EPFL

The Quest for Trusted Information

Karl Aberer

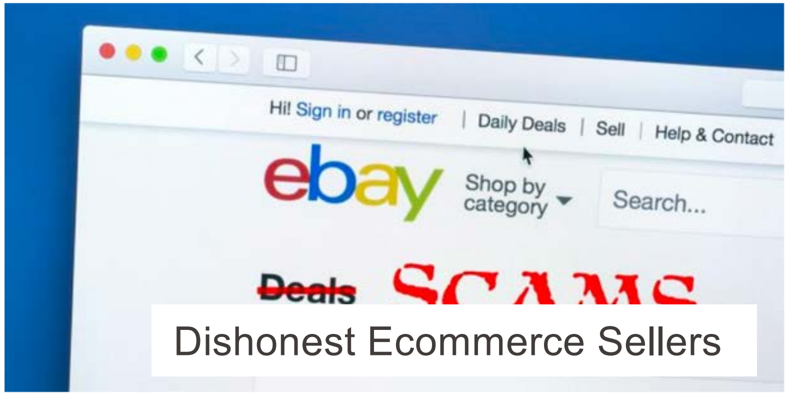
Celebrating and commemorating

Conference on Erwin Engeler's 90th birthday and Ernst Specker's centenary

21–22 February 2020

The Internet is a dangerous place

Fake News



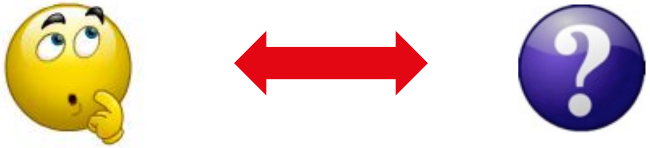
Dishonest Ecommerce Sellers

Unreliable Crowd-Workers

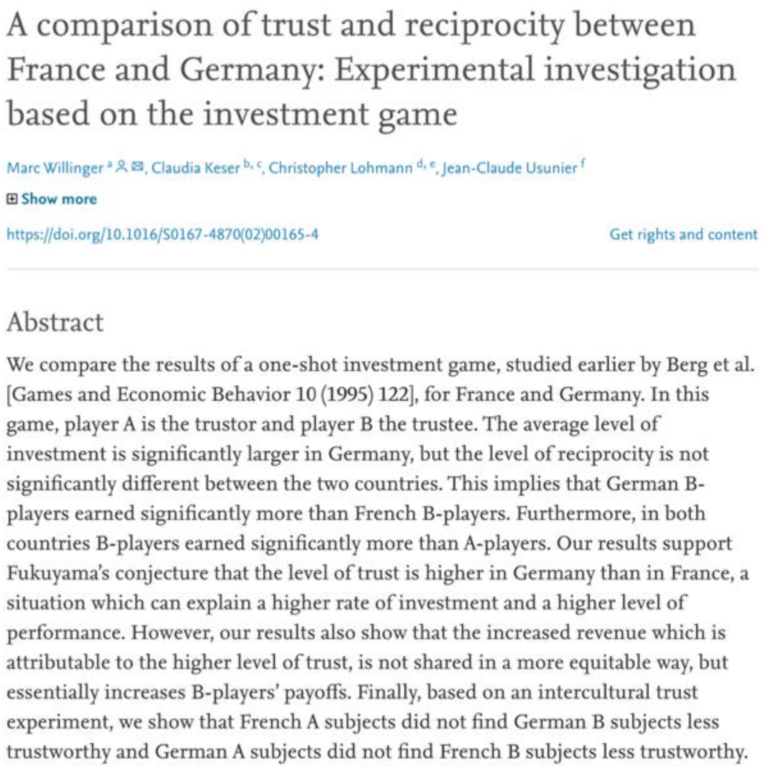
Trolls on Social Networks

Trust

- **Trust:** the extent to which an agent believes an other agent will cooperate
  - Gain is to be shared with the other agent
  - But the agent is exposed to a risk of loss



Is Trust Important?



A comparison of trust and reciprocity between France and Germany: Experimental investigation based on the investment game

Marc Willinger [a], Claudia Keser [b, c], Christopher Lohmann [d, e], Jean-Claude Usunier [f]

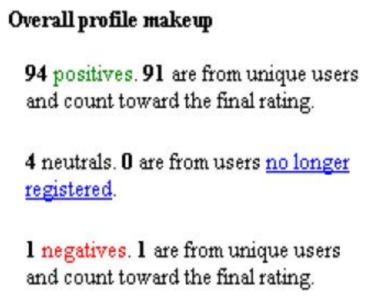
⊞ Show more

https://doi.org/10.1016/S0167-4870(02)00165-4 Get rights and content

Abstract

We compare the results of a one-shot investment game, studied earlier by Berg et al. [Games and Economic Behavior 10 (1995) 122], for France and Germany. In this game, player A is the trustor and player B the trustee. The average level of investment is significantly larger in Germany, but the level of reciprocity is not significantly different between the two countries. This implies that German B-players earned significantly more than French B-players. Furthermore, in both countries B-players earned significantly more than A-players. Our results support Fukuyama's conjecture that the level of trust is higher in Germany than in France, a situation which can explain a higher rate of investment and a higher level of performance. However, our results also show that the increased revenue which is attributable to the higher level of trust, is not shared in a more equitable way, but essentially increases B-players' payoffs. Finally, based on an intercultural trust experiment, we show that French A subjects did not find German B subjects less trustworthy and German A subjects did not find French B subjects less trustworthy.

- Example: eBay
  - eBay reputation profiles predictive of future performance [Resnick et al., 2002]
  - Prices positively correlated with the feedback [Melnik and Alm, 2002]

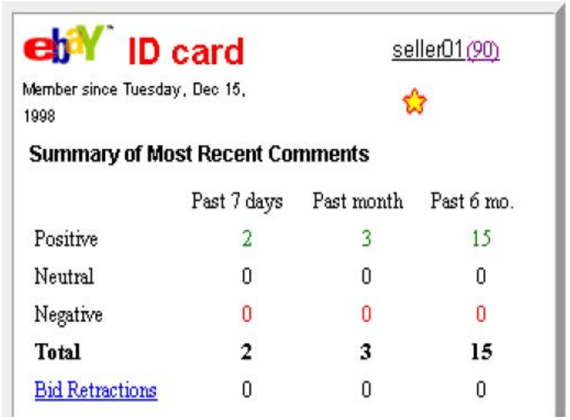


Overall profile makeup

94 positives. **91** are from unique users and count toward the final rating.

4 neutrals. **0** are from users no longer registered.

1 negatives. **1** are from unique users and count toward the final rating.



seller01(90)

Member since Tuesday, Dec 15, 1998

Summary of Most Recent Comments

	Past 7 days	Past month	Past 6 mo.
Positive	2	3	15
Neutral	0	0	0
Negative	0	0	0
Total	**2**	**3**	**15**
Bid Retractions	0	0	0

Trust Model: The way in which the trust belief is inferred

- Trust is a property of an agent A
  - i.e. label agent as trustworthy / not trustworthy

- Typical inferences could be
  1. If A cooperated with ME in the past, it will also cooperate with ME in the future
  2. If A cooperates with B then A will also cooperate with ME with high probability

- Basis of a large body of work on **reputation-based trust**
  - Building **predictive models** based on historical data (reports)
  - **Signalling** approach to trust management

Reputation-based Trust

- Requires
  1. that we obtain reports and
  2. that we are able to interpret them

- Problem 1: have no reports about an unknown agent
  - Solution: share reports in a community

- Problem 2: we cannot interpret the report
  - Solution: perform inferences on sets of reports

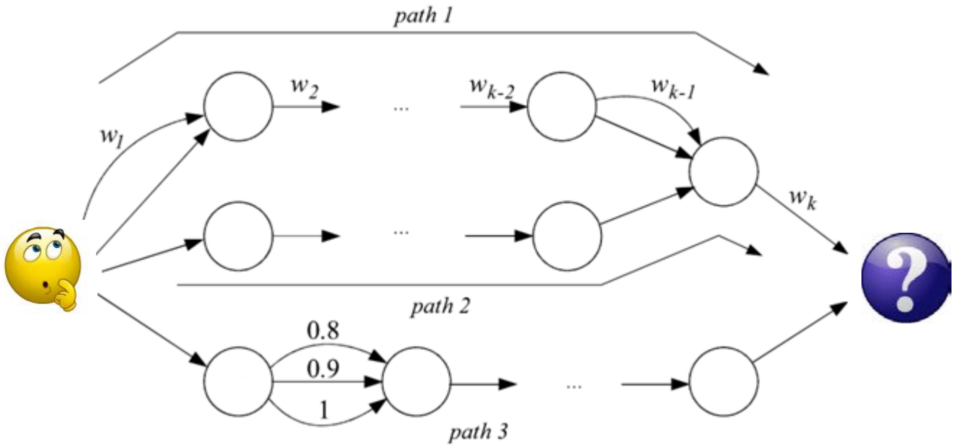
Overview

- Entrée: Community-based Trust (Problem 1)
- Main Course: Inference on Trust Signals (Problem 2)
  - Checking consistency
  - Learning from history
- Dessert: Rationality

Entrée:
Community-based Trust

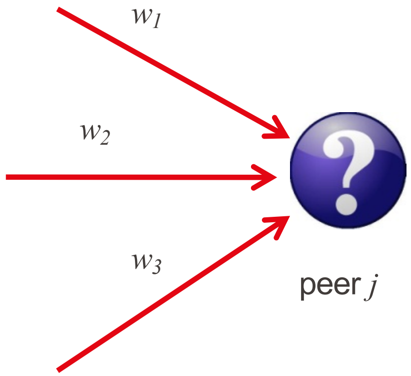
Model

- Original motivation: peer-to-peer systems
- Peers provide services to other peers
  - Know other peers they interacted with
  - Store feedback w_i on interactions with peer i
  - Feedback on service and reporting
  - A social network (trust graph)



Social Networking Approach

- Compute global trust values t_j for peer j
  - By aggregating feedbacks
  - Weighting by the trust in the recommender
  - Similar to PageRank computation



$$t_j = \sum_{i \in in(j)} w_i \frac{t_i}{\sum_{k \in in(j)} t_k}$$

Xiong, Li, and Ling Liu. "Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities." IEEE transactions on Knowledge and Data Engineering 16.7 (2004): 843-857.

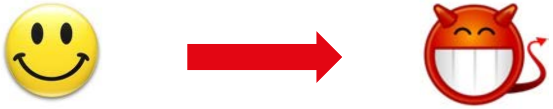
Social Networking Assessment

- Problems solved
  1. Automatic aggregation of reputation data
  2. Does not require central authority
  3. Possibility of misreporting considered
- Shortcomings
  - Costly (global) computation
  - Trust values have no further meaning but ranking
  - No distinction between propensity to provide poor service and to misreport

Probabilistic Estimation Approach

- Assume probabilistic peer behavior, e.g.
  - P[peer j performs service honestly] = θ_j
  - P[peer k lies when reporting] = l_k
- Probability of report y_k on peer j from peer k

$$P[Y_k = y_k] = \begin{cases} l_k(1-\theta_j) + (1-l_k)\theta_j & if\ \ y_k = 1 \\ l_k\theta_j + (1-l_k)(1-\theta_j) & if\ \ y_k = 0 \end{cases}$$



peer k
l_k = 0

report y_k = 0

peer j
θ_j = 0

Despotovic, Zoran, and Karl Aberer. "P2P reputation management: Probabilistic estimation vs. social networks." Computer Networks 50.4 (2006): 485-500.

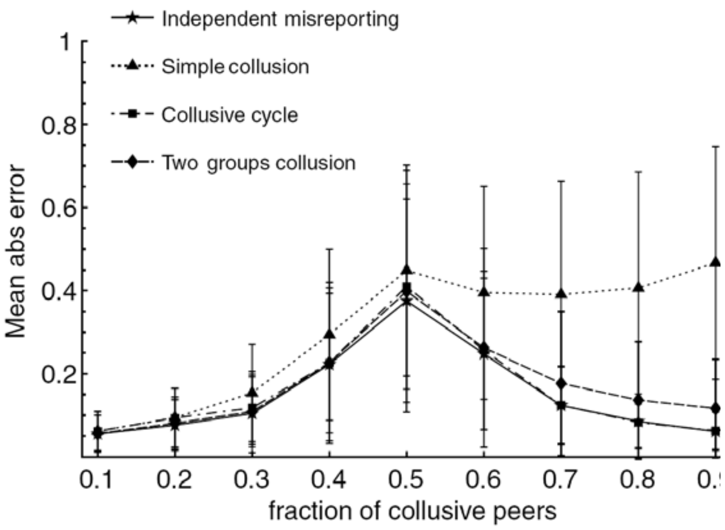
Probabilistic Estimation Approach

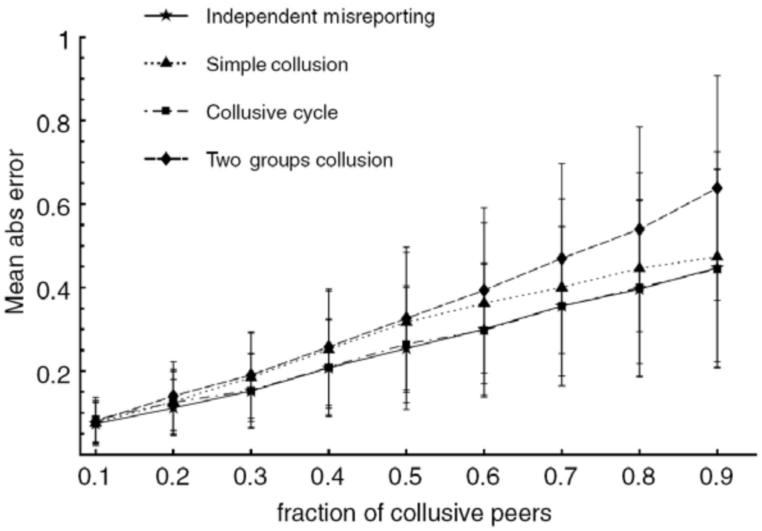
- Maximum Likelihood Estimation
  - Determine l_k by checking reports on own performance
  - Collect all reports $y_1, y_2, \ldots, y_n$ on peer j
  - Select θ_j that maximizes the probability to obtain the observed reports

$$L(\theta_j) = P[Y_1 = y_1]P[Y_2 = y_2]\cdots P[Y_n = y_n]$$

Comparative Evaluation

- Precision in correctly assessing misbehavior





Maximum Likelihood Estimation

Social Networking

Implementation Questions

- Central or distributed storage of trust reports
  - Cost of updating and retrieving reports

- Manipulation of stored data
  - Replication

Main Course:
Inference on Trust Signals

Chapter 1: Checking Consistency

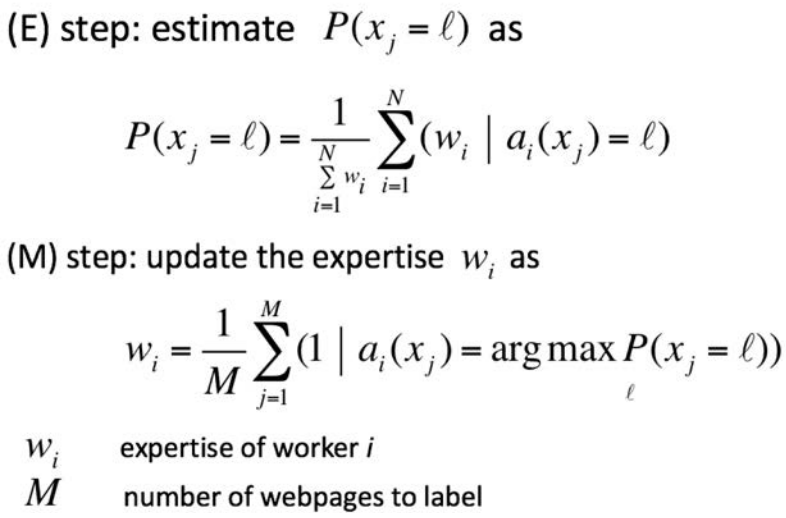
Typical Example: Crowd-Sourcing

- Crowd-Sourcing Platform
  - Crowd-workers label data

- Can the labels be trusted?
  - Use multiple labels to check their consistency
  - Assign trust value to crowd-workers

Classical Approach

- Expectation maximization
- Iterates in two steps
  1. E-Step: estimate the labels from the answers of workers
  2. M-Step: estimate the reliability of workers from the consistency of answers



(E) step: estimate $P(x_j = \ell)$ as

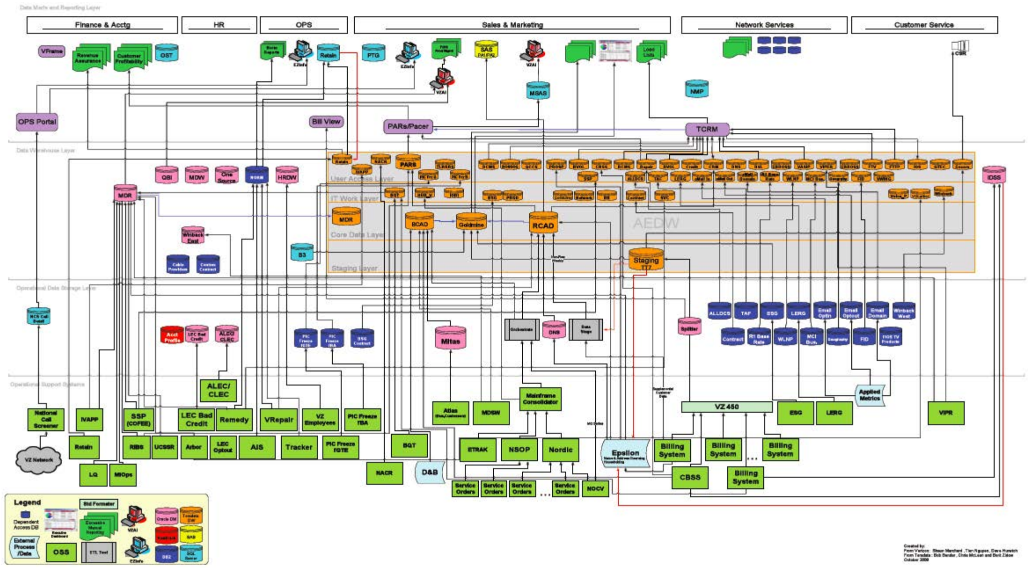
$$P(x_j = \ell) = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} (w_i \mid a_i(x_j) = \ell)$$

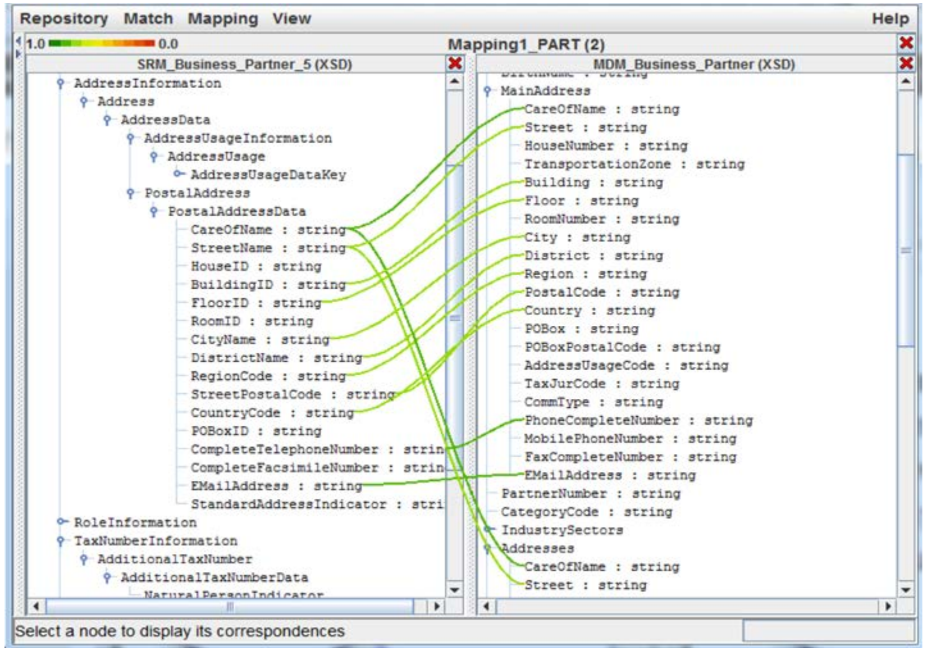
(M) step: update the expertise w_i as

$$w_i = \frac{1}{M} \sum_{j=1}^{M} (1 \mid a_i(x_j) = \arg\max_{\ell} P(x_j = \ell))$$

w_i	expertise of worker *i*
M	number of webpages to label

Data Integration





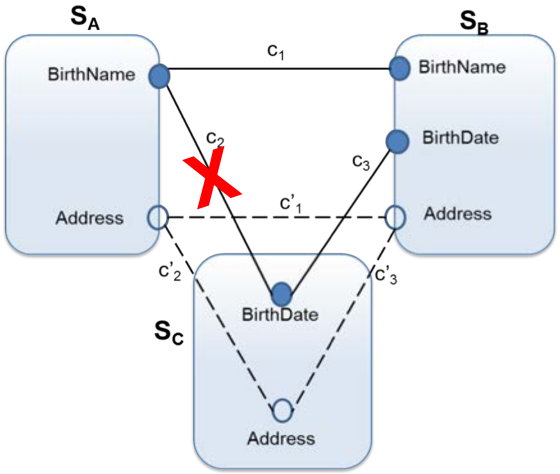
```
AddressInformation
  Address
    AddressData
      AddressUsageInformation
        AddressUsage
          AddressUsageDataKey
      PostalAddress
        PostalAddressData
          CareOfName : string
          StreetName : string
          HouseID : string
          BuildingID : string
          FloorID : string
          RoomID : string
          CityName : string
          DistrictName : string
          RegionCode : string
          StreetPostalCode : string
          CountryCode : string
          POBoxID : string
          CompleteTelephoneNumber : strin
          CompleteFacsimileNumber : strin
          EMailAddress : string
          StandardAddressIndicator : stri
RoleInformation
TaxNumberInformation
  AdditionalTaxNumber
    AdditionalTaxNumberData
      NaturalPersonIndicator
```

```
MainAddress
  CareOfName : string
  Street : string
  HouseNumber : string
  TransportationZone : string
  Building : string
  Floor : string
  RoomNumber : string
  City : string
  District : string
  Region : string
  PostalCode : string
  Country : string
  POBox : string
  POBoxPostalCode : string
  AddressUsageCode : string
  TaxJurCode : string
  CommType : string
  PhoneCompleteNumber : string
  MobilePhoneNumber : string
  FaxCompleteNumber : string
  EMailAddress : string
PartnerNumber : string
CategoryCode : string
IndustrySectors
Addresses
  CareOfName : string
  Street : string
```

Select a node to display its correspondences

The Problem

- Multiple database schemas to be mapped pair-wise by attribute correspondences

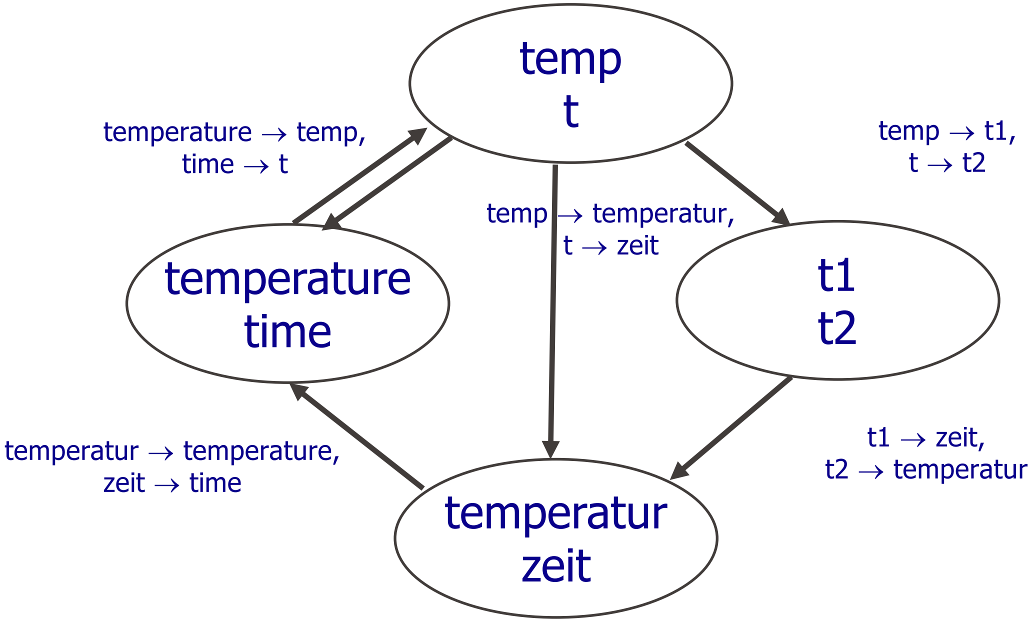


Initial mappings generated by humans or automated schema matchers

Can we trust them?

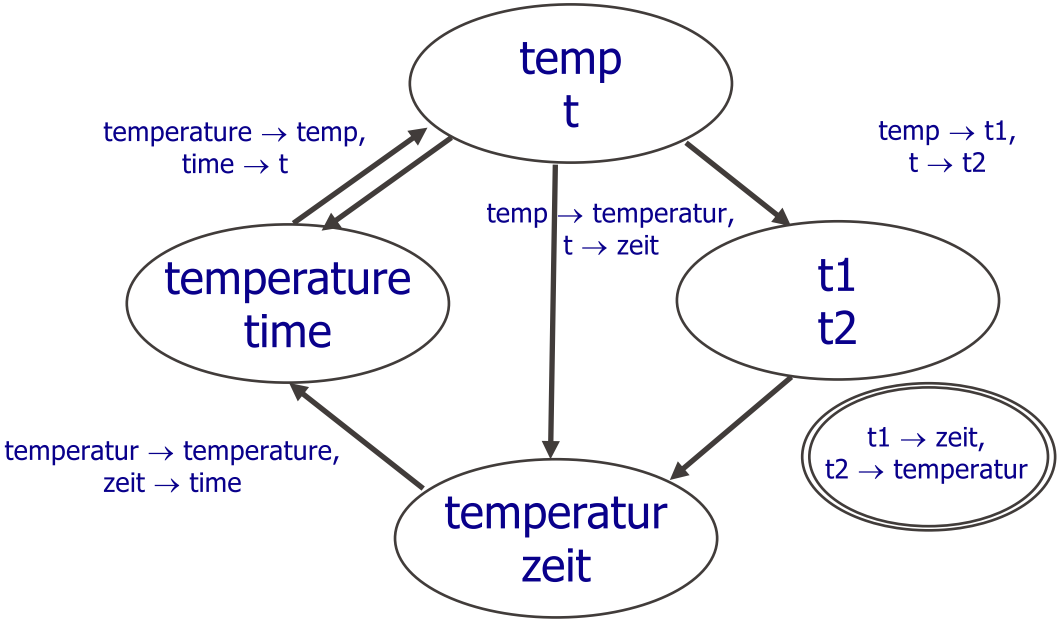
Semantic Gossiping

- Different schemas on similar data will be related to each other



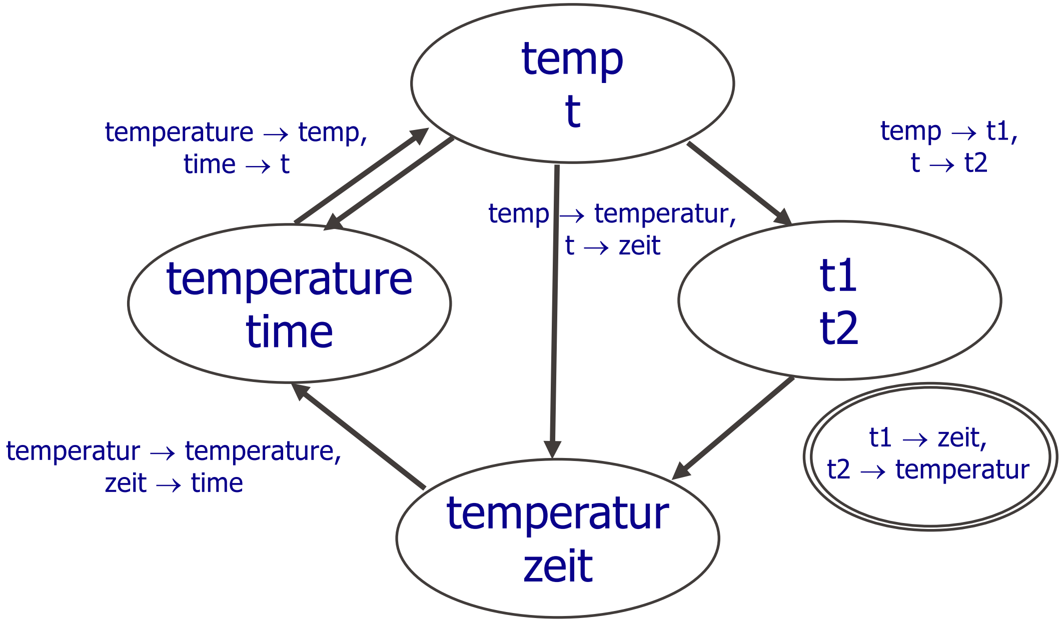
Erroneous Mappings

- Relationships among schemas will be either manually or automatically created: both methods are (highly) error-prone



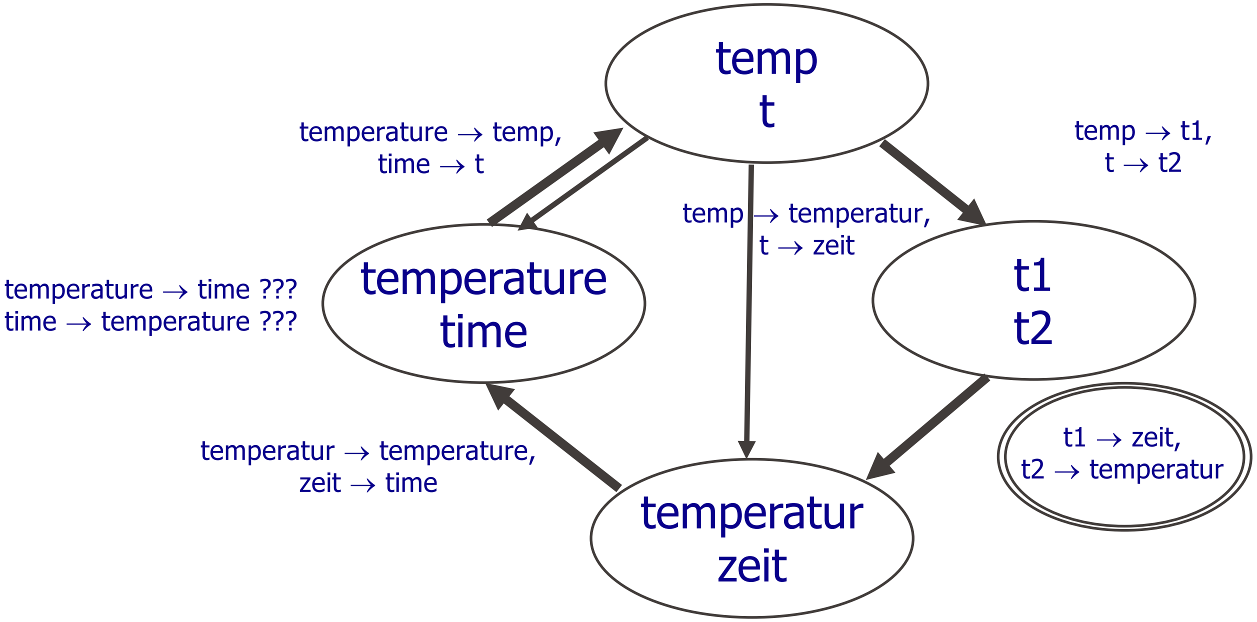
Erroneous Mappings

- Problem: Is it possible to learn from the network about the correctness of relationships?



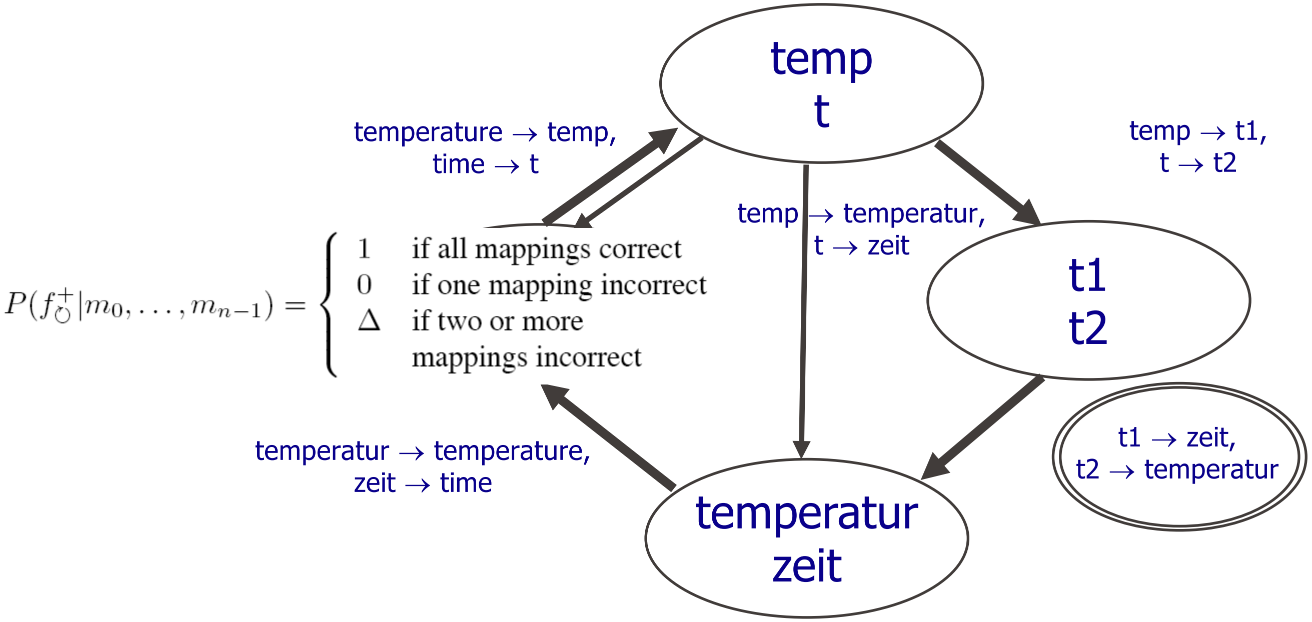
Erroneous Mappings

- Mappings among concepts can be composed
  - Allows to identify redundant compositions, e.g. cycles
  - Those can be checked for **consistency**



Probabilistic Reasoning

- Probabilistic model on the correctness of mappings m_i in a cycle using feedback f_j from mapping composition
  - Basis for probabilistic inference of correctness of mappings



$$P(f^{+}_{\circlearrowleft}|m_0,\ldots,m_{n-1}) = \begin{cases} 1 & \text{if all mappings correct} \\ 0 & \text{if one mapping incorrect} \\ \Delta & \text{if two or more mappings incorrect} \end{cases}$$

Cudre-Mauroux, Philippe, Karl Aberer, and Andras Feher. "Probabilistic message passing in peer data management systems." 22nd International Conference on Data Engineering (ICDE'06). IEEE, 2006.

Computing a Marginal for One Mapping

- Goal: determine $P(m_i \mid f_0)$

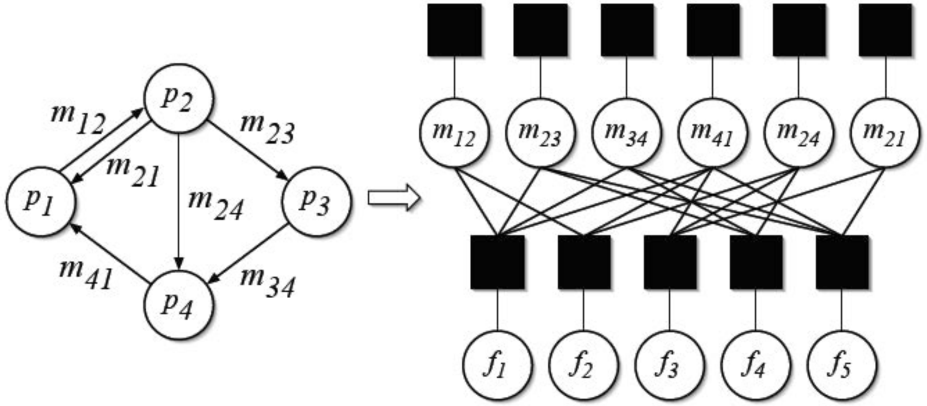
$$P(m_0 \mid f_0) = \sum_{m_3 m_4} P(m_0, m_3, m_4, f_0) P(f_0)^{-1} =$$

$$\sum_{m_3 m_4} \underbrace{P(m_0)P(m_3)P(m_4)}_{\text{a priori probability}} \underbrace{P(f_0 \mid m_0, m_3, m_4)P(f_0)^{-1}}_{\text{known}}$$

- Maximum entropy principle: a priori probability $P(m_i) = 0.5$
- Bayes Theorem $P(m \mid f) = \dfrac{P(m)P(f \mid m)}{P(f)}$ and marginalization

Correlated Feedback

- But: feedbacks on different cycles are correlated
  - One wrong mapping will affect several cycles/paths
  - Need to express a global probabilistic model for the mapping graph



Deriving PDMS Factor-Graphs from Mapping Network

Message Passing on Factor Graphs

- General Formulation (n() neighbors in factor graph)

variable x to local factor f:

$$\mu_{x\to f}(x) = \prod_{h\in n(x)\setminus\{f\}} \mu_{h\to x}(x)$$

local factor f to variable x

$$\mu_{f\to x}(x) = \sum_{\sim\{x\}} \left(f(X) \prod_{y\in n(f)\setminus\{x\}} \mu_{y\to f}(y) \right)$$

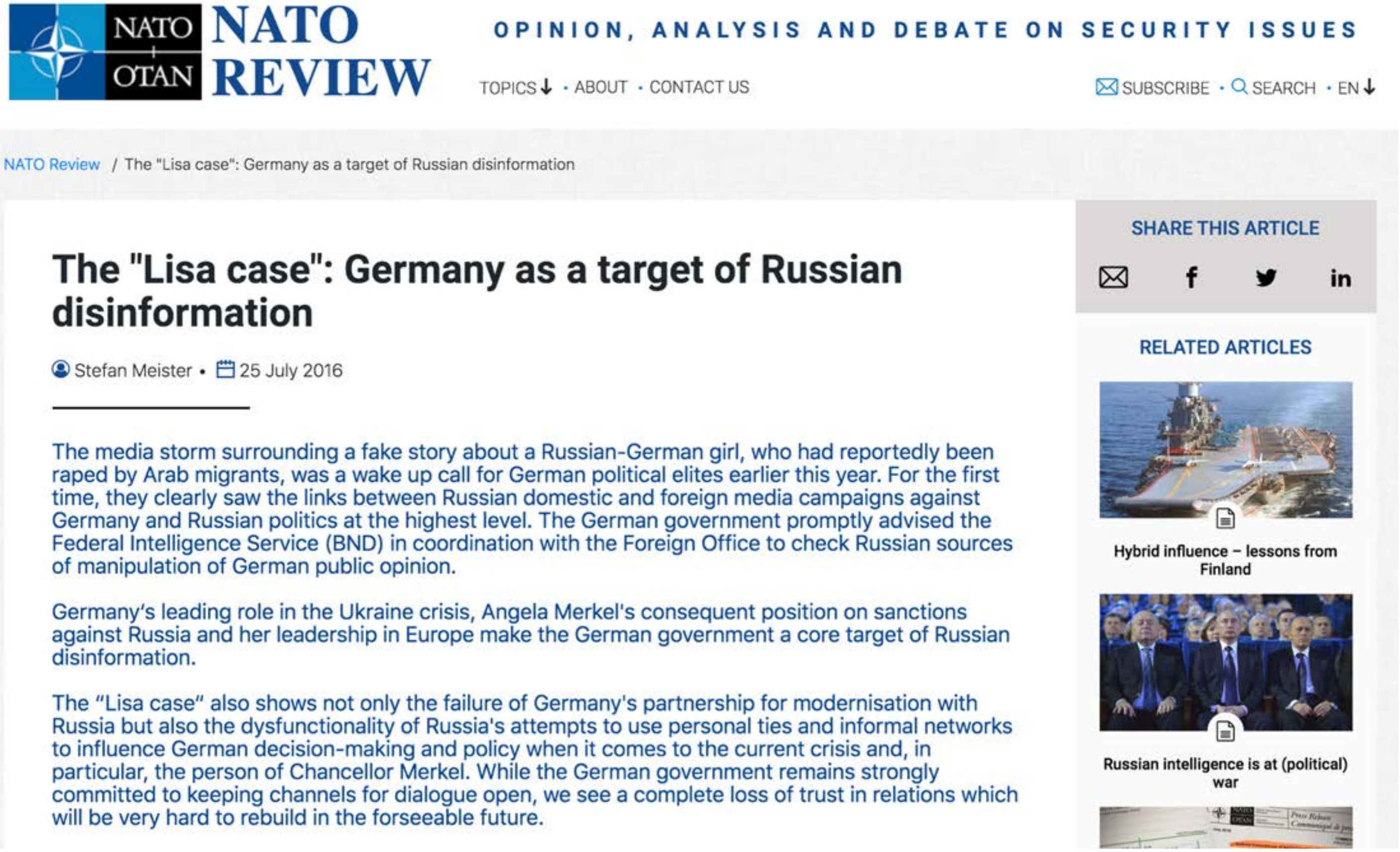
- Works correctly on trees

Summary

- Checking consistency works well if you have many redundant reports (dense feedback)
- What to do if feedback is sparse?

Chapter 2: Learning from History

Fake News


NATO OTAN
NATO REVIEW
OPINION, ANALYSIS AND DEBATE ON SECURITY ISSUES
TOPICS · ABOUT · CONTACT US
SUBSCRIBE · SEARCH · EN
NATO Review / The "Lisa case": Germany as a target of Russian disinformation
SHARE THIS ARTICLE
f
in
RELATED ARTICLES
Hybrid influence – lessons from Finland
Russian intelligence is at (political) war

The "Lisa case": Germany as a target of Russian disinformation

Stefan Meister • 25 July 2016

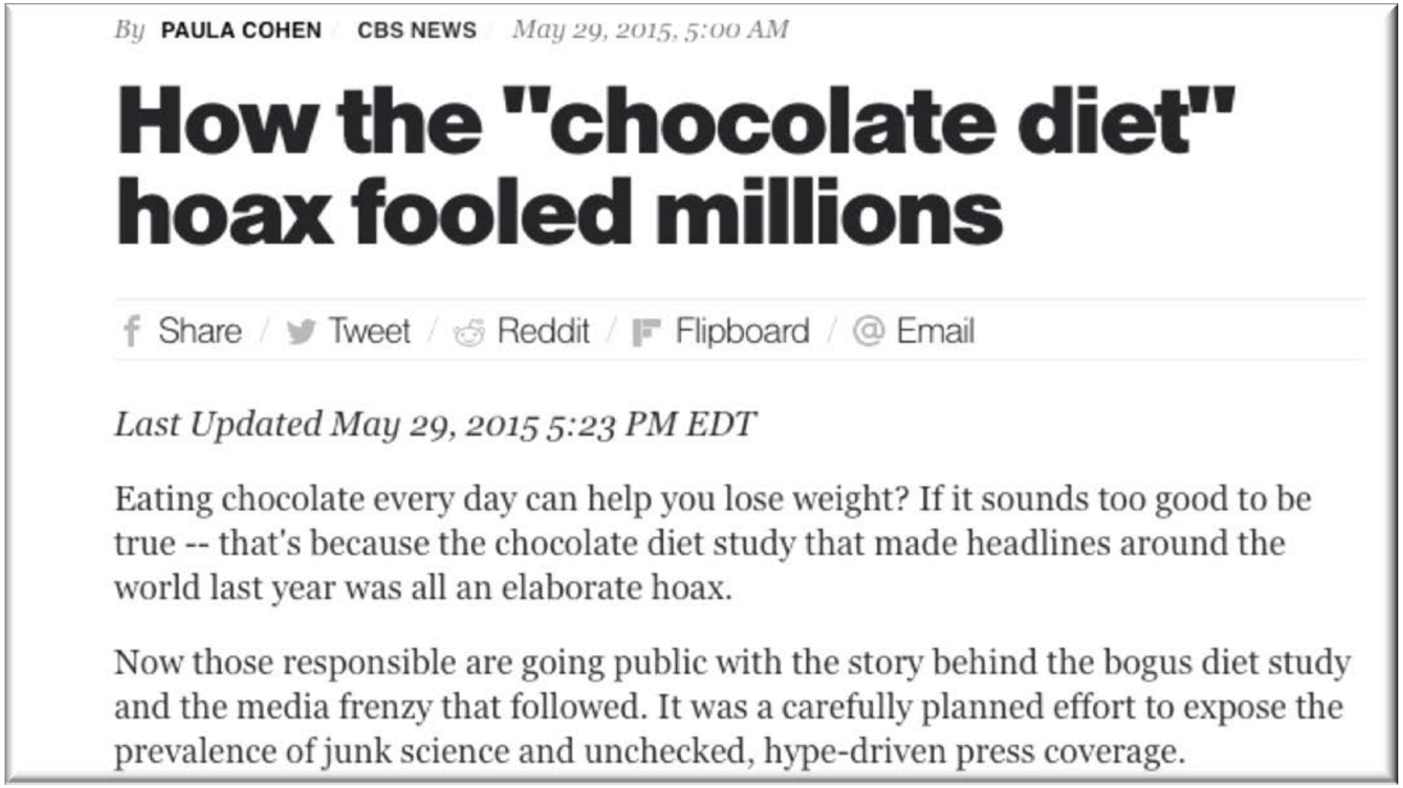
The media storm surrounding a fake story about a Russian-German girl, who had reportedly been raped by Arab migrants, was a wake up call for German political elites earlier this year. For the first time, they clearly saw the links between Russian domestic and foreign media campaigns against Germany and Russian politics at the highest level. The German government promptly advised the Federal Intelligence Service (BND) in coordination with the Foreign Office to check Russian sources of manipulation of German public opinion.

Germany's leading role in the Ukraine crisis, Angela Merkel's consequent position on sanctions against Russia and her leadership in Europe make the German government a core target of Russian disinformation.

The "Lisa case" also shows not only the failure of Germany's partnership for modernisation with Russia but also the dysfunctionality of Russia's attempts to use personal ties and informal networks to influence German decision-making and policy when it comes to the current crisis and, in particular, the person of Chancellor Merkel. While the German government remains strongly committed to keeping channels for dialogue open, we see a complete loss of trust in relations which will be very hard to rebuild in the forseeable future.

https://www.nato.int/docu/review/articles/2016/07/25/the-lisa-case-germany-as-a-target-of-russian-disinformation/index.html

Boosting Bad Science



By PAULA COHEN / CBS NEWS / May 29, 2015, 5:00 AM

How the "chocolate diet" hoax fooled millions

Last Updated May 29, 2015 5:23 PM EDT

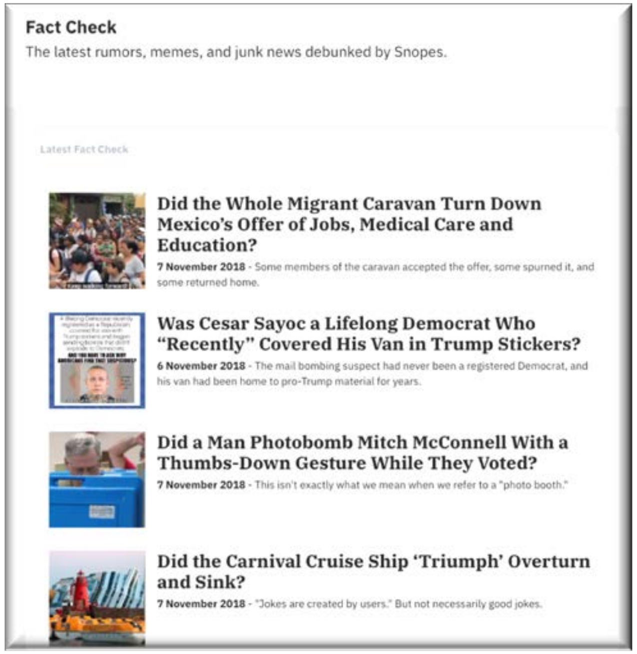
Eating chocolate every day can help you lose weight? If it sounds too good to be true -- that's because the chocolate diet study that made headlines around the world last year was all an elaborate hoax.

Now those responsible are going public with the story behind the bogus diet study and the media frenzy that followed. It was a carefully planned effort to expose the prevalence of junk science and unchecked, hype-driven press coverage.

https://www.cbsnews.com/news/how-the-chocolate-diet-hoax-fooled-millions/

Credibility Evaluation Today

Manual fact checking by experts



Fact Check

The latest rumors, memes, and junk news debunked by Snopes.

Latest Fact Check

Did the Whole Migrant Caravan Turn Down Mexico's Offer of Jobs, Medical Care and Education?

7 November 2018 - Some members of the caravan accepted the offer, some spurned it, and some returned home.

Was Cesar Sayoc a Lifelong Democrat Who "Recently" Covered His Van in Trump Stickers?

6 November 2018 - The mail bombing suspect had never been a registered Democrat, and his van had been home to pro-Trump material for years.

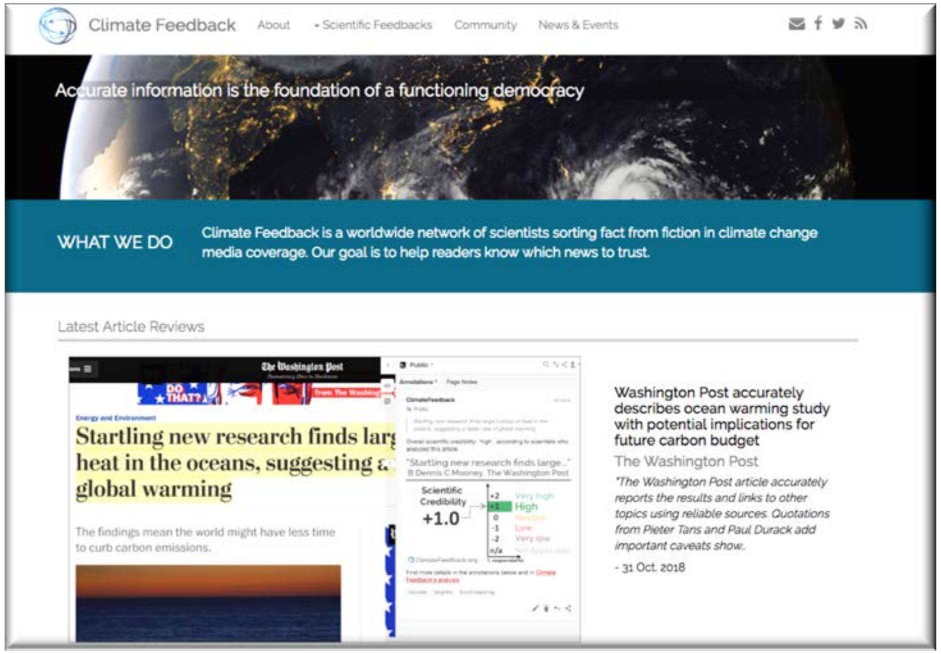
Did a Man Photobomb Mitch McConnell With a Thumbs-Down Gesture While They Voted?

7 November 2018 - This isn't exactly what we mean when we refer to a "photo booth."

Did the Carnival Cruise Ship 'Triumph' Overturn and Sink?

7 November 2018 - "Jokes are created by users." But not necessarily good jokes.

General Fact Checking Sites (mostly journalists)



Climate Feedback   About   • Scientific Feedbacks   Community   News & Events

Accurate information is the foundation of a functioning democracy

WHAT WE DO

Climate Feedback is a worldwide network of scientists sorting fact from fiction in climate change media coverage. Our goal is to help readers know which news to trust.

Latest Article Reviews

The Washington Post

Energy and Environment

Startling new research finds larg heat in the oceans, suggesting a global warming

The findings mean the world might have less time to curb carbon emissions.

Public   Annotations   Page Notes

ClimateFeedback

"Startling new research finds large..."

Scientific Credibility +1.0

+2 Very high
+1 High
0 Neutral
-1 Low
-2 Very low
n/a

Washington Post accurately describes ocean warming study with potential implications for future carbon budget

The Washington Post

"The Washington Post article accurately reports the results and links to other topics using reliable sources. Quotations from Pieter Tans and Paul Durack add important caveats show.

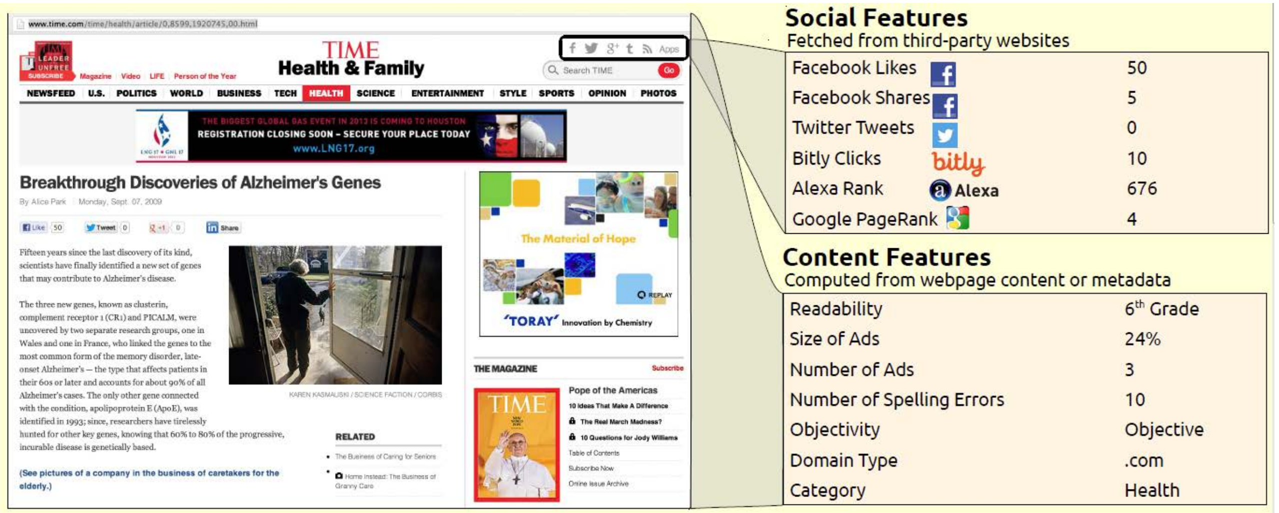
- 31 Oct. 2018

Specialized Science Fact Checking Sites (mostly scientists)

Automated Web Credibility Evaluation?

Insights

- There exist features that correlate with credibility
- Combination of content and social features works best
- **Quality of classification depends on quality of training data = ground truth**


www.time.com/time/health/article/0,8599,1920745,00.html
TIME LEADER UNFREE
SUBSCRIBE
Magazine | Video | LIFE | Person of the Year
TIME
Health & Family
Apps
Search TIME
Go
NEWSFEED | U.S. | POLITICS | WORLD | BUSINESS | TECH | HEALTH | SCIENCE | ENTERTAINMENT | STYLE | SPORTS | OPINION | PHOTOS
THE BIGGEST GLOBAL GAS EVENT IN 2013 IS COMING TO HOUSTON
REGISTRATION CLOSING SOON – SECURE YOUR PLACE TODAY
www.LNG17.org
LNG 17 ★ GNL 17
Like 50
Tweet 0
+1 0
Share
The Material of Hope
REPLAY
'TORAY' Innovation by Chemistry
TIME

Breakthrough Discoveries of Alzheimer's Genes

By Alice Park | Monday, Sept. 07, 2009

Fifteen years since the last discovery of its kind, scientists have finally identified a new set of genes that may contribute to Alzheimer's disease.

The three new genes, known as clusterin, complement receptor 1 (CR1) and PICALM, were uncovered by two separate research groups, one in Wales and one in France, who linked the genes to the most common form of the memory disorder, late-onset Alzheimer's — the type that affects patients in their 60s or later and accounts for about 90% of all Alzheimer's cases. The only other gene connected with the condition, apolipoprotein E (ApoE), was identified in 1993; since, researchers have tirelessly hunted for other key genes, knowing that 60% to 80% of the progressive, incurable disease is genetically based.

KAREN KASMAUSKI / SCIENCE FACTION / CORBIS

(See pictures of a company in the business of caretakers for the elderly.)

RELATED

- The Business of Caring for Seniors
- Home Instead: The Business of Granny Care

THE MAGAZINE Subscribe

Pope of the Americas

10 Ideas That Make A Difference

The Real March Madness?

10 Questions for Jody Williams

Table of Contents

Subscribe Now

Online Issue Archive

Social Features
Fetched from third-party websites

Facebook Likes	50
Facebook Shares	5
Twitter Tweets	0
Bitly Clicks bitly	10
Alexa Rank Alexa	676
Google PageRank	4

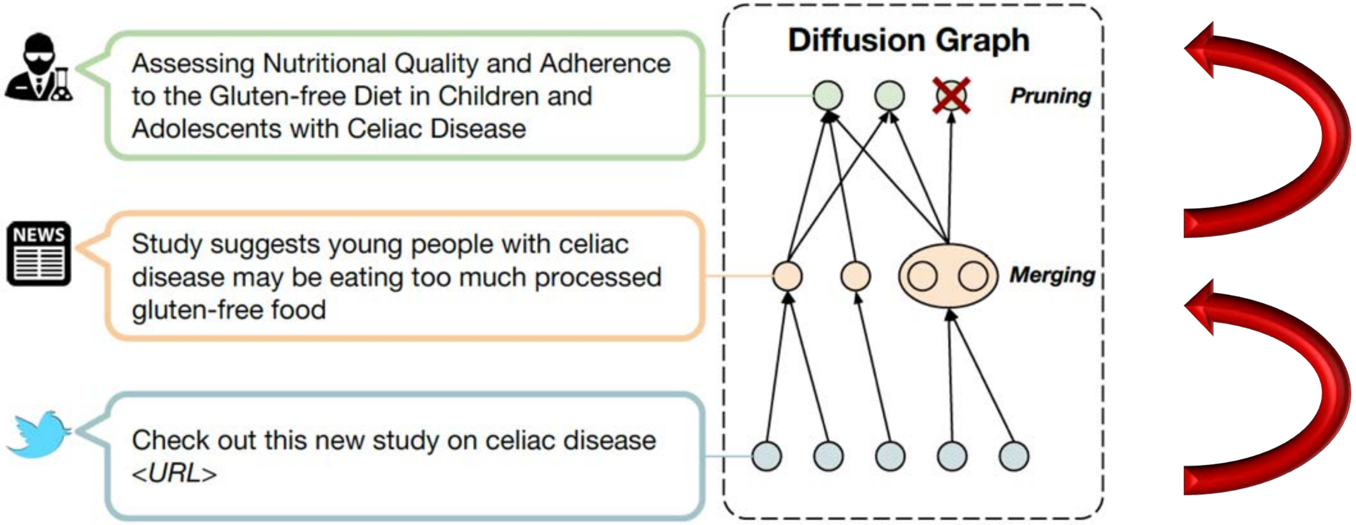
Content Features
Computed from webpage content or metadata

Readability	6th Grade
Size of Ads	24%
Number of Ads	3
Number of Spelling Errors	10
Objectivity	Objective
Domain Type	.com
Category	Health

Olteanu, A., Peshterliev, S., Liu, X., & Aberer, K. (2013, March). Web credibility: features exploration and credibility prediction. In *European conference on information retrieval* (pp. 557-568). Springer, Berlin, Heidelberg.

SciLens: Evaluating Quality of Scientific News

Investigate whether scientific findings are adequately presented in science news: Analyze the news in their **context**



Assessing Nutritional Quality and Adherence to the Gluten-free Diet in Children and Adolescents with Celiac Disease

Study suggests young people with celiac disease may be eating too much processed gluten-free food

Check out this new study on celiac disease *<URL>*

Diffusion Graph

Pruning

Merging

Difficult:
Resolve Quotes

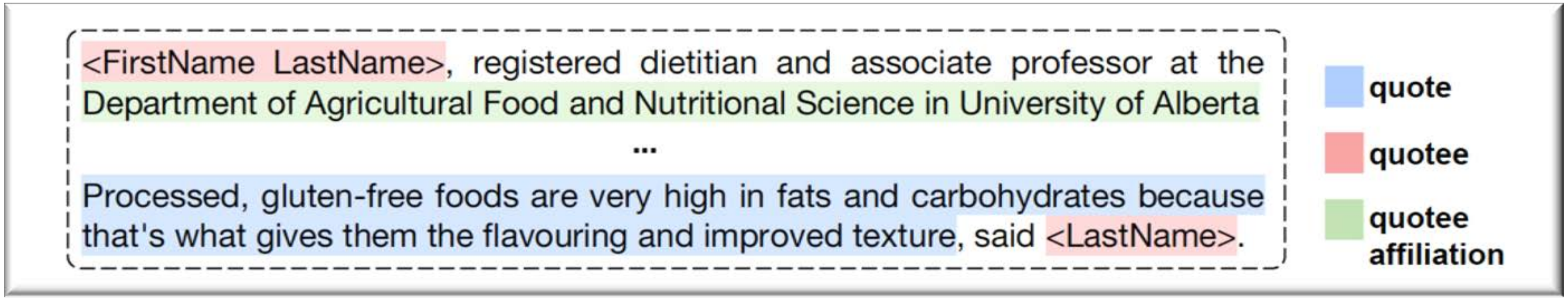
Doable:
Resolve URLs

Panayiotis Smeros, Carlos Castillo, Karl Aberer: Evaluating the Quality of Scientific News Reporting from Social and Content Indicators, In *The International World Wide Web Conference 2019.*

Resolving Quotes

Linking news to science articles: bootstrapping approach

- Create a set of reporting verbs (“say”, “claim”, ”prove”,…), studies (“survey”, “analysis”) and scientists (“researcher”, ”analyst”)
- Extend this set with semantic similar words using word embeddings
- Identify indirect speech using NLP processing
- Build regular expression patterns over word classes



<FirstName LastName>, registered dietitian and associate professor at the Department of Agricultural Food and Nutritional Science in University of Alberta

...

Processed, gluten-free foods are very high in fats and carbohydrates because that's what gives them the flavouring and improved texture, said <LastName>.

🟦 quote

🟥 quotee

🟩 quotee affiliation

90% precision at 50% recall

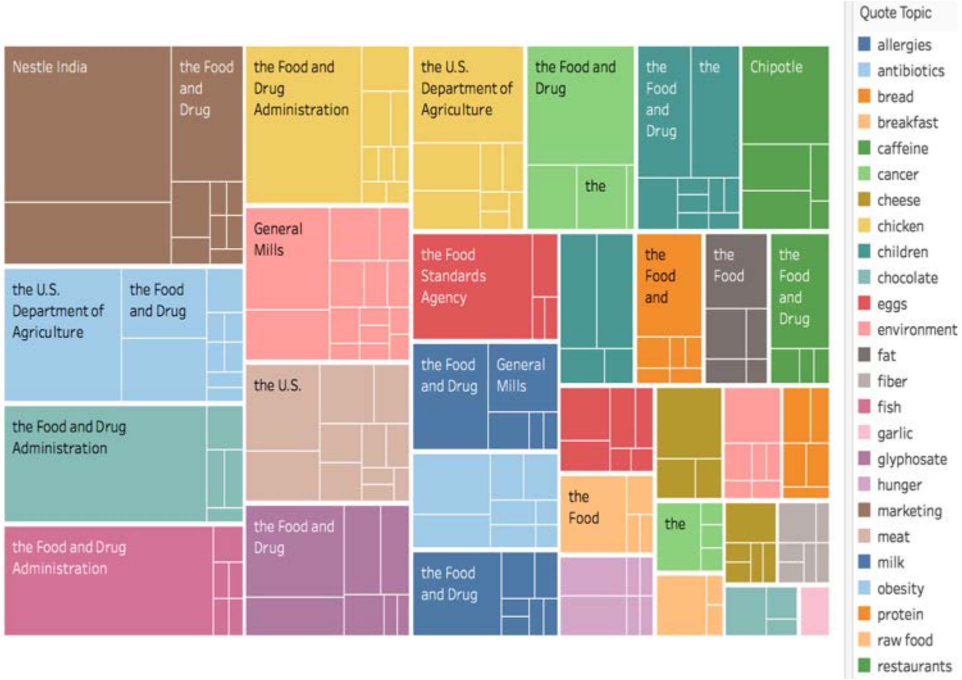
Data Collection in the Nutrition domain

Start from social media posts

then identify news,

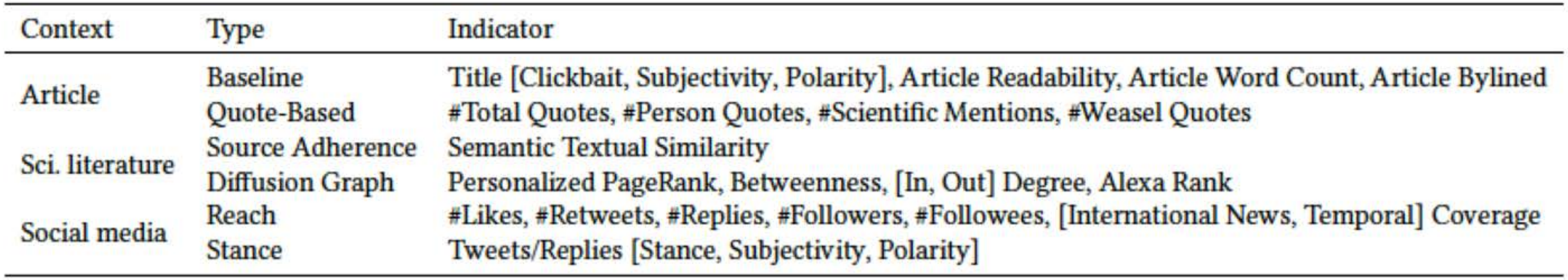
finally find quotations of science articles

- 11’000 social media posts
- 7’500 news articles
- 1’300 papers

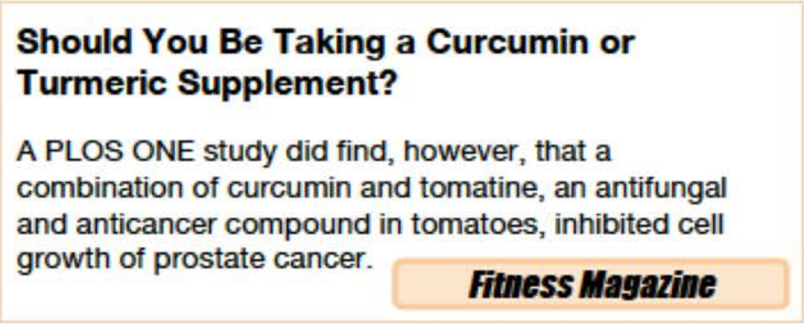


Most Quoted Organization per Topic

Quality Features



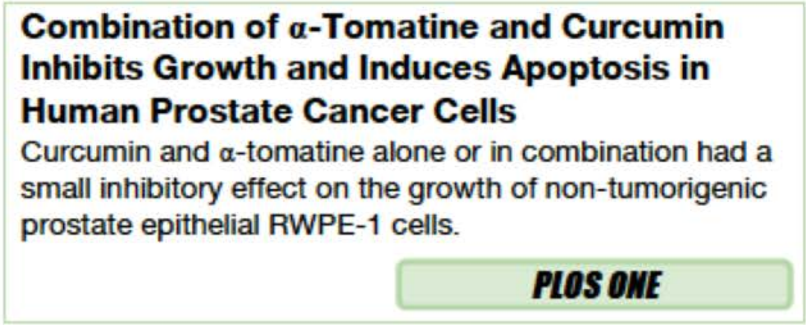
Context	Type	Indicator
Article	Baseline	Title [Clickbait, Subjectivity, Polarity], Article Readability, Article Word Count, Article Bylined
	Quote-Based	#Total Quotes, #Person Quotes, #Scientific Mentions, #Weasel Quotes
Sci. literature	Source Adherence	Semantic Textual Similarity
	Diffusion Graph	Personalized PageRank, Betweenness, [In, Out] Degree, Alexa Rank
Social media	Reach	#Likes, #Retweets, #Replies, #Followers, #Followees, [International News, Temporal] Coverage
	Stance	Tweets/Replies [Stance, Subjectivity, Polarity]



Should You Be Taking a Curcumin or Turmeric Supplement?

A PLOS ONE study did find, however, that a combination of curcumin and tomatine, an antifungal and anticancer compound in tomatoes, inhibited cell growth of prostate cancer.

Fitness Magazine



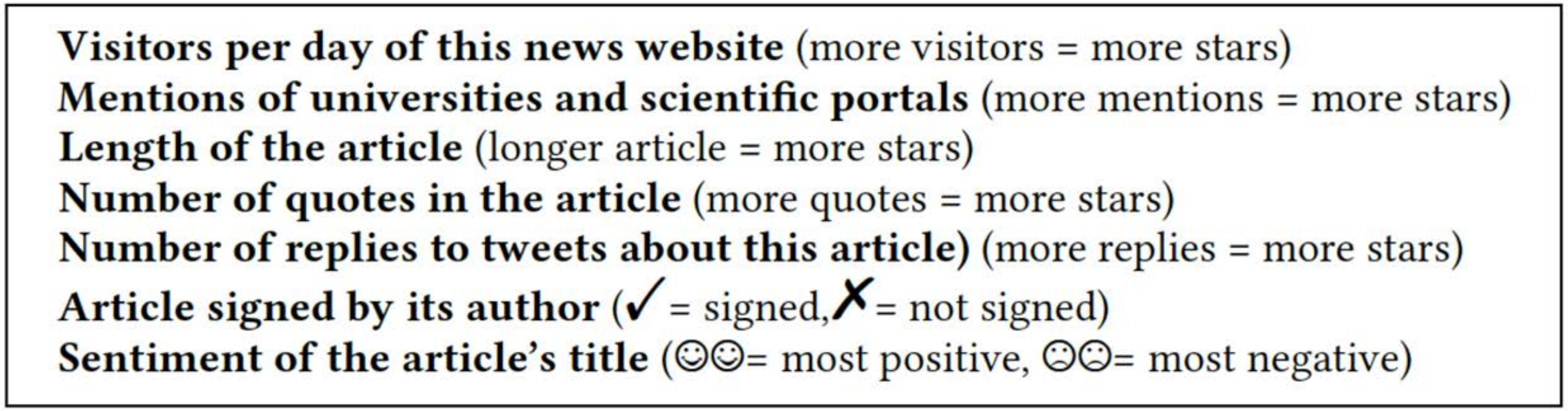
Combination of α-Tomatine and Curcumin Inhibits Growth and Induces Apoptosis in Human Prostate Cancer Cells

Curcumin and α-tomatine alone or in combination had a small inhibitory effect on the growth of non-tumorigenic prostate epithelial RWPE-1 cells.

Example: Source Adherence (semantic textual similarity feature for News article)

Non-Expert Evaluation

Hypothesis: when non-experts see the quality features they can do a better assessment

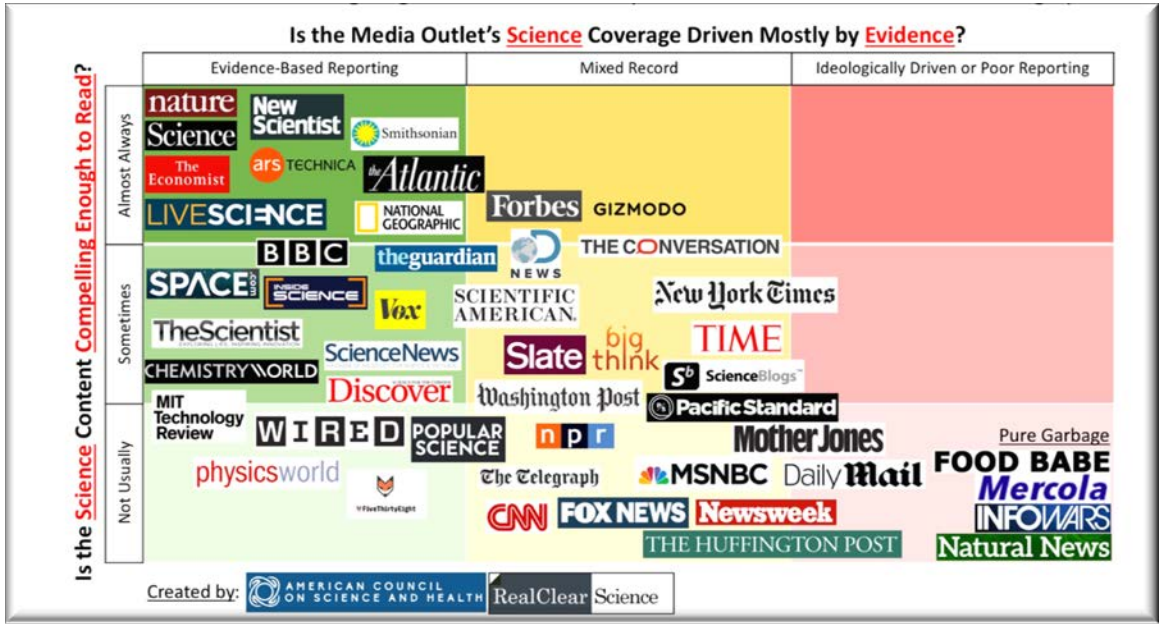


Visitors per day of this news website (more visitors = more stars)
Mentions of universities and scientific portals (more mentions = more stars)
Length of the article (longer article = more stars)
Number of quotes in the article (more quotes = more stars)
Number of replies to tweets about this article) (more replies = more stars)
Article signed by its author (✓ = signed, ✗ = not signed)
Sentiment of the article's title (☺☺= most positive, ☹☹= most negative)

Weak Supervision

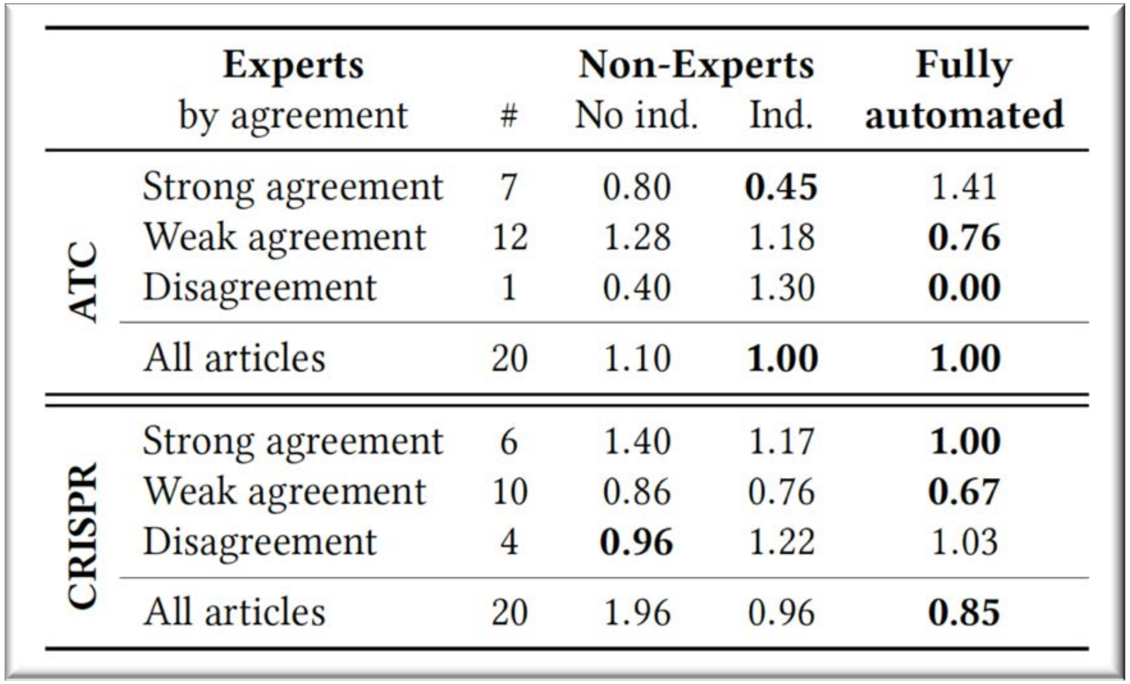
Building a machine learning classifier using the quality features

- In absence of ground truth we use weak supervision
- Instead of judging the quality of articles we judge quality of outlet



Expert assessment of quality of journals

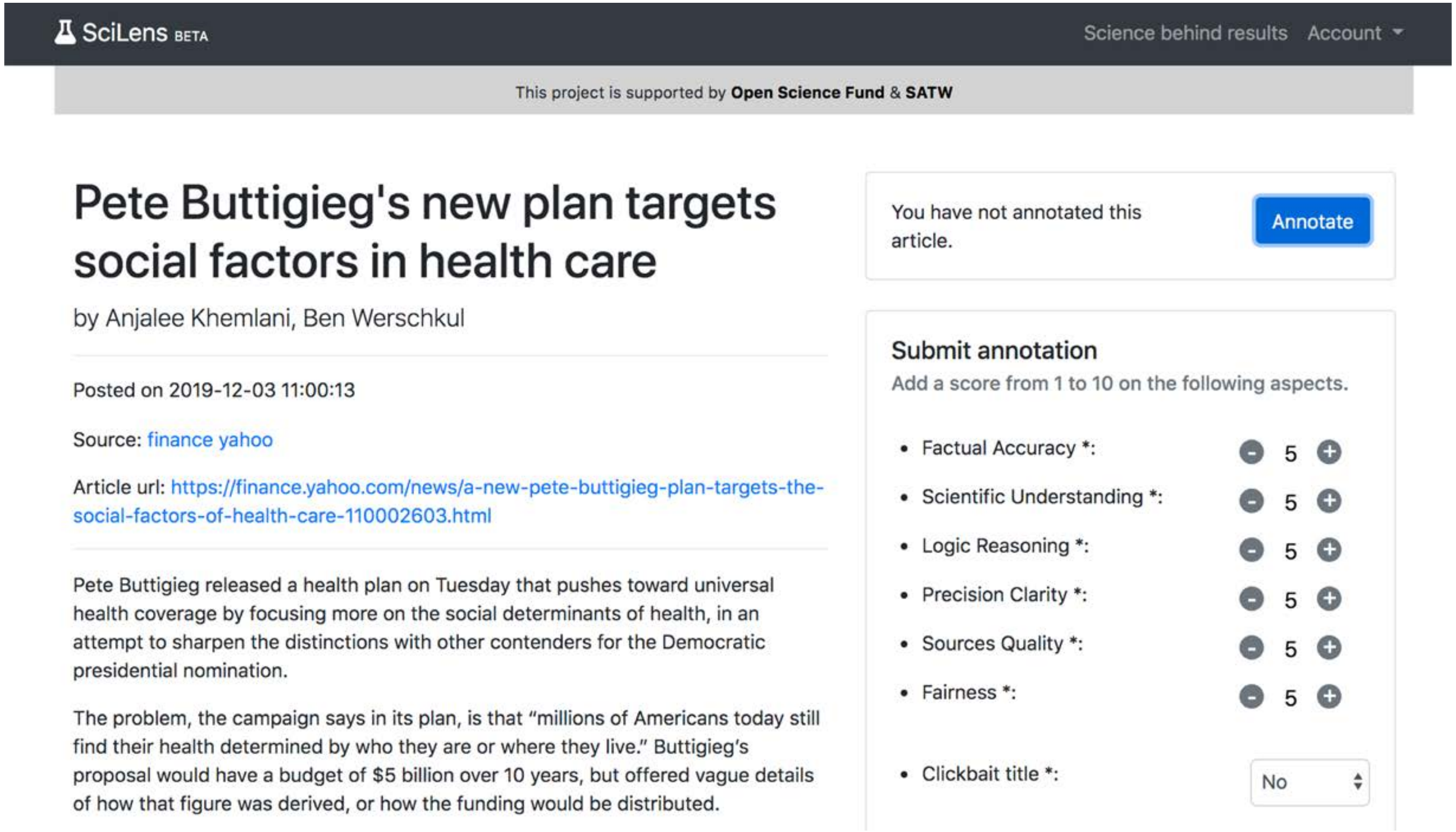
Results



	Experts by agreement	#	Non-Experts No ind.	 Ind.	Fully automated
ATC	Strong agreement	7	0.80	**0.45**	1.41
	Weak agreement	12	1.28	1.18	**0.76**
	Disagreement	1	0.40	1.30	**0.00**
	All articles	20	1.10	**1.00**	**1.00**
CRISPR	Strong agreement	6	1.40	1.17	**1.00**
	Weak agreement	10	0.86	0.76	**0.67**
	Disagreement	4	**0.96**	1.22	1.03
	All articles	20	1.96	0.96	**0.85**

- When displaying features to non-experts, the quality of their evaluation increases significantly
- A weakly-supervised classifier performs better than non-experts!

Platform



This project is supported by **Open Science Fund** & **SATW**

Pete Buttigieg's new plan targets social factors in health care

by Anjalee Khemlani, Ben Werschkul

Posted on 2019-12-03 11:00:13

Source: finance yahoo

Article url: https://finance.yahoo.com/news/a-new-pete-buttigieg-plan-targets-the-social-factors-of-health-care-110002603.html

Pete Buttigieg released a health plan on Tuesday that pushes toward universal health coverage by focusing more on the social determinants of health, in an attempt to sharpen the distinctions with other contenders for the Democratic presidential nomination.

The problem, the campaign says in its plan, is that “millions of Americans today still find their health determined by who they are or where they live.” Buttigieg's proposal would have a budget of $5 billion over 10 years, but offered vague details of how that figure was derived, or how the funding would be distributed.

You have not annotated this article.

Submit annotation

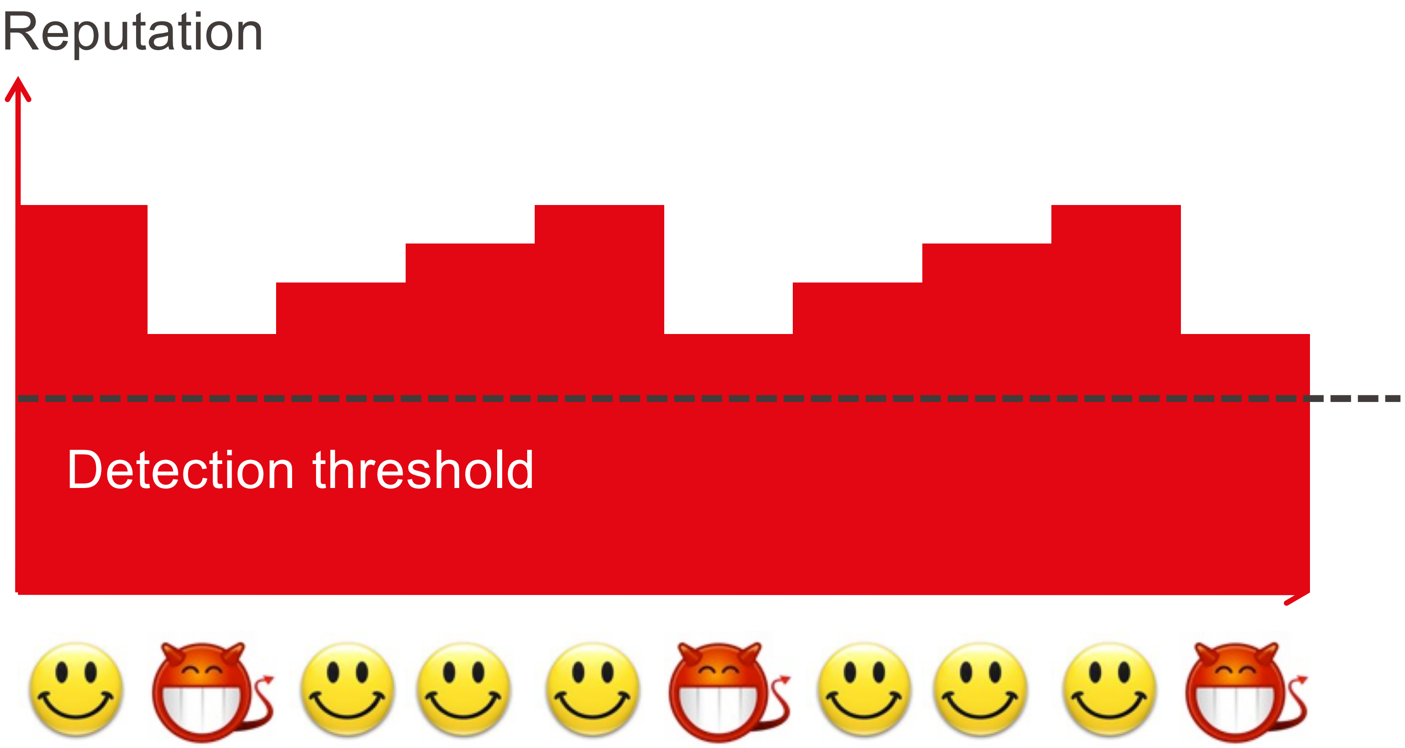
Add a score from 1 to 10 on the following aspects.

- Factual Accuracy *: 5
- Scientific Understanding *: 5
- Logic Reasoning *: 5
- Precision Clarity *: 5
- Sources Quality *: 5
- Fairness *: 5
- Clickbait title *: No

scilens.epfl.ch

Dessert: Rationality

A Rational Agent



Trust is not an inherent property of an agent!

Rationality

- Actions of peers have associated utility
- Example: prisoners dilemma

u_1,u_2	Cooperate	Cheat
Cooperate	5, 5	0, 6
Cheat	6, 0	1, 1

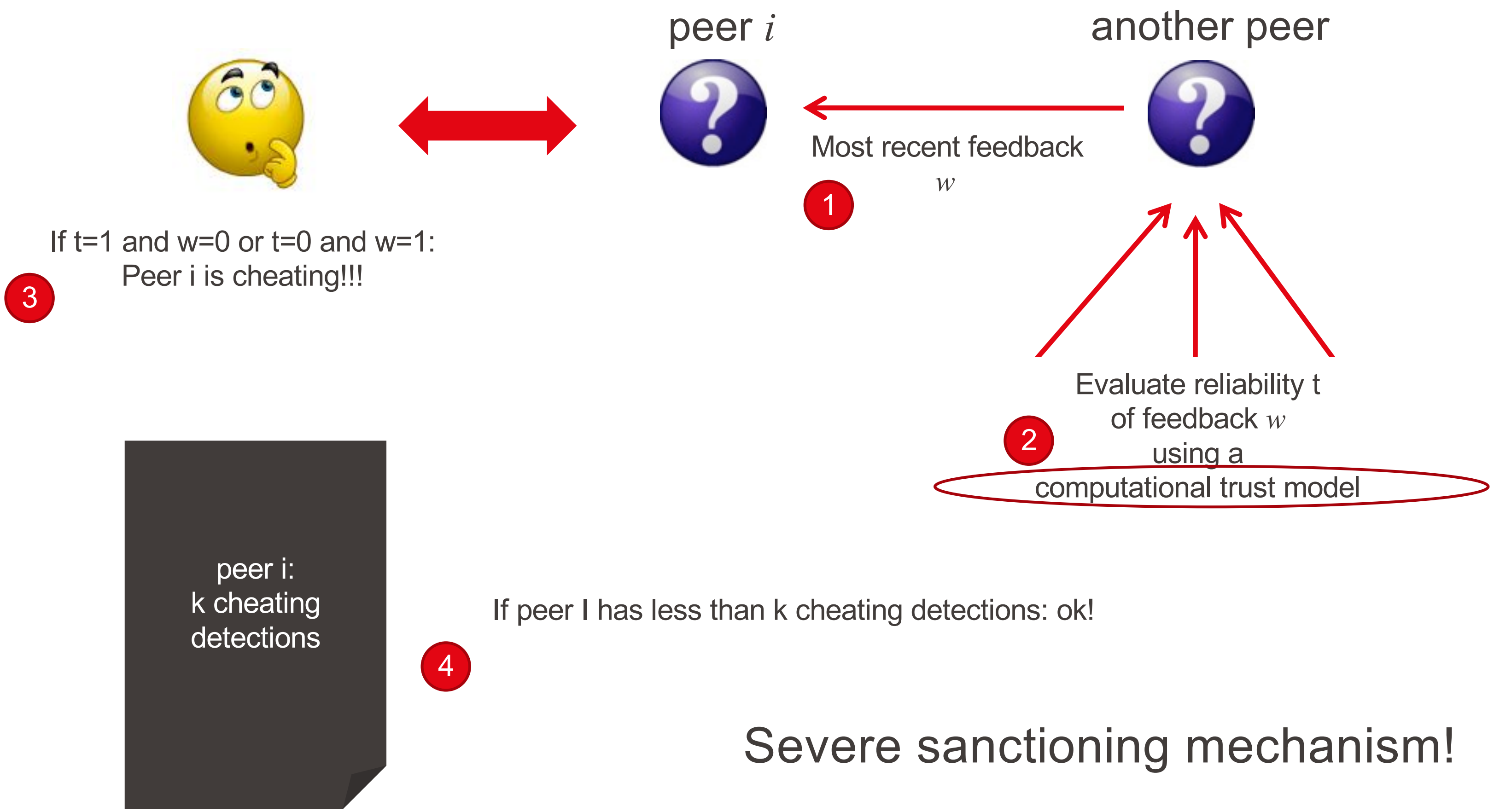
- Main insights (2 Nobel prices)
  - One shot game: no cooperation (Nash)
  - Repeated game: cooperation – tit-for-tat (Axelrod)
- Common explanation of the concept of trust

The Setting

- Model (the community model)
  - Feedback w_i on peer i's service
  - Peer i has incentive to cheat:
    legal gain u_i < illegitimate gain $u_i + v_i$
- Peer behavior
  - Honest peers never cheat
  - Malicious peers cheat probabilistically
  - Rational peers optimize their utility

Vu, Le-Hung, and Karl Aberer. "Effective usage of computational trust models in rational environments." ACM Transactions on Autonomous and Adaptive Systems (TAAS) 6.4 (2011): 1-25.

Sanctioning



peer i

another peer

Most recent feedback w

1

Evaluate reliability t of feedback w using a computational trust model

2

If t=1 and w=0 or t=0 and w=1:
Peer i is cheating!!!

3

peer i:
k cheating
detections

If peer I has less than k cheating detections: ok!

4

Severe sanctioning mechanism!

Computational Trust Model

- May use several information sources
  - past performance of sellers, trusted sources, own belief on environment's vulnerability, relations between peers, ratings from other agents
- May use variety of statistical models/heuristics
  - probabilistic approaches, e.g. EM, collaborative filtering, clustering of ratings, etc.
- Accuracy measure (known to peers)
  - $P[est+| real-] < \varepsilon$, $P[est - | real+] < \varepsilon$

Trust Accuracy vs. Cooperation

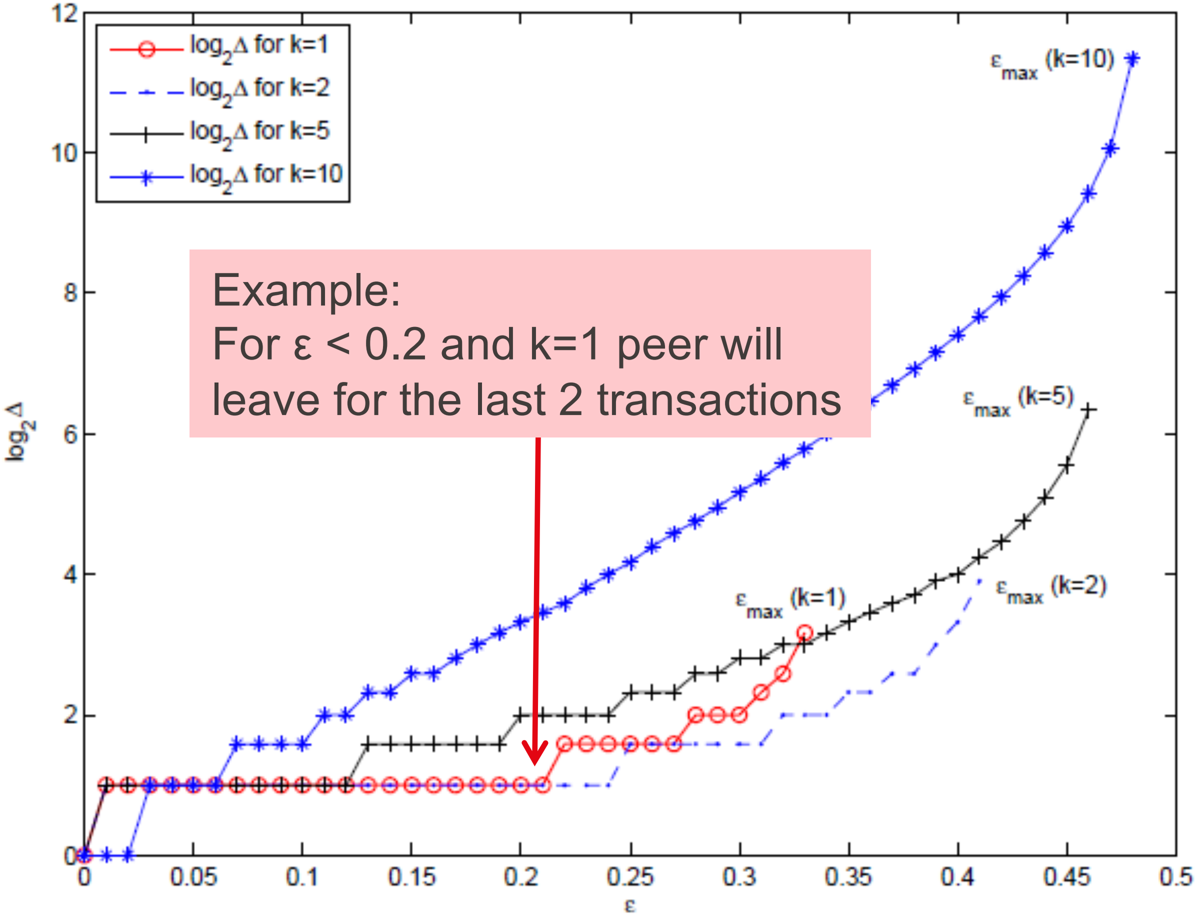
Theorem: if computational trust model sufficiently accurate and gains are bounded rational peers cooperate in all but the last Δ transactions.

Bounded gains $(u_* < u < u^*,\ v_* < v < v^*)$

$$\Delta = max\{1, \lceil \frac{\ln\left[1 - \frac{v^* \varepsilon^k}{u_*((1-\varepsilon)^k - \varepsilon^k)}\right]}{\ln\left(1 - \varepsilon^k\right)} \rceil\}$$

$$\varepsilon < \varepsilon_{max}(k) = 1/(1 + \sqrt[k]{1 + v^*/u_*})$$

Incentive to Leave (Δ)



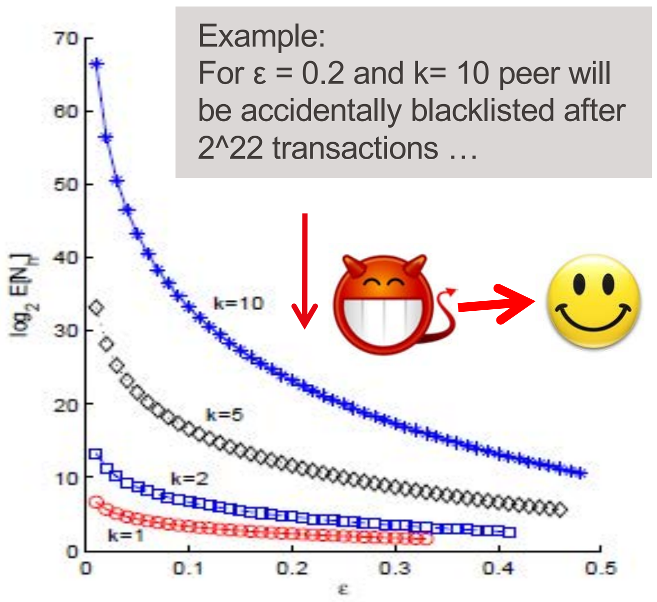
Example:
For ε < 0.2 and k=1 peer will leave for the last 2 transactions

accuracy of computational trust model

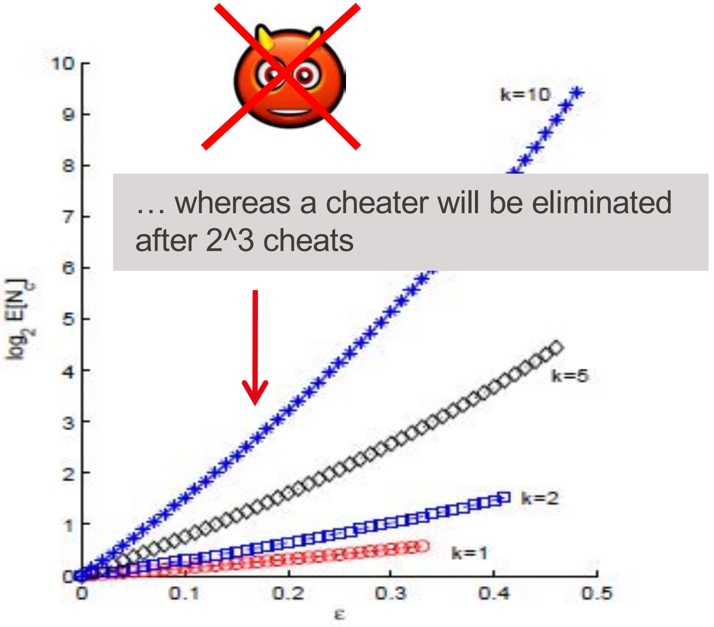
Emergence of Cooperation

- Cooperation is enforced if peers stay infinitely or long enough given ε sufficiently small
  - resilient against rating manipulation
  - malicious peers are eliminated

#transactions of an honest seller till mistakenly blacklisted



Example:
For ε = 0.2 and k= 10 peer will be accidentally blacklisted after 2^22 transactions …



… whereas a cheater will be eliminated after 2^3 cheats

#cheats of a malicious seller till being eliminated

How to Prevent Whitewashing?

- If peers that have been sanctioned can return with a new identity, this mechanism will not work

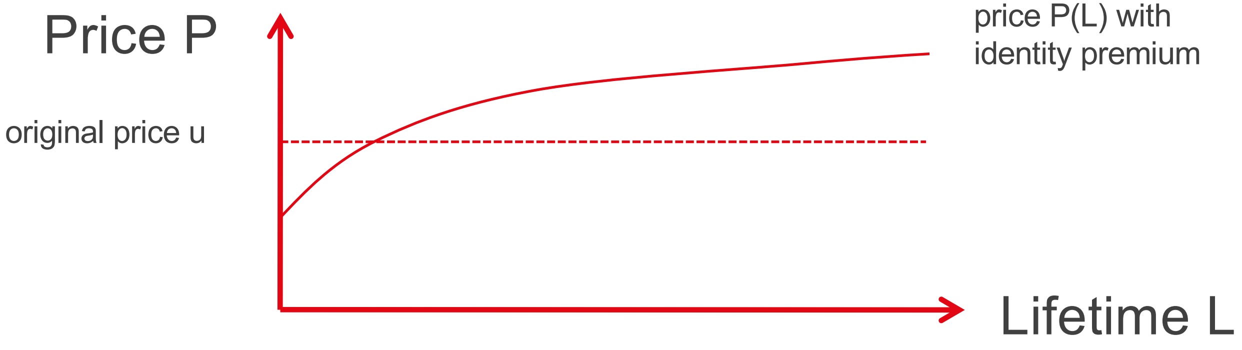
- Idea in Ebay: sellers that stay longer in the system ask for higher prices
  - Makes it unattractive to leave the system for whitewashing
  - Clients on eBay are willing to pay higher prices for reliable seller [Resnick 2006]

- *Does it work?*

Identity Premium

$$P(L) = u\,(1 - \Phi) + f(L)$$

- L lifetime of seller
- f(0) = 0, f monotonically increasing
- $0 < \Phi < 1$, initial price below original price u

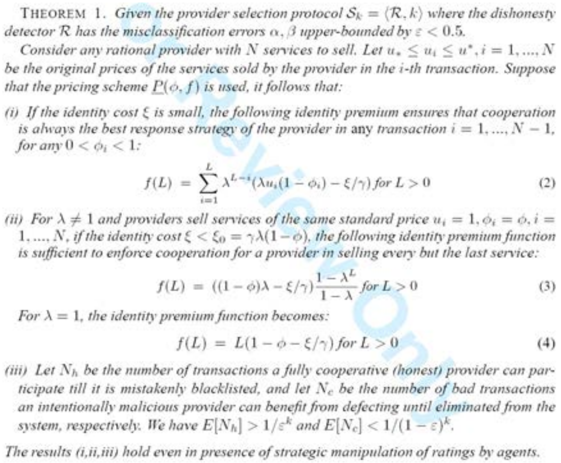


Vu, Le-Hung, Jie Zhang, and Karl Aberer. "Using identity premium for honesty enforcement and whitewashing prevention." Computational Intelligence 30.4 (2014): 771-797.

Cooperation Enforcement

- **Theorem:** If the identity cost is sufficiently small there exists an identity premium function such that a rational provider will cooperate in every but the last interaction.

- Bound on identity cost and premium depends on
  - Accuracy of dishonesty detector
  - Potential cheating gain
  - Initial price



THEOREM 1. *Given the provider selection protocol $\mathcal{S}_k = \langle \mathcal{R}, k \rangle$ where the dishonesty detector $\mathcal{R}$ has the misclassification errors α, β upper-bounded by $\varepsilon < 0.5$.*

Consider any rational provider with N services to sell. Let $u_* \le u_i \le u^*, i = 1, ..., N$ be the original prices of the services sold by the provider in the i-th transaction. Suppose that the pricing scheme $\underline{P}(\phi, f)$ is used, it follows that:

(i) If the identity cost ξ is small, the following identity premium ensures that cooperation is always the best response strategy of the provider in any *transaction $i = 1, ..., N-1$, for any $0 < \phi_i < 1$:*

$$f(L) = \sum_{i=1}^{L} \lambda^{L-i}(\lambda u_i(1-\phi_i) - \xi/\gamma) \text{ for } L > 0 \qquad (2)$$

(ii) For $\lambda \neq 1$ and providers sell services of the same standard price $u_i = 1, \phi_i = \phi, i = 1, ..., N$, if the identity cost $\xi < \xi_0 = \gamma\lambda(1-\phi)$, the following identity premium function is sufficient to enforce cooperation for a provider in selling every but the last service:

$$f(L) = ((1-\phi)\lambda - \xi/\gamma)\frac{1-\lambda^L}{1-\lambda} \text{ for } L > 0 \qquad (3)$$

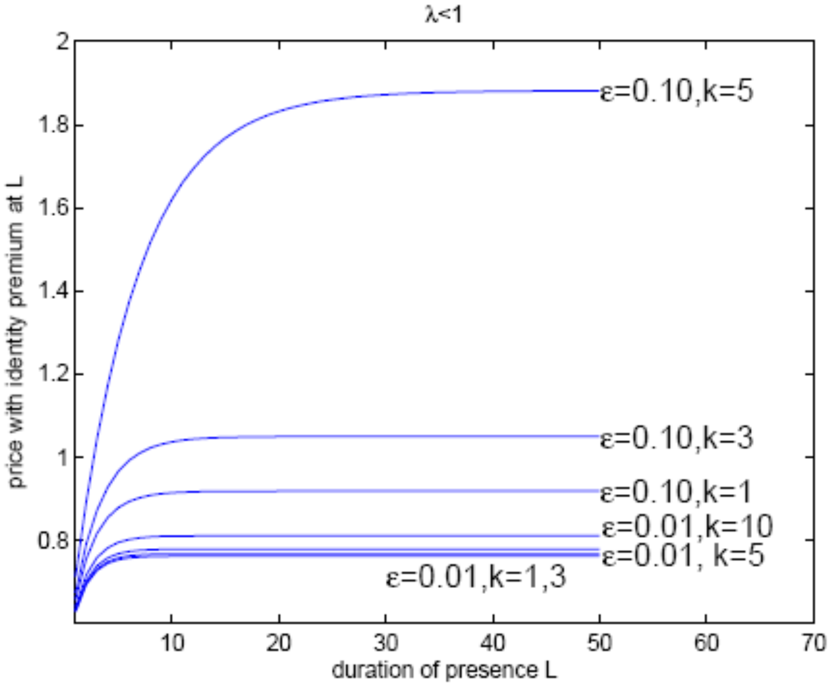
For $\lambda = 1$, the identity premium function becomes:

$$f(L) = L(1 - \phi - \xi/\gamma) \text{ for } L > 0 \qquad (4)$$

(iii) Let N_h be the number of transactions a fully cooperative (honest) provider can participate till it is mistakenly blacklisted, and let N_c be the number of bad transactions an intentionally malicious provider can benefit from defecting until eliminated from the system, respectively. We have $E[N_h] > 1/\varepsilon^k$ and $E[N_c] < 1/(1-\varepsilon)^k$.

The results (i,ii,iii) hold even in presence of strategic manipulation of ratings by agents.

Example



Limit price above original price:
Acceptable to buyers?

Inefficiency?

Limit price below original price:
Acceptable to sellers?

- original price u=1
- cheating gain 50% of original price
- Initial price 0.5 (Φ = 0.5)

Eliminating Inefficiency

- If the cheating gain is sufficiently small and the dishonesty detector is sufficiently accurate the price remains bounded and can approach any limit price by properly choosing Φ
  - Thus for a finite number of services this inefficiency can be eliminated
  - For infinite number it can be kept extremely small (order of provisioning a few services)

- Rationale to accept the scheme
  - Providers: without premium no trade at all
  - Consumers: no risk if providers are rational (provably)

Summary

- Establishing trust is a fundamental challenge in information systems

- After 20 years of work related to trust, what can we say?
  - There is not a single model of trust or universal trust mechanism
    - Assumptions
    - Context
    - Availability of data
    - Identity
  - Establishing trust remains a never ending challenge

Acknowledgements

Zoran Despotovic
Principal Engineer
Huawei

Philippe Cudre-Mauroux
Associate Professor
University Fribourg

Le-Hung Vu
R&D Engineer
NextThink

Alexandra Olteanu
Social Good Fellow
IBM Watson Research Center

Panayiotis Smeros
PhD Student
EPFL

References

- [Li et al 2004] Xiong, Li, and Ling Liu. "Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities." IEEE transactions on Knowledge and Data Engineering 16.7 (2004): 843-857.
- [Despotovic et al 2004] Despotovic, Zoran, and Karl Aberer. "P2P reputation management: Probabilistic estimation vs. social networks." Computer Networks 50.4 (2006): 485-500.
- [Cudre-Mauroux et al, 2006] Cudre-Mauroux, Philippe, Karl Aberer, and Andras Feher. "Probabilistic message passing in peer data management systems." 22nd International Conference on Data Engineering (ICDE'06). IEEE, 2006.
- [Vu et al 2011] Vu, Le-Hung, and Karl Aberer. "Effective usage of computational trust models in rational environments." ACM Transactions on Autonomous and Adaptive Systems (TAAS) 6.4 (2011): 1-25.
- [Olteanu et al 2013] Olteanu, Alexandra, et al. "Web credibility: Features exploration and credibility prediction." European conference on information retrieval. Springer, Berlin, Heidelberg, 2013.
- [Vu et al 2014] Vu, Le-Hung, Jie Zhang, and Karl Aberer. "Using identity premium for honesty enforcement and whitewashing prevention." Computational Intelligence 30.4 (2014): 771-797.
- [Smeros et al 2019] Panayiotis Smeros, Carlos Castillo, Karl Aberer: Evaluating the Quality of Scientific News Reporting from Social and Content Indicators, In *The International World Wide Web Conference 2019.*