European Territorial Cooperation Programme
Greece - Italy
2007-2013
INVESTING IN OUR FUTURE

Co-funded by the European Union (ERDF) and by National Funds of Greece & Italy

INTERSOCIAL-I1-1.2, Subsidy Contract No. <I1-12-03>, MIS Nr 902010

European Territorial Cooperation Programme

Greece-Italy 2007-2013

INTERSOCIAL: Unleashing the Power of Social Networks for Regional SMEs

Deliverables D4.4.1: Deployment of tools for social-based promotion of SMEs

Action 4.4: Use of Social Promotion Mechanisms

WP4: Deployment and Evaluation of Innovation Devices in Specific SMEs

Priority Axis 1: Strengthening competitiveness and innovation
Specific Objective 1.2: Promoting cross-border advanced new technologies

Financed by the European Territorial Cooperation Operational Programme "Greece-Italy" 2007-2013, Co-funded by the European Union (European Regional Development Fund) and by National Funds of Greece and Italy

Deployment of tools for social-based promotion of SMEs

<u>Deliverable D4.4.2 Action 4.4</u>

Workpackage WP4: Deployment and Evaluation of Innovation Devices in Specific SMEs

Responsible Partner:	UOI (LP)		
Participating Partner(s):			
SAT:	UPATRAS		
WP/Task No.:	WP4	Number of pages:	18
Issue date:	2013/8/31	Dissemination level:	Public

Purpose: Investigating the role of social media for promotion of SMEs
Results: We have studied the use of location based social media for promotion of SMEs. We have performed collection and analysis of foursquare data by building a graph representing the social connections. Subsequently we have carried out analysis of data and development of prediction policies.
Conclusion: Based on the analysis of data from location based social networks we can suggest effective and cost efficient promotion strategies using social media for SMEs.

Approved by the project coordinator: Yes
Date of delivery to the JTS/MA: 20/10/2013

Document history

When	Who	Comments
2013/10/31	Ioannis Fudos, Evaggelia Pitoura, Ioannis Kotrotsios	Initial version
2013/12/31	Ioannis Fudos	Minor administrative corrections

Table of Contents

Section 1 Introduction

The last five years we have observed an unprecedented increase of the use of social media. In 2013, around 70% of adults use social media of some kind, whereas 40% use multiple social networking sites. Today, smartphones with georeference capabilities (GPS) are commodity hardware, thus enabling users to share besides multimedia (text, images, videos), their physical location. In this context we have studied the role of location based social networks such as four square for social-based promotion of SMEs.

Foursquare is a location-based social network through which users can connect and befriend other users. Moreover, it enables users to *check-in* at a location (called *venue*), let their friends know where they are, browse through the history of venue visits for themselves and their friends. Location sharing is performed through the Foursquare social network, but it can be propagated to friends through other social networks such as Facebook and Twitter.

Location-sharing is the main feature that makes foursquare a powerful SME promotion tool. Such location-based techniques can be enhanced with reward schemes, commonly used in online games, to make the process more interesting for the user. Every user that checks-in collects points and may receive badges or become a mayor for a certain location. A user is appointed as mayor if he/she has the largest number of check-ins during a period of 60 days by taking into account a single check-in per day. A venue has only one mayor at each time instance. A user may have multiple mayor appointments at the same time.

We have studied the characteristics of foursquare and the behavior of the users in Ioannina. More specifically we have:

- collected data for venues at Ioannina.
- found the mayor for every venue and then we have built a bi-partite graph for visualizing the venue-mayor relationship.
- collected the "friends" info for every mayor and we have built the corresponding graph representing their social network.
- detected homophily among mayors and we have tried to predict friendships based on the type of venues where each one is mayor.
- performed automatic detection of popular areas based on venue check-ins. Detected popular venue types and associated them with popular areas.
- analyzed data collected from venue check-ins, mayors, venue types, proximity of venues so as to statistically verify prediction strategies that may promote SMEs by increasing the number of check-ins (and therefore the number of clients).

- We have reported on the development of this innovative tool suite and the implementation issues that we have encountered.

Section 2 Related Work

In this section we present prior approaches to analyzing data collected from location-centric social networks.

In [3] the authors study location sharing services from social networks such as Foursquare and Gowalla. Due to the limitations set by the social network APIs the authors extracted check-ins from published twitter check-ins. The authors have considered 22 million check-ins and reported on a quantitative assessment of human mobility patterns by analyzing spatial, temporal, social and textual aspects associated with these footprints. The have used three statistical properties: user displacement, radius of gyration and returning probability.

[5] focuses on Foursquare by collecting data through Twitter by 700k users in a period of 100 days. They consider the spatial arrangement of data (the check-in locations) and their temporal relation (when check-ins occurred) to detect meaningful spatiotemporal patterns. Such patterns may suggest daily user activity, weekend user activity, sequence of tasks carried out by users. With appropriate statistical processing this may be an invaluable asset for developers of tools targeting web-based promotion of SMEs.

Section 3 Data Collection

The first step in the analysis is the data collection phase. To acquire the required data, we used the API (Application Programming Interface) offered by Fourquare. The Foursquare API gives to the programmers access to the world-wide places databases of Foursquare. It also provides the ability to interact with Foursquare users.

In particular, Foursquare API consists of four components: (a) the Core API, (b) the real-time API, (c) the venues API and (d) the Merchant API. The *Core API* provides users with all the functionality of a mobile application and website. They can check in, view their history, see where their friends are, create tips and lists, search for and learn about venues and access specials and recommendations. The *Real-time API* notifies venue managers when users check in their venues and developers when their users check in anywhere. The *Venues Platform* allows developers to search for places and access a variety of related data, such popularity, tips, photos and addresses. Finally, the *Merchant Platform* allows developers to write applications that help venue owners to manage their Foursquare presence.

In this report, we report the results of using the Core and Venue APIs to collect information about venues in the city of Ioannina, Greece and using it for social-based analysis.

Some Implementation Details

Foursquare as most social networks APIS uses the oauth2 protocol for authentication.

As a first step, we created and registered with Foursquare a web application. Data were collected in JSON format and stored for further processing. For this task, Python scripts were used.

Foursquare as most social networks APIS uses the oauth2 protocol for authentication. For this task we used the oauth library of *pyfoursquare library*. We also used the *pyfoursquare library* for querying, parsing and storing the collected data in a dictionary data structure.

API Restrictions

For security and reliability, Foursquare imposes some restrictions on data collection.

In particular, Foursquare sets specific limits on the volume of data that can be collected. Specifically, an application can make up to 5,000 userless requests per hour to venues or endpoints and up to 500 userless request to all other endpoints. Authenticated requests can make up to 500 requests per hour per oauth token. For example, if an application has

3 connected users, it can make at most 500 requests for each of them for a maximum of 1,500 authenticated requests per hour.

Furthermore, each request returns only 50 venues.

In terms of the type of data that can be collected, for privacy reasons, it is not possible to retrieve the users checkins. This restricts the form of analysis that can be performed.

Collecting venues and mayors in Ioannina

To overcome the limitation of 50 venues per request and collect all venues in the city of Ioannina, we submitted multiple requests with different geographical coordinates. In particular, we divided the city in small regions with different radii, namely, 150 m for the city center and 1.2km for the outskirts of the city.

We collected 2,023 venues. For each venue, we also retrieved additional information including the venueID, name, coordinates, category, number of checkins and information about its mayors. The mayor of an event is the user with the largest number of checkins for a period of two months.

For the mayors of the collected events, we retrieved related information including the list of their friends.

Section 4 Data Analysis

In this section, we present the results of our analysis. The analysis provides useful information about the venues and the mayors. Mayors are the users that visit a venue the most and identifying those users offers insight about what makes a venue popular.

Overall, we collected 2,023 venues in the city of Ioannina, of which 1,492 had a mayor. There were 422 distinct mayors. The total number of checkins was 106,277. For the mayors, we also collected all their friends.

The Mayors-Venues Graph

First, we studied the distribution of mayors to venues. To this end, we construct the bipartite graph between mayors and venues. The nodes of the graph are the venues and the mayors. There is an edge between a node u_1 representing a mayor m and a node u_2 representing a venue v, if and only if, m is the mayor of u.

The graph is shown in Figure 3. The distribution is highly skewed. Most of the users hold 2 mayorships and a small number more than 10 mayorships.

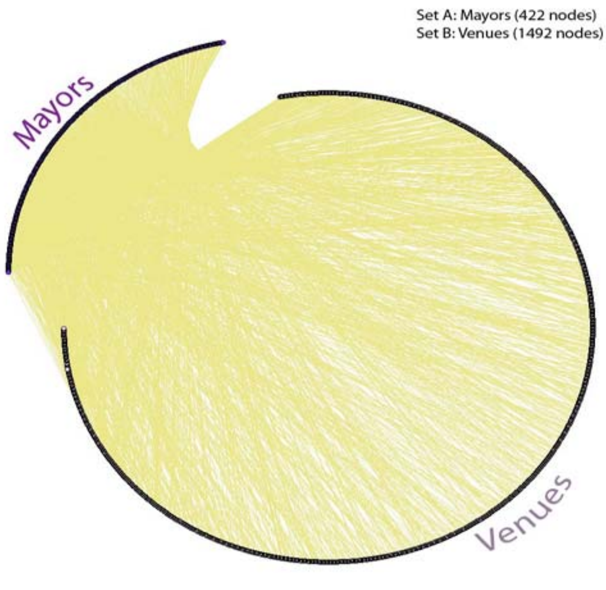

Figure 4.1: The Mayors-Venues Bipartite graph

The Friendship Network among Mayors

For the distinct mayors, we collected all their friendships. Then, we looked into whether mayors were friend with each other. This can be seen as an indication of whether the users that frequent places know each other, forming a small group of social media fans.

The network of friendships amongst mayors is shown in Figure 4.2. The nodes are the mayors and there is an edge between them, if and only if, they are friends.

Most of the mayors, 358 of the 422, have at least one other mayor as their friend. A mayor has on average 13.2 other mayors as friends which shows that mayors know a relatively large number of other mayors. Mayors are also tightly connected with each other as the average distance between any two of them is 3.146. The average clustering coefficient of the mayor friendship network is 0,383 and the diameter is equal to 8. There is a total of 2,379 edges in the network.

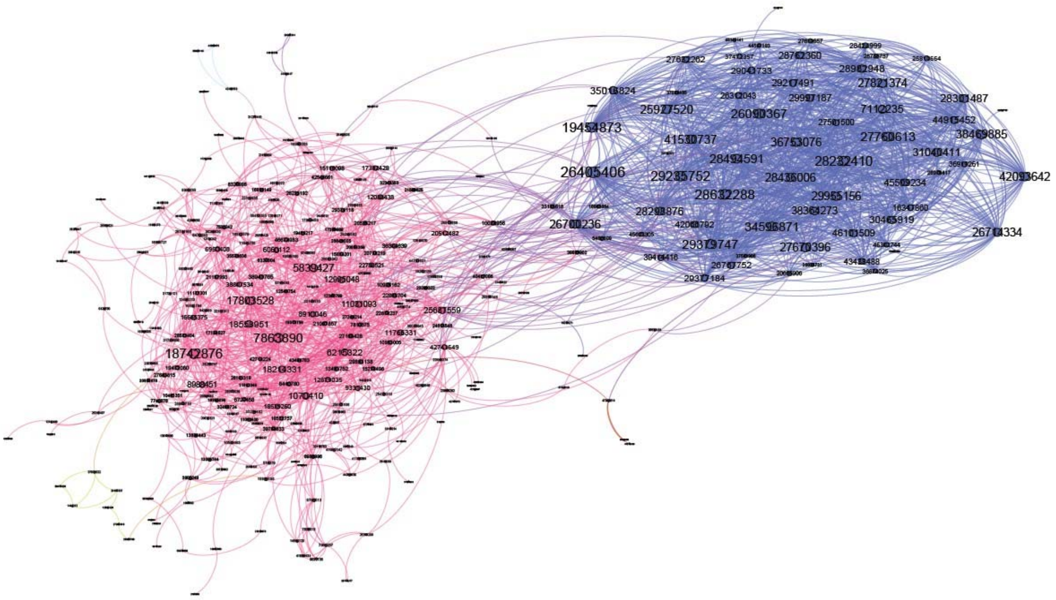

Figure 4.2: The friendship network for mayors

Looking closely into the mayor friendship network, we can distinguish between two communities of mayors, depicted with blue and pink in Figure 4.2. The blue community involves a smaller number of mayors than the pink community but the blue community is a very dense one, since all its members are tightly connected to each other, with each user having almost the same number of friends in the community. The pink community is larger with a less uniform in-community degree distribution. In particular, there are three nodes with a large number of connections, with the other nodes having just 1 or 2 connections.

The cumulative degree distribution for the friendship network amongst mayors follows a power law with exponent a = 1.38 as shown in Figure 4.3 (left).

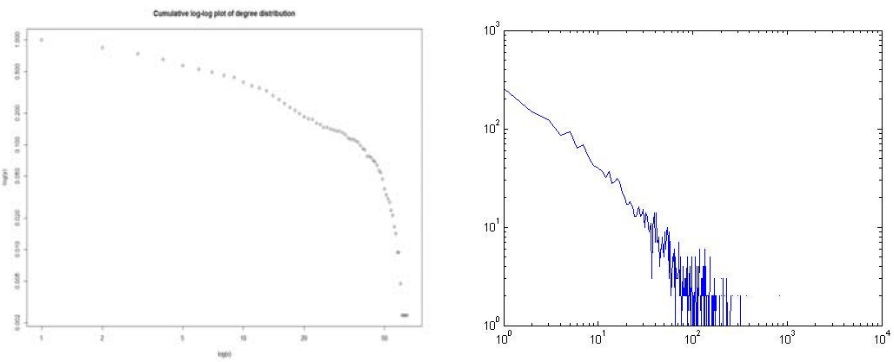

Figure 4.3: (left) Mayor Friendship CDF (right) distribution of checkins per venue (log-log scale)

The cumulative degree distribution for the friendship network amongst mayors follows a power law with exponent a = 1.38 as shown in Figure 4.3 (left).

Distribution of Checkins

Next, we study the popularity of venues measured by the number of checkins they have received. As expected, the popularity of venues is highly skewed, with most venues being moderately popular and a small number of them being very popular. The distribution of checkins per venue is shown in Figure 4.3 (right).

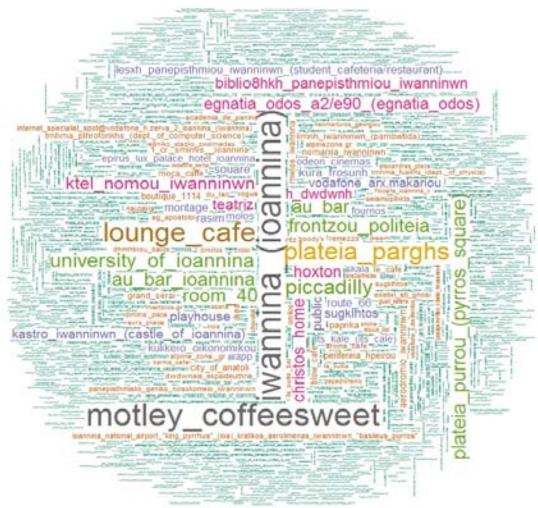

Figure 4.4: Wordcloud of the venues in Ioannina

In Figure 4.4, we show the wordcloud of the venues based on their checkins: the size of the venue-word is proportional to the number of its checkins. The exact number of checkins for the most popular ones is depicted in Table 4.1.

Popular Venue Categories

When a venue is created, the user may specify one or more categories for the venue. In Table 4.2, we present information about 8 general categories. The most popular category is by far the Food category, having the most venues and checkins. This is consistent with related studies in other cities in the world. The next most popular venues in terms of checking belong to the Nightlife category.

Table 4.1: Most popular venues in Ioannina

Venues	Checkins
Ioannina	1857
Motley Coffeesweet	1772
Lounge Cafe	1528
Pargis Square	1381
Piccadilly	1120
University of Ioannina	1063
Pyrros Square	1032
Au bar Ioannina	1006
Room 40	1005
Frontzou Politia	994

Popular Areas in Ioannina

Using Google Maps, we make a map of the city to highlight the areas that attract the most checkins.

Table 4.2: Popularity of Venue Categories

Category	Venues	Checkins
Food	463	32489
Nightlife	195	15372
Buildings	414	14169
Parks and Outdoors	151	11875
Shops	362	10510
Education	119	8407
Transportation	85	5643
Arts and entertainment	110	3927

In Figure 4.5, the number in the cycles indicate the number of checkins and the color their distribution. Most of the popular venues are in the center of the city.

Next, in Figures 4.6, 4.7 and 4.8, we show the distribution of venues by category. It is evident that specific areas in the city attract specific categories of venues. This is useful information for both users and venue owners.

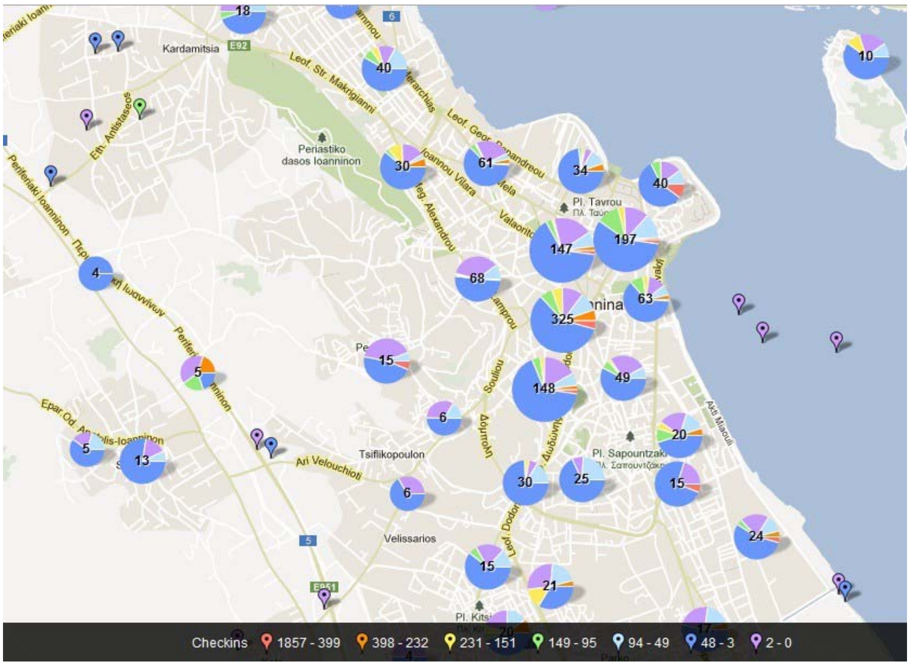

Figure 4.5: Distribution of checkins

In Figure 4.6 (left), we see the distribution of Food-related venues. Most of them are in the center and more particularly they are concentrated on specific streets and areas of the center. The same holds for Nightlife-related events as shown in Figure 4.6 (right). Again, the center attracts the most such events both in terms of number of venues and in terms of popularity, i.e., number of checkins. Furthermore, although for Food, there are a few venues in the outskirts in the city, Nightlife seems to be found in the center of the city almost exclusively.

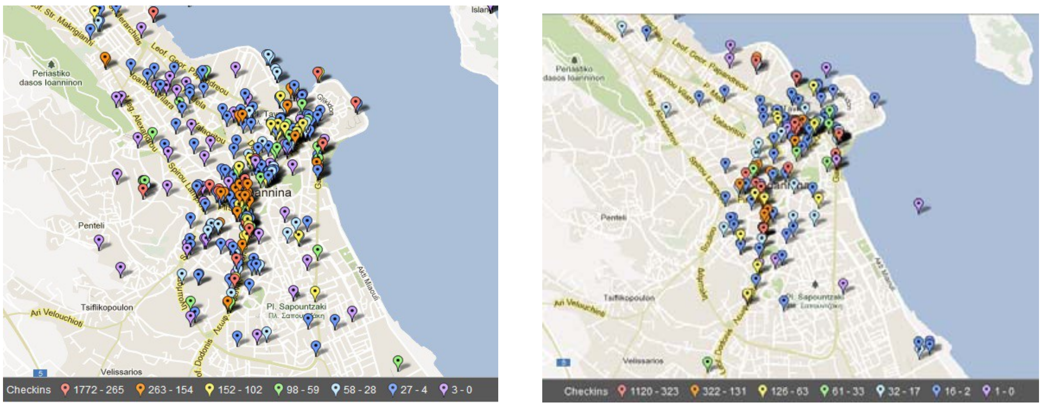

Figure 4.6: Distribution of checkins per category: (left) Food and (right) Nightlife

In Figure 4.7, we see the distribution of venues for the Building category, Figure 4.7 (left), and the Education category, Figure 4.7 (right). Buildinds are spread all over the map. There are many such venues, but of them have a small number of checkins. This is because venues in this category are mainly homes created by their owners. There are a few exceptions, most notably the venues that correspond to the two Hospitals (one in the north part of the city and the other at the University at the south-west). For education, checkins are divided in two areas, the center (corresponding to schools) and the University of Ioannina area at the south-west corner.

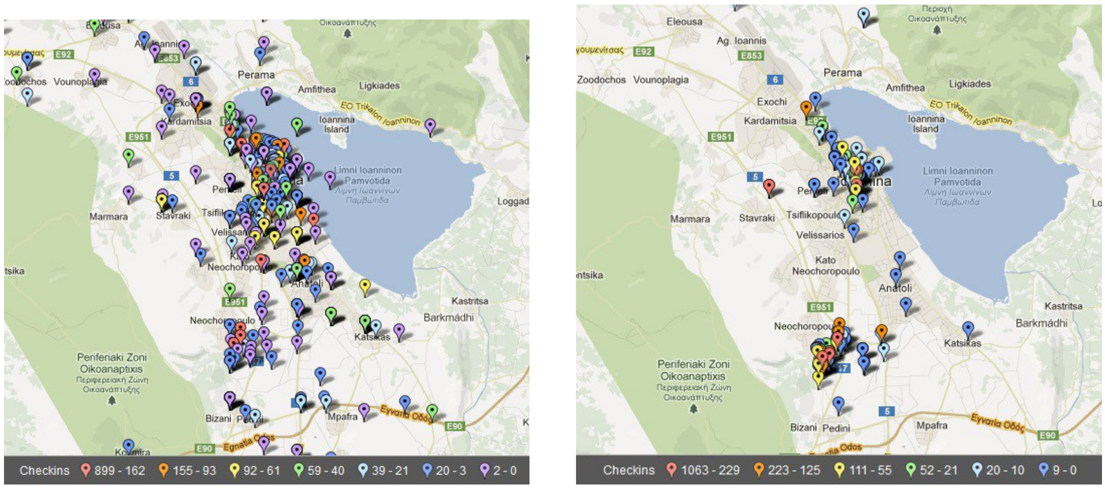

Figure 4.7: Distribution of checkins per category: (left) Buildings and (right) Education

Finally, in Figure 4.8, we see the distribution of venues for the less popular Transportation (airplanes, buses, etc), Figure 4.8 (left), and Arts&Entertainment, Figure 4.8 (right), categories. The Transportation events are very few, since Ioannina is a rather small city. Arts&Entertainment are scattered, corresponding to the location of various museums and other points of interest.

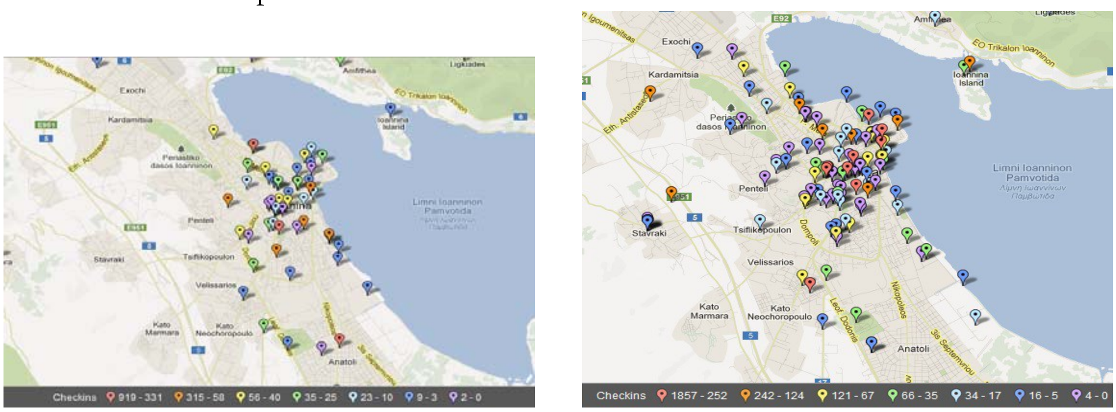

Figure 4.8: Distribution of checkins per category: (left) Travel and (right) Arts&Entertainment

Link Prediction

Finally, we also performed link prediction based on the collected data. In particular, we predict friendship between mayors based on the category of the events they have checkins.

In particular, we used the Jaccard similarity [6] to compute a similarity score for all pair of mayors. We assume that two mayors are more similar, if they are mayors in venues belonging to the same category. We compute for each pair the rate of same-category mayorships over their mayorships. We call this rate, score.

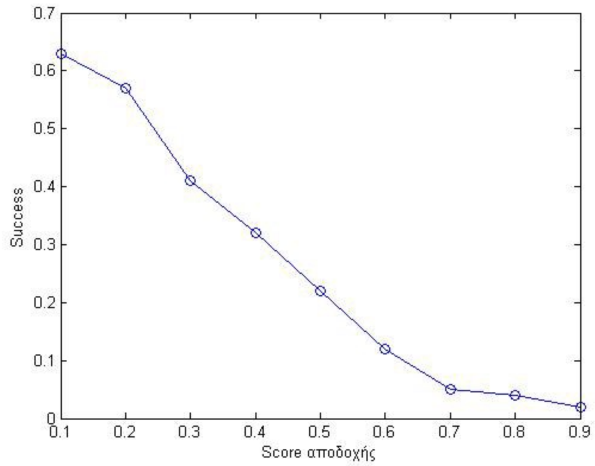

Figure 4.9: Prediction rate

In Figure 4.9, we plot the success rate, if we predict that two mayors are friends if theirs cores is above a threshold value of 0.1 up to 0.9.

Section 5 Implementation issues

We have used several APIs, libraries and applications for developing the tools that perform collection of data, statistical processing and analysis:

- RStudio: for statistical computing and for building the word cloud.
- Gephi: For graph building, analysis, manipulation and visualization
- Matlab: for further computations and graphs (clustering, spatiotemporal processing)
- Batchgeo [2]: For creating maps
- Foursquare API [4] and python wrapper for Foursquare [7]: for collecting check-ins directly from foursquare.

The first issue that has not been addressed effectively in the literature is overcoming the Foursquare API limitations regarding the volume of data and the type of data that may be captured. The limitation on data volume has made it difficult to collect data over a long period of time for the same venues/users. The type of data made available may restrict significantly the type of analysis that we can perform.

More technical issues include the API interface that requires authentication and data encryption/decrtyption. To this end we have used pyfoursquare python wrapper that facilitates this process if it is enhanced by a process that changes the token every few queries that we have developed for this purpose.

A major technical challenge was storing data so that they can be retrieved efficiently. We have used hashing, python dictionaries and external storage using the pickle library for serializing and de-serializing a python object structure.

Section 6 Conclusions

We have reported on the deployment of tools for analyzing and predicting the behavior of users in location based social media. This information is an important asset for the promotion of SMEs.

In our example cases we have determined that:

Established relations between venues and mayors in Ioannina. Most mayor in Ioannina have more than one mayorships. Only a few mayors have more than 10 mayorships.

Regarding the mayor friend network, we have observed that there are two large communities with distinct characteristics.

We have detected and visualized the most popular venues and venue categories: The largest number of check-ins are in selected restaurants and cafés. The most popular venue category is "Food".

We have predicted the existence of friendship between two mayors based on the type of venues with very good success rate. This is a useful feature for promoting SMEs with advertising or offers targeted to specific users.

We have visualized in maps the distribution of venues/check-ins around Ioannina. More venues are in specific areas in the city center or adjacent to the city center. The distribution of venues may change according to season (summer, winter) or based on current events or trends.

References

[1] About foursquare. http://foursquare.com/about. Referenced January, 2013

[2] BatchGeo. http://batchgeo.com/, January, 2013

[3] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring Millions of Footprints in Location Sharing Services. In Proc. of the 5th Int'l AAAI Conference, on Weblogs and Social Media (ICWSM), Barcelona, July 2011.

[4] Foursquare API. https://developer.foursquare.com/.Referenced January, 2013

[5] A. Noulas, S. Scellato, C. Mascolo, M. Pontil. An empirical study of geographic user activity patterns in foursquare, In Proc. of the 5th Int'l AAAI Conference on Weblogs and Social Media (2011), pp. 570-573

[6] Liben-Nowell, David and Kleinberg, Jon. The link-prediction problem for social networks. J. Am. Soc. Inf. Sci. Technol., May 2007, 58, 7

[7] pyFoursquare. https://github.com/marcelcaraciolo/foursquare