2012 International Conference on Social Informatics

Tourist Tour Planning Supported by Social Network Analysis

Lule Ahmedi, Korab Rrmoku and Kadri Sylejmani
Computer Engineering Department, University of Prishtina
Kodra e diellit pn, 10000 Prishtina, Kosova
lule.ahmedi@uni-pr.edu, korab.rrmoku@gmail.com, kadri.sylejmani@uni-pr.edu

Abstract—**In line with the just recent "social machines" phenomena, this work proposes an extension based on social network analysis (SNA) to the tourist tour planning. It is able to estimate the tourist's satisfaction with individual Point of Interest (POI), and accordingly recommend or not that POI in the tour in view for that tourist. We first provide a model of developing a social network comprised of tourists and reviewers including their personal attributes (like age or gender in their social profile), preferences of reviewers for certain POIs, and tourists' preferences for certain types or categories of POIs (say archeology) in a given touristic destination. Then in the second part, an algorithm for *grouping* into "islands" of most similar reviewers to a certain tourist is defined. Additionally, a *ranking* algorithm based on authority centrality is adopted to identify the highest ranked reviewer within the island and recommend his preferred POI to a given tourist. The results of the evaluation tests prove our approach as feasible in estimating the tourist's satisfaction with individual POIs. Moreover, it is already promising since acting within a social network as opposed to its counterparts of less appreciation for the social dimension of user.**

Keywords-social network analysis; tourist tour planning; satisfaction factor; models and algorithms.

I. Introduction

We are just recently entering the new area of so-called "social machines" [1] which adapt to the social behavior of everyday user as opposed to traditional computational machines of less appreciation for the social dimension of the user. In line with that trend, this work introduces a novel idea of utilizing social network analysis (SNA) [2][3] for tourist tour planning. To our knowledge, there is no evidence of such an approach to date.

Mostly, during the pre-trip phase, tourists are engaged in planning their activities to be conducted at the touristic destination of their choice. Even though, with help of various tools or communities, they are able to make some preparations (such as e.g., accommodation booking, purchasing travel tickets, or entry tickets for visiting a certain Point of Interest - POI), they still find it difficult to get detailed information about which POIs to visit that meet their personal preferences. The problem thus lies in (1) estimating the tourist's satisfaction with individual POIs, and then (2) finding an optimal route to visit those POIs under some user set constraints [4]. Here we focus on the former, to then provide recommendations to the tourist tour planning system on which POIs to cover when generating an optimal tourist route for a given tourist.

This work is partially supported by a national research grant of the Republic of Kosova for the project "Tourist Tour Planning and Social Network Analysis".

A. *Our approach at a glance*

Our SNA-based approach of estimating the tourist's satisfaction with individual POIs consists of two main steps. First, we develop a social network comprised of tourists and reviewers for certain POIs in the given touristic destination. It takes into account some personal attributes of both tourists and reviewers, like age or gender, as well as their touristic tastes, like type or category of the POIs (say archeology) [5] they like / dislike, or alone certain POIs and their ranking. Then in the second step, several metrics of social networks analysis (SNA) over the generated tourist-reviewer network are considered, and the most adequate ones are adopted and evaluated for estimating the tourist's interest in a given POI.

Since the evaluation results show that the percentage of correct estimation of tourist's satisfaction with individual POIs is above 75% in all cases, we consider our approach as valuable for application in different tourist trip planning systems. Although compliant with the existing e-Tourist solutions, it seeks to further gain from a typical behavior of a user on the Web 3.0 which acts within a social network rather than isolated in whatever online activity (alike the selection of the destination POIs).

B. *Related work*

Many tourist trip planning systems have been developed [4] to date which support an optimal route planning to visit POIs of highest satisfaction values for a tourist in a given destination, Next, we review some recent systems in that domain, and compare them with our approach.

The City Trip Planner [6] prepares city trip itineraries that are tailored to user personal interests. Prediction of personal interest score is supported for each POI and bases on the type score, the category score, and the keyword search score. The first two scores use a rather simple formula for matching user given preferences, while the third one is based in the vector space model [7]. In another touristic trip planning system [8], the estimation of the tourist satisfaction in respect to tourist preferences is less considered since the focus is put in the route generation and customization. Further, Kurata in [9] takes a more interactive approach by allowing a tourist to customize the itinerary until she/he is satisfied. The estimation of tourist satisfaction for individual POIs is done by a simple function that uses a weighted sum of matching criteria for five POI categories: popularity, education, art, nature and amusement.

978-0-7695-5015-2/12 $26.00 © 2012 IEEE
DOI 10.1109/SocialInformatics.2012.89

Recently, a restaurant recommender system for tourists based on social network analysis is introduced [10]. It suggests utilizing information in social networks, including the tourist's own profile and preferences, item's general acceptance, and influence from social friends to improve the performance of recommender systems. Experiments they conducted on a real online social network revealed that friends have a tendency to select same items and give similar ratings. Another work [11] proposes the representation of collaborative relationships in recommender systems as a social network, and the deployment of several measures of SNA to analyze collaborations in order to optimize the information exchange in collaborative recommender systems.

To the best of our knowledge, none of these existing systems considers SNA techniques to suggest POIs of interest to a given tourist. Results obtained in [10][11], as well as the existence of well-defined theories on calculating a rich set of metrics which characterize social networks, like group metrics, or centrality measures for ranking, motivated us in designing a new paradigm for estimation of tourist interest, and then utilize it for a tourist tour planning which may take advantage of information in social networks, as will be detailed in the sections to follow.

II. BACKGROUNG ON RELEVANT SNA CONCEPTS

With the birth of the web 2.0 technologies, the social network data have emerged in popularity [12; 13; 14; 15; 16; 17; 18] to the extent that numerous social network services like Facebook, LinkedIn or Twitter are now supporting users to share their interests and/or activities in form of social network structures. You can publish and share your data and services of arbitrary topic domain including tourism thanks to dedicated social network applications.

A. SNA main metrics

Social network analysis (SNA) has attracted considerable interest in recent years and plays an important role in many disciplines [19]. It provides theories and techniques that prove the effects of an individual or a group of individuals belonging to a given network into some outcomes related to that individual or group. The position of the participant in the network determines his / her favorable or constraining role in the network in terms of the outcomes under consideration.

The authors in [20] were the first to represent social networks in form of a sociogram, i.e., a graph with vertices (nodes) representing social actors and the graph links (ties, lines) representing relationships between social actors. The graph-based representation allows researcher to apply graph theory to the analysis of social networks [3][2]. Following are some main SNA metrics and their categorization:

- Global graph metrics: seek to describe the characteristic of a social network as a whole, for example the graphs diameter, mean node distance, the number of components (fully connected subgraphs), cliques, clusters, small-worldness, etc.
- Individual actor properties: relate to the analysis of the individual properties of network actors, e.g. actor status as central (degree, closeness, or betweenness centrality) or authoritative (eigenvector, PageRank, SALSA, HITS, or weighted HITS), distance, and position in a cluster.

B. Dyadic two-mode network and the homophily

Definition 1 (Dyadic two-mode network). A dyadic two-mode network $\mathrm{N} = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W})$ consists [2] of:

- a *graph* $\mathrm{G} = (\mathcal{V}, \mathcal{L})$ where $\mathcal{V} = \top \cup \bot$ is the set of *vertices* with $\top$ representing the set of top vertices, and $\bot$ the set of bottom vertices, whereas $\mathcal{L} \subseteq \top \times \bot$ is the set of *links*. The difference with classical graphs lies in the fact that the vertices are in two disjoint sets, and that the links are always functions of dyads in which the first vertex and the second vertex in the dyad are from the different sets. In other words, there cannot be any link between two vertexes in the same set. Undirected links $\mathcal{E}$ are called *edges*, and directed links $\mathcal{A}$ are called *arcs*. $n = \mathrm{card}(\mathcal{V})$, $m = \mathrm{card}(\mathcal{L})$.
- $\mathcal{P}$ *vertex properties*: $p: \mathcal{V} \to A$
- $\mathcal{W}$ link weights: $w: \mathcal{L} \to B$. ▫

An example dyadic two-mode network is a network containing data on authorships (i.e., publications authored by authors).

Definition 2 (Homophily). An application of two-mode network analysis is based on the idea of *homophily* (Greek, "love of the similar")—the idea that people who share interests or attributes are more likely to form ties than people who are very different [21]. The theory seems to stand in some cases, and fails in some others, so it's not universal by any means. ▫

Homophily is the tendency to relate people with similar characteristics (status, beliefs, etc.). Thus, in applications where such rationale may be the case, treating an interest or attribute matrix as a two-mode network can be a useful mechanism.

C. A simple analysis using link-cuts: the islands in the net method

Definition 3 (Link-cut of a network). The link-cut [22] of a network $\mathrm{N} = (\mathcal{V}, \mathcal{L}, w)$, $w: \mathcal{L} \to \mathbb{R}$, at selected level t is a subnetwork $\mathrm{N}(t) = (\mathcal{V}(\mathcal{L}'), \mathcal{L}', w)$, determined by the set

$$\mathcal{L}' = \{l \in \mathcal{L}: w(l) \geq t\}$$

and $\mathcal{V}(\mathcal{L}')$ is the set of all endpoints of the links from $\mathcal{L}'$. ▫

We look at the components of $\mathrm{N}(t)$. Their number and sizes depend on t. There are usually many small components. Often only components of size at least k and not exceeding K are considered. The components of size smaller than k are discarded as 'non-interesting'; and the components of size larger than K are cut again at some higher level.

The values of thresholds t, k and K are determined by inspecting the distribution of vertex/link-values and the distribution of component sizes, and considering additional knowledge on the nature of network or goals of analysis.

Definition 4 (*Islands in the net* method). The "islands in the net" method [22] bases on link-cuts of a network (see Def. 3). In the reduced network $N(t)$ for a selected threshold t, also referred to as the "water level" in the literature [23], (weakly) connected components are determined. Each component of size in range $k..K$ represents an island since:

- they are connected and of selected size,
- all links linking them to their outside neighbors have weight lower than t, and,
- each vertex of an island is linked with some other vertex in the same island with a link weight at least t.

Whenever the threshold t (i.e., the water level) is raised at a higher value, it is possible that:

- The given island splits into smaller islands (connected components) with a couple of "valleys flooded", or
- The entire island is "flooded" leaving just a couple of isolates (connectionless components) survive, and thus the island disappears.

Therefore, the islands method needs to be applied judiciously to reveal meaningful results. ▫

Islands are characterized with stronger internal cohesion in terms of link weights relatively to its neighborhood. Therefore, the *islands in the net* method may serve well in finding potential groupings of people. Each island is identified with its *port* – its lowest link weight. The main problem is the links at the same weight level – flat regions, which might require additional modeling constraints as will be discussed later in the context of building our approach.

III. SYSTEM ARCHITECTURE

In Fig. 1, the architecture of a tourist tour planning system supported by our social network analysis (SNA) module is depicted. At the input, the system is supplied with the following types of data:

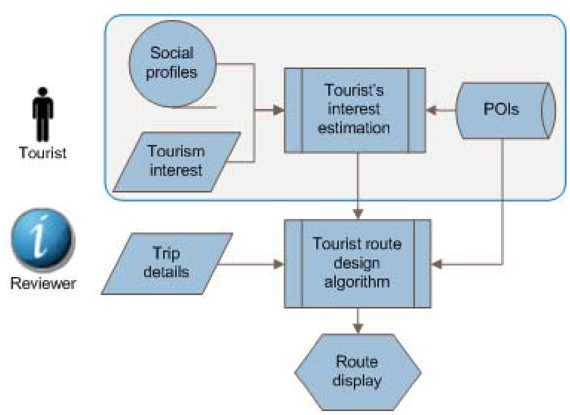

Figure 1. The architecture of a tourist tour planning system supported by SNA.

- **Trip details** (e.g. date, number of days, number of tourists, start location, etc.).
- **POIs** that exist in a certain city/region. The attributes of POI include type and (e.g. archeology, architecture, classical art etc.) category (e.g. monument, castle, park etc.) according to [5], coordinates, entry fee, typical duration, working hours, etc.
- **Social profile** data of both tourists and reviewers (see Table 1) – gathered by default from social networks (their accounts in Facebook, Twitter, etc.); For tourists, if no social network data available, an interface to enter such data (e.g., age, gender, country, occupation) is provided; The "reviewers" are people that are either professionals in ranking POIs, or they are past tourists.
- **Tourism interests** (cf. Table 1) of reviewers for certain POIs, as well as of tourists for different POI types and categories (e.g., archeology, nature, or entertainment) [5] - social network applications posted and gathering opinions from reviewers on POIs, which indirectly provides review data for types and categories too given each POI is by default known to which type and category it belongs. Tourists enter their preferences for type and category explicitly either through social network applications, or through a Web interface.

As one may notice, two main types of users which interact with the system are tourists and reviewers (Fig. 1). Tourists provide some basic input in relation to their touristic taste in general, like *type and category* of POIs they like, to meet in the given destination, whereas reviewers provide their *preferences* for individual POIs they like. The tourist's personal profile, as well as the *profiles* of reviewers may also aid in recommending that tourist's destination POIs.

Tourists' interest estimation. The personal attributes and interests on tourism that tourists and reviewers share, like same gender or nationality, and same type of POIs they like / dislike, may guide deducing a preferable POI of the reviewer become the preferable POI of the tourist (Table 1). It is the role of the tourists' interest estimation component to calculate the preferable POIs of the tourist, and recommend them further to the component which designs the tourist's optimal route for visiting those POIs.

To better understand how the system works as a whole and where is the contribution of our SNA-based module which estimates the tourists' interest on each POI to visit, let us suppose that *n* tourists visit the region (or city) A for individual *n* tours each. The region A has *m* POIs that tourists may visit during their stay. Each POI has *n* satisfaction values Sij, where each of them represents the level in the range 0 to 3 (the lowest 0 to the highest one 3) of individual tourist satisfaction in respect to his or her preferences. The meaning of each of four satisfaction values 0, 1, 2, and 3 is as follow: 0 – means "I do not like it", 1 – means "I like it to a certain extent", 2 – means "I like it", and 3 – means "I like it very much".

In order to have a more clear view about the output of our system, let us assume a simple trip with the following details: n=3, m=5, Sij (i=1... n, j=1... m), and satisfaction values, POIs' attributes, and travel distances between POIs as provided in the figure below.

P_2	P_3	P_4	P_5	P_{SE}	
60.83	60.00	72.80	94.34	10.00	P_1
	10.00	72.11	94.34	70.71	P_2
		80.62	80.62	60.00	P_3
			31.62	72.80	P_4
				100.00	P_5

Tourist	P_1	P_2	P_3	P_4	P_5
T_1	2	1	3	**3**	1
T_2	1	1	2	0	**3**
T_3	2	1	2	**3**	2

POI	Latitude	Longitude	Visit duration	Open time	Closing time
P_1	90	10	30	0	120
P_2	**80**	**70**	15	30	120
P_3	**90**	**70**	20	30	105
P_4	20	30	30	60	120
P_5	10	60	25	0	90
P_{SE}	90	0	0	0	330

Figure 2. Tourists' satisfaction values (left), basic details of POIs (bottom), and traveling distances in minutes between POIs (right).

Now, assuming the limited duration of 330 minutes (5 and half hours) for the trip, and aiming for the maximal overall tourist's satisfaction with the provided POIs along the route, the system may prepare a trip itinerary, say P1P3P4 is the best route for tourist T1 with overall satisfaction 8 and duration 303.42 min, P1P3P5 for tourist T2 with overall satisfaction 6 and 325.62 min, and P1P4P5 for tourist T3 with overall satisfaction 7 and duration 299.42 min. The contribution of our SNA-based module is to provide the tourists' satisfaction values, or interests' values of each given tourist for each POI in the region to visit. The taboo search algorithm for tourist planning introduced in [24] uses a rather simple method of computing the satisfaction factor.

The algorithm we developed which makes use of the input data in the system to estimate the tourists' interest on individual POIs will be described later once we have introduced the modeling of input data in our approach to then apply the algorithm.

TABLE I. REVIEWERS' AND TOURISTS' INPUT DATA TO CALCULATE THE PREFERABLE POIS OF TOURISTS AT THE OUTPUT.

	Reviewer	**Tourist**
Social profile attributes (e.g., of the Facebook profile)	Age Gender Nationality Resident country Occupation Hobby	Age Gender Nationality Resident country Occupation Hobby
Interests on tourism	Type Category POI review: 0,1,2, or 3	Type preferences: 0,1,2, or 3 Category preferences: 0,1,2, or 3 Preferable POI = ?

IV. TOURIST-REVIEWER NETWORK

Given the above rationale, tourists and reviewers are modeled as actors in a social network, we refer to it here as the tourist-reviewer network, where the links and link weights between tourists and reviewers in the network are drawn based on the factual commonalities in tourist's and reviewer's touristic tastes (i.e., common type and/or category they like), and to some extent if common attributes in their personalities (profiles). The tourist-reviewer network provides a good modeling base to start with analysis in order to infer as close as possible the tourist's taste for particular POIs.

Next we provide few definitions of our modeling approach of the tourist-reviewer network.

Definition 5 (Tourist-reviewer network). A tourist-reviewer network is a dyadic two-mode network, we denote here as $\mathrm{N_{t,r}} = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W})$, where:

- A *graph* $\mathrm{G} = (\mathcal{V}, \mathcal{L})$ consists of the set of vertices $\mathcal{V} = \mathrm{T} \cup \mathrm{R}$, with T representing the set of tourists, and R the set of reviewers for a given set of POIs, whereas $\mathcal{L} \subseteq \mathrm{T} \times \mathrm{R}$ is the set of *links*.
- $\mathcal{P}$ is the set of *vertex properties*, namely their personal attributes and tourism interests as given in Table 1: $p: \mathcal{V} \rightarrow A$.
- $\mathcal{W}$ is the set of link weights: $w: \mathcal{L} \rightarrow B$ calculated following the definition of so-called *tourist-reviewer similarity weight* we provide below to best reflect the influence of reviewers' profile and his / her touristic taste to the given tourist.
- $\mathcal{L} \subseteq \mathrm{T} \times \mathrm{R}$ is the set of *links* l: There is a link l between a tourist and a reviewer $\mathrm{T} \times \mathrm{R}$ if its link weight $w(l)$ is at least a predefined minimum weight value, i.e., $l \in \mathcal{L} : w(l) \geq w_{min}$.

Definition 6 (Tourist-reviewer similarity weight). Given a set of vertices $\mathcal{V} = \mathrm{T} \cup \mathrm{R}$, where T is the set of tourists and R is the set of reviewers, as well as a set $\mathcal{P}$ of properties (personal and touristic) for that set $\mathcal{V}$ of vertices, we calculate the set $\mathcal{W}$ of similarity weights w for each T × R pair based on their personal similarity and touristic interests similarity as follows (cf. the algorithm in Fig. 3):

1) We first calculate the *touristic interests similarity weight* w_{int} (lines 1 to 4) which is the sum of:
   a) the touristic interests values of a tourist $Int_j(T)$ out of the set {0, 1, 2, 3} for *types* of POIs if both tourist and reviewer have assigned exactly the same touristic interests values for that same type of POIs, and,
   b) the touristic interests values of a tourist $Int_j(T)$ out of the set {0, 1, 2, 3} for *categories* of a POIs if both tourist and reviewer have assigned exactly the same touristic interests values for that same category of POIs.
2) If the *touristic interests similarity weight* w_{int} *is greater than 0*, we continue the calculation; otherwise it makes no sense to continue with the weight calculation, and hence return 0 as the similarity weight value w (lines 5 to 7).
3) We next calculate the *personal similarity weight* w_{per} (lines 8 to 11) which is initially set to 0, and then incremented by 1 each time there is an exact match in values between the tourist and the reviewer for a same given personal attribute Per_i under consideration.
4) We *sum up* (line 12) the values of the touristic interests similarity weight and the personal similarity weight to yield the total weight for a given tourist-reviewer pair.

5) Finally, we *normalize* the above weight value to a certain minimal value, e.g. $w_{min} = 5$, and subtract that value (5) from the above weight, as well as avoid "loosely coupled" tourists and reviewers influence our algorithm of preferring the POIs to visit for a given tourist: their weights reset to 0 if not greater than 5. One might of course select a different w_{min} as to reach a perfect match for his / her domain of application. Recall that following Definition 5, there is a link between a tourist and a reviewer if its link weight is at least a predefined minimal weight: in our case, we have set that minimum value to 5.

Algorithm(SimilarityWeight(T, R, P(T),P(R)): T $\vec{w}$ R

```
INPUT:
 • A tourist T in T, and a set of its properties P(T) as:
          P(T) = Per(T) U Int(T), where

/* Per(T) stands for personal attributes of tourist T, like his age as Per1(T), his
gender as Per2(T), etc. */
          Per(T) = {Per1(T), .., Peri(T), .., Perm(T)}

/* Int(T) stands for touristic interests of tourist T, like he is interested for the POIs
of type archeology (Int1(T)=2), or not at all interested for the POIs which fall into
the category of monuments (Int2(T)=0), etc. */
          Int(T) = {Int1(T), .., Intj(T), .., Intn(T)}

 • Similar to tourists, a reviewer R in R, and a set of its properties P(R) as:
          P(R) = Per(R) U Int(R).

OUTPUT:
 • Similarity weight w of the T w R link for a given T X R pair

/* We first calculate the weight wint of the touristic interests similarity */
   Let initially wint =0
/* For each same type or category of POIs, say, for the type "archeology" as Int1 */
   FOREACH Intj (j=1 to n)
      IF ( Intj(T) == Intj(R) ) // e.g., both T & R are equally interested in archeology
           THEN wint = wint + Intj(T)
   IF (wint == 0 )
       THEN RETURN 0      // w = 0
   ELSE
/* We next calculate the weight wper of the personal similarity as follows: */
   Let initially wper =0
/* For each same personal attribute, say age as Per5 */
   FOREACH Peri (i=1 to m)
      IF ( Peri(T) == Peri(R) )
           THEN wper ++
/* Similarity weight is the sum of the personal similarity weight and the touristic
interests similarity weight */
   w = wint + wper
/* Finally, we normalize the weight to a certain minimal value, and discard links of
negligible weights for our domain */
   IF (w > wmin )     // predefined minimum similarity weight wmin , e,g, set to 5
        THEN RETURN w – wmin
   ELSE RETURN 0
```

Figure 3. The SimilarityWeight algorithm.

Essentially, what this means is that individual's shared interests, and their personal affinities, are shaped by weights in links relating them. The weight *w* is higher for pairs T X R if closer in profile and taste as followed in the algorithm.

Example 1 (Similarity weight). Following is an example of calculating the similarity weight following the SimilarityWeight algorithm. As shown in Table II, given is a tourist T1 and two reviewers R1 and R2, each with its set of properties, both personal (Per_1 to Per_6) and touristic interests properties (Int_1, Int_2). The two last rows in the table show the values obtained for the personal similarity weight, w_{per}, the touristic similarity weight w_{int}, and finally the weight value w after normalization for a given tourist-reviewer pair, (T1, R1) and (T1, R2) respectively.

TABLE II. AN EXAMPLE OF CALCULATING THE SIMILARITY WEIGHT

	Per_1: Age	Per_2: Gender	Per_3: Nationality	Per_4: Resident country	Per_5: Occupation	Per_6: Hobby	w_{per}	Int_1: Type of POIs	Int_2: Category of POIs	w_{int}	w
T1	22	Female	DE	DE	Engineer	Sports		2	3		
R1	29	Female	KS	DE	Engineer	Sports		2	1		
R2	40	Male	MK	DE	Student	Reading		1	2		
T1,R1	1	1	0	1	1	1	5	2	0	2	(5+2) -5=2
T1,R2	0	0	0	1	0	0	1	0	0	0	1+0= 0

V. ESTIMATION OF TOURISTS' INTEREST USING SNA

Although the tourist-reviewer network provides an orientation already of which tourists and reviewers are more similar to each other in terms of common tourist's interests and personal affinities, due to likely large number of links relating a tourist to multiple reviewers, it is cumbersome to recognize the most similar reviewer to a given tourist and deduce the tourist's interest.

In this work, we have considered several metrics of Social Networks Analysis (SNA) [2][3] to find the most strongly linked tourist-reviewer groups, i.e., tourist-reviewer groups with links of highest similarity weights. The **islands in the net** group analysis, proved to suit best for our touristic domain. Islands are characterized with stronger internal cohesion in terms of similarity weights relatively to its neighborhood. Therefore, the *islands in the net* method may serve well in finding potential groupings of tourists and reviewers.

Finally, if there are yet ambiguities in selecting the most similar reviewer to a given tourist within the island, the **ranking** by the simple in-degree centrality, and by the authority centrality applied over "concurrent" reviewers in the island may resolve the situation.

A. The tourist-reviewer islands and the algorithm

In order to identify tourist-reviewer groups with links of highest similarity weights, we adopt the "islands in the net" method to our tourist-reviewer network $N_{t,r}$ as detailed in the following definition.

Definition 7 (*Islands in the tourist-reviewer net* method). The "islands in the tourist-reviewer net" method bases on link-cuts (see Def. 3) of a tourist-reviewer network where the threshold or the so-called "water level" is adjusted to a certain similarity weight value. In the reduced tourist-reviewer network $N_{t,r}(t)$ for a selected similarity weight threshold t, tourist-reviewer islands are determined such that *each tourist of an island is linked with some reviewer in that same island with a similarity weight at least t*.

The value of water level t is determined by inspecting the distribution of tourists in relation to reviewers within islands, and how strong their links are. Tourist's preferences for individual POIs will be estimated by a set of reviewers. The process of suggesting POIs to tourists in the algorithm consists of mainly raising the "water level" incrementally and with care as to reveal meaningful estimations, as detailed through the

RecommendPOIs algorithm whose pseudo-code is provided next in Fig. 4 and will be explained through an example to follow.

```
Algorithm(RecommendPOIs(N_t,r): Ti ← Rj
INPUT:  A tourist-reviewer network N_t,r

1   Let t=0                      // t stands for threshold, or "water level"
2   WHILE Tourists set is not empty
3     FOREACH (Ti in tourists set)
4       WHILE not(Ti ← Rj)
5       t++
6       IF (Ti is isolated)   // no links to reviewers
7          THEN t--    UNTIL Ti links to at least one reviewer
8                         CONTINUE
9       ELSE IF (Ti links to a single reviewer, say to Rj)
10          THEN RETURN Ti ← Rj
11      ELSE // remains that Ti links to more reviewers Rj in the concurrent set Rc
12          Apply ReviewersRanking(Rc) algorithm
13          RETURN Ti ← HighestRanked(Rj)
```

Figure 4. The islands algorithm adopted to our tourist-reviewer network.

Example 2 (Tourist-reviewer islands). Let us consider a tourist-reviewer network depicted in Fig. 4 which is composed of ten tourists and fifty reviewers both with their personal and touristic interests properties (cf. Definition 5), as well as their links of weights calculated following the SimilarityWeight algorithm (cf. Fig. 3).

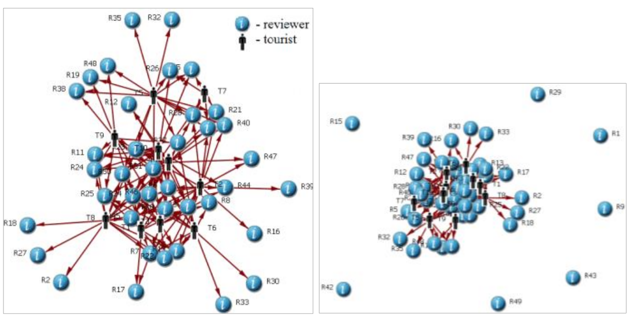

Figure 5. Network graph of our Example 2 with 50 reviewers R1 to R50, and 10 tourists T1 to T10: (left) the core component without isolates; (right) the whole graph including isolated nodes, i.e., reviewers R1, R9, R43, R49, R42, R15, and R29, which connect to no tourists.

If we try to find out which reviewers influence most the given tourist, say tourist T1, in terms of satisfactory POIs, then referring to Fig. 7, we will hardly manage to infer that knowledge from the current dense graph: the tourist T1 links to a number of reviewers.

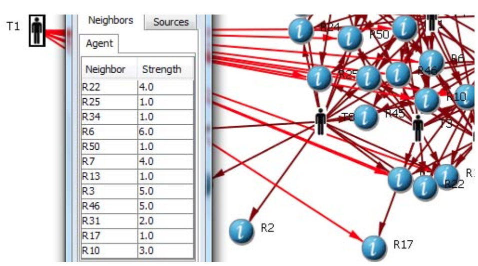

Figure 6. Tourist T1 linking to its neighboring reviewers in our Example 2.

Therefore we apply the RecommendPOIs algorithm (Fig. 4) for the tourist T1 by incrementing the water level threshold (line 5) if yet no recommendation provided to tourist T1 (line 4) unless an optimal number of links remain for T1 (or alternatively lines 6 to 8), sufficient to either determine the sole influencing reviewer (lines 9 and 10), or further consult the ranking algorithm ReviewersRanking (lines 11 to 13) whose pseudo-code is provided in Fig. 5 to resolve the dilemmas among several reviewers remaining.

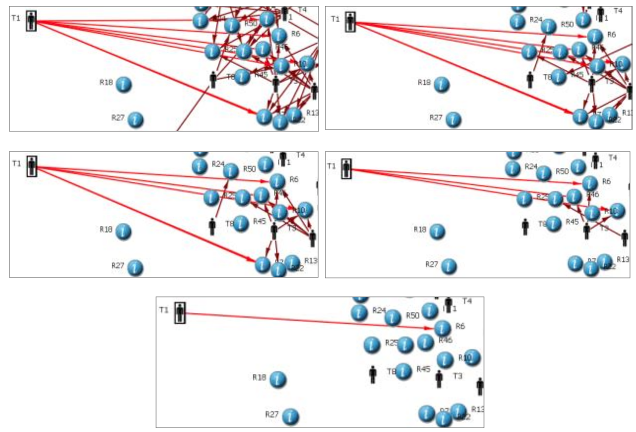

Figure 7. Tourist's T1 decreasing number of links to reviewers while raising the waterlevel (i.e., link weight) threashold from 0 in Fig. 6 (nr of links is 12) to: > 1 (number of links decreases to 6), > 2 (nr of links decreases to 5), > 3 (nr of links decreases to 4), > 4 (nr of links decreases to 3), and finally > 5 (remains a single link only).

For the tourist T1 in the example, we have incremented the water level from its initial value 0 to slightly greater than 1 (say 1.1), then to 2, 3, 4 and finally slightly greater than 5 which satisfies the condition for T1 to link to a single reviewer (line 9 in Fig. 4), reviewer R6. Figure 7 shows clearly the weight of links, referred to as strength, of tourist T1 to each of its 12 neighboring reviewers: T1 has link of strength (i.e., similarity weight) 6 and is therefore the most influencing reviewer whose recommendation for POIs shall be preferred to tourist T1 (line 10 in Fig. 4).

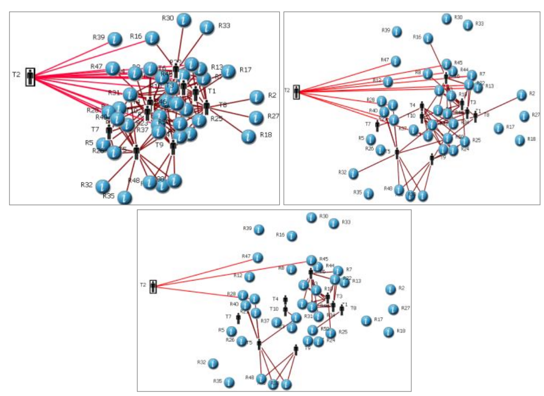

Figure 8. Tourist's T2 decreasing number of links to reviewers while raising the waterlevel up to > 2. Three equaly weighted links (weight 3.0) to R47, R45, and R4 remaining at the end.

In general when applying the RecommendPOIs algorithm over a tourist-reviewer network, the *best case scenario* happens when at a given water level, more pairs are defined at one hand, i.e., more than just one tourist link to their corresponding single reviewers at that same actual water level (condition in line 9). In contrary, the *worst case scenario* is when after raising the water level, a certain tourist "sunks" although not yet paired with any reviewer. In such situations, we shall return back to a lower "water level" (lines 7 and 8) to see for any matching options of a given tourist.

B. Author centrality ranking of reviewers and the algorithm

As stated in [22], the main problem with the islands method is the remaining links of the same weight – so-called "flat islands". In our domain, flat islands make a tourist link with the same weight to multiple reviewers and hence cause recommending multiple POIs to a given tourist. If instead a single POI recommendation is expected for each given tourist (recall our RecommendPOIs algorithm), further algorithmic processing is required to resolve ambiguous recommendations. We have therefore developed an algorithm, ReviwersRanking (Fig. 9), invoked from within the RecommendPOIs algorithm, which takes a set of same-weight reviewers at the input, and returns a single highest ranked reviewer according to both the in-degree centrality and the authority centrality ranking.

Note that authority centrality equals to the simple in-degree centrality in our network since each inbound traversal to a tourist which may influence that tourist's authority consists of a single directed reviewer-tourist link in the network. Recall that [2][3] a node's degree is used as a measure in assessing which vertexes are central with respect to influencing others in their immediate 'neighborhood', and in our domain there exists nothing but *immediate* neighbors. Other frequent measures of centrality in SNA today for two-mode networks like closeness and betweenness centralities [25] are not applicable to our domain since again, the information spreading in our tourist-reviewer network is rather limited to single link traversals and is hence insufficient.

The example which follows illustrates best the application of the ReviewersRanking algorithm.

```
Algorithm(ReviewersRanking(Rc))
FOREACH Rj from the concurrent set Rc of reviewers
Rank Rc according to
    In Degree Centrality
    Authority Centrality
RETURN HighestRanked(Rj)
```

Figure 9. The ranking algorithm over same-weight concurrent reviewers.

Example 3 (Reviewer's Ranking). Let us continue with the same example tourist-reviewer network as in the previous example (Example 2), but consider this time tourist T2. Following the RecommendPOIs algorithm for tourist T2 (see Fig. 8), the water level is incremented gradually by 1 which causes the decrease in number of links of tourist T2 to reviewers in the network as depicted in Fig. 9. At the end, when the water level is set to a value slightly greater than 2, three reviewers of equal weight remain connected to tourist T2, namely reviewers R47, R45, and R4. Therefore, the ranking algorithm ReviewersRanking (Fig. 5, lines 11 to 13) is next applied over three remaining reviewers, and returns R4 as the highest ranked reviewer among them. The RecommendPOIs algorithm responds with finally recommending reviewer R4 (namely its recommended POI) to tourist T2.

VI. Preliminary Results

We have evaluated our approach with running a test over around 200 POIs of the Republic of Kosova. Three different tourist-reviewer network instances are developed for the sake of the test as explained in Table III, 1st column. Note that different number of tourists involved in distinct instances means also completely different set of tourists involved within distinct instances. For each instance network, in addition to tourists and reviewers, data about their properties both personal and their interests in tourism are gathered. Then the similarity weights are calculated in all three network instances separately.

Finally, the analysis of the similarity of tourist-reviewer pairs based on the islands method, as well as, whenever required, ranking of "concurrent similar reviewers" is performed. The recommended POIs to tourists by their most similar reviewers which resulted out of these islands-based analysis and ranking are checked against preferred POIs by tourists who have again been interviewed to evaluate their satisfaction with the recommended POIs. The preliminary evaluation results on the satisfaction factor (SF) of tourists with the recommended POIs are summarized in Table III.

TABLE III. AVERAGE SATISFACTION FACTOR OF TOURISTS

	nr of reviewers, tourists	**islands in the net**	**islands in the net & indegree centrality**	**islands in the net & authority centrality**
1st instance	50, 10	76.94%	84.44%	84.44%
2nd instance	50, 7	75.39%	78.57%	78.57%
3rd instance	50, 5	80.00%	83.33%	83.33%
average		77.44%	**82.11%**	82.11%

Obviously, the preliminary results on tourist's satisfaction which vary in the range 75.39% to 84.44% prove the approach as feasible and promising, be it without (2nd column) or with (3rd and 4rth column) application of the centrality ranking algorithm. Nevertheless, when comparing the results before and after applying the ranking algorithms, following remarks may be drawn:

- In the 1st instance, we have reached an improvement of 7.50% in result after ranking is applied.
- In the 2nd instance, we have reached an improvement of 3.18% in result after ranking is applied.
- In the 3rd instance, we have reached an improvement of 3.33% in result after ranking is applied.

It is also obvious from Table III that in-degree centrality and authority centrality measures yield equal values, as also theoretically argued in the earlier section.

A detailed overview on the evaluation results of our approach over the same dataset is provided in Table IV, Table V, and Table VI.

As an illustration, let us explain the evaluation results over the 1st network instance (cf. Table IV) of a total of 1o distinct tourists T1 to T10 interviewed. Once the islands method is applied (the left-hand side columns), there are in general cases (rows) that a tourist links:

- to a single peak (most similar) reviewer suggesting him a POI (the rows with the shaded 2nd column), or
- to more than just a single peak reviewer suggesting him a POI (the rows with no shadow).

If considering the tourist T1 of this instance, only a single reviewer, the reviewer R6, results into a peak reviewer once the islands method is applied. No ranking algorithm is further applicable in that case. On the interview afterwards, the tourist T1 states he / she is very satisfied (value 3 in the 3rd column, i.e., "I like it very much") with the suggested POI by the reviewer R6, i.e., a 100% satisfaction factor (column 4) achieved through our *islands in the net* algorithm.

As another typical case, if considering the tourist T2, several (more than just one) reviewers, R4, R45, and R47, result into peak reviewers to T2 according to our islands method. The tourist's satisfaction factor with each reviewer R4, R45, and R47 is 2 ("I like it"), 3 ("I like it very much"), and 2 ("I like it") respectively, which in average yields a satisfaction value 7/3 = 2.33 (slightly above "I like it"), or expressed in percentage it is 2.33*100/3 = 77.77%. If a ranking algorithm (in-degree, or centrality) is further applied to pick up only a single reviewer as the highest ranked peak reviewer (most similar) reviewer among R4, R45, and R47, which in this case is the reviewer R4, the satisfaction factor is improved to 3, i.e., expressed in percentage equals to 100%. As one may notice when comparing the satisfaction in percentage of tourist T2 *before and after* the ranking algorithm is applied, a significant improvement is achieved at the level of more than 20% of the tourist T2 satisfaction with the suggested POI after ranking of concurrent peak reviewers.

The same rationale as for the tourists T1 and T2, and the calculation of their satisfaction factors against the suggested POIs is valid also for the rest of the tourists, T3 to T10, in this instance (Table IV), as well as for the whole set of tourists in other two instances evaluated in this work (Table V and Table VI).

TABLE IV. INSTANCE 1 NETWORK

Islands in the Net				Ranking Algorithms			
Tourists	Peak reviewers and their POIs suggested	Overall SF over "peak" POIs	Average SF in percentage	Tourists	Highest ranked peak reviewer and its POI	SF over suggested POI	SF in percentage
Tourist 1	R6	3	**100%**	Tourist 1	NA	3	100%
Tourist 2	R4, R45, R47	2+3+2 = 7	77.77%	Tourist 2	R4	3	**100%**
Tourist 3	R3, R6, R46	2+3+3 = 8	88.88%	Tourist 3	R6	3	**100%**
Tourist 4	R11, R21, R23, R24, R34, R25	2+2+3+3+3+2 = 15	83%	Tourist 4	R34	3	**100%**
Tourist 5	R19, R21, R38, R48	2+3+2+2 = 9	75%	Tourist 5	R21	3	**100%**
Tourist 6	R34	2	**66.66%**	Tourist 6	NA	2	66.66%
Tourist 7	R21	2	**66.66%**	Tourist 7	NA	2	66.66%
Tourist 8	R50	3	100%	Tourist 8	NA	3	**100%**
Tourist 9	R19, R38, R48	2+3+2 = 7	**77.77%**	Tourist 9	R19	2	77.77%
Tourist 10	R31	1	**33.33%**	Tourist 10	NA	1	33.33%

TABLE V. INSTANCE 2 NETWORK

Islands in the Net				Ranking Algorithms			
Tourists	Peak reviewers and their POIs suggested	Overall SF over "peak" POIs	Average SF in percentage	Tourists	Highest ranked peak reviewer and its POI	SF over suggested POI	SF in percentage
Tourist 1	R6, R13, R16, R22, R30, R49	2+3+3+3+1+2 = 14	77.77%	Tourist 1	R13	3	**100%**
Tourist 2	R1	2	**66.66%**	Tourist 2	NA	2	66.66%
Tourist 3	R3, R46, R48	3+3+3 = 9	100%	Tourist 3	R46	3	**100%**
Tourist 4	R33, R34	1+2 = 3	**50%**	Tourist 4	R34	1.5	50%
Tourist 5	R10, R23	2+2 = 4	**66.66%**	Tourist 5	R10	2	66.66%
Tourist 6	R46	3	**100%**	Tourist 6	NA	3	100%
Tourist 7	R3, R6, R46, R48	1+2+2+3 = 8	**66.66%**	Tourist 7	R46	2	66.66%

TABLE VI. INSTANCE 3 NETWORK

Islands in the Net				Ranking Algorithms			
Tourists	Peak reviewers and their POIs suggested	Overall SF over "peak" POIs	Average SF in percentage	Tourists	Highest ranked peak reviewer and its POI	SF over suggested POI	SF in percentage
Tourist 1	R33	2	**66.66%**	Tourist 1	NA	2	66.66%
Tourist 2	R46	3	**100%**	Tourist 2	NA	3	100%
Tourist 3	R7, R22	3+2 = 5	83.33%	Tourist 3	R7	3	**100%**
Tourist 4	R7, R22	2+3 = 5	**83.33%**	Tourist 4	R7	2.5	83.33%
Tourist 5	R10	2	**66.66%**	Tourist 5	NA	2	66.66%

VII. CONCLUSION AND FUTURE WORK

This work introduces a novel idea of utilizing social network analysis (SNA) [2][3] for tourist tour planning. It is able to estimate the tourist's satisfaction with individual Point of Interest (POI), and accordingly recommend or not that POI in the tour in view for that tourist. To the best of our knowledge, there is no evidence of such a SNA-based approach to date to suggest POIs of interest to a given tourist.. There are rather simple solutions of assuming preferable POIs to a given tourist in existing tourist tour planning systems [6][7][8][9].

Further, there are few approaches which have adopted recommended systems for estimating tourist's interests [10][11], but none of them have considered utilizing the existing well-defined SNA theories in designing a new paradigm for estimation of tourist interest, and then utilize it for a tourist tour planning which may take advantage of information in social networks, as is elaborated in earlier sections.

To summarize, following are the main contributions of this work:

- A novel model is designed of developing a social network comprised of tourists and reviewers including their personal attributes (like age or gender in their social profile), preferences of reviewers for certain POIs, and tourists' preferences for certain types or categories of POIs (say archeology) [5] in a given touristic destination.
- The algorithm for *grouping* into "islands" of most similar reviewers to a certain tourist is developed.
- A *ranking* algorithm based on authority centrality is adopted in order to identify the highest ranked reviewer within the island and recommend his / her preferred POI to a given tourist.

The results of the evaluation tests run over dozens of POIs and several instances of tourist-reviewer network, prove our approach as feasible in estimating the tourist's satisfaction with individual POIs (is above 75% in all cases). Moreover, it is already promising as opposed to its counterparts since acting within a social network which have already gained their momentum and might be easily utilized to gather useful data like in tourism, and of course social profile data.

There is already work in place going on to enlarge the set and diversity / coverage of input data to the tourist's interest estimation module to finalize the evaluation of the system, and occasionally transform them as to enable the comparison of this system to other tourist tour systems like those which base on recommender systems [10][11].

REFERENCES

[1] Hendler, J., & Berners-Lee, T. (2010). From the Semantic Web to social machines: A research challenge for AI. *Artif. Intell. , 174* (2).

[2] Wasserman, S., Faust, K., Iacobucci, D., & Granovetter, M. (1994). *Social network analysis: Methods and applications.* Cambridge University Press.

[3] Scott, J. (2000). Social Network Analysis: A Handbook. London, UK: SAGE Publications.

[4] Souffriau, W., & Vansteenwegen, P. (2010). Tourist Trip Planning Functionalities: State of the Art and Future. (F. Daniel, & F. M. Facca, Eds.) *LNCS , 6385*, pp. 474-485.

[5] Souffriau, W., Maervoet, J., Vansteenwegen, P., Berghe, G. V., & Oudheusden, D. V. (2009). A mobile tourist decision support system for small footprint devices. *IWANN 2009, Part I, LNCS 5517* , 1248-1255.

[6] Vansteenwegen, P., Souffriau, W., Berghe, G. V., & Oudheusden, D. V. (2011). The City Trip Planner: An expert system for tourists. *Expert Syst. Appl. , 38* (6), 6540-6546.

[7] Baeza-Yates, R. A., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval.* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

[8] Garcia, A., Arbelaitz, O., Linaza, M. T., Vansteenwegen, P., & Souffriau, W. (2010). Personalized tourist route generation. In F. M. Facca (Ed.), *ICWE'10 Proceedings of the 10th international conference on Current trends in web engineering* (pp. 486-497). Springer-Verlag Berlin, Heidelberg @2010.

[9] Kurata, Y. (2011). CT-Planner2: More Flexible and Interactive Assistance for Day Tour Planning. *ENTER, Information and Communication Technologies in Tourism 2011* (pp. 25-37). Springer Vienna.

[10] He, J., & Chu, W. W. (2010). A Social Network-Based Recommender System (SNRS). *Annals of Information Systems , 12*, pp. 47-74.

[11] Palau, J., Montaner, M., López, B., & De La Rosa, J. L. (2004). Collaboration Analysis in Recommender Systems Using Social Networks. *Lectures Notes in Computer Science , 3191*, pp. 137--151.

[12] Ahmedi, L., Abazi-Bexheti, L., & Kadriu, A. (2011). A Uniform Semantic Web Framework for Co-authorship Networks. *IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, (pp. 958-965). Sydney, Australia.

[13] Aleman-Meza, B., Nagarajan, M., Ding, L., Sheth, A. P., Arpinar, I. B., Joshi, A., et al. (2008). Scalable semantic analytics on social networks for addressing the problem of conflict of interest detection. *TWEB , 2* (1).

[14] De Castro, R., & Grossman, J. W. (1999). Famous trails to Paul Erdös. *MATHINT: The Mathematical Intelligencer , 21*, 51--63.

[15] Elmacioglu, E., & Lee, D. (2005). On six degrees of separation in DBLP-DB and more. *SIGMOD Record , 34* (2), 33--40.

[16] Erétéo, G., Buffa, M., Gandon, F., Grohan, P., Leitzelman, M., & Sander, P. (n.d.). A State of the Art on Social Network Analysis and its Applications on a Semantic Web. SDoW2008, Workshop at ISWC'2008.

[17] [17] Erétéo, G., Gandon, F., & Buffa, M. (2009). Semantic Social Network Analysis. In *Web Science.* Athens, Greece.

[18] Hendler, J. A., & Golbeck, J. (2008). Metcalfe's law, Web 2.0, and the Semantic Web. *J. Web Sem , 6* (1), 14--20.

[19] Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Inf. Process. Manage. , 41* (6), 1462--1480.

[20] Moreno, J. L. (1933). Emotions mapped by new geography. New York Times.

[21] Breiger, R. (1974). Duality of Persons and Groups. *Social Forces* (53), 181-190.

[22] Batagelj, V. (2003). Analysis of large networks - Islands. Dagstuhl, Germany: Dagstuhl Seminar, Algorithmic Aspects of Large and Complex Networks.

[23] Tsvetovat, M., & Kouznetsov, A. (2011). *Social Network Analysis for Startups: Finding connections on the social web.* O'Reilly Media.

[24] Sylejmani, K., & Dika, A. (2010). A taboo search algorithm for touristic trip planning. *Workshop on Information Technology and Turism.* Edinburgh, Scotland: Journal of Information Technology and Tourism (JITT).

[25] Borgatti, P. S. (2009). *2-Mode Concepts in Social Network Analysis.* Robert A. Meyers (ed) Encyclopedia of complexity and systems science New York : Springer.

[26] Nascimento, M. A., Sander, J., & Pound, J. (2003). Analysis of SIGMOD's co-authorship graph. *SIGMOD Rec. , 32* (3), 8--10.