

SOCIAL METRICS APPLIED TO SMART TOURISM

Ofelia Cervantes*, Ernesto Gutiérrez, Francisco Gutiérrez, J. Alfredo Sánchez

Universidad de las Américas Puebla
Ex-Hacienda Santa Catarina Mártir, Cholula
72810 Puebla Mexico

(ofelia.cervantes, ernesto.gutierrezca, francisco.gutierrez, alfredo.sanchez)@udlap.mx

KEY WORDS: Smart City, Smart Recommendation, Social Network Analysis, Graph Processing, User Model, Object of Interest Model

ABSTRACT:

We present a strategy to make productive use of semantically-related social data, from a user-centered semantic network, in order to help users (tourists and citizens in general) to discover cultural heritage, points of interest and available services in a smart city. This data can be used to personalize recommendations in a smart tourism application. Our approach is based on flow centrality metrics typically used in social network analysis: flow betweenness, flow closeness and eccentricity. These metrics are useful to discover relevant nodes within the network yielding nodes that can be interpreted as suggestions (venues or services) to users. We describe the semantic network built on graph model, as well as social metrics algorithms used to produce recommendations. We also present challenges and results from a prototypical implementation applied to the case study of the City of Puebla, Mexico.

1. INTRODUCTION

A smart city is characterized by a holistic approach in the use of information and communication technologies for improving urban services in at least one of six dimensions: people, government, economy, mobility, environment and living (Anthopoulos et al., 2015). In this context, the tourism sector can be classified as an overlapping subproblem of mobility and economy dimensions since tourism has the potential of enhancing the urban competitiveness of the city (Singhal et al., 2013) and represents a challenge in terms of mobility in certain periods of the year.

The widespread range of venues and services provided by a smart city makes it difficult for tourists to select an appropriate venue to visit according to their preferences. Tourists typically follow the “closest next” strategy to visit a point of interest (POI) or to use a service. However, they may be also interested in understanding how POIs are semantically (or conceptually) related (Wolff and Mulholland, 2015).

Several approaches have been proposed to tackle the problem of automatically selecting, from a list of items, those that really contribute to satisfying the needs of end users. Approaches based on demographics or user profile models are oriented to exploit user features and preferences for filtering available choices. Key contributions of this paper include a user model and an object model based on linguistic features to represent user preferences and item characteristics, where items in our application domain refer to points of interest.

Modeling social and semantic networks using graphs has opened opportunities for exploring alternatives for implementing recommender systems. Social metrics, such as flow centralities that are calculated on graph-based models provide interesting measures to represent the semantic predominance of concepts featuring user’s preferences as well as item characteristics. We provide a detailed description of our model as well as the challenges to implement our approach as a complete recommendation strategy that can be integrated into recommendation systems.

The remainder of the paper is structured as follows: Section 2 presents related work. Then, Section 3 introduces the graph-

based model we propose to represent user preferences and item characteristics as a semantic network. Section 4 discusses selected graph algorithms used to calculate social metrics, particularly flow centralities. Next, Section 5 describes a prototype we developed in order to validate our approach. Performance challenges are discussed. Finally, in Section 6 we report the main results we have obtained thus far and discuss future work.

2. RELATED WORK

The ever increasing amount of information that has been accumulated in social media can be utilized in order to improve services in a smart city. The rich knowledge that can be extracted from social media can be used, for example, to enhance recommendation systems, or to improve citizen experience, or to generate novel services. In this paper, we focus on extracting knowledge to enhance recommendation systems, specifically, for a smart tourism application.

Recommendation systems that are based on knowledge use information about how selected items meet user needs (Bobadilla et al., 2013). In this type of system, the knowledge extracted is used to build a relationship between users and points of interest. In general, a recommender system comprises five fundamental components: user model, community (social network), object of interest model, recommendation algorithm and interaction strategy (Zanker and Jessenitschnig, 2009). Here, we present a brief overview of related work regarding knowledge-based recommender systems focused on the building components in which we have made a contribution.

2.1 User Modeling

In order to personalize recommendations, it is necessary to know information about each user. User models are representations of user needs, goals, preferences, interests, and behaviors along with demographic characteristics (Schiaffino and Amandi, 2009). Several user modeling approaches have been proposed, from typical weighted vectors to domain ontologies. In (Anand and Mampilli, 2014), for example, the authors define a user model based on

fuzzy logic and proposed an approach to infer the degree of genre presence in a movie by considering the tags assigned by users. In (Zanker and Jessenitschnig, 2009), the authors present a simple attribute-value pair dictionary to model the user through the explicit elicitation of user requirements. A richer user model is presented in (Eyharabide and Amandi, 2012), where the authors used a machine learning process to capture the user profile and context into a domain ontology.

Our work tries to balance between simple (Zanker and Jessenitschnig, 2009) and complex models (Eyharabide and Amandi, 2012) with the goal of having an efficient but still rich user model. Other works, like Cantador et al. and Moahedian et al. (Cantador et al., 2008, Movahedian and Khayyambashi, 2014), are similar to our proposed user model, since we use tags and keywords to build a lax ontology.

2.2 Recommendation Algorithms and Techniques

A wide range of recommendation algorithms and techniques have been reported in the literature. They vary mostly in data availability, recommender filtering type as well as user and object representations. Various methods have demonstrated acceptable performance, including: Bayesian networks (de Campos et al., 2010), nearest neighbors (Bobadilla et al., 2011), genetic algorithms (Hwang et al., 2010), neural networks (Bobadilla et al., 2012), clustering (Shinde and Kulkarni, 2012), association rule learning (Zanker and Jessenitschnig, 2009), and latent semantic features. More details on these and other methods can be found in (Bobadilla et al., 2013).

In this work, we rely on graph centrality metrics commonly used in social network analysis (Thovex and Trichet, 2013). We propose semantic social network analysis that integrates semantic methods of knowledge engineering and natural language processing with classic social network analysis. Advantages of semantic social network analysis include its knowledge foundation and its non-probabilistic nature. In contrast, one disadvantage is its computational cost. Enhancement techniques are thus needed in order to process graph centrality metrics more efficiently.

2.3 Information Extracted from Social Networks

With the success of emerging Web 2.0 and various social network websites, recommender systems are creating unique opportunities to assist people in finding relevant information when browsing the web and making meaningful choices. In (Chang and Chu, 2013), the author has proposed a novel approach for recommendation systems based upon data collected from social networks.

In the work of (Wang et al., 2013) the problem of recommending new venues to users who participate in location-based social networks (LBSNs) is studied. They propose algorithms that create recommendations based on past user behavior (visited places), the location of each venue, the social relationships and the similarity among users.

In the work of (Ye et al., 2010) the social and geographical characteristics of users and locations to research issues for offering location recommendation services for large scale location based social networks are utilized. They observed the strong social and geospatial ties among users and their favorite locations in the system via the analysis of datasets collected from Foursquare.

Similar to our work, in (Saiph Savage et al., 2012) the design of a more complete, ubiquitous location-based recommendation algorithm that is based on a text classification problem is investigated. The system learns user preferences by mining a person's

social network profile. The author also defined a decision-making model, which considers the learned preferences, physical constraints, and how the individual is currently feeling.

We can state that novel approaches rely mainly on the fusion of information inferred from a user's social network profile and other data sources (e.g. mobile sensors). In this sense, it is necessary to develop new strategies that produce recommendations from rich but still incomplete information.

3. GRAPH MODEL

Our approach is based on a graph representation of users and points of interest linked through concepts (denoted as terms). Figure 1 shows the graph model where every node falls into one of three categories: *User*, *Term* or *Point of Interest*, whereas every edge represents the semantic relation between nodes: *Predominance*, *Similarity* or *Friendship*. Every *Term* node of the graph in Figure 1 acts as a semantic descriptor of both *Users* and *Points of Interest*. In other words, every user and every POI are correspondingly described by the terms linked to them. In general, users are described by their tastes, preferences, and interests (user model) whereas POIs are described by tags and keywords (here we refer to POIs as object model). In this manner, when a term is shared between a user and a POI, it shows the possibility that the user could be interested in that particular POI, even though the POI has never been seen or rated by the user.

A graph-based representation allows us to apply graph algorithms (e.g. centrality metrics) to discover topological features, key relationships, and important (prestigious) nodes. Then, with these features we can make relevant recommendations to users, such as suggesting friends or venues. Therefore, the foundation of our recommender system relies on a knowledge base constructed from both a **user model** (see section 3.3) and an **object model** (see section 3.4).

In order to construct the user and object models, we applied a linguistic analysis over user and object text descriptions. Basically, we conducted pre-processing (removal of stopwords and selection of most descriptive words) and statistical linguistic analysis (using weighting schemes: tf-idf and okapi BM25) to define a bond between text descriptions and semantic relations represented in the graph (see section 3.2). It is possible to obtain user and object descriptions from social networks (Facebook, Twitter, Foursquare), web pages (Wikipedia, web search results, etc.), human experts contributions or other textual resources.

3.1 Weighted Graph Definition

Formally, we define a weighted graph $G = (V, E, f_E)$ where $V = \{v_1, \dots, v_n\}$ is a set of vertices, $E = \{e_1, \dots, e_n\} \subset \{\{x, y\} \mid x, y \in V\}$ is a set of edges, and $f_E : E \rightarrow \mathbb{R}$ the function on weights for every edge. In our recommender system, $V = U \cup T \cup O$ where U is the set of users, T is the set of terms, and O is the set of objects of interest. And $E = P \cup S \cup F$ where P is the set of predominance edges, S the set of similarity edges, and F is the set of friendship edges. Function f_E is adapted according to each type of edge. For instance, we can obtain the sub-graph of users as $G_U = (U, F, f_F)$ (see User layer in Figure 1), the sub-graph of objects as $G_O = (O, S, f_S)$ (see Points of Interest layer in Figure 1), and the sub-graph of user and object profiles as $G_{U \cup T \cup O} = (U \cup T \cup O, P, f_P)$.

3.2 Semantic Relations

In order to build the semantic relations of the graph, it is necessary to obtain text descriptions of users and POIs. As a result,

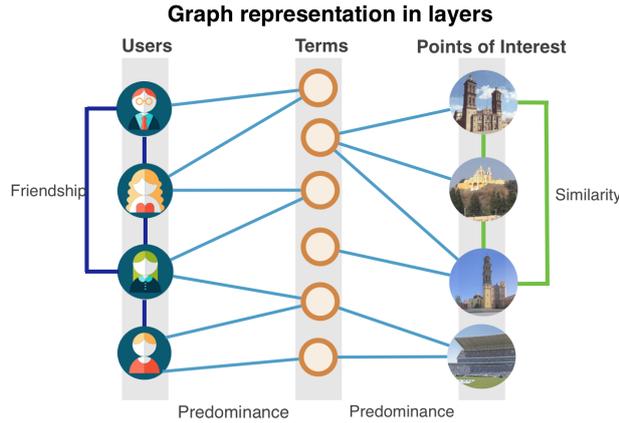


Figure 1: Graph model. Node layers: User, Term, and Points of Interest linked through Friendship, Predominance and Similarity edges.

we have two collections: the users text collection (UTC) and the POIs text collection (POITC), where each text description is considered a document D in a vector space model.

We define three types of semantic relations (edges of the graph): predominance, similarity, and friendship. Each semantic relation links different types of nodes and has a different weighting function. Predominance is the edge between a user or an object and a term, similarity is the edge between two objects and friendship is the edge between two users.

Predominance is the semantic relation between a term and a user or an object. A term acts as a descriptor of users and POIs. We define a weighting function over the edge of predominance based on linguistic analysis. We apply the Okapi BM25 ranking function (Robertson and Zaragoza, 2009) to each independent document collection (UTC and POITC) using Equation 3.

In Equation 3, $pred$ is the predominance of the term T in document D , I_{doc} is the number of indexed documents (size of collection), T_{doc} is the number of documents containing term T , TF is the term frequency relative to document D , DL is the document length, $avgDL$ is the average document length among the entire collection, K and B are free parameters (usually $K = 1.2$ and $B = 0.75$).

$$TF_{Normalized} = \left(\frac{TF * (K + 1)}{TF + K * ((1 - B) + B * \frac{DL}{avgDL})} \right) \quad (1)$$

$$IDF = \log_{10} \left(\frac{I_{doc} + 0.5}{T_{doc} + 0.5} \right) \quad (2)$$

$$pred_{(D,T)} = TF_{Normalized} * IDF \quad (3)$$

Similarity is the semantic relation between two POIs. This measure indicates the degree of affinity between POIs. We apply the cosine similarity measure (Equation 4) to obtain this value. The Similarity is calculated after the predominance, since it relies on shared terms. Then, every object is a vector of predominances as shown below Equation 4.

$$Similarity_{(A,B)} = \cos \Theta = \frac{A \cdot B}{\|A\| \|B\|} \quad (4)$$

$$POIA = [pred_{(A,T_1)}, pred_{(A,T_2)}, \dots, pred_{(A,T_n)}]$$

$$POIB = [pred_{(B,T_1)}, pred_{(B,T_2)}, \dots, pred_{(B,T_n)}]$$

In Equation 4, the similarity between POI A and POI B is determined by the weights of the terms they have in common. In this manner, a high similarity value indicates a higher semantic correspondence between POIs.

Friendship is the semantic relation between two users. This measure indicates the degree of affinity between two users. Our current model does not distinguish between close friends, friends or acquaintances. Therefore, the users' sub-graph is only a friend-of-a-friend (FOAF) node-link type.

3.3 User Model

As part of the graph-based representation, users are defined as sub-graphs. A user model is composed of two sub-graphs: user profile $G_u = (u, P, f_P)$ and user FOAF network $G_U = (U, F, f_F)$. Figure 2A shows the user profile network, whereas 2B illustrates the user's FOAF network.

In the user profile sub-graph, each user is linked to a set of terms that indicate tastes, preferences and interests. Tastes are general inclinations of user towards some entities and they are generally expressed with actions such as likes (e.g. Foursquare, check-ins and Facebook likes *football, beer, steak, coffee, etc*). Preferences are user inclinations towards taste features. Preferences are more fine-grained than tastes and are usually expressed in users' reviews and ratings (e.g. starred reviews: *I like the double espresso, I don't like diet soda*). Interests are defined as contextual user inclinations or intentions (e.g. *I want to try Chinese food, I'm going to watch a minions movie*).

$$f_P = \begin{cases} pred_{(U,T)} & \text{Case A} \\ 1 & \text{Case B} \\ \#stars - 2 * \frac{1}{3} & \text{Case C} \end{cases} \quad (5)$$

Our scheme to weight edges within a user profile is indicated in Equation 5. Case A occurs when only text descriptions are used; this means that terms are weighted according to the Okapi measure ($pred$, as shown in Equation 3). Case B occurs when explicit likes are found in Foursquare or Facebook. Case C occurs when terms extracted from starred reviews are used to describe a user. In addition to Equation 5, we use a threshold value to limit the number of terms connected with a given user. In fact, we use the first quartile as threshold value. An example of user profile is shown in Figure 2A, where, it is possible to notice that a user likes football, rock and coffee, and is likely that the user is a student. In user friendship networks, as mentioned earlier, there are no differences among friendship types. Then, in FOAF network all weights are equal to 1 ($f_F = 1$). A user FOAF network is shown in Figure 2B.

3.4 Object Model (Points of Interest)

In this section we generalize the notion of Point of Interest as Objects of interest. Objects of Interest are sets of items that can be of potential interest to a user. Depending upon the application, objects can be of different grain size. For instance, they can

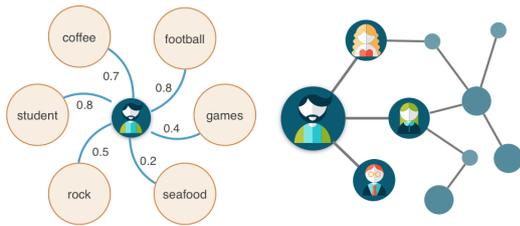


Figure 2: a) User profile and b) User FOAF Network

be coarse-grained as a point of interest (POI) or fine-grained as items inside places. An object is described by the category to which it belongs (e.g. Cathedral is a Christian Church). It is also described by tags designed by users or keywords found in object's description. The object model consists of two sub-graphs: object profile $G_o = (o, P, f_P)$ and object similarity sub-graph $G_O = (O, S, f_S)$. An object profile is built with data gathered from Foursquare, Wikipedia and results from web searches. The weights of edges that link objects and terms are calculated using the predominance formula shown in Equation 3. This means that the weight function on edges is $f_P = pred_{(O,T)}$.

3.5 User Global and Local Network

In order to apply social metrics (centrality measures) and relate them to pertinent recommendations, we defined two networks from user perspective: a user global network (UGN_u) and a user local network (ULN_u). By user global network we refer to the whole graph (all nodes: users U , terms T and objects O and all edges: similarities S , predominances P and friendships F) centered in current user. Therefore, $UGN_u = (U \cup T \cup O, S \cup P \cup F)$ (see Figure 1). Whereas user local network is the sub-graph defined by current user node u , term nodes adjacent to user T_u and object nodes adjacent to terms node O_T linked through predominance edges from user P_u and from objects P_O . Hence, $ULN_u = (u \cup T_u \cup O_T, P_u \cup P_O)$ (see Figure 3). It is important to highlight the difference between user global and local networks, since it will lead to different semantics interpretations when calculating centrality measures over them.

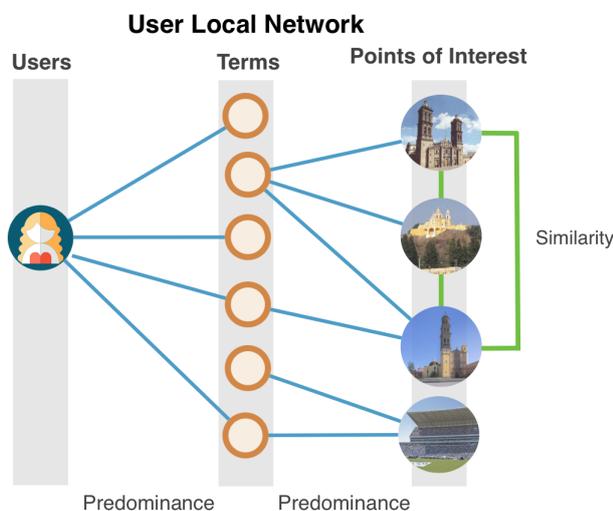


Figure 3: User Local Network

4. CENTRALITY MEASURES AND RECOMMENDER ENGINE

Centrality measures have been used extensively to exploit networks and discover the relevancy of nodes in a graph. In social network analysis (SNA) graph centralities are used to identify the most important persons, communities and detect strange behaviors in the network. However, given the popularity of social networks, people have increased their interaction not only to meet people and friends but also to search things they like, express their opinions, and find points of interest and objects of interest. These spatio-temporal interactions can also be represented in a graph, thus, an accurate user profiling representation can give us a great deal of insight about the user behavior. Because of these approaches in utilizing graph measures we have explored the use of centralities to take advantage of the topological structure of our users' global network and flow centrality to get the most of our weighted global and local networks in terms of a recommender engine.

4.1 Centrality Algorithms

Centrality in graphs is widely used to measure the importance of a node in a graph, especially in SNA (Le Merrer and Trédan, 2009). Our recommender engine implements these centralities to measure the relevance of people in the social network. Some centrality measures like closeness and betweenness are based on the calculation of the shortest distance to reach all other nodes in the graph. Our algorithms to calculate centralities are applied to the network of persons so we can infer the most popular nodes (degree), the capacity of a node to reach any other in the network (closeness), and to identify the leaders interconnected within a neighborhood in the graph (betweenness) (Newman, 2005). Degree centrality is a measure that counts the direct relationships a node has, and thus, the nodes that are in direct contact. Closeness is defined as the inverse sum of the shortest paths between any two nodes and betweenness is defined as the number of shortest paths from all vertices to all others that pass through that node. Centrality measures are calculated over the network at a topological level given a scale-free graph of persons. Thus, these measures are not exploiting our weighted graph, they are applied only at a social-network level. Terms and objects of interest can be seen as sub-graphs of the global network that can be exploited by using flow-based centrality measures.

4.2 Flow-Centrality Algorithms

We are using flow centralities (Newman, 2005) to measure the betweenness, closeness, and eccentricity between the objects of interest, terms, and people profiles. Flow centralities allow us to exploit the semantic relationships between the user and the profiles of the objects of interest. Flow centralities reveal the most relevant nodes in the graph given their weights. For instance, given a set of terms associated to a user profile, we can better understand user preferences and give a better recommendation.

4.2.1 Flow Betweenness: In SNA, betweenness is one of the most common referenced centralities. Let m_{jk} be the amount of flow between vertex j and vertex k which must pass through i for any maximum flow. Flow betweenness of vertex i , (see Equation 6) as defined in (Freeman et al., 1991), is the sum of all m_{jk} where i, j and k are distinct and $j < k$. The flow betweenness is therefore a measure of the contribution of a vertex to all possible maximum flows. A node with a high flow betweenness centrality has a large influence in the network because of the flow that passes through it. Due to the relevance of a node with high

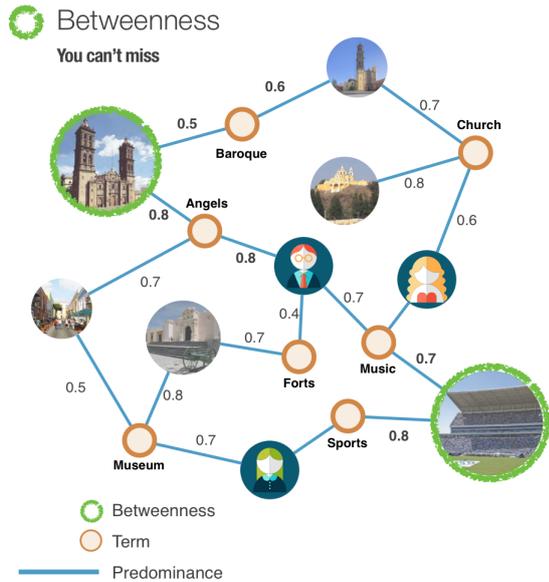


Figure 4: “You can’t miss” as a result of computing flow betweenness.

betweenness, in our recommender model, a node with high betweenness should be recommended as the things the user cannot miss (see Figure 4).

$$F_b(i) = \frac{\sum_{j < k}^n m_{jk}(x_i)}{\sum_{j < k}^n m_{jk}} \quad (6)$$

4.2.2 Flow Closeness: Closeness is just a measure of distance and is defined as the inverse of the average distance to other vertices. A node with high flow closeness centrality has a fast communication within all the nodes in the graph. In equation 7, flow closeness is defined as the inverse sum of the max flow to every other resource. In our recommender model, elements with high flow closeness should be recommended to the user as things that could be interesting, because those weighted elements are close to the user profile (see Figure 5).

$$F_c(i) = \frac{1}{\sum_{j < k}^n m_{jk}(x_i)} \quad (7)$$

4.2.3 Eccentricity: On the other hand, eccentricity is the maximum distance taking into consideration the weighted paths of the network. Eccentricity lets us find the nodes that are far away from the most central node in the network. In Equation 8 eccentricity is defined as the maximum distance between pairs of nodes given their maximum flow in the network. In our recommender model, an item with high eccentricity should be recommended if the user has nothing left to do and would like to discover something different (see Figure 6).

$$F_e(i) = \max_{j \in v} \sum m_{jk}(x_i) \quad (8)$$

4.3 Graph recommendations

As we have shown, our recommender model relies on the continuous computation of predominance and similarity between items in the graph. As the graph evolves from interactions between the user and objects of interest, recommendations get more accurate

Table 1: Centrality-based recommendations.

Measure	Recommendation
Similar nodes	What else?
Flow betweenness	You can’t miss
Flow closeness	Also discover
Eccentricity	Be different

over time. However, in order to give recommendations, the computation of centralities is required. As shown in Table 1, we can recommend similar items if the user is asking “What else?”, then we can suggest items that are similar to the recommended item. Flow betweenness is used to recommend things the user “can’t miss” because of their relevance in the network. Flow closeness is used to recommend central items that the user “would like to discover”, whereas eccentricity is used to show items that are far away from the more central nodes in the network and could cause a “being different” impression.

5. IMPLEMENTATION

We have implemented the model discussed above as well as graph measures calculation in a weighted graph. We explored different graph databases and graph processing frameworks in order to select tools that could give us flexibility to calculate those metrics with ease when building a graph processing framework.

5.1 The Census framework

We defined an architecture (see Figure 7) based on the graph model discussed previously. Our graph processing framework (which we named “Census”) has been built with the Play framework¹ and is intended to have multiple instances of Signal/Collect while processing our graph in the Google Compute Engine². Census uses the Neo4j³ Graph database to store the graph. Neo4j provides flexibility to issue queries over the computed network

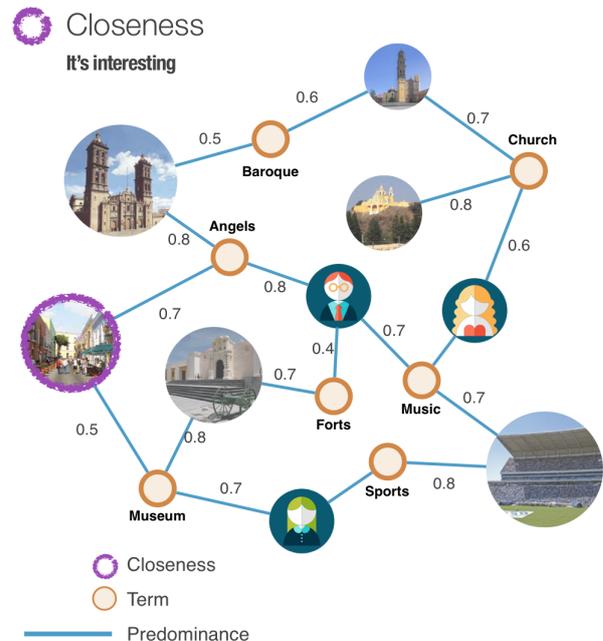


Figure 5: “Could be interesting” as a result of computing closeness.

¹<https://www.playframework.com/>

²<https://cloud.google.com/compute/>

³<http://neo4j.com/>

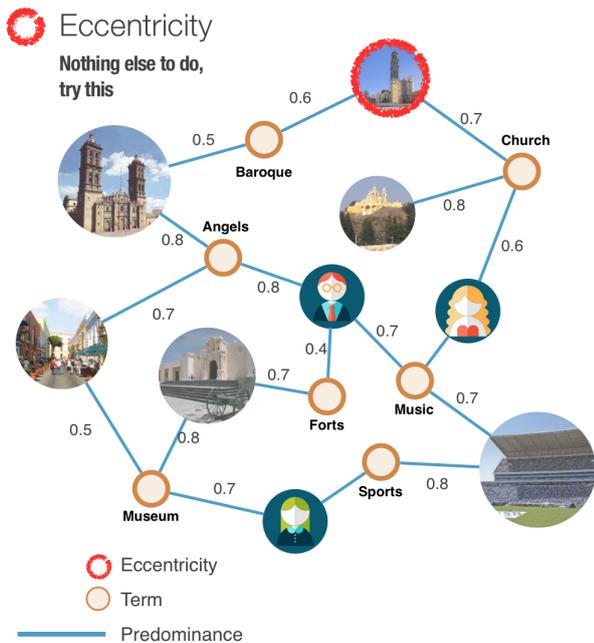


Figure 6: “Nothing else to do? Try this” as a result of computing betweenness.

through custom plugins that serve queries through a REST API. Census processes requests from Census Control which uses an orchestrator to administrate compute requests and instances of Census in the graph.

5.2 Proof of concept

With Census we explored a first approach to implement our recommender model. The graph database was populated with nodes of persons and points of interest from the city of Puebla, Mexico, then we selected documents from the Web to create a profile of the points of interest using our semantic approach described before. We created different users with their respective profiles setting them with random characteristics as weights in the relationships of the graph. We calculated similarities between those points of interest and the terms describing them. The result was a large graph database with approximately 3,000 nodes and 10,000

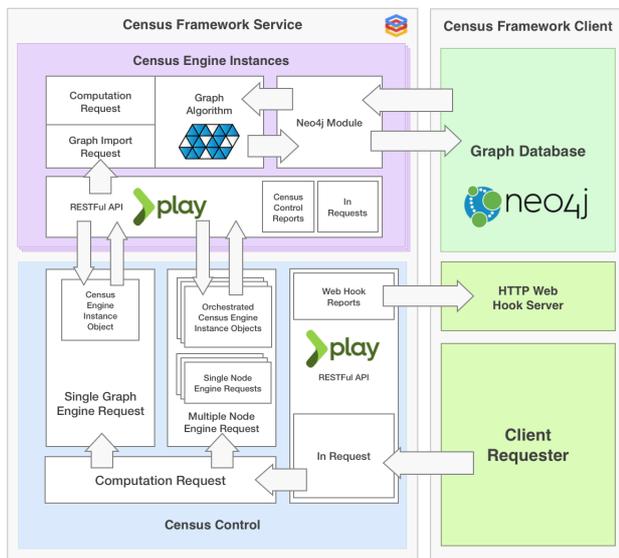


Figure 7: Architecture of Census, graph processing framework.

Table 2: Centrality computation results

Point of Interest	Flow		
	Betweenness	Closeness	Eccentricity
Centro Expositor	0.00	0.00	0.39
African	1.05	0.04	0.28
Museo Revolución	6.00	2.08	0.46
Convento Sta. Monica	2.21	0.73	0.46
Museo Amparo	0.00	0.00	0.65
Museo Regional Puebla	2.77	1.04	0.65
Fuerte Loreto	0.00	0.00	0.78
Casa Alfeñique	0.00	0.00	0.78
Catedral Puebla	7.13	1.01	0.31
Exconvento Calpan	0.00	0.00	0.40
Convento San Gabriel	0.00	0.00	0.84
Capilla Rosario	0.00	0.00	0.84
Fuerte Guadalupe	0.00	0.00	0.23
Plazuela Sapos	2.04	0.16	0.35
Barrio Analco	0.00	0.00	0.41
Paseo San Francisco	0.00	0.00	0.25
Barrio Artista	0.00	0.00	0.25
Estadio Cuauhtémoc	1.05	0.12	0.85

weighted relationships. Over the global network we calculated the graph measures to discover the relevant nodes in the graph.

5.3 Results

After running all the algorithms, we focused our attention on the local network of a particular user. Results of centrality computation are presented in Table 2. We can notice that higher centrality measure values allow us to suggest the most relevant points of interest for this user. For example we can see that “Catedral Puebla” is an element with high betweenness, which means that it is a relevant place in the city and that element should be recommended as “You can’t miss”. Another relevant element is the “Museo Revolución” because it shows the highest flow closeness. In the case of flow eccentricity, we can see the elements that are far away from user preferences giving the opportunity to explore new things and try something different.

6. CONCLUSIONS

We have presented a first approach to a graph-based recommendation model that takes advantage of social metrics and recommends points of interest to citizens and visitors of a smart city. The proposed model expresses the semantics of relationships that exist between users and points of interest through terms that define a profile for the items. This novel approach, using particularly flow centralities, considers semantic predominance of terms for defining and exploiting the relationships among user profile preferences as well as the descriptive characteristics of points of interest. Recommendations can then be extracted based on the knowledge represented in the graph.

In order to validate the recommendation model, a recommendation engine was implemented and has shown that interesting recommendations could be suggested to users, considering not only their preferences, but also taking into account suggestions coming out from preferences of other members of the social network related to them by the friendship relationship. The graph-based recommendation model also proposes to explore points of interest that are very different to user preferences, inviting them to explore new points of interest in the city.

The implementation of the recommendation engine is a challenging task because of the data volume and the complexity of required calculations to evaluate flow centralities and semantic predominance. This challenge not only has raised new questions

but also opened interesting opportunities for dealing with performance issues. Preliminary results were presented, showing that the use of social metrics in any real recommendation system must include specialized components for solving distributed and concurrent processing tasks. Even though nowadays there are advanced and efficient solutions for managing big data, adequate use of graph-based solutions for modeling social networks still remains as the core problem of a recommendation engine.

The framework presented in this paper was tested through a prototype that demonstrated the validity of our proposal. The recommendation engine is available through a REST APIs. These web services can be easily integrated into web or mobile apps. Application domains include intelligent tourism, (as in those described in this paper), as well as other areas of interest for citizens, such as administrative services in a smart city.

7. FUTURE WORK

The prototype will be extended and adapted to include specific information on the cities of Puebla in Mexico and Shanghai in China. In addition, different aspects could still be improved in the recommendation engine to contribute to enrich the user experience in a smart city:

- Incorporation of new semantic filters to propose lists of objects of interest; proposing for instance only the points of interest in the proximity of the user's location and considering the time when the user queries the recommendation system.
- In absence of explicit evaluation of user preferences, we will explore the integration of a Sentiment Analysis component as the one used in (Gutiérrez et al., 2015) to offer the possibility to add open comments and to evaluate automatically their polarity.
- Routes recommendation: from the list of recommended POIs, different alternative routes can be built. A prototype of a mobile application has been already developed for evaluating the interaction with users (Pedraza, 2015). The prototype relies on data from Foursquare and recommends points of interest based on ratings made by users. The integration with the recommendation engine needs to be completed.
- Performance evaluation and recommendation results validation as well as evaluation of current user interfaces with actual and potential users.

8. ACKNOWLEDGMENTS

Special thanks to our colleagues Christophe Thovex, Francisco Arámburo and David Báez, for their significant contributions in this work. This project was partially supported by CONACYT project no. 0192321.

REFERENCES

Anand, D. and Mampilli, B. S., 2014. Folksonomy-based fuzzy user profiling for improved recommendations. *Expert Syst. Appl.* 41(5), pp. 2424–2436.

Anthopoulos, L. G., Janssen, M. and Weerakkody, V., 2015. Comparing smart cities with different modeling approaches. In: *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp. 525–528.

Bobadilla, J., Hernando, A., Ortega, F. and Bernal, J., 2011. A framework for collaborative filtering recommender systems. *Expert Syst. Appl.* 38(12), pp. 14609–14623.

Bobadilla, J., Ortega, F., Hernando, A. and Bernal, J., 2012. A collaborative filtering approach to mitigate the new user cold start problem. *Know.-Based Systems* 26, pp. 225 – 238.

Bobadilla, J., Ortega, F., Hernando, A. and Gutiérrez, A., 2013. Recommender systems survey. *Know.-Based Syst.* 46, pp. 109–132.

Cantador, I., Fernández, M., Vallet, D., Castells, P., Picault, J. and Ribière, M., 2008. A multi-purpose ontology-based approach for personalised content filtering and retrieval. In: M. Wallace, M. Angelides and P. Mylonas (eds), *Advances in Semantic Media Adaptation and Personalization*, Studies in Computational Intelligence, Vol. 93, Springer Berlin Heidelberg, pp. 25–51.

Chang, C.-C. and Chu, K.-H., 2013. A recommender system combining social networks for tourist attractions. In: *Computational Intelligence, Communication Systems and Networks (CI-CyN)*, 2013 Fifth International Conference on, pp. 42–47.

de Campos, L. M., Fernández-Luna, J. M., Huete, J. F. and Rueda-Morales, M. A., 2010. Combining content-based and collaborative recommendations: A hybrid approach based on bayesian networks. *Int. J. Approx. Reasoning* 51(7), pp. 785–799.

Eyharabide, V. and Amandi, A., 2012. Ontology-based user profile learning. *Applied Intelligence* 36(4), pp. 857–869.

Freeman, L. C., Borgatti, S. P. and White, D. R., 1991. Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks* 13(2), pp. 141–154.

Gutiérrez, E., Cervantes, O., Báez-López, D. and Alfredo Sánchez, J., 2015. Sentiment groups as features of a classification model using a spanish sentiment lexicon: A hybrid approach. In: J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, J. H. Sossa-Azuela, J. A. Olvera López and F. Famili (eds), *Pattern Recognition*, Lecture Notes in Computer Science, Vol. 9116, Springer International Publishing, pp. 258–268.

Hwang, C.-S., Su, Y.-C. and Tseng, K.-C., 2010. Using genetic algorithms for personalized recommendation. In: *Proceedings of the Second International Conference on Computational Collective Intelligence: Technologies and Applications - Volume Part II*, ICCCI'10, Springer-Verlag, Berlin, Heidelberg, pp. 104–112.

Le Merrer, E. and Trédan, G., 2009. Centralities: Capturing the fuzzy notion of importance in social graphs. In: *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, SNS '09, ACM, New York, NY, USA, pp. 33–38.

Movahedian, H. and Khayyambashi, M. R., 2014. Folksonomy-based user interest and disinterest profiling for improved recommendations: An ontological approach. *J. Inf. Sci.* 40(5), pp. 594–610.

Newman, M. J., 2005. A measure of betweenness centrality based on random walks. *Social Networks* 27(1), pp. 39 – 54.

Pedraza, D., 2015. Rox: Sistema de recomendación basado en preferencias. Master's thesis, Universidad de las Américas Puebla, Ex-Hacienda Santa Catarina Martir, Cholula, Puebla.

Robertson, S. and Zaragoza, H., 2009. The probabilistic relevance framework: Bm25 and beyond. *Found. Trends Inf. Retr.* 3(4), pp. 333–389.

Saiph Savage, N., Baranski, M., Elva Chavez, N. and Höllerer, T., 2012. I'm feeling loco: A location based context aware recommendation system. In: G. Gartner and F. Ortog (eds), *Advances in Location-Based Services*, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, pp. 37–54.

Schiaffino, S. and Amandi, A., 2009. Intelligent user profiling. In: M. Bramer (ed.), *Artificial Intelligence An International Perspective*, Lecture Notes in Computer Science, Vol. 5640, Springer Berlin Heidelberg, pp. 193–216.

Shinde, S. K. and Kulkarni, U., 2012. Hybrid personalized recommender system using centering-bunching based clustering algorithm. *Expert Syst. Appl.* 39(1), pp. 1381–1387.

Singhal, S., McGreal, S. and Berry”, J., 2013. Application of a hierarchical model for city competitiveness in cities of india. *Cities* 31, pp. 114 – 122.

Thovex, C. and Trichet, F., 2013. Semantic social networks analysis. *Social Network Analysis and Mining* 3(1), pp. 35–49.

Wang, H., Terrovitis, M. and Mamoulis, N., 2013. Location recommendation in location-based social networks using user check-in data. In: *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL’13, ACM, New York, NY, USA, pp. 374–383.

Wolff, A. and Mulholland, P., 2015. Navigation strategies in the cityscape/datascape. In: *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC’15 Adjunct, ACM, New York, NY, USA, pp. 1143–1151.

Ye, M., Yin, P. and Lee, W.-C., 2010. Location recommendation for location-based social networks. In: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS ’10, ACM, New York, NY, USA, pp. 458–461.

Zanker, M. and Jessenitschnig, M., 2009. Case-studies on exploiting explicit customer requirements in recommender systems. *User Modeling and User-Adapted Interaction* 19(1-2), pp. 133–166.