DS 707
Association Rules Deliverable: Team 23
Dating Analytics

Rishabh Manoj (IMT2013035)
Revanuru Karthik (IMT2013033)
Nigel Fernandez (IMT2013027)
Anirudh Ravi (IMT2013005)

November 25, 2016

Contents

List of Tables

List of Figures

1 Predictive Analytics using Association Rules

We used association rules to find patterns (co-occurrences of features) in the decision process of both men and women. Since men and women may have different methods in accepting a potential partner, we divided our data set gender wise and mined for association rules in these segments (men and women) separately.

Our first step was to convert the data type of our attributes from ordinal to nominal. We converted the continuous rating scale from $[0, 100]$ to three segments $[0, 2/3 \cdot \mu)$, $[2/3 \cdot \mu, 4/3 \cdot \mu)$ and $[4/3 \cdot \mu, 100]$ where μ is the average. These segments were labelled as *low (L), medium (M) and high (H)* thereby converting the data type to categorical.

After this we divided our data type to two segments, namely, (1) men deciding on women and (2) women deciding on men. These two segments were further divided into two parts, namely, (1) decision is yes and (2) decision is no. Association rules were mined for each of the four segments. The top 20 rules (by lift) in each segment were visualized using *scatter, matrix, graph* and *parallel* plots.

2 Men Decide

These association rules were obtained from the male segment of the data. Since we were interested in finding patterns in the decision process, we constrained the RHS of the rule to include only the decision attribute dec_o which takes values 0 and 1 corresponding to *no* and *yes* decisions respectively.

2.1 When Men Say Yes

The RHS was constrained to dec_o = 1

Table 1: Rules when Men Say Yes

S No	LHS	RHS	Support	Confidence	Lift
1	{need_fun:H, theatre:M, movies:M, samerace:0}	{dec_o:1}	0.010	0.933	1.973
2	{need_fun:H, sports:H, theatre:M, movies:M, samerace:0}	{dec_o:1}	0.010	0.933	1.973
3	{hiking:H, theatre:M, movies:M}	{dec_o:1}	0.010	0.931	1.970
4	{hiking:H, movies:M, concerts:M}	{dec_o:1}	0.010	0.931	1.970
5	{hiking:H, theatre:M, movies:M, concerts:M}	{dec_o:1}	0.010	0.931	1.970

Interpretation: In rule 2 in table 1, men needing high fun in their partner, women with high sports interest, medium theatre interest, medium movies interest and men fine with a different race co-occurred with a yes decision by men on women.
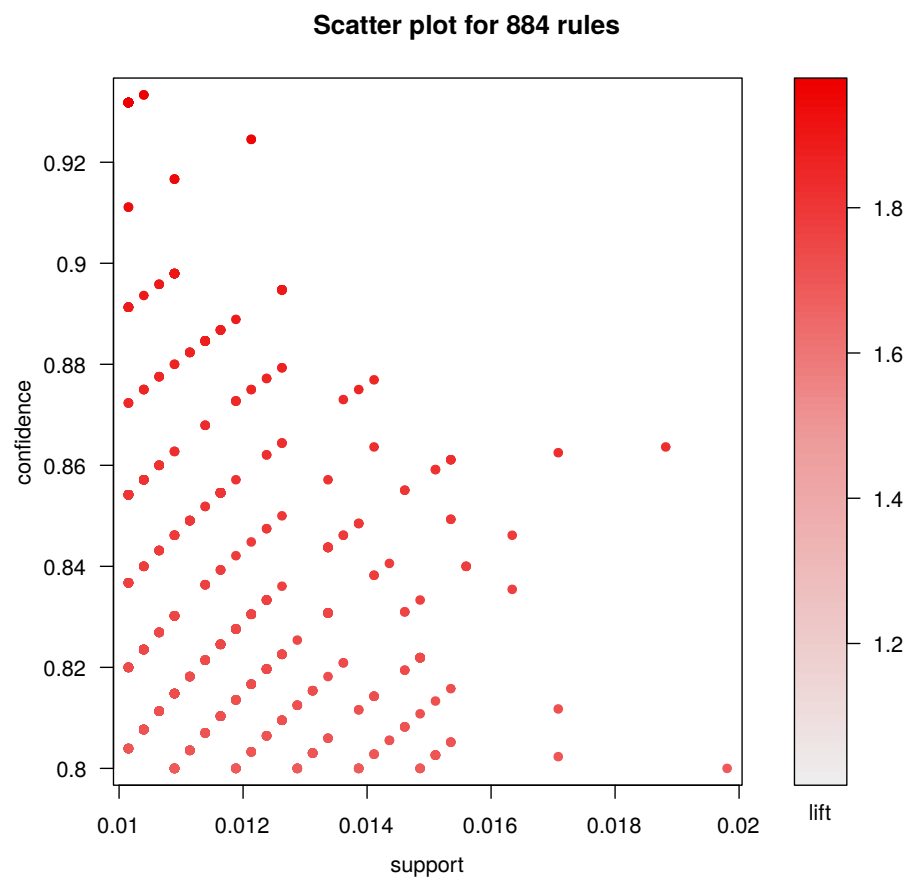
Figure 1: Scatter Plot: Rules when Men Say Yes

Grouped matrix for 884 rules

size: support
color: lift

LHS

{f_movies=M, +9 items} - 113 rules
{f_tvsports=H, +9 items} - 5 rules
{f_movies=M, +17 items} - 34 rules
{f_tvsports=H, +23 items} - 32 rules
{samerace=0, +25 items} - 24 rules
{f_exercise=H, +21 items} - 14 rules
{intel1_1.1=M, +34 items} - 28 rules
{f_hiking=H, +33 items} - 62 rules
{f_hiking=H, +36 items} - 46 rules
{f_hiking=H, +38 items} - 37 rules
{f_reading=M, +41 items} - 66 rules
{f_sports=H, +44 items} - 67 rules
{f_art=M, +42 items} - 82 rules
{f_exercise=H, +38 items} - 35 rules
{f_gaming=M, +30 items} - 13 rules
{f_music=H, +19 items} - 8 rules
{f_concerts=M, +42 items} - 46 rules
{sinc3_1=H, +41 items} - 60 rules
{f_tvsports=H, +46 items} - 60 rules
{f_art=M, +41 items} - 52 rules

RHS

{dec_o=1}

Figure 2: Grouped Matrix Plot: Rules when Men Say Yes

6

Graph for 20 rules

size: support (0.01 - 0.01)
color: lift (1.97 - 1.974)

samerace=0

f_exercise=H

fun3_1=H

f_theater=M

f_reading=H

dec_o=1
f_movies=M

f_tv=M

f_sports=H

f_hiking=H

f_concerts=M
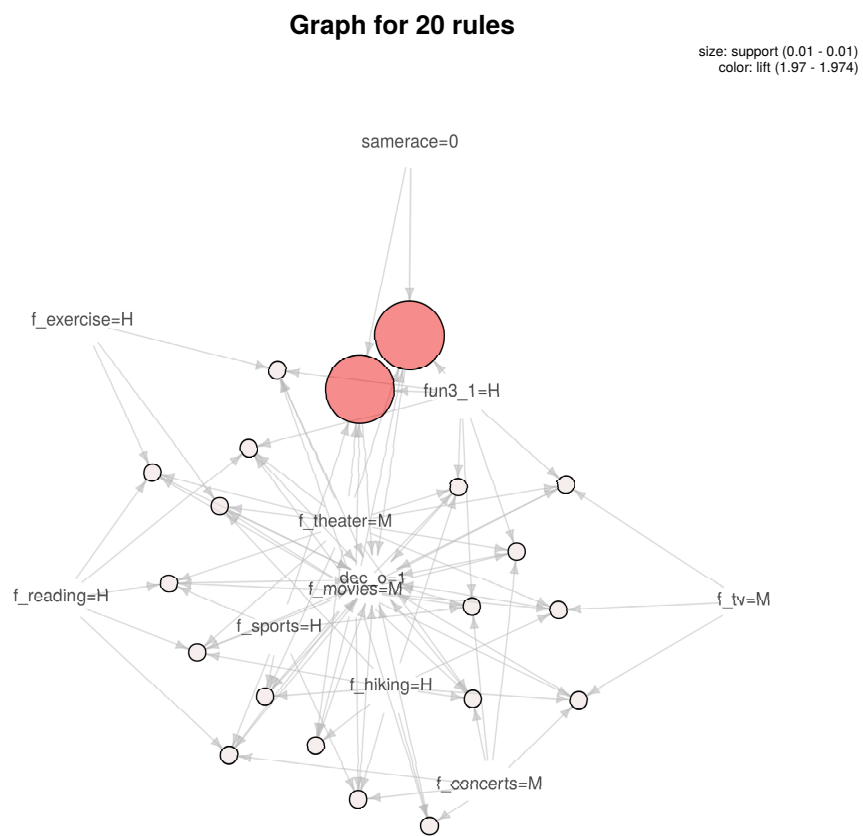
Figure 3: Graph Plot: Rules when Men Say Yes

Parallel coordinates plot for 20 rules

dec_o=1
samerace=0
f_exercise=H
f_reading=H
f_sports=H
f_tv=M
fun3_1=H
f_concerts=M
f_movies=M
f_theater=M
f_hiking=H

Position
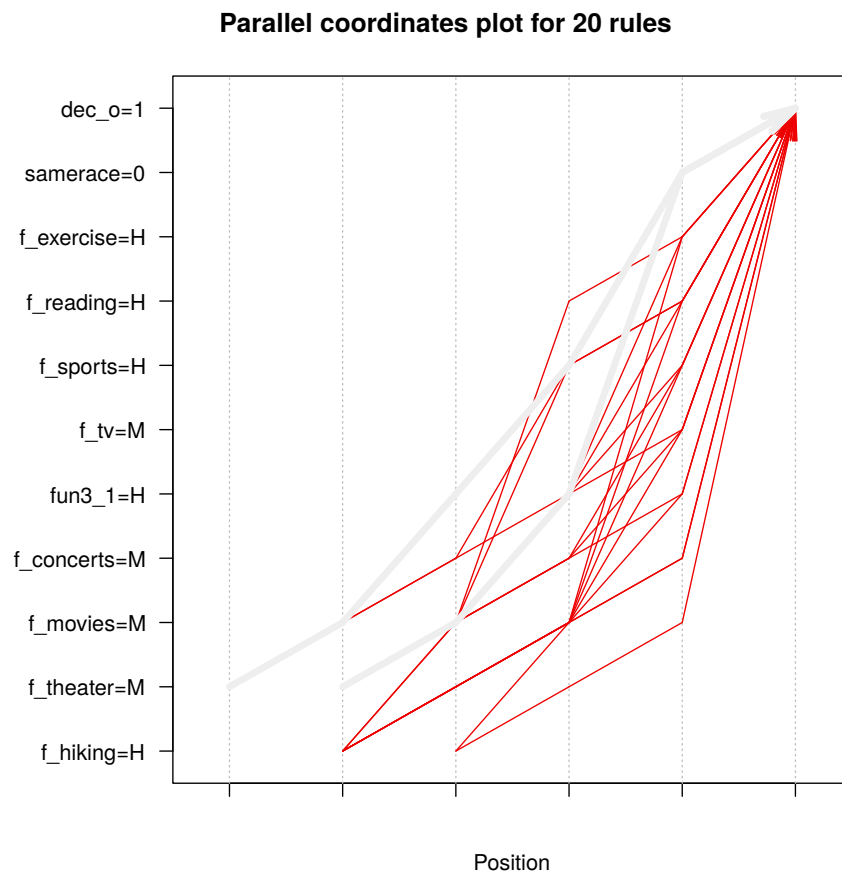
Figure 4: Parallel Plot: Rules when Men Say Yes

2.2 When Men Say No

The RHS was constrained to dec_o = 0

Table 2: Rules when Men Say No

S No	LHS	RHS	Support	Confidence	Lift
1	{sports:H, tvsports:L, museums:M, yoga:L}	{dec_o:0}	0.010	0.954	1.810
2	{sports:H, tvsports:L, art:M, yoga:L}	{dec_o:0}	0.010	0.954	1.810
3	{sports:H, tvsports:L, art:M, concerts:H}	{dec_o:0}	0.010	0.954	1.810
4	{sports:H, tvsports:L, museums:M, art:M, yoga:L}	{dec_o:0}	0.010	0.954	1.810
5	{sports:H, tvsports:L, museums:M, art:M, concerts:H}	{dec_o:0}	0.010	0.954	1.810

Interpretation: In rule 1 in table 2, women with high sports interest, low tv sports interest, medium museums interest and low yoga interest co-occurred with a no decision by men on women.

Scatter plot for 2307 rules
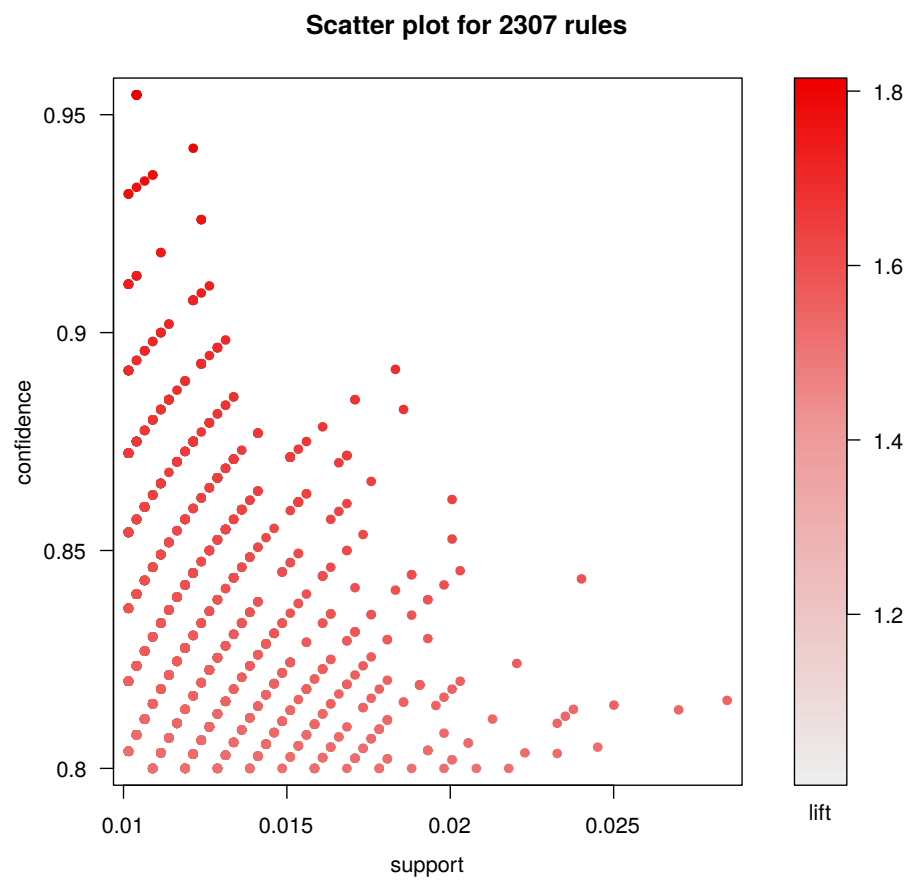
Figure 5: Scatter Plot: Rules when Men Say No

Grouped matrix for 2307 rules

size: support
color: lift
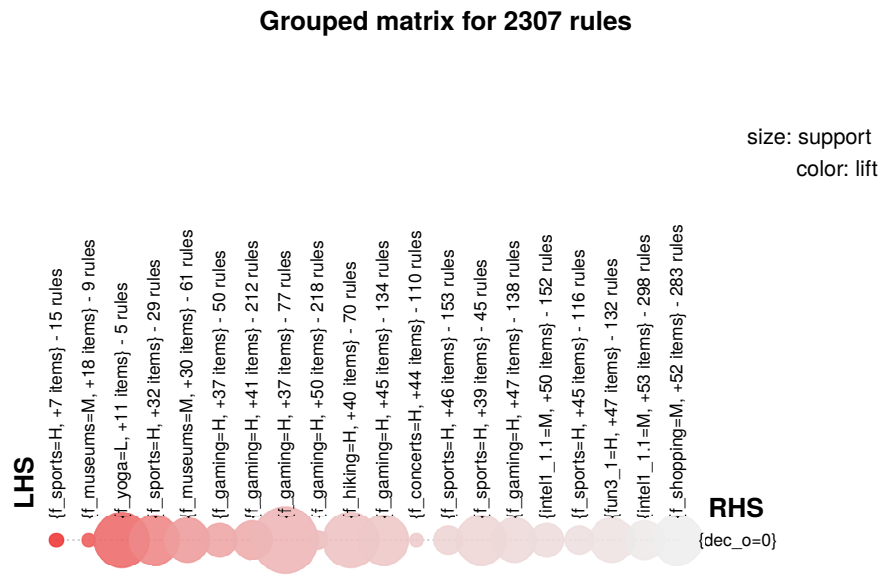
Figure 6: Grouped Matrix Plot: Rules when Men Say No

Graph for 20 rules

size: support (0.01 - 0.012)
color: lift (1.771 - 1.811)

intel1_1.1=M

f_gaming=L

fun3_1=H

f_art=M

f_concerts=H

f_sports=H
f_tvsports=L

f_yoga=L

dec_o=0

attr3_1=H

f_museums=M

f_movies=H

f_clubbing=H

attr1_1.1=M

shar1_1.1=M

f_gaming=H

f_hiking=H
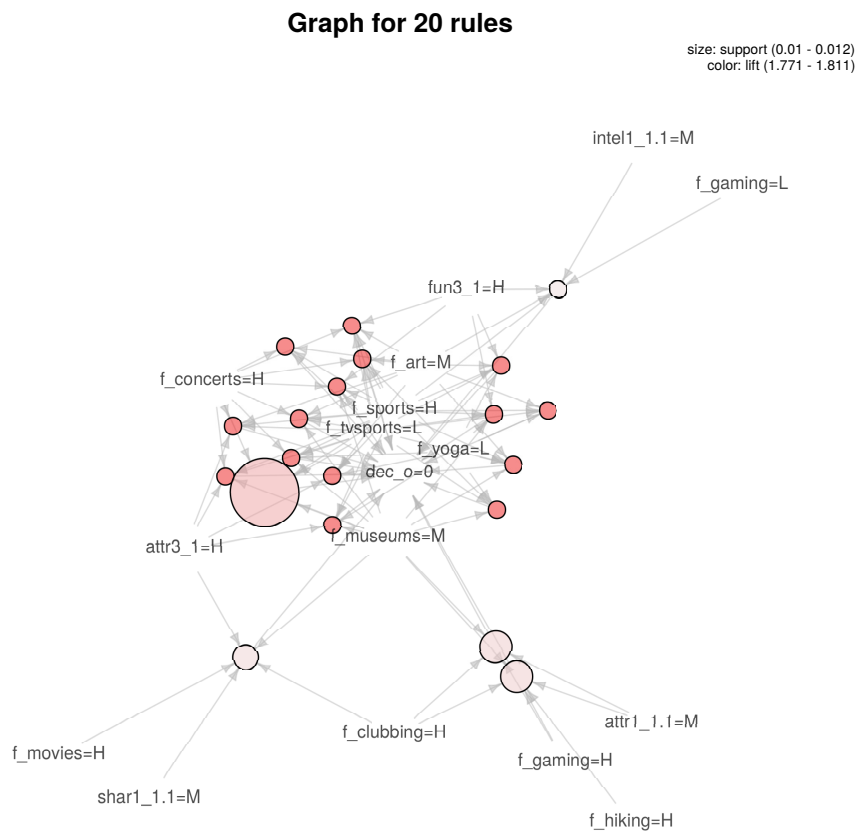
Figure 7: Graph Plot: Rules when Men Say No

Parallel coordinates plot for 20 rules

Figure 8: Parallel Plot: Rules when Men Say No

3 Women Decide

These association rules were obtained from the female segment of the data. Since we were interested in finding patterns in the decision process, we constrained the RHS of the rule to include only the decision attribute *dec_o* which takes values 0 and 1 corresponding to *no* and *yes* decisions respectively.

3.1 When Women Say Yes

The RHS was constrained to dec_o = 1

Table 3: Rules when Women Say Yes

S No	LHS	RHS	Support	Confidence	Lift
1	{need attractivesness:L, need sincerity:M, need intelligence:M, need fun:L, need ambition:M}	{dec_o:1}	0.005	0.960	2.630
2	{reading:M, movies:H, yoga:M, need_sincerity:M, need ambition:H}	{dec_o:1}	0.005	0.916	2.511
3	{sincerity:H, exercise:M, gaming:H, yoga:H, need ambition:M}	{dec_o:1}	0.005	0.913	2.501
4	{tvsports:L, dining:H, tv:M, movies:H, need ambition:H}	{dec_o:1}	0.005	0.913	2.501
5	{tvsports:H, clubbing:H, reading:M, need sincerity:M, need ambition:H}	{dec_o:1}	0.005	0.884	2.423

Interpretation: In rule 3 in table 3, men with high sincerity, medium exercise interest, high gaming interests, high yoga interest and women needing medium ambition co-occurred with a yes decision by women on men.
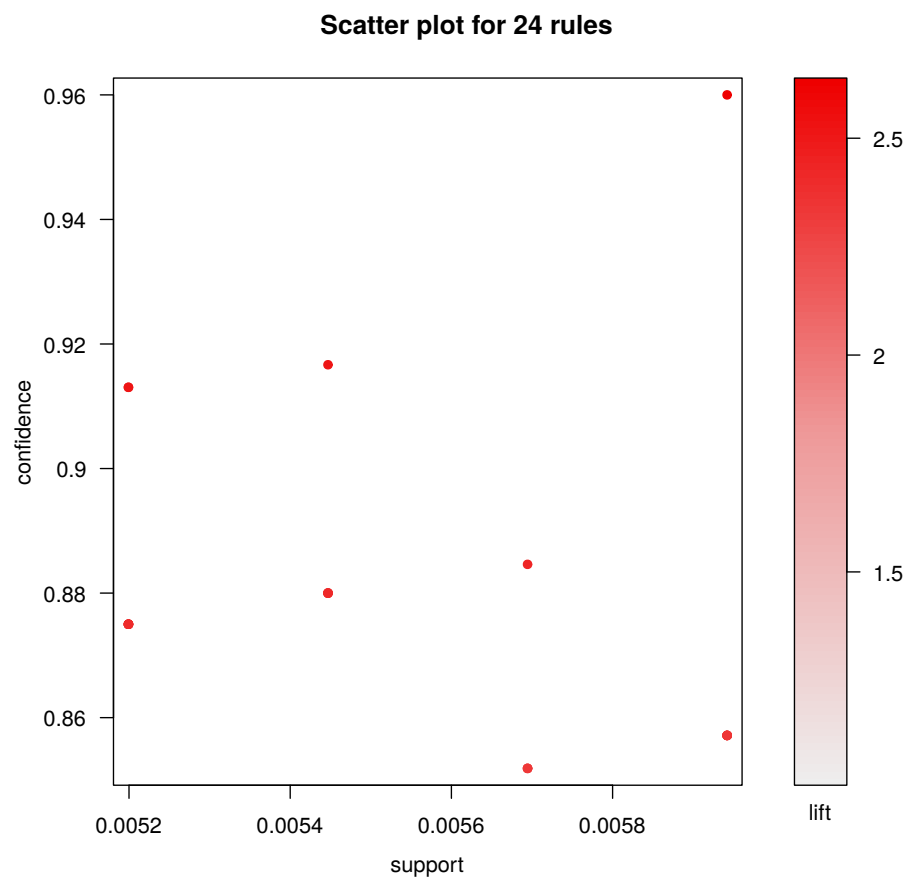
Figure 9: Scatter Plot: Rules when Women Say Yes

Grouped matrix for 24 rules
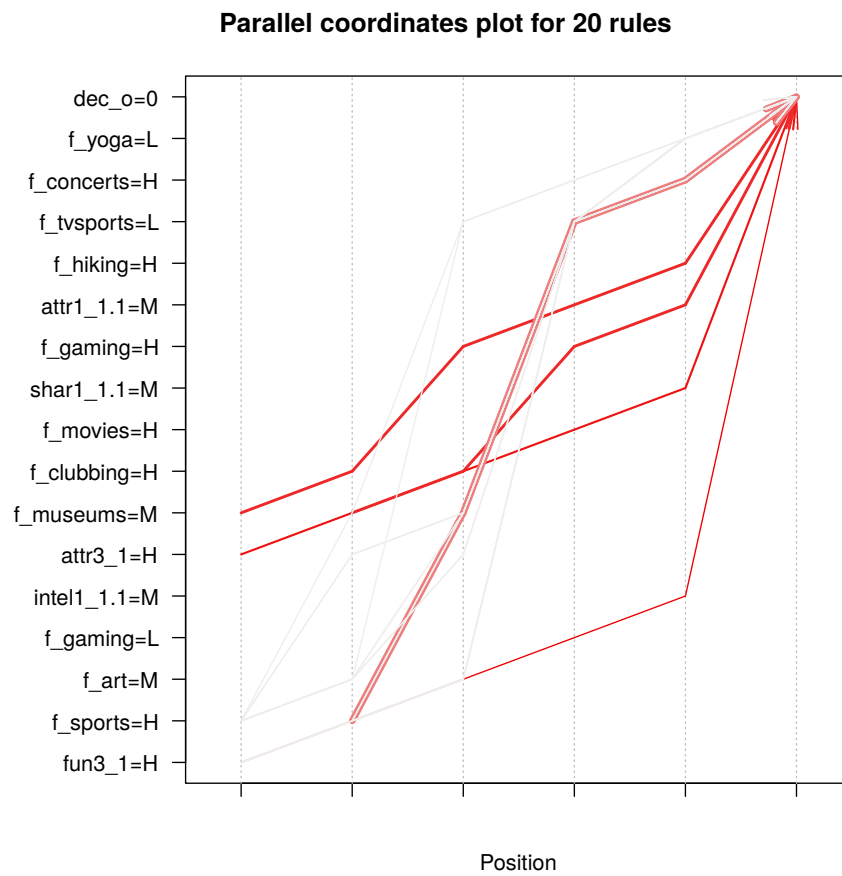
size: support
color: lift

{attr1_1.1=L, +4 items} - 1 rules
{m_reading=M, +4 items} - 1 rules
{sinc3_1=H, +9 items} - 2 rules
{m_tvsports=H, +4 items} - 1 rules
{m_yoga=M, +11 items} - 4 rules
{amb1_1.1=H, +14 items} - 6 rules
{m_hiking=H, +11 items} - 6 rules
{amb1_1.1=H, +8 items} - 3 rules

LHS

RHS

{dec_o=1}
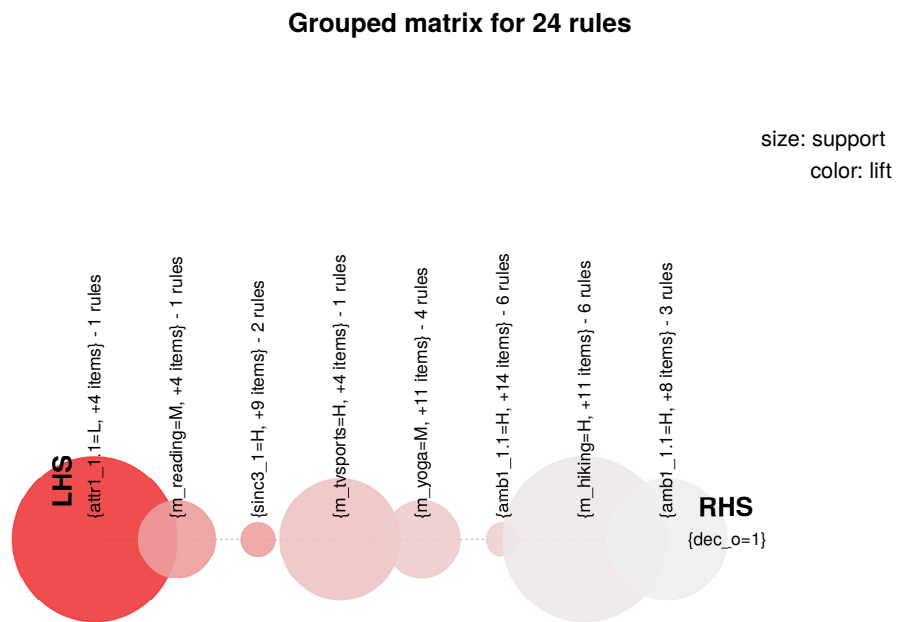
Figure 10: Grouped Matrix Plot: Rules when Women Say Yes

16

Graph for 20 rules

size: support (0.005 - 0.006)
color: lift (2.349 - 2.631)
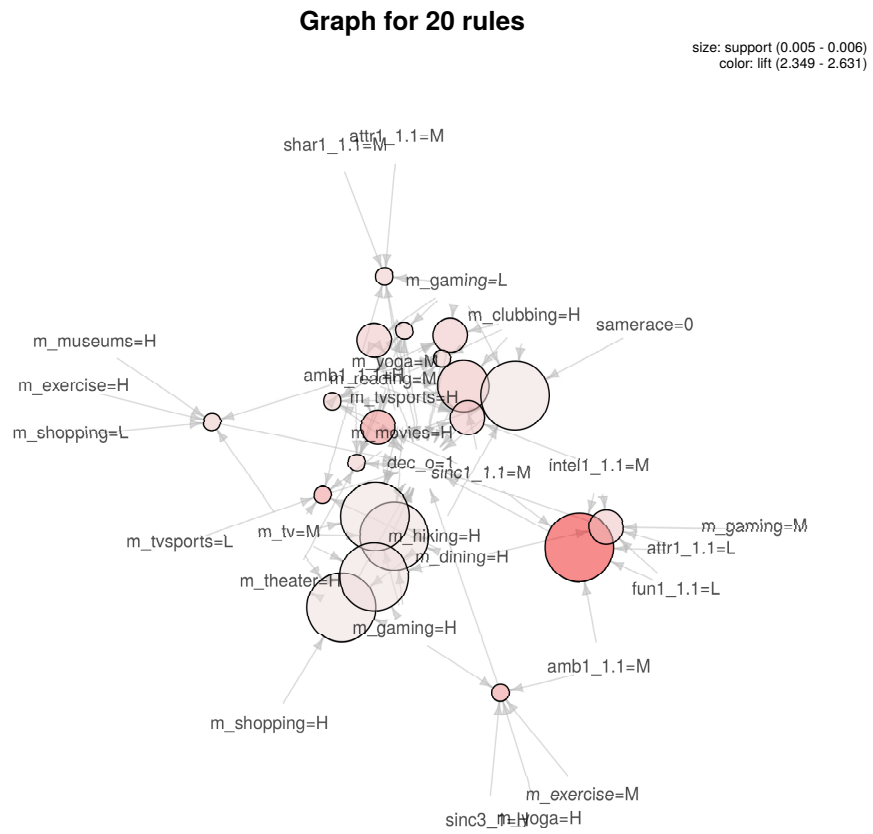
Figure 11: Graph Plot: Rules when Women Say Yes
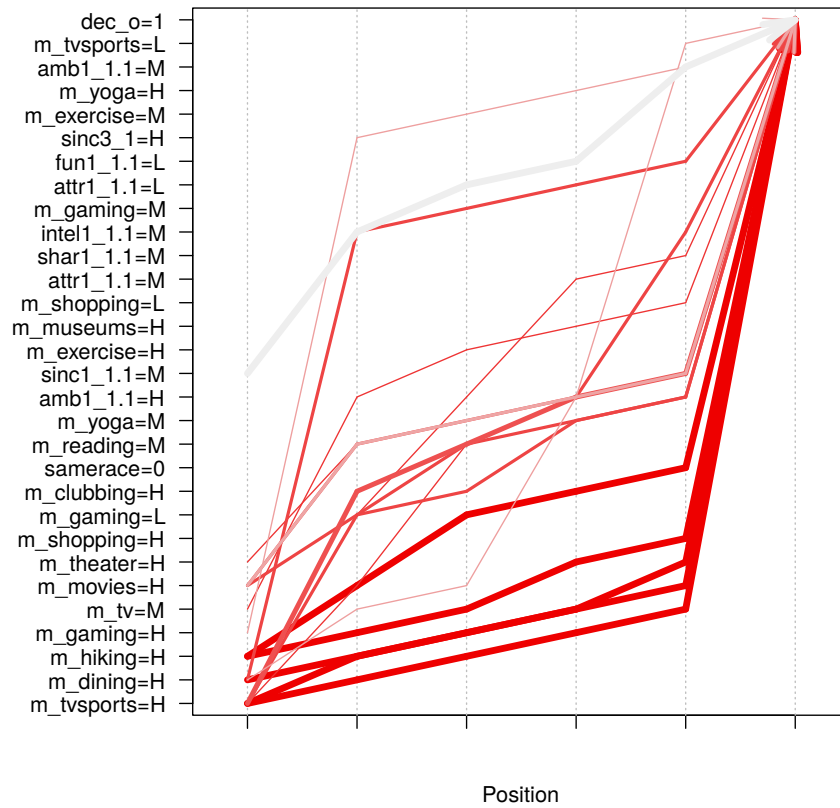
Figure 12: Parallel Plot: Rules when Women Say Yes

3.2 When Women Say No

The RHS was constrained to dec_o = 0

Table 4: Rules when Women Say No

S No	LHS	RHS	Support	Confidence	Lift
1	{sports:L, museums:L}	{dec_o:0}	0.005	1.000	1.574
2	{museums:L, yoga:H}	{dec_o:0}	0.005	1.000	1.574
3	{museums:L, gaming:H}	{dec_o:0}	0.005	1.000	1.574
4	{museums:L, shopping:H}	{dec_o:0}	0.005	1.000	1.574
5	{museums:L, theatre:H}	{dec_o:0}	0.005	1.000	1.574

Interpretation: In rule 1 in table 4, men with low sports interest and low museums interest co-occurred with a no decision by the women on men.

Scatter plot for 130717 rules
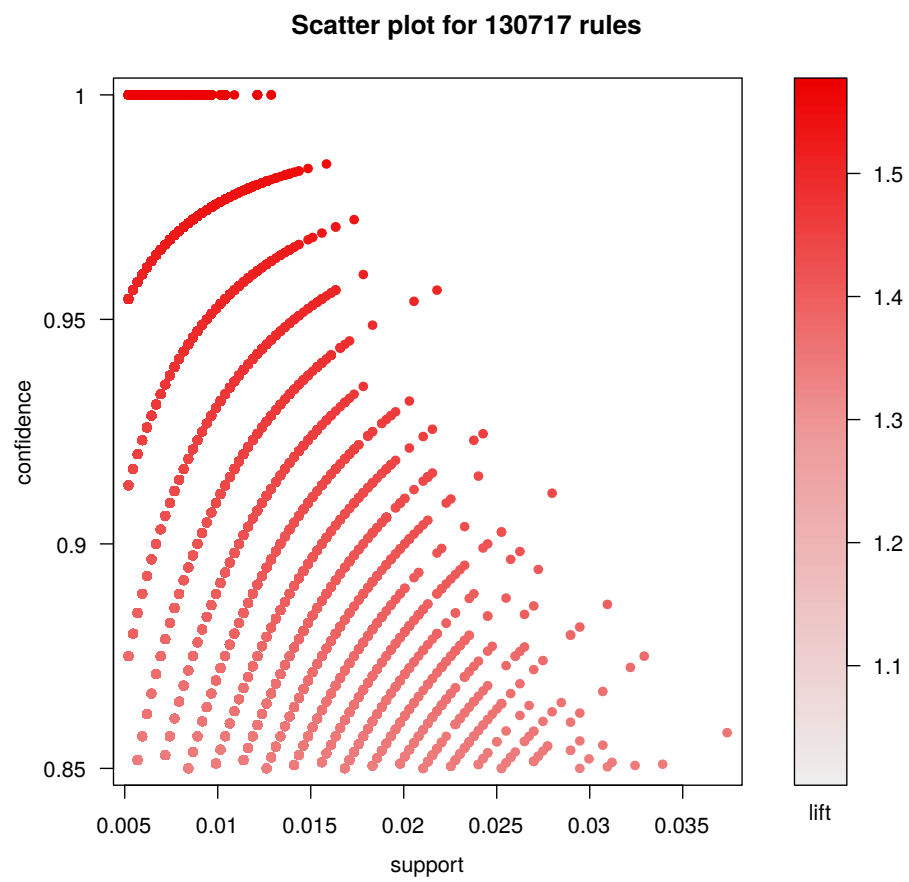
Figure 13: Scatter Plot: Rules when Women Say No

Grouped matrix for 130717 rules

size: support
color: lift

LHS

{m_yoga=H, +69 items} - 11611 rules
{m_sports=L, +58 items} - 2647 rules
{fun3_1=M, +67 items} - 5761 rules
{m_exercise=L, +73 items} - 5238 rules
{attr3_1=M, +66 items} - 1646 rules
{m_sports=M, +67 items} - 3007 rules
{m_sports=M, +71 items} - 5345 rules
{m_sports=M, +75 items} - 8355 rules
{m_sports=M, +75 items} - 8900 rules
{m_sports=M, +73 items} - 8968 rules
{m_sports=M, +74 items} - 9314 rules
{m_sports=M, +74 items} - 6944 rules
{m_sports=M, +73 items} - 8262 rules
{m_art=M, +75 items} - 10321 rules
{m_sports=M, +75 items} - 7416 rules
{m_sports=M, +71 items} - 2819 rules
{m_sports=M, +76 items} - 7630 rules
{intel1_1.1=M, +76 items} - 6121 rules
{m_sports=M, +75 items} - 7947 rules
{m_sports=M, +71 items} - 2465 rules

RHS

{dec_o=0}

Figure 14: Grouped Matrix Plot: Rules when Women Say No

Graph for 20 rules

size: support (0.005 - 0.005)
color: lift (1.575 - 1.575)

m_clubbing=M

m_tvsports=L

m_gaming=H

m_concerts=H

fun3_1=M

m_sports=L

dec_o=0
m_museums=L
m_art=L

m_exercise=M

m_tv=H

m_theater=H

m_yoga=H

m_shopping=H
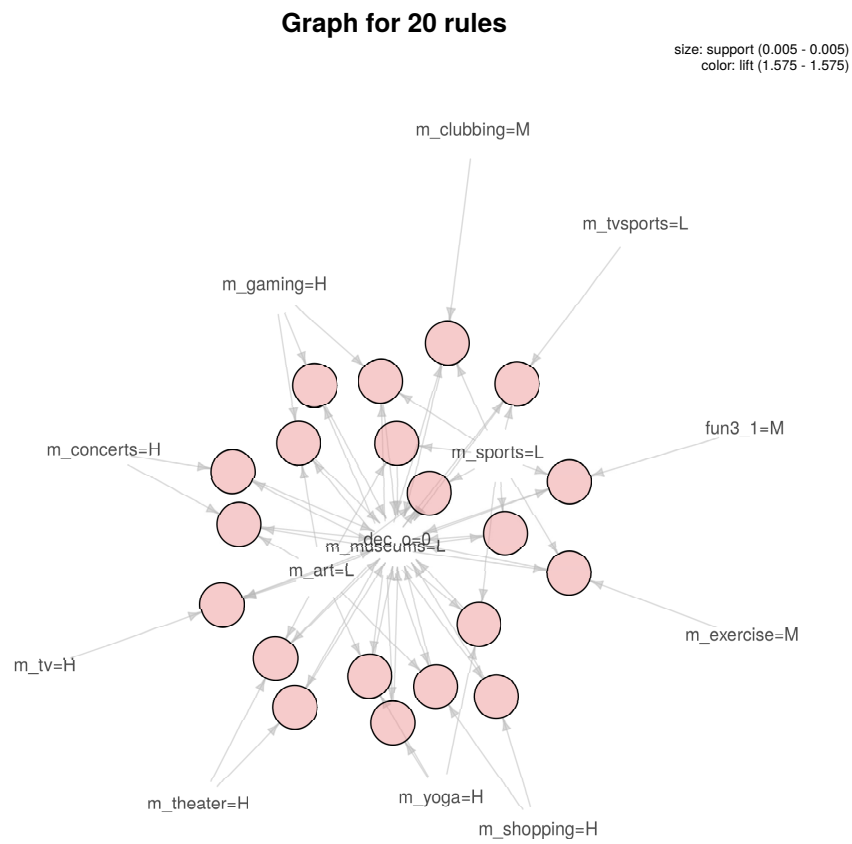
Figure 15: Graph Plot: Rules when Women Say No

22

Parallel coordinates plot for 20 rules

dec_o=0
m_exercise=M
m_clubbing=M
m_tvsports=L
fun3_1=M
m_tv=H
m_art=L
m_concerts=H
m_theater=H
m_shopping=H
m_gaming=H
m_yoga=H
m_museums=L
m_sports=L

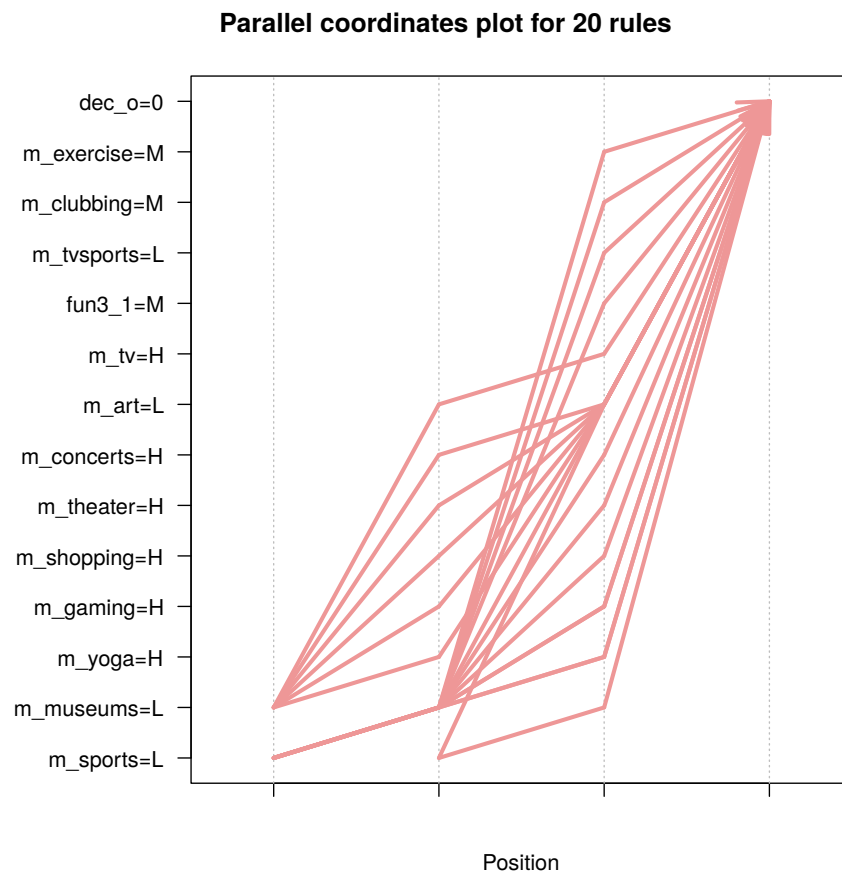Position

Figure 16: Parallel Plot: Rules when Women Say No

4 Other Interesting Association Rules

There were no constraints imposed on attribute inclusion in these rules.

4.1 Concerning Men

Table 5: Interesting Rules about Men

S No	LHS	RHS	Support	Confidence	Lift
1	{museums:M, hiking:M, theater:H}	{art:M}	0.059	1.000	2.983
2	{attractiveness:H, museums:M, yoga:L, dec_o:0}	{art:M}	0.055	0.991	2.956
3	{intelligence:H, art:M, clubbing:H, theater:H}	{hiking:M}	0.050	1.000	2.800
4	{museums:H, gaming:H, shopping:H}	{art:H}	0.051	1.000	2.216
5	{attractiveness:H, museums:H, gaming:H, shopping:H}	{art:H}	0.0517	1.000	2.216

Interpretation: In rule 1 in table 5, men with medium museums interest, medium hiking interest, high theatre interest co-occurred with medium art interest.
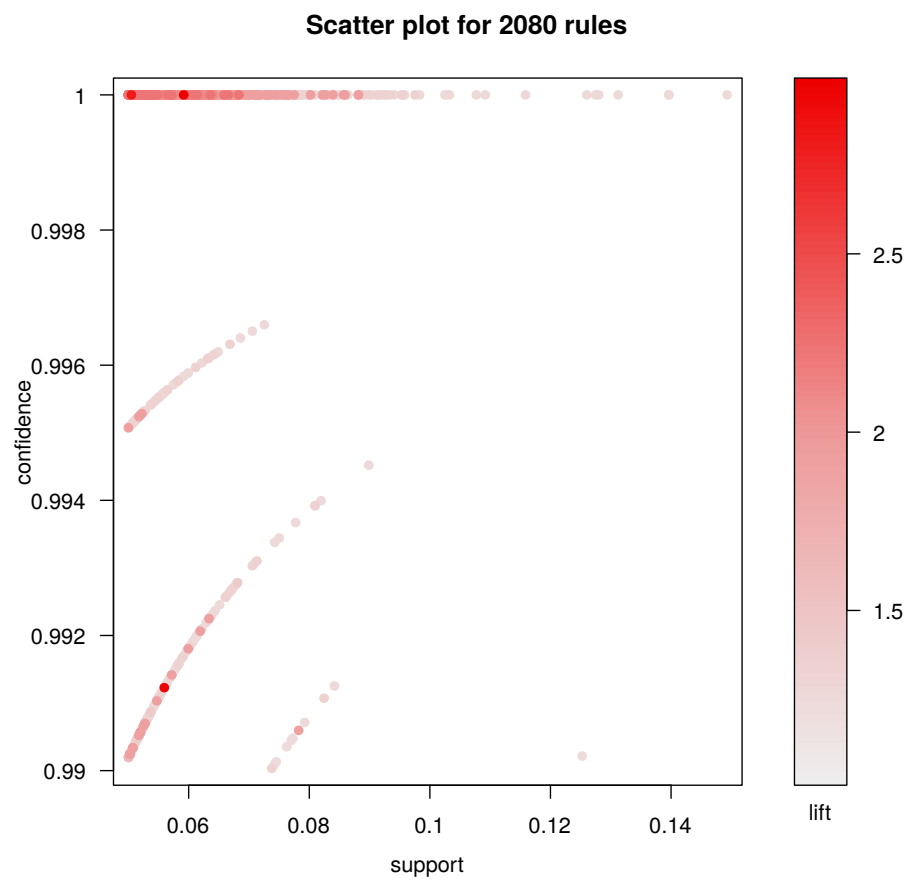
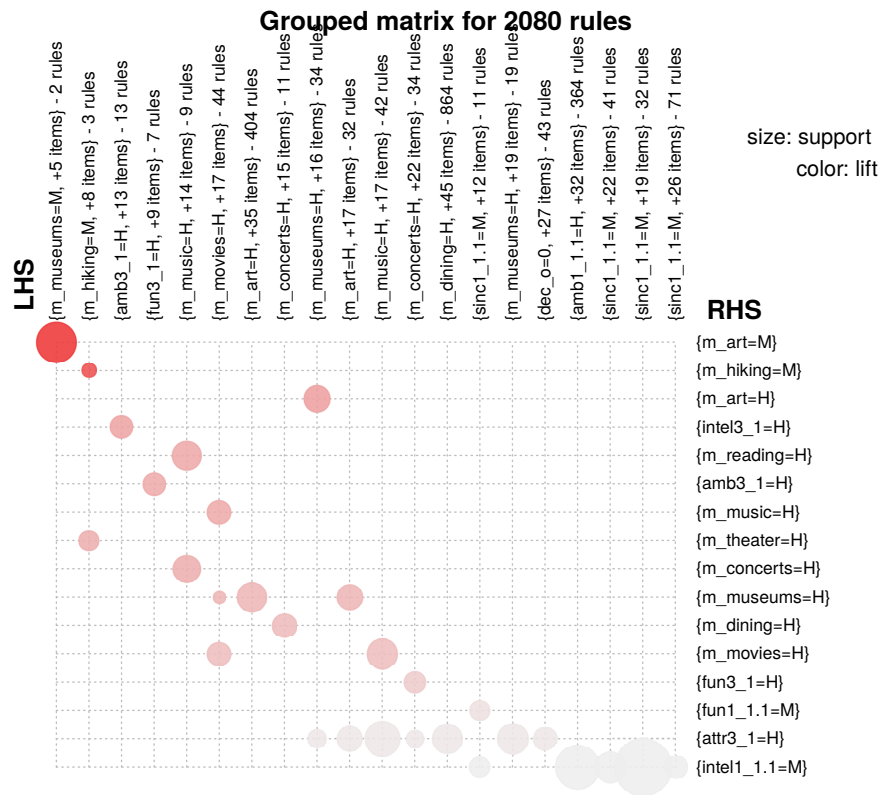Figure 17: Scatter Plot: Interesting Rules about Men

Grouped matrix for 2080 rules

LHS

size: support
color: lift

Columns (LHS):
- {m_museums=M, +5 items} - 2 rules
- {m_hiking=M, +8 items} - 3 rules
- {amb3_1=H, +13 items} - 13 rules
- {fun3_1=H, +9 items} - 7 rules
- {m_music=H, +14 items} - 9 rules
- {m_movies=H, +17 items} - 44 rules
- {m_art=H, +35 items} - 404 rules
- {m_concerts=H, +15 items} - 11 rules
- {m_museums=H, +16 items} - 34 rules
- {m_art=H, +17 items} - 32 rules
- {m_music=H, +17 items} - 42 rules
- {m_concerts=H, +22 items} - 34 rules
- {m_dining=H, +45 items} - 864 rules
- {sinc1_1.1=M, +12 items} - 11 rules
- {m_museums=H, +19 items} - 19 rules
- {dec_o=0, +27 items} - 43 rules
- {amb1_1.1=H, +32 items} - 364 rules
- {sinc1_1.1=M, +22 items} - 41 rules
- {sinc1_1.1=M, +19 items} - 32 rules
- {sinc1_1.1=M, +26 items} - 71 rules

RHS
- {m_art=M}
- {m_hiking=M}
- {m_art=H}
- {intel3_1=H}
- {m_reading=H}
- {amb3_1=H}
- {m_music=H}
- {m_theater=H}
- {m_concerts=H}
- {m_museums=H}
- {m_dining=H}
- {m_movies=H}
- {fun3_1=H}
- {fun1_1.1=M}
- {attr3_1=H}
- {intel1_1.1=M}

Figure 18: Grouped Matrix Plot: Interesting Rules about Men

26

Graph for 20 rules

size: support (0.05 - 0.068)
color: lift (2.217 - 2.983)
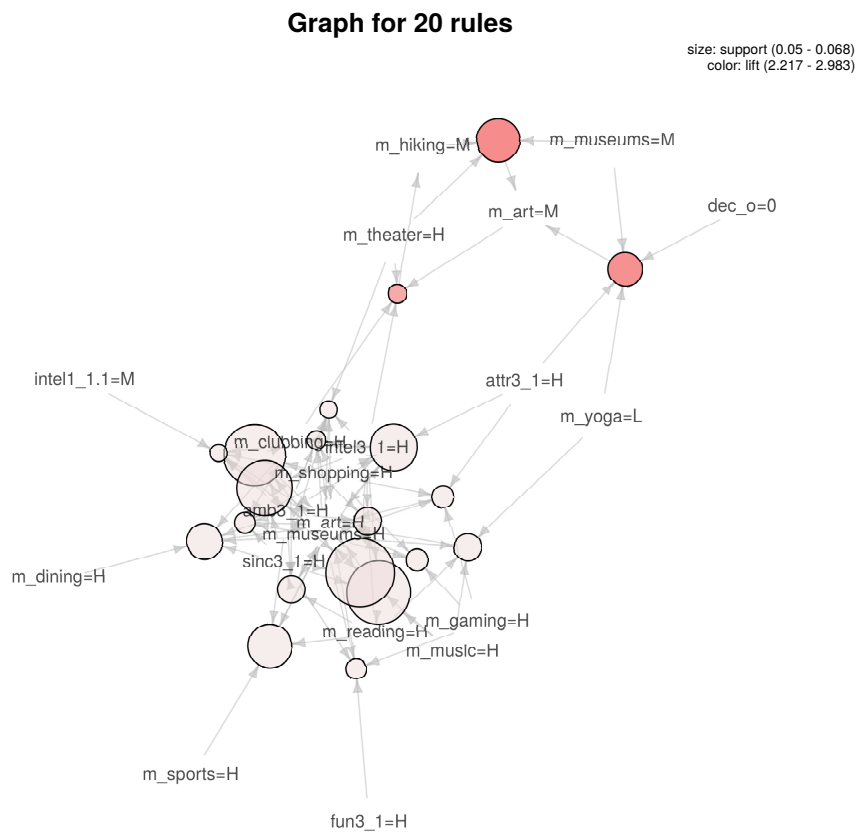
Figure 19: Graph Plot: Interesting Rules about Men

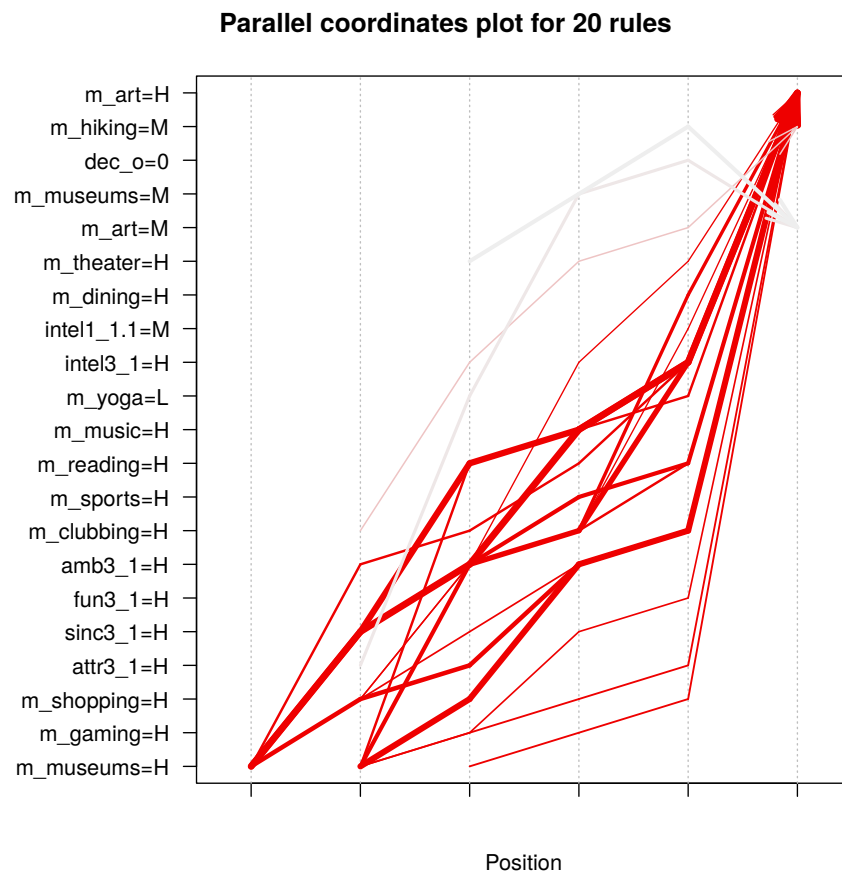**Parallel coordinates plot for 20 rules**

Figure 20: Parallel Plot: Interesting Rules about Men

4.2   Concerning Women

Table 6: Interesting Rules about Women

S No	LHS	RHS	Support	Confidence	Lift
1	{attractiveness:H, exercise:H, museums:M}	{art:M}	0.055	1.000	5.563
2	{attractiveness:H, ambition:H, art:H, clubbing:H, concerts:H, shopping:H}	{sincerity:H}	0.053	1.000	2.500
3	{attractiveness:H, ambition:H, museums:H, art:H, hiking:H, concerts:H}	{sincerity:H}	0.057	1.000	2.500
4	{exercise:H, movies:H, shopping:H, yoga:H}	{reading:H}	0.051	1.000	
5	{exercise:H, art:H, theatre:H, movies:H, shopping:H}	{reading:H}	0.061	1.000	2.458

Interpretation: In rule 1 in table 6, women who rate themselves highly attractive, with high exercise interest and medium museums interest co-occurred with medium art interest.
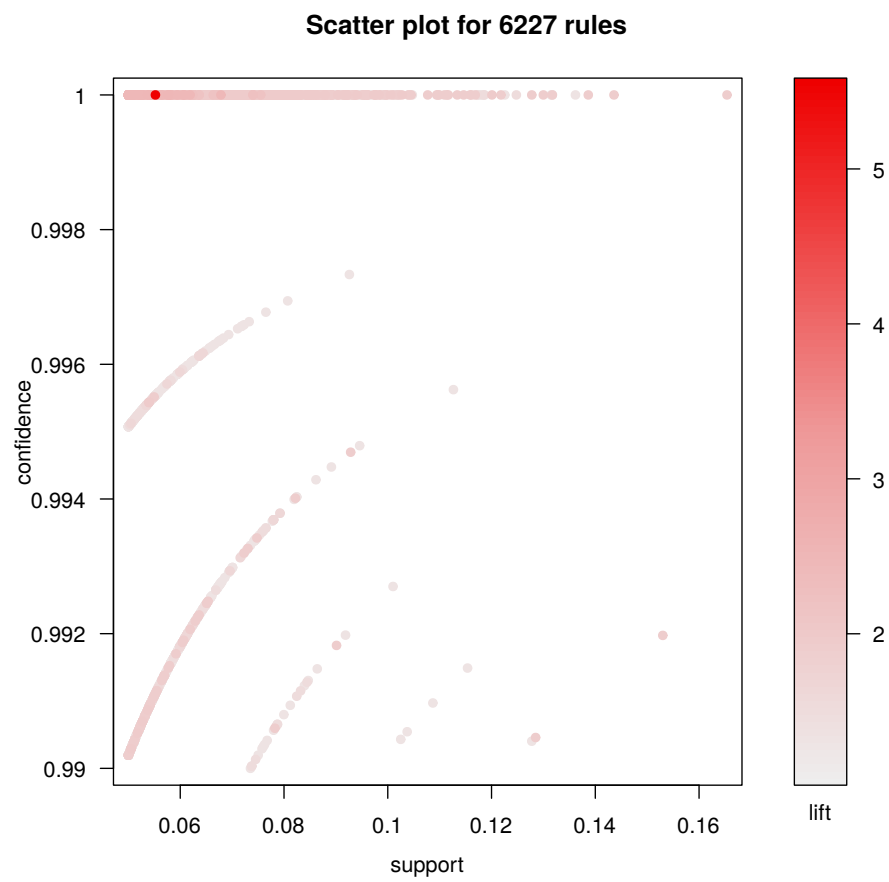
Scatter plot for 6227 rules

Figure 21: Scatter Plot: Interesting Rules about Women

Grouped matrix for 6227 rules

size: support
color: lift

LHS

RHS
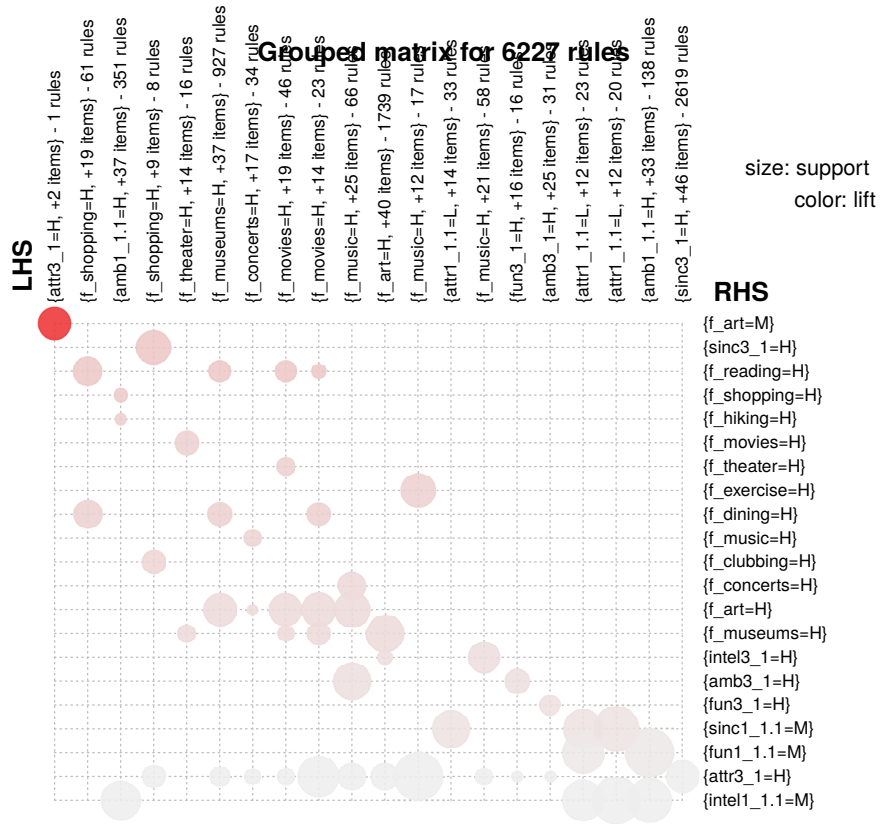
Figure 22: Grouped Matrix Plot: Interesting Rules about Women

Graph for 20 rules

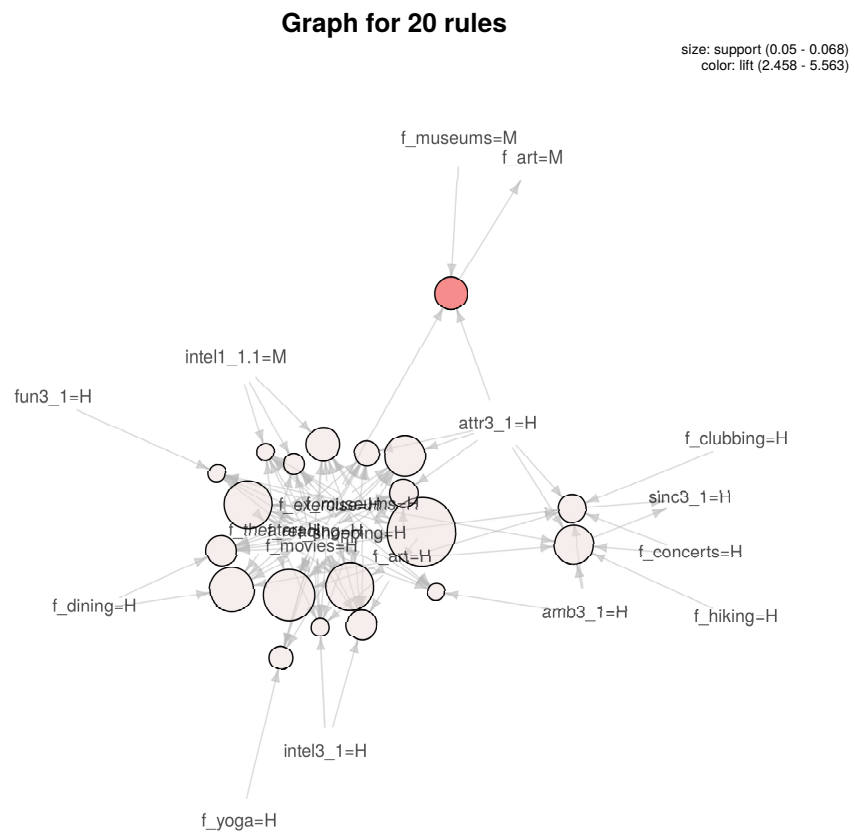size: support (0.05 - 0.068)
color: lift (2.458 - 5.563)

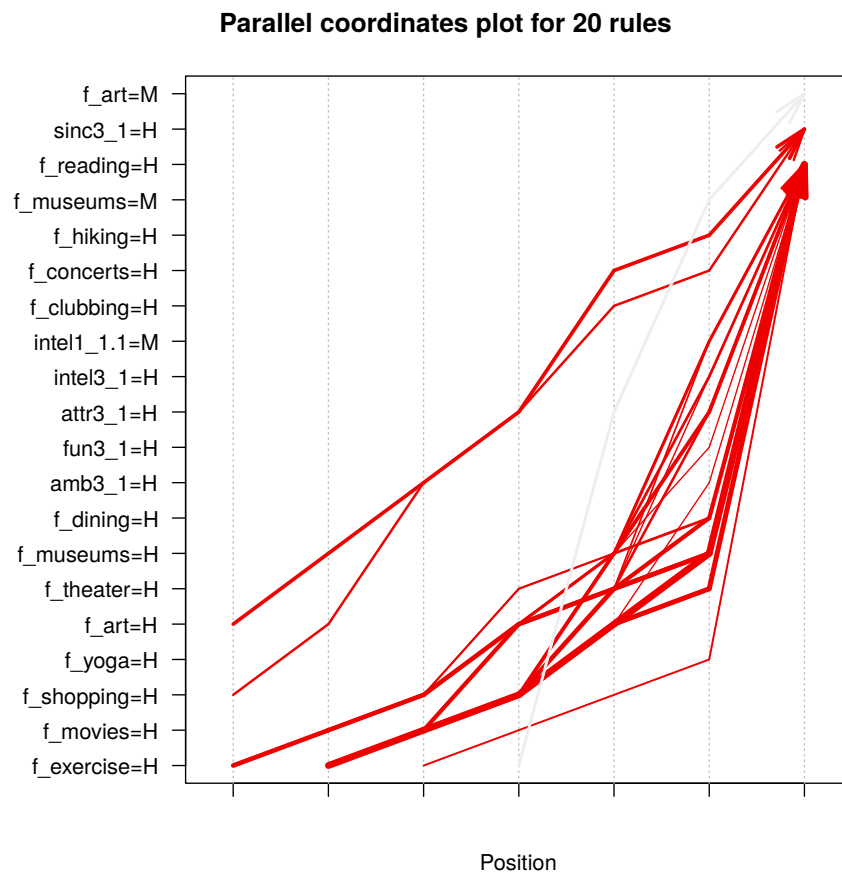
Figure 23: Graph Plot: Interesting Rules about Women

Figure 24: Parallel Plot: Interesting Rules about Women

5   Conclusion

In the association rules obtained we found that men were concerned with the interest in hobbies of their partner (with sports and hiking being popular) while women were concerned of the qualities of their partner (with sincerity and ambition being important).