Information Sciences 606 (2022) 742–762

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Whom to trust in a signed network? Optimal solution and two heuristic rules

Fanyuan Meng [a], Matúš Medo [b,a], Berno Buechel [c,*]

[a] Department of Physics, University of Fribourg, 1700 Fribourg, Switzerland
[b] Department of Radiation Oncology, Inselspital, Bern University Hospital and University of Bern, 3010 Bern, Switzerland
[c] Department of Economics, University of Fribourg, 1700 Fribourg, Switzerland

ARTICLE INFO

Article history:
Received 20 December 2021
Received in revised form 23 March 2022
Accepted 18 May 2022
Available online 23 May 2022

Keywords:
Signed graphs
Social networks
Trust
Structural balance
Learning

ABSTRACT

Signed networks capture both positive and negative relationships in a social network. We assume that the network nodes are of two different types and, in agreement with social balance theory, link signs correlate with node types. Given a signed network and information on the type of some source nodes, we study the problem of judging the type of a target node. We first formulate a globally optimal Bayes solution to this problem. As this optimal solution is too complex to be used by humans, we proceed by introducing a heuristic based on the shortest paths between the source nodes and the target node. We prove that this heuristic is weakly better than a previously introduced heuristic based on random walks, and that it coincides with the optimal rule for star-like networks. With simulations, we assess the accuracy of the three type-judging rules and find that our heuristic is particularly accurate for networks with short distances and when there are multiple source nodes. Our work contributes to the active field of inference in complex networks and puts emphasis on methods that could be applied by humans facing network data.

1. Introduction

The extensive study of social networks across several disciplines has revealed many patterns of how people (or organizations) are related to each other [16]. While in most studies links are assumed to be positive (*e.g.*, friendship), negative links (*e.g.*, hostility) also exist in many applications. Signed networks capture both positive and negative links and thus allow us to represent structures that could not be represented with a standard network. Relative to their importance, signed networks are considered as under-studied [15]. While multiple common complex network problems, such as node ranking and community detection, are studied also for signed networks, this is rarely the case for the problem of node classification [30].

In this paper, we build upon [22] which assumes that link signs are correlated with hidden node types (while link presence is independent of the signs) and introduces a model of an observer learning node types from link signs and information about the types of some source nodes. The source nodes represent the information that the observer has obtained in other ways, from prior direct experience, for example. Whilst [22] mostly focus on a local heuristic (the *random neighbor rule*) to infer the node types, we explicitly derive the globally optimal inference, which we call the *Bayesian rule*. This rule provides a natural benchmark for how much there is to learn. However, the rule has high computational costs, as it involves considering

* Corresponding author
E-mail address: berno.buechel@unifr.ch (B. Buechel).
URL: http://www.berno.info (B. Buechel).

https://doi.org/10.1016/j.ins.2022.05.072
0020-0255/

2^N states where N is the number of nodes in the network. It is thus unrealistic that a human observer would apply this rule when making a judgment. We propose a much simpler heuristic that is based on the shortest paths between source nodes and target nodes. This *shortest path rule* considers just one shortest path between a source node and a target node and uses the information of the link signs along this path to judge the target node. For multiple source nodes it aggregates the information as if the paths were non-overlapping. We derive the expected accuracy of the shortest path rule and compare it to the Bayesian rule and the random neighbor rule.

As the main theoretical result, we show that the three rules can be ordered in terms of expected accuracy: the random neighbor rule is weakly dominated by the shortest path rule, which is weakly dominated by the Bayesian rule (Proposition 1). Only in a small class of networks (that includes trees when there is one source node), the three rules lead to the same judgment, which means that the shortest path rule and the random neighbor rule are also optimal (Proposition 2). In a larger class of networks that can be described as a star-like structure with the target node at the center, the shortest path rule is still optimal, but the random neighbor rule is in general not (Claim 1). These results show how the two simple heuristics work and when they fail.

We use simulations to quantify the differences in accuracy between the three rules. We show that the crucial network statistic is the average distance and show how accuracy depends on the shortest path length. For large networks, the heuristics fail to produce accurate results, except for the shortest path rule if the number of source nodes grows with network size and the correlation between node types and link signs is high enough.

The remainder of the paper is organized as follows. After briefly relating it to the literature in Section 2, we introduce the model in Section 3. In Section 4 we present the theoretical results; in Section 5 the simulation results. We conclude with Section 6.

2. Related literature

This work directly builds on [22] where the problem of judging node types in a signed network has been considered for a special case of the set-up that we consider here. Besides this directly related work, there are several different strands of literature that our work is influenced by and speaks to.

Research on signed networks has always had an emphasis on structural balance [13,4,6] which asserts that triads (or longer loops) in signed networks tend to be *balanced*.[1] Investigations of structural balance in signed networks spans from data on relations among countries [14] to recent analyses of data from online social networks [28,19]. In [20], the author argues that common tests of balance theory fail to disentangle the observed signed structures from the probability of nodes to connect with each other, and proposes a new way to assess the level of structural balance. Structural balance is not only a static network property but it is also closely connected with the dynamics and formation of signed networks [31,8,32,26].

Our work is also related to the literature on node classification in a signed network [33,25,29]. With respect to these works, we not only provide a new algorithm that classifies nodes in a signed network into two types, but also solve for the optimal solution within the setting of our model. At the same time, there are also clear differences between our work and this literature: Instead of using a data-driven and computational approach, we study the theoretically optimal solution and compare it to simpler heuristics that could mimic cognition of humans. We show *why* classification of node types works in certain settings and fails in others. Our work thus also contributes to the literature on human learning, in particular on various adverse outcomes as a result of bounded rationality (see [1,11,12], for recent contributions), which differentiates it clearly from recent contributions in computer science [27,23] where computational complexity limitations are low or absent.

3. Model

We first provide a practical example for learning from link signs. We then define the basic set-up, introduce the inference rules, and illustrate their application re-using the first example.

3.1. A motivational example

Consider the following situation:

> You are a manager in an organization and have to assess whether to promote a certain candidate t that you do not know first-hand. Besides other data and means of assessment, you do know two employees s_1 and s_2 who work in the same unit as t. For instance, you think that s_1 is someone who cooperates well, while s_2 is rather not cooperative (say s_1 makes very constructive contributions to work teams, while s_2's contributions are rather impeding progress). Moreover, you know the relationship among employees, as illustrated in Fig. 1. Now, the fact that candidate t holds a positive relation to the cooperative employee s_1 might yield a positive signal about the cooperativeness of t. Moreover, this belief could be reinforced

[1] Basically, if a friend of a friend is my friend, this 3-cycle is balanced, but if a friend of a friend is my enemy, this 3-cycle is not balanced. More generally and more formally, a signed network is called *balanced* if for every cycle the number of negative links is even [4].

by t's negative relation to employee j, who is positively related to the non-cooperative employee s_2. Finally, the positive view might be somewhat challenged by the fact that candidate t holds a positive relation with employee i, who is, in turn, negatively related to the cooperative employee s_1. Overall, the way candidate t is related to s_1 and s_2 might be informative about his own cooperativeness, here yielding rather positive signals.

More generally, for agents (nodes) whose qualities the observer does not know first hand, the observer can still try to infer them from other nodes and the link signs between them in the network. This is particularly important for finding hidden qualities of agents like trustworthiness, endurance, cooperativeness, honesty, and compliance to norms, for example. There are several applications with such a characteristic: Networks among employees (as in the example); inter-firm networks (e.g., which other small businesses to trust); criminal networks (who collaborates with the mafia and who acts legally); friendship networks (e.g., is someone a suitable roommate); two political camps (e.g., tell revolutionists from loyalists); alliances between countries (see, e.g., [14]); networks of artists (e.g., whether a certain book/film/song probably fits one's taste).

To elaborate slightly more on a second application, let us consider a society consisting of "loyalists" and "revolutionists." Loyalists sympathize with the government's ideology whilst revolutionists would like to overthrow it. When recruiting for a position (in the state or in the revolution), the recruiter will be interested in the type of a candidate. Again, knowing about the positive and negative relations that the candidate maintains can be quite informative. Information about the network can come from observed friendship and enmity, but also from social media debates (such as in a recent study with Twitter data, [18]).[2]

3.2. Set-up

We extend here the model introduced in [22] that can be used to assign trust to nodes in an arbitrary signed network. A signed network is constructed as follows. Let (V,E) be a simple graph of $N := |V|$ vertices (or nodes) and with the set of edges (or links) E. The graph is assumed to be connected. For each node $i \in V$, the type $\theta_i \in \{-1,+1\}$ is drawn independently at random with equal probability such that $P[\theta_i = +1] = P[\theta_i = -1] = \frac{1}{2}$. A signed network can be represented by a symmetric matrix $\boldsymbol{R}$ with $R_{ij} \in \{-1,0,+1\}$, where $R_{ij} = 0$ if there is no link between i and j. For each link $(i,j) \in E$, there is either a positive sign $R_{ij} = +1$ or a negative sign $R_{ij} = -1$. The link signs are drawn independently and conditionally on node types as follows: between nodes of the same type (different types) the probability of a positive (negative) link sign is $r \in (0.5,1]$, i.e., for any link (i,j), $P\left[R_{ij}\theta_i\theta_j = +1\right] = r$. We refer to parameter r as *link reliability*. This is the fundamental assumption of our model: As $r > 0.5$, link signs are related to node types.[3]

Some nodes $S \subset V$ are pre-determined as source nodes. For each source node $i \in S$, the observer receives an independent signal $\sigma_i \in \{-1,+1\}$ about its type. Each signal correctly indicates the node type with probability $q \in (0.5,1]$, i.e., $P[\sigma_i\theta_i = +1] = q$. Parameter q is referred to as source *signal quality*. We collect all initial signals in a vector $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_N)$, where $\sigma_i \in \{-1,+1\}$ for $i \in S$, and $\sigma_i = 0$ for $i \notin S$. We assume $r,q > 0.5$ because the signed network contains no information when either r or q is 0.5. Table 1 summarizes the notation including some symbols that will be introduced later.

Given a signed network $\boldsymbol{R}$ (*e.g.*, the positive and negative links in Fig. 1) and signals $\boldsymbol{\sigma}$ (*e.g.*, the prior information about two employees in Fig. 1), an observer wants to judge the quality of some target node $t \in (V \setminus S)$. To this end, the observer might apply one of the following three rules.

3.3. Three inference rules

3.3.1. Optimal benchmark: the Bayesian rule (Bayes)

The Bayesian rule properly takes into account all available information and forms the corresponding posterior belief. Each state of the world is one combination of node types $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_N) \in \Theta = \{-1,+1\}^N$. As $\theta_i \in \{-1,+1\}$ is binary, there are 2^N states of the world. Since node types are independently drawn with equal probability, the prior probability of each state $\boldsymbol{\theta}$ is the same, $P[\boldsymbol{\theta}] = \left(\frac{1}{2}\right)^N$. The conditional probability that, given a true state $\boldsymbol{\theta}$, we observe signals $\boldsymbol{\sigma}$ and link signs $\boldsymbol{R}$ is

$$P[\boldsymbol{\sigma},\boldsymbol{R}|\boldsymbol{\theta}] = q^{z_1(\boldsymbol{\theta})}(1-q)^{z_2(\boldsymbol{\theta})}r^{z_3(\boldsymbol{\theta})}(1-r)^{z_4(\boldsymbol{\theta})},$$

where $z_1(\boldsymbol{\theta}) := |\{\, i \in S|\sigma_i\theta_i = +1 \,\}|$ is the number of source nodes for which the signal is correct; $z_2(\boldsymbol{\theta}) := |\{\, i \in S|\sigma_i\theta_i = -1 \,\}|$ is the number of source nodes for which the signal is wrong; $z_3(\boldsymbol{\theta}) := |\{\, (i,j) \in E|R_{ij}\theta_i\theta_j = +1 \,\}|$ is the number of links for which the link sign is according to the node types; and $z_4(\boldsymbol{\theta}) := |\{\, (i,j) \in E|R_{ij}\theta_i\theta_j = -1 \,\}|$ is the number of links for which the link sign is not according to the node types. Note that $z_1(\boldsymbol{\theta}) + z_2(\boldsymbol{\theta}) = N_S$, where $N_S := |S|$ is the number of source nodes; and $z_3(\boldsymbol{\theta}) + z_4(\boldsymbol{\theta}) = |E|$.

Bayes' rule gives us the conditional probability that, given observations $\boldsymbol{\sigma},\boldsymbol{R}$, the true state of the world is $\boldsymbol{\theta}$:

[2] These ties and debates need not at all be dedicated to questions of government and revolution; only their sign (positive or negative) is assumed to correlate with type, e.g having an online argument with someone who is known to be a loyalist indicates rather the revolutionist than the loyalist type.

[3] The classic theory of structural balance provides a foundation for this assumption [4].

Table 1
Summary of notation.

Symbol	Values	Meaning
r	$\in (0.5, 1]$	Link reliability
q	$\in (0.5, 1]$	Signal quality
R	$N \times N$-matrix	Signed relation network
R_{ij}	$\in \{-1, 0, +1\}$	Relation of nodes i and j
$\boldsymbol{\theta}$	$= (\theta_1, \ldots, \theta_N)$	Profile of node types
θ_i	$\in \{-1, +1\}$	Type of node i
$\boldsymbol{\sigma}$	$= (\sigma_1, \ldots, \sigma_N)$	Profile of signals
σ_i	$\in \{-1, +1\}$	Signal about node i's type
N	$\in \mathbb{N}; N < \infty$	Number of nodes
N_S	$\in \mathbb{N}; N_S < N$	Number of source nodes
$b(t)$	$\in [0, 1]$	Belief that node t is of positive type
$g(t)$	$\in \{-1, +1\}$	Best guess about node t's type
$A(t)$	$\in \{0, 1\}$	Accuracy
$E[A]$	$\in [0, 1]$	Expected accuracy

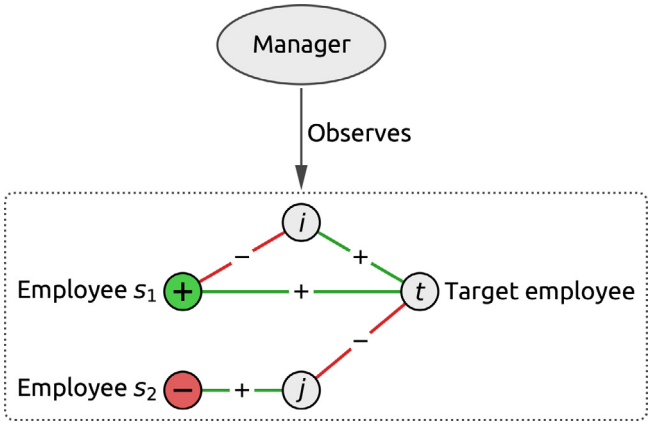

Fig. 1. An example: A manager who knows employees s_1 and s_2 attempts to learn about the quality of a target employee t from observing the positive and negative relations among employees.

$$P[\boldsymbol{\theta}|\boldsymbol{\sigma}, \boldsymbol{R}] = \frac{P[\boldsymbol{\sigma}, \boldsymbol{R}|\boldsymbol{\theta}] \cdot P[\boldsymbol{\theta}]}{P[\boldsymbol{\sigma}, \boldsymbol{R}]} = \frac{q^{z_1(\boldsymbol{\theta})}(1-q)^{z_2(\boldsymbol{\theta})} r^{z_3(\boldsymbol{\theta})}(1-r)^{z_4(\boldsymbol{\theta})} \cdot \left(\frac{1}{2}\right)^N}{\sum_{\boldsymbol{\theta}' \in \Theta} q^{z_1(\boldsymbol{\theta}')}(1-q)^{z_2(\boldsymbol{\theta}')} r^{z_3(\boldsymbol{\theta}')}(1-r)^{z_4(\boldsymbol{\theta}')} \cdot \left(\frac{1}{2}\right)^N}. \tag{1}$$

While the numerator uses the likelihood of a given state $\boldsymbol{\theta}$, the denominator uses all possible states, $\boldsymbol{\theta}' \in \Theta$, to generate the probability of the actual observation $\boldsymbol{\sigma}, \boldsymbol{R}$. Note that the terms $\left(\frac{1}{2}\right)^N$ cancel out.

Finally, the probability that target node t is of the positive type is the probability that we are in a state of the world with $\theta_t = +1$. Let $\Theta(t)$ be all 2^{N-1} states with $\theta_t = +1$. Then the posterior belief that target node t is of positive type, given observations $\boldsymbol{\sigma}, \boldsymbol{R}$, is

$$b^{Bayes}(t) := P[\theta_t = +1|\boldsymbol{\sigma}, \boldsymbol{R}] = \sum_{\boldsymbol{\theta} \in \Theta(t)} P[\boldsymbol{\theta}|\boldsymbol{\sigma}, \boldsymbol{R}], \tag{2}$$

where $P[\boldsymbol{\theta}|\boldsymbol{\sigma}, \boldsymbol{R}]$ is given by Eq. (1). From the posterior belief, we can also derive the expected node type: $E[\theta_t|\boldsymbol{\sigma}, \boldsymbol{R}] = P[\theta_t = +1|\boldsymbol{\sigma}, \boldsymbol{R}] \times 1 + (1 - P[\theta_t = +1|\boldsymbol{\sigma}, \boldsymbol{R}]) \times (-1)$.

For an observer (who knows which model has produced $\boldsymbol{\sigma}, \boldsymbol{R}$ and knows the values of q and r), the Bayesian rule is the optimal way to process the available information. The downside is that its application involves 2^N states which makes its direct use prohibitively computationally demanding and an unlikely candidate for mimicking human cognition. We apply the Bayesian rule theoretically and restrict simulations with the Bayesian rule to N up to 20.

3.3.2. A new heuristic: the shortest path rule (ShPath)

Instead of going through all complexities that the Bayesian rule necessitates, the observer might simply consider how the target node is connected to a given source node by a shortest path. We propose a new rule based on this idea. The *shortest path rule* is defined as follows:

1. Select one source node $s \in S$. Identify all shortest paths from the source node s to the target node t and choose one at random. Label nodes on the path "$s \to t$" by $n_0, n_1, \ldots, n_L$ where $n_0 := s, n_L := t$ and L is the path length.
2. Compute the *path orientation* produced by both the signal and link signs along the path

$$o_{s\to t} := \sigma_s \prod_{m=0}^{L-1} R_{n_m, n_{m+1}}. \tag{3}$$

3. Compute the probability that the shortest path has an even number of errors (i.e., links whose sign does not match the node types)[4]

$$\pi_{s\to t} := r^L + \binom{L}{2} r^{L-2}(1-r)^2 + \ldots = \frac{1}{2}\left[1 + (2r-1)^L\right]. \tag{4}$$

4. Determine the belief derived from this path as

$$b_{s\to t} := \begin{cases} q\pi_{s\to t} + (1-q)(1-\pi_{s\to t}) & \text{if } o_{s\to t} = +1, \\ 1 - q\pi_{s\to t} - (1-q)(1-\pi_{s\to t}) & \text{if } o_{s\to t} = -1. \end{cases} \tag{5}$$

We can interpret $b_{s\to t}$ as the subjective probability that node t is of positive type, as it can be derived from the path from s to t. Consider, for instance, the first case, where the path's orientation is positive, i.e., where $o_{s\to t} = +1$. There the former term, $q\pi_{s\to t}$, is the probability that the source node signal is correct *and* the number of erroneous links between s and t is even which together imply that $o_{s\to t} = \theta_t$. The latter term is the probability that the source node signal is incorrect *and* the number of erroneous links between s and t is odd, which again implies that $o_{s\to t} = \theta_t$. Hence, both terms together yield the probability that the path orientation, which is positive, coincides with the type of the target node t. Analogously, the second case yields the probability that the path orientation, which is negative, does not coincide with the type of the target node.

5. Repeat steps 1–4 for all source nodes S. Compute the belief (*i.e.*, subjective probability) that target node t is of positive type as

$$b^{ShPath}(t) := \frac{\prod_{s\in S} b_{s\to t}}{\prod_{s\in S} b_{s\to t} + \prod_{s\in S}(1 - b_{s\to t})}, \tag{6}$$

which aggregates beliefs $b_{s\to t}$ from each path, as if they were containing independent information. For one source node ($N_S = 1$), we have $b^{ShPath}(t) := b_{s\to t}$ as given by Eq. (5).

The application of this rule is actually quite simple. Suppose first that there is only one source node, i.e., $S = \{s\}$ and $N_S = 1$. Then the *path orientation* $o_{s\to t}$, which is computed in step 2, already indicates the best guess about the target node's type: A positive (negative) path orientation leads to the best guess that the target node is rather of the positive (negative) type. Steps 3 and 4 attach a probability to this guess. Eq. (5) (in step 4) accounts for the probability q that the source node's signal is accurate. Finally, in step 5 we obtain belief $b^{ShPath}(t) := b_{s\to t}$ when there is only one source node. For multiple source nodes the procedure is repeated for each source node. Eq. (6) in step 5 then aggregates information from all source nodes *as if* the shortest paths were independent.[5]

The shortest path rule is motivated by real agents who cannot be assumed to use the Bayesian rule. In particular, the shortest path rule is much simpler as it ignores information from other shortest paths (if multiple shortest paths exist between s and t), from other paths that are not shortest, and the fact that the multiple shortest paths might be overlapping. In spirit, this is similar to imperfections in belief formation and social learning. For instance, there are several well-documented biases when people learn about probabilities (e.g., [1]); including correlation neglect, i.e., assuming that signals are independent even if they are correlated [9]. Also non-Bayesian models of social learning (or opinion dynamics) are motivated by naïveté in updating, e.g., assuming that people do not sufficiently account for redundancy of information.[6]

3.3.3. A known heuristic: the random neighbor rule (RNeighbor)

In a recent paper, [22] introduced a closely-related framework and studied the *random neighbor rule*. Our model framework translates to theirs when we set link reliability $r \equiv 1 - \beta$ where β is their noise parameter and fix signal quality to be $q \equiv 1$. The other rule studied in [22], the Majority rule, is less founded than the random neighbor rule, so we omit it from our

[4] The closed-form expression is obtained by rewriting the sum as $[r + (1-r)]^L/2 + [r - (1-r)]^L/2$.

[5] This formula weighs the likelihood that all beliefs are correct against the likelihood that all are correct *or* all are wrong. For example, if one source node would lead to a belief of 0.8 and the other source node to a belief of 0.6, then the belief from both source nodes is $(0.8 * 0.6)/[0.8 * 0.6 + (1-0.8) * (1-0.6)] \approx 0.857$. Several simple examples that illustrate the application of all three rules are presented in Sections 3.4 and Appendix B.

[6] For instance, [7] justify the DeGroot model with "persuasion bias," a tendency to ignore that several pieces of information may be redundant, e.g., stemming from the same source.

analysis. In the random neighbor rule, the observer uses signals on the source node types to form opinions $x_i \in \{-1, +1\}$ for all other nodes. The opinion formation process is step-wise:

1. Judge every source node $s \in S$ according to its signal: $x_s = \sigma_s$.
2. Choose node j to judge at random from all nodes that have not been judged yet. If no neighbor of j has been judged yet, repeat the choice.
3. From the judged neighbors of node j, choose neighbor i at random. The judgment of node j is then $x_j = x_i R_{ij}$.
4. If target node t has not been judged yet, continue at step 2. If the target node has been judged, then we define the best guess about the target node's type as $g^{RNeighbor}(t) := x_t$.

There is a difference between the random neighbor rule and the two other rules in terms of the output. The *Bayesian rule* and the *shortest path rule* produce beliefs $b(t)$ that lie in the interval $[0, 1]$ and express the observer's (subjective) probability that the target node is of the positive type $\theta_t = +1$. The *random neighbor rule* makes a judgment x_t about the target node that is either $+1$ or -1 without attaching a level of confidence to it, i.e., $g^{RNeighbor}(t) \in \{-1, +1\}$. To allow for a fair comparison between the three rules, we map each continuous belief $b(t) \in [0, 1]$ into a "best guess" $g(t) \in \{-1, +1\}$. In particular, let $g(t) = +1$ for $b(t) > 0.5, g(t) = -1$ for $b(t) < 0.5$. In the knife-edge case $b(t) = 0.5$, the best guess is chosen from $\{-1, +1\}$ at random, with equal probability. Note that all three rules can be applied also for $t \in S$ (i.e., when the target node is also a source node).[7]

We stress again that the cognitive complexity (and the required amount of information) is substantially higher for the Bayesian rule than for the other two rules. By contrast, the random neighbor rule is the easiest of the three to apply as it allows the opinion on all nodes to be formed in $N - N_S$ elementary steps. The shortest path rule requires shortest paths to be found; using breadth-first search, the worst-case time complexity for one target node is $O(N + |E|)$ where $|E|$ is the number of links (positive or negative) in the relation network. Assuming a sparse network ($|E| \sim N$), the time complexity of the shortest path rule is thus $O\left(N^2\right)$ compared to $O(N)$ of the random neighbor rule.

3.4. Illustration of how the rules are applied

Before we systematically compare the three rules, let us first illustrate how they are applied in two simple examples. Both examples are subgraphs of the signed network among employees illustrated in Fig. 1.

Example 1. [A Triangle] Focus on the top part of Fig. 1, i.e., on the subgraph $V' = \{s_1, i, t\}$ with edges $E' = \{s_1 t, s_1 i, it\}$. The source node's signal is $\sigma_{s_1} = +1$ and the link signs are $R_{s_1 t} = R_{it} = +1$ and $R_{s_1 i} = -1$.

The Bayesian rule considers all $2^3 = 8$ states of the world. For instance, state $\boldsymbol{\theta} = (\theta_{s_1}, \theta_i, \theta_t) = (+1, +1, +1)$, in which all nodes are of positive type, would generate the observations $\boldsymbol{\sigma}, \boldsymbol{R}$, with probability $P[\boldsymbol{\sigma}, \boldsymbol{R}|\boldsymbol{\theta}] = qr^2(1 - r)$, as the signal for s_1 agrees with its type and two out of three links agree with their node types. Using Eq. (2), we obtain the Bayesian posterior belief that target node t is of positive type, given observations $\boldsymbol{\sigma}, \boldsymbol{R}$, and simplify it to

$$b^{Bayes}(t) = \frac{(1+q)r^2 + (1-q)(1-r)^2}{3r^2 + (1-r)^2}.$$

Next, let us apply the shortest path rule, again to the subgraph (V', E'). The only shortest path from source to target node is obviously s_1, t of length 1. The path's orientation is $o_{s_1 \to t} = \sigma_{s_1} R_{s_1,t} = (+1) \times (+1) = 1$, which means that the best guess is that the target node is rather of the positive type. The probability that the shortest path has an even number of errors is $\pi_{s_1 \to t} = r$ by Eq. (4). Hence, the belief derived from this path by Eq. (5) is $b_{s_1 \to t} = qr + (1 - q)(1 - r)$, which does not change by Eq. (6).

$$b^{ShPath}(t) = qr + (1 - q)(1 - r).$$

Finally, let us apply the random neighbor rule. In the first step, the judgment of source node s_1 is produced by its signal as $x_{s_1} = \sigma_{s_1} = +1$. Now, a next node is chosen from $\{i, t\}$ at random. If this node is t, we receive the judgment $x_t = x_{s_1} R_{s_1 t} = (+1) \times (+1) = +1$ and we have the best guess $g^{RNeighbor}(t) = +1$. Otherwise, i is judged next with $x_i = x_{s_1} R_{s_1 i} = (+1) \times (-1) = -1$. Then the next node must be t. It is selected at random whether the judgment on t comes from neighbor s_1 or from neighbor i (who are both already judged). In the former case we have $x_t = x_{s_1} = 1$; in the latter case we have $x_t = x_i = -1$. Overall, we have a 3/4 chance to reach $g^{RNeighbor}(t) = +1$ (when the judgment about t is derived from s_1 directly) and a 1/4 chance to reach $g^{RNeighbor}(t) = x_i = -1$ (when the judgment about t is derived from s_1 via i).

[7] With the Bayesian rule, for example, one can then assess if the signal is false for one source node, given that there are other source nodes.

Beliefs derived in Example 1 satisfy $0.5 < b^{Bayes}(t) < b^{ShPath}(t)$. Hence, we can observe that, in this example, the *Bayesian rule* and the *shortest path rule* lead to the same best guess that the target node is of the positive type ($g^{Bayes}(t) = g^{ShPath}(t) = +1$). The random neighbor rule agrees in most, but not all cases. Interestingly, the confidence level of the *Bayesian rule* is lower than that of the *shortest path rule*, i.e., $b^{Bayes}(t) < b^{ShPath}(t)$, since it takes into account that besides the direct shortest path with a positive orientation, there is a second path with a negative orientation. The next example shows how the rules are applied when there is a single longer path.

Example 2. [A Line] Focus now on the lower part of Fig. 1, i.e., on the subgraph $V'' = \{s_2, j, t\}$ with edges $E'' = \{s_2 j, jt\}$. The source node's signal is $\sigma_{s_2} = -1$ and link signs are $R_{s_2 j} = +1$ and $R_{jt} = -1$.

The Bayesian rule considers again 2^3 states of the world and yields

$$b^{Bayes}(t) = \Big[qr^2 + q(1-r)^2 + 2(1-q)r(1-r)\Big].$$

The shortest path from source to target node is s_2, j, t and has length 2. The path's orientation is positive as $o_{s_2 \to t} = \sigma_{s_2} R_{s_2 j} R_{jt} = (-1) \times (+1) \times (-1) = 1$. The probability that the shortest path has an even number of errors is $\pi_{s_2 \to t} = r^2 + (1-r)^2$ by Eq. (4). Hence, we receive the final belief

$$b^{ShPath}(t) = q\Big[r^2 + (1-r)^2\Big] + (1-q)\Big[1 - \Big(r^2 + (1-r)^2\Big)\Big],$$

and we find that $b^{Bayes}(t) = b^{ShPath}(t)$.

The random neighbor rule starts with setting $x_{s_2} = \sigma_{s_2} = -1$. Then it chooses a node from $\{j, t\}$ at random. If the node is t, it chooses again as the only neighbor of t, j, has not been judged yet. If the node is j, it makes the judgment $x_j = x_{s_2} R_{s_2 j} = (-1) \times (+1) = -1$. Finally, it judges node t as $x_t = x_j R_{jt} = (-1) \times (-1) = +1$. Hence, $g^{RNeighbor}(t) = +1$.

Computing the beliefs in Example 2 yields $0.5 < b^{Bayes}(t) = b^{ShPath}(t)$, while $g^{Neighbor}(t) = +1$. We see that: (1) All rules lead to the same best guess that the target node t is rather of the positive type: $g^{Bayes}(t) = g^{ShPath}(t) = g^{RNeighbor}(t)$; (2) the shortest path rule leads to exactly the same belief as the Bayesian rule and is hence optimal. Both observations are not coincidences, as we prove in Claims 1 and 2 in the next subsection.

4. Theoretical results

4.1. Comparison of the three rules

The three seemingly very different rules are in fact strongly related to each other.

4.1.1. Comparison of shortest path rule with Bayesian rule

The example of a line graph above (Example 2) already indicates that sometimes the shortest path rule leads to the same belief as the Bayesian rule. The first proposition provides a sufficient condition for this conclusion. Recall that a path is defined as a sequence of distinct nodes such that each neighboring pair in the sequence is linked by an edge.

Claim 1 (*Equivalence*). Suppose that in (V, E) between each source node $s \in S$ and the target node t there is a *unique* path that is *non-overlapping* with the paths of the other source nodes. Then the Bayesian rule and the shortest path rule are equivalent: for any observation $\boldsymbol{\sigma}, \boldsymbol{R}$,

$$b^{Bayes}(t) = b^{ShPath}(t).$$

The condition of the result (*unique* path that is *non-overlapping*) is equivalent to the following: Between every pair of source nodes, there is one single path and the target node t lies on this path. Hence, these networks have a star-like structure, where target node t is the center and each source node s is on a unique ray (of arbitrary length), plus potentially subgraphs that are "pockets", as illustrated in Fig. 2.

Conversely, if these conditions (*unique* path that is *non-overlapping*) are not met, the shortest path rule may be suboptimal. We demonstrate this by three examples which are illustrated in Fig. 3, and solved in Appendix B. For concreteness, we suppose in all examples that the source node signals are positive. In Example B.1, illustrated in Panel (A) of Fig. 3, there is a unique shortest path between source node and target node, but there are additional (non-shortest) paths. The shortest path rule only uses the direct path, which has a positive orientation. However, the Bayesian rule uses all information including the two paths of length 2, which are both negatively oriented. Hence for r not too small, the best guess of the Bayesian rule is negative, while the shortest path rule comes to the opposite conclusion.

In Example B.2, illustrated in Panel (B) of Fig. 3, there are multiple shortest paths between source node and target node. The Bayesian rule always leads to a belief $b^{Bayes}(t) > 0.5$ and corresponding best guess of $+1$, as two out of three paths are

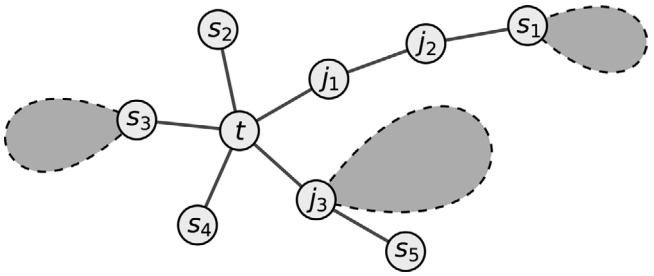

Fig. 2. A star-like network with five source nodes ($s_1, \ldots, s_5$) and subgraphs attached to nodes s_1, s_3, and j_3. As the path between each source node and target node t is unique and does not overlap with other paths, the condition of Claim 1 holds. Hence, the shortest path rule and the Bayesian rule yield identical beliefs about t (for any $\boldsymbol{\sigma}$ and R).

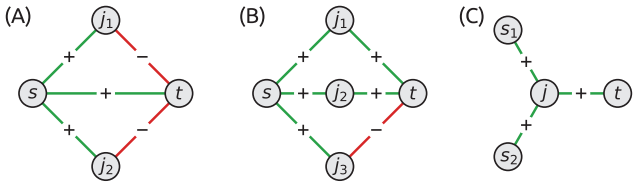

Fig. 3. Illustration of signed networks in Examples B.1,B.2,B.3 (see Appendix B for the solutions). These networks do not satisfy the condition of Claim 1, the shortest path rule hence leads to suboptimal beliefs.

positively oriented. The shortest path rule selects one of the three shortest paths at random. Hence, with probability 2/3 it selects a positively oriented path and induces a best guess of $+1$, but with probability 1/3 it induces a best guess of -1.

Finally, in Example B.3, illustrated in Panel (C) of Fig. 3, there are two source nodes whose shortest paths to the target node overlap. For $r, q \in (0.5, 1)$, we receive $b^{ShPath}(t) > b^{Bayes}(t)$. The suboptimality of the shortest path heuristic in this example enters in step 5 of the procedure when multiple paths are aggregated. The shortest path rule aggregates information of both paths as if they were independent, ignoring that they are correlated by having link jt in common. This mistake is very similar in spirit to persuasion bias in the DeGroot model [7] and to correlation neglect [9]. We observe that ignoring this overlap leads to overconfidence, as the subjective probability is higher than the objective probability: $b^{ShPath}(t) > b^{Bayes}(t) > 0.5$.

As we can see in these three examples, the reason for potential suboptimality of the shortest path rule is that it aggregates information from the signed network *as if* all paths from any source node to the target node were unique and independent.

4.1.2. Comparison of shortest path rule with random neighbor rule

The line graph discussed as Example 2 already indicates that the random neighbor rule can sometimes lead to the same best guess as the shortest path rule. The next proposition provides a sufficient condition for this conclusion.

Claim 2 (*Same Best Guess*). Suppose that in (V, E) there is a unique path between the only source node s and the target node t. Then for any observation $\boldsymbol{\sigma}, \boldsymbol{R}$, we have $g^{ShPath}(t) = g^{RNeighbor}(t)$.

For the interpretation of Claim 2 it is helpful to re-interpret the random neighbor rule in terms of paths. When we apply the random neighbor rule using source node s, the final judgment of node t is ultimately based on a randomized path between the source node and the target node. This randomized path's orientation $o_{s\to t}$ determines the best guess, similarly to the shortest path rule, where the orientation of the shortest path determines the best guess. Hence, when there is a unique path between s and t, then this path is also the shortest path and the two rules lead to the same best guess.[8]

Conversely, when there are multiple paths between the single source node s and the target node t; or when there is more than one source node, the random neighbor rule does not always lead to the same best guess as the shortest path rule. For instance, in Example 1, illustrated in the upper part of Fig. 1, there are multiple paths and the random neighbor rule may lead to a best guess of -1, while the shortest path rule leads to $+1$. In Example B.3, illustrated in Panel (B) of Fig. 3, there are multiple paths of the same length. It might happen that we have a negative signal for source node s_1 and a positive signal for s_2 and that the random neighbor rule bases the judgment on the path from s_1, while the shortest path rule randomly selects the other shortest path.[9]

[8] Moreover, since the conditions of Claim 2 are stronger than those of Claim 1, it actually follows that all three rules lead to the same best guess under the conditions of Claim 2, i.e., in the special case that there is only one source node and a unique path between source node and target node. This insight will lead to Proposition 2 below.

[9] However, in this network the two rules would lead to the same quality of judgment in expectations – a topic that we address in the next subsection.

4.2. Comparison of accuracy

We now theoretically assess to which extent the three rules deliver a correct judgment.

Definition 1 (*Accuracy*). Let $g(t)$ be the observer's best guess of type of node t obtained using a certain rule, while true node types are $\boldsymbol{\theta}$.[10] The rule's accuracy for target node t is defined to be 1 for a correct judgment and 0 for a wrong judgment:

$$A(t) = \begin{cases} 1 & \text{if } g(t)\theta_t = +1, \\ 0 & \text{if } g(t)\theta_t = -1. \end{cases}$$

This definition allows us to assess the obtained opinion for any given realization of node types, source signals, and link signs. However, such a realization is not necessarily representative. To measure the performance of an inference rule, we thus assess its expected accuracy.

Definition 2 (*Expected Accuracy*). Fix a network (V, E) with source nodes S and target node t and also fix the model parameters q and r. A rule's *expected accuracy*, $E[A]$, is the expected value of accuracy over all different realizations (states $\boldsymbol{\theta}$, signals $\boldsymbol{\sigma}$ and link signs R) and all different judgment outcomes.[11]

Expected accuracy can be interpreted as the ex ante probability that the final judgment of the target node will be correct. If our belief $b(t)$ correctly reflects the probability that target node t is of the positive type, then $E[A(t)] = b(t)$ if $b(t) \geqslant 0.5$ and $E[A(t)] = 1 - b(t)$ if $b(t) < 0.5$.

While expected accuracy $E[A]$ can in principle lie anywhere in the range $[0, 1]$, already with random guessing of target node types, we achieve an expected accuracy of $E[A] = 0.5$.[12] Hence, 0.5 is a lower bound on the performance of a sensible rule. At the other extreme, expected accuracy of any rule is bounded from above by the initial information given through the source nodes. For the case of one source node, $N_S = 1$, we have in particular that $E[A] \leqslant q$ for any rule; the equality is achievable when $r = 1$.

The newly defined expected accuracy allows us to understand how is the performance of the shortest path rule affected by positions of source and target nodes in the network.

Claim 3 (*Shortest Path Accuracy*). Let $N_S = 1$ and suppose that the shortest path between source node s and target node t has length L. The expected accuracy of the shortest path rule is then

$$E\left[A^{ShPath}\right] = \frac{1}{2} + \left(q - \frac{1}{2}\right)(2r-1)^L. \tag{7}$$

As $o_{s\to t}$ is the best guess of θ_t for the shortest path rule, the expected accuracy introduced by Definition 2 is thus equal to $P[o_{s\to t} = \theta_t]$. Claim 3 then follows by combining with Eqs. (4) and (5). The obtained expected accuracy depends on the model parameters, q and r, and the shortest path length, L. When $r = 1$ (fully reliable links), $E\left[A^{ShPath}\right] = q$ for any L. When $r < 1, E\left[A^{ShPath}\right]$ decreases as L grows.[13] Note that Eq. (7) does not hold exclusively for a shortest path – it can be used to compute the expected accuracy when a path of an arbitrary length L is used by any of the three rules.

We are now in the position to state the main result. It establishes how the three rules are ordered in terms of expected accuracy.

Proposition 1 (*Ordering*). *For every given network (V, E), set of source nodes S and target node t, the accuracy expectations of the three rules are ordered as*

$$E\left[A^{Bayes}\right] \geqslant E\left[A^{ShPath}\right] \geqslant E\left[A^{RNeighbor}\right].$$

The first inequality is a simple implication of the fact that the Bayesian rule computes the correct posterior belief for any realization of source node signals $\boldsymbol{\sigma}$, and link signs R. When the observer uses this rule, then her belief (i.e., subjective probability) equals exactly the objective probability that the target node t is of positive type. Hence, the expected value of accuracy after realization of observations $\boldsymbol{\sigma}, R$ (we might call this "ad interim expected accuracy") is maximal. Since this holds for every realization of observations, also the ex ante expected value of accuracy, that is expected accuracy $E\left[A^{Bayes}\right]$, is maximal.

[10] Recall that for any belief $b(t) > 0.5$ the best guess is $+1$, for any belief $b(t) < 0.5$ the best guess is -1, and for $b(t) = 0.5$ the best guess is chosen from $\{-1, +1\}$ with equal probability.

[11] The probabilistic element in the shortest path rule is which shortest path to use if there are multiple. The probabilistic element in the random neighbor rule is the order of judgment. The Bayesian rule is deterministic.

[12] In the work of [22], a closely-related metric, opinion consistency, is used to assess the opinions formed on several nodes. Since a simple linear relation exists between accuracy $A(t)$ and node t's contribution to opinion consistency, $C(t) = 2A(t) - 1$, the results are qualitatively the same for both of them.

[13] When $q = 0.5$ (the source signals are not informative) or $r = 0.5$ (the links are not informative), $E\left[A^{ShPath}\right] = 0.5$ for any shortest path length L.

The comparison between shortest path rule and random neighbor rule is based on the previously mentioned observation that the random neighbor rule effectively uses a randomized path between some source node s and the target node t. If this path is a shortest path, then the random neighbor rule matches the accuracy of the shortest path rule (when there are no other source nodes). If the path used by the random neighbor rule is not a shortest path, then the shortest path rule's expected accuracy is weakly higher due to Eq. (7). Moreover, with multiple source nodes, the random neighbor rule only uses one of them, while the shortest path rule aggregates information obtained from all source nodes. Hence, the shortest path rule always weakly dominates the random neighbor rule in terms of expected accuracy.

Taken together, the Bayesian rule weakly dominates the shortest path rule, which in turn weakly dominates the random neighbor rule. To be clear, this does not mean that there are no instances, where a dominated rule is accurate and a undominated rule is not. For instance, we can have, e.g., $A^{RNeighbor} = 1$ while $A^{ShPath} = 0$ in some realization, but ex ante the ordering $E\left[A^{ShPath}\right] \geqslant E\left[A^{RNeighbor}\right]$ holds for any network.

Since Proposition 1 only establishes a weak dominance relation between the three rules, the consequent question is when the three rules do not differ in accuracy. The next result identifies a condition for all three rules to be equally accurate.

Proposition 2 (*Unique Path*). *Let $N_S = 1$. Suppose that in network (V, E) there is a unique path between source node s and the target node t. Then all three rules (Bayes, ShPath, RNeighbor) lead to the same expected accuracy:*

$$E\left[A^{Bayes}\right] = E\left[A^{ShPath}\right] = E\left[A^{RNeighbor}\right].$$

This proposition follows from the Claims 1 and 2, which establish when the shortest path rule leads to the same belief as the Bayesian rule and when it leads to the same best guess as the random neighbor rule. Indeed, the condition of Proposition 2, i.e., there is a unique path between the only source node and the target node, is equal to the condition of Claim 2, which is stronger than the condition of Claim 1. Hence, all three rules must lead to the same best guess and hence to equal accuracy under this stronger condition. As the three accuracy values are equal for each realization of states, signals, and link signs, they are also equal on average.

The intuition for this result is illustrated by Fig. 4. First, the existence of a unique path from s to t implies that the network must look as shown in this figure, where subgraphs are attached to individual nodes of the path. Second, these subgraphs do not affect the inference of the three considered rules. Finally, all three rules lead to the same best guess when only the path from s to t is considered.

Proposition 2 applies in particular to the class of tree networks, as trees are characterized by a unique path between any two nodes. The opposite direction of the statement – the three rules lead to the same expected accuracy *only if* there is a unique path between the source node(s) and the target node – seems intuitive, but does not hold in general.[14]

In this section we have theoretically established the order of the three rules according to expected accuracy. To quantify these differences between the three rules, we will now turn to simulation results.

5. Simulation results

We first define the simulation procedure.

5.1. Simulation procedure

For each of the three rules (*Bayesian rule*, *shortest path rule*, *random neighbor rule*), the simulations proceed as follows:

1. Generate a synthetic random network of a given class with N nodes and mean degree z. If it is not connected, generate a new network.[15]
2. Choose N_S nodes at random to serve as the source nodes.
3. Draw node types, link signs, and source node signals according to the model introduced in Section 3.2.
4. For each node that is not a source node, designate it as target node, apply a given rule (*Bayes*, *ShPath*, or *RNeighbor*) according to Section 3.3, and measure its accuracy.
5. Repeat the above steps $1,000$ times. From the obtained accuracy values, compute mean accuracy, $\overline{A}$, and the standard error of the mean (SEM), $\sigma_{\overline{A}}$. (Figures reporting simulation results depict the mean with error bars that are twice the SEM.)

We apply this procedure for various model parameters (q, r), network topologies, and network sizes N.

[14] For instance, in Panel (C) of Fig. 3, the two paths either have the same orientation and the best guess thus coincides for all three rules, or the paths have opposite orientations and the probability of an accurate judgment is thus one half for all three rules.

[15] Further network generation details are provided in Appendix C.1.

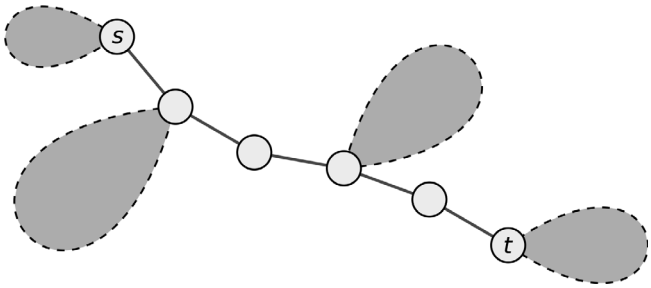

Fig. 4. Illustration of Proposition 2. When there is a unique shortest path between source node s and target node t, the three rules induce the same accuracy.

5.2. Varying signal quality and link reliability

From the theoretical results we know the ordering of the three rules: $E\left[A^{Bayes}\right] \geqslant E\left[A^{ShPath}\right] \geqslant E\left[A^{RNeighbor}\right]$ (Proposition 1) and that they coincide in special cases (Proposition 2). Let us now assess how large the differences between the three rules are. Fig. 5 illustrates how accuracy depends on the applied rule for different levels of signal quality q and link reliability r. The illustrated simulations use Erdös-Rényi random networks of $N = 20$ nodes with mean degree $z = 5$. The left panels show results for $N_S = 1$ source node, the right panels for $N_S = 5$ source nodes.

Clearly, accuracy of all rules is weakly increasing in signal quality q and weakly increasing in link reliability r. Fig. 5 illustrates that, for intermediate values of link reliability r, the rules significantly differ in accuracy. For instance, in the top left panel at $r = 0.8$, accuracy of the Bayesian rule (around 80%) is substantially higher than accuracy of the shortest path rule (around 70%), which in turn is substantially higher than accuracy of the random neighbor rule (slightly above 60%). Considering the other depicted settings of signal quality q and number of source nodes N_S, there are again levels of link reliability r with significant differences between the three rules. Hence, we receive the following first result:

Result 1. Differences in accuracy between the three rules can be substantial.

Accuracy of all rules and hence also differences are smaller for low levels of signal quality q and they would vanish when quality approaches $q = 0.5$. Similarly, differences also vanish when link reliability approaches $r = 0.5$.

Let us now explore how the presence of multiple source nodes affects the comparison of the three rules. The simulations indicate that the shortest path rule may be closer to the random neighbor rule for one source node, but closer to the Bayesian rule when there are multiple source nodes. Comparing the left and right panel of the figure reveals that multiple source nodes improve accuracy of the Bayesian and the shortest path rule. For the random neighbor rule, multiple source nodes do not substantially improve accuracy. Recall that this rule effectively uses only one source node. Thus, multiple source nodes do not help to aggregate information, but only to reduce the length of the used path, on average. This disadvantage of the random neighbor rule can also be seen for the case of perfectly reliable links, $r = 1$, and understood as follows.

In the extreme case of $r = 1$, the signed network has the *two-camp structure* such that there are only positive links within camps and negative links across [4]. In other words, the network is socially balanced [13]. Hence, there are only two possible states of the world, which differ by which camp is of the positive type. Source nodes provide information about the likelihood of these two states. In particular, if there is only one source node ($N_S = 1$), the expected accuracy of any of the three rules is q. For more than one source node, the expected accuracy is higher for the Bayesian and the shortest path rule, but not for the random neighbor rule. Accuracy of the random neighbor rule stays at q, no matter how the number of source node changes. This is the intuition for the second result:

Result 2. The difference in accuracy between the shortest path rule and the random neighbor rule is particularly pronounced for multiple source nodes.

Moreover, we can observe in Fig. 5 that r is not an upper bound for expected accuracy of the Bayesian rule. For instance at $r = 0.9$, the Bayesian rule's expected accuracy is larger than 0.9. The reason is that multiple paths with link reliability of r can yield information that exceeds r. For the other two rules, the shortest path rule and the random neighbor rule, link reliability r is in fact an upper bound for expected accuracy when we have only one source node. The reason is that only one path is actually used such that information of multiple paths from one source node are not aggregated by these two rules (cf. Example 1).

5.3. Different network topologies

We now investigate how the network topology affects accuracy. We have so far used Erdös-Rényi (ER) random networks, now we consider three additional network classes: regular graphs (RG), Barabási-Albert (BA) scale-free networks, and Watts-Strogatz small world (WS) networks (see [24], for more information on these basic network models). These networks differ

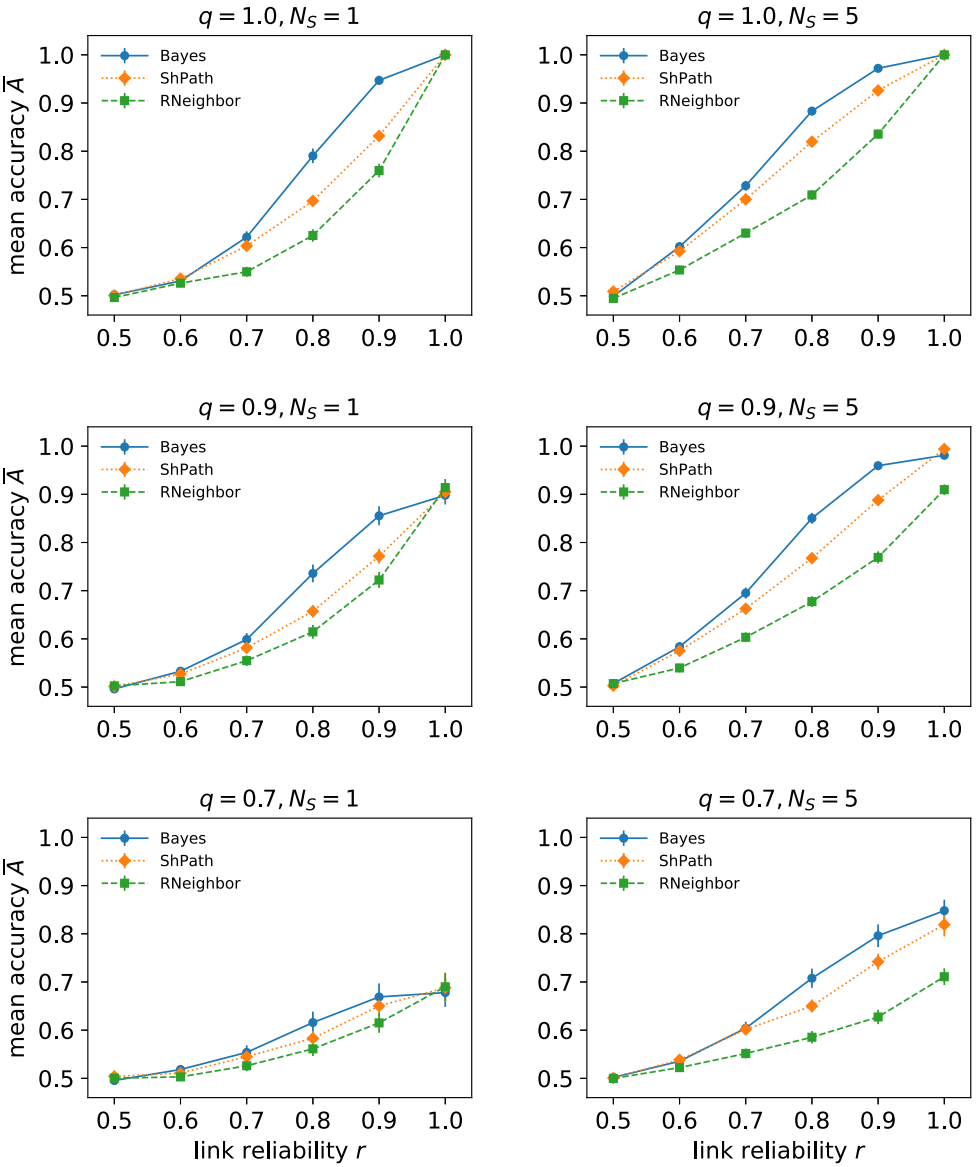

Fig. 5. Mean accuracy of the three rules for different levels of link reliability r. Signal quality is $q = 1$ (top row), $q = 0.9$ (middle row), $q = 0.7$ (bottom row). Results are for Erdös-Rényi random networks with $N = 20, z = 5, N_S = 1$ (left) and $N_S = 5$ (right).

in several characteristics: shape of the degree distribution (regular graphs have the most homogeneous degree distribution, BA networks the most heterogeneous), clustering (WS networks have high clustering coefficient when the rewiring probability is small), and the distances, as we will discuss.

First, we check whether our simulation results are robust to changing the class of network. Fig. 9 in Appendix C.2 re-examines the differences between the three rules of Fig. 5 when the Erdös-Rényi random network generation process is replaced with RG, BA, and WS. Indeed, the differences between the three rules are very similar for all topologies and hence our two empirical results R1 and R2 (see Section 5.2) are robust.

Second, we address how two crucial features of a network structure – density and distances – affect accuracy.

Concerning density, Fig. 6 shows how expected accuracy in the three rules increases with mean degree z. Intuitively, denser networks provide more information and shorter paths, which should increase expected accuracy of all three rules. We see that this is the case for the Bayesian rule and, to a lesser extent, for the shortest path rule with decreasing gains in accuracy. For the random neighbor rule, however, the increase of expected accuracy is insubstantial for $N_S = 5$ and it saturates early (at $z \approx 4$) for $N_S = 1$. This agrees with the results presented in [22] where an insubstantial dependence on z was observed too. Hence, we conclude:

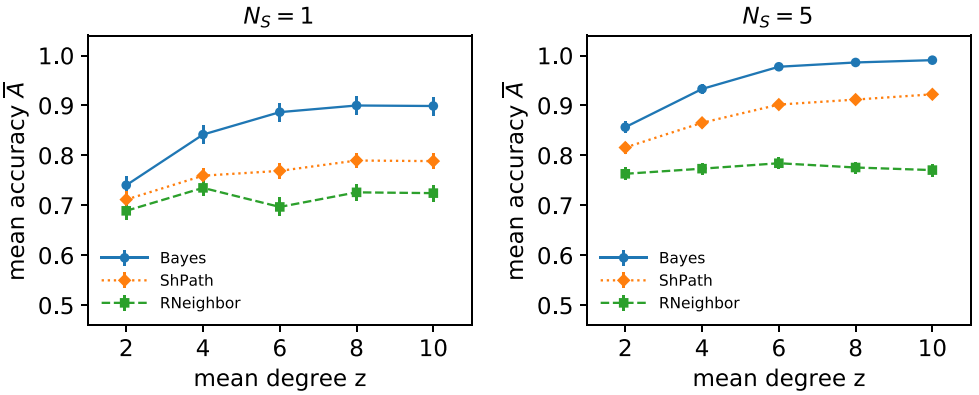

Fig. 6. Mean accuracy of the three rules for increasing mean degree z. Results are for Erdös-Rényi random networks with $N = 20, q = 0.9, r = 0.9, N_S = 1$ (left) and $N_S = 5$ (right).

Result 3. Accuracy of the Bayesian rule and the shortest path rule weakly improves with the network density.

Let us now address the effect of distances. To vary the distances, without changing the density of a network, we study the different classes of networks – ER, RG, BA, WS – all for the same density and measure the distances their realizations have. For WS networks we also vary the rewiring probability β, as this parameter affects distances, but not density. Our simulated networks satisfy $d_{BA} < d_{ER} < d_{WS(\beta=0.5)} \approx d_{RG} < d_{WS(\beta=0.1)}$ (empirically determined shortest lengths are $3.54 \pm 0.03, 4.20 \pm 0.03, 4.51 \pm 0.02, 4.55 \pm 0.01$, and 6.85 ± 0.05, mean $\pm$ twice the SEM, respectively) for $N = 300$ and $z = 4$ in Fig. 7.

Fig. 7 shows that the mean accuracy achieved by the shortest path rule is the lowest for the WS networks with low rewiring probability, $\beta = 0.1$, where the average shortest path length is considerably longer than in the other networks. By contrast, the highest mean accuracy (for fixed link reliability, r) is achieved in BA networks where distances are shortest. These observations agree with Claim 3 which identifies the shortest path length as the key determinant of the expected accuracy for the shortest path rule. These findings strongly suggest that the average path length of a network is a key determinant for accuracy of the shortest path rule.

Result 4. The shorter the average distances in a network, the higher the accuracy of the shortest path rule.

The same ordering of network topologies has been reported for the random neighbor rule in [22]. Appendix C.2 discusses similar results for the Bayesian rule, suggesting that the shortest path length is important also for this optimal rule.

5.4. Increasing the network size

We finally analyze how results scale when network size N becomes larger. To focus on the effect of growing network size only, we keep mean degree fixed. We consider two settings. Either the number of source nodes is kept constant or it grows proportionally to network size N.[16] When $N_S \sim N$, the expected accuracy of the random neighbor rule quickly approaches a limit value greater than 0.5. We now explore accuracy of the shortest path rule in comparison to the random neighbor rule for growing N with simulations. (We also include results for the Bayesian rule, although this exercise is limited, as the rule's computational complexity restricts our simulations to $N \leqslant 20$.).

Fig. 8 shows the mean accuracy against the network size for several settings of q and r, using Erdös-Rényi random networks with mean degree $z = 5$. Comparing the different panels, mean accuracy in the top row is slightly lower than in the middle row because it has a lower signal quality of $q = 0.9$ (versus $q = 1$), while link reliability is $r = 0.9$ in both. Accuracy in the bottom row is lower than in the middle row because it has a lower link reliability $r = 0.7$ (versus $r = 0.9$), while signal quality is perfect, i.e., $q = 1$, in both.

Besides confirming the previously reported results for the random neighbor rule [22], there are a number of observations to make from Fig. 8. First, the shortest path rule performs significantly better than the random neighbor rule, in particular when there are several source nodes ($N_S = 5$, middle column) or when the number of source nodes grows with the number of nodes ($N_S = N/10$, right column). This confirms Results 1 and 2 from above.

Second, when N_S is fixed, the mean accuracy of all three rules decreases with N, albeit at different rates. The rate of the shortest path rule, like for the random neighbor rule, can be derived analytically for a single source node, as detailed in

[16] The dependence of the accuracy on the network size has been a major topic in [22]. They show that the random neighbor rule's expected accuracy slowly converges to 0.5 (equivalently, in their terms the rule's consistency convergences to 0) as N grows when the number of source nodes is constant.

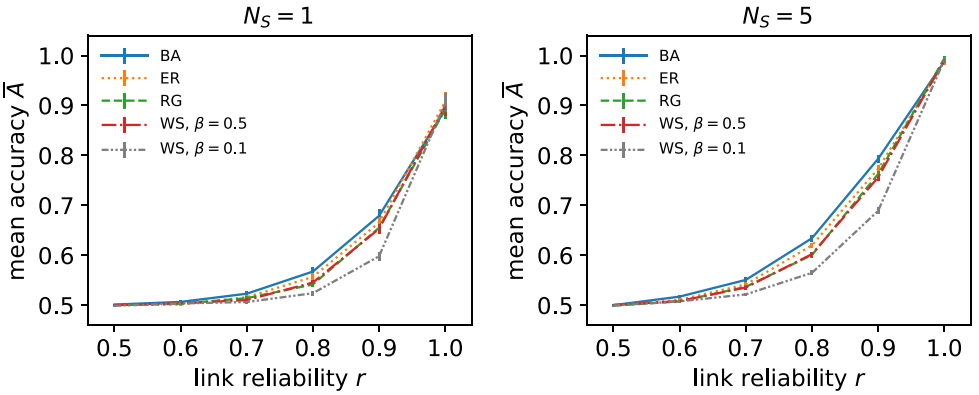

Fig. 7. Mean accuracy of the shortest path rule for various classes of random networks and different levels of link reliability r. Results are for $N = 300, z = 4, q = 0.9, N_S = 1$ (left) and $N_S = 5$ (right).

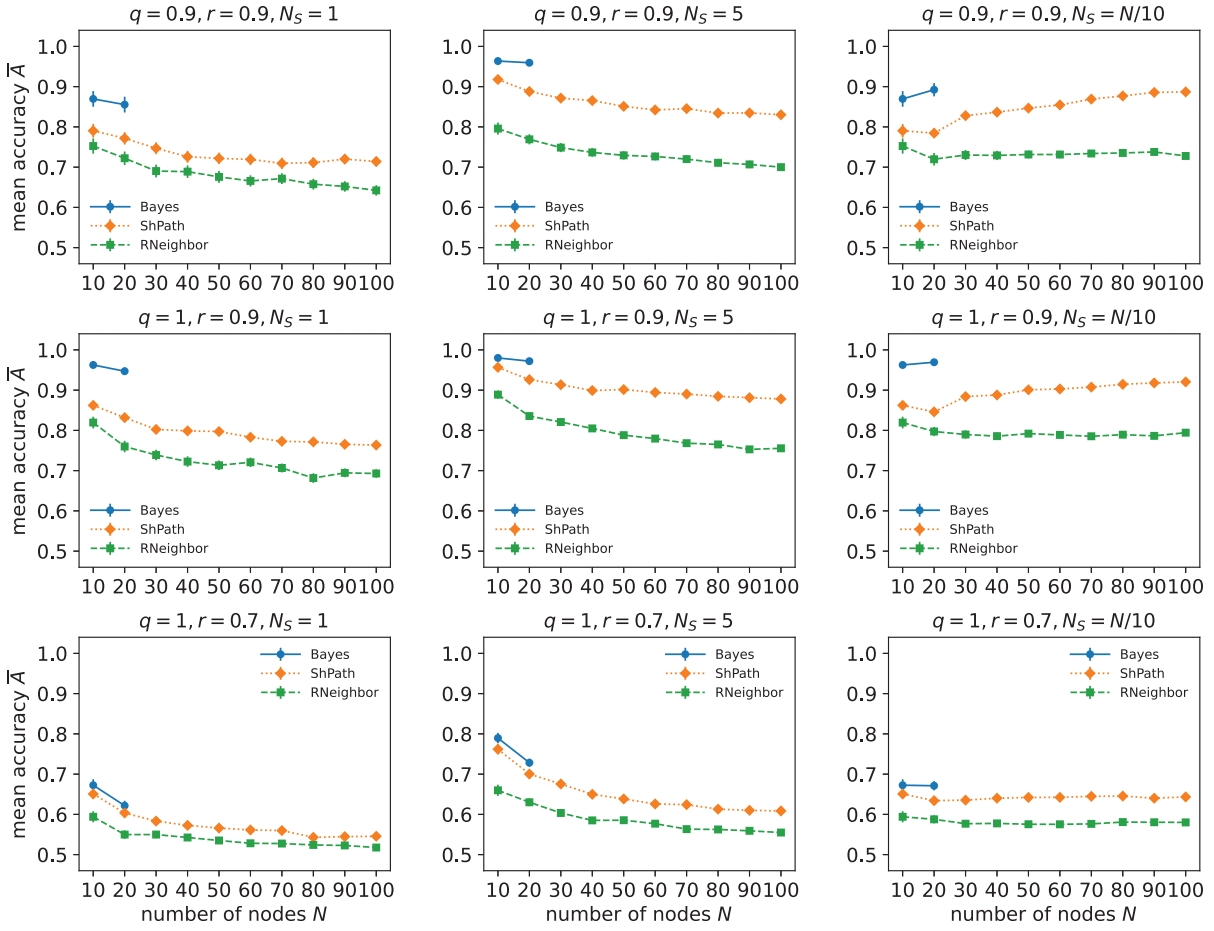

Fig. 8. Mean accuracy of the three rules for growing network size N. Results are for Erdös-Rényi random networks with $z = 5, q = 0.9$ and $r = 0.9$ (top row), $q = 1$ and $r = 0.9$ (middle row), $q = 1$ and $r = 0.7$ (bottom row), $N_S = 1$ (left column), $N_S = 5$ (middle column), and $N_S = N/10$ (right column).

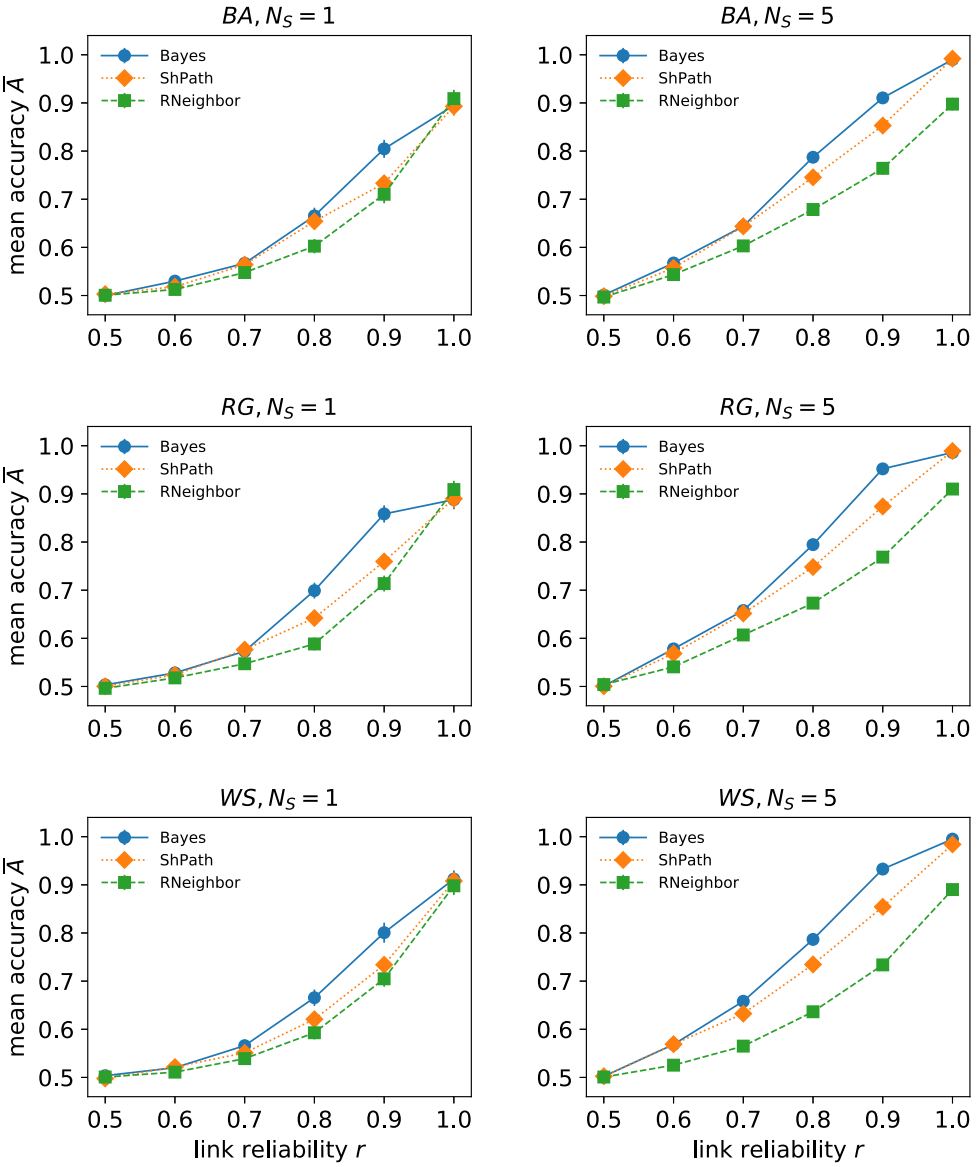

Fig. 9. Mean accuracy of the three rules for various classes of random networks. Setting signal quality $q = 0.9$, network size $N = 20$, average degree $z = 4$, and for different values of link reliability r. Top panels use BA, center panel RG, lower panel WS (rewiring probability is $\beta = 0.1$). Left panels have $N_S = 1$ source node; right panels have $N_S = 5$ source nodes.

Appendix C.3. In particular, the classical result for the average distance between two nodes chosen at random in Erdös-Rényi random networks, $\bar{L} \approx \ln(N)/\ln(z)$ (e.g., [3]), can be plugged in Eq. (7) to obtain

$$E\left[A^{ShPath,ER}\right] = \frac{1}{2} + \left(q - \frac{1}{2}\right)(2r-1)^{\ln(N)/\ln(z)} \tag{8}$$

which holds for the shortest path rule and one source node. This $E\left[A^{ShPath,ER}\right]$ approaches 0.5 as $N \to \infty$ for any $r < 1$. The reason for the decrease in expected accuracy is the effect of longer distances (cf. Result 4), but the speed of convergence to 0.5 is lower than for the random neighbor rule.

Result 5. When the number of source nodes is fixed, mean accuracy of the shortest path rule converges to 0.5 slower than mean accuracy of the random neighbor rule.

Third, when the number of source nodes grows with N, the mean accuracy can even increase with N (see results for the shortest path rule for $r = 0.9$ and $N_S = N/10$) – the information gained from a growing number of source nodes is sufficient to overweight the shortest paths that become longer, on average, as N grows. This occurs for sufficiently high link reliability $r = 0.9$. Theoretically, this effect is clear for $r = 1$, where additional source nodes mean additional information without any loss due to long distances. This argument also applies to the Bayesian rule, for which expected accuracy is weakly above the shortest path rule's.

Result 6. Given that link reliability r is large enough, when the number of source nodes grows proportionally with network size N, accuracy of the Bayesian rule and the shortest path rule increases, in contrast to the random neighbor rule.

Fourth, although the Bayesian rule produces substantially higher mean accuracy when $r = 0.9$, the difference between the Bayesian rule and the shortest path rule becomes much smaller when $r = 0.7$ (bottom row). This holds for all three considered numbers of source nodes: $N_S = 1, N_S = 5$, and $N_S = N/10$.

Result 7. For low link reliability r accuracy of all three rules remains relatively low regardless of the number of source nodes.

This last result highlights the importance of link reliability: low r cannot be compensated by the source information.

6. Conclusion

We have investigated how an observer can learn from a signed network where node types correlate with link signs. In such a setting, information about the type (e.g., trustworthiness) of source nodes is informative for the type of a target node. While we start by formulating the optimal Bayes belief for the target node type, our chief interest is in heuristic rules whose cognitive complexity is much lower. We prove that in stylized networks, including trees, two relatively simple heuristics can lead to optimal learning. Studying less-stylized random networks with a sequence of simulations, we find that differences in accuracy between the two heuristics and the Bayesian rule is often substantial. The heuristic that was studied previously [22], the random neighbor rule, performs poorly when there are multiple source nodes. The heuristic that we introduce in this paper, the shortest path rule, avoids this by aggregating the information obtained using all source nodes. Its accuracy, however, relies on short average distances.

Our model could be extended in several directions. First, we focus on two node types, while we could integrate k node types as follows. Instead of a single link reliability parameter r, consider a symmetric $k \times k$ matrix specifying the probability that nodes of each type are connected by a positive link. Our current model is then nested by using a two-by-two matrix where r is on the diagonal and $1 - r$ is off the diagonal. Nevertheless, one justification for two node types comes from the property of structural balance that is equivalent to the two-camp structure [13].

Second, to keep our model simple, we assumed that positive and negative links are equally informative, as well as that positive and negative signals are equally informative. A more general model could introduce asymmetry, e.g., to capture that a positive signal form a source node with a positive link to the target node is more informative than a negative signal from a source node with a negative link.

Third, our model uses undirected networks. In some applications, the signed networks are directed and hence this is a considerable extension.

Fourth, one implicit assumption of our model is that the observer knows the whole network and all link signs. This is in particular necessary to apply the Bayesian rule. For the application of the shortest path rule, the observer needs to at least identify the shortest paths between the source node(s) and the target node, which is trivial in some examples, but a non-trivial task when the target node is distant from the source node(s). The research of navigability in complex networks [2,21] is relevant in this respect, as it seeks to understand which features of complex networks are helpful for finding shortest paths in them.

Fifth, our model supposes that only link signs depend on node types, while the presence or absence of a link does not. This makes the analysis and interpretation clean, but it is a limitation, as in reality also the presence of a link can depend on link types. For instance, homophily in link formation can influence the presence of links (e.g., [5]) and evidence on homophily is abundant (e.g., [17]). With this perspective, the vast literature on detecting communities in non-signed networks [10] serves the goal of learning about node types from the presence of links. Complementary to that, we propose a model of learning about node types from the sign of links. A new research avenue would combine these two approaches in a framework that accommodates link presence and link signs that depend on link type.

Finally, empirical studies could be used to investigate the opinion and belief formation of human beings on signed network data. Our model can help to measure and understand accuracy of given judgment behavior, and it can potentially also help to improve accuracy in judgment, for instance in the context of a manager who wants to promote an employee.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Manuel Mariani, Stefan Kloessner, and the audience of the Networks 2021 (the joint Sunbelt and NetSci conference) for helpful comments. Fanyuan Meng thanks the China Scholarship Council for financial support.

Appendix A. Proofs

A.1. Proof of Claim 1 (Equivalence)

The condition of Claim 1 (a unique path between t and each $s \in S$ and these paths are non-overlapping) implies a network structure as exemplified in Fig. 2 where t is connected with source nodes by non-overlapping rays. Besides this tree structure, the network can contain arbitrary subgraphs appended to individual network nodes as they do not affect the uniqueness of paths between source nodes and t (see the subgraph attached to node j_3 and the path between s_5 and t in Fig. 2, for example).

We first show that if the network consists only of the path between one source node, s, and target node, t, the Bayesian rule represented with Eq. (1) simplifies to

$$P[\boldsymbol{\theta}|\boldsymbol{\sigma}, R] = q^{z_1(\boldsymbol{\theta})}(1-q)^{z_2(\boldsymbol{\theta})} r^{z_3(\boldsymbol{\theta})}(1-r)^{z_4(\boldsymbol{\theta})}$$

as the denominator of Eq. (1) (after canceling $1/2^N$) is then

$$\Omega := \sum_{\boldsymbol{\theta}' \in \Theta} q^{z_1(\boldsymbol{\theta}')}(1-q)^{z_2(\boldsymbol{\theta}')} r^{z_3(\boldsymbol{\theta}')}(1-r)^{z_4(\boldsymbol{\theta}')} = 1.$$

We prove this by induction. For the elementary graph with two nodes, s and t, and one link, the sum over four possible states yields $\Omega = qr + q(1-r) + (1-q)r + (1-q)(1-r) = 1$. Here the first term, for example, corresponds to correct source signal and type-conforming link between the two nodes. In one induction step, we append a node of degree one to one end of the current network for which $\Omega = 1$ holds (adding a link elsewhere would create a network outside the considered class). In the new denominator Ω' each term of the original denominator Ω is paired with two possible states of the added node, corresponding to forming a node-type conforming link (with probability r) and a nonconforming link (probability $1-r$), respectively. We can thus write $\Omega' = \Omega r + \Omega(1-r) = \Omega = 1$ which completes this step of the proof.

As the second step, we show that the belief derived with the Bayesian rule in Eq. (2) coincides with the belief derived with the shortest path rule in Eq. (5) when the network consists only of the shortest path between s and t. Eq. (2) states that

$$b^{Bayes}(t) = \sum_{\boldsymbol{\theta} \in \Theta(t)} P[\boldsymbol{\theta}|\boldsymbol{\sigma}, R],$$

where $\Theta(t)$ are all 2^{N-1} states with $\theta_t = +1$. When $\boldsymbol{\sigma}$ and R are given, the mapping between states of the world $\boldsymbol{\theta}$ and the values of z_1 and z_2 together with the conformity information for each link is bijective. The direction from $\boldsymbol{\theta}$ to z_1 and conforming/nonconforming links is obvious as it is used by the Bayesian rule. For the opposite direction, σ_s combined with z_1 determines the type of s (if $z_1 = 1, \theta_s = \sigma_s$, otherwise $\theta_s = -\sigma_s$). Progressing from the source node, the next node's state is $\theta_i = R_{ij}\theta_j$ if the link is conforming to the node types and $\theta_i = -R_{ij}\theta_j$ if the link is nonconforming. Each term in the sum in Eq. (2) thus appears also in Eq. (5) (and vice versa). When $o_{s\to t} = +1$, for instance, terms proportional to q (correct source signal) will be proportional to $(1-r)^n$ where n is even because an odd number of incorrect links would lead to $\theta_t = -1$ which does not contribute to the sum in Eq. (2). This step of the proof is thus concluded.

As the third step, we prove that attaching a subgraph to a single node in a given network does not change the result obtained with the Bayesian rule. Upon attaching a subgraph with link signs R' to a node, the space of all states grows from Θ to $\Theta \times \Psi$ where Ψ is the space of states $\boldsymbol{\psi}$ of the subgraph nodes. We can write $P(\boldsymbol{\theta}|\boldsymbol{\sigma}, R) = \sum_{\boldsymbol{\psi}\in\Psi} P(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{\sigma}, R, R')$. The numerator of $P(\boldsymbol{\theta}|\boldsymbol{\sigma}, R)$ can be written as

$$\frac{q^{z_1(\boldsymbol{\theta})}(1-q)^{z_2(\boldsymbol{\theta})} r^{z_3(\boldsymbol{\theta})}(1-r)^{z_4(\boldsymbol{\theta})}}{2^{N+N'}} \sum_{\boldsymbol{\psi}\in\Psi} r^{z_3(\boldsymbol{\psi})}(1-r)^{z_4(\boldsymbol{\psi})},$$

where N' is the number of nodes in the subgraph. Crucially, the sum over $\boldsymbol{\psi} \in \Psi$ is independent of $\boldsymbol{\theta}$ as the subgraph is attached to a single node. If the attachment node is of a positive type and $\boldsymbol{\psi}$ yields some $z_3(\boldsymbol{\psi})$ and $z_4(\boldsymbol{\psi})$, then $-\boldsymbol{\psi}$ would yield the same z_3 and z_4 for a negative type of the attachment node.[17] The terms $1/2^{N+N'}$ and $\sum_{\boldsymbol{\psi}\in\Psi} r^{z_3(\boldsymbol{\psi})}(1-r)^{z_4(\boldsymbol{\psi})}$ cancel out and we find that the obtained $P(\boldsymbol{\theta}|S, R)$ is the same as without the attached subgraph. If adding subgraphs does not change the Bayesian

[17] If, by contrast, we attach node x to two different nodes of an existing graph, the sum over two possible states of node x is $r^2 + (1-r)^2$ when the attachment nodes have the same types whereas it is $2r(1-r)$ when the attachment nodes have different types. The sum over $\boldsymbol{\psi} \in \Psi$ thus depends on $\boldsymbol{\theta}$ and does not factor out from $P(\boldsymbol{\theta}|\boldsymbol{\sigma}, R)$.

inference, removing them has no effect as well. If only the shortest path between s and t remains after removing all subgraphs attached to single nodes, the Bayesian rule thus reduces to the shortest path rule as we wanted to prove.

We have shown that for every network with a unique path from source node to target node both Eq. (2) and Eq. (5) correctly determine the posterior belief about the target node's type. Consider now a network with multiple source nodes that have unique, non-overlapping, paths to the target node as illustrated in Fig. 2. Each subgraph with a source node yields independent information about the target node. Since Eq. (6) is the correct aggregation formula for independent signals, we can conclude that the introduced shortest path rule indeed coincides with the Bayesian rule for the given class of networks, which finally completes our proof.

A.2. Proof of Claim 2 (Same Best Guess)

We first show that according to the shortest path rule, the best guess coincides with the path orientation, when there is only one source node. The shortest path rule uses path orientation defined by Eq. (3), $o_{s\to t} := \sigma_s \prod_{m=0}^{L-1} R_{n_m,n_{m+1}}$ where $n_0,\ldots,n_L$ are the nodes along the path, $n_0 := s$, and $n_L := t$. It further computes the probability that the path orientation agrees with the type of node t as $P(o_{s\to t} = \theta_t) = q\pi_{s\to t} + (1-q)(1-\pi_{s\to t})$ (see Eq. (5) and the discussion thereafter). Here $\pi_{s\to t}$ is the probability that the path between s and t has an even number of errors. Since $r > 0.5$, it follows from Eq. (4) that $\pi_{s\to t} > 0.5$ for any finite L. We can now rewrite

$$P(o_{s\to t} = \theta_t) = 1 - \pi_{s\to t} + q(2\pi_{s\to t} - 1)$$

which is increasing in q. Since $P(o_{s\to t} = \theta_t) = 0.5$ when $q = 0.5, P(o_{s\to t} = \theta_t) > 0.5$ for $q > 0.5$. The computed probability that $o_{s\to t}$ correctly identifies the type of node t is more than 0.5, so the best guess of t's type produced by the shortest path rule is thus always the path orientation. Note that we assumed here one source node but not the path uniqueness, so the last statement holds for the shortest path rule on any network where $N_S = 1$.

When the path between s and t is unique, the random neighbor rule is deterministic as the path over which the opinion on t is formed is given. According to this rule, the judgment of node n_1, which is adjacent to $n_0 := s$, is a product of the source signal and the sign of the link between the two nodes. The judgment of node n_2 is further multiplied by $R_{n_1 n_2}$, and so on, until the judgment of t is determined as a product of the source signal with link signs along the path which is precisely the path orientation $o_{s\to t}$. The two rules thus coincide when $N_S = 1$ and the path between s and t is unique.

A.3. Proof of Claim 3 (Shortest Path Accuracy)

Assume that node states $\boldsymbol{\theta}$ are fixed. As we showed in Section A.2, $o_{s\to t}$ is the best guess of θ_t for the shortest path rule. The probability that $o_{s\to t} = \theta_t$ is $q\pi_{s\to t} + (1-q)(1-\pi_{s\to t})$ where the first term accounts for a correct source signal *and* an even number of errors along the path from s to t, and the second term accounts for an incorrect source signal *and* an odd number of errors along the path from s to t, respectively. Using Eq. (4), this probability can be simplified to the form

$$P(o_{s\to t} = \theta_t) = \frac{1}{2} + \left(q - \frac{1}{2}\right)(2r-1)^L.$$

This result already takes into account the limited reliability of source signals and link signs: it averages accuracy over different realizations of $\boldsymbol{\sigma}$ and R. As it does not depend explicitly on node states $\boldsymbol{\theta}$, it is straightforward to average it over $\boldsymbol{\theta}$ and thus obtain the expected accuracy in line with Definition 2.

A.4. Proof of Proposition 1 (Ordering)

The Bayesian rule uses all available information, source signals $\boldsymbol{\sigma}$ and link signs R, to correctly compute the objective probability that target node t is of positive type. It is thus the optimal rule by construction and its expected accuracy is the highest of all possible rules, hence $E\left[A^{Bayes}\right] \geqslant E\left[A^{ShPath}\right]$ and $E\left[A^{Bayes}\right] \geqslant E\left[A^{RNeighbor}\right]$.

When $N_S = 1$, the ordering between the shortest path rule and the random neighbor rule can be established as follows. Denote the shortest path length between s and t as L; the expected accuracy of the shortest path rule can be obtained using Eq. (7). As already explained, the judgment formed using the random neighbor rule corresponds to a path between s and t. Denote this path's length as L'. For a given path between s and t, the best guess of the shortest path rule is the same as the best guess formed by the random neighbor rule (both are equal to the path orientation $o_{s\to t}$). The expected accuracy of the random neighbor rule can be thus computed using Claim 3 with the only difference being that the shortest path length L is replaced by the actual path length L'. Now, every path length L' must be at least as long as the shortest path length, $L' \geqslant L$. As the expected accuracy decreases with L' [when $r \in (0.5, 1)$], we immediately obtain that $E\left[A^{ShPath}\right] \geqslant E\left[A^{RNeighbor}\right]$.

When $N_S > 1$, the judgment of t formed using the random neighbor rule is still formed using a path of length L' between one source node and target node t. By contrast, the shortest path rule uses shortest paths from each source node and aggregates beliefs resulting from Eq. (5) using Eq. (6). While these paths have different lengths, L_i, the shortest of them cannot be longer than $L' : \min L_i \leqslant L'$. This path alone results in expected accuracy of the shortest path rule which is larger or equal than

expected accuracy of the random neighbor rule. The additional paths used by the shortest path rule, some of which may be longer than L', represent useful information that is taken into account by this rule and thus further improve its expected accuracy. We can thus conclude that $E\left[A^{ShPath}\right] \geqslant E\left[A^{RNeighbor}\right]$ holds also for $N_S > 1$.

A.5. Proof of Proposition 2 (Unique Path)

According to Claim 1, a unique non-overlapping path between every source node s and the target node t implies that the best guess obtained with the Bayesian rule and the shortest path rule are the same. If, in addition, there is only one source node, Claim 2 states that the best guess of the shortest path rule and the random neighbor rule are the same. When $N_S = 1$ and the path between s and t is unique, the three rules thus yield the same best guess for any realization of $\boldsymbol{\theta}, \boldsymbol{\sigma}, R$; their accuracy values are thus the same. As a consequence, the expected accuracy values obtained by averaging over $\boldsymbol{\theta}, \boldsymbol{\sigma}, R$ are also the same.

Appendix B. Examples from Fig. 3

Example B.1. [Two Triangles] Let $(V, E) = (\{s, j_1, j_2, t\}, \{st, sj_1, sj_2, j_1t, j_2t\})$. Let $R_{st} = R_{sj_1} = R_{sj_2} = +1$ and $R_{j_1t} = R_{j_2t} = -1$ and let $\sigma_s = +1$, as illustrated in Panel (A) of Fig. 3. The *Bayesian rule* yields

$$b^{Bayes}(t) = \frac{4qr^3(1-r) + 2(1-q)r^2(1-r)^2 + (1-q)\left[r^4 + (1-r)^4\right]}{4r^3(1-r) + 2r^2(1-r)^2 + \left[r^4 + (1-r)^4\right]},$$

whereas the shortest path rule yields

$$b^{ShPath}(t) = qr + (1-q)(1-r).$$

It can be checked that $b^{Bayes}(t) < b^{ShPath}(t)$ for any $q, r \in (0.5, 1)$. This is because the shortest path rule only uses the direct path, which has a positive orientation, whereas the Bayesian rule uses all information including the two paths of length 2, which are both negatively oriented.

Example B.2. [Multiple Shortest Paths] Let $(V, E) = (\{s, j_1, j_2, j_3, t\},$
$\{sj_1, sj_2, sj_3, j_1t, j_2t, j_3t\})$. Let $R_{ij} = +1$ for all links in E except for j_3t ($R_{j_3t} = -1$) and $\sigma_s = +1$, as illustrated in Panel (B) of Fig. 3.

The Bayesian rule always leads to a belief $b^{Bayes}(t) > 0.5$, hence the best guess is $g^{Bayes}(t) = +1$.

The shortest path rule selects one of the three shortest paths at random. Hence, with probability 2/3 it induces belief $b^{ShPath}(t) = qr + (1-q)(1-r) > 0.5$ and with probability 1/3 it induces belief $b^{ShPath}(t) = q(1-r) + (1-q)r < 0.5$.

Example B.3. [Overlapping Paths] Let $(V, E) = (\{s_1, s_2, j, t\}, \{s_1j, s_2j, jt\})$. Let $R_{ij} = +1$ for $ij \in E$ and let $\sigma_{s_1} = \sigma_{s_2} = +1$, as illustrated in Panel (C) of Fig. 3.

The Bayesian rule yields

$$b^{Bayes}(t) = \frac{q^2\left[r^3 + (1-r)^3\right] + 2q(1-q)r(1-r) + (1-q)^2r(1-r)}{\left[q^2 + (1-q)^2\right]\left[(1-r)^3 + r^3\right] + 4q(1-q)r(1-r) + \left[q^2 + (1-q)^2\right]r(1-r)},$$

whereas the shortest path rule yields

$$b^{ShPath}(t) = \frac{b_{s_1\to t}b_{s_2\to t}}{b_{s_1\to t}b_{s_2\to t} + (1 - b_{s_1\to t})(1 - b_{s_2\to t})},$$

where $b_{s_1\to t} = b_{s_2\to t} = q\left[r^2 + (1-r)^2\right] + 2(1-q)r(1-r)$. One can check that $b^{ShPath}(t) > b^{Bayes}(t)$ for any $q, r \in (0.5, 1)$. The excess confidence of the shortest path rule is due to the assumption of independence when signals from the two paths are combined; ignoring that they are correlated by having link jt in common.

Appendix C. Simulations

C.1. Simulation details

Erdös-Rényi (ER) random networks have two parameters: network size N and probability p which determines the connection probability for every pair of nodes in the network. We set $p = z/(N-1)$ to obtain mean node degree z.

Regular graphs (RG) are random graphs where each node has z links. They are constructed by assigning z "stubs" to each node and matching node stubs at random until no unmatched stubs remain. Loops and duplicate links are avoided.

Barabási-Albert (BA) random networks with N nodes are grown by gradually adding nodes to an initial empty network with m nodes. Every new node with m edges is attached to existing nodes using the preferential attachment mechanism (i.e., the probability of choosing a node is directly proportional to the node's degree). We set $m = \left[\left(N - \sqrt{N^2 - 2zN}\right)/2\right]$ to obtain mean node degree z (approximately).

Watts-Strogatz (WS) random networks have three parameters: network size N, mean degree z (this is assumed to be an even integer), and rewiring probability $\beta \in [0, 1]$. The networks are generated as follows:

1. Construct a regular ring lattice where each node is connected to its z nearest neighbors.
2. Go over all nodes $i = 1, \ldots, N$ and with probability β rewire each of the links to their right neighbors (i.e, neighbors $j = (i + k) \bmod N$ where $k = 1, \ldots, z/2$) to a node chosen at random. Loops and duplicate links are avoided.

In our simulations, we use the network generator functions provided by the LightGraphs v1.3.3 package of Julia programming language.[18]

C.2. Robustness of simulations with respect to network topology

Fig. 9 shows that our main results hold also for different network topologies. Simulations on small networks ($N = 20$, not shown) further indicate that the average shortest path length affect the Bayesian rule's mean accuracy (Fig. 7 shows the same for the shortest path rule).

C.3. Shortest Path Rule's Accuracy in Erdös-Rényi Random Networks

As explained in the main text, the average distance between two random nodes in Erdös-Rényi random networks is $\overline{L}_{ER} \approx \ln(N)/\ln(z)$ (e.g., [3]). By plugging this $\overline{L}_{ER}$ for L in Eq. (7), we obtain the expected accuracy of the shortest path rule in Erdös-Rényi random networks (Eq. (8) in the main text). As shown in Fig. 10, this analytical result agrees well with numerical simulations.

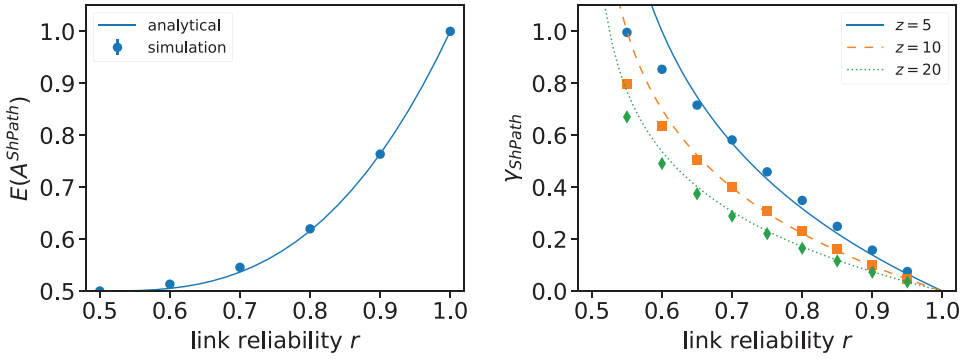

Fig. 10. Left panel: Mean accuracy of the shortest path rule for Erdös-Rényi random networks with $N = 100, z = 5, q = 1$. Right panel: Estimated exponents of the scaling relation $E\left(A^{ShPath}\right) - 0.5 \sim N^{-\gamma_{ShPath}}$ for numerical simulations on Erdös-Rényi random networks with various mean degree values (symbols) and analytical results obtained using Eq. (9) (lines). The exponents were estimated from simulation results using least squares fits between $\ln\left[E\left(A^{ShPath}\right) - 0.5\right]$ and $\ln[N]$ for N between 50 and 500.

[18] https://github.com/JuliaGraphs/LightGraphs.jl.

In [22], a relation similar to Eq. (8) has been derived for the random neighbor rule. This relation states that the "effective" accuracy, $E(A) - 0.5$, of the shortest path rule is proportional to $N^{-2(1-r)}$ when N_S is fixed. Eq. (8) allows us to derive here a similar scaling relation for the shortest path rule. We get $E\left[A^{ShPath}\right] - 0.5 \sim N^{-\gamma_{ShPath}}$ where

$$\gamma_{ShPath} = -\ln(2r-1)/\ln(z). \tag{9}$$

We see that unlike for the random neighbor rule where the scaling exponent $2(1-r)$ depends only the link reliability (see [22], for a derivation), here both link reliability and mean degree contribute. As mean degree increases, the scaling exponent γ decreases which means that $E\left[A^{ShPath}\right] - 0.5$ vanishes slower with N, i.e., $E\left[A^{ShPath}\right]$ approaches 0.5 slower. At the same time, γ_{ShPath} is independent of q: While signal quality q strongly influences the absolute value of expected accuracy $E\left[A^{ShPath}\right]$, it does not affect its decay with N.

Right panel of Fig. 10 shows that while Eq. (9) agrees well with numerical simulations for $r \gtrsim 0.65$, significant deviations appear for low r values (in particular, the analytical γ diverges as $r \to 0.5$). The main reason for the deviations is that to derive Eq. (8), we plugged the average shortest path length in Eq. (7). While this is justified by relatively narrow distributions of the shortest path length in Erdös-Rényi networks, it does not produce sufficiently precise results when r is close to 0.5: the term $2r-1$ in Eq. (7) is then small and A^{ShPath} thus depends strongly on L. If instead of using the average shortest path length, we numerically determine the distribution of the shortest path lengths in Erdös-Rényi networks and use this empirical distribution to average over various shortest path lengths in Eq. (7), we obtain scaling exponents that do not diverge as $r \to 0.5$ and agree well with numerical simulations for all r (results can be requested from the authors).

References

[1] Daniel J Benjamin, Errors in probabilistic reasoning and judgment biases, Handbook of Behavioral Economics: Applications and Foundations 1, vol. 2, Elsevier, 2019, pp. 69–186.
[2] Marian Boguna, Dmitri Krioukov, Kimberly C Claffy, Navigability of complex networks, Nat. Phys. 5 (1) (2009) 74–80.
[3] Béla Bollobás, Random graphs, Cambridge University Press, 2001.
[4] Dorwin Cartwright, Frank Harary, Structural balance: A generalization of Heider's theory, Psychol. Rev. 63 (5) (1956) 277.
[5] Sergio Currarini, Matthew O. Jackson, Paolo Pin, An economic model of friendship: Homophily, minorities, and segregation, Econometrica 77 (4) (2009) 1003–1045.
[6] James A. Davis, Structural balance, mechanical solidarity, and interpersonal relations, Am. J. Sociol. 68 (4) (1963) 444–462.
[7] Peter M. DeMarzo, Dimitri Vayanos, Jeffrey Zwiebel, Persuasion bias, social influence, and unidimensional opinions, Q. J. Econ. 118 (3) (2003) 909–968.
[8] Tyler Derr, Charu Aggarwal, Jiliang Tang, Signed network modeling based on structural balance theory, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 557–566.
[9] Benjamin Enke, Florian Zimmermann, Correlation neglect in belief formation, Rev. Econ. Stud. 86 (1) (2019) 313–332.
[10] Santo Fortunato, Darko Hric, Community detection in networks: A user guide, Phys. Rep. 659 (2016) 1–44.
[11] Mirta Galesic, Henrik Olsson, Jonas Dalege, Tamara van der Does, Daniel Stein, Integrating social and cognitive aspects of belief dynamics: towards a unifying framework, J. R. Soc. Interface 18 (176) (2021).
[12] Nika Haghtalab, Matthew O. Jackson, Ariel D. Procaccia, Belief polarization in a complex world: A learning theory perspective, Proc. Natl. Acad. Sci. 118 (19) (2021).
[13] Frank Harary, On the notion of balance of a signed graph, Michigan Math. J. 2 (2) (1953) 143–146.
[14] Frank Harary, A structural analysis of the situation in the Middle East in 1956, J. Conflict Resolut. 5 (2) (1961) 167–178.
[15] Nicholas M. Harrigan, Giuseppe J. Labianca, Filip Agneessens, Negative ties and signed graphs research: Stimulating research on dissociative forces in social networks, Soc. Networks 60 (2020) 1–10.
[16] Matthew O. Jackson, Social and economic networks, Princeton University Press, 2010.
[17] Matthew O. Jackson, The human network: How your social position determines your power, beliefs, and behaviors, Vintage (2019).
[18] Anna Keuchenius, Petter Törnberg, Justus Uitermark, Why it is important to consider negative ties when studying polarized debates: A signed network analysis of a Dutch cultural controversy on Twitter, PLoS ONE 16 (8) (2021) e0256696.
[19] Alec Kirkley, George T. Cantwell, Mark E.J. Newman, Balance in signed networks, Phys. Rev. E 99 (1) (2019) 012320.
[20] Jürgen Lerner, Structural balance in signed networks: Separating the probability to interact from the tendency to fight, Soc. Networks 45 (2016) 66–77.
[21] Yury A. Malkov, Dmitry A. Yashunin, Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs, IEEE Trans. Pattern Anal. Mach. Intell. 42 (4) (2018) 824–836.
[22] Matúš Medo, Manuel S. Mariani, Linyuan Lü, The fragility of opinion formation in a complex world, Commun. Phys. 4 (1) (2021) 1–10.
[23] Pedro Mercado, Jessica Bosch, Martin Stoll, Node classification for signed social networks using diffuse interface methods, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2019, pp. 524–540.
[24] Mark E.J. Newman, Networks, Oxford University Press, 2018.
[25] F. Javier Ortega, José A. Troyano, Fermín L. Cruz, Carlos G. Vallejo, Fernando Enríquez, Propagation of trust and distrust for the detection of trolls in a social network, Comput. Netw. 56 (12) (2012) 2884–2895.
[26] F. Rabbani, Amir H Shirazi, G.R. Jafari, Mean-field solution of structural balance dynamics in nonzero temperature, Phys. Rev. E 99 (6) (2019) 062302.
[27] Pengfei Shen, Shufen Liu, Ying Wang, Lu. Han, Unsupervised negative link prediction in signed social networks, Mathematical Problems in Engineering (2019).
[28] Michael Szell, Stefan Thurner, Measuring social dynamics in a massive multiplayer online game, Soc. Networks 32 (4) (2010) 313–329.
[29] Jiliang Tang, Charu Aggarwal, Huan Liu, Node classification in signed social networks, in: Proceedings of the 2016 SIAM International Conference on Data Mining, SIAM, 2016, pp. 54–62.
[30] Jiliang Tang, Yi Chang, Charu Aggarwal, Huan Liu, A survey of signed network mining in social media, ACM Comput. Surv. 49 (3) (2016) 1–37.
[31] Arnout Van de Rijt, The micro-macro link for the theory of structural balance, J. Math. Sociol. 35 (1–3) (2011) 94–113.
[32] Xuehua Zhao, Bo Yang, Xueyan Liu, Huiling Chen, Statistical inference for community detection in signed networks, Phys. Rev. E 95 (4) (2017) 042313.
[33] Dengyong Zhou, Jiayuan Huang, Bernhard Schölkopf, Learning from labeled and unlabeled data on a directed graph, in: Proceedings of the 22nd International Conference on Machine Learning, 2005, pp. 1036–1043.