# On Agent Prioritization in Trust Networks *

**Jonathan Ben-Naim, Emil Weydert**

University of Luxembourg

**Abstract**

In multi-agent communities, the subjective attitudes of agents towards other agents are of central importance. If the agents report these attitudes, then an external observer may use this possibly partial social assessment to estimate in particular the trustability of the agents. We present a simple framework to specify procedures for deriving agent prioritizations from binary support relations among agents. We propose a number of rationality postulates, describe two specific approaches, and investigate to what extent they satisfy these principles.

## 1 Introduction

In multi-agent communities, the subjective attitudes of agents towards other agents are of central importance. If the agents report these attitudes, then an external observer may use this usually partial social assessment to estimate some corresponding objective qualities of the agents. In particular, one can exploit the positive feedback on agents to evaluate their actual reliability or trustability. This may be done by taking into account not only the immediate neighbours, but the whole backyard, if not the entire network, even including disconnected parts (in the most general approaches).

Several methods for estimating trust have been proposed in the literature. Most of them are numerical, for instance, the Bayesian reputation systems, e.g. [5, 8]. They take binary ratings (either positive or negative) as their input and derive from them a reputation score represented by a beta probability density function, possibly accompanied by a confidence parameter.

Another example is based on the belief model in [3, 4], which defines an opinion as a tuple $(b, d, u, a)$ where $b$ represents belief, $d$ disbelief, $u$ uncertainty, and $a$ the base rate probability in the absence of evidence. Then Subjective Logic [4] is used to derive trust from such opinions.

Some approaches, like [7], represent a trust network by a labelled directed graph whose nodes represent agents, and where a directed edge from $A$ to $B$ with label $d$ means that $A$ rates $B$ with $d$. Then the trust score that an agent $A$ attaches to an unknown agent $B$ is defined to be the weighted mean of the ratings given by those who had interactions with $B$.

One problem of these and similar numerical approaches is that they are based on several parameters, which may be hard, if not impossible, to identify correctly. Furthermore, a more cumbersome numerical perspective may also hide structural features and relationships, which could pave the way to a better understanding of the inferential processes. So there are a priori good reasons to investigate also more qualitative approaches.

A common strategy in this field has been to model the beliefs of the agents to derive trust relations between them, be it by using logical formalisms, like e.g. [2, 6], or quasi-probabilistic formalisms, like e.g. [3, 4]. However, we will take a slightly different direction and work within a framework which is – at least on the formal level – closer to that of Social Choice (see for instance [1], which provides some interesting impossibility theorems). There one typically has a collection of agents, each of them ranking a set of alternatives. The main question is how to aggregate these local, personal rankings into a global, social one.

In the present work, we take a look at the particular case where the alternatives consist of the agents themselves, and where the personal rankings are binary, encoding high, respectively low trust. If agent A attributes the high rank to agent B, this means that A supports B and is willing to provide positive feedback

*This is an updated version of the paper of the same title published in *the 19th Belgian-Dutch Conference on Artificial Intelligence*.

on B. On the other hand, if A attributes the low rank, this only means that A does not support B, but not that there is explicit distrust, or actual negative feedback.

The paper is structured as follows. After introducing our basic framework, as well as our support tree concept, we formulate some basic requirements for preference relations over these trees. We are then going to describe two specific methods for deriving agent prioritizations from binary support relations. To conclude, we investigate whether they verify our postulates.

# 2 Basic framework

We start with a very simple model consisting of a set of agents $\mathcal{A}$ and a binary support relation $\triangleright$ on $\mathcal{A}$. Intuitively, $\triangleright$ is meant to represent the explicit, visible, or reported support links among agents, with $A \triangleright B$ telling us that $A$ supports $B$. The question we are going address is the following: how can we use the support information encoded in the relation to estimate and compare the trustability of arbitrary agents. First some notation.

**Definition 1 (Support structure)** *A* support structure *is an ordered pair* $(\mathcal{A}, \triangleright)$*, where* $\mathcal{A}$ *is a non-empty finite set, and* $\triangleright$ *a binary relation on* $\mathcal{A}$*. The elements of* $\mathcal{A}$ *are called agents, and* $\triangleright$ *is called the support relation.*

**Definition 2 (Support chain)** *Given a* support structure $(\mathcal{A}, \triangleright)$*, an agent sequence* $(A_0, \ldots, A_n)$ *with* $A_i \triangleright A_{i+1}$ *for* $i \leq n-1$ *is called a* support chain *of length* $n$ *from* $A_0$ *to* $A_n$ *over* $(\mathcal{A}, \triangleright)$*. We say that* $A_n$ *receives (non-exclusively)* $n$*th-degree support from* $A_0$*.*

Our basic observation is that agents appear to be more trustable if they get support, not only from more agents, but also from agents which are themselves backed by other agents. In other words, more and/or longer support chains suggest more trustability. However, the impact of length and number can be limited by admissibility criteria. Consider for instance a support chain $(A_0, \ldots, A_n)$ which starts and finishes with the same agent $A_0 = A_n$. Of course, if $n = 1$, we tend to reject the chain as an instance of ruthless self-promotion. If $n > 1$, the situation is less obvious. Intuitively, it seems that while $(A_0, \ldots, A_n)$ may still provide slightly more support than $(A_1, \ldots, A_n)$, because of the circular dependency, it appears to be considerably less supportive than $(A'_0, A_1, \ldots, A_n)$, for $A'_0 \neq A_0$. The actual contribution of circular support, from none to substantial, depends here on the exact interpretation of the support relation.

Nevertheless, for the basic approach, we are going to take a strict stance and ignore any – direct and indirect, i.e. higher degree – support from an agent to itself in the context of its support evaluation. That is, we consider a support chain from $A_0$ to $A_n$ to be inadmissible if $A_0 = A_n$. Note however that this does not mean that we completely disregard the link from $A_0$ to $A_1$. It may still have an impact on the support for other agents.

Furthermore, the $n$th-degree support represented by the full chain may be seen as the combination of the $i$th-degree support for $A_i$ encoded by $(A_0, \ldots, A_i)$ and the $(n-i)$th-degree support for $A_n$ modeled by $(A_i, \ldots, A_n)$. So, if $(A_0, \ldots, A_n)$ is an admissible support chain, we would expect this to be true for $(A_0, \ldots, A_i)$ and $(A_i, \ldots, A_n)$ as well. It follows from this downwards closure condition that admissible support chains are not allowed to include an agent twice (at least in our basic framework). From now on, when we are going to refer to support chains, we will always mean admissible ones.

**Definition 3 (Admissibility)** *Let* $(\mathcal{A}, \triangleright)$ *be a support structure. A support chain* $(A_0, \ldots, A_n)$ *is said to be* admissible *iff for all* $i \neq j$*, we have* $A_i \neq A_j$*.*

Our strategy is to evaluate and compare the trustability of agents based on the admissible support chains over $(\mathcal{A}, \triangleright)$ targeting them. Since the collection of support chains in favour of an agent $A$ is closed under final segments, we can describe it in a natural and concise way by a tree, which we denote $\mathcal{T}^A$. Its nodes are the admissible support sequences with endpoint $A$, and its root is obviously $(A)$. The sons of $(A)$ are the admissible support chains of the form $(A_1, A)$, its grandsons those of the form $(A_2, A_1, A)$, and so on. More formally:

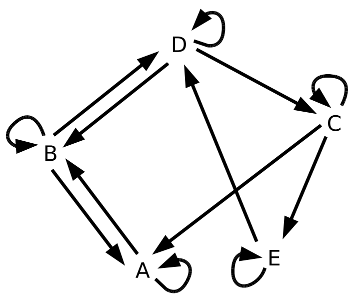
**Definition 4 (Support tree)** *Let* $(\mathcal{A}, \triangleright)$ *be a support structure and* $A \in \mathcal{A}$*. The support tree of* $A$ *over* $(\mathcal{A}, \triangleright)$*, denoted by* $\mathcal{T}^A$*, is defined as* $\mathcal{T}^A = (N_A, R)$*, where*

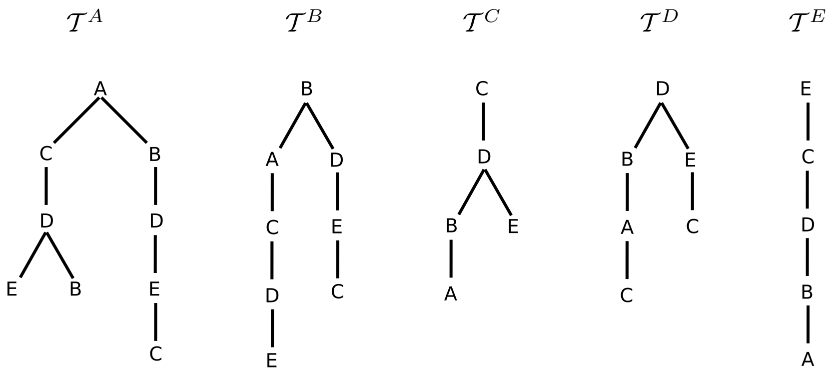
- $N_A$ *(the set of nodes) consists of the admissible support chains* $(A_0, \ldots, A_n)$ *over* $(\mathcal{A}, \triangleright)$ *with* $A_n = A$*,*

- $R$ *(the edge/parent relation) is given by* $(A_n, \ldots, A_0)R(A'_m, \ldots, A'_0)$ *iff* $m = n+1$*,* $A_i = A'_i$ *for* $i \leq n$*, and* $A_0 = A'_0 = A$*.*

If we label each node $(A_n, \ldots, A_0)$ by its initial element $A_n$, we see that it encodes the sequence of labels decorating the path connecting it to the root. Consequently, any support chain in favour of an agent $A$ $(= A_0)$ corresponds to a certain path in $\mathcal{T}^A$ from a node to the root, and vice versa. In other words, the support tree for an agent $A$, with the above natural labeling, allows us to represent in a consise manner all the support chains in favour of $A$.

Here is an example with five agents linked by different support or trust relationships. First we give the initial support structure, and then we extract the resulting support trees for each agent. Note that in the tree representation, support is upwards directed.




The next task will now be to see how we can use these support trees to prioritize agents in the context of a support structure.

# 3 Support tree comparison

We are now going to consider methods to estimate and compare the support which agents receive in the context of a given support structure. More specifically, we intend to achieve this through the evaluation of the support trees associated with agents. Their simple structure should allow us to get a better understanding of the basic prioritization issues, and also to pave the way for more fine-grained accounts, e.g. with numerical tags or processing circular support, which are beyond the scope of this paper.

Let $\vec{\mathcal{T}} = (\mathcal{T}^{\mathcal{A}} | A \in \mathcal{A})$ be the family of support trees derived from a given support structure $(\mathcal{A}, \triangleright)$. The main step is to determine a suitable pre-order $\preceq_{\vec{\mathcal{T}}}$ on support trees which reflects the overall amount of support for their respective root agents, and thereby provides a prioritization of the agents in $\mathcal{A}$. We observe that the agent labeling encodes dependency information which we may want to exploit. For instance, if an agent gets support from two other agents which are themselves supported, either by the same, or two different agents, its overall support seems to be lower in the former scenario. Within our present framework, keeping the evaluation elementary and tree-oriented, we ignore all the other, intrinsic properties of the labeling agents (except their identity). Note that the sub-tree structures below nodes sharing the same agent label may well diverge.

So the task is to find – in accordance with the support interpretation – intuitively appealing preference pre-orders $\preceq$ over finite rooted trees whose nodes are labelled by agents $A \in \mathcal{A}$. But before we are going to propose specific approaches, it may be helpful to formulate some general principles describing or guiding support tree comparison. We see several fundamental requirements for $\preceq$.

First, as we pointed out before, we restrict ourselves to exploit the structure of the support relation, but not any further characteristics of the labeling agents. Therefore, we should expect the preference relation to be invariant under renaming, i.e. a one-to-one exchange of labels. Accordingly, we propose the following structurality principle. For a function $f$ on labels, and a finite rooted labelled tree $\mathcal{T}$, let $\mathcal{T}_f$ be the tree obtained by replacing each label $l$ by its image $f(l)$.

**Invariance (I).** Let $\mathcal{T}$ and $\mathcal{T}'$ be two finite rooted labelled trees, and $\pi, \pi'$ be bijections on the full set of labels. Then $\mathcal{T} \preceq \mathcal{T}'$ implies $\mathcal{T}_\pi \preceq \mathcal{T}'_{\pi'}$.

Thus, an exchange of labels does not affect the preference of trees. Of course, if we want to exploit the fact that the trees to be compared share certain labels, we have to consider a more permissive – because invariant under a smaller set of permutation pairs – version.

**Weak Invariance (WI).** Let $\mathcal{T}$ and $\mathcal{T}'$ be two finite rooted labelled trees, and $\pi$ be a bijection on the full set of labels. Then $\mathcal{T} \preceq \mathcal{T}'$ implies $\mathcal{T}_\pi \preceq \mathcal{T}'_\pi$.

Of course, these conditions mainly reflect our structural perpective, and much less the nature of support. Here one basic intuition would be that if an agent is directly supported by a bunch of other agents, and some of these get more support, then the overall support for the initial agent should increase as well, or at least stay the same. Let $\mathcal{T}^A[\mathcal{T}^1, \ldots, \mathcal{T}^m]$ be the combined tree obtained by connecting the roots of the $m$ rooted labelled trees $\mathcal{T}^j$ to a new superroot labelled by a new agent $A$.

**Structural Monotony (SM).** If $\mathcal{T}_1^j \preceq \mathcal{T}_2^j$ holds for each $1 \leq j \leq m$, and the $\mathcal{T}_1^j$ are distinct iff the $\mathcal{T}_2^j$ are, then $\mathcal{T}^A[\mathcal{T}_1^1, \ldots, \mathcal{T}_1^m] \preceq \mathcal{T}^A[\mathcal{T}_2^1, \ldots, \mathcal{T}_2^m]$.

As usual, we define $\mathcal{T}_1 \prec \mathcal{T}_2$ iff $\mathcal{T}_1 \preceq \mathcal{T}_2$ and not $\mathcal{T}_2 \preceq \mathcal{T}_1$.

**Strict Structural Monotony (SSM).** Suppose SM holds. If $\mathcal{T}_1^j \preceq \mathcal{T}_2^j$ holds for each $1 \leq j \leq m$, and the $\mathcal{T}_1^j$ are distinct iff the $\mathcal{T}_2^j$ are, and $\mathcal{T}_1^j \prec \mathcal{T}_2^j$ for at least one $j$, then $\mathcal{T}^A[\mathcal{T}_1^1, \ldots, \mathcal{T}_1^m] \prec \mathcal{T}^A[\mathcal{T}_2^1, \ldots, \mathcal{T}_2^m]$.

In some sense, SSM implements a more additive philosophy for support. To take off, these requirements need however another core principle which roughly states that support is preferable to non-support.

**Non-triviality (NT).** Let $\mathcal{T}^B$ be a labelled tree consisting of a single node. If $A \neq B$, then $\mathcal{T}^A[\mathcal{T}^B] \prec \mathcal{T}^B$.

In a first step, we would propose the following – non-exhaustive – list of basic rationality postulates for $\preceq$:

- WI, SSM (hence also SM), and NT.

# 4 Comparison strategies

We are now going to present two simple methods for evaluating the relative trustability of agents by comparing their support trees in the context of a given support structure $(\mathcal{A}, \triangleright)$. In the present paper, this amounts to specify preference relations over finite labelled rooted trees which are compatible with our intuitions about support/trust. We emphasize that these proposals are mainly meant to illustrate our conceptual framework without too much technical luggage. Both procedures are based on lexicographic and cardinality considerations. This has the advantage that we may not have to construct the entire trees to compare them.

**1. Agent Strategy**

The first approach compares the number of agents supporting (directly/indirectly) the root agent, but focusing on the lowest-degree support, which intuitively appears to be the most reliable (proximity to the object of judgment), and therefore the most important one. For two agents $A, B \in \mathcal{A}$, let $deg(A, B)$ be the length of the smallest support chain from $B$ to $A$, if there is one. If there is none, we set $deg(A, B) = \infty$ if $A \neq B$, and $deg(A, B) = 0$ if $A = B$. Let $S_A^i = \{B \in \mathcal{A} \mid deg(A, B) = i\}$ be the set of agents which provide

$i$th-degree – but not $j$th degree for $j < i$ – support to $A$. The inductive construction of this layering is unproblematic. Obviously, the definition applies to arbitrary labelled trees.

The idea is now to say that the contribution of a single agent providing $i$th-degree support to the trustability of a root agent outweighs the contributions of an arbitrary number of agents providing $j$th-degree support, for $j > i$, to the trustability of any root agent. This gives us a dominance relation across support trees. Consequently, according to this view, $A$ is more trustable than $B$ if and only if there exists an $i$ such that $S^i_A$ has more elements than $S^i_B$, and for each $j < i$, $S^j_A$ has the same cardinality as $S^j_B$. Here is a formal definition based on support trees:

**Definition 5 (Agent strategy)** *Let $\mathcal{T}^A, \mathcal{T}^B$ be two rooted labelled trees. We say that $\mathcal{T}^A$ is preferred to $\mathcal{T}^B$ w.r.t. the* agent-strategy*, in symbols $\mathcal{T}^A \prec_a \mathcal{T}^B$, iff there is an $i$ with $|S^i_A| < |S^i_B|$, and for all $j < i$, $|S^j_A| = |S^j_B|$. We set $\preceq_a = \prec_a \cup =$.*

**Definition 6 (Agent-trustability)** *Let $(\mathcal{A}, \triangleright)$ be a support structure, $A, B \in \mathcal{A}$. We say that $A$ is* more trustable *than $B$ w.r.t. the* agent-strategy*, in symbols $A <_a B$, iff $\mathcal{T}^A \prec_a \mathcal{T}^B$.*

It is easy to see that $<_a$ provides a stratification of the agents reflecting their trustability. For instance, in the previous example there are five layers: first $B$, second $D$, third $A$, fourth $C$, and fifth $E$.

**2. Chain strategy**

Another possible strategy is to consider each support chain in favour of an agent in itself as a contribution to the agent's trustability. In other words, instead of counting the number of agents supporting $A$ or $B$, we may count the number of chains targeting $A$ or $B$. The idea here is similar to the previous one insofar as it says that the contribution of the starting point of a single chain of length $i$ supporting $A$ outweighs the contributions of the starting points of an arbitrary number of chains of length $j > i$ supporting $B$. More formally:

**Definition 7 (Chain strategy)** *Let $\mathcal{T}^A, \mathcal{T}^B$ be two rooted labelled trees. We say that $\mathcal{T}^A$ is preferred to $\mathcal{T}^B$ w.r.t. the* chain strategy*, in symbols $\mathcal{T}^A \prec_c \mathcal{T}^B$, iff there is an $i$ such that the number of nodes at depth $i$ in $\mathcal{T}^A$ is larger than that in $\mathcal{T}^B$, and for each $j < i$, there is an equal number of nodes at depth $j$ in $\mathcal{T}^A$ and $\mathcal{T}^B$. We set $\preceq_c = \prec_c \cup =$.*

**Definition 8 (Chain trustability)** *Let $(\mathcal{A}, \triangleright)$ be a support structure, $A, B \in \mathcal{A}$. We say that $A$ is* more trustable *than $B$ w.r.t. the* chain strategy*, in symbols $A <_c B$, iff $\mathcal{T}^A \prec_c \mathcal{T}^B$.*

Again $<_c$ provides a stratification of the agents reflecting their trustability. In the above example, we get again five layers, but different ones: first $A$, second $B$, third $D$, fourth $C$, and fifth $E$

To conclude, we want to see to what extent these two elementary approaches verify the principles we formulated for the comparison of support trees.

**Proposition 9 (Principles)**

1. *$\preceq_a$ verifies NT, I, WI, but not SM, SSM.*
2. *$\preceq_c$ verifies NT, I, WI, SM, SSM.*

While both approaches have some technical and conceptual drawbacks, which may be expected given the straightforward definitions, the chain strategy is at least in line with our basic postulates.

As far as we can tell, there is no model in the literature producing exactly the same inferential results, in particular because of our handling of cycles. The exact relationship with other accounts, as well as the evaluation of competing proposals within our framework, will be addressed elsewhere.

# References


[1] A. Altman and M. Tennenholtz. On the Axiomatic Foundations of Ranking Systems. In *IJCAI'05*, 2005.

[2] C. Castelfranchi and R. Falcone. Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In *International Conference on MultiAgent Systems (ICMAS98)*, p.7279, 1998.

[3] A. Jøsang. Trust-Based Decision Making for Electronic Transactions. In *4th Nordic Workshop on Secure Computer Systems (NORDSEC'99)*, 1999.

[4] A. Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279-311, 2001.

[5] A. Jøsang and R. Ismail. The Beta Reputation System. In *15th Bled Electronic Commerce Conference*, 2002.

[6] Cj. Liau. Belief, information acquisition, and trust in multi-agent systems: a modal logic formulation. In *Artificial Intelligence*, 149(1):31-60, 2003.

[7] P. Massa, P. Avesani, and R. Tiella. A Trust-enhanced Recommender System application: Moleskiing. In *ACM SAC TRECK Track*, 2005.

[8] A. Withby, A. Jøsang, and J. Indulska. Filtering Out Unfair Ratings in Bayesian Reputation Systems. In *7th Int. Workshop on Trust in Agent Societies (at AAMAS'04)*, 2004.