Improving Judgment Reliability in Social Networks via Jury Theorems

Paolo Galeazzi[1] and Rasmus K. Rendsvig[1] and Marija Slavkovik[2]

[1] Center for Information and Bubble Studies, University of Copenhagen
pagale87@gmail.com, rasmus@hum.ku.dk
[2] Department of Information Science and Media Studies, University of Bergen
marija.slavkovik@uib.no

Abstract. Opinion aggregators—such as 'like' or 'retweet' counters—are ubiquitous on social media platforms and often treated as implicit quality evaluations of the entry liked or retweeted, with higher counts indicating higher quality. Many such aggregators are poor quality evaluators as they allow disruptions of the conditions for positive wisdom-of-the-crowds effects. This paper proposes a design of theoretically justified aggregators that improve judgment reliability. Interpreting states of diffusion processes on social networks as implicit voting scenarios, we specify procedures for isolating sets of independent voters in order to use jury theorems to quantify the reliability of network states as quality evaluators. As real-world networks tend to grow very large and independence tests are computationally expensive, a primary goal is to limit the number of such tests. We consider five procedures, each trading a degree of reliability for efficiency, the most efficient requiring a low-degree polynomial number of tests.

1 Introduction

Web2.0 introduced the possibility for internet consumers to also become content creators [6]. This change led to an unprecedented amount of information being shared online, but also to a deterioration in the quality control of the information that is being shared. In some areas, such as medicine, the quality of the information online has been a now long-standing concern (e.g., [7, 11]), leading to the development of a code of conduct and certification standards [12]. Recently, we are witnessing the problem of low quality or even damaging content spreading in many areas, with the problem of so-called "fake news" being particularly worrisome [14, 18].

With the content quality problem being an all-topic concerning issue, the challenge is to find a way to automatically separate the reliable content from bad quality, unreliable, offensive and even illegal content. As a possible improvement of the current situation, here we propose that a good use of the consumers' behavior can help ascertain the quality of a post.

A ubiquitous feature of social media platforms is the ability of users to directly express their impressions and opinions about media content making its rounds on the platform. Beyond the option of reporting content for violations of law or community standards, there are three main, widely available channels for such expressions. One is by reaction buttons such as Twitter and Instagram's ♡ buttons, Reddit or YouTube's up- and downvotes, or Facebook's six choices of emoticons. The second is by textual reply, such as Twitter's reply option or Facebook's comments. The third, and the focus of this paper, is by sharing the content with one's social network, as e.g. by Twitter's retweet option.

In this paper, we assume that the aforementioned "judgment" expressions—and especially the decision on whether to further propagate content through sharing—may be seen as quality signifiers, while remaining aware that this is not necessarily always the case in social media, where an upvote may be a herding reaction [19] and sharing (or not) may be influenced more by emotional response than by sound quality judgment [5].

Interpreting decisions to share or not as quality signifiers, we aim to design procedures which aggregate such judgment expressions into a reliable collective quality judgment through majority rule. Yet, a collection of judgment expressions is not sufficient to provide a good and reliable collective judgment. Some users may not be competent at ascertaining content quality, if for no other reason than by not being human [10]. Even when judgment expressions are from competent users, the majority verdict cannot simply be taken as a theoretically sound quality assessment: due to herding effects or the mimicking of salient users, influenced signals may stop carrying information about a competent quality judgment but opaquely repeat already accounted for signals. Such dependencies between judgments invalidate the premises of the Condorcet Jury Theorem [8] and its many generalizations (e.g., [4, 13, 17, 20]) and thus disrupt the theoretical foundations of positive wisdom-of-the-crowds effects. When dependent signals are present, majority polling can no longer be trusted to reach the correct evaluation with high probability. To reestablish the positive wisdom-of-the-crowds effects of jury theorems, a main focus of the paper is on the identification of sets of independent users, whose majority vote may be trusted to be correct with high probability.

Our main contribution is the new approach to providing a quality evaluation of a post, given the current state of a diffusion process of the post in a social network. We suggest to obtain quality evaluations by considering a subset (the jury) of agents that have been exposed to the post through its diffusion in the network. The majority vote in the jury together with its correctness probability then constitutes the quality evaluation of the post (Sec. 3). How to select a jury among all agents exposed to the post is the main consideration of the paper. For simple jury theorems to apply (Sec. 3), the jury must be independent. Independence may be established statistically (Sec. 4), given a presumed set of diffusion processes serving as background data (Sec. 2). Yet, independence tests are computationally expensive: as real-world networks tend to grow large, the number of tests required to find a satisfactory jury should ideally be minimized for efficiency. In Sec. 5, we consider five jury selection procedures, each trading a degree of reliability for efficiency, and analyze the number of required tests of each of them, ranging from being exponential in the size of the set of exposed agents to being low-degree polynomial.

In designing the procedures, we need to make several choices regarding our priorities. Ideally, we desire a quality evaluation procedure that *i*) is theoretically justifiable, *ii*) is computationally efficient, *iii*) uses only data endogenous to the social network on which it is implemented, and *iv*) is conservative in its estimates, so that in choosing between two evils, we rather want a measure overly cautious with respect to estimating an entry as high quality than *vice versa*. The last point entails that we rather ignore the voice of a juror that might be beneficial than include the vote of a juror that may be misleading. Finally, in this paper, we prioritize *i*) over *ii*), retaining the use of expensive independence tests. We hope future work on such problems will improve efficiency.

2 Social Networks and Diffusion

A **social network** is specified by a set of links $N \subseteq \mathcal{A}^2$ connecting finitely many **agents** $\mathcal{A}$. We interpret $(a,b) \in N$ to mean that a sees the content shared by b. Let $N(a) = \{b \in \mathcal{A} : (a,b) \in N\}$. We assume no properties of N.

In the following, each network N is associated with a **topic**, identified with a set of **entries** $E = \{e_1, e_2, ...\}$. We think of N as used by the agents to discuss the topic E with the desire to evaluate the quality of each entry. We assume that entries bear an objective quality, or truth value, given by an exogenous **quality valuation** $V : E \to \{\bot, \top\}$, unknown to the agents. We interpret agents as casting votes on the quality of each entry. Given an entry, $\top_a$ is the event that agent a votes for the entry being of high quality, and $\bot_a$ that a votes for it being of low quality. As it is common in jury theorems, under the assumption of independence, each agent $a \in \mathcal{A}$ is assumed to have the same individual **correctness probability** $c \in [0,1]$ in their assessment, formally expressed in terms of the following conditional probabilities: $\forall a \in \mathcal{A}$,

$$p(\top_a \mid \top) = p(\bot_a \mid \bot) = c = 1 - p(\top_a \mid \bot) = 1 - p(\bot_a \mid \top).$$

Remark 1. The homogeneous correctness probability is a fundamental assumption for the jury theorem applied throughout, and deserves comment. First, we use homogeneous correctness probabilities for simplicity: jury theorems heterogeneous correctness probabilities also exist—see e.g. [13]. The procedures introduced below assume knowledge of the correctness probability of the individuals, and that this correctness probability is homogeneous for all users. In this respect, we remark that knowledge of the correctness probability c may be obtained through lab experiments where single individuals are tested in isolation and separately from the network interaction, just as individual utility functions in economics are elicited through lab tests, separately from interactive game-theoretic scenarios (see e.g. [21]). This would provide the aforementioned assumptions with testable behavioral foundations. The details of this process are outside of the scope of this work.

To describe the information flow through the network, define a **state** s of N as a pair (e, L_i) with e an entry and $L_i : \mathcal{A} \to \{\mathtt{S}, \mathtt{R}, \mathtt{U}\}$ a labeling map. Throughout, let $S_i := L_i^{-1}(\mathtt{S})$, $R_i := L_i^{-1}(\mathtt{R})$ and $U_i := L_i^{-1}(\mathtt{U})$. Agents in S_i have chosen to *share* e, those in R_i have been *reached* by e (e.g., by having a neighbor in S_i), and agents in U_i are *unreached* by e. An **initial state** $s_0 = (e, L_0)$ satisfies $S_0 = \emptyset$. A state $s_i = (e, L_i)$ can then **transition** to state $s_j = (e', L_j)$ if and only if

1. $e' = e$
2. $S_i \subseteq S_j$ and $S_j \backslash S_i \subseteq R_i$
3. $R_j = (R_i \backslash S_j) \cup \{a \in U_i : \exists b \in S_j \text{ and } b \in N(a)\}$
4. $U_j = U_i \backslash R_j$

I.e.: 1. the entry stays fixed, 2. agents never un-share and only reached agents can start sharing, 3. sharing agents stop being reached, but agents with a sharing neighbor become reached, and 4. else agents remain unreached.

A **diffusion process** is a sequence of states $d = s_0, s_1, ...$ such that i) s_0 is initial, ii) all other states are transitions from the previous state, and iii) if $s_k = s_{k+1}$, then $s_k = s_{k+n}$ for all $n \in \mathbb{N}$. These transition rules ensure that any diffusion process reaches a fixpoint s_k with $s_k = s_{k+n}$ for all $n \in \mathbb{N}$, called the **terminal state** of d.

Proposition 2. *For any network N, for any initial state s_0, any diffusion process $d = s_0, s_1, ...$ reaches a fixpoint.*

Proof. For any two states $s_k = (e, L_k), s_{k+1} = (e, L_{k+1})$ for which s_k can transition to s_{k+1}, points 2. and 4. of ensure that $S_k \subseteq S_{k+1}$ and $U_{k+1} \subseteq U_k$. Hence the proposition follows as N is finite.

We interpret a state in a diffusion process as an implicit, possibly partial, cast of votes. For a state $s_n = (e, L_n)$, $n \geq 1$, we take the set of agents that have voted on the quality of e to be the **jury** $J_n = R_{n-1} \cup S_n$ at time n. The jury J_n does not include newly reached agents $R_n \backslash (R_{n-1} \cup S_n)$ as they have not yet had the opportunity to choose whether to share or not. The jury J_n gives rise to a **voting profile** $(v_a)_{a \in J_n}$ with $v_a = \bot_a$ if $a \in R_n$ and $v_a = \top_a$ if $a \in S_n$. Hence, the agents that have been reached in previous states but have not shared vote for the low quality of e, while those that have shared vote for the high quality. Agents in U_n are excluded from the jury: unexposed to e, they have not had the chance to share. We use lower-case j's to refer to jury cardinality, such that $j = |J|, j_n = |J_n|$, etc., for J, J_n given by context. We assume diffusion processes are observed, so we know the progress of each entry.

The model makes at least the following idealized assumptions: All agents i) pay attention to all entries that reach them, and ii) to the best of their ability decide to share or not to with the only aim of proliferating high quality content. For i), moving agents from U to R could be done based on logged screen activity, as social media users may not always pay attention to all content shared by neighbors. The approach suggested here is not apt to extract reliable information if one does not assume ii).

3 Voting and a Jury Theorem

Jury theorems provide a mathematical argument for larger groups being beneficial in collective decision making. Under some assumptions on correctness probability ($c > 1/2$) and on independence of the voters, the majority vote in larger groups is probabilistically more accurate than in smaller groups (see e.g. [2, 3, 13]). Accepting social network diffusions as implicit voting scenarios and aiming to extract information on the quality of the entries, jury theorems thus constitute a natural basis for the selection of informative juries.

Given a quality evaluation $V(e)$ and a voting profile $(v_a)_{a\in J}$, for conciseness we write $v_a = 1$ if agent a's vote is correct about e (i.e., if $V(e) = \top$ and $v_a = \top_a$, or $V(e) = \bot$ and $v_a = \bot_a$), and $v_a = 0$ otherwise. For a state s with jury J, the **majority vote** is whichever of $\bot$ and $\top$ that gets more votes (or, in case of a tie, either $\bot$ or $\top$ is chosen by a fair coin toss). The majority vote of voting profile $(v_a)_{a\in J}$ on entry e is **correct** if it coincides with the quality valuation $V(e)$.

At a state s, if all j jurors in J vote independently, the probability that the majority vote is correct is given by $M(j)$ below, where the first term captures the tie-breaking rule, and the second is the probability of correctness of a strict majority.

$$M(j) = \frac{\ell}{2}\binom{j}{\frac{j}{2}}c^{\frac{j}{2}}(1-c)^{\frac{j}{2}} + \sum_{k=m_j}^{j}\binom{j}{k}c^k(1-c)^{j-k}$$

with $\ell = 0$ and $m_j = (j+1)/2$ for j odd, and $\ell = 1$ and $m_j = j/2+1$ for j even.

The simplest jury theorems, to which we stick here, concern the probability that a group makes the correct decision under *majority rule*. One statement of the classic ***Condorcet Jury Theorem*** is: if $c > 1/2$ and all jurors vote independently, then

1. the probability of a correct majority vote goes to one as the jury size goes to infinity: $\lim_{j\to\infty} M(j) = 1$.
2. the probability of a correct majority vote increases under the addition of two jurors: for $j + j'$ with $j' = (0 \mod 2)$, $M(j) < M(j + j')$.

Points 1 and 2 are sometimes referred to as the *asymptotic* and the *non-asymptotic* part of the theorem, respectively, with the former possibly taken to show that huge groups are infallible, while the latter shows that larger groups are better truth-trackers than small ones [9]. Note that the addition of two jurors is essential in the non-asymptotic part: moving from an odd to an even jury by adding a single juror may cause a drop in the probability of correctness due to the tie-breaking rule. As customary, in the following we simplify matters by limiting attention to juries of odd size.

Similar jury theorems also exist for juries which exhibit patterns of dependence and correlation among the voters (see [4,15,20]). Using independent juries for quality assessments may thus cause an information loss: an independent set of jurors can possibly be extended by the addition of dependent jurors while improving the majority vote precision. An alternative to our approach of seeking independent juries is thus to look for juries that exhibit patterns of correlation which do not negatively affect the majority correctness probability. However, the approaches of [4] and [15] requires to calculate the majority correctness probability by an expression exponential in jury size (The results of [20] are not applicable, being for the asymptotic case.) For this reason, we here make use of the simpler jury theorem requiring independent juries and suggest to use the majority vote of an independent jury as quality evaluation, presented together with its correctness probability.

4 Assessing Independence

Given a set of voters, whether their votes are independent and therefore warrant an application of the Condorcet Jury Theorem is an unobservable empirical matter. On this matter, we cannot supply a theoretical guarantee, but only hypothesize from other observables. One could hypothesize from network structure: If no voters in the set are connected, then conclude the set independent. Or one could hypothesize from personal or demographic traits: if all voters in the set are 'different enough', then conclude the set independent. These approaches cannot ensure independence: agents may be influenced by another through long chains in the network, and trait differences may lead to negative correlation in voting.

A third is a history-based, statistical approach: if the set of voters have not previously shown stochastically dependent behavior, then conclude the set independent. This approach proceeds via an independence test, e.g. the χ^2 test, to check if the voters' previously observed votes were stochastically independent.[3] We follow this approach.

The χ^2 test assesses how compatible some observed frequencies are with a theoretical probability distribution. In our case, the theoretical distribution is the distribution of votes obtained under the assumption of independent jurors. Having a network N with correctness probability c, and a set D of diffusion processes with terminal states $T = (t_1, ..., t_m)$ over the topic $E = \{e_1, ..., e_m\}$ valuated by V, we can then use the χ^2 test to compare the theoretical distribution with the observed distribution of votes in T to assess the stochastic independence of any subset of voters $J \subseteq \mathcal{A}$. Again, this does not guarantee independence: not rejecting the null hypothesis that the agents are independent does not prove that they are independent, but at least tells us that they cannot be shown dependent beyond any reasonable doubt (the P-value). The details follow.

Theoretical Distribution. In a terminal state t where $U = \emptyset$, the probability distribution p of vote profiles $v = (v_1, ..., v_n) \in \{0,1\}^{|\mathcal{A}|}$, given independent voters with correctness probability c and quality valuation V, is

$$p(v_1, ..., v_n) = \prod_{i \in \mathcal{A}} x_i \text{ with } x_i = \begin{cases} c & \text{if } v_i = 1 \\ 1 - c & \text{if } v_i = 0. \end{cases}$$

[3] We use the classic χ^2 test just to exemplify our procedure, but other alternatives are also possible, e.g. the G-test.

This distribution does not apply when $U \neq \emptyset$, since agents in U are just unreached by the entry and have not had the chance to vote. For such cases, we have to find the appropriate theoretical distribution by taking the marginal of p on the set of exposed agents $S_t \cup R_t$. For a subset of voters $J \subseteq \mathcal{A}$, let $T_J \subseteq T$ be the set of terminal states such that $J \subseteq S_t \cup R_t$ for all $t \in T_J$. Then the marginal on J of p is

$$p_J(v_J) = \sum_{v_{-J} \in \{0,1\}^{|\mathcal{A} \setminus J|}} p(v_J, v_{-J})$$

with $v_J = (v_i)_{i \in J}$ and $v_{-J} = (v_i)_{i \in \mathcal{A} \setminus J}$. The distribution p_J hence gives the probability that one should expect, under the assumption of independent voters, on the votes by the agents in J who are jointly exposed to the entries in T_J.

The χ^2 Test for Independence. Given the theoretical distribution p, for each subset $J \subseteq \mathcal{A}$, we can then test for independence between its members. Running a χ^2 test amounts to the following: 1. Select a significance (e.g., .1, .05, .01) for rejecting the null hypothesis that the tested variables (votes) are dependent. 2. For each outcome i, $0 \leq i \leq n$, find the number E_i of occurrences estimated by the theoretical distribution given the bounds set by the data. 3. Compare E_i to the number O_i of observed occurrences of i by finding the χ^2 statistic:

$$\chi^2 := \sum_{i=0}^{|J|+1} \frac{(O_i - E_i)^2}{E_i}$$

4. Compare χ^2 to the upper-tail critical values of the χ^2 distribution for the selected significance level and the appropriate number of degrees of freedom, and reject the null hypothesis if the P-value is less than the chosen significance level.

[1] shows that there exists an algorithm for testing the independence of j random variables with time complexity

$$O\left(\left(\prod_{i=1}^{j}[i]\right)^{1/2} + \sum_{i=1}^{j}[i]\right)$$

where $[i]$ denotes the number of possible values that random variable i can take. The exponential complexity arises as a direct consequence of the exponential growth of the number outcomes, as going through the whole outcome space is necessary to assess the independence of a given set of random variables. In our social network scenario, the (votes of the) agents in jury J are the random variables whose mutual independence we are interested in, and the number of possible outcomes of the voting process among the agents in J is the number of possible voting profiles, $2^{|J|}$. Therefore, given the result in [1], the time complexity in our case is $O\left(2^{|J|/2} + 2|J|\right)$.

With independence tests being computationally expensive, we cannot readily design a procedure that both checks independence and is effective: any procedure based on the current state-of-the-art will be exponential in the size of the jury. We then look for procedures that limit the required number of tests.

Remark: Family-wise Error and Bonferroni Correction. Seeking procedures that limit the required number of tests highlights the expectation that it will often be necessary to run multiple tests before settling on a jury. This implies that we cannot fix a general significance level to be used in all tests, for the risk of committing type I errors. E.g. setting a significance level of 0.1 while running 1000 tests makes the family-wise error rate 1, theoretically guaranteeing one false positive—one jury deemed independent when it is not.

To control the family-wise error rate, we throughout assume the significance level of the individual tests is adjusted using **Bonferroni correction**: with an overall desired significance of α, Bonferroni correction tests the individual hypotheses at α/m, with m the total number of hypotheses to be tested. As each of the procedures discussed below specifies such an m and as the significance level does not affect the complexity of a χ^2-test, Bonferroni correction does not affect the overall test complexity.

Compared to alternative methods, Bonferroni correction fits present purposes well: First, it is conservative in its estimates, fitting well with the conservatism of point *iv*) from the introduction. Second, it is computationally trivial, compared e.g. to the stronger *Holm-Bonferroni method* which requires finding all m P-values prior to running any tests.

5 Jury Finding Problems

Given a network state s_n with jury J_n, we seek to present the users with a conservative estimate of the quality of the current entry e in s_n. Due to possible dependencies between jurors, the majority decision of J_n need not be as trustworthy as stochastic independence would imply. Therefore, we look for subsets $J^* \subseteq J_n$ that retain independence. However, we also seek to constantly improve correctness of the assessment that we can extract from the agents as the information flows through the network. To this end, we look for a jury J^* in relation to a jury $J' \subseteq J_{n-1}$ assumed found independent in the previous round $n-1$. For the sake of brevity, we refer to J_n, J' and J^* with the above properties implicit throughout this section.

5.1 Optimal Juries

The optimal choice of J^* is any largest set of independent agents that we can find among those in J_n, as this maximizes collective correctness probability. An optimal jury may be found by solving the following problem for increasing values of correctness probability, C:

Problem 3. C**-precise Jury** is the decision problem
Instance: A social network $(\mathcal{A}, N)$, its network state s_n with jury J_n, a jury J' for state s_{n-1} and a number $C \in [c, 1]$.
Question: Is there an independent subset $J^* \subseteq J_n$ such that $M(J^*) \geq C$?

Finding an optimal choice of J^* corresponds to finding a solution to the optimization version of C-precise Jury. Using binary search, we need only solve the decision problem for $O(\log(j_n - j'))$ different values of C, as j' is a lower bound on the jury size of interest and the number of possible values of C is limited to $\{M(k) \colon 1 \leq k \leq j_n\}$.

The C-precise jury problem is NP-hard, as may be shown by reduction to the *Independent Set problem*, cf. e.g. [16]. Given a graph $G = (V, E)$, call a set $A \subseteq V$ `independent`* if no $v, v' \in A$ are connected by an edge in E. Then:

Problem 4. **Independent Set** is the decision problem
Instance: A graph $G = (V, E)$, and an integer $k \leq |V|$.
Question: Does G contain an `independent`* set of size at least k?

Proposition 5. *C-precise Jury is NP-complete.*

Proof. We show hardness by reduction to Independent Set which is NP-hard cf. e.g. [16]. Identify the graph (V, E) with a network $(\mathcal{A}, N)$ given by $\mathcal{A} = V$ and $N = E$. Assume a non-initial diffusion state s_n of $(\mathcal{A}, N)$ such that $J_n = \mathcal{A}$. We make no special use of J', so let $J' = \{a\}$ for some $a \in J$. Checking that an `independent`* set of G of size k exists is a special case of checking whether a jury J^* with $M(j^*) \geq C$ exists. Let $C = M(k)$. It is always possible to find a probability distribution p on voting profiles $(v_i)_{i \in \mathcal{A}}$ such that two agents i,j are not independent under p if and only if they are connected by an edge in N, and where p additionally satisfies that for all $A \subseteq \mathcal{A}$, if all elements of A are pairwise independent, then A' is mutually independent. For such p, `independence`* in $(\mathcal{A}, N)$ implies mutual independence in $(\mathcal{A}, N)$. As A can only be mutually independent if it is pairwise independent, mutual independence also implies `independence`*. Hence an independent jury $J^* \subseteq J_n$ satisfying $M(j^*) \geq C$ is also an `independent`* set of size at least k. *Inclusion:* A simple guess and check algorithm can be constructed: if we guess a subset $J^* \subseteq J_n$, we can check in polynomial time whether $M(J^*) \geq C$.

Finding an optimal jury at state s_n requires finding the maximal C for which the C-precise Jury is solved in the positive. A naive brute force algorithm examining every candidate subset $J \subseteq J_n, j > j'$, in the worst case requires running $2^{j_n - 1}$ independence tests. By the result of [1], finding a optimal jury is in $O\left(2^{j_n-1} \cdot \left(2^{j_n/2} + 2j_n\right)\right)$. The naive aspect of this complexity may be diminished be the algorithm of [22] which finds a maximum `independent`* set in $O(1.1996^{j_n})$ using polynomial space.

5.2 Error-Diminishing Juries

Due to the exorbitant number of independence tests required, the optimality of the outcome jury has to be foregone in the interest of computational efficiency. A first alternative is to look for improvements in the collective correctness probability, without aiming to identify the best possible jury. In the next three subsections we follow this path.

One option in this direction is to seek a jury that diminishes the error of the current jury by a given percentage:

Definition 6. *A jury $J^* \subseteq J_n$ is **error-diminishing** by $h\%$ with respect to $J' \subseteq J_{n-1}$ if its probability of an incorrect majority vote is $h\%$ lower than that of J'. I.e., if $1 - M(j^*) \leq \frac{h}{100}(1 - M(j'))$.*

In finding an error-diminishing jury J^*, the required size depends on the size of J', as well as on the correctness probability c and the increment h. The number of necessary tests, however, has not been reduced much by this approach:

Proposition 7. *Whether a jury J^* error-diminishing by $h\%$ exists can be determined by testing at most $\binom{j}{\frac{j+1}{2}}$ subsets for independence. Worst case, none of these tests are redundant.*

Proof. With a fixed error-diminishing degree $h\%$, there is some $j^* \in \mathbb{N}$ given as a function of the size of J' such that only juries of size at least j^* will be precise enough. It then suffices to seek through the $\binom{j}{j^*}$-many size j^* subsets of J for an independent jury: no smaller sets will do, and every larger set will be non-independent if all size j^* are. With $h^* := 1 - \frac{h}{100}(1 - M(j'))$, this is the smallest j^* such that $h^* \leq M(j^*)$. This j^* may be approached from j' using binary search, but checking higher values is more expensive: the inequality needed checked for a value m has a fixed left-hand side, but a right-hand side increasing linearly in m. Finding j^* is thus in $O(j)$. Second, we seek for a suitable $J^* \subseteq J$. Worst case, j^* is $\arg\max \binom{j}{x} = \{\frac{j-1}{2}, \frac{j+1}{2}\}$, providing an upper bound of $\binom{j}{\frac{j+1}{2}}$ tests before concluding. The lower bound is established by the worst case where each of the $\binom{j}{\frac{j+1}{2}}$ subsets may be non-independent due to just one agent. In this case, none of the tests are redundant.

For reference, we remark that $f(x) = \binom{x}{\frac{x+1}{2}}$ is not a slow-growing function. In fact, it grows as fast $a^x, a > 1$.

5.3 Incrementally Improved Juries

Since the required number of tests has not appreciably decreased by looking for error-diminishing juries, we turn to the alternative of improving the collective correctness probability by a fixed percentage. For a 5% increment, for instance, the number of required additional jurors is illustrated in Table 1.

	.60	.65	.70	.75	.80	.85	.90	.95
$c = .6$	1	3	7	11	17	27	41	65
$c = .75$	1	1	1	1	3	5	5	9

Table 1. The number of agents of correctness probability c (in the rows) needed to reach a certain collective correctness probability (in the columns). Even when the individual correctness c is relatively low, e.g. $c = .6$, only 65 independent jurors are needed to reach a collective correctness probability of 0.95.

Definition 8. *A jury $J^* \subseteq J_n$ is* ***incremental*** *by $h\%$ with respect to $J' \subseteq J_{n-1}$ if the probability of a correct majority vote is $h\%$ higher than that of J'. I.e., if $(1 + \frac{h}{100})M(j') \leq M(j^*)$.*

Note that while the error-diminishing requirement above may fail to be satisfied for some jury J' just because there is no sufficiently large independent subset of the currently exposed agents, finding a jury of improved correctness probability by a fixed increment may be impossible also because the current jury's correctness probability cannot be raised by $h\%$ without exceeding 1. One may therefore expect that this could reduce the search for improved juries as compared to the previous case. However, the required number of tests does not change:

Proposition 9. *Whether a jury J^* incremental by $h\%$ exists can be determined by testing at most $\binom{j}{\frac{j+1}{2}}$ subsets for independence. Worst case, none of these tests are redundant.*

Proof. With a fixed desired increment, there is some $j^* \in \mathbb{N}$ given as a function of the size of J' such that only juries of size at least j^* will be precise enough. As in the proof of Prop. 7, finding j^* is in $O(j)$ using binary search, but here we find the smallest j^* satisfying $h^* \leq M(j^*)$ for $h^* := (1 + \frac{h}{100})M(c, j')$ fixed. Again as in the proof of Prop. 7, it suffices to seek through the $\binom{j}{j^*}$-many size j^* subsets the same non-redundancy argument applies.

5.4 Monotonic Juries

Looking thus for even simpler approaches, the next possible simplification is to merely seek any improvement in the correctness probability of the jury. A corresponding formal requirement is then the following:

Definition 10. *A jury $J^* \subseteq J_n$ is* **monotonic** *with respect to $J' \subseteq J_{n-1}$ if the probability of a correct majority vote is strictly higher than that of J'. I.e., if $M(j') < M(j^*)$.*

Proposition 11. *Whether a jury J^* monotonic with respect to a fixed jury J' exists can be determined by testing at most $\binom{j}{j'+2}$ subsets for independence. When J' may grow with J, it can be determined by testing at most $\binom{j}{\frac{j+1}{2}}$ subsets for independence. Worst case, none of these tests are redundant.*

Proof. For a fixed size J', searching through juries of size $j^* = j' + 2$ is sufficient: if an independent jury J^* of size j^* is found, it will satisfy $M(c, j') < M(c, j^*)$; if no independent size j^* jury exists, then every larger set will also be non-independent. Testing each of the $\binom{j}{j'+2}$ size $j' + 2$ subsets may also be necessary, as each of the subsets may be non-independent due to just 1 agent. When J' may grow with J, the upper bound is established by the worst case number of tests, $\arg\max \binom{j}{x} = \{\frac{j-1}{2}, \frac{j+1}{2}\}$. The lower bound is established by the argument used for Prop. 9.

5.5 Inflationary Juries

A common cause for the large search space—and hence the many required independence tests—across the hitherto considered procedures is that they make little use of the jury J' assumed found in the previous step: only its *size* matters, used as a lower bound of the size of an improved jury J^*. The simplification we now propose is instead based on the idea of looking only for additions to the current jury J', rather than throwing it away and starting the search anew. In other words, rather than focusing on a general improvement in the correctness probability as we have done in subsections 5.1–5.4, one may instead focus exclusively on extensions of the current jury J'.

On the one hand, this approach may be considered the farthest from Section 5.1's search for an optimal jury, in that an unfortunate start might lead to a maximal independent set of agents much smaller than an *de facto* optimal jury. On the other hand, however, its convenience arises precisely from favoring simplicity over optimality. As we have seen in Table 1 above, small independent juries already suffice to achieve a high collective correctness probability, even when the individual correctness c is low. Given nowadays dimensions of social networks, a set of a few tens or hundreds of agents is but a minimal fraction of the total number of users. Hence, even when holding the actual independent jury J' fixed, one may reasonably hope to be able to find another pair of agents independent of J' among the many available users.

The corresponding requirement that the next jury J^* should satisfy is then the following:

Definition 12. *A jury $J^* \subseteq J_n$ is **inflationary** with respect to $J' \subseteq J_{n-1}$ if it extends J' and the probability of a correct majority vote is strictly higher than that of J'. I.e., if $J' \subseteq J^*$ and $M(j') < M(j^*)$.*

By the Condorcet Jury Theorem, to find an inflationary jury, it is sufficient to find a pair of agents $a, a' \in J \backslash J'$ such that $J' \cup \{a, a'\} = J^*$ is a jury of mutually independent agents. We are thus able to greatly reduce the required number of tests, as stated by the following:

Proposition 13. *Whether an inflationary jury J^* exists can be determined by testing at most $\binom{j-j'}{2}$ subsets for independence. Worst case, none of these tests are redundant.*

Proof. Given J', by the Condorcet Jury Theorem, it suffices to find J^* with $j^* = j' + 2$ and $J' \subseteq J^*$. There are $\binom{j-j'}{2}$ candidates of pairs to add to J'. Testing each is sufficient; testing each may also be necessary, as each may be non-independent due to just 1 agent.

As $\binom{n}{2} = \sum_{k=1}^{n-1} k = \frac{(n^2-n)}{2}$, the number of tests required to find an inflationary jury is bounded above by a degree-2 polynomial. The inflationary jury procedure thus considerably reduces the number of required tests.

6 Conclusion

We have considered how states of diffusion processes in social networks may be used as quality evaluations of shared content. We have noted that establishing independence of juries is essential to rely on wisdom-of-the-crowds results from jury theorems and to ensure a theoretically sound evaluation, but that independence testing is computationally expensive. For this reason, we have sought jury selection procedures that reduce the number of necessary independence tests.

Of the five selection procedures introduced here, only the inflationary jury procedure requires a number of tests bounded by a polynomial (of degree 2). While, given the current algorithms, the time complexity of testing for independence remains exponential in the size of the inflationary jury, we have also shown that there is hope for tractably using the inflationary jury procedure in practice, as the number of independent jurors needed to achieve a high collective correctness probability is, even for low individual correctness probability, rather small, cf. Table 1.

Several fundamental questions remain unexplored, and core elements may be chosen differently. One question pertains to the amount of data required to conduct the χ^2 tests. Highly competent voters will often vote alike, wherefore a large set of previous diffusion processes will be required to determine whether their voting pattern significantly differs deform from the theoretical distribution under independence. We do not know how this required data grows with competence, and it may thus introduce computational hindrances. Related is the use of the χ^2 test itself. Possibly, alternative statistical approaches may lead to stronger conclusions about independence. That field should be surveyed, with complexity issues in mind.

The results presented do not tell us much about the practical difficulty of the proposed approach. It could be informative to develop a a randomized algorithm, or applying reduction to SAT to use one the excellent SAT solving algorithms developed in recent years. Currently, we do not know if the problems posed in this paper are highly approximate, or exactly solvable for all practical problems.

Finally, it would be instructive to perform empirical evaluations of several aspects of the proposed approach, to gauge both its efficiency (cf. the above) and its necessity. As real-life social networks tend to grow large, the continuous and global observation assumed here may be unfeasible. Due to the large size of networks, it could also be the case that random sampling of users or other selection methods *de facto* provide a way to obtain a correct aggregated judgment with sufficiently high frequency. Empirical studies could thus be instructive in determining how to best improve judgment reliability in social networks via jury theorems.

Acknowledgments. We thank the reviewers of *LORI-VII* and the participants of the *Social Interactions in Epistemology and in Economics* conference (Cph, 29-31/5/2019) for insightful comments. The Center for Information and Bubble Studies is funded by the Carlsberg Foundation. RKR was partially supported by the DFG-ANR joint project *Collective Attitude Formation* [RO 4548/8-1].

References

1. Acharya, J., Daskalakis, C., Kamath, G.: Optimal Testing for Properties of Distributions. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R.

(eds.) Advances in Neural Information Processing Systems 28, pp. 3591–3599. Curran Associates, Inc. (2015)
2. Ben-Yashar, R., Paroush, J.: A nonasymptotic Condorcet jury theorem. Social Choice and Welfare **17**(2), 189–199 (2000)
3. Berend, D., Sapir, L.: Monotonicity in Condorcet Jury Theorem. Social Choice and Welfare **24**(1), 83–92 (2005)
4. Berend, D., Sapir, L.: Monotonicity in Condorcet's Jury Theorem with dependent voters. Social Choice and Welfare **28**(3), 507–528 (2007)
5. Berger, J., Milkman, K.L.: What Makes Online Content Viral? Journal of Marketing Research **49**(2), 192–205 (2012)
6. Brake, D.R.: Are We All Online Content Creators Now? Web 2.0 and Digital Divides*. Journal of Computer-Mediated Communication **19**(3), 591–609 (2014)
7. Chumber, S., Huber, J., Ghezzi, P.: A methodology to analyze the quality of health information on the internet: The example of diabetic neuropathy. The Diabetes Educator **41**(1), 95–105 (2015)
8. Condorcet, M.M.d.: Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix. Paris (1785)
9. Dietricht, F., Spiekermann, K.: Jury Theorems. In: Fricker, M., Graham, P.J., Henderson, D., Pedersen, N., Wyatt, J. (eds.) The Routledge Companion to Social Epistemology. Routledge (2019)
10. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Commun. ACM **59**(7), 96–104 (2016)
11. Gagliardi, A., Jadad, A.R.: Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. BMJ **324**(7337), 569–573 (2002), `https://www.bmj.com/content/324/7337/569`
12. Greenberg, L., D'Andrea, G., Lorence, D.: Setting the public agenda for online health search: a white paper and action agenda. Journal of Medical Internet Research **6**(2), e18 (2004)
13. Grofman, B., Owen, G., Feld, S.L.: Thirteen theorems in search of the truth. Theory and Decision **15**(3), 261–278 (1983)
14. Hendricks, V.F., Vestergaard, M.: Reality Lost: Markets of Attention, Misinformation and Manipulation. Springer Nature (2019)
15. Kaniovski, S.: Aggregation of correlated votes and Condorcet's Jury Theorem. Theory and Decision **69**(3), 453–468 (2010)
16. Kleinberg, J., Tardos, E.: Algorithm Design. Pearson (2005)
17. Ladha, K.K.: Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. Journal of Economic Behavior & Organization **26**(3), 353–372 (1995)
18. Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J., Zittrain, J.L.: The science of fake news. Science **359**(6380), 1094–1096 (2018)
19. Muchnik, L., Aral, S., Taylor, S.J.: Social Influence Bias: A Randomized Experiment. Science **341**(6146), 647–651 (2013)
20. Pivato, M.: Epistemic democracy with correlated voters. Journal of Mathematical Economics **72**, 51–69 (2017)
21. Wakker, P.: Prospect Theory. Cambridge University Press (2010)
22. Xiao, M., Nagamochi, H.: Exact algorithms for maximum independent set. Information and Computation **255**, 126–146 (2017)