MTA SZTAKI datamining search

SzTAKI@ImageCLEF 2009

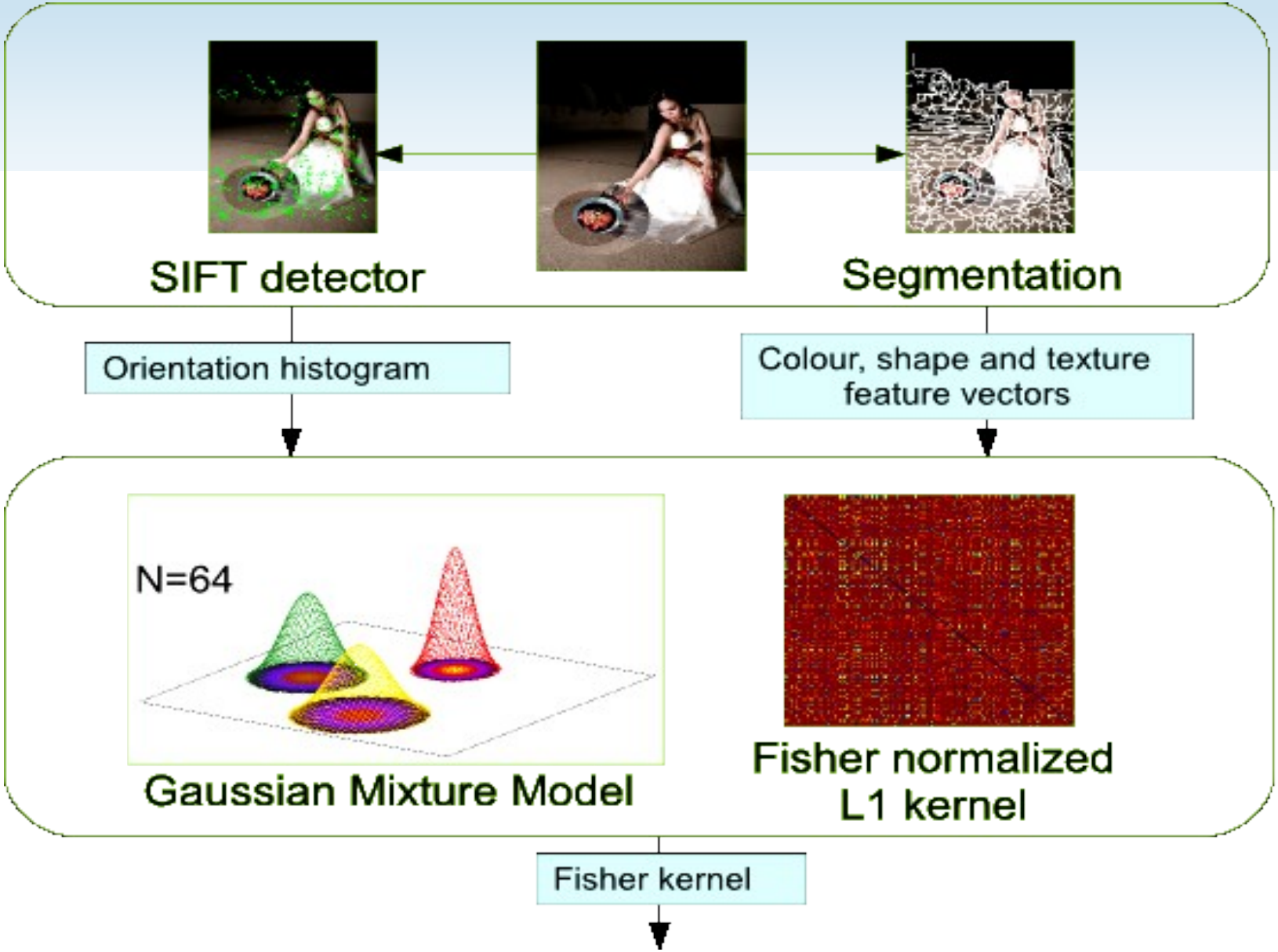
Using Combination and Segmentation for automatic concept detection

István Petrás, Bálint Daróczy, Zsolt Fekete

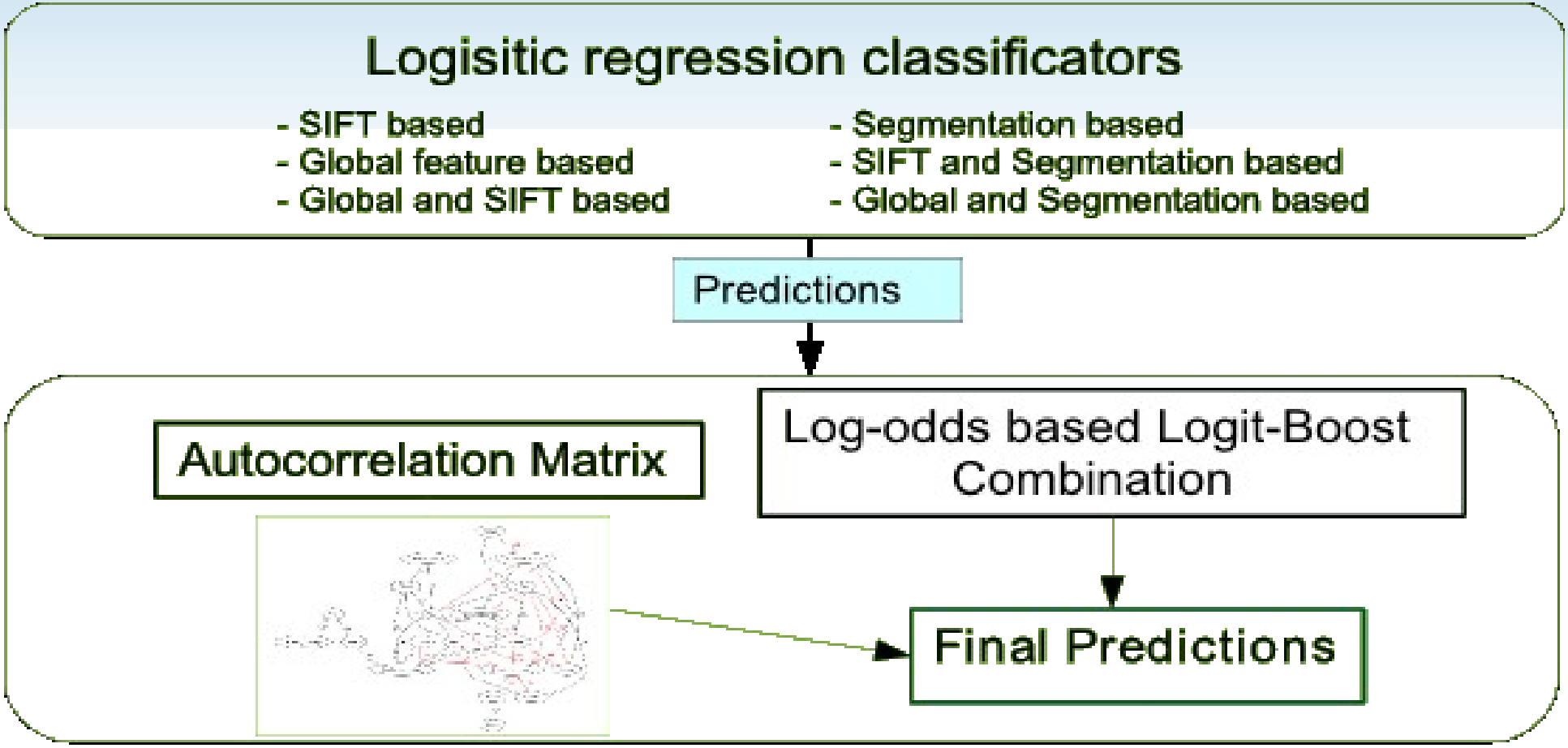
Data Mining and Web Search Group

petras@sztaki.hu
http://websearch.sztaki.hu

Flowchart of the processing



Flowchart of the processing



Final results

	EER	AUC
SIFT + Segmentation + Cross	0.291718	0.773133
SIFT + Segmentation	0.296315	0.771324
Log Odds combination	0.304113	0.746300
SIFT	0.322632	0.733264
Segmentation	0.346106	0.707860

Final results

Nr.	Concept	AUC	AUC comb.	LogOdds c.	Nr.	Concept	AUC	AUC comb.	LogOdds c.
0	Partylife	0.73957	0.72941	0.68350	27	Day	0.74645	**0.76309**	**0.76555**
1	Familiy_Friends	0.73815	0.73811	0.71786	28	Night	0.86187	0.86187	**0.87458**
2	Beach_Holidays	0.81424	**0.85588**	0.79171	29	No_Visual_Time	0.74712	**0.75651**	0.74943
3	Building_Sights	0.80933	0.80933	0.79170	30	Sunny	0.72356	**0.73187**	0.72782
4	Snow	0.80294	0.75331	0.79255	31	Sunset_Sunrise	0.92588	0.92588	0.92026
5	Citylife	0.78624	0.78624	0.75672	32	Canvas	0.73779	0.73779	0.73235
6	Landscape_Nature	0.89293	0.89293	0.88607	33	Still_Life	0.74807	**0.75257**	0.72224
7	Sports	0.59259	**0.59820**	0.51282	34	Macro	0.70740	0.70740	0.69002
8	Desert	0.83242	0.83242	0.74518	35	Portrait	0.77825	0.77825	0.73150
9	Spring	0.75002	0.75002	0.68572	36	Overexposed	0.76191	0.76191	0.72049
10	Summer	0.77211	**0.77541**	0.73943	37	Underexposed	0.86044	0.86044	**0.88017**
11	Autumn	0.77894	0.77894	0.73783	38	Neutral_Illumination	0.79114	0.79114	**0.79946**
12	Winter	0.79843	0.79843	**0.80648**	39	Motion_Blur	0.66064	0.66064	0.62978
13	No_Visual_Season	0.76659	0.75141	0.69890	40	Out_of_focus	0.79453	0.74420	0.72747
14	Indoor	0.72815	0.69309	0.71512	41	Partly_Blurred	0.82296	0.82296	0.78099
15	Outdoor	0.81166	0.80824	0.80551	42	No_Blur	0.79566	0.79566	0.73280
16	No_Visual_Place	0.72283	0.72283	0.64580	43	Single_Person	0.71757	0.71757	0.62987
17	Plants	0.78703	**0.79857**	0.78064	44	Small_Group	0.66305	0.66305	0.66832
18	Flowers	0.79807	**0.81733**	**0.80806**	45	Big_Group	0.78595	0.78595	0.72330
19	Trees	0.82024	**0.83775**	**0.85227**	46	No_Persons	0.72840	0.72840	0.72431
20	Sky	0.87776	0.87776	0.85901	47	Animals	0.74320	0.74320	0.70871
21	Clouds	0.88680	**0.89788**	0.88929	48	Food	0.83345	**0.84414**	0.83605
22	Water	0.80937	0.80937	0.79304	49	Vehicle	0.72292	**0.73282**	0.70768
23	Lake	0.79240	**0.85950**	0.73959	50	Aesthetic_Impression	0.63817	0.63817	0.62796
24	River	0.74148	0.74148	0.70190	51	Overall_Quality	0.63330	0.63330	0.55105
25	Sea	0.87256	**0.89580**	0.87399	52	Fancy	0.56859	0.56859	0.52613
26	Mountains	0.85911	0.85911	0.85493					
	mean AUC	**0.7713**	**0.7731**	**0.7463**		**Mean increase:**		**0.0018**	**-0.0250**

Feature extraction

Graph based Segmentation

- Undirected weighted graph over neighboring pixels
- Bottom-up clustering with dynamic thresholds
- Sobel gradient image to select important edges

Local features

- SIFT descriptors

Feature extraction: vectors

From graph based segmentation

Average RGB HSV	RGB, HSV Histogram	Contrast

Zig-Zag Fourier

From local feature points

128 SIFT descriptor

The modeling: Fisher vectors

Bag of visual words. Images are represented by gradient vector of a generative probability model (GMM) [Perronnin, Dance.]

Fitting a GMM with parameters $\Phi = \{w_i, \mu_i, \Sigma_i,\ i = 1 \ldots N\}$

Gradient of the log-likelihood describes the direction in which parameters should be modified to best fit the data:

$$\nabla_\Phi \log p(I|\Phi)$$

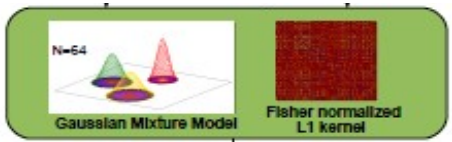
Fisher vector (normalized):

$$\mathbf{f}_I = F_\Phi^{-1/2} \nabla_\Phi \log p(I|\Phi)$$

$$F_\Phi = E_X P\left[\nabla_\Phi \log p(I|\Phi) \nabla_\Phi \log p(I|\Phi)^T\right]$$

Input to the classifier:

$$K(\mathbf{f}_I, \mathbf{f}_J) = \sum_{t=1}^{T} ||\tilde{\mathbf{f}}_I - \tilde{\mathbf{f}}_t||_1 \cdot ||\tilde{\mathbf{f}}_J - \tilde{\mathbf{f}}_t||_1$$



Classification

Logistic regression based classifier

- Computed for each concept using the Liblinear package

Combining the results of the different concepts

- *Method 1*: log odds-logitBoost
- *Method 2*: exploiting information embedded in training annotation labels

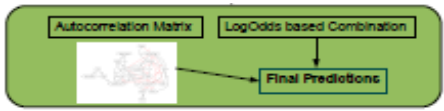
Combining 1

Log odds based combining

Log odds features: computes new features from the outputs of the different classificators. Creates a new re-reanking of results.

Boosting based learning on the re-ranked results

	EER	AUC
SIFT + Segmentation + Cross	0.291718	0.773133
SIFT + Segmentation	0.296315	0.771324
Log Odds combination	0.304113	0.746300
SIFT	0.322632	0.733264
Segmentation	0.346106	0.707860

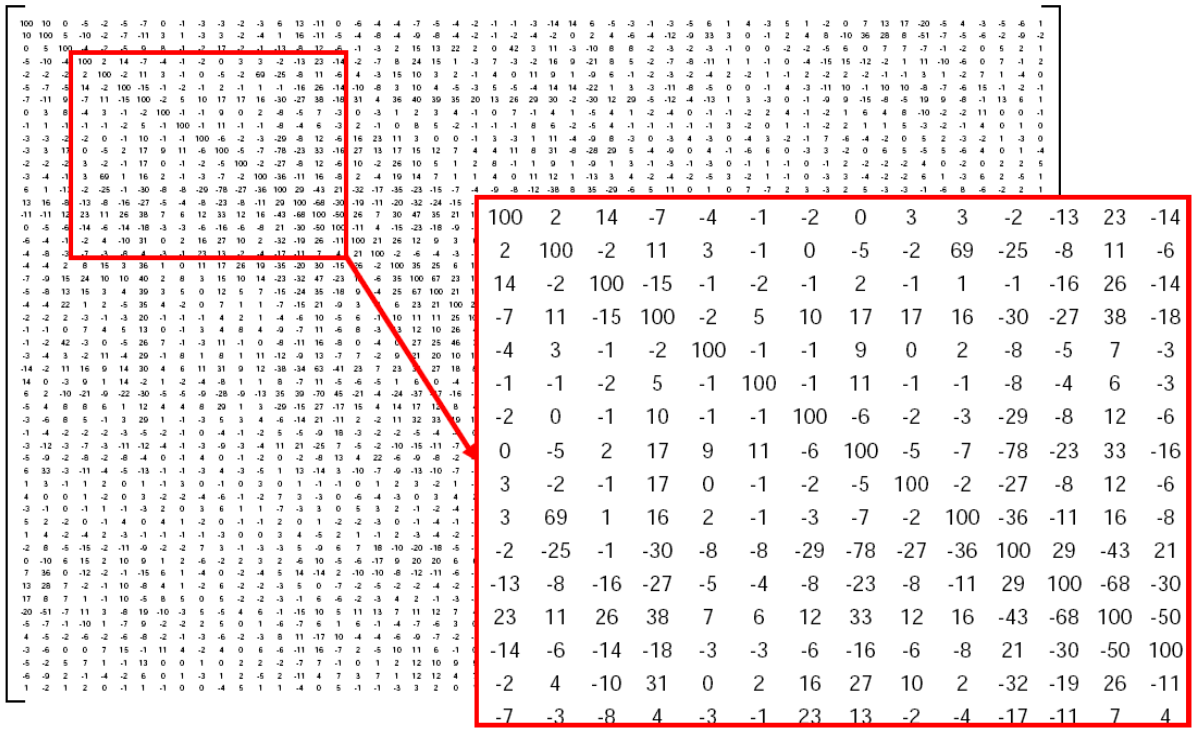


Combining 2

Correlation matrix from training annotations

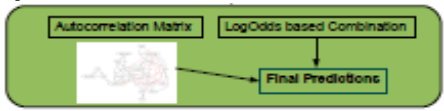
0 Partylife
1 Familiy_Friends
2 Beach_Holidays
3 Building_Sights
4 Snow
5 Citylife
6 Landscape_Nature
7 Sports
8 Desert
9 Spring
10 Summer
11 Autumn
12 Winter
13 No_Visual_Season
14 Indoor
15 Outdoor
16 No_Visual_Place
17 Plants
18 Flowers
19 Trees
20 Sky
21 Clouds
22 Water
23 Lake
24 River
25 Sea
26 Mountains
27 Day
28 Night
29 No_Visual_Time
30 Sunny
31 Sunset_Sunrise
32 Canvas
33 Still_Life
34 Macro
35 Portrait
36 Overexposed
37 Underexposed
38 Neutral_Illumination
39 Motion_Blur
40 Out_of_focus

C =



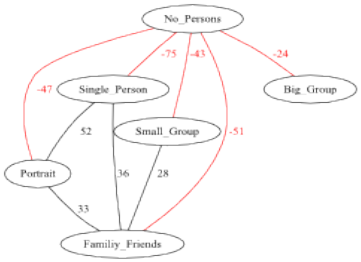
c_{ij} = corr (a_i; a_j), a_i is the ith column of A,

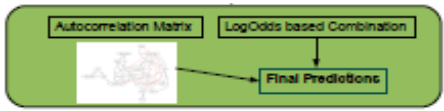
corr (x; y) = cov (x; y) = (std (x) std (y))



Combining 2

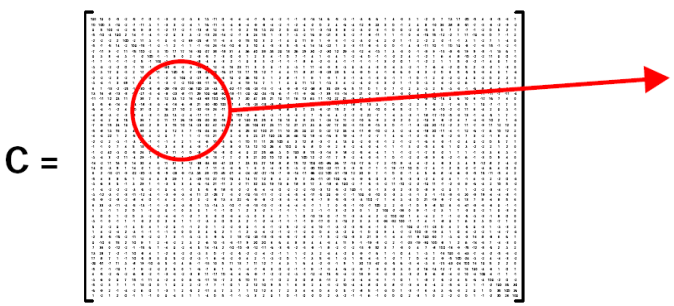
Correlation matrix from training annotations represented as a graph

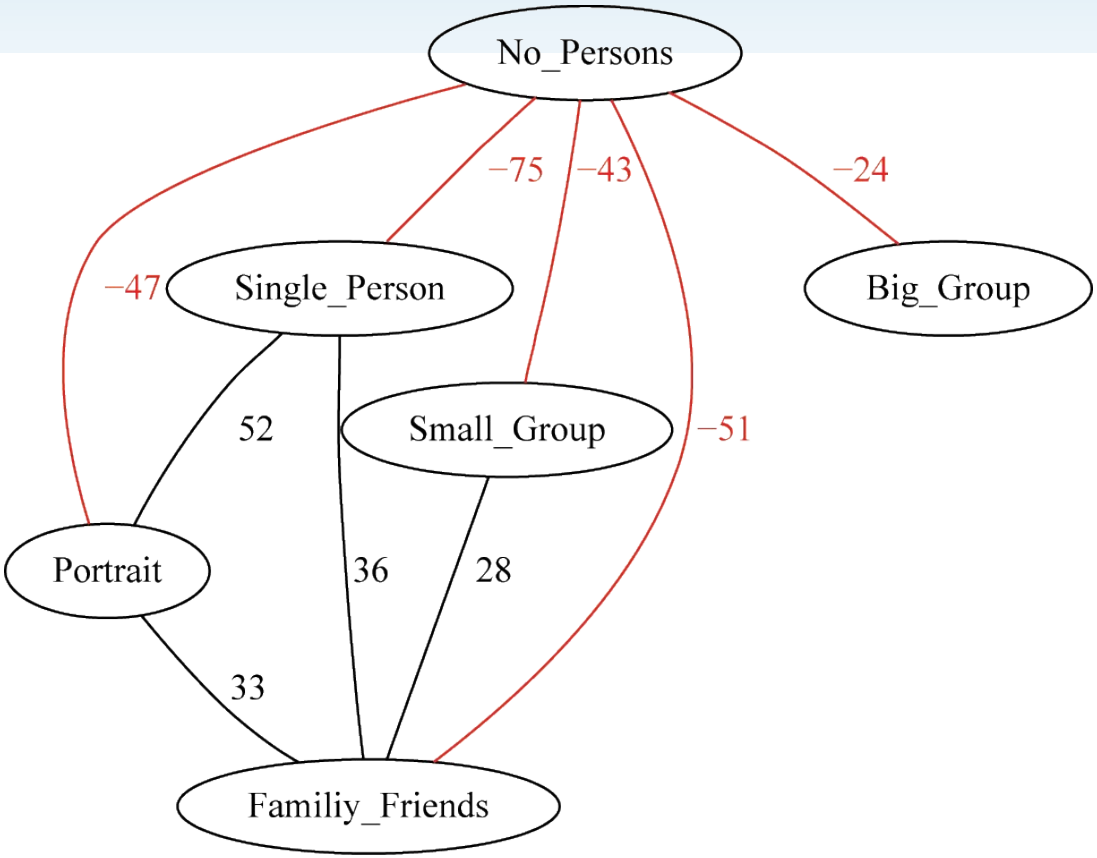


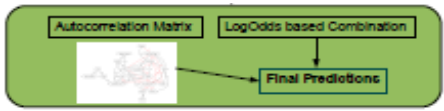


Combining 2

Correlation matrix subgraph

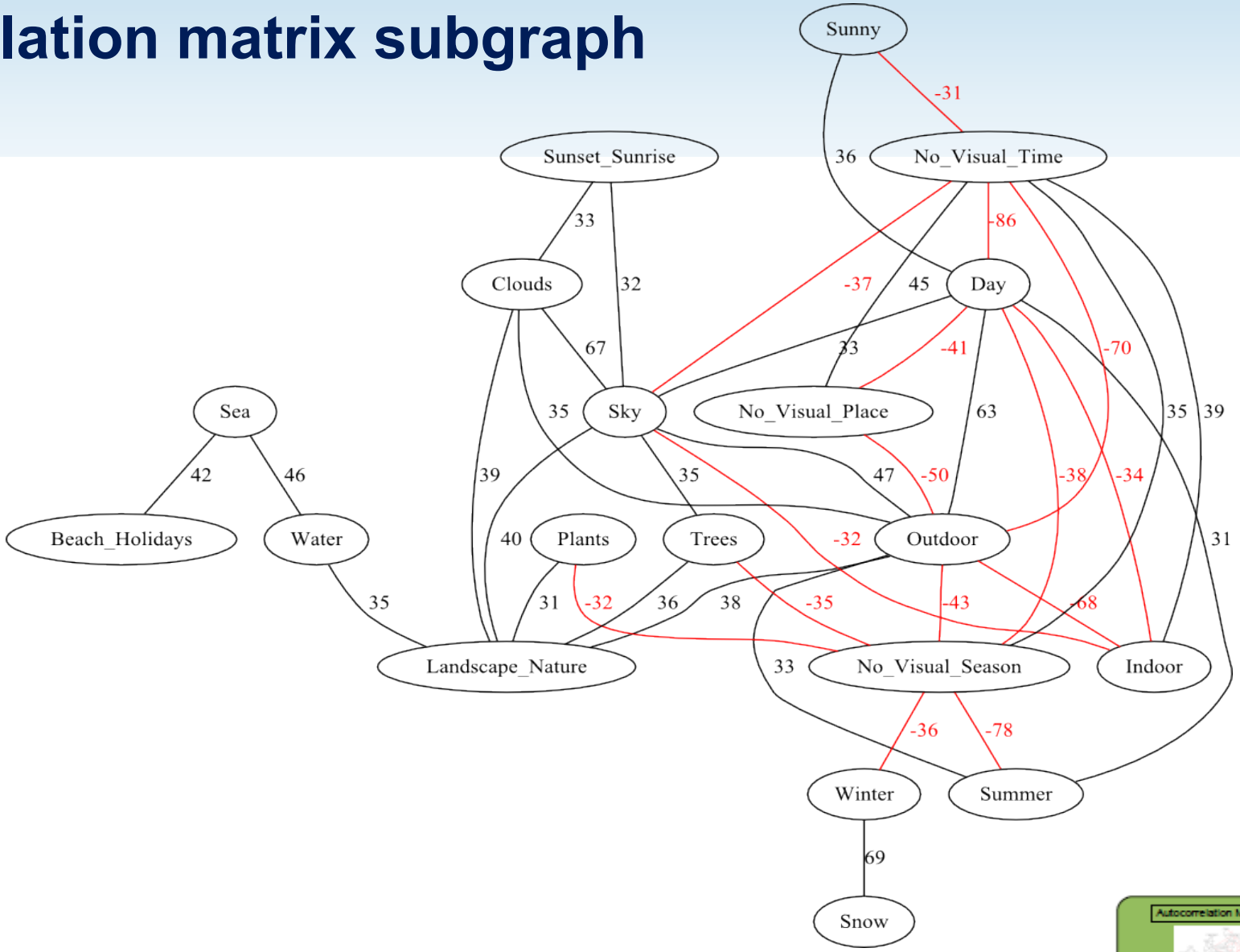


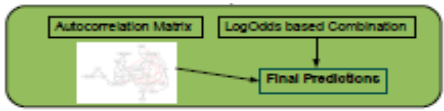




Combining 2

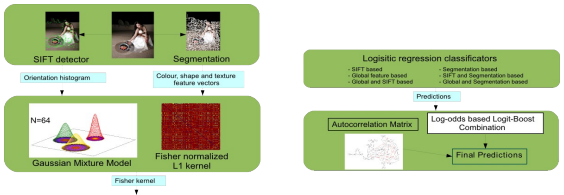
Correlation matrix subgraph



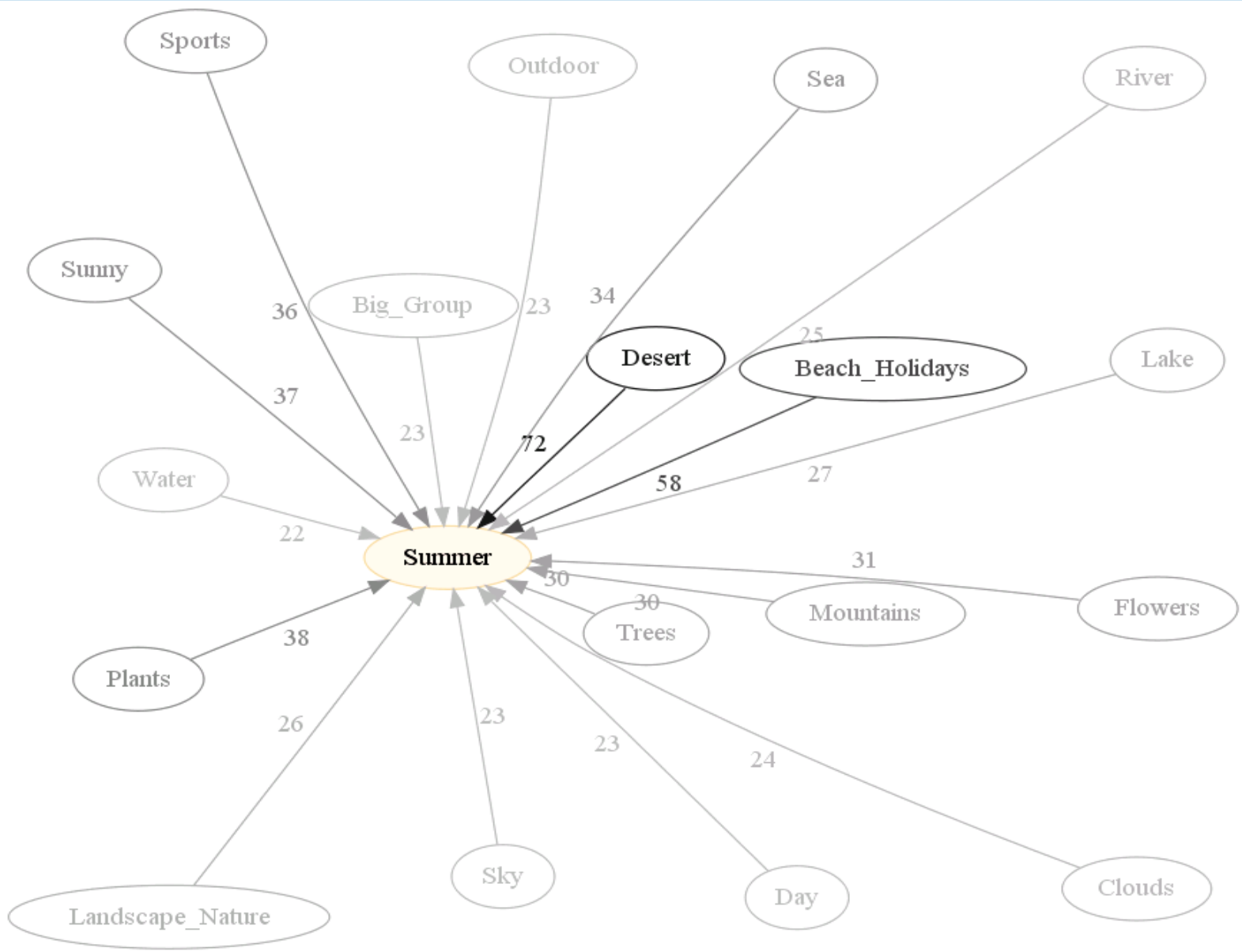


Final results

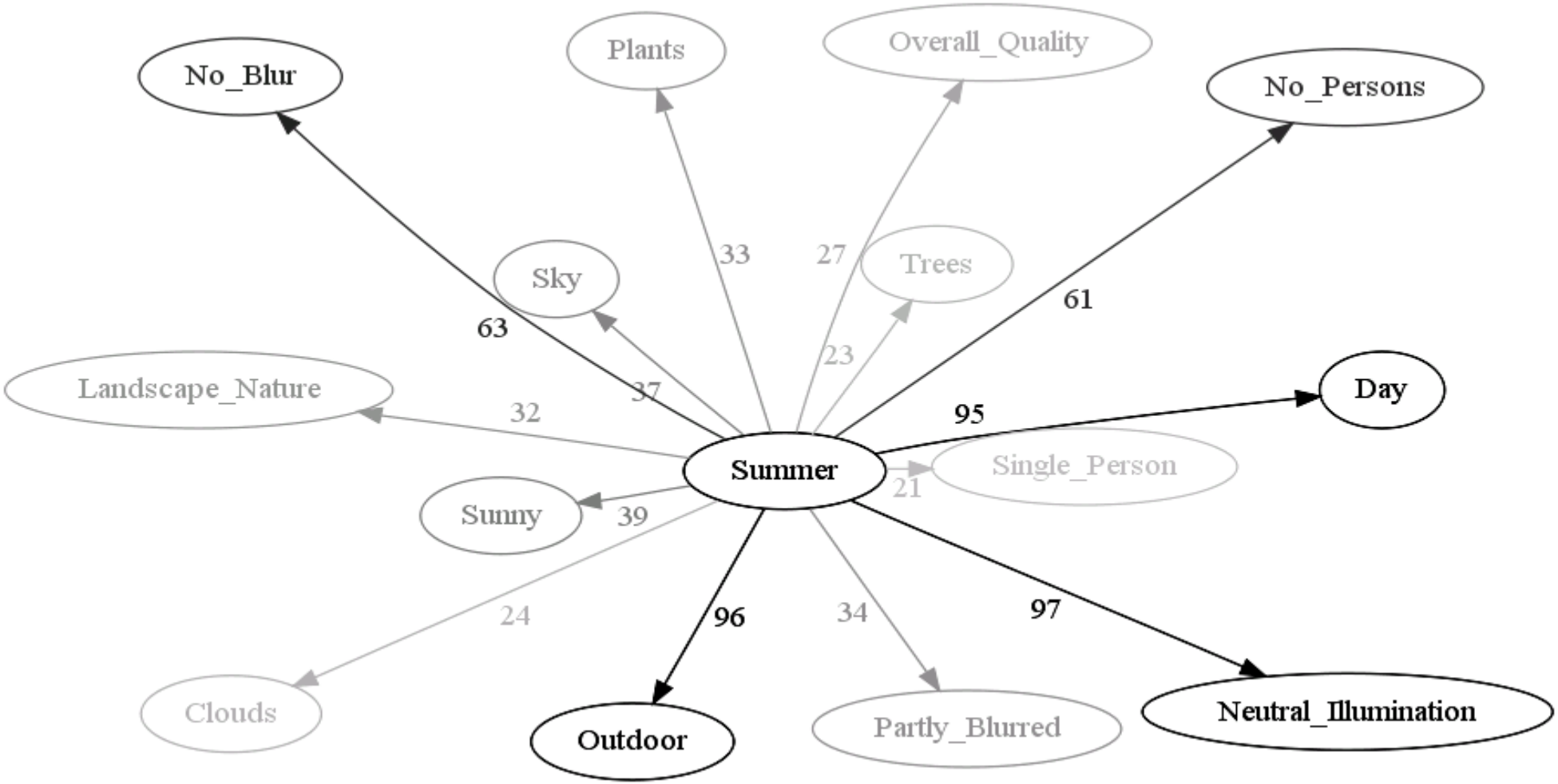
	EER	AUC
SIFT + Segmentation + Cross	0.291718	0.773133
SIFT + Segmentation	0.296315	0.771324
Log Odds combination	0.304113	0.746300
SIFT	0.322632	0.733264
Segmentation	0.346106	0.707860



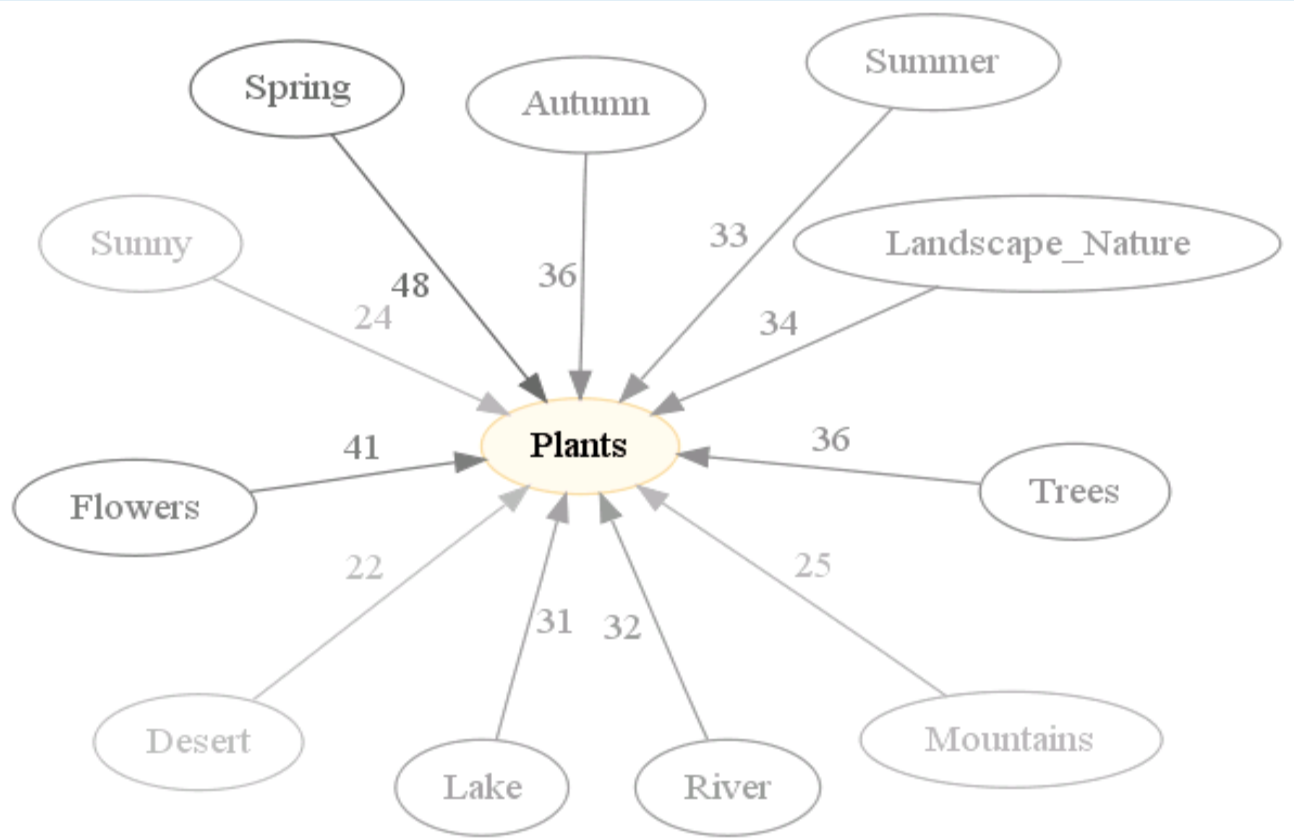
What more is in annotations?



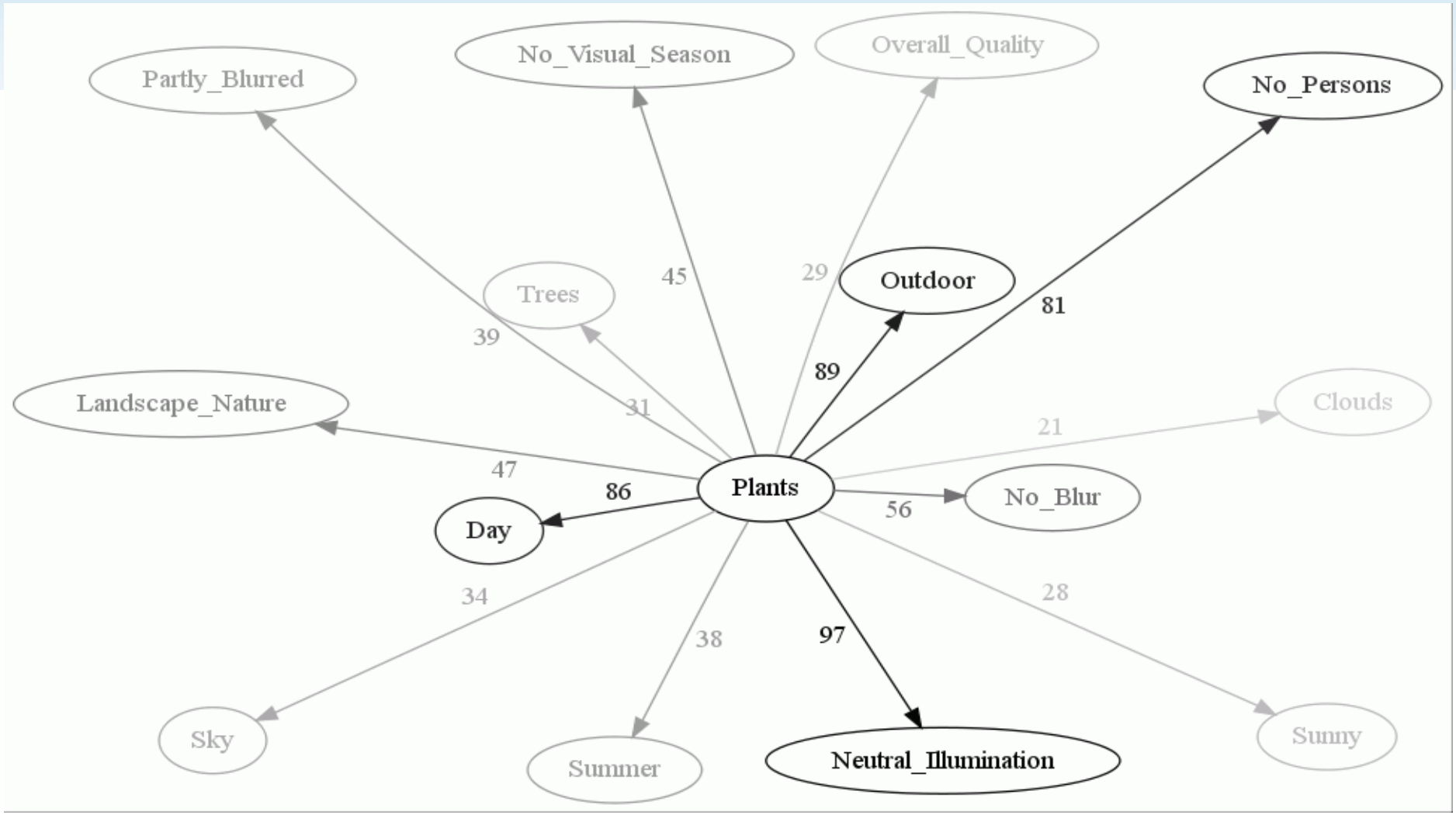
What more is in annotations?



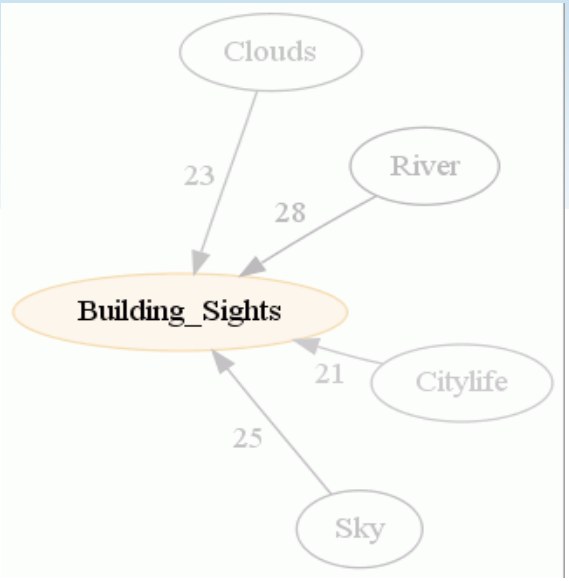
What more is in annotations?

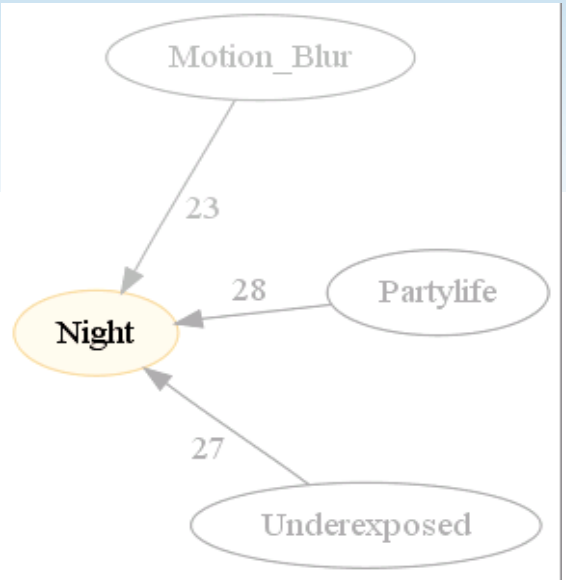


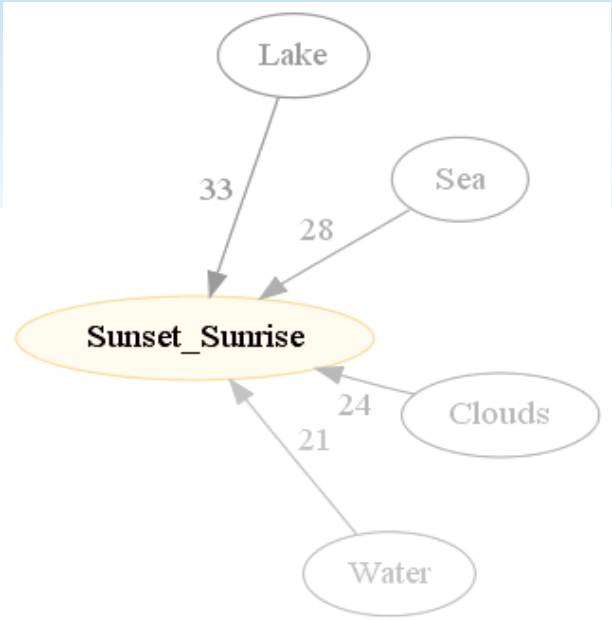
What more is in annotations?

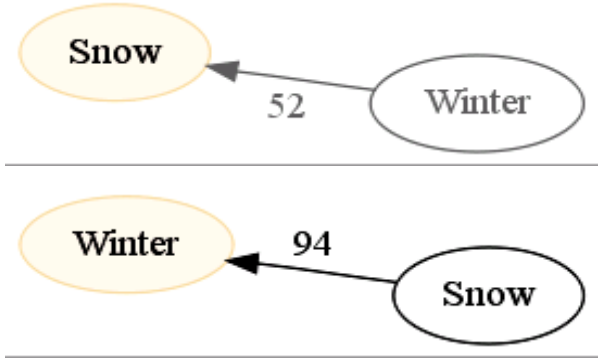


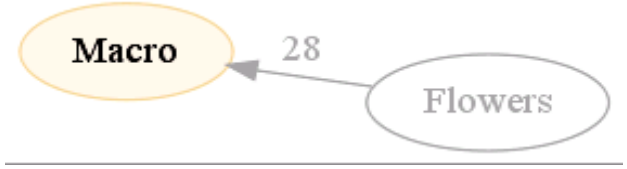
Thank You!











István Petrás, Bálint Daróczy, Zsolt Fekete