

Lecture #13: Reputation Systems

Prof. Dr. Sven Seuken
26.4.2012

Housekeeping

- Questions? Concerns?
- Mturk assignment:
 - Bug → new version online
 - Extension until next Thursday
- Last theory assignment
 - Will return at the end of the class
 - Study our solution in detail!
- Course evaluation

Recap: Social Choice

- Mechanism Design vs. Social Choice
- Condorcet Consistency

Theorem 12.3 (Arrow's impossibility theorem). *If $|A| \geq 3$, any social ranking rule r that satisfies unanimity and IIA is dictatorial.*

- Relax IIA \rightarrow other axioms
- Get as close to Condorcet consistency as possible (swap distance)
- Dodson rule, Kemeny rule
- Schulze rule

Outline: Reputation Systems

1. Moral Hazard vs. Adverse Selection
2. Discuss the examples
3. Discuss hybrids
4. Distinguish from other systems
5. Why do you use rep systems?
6. Why do you leave feedback?
7. eBay Re-Design

Moral Hazard

- “hidden action”
 - agent does not bear full consequences
- Examples: car insurance, eBay
- Possible solution: sanctioning mechanism, i.e., punish bad behavior
- Assumption: each agent is capable of good and bad behavior

Adverse Selection

- “information asymmetry”
 - one agent does not know the “true” quality
 - is not willing to pay for high quality
 - high-quality sellers leave the market
 - only low quality left
- Examples: used cars, products on Amazon
- Possible solution:
signaling/feedback mechanism

Reputation Systems vs. ...

- vs. Recommender Systems
- vs. Transitive-Trust Systems

Examples

- Is this a serious problem?
- What is the reason for the problem?
- What is the result of the problem?
- How is it being handled right now?
- What would be a better solution?

Moral Hazard Examples

- Fund manager (Robert)
- Banking crisis (Mengia)
- Fixed-salary employment (Martin, Andrea)
- Mountain climbing (Basil)
- Human Computation (Patrick)
- German officials (Malte)
- Parking your car in Italy (Andras)
- Leaving the stove on (Jessica)
- Cheating in a relationship
- Insurance fraud (pretending you had an accident)
- Going around eBay/Ricardo to avoid fees
- Discounter repacking goods, selling after expiration date
- Company promising certain benefits

Adverse Selection Examples

- Insurance market (Robert, Martin, Andrea, Jessica)
- Going to a concert (Andrin)
- Video game market (Balthasar)
- Booking a hotel (Andras)
- Shopping on a market in Istanbul (Jan)
- Job market (employers with different skills)
- Cities with different tax rates
- Sloppy website design for clients
- Employers offering fixed salaries

Other Examples

- IMDB
- Amazon
- Facebook
- Google+
- Testeo
- Holidaycheck, Tripadvisor
- Stackoverflow

Why do you leave feedback?

- I don't, lazy 😊
- Leaving feedback on Stackoverflow ultimately helps myself
- Leaving feedback on eBay/Ricardo helps other buyers/sellers
- I rate items on video game website to help other players find good games
- I want to help other people make good decisions
- I like giving back to the community

Whitewashing

- Problem
- Solutions...

The eBay Reputation System

- Status quo 2007:
 - Conventional feedback by buyers and sellers
 - Feedback is published immediately
 - Can only be removed by court ruling, when the buyer doesn't pay, or when both traders agree

Ebay's Problems

- Reciprocal Feedback
 - Retaliation
- Fear that this reduces informativeness

High Degree of Reciprocation

TABLE 1: FEEDBACK GIVING AND CONTENT, CONDITIONAL PROBABILITIES AND CORRELATIONS

Feedback giving probability	Partner did not yet give FB	Partner gave FB already
Buyer	68.4%	74.1%
Seller	51.4%	87.4%

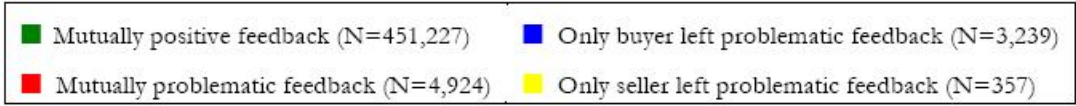
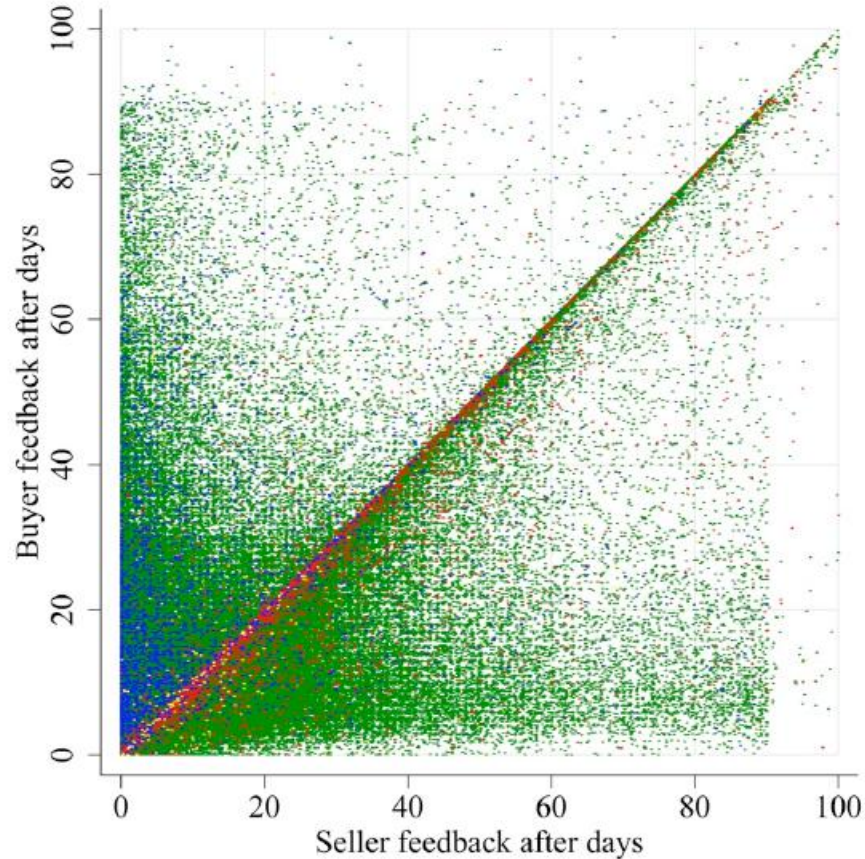
Kendall's tau correlations between seller's and buyer's feedback

Country	FB content correlation						FB giving correlation	
	All cases		Buyer gave FB second		Seller gave FB second			
	N	tau	N	Tau	N	tau	N	tau
All	458,249	0.710	139,772	0.348	318,477	0.884	725,735	0.693
Australia	20,928	0.746	6,040	0.340	14,888	0.928	31,990	0.752
Belgium	8,474	0.724	3,097	0.464	5,377	0.880	12,301	0.684
France	24,933	0.727	8,095	0.423	16,838	0.883	39,104	0.703
Germany	133,957	0.656	45,836	0.331	88,121	0.840	192,565	0.644
Poland	457	1.000	172	-	285	1.000	1,134	0.783
U.K.	93,266	0.694	31,316	0.379	61,950	0.875	143,877	0.692
U.S.	176,009	0.746	45,133	0.313	130,876	0.911	302,213	0.701

Notes: Observations where feedback was eventually withdrawn are not included in correlations. In the cell marked with “-“, the standard deviation is zero. All other correlations are highly significant.

Timing

FIGURE 1: CONTENT AND TIMING OF MUTUAL FEEDBACK ON EBAY



Notes: The scatter plot reports about 460,000 observations where both transaction partners gave feedback. 'Problematic' feedback includes neutral or withdrawn feedback.

Two Proposals

- Blind Feedback
- Detailed Seller Rating (DSR)

Informativeness?

- How do we compare the quality of two reputation systems?
 - What is the goal (here: on eBay)?
- Differentiate between “good” and “bad”!
What does that mean?

Informativeness

- Idea:
 - Experimental study, induce “true scores”
 - Measure **correlation** between “reputation scores and true scores”

 - Alternatively: let users trade
 - Measure **correlation** between “reputation scores” and future seller behavior

eBay Lab Study

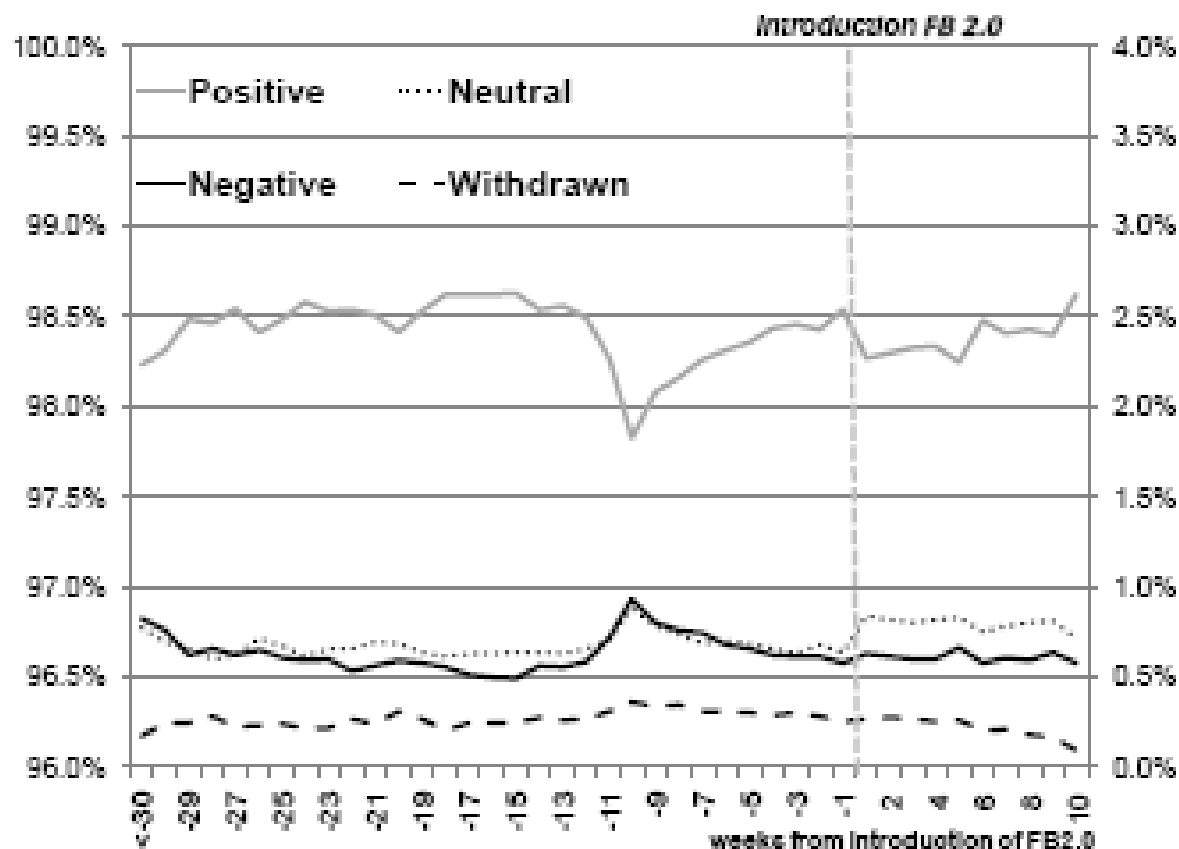
TABLE 7: DETERMINANTS OF SELLER AVERAGE PROFIT, TOBIT COEFFICIENT ESTIMATES
(ROBUST STANDARD ERRORS CLUSTERED ON MATCHING GROUP, ROUNDS 1 TO 10)

Dep var Model	Seller average future profit			
	1		2	
	Coeff	(StdErr)	Coeff	(StdErr)
Constant	73.61 ***	(4.128)	70.45 ***	(4.442)
S FScore			3.04 ***	(0.489)
S FScore* <i>Blind</i>			1.36 *	(0.748)
S FScore* <i>DSR</i>			-2.30 ***	(0.763)
S DSR Avg			3.92 ***	(1.083)
Quality* <i>Baseline</i>	0.079	(0.083)	-0.019	(0.056)
Quality* <i>Blind</i>	0.175 **	(0.082)	0.098	(0.062)
Quality* <i>DSR</i>	0.179 ***	(0.0478)	0.034	(0.042)
Nosale	-53.92 ***	(5.00)	-43.75 ***	(6.533)
N	2400		2400	
Restricted LL	-11398.2		-11180.5	

Note: *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively. *Blind* and *DSR* are treatment dummies. S FScore denotes the feedback score of the seller, and S DSR Avg the average DSR score. The Period variable is omitted because the associated coefficient is small and insignificant.

Evolution of CF

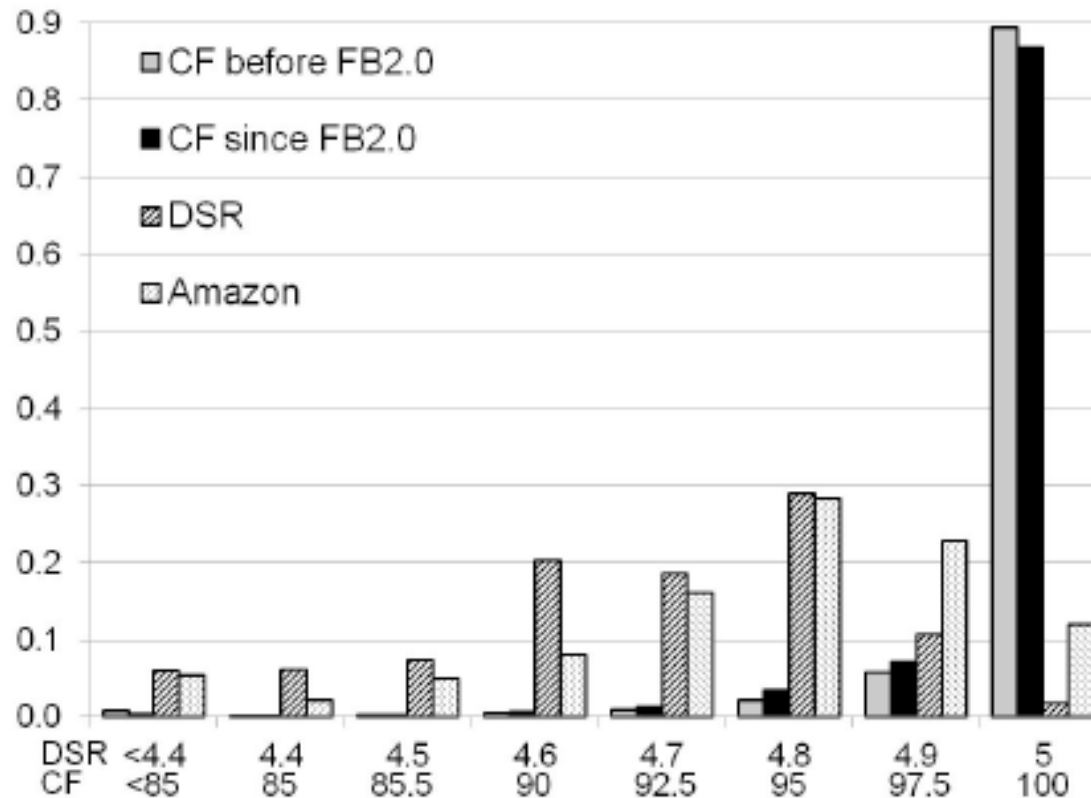
FIGURE 5: EVOLUTION OF POSITIVE, NEUTRAL, NEGATIVE AND WITHDRAWN FEEDBACK BEFORE AND AFTER INTRODUCTION OF FEEDBACK 2.0



Notes: The figure is based on about 7 and 3 million individual feedbacks in the 30 weeks before and the first 10 weeks after introduction of Feedback 2.0, respectively, in the pilot countries Australia, Belgium, France, Poland and UK. Positive feedback is plotted on the left y-axis, all other feedback on the right y-axis.

Distribution of Scores

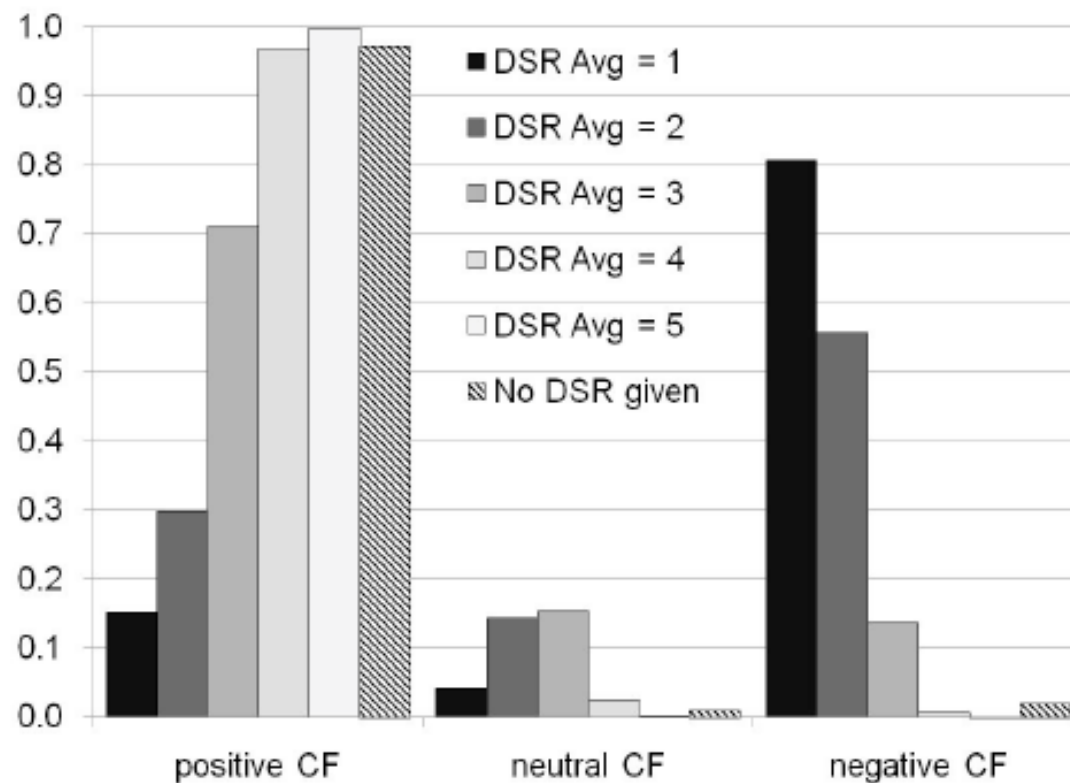
FIGURE 6: DISTRIBUTION OF AVERAGE CF AND DSR SCORES IN MEMBER PROFILES



Notes: DSR and Amazon.com's 1-5 range and CF percent positive's 0-100 range are divided in the same number of categories and are aligned at the x-axis. EBay data is based on the feedback of the same 27,759 members from Australia, Belgium, France, Poland and UK, received as seller in Jan/Feb 2007 and March/Apr/May 2007, respectively. Inclusion criterion was more than 10 DSRs in at least one DSR category. Amazon data is based on 9,741 Amazon market place sellers.

Distribution

FIGURE 7: DISTRIBUTION OF CF CONDITIONAL ON AVERAGE OF CORRESPONDING DSRs



Notes: To calculate the DSR average we take all available of the up to four DSR ratings per feedback, average and round to integer. Thus, a DSR average of 1 implies two or three ratings of 1 and at most one rating of 2.

Market Design Takeaways

- Strategic behavior is present!
- Details matter a lot (e.g., timing)
- Statistical analysis to detect problems
- Field studies to evaluate alternatives
- Lab experiments to evaluate alternatives
- Other important concerns in the real world..