“Current City” Prediction for Coarse Location Based Applications on Facebook

Wipada Chanthaweethip*, Xiao Han*, Noël Crespi*, Yuanfang Chen*, Reza Farahbakhsh* and Ángel Cuevas†*
* Institut-Mines Télécom, Télécom SudParis,
{wipada.chanthaweethip, han.xiao, noel.crespi, yuanfang.chen, reza.farahbakhsh}@it-sudparis.eu
† Universidad Carlos III de Madrid,
{acrumin}@it.uc3m.es

***Abstract*—Location-Based services with social networks improve users’ experience and enrich people’s social live. However, location information is often inadequate due to privacy and security concerns. We seek to infer users’ ‘Current City’ on Facebook for coarse location based applications. We first extract users’ multiple explicit and implicit location attributes, and analyze correlations of these attributes from two perspective: user-centric and user-friends. We observe that both user-centric and user-friends location attributes tightly correlate to a user’s Current City (e.g., 60% of users stay in their hometown, 60% of users live in the same city as 50% of their friends). Based on extensive analysis and observations on location attributes correlations, we have constructed a Current City Prediction model (CCP) using artificial neural network (ANN) learning frameworks. The experimental results indicate that we achieve accuracy levels of 84% for city-level prediction and 98% for country-level which are increases of 9% and 18%, respectively than what is possible with Tweecalization.**

***Keywords*—LBA; Coarse Location; Location Prediction;**

I. INTRODUCTION

Online social networks (OSN) combined with their users’ profiles and preferences provide the impetus for a number of compelling personalized applications and services that can significantly improve users’ experience and enriches people’s social life. Location-Based services (LBS) are one of the most popular and pervasive categories of services that make use of users’ location information from online social networks or from mobile GPS devices. The LBS can be categorized in two types: fine location based services and coarse location based services. The former requires a high precision of location with longitude and latitude generally, while the latter only asks for the coarse location or region. There are number of fine location based applications and services that people are becoming used to implementing: foursquare helps people search restaurants and public transportation nearby; friend finder provides people a chance to easily find new friends around them and enlarge their social life.

Most of the previous work has focused on issues related to the above-mentioned precise-location based applications, while much fewer have paid attention to the coarse location based services. However, there is indeed the potential for applications and services that could serve people based on their region or area rather than on their precise location. For instance, in a mayor’s election campaign, mayoral candidates could announce their activities and policies through a city-based application to attract and inform citizens. Municipal governments could invite bureau of meteorology push local weather report to its citizens as a service to tempt people’s heart. Also local matrimonial agency could broadcast and advertise their coming date party for singles in the same city therefore to achieve the goals of enhancing its reputation and earning money.

Providing or extracting users’ coarse location for coarse location based applications is basic but not a trivial task. Sampling and using the exact location position by means of check-in data from OSN or mobile GPS devices is not enough for non-precise-location based applications. Most people do not change their coarse region (e.g. city, state or country) frequently but move quite a bit inside their region. In this case, neither we do not need to change the region information even when a user’ precise location changes, nor is it viable use of resources to obtain a coarse location by sampling, analyzing and pre-processing the massive amount of location-specific data. Even though some online social networks allow users to register their static region for the vast number of users, location information is too sparse. Only a few users (about 16%) register their city-level locations on Twitter [1], and an even smaller fraction of Facebook users (about 6%) input their home addresses [2].

We address the issue of extracting and predicting users’ coarse location from users’ profiles and their social circles, from the perspective of requirements and challenges. We focus on Facebook for three reasons. First, Facebook is the most famous and active social network attracting over one billion active users by September 2012, and the number of applications and websites that integrate Facebook have already increased to two million since Dec. 2011. Second, Facebook not only provides one important coarse location attribute of each user’s current city which is very practical and easy to utilize by coarse location based applications and services, it also provides other potential explicit and implicit attributes which are informative for location prediction. Third, as mentioned above, location information is too sparse on Facebook.

To predict users’ current city on Facebook, we first extract exhaustive explicit and implicit location-relative attributes from users’ profiles. In addition to current city, hometown is another explicit location on Facebook. We can also infer and translate location information from two other Facebook attributes of high school and employer. Based on crawling 345,506 Facebook users, we study and analyze the correlation between current city and other location attributes from two perspectives: user-centric and user-friends. User-centric analysis utilizes the users' own other location information and user-friends analysis concentrates on the location information from users' friendships. From our comprehensive observations and analysis, we observe that both user-centric and user-friends location attributes correlate strongly to user's current location (e.g., 60% of users stay in their hometown, 60% of users live in the same city as 50% of their friends). Finally by applying artificial neural network learning mechanism, we build up a Current Location Prediction model and predict coarse location at two levels: city and country. The experimental results indicate that we achieved 96% of overall accuracy where the model can correctly predict 84% for city-level prediction and 98% for country-level which are increases of 9% and 18%, respectively than what is possible with Tweecalization.

To best of our knowledge, there is still no work reported on Current City prediction for providing coarse location information to various applicable services on Facebook. Moreover, our approach departs from other location-prediction method in the following ways:

(i) We do not merely extract the explicit location information but also infer and translate implicit location from education and work attributes;

(ii) From two perspectives, user-centric and user-friends, we performed extensive location correlation analyses based on 345,506 users, and correctly predict their current city with accuracy of 84% for city-level and 98% for country-level.

(iii) We construct a Current City Prediction (CCP) model, and predict users' location at two precision levels: city and country.

The structure of this paper is organized as follow. Section II provides an overview of the related work. The data analysis is presented in section III, including the data processing and attributes' implications. Section IV proposes our Current Location Prediction model based on ANN-based learning. The results and the accuracy of our prediction model are discussed in section V. Finally, our conclusion is presented in section VI.

II. Related work

User's location becomes essential information in large-scale applications such as content-based delivery networks, location-based recommendation systems [3] and personalized services. Even with a billion people already connected to online social networks (OSNs), the problem of location information deficiency still exists. Several studies have proposed prediction methods to solve this problem, as they determined how user's location can be inferred from implicit information. Reference [4] confirmed that geographical location has an influence on our social network structure. There is a high probability that friends in our social network are from our physical social circle such as our school friends or people who share the same geographical location. Cho *et al.* discovered that user's movements have a pattern that indicates geographic and social constraints; people will move a long distance when they have a friend in another location [2]. Other researchers created a model that describes the interactions between geography and social relationship, which they then utilized to predict user location [5]. Using the maximum likelihood approach, they predicted the location of Facebook users in United States from a given location. That method used latitude and longitude, while our work uses the translation of location ID information to indicate city and country. A probabilistic framework that uses tweet content to estimate user location at the city level was implemented by Cho et al. [2]. Researchers in [6] built a prediction model for the home location of Twitter users based on their friendship network. The prediction methods of each of the studies above use a single perspective of implicit information.

However, there are other methods that use multiple perspectives of implicit information. Reference [7] employed the integration of both a social network and user's centric information to predict user's home location. They extracted the location from a user's tweet content, as it is called user-centric. In addition, they utilized the relationship between the followers and venue of that tweet content. They claimed that their work can predict only one location even where a user may involve multiple locations in their tweet content. They therefore proposed multiple-location profiling that considers all the sets of a user's location information [1]. It is obvious that friendship information can derive a pattern to predict a target output, such as user's location.

III. Data Description and Analysis

A. Data Description

Facebook is the most popular social network that was success to attract attention from all over the world, including celebrities, merchants, politicians, NGOs, artists and demographic researchers. As Facebook allows users to set their own privacy policies which can keep themselves only exposed to certain viewers, we merely intend to extract information from users who make their profile public as our dataset resources. We first selected some active users who make their profile accessibly and provide relatively more information public as root users. Started from these root users, we crawled their friends and their friends of friends by using a Breadth First Search (BFS) approach. We crawled Facebook from March to June in 2012 and collected 345,506 users' profiles which consists of their user ID, name, age, gender, current city, hometown, high school and employer, interests, friendship, etc.

Even though the users in our database make their profile open to public, we observed that most them only expose some of their attributes. For instance, Bob is a user in our database. He exposes attributes of high school and employer but leaves his age and current city empty. However, in general people share many similar attributes with their friends, and parts of people's attributes could reflect or have implication on other parts of their attributes. If Bob and his friend Jim were classmates in high school and Jim exposes his age, we could infer they are almost the same age. Similarly, the company which Bob is serving for is located in Paris might indicate that Bob's current city is Paris as well.

As we are interested in predicting users' current city for coarse location based applications in this paper, we tend to extract the exhaustive explicit and implicit location-related information both from users' themselves and their friends. In addition to current location, hometown is another explicit location on Facebook. Although it is a historical location, people often would prefer to go back their hometown after their university, we expect a correlation between hometown and current residence location. Generally people live in the same city as where they work, we also hope to find a tight correlation between users' current location and employer. A users' educational institution could be another location indication.

Another issue that we address is translating the non-location attributes to their corresponding locations. The Facebook Graph API[1] provides services for location information by inputting the Facebook ID of a school or employer. However, the location information can only be obtained if the organization provides their location on its Facebook site. By this approach we input 29,505 schools and 68,205 employers, and obtain 17,786 school locations and 13,389 employer addresses.

Facebook supports user-generated profiles and attributes, we must overcome the problem of diverse identification of the same location. For example, Bob is from Paris, France and his classmate Jim claims himself from Paris as well. Although both do come from Paris, Facebook generate two Facebook location IDs for those two places and identifies them as different places. To alleviate this problem, we use Yahoo's Geocode API [2] to translate and unify places into their corresponded cities and countries.

Table I presents the number of profiles with different available attributes in our dataset. For instance, we have 120,192 users explicitly presenting their hometown and 111,339 users whose hometowns are seen publicly. We also calculate number of users who have friends that publicize their location attributes. If a user has at least one friend that indicates their current location, we add 1 to "U has friends with Current Location" in Table I, and do the same to the other attributes. Since we are interested learning the correlations between other location attributes and current location. The study is conducted based on those users who have both attributes. For instance, we study the correlation between users' hometown and current location based on the 83,258 users who present both hometown and current location.

B. Attributes' Correlation

In this subsection, we study attributes' implications to users' current location from two perspectives: user-centric and user-friends' attributes. The analysis is divided in two parts: firstly we observe how users' current city (i.e. CC) correlates to their hometown (i.e. HT), high school (i.e. HS) and employer (i.e. EM) locations. Second, we note how their friends' location information correlate to their own current city, including current city, hometown, employer and high school. We evaluate all of the following studies and analyses with the goal of determining two precise levels: city and country.

1) Correlation of User-centric Attributes: First, we analyze and explore how users' other location information could be helpful to infer their current residence location. We calculate the Average Correlation between users' current location and other attributes by means of equation 1, where M is the total number of users; u_{CC} and u_{ai} represent users' current location value and any of the other location attributes value. If u_{CC} is equal to u_{ai}, then $F(u_{CC}, u_{ai}) = 1$; otherwise $F(u_{CC}, u_{ai}) = 0$.

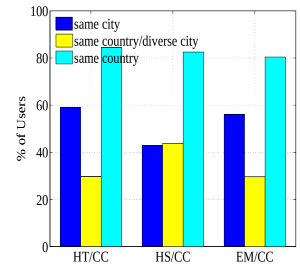


Fig. 1. User-centric Correlations

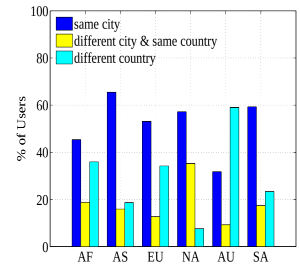


Fig. 2. Correlation of CC and HT by Continent

$$\overline{Cor_u} = \frac{\sum_{i=1}^{M} F(u_{CC}, u_{ai})}{M} \quad (1)$$

We start with how much we can infer about users' current city from their hometown by computing the percentage of users who indicate both locations (hometown and current locations) with three different properties: i) same city, ii) same country but diverse city, and iii) same country. Moreover, we use a similar approach to study the users' current location correlations with their high school and employer based on the users' communities where both of current location and high school are indicated, or both current location and employer.

From the results in Fig.1, we can see that around 60% of people live in the same city as their hometown. On the contrary, employer location does not match current city with a high probability as our expectation. One possible reason could be many large companies have branches all over the world but only indicate the address of the headquarters on their websites. However, we still find 56% of users have the same employer city as their current city. For high school, we matched the high school's city and current city with 42.8% of users. At country level, we note that more than 80% of users stay in the country of their hometown, employer and high school.

We further observe that people move in different patterns according to the continent they belong to, since we could consider that people from immigrant countries would be more likely to move and people from the countries that have more static populations would travel less. We intuited, people from different continents present different relocation patterns, shown in Fig.2 where we use AF, AS, EU, NA, AU and SA represent Africa, Asia, Europe, North America and South America. Almost 70% people in Australia moved to a new place by knowing their hometown, therefore it might much harder to predict Australians current location from their hometown. However, more than 60% of people in Asia and South America remain in their hometown. The correlations of current city and hometown are much higher in these continents.

2) Correlation of User-friends' Attributes: With OSNs people can make friends without physical interaction, various studies have indeed established the similarities and relevance among OSN friends. In this section, therefore, our objective is to understand whether the location information available from

[1]https://developers.facebook.com/tools/explorer

[2]http://developer.yahoo.com/boss/geo/

TABLE I. AVAILABLE USERS WITH DIFFERENT LOCATION INFORMATION

Category	#Users	%	Category	#Users	%
Users in all	345506				
U with Current Location	120192	34.8	U has friends with Current Location	98925	28.6
U with Hometown	111339	32.2	U with (Hometown & Current Location)	83258	24.1
U with High School	34962	10.1	U with (High School & Current Location)	20081	5.8
U with Employer	12581	3.6	U with (Employer & Current Location)	7232	2.1
U with Friends	211042	61.8	U with (Friends & Current Location)	94512	27.4
U has friends with Hometown	57174	16.5	U with Current Location & U has friends with Hometown	56210	16.3
U has friends with High School	10967	3.2	U with Current Location & U has friends with High School	10866	3.1
U has friends with Employer	4163	1.2	U with Current Location & U has friends with Employer	4134	1.2

friends can give an indication of where users are from. In particular, people are generally friends with their classmates or others from the same high school, colleagues those who may work in the same company or organization, and with people who participate in the same activities (who may all be in the same city). We define the attribute correlation between friends and a user as the rate of friends who share the same value of a corresponding attribute, equation 2. In this equation, N is the number of users' friends; u_{ai} and u_{fj} stand for users' location attribute value and friend' attribute value. If u_{ai} is equal to u_{fj}, then $F(u_{ai}, f_{aj}) = 1$; otherwise $F(u_{ai}, u_{fj}) = 0$

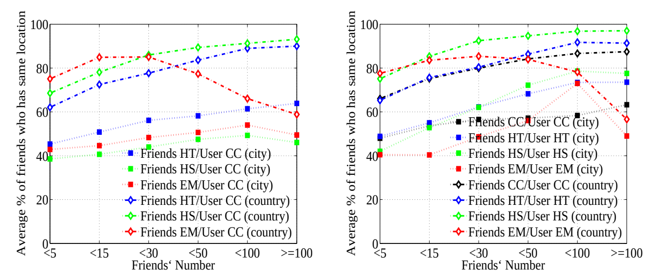
$$Cor_{u_{ai}f_{aj}} = \frac{\sum_{i=1}^{n} F(u_{ai}, f_{aj})}{N} \tag{2}$$

First, we calculate what percentage of a users' friends have the same location (current location, hometown, employer, high school) as the users' corresponding locations. At city level (Fig.3(a)), we have around 50% of users who work in the same city as more than 50% of their friends; around 60% of users went to the same school as more than 50% of their friends; and 60% of users come from the same city as more than 60% of their friends. At country level the number of location correlations goes even higher – more than 70% of users are in the same country as 80% their friends (Fig.3(b)).

We also compute the interrelation of users' current location and their friends' hometown, employer, and high school respectively, shown in Fig.3(c) and Fig.3(d). Viewed from the city level, more than 50% of users are in the city that 60% of their friends come from; and around 40% of users live in the same city that 50% of their friends work in or where they went to high school. And again, at the country level, about 70% of users live in the same country as 80% of their friends.

Another analysis was done based on the assumption that location can be inferred more precisely when we have more knowledge about a user's friends. We classify users into six groups based on the number of their friends whose location is available $\{< 5, < 15, < 30, < 50, < 100, >= 100\}$. We compute the average percentage of friends who have the same location as their users for each group, shown in Fig.4.

Fig.4(a) compares the correlation of the levels of identical location between users and their friends. In Fig.4(b) we calculate the average percentage of users in each group who have the same current location as their friends' hometown, high school and employer. As we had expected, the higher the percentage of a user' friends who has the information of hometown and high school, the higher the percentage of users share similar locations with their friends. However, the improvement decreases when the number of friends is greater than 100. Meanwhile, the employer country projects a different situation. We propose two possible reasons. The first, similar to the explanation mentioned earlier is that a company could have many addresses for branches running the business but only register one location on Facebook website. Moreover, international cooperation between companies and branches of a company means that colleagues may work in different places.



(a) Correlation between users' current city and friends' location

(b) Correlation between corresponding location

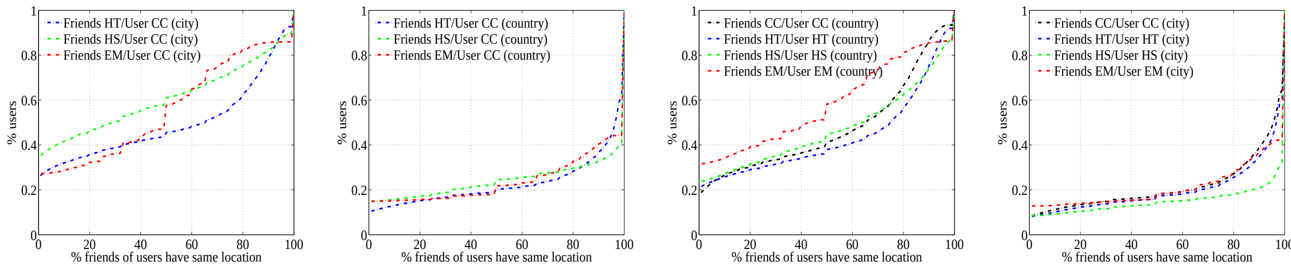
Fig. 4. Implication of Friends' Number

Based on the above extensive analysis and observations of location attributes correlation, we can say that: (i) users' current locations correlate to their hometown, high school and employer locations to a certain degree; and (ii) friends' explicit and implicit location attributes reflect users' current location to some extent.

IV. CURRENT CITY PREDICTION MODEL

In this section, we introduce current city prediction (CCP) model. Guided by the tight correlations between different locations, CCP applies all the relative attributes as input. The goal of CCP is to predict users' current city by two levels: city level and country level. We train CCP by exploiting artificial neural network (ANN) learning framework. We rely on two main reasons to use ANN: ANN is a supervised learning method which can address complex and unstructured input well [8]; and also it allows both multiple inputs and multiple outputs.

Based on the analysis on the ground truth data from Facebook, we have two sets of attributes which are user-centric and user-friends respectively. These attributes involve four pieces of information which are current city (CC), hometown (HT), employer (EM) and high school (HS) locations for both users and their friends. Since we used multiple attributes



(a) Correlation between users' current city and friends' location (city)

(b) Correlation between users' current country and friends' location (country)

(c) Correlation between corresponding location (city)

(d) Correlation between corresponding location (country)

Fig. 3. Correlation between Users and Their Friends

which in some cases had no location value thus, we needed to translate them into a location at two levels: city and country as explained in section III-A. The analysis of user-centric attributes presents that user current location can be inferred from other attributes including the similarities in locations between users and their friends. To abstract this information, we define a score by comparing each pair of attributes: CC-HT, CC-EM, and CC-HC. We also denote a score between two users who share similar location value. For a user u_i who has friends' network F_i which consists of $\{f_1, f_2, \dots, f_n\}$, the score is defined by comparing each pair of locations between user(u_i) and friend(f_j), where a location $l = \{c, p\}$ consists of a city $c \in C$ and country $p \in P$. For example, if both a user and that user's friend have a current location in France the score will equal to 1 but if they are in Paris, France the score will increase to 2. The maximum score will be assigned when two users who are friends share similar values for city and country. The comparison is done for all pairs of CC, HT, EM and HS.

$$E \in \{CC, HT, EM, HS\}$$
$$Scr(E)_{(u_i,f_j)} = \begin{cases} 2: & c_{u_i} = c_{f_j} \wedge p_{u_i} = p_{f_j} \\ 1: & c_{u_i} \neq c_{f_j} \wedge p_{u_i} = p_{f_j} \\ 0: & \text{otherwise} \end{cases} \tag{3}$$

To predict a user's current location we generate a location candidates list from a friend's network, F_i which is a set of friends of user u_i. We first finalized the location of each friend. The primary location to be selected is the CC location, if this information is not available,then the HT, EM and HS location will replace in that order.

The locations candidate list L_{u_i} is generated based on the frequency of its occurrence within a user's friend network. Where $L_{u_i} = \{l_1, l_2, \dots, l_n\}$ is a locations candidate list of user u_i, while a location $l = \{c, p\}$. Similar country and city names are grouped together and the percentage of occurrence for each l_j will be calculated. Equation 4 and 5 compute percentage of the occurrences of a country and of a city respectively. For each user, the total number of friends in their network, the total number of friends who have share the same country value and who share the same city value are computed.

$$Per_{p_i} = \frac{\sum_{j=1}^{n} d_{p_i}(f_j)}{n} \quad d_p(f) = \begin{cases} 1: & \text{if } f_j = p_i \\ 0: & \text{otherwise} \end{cases} \tag{4}$$

$$Per_{c_i} = \frac{\sum_{j=1}^{n} b_{c_i}(f_j)}{\sum_{j=1}^{n} d_{p_i}(f_j)} \quad b_c(f) = \begin{cases} 1: & \text{if } f_j = c_i \\ 0: & \text{otherwise} \end{cases} \tag{5}$$

For each location in the candidate list, a score whose value vaires from 0 to 1 (as in equation 6) is computed. The score abstracts the level of similarity of location between two users. For example, for the candidate location of Paris, France l_j and we have 6 friends of a user whose final location is Paris, France. The score is computed by comparing each pair of locations between a user and their friends as in equation 3 for all locations: CC, HT, EM and HS. These 6 friends could have a score of HT=7/12 as 3 of them share the same HT location with the user for both city and country, while one only shares the same country and other two have not disclosed their HT. The score for HT is equal to 0.58.

$$S_{(E_i)} = \frac{\sum_{i=1}^{n} Scr(E_i)}{\sum_{j=1}^{n} b_{c_i}(f_j)) \times 2} \tag{6}$$

The locations candidate list consist of 16 parameters as listed in table II. They are used as input variables to ANN. This paper use well-known neural network architecture which is multilayer perceptron (MLP). MLP composes of input, hidden and output layers. At the input layer, variables are fed through the network during training and then they compute a result at output layer. In training process, the predicted value is compared to the actual value, the different between these two values is propagated back to the network which adjusts its calculations to improve prediction result. Finally, the output is defined by comparing location with the actual location of that user, so that the two classes of output are: a) "City" or b) "Country". The meaning of a) is that the location in the candidate list is correct at both the city and country levels compared to the actual user location. While b) indicates that the location is confirmed only at the country level (i.e. the city names differ).

V. Results and Evaluation

We use real users' information from Facebook to evaluate our proposed CCP model. Data is partitioned into two sets: training and testing, proportioned 70:30. The training dataset is used to build the model and then 30% of the data is used for evaluation as it appears as new data to the model. We first

TABLE II. PARAMETERS AND DIRECTION TO ANN

Parameter	Direction	Parameter	Direction
User's gender	Input	Score CC-HT	Input
# friend from profile	Input	Score CC-EM	Input
% of friend who have location	Input	Score CC-HS	Input
% of people moving city	Input	Score CC	Input
% of people moving country	Input	Score HT	Input
# friend who we have their profile	Input	Score HS	Input
# friend who have location	Input	Score EM	Input
% of country in candidate list	Input	Result class	Output
% of city in candidate list	Input		

TABLE III. CONTINGENCY MATRIX

	(Actual) City	(Actual) Country
(Predicted as) City	TP	FP
(Predicted as) Country	FN	TN

TABLE IV. PERFORMANCE EVALUATION RESULTS

Parameters	Formulae	Result
Sensitivity	$\frac{TP}{TP+FN}$	83.6%
Sensitivity	$\frac{TN}{FP+TN}$	97.9%
Positive predictive value (PPV)	$\frac{TP}{TP+FP}$	85.0%
Negative predictive value (NPV)	$\frac{TN}{FN+TN}$	97.6%
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$	96.1%

evaluate our model by using contingency matrix as presented in table III to measure specificity, sensitivity, PPV, NPV and accuracy. Then we check the different prediction capabilities when the number of friends increases.

From our ground truth dataset, the model achieved a very high performance for identifying users' current city. The model can produce high sensitivity, known as recall rate which is the ratio of correctly predicted locations at City level (84%). Specificity is also called true negative rate where the locations correctly recognized as country of all locations without known city (98%). The result of model evaluation is presented in table IV. We note that CCP model could achieve overall accuracy of 96% for predicting location at city and country level. The ability to predict locations at City level is 85% while 98% at Country level. Comparing to the existing model [6] that has a 75% accuracy level for city and 80% for country prediction. We achieve a better performance because of our combination of explicit and implicit location information from two aspects: user-centric and user-friends' attributes, while other works focuses only on friendship information. Thus, the multiple location-related attributes and a strong tie of friendship network allows us to improve user location prediction.

We have done further analysis focusing on the number of friends comparing with prediction result. Fig.5 shows that when number of friends in the network is large, location can be more precisely identify (at city level). Corresponding to our analysis that when the number of friends increases, the percentage of users who share similar location increases. In contrast, if the number of friends decreases location can be defined at country level. This explains that when we have strong tie in our friends' network we are able to locate users' "Current City" more precisely.

VI. CONCLUSION

In this paper, we propose CCP model for predicting user's current city on Facebook. We first analyze Facebook users

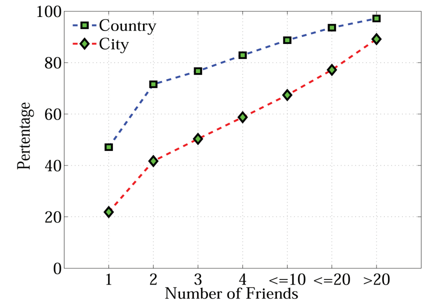


Fig. 5. Prediction Capacities

with multiple explicit and implicit location-related attributes. We found that users' current city strongly correlates to other location-related attributes from both of user-centric and user friends' perspective. We derive the correlations by generating locations candidate list and use ANN learning mechanism to train our model. With the proposed model, coarse location based application are able to obtain current city information without asking from users. In addition, the current city that obtain from our method is more accurate and simple than using IP address which currently difficult to render geolocation. We achieved 96% of overall accuracy where the model can correctly predict 84% for city-level prediction and 98% for country-level.

VII. ACKNOWLEDGMENTS

The research leading to these results was funded by the European Union under the project eCOUSIN (EU-FP7-318398) and the project TWIRL (ITEA2-Call 5-10029).

REFERENCES

[1] R. Li, S. Wang, and K. C.-C. Chang, "Multiple location profiling for users and relationships from social network and content," *CoRR*, vol. abs/1208.0288, 2012.

[2] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: improving geographical prediction with social and spatial proximity," in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10, 2010, pp. 61–70.

[3] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '11, 2011, pp. 1082–1090.

[4] G. Pui, C. Fung, K.-O. Chow, and T.-L. Wong, "What is the role of physical location in our online social networks?" in *Machine Learning and Cybernetics (ICMLC), 2012 International Conference on*, vol. 4, July, pp. 1252–1257.

[5] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, ser. CIKM '10, 2010, pp. 759–768.

[6] S. Abrol, L. Khan, and B. Thuraisingham, "Tweecalization: Efficient and intelligent location mining in twitter using semi-supervised learning," in *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2012 8th International Conference on*, Oct. 2012, pp. 514–523.

[7] R. Li, S. Wang, H. Deng, R. Wang, and K. C.-C. Chang, "Towards social user profiling: unified and discriminative influence model for inferring home locations," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '12, 2012, pp. 1023–1031.

[8] E. Turban, R. Sharda, and D. Delen, *Decision Support and Business Intelligent Systems*. Upper Saddle River, New Jersey: Pearson, 2011.