqrt{\sum_{n=0}^{N}(S_{in}-\overline{S_i})^2}.\sqrt{\sum_{n=0}^{N}\left(S_{jn}-\overline{S_j}\right)^2}} \right| \quad (13)$$

$$sim_F(i,j) = \left| \frac{\sum_{k=0}^{K}\left[(F_{ik}-\overline{F_i}).\left(F_{jk}-\overline{F_j}\right)\right]}{\sqrt{\sum_{k=0}^{K}(F_{ik}-\overline{F_i})^2}.\sqrt{\sum_{k=0}^{K}\left(F_{jk}-\overline{F_j}\right)^2}} \right| \quad (14)$$

The following is the same as in the method S for combining the similarity values and for computing the prediction scores.

4.4 Katz algorithm on frequentation and social (method KatzFS)

This algorithm is proposed in [12], it is used originally for the recommendation of groups, but we adapt it here for the recommendation of places in order to compare it to our proposition. In this algorithm, the social graph and the frequentation graph are merged in one graph. The idea is to propagate weights into this graph in order to highlight non-obvious relations.

The matrices S and F are merged in one unique graph represented by the matrix C such as:

$$C = \begin{pmatrix} \lambda S & F \\ F^T & 0 \end{pmatrix}, 0 \leq \lambda \leq 1 \quad (15)$$

λ represents the influence degree of the social graph in the merged graph. F has been normalized before integrating C. Weights are then propagated into the graph by the Katz measure as follows:

$$Katz(C) = \beta C + \beta^2 C^2 + \beta^3 C^3 + \cdots, 0 \leq \beta \leq 1 \quad (16)$$

Since we are only interested in the users-places block, the computation is limited to the block $\mathrm{Katz(C)}_{12}$. For the benchmarks, we will stop the computation at the third rank.

The recommendations are then made according to the newly obtained users-places weights by selecting those with the highest scores.

4.5 Geographic method (method G)

This method called geographic influence recommendation is described in [14]. The idea is to take only into account the geographic coordinates of the visited places in order to deduce unvisited places that are at interesting distances for users. They first find the check-ins distribution function for each user, which is a power law function $f(x) = ax^b$, where a and b are the parameters found from the real distribution of check-ins.

Then, by using the naïve Bayesian method, it is possible to calculate the probability $Pr\ (l_i)$ that a place l_i is interesting relative to its distance to the other visited places:

$$Pr\ (l_i) = \prod_{l_j \in L_k} f\left(d(l_i, l_j)\right) = \prod_{l_j \in L_k} a \times d(l_i, l_j)^b, \qquad (17)$$

L_k is the set of visited places and l_j is a visited place from this set. $d(l_i, l_j)$ is the distance between the two places l_i and l_j .

Finally, the prediction matrix will be composed of all these probabilities that have been computed for each user (the distribution function has to be found for each user).

4.6 Mixing methods

The methods described previously can be fused easily to take into account all aspects: social, frequentation, geographic.

To do this, for each user, their prediction vectors (the lines of the prediction matrices corresponding to the user) are normalized and are summed in a weighted way to give more or less importance to each method.

An example of fusing methods is as follows. Let F'_1 be the prediction matrix obtained by the method G (*section 4.5*). Let F'_2 be the prediction matrix obtained by the method KatzFS (*section 4.4*). For each user i, the prediction scores F'_{ik} for each place k are calculated such as:

$$F'_{ik} = \frac{\alpha}{\max(F'_{1_i})} F'_{1_{ik}} + \frac{\beta}{\max(F'_{2_i})} F'_{2_{ik}} \qquad (18)$$

α and β are the coefficients that define the influence degree of the matrices F'_1 and F'_2 in the final matrix. $\max(F'_{1_i})$ and $\max(F'_{2_i})$ allow to normalize the values obtained in each matrix on each line. Finally, the prediction matrix is the matrix F' connecting users and places with a score.

Then, in a similar way to this example, mergers of the methods described above will be proposed. We will test (**F+S**), (**KatzFS+G**), (**FuseFS+G**) and (**F+S+G**). This latter is the method called unified collaborative POI recommendation in [14].

5. EXPERIMENTATION

5.1 Dataset

In order to test our algorithm, we are using a Gowalla dataset. It consists in a dataset of 196591 users, 950327 friendship relations and 6442890 check-ins of these users in 1279228 different places. The period of check-ins goes from Februray 2009 to October 2010. Gowalla was a location-based social network which closed in March 2012 after being acquired by Facebook in late 2011. This dataset has been chosen because it is publicly available and often used in research papers. To our knowledge, no LBSN dedicated to shopping is available.

Given that the computation time of our algorithm is important, we are proposing to only rely on some geographic areas to test it. We will restrain areas to a maximum size of 500×500 kilometers. Arbitrarily, the selected areas are: San Francisco, Chicago, Ireland, South Louisiana and Paris.

The dataset is divided into time periods of one week from 2009/12 to 2010/10. Each week is used as training dataset and the next week is used as a test dataset. In order to measure the quality of the results, recall and precision measures are done. They are based on the recommendations computed on the training dataset compared to the real check-ins in the test dataset. The choice of taking one week periods is mostly influenced by the limited memory available to handle big matrices on the test machine.

The measures are averaged on the set of periods in order to have an average value of the recall and of the precision for a one week period. This is done for every selected area.

5.2 Results

The different benchmarks are conducted on the methods presented in the previous section, in addition to our method (KatzFSG). As it is frequent in this kind of benchmarks, we add also a method "popularity" which recommends the most popular places from the frequentation matrix.

The composed methods (which aggregate several matrices for their computation) need to define parameters for the influence of the frequentation, the social and/or the geographic proximity. However, we want here to compare the methods in optimal conditions such that they give the best possible results. Thus, the best values of these parameters are found for each period of time in order to obtain in each case the best possible values for the recall and the precision in each method. The number of recommendations for each user will take the following values: 5, 10, 20.

Table 1, Table 2 and Table 3 show respectively the average value of the recall for each method with 5, 10 and 20 recommendations by user. We can see that our method (KatzFSG) gives the best values for every tested area, in average. Beyond this, these results show again the interest of using social and/or geographical information for making recommendations. Indeed, we can see that the method (F+S) gives better results than (F), and (F+S+G) gives better results than (F+S). It is also confirmed with the other composed methods.

Figure 2 presents the recall gain of the method (KatzFSG) compared to (F+S+G), since we consider this method as the reference one for recommending places in LBSNs. In average, the gain of recall is about 40% on the different tested areas.

Table 1. Average recall in % for 5 recommendations in each area

	San Francisco	Chicago	Ireland	South Louisiana	Paris
Popularity	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**

Table 2. Average recall in % for 10 recommendations in each area

	San Francisco	Chicago	Ireland	South Louisiana	Paris
Popularity	1,23	2,15	4,12	6,73	5,09
G	2,32	5,37	7,13	9,09	11,73
S	4,1	6,12	17,21	16,54	9,35
F	5,92	9,19	15,31	18,31	15,36
KatzFS	5,71	8,11	17,62	21,93	12,85
FuseFS	6,69	10,07	19,44	22,59	22,35
F+S	7,44	11,68	22,24	26,75	21,79
KatzFS+G	5,76	8,91	18,25	25,44	27,66
FuseFS+G	6,73	10,86	22,71	27,89	29,35
F+S+G	7,46	12,43	23,42	28,64	25,51
KatzFSG	**9,34**	**18,21**	**32,28**	**38,31**	**37,53**

Table 3. Average recall in % for 20 recommendations in each area

	San Francisco	Chicago	Ireland	South Louisiana	Paris
Popularity	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**

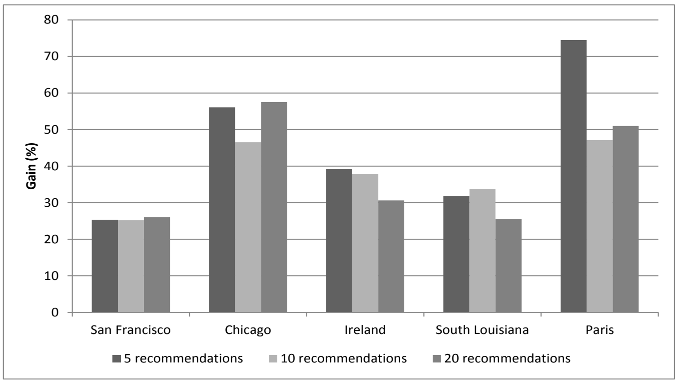

Figure 2. Recall gain of (KatzFSG) compared to (F+S+G)

For the precision values, we present here only the gain of precision of (KatzFSG) compared to (F+S+G) in Figure 3. The gain of precision is about 33% in average on the different areas.

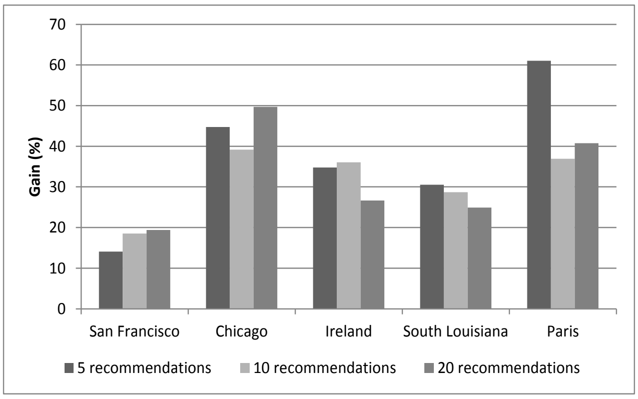

Figure 3. Precision gain of (KatzFSG) compared to (F+S+G)

These results show clearly the interest of using our solution to improve the quality of the recommendations in a LBSN.

Nevertheless, before choosing our solution, it is necessary to wonder how far we are willing to sacrifice the computational cost for the quality improvement. Indeed, for example in San Francisco for 5 recommendations, we can see that the gain in recall is about 20% but the value is still weak (about 6% in Table 1). In this kind of cases, is it worth using a costly algorithm since it still gives a bad result? The answer would depend on the system/application constraints.

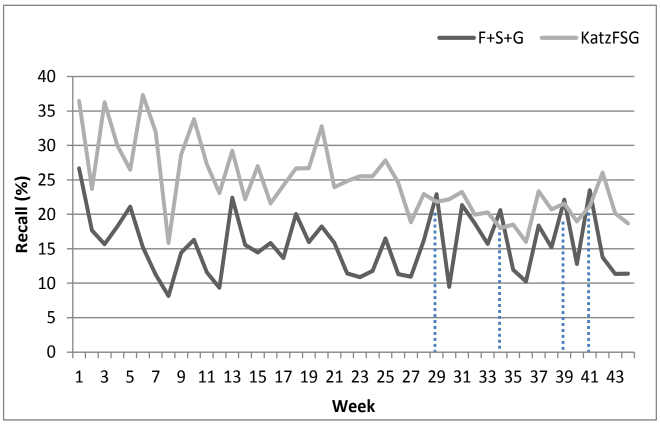

Figure 4. Recall of (KatzFSG) and (F+S+G) for Chicago over all the periods, with 20 recommendations

Moreover, we have to keep in mind that we show here only average values. In some cases, our algorithm is worse than the (F+S+G) algorithm. For instance, for Chicago, the Figure 4 shows the variation of recall over the periods for 20 recommendations, and we can see that the recall given by (KatzFSG) is sometimes smaller than (F+S+G). We do not show the variation of precision because it is quite similar.

These variations are difficult to explain, but when looking at the variation of the densities of the frequentation and social matrices in Figure 5, it seems that more the density of the matrices are weak, more the probability of having worse results from (KatzFSG) increases.

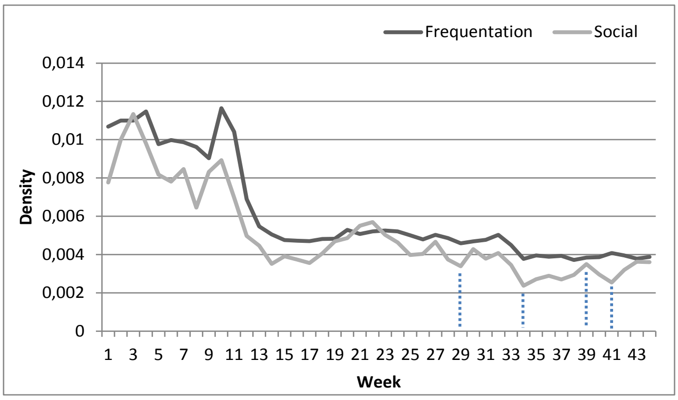

Figure 5. Variation of the density of the matrices over all the periods, for Chicago

However, it is not sufficient to explain these variations. Let's take a week where (KatzFSG) gives worse recall than (F+S+G). The Table 4 shows the recall values of each method for the week 34. We can see that the method (F+S) gives the same results as (F+S+G). Moreover, we show on this table the results of (F+G) and (S+G). We can see that (F+G) gives the same results as (F) and that (S+G) improves the results of (S) and (G).

Table 4. Recall values in % for the area of Chicago for a selected week where (KatzFSG) is worse than (F+S+G)

	Recall @5	Recall @10	Recall @20
Popularity	0,25	0,25	0,84
S	4,56	4,56	6,84
G	4,79	6,00	9,54
F	9,63	13,61	20,96
KatzFS	7,12	9,64	12,73
FuseFS	9,63	13,93	21,93
F+S	**11,57**	**16,13**	**23,48**
F+G	9,63	13,61	20,96
S+G	8,27	9.0	13,71
KatzFS+G	7,12	9,64	12,73
FuseFS+G	9,63	13,93	21,93
F+S+G	**11,57**	**16,13**	**23,48**
KatzFSG	8,33	14,09	21,19

Given these observations, it appears that the geographic part does not give any new good information to improve the recommendations in this period because they are already available in the frequentation part. We can see the similar phenomenon in other cases where (KatzFSG) is worse than (F+S+G). Thus, we advance the hypothesis that, in these special cases, the frequentation part and the geographic part are "redundant", so that the geographic dimension does not bring extra knowledge: correlated locations are also close in distance. Are there any special events in these moments? What is the bias introduced by the way we generate the geographic graph?

These experiments show that our algorithm generally outperforms significantly the other algorithms. Let us note that even in the worse cases in Gowalla dataset, (KatzFSG), (F+S+G) and (F+S) give comparable results with no significant difference in terms of recall and precision.

6. CONCLUSIONS

This article presents a method to enhance the user experience when she is searching for shopping places to go. The purpose is to provide recommendations of shopping places that should interest her. For this, we suppose that her previous visited shopping places and her social network are known. Thus, this kind of recommendation problem is a LBSN recommendation problem because we can consider social relations and visits of users in different geographic locations for making recommendations. This article proposes an algorithm combining the social graph, the frequentation graph and a geographic graph in one unique graph, in order to then realize a proximity computation between users and places in the graph using the Katz measure. This process allows finding places to recommend to users according to their frequentations, their friends frequentations, and the geographic distance between the places. It is compared with well-known algorithms of the field of recommendation and especially the algorithm of [14], that we consider as the reference for making recommendations in LBSNs. A Gowalla dataset is used to realize the comparisons.

First of all, the tests done on different areas demonstrate again the statements of the literature telling that the social and/or geographical information have a real value for improving the recommendations.

Then, the results show that our method generally improves significantly the recommendation quality in terms of recall and precision, whatever the number of recommendations. Nevertheless, even if our method is better in average than the other methods tested, it happens sometimes that it is not the case as we have shown on an example on Chicago. We tried to understand why there are these phenomenon and we saw that it happens more often when the densities of the social and the frequentation matrices are weak. After some extensive testing, we have highlighted the fact that this could happen when the geographic part and the frequentation part are too correlated so that the recommendations extracted from the frequentation part contain already the recommendations extracted from the geographic part. In the future, we will be interested in making more experiments to know how to identify these special cases.

Moreover, we are interested in the parameters of the algorithms. Indeed, in the comparison part, we are comparing the algorithms in the optimal conditions where the parameters are best for the tested periods. For these comparisons, these parameters are found by varying them and looking the quality of the recommendations relative to the check-ins in the following period. Nevertheless, it is not possible to do this in a real use and the parameters have to be fixed a priori. For determining them, a study has to be done to approach the optimal parameters according to the features of the graphs, like their density.

Finally, for future works, as we have included in this work the spatial aspect to the recommendation process, we would like to go further by including the temporal aspect in order to generate recommendations based on spatio-temporal regularities.

7. ACKNOWLEDGMENTS

This work has been supported by the French FUI Project Pay2You Places, "labelled" by three Pôles de Compétitivité (French clusters).

8. BIBLIOGRAPHIE

[1] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks.," in *4th ACM RecSys Conf.*, 2010.

[2] I. Konstas, V. Stathopoulos and J. M. Jose, "On social networks

and collaborative recommendation," in *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, New York, USA, 2009.

[3] D. Brockmann, L. Hufnagel and T. Geisel, "The scaling laws of human travel," *Nature 439,* pp. 462-465, 2006.

[4] N. Couldry, "Mediaspace: Place, scale and culture in a media age," *Routledge,* 2004.

[5] B. Jiang, J. Yin and S. Zhao, "Characterizing human mobility patterns in a large street network," *arXiv preprint arXiv:0809.5001,* 2008.

[6] A. Noulas, S. Scellato, C. Mascolo and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," *ICWSM,* 2011.

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

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

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

[10] H. Ma, D. Zhou, C. Liu, M. R. Lyu and I. King, "Recommender Systems with Social Regularization," in *The fourth ACM international conference on Web search and data mining (WSDM '11),* New York, 2011.

[11] P. Symeonidis, E. Tiakas and Y. Manolopoulos, "Product Recommendation and Rating Prediction based on," in *The fifth ACM conference on Recommender systems (RecSys 2011)*, New York, USA, 2011.

[12] V. Vasuki, N. Natarajan, Z. Lu and I. Dhillon, "Affiliation Recommendation using Auxiliary Networks," in *The fourth ACM conference on Recommender systems (RecSys 2010)*, 2010.

[13] A. Noulas, S. Scellato, N. Lathia and C. Mascolo, "A Random Walk Around the City: New Venue Recommendation in Location-Based Social Networks," in *IEEE Internationcal Conference on Social Computing (SocialCom 2012)*, Amsterdam, The Netherlands, 2012.

[14] M. Ye, P. Yin, W.-C. Lee and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *The 34th international ACM SIGIR conference on Research and development in Information Retrieval*, New York, NY, USA, 2011.

[15] J. S. Breese, D. Heckerman and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, 1998.

[16] H. Ma, M. R. Lyu and I. King, "Learning to recommend with trust and distrust relationships," in *The third ACM conference on Recommender systems (RecSys 2009)*, 2009.

[17] A. Noulas, S. Scellato, N. Lathia and C. Mascolo, "Mining User Mobility Features for Next Place Prediction in Location-based Services," in *IEEE Internationcal Conference on Data Mining (ICDM 2012)*, Brussels, Belgium, 2012.

[18] J. Golbeck, "Trust and nuanced profile similarity in online social networks," *ACM Trans. Web,* pp. 1-33, 2009.

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

[20] J. He and W. Chu, "A Social Network-Based Recommender System (SNRS)," *Annals of Information Systems, Special Issue on Data Mining for Social Network Data,* vol. 12, pp. 44-74, 2010.

[21] K. C. Foster, S. Q. Muth, J. J. Potterat and R. B. Rothenberg, "A Faster Katz Status Score Algorithm," *Comput. Math. Organ. Theory,* vol. 7, no. 4, pp. 275-285, 2001.