Cluster-Based Analysis and Recommendation of Sellers in Online Auctions

Mikolaj Morzy Juliusz Jezierski

Institute of Computing Science
Poznan University of Technology
Piotrowo 3A, 60-965 Poznan, Poland

3rd International Conference on
Trust, Privacy, and Security in Digital Business
TrustBus 2006
Krakow, Poland, September 2006

Outline

Some Numbers...

- 63% of online population engaged in e-commerce in 2006
- 18% of global sales in 2006
- over 250 online auction sites (C2C business)
- over 1.3 million transactions committed daily
- the size of eBay
  - 95 million registered users
  - 5 million transactions per week
  - 12 million items posted at any given time
  - net revenues of $1.1 billion (40% increase, Q2 2005)
  - operating income of $380 million (49% increase, Q2 2005)
  - net income of $290 million (53% increase, Q2 2005)

Success factors

- no constraints on time
- no constraints on place
- reduced prices due to abundance of sellers and buyers
- business model of 24/7/365
- varitety of auction protocols and offered goods
- gambling experience

Online Auction Fraud

First some numbers

- 73% of unconvinced: security of payment, delivery issues, warranty terms (EuroBarometer)
- 48% of complaints concerning e-commerce involve online auction fraud (FTC)
- total loss of $ 437 million in one year
- 63% of complaints about Internet fraud concerned online auctions, $ 478 per capita
- popular methods: bid shielding, bid shilling, accumulation

Current solution

“positive”, “neutral”, and ”negative” feedbacks, but . . .

- virtual bidders drive up reputation score (ballot stuffing)
- sellers create cliques of bidders
- “bad-mouthing” can be beneficial
- reputation of buyers is of little importance
- sellers and buyers exposed to different types of risk

Contribution

Our contribution

- new measure of reputation for sellers in online auctions
- clustering of densely connected sellers
- automatic recommendation generation
- experimental evaluation of the proposal

Related Work

- reputation systems: develop long-term relationships (Resnick et al.)
- deficiencies of feedback-based reputation systems (Malaga)
- complaint-only trust model (Aberer et al.)
- recursive definition of credibility (Morzy et al.)
- a trusted third party (Ba et al., Snyder)
- using trust and distrust statements between individuals (Guha et al.)

Basic Definitions

- given a set of sellers $S = \{s_1, s_2, \ldots, s_m\}$
- sellers s_i and s_j are *linked* if
  - at least *min_buyers* bought from both s_i and s_j
  - the closing price of each auction was at least *min_price*
- *strength* of a link, denoted $link(s_i, s_j)$, is the number of connecting buyers
- *neighborhood* of a seller s_i, denoted $N(s_i)$, is the set of sellers $\{s_j\}$ who are linked to s_i
- *density* of a seller s_i, denoted $density(s_i)$, is the cardinality of seller's neighborhood $N(S_i)$

Rationale

How the thresholds are used?

- *min_buyers*: selects sellers with significant number of sales
- *min_price*: prunes low-value transactions

Rationale

How the thresholds are used?

- *min_buyers*: selects sellers with significant number of sales
- *min_price*: prunes low-value transactions

Rationale behind the measure

- buyer b_k buying from sellers s_i and s_j acknowledges both sellers
- unexperienced buyers do not link many sellers
- a link indicates similar or complementary offers (although it might be coincidental)
- clusters uncover natural groupings around product categories

Score Measure

Score

Density measure does not consider the strengths of links between sellers

$$score(s_i) = \sum_{s_j \in N(s_i)} density(s_j) * \log_{min\_buyers} link(s_i, s_j)$$

Resistance to Fraud

Density measure is very resistant to fraud

- linking to a single seller induces a cost of $min\_buyers * min\_price$
- linking to multiple sellers repeats the above procedure several times
- other sellers used to rate a current seller - harder to influence (!)

Recommendations

- let R denote a set of target n sellers
- let $d(s_i, s_j)$ denote the distance between s_i and s_j (the lenght of the shortest path between s_i and s_j)

Group Density

$$density(R) = \frac{\sum_{s_r \in R} density(s_r)}{\sum_{(s_p, s_q) \in R \times R} d(s_p, s_q)}$$

Recommendations

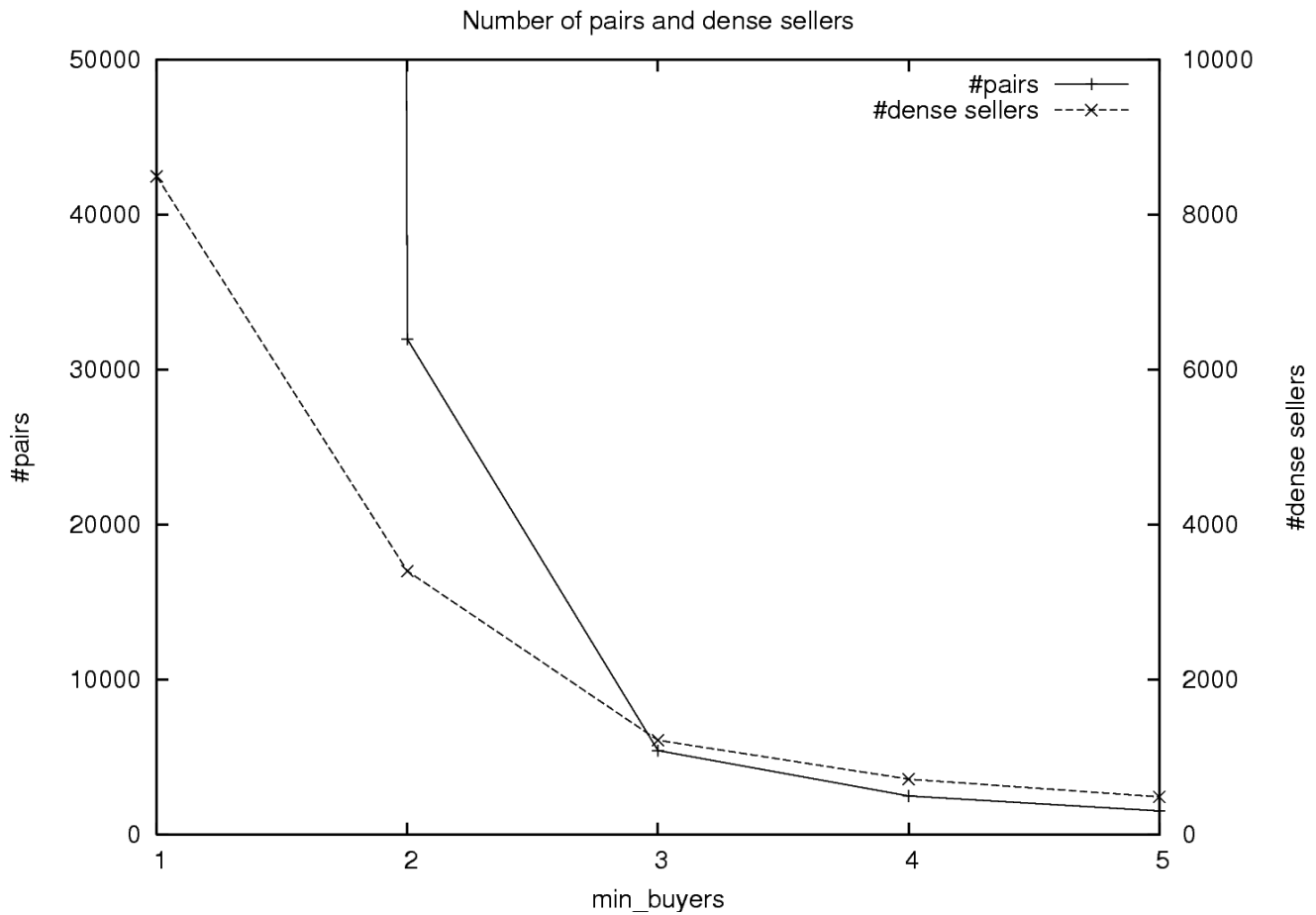
When displaying top n sellers as a recommendation for currently selected seller s_i we are trying to find the set $R(s_i)$ of sellers who are characterized by high group density and who are close to a given seller s_i

$$R(s_i) = \arg\max_{R} \frac{density\,(R)}{\sum_{s_r \in R} d(s_i, s_r)}$$

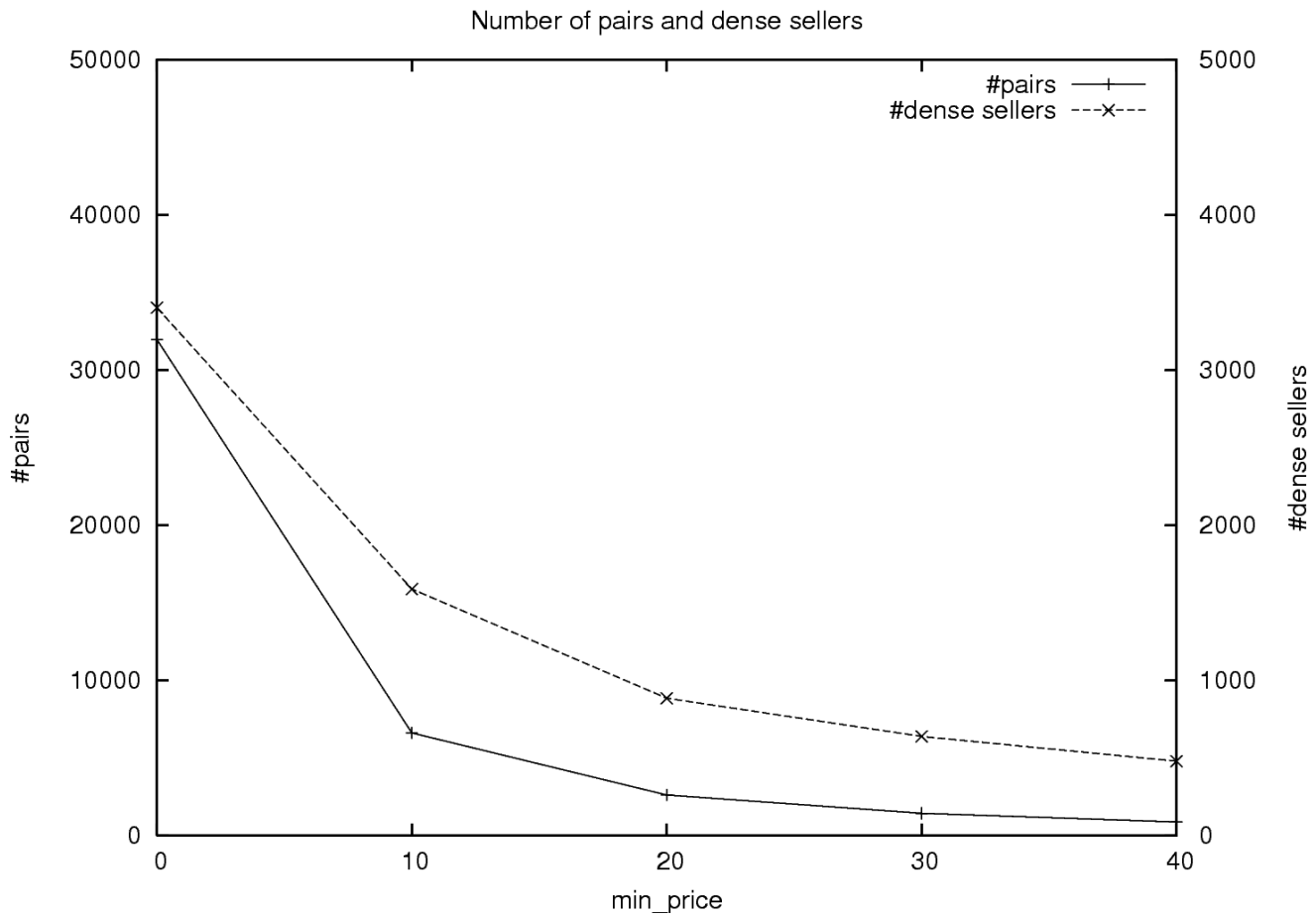
Synthetic Datasets

- `www.allegro.pl`
- 440 000 participants
- 400 000 auctions
- 1 400 000 bids
- analysis: 10 000 sellers, 10 000 buyers, 6 months of data

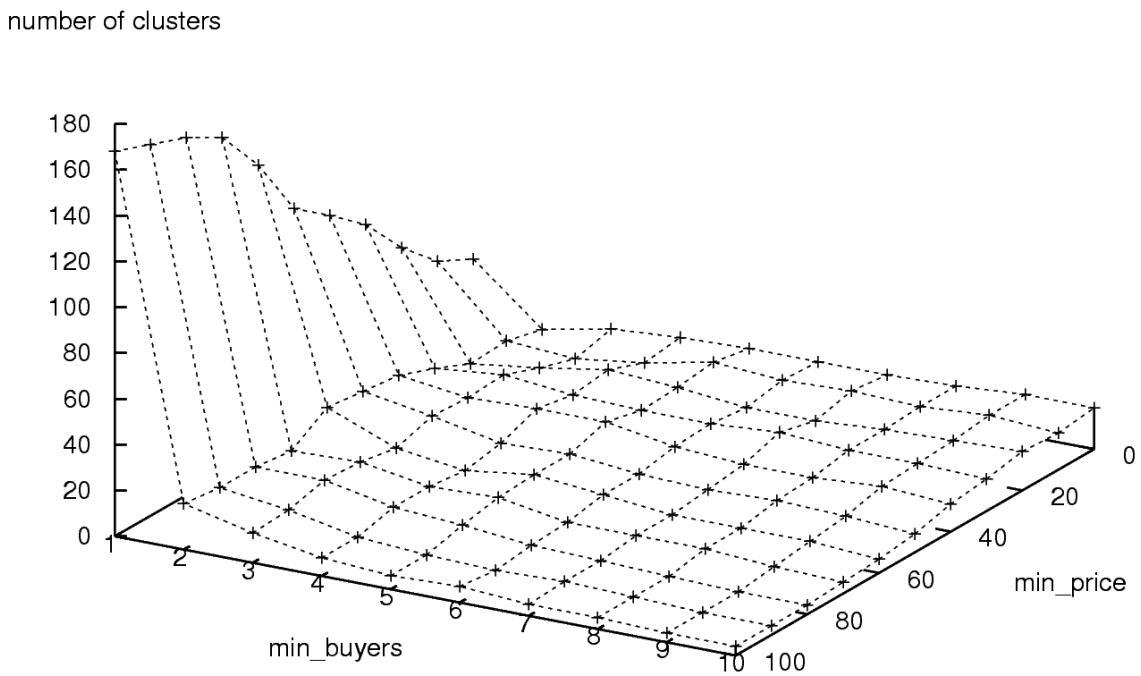
Number of pairs and dense sellers w.r.t. *min_buyers* threshold



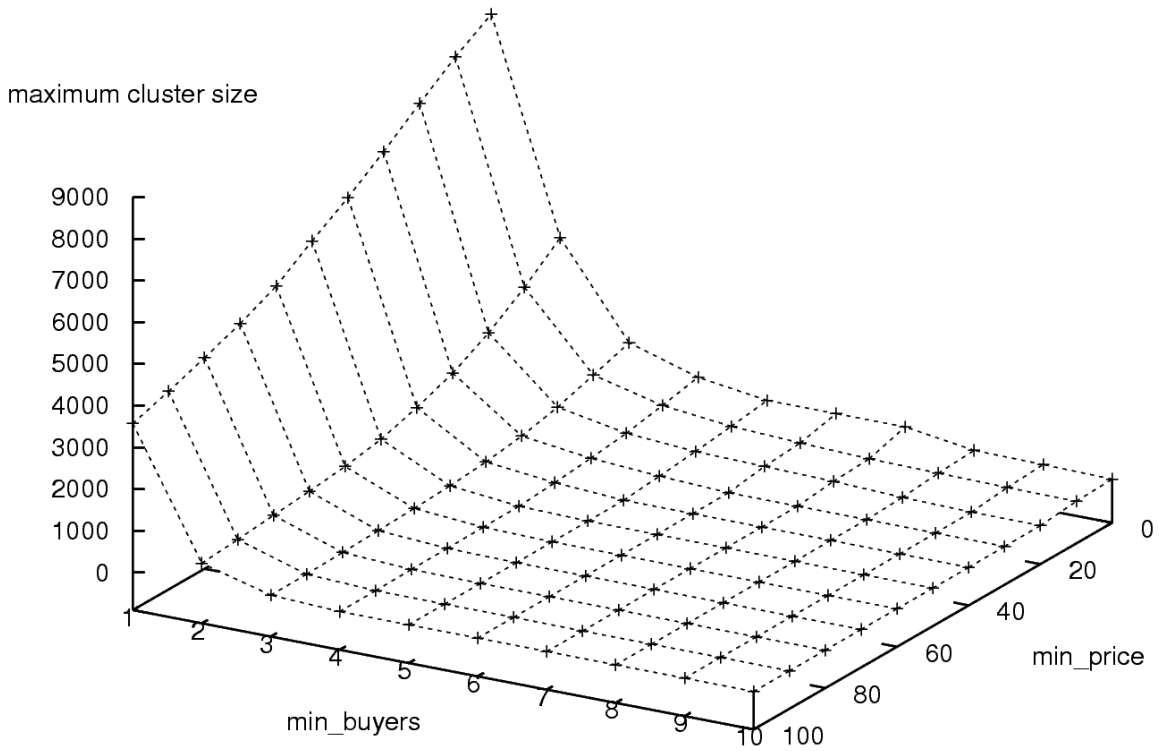
Number of pairs and dense sellers w.r.t. *min_price* threshold



Number of discovered clusters

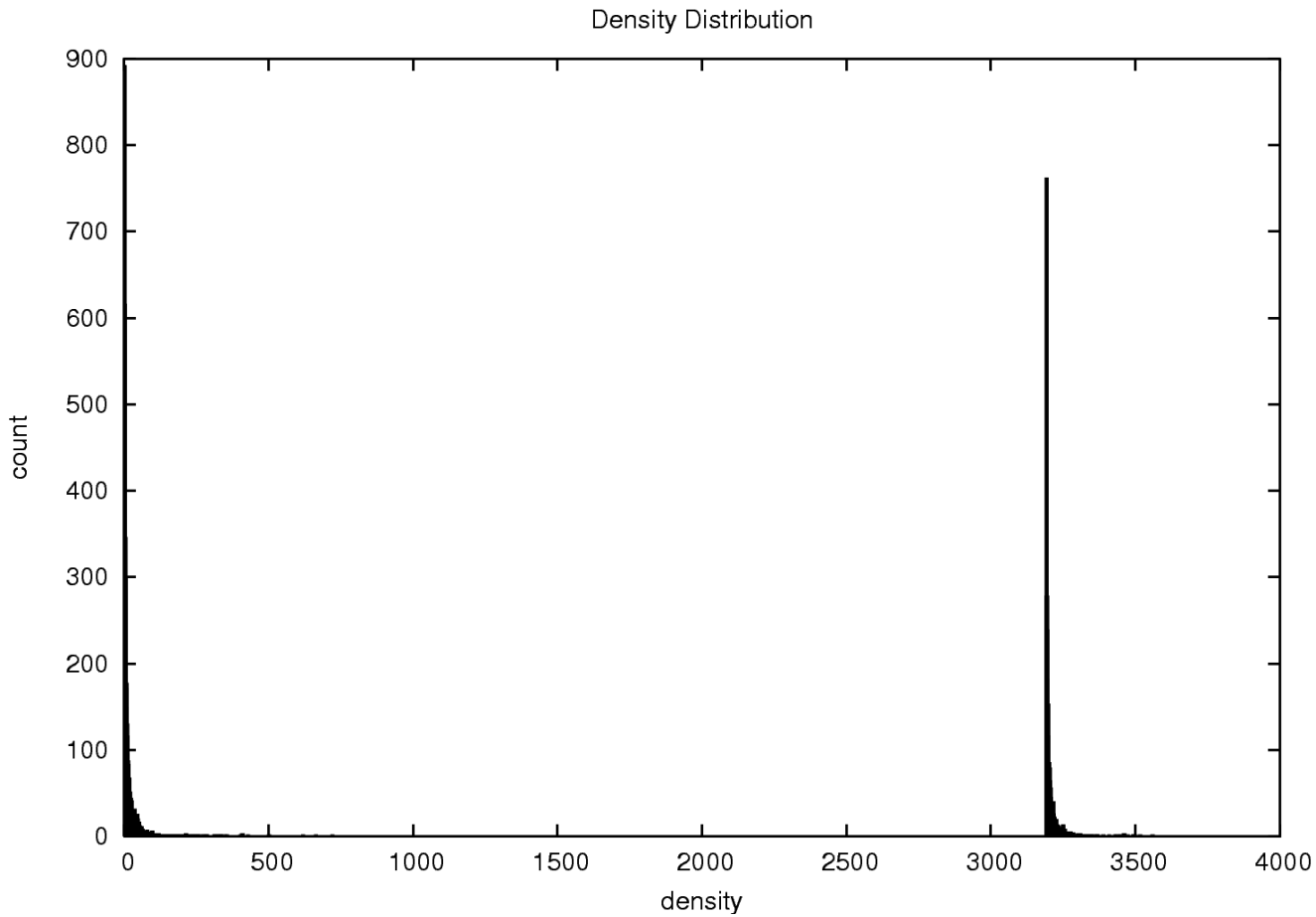


Maximum cluster size



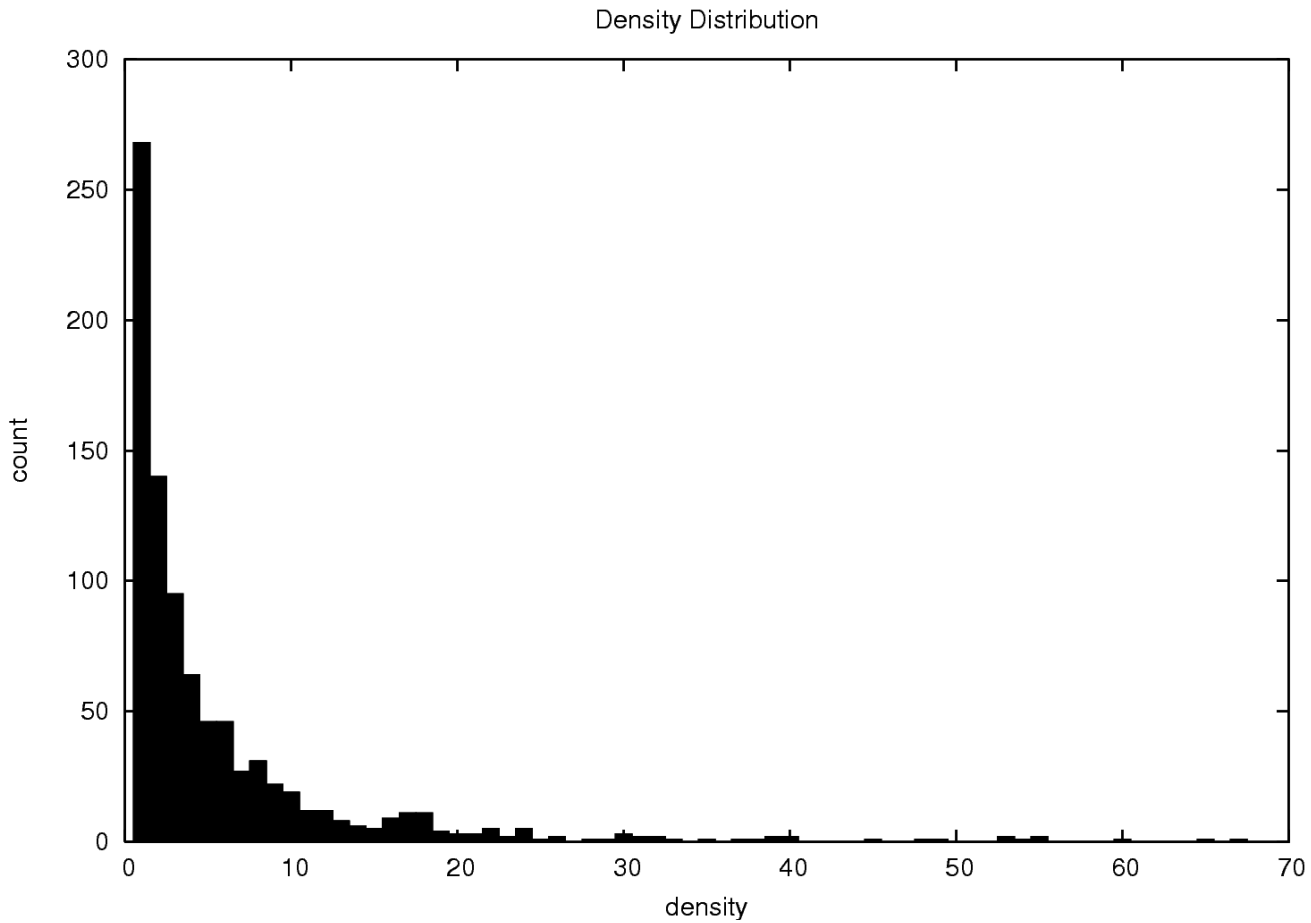
Density distribution

No constraints on *min_buyers* and *min_price*



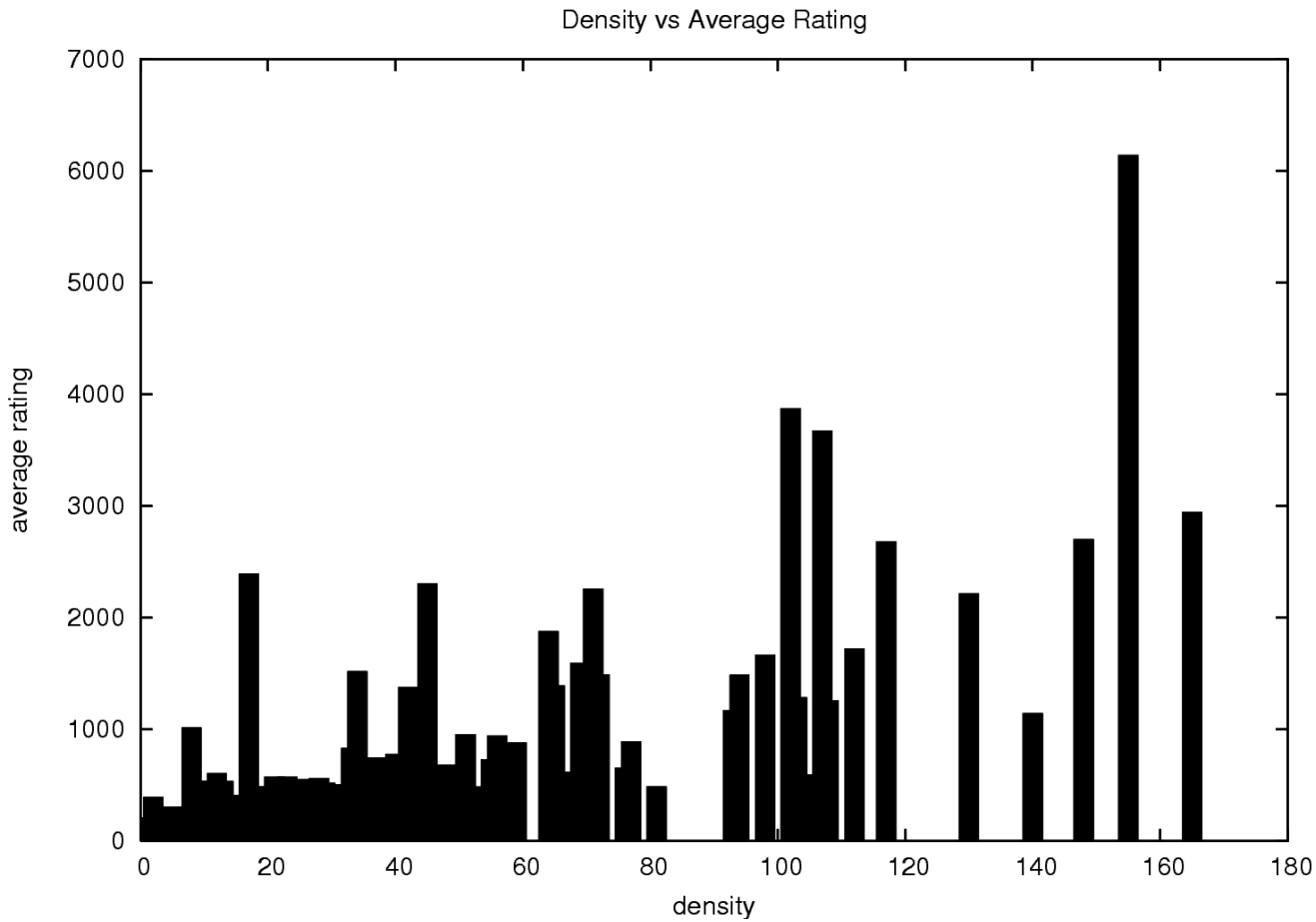
Density distribution

$min\_buyers = 2$, $min\_price = \$\,20$



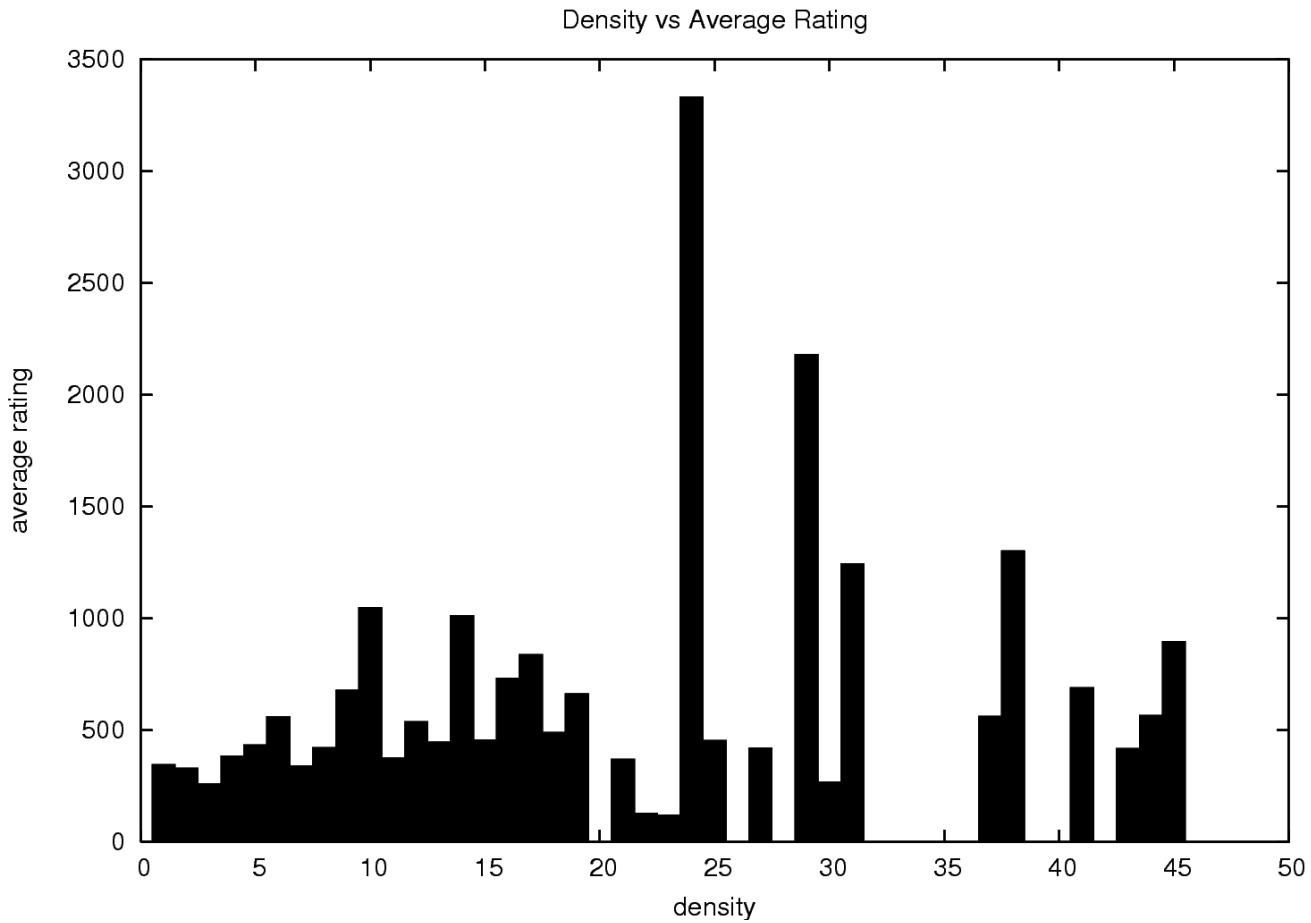
Average rating w.r.t. density

$min\_buyers = 3$, $min\_price = \$\,0$



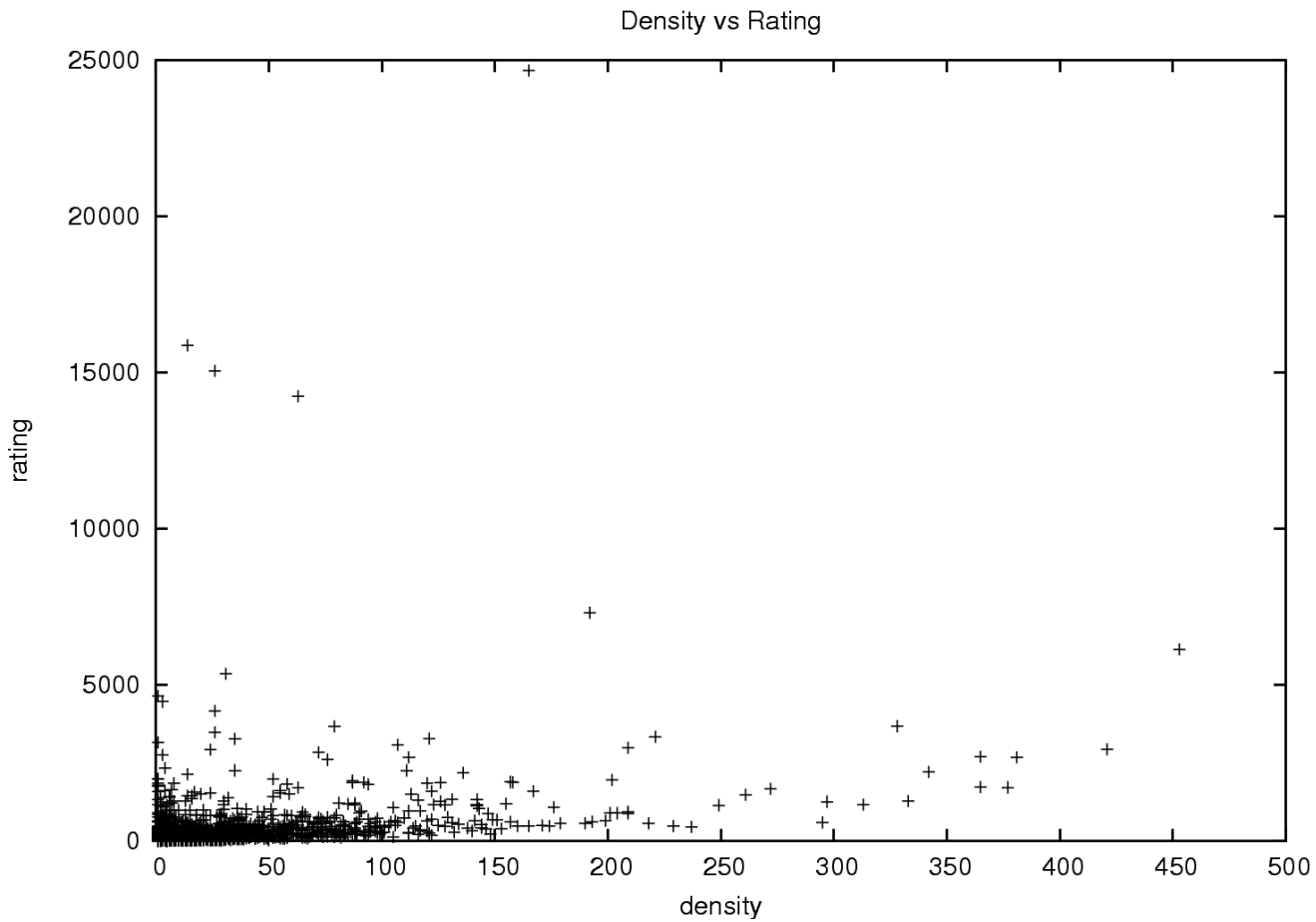
Average rating w.r.t. density

$min\_buyers = 2$, $min\_price = \$\,30$



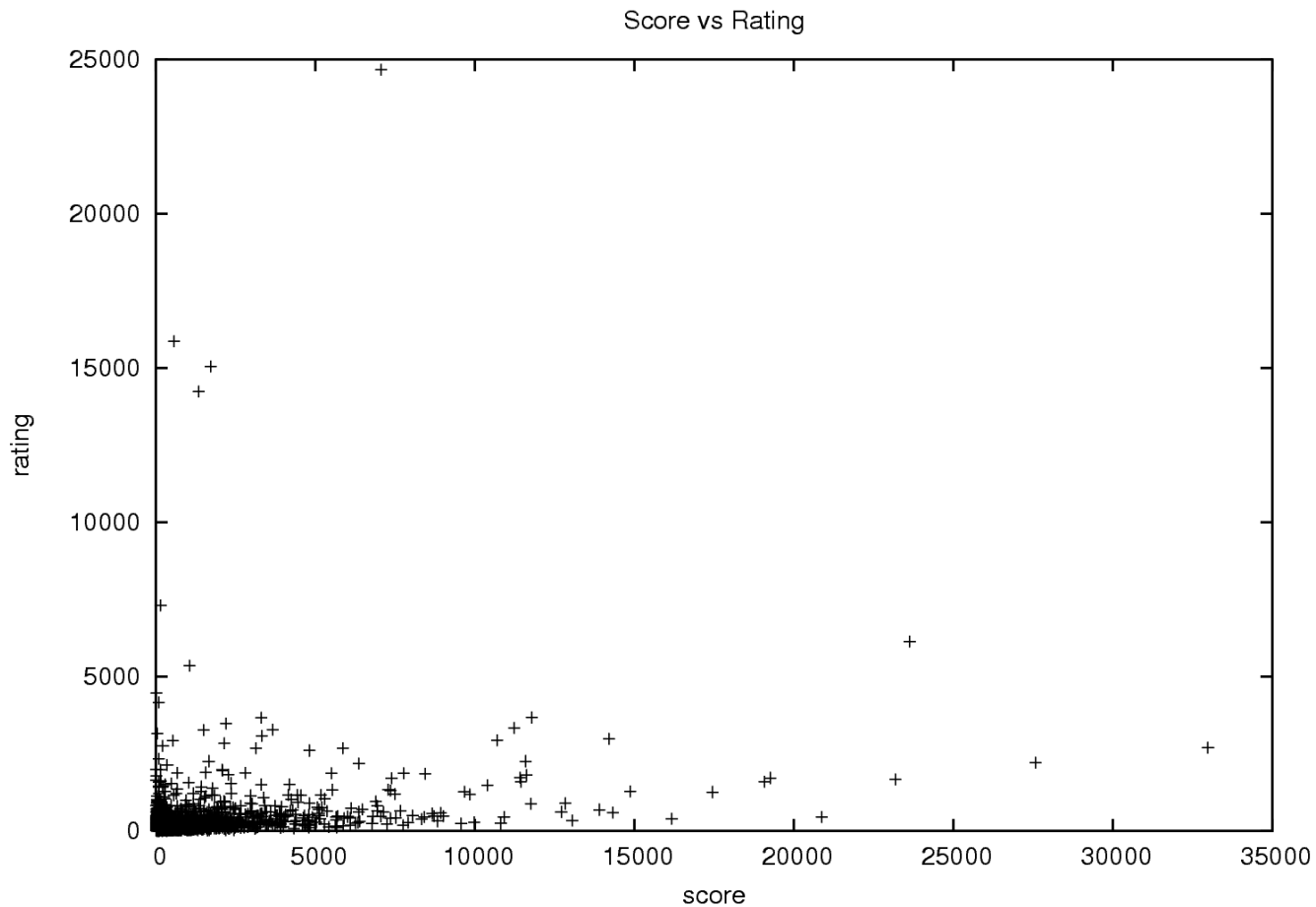
Projection of density on rating

$min\_buyers = 2$, $min\_price = \$\,0$



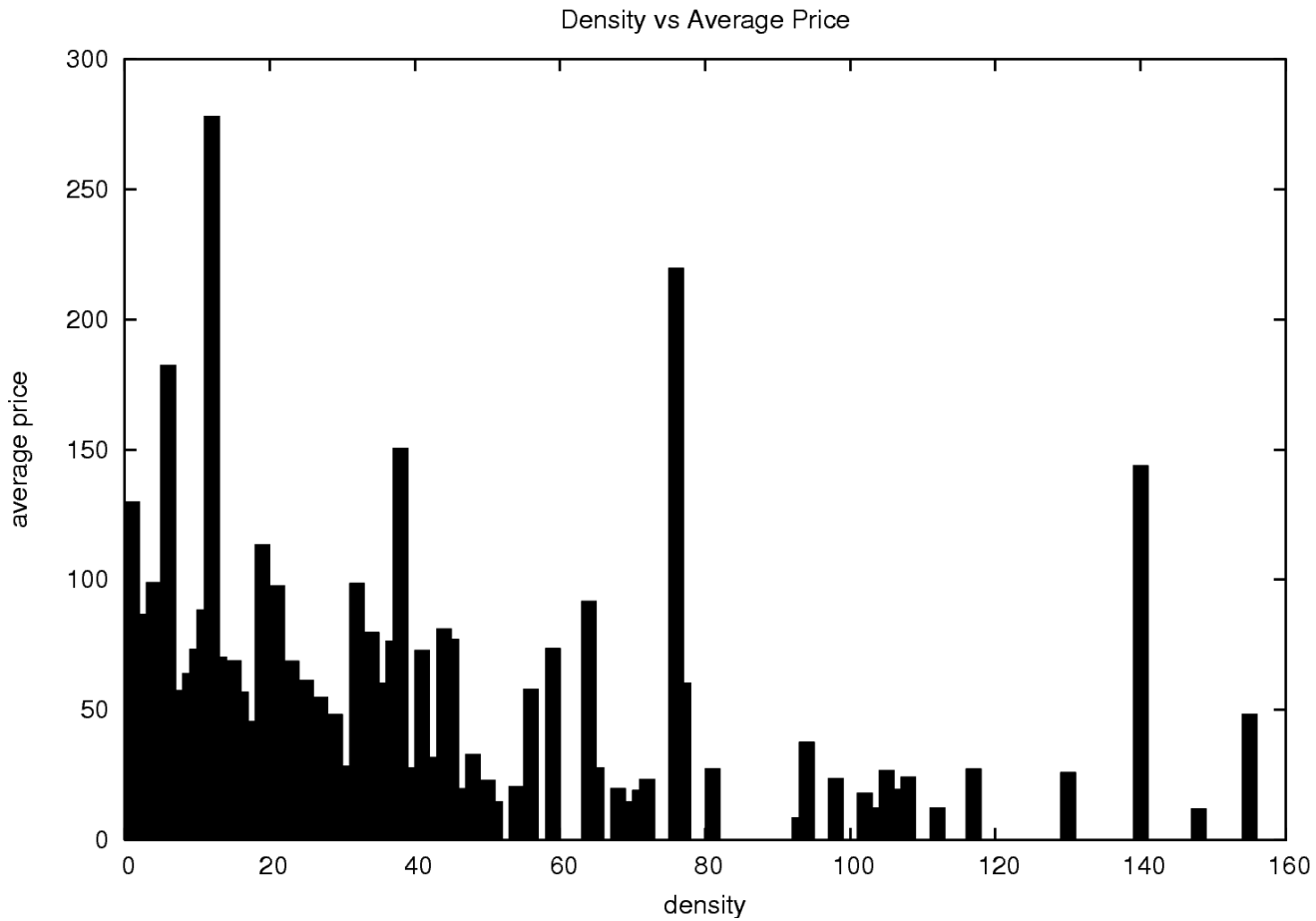
Projection of score on rating

$min\_buyers = 3$, $min\_price = \$\,0$



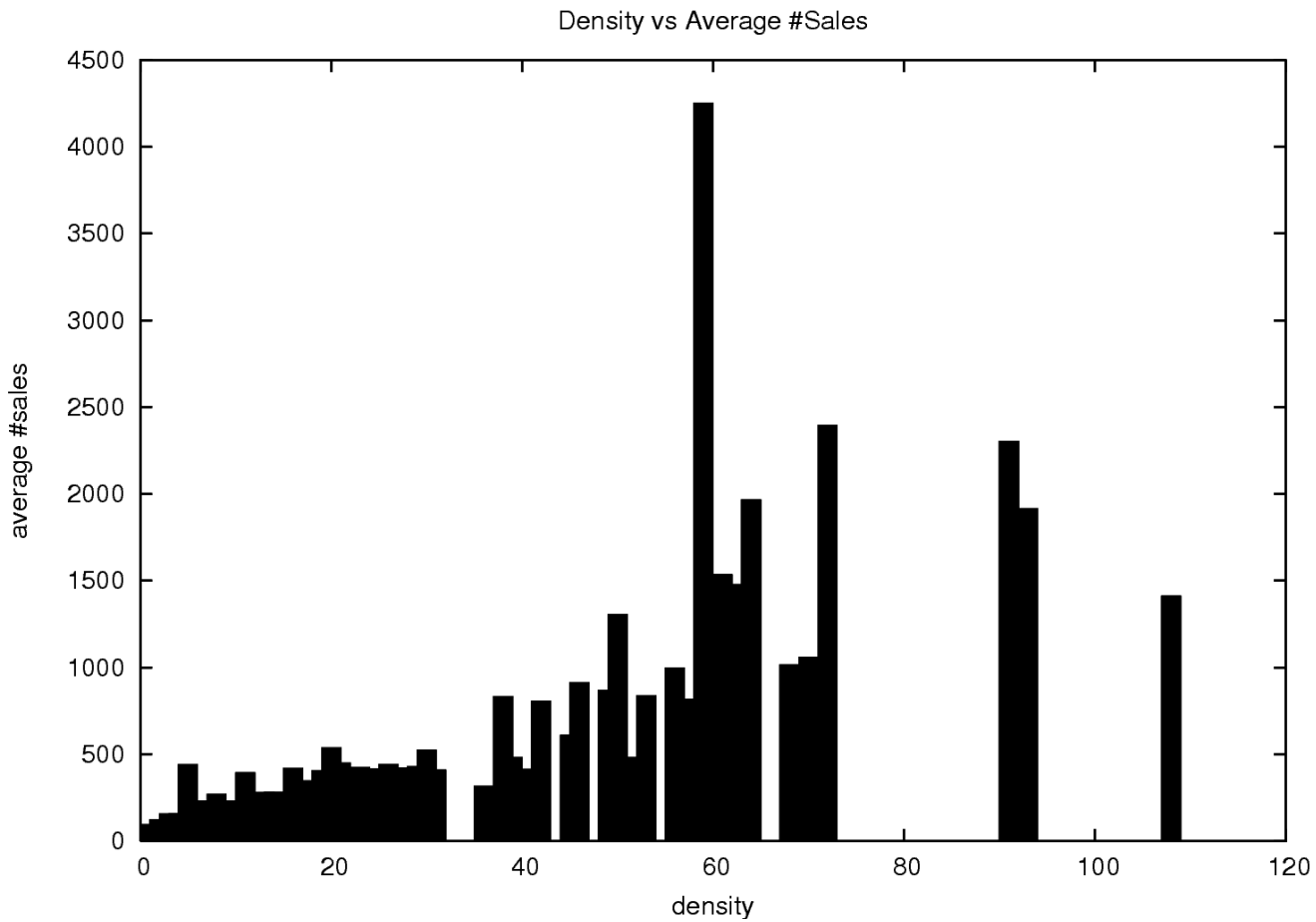
Average price w.r.t. density

$min\_buyers = 3$, $min\_price = \$\,0$



Average number of sales w.r.t. density

$min\_buyers = 4$, $min\_price = \$\,0$



Conclusions and Future Work

Conclusions

Discovered clusters of densely connected sellers

- predict future behavior of sellers
- allow description-independent and taxonomy-independent recommendations
- resist fraud and manipulation

Conclusions and Future Work

Future Work

- effective use of negative and missing feedbacks
- context-aware recommendations
- further investigation of clusters' properties