

The original article is available at [www.iospress.nl](http://www.iospress.nl).

# A Generic Model of Distrust Behaviour in Online Communities of Trust

David ZEJDA<sup>a</sup>

<sup>a</sup>*Faculty of Informatics and Management, UHK, Czech Republic*

**Abstract.** With boost of interest in multi-user networking ranging from Web 2.0 technologies to agent systems and smart environments, various aspects of mutual interactions between system components, including but not limited to human users themselves, have to be considered. Trust is one of core dimensions of vital, multilaterally beneficial interactions. In the first section of the paper we describe the trust itself and its characteristics, such as multidimensionality, contextuality, asymmetry, transitivity, scope, disproportion, and dynamics. Further we present a niche type of community where users trust each other as default and where the trust loses most of its subjective flavour, which we call a community of trust. Common trust models do not apply for the community of trust well, because they build upon different presumptions about trust. As a main contribution of the paper, we introduce a simple generic and extensible model of distrust for the community of trust. The model is based on activity and distrust matrices arranged into vectors in order of time of occurrence. Various derived characteristics, such as harmful/harmless or distrusting/distrusted ratios provide additional insight on the model. Tests in simulated scenarios and with real human users will follow in our future research in order to concretize and evaluate the model.

**Keywords.** Distrust, trust, community of trust, social networks, interactions

## Introduction

Both dynamics of our social relations and also individual social interactions are highly influenced, if not even governed, by trust. Trust may be defined as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.” [1] For our work we decided to adopt slightly simpler definition, by Goldbeck et al. [2]: “Trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome.”

In real life trust emerges primarily from our experiences with others, their acts, words, their willingness to help us in difficulties, their promises which have been kept. Our trust may be also based on recommendation or guarantee from those, who we trust already. With trusted people we deal differently than with strangers. The level of trust which we feel toward someone helps us to decide whether to rely on his promises or whether to entrust him certain information, task or other person to care for. In general, trust grows slowly, but falls sharply [3]. It may take months or years before we credit someone, whereas a single act of betrayal destroys the trust from the roots.

We all belong to a global-world village. As expressed in the small world phenomenon, everyone is connected with anyone else through only several steps of

relations [4]. New social strategies are necessary to cope with the social and information overload [3]. Web becomes not only bigger every year, but also semantically richer and more driven by its users, who are no longer consumers only, but consumers and producers in parallel. Besides acting as a space for implicit socialization [5], the web provides variety of purposely social spaces, such as dating sites, community portals and social networking sites. As the web fades into our lives, new dangers arise. Actually, reputation of the sites has been affected by many incidents already. For example, according to Notts Police [6] there was a 346% rise in the number of incidents reported to them which involve Facebook during only first three months of 2010 in comparison with the whole year 2009. The resulting charges include harassment, attempted rape and other crime. In other cases paedophiles or murderers met their victims online.

Besides the social networks, we may identify many other contexts where trust is essential. One of them is cloud computing and web services, where for example a confidential information has to flow through various systems, owned and maintained by different parties, to achieve a desired effect. Even in smaller contexts, such as instalments of ambient intelligence in residential, business, healthcare or education smart environments, importance of trust solutions grow in parallel with grow of complexity of the previously mentioned systems. In many applications another actors besides humans, artificial software agents or smart pieces of hardware enter the environments, reside there, interact with each other. Human users are not the only participants in trust relationships to be considered. The same question whom to trust appears again and again, sometimes in new situations or new ways. The impact of decisions taken on the trust evaluation may range from negligible, if the result of the decision is e.g. which song should the trust evaluator consider for inclusion in his playlist, to very consequential, if money, emotions, or health or even someone's life is entrusted to the other party.

Online trust issues bring more attention nowadays in all the mentioned contexts. Efficient and appropriate solutions able to assist in establishing and maintaining trust among participants in online networks on one hand and able to reveal distrust behaviour and propagate it to those who are concerned on the other hand are in need. As a core of the trust solutions there is usually a model – a set of rules which reflect characteristics of trust, both general and specific for a particular scenario. Further, both the model itself and the whole trust system, which implements model in a network, has to be reliable and trusted by the involved participants. In this sense, trust among participants (nodes, users, agents) and trust in the network itself (typically represented in a web interface) and its provider are significantly interrelated and cannot be considered separately – trust determined by a distrusted system has no value.

In the paper we describe trust first, then we follow with introduction of the community of trust, where trust among all users is a default state. Finally, we bring a simple generic model of trust designed specifically to respect characteristics of the community of trust.

## **1. Trust Characteristics**

If we wish to formulate a model of trust, let's discuss the trust itself first. Interactions which occur in an artificial environment (the web, an ambient intelligence instalment,

or any other network) are not much more than just technically captured or flavoured reflection of interactions of the real world [7]. So, trust in the systems mirrors general trust characteristics in a great degree too. Meo et al. [8] define three aspects of trust, which could be called multidimensionality, contextuality and scope of relevance. Goldbeck et al. [2] identify transitivity, asymmetry and personalization. Besides the concepts mentioned, to get a more complex view, we added the disproportion of impacts and the dynamics [9]:

*Multidimensionality.* There is no single source of trust, on the contrary various factors may be considered to evaluate trust, such as honesty, experience, precision, efficiency, or cooperativeness of the party. Actually we may mix the indices to get more accurate results. Broader social space brings further dimensions. In a virtual space on one hand we miss many relevant non-verbal indices - we do not see the person in real, sometimes we even do not see him at all, it is also more likely that there are no other trustful people around who could share their opinions based on their direct personal experiences. But on the other hand we may take the whole community and a whole record of previous interactions into account to overcome the drawback.

*Contextuality.* Social context and purpose of trust evaluation affect the trust requirements and the process of trust formation. For example, when we search for reliable advices on particular topic, we prefer experts on the domain, so our trust is contextually-dependant and determined by both moral credit of the party, but also by its experiences and abilities.

*Asymmetry.* Even though trust in one direction may encourage formation of trust in the reverse direction, trust is not necessarily reciprocated. Trust of one to another does not imply trust in reverse direction. Graph of trust is usually directed, matrix of trust is not necessarily symmetrical.

*Transitivity.* How to establish trusty relationships in an online social network, where no existing trusted social network is present in the background? How to measure the reliability of advices provided by strangers? Assuming the trust is transitive, it gives sense to follow trust relations to infer trust among those who do not trust each other yet. Multiplication along the path [10] effectively discounts the resulting value, thus those whom the user trusts already are being taken as a more reliable source of recommendation whom else to trust. Many refined the basic idea to fit different scenarios or to overcome some of its weaknesses. For example Walter et al. [11] introduced an algorithm which does not reduce cycles in a graph before computation as most other algorithms do. Hales et al. [12] use similar algorithm to find cooperative routes among selfish agents acting as players in prisoner's dilemma, in an environment with no central authority. A lot of research has been done on the trust inference, such as [13] [14] [15] [16].

*Scope of relevance, personalization.* There is a difference between subjective and objective trust. Many current models treat trust as inherently subjective [2]. Meo et al. [8] distinguish 'trust', 'reputation' and 'reliability' levels. Objectivity of trust may be also scaled, from a community-wide reputation reflecting the nature of the community, through a system-wide trustworthiness, keeping only the virtue of "being good", further up to a world-wide trust identity. Trust identity exceeds borders of systems to be shared among them, such as if certain user trustworthy on Facebook would be automatically considered trustworthy on Twitter and LinkedIn too and also in various local contexts, including his smart office or his house equipped with ambient intelligence. We are still lacking solutions for this level of trust [17]. There are strong relations between the levels, such as that objective trustworthiness implies quicker subjective trust formation

and reversely objective trustworthiness may be constructed out of subjective trusts. Instead of simply summing and normalizing individual subjective trusts, back to the idea of trust transitivity, e.g. eigenvector-based<sup>1</sup> algorithms may be used to weigh individual trust expressions according to user's own trust, so the trustor's objective trust acts as a confidence measure of his own trust expressions [18]. In result, the objective trust of certain user is dependent on objective trust of his neighbours in the graph of trust [11].

*Disproportion of impacts.* We may identify two complementing types of errors in the process of trust emergence. First, an user may be too suspicious that he indeed does trust virtually nobody even if there are objectively sufficient positive indices. We may refer to this type of error as to 'an excessive prudence'. The complementary error, 'a trust to deceiver', occurs if an user is either careless and he does not check proclamations of others enough and believes quickly. While retarded formation of trust (the first type of error in cogitation) brings certain losses, such as lowered convenience, potential harm of the latter type of error is typically much higher. Abuser who enjoys undeserved confidence has wide opportunities for attack and the overconfident user is prone to the fraud attempts. Actually it was the strategy of deceivers to gain confidence first, which has led to the most severe abuse incidents on social networking sites mentioned in the introduction.

*Dynamics.* Caverlee et al. [19] recommend to fold two main sources of information in a well-designed trust metric – a relatively static network topology and users' behaviour. A feedback mechanism may capture the trust-related aspects of interactions, thus provide the necessary dynamics. Moghaddam et al. [20] introduce a model for rapidly evolving networks, with emphasis on feedback as a primary source of trust. Trust undergoes transitions between various states, it may be gained, lowered, or even lost. Goldbeck et al. [21] offer conceptual representations of failures of trust, such as distrust, mistrust, untrust and ignorance. Explicit distrust may be utilized by social networking site maintainers to reveal malicious users, scammers or other betrayals. The loss of trust is not necessarily terminal, when a sufficient regret is followed by forgiveness, which effectively means that the trust lost once is being regained.

The characteristics mentioned above are mutually interrelated. E.g. contextuality brings further dynamics to the model. Yan et al. [22] redefine the trust itself to fold some of its characteristics, the contextuality and scope of relevance in particular: “trustor A trusts trustee B for purpose P under condition C based on root trust R”. The characteristics are reflected in the component C, the condition to trust. Among other characteristics, not discussed in detail in our paper we may mention at least uncertainty, which is stressed e.g. in the fuzzy model of cooperation by Carbó et al [23].

## **2. Trust in Virtual Interactions and the Community of Trust**

What's the source of trust in social networking systems? Trusted friendships may arise out of vital interactions within the context of the artificial interaction space, usually during a sufficient period of time and based on a sufficient intensity of harmless activity within the period. Besides this, existing trust may be captured from a real social background and mapped into the system [3]. For example, if you personally invite someone to join a networking site, you probably know him and trust him, at least

---

<sup>1</sup>One of well-known eigenvector-type algorithms is PageRank by Google.

at certain level. The trust has been established in advance, based on personal experiences and you do not ask the system to help you to evaluate trustfulness of the user. Reversely, you are capable to provide trust indices to the system. The situation is easier in smaller communities with restricted membership.

This model of trust, where trust does not exist at the beginning, but gradually emerges from interactions or enters the system in a form of previous subjective guarantee, is suitable for vast amount of scenarios. For example, it suits very well to big and markedly open communities inhabiting general purpose social networks. However, virtual social spaces are not homogeneous. Advancements in online social networks, availability of hardware, convergence between online social networks, artificial intelligence and smart environments and other trends allow further diverse virtual socialization. Variety of existing communities, such as work groups, tutors and their classes, patients and their attending physicians, moved online already or are in the transition. Other completely new communities emerged as fully virtual from the beginning. All the communities have their own specifics. Perception of virtue of trust is not unique among them, actually though there is a degree of similarity, it differs from community to community. Generally, people are willing to make only the effort, which brings obvious reward to them. Talking about trust, users should be allowed to express their trust in situations and in a way which reflects their pattern of thinking or their habitual approach, which differs per community, otherwise they could be confused or bored, which could lead to lack of cooperation from them. And the lack of cooperation is something least desirable in something so delicate as a system of trust. So, different models of trust are needed to reflect the needs.

From the variety we selected a class of niche communities which we call the *community of trust* as a matter of our particular interest. The main difference is that whereas usually trust is to be gained, in the community of trust its users are being considered trustworthy as default. The scenario reflects groups of people who know each other in advance enough to trust each other deeply (they just trust each other and it is not necessary to provide any indices to the system how firm is the trust), or communities bonded with certain strong principles and rules, which consist the base for their mutual trust. Virtually, trust acts as a key power in these communities, distrustful behaviour is rare there and if occurs, results in expulsion from the community, so the trust there may be preserved. The community of trust may arise e.g. among devoted volunteers sharing particular common interest, in a church which provides strong holds for lives of its members, among relatives, among close friends who know each other for a long time. The fact that the community of trust may emerge in these settings does not mean more than it *may* arise there. Strong and deep trust is indeed not so widespread.

While in other models a trust functionality primarily assists to distinguish who deserves user's trust and who does not, in the community of trust all users trust each other, so the model fulfils rather different purposes:

- *Motivates to good activity* – The sense of trust gently positively fosters fair interactions among the community, which brings deeper feeling of reliance and connectedness in result.
- *Discourages from inappropriate* – Healthful fear of possible consequences negatively motivates users to avoid any kind of bad behaviour.
- *Keeps the community clean* – Trust evaluation and related mechanisms allows to reveal possible intruders, impostors or those who turned bad, in order to cut them out.

According to Goldbeck et al. [2] trust is inherently a personal opinion, so each node has different levels of trust for each other node [8]. For most cases this is perfectly valid, but [2] admit, that systems of absolute (objective) may exist. The community of trust is the case. The community of trust is so tightly coupled that the trust loses most of its subjective flavour there. If someone belongs to the community, all other members trust him. If he loses the trust of one, nobody trusts him more. The trust in the community of trust is objective. The objectivity of trust brings implications for transitivity. Usually it gives sense to infer trust and distrust from the graph of trust relations following paths of transitivity in a given community, but in the community of trust it does not make sense, because transitivity of trust (as well as transitivity of distrust) tends to infinity. The trust becomes globally valid trait of each node. Transitivity of trust in a pure community of trust is total.

Further, in most models the trust is highly dynamic. The dynamics is driven by both changes in network topology and by interactions of users (nodes) [19]. Model for the community of trust does not have to count with so high level of dynamics, because distrust behaviour is usually rare there. But distrust, if occurs, has to be propagated as quickly as possible to the whole community, otherwise the pivotal virtue of trust within the community could be lost. Table 1, published also in [9], outlines the characteristics described above.

**Table 1:** The community of trust in comparison with a common community

<b>a common community</b>	<b>the community of trust</b>
model of trust	model of distrust
distrustful behaviour relatively common	bad behaviour rare, propagate distrust quickly
users are notably cautious	users are not much cautious
pre-validation of users not necessary / possible	users have to prove their membership first
users may express trust or both trust and distrust	users may express distrust or confirm trust
trust is important	trust is pivotal
trust is to be gained	trust is default state
trust is subjective	trust is objective
trust is dynamic	trust is not too dynamic
trust is transitive	distrust is totally transitive

A model of trust itself should be trusted by users, which implies that it should be also understandable. Because trust is so important within the community of trust, it further underlines the requirement to bring a well-designed model and to keep it understandable.

### **3. The Activity**

A community would have no sense without interactions, more to the contrary, users are usually part of a community just to interact – to send and read messages, to post content, to chat, to comment, or else. Interaction is a kind of action with more actors involved, thus the term may cover wide range of activities, from simple to complex. For our purpose we understand interaction as an atomic activity of certain user aiming

other user. Response constitutes other interaction, with roles of participants switched. Each interaction in our model is one-to-one, so action aiming more users, e.g. chat post in a room with more users present, constitutes a set of one-to-one interactions. All interactions within the whole system form a *record of interactions*.

Let's have an *interaction value function*  $v_i$ , which assigns a finite non-negative real value to each interaction, regardless to the time of the occurrence of the interaction. Assuming a single interaction of a particular user to an other particular user at a particular time, the whole interaction space may be captured as a record of  $\text{interaction}_{mnt}$  where the user  $m$  interacted to the user  $n$  at the time  $t$ . Applying the interaction value function, we receive a space of interaction values with the same dimensions  $(m,n,t)$ :

$$v_i(\text{interaction}_{mnt}) = i_{mnt} \quad (1)$$

*Record of interactions of the user  $m$  to the user  $n$*  is  $\text{interaction}_{mn*}$  and similarly record of interactions of the user  $n$  to the user  $m$  is  $\text{interaction}_{nm*}$ , where the asterisk symbol (\*) denotes all recorded values of the dimension. The interaction value has been defined as independent on time, but recent interactions are usually perceived as of a greater value than those which happened long time ago. So, let's have a *discount function* which assigns a discount for the time which passed, and which is defined as:

$$\begin{aligned} \text{for } t = \text{now } d_t &= 1 \\ \text{for } t < \text{now } d_t &\in (0,1), \text{ creasing alongside } t \end{aligned} \quad (2)$$

*Activity of  $m$  towards  $n$  till the time  $t$*  is a discounted sum of single interactions:

$$\begin{aligned} \text{for all defined interactions where } j \leq t \\ a_{mnt} = \sum d_j i_{mnj} \end{aligned} \quad (3)$$

Discounted sum of all interactions within the community till the time  $t$  constitutes a square *matrix of activity* of size  $s$ , where  $s$  is count of users in the system:

$$\begin{aligned} \text{for } i,j \in 0..s \\ A_t = (a_{ijt})_{s,s} \end{aligned} \quad (4)$$

Bearing the total transitivity of trust in our model on mind, row sums represent the activity of a certain user till a certain time (e.g.  $a_{m*t}$ ) and column sums represent an activity aimed on a certain user till a certain time (e.g.  $a_{*nt}$ ). For each  $t$  the corresponding activity matrix captures all activities which occurred till the time  $t$ , thus the line of the matrices  $A_0, A_1, \dots, A_{\text{now}}$  reflects the overall dynamics of interactions within the system.

#### 4. The Distrust

The sole fact that an user is a member of the community is enough for others to trust him. It is not necessary to define a model of trust, harmless interactions of users just confirm their good intentions, thus prove their trustworthiness, pre-assumed by others.

But, bad behaviour may occur even in the community of trust. So, what we need is a model to capture the cases and deal with them – we are in need of a model of distrust.

We could distinguish misbehaviour of directly general scope (when an user directly attacks the main values which bonds the community together) to a misbehaviour which is primarily subjective in nature (a provocative message, a personal offence, an improper pushing). But because, as defined above, the trust is objective in the community of trust, the two kinds of misbehaviour do not require a conceptually different treating. Simply, if anyone perceive an interaction as unpleasant, harmful, distrustful or generally speaking “bad”, either as sole or in the context of previous interactions, he may express his dissent together with severity of the dissent. Should we just record the distrust expressions in the activity matrices? For example Guha et al. [15] discussed the distrust as a concept in a great detail. According to them, in many cases the distrust is just the other end of the trust continuum. But, in our model, the interactions which induce a distrust are exceptional, they do not belong to the common flow of interactions. So, in compliance with the slightly shifted interpretation of distrust, we decided to log distrust separately, as described below. All the single distrust expressions form *a record of distrust* in the model.

Let's have a *distrust function*  $v_d$  which assigns a finite non-negative real value to each distrust expression from the record of distrust. The value of distrust expressed by the user  $n$  to the user  $m$  for interaction of the user  $m$  to the user  $n$  performed at the time  $t$  is:

$$v_d(\text{distrust}_{mnt}) = b_{mnt} \quad (5)$$

where  $\text{distrust}_{mnt}$  belongs to the interaction  $i_{mnt}$  and respectively  $b_{mnt}$  belongs to  $i_{mnt}$ . Our current aim is to define a generic model, so we do not provide any particular function definitions. But, to reflect the fact that distrust is of higher impact than harmless activity and the fact that trust is being perceived as a virtue which falls sharply, distrust values returned by  $v_d$  function are to be much higher than activity values returned by  $v_i$  function to make the calculations introduced later meaningful.

If there is no distrust expressed for an interaction, which is true for most cases, the assigned distrust value is 0. Analogous to the activity, to get *a distrust of user  $m$  towards user  $n$  till the time  $t$* , distrust may be summed, using a discount function:

$$\begin{aligned} &\text{for all defined distrust expressions where } j \leq t \\ b_{mnt} &= \sum d_j i_{mnj} \end{aligned} \quad (6)$$

All distrust within the community till the time  $t$  constitutes the square *matrix of distrust*, again of size reflecting count of users in a system:

$$\begin{aligned} &\text{for } i, j \in 0..s \\ B_t &= (b_{ijt})_{s,s} \end{aligned} \quad (7)$$

Row sums represent the distrust expressed *to* a certain user for his bad interactions till a certain time (e.g.  $b_{m^*t}$ ) and column sums represent the distrust expressed *by* a certain user till a certain time (e.g.  $b_{*nt}$ ). For each  $t$  the distrust matrix captures all distrust expressed till the time  $t$ , thus the line of the matrices  $B_0, B_1, \dots, B_{\text{now}}$  reflects the whole dynamics of distrust within the system.

## 5. Derived Characteristics

With the activity matrix A and the distrust matrix B we defined the main components of the model. The matrices allow us to calculate various ratios. *Harmful/harmless ratio of the user m* sums harmfulness of the user weighed by his other correct behaviour:

$$r1_{m^*t} = b_{m^*t} / a_{m^*t} \quad (8)$$

The user who is new in the system and who started behaving badly, gets high r1 rating quickly, which signalizes that he really should not be trusted. Another user who acted correctly many times and fails first time may be still considered as trustful according to the ratio. The ratio may be used also in its subjective variant, *the harmful/harmless ratio of the user m to the user n*, but that is likely to be less usable because of the objective nature of the trust in the community of trust:

$$r1_{mnt} = b_{mnt} / a_{mnt} \quad (9)$$

*Distrusting/distrusted ratio of the user m* expresses how much the user distrusts to others and how significant is the amount of his distrust if we consider his own distrustful behaviour expressed by others as a weigh:

$$r2_{m^*t} = b_{m^*t} / b_{m^*t} \quad (10)$$

The ratio may be helpful to determine which users are objectively likely deceivers. Again, subjective variant of the same ratio, *distrusting/distrusted ratio of the user m to the user n*, may be considered:

$$r2_{mnt} = b_{mnt} / b_{mnt} \quad (11)$$

The subjective variant may be helpful e.g. to resolve minor disputes between particular users, but again, because we consider the trust as objective the potential for usage is limited. Besides the simple ratios the full record of activities and their interpretation in terms of trust and distrust allows us to calculate more sophisticated characteristics of individual users, combining their vital activity, incredulousness (as a tendency to consider a behaviour as distrustful), and their own trustfulness, as these develop in time. Our aim is to define these later, as soon as we have sufficient experimental results with the model.

## 6. Possible Refinements

The model as described above has been intentionally defined as simple and generic, as a basis for further evaluation, extensions and refinements. For example, a basic variant of the interaction value function  $v_i$  may simply assign certain value to each of the interaction classes, such as  $v_i(\text{"message"}) = 50$ ,  $v_i(\text{"chat post"}) = 2$  etc. Or the function may be refined to take further properties of interactions into account, e.g. the length, if the message is a reply on a previous message, or so.

As another example, the distrust value function  $v_d$  may simply return the distrust severity expressed by the reporting user as the distrust value. But, we may also refine the function to follow characteristics of users involved – the distruster and his distrustee. So, for example, an expressed distrust may be weighted e.g. by the harmful/harmless ratio of the distruster:

$$v_d'(distrust_{mnt}, r1_{m^*u}) = b_{mnt} \quad (13)$$

where u is closest lesser than t

The r1 ratio in the definition above does involve only previous events (those which happened already), because otherwise we would get a circular definition. Step further, we may take many other aspects of users' previous history into account, both mutual and overall:

$$v_d''(distrust_{mnt}, a_{m^*u}, a_{*nu}, b_{m^*u}, b_{*nu}, a_{mnu}, b_{mnu}, a_{nmnu}, b_{nmnu}) = b_{mnt} \quad (14)$$

The extended distrust value function allows to reflect statements such as:

- “A more active user has more experiences, so he is also more competent to evaluate behaviour of others.”
- “Claims of the user who received low or no distrust are more valid.”
- “If an user has been reported repeatedly for his bad behaviour it should be considered really seriously.”
- “If the abuse record has been preceded with a long interaction record between the involved users, it has more likely serious grounds.”

## 7. Future Work

The model introduced above needs both further concretization and evaluation. We are going to work on both tasks in parallel in our further intended research. First of all we would like to simulate the model on a virtual network test bed with artificial agents interacting with each other. All interactions will be logged into the activity matrices, all occurrences of distrust behaviour will be logged into the distrust matrices. Various tests will be performed on populations of different sizes and characteristics. The experiments will help to formulate particular shapes of the functions used in the model including their coefficients. Tests with bigger and more lively populations may help to check the scalability and reveal potential technical bottlenecks of the model. Further various implementation-related questions may be answered as well in this stage, such as how often to calculate the matrices and relevant ratios, how to store them, when an action should be triggered and so. Finally, the trust model will be turned into a trust architecture, applicable in a virtual social environment with real users.

As a next step, we would like to apply the trust architecture in a particular lively community of trust, supported by social networking system which we developed earlier. In a production environment users will have to be checked (authenticated, validated) prior entering the community. Even though the authentication mechanism is out of scope of our current and intended research, it can not be omitted. We already mentioned, that users should be provided with the means which reflect their pattern of thinking in the specific context. In the community of trust users do not wish to be

annoyed by requests to evaluate trust of each other – they trust each other as default. They just wish to use the service and have something at hand to react, to defend themselves and warn others, if matters go wrong. We assume that only minor changes in the user interface itself will be necessary, but various trust-related functions, such as a “report abuse” button, will be linked to the new trust architecture working behind the scenes.

As well as in the stage with artificial agents, we are going to store activity values related to every interaction and any distrustful behaviour reported by users in the respective matrices A and B. But, besides this, we would like to keep a record of related interactions and distrust expressions themselves, including narratives and other details. Doing so in a real scenario, we may identify and statistically examine various trends within the system, alongside various dimensions (interaction class, distrust expression severity, belonging of users involved to certain cluster or clique, key words used, and other). The trends in activity and distrust may prove as helpful, e.g. to reveal where to target actions to fight with any further distrustful behaviour or how to further refine the model. We are going to deal with a potentially sensitive data, so privacy and security questions will have to be considered as well. The phase of evaluation should finally conclude in a statistical examination of effectiveness of the model itself according to the users in comparison with different trust models.

## 8. Conclusions

In the paper we outlined state-of-the-art trust models for virtual communities, either on-line social networks or systems from the field of ambient intelligence. Besides multidimensionality, contextuality, asymmetry, transitivity, personalization, further characteristics of trust have been identified - the disproportion of impacts and trust dynamics. In the section we mentioned basic ideas of trust processing and inference in models with transitive trust, and levels of trust scope.

The community of trust is a concept to describe niche tightly coupled communities, where trust has an objective instead of subjective character and users trust each other as default. As a main contribution of the paper we outlined a simple model of distrust for communities of trust. Assuming objective trust with total transitivity in our model, the idea of tracking particular paths of trust among users, common in most other models, has been omitted fully. Our model captures activity and distrust dynamics in lines of activity and distrust matrices. Characteristics derived from the matrices provide a lens to keep watch on various aspects of the interaction dynamics in the system. Our aim was to design a generic model, simple and highly extensible. Further work, tests with artificial agents and an evaluation in a real community in particular, will follow.

## References

- [1] R. Mayer, J. Davis, and D. Schoorman, “An Integrative Model of Organizational Trust,” *The Academy of Management Review*, vol. 20, no. 3, pp. 734, 709, 1995.
- [2] J. Golbeck and J. Hendler, “Inferring binary trust relationships in Web-based social networks,” *ACM Trans. Internet Technol.*, vol. 6, no. 4, pp. 497-529, 2006.
- [3] F. Walter, S. Battiston, and F. Schweitzer, “A model of a trust-based recommendation system on a social network,” *Autonomous Agents and Multi-Agent Systems*, vol. 16, no. 1, pp. 57-74, nor. 2008.

- [4] D. Pavlovic, "Dynamics, Robustness and Fragility of Trust," in *Formal Aspects in Security and Trust: 5th International Workshop, FAST 2008 Malaga, Spain, October 9-10, 2008 Revised Selected Papers*, Springer-Verlag, 2009, pp. 97-113.
- [5] P. O. Wennerberg and T. Oellinger, "Ontology Based Modelling and Visualization of Social Networks for the Web: Discovering Security Related Information from Online News Sites," 2006.
- [6] G. Woodford, "Police see huge rise in Facebook 'incidents'," 2010. [Online]. Available: <http://www.thisisnottingham.co.uk/news/Police-huge-rise-Facebook-incidents/article-1949266-detail/article.html>. [Accessed: 01-May-2011].
- [7] C. Dwyer, S. Hiltz, and K. Passerini, "Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace," in *Proceedings of the Thirteenth Americas Conference on Information Systems*, Keystone, CO, USA, 2007.
- [8] P. D. Meo, A. Nocera, G. Quattrone, D. Rosaci, and D. Ursino, "Finding reliable users and social networks in a social internetworking system," in *Proceedings of the 2009 International Database Engineering & Applications Symposium*, Cetraro - Calabria, Italy, 2009, pp. 173-181.
- [9] D. Zejda, "Characteristics of Trust in Online Social Networks and Community of Trust as a Special Case of Online Community," *WEBIST 2011 Proceedings (in press)*, 2011.
- [10] J. Huang and M. S. Fox, "An ontology of trust: formal semantics and transitivity," in *Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet*, Fredericton, New Brunswick, Canada, 2006, pp. 259-270.
- [11] F. E. Walter, S. Battiston, and F. Schweitzer, "Personalised and dynamic trust in social networks," in *Proceedings of the third ACM conference on Recommender systems*, New York, New York, USA, 2009, pp. 197-204.
- [12] D. Hales and S. Arteconi, "Friends for Free: Self-Organizing Artificial Social Networks for Trust and Cooperation," 2005.
- [13] C.-N. Ziegler and G. Lausen, "Spreading Activation Models for Trust Propagation," in *Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04)*, 2004, pp. 83-97.
- [14] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The Eigentrust algorithm for reputation management in P2P networks," in *Proceedings of the 12th international conference on World Wide Web*, Budapest, Hungary, 2003, pp. 640-651.
- [15] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *Proceedings of the 13th international conference on World Wide Web*, New York, NY, USA, 2004, pp. 403-412.
- [16] M. Richardson, R. Agrawal, and P. Domingos, "Trust Management for the Semantic Web," in *The SemanticWeb - ISWC 2003*, 2003, pp. 351-368.
- [17] D. Zejda, "From Subjective Trust to Objective Trustworthiness in On-line Social Networks: Overview and Challenges," *Journal of Systems Integration*, vol. 2010, no. 1, pp. 16-22, 2010.
- [18] Z. Yan and S. Holtmanns, "Trust Modeling and Management: from Social Trust to Digital Trust," *book chapter of Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions*, 2007.
- [19] J. Caverlee, L. Liu, and S. Webb, "Towards robust trust establishment in web-based social networks with socialtrust," in *Proceeding of the 17th international conference on World Wide Web*, Beijing, China, 2008, pp. 1163-1164.
- [20] S. Moghaddam, M. Jamali, M. Ester, and J. Habibi, "FeedbackTrust: using feedback effects in trust-based recommendation systems," in *Proceedings of the third ACM conference on Recommender systems*, New York, New York, USA, 2009, pp. 269-272.
- [21] J. Golbeck, *Computing with Social Trust*, 1st ed. Springer, 2008.
- [22] Z. Yan and P. Cofta, "A Mechanism for Trust Sustainability Among Trusted Computing Platforms," in *Trust and Privacy in Digital Business*, 2004, pp. 11-19.
- [23] J. Carbó, J. M. Molina, and J. Dávila, "Fuzzy referral based cooperation in social networks of agents," *AI Communications*, vol. 18, no. 1, p. 1-13, 2005.