

## Research Article

# A Methodology for Evaluating Algorithms That Calculate Social Influence in Complex Social Networks

Vanja Smailovic <sup>1,2</sup>, Vedran Podobnik <sup>2,3</sup> and Ignac Lovrek <sup>2,3</sup>

<sup>1</sup>Sandvik Machining Solutions AB, Stockholm, Sweden

<sup>2</sup>Social Networking and Computing Laboratory (socialLAB), Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia

<sup>3</sup>Department of Telecommunications, Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia

Correspondence should be addressed to Vedran Podobnik; [vedran.podobnik@fer.hr](mailto:vedran.podobnik@fer.hr)

Received 22 December 2017; Revised 8 June 2018; Accepted 19 June 2018; Published 8 August 2018

Academic Editor: Xiuzhen Zhang

Copyright © 2018 Vanja Smailovic et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Online social networks are complex systems often involving millions or even billions of users. Understanding the dynamics of a social network requires analysing characteristics of the network (in its entirety) and the users (as individuals). This paper focuses on calculating user's social influence, which depends on (i) the user's positioning in the social network and (ii) interactions between the user and all other users in the social network. Given that data on all users in the social network is required to calculate social influence, something not applicable for today's social networks, alternative approaches relying on a limited set of data on users are necessary. However, these approaches introduce uncertainty in calculating (i.e., predicting) the value of social influence. Hence, a methodology is proposed for evaluating algorithms that calculate social influence in complex social networks; this is done by identifying the most accurate and precise algorithm. The proposed methodology extends the traditional ground truth approach, often used in descriptive statistics and machine learning. Use of the proposed methodology is demonstrated using a case study incorporating four algorithms for calculating a user's social influence.

## 1. Introduction

In 2017, more than 2.5 billion people participated in online social networking, with more than two billion of them using Facebook as one of the largest online social networking platforms [1]. In a broader sense, social networks are not just structures of interconnected humans based on their participation in such platforms. Social networks can also be built around other digital products such as telecommunication network operator services (e.g., mobile phone calls and text messaging) or even nonhuman users such as networked objects and smart devices (i.e., forming the so-called Social Internet of Things) [2]. Finally, overarching social networks can be built by combining membership and activities in multiple social networks, thus creating even more complex social networks characterised by not only millions or billions of (human and nonhuman) users but also a very rich set of possible relationships between social network users.

Importantly, understanding the dynamics within a social network requires calculating different properties of complex networks. This paper will focus on properties that describe social networks at the level of the individual user. Though two types of network properties from the aspect of the individual user can be calculated—*key actors* and *key relationships*—they differ significantly in the approach to calculating them. The property *key actors* (such as influence [3–6]) represents *global user properties* as it depends on (i) the *global positioning* of the user within the entire social network and (ii) interactions between the user and *all other users* in the social network (i.e., the property  $1:N$ , where  $N$  is the size of the social network). On the other hand, the property *key relationships* (such as trust [7, 8]) represents *local user properties*, given that they depend on local dynamics between pairs of individual users (i.e.,  $1:1$  property).

Today, there are algorithms for calculating both global and local user properties in social networks [9]. Nevertheless, evaluating the algorithms varies significantly. In evaluating

local user properties, the *ground truth approach* can be applied, which is a traditional approach often used in statistics and machine learning. The basic idea behind the ground truth approach is to collect proper objective data on the modelled property and compare the result obtained from the evaluated algorithm with the result found in ground truth data. For example, when modelling the trust relationship between social network users, ground truth data can be collected using a questionnaire where the number of social network users determines the level of trust between them and other social network users [10, 11]. Given that social trust is a 1:1 user property, surveyed users may answer questions about their level of trust towards other social network users, and consequently, this provides the ground truth data. However, the same approach for evaluating global user properties is not applicable as those properties are 1:N user properties, and only users who have full knowledge of all other social network members are able to answer the ground truth questions. Considering that today's online social networks are quite sparse [12, 13] and only social network platform operators have comprehensive data on its respective users [14], new methods are obviously needed for evaluating the modelling of global user properties in complex social networks.

This paper is a contribution to existing literature in that it proposes a novel methodology for evaluating algorithms that calculate social influence in complex social networks. The proposed methodology (i) compares algorithms that rely solely on available ego-user data for calculating ego-user social influence and (ii) identifies the most accurate and precise algorithm for predicting social influence. To the best of our knowledge, there are no other methodologies for evaluating algorithms that calculate social influence in complex social networks which are in addition able to identify the most accurate and precise calculation algorithm. The paper demonstrates different phases of the proposed methodology using a case study to calculate social influence by evaluating accuracy and precision of four different algorithms that calculate social influence.

The paper follows a specific structure. Section 2 presents the concept of *social influence* in online social networks and related work in the respective field, including the use of SmartSocial Influence algorithms. In Section 3, a methodology for evaluating the method of calculating social influence in complex social networks is introduced, and its use is demonstrated in Section 4. Next, Section 5 discusses the impact of the proposed evaluation methodology and elaborates on possible implications of identifying the best-performing social influence algorithm. Section 6 provides a conclusion, focusing on constraints of the proposed approach as well as further work in the field. The questionnaires used in method for evaluating social influence are provided in the appendix to this paper.

## 2. Background on Previous Work

Looking back on previous work, the paper first explains the concept of social influence in online social networks and provides examples of the main services stemming from

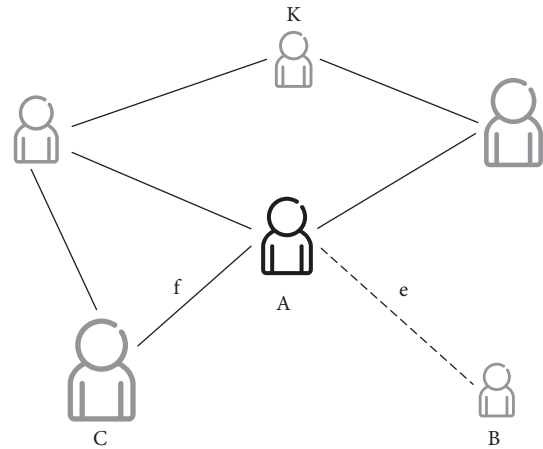


FIGURE 1: Graphical illustration of influence in a social network.

social influence. The second part in this section introduces SmartSocial Influence algorithms, a specific class of algorithms for calculating social influence.

**2.1. Social Influence in Online Social Networks.** Social influence is “a measure of how people, directly or indirectly, affect the thoughts, feelings and actions of others” [15]. It is a topic of interest in both sociology and social psychology, and more recently in information and communication technology (ICT), computer science, and related fields. Social influence in online social networks has seen a great rise with services such as Klout [16], Kred [17], PeerIndex [18], or Tellagence [19], all of which have demonstrated the central role of empowered users in everyday lives of ordinary people [20]. With over 620 million users scored and serving over 200 thousand business partners, Klout is an important service that is aimed at bringing influencers and brands together. Klout defines influence as “the ability to drive action” and measures it on a scale from 1 to 100, based on data from more than ten of the most popular social networking services (SNSs). As of 2017, the two most influential Klout users are Barack Obama and Justin Bieber with Klout scores of 99 and 92, respectively [21]. Figure 1 illustrates the concept of social influence using an example of six users interconnected in a social network through two types of connections. Ego-user *User A* has a greater social influence than *User B*, but less than *User C*, as denoted by the size of graphical symbols representing them. Users in the network are connected through different types of connections (e.g., *User A* and *User C* are Facebook friends, while *User A* and *User B* communicate using a text messaging service).

Numerous studies, tests, experiments, and research over a period of more than 50 years have led to various approaches in elaborating social influence [22–27]. Although rooted in social psychology and sociology, the topic of social influence has independently spread to modern online social networks with the rise of the Internet era [28].

**2.2. SmartSocial Influence Algorithms.** The paper compares the prediction accuracy and precision of four social influence algorithms—SLOF, SAOF, SMOF, and LRA—which all

TABLE 1: Comparison of Klout, Kred, and the *SmartSocial Influence*.

	Klout	Kred	SmartSocial Influence model
Scope of observation	“Big Brother”	“Big Brother”	Ego-network
Scale	1 to 100	1 to 1000	0 to 100
Support for multiple SNSs	✓	✗	✓
Openly published algorithm	✗	✓	✓
Telco network data-source	✗	✗	✓

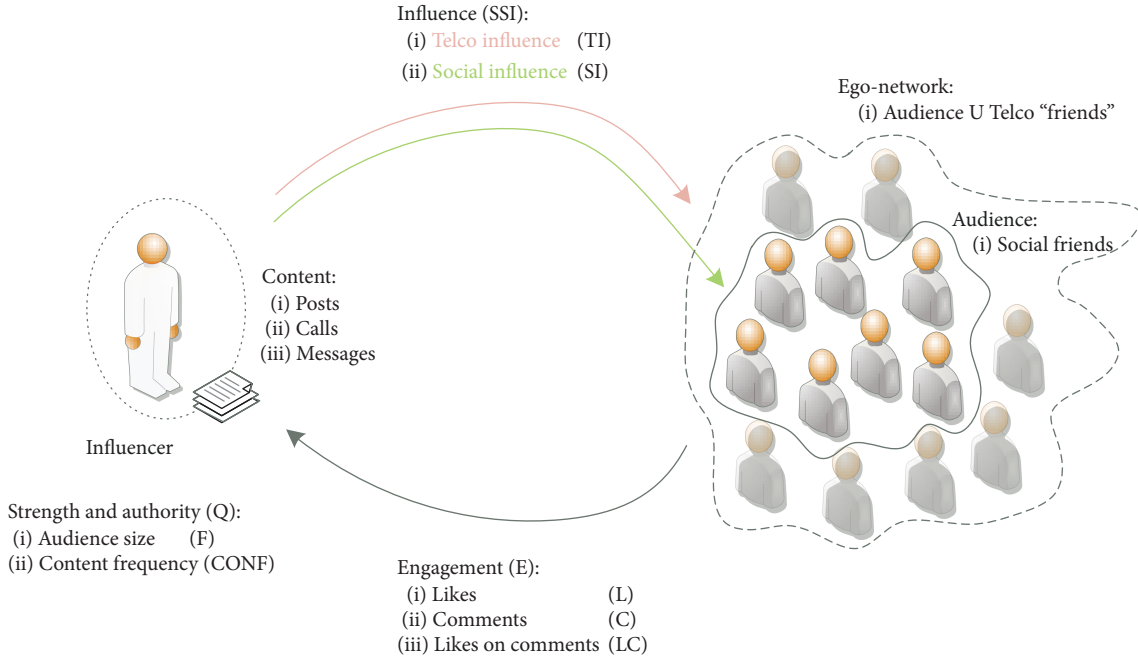


FIGURE 2: The SmartSocial Influence model.

belong to the *SmartSocial Influence* class of algorithms. SmartSocial Influence [4] is an approach to social influence modelling which takes into account the following goals: (i) inferring social influence of users based on their data retrieved from multiple, heterogeneous data-sources, namely, data on social networking services combined with data from telecommunication operators, and (ii) a multidisciplinary approach rooted in previous approaches to social influence modelling in the fields of social psychology and sociology, as well as ICT. The important difference to common approaches in social influence modelling (e.g., Klout and Kred) is the scope of observation. Unlike the SmartSocial Influence approach, the approach common to both Klout and Kred is their “Big Brother” scope of observation—they endeavor to collect vast amounts of user data to model influence that may expand beyond activities in a user’s first-degree ego-network (Table 1). Moreover, the SmartSocial Influence approach operates on smaller datasets as its scope of observation is limited to the user’s ego-network alone (Figure 1—*User B* and *User C* are in the ego-network of *User A*, but the same is not true for *User K*).

Furthermore, SmartSocial Influence explores social influence in social networks both from the *structural*

(*Structural models* analyse network structure using metrics such as degree, betweenness, and closeness centrality [29, 30], as well as eigenvector centrality [31]) and *behavioural* (*Behavioural models* analyse interaction among users, e.g., how connected users propagate or repost content, how many of them like or comment on it, or the way they engage in conversations [32, 33]) perspective—by analysing node degree (i.e., audience size), content type (i.e., quality), and content frequency (i.e., time-based longitudinal quantity) of interactions between users. Figure 2 illustrates this by identifying the main SmartSocial entities:

- (i) *Influencer*—the ego-user exerting the influence
- (ii) *Content*—items (SNS posts, calls, or messages) created by the Influencer in the SNS or telecom network
- (iii) *Ego-network*—all users who communicate with the Influencer
- (iv) *Audience*—users of a SNS who observe and *engage* with the Influencer’s *content*, a subset of the Influencer’s *Ego-network*

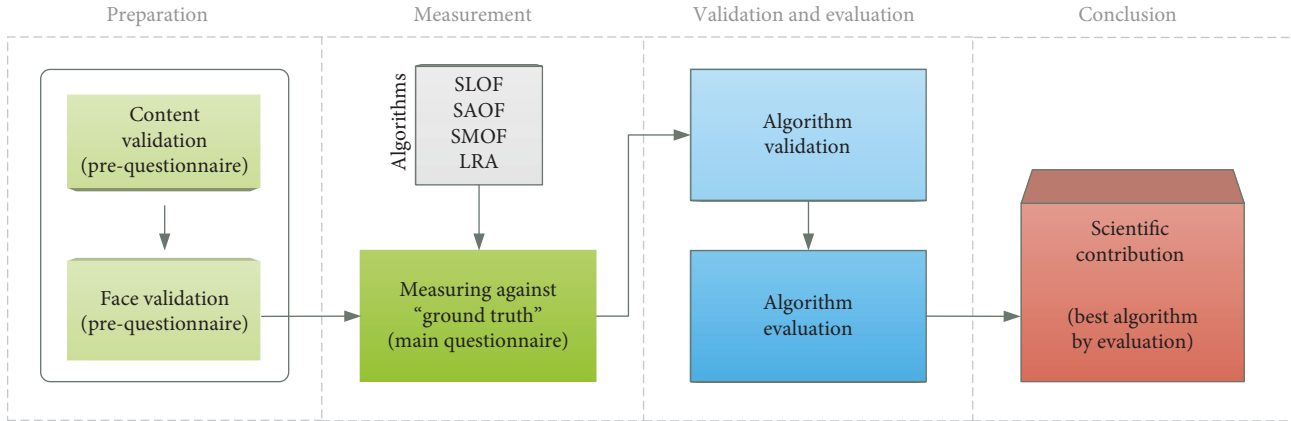


FIGURE 3: Proposed methodology for evaluating algorithms that calculate social influence in complex social networks with its four distinct phases.

An important feature is the difference between the Influencer’s *Ego-network* and the *Audience*. The Audience comprises users *connected to* the Influencer through the same SNS. The Influencer may have multiple Audiences, but for a single SNS there is only one. On the other hand, a user’s *Ego-network* comprises all users with whom the Influencer has *communicated* in the combined telecom network and SNSs. Hence, the Audience is a subset of the Influencer’s *Ego-network*. Definitions of the *relationships* between Smart-Social entities are as follows (Figure 2):

- (i) *Influence (SSI)*—*SmartSocial Influence* comprised of *TI* and *SI*
- (ii) *Telco Influence (TI)*—Influencer’s effect on the respective *Ego-network*
- (iii) *Social Influence (SI)*—Influencer’s effect on the respective *Audience*
- (iv) *Engagement*—the action taken towards the Influencer’s *content* by the *Audience* through the SNS (in the form of *likes, comments, or likes on comments*)

In short, the purpose of SmartSocial Influence algorithms is to quantify the number of engagements or interactions for a user’s publication or post (e.g., likes) with respect to the size of the audience (i.e., number of friends). In other words, a highly influential user of a SNS will have numerous posts and will be massively engaged by a large share of the respective audience.

Let us further explain the SmartSocial Influence concept on the social graph shown in Figure 1. The Influencer (or ego-user) is *User A*, connected to other users in the respective *Ego-network* (e.g., to *User B* and *User C*). *User C* is part of *User A*’s Audience; *User B* is not. Therefore, *User B* is not able to “perceive” *A*’s influence—but merely contributes to it. *User A*’s influence is defined as a property of node *A*, exerting influence on all other users in the respective Audience (part of the *Ego-network*) and described as a 1 : N relationship. This means that *User K* (not part of the Audience or *Ego-network*) is not able to “perceive” *A*’s influence.

If *User A* and *User K* were connected through the same SNS, this would then be possible. Influence is graphically represented through the size of the graphical symbol, with *User C* being the most influential in *User A*’s *Ego-network* (and Audience). In other words, Influencer’s influence is “perceivable” only by members of the Audience, whereas for the entire *Ego-network* it is “a result of contribution.” Nonaudience users of the *Ego-network* cannot “perceive” influence since they do not possess the means to do so.

More details on calculating SmartSocial Influence, along with pseudocodes of algorithms SLOF, SAOF, SMOF, and LRA, are available in [4].

### 3. Proposed Methodology

As previously mentioned, the SLOF, SAOF, SMOF, and LRA algorithms produce meaningful and usable results regarding one’s social influence [4, 34, 35]. However, to prove that the results hold true, they have to be *validated*.

*Validity* is the degree to which *evidence* supports interpretations of test scores [36]. In other words, validation reveals whether the respective algorithm produces *correct* results (that hold evidence of being truthful in the largest amount of cases) for social influence. Subsequently, evaluation leads to discovery of the best algorithm, that is, the most *accurate* and *precise* algorithm. Differences between these two terms are explained in detail in Section 3.3. In short, the methodology for evaluating algorithms provides insights into identifying the best social influence algorithm.

The proposed methodology takes place in *four phases* (Figure 3): (i) the first phase is a *preparatory* step; (ii) the second phase involves taking *measurements* of the performances of algorithms with respect to “ground truth”; (iii) the third phase is *validatory* and *evaluatory* regarding the algorithms; and (iv) the last phase is *conclusive*.

Namely, the first phase involves *pre-questionnaires*, essential to forming the *main questionnaire* in a scientifically valid manner in the second phase. The third phase uses the *main questionnaire* to validate the algorithms, and the fourth phase provides a conclusion by identifying the *best* algorithm.

The four phases of the proposed methodology for evaluating algorithms that calculate social influence in complex social networks are described in more detail further on.

**3.1. Evaluation of Social Influence Calculation: Preparatory Phase.** The *main questionnaire* (MQ) is employed to validate social influence results produced by each of the algorithms. Just as any other questionnaire, the MQ is a test given to respondents in form of *questions*. Each question represents a test *item* ( $Q_i$ ). To make sure the (MQ) measures what it is *supposed to measure*, two different facets of validity have to be satisfied for each of the selected *items* (questions). First is *content validity* and second is *face validity*. These are tools incorporating a rigorous scientific method—validation of an artefact, in this case, the *main questionnaire* (MQ).

**3.1.1. Content Validity Test.** Content validity [37], also known as *logical validity*, indicates to what degree each of the test *items* measures what it should be measuring (i.e., test content). A test created by a single author may or may not be content valid, given that an author may be *biased* and create a test that does not measure what it is supposed to.

Therefore, as the content validity test, a number of individuals who are sociology/psychology researchers were asked to validate questions *directly*, after being provided with definitions of social influence.

**3.1.2. Face Validity Test.** Items that pass the content validity process are advanced into the *face validity* process. In contrast to content validity, face validity does not show how good the test measures what it is supposed to measure, but what it *actually appears to measure*. In other words, despite the scientific rigour of content validity, it is face validity that ensures correctness of the *interpretation* of questions and their *relevance* of the participants' answers. Some researchers argue that *face validity* is somewhat unscientific [38]; nonetheless, the test is face-valid if it *seems* valid and meaningful to the participants taking the test, decreasing its overall bias levels [39].

For that purpose, after establishing content validity with the *content validity pre-questionnaire* ( $PQ_{CV}$ ), an additional pre-questionnaire should be used for establishing face validity ( $PQ_{FV}$ ) of items  $Q_{CV_i}$ . The basic principle remains the same as with content validity test, but the implementation is somewhat different. Since those who are not sociology/psychology researchers are not familiar with definitions and concepts of social influence, asking them to validate questions *directly* is inappropriate. Providing them with *definitions of social influence* beforehand, as is the case with the sociology/psychology researchers, may *distort* the responses and undermine face validity. (This design approach, to the best of its ability, endeavors to mitigate the *Hawthorne* (or *Reactivity*) effect [40], the *Observer-expectancy effect* [41], and to the greatest extent the bias resulting from the *Demand characteristics* [42].) Therefore, as the face validity pre-questionnaire tests a number of nonexpert individuals who are not sociology/psychology researchers, they were asked to validate questions *indirectly*, without being provided

with definitions of social influence beforehand in order to avoid bias.

**3.2. Evaluation of Social Influence Calculation: Measurement Phase.** The results of the content-validity and face-validity tests are the basis for compiling the *main questionnaire* (MQ). The MQ serves as the ground truth or the “golden standard”—its purpose is to validate and evaluate algorithms SLOF, SAOF, SMOF, and LRA. Each question  $Q_i$  in the MQ requires the participant to read an “imaginary Facebook post” and choose between Facebook friends who exert a greater personal influence (either on *emotions*, *actions*, or *behaviours* as described in the question).

What each question  $Q_i$  (the total number of questions in the questionnaire MQ is denoted as  $|MQ|$ ) explores is, in fact, the *greater social influencer* among two Facebook friends in each pair. All of the questions pose the same question *indirectly*—which of the two Facebook friends has greater social influence? A total of  $|Pair|$  Facebook-friend pairs are offered as answers to each question. These pairs are permuted between questions, to avoid participant boredom and fatigue. All  $|Pair|$  friend pairs in  $|MQ|$  questions equal  $|Pair| \times |MQ|$  observations *per participant*. Combined with  $|PAR|$  participants, there are a total of  $|Pair| \times |MQ| \times |PAR|$  observations *per algorithm*. Observations were carried out in the manner described below.

First, consider a single participant, denoted as  $PAR_j$ . For each Pair offered as answers to questions, there are two Facebook friends—left FB friend and right FB friend. Each Facebook friend has four social influence scores attached to it, as calculated per respective algorithm ALGO— $SI_{SLOF}$ ,  $SI_{SAOF}$ ,  $SI_{SMOF}$ , and  $SI_{LRA}$ . Calculating the difference between social influence scores (SI) of the left and right Facebook friends yields a new measure defined as

$$\begin{aligned} \Delta_p(\text{Pair}, \text{ALGO}) &= SI_{\text{ALGO}}(\text{left FB friend}) \\ &\quad - SI_{\text{ALGO}}(\text{right FB friend}), \end{aligned} \quad (1)$$

$$p \in \{1, 2, \dots, |Pair| \times |MQ| \times |PAR|\},$$

where

$$\begin{aligned} \text{ALGO} &\in \{\text{SLOF}, \text{SAOF}, \text{SMOF}, \text{LRA}\}, \\ (\text{left FB friend}, \text{right FB friend}) &\in \text{Pair}, \end{aligned} \quad (2)$$

$$\text{Pair} \subseteq \text{FB Friends}(PAR_j), j \in \{1, 2, \dots, |PAR|\}.$$

Since social influence scores (SI) attain values between 0 and 100,  $\Delta_p$  attains values between  $-100$  and  $100$ . The value  $\Delta_p$  in fact represents “measurement of certainty” with which the respective algorithm determines that the left FB friend has *greater* social influence than the right FB friend has, or vice versa. For example,  $\Delta_p = -42$  means that “the right FB friend is more influential than the left FB friend by 42.”

An algorithm that *correctly* measures a *more influential* Facebook friend in a Pair (with respect to the participant's answer) gets *rewarded*, whereas the algorithm that *incorrectly* measures it gets *punished*. This means that  $\Delta_p$  is a *single measurement*.

How do algorithms get *rewarded* or *punished* with respect to a *correct* or *incorrect* measurement? Let us define the measurement score of a Pair as

$$ms_p = \varepsilon_p \cdot \frac{\Delta_p}{100}, \quad (3)$$

where for each Pair of Facebook friends found in the “ground truth,”

$$\varepsilon_p = \begin{cases} -1, & \text{if more influential is the friend on the right} \\ 1, & \text{if more influential is the friend on the left.} \end{cases} \quad (4)$$

This simply means that for *correctly* measuring the more influential Facebook friend in a given Pair, an algorithm receives a measurement score of  $ms_p = +|\Delta_p|/100$ . In contrast, it receives  $ms_p = -|\Delta_p|/100$  for an *incorrect* measurement. (One might argue whether this approach is justified. Replace the “algorithm” with a Geiger instrument for measuring radioactivity and consider the logic of “measurement confidence” as follows. If the Geiger instrument is correct, it should be rewarded. If not, it should be punished. Now, imagine an instrument that measured  $\Delta_p = 100$  between two people, determining the person on the left +100 more radioactive than the person on the right. If incorrect, the algorithm should be severely punished—for potentially endangering the person on the right. If correct, it should be maximally rewarded for saving the life of the person on the left. The same holds true for smaller measurements (e.g., moderate punishment/reward for  $\Delta_p = 5$ ) and all other variations.)

**3.3. Evaluation of Social Influence Calculation: Validation and Evaluation Phase.** In the phase that follows, it is important to distinguish between the two constructs—validation and evaluation of algorithms. Validation yields *proof* that the algorithm produces *sound* and *truthful* social influence scores with respect to participants’ answers, which are taken as the “ground truth.”

The single criterion for validating an algorithm is as follows:

- VI. The overall amount of *correct* measurements (from the measurement phase) is greater than half (50%) with respect to participants’ answers.

In other words, the ALGO algorithm is valid if its *average* measurement score  $ms_p$  is greater than zero by a statistically significant margin. Statistically speaking, this shows that the algorithm did not *bet* and correctly determined the greater social influencers by *sheer chance* alone, but by being aligned with the ground truth found in the participants’ answers. Since validation is a binary variable, an algorithm can either be *valid* or *invalid*. There is no comparison between the algorithms in terms of their validity; one cannot be *more valid* than the other.

Evaluation, on the other hand, enables ranking of the algorithms. As can be seen, the algorithm with the *greatest* amount of both correct and “confident measurements” (utilising greater  $|\Delta_p|$ ) is declared the *most truthful*.

Averaging over all of the Facebook friend pairs, the *most truthful* algorithm can be identified using the evaluation criteria prioritized as follows:

- E1. The greatest *average* measurement score  $ms_p$
- E2. The smallest *spread* (also known (in statistics) as variability, scatter, or dispersion) of measurement scores  $ms_p$  in the distribution

To paraphrase using statistics vocabulary, the criteria for the *most truthful* algorithm would be as follows:

- E1. The algorithm with the greatest *accuracy*
- E2. The algorithm with the greatest *precision*

The first criterion assumes the *average* to be *true* as a point-estimation through a sufficient amount of data points (in our case, exactly 1,152 measurement scores per algorithm (12 Facebook friend pairs in 6 questions given to 16 participants)) Let us be clear that each algorithm is completely precise with respect to repeating a *single* measurement; that is, repeating the measurement of the same Pair will always return an *identical* value. Precision is *not* used in the sense of an internally intrinsic measure, but in comparing against the ground truth. It is a question of how precise an algorithm is when put up against participants’ answers in the real world.

**3.4. Evaluation of Social Influence Calculation: Conclusion Phase.** Importantly, the underlying research problem should be evident—to correctly determine the *more influential* of the two Facebook users, with the ultimate goal of ranking them according to their social influence score SI. Knowing a certain SI score is inadequate per se unless comparable to another SI score. In the most general sense, this approach to evaluating relates to *maxDiff* and *best-worst choice* methodologies [43, 44] and is used to establish which of the algorithms produces the best results in a *relative (ranked)*, not *absolute (nonranked)* manner.

## 4. Methodology in Practice—Evaluating the SmartSocial Algorithms

In the previous section, four phases of the proposed methodology for evaluating calculation of global user properties in complex social networks were explained. In this section, use of the proposed methodology will be demonstrated using a case study of calculating social influence by evaluating the accuracy and precision of four social influence algorithms—SLOF, SAOF, SMOF, and LRA—which all belong to the SmartSocial Influence class of algorithms.

**4.1. Evaluation of the SmartSocial Algorithms: Preparation Phase**

**4.1.1. Content Validity Test.** To avoid bias in selecting questions for the MQ, a *content validity pre-questionnaire* (PQ<sub>CV</sub>) has to be employed. In our case, the PQ<sub>CV</sub> was given to a group of 22 experts (All the experts were graduates from the Faculty of Humanities and Social Sciences, University of Zagreb, familiar with the field of social influence through (social) psychology and sociology classes and research.) (EXP<sub>k</sub>) on the subject of *social influence*. Before answering questions, experts were shown important

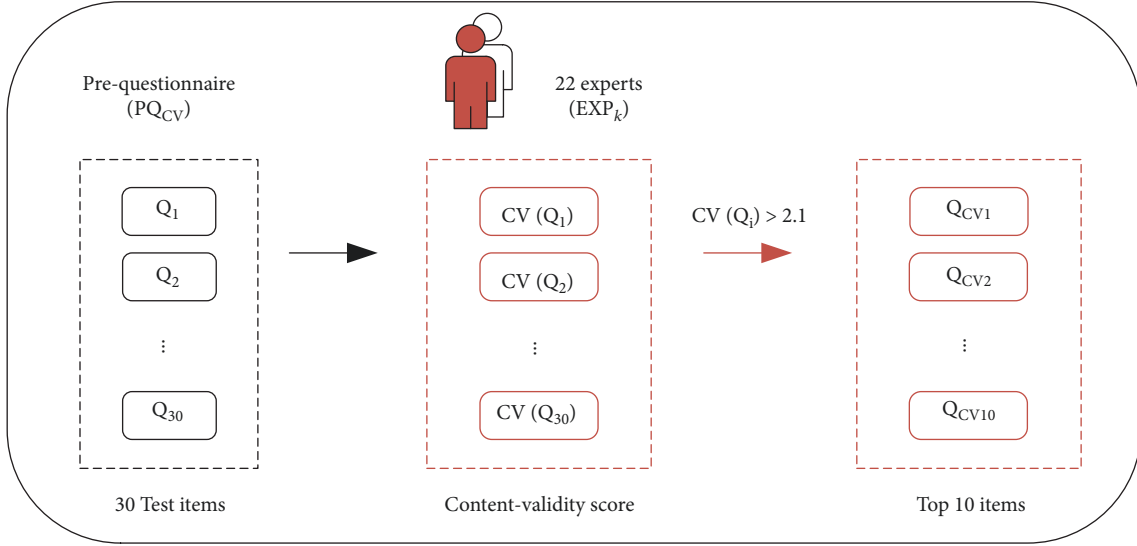


FIGURE 4: The process of content validation for questions  $Q_i$ .

definitions of social influence, which ensured that all of them utilize the same underlying concept.

The content validation process is shown in Figure 4. Of the 30 questions ( $Q_i$ ) from the total in  $PQ_{CV}$ , only the *top best-rated 10* passed through to the next step of validation. An expert was given the opportunity to score each question  $Q_i$  on a scale of 1 to 5, depending on how well it explored social influence in line with the given definitions. After  $PQ_{CV}$  was finished, each question score was averaged across all experts. This produced the content-validity score for a particular item, denoted as  $CV(Q_i)$ .

According to [37], for a group of 22 experts, each item has to be rated above 0.42 out of a maximum of 1 in order to pass as *valid for content*. On a scale of 1 to 5, this equates to 2.1, which is the threshold for selecting a question  $Q_i$  as content-valid. In other words, the statement  $CV(Q_i) > 2.1$  must hold true for each of the questions  $Q_i$  to be content-valid.

All questions  $Q_i$ , as well as their respective  $CV(Q_i)$  scores, can be found in Appendix B. *Pre-questionnaire (content validity)*. Of the 30 questions in the  $PQ_{CV}$ , 29 questions passed the content validity test and the top 10 with the highest  $CV(Q_i)$  scores were selected for the next phase—the face validity test ( $PQ_{FV}$ ).

**4.1.2. Face Validity Test.** In this phase, a pre-questionnaire of top 10 questions that passed  $PQ_{CV}$  was given to 22 individuals who were not experts ( $NEX_x$ ) on the subject of *social influence*. As is evident in Appendix C, these questions do not address *social influence* per se in any shape or form but ask the nonexpert to read an “imaginary Facebook post,” and each time a different one. The “post” is followed by a description regarding the effect either on personal *emotions*, *actions*, or *behaviours* with respect to a given imaginary Facebook post. Next, the nonexpert is instructed to choose which Facebook friend would cause a greater effect either on *emotions*, *actions*, or *behaviours* as described in the question. Facebook friends are presented in pairs, with each question

holding the *identical* four Facebook-friend pairs as answers. The face validation process is shown in Figure 5.

A note here is that pairs themselves are not important in this phase; the point of  $PQ_{FV}$  lies in a “hidden” 11th question which reveals itself to the nonexpert once  $PQ_{FV}$  is finished. This last question provides the necessary *definitions of social influence* and then asks the nonexpert to choose—in accordance with the provided definitions—the *more influential* friend among the *same* four Facebook-friend pairs used beforehand. In essence, it provides a filter of “correct answers” for all of the previous 10 questions. Details about the face validity test and face validity scores  $FV(Q_{CVi})$  with respect to the 10 questions in  $PQ_{FV}$  can be found in Appendix C.

Exactly *four* Facebook-friend pairs are offered as answers in each  $Q_{CVi}$  because questions can have anything between 0 and 4 “correct answers,” based on “criteria” in the 11th question. Upon shifting the scale by +1, this yields a scale from 1 to 5, which corresponds directly to the previously used scale in  $PQ_{CV}$ , which is important for equal treatment of both *content-* and *face-validity*. Again, each question is given a score  $FV(Q_{CVi})$  as an average across all scores of the 22 nonexperts.

Finally, the top 5 questions were chosen for MQ, with an additional  $Q_{MQ6}$ . This additional question was important for MQ as it involved a topic referring to the *mobile telecommunication operator*. In fact, it is both content- and face-valid (see Appendix B and Appendix C).

**4.2. Evaluation of the SmartSocial Algorithms: Measurement Phase.** To avoid fatigue [38], participants in the *main questionnaire* MQ were asked 6 questions, leading to  $|MQ| = 6$ . The highest scored questions that passed *content validity* as well as *face validity* pre-questionnaires were chosen to be part of the MQ, as described in the previous subsection. A total of 16 participants participated in the MQ, leading to  $|PAR| = 16$ . A total of 12 Facebook-friend pairs were offered as answers to

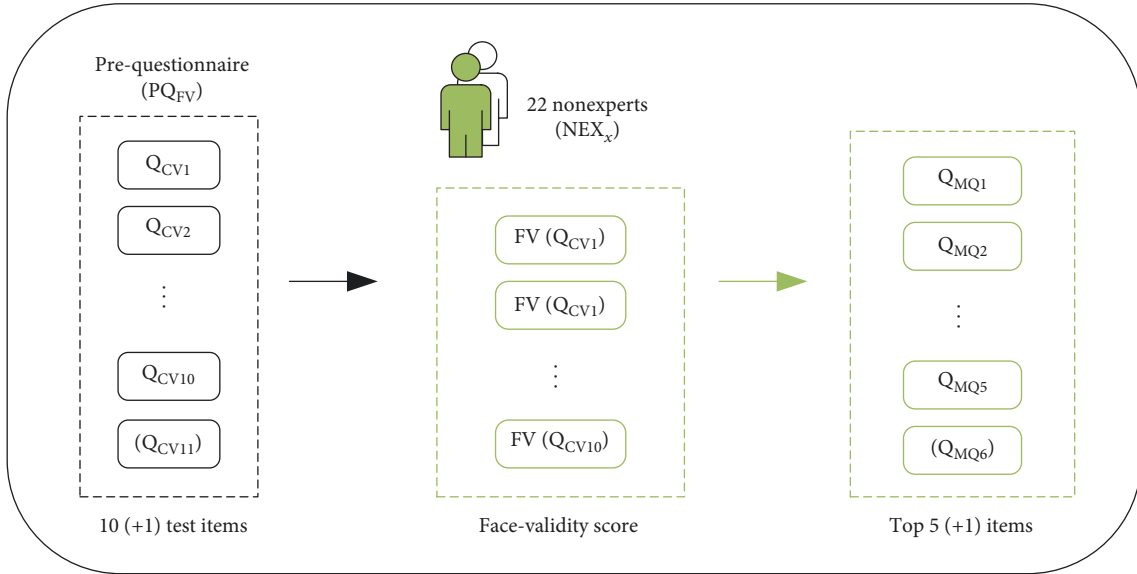


FIGURE 5: The process of face validation for questions  $Q_{FV,i}$ .

each question, leading to  $|\text{Pair}| = 12$ . All 12 friend pairs in 6 questions equal 72 observations *per participant*. Combined with 16 participants, there are a total of 1152 observations *per algorithm*.

More details about the specific questions which were part of the MQ are given in Appendix D, while more details about the metrics used in the measurement process are given in Section 3.2.

#### 4.3. Evaluation of the SmartSocial Algorithms: Validation and Evaluation Phase

**4.3.1. Validation Using Measurement Scores.** Figure 6 shows the distribution of final measurement scores  $ms_p$  for the SLOF algorithm. Individual measurement scores are retrieved for *each pair* of Facebook friends and can attain values in the range  $[-1, 1]$  (i.e.,  $+\Delta_p/100$  or  $-\Delta_p/100$  for a certain pair). Given that there are 6 questions with 12 pairs across 16 participants, the distribution shows a *total of 1152 measurement scores*.

At the given resolution, it becomes evident that the SLOF  $ms_p$  distribution is multimodal, having five *modes*. This observation holds true for other (SAOF, SMOF, and LRA)  $ms_p$  distributions as well. The reason lies in the somewhat nonrandom method of selecting Pairs and their respective differences in SI, which produces a nonnormally distributed  $\Delta_p$  that sometimes overlaps or repeats, producing several *modes*. (Although desirable, it was not feasible to select truly random values of  $\Delta_p$  due to the fact that the SI score distributions from SmartSocial Influence algorithms are not normal. Particularly in the case of the SLOF algorithm, a high-kurtosis distribution of SI scores exists, resulting in the measurement score  $ms_p$  distribution displaying “groups” based on similar  $\Delta_p$ .)

It becomes evident that the majority of *measurement scores* ( $ms_p$ ) are greater than zero. To be exact, 58% of them

are positive. This means that SLOF correctly determined the greater influencer in 668 out of 1152 pairs. Validity is similar to SLOF for SAOF (Figure 7), SMOF (Figure 8), and LRA (Figure 9) as well. They correctly determined 61%, 61%, and 64% of greater influencers in pairs, respectively.

To prove the validity of each algorithm, let us formally use statistical hypothesis testing in the following manner. Consider the statement “SLOF algorithm works by sheer guessing of the correct measurements” as the *null hypothesis* ( $H_0$ ) being tested. The *test statistic* is “the number of correct measurements.” Let us set the *significance level* ( $\alpha$ ) at 0.01. The *observation* is “668 correct measurements out of 1,152.”

Therefore,  $p$  value is the probability of observing between 668 and 1152 *correct* measurements with the null hypothesis being true. Calculation of  $p$  value is as follows [45]:

$$p \text{ value} = \left(\frac{1}{2}\right)^{1152} \cdot \sum_{d=668}^{1152} \binom{1152}{d}, \quad (5)$$

which equals approximately  $3.28 \cdot 10^{-8}$ . In other words, guessing more than 58% out of 1152 measurements correctly ( $p$  value) is statistically very improbable. Since  $p$  value  $\ll \alpha$ , the null hypothesis is strongly rejected.

Therefore, the *logical complement* of the null hypothesis ( $\neg H_0$ ) can be accepted, stating that “the SLOF algorithm does *not* work by the sheer guessing of correct measurements,” which validates the algorithm. Considering that the other algorithms (SAOF, SMOF, and LRA) have even greater *test statistics*, the null hypothesis can be safely rejected for them as well. The summary is shown in Table 2.

To summarise, all of the algorithms were *successfully validated* by satisfying the single criterion for validation (V1). Note that the percentages of correct measurements are *not* comparable across the algorithms—which may be 58% percent of “correct pairs” for SLOF, and is not comparable with 64% of “correct pairs” for LRA, given that pairs are associated with different “weights” ( $\Delta_p$ ) to them. This is the reason, for



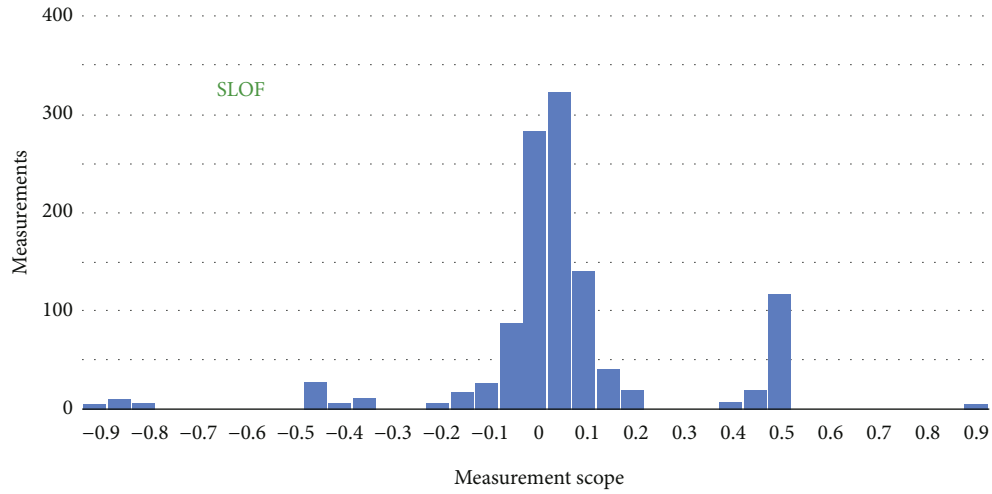


FIGURE 6: Distribution of measurement scores for the SLOF algorithm. Extreme values and outliers are not shown; the distribution shows 936 (of 1152) or 81% of all measurement scores for SLOF.

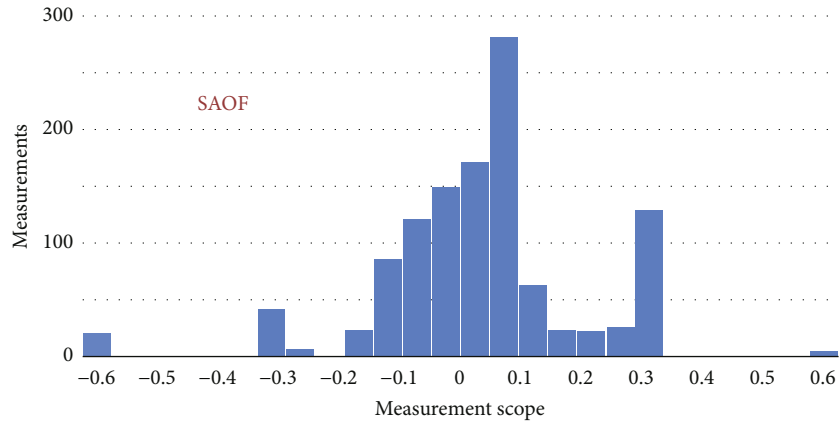


FIGURE 7: Measurement score distribution for the SAOF algorithm.

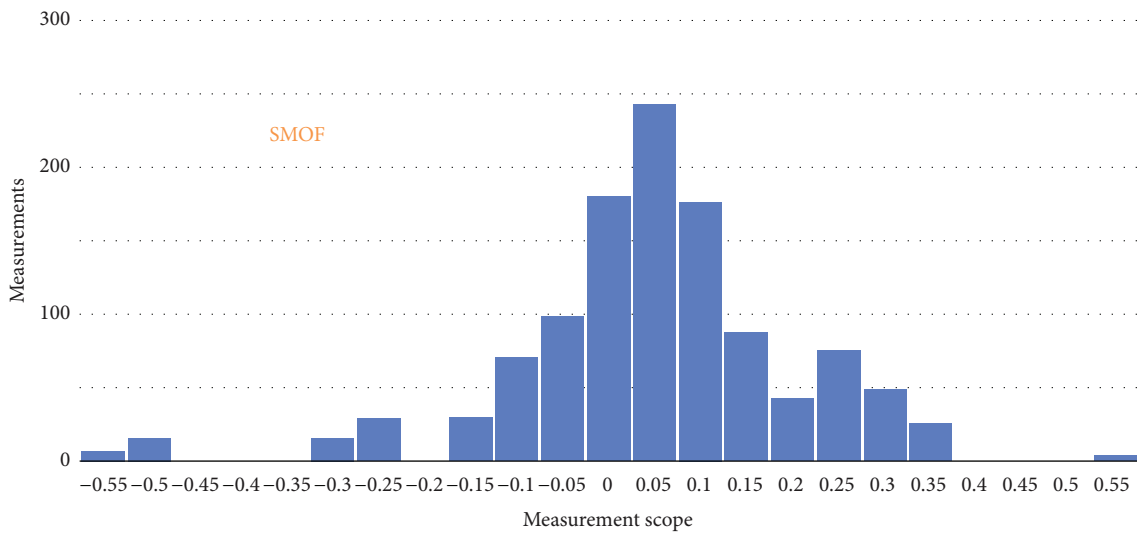


FIGURE 8: Measurement score distribution for the SMOF algorithm.

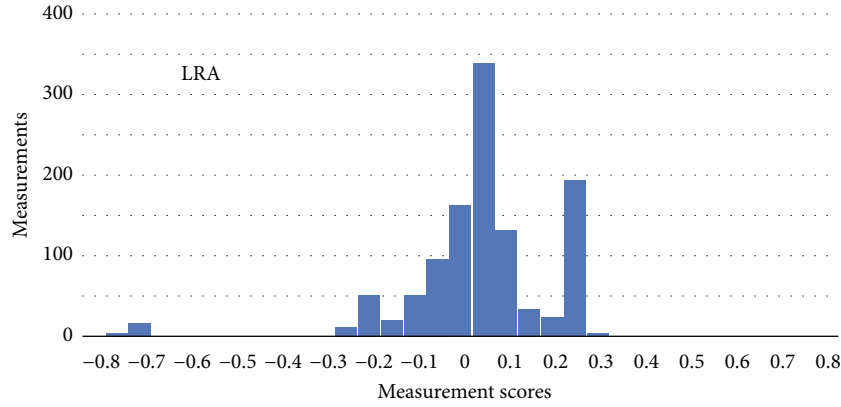


FIGURE 9: Measurement score distribution for the LRA algorithm.

TABLE 2: Summary of statistical hypothesis testing with the goal of social influence algorithm validation.

Testing parameters	SLOF	SAOF	SMOF	LRA
$H_0$ (null hypothesis)	Algorithm being tested works by the sheer guessing of correct measurements			
Test statistic	Number of correct measurements			
Significance level ( $\alpha$ )	0.01			
Observation (number of correct measurements)	668 in 1152	706 in 1152	706 in 1152	733 in 1152
$p$ value	$3.28 \cdot 10^{-8}$	$9.08 \cdot 10^{-15}$	$9.08 \cdot 10^{-15}$	$8.52 \cdot 10^{-21}$
Conclusion ( $\neg H_0$ )	Algorithm being tested does <i>not</i> work by the sheer guessing of correct measurements.			
Validation successful	✓	✓	✓	✓

example, that LRA is not *more valid* than SLOF. The mentioned challenge of ranking is a task for *evaluation*, not validation, as will be explained in detail in the following subsection.

**4.3.2. Evaluation by Comparison.** Figure 10 shows a boxplot of *measurement scores*  $ms_p$  for each algorithm. Although all four algorithms belong to the same SmartSocial Influence class of algorithms, LRA is denoted with a different color (light blue) since it is the only solely literature-based algorithm (i.e., the benchmark algorithm) and the predecessor to SLOF, SAOF, and SMOF (which are the upgraded versions [4]). The measurement scores are retrieved *per pair*, as either correct ( $+\Delta_p/100$ ) or incorrect ( $-\Delta_p/100$ ). A summary of the boxplot is given in Table 3.

Let us first consider the *first criterion* for evaluation (E1)—the greatest *average* measurement score ( $\overline{ms}_p$ ), denoted with a “+” symbol in Figure 10. The greatest  $\overline{ms}_p$  is found in SLOF and equals 0.0358. The smallest  $\overline{ms}_p$  is found in SMOF and equals 0.0240. In between are SAOF with 0.0271 and LRA with 0.0250  $\overline{ms}_p$ , respectively. Observing the averages, SLOF and SAOF are evaluated as more truthful, while SMOF as less truthful than their predecessor LRA—showing a +43.4%, +8.7%, and −3.8% difference in  $\overline{ms}_p$ , respectively. Based on the *first criterion* used for evaluation (E1), the two algorithms—SLOF and SAOF—demonstrated and clearly showed significant improvements

over their predecessor, the LRA algorithm, and provided a scientific contribution. In other words, this means that, on average, SLOF and SAOF surpass LRA (accuracy) in correctly determining the *greater* influencer between the two—while considering the *differences* in their respective SI scores.

Let us now consider the *second criterion* for evaluation (E2)—the smallest *spread* of *measurement scores*. Statistically speaking, there are various estimators that estimate the *spread* of values across a distribution. They are called *estimators of scale*, in contrast to *estimators of location* (i.e., such as mean or median) [46–48]. The view is that the first criterion used for evaluation utilized the sample mean (average) as an *estimator of location* to rank the algorithms.

When dealing with a large amount of data or variable measurements, *outliers* and *extreme values* are common, along with certain departures from parametric distributions. To be “resistant” to outliers or underlying *parameters* of a distribution (namely nonnormality, asymmetry, skewness, and kurtosis), robust estimators of scale have to be employed [49]. In such situations, performance of robust estimators tends to be *greater* than their nonrobust counterparts (such as *standard deviation* or *variance*) [50].

On the other hand, statistical efficiency (In (descriptive) statistics, *efficiency* of an estimator is its performance with regards to the (minimum) necessary number of observations. A more efficient estimator needs fewer observations; given that the amount of observations is not an issue with measurement scores, lower efficiency is not problematic.) of robust

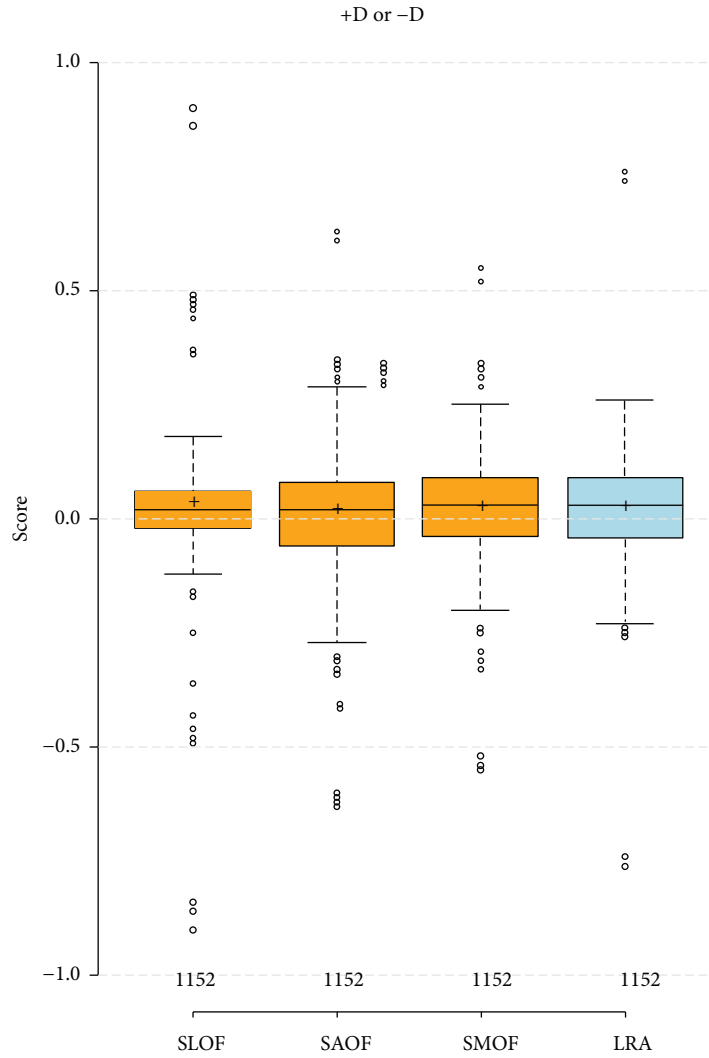


FIGURE 10: Boxplot of measurement scores for each algorithm. Boxplot uses values that are less than a 1.5x interquartile range from the 1st and/or 3rd quartile for the lower and upper whiskers (as defined by Tukey [51]); box lower-bound is the 25th percentile, middle-bound is the median, and upper-bound is the 75th percentile, and “+” denotes an average (mean) value. Plotted using BoxPlotR [52].

TABLE 3: Summary of measurement scores as boxplot statistics.

Boxplot statistic	SLOF	SAOF	SMOF	LRA
Number of measurement scores	1152	1152	1152	1152
Maximum value	0.90	0.63	0.55	0.76
Upper whisker	0.18	0.29	0.25	0.26
75th percentile	0.06	0.08	0.09	0.09
Average (mean)	<b>0.0358</b>	<b>0.0271</b>	<b>0.0240</b>	<b>0.0250</b>
Median	0.02	0.03	0.03	0.03
25th percentile	-0.02	-0.06	-0.04	-0.04
Lower whisker	-0.12	-0.27	-0.20	-0.23
Minimum value	-0.90	-0.63	-0.55	-0.76

estimators tends to be smaller. Caution should be used when seeking “resistance” to outliers—sometimes, they carry *very important* information, such as the early onset of *ozone holes* which were initially rejected as outliers [53]. Since

*measurement scores* are a large amount of nonparametrically distributed data containing outliers, utilization of *robust estimators of scale* is mandatory.

A thorough description of all estimators is beyond the scope of this paper; instead, only appropriate estimators are selected together with an explanation for selecting them. The estimator needs to be *appropriate* for comparing *spread* between *measurement score* distributions. The appropriate estimator successfully avoids all the “pitfalls” of the characteristics in *measurement score* distributions and additionally [48, 49, 54]

- (i) is applicable to variables using *interval* scale and not just ratio scale (*Ratio scales* (e.g., Kelvin temperature, mass, or length) have a nonarbitrary, meaningful, and unique zero value. *Interval scales* (e.g., Celsius temperature) explain the degree of difference, but not the ratio between the values. A measurement score of 0.4 is greater than that

of  $-0.1$ , but not proportionally so. Additionally, a measurement score of  $0.0$  does not indicate “no determination.” Hence, measurement scores use an interval scale.)

- (ii) is applicable to variables containing both negative and positive values
- (iii) is insensitive to mean (average) value close to or approaching zero
- (iv) is insensitive to variables of which the mean (average) value can be zero
- (v) is invariant (robust) to underlying distribution of the variable (i.e., nonparametric)
- (vi) is invariant (robust) to a small number of outliers
- (vii) is invariant (robust) to asymmetry of the distribution and location estimate (or choice of central tendency, e.g., mean or median)
- (viii) has the best possible breakdown point (The breakdown point of an estimator is the proportion of incorrect observations an estimator can handle before producing incorrect results [55]. For example, consider the median; its breakdown point is 50% because that is the amount of incorrect observations introduced for it to have an incorrect median. The maximum achievable breakdown point is 50%, since that is the threshold at which it becomes impossible to discern correct from incorrect data. IQR has a breakdown point of 25%; Rousseeuw-Croux  $S_n$  and  $Q_n$  achieve 50%. The higher the breakdown point of an estimator, the greater its robustness.)

The interquartile range (IQR) is the difference between the upper and lower quartiles; also, it is the “height” of the box in a boxplot [56]. The coefficient of quartile variation (CQV) equals IQR divided by the sum of lower and upper quartiles [47]. Although IQR does not satisfy the criterion (viii), it is an *appropriate* statistic because it satisfies all of the other (more important) criteria; the *breakdown point* of the IQR is not critically low and equals 25%, together with the CQV for which the same reasoning of *appropriateness* applies. Furthermore, Rousseeuw-Croux estimators  $S_n$  and  $Q_n$  [57] offer breakdown points of 50%, do not assume distribution *symmetry*, and work independently of the choice of central tendency (*mean* or *median*)—all highly favourable traits. Notably, the median absolute deviation (MAD), as a robust measure of spread, was considered a serious contender due to its clear benefits, for example, over standard deviation as defined and elaborated in [50]. However, an important drawback of classical MAD with regard to criterion (vii) is its sensitivity to distribution asymmetry, a behaviour measurement score distribution definitely evident as shown in Figure 10. Therefore, IQR, CQV,  $S_n$ , and  $Q_n$  form a group of selected, *appropriate estimators of scale*.

To conclude evaluation of the algorithms, a summary of boxplot parameters (*measurements scores*) and appropriate

estimators is given in Table 4. Next to each estimator is the criterion which the estimator is attached to; criterion (E1) bears *one* and criterion (E2) bears *four* estimators altogether.

All of the *appropriate estimators* gave their output in the form of a single number (i.e., values in brackets); these numbers were compared, and algorithms *ranked* accordingly (for the criterion (E1), greater values are better (more is better); for the criterion (E2), the opposite is true—smaller values are better). Ranks reflect *true positions* with respect to each estimator’s output, respectively. Some ranks exhibit a “tie” (e.g., as with  $S_n$ ), where three algorithms came in 2nd, and only one came in 1st.

**4.4. Evaluation of the SmartSocial Algorithms: Conclusion Phase.** The *last row* (evaluation rank) in Table 4 declares the final, total rankings of algorithms with respect to evaluation. The final rank was produced as an arithmetic mean of the ranking of evaluation criteria (E1 and E2), the ranks of which were produced as arithmetic means of the respective evaluators. SLOF is compared to LRA in bold. As with criterion (E1), SLOF reigns supreme over the other algorithms along with criterion (E2) as well. In other words, SLOF is the most accurate and precise algorithm of the four analysed SmartSocial Influence algorithms. Evaluation clearly demonstrates that SLOF exhibits significant improvements over its predecessor, the LRA, and provides an original scientific contribution.

SAOF shows a minor improvement, whereas SMOF shows no improvement in the overall rankings, while SAOF is more *accurate* and SMOF is more *precise* than LRA. An interesting notice is that they are ranked (throughout the criteria) very closely to LRA, lacking the demonstrative power of improvement as exhibited by SLOF.

It seems that SMOF would greatly benefit from increasing its *accuracy*, as its precision is already on par with that of LRA. Likewise, SAOF would greatly benefit from increasing its *precision*, as it is already more accurate than LRA. Nonetheless, future research and additional work are necessary to uncover as to why the algorithms rank as they do—and motivation in answering this question lies in further experimentation and auxiliary analysis which may very well shed some additional light on a potentially decisive answer.

## 5. Discussion

This section discusses the impact of the proposed methodology and possible implications of SLOF as the best-evaluated algorithm. But first, to avoid any misconceptions, let us explain what validation and evaluation *are*, and what they are *not*—in terms of their respective goals.

Validation proves that all of the four SmartSocial influence algorithms do *not* work by the sheer *guessing* of correct measurements. The alternative hypotheses may be either true, or false—one cannot reason as to *how much* the algorithms produce “correct, meaningful and truthful” results; only that they do not produce random results (as is the case with guessing), when compared against the ground truth or “golden standard.” Validity is proven by ignoring the “pair weights” ( $\Delta_p$ ) associated with each measurement and looking

TABLE 4: Summary of criteria ranks and evaluation conclusion.

Statistic or estimator	Criterion	SLOF	SAOF	SMOF	LRA
Average (mean) rank	Accuracy (E1)	<b>1st (0.0358)</b>	2nd (0.0271)	4th (0.0240)	<b>3rd (0.0250)</b>
IQR rank	Precision (E2)	<b>1st (0.08)</b>	3rd (0.14)	2nd (0.13)	<b>2nd (0.13)</b>
CQV rank	Precision (E2)	<b>1st (2.0)</b>	3rd (7.0)	2nd (2.6)	<b>2nd (2.6)</b>
Rousseeuw-Croux $S_n$ rank	Precision (E2)	<b>1st (0.0835)</b>	2nd (0.1073)	2nd (0.1073)	<b>2nd (0.1073)</b>
Rousseeuw-Croux $Q_n$ rank	Precision (E2)	<b>1st (0.0885)</b>	2nd (0.1106)	2nd (0.1106)	<b>2nd (0.1106)</b>
Evaluation rank	(E1&E2)	<b>1st</b>	<b>2nd</b>	<b>4th</b>	<b>3rd</b>

at the percentage of correct measurements, as opposed to incorrect measurements.

Evaluation proves that SLOF is the best-ranked algorithm according to a pre-given set of criteria—namely, *accuracy* and *precision*. For each algorithm, accuracy is calculated using the mean (average) measurement score (as an *estimator of location*), and precision is calculated using measurement score spread (or dispersion, using robust *estimators of scale*). The algorithm with the greatest accuracy and precision emerges as the winner.

Additionally, evaluation does not enable any kind of statistical inference—the goal of validation and evaluation is not generalizability. The experiment, by its very design, did not (representatively) sample a predetermined population (One might define the population as mostly those between 20 and 30 years of age, predominantly highly educated (mostly from Zagreb, Croatia), with university degrees in information technology, medicine, psychology, or sociology.); doing so would greatly lower the amount of Facebook friendships in a sample graph, making the job of comparing algorithms all the more difficult—which is exactly what the purpose of the evaluation was in the first place.

The definition of *social influence* has been from social psychology, which is reflected to a certain degree in the design of the algorithms. On the other hand, there is no guarantee as to how much *social influence* measured by the algorithms fits *social influence* as measured by social psychologists. In other words, social influence in the “digital” realm may or may not correspond to (or be associated with) with that in the “physical, real world”—it is solely a best-effort model of it [4, 34, 35].

An analysis was conducted on the age and number of Facebook friends totalling 361 SmartSocial Influence experiment participants (The SmartSocial influence experiment was conducted in the period from September 2014 until May 2015. A total of 465 user profiles were created. Of these, 104 contained only telecommunication data, as these users did not provide their Facebook data. Consequently, the SmartSocial real-world sample comprised the remaining 361 profiles with complete, personal multisource data necessary for SmartSocial algorithms to run—both Facebook and telecommunication personal data.) (these are not the same participants who participated in the evaluation questionnaire (The SmartSocial Influence evaluation questionnaire was conducted in the period from 21st February 2016 until 14th March 2016. The first phase (pre-questionnaire) had 22 experts and 22 nonexperts as the participants. The second

phase (main questionnaire) had 16 participants.) although some may overlap). Analysis of age draws some interesting conclusions (Figure 11). Up until SI of 61, there is a slowly rising trend of age with respect to the social influence scores of participants. However, as SI approaches (60, 70], there is a sharp increase in the age of the participants, as there is a much greater representation of 30-year-olds in the sample. More interestingly, highly influential participants (SI > 80) were all 25 years of age and younger, with the most influential ones (SI > 90) being below 21.5 years of age. According to SLOF, the youth is more socially influential.

What is most surprising is the results from analysing the number of friends (Figure 12). Once more, a group of participants with SI = (60, 70] shows specific characteristics. As observed with *age*, this group predominantly comprises those older than 30 years of age; they have the average number of friends that strongly correlated to *age*. The number of friends in all other groups of influencers equals a constant 475 to 575, while the 30-year-olds, of whom 50% are female, average 160 Facebook friends.

What follows are certain specifics of SLOF, the most truthful algorithm, with regard to the sample of experiment participants described in [4]. It is important to keep in mind that SI score groups do not hold an equal number of participants—this is easily observed in the SLOF distribution of SI scores [4]. A group of SI = (0, 10] contains as much as 65% of the participants; SI = 0 holds 11% and SI = (10, 20] holds 15% of the participants. The remaining 9% of participants altogether form a great minority with SI > 20. As is expected of a score such as SI, it follows a power law with a minority of participants being responsible for the majority of social influence. Therefore, no definitive conclusions regarding *gender*, *age*, or *number of friends* with respect to social influence on Facebook can be drawn; instead, a larger, more diverse *real-world sample of participants* is needed.

Comparing the specifics of SLOF to the state-of-the-art influence algorithm Klout would be noteworthy, but impossible as Klout has been a “black box” ever since official launch in 2008, meaning its proprietary method and processing details have been unknown and remain a secret. Only recently has Klout received attention from the scientific community with their paper outlining the principles and basic mechanism of calculating social influence combined with nine other SNSs [58]. The paper does not enable direct comparison of the Klout algorithm to SmartSocial Influence algorithms because (i) validation of Klout scores in the paper

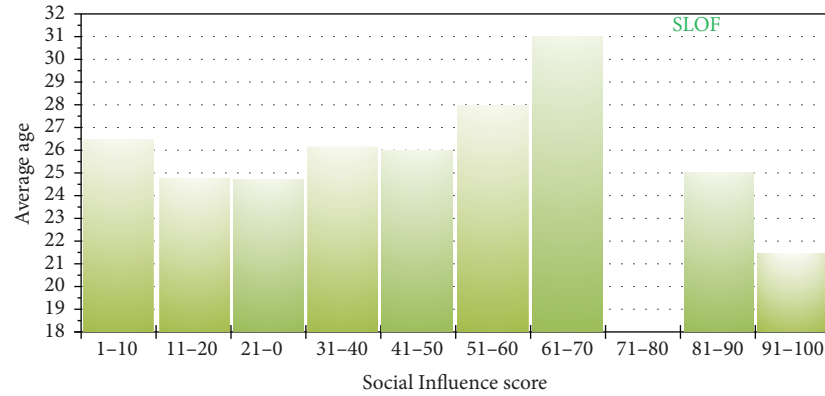


FIGURE 11: Average age with respect to SI scores of SLOF.

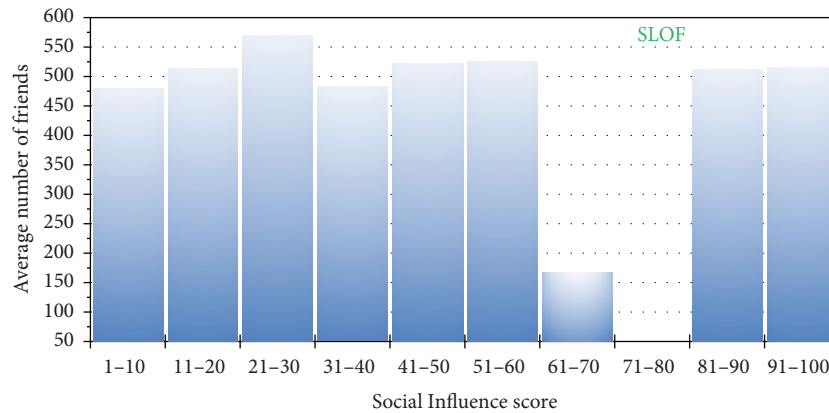


FIGURE 12: Average number of friends with respect to SI scores of SLOF.

is not as formal as the validation provided in this paper; (ii) validated scores include the top twenty people in specific categories (i.e., best ATP Tennis Players and Forbes Most Powerful Women); and (iii) it would be difficult to collect Klout scores of all 361 participants, since the Klout API as of 2017 does not yet enable fetching of Klout scores programmatically in a streamlined fashion. Klout's previous publications of Klout score distributions are obsolete due to several (major) revisions of the algorithm in the meantime. When taking everything into consideration, Klout is an impressive SNS for calculating social influence, but more transparency regarding the Klout algorithm is needed for a fair and direct comparison with alternative approaches.

## 6. Conclusion

This paper contributes to existing literature by proposing a new methodology for evaluating algorithms that calculate social influence in complex social networks. The paper has demonstrated the use of the proposed methodology using a case study in evaluating the accuracy and precision of four

social influence calculation algorithms from the class of SmartSocial Influence algorithms. The concept and details of SmartSocial Influence algorithms have already been presented in [4, 34, 35]; the proposed methodology validates all of them and has determined that the SmartSocial Influence algorithm (SLOF) is the most accurate and precise among them. This paper also contributes to existing literature by identifying the social influence calculation algorithm that offers higher accuracy and precision as benchmarked against the state-of-the-art LRA algorithm.

More broadly, the paper deals with a novel approach to social network user profiling with the goal of utilising multi-source, heterogeneous user data in order to infer new knowledge about users in terms of their social influence. By doing so, the paper addresses an ongoing research challenge in utilising such vast amounts of multisource, heterogeneous user data with the goal of identifying key, socially influential actors in the process of provisioning information and communication services. These actors are users equipped with smartphones, which reveals new information in regard to their social influence. This new information about a mobile

smartphone user has not only scientific but also industrial applications. For example, the best-evaluated novel algorithm for calculating a user's social influence (i.e., SLOF) can be used by telecommunication operators for churn prevention and prioritizing customer care, or by social networking services for digital advertising and marketing campaigns.

Some constraints in the proposed approach do exist. First, while the proposed methodology evaluates social influence algorithms, the question remains as to how to evaluate the very proposed methodology in return. To the authors' best knowledge, this approach is the first methodology to compare algorithms when calculating social influence based solely on available ego-user data rather than complete data on all social network users. That said, the authors of this paper will pursue encouragement of other similar research groups to develop alternative methodologies for evaluating algorithms that calculate social influence or more general global user properties, in online social networks. Second, the proposed methodology in this paper was applied on four algorithms from the SmartSocial Influence algorithm class. One of those—LRA—is a state-of-the-art benchmarking algorithm, while the other three—SLOF, SAOF, and SMOF—were previously developed by the authors of this paper. A more robust demonstration of the proposed methodology would include applying it on algorithms other than SmartSocial Influence class algorithms. This was not possible in this paper as the authors did not have access to (pseudo) code, test data, and ground truth data for other algorithms that solely use ego-user data for calculating ego-user social influence. However, they do hope that other research groups

developing such algorithms will apply the proposed methodology, presented in this paper, for benchmarking their algorithms against the SmartSocial Influence class of algorithms.

For future work, the authors plan to demonstrate applicability of the proposed evaluation methodology to other global user properties in complex social networks extending beyond social influence. Furthermore, they plan to adapt the methodology such that it is directly applicable to other social networks other than Facebook and other types of social network users beyond humans, such as networked objects and smart devices forming the Social Internet of Things.

## Appendix

### A. Questionnaires

The following questionnaires were developed and carried out using *Google Forms* (<https://docs.google.com/forms>). The content of the questionnaires below has been translated into English, as originally the questionnaires were given to participants in their native Croatian language.

### B. Pre-Questionnaire (Content Validity)

This pre-questionnaire was given to 22 *experts* in the form of 30 questions (items); each item is scored between [1.0, 5.0], with the threshold for *passing* content validity  $>2.1$ . Next to each question  $Q_i$  is its score  $CV(Q_i)$ . Questions marked as *chosen* are used for the next step (face validity pre-questionnaire).

TABLE 5: Pre-questionnaire (content validity) given to experts.

$Q_i$	Question text	Available answers	$CV(Q_i)$	Passed	Chosen
$Q_1$	You've noticed a post "Dangerous levels of chlorine detected in our hot water used for showering." A greater impression on you would leave a post by your Facebook friend: _____ or _____.	1 2 3 4 5	2.95	✓	
$Q_2$	You've noticed a post "If every one of us recycled, we would have CO2 emissions and receive state/country stimulus for it." You would recycle more frequently if it were posted by your Facebook friend: _____ or _____.	1 2 3 4 5	3.36	✓	
$Q_3$	You've noticed a post "Disaster has struck, the Nepalese are left with no food, water and electricity. I've donated money, here are instructions for you to do the same." You would donate a greater amount if it were posted by your Facebook friend: _____ or _____.	1 2 3 4 5	3.05	✓	

Instructions. The pre-questionnaire contains 30 questions which you need to score 1 to 5. The goal is to explore which questions are most suitable in determining social influence. A very suitable question is given a score of 5; the least suitable is given a 1. Best-scored questions in this pre-questionnaire will be used in compiling a new questionnaire which will subsequently be forwarded to participants. The scores you provide directly affect the process of screening for the most suitable questions. By participating in this pre-questionnaire as an expert, you are providing support to the final phase of Vanja Smailović's PhD research. It is important to understand the meaning of social influence. While scoring the questions, keep these definitions in mind at all times: - Social influence is a measure of how people, directly or indirectly, affect the thoughts, feelings, and actions of others; - social influence is the ability to drive action; and - social influence occurs when a person's emotions, opinions, and behaviours are affected by other persons. The newly compiled questionnaire (using some of the questions below) will be given to nonexperts who initially will not be aware of its purpose (surveying social influence). They, that is, the nonexperts, will have 4 pairs of their Facebook friends offered as answers, unlike the scale of 1 to 5 noted here. In other words, each nonexpert will choose the more influential friend in a pair—without being directly asked about their respective social influence. Your task is to score the given "criterion" in each of the questions, with respect to how much it conforms to the definitions of social influence (given above). The pre-questionnaire, unlike the next questionnaire, does not show you Facebook-friend pairs as answers, because they will be tailored and specific for each nonexpert. Your role as an expert is to focus on the questions, not the answers given to future nonexpert participants.

TABLE 5: Continued.

Q <sub>4</sub>	You are dissatisfied with your mobile operator. You've noticed a post "I've moved to my new telco X, I think they are better." You would more likely change your mobile operator if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.59	✓	✓✓
Q <sub>5</sub>	You've noticed a post "Gas station X has the best fuel." You would more likely refuel more frequently at the mentioned gas station if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.36	✓	
Q <sub>6</sub>	You've noticed a post "Smoking while pregnant greatly increases chances of health issues in a child." You would more likely spread this information if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	2.55	✓	
Q <sub>7</sub>	You've noticed a post "Electricity bills will soon go up." You would search for more details if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.50	✓	
Q <sub>8</sub>	You've noticed a post "Video out showing New Zealand's prime minister slipping on a banana." You would more likely watch the video if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	2.82	✓	
Q <sub>9</sub>	You've noticed a post "Hidden camera caught them in adultery." You would watch the video to a greater length if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	2.09	✗	
Q <sub>10</sub>	You've noticed a post "World leaders at their last meeting decided to increase nuclear armament." You would search for more details if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.68	✓	✓✓
Q <sub>11</sub>	You've noticed a post "Toothpaste X discounted in all shopping malls." You more likely go and buy the toothpaste if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	2.41	✓	
Q <sub>12</sub>	You've noticed a post "Travel agency X offers phenomenal discounts for Asia." You would search for more details if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	2.86	✓	
Q <sub>13</sub>	You've noticed a post "Concert tickets to see X selling out soon." You would search for more details if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	2.91	✓	
Q <sub>14</sub>	You've noticed a post "I'm calling everyone to join a public protest against getting rid of future generation pensions." You would more likely join this protest if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	4.0	✓	✓✓
Q <sub>15</sub>	You've noticed a post that explains the proven downside of your preferred political party. You would more likely change your vote if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	4.27	✓	✓✓
Q <sub>16</sub>	You've noticed a post "People killed in a terrorist attack in Ireland." You would search for more details if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.23	✓	
Q <sub>17</sub>	You've noticed a post "Home visits by National TV bill collectors more frequent in the following month." You would less likely open your doors to strangers if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.14	✓	
Q <sub>18</sub>	You've noticed a post "Quickly pay your monthly bills, otherwise fines follow within 24 hours according to latest news." Reading the post would to a greater extent cause restlessness if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.77	✓	✓✓
Q <sub>19</sub>	You've noticed a motivational post with thoughts about a brighter future, more jobs and possibilities, and greater salaries where you live. Reading the post would more likely cause peacefulness in you if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	4.0	✓	✓✓
Q <sub>20</sub>	You've noticed a post "Tensions between Balkan EU members might lead to war." Reading the post would more likely cause restlessness if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.64	✓	✓✓
Q <sub>21</sub>	You've noticed a post "Immigrants in EU constantly on the rise." Reading the post would more likely spark an interest if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.91	✓	✓✓
Q <sub>22</sub>	You've noticed a post "Parliament representatives physically confront each other at the morning sitting." Reading the post would more likely surprise you if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	2.68	✓	
Q <sub>23</sub>	You've noticed a motivational post about exercise, more physical activity, and health benefits. Reading the post would more likely motivate you if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.95	✓	✓✓
Q <sub>24</sub>	You are planning on seeing the movie X. You've noticed a post "I've seen X, it's horrible." Reading the post would more likely dissuade you from watching the film if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	4.41	✓	✓✓



TABLE 5: Continued.

Q <sub>25</sub>	You are planning a trip to a neighboring country/state X, it's snowing outside. You've noticed a post "X's police officers fine drivers without winter tires." Reading the post would more likely persuade you to buy winter tires if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.23	✓
Q <sub>26</sub>	You are driving on X section of road. You've noticed a post "Police radar-tracking speed at section X." Reading the post would more likely cause you to obey the speed limit if it were posted by your Facebook friend: _____ or _____"	1 2 3 4 5	3.0	✓
Q <sub>27</sub>	You are driving on the road section X. You've noticed a post "Terrible traffic accident at section X." Reading the post would more likely disturb you if it were posted by your Facebook friend: _____ or _____"	1 2 3 4 5	2.64	✓
Q <sub>28</sub>	You are planning a trip to city X. You've noticed a post "City X caught in bad weather." Reading the post would more likely worry/disappoint you if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	2.27	✓
Q <sub>29</sub>	You are planning a trip to city X. You've noticed a post "City X has seen a rise in crime-rates in recent years." Reading the post would more likely worry you if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.18	✓
Q <sub>30</sub>	You are planning a trip to city X. You've noticed a post "City X has seen a rise in crime-rates in recent years." Reading the post would more likely persuade you to re-plan the trip if it were posted by your Facebook friend: _____ or _____	1 2 3 4 5	3.09	✓

### C. Pre-Questionnaire (Face Validity)

This pre-questionnaire was given to 22 *nonexperts* in form of 10 questions (items); each item is scored between [1.0, 5.0], with the top 5 best (plus one fixed) questions chosen for the main questionnaire. Next to each question Q<sub>i</sub> is its score FV(Q<sub>i</sub>).

### D. Main Questionnaire (Algorithm Validity)

The main questionnaire was given to 16 *participants* with the goal of obtaining measurement scores for each algorithm, used in their validation and evaluation. The main questionnaire uses questions which "passed" both validities in pre-questionnaires; they are both content-valid and face-valid.

TABLE 6: Pre-questionnaire (face validity) given to nonexperts.

Q <sub>i</sub>	Question text	Available answers	FV(Q <sub>i</sub> )	Passed	Chosen
Q <sub>4</sub>	You are dissatisfied with your mobile operator. You've noticed a post "I've moved to my new telco X, I think they are better." You would more likely change your mobile operator if it were posted by your Facebook friend: _____ or _____	A or H B or G C or F D or E	3.32	✓	✓✓
Q <sub>10</sub>	You've noticed a post "World leaders at their last meeting decided to increase nuclear armament." You would search for more details if it were posted by your Facebook friend: _____ or _____	A or H B or G C or F D or E	3.73	✓	✓✓
Q <sub>14</sub>	You've noticed a post "I'm calling everyone to join a public protest against getting rid of future generation pensions." You would more likely join this protest if it were posted by your Facebook friend: _____ or _____	A or H B or G C or F D or E	3.45	✓	
Q <sub>15</sub>	You've noticed a post that explains the proven downside of your preferred political party. You would more likely change your vote if it were posted by your Facebook friend: _____ or _____	A or H B or G C or F D or E	3.86	✓	✓✓
Q <sub>18</sub>	You've noticed a post "Quickly pay your monthly bills, otherwise fines follow within 24 hours according to latest news." Reading the post would to a greater extent cause restlessness if it were posted by your Facebook friend: _____ or _____	A or H B or G C or F D or E	3.55	✓	
Q <sub>19</sub>	You've noticed a motivational post with thoughts about a brighter future, more jobs and possibilities, and greater salaries where you live. Reading the post would more likely calm you down if it were posted by your Facebook friend: _____ or _____	A or H B or G C or F D or E	3.59	✓	✓✓

TABLE 6: Continued.

Q <sub>20</sub>	You've noticed a post "Tensions between Balkan EU members might lead to war." Reading the post would more likely cause restlessness if it were posted by your Facebook friend: _____ or _____	<i>A or H B or G C or F D or E</i>	3.77	✓	✓✓
Q <sub>21</sub>	You've noticed a post "Immigrants in EU constantly on the rise." Reading the post would more likely spark an interest if it were posted by your Facebook friend: _____ or _____	<i>A or H B or G C or F D or E</i>	3.45	✓	
Q <sub>23</sub>	You've noticed a motivational post about exercise, more physical activity, and health benefits. Reading the post would more likely motivate you if it were posted by your Facebook friend: _____ or _____	<i>A or H B or G C or F D or E</i>	3.45	✓	
Q <sub>24</sub>	You are planning on seeing the movie X. You've noticed a post "I've seen X, it's horrible." Reading the post would more likely dissuade you from watching the film if it were posted by your Facebook friend: _____ or _____	<i>A or H B or G C or F D or E</i>	3.55	✓	✓✓
Q <sub>+</sub>	You have reached the final question. It is unique and very important. You will reply to it in the same manner as you replied to the questions earlier. Once more, you will choose a Facebook friend from the 4 pairs offered in answers—only this time, pay attention to the definitions of social influence below: - social influence is a measure of how people, directly or indirectly, affect the thoughts, feelings, and actions of others; - social influence is the ability to drive action; and - social influence occurs when one's emotions, opinions, and behaviours are affected by others. Read the definitions of social influence above. For each of the pairs, choose the Facebook friend whom you consider has the GREATER social influence on you on Facebook. While doing so, try to encompass all 3 definitions above as best as you can.	<i>A or H B or G C or F D or E</i>	—	—	—

TABLE 7: Main questionnaire given to participants.

The questionnaire contains 6 questions and takes 10 minutes to complete. Your answers will be anonymized and analysed collectively for all publishing or discussion purposes. By participating, you are supporting the final phases of Vanja Smailović's PhD research. Each of the 6 questions requires reading an imaginary Facebook post. Each question offers several *pairs* of your Facebook friends which are offered as answers. Your task, for each of the pairs, is to choose the Facebook friend which you consider to be the correct answer for a given question. If the question seems absurd or inapplicable, choose the Facebook friend whom you consider to be MORE correct. Read all the questions in advance—it is advisable to at least skim through them all before proceeding. *Important:* If the pairs repeat among the 6 questions—choose your answer always while paying attention to the *question*. On the other hand, watch out for pairs that repeat in a *single* question—those pairs require the same answer, because their goal is to check consistency. In other words, the same pairs between *different* questions are allowed to (and can) have a different answer—same pairs within a *single question cannot!* Do not communicate or consult with others while filling out the questionnaire and remain concentrated. You are allowed to go back in steps—the questionnaire is finalized, submitted, and locked only after you press Submit. In case of any questions or doubts, please call Vanja at [telephone number provided] to avoid making mistakes or errors.

Q <sub>i</sub>	Question text	Available answers
Q <sub>4</sub>	You are dissatisfied with your mobile operator. You've noticed a post "I've moved to my new telco X, I think they are better." You would more likely change your mobile operator if it were posted by your Facebook friend: _____ or _____	12 Facebook-friend pairs
Q <sub>10</sub>	You've noticed a post "World leaders at their last meeting decided to increase nuclear armament." You would search for more details if it were posted by your Facebook friend: _____ or _____	12 Facebook-friend pairs
Q <sub>15</sub>	You've noticed a post that explains the proven downside of your preferred political party. You would more likely change your vote if it were posted by your Facebook friend: _____ or _____	12 Facebook-friend pairs
Q <sub>19</sub>	You've noticed a motivational post with thoughts about a brighter future, more jobs and possibilities, and greater salaries where you live. Reading the post would more likely cause peacefulness in you if it were posted by your Facebook friend: _____ or _____	12 Facebook-friend pairs
Q <sub>20</sub>	You've noticed a post "Tensions between Balkan EU members might lead to war." Reading the post would more likely cause restlessness if it were posted by your Facebook friend: _____ or _____	12 Facebook-friend pairs
Q <sub>24</sub>	You are planning on seeing the movie X. You've noticed a post "I've seen X, it's horrible." Reading the post would more likely dissuade you from watching the film if it were posted by your Facebook friend: _____ or _____	12 Facebook-friend pairs

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

## Acknowledgments

The authors acknowledge the support of research projects “Managing Trust and Coordinating Interactions in Smart Networks of People, Machines and Organizations,” funded by the Croatian Science Foundation under the Grant UIP-11-2013-8813; “Ericsson Context-Aware Social Networking for Mobile Media,” funded by the Unity through Knowledge Fund; and “A Platform for Context-aware Social Networking of Mobile Users,” funded by Ericsson Nikola Tesla. This research has also been partly supported by the European Regional Development Fund under the Grant KK.01.1.1.01.0009 (DATACROSS). Furthermore, the authors would like to thank all participants who provided their personal data by installing the SmartSocial Android application and participating in questionnaires by which they greatly contributed to this research.

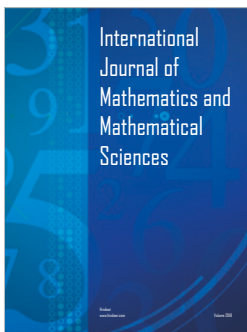
## Supplementary Materials

The paper is supplemented with the Excel file named “Smart-Social Influence evaluation dataset (anonymised),” which contains detailed evaluation results. Data in the file is anonymised to assure the privacy of individuals who participated in the evaluation. This dataset is also available at <http://socialab.science/datasets>. (*Supplementary Materials*)

## References

- [1] Statista August 2017 <http://www.statista.com>.
- [2] M. Pticek, V. Podobnik, and G. Jezic, “Beyond the internet of things: the social networking of machines,” *International Journal of Distributed Sensor Networks*, vol. 12, no. 6, Article ID 8178417, 2016.
- [3] Y. Liu, D. Pi, and L. Cui, “Mining community-level influence in microblogging network: a case study on Sina Weibo,” *Complexity*, vol. 2017, Article ID 4783159, 16 pages, 2017.
- [4] V. Smailovic and V. Podobnik, “Mining social networks for calculation of SmartSocial Influence,” *Journal of Universal Computer Science*, vol. 22, no. 3, pp. 394–415, 2016.
- [5] Q. Wang, X. Yu, and X. Zhang, “A connectionist model-based approach to centrality discovery in social networks,” in *Behavior and Social Computing*, L. Cao, H. Motoda, J. Srivastava, E.-P. Lim, I. King, P. S. Yu, W. Nejdl, G. Xu, G. Li, and Y. Zhang, Eds., vol. 8178 of Lecture Notes in Computer Science, Springer, Cham, 2013.
- [6] C. Kiss and M. Bichler, “Identification of influencers - measuring influence in customer networks,” *Decision Support Systems*, vol. 46, no. 1, pp. 233–253, 2008.
- [7] J. Golbeck and J. Hendler, “Inferring binary trust relationships in web-based social networks,” *ACM Transactions on Internet Technology*, vol. 6, no. 4, pp. 497–529, 2006.
- [8] V. Podobnik, D. Striga, A. Jandras, and I. Lovrek, “How to calculate trust between social network users?,” in *SoftCOM 2012, 20th International Conference on Software, Telecommunications and Computer Networks*, Split, Croatia, September 2012.
- [9] M. Stupalo, J. Ilić, L. Humski, Z. Skočir, D. Pintar, and M. Vranić, “Applying the binary classification methods for discovering the best friends on an online social network,” in *2017 14th International Conference on Telecommunications (ConTEL)*, pp. 155–162, Zagreb, Croatia, June 2017.
- [10] J. Ilic, L. Humski, D. Pintar, M. Vranic, and Z. Kocir, “Proof of concept for comparison and classification of online social network friends based on tie strength calculation model,” in *Proceedings ICIST 2016. Belgrade, Serbia: Society for Information Systems and computer networks*, pp. 159–164, Belgrade, Serbia, 2016.
- [11] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, *The Anatomy of the Facebook Social Graph*, 2011, <https://arxiv.org/abs/1111.4503>.
- [12] M. Gabielkov, A. Rao, and A. Legout, “Studying social networks at scale: macroscopic anatomy of the twitter social graph,” in *The 2014 ACM International Conference on Measurement and Modelling of Computer Systems*, pp. 277–288, ACM, Austin, USA.
- [13] S. Bhagat, M. Burke, C. Diuk, I. O. Filiz, and S. Edunov, “Three and a half degrees of separation,” 2016, November 2017, <https://research.fb.com/three-and-a-half-degrees-of-separation>.
- [14] A. Landherr, B. Friedl, and J. Heidemann, “A critical review of centrality measures in social networks,” *Business & Information Systems Engineering*, vol. 2, no. 6, pp. 371–385, 2010.
- [15] J. C. Turner, *Social Influence*, Brooks/Cole, Pacific Grove, Calif, 1st ed edition, 1991.
- [16] Klout, Inc., “Klout|Be Known For What You Love,” Klout,” 2016, July 2017, <https://klout.com/home>.
- [17] T. L. D. Kred, “Kred|The Home of Influence,” 2017, July 2017, <http://home.kred>.
- [18] Runtime Collective Limited (Brandwatch), “Peer Index: Social Analytics & Influence Tools-Brandwatch,” 2017, July 2017, <https://www.brandwatch.com/peerindex-and-brandwatch..>
- [19] Tellagence, “Tellagence,” 2017, July 2017, <http://www.tellagence.com>.
- [20] Openinfluence.Com, “Creative + Data-Driven Influencer Marketing Services,” previously instabrand.com,” 2017, August 2017, <http://www.openinfluence.com>.
- [21] Klout, Inc., “Klout Score,” Klout,” 2017, August 2017, <https://klout.com/corp/score>.
- [22] M. Deutsch and H. B. Gerard, “A study of normative and informational social influences upon individual judgment,” *The Journal of Abnormal and Social Psychology*, vol. 51, no. 3, pp. 629–636, 1955.
- [23] S. E. Asch, “Opinions and social pressure,” in *Scientific American*, vol. 193, pp. 31–35, W. H. Freeman and Co., San Francisco, California, 1955.
- [24] S. Milgram, “Behavioral study of obedience,” *The Journal of Abnormal and Social Psychology*, vol. 67, no. 4, pp. 371–378, 1963.
- [25] C. Haney, C. Banks, and P. Zimbardo, “Interpersonal dynamics in a simulated prison,” *International Journal of Criminology and Penology*, vol. 1, pp. 69–97, 1973.
- [26] B. Latané, “The psychology of social impact,” *American Psychologist*, vol. 36, no. 4, pp. 343–356, 1981.
- [27] R. B. Cialdini, *Influence: Science and Practice*, Allyn & Bacon, Boston, MA, 4 edition edition, 2000.
- [28] M. J. Rosenfeld and R. J. Thomas, “Searching for a mate: the rise of the internet as a social intermediary,” *American Sociological Review*, vol. 77, no. 4, pp. 523–547, 2012.

- [29] M. Fink and J. Spoerhase, "Maximum betweenness centrality: approximability and tractable cases," August 2010, <https://arxiv.org/abs/1008.3503>.
- [30] K.-I. Goh, E. Oh, B. Kahng, and D. Kim, "Betweenness centrality correlation in social networks," *Physical Review E*, vol. 67, no. 1, article 017101, 2003.
- [31] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.
- [32] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: the million follower fallacy," in *Fourth International AAAI Conference on Weblogs and Social media*, Washington, DC, USA, 2010.
- [33] A. A. Rad and M. Benyoucef, "Towards detecting influential users in social networks," in *E-Technologies: Transformation in a Connected World*, G. Babin, K. Stanoevska-Slabeva, and P. Kropf, Eds., pp. 227–240, Springer Berlin Heidelberg, 2011.
- [34] V. Smailovic, D. Striga, D.-P. Mamic, and V. Podobnik, "Calculating user's social influence through the smart social platform," in *Proceedings of Software, Telecommunications and Computer Networks Conference (Soft COM)*, pp. 383–387, Split, Croatia, 2014.
- [35] V. Smailovic, D. Striga, and V. Podobnik, "Advanced user profiles for the smart social platform: reasoning upon multi-source user data," in *ICT Innovations 2014 Web Proceedings*, Ohrid, Macedonia, 2014.
- [36] American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), *The Standards for Educational and Psychological Testing*, American Educational Research Association, Washington, DC, USA, 2014.
- [37] C. H. Lawshe, "A quantitative approach to content validity," *Personnel Psychology*, vol. 28, no. 4, pp. 563–575, 1975.
- [38] F. J. Gravetter and L.-A. B. Forzano, *Research Methods for the Behavioral Sciences*, Wadsworth, Australia; Belmont, CA, 4th edition, 2012.
- [39] I. B. Weiner and W. E. Craighead, Eds., *The Corsini Encyclopedia of Psychology*, Wiley, Hoboken, NJ, 4th edition, 2010.
- [40] R. McCarney, J. Warner, S. Iliffe, R. van Haselen, M. Griffin, and P. Fisher, "The Hawthorne effect: a randomised, controlled trial," *BMC Medical Research Methodology*, vol. 7, no. 1, p. 30, 2007.
- [41] D. L. Sackett, "Bias in analytic research," *Journal of Chronic Diseases*, vol. 32, no. 1-2, pp. 51–63, 1979.
- [42] R. Rosenthal and R. L. Rosnow, *Artifacts in Behavioral Research Robert Rosenthal and Ralph L. Rosnow's Classic Books: A Re-Issue of Artifact in Behavioral Research, Experimenter Effects in Behavioral Research and the Volunteer Subject*, Oxford University Press, New York, 2009.
- [43] A. A. J. Marley and J. J. Louviere, "Some probabilistic models of best, worst, and best-worst choices," *Journal of Mathematical Psychology*, vol. 49, no. 6, pp. 464–480, 2005.
- [44] J. J. Louviere, T. N. Flynn, and A. A. J. Marley, *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge, Cambridge University Press, New York, 2015.
- [45] A. Papoulis, "Bernoulli trials," in *Probability, Random Variables, and Stochastic Processes*, pp. 57–63, McGraw-Hill, New York, 2nd edition, 1984.
- [46] R. A. Maronna and R. H. Zamar, "Robust estimates of location and dispersion for high-dimensional datasets," *Technometrics*, vol. 44, no. 4, pp. 307–317, 2002.
- [47] D. G. Bonett, "Confidence interval for a coefficient of quartile variation," *Computational Statistics & Data Analysis*, vol. 50, no. 11, pp. 2953–2957, 2006.
- [48] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, 1993.
- [49] P. J. Huber and E. Ronchetti, *Robust Statistics*, Wiley, Hoboken, N.J, USA, 2nd edition, 2009.
- [50] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764–766, 2013.
- [51] M. Krzywinski and N. Altman, "Points of significance: visualizing samples with box plots," *Nature Methods*, vol. 11, no. 2, pp. 119–120, 2014.
- [52] M. Spitzer, J. Wildenhain, J. Rappsilber, and M. Tyers, "BoxPlotR: a web tool for generation of box plots," *Nature Methods*, vol. 11, no. 2, pp. 121–122, 2014.
- [53] J. Masters, "Ozone Hole FAQ|Weather Underground," *Weather Underground*, 2011, March 2016, <http://www.wunderground.com/climate/holefaq.asp>.
- [54] C. Gourieroux and A. Monfort, *Statistics and Econometric Models, Vol. 1. Cambridge [England]*, Cambridge University Press, New York, NY, USA, 1995.
- [55] X. He, D. G. Simpson, and S. L. Portnoy, "Breakdown robustness of tests," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 446–452, 1990.
- [56] W. C. Navidi, *Statistics for Engineers and Scientists*, McGraw-Hill, New York, 3rd edition, 2011.
- [57] P. Wessa, *Free Statistics Software (version 1.1.23-r7)*, Office for Research Development and education, 2016, March 2016, <http://www.wessa.net>.
- [58] A. Rao, N. Spasojevic, Z. Li, and T. D. Souza, "Klout score: measuring influence across multiple social networks," 2015, <https://arxiv.org/abs/1510.08487>.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

