

Combining Evidence for Social Situation Detection

Georg Groh, Christoph Fuchs, Alexander Lehmann
Institut für Informatik, Technische Universität München, Germany
{grohg,fuchsc,lehmann}@in.tum.de

Abstract—Social Situations are social context models indicating n-ary social interaction on small spatio-temporal scales detected by means of Social Signal Processing. We discuss the problem of how to exchange and combine evidence from several sensor-sources for the benefit of algorithmic assessment of Social Situations in a distributed agent-based Social Networking scenario. We propose a solution based on Subjective Logic that mediates between exchanging&processing of raw low level sensor data, of intermediate results of 'sub-symbolic' probabilistic models typically used for Social Signal Processing and of the final 'symbolic' Social Situation models. We evaluate key aspects of the approach on the basis of a social experiment, combining audio-based and geometry-of-interaction-based methods for Social Situation detection.

I. INTRODUCTION

Social situations are an important part of most people's lives [1] and algorithmically detecting and modeling social situations can be a useful form of social context for Social Networking (SN) and Mobile Social Networking (MSN) services [2]. Sensors in mobile devices are especially useful for measuring temporal patterns of behavioral cues [3] for the detection of social situations, because of their deep and natural integration into people's every-day lives. The fields of Social Signal Processing [3] and Reality Mining [4] have made substantial contributions towards modeling social and individual aspects of human behavior, partly using mobile device sensors.

While most activities in the field of Social Signal Processing (SSP) aim at *understanding* aspects of social situations arriving at interpretative models of social signals and social behavior, our model of social situations mainly focuses on indicating and describing the limitations of the *existence* of social situation: A Social Situation is a four tuple $S = (P, T, X, K)$ of a set P of socially interacting persons exhibiting full, synchronous mutual awareness of the ongoing interaction, a temporal reference T (connected proper subset of \mathbb{R}), a spatial reference X (connected proper subset of \mathbb{R}^3 and a set of keywords K describing the semantics of the situation [5]. T and X are projections of a more general spatio-temporal reference \tilde{X} , a connected proper subset of \mathbb{R}^4 . Full mutual awareness [5] is a prerequisite for a proper definition of S [5]. We will focus on co-located social interaction, excluding mediated interaction (e.g. via an electronic conference system). Although this general model allows such Social Situations to be nested, (e.g. a Social Situation model of a conversation $S_1 \subset S_2$ between three people at a party of 70 people, modeled by S_2), we will restrict ourselves to immediately face-to-face Social Situations thus employing a narrow interpretation

of the demand of co-location and full, synchronous mutual awareness. This effectively excludes those Social Situations from significantly overlapping (compare [5]).¹

Social Situation models may be computed by algorithmically analyzing human interaction geometry (e.g. interpersonal distances and body angles, won by analyzing orientation and position data from mobile devices) [6] [5] or turn taking patterns won by analyzing audio data of conversations from mobile device microphones etc. We assume a decentralized MSN scenario, where each user is associated with a software agent running on her mobile device (see [7] for a discussion on why decentralized SN is an important alternative for conventional platform based SN). Akin to higher level subjective human perspectives that each individual may acquire on a social setting, each such agent will measure different sensor data and may thus algorithmically analyze the social setting differently. In order to get more accurate or more complete representation of the social setting (e.g. in the form of Social Situation models), these agents may profit from exchanging sensor data or their interpretation models.

In this contribution we will thus discuss the question of how independent user agents running on or using the sensors of mobile devices may exchange data, analysis results and Social Situation models in order to compute more accurate and more complete subjective Social Situation models or to agree on a common inter-subjective current Social Situation model.

II. FOUNDATIONS AND RELATED WORK

In this section we will put our main focus on some fundamental results on protocols for reaching agreements with an emphasis on Subjective Logic which will be a key element of our suggestions in view of our research question. Due to space restrictions, we will have to restrict ourselves to few pointers to basic literature with respect to most other related fields.

As for the related work on *social situations from a social psychology perspective*, motivating our model, we refer the reader to the respective sections of [5]. Decentralized SN is discussed in [7], [8] (from a Social Semantic Web perspective) and [9] (from a P2P perspective). Vinciarelli et al. [3] give a good overview on SSP in general. Furthermore, refer to [10] and [11] for an overview and example of *Reality Mining*.

In the classical theory of *Sensor Fusion* (SF) [12][13], sensors are usually assumed to output a value $V(t)$ together

¹ $\forall i, j : S_i = (P_i, T_i, X_i, K_i) \wedge S_j = (P_j, T_j, X_j, K_j) \rightarrow [T_i \cap T_j \neq \emptyset \rightarrow (P_i \cap P_j = \emptyset \wedge X_i \cap X_j = \emptyset)]$

with a confidence interval $\epsilon(t)$ containing the true value with confidence $1 - \alpha$, modeling uncertainty. Sensors are (compare [12], p. 32ff.), either complimentary (no fusion necessary), competitive (classical case for SF) or cooperative. In our case all three situations can occur, but the competitive case is most central: consider one sensor measuring the social interaction probability of two people on the basis of audio (conversation analysis) and another sensor on the basis of interaction geometry. [12] distinguishes several basic families of approaches in dealing with uncertainty: Directly fusing uncertainty intervals (compare [13]), Dempster-Shafer based approaches (see below), statistical / probability based approaches such as DBNs (see below) and fuzzy set theory.

Probabilistic reasoning via Bayesian Networks (BN) [14] is well suited to interpret, reason about ², and general fusing (see e.g. [15] [16]) of temporal sequences of measurements from different sensors. Statistical learning of parameters of such Bayesian networks (e.g. learning sensor and transition models via Expectation Maximization methods) can be perceived as an act of sub-symbolic modeling of the general probabilistic relations between low-level sensor data and higher level social signals or Social Situation models via appropriate training examples. Here, the assumption has to be made that the respective aspect of human social behavior is general enough and structurally uniform enough, to ensure that the models of this behavioral aspect are sufficiently generally applicable and can be trained from finite labeled (often experimental) training data [6]. The resulting BN model may then e.g. be readily built into the agents and used without the need for further training. As such, BN are essential for SSP agents to *internally* instantiate higher level subjective social models (such as Social Situation models) from data of various low level sensors (see [3]) or combinations of such sensors (compare [16]) or even sensor data communicated by other SSP agents. However, for processing and formulating intermediate results communicated by a dynamic number of other (potentially heterogeneous) SSP agents, BN appear to be too expressive and thus less ideal, since a respective BN would very probably have to involve a high number of probability-parameters to deal with foreign agent's subjectivity which are hard to control and learn without large amounts of training data.

We thus suggest a simpler problem specific collection of sub-symbolic and symbolic methods, which makes use of Subjective Logic to process and formulate agent opinions e.g. concerning Social Situations. (see section III for a deeper discussion).

Subjective Logic (SL) [17] [18] is a variant or further development of Dempster-Shafer theory of evidence [19] [20] with the goal to pragmatically model uncertainty and subjectivity of propositions. Assume a defined part A of a world of discourse ('frame of discernment') which is in exactly one state out of a state space $\Theta = \{x_1, x_2, \dots, x_n\}$. Classic

²e.g. filtering, predicting, smoothing (e.g. via variations of forward-backward algorithms (relevant example: Kalman Filters [14])), or computing most likely explanations (e.g. via Viterbi)

logic, where a proposition made by an agent about A is either true or false, allows for subjectivity since agents may communicate propositions which may be combined (reasoned over) but has no standard means to model uncertainty. Classic probability theory that formulates a probability measure M over a σ -algebra over Θ , allows to model uncertainty but has no standard means to deal with subjectivity (different probability measures communicated by agents). SL allows for modeling both and is compatible with both classic logic and probability theory (see e.g. [21], section 6 or [22] p. 183).

In SL, a *Belief Mass Assignment* (BMA) $m : 2^\Theta \mapsto [0, 1]$ assigns a *belief mass* $m(\theta)$ to a subset $\theta \subseteq \Theta$, such that $m(\emptyset) = 0$ and $\sum_{\theta \in 2^\Theta} m(\theta) = 1$. The *belief* $b(\theta)$ that an agent may have in 'A is in a state in θ ', is defined as $b(\theta) = \sum_{\theta' \subseteq \theta} m(\theta')$.

A simplifying *Dirichlet* BMA (DBMA) assigns only belief mass to single states and Θ as a whole: $m(\theta) \neq 0 \rightarrow (\theta = \Theta) \vee (|\theta| = 1)$. Thus for a DBMA an agent has beliefs $b(x_i) = m(x_i) \neq 0$ for all single states and a general *uncertainty* $u = m(\Theta)$. The *base rates* $a : \Theta \mapsto [0, 1]$ with $a(\emptyset) = 0$ and $\sum_i a(x_i) = 1$ allow for modeling the equivalent of a-priori probabilities for each state. A *multinomial opinion* ω is a tuple $\omega = (b, u, a)$ based on a DBMA. The value of the *probability expectation function* concerning a state x_i is computed as $p(x_i) = b(x_i) + a(x_i)u$ modeling the equivalent of an a-posteriori probability. A *binary opinion* $\omega = (b = b(x), d = b(\bar{x}), u, a = a(x))$ is a multinomial opinion over a binary set $\Theta = \{x, \bar{x}\}$; $b(\bar{x})$ is interpreted as disbelief in x and denoted as d .

Opinions may also be formulated in an *evidence notation* $w = (r, a)$ where $r(x_i)$ models the collected evidences for state x_i (and a Dirichlet probability distribution equivalent to the DBMA is required (see [17][23]) to compute the expectation value). The two notations are equivalent [17]: For $u \neq 0$,

$$\begin{aligned} b(x_j) &= \frac{r(x_j)}{C + \sum_{i=1}^k r(x_i)} \\ u &= \frac{r(x_j)}{C + \sum_{i=1}^k r(x_i)} \end{aligned} \quad (1)$$

(for the case $u = 0$ and the inverse expressions refer to [17]). The constant C acts as a Laplace estimator (compare the Rule of Succession in general statistics and machine learning) and can be interpreted as a weight associated with uncertainty.

SL provides several operators to combine opinions into a *consensus* [24].

Cumulative fusion combines independent opinions ω^A of agent A and ω^B of agent B over ω into $\omega^{A \circ B} = \omega^A \oplus \omega^B$. *Averaging fusion* combines dependent opinions into $\omega^{A \diamond B} = \omega^A \oplus \omega^B$. As an example of independence / dependence consider the two opinions resulting from observations in disjoint time intervals / the same time interval.

In evidence notation, $\omega^{A \circ B} = (r^{A \circ B}, a^{A \circ B})$ corresponds to summing up evidence, whereas $\omega^{A \diamond B} = (b^{A \diamond B}, u^{A \diamond B}, a^{A \diamond B})$

corresponds to averaging evidence:

$$\begin{aligned} r^{A \circ B}(x_j) &= r^A(x_j) + r^B(x_j) \quad [24] \\ r^{A \oslash B}(x_j) &= \frac{1}{2}(r^A(x_j) + r^B(x_j)) \quad [24] \quad (2) \\ a^{A \circ B}(x_j) &= \frac{a^A(x) \cdot (\sum_i r^A(x_i)) + a^B(x) \cdot (\sum_i r^B(x_i))}{(\sum_i r^A(x_i)) + (\sum_i r^B(x_i))} \end{aligned}$$

where the expression for $a^{A \circ B}$ is a straightforward generalization of [18], Def.13.3.: in the weighted average, the total number of evidences are used as weights.

In belief notation, we see that for combined beliefs the beliefs of one agent are weighted with the beliefs of the other agent (weighted arithmetic mean), combined uncertainty is won as harmonic mean and combined base rates can be simply averaged: for $u^A, u^B \neq 0$ we have

$$\begin{aligned} b^{A \circ B}(x_j) &= \frac{b^A(x_j) u^B + b^B(x_j) u^A}{u^A + u^B} \quad [24] \\ u^{A \circ B} &= \frac{2u^A u^B}{u^A + u^B} \quad [24] \quad (3) \\ a^{A \circ B} &= \frac{a^A + a^B}{2} \quad [22], \text{p.184.} \end{aligned}$$

For $u^A, u^B = 0$ and the analogous case of $b^{A \circ B}$ and $u^{A \circ B}$ see [24]. For a suggestion on $a^{A \circ B}$ see [21].

SL can also be used to model trust, which is also an important feature for our purposes. [22] propose several possibilities, one of which is especially relevant for us. Consider agent B having a binary opinion ω_B^B concerning some frame of discernment Θ . Agent A has a binary opinion concerning agent B also expressed as a binary opinion ω_B^A which models trust in B . In *Uncertainty Favouring Discounting*, A will combine the opinion ω_B^B communicated from B to A with his own trust ω_B^A in B as $\omega_{\Theta}^{A:B} = \omega_B^A \otimes \omega_B^B$ where $b_{\Theta}^{A:B} = b_B^A b_{\Theta}^B$, $d_{\Theta}^{A:B} = b_B^A d_{\Theta}^B$, $u_{\Theta}^{A:B} = d_B^A + u_B^A + b_B^A u_{\Theta}^B$, and $a_{\Theta}^{A:B} = a_{\Theta}^B$. That is, beliefs and disbeliefs are weighted with trust. A 's uncertainty concerning Θ arises because of distrust in B , uncertainty concerning trust in B and B 's uncertainty weighted with trust in B . The resulting weight trusted communicated opinion $\omega_{\Theta}^{A:B}$ may then be combined with some own opinion that A may have concerning Θ with one of the aforementioned operators \oplus and \ominus .

Bamberger et al. show [25], how SL can be used to model trust and subjectivity in a car-to-car scenario. Cars exchange opinions concerning some Θ and built up trust concerning other cars via comparing their communicated opinions with the own opinion after sufficient evidence has been collected. This is done by formulating two trust components: An SL opinion estimating the *competence* of an other car (a full averaging balance concerning the complete (time-weighted) exchange history) and another similar opinion, formalizing a more outlier-sensitive estimation of the *predictability* of the other car (see [25] for details). Furthermore, *reputation* is computed by collecting other car's trust opinions on other car's. They use a variant of \oplus to combine opinions.

Besides consensus operators from SL, numerous other techniques for *reaching agreements between agents* exist, stemming from the overlapping fields of Multi-Agent-Systems

/ Distributed AI, Robotics, CSCW, Game-Theory and Economics (compare [26]). These include approaches from voting / social choice theory (mostly geared toward finite discrete sets of choices) or bargaining approaches like Monotonic Concession Protocol (MCP) + Zeuthen-strategy or the approach of deGroot [27]. While basic MCP + Zeuthen allows two agents having certain utility functions to reach a consensus on some values in several rounds of bargaining, deGroot's approach finds a consensus between n agents with respect to probability distributions on some parameter. The distributions are iteratively combined using a matrix of trust values, ultimately leading to a stationary solution if the matrix fulfills certain properties (e.g. corresponds to an irreducible and aperiodic Markov chain). A simple one-step approach for reaching an agreement in case agents A_i can formulate a utility function $u(s)$ over some fixed set of alternatives S which was proposed by [28] (p.49f) simply maximizes the utility product to agree on some s' : $s' = \text{argmax}_s \prod_i u_i(s)$.

Possible data sources / sensors which can be used to detect Social Situations include e.g. audio and geometry. Analysis of conversations recorded by microphones for analyzing social settings has been proposed in e.g. [29][30][31][32][33][34]. [35] and [3] review visual analysis methods for social analysis, which are difficult to use in 'real-life' mobile devices, since they are usually not carried or worn in a way that would permit the frequent use of its cameras. Sensors allowing for capturing aspects of the geometry of interaction of the persons carrying them (such as gyroscopes, compasses, ultra-sound-sensors and accelerometers) may also be used for detecting Social Situations [6] [5].

III. APPROACH

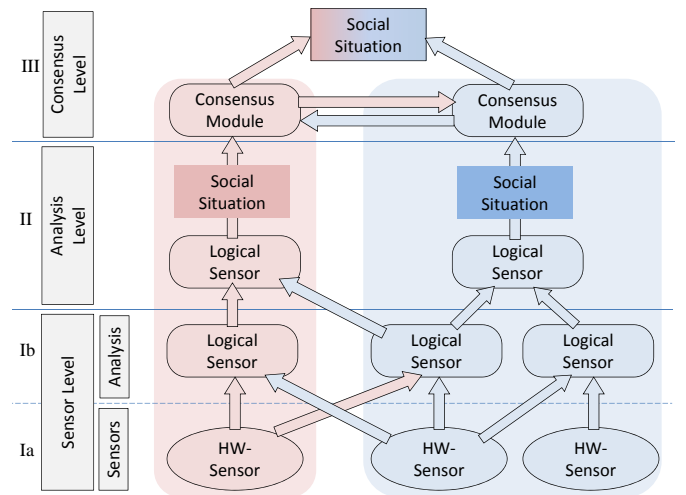


Figure 1. Basic architectural elements: exemplary setting involving two agents (light blue and light red)

Figure 1 shows basic architectural elements of our approach in an exemplary setting involving two agents. Any agent A_i has a number of *hardware sensors* H_j^i at layer Ia (e.g. two

microphones $H_1^i = H_{\text{Mic1}}^i$ and $H_2^i = H_{\text{Mic2}}^i$ and a gyroscope $H_3^i = H_{\text{Gyro}}^i$) providing raw measurements. *Logical sensors* L_j^i (at layers Ib and II) analyze output from several types of own or other agent's logical and hardware sensors. An example of a layer Ib logical sensor could be a $L_{\Theta_1}^i$ that analytically combines measurements from a gyroscope H_{Gyro}^i , an accelerometer H_{Acc}^i and a compass H_{Comp}^i into a device-orientation using Kalman filters for smoothing. Another example would be a $L_{\Theta_2}^i$ that combines an L_{Status}^i which outputs wearing or carrying status of the device and the aforementioned $L_{\Theta_1}^i$ into an assessment of the body orientation of the associated user. The layers are included in the diagram 1 for conceptual purpose only. Some logical sensor may be better characterized as an abstraction of a group of hardware sensors (and thus classified into layer Ib) and others as encapsulating more abstract, higher levels of analysis (thus layer II). In terms of Social Signal Processing, level II logical sensors output assessments of behavioral cues and social signals of varying degrees of abstraction. In our case, each agent is assumed to have one logical sensor L_S^i whose output is the agent's subjective assessment of the agent user's current Social Situation. Agents may finally reach an agreement on a common Social Situation using their *consensus modules* (conceptually in layer III). Each sensor may communicate its output to other agent's sensors in the form of binary or multinomial *Subjective Logic (SL) opinions*.

In order to further illustrate these structures, consider the following example:

At level I, agent A_{i_1} may have $L_{\Theta}^{i_1}$, providing absolute body angles $\theta_{i_1}(t)$, and $L_{\delta d; i_1, i_2}^{i_1}$, providing interpersonal distances $\delta d_{i_1 i_2}(t)$ from A_{i_1} to other agents A_{i_2} in communication distance. $L_{\delta d; i_1, i_2}^{i_1}$ may use an ultrasound hardware sensor for distance measurements and $L_{\Theta}^{i_1}$ may use a gyroscope, accelerometer and compass. Logical sensor $L_{\delta \theta; i_1, i_2}^{i_1}$ may then assess the relative body angles $\delta \theta_{i_1, i_2}(t)$ of A_{i_1} to other agents A_{i_2} , by combining own $L_{\Theta}^{i_1}$ and communicated $L_{\Theta}^{i_2}$ outputs from other agents. This is a case of cooperative fusion of logical sensors. Another case could be $L_{\text{Talk}}^{i_1}$ with binary output of ('user talking', 'user not talking') at time t .

We show in [6] and [5] how level II logical sensors $L_{\text{GEO}; \{i_1, i_2\}}^{i_1}$ can be constructed that use $L_{\delta d; i_1, i_2}^{i_1}$ and $L_{\delta \theta; i_1, i_2}^{i_1}$ (together assessing the basic geometry of interaction) to determine, from A_{i_1} 's perspective, for each pair of agents $\{A_{i_1}, A_{i_2}\}$ the probability $p_{\text{GEO}; \{i_1, i_2\}}^{\oplus}(t)$ that the users corresponding to these agents are socially interacting and the probability $p_{\text{GEO}; \{i_1, i_2\}}^{\ominus}(t)$ that they are not interacting at time t with the help of Gaussian Mixture Models. Furthermore, in [34] we show how a level II logical sensor $L_{\text{TT}; \{i_1, i_2\}}^{i_1}$ may be constructed that outputs $(p^{\oplus}, p^{\ominus})_{\text{TT}; \{i_1, i_2\}}(t)$ by analyzing time-series of output from $L_{\text{Talk}}^{i_1}$ and $L_{\text{Talk}}^{i_2}$, corresponding to turn taking patterns, centered around t with the help of Hidden Markov Models. More pragmatic sensors may e.g. use Bluetooth encounter events as evidences r and use equations (1) to compute $p^{\oplus} = b$ and $p^{\ominus} = 1 - b - u$.

In this contribution we cannot not provide an in depth

discussion on Social Signal Processing techniques involved in layer II logical sensors (see [3]). Furthermore, instead of discussing layer I fusion / data analysis of logic and hardware sensors, we will focus on higher level fusion of layer II logical sensors in view of Social Situation assessment.

We assume that each agent A_{i_1} has several level II logical sensors $L_{j; \{i_1, i_2\}}^{i_1}$ that output pairs of probabilities $(p^{\oplus}, p^{\ominus})_{j; \{i_1, i_2\}}^{(i_1)}(t)$ assessing the probability that A_{i_1} is / is not in a Social Situation (abbreviated by SS from now on) with other perceived agents A_{i_2} (compare the previous examples). In order to compute these probability pairs, the $L_{j; \{i_1, i_2\}}^{i_1}$ may rely upon different sets of sensors of lower levels of abstraction.

As has been mentioned before, each agent A_{i_1} furthermore is assumed to have one logical Sensor $L_S^{i_1}$ assessing the SSs S_m that the perceived agents (including A_{i_1} itself) are currently in, making use of the outputs of own $L_{j; \{i_1, i_2\}}^{i_1}$ and of the communicated output of other agent's $L_{j; \{i_2, i_3\}}^{i_2}$.

The sensors $L_{j; \{i_k, i_l\}}^{i_k}$ communicate their results as binary SL opinions $\omega_{j; \{i_k, i_l\}}^{(i_k)}(t) = (b, d, u, a)(t)$ for each time step t , where belief b , disbelief d and uncertainty u are functions of $(p^{\oplus}, p^{\ominus})_{j; \{i_k, i_l\}}^{(i_k)}(t)$. Taking into account appropriate Laplace estimators, the base rate a may be determined as $a = (x + 1)/(y + 2)$, where x denotes the number of past encounters of A_{i_k} with A_{i_l} in SSs and y denotes the total number of SSs that agent A_{i_k} has recorded (both in a time window $[t - \Delta t, t]$). We will elaborate more on alternatives for these functions in section IV.

Each agent A_{i_1} can then combine opinions $\omega_{j_1; \{i_1, i_2\}}^{(i_1)}(t)$ from his own sensors and opinions $\omega_{j_2; \{i_1, i_2\}}^{(i_2)}(t)$ communicated by sensors of perceived agents A_{i_2} with the help of SL fusion operators:

$$\omega_{\{i_1, i_2\}}^{(i_1)}(t) = \left[\underbrace{\oplus}_{j_1 \in J_1} \left(\omega_{L_{j_1; \{i_1, i_2\}}^{i_1}}^{(i_1)}(t) \otimes \omega_{j_1; \{i_1, i_2\}}^{(i_1)}(t) \right) \right] \underbrace{\oplus}_{j_2 \in J_2} \left[\underbrace{\oplus}_{j_2 \in J_2} \left(\omega_{L_{j_2; \{i_1, i_2\}}^{i_2}}^{(i_2)}(t) \otimes \omega_{j_2; \{i_1, i_2\}}^{(i_2)}(t) \right) \right] \quad (4)$$

Other agents are 'perceived' by A_{i_1} if located within its social sphere [36] (e.g. as a simple approximation are within range of near-field communication means (such as Bluetooth)). The operator $\underbrace{\oplus}$ is the Averaging Fusion operator, employed to the fusion of, in this case, competitive sensors with dependent opinions (indexed by j). The operator \otimes is the Uncertainty Favoring Discounting operator. The pragmatics in using these operators for our purposes was discussed in section II. The opinions $\omega_{L_{j_1; \{i_1, i_2\}}^{i_1}}^{(i_1)}(t)$ model trust in the own sensors $L_{j_1; \{i_1, i_2\}}^{i_1}$. Trust in the foreign agent's sensors $\omega_{L_{j_2; \{i_1, i_2\}}^{i_2}}^{(i_2)}(t)$ may be summarily replaced by an opinion $\omega_{A_{i_2}}^{(i_1)}(t) = (b, d, u, a)(t)$, modeling general trust in the agents A_{i_2} as explained in section II. Here, the belief b , disbelief d and uncertainty d can be won by statistically analyzing the distance of agent A_{i_2} 's communicated subjective assessment of the SSs from the consensus SSs in a time window $[t - \Delta t, t]$ (see discussion below). As a value for the base rate, $a = 0.5$ represents an unbiased choice.

Each agent A_{i_1} then requests all perceived agents A_{i_2} to send their opinions $\omega_{\{i_2, i_3\}}^{(i_2)}(t)$ on being in SS with further agents $A_{i_3} \neq A_{i_1}$. Analogous to equation (4), these communicated opinions may also be \otimes -weighted with general trust in these agents:

$$\omega_{\{i_2, i_3\}}^{(i_1)}(t) = \omega_{A_{i_2}}^{(i_1)}(t) \otimes \omega_{\{i_2, i_3\}}^{(i_2)}(t) . \quad (5)$$

Together with opinions from other sensors (e.g. addressing the agent's positions), $L_S^{i_1}$ can then make use of these $\omega_{\{i_k, i_l\}}^{(i_1)}(t)$ in order to output A_{i_1} 's current subjective estimation of the SSs $S_m = (P_m, T_m, X_m, K_m)$ among the perceived agents. Here, due to lack of space, we will mainly focus on methods for $L_S^{i_1}$'s estimation of P_m (the persons involved).

The estimation of P_m proceeds as follows: the $\omega_{\{i_k, i_l\}}^{(i_1)}(t)$ allow $L_S^{i_1}$ for every t to construct a complete graph $G_{i_1}(t) = (V, E, w)(t)$ summarizing A_{i_1} 's assessments of the SS involvement probabilities of perceived agents. The nodes V correspond to the set of perceived agents and the weights $w(E, t)$ of the edges $E = (i_k, i_l)$ are the probability expectations of the opinions $\omega_{\{i_k, i_l\}}^{(i_1)}(t)$. Thus $G_{i_1}(t)$ represents A_{i_1} 's estimation of the probabilistic, situational social network between the perceived agents (including A_i itself). $L_S^{i_1}$ may then use techniques for identifying dense sub-networks of $G_{i_1}(t)$ to compute a partition of set of perceived agents. This partition is then A_{i_1} 's estimation of the P -components of the SSs surrounding and involving A_{i_1} at time t . Any non-overlapping graph clustering approach (see [37] or [38]) may in principle be applied. We assume that SSs of the type we focus will not significantly overlap (as explained in section II).

In chapter IV we will discuss and evaluate several alternatives.

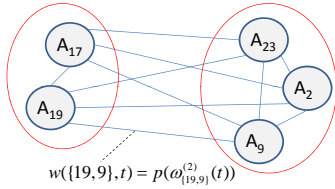


Figure 2. Example of agent A_2 's graph $G_2(t)$. The edge's lengths are depicted as inversely proportional to their weights. The current graph may give rise to assuming two SSs involving $P_1 = \{A_{17}, A_{19}\}$ and $P_2 = \{A_{23}, A_2, A_9\}$

At this point of analysis, each agent A_{i_1} has some own current subjective estimation of the SSs $\{S_m\}$ among the perceived agents and in particular its *own* SS (the unique SS $S(A_{i_1})$ it assumes it is in) Agents then deliberate on the (inter-subjectively) 'true' local landscape of SSs at each t , ultimately finding a consensus view. This consensus view can be used by each agent to assess the differences between the consensus and its own estimation and the differences between the consensus and estimations of other agents. These differences may then be used to determine trust values with respect to other agents or may be used for correcting $S(A_{i_1})$. Finding a consensus may be broken down to two problems: (1) establishing a set

of agents likely to be in a common SS on the basis of their subjective assessments and all having full knowledge of the set's to deliberate consensus (2) reaching an agreement among these agents on a consensus SS.

Our basic approach for solving (1) is as follows:

Time synchronization between agents is accomplished using an absolute reference time. The consensus protocol is periodically executed at fixed points in time, with a heuristic period τ properly reflecting the maximal dynamics of human SSs (e.g. $\tau = 5$ sec).

[Step1]: Each agent A_{i_1} pushes its $[t - \tau, t]$ -smoothed subjective estimation $S(A_{i_1}) = (P_{S(A_{i_1})}, T_{S(A_{i_1})}, X_{S(A_{i_1})}, K_{S(A_{i_1})})$ to all agents $A_{i_k} \in P_{S(A_{i_1})}$ and also requests all of their smoothed estimations $S(A_{i_k})$. Including both ensures that each agent acquires a sufficiently broad view, i.e. ensures that all candidate SS are actually acquired. Smoothing is done for each component separately.

[Step2]: Each agent A_{i_1} requests all agents $A_{i_n} \in P(\mathcal{S}_{\text{socSphere}} = \mathcal{S}_{\text{req}} \cup \mathcal{S}_{\text{push}} \cup \{S(A_{i_1})\})$ to enter into a consensus process. $\mathcal{S}_{\text{req}} = \cup_k \{S(A_{i_k})\}$ is the set of SSs received on request, $\mathcal{S}_{\text{push}} = \cup_m \{S(A_{i_m})\}$ is the set of SSs received by push. $\mathcal{S}_{\text{socSphere}}$ denotes the candidate social situations in agent's social sphere.

[Step3]: Each agent A_{i_1} accepts or declines these requests using a decision function df . df operates on distances/similarities between SSs in $\mathcal{S}_{\text{socSphere}}$ effectively performing a clustering of $\mathcal{S}_{\text{socSphere}}$, with $df = \text{'accept'}$ for the agents in the cluster(s) that contain the most candidate SSs that A_{i_1} is also in.

[Step4]: Each agent A_{i_1} thus has a star-shaped graph G induced by mutual agreement. If its corresponding user is not in a SS, its graph will have few edges or no edges at all.

[Step5]: Each agent A_{i_1} uses the distributed community detection algorithm MODULARITY described in [39] to collaboratively determine with the other perceived agents the set of agents X it will be entering into consensus deliberations with.

With respect to problem (2), in order to determine a consensus SS in $[t - \tau, t]$, the agents in each set X may then employ Averaging Fusion of their respective SS assessment adequately formulated as SL opinions via evidence based formulation, if the SS assessment of the set is sufficiently homogeneous or else employ the approach of Rosenschein and Zlotkin (see chapter II).

While this approach cannot absolutely guarantee that each agent will deliberate consensus only in one set X , social reality and locally restricting communication will ensure that the number of deliberations will remain low and each agent can decide on the basis of its results and df which consensus it accepts.

For this approach of finding a consensus as well as for our proposed variant of determining trust in other agents, a distance measure $d(S_a, S_b)$ (or respective similarity measure $s = 1/(1 + d)$) between SSs is required. We propose a weighted Manhattan distance $d(S_a, S_b) = wPd(P_a, P_b) +$

$w_X d(X_a, X_b) + w_T d(T_a, T_b) + w_K d(K_a, K_b)$ where the weights w can be determined by each agent according to its preferences. For the component distances the basic Jaccard coefficient is used (e.g for person distance $d(P_a, P_b) = |P_a \cap P_b|/|P_a \cup P_b|$). More sophisticated distance / similarity measures are also possible. E.g. for the semantic similarity $s(K_a, K_b)$, the approaches in [40] or [36] may be used. Spatio-temporal similarities $s(\tilde{X}_a, \tilde{X}_b)$ can be determined by a density-based approach, generalizing Jaccard: $s = \int d^4x \min(\rho_a(x), \rho_b(x)) / \int d^4x \max(\rho_a(x), \rho_b(x))$ where $\rho(x)$ are spatio-temporal densities for SSs on \mathbb{R}^4 . These densities $\rho_{\tilde{X}_a}$ may be discretely approximated by A_i assessing $S_a = S(\tilde{A}_i)$ via equi-temporally collecting location measurements from the $A_j \in P_a$ or, less privacy invasive, via equi-temporally heuristically estimating the spatial extension of its social circle (see [36]).

The basic ACL primitives, communication protocols and further questions concerning temporal and spatial measurements, synchronization and references involved in our approach are specified in [41].

IV. EVALUATION

A key element of our approach is each agent A_{i_1} 's logical sensor $L_S^{i_1}$, assessing the SSs S_m in A_{i_1} 's social sphere, operating on SL opinions $\omega_{\{i_k, i_l\}}^{(i_1)}(t)$. While a realistic evaluation of the full consensus process and trust mechanism is out of scope of this contribution, we will discuss an evaluation of options for construction of $L_S^{i_1}$ with the help a social scenario data-set, which we will also use to evaluate our concept of combining opinions from several sensors $L_{j;\{i_1, i_2\}}^{i_1}$.

To acquire this data set (which is more precisely described in [6] [5]), conversational audio and precise interaction geometry $(\delta d, \delta \theta)_{i_1, i_2}(t)$ for nine persons were recorded for ≈ 32 minutes in an interactive social scenario in which SSs changed frequently. Geometry was measured at a rate of 6 frames/sec. As discussed in section III, we constructed logical sensors $L_{\text{GEO};\{i_1, i_2\}}^{i_1}$ which compute the interaction probabilities $(p^\oplus, p^\ominus)_{\text{GEO};\{i_1, i_2\}}^{(i_1)}(t)$ on the basis of interaction geometry (see [6] [5]).

In order to detect SS on the basis of conversational audio, won by low-fi MP3-player microphones worn around the neck, we constructed a simple direct approach $L_{\text{AUDIO};\{i_1, i_2\}}^{i_1}$ bypassing turn-taking patterns and HMM model generation, which computes $p_{\text{AUDIO};\{i_1, i_2\}}^{\oplus, (i_1)}(t)$ directly on the basis of correlations of low-level MFCC audio coefficients of two speakers U_{i_1} and U_{i_2} provided by level Ib logical sensors $L_{\text{MFCC}}^{i_1}$ and $L_{\text{MFCC}}^{i_2}$. The correlation is estimated twice per second as either corresponding to an SS or not corresponding to an SS with a KNN classifier whose output is used to compute $p_{\text{AUDIO, frame};\{i_1, i_2\}}^{\oplus, (i_1)}(t)$, and averaging over a sliding window time-interval of 60 seconds (= 120 frames) yielding $p_{\text{AUDIO};\{i_1, i_2\}}^{\oplus, (i_1)}(t)$. The approach (described in [42]) is semi-supervised, only requiring an initial choice between elements of generally trained set of rough profiles of the audio situation (e.g. whether indoor or outdoor), which can be algorithmically determined.

In order to investigate options for basically constructing binary SL opinions $\omega_{j;\{i_1, i_2\}}^{(i_1)}(t) = (b, d, u, a)$ from the $(p^\oplus, p^\ominus)_{j;\{i_1, i_2\}}^{(i_1)}(t)$ and to evaluate the fusion of opinions (4), we will first investigate the classifier performances of $\omega_{\text{GEO}} = \omega_{\text{GEO};\{i_1, i_2\}}^{(i_1)}(t) = \omega_{\text{GEO};\{i_1, i_2\}}^{(i_1)} \oplus \omega_{\text{GEO};\{i_1, i_2\}}^{(i_2)}$ and of $\omega_{\text{GEO}} = \omega_{\text{GEO};\{i_1, i_2\}}^{(i_1)}$ against the manual annotations with the help of various decision functions f_i . Since our experiment is too limited to provide grounds for computing trust values for other agents, we have to drop the \otimes operations involving trust. Basic variants considered for constructing binary SL opinions are V1 with $b = p^\oplus, d = p^\ominus, u = 1 - b - d, a$ and V2 with $b = 0.75p^\oplus / (p^\oplus + p^\ominus), d = 0.75p^\ominus / (p^\oplus + p^\ominus), u = 0.25, a$, where the base rates a are determined as explained in section III.

Table I shows the results. \boxplus indicates a decision in favor of social interaction and \boxminus against social interaction.

Decision-function (DF)	What	Version	Accuracy	Precision	Recall
$f_1 = \begin{cases} \boxplus, & \text{if } b \geq d \\ \boxminus, & \text{else} \end{cases}$	ω_{GEO}	V1	0.731	0.676	0.764
		V2	0.761	0.721	0.758
		V1, V2	0.757	0.719	0.748
$f_2 = \begin{cases} \boxplus, & \text{if } b - d > d/2 \\ \boxminus, & \text{else} \end{cases}$	ω_{GEO}	V1	0.739	0.724	0.669
		V2	0.772	0.796	0.656
		V1, V2	0.760	0.768	0.664
$f_3 = \begin{cases} \boxplus, & \text{if } b - d > (b + d)/2 \\ \boxminus, & \text{else} \end{cases}$	ω_{GEO}	V1	0.723	0.802	0.503
		V2	0.746	0.874	0.502
		V1, V2	0.736	0.841	0.502

Table I
CLASSIFICATION PERFORMANCE OF GEO PER FRAME OF THREE VARIANTS OF DECISION FUNCTIONS. PRECISION AND RECALL ARE EVALUATED WITH RESPECT TO CLASS \boxplus

While f_1 is the most straightforward direct classifier (for all variants, we neglect class priors for the sake of simplicity, since in our experiment $|\boxplus| \approx |\boxminus|$), f_2 and f_3 demand that p^\oplus must be significantly larger than p^\ominus for a \boxplus decision. While the stricter decision functions will in most cases achieve a better precision at the cost of a lower recall as expected, the overall accuracy of all f_i does not differ significantly. In general, we see that V2 is more accurate and precise than V1. Furthermore, if the absolute values of p^\oplus and p^\ominus are small as a consequence of the model (as in GEO), u may become very large in V1, influencing probability expectation calculations in the expression 4 when involving several agents. This may be avoided by scaling as in V2 where $u = 0.25$ reflects the approximate overall accuracies of table I. Thus V2 may generally be preferred over V1. Average Fusioning the V2 opinions of the two agents (case $\omega_{\text{GEO}}, V2$ yields the best results in all cases, thus V2 will be used from now on.

Furthermore evaluating the classifier performances of $\omega_{\text{AUDIO}} = \omega_{\text{AUDIO};\{i_1, i_2\}}^{(i_1)}(t) = \omega_{\text{AUDIO};\{i_1, i_2\}}^{(i_1)} \oplus \omega_{\text{AUDIO};\{i_1, i_2\}}^{(i_2)}$ and of an Average Fusion of GEO and AUDIO

$$\omega_{\text{AUDIO} \oplus \text{GEO}} = \omega_{\text{AUDIO} \oplus \text{GEO};\{i_1, i_2\}}^{(i_1)}(t) = (\omega_{\text{AUDIO};\{i_1, i_2\}}^{(i_1)} \oplus \omega_{\text{GEO};\{i_1, i_2\}}^{(i_1)}) \oplus (\omega_{\text{AUDIO};\{i_1, i_2\}}^{(i_2)} \oplus \omega_{\text{GEO};\{i_1, i_2\}}^{(i_2)})$$

in the same fashion yields the results in table II. The results especially show that combining different sensors for SS detec-

DF	What	Version	Accuracy	Precision	Recall
f_1	ω_{AUDIO}	V2	0.737	0.709	0.695
	$\omega_{\text{AUDIO} \oplus \text{GEO}}$	V2	0.785	0.756	0.767

Table II
CLASSIFICATION PERFORMANCE PER FRAME OF AUDIO AND OF
AVERAGE FUSION OF GEO AND AUDIO

tion can lead to better results than any of the sensors alone. While these results may appear obvious for human observers of social interaction, they support the usefulness of SL fusion for a frame-based social interaction classification as a basis for our proposed algorithmic procedure for SS detection.

We will now further evaluate this procedure and certain options for the clustering part described in section III which is applied after having collected and processed $\omega_{\{i_2, i_3\}}^{(i_2)}(t)$ from the other perceived agents A_{i_2} . We evaluate the algorithm for our different sensor classes separately which means separately evaluating the respective performances of $L_{\text{GEO};S}^{i_1}$ and $L_{\text{AUDIO};S}^{i_1}$. We compare these to the performance of an $L_{\text{AUDIO} \oplus \text{GEO};S}^{i_1}$ using an average of opinions won from GEO sensors and opinions won from AUDIO sensors.

The weights $w(E = \{i_k, i_l\}, t) = w(\omega_{\{i_k, i_l\}}^{(i_1)}(t))$ of the subjective situational social network $G_{i_1}(t) = (V, E, w)(t)$ in an agent's social sphere are computed as the SL probability expectation of $\omega_{\{i_k, i_l\}}^{(i_1)}(t)$. For the evaluation, we uniformly employ $a = 0.5$ as SL priors for all pairs due to the limited scope of the experiment. For computing non-overlapping clustering partitions of $G_{i_1}(t)$, we employ agglomerative Single Link (SiL), Complete Link (CoL) and Average Link (AvL) approaches (see [37]) with weighted Maximum Modularity [43] as cluster validity index to determine the optimal cutting level of the dendrogram. Furthermore we also employ direct greedy maximization (grMM) of Maximum Modularity as proposed in [43].

The manual annotation $\mathcal{A}(t)$ [6] denotes the SSs for every t (including pseudo-SS with singleton sets P for persons not in a SS at time t) and thus also effectively partitions the set of agents. In order to compare the annotation with the partition $\mathcal{C}(t)$ of the nodes (agents) resulting from the graph clustering, we employ the RAND index [44] $R(t) = R(\mathcal{A}(t), \mathcal{C}(t)) = (a + b) / \binom{n}{2}$ where a denotes the number of pairs of elements that are in the same cluster in $\mathcal{A}(t)$ and in the same cluster in $\mathcal{C}(t)$ too, b denotes the number of pairs of elements that are in different clusters in $\mathcal{A}(t)$ and in different clusters in $\mathcal{C}(t)$. Table III shows the results of averages $\langle R(t) \rangle_t$ of $R(t)$ over t for the different variants. Using random partitions $\mathcal{C}_{\text{rndm}}(t)$ as a baseline, the ordinary RAND index will, in our case, produce vales of $\langle R(\mathcal{A}(t), \mathcal{C}_{\text{rndm}}(t)) \rangle_t \approx 0.5$, also shown in table III. For large numbers of agents n , ordinary RAND index between random partitions $\langle R(\mathcal{C}_{\text{rndm}}(t), \mathcal{C}'_{\text{rndm}}(t)) \rangle_t$ will even $n \rightarrow \infty$ 1 because of the large probability of many singleton clusters. Thus, an Adjusted RAND index R_{adj} was proposed (see [45]) that compensates for this non-zero random expectation and is also shown in table III.

#	Variant	$\langle R(\mathcal{A}(t), \mathcal{C}(t)) \rangle_t \pm \sigma_t$		
		GEO	AUDIO	GEO \oplus AUDIO
1	AvL	0.766 \pm 0.198	0.737 \pm 0.306	0.777 \pm 0.218
2	SiL	0.751 \pm 0.220	0.690 \pm 0.342	0.777 \pm 0.257
3	CoL	0.761 \pm 0.201	0.740 \pm 0.312	0.776 \pm 0.218
4	grMM	0.755 \pm 0.205	0.741 \pm 0.289	0.771 \pm 0.212
6	Random	$\langle R(\mathcal{A}(t), \mathcal{C}_{\text{rndm}}(t)) \rangle_t = 0.524 \pm 0.233$		

#	Variant	$\langle R_{\text{adj}}(\mathcal{A}(t), \mathcal{C}(t)) \rangle_t \pm \sigma_t$		
		GEO	AUDIO	GEO \oplus AUDIO
1	AvL	0.529 \pm 0.374	0.574 \pm 0.415	0.567 \pm 0.388
2	SiL	0.513 \pm 0.370	0.536 \pm 0.435	0.603 \pm 0.388
3	CoL	0.518 \pm 0.380	0.581 \pm 0.425	0.561 \pm 0.392
4	grMM	0.507 \pm 0.395	0.564 \pm 0.402	0.549 \pm 0.390
6	Random	$\langle R_{\text{adj}}(\mathcal{A}(t), \mathcal{C}_{\text{rndm}}(t)) \rangle_t = 0.022 \pm 0.181$		

Table III
MEANS AND STD.-DEV. OF THE RAND INDEX AND ADJUSTED RAND
INDEX FOR DIFFERENT COMBINATIONS OF CLUSTERING APPROACHES /
WEIGHTING SCHEMAS.

Keeping in mind that the adjusted Rand index can be considered a not very optimistic estimation of partition matches, the overall results show that a satisfactory level of recovery of the annotated (real) SS can be reached with the proposed approach.

The most important finding from these experiments is that the approach working with the fused opinions GEO \oplus AUDIO works significantly ³ better for most cases than GEO and AUDIO alone. Together with the previously discussed classifier results, this shows that combining two different social signal sources may provide a clearer overall picture of the SSs.

This analysis is justified because a deeper analysis of the variants where $\langle \mathcal{R}_{\text{adj}; \text{AUDIO}} \rangle_t$ is slightly better than $\langle \mathcal{R}_{\text{adj}; \text{GEO} \oplus \text{AUDIO}} \rangle_t$ reveals that for a high number of frames corresponding to a common SS of all participants, the better values of $\mathcal{R}_{\text{adj}; \text{AUDIO}}(t)$ are artifacts: As also supported by table II, for a substantial number of frames, the pairwise interaction probabilities of AUDIO provide a diffuse picture, which leads to a maximum of Modularity for the 1-clustering of $G(t)$. Averaged over all variants, these cases occur over 4 times as often for AUDIO than for GEO \oplus AUDIO. Since for over 2000 frames, common SSs of all participants corresponding to a real 1-clustering occur in our experiment, these cases lead to a coincidental agreement between AUDIO and the annotation. If these cases are left out, GEO \oplus AUDIO is substantially better than AUDIO for *all* variants $V = \{\text{AvL}, \text{SiL}, \text{CoL}, \text{grMM}\}$ ($\langle \langle \mathcal{R}_{\text{adj}; t} \rangle_V \rangle_t = 0.062$).

V. CONCLUSION

We presented an approach for combining evidences for Social Situation detection from various sources (e.g. sensors from other agents or working on different types of data) based

³e.g. Hypotheses 'Mdn($\mathcal{R}_{\text{AVL}; \text{GEO} \oplus \text{AUDIO}}$) = Mdn($\mathcal{R}_{\text{AVL}; \text{GEO}}$)' and 'Mdn($\mathcal{R}_{\text{adj}; \text{SiL}; \text{GEO} \oplus \text{AUDIO}}$) = Mdn($\mathcal{R}_{\text{adj}; \text{SiL}; \text{AUDIO}}$)' are rejected (Wilcoxon Rank Sum Test, $n = 11467$, $\alpha = 0.05$). For $n = 11467$ random values from corresponding $\mathcal{N}(\mu, \sigma)_{\text{GEO} \oplus \text{AUDIO}}$ and $\mathcal{N}(\mu, \sigma)_{\text{GEO}}$ or AUDIO , a T-Test rejects ' $\mu_{\text{GEO} \oplus \text{AUDIO}} = \mu_{\text{GEO}}$ ' and ' $\mu_{\text{GEO} \oplus \text{AUDIO}} = \mu_{\text{AUDIO}}$ ' at $\alpha = 0.05$ -level.

on Subjective Logic opinions. This approach allows to mediate between the difficulties of a low level sensor data or model parameter exchange on the one hand and a simple exchange of end-results. We evaluated key aspects of the approach using a data set of Social Situations and, besides positively evaluating the suitability of the approach for Social Situation detection, showed that combining sensors from other agents and different signals sources with SL fusion operators is able to provide a better (clearer) picture of the landscape of Social Situations than each of the signal sources alone. Future work will continue to further develop and evaluate the consensus finding part of the approach, which cannot reasonably be investigated with the current data-set. It may also be interesting to investigate the suitability of the method for a deeper Social Situation understanding (e.g. investigating emotions of the participants).

REFERENCES

- [1] M. Argyle, A. Furnham, and J. A. Graham, *Social Situations*. Press Syndicate of the University of Cambridge, 1981.
- [2] G. Groh, A. Lehmann, T. Wang, S. Huber, and F. Hammerl, "Applications for social situation models," in *Proc. WAC2010, Freiburg, Germany*, 2010.
- [3] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743 – 1759, 2009.
- [4] A. Pentland, N. Eagle, and D. Lazer, "Inferring social network structure using mobile phone data," *PNAS* 106(36), pp. 15274–15278, 2009.
- [5] G. Groh and A. Lehmann, "Deducing evidence for social situations from dynamic geometric interaction data," *International Journal of Social Computing and Cyber-Physical Systems (IJSCCPS)*, (to appear), 2011.
- [6] G. Groh, A. Lehmann, J. Reimers, R. Friess, and L. Schwarz, "Detecting social situations from interaction geometry," *Proc. IEEE SocialCom 2010, Minneapolis USA*, 2010.
- [7] C. A. Yeung, I. Lliccardi, K. Lu, O. Seneviratne, and T. Berners-Lee, "Decentralization: The future of online social networking," in *W3C Workshop on the Future of Social Networking*, 2009.
- [8] J. Breslin, A. Passant, and S. Decker, *The social semantic web*. Springer Verlag, 2009.
- [9] S. Buchegger, D. Schiöberg, L. Vu, and A. Datta, "PeerSoN: P2P social networking: early experiences and insights," in *Proc. 2nd ACM EuroSys Workshop on Social Network Systems*, 2009, pp. 46–52.
- [10] N. Eagle and A. Pentland, "Reality mining: sensing complex social systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, 2006.
- [11] A. Madan and A. Pentland, "Modeling social diffusion phenomena using reality mining," in *AAAI Spring Symposium on Human Behavior Modeling, Palo Alto, CA*, 2009.
- [12] R. Brooks and S. Iyengar, *Multi-sensor fusion: fundamentals and applications with software*. Prentice-Hall, 1998.
- [13] H. Qi, S. Iyengar, and K. Chakrabarty, "Distributed sensor networks—a review of recent research* 1," *Journal of the Franklin Institute*, vol. 338, no. 6, pp. 655–668, 2001.
- [14] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Springer, October 2007.
- [15] S. Ferrari and A. Vaghi, "Demining sensor modeling and feature-level fusion by Bayesian networks," *Sensors Journal, IEEE*, vol. 6, no. 2, pp. 471–483, 2006.
- [16] Y. Zhang and Q. Ji, "Active and dynamic information fusion for multi-sensor systems with dynamic Bayesian networks," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 36, no. 2, pp. 467–472, 2006.
- [17] A. Jøsang, "Probabilistic logic under uncertainty," in *Proceedings of the thirteenth Australasian symposium on Theory of computing-Volume 65*. Australian Computer Society, Inc., 2007, pp. 101–110.
- [18] A. Josang, "A logic for uncertain probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 3, pp. 279–311, 2001.
- [19] G. Shafer, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.
- [20] A. Dempster, "A generalization of Bayesian inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no. 2, pp. 205–247, 1968.
- [21] A. Josang, "The consensus operator for combining beliefs," *Artificial Intelligence*, vol. 141, no. 1-2, pp. 157–170, 2002.
- [22] A. Jøsang, S. Marsh, and S. Pope, "Exploring different types of trust propagation," *Trust Management*, pp. 179–192, 2006.
- [23] A. Jøsang, "The Cumulative Rule for Combining Beliefs," <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.70.6931> (April,2011), 2006.
- [24] A. Jøsang, J. Diaz, and M. Rifqi, "Cumulative and averaging fusion of beliefs," *Information Fusion*, vol. 11, no. 2, pp. 192–200, 2010.
- [25] W. Bamberger, J. Schlittenlacher, and K. Diepold, "A trust model for intervehicular communication based on belief theory," *IEEE Socialcom 2010*, 2010.
- [26] M. Wooldridge, *An introduction to multiagent systems*. Wiley, 2009.
- [27] M. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.
- [28] J. Rosenschein and G. Zlotkin, *Rules of encounter: designing conventions for automated negotiation among computers*. The MIT Press, 1994.
- [29] D. Wyatt, T. Choudhury, and J. Bilmes, "Conversation detection and speaker segmentation in privacy sensitive situated speech data," in *Proceedings of Interspeech*, 2007, pp. 586–589.
- [30] D. Wyatt, T. Choudhury, and J. Bilmes, "Creating Social Network Models from Sensor Data," in *Proc. NIPS Network Workshop*, 2007.
- [31] D. Wyatt, J. Bilmes, T. Choudhury, and J. Kitts, "Towards the automated social analysis of situated speech data," in *Proc. 10th Int'l Conf. on Ubiquitous Computing*. ACM, 2008, pp. 168–171.
- [32] D. Wyatt, T. Choudhury, and H. Kautz, "Capturing spontaneous conversation and social dynamics: A privacy-sensitive data collection effort," in *Proc. IEEE ICASSP 2007*, 2007.
- [33] O. Brdiczka, J. Maisonnasse, and P. Reignier, "Automatic detection of interaction groups," in *Proceedings of the 7th international conference on Multimodal interfaces*. ACM, 2005, pp. 32–36.
- [34] G. Groh, A. Lehmann, and M. deSouza, "Detection of social situations with turn-taking patterns," (to appear), 2011.
- [35] D. Gatica-Perez, "Automatic nonverbal analysis of social interaction in small groups: A review," *Image and Vision Computing*, vol. 27, no. 12, pp. 1775–1787, 2009.
- [36] G. Groh, "Groups and group instantiations – detection, modeling and applications," *Proc. AAAI International Conference on Weblogs and Social Media (ICWSM07), Boulder, Co, USA*, 2007.
- [37] M. Gaertler, "Clustering," in *Brandes, U.; Erlebach, T. (eds.): Network Analysis, LNCS3418*. Springer, 2005, pp. 178–215.
- [38] M. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, p. 8577, 2006.
- [39] P. Hui, E. Yoneki, S. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture*. ACM, 2007, p. 7.
- [40] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme, "Evaluating similarity measures for emergent semantics of social tagging," *Proc. 18th WWW conference*, pp. 641–650, 2009.
- [41] C. Fuchs, *Kombination multipler Evidenzen zur Detektierung sozialer Situationen*. Master's Thesis, TU-München, SS 2010; Supervisor: Georg Groh, 2010.
- [42] C. Peppmeier, *Speaker Diarization for Social Situation Detection*. Diploma Thesis, TU-München; Supervisor: Georg Groh, 2011.
- [43] M. Newman, "Analysis of weighted networks," *Physical Review E*, vol. 70, no. 5, p. 56131, 2004.
- [44] W. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical association*, vol. 66, no. 336, pp. 846–850, 1971.
- [45] L. Hubert and P. Arabie, "Comparing partitions," *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.