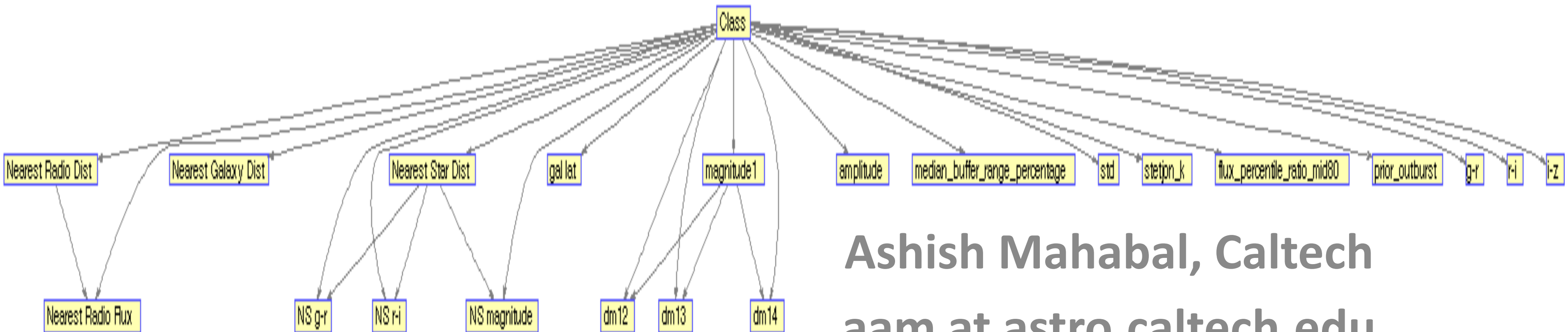
Connecting Bayesian Networks and Semantics
(as a step towards complete classification)



Ashish Mahabal, Caltech

aam at astro.caltech.edu

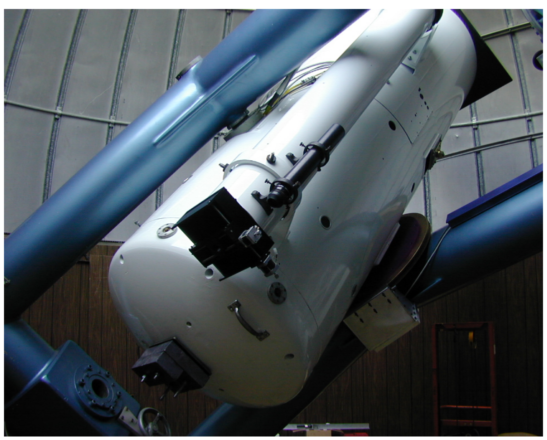
AstroInformatics 2012, MSR, Redmond

With Alex Ball, George Djorgovski, Ciro Donalek, Andrew Drake, Matthew Graham, Baback Moghaddam, Mike Turmon, Roy Williams, NS Philip, …

Transient astronomy has been gaining traction

+/- 5 deg ecliptic	-25 < Dec < +70	-80 < Dec < -25
1.2	8.1	4.2
21.5	19.5	19.0

Catalina Real-time Transient Survey
2,500 deg^2 / night (total 30000+)



GALEX, Spitzer, FIRST, …

Recent, current and future multiepoch surveys

LIGO + IndIGO …

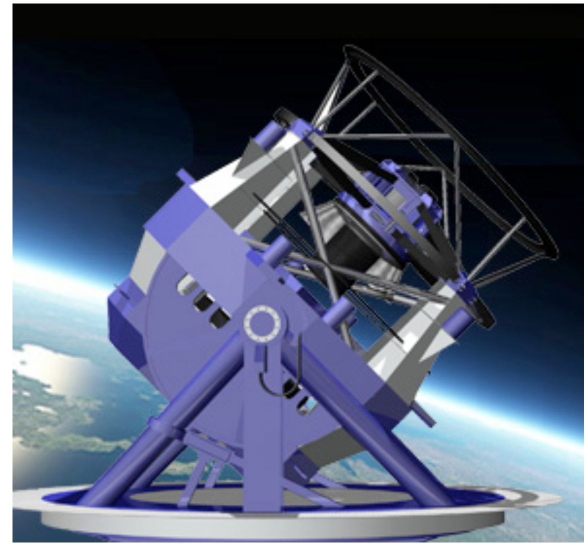
PTF

Skymapper

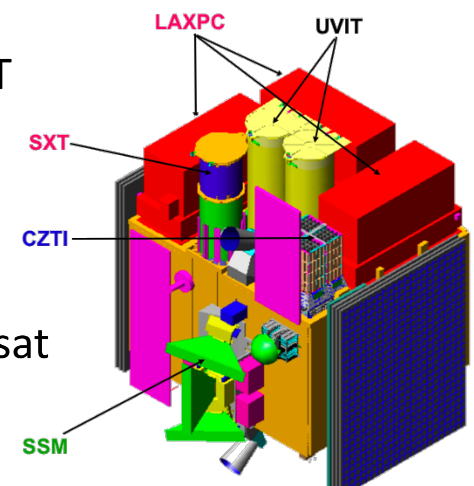
Pan-STARRS

Data capacities orders of magnitudes different.

Ashish Mahabal

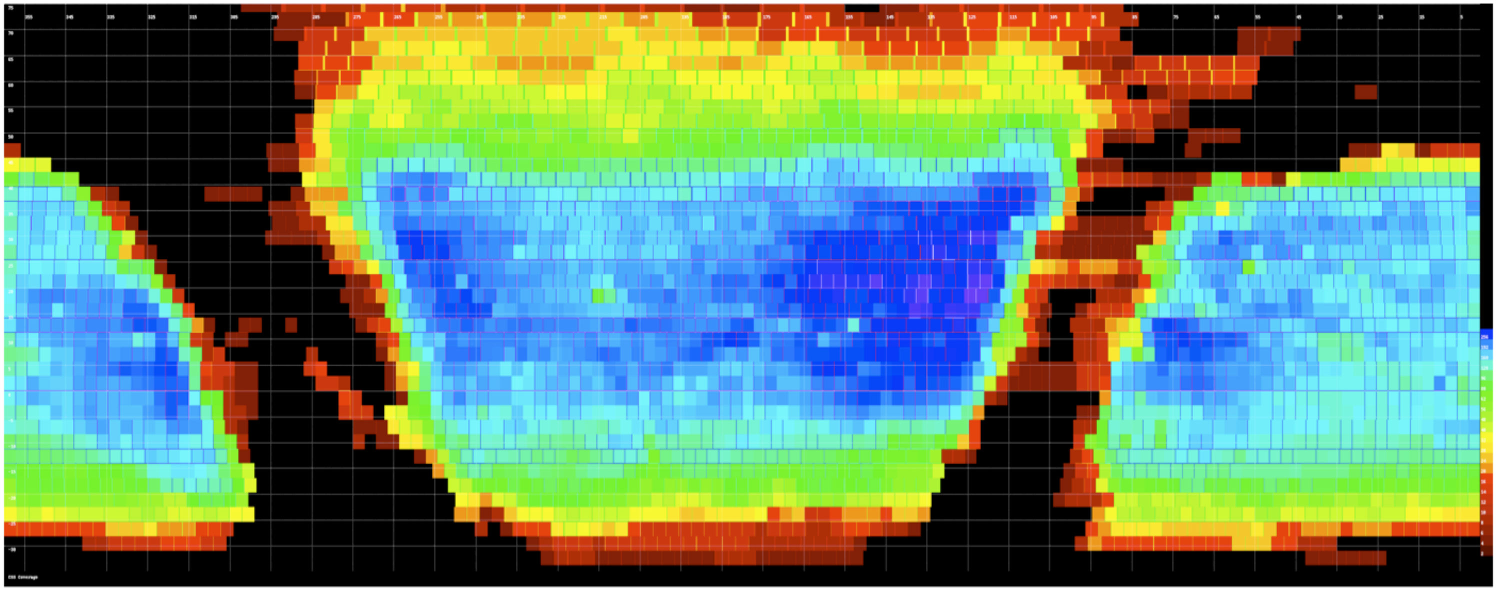


LSST



Astrosat

CRTS has discovered over 6000 transients



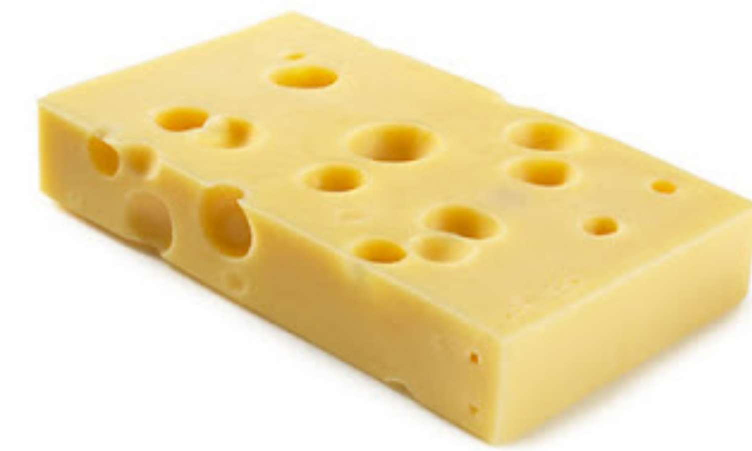
Telescope	All OTs	Supernovae	Cataclysmic Variables	Blazars	Asteriods/Flares	CV or SN	AGN	Other
CSS	2773	826	612	151	228	372	337	324
MLS	2729	374	53	29	195	499	1345	387
SSS	533	78	199	12	8	87	24	130
SNhunt	146	146	0	0	0	0	0	0
Total	6181	1424	864	192	431	958	1706	841

Associated with the transients are several types of parameters.

But we have a Variety of parameters that can be used

- Discovery: magnitudes, delta-magnitudes
- Contextual:
  - distance to nearest star
  - Magnitude of the star
  - color of that star
  - normalized distance to nearest galaxy
  - Distance to nearest radio source
  - Flux of nearest radio source
  - Galactic latitude
- Follow-up
  - Colors (g-r, r-I, i-z etc.)
- Prior classifications (event type)
- Characteristics from light-curve
  - amplitude
  - Median buffer range percentage
  - Standard deviation
  - Stetson k
  - Flux percentile ratio mid80
  - Prior outburst statistic

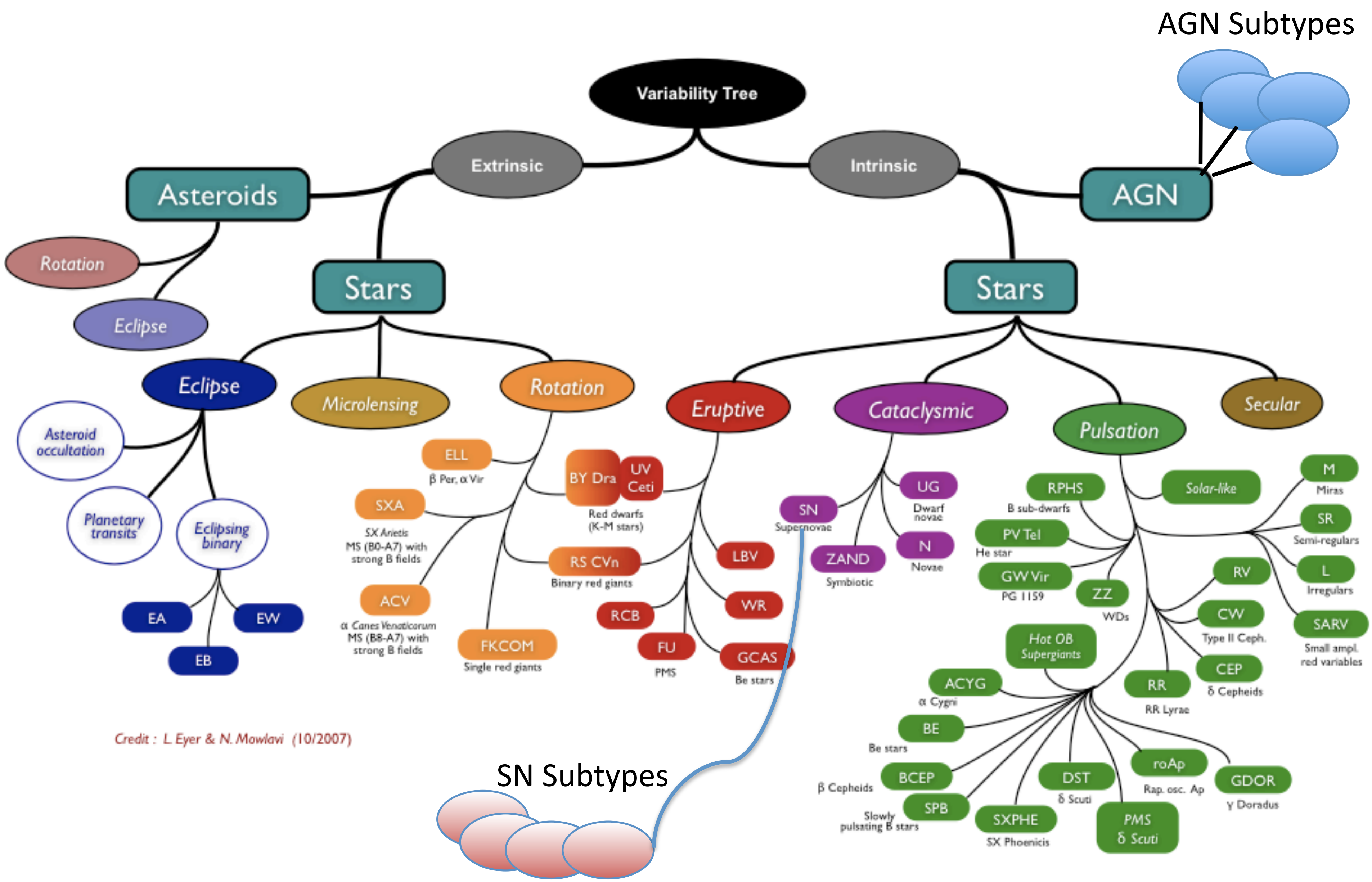
Not all parameters are always present leading to swiss-cheese like data sets



http://ki-media.blogspot.com/

Bayesian Networks best to deal with such datasets as they can deal with missing data and the structure can be learnt from the data – at least in principle

We are interested in all types of Transients

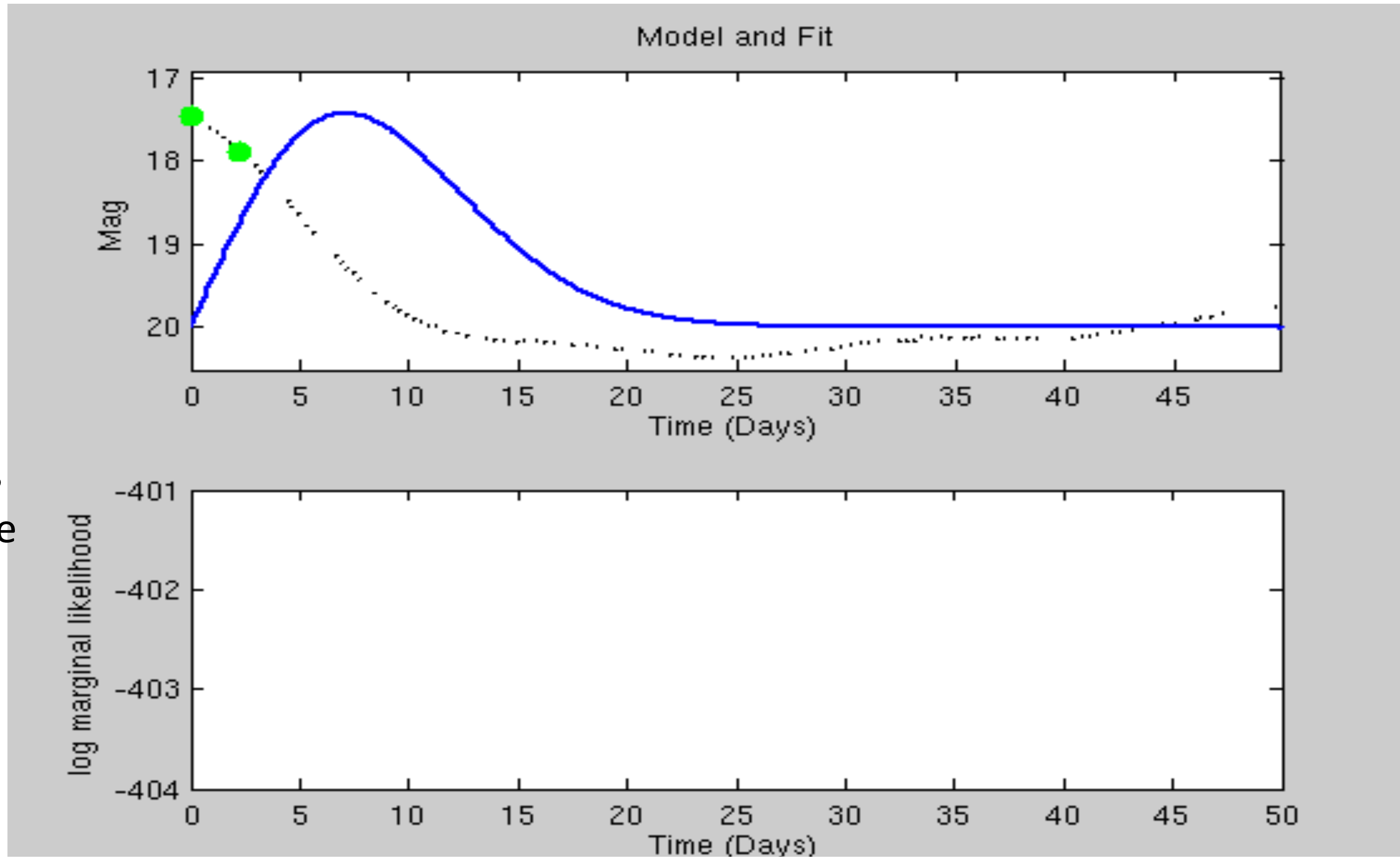


GPRs can be useful for this

Given several epochs and corresponding magnitudes, estimate the likelihood of a particular magnitude for a new epoch (using some covariance function)

B Moghaddam

hyperparameters are “free” and are varied (all 3)



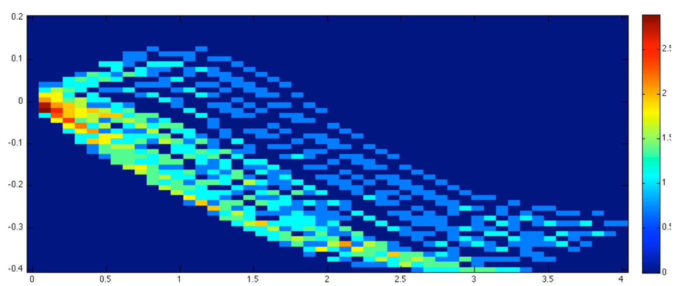
$$Cov(f(x_p), f(x_q)) = k_y(x_p, x_q) = \sigma_f^2 e^{-\frac{1}{2} l^2 (x_p - x_q)^2} + \sigma_n^2 \delta_{pq}$$

dm/dt Classifier Architecture

Decision Tree decomposes this multi-class classifier into a series of binary discrimination tasks.

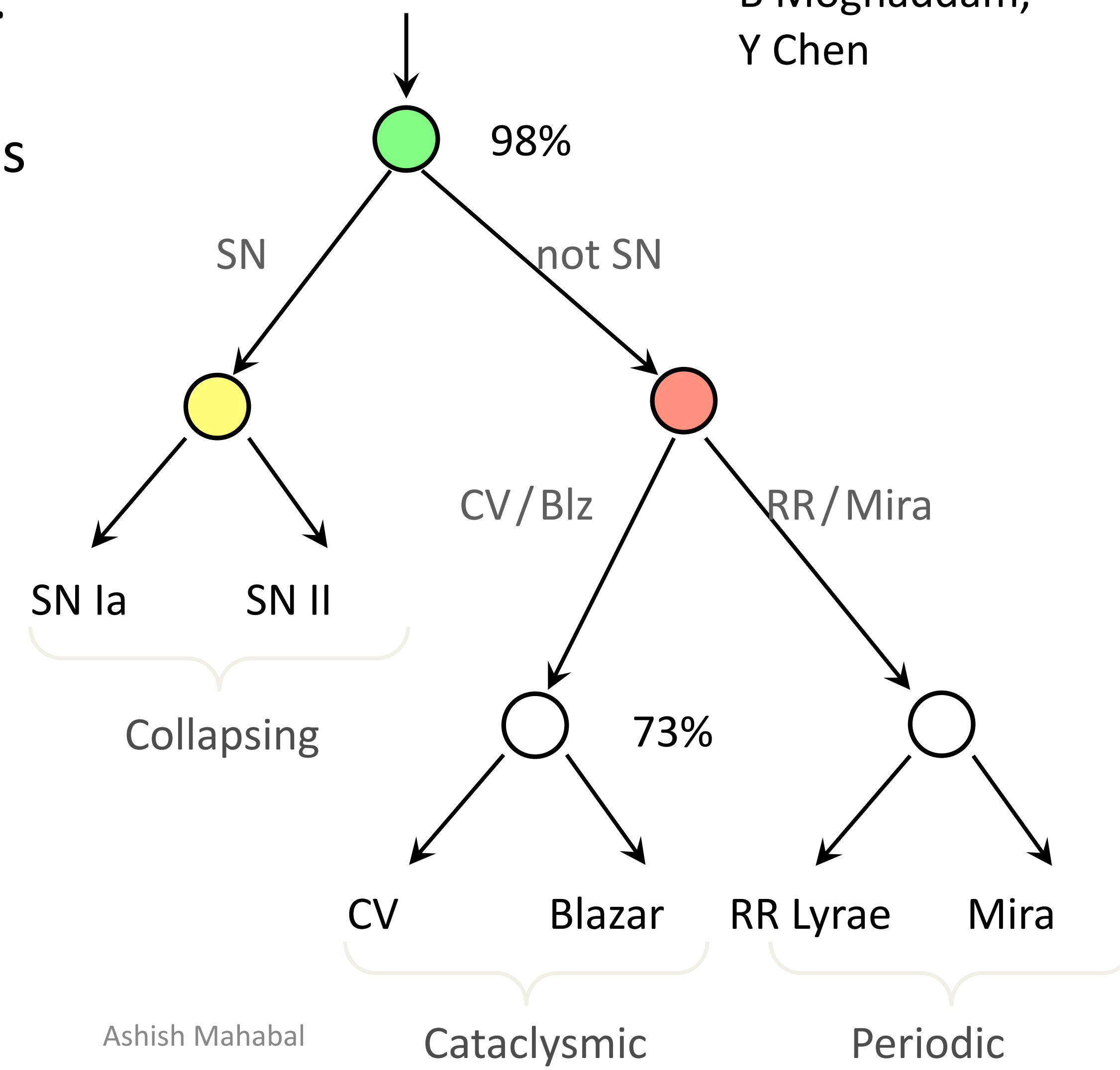
This specific DT follows the stratification that seems natural to astronomers.

All nodes shown were Implemented via "jump histogram" binary classifiers.

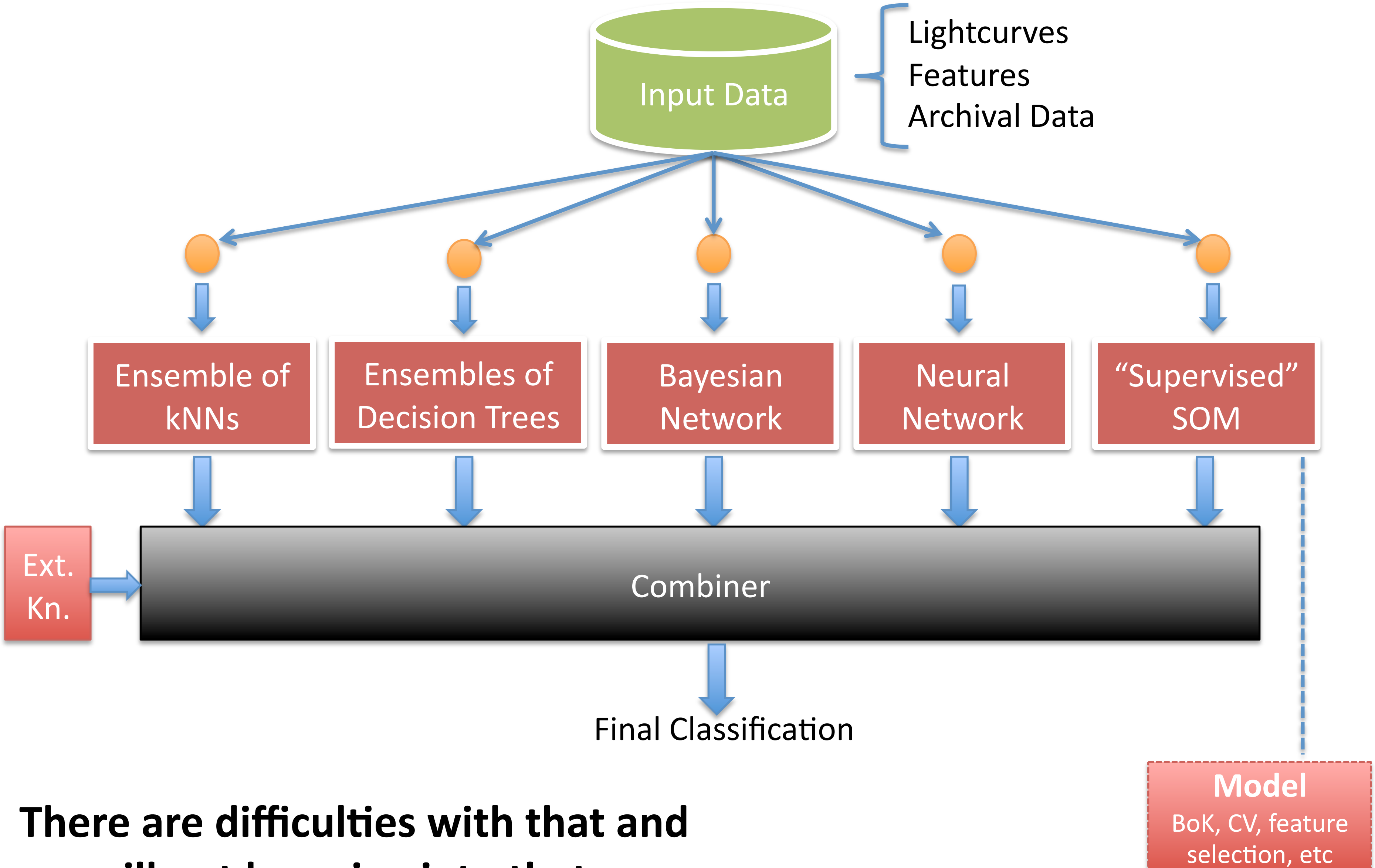


With GPRs and dm/dt, the more epochs you have the better the results are

B Moghaddam, Y Chen



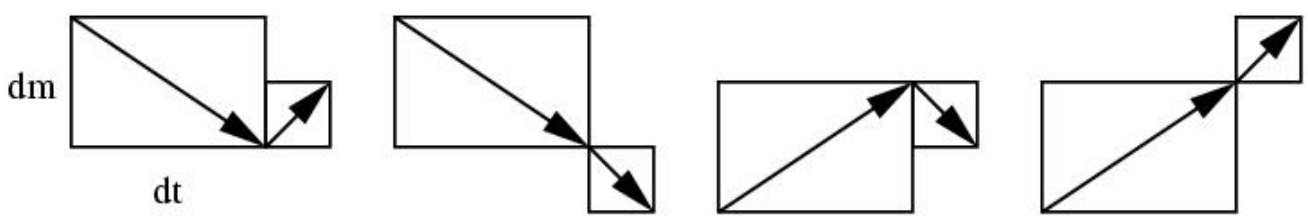
Combiner



There are difficulties with that and we will not be going into that.

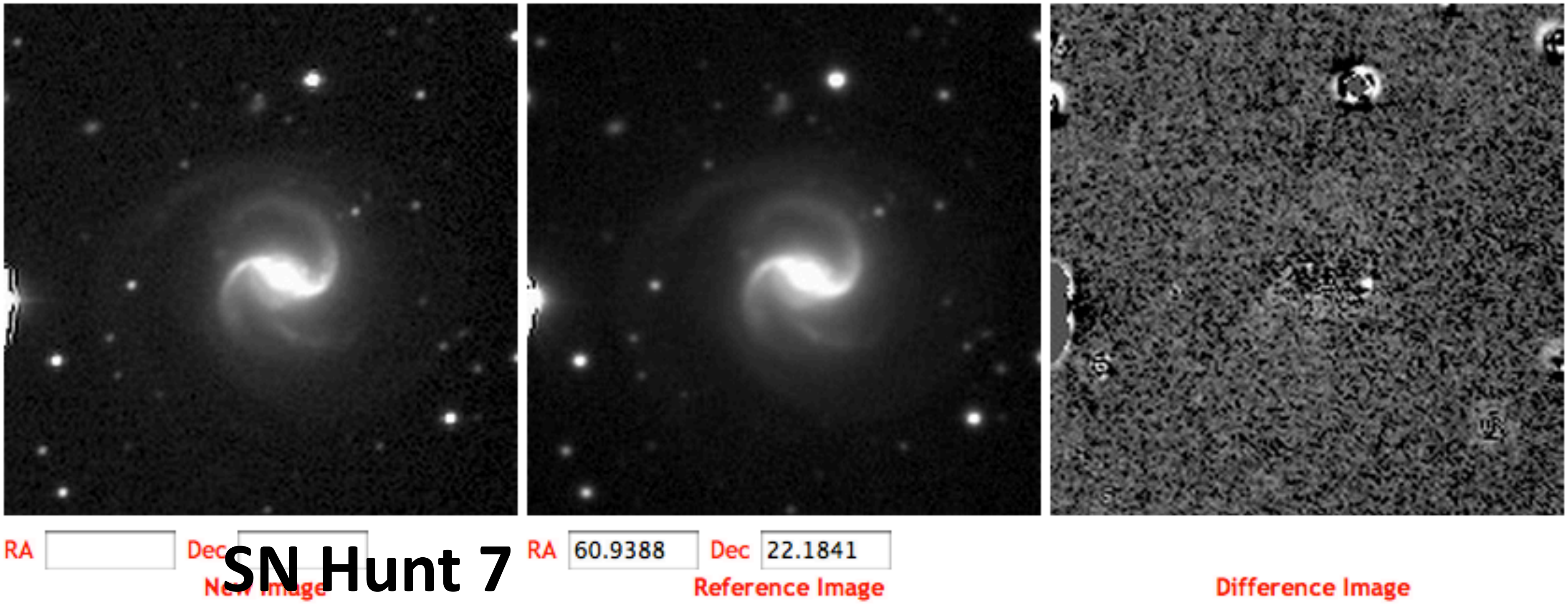
As part of the general theme of classifying transients and variables, it is important to separate SNe from non-SNe

- Many bits of information can be used towards that
  - Broadly speaking SNe rise only once (not strictly, but we are talking large delta-mag in small delta-time here), allowing for early characterization



Just three points can be used in principle to say if an object is a supernova or not

SN Hunt 7

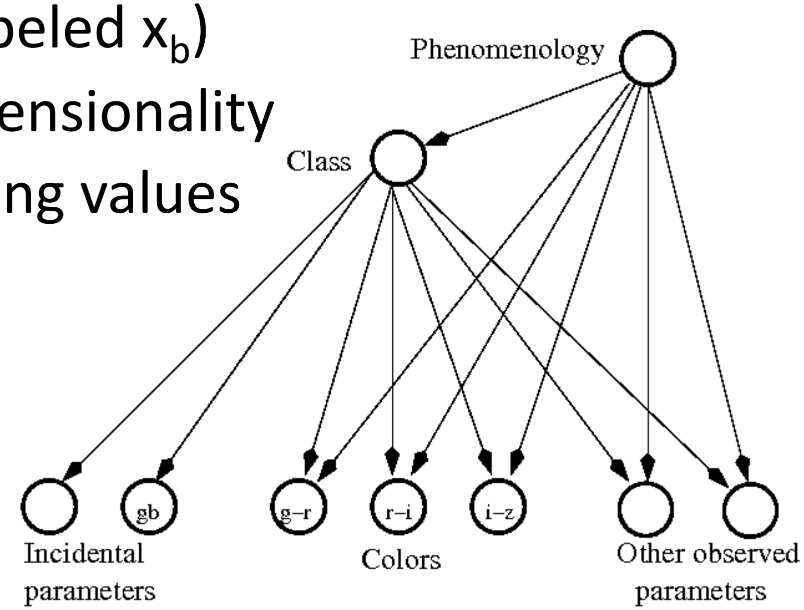


- Many surveys are primarily after SNe
  - Often using image subtraction (and thus targeting known galaxies)
  - One problem is that dwarf galaxies are missed
    - And there are many such

Naïve Bayes

$$P(y=k \mid x) = P(x \mid y=k)P(k)/P(x) \propto P(k)P(x \mid y=k) \approx P(k)\prod_{b=1}^{B} P(x_b \mid y=k)$$

- x: feature vector of event parameters
- y: object class that gives rise to x (1<y<k)
- Certain features of x known: (position, flux)
- Others will be unknown: (color, delta-mag)
- Assumption: based on y, x is decomposable into B distinct independent classes (labeled x_b)
- This helps with the curse of dimensionality
- Also allows us to deal with missing values



Radio, for instance

BN with P60 follow-up colors:
CV/SN classification ~80% with single epoch

BNs allow us to state dependencies and independencies between variables (nodes)

Very broadly speaking 5 flavors of BNs possible

- Naïve
- TAN
- constructed (semantics, expert knowledge etc.)
- single winner from several naïve
- Fully learned from data

Naive Bayesian Network is good because its easier

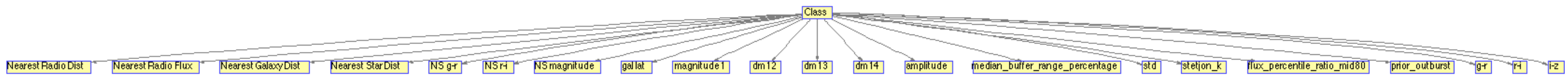
Learning from data is time consuming

Missing values complicate matters

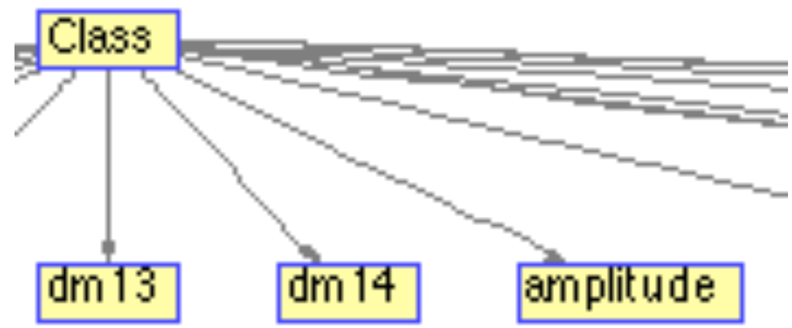
TAN often "learns" wrong connections

As number of parameters grow, search space increases super-exponentially

n	G(n)
1	1
2	3
3	25
4	543
5	29,281
6	3,781,503
7	1.1 x 10^9
8	7.8 x 10^11
9	1.2 x 10^15
10	4.2 x 10^18

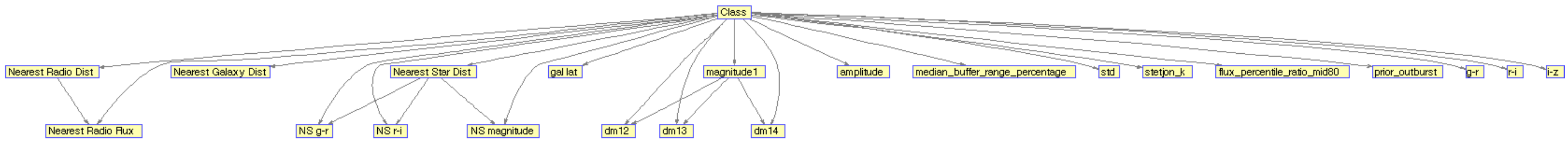


Output table from Naïve BN

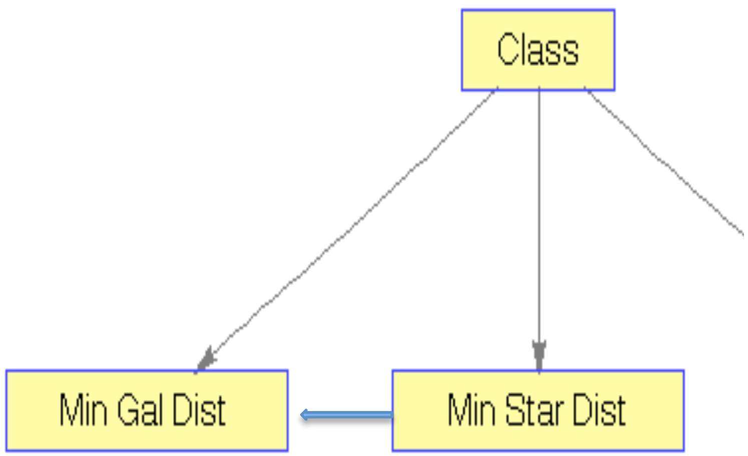


Predicted\Real	CV	SN	Blazar	AGN	UVCeti	Mira	HPMS	Total
CV	215	14	14	6	4	6	1	260
SN	48	388	15	25	5	0	12	493
Blazar	20	4	57	5	0	1	0	87
AGN	8	21	14	81	0	1	0	125
UVCeti	0	0	1	0	41	0	2	44
Mira	0	1	0	0	0	8	0	9
HPMS	1	3	0	0	0	0	83	87
Total	292	431	101	117	50	16	98	1105
Completeness	0.736	0.900	0.564	0.692	0.820	0.500	0.847	
Contamination	0.173	0.213	0.345	0.352	0.068	0.111	0.046	

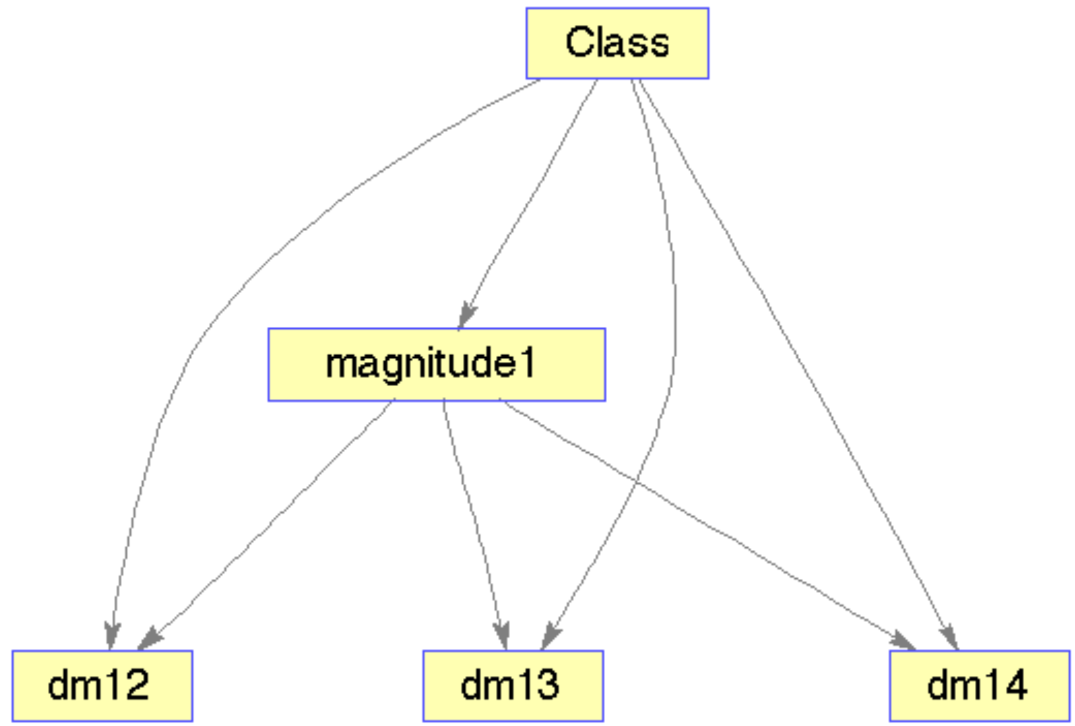
A Ball



Tree Augmented Network



Wrongly connected

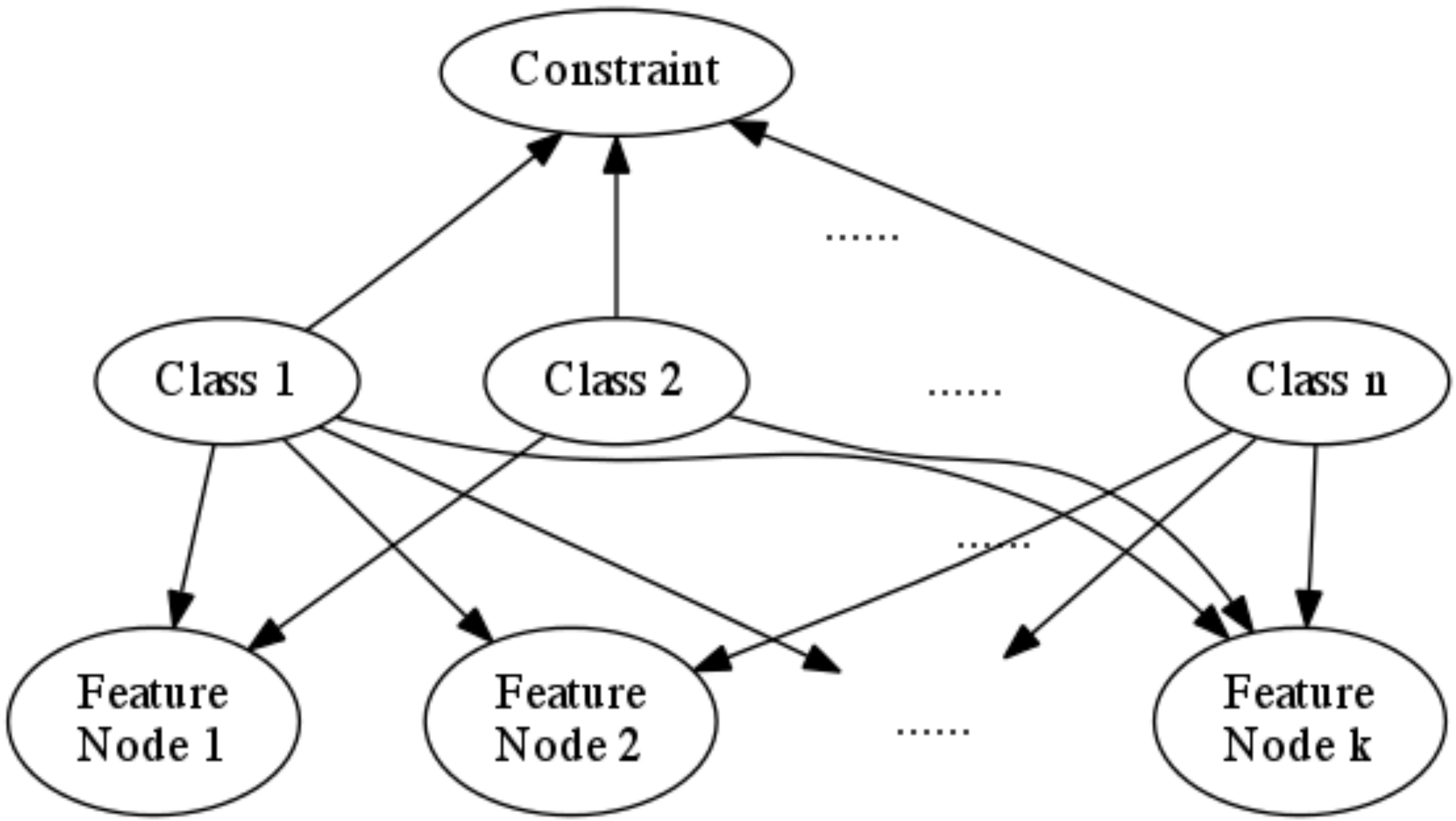


TAN can sometimes “learn” wrong connections

Output table for TAN BN (clearly worse than naïve)

Predicted\Real	CV	SN	Blazar	AGN	UVCeti	Mira	HPMS	Total
CV	208	34	14	4	6	4	4	274
SN	57	360	16	32	4	0	22	491
Blazar	25	10	37	11	1	3	1	88
AGN	4	16	9	80	3	1	0	113
UVCeti	2	0	0	0	47	0	1	50
Mira	2	2	0	0	4	8	0	16
HPMS	5	4	0	0	1	0	63	73
Total	303	426	76	127	66	16	91	1105
Completeness	0.686	0.845	0.487	0.630	0.712	0.500	0.692	
Contamination	0.241	0.267	0.580	0.292	0.060	0.500	0.137	

The aim …



Constraint
……
Class 1
Class 2
……
Class n
……
Feature Node 1
Feature Node 2
……
Feature Node k

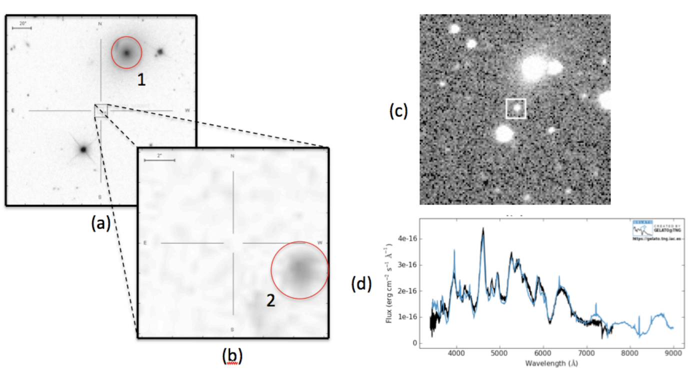
As part of the general theme of classifying transients and variables, it is important to separate SNe from non-SNe

Simple parameters that can be used:

- Distance to nearest star (and perhaps color)
- Galaxy proximity (normalized)
- Archival lightcurve (including with upper limits)

Proximity to a galaxy is useful in marking SNe

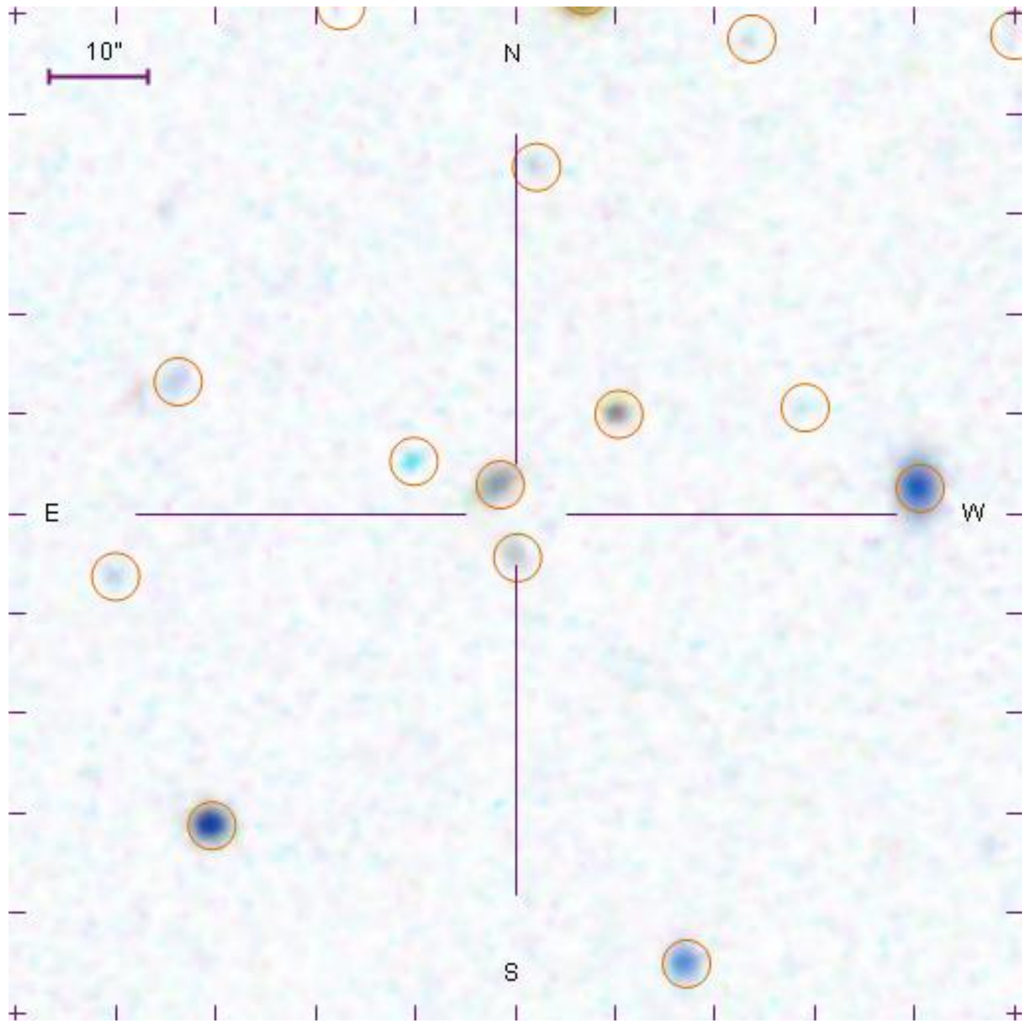
- Often there is confusion when more than one galaxy is present nearby
- As a result distance has to be normalized
  - Which radius to consider is also an important consideration
- Misclassification of S/G possible, so look at nearest stars
- Coincident stars will be a direct NO
- Dwarf galaxies (low SNR – not catalogued – have to be considered)



Only a small fraction have radio-counterparts – so that too is a NO (but if yes, its likely to be very interesting)

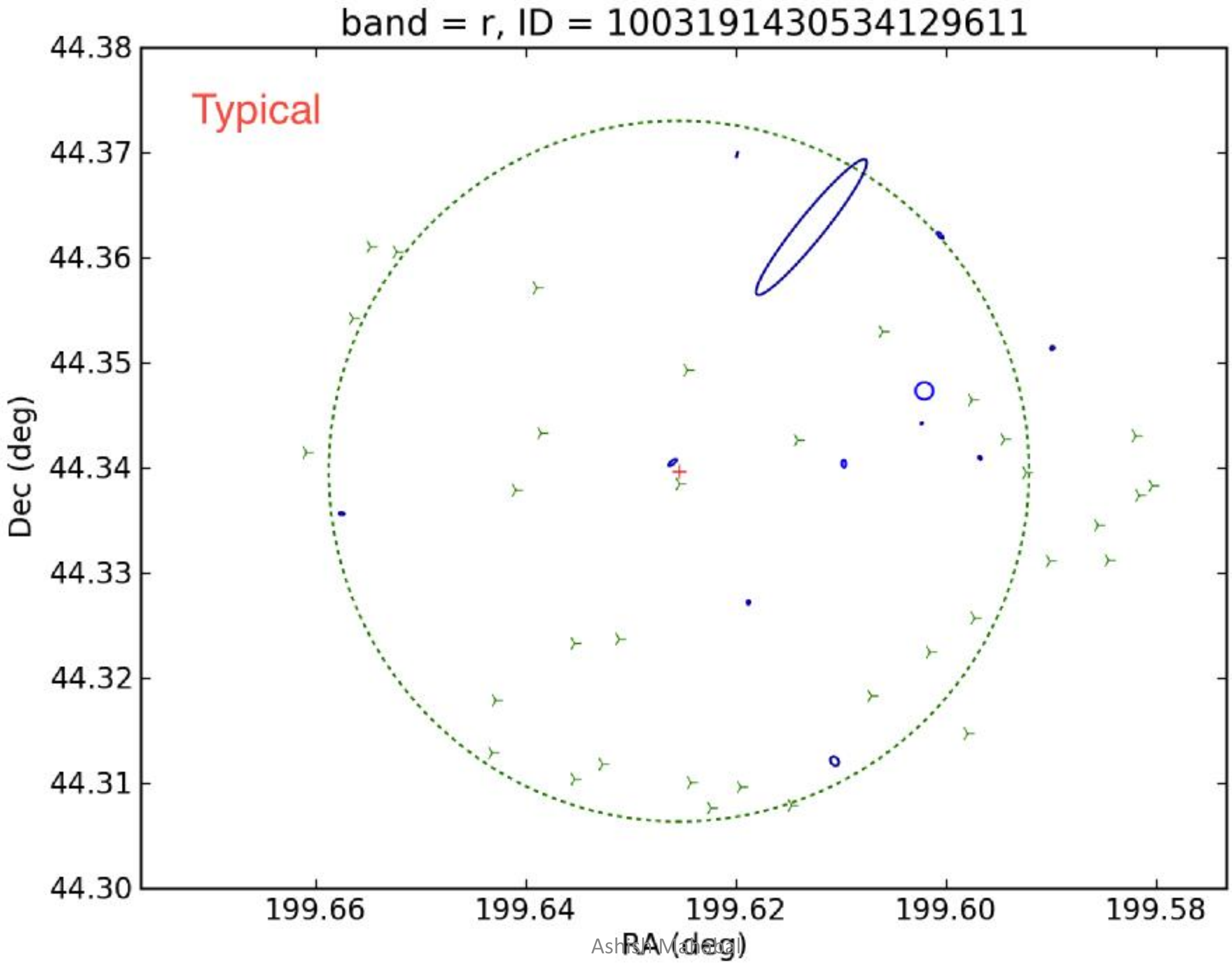
Definite SN – but in which galaxy?

The transient (not seen) is at the center)

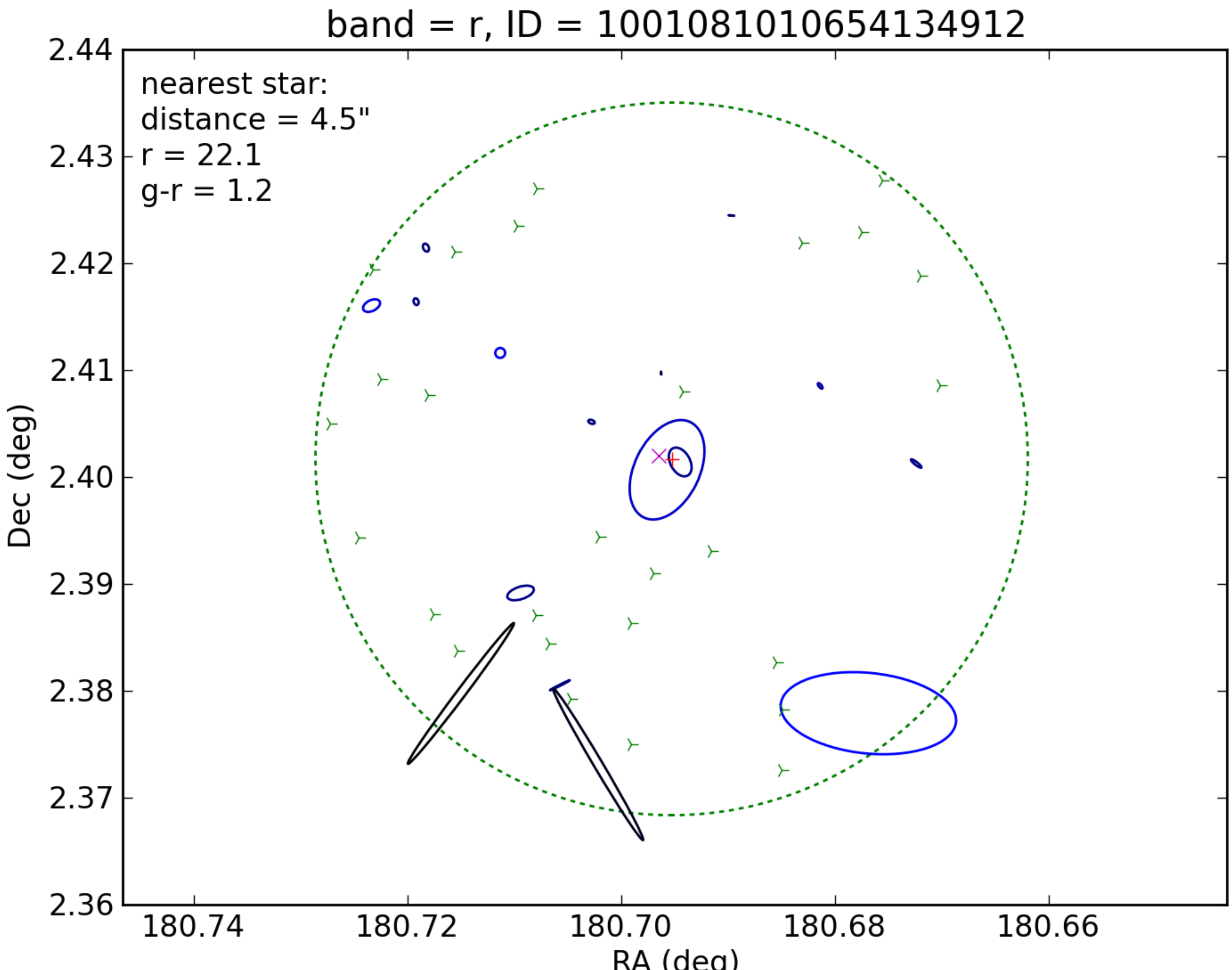


Ashish Mahabal

The corresponding schematic

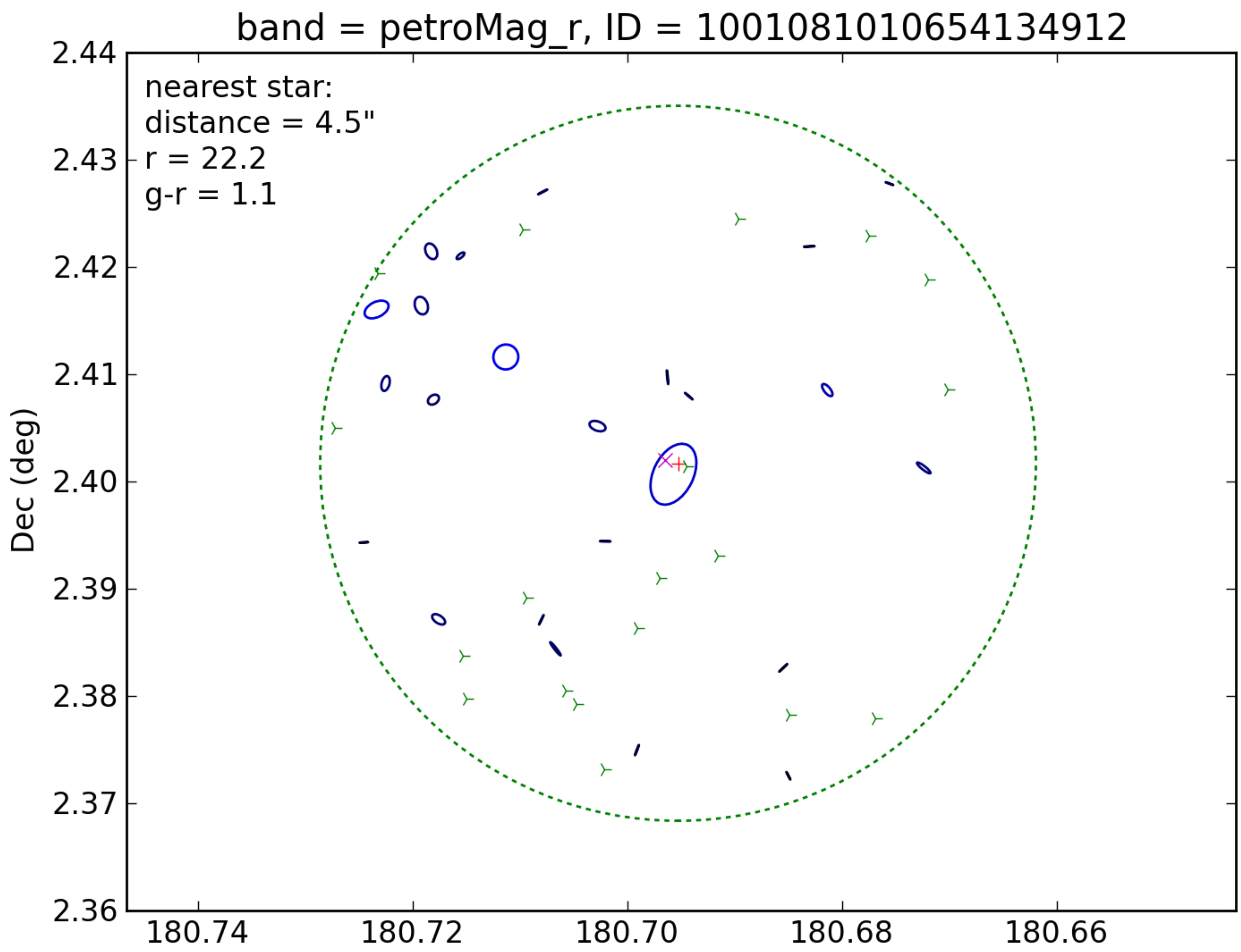


De Vaucouleurs' radius (and associated ellipticity) often misleading for some types of galaxies.



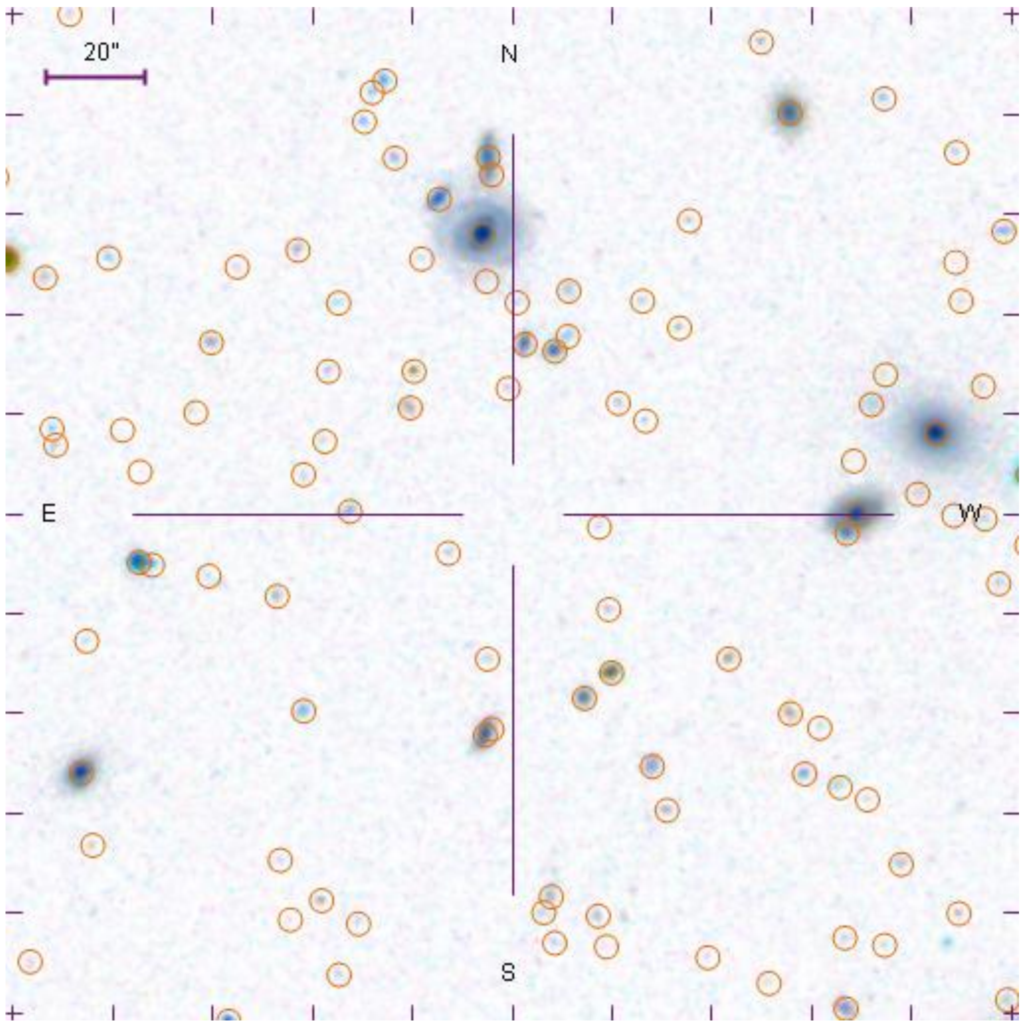
Ashish Mahabal

Petrosian radius often more representative

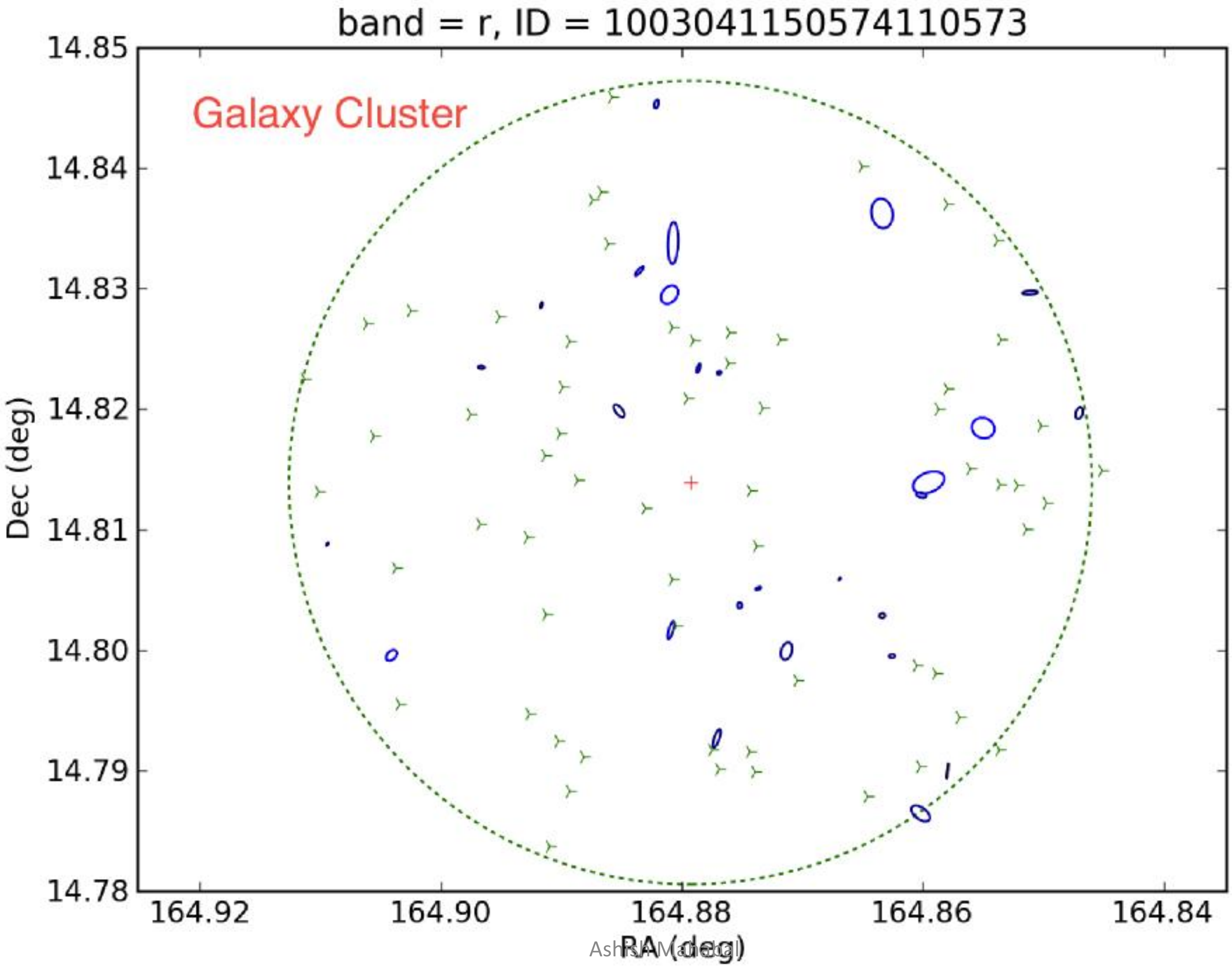


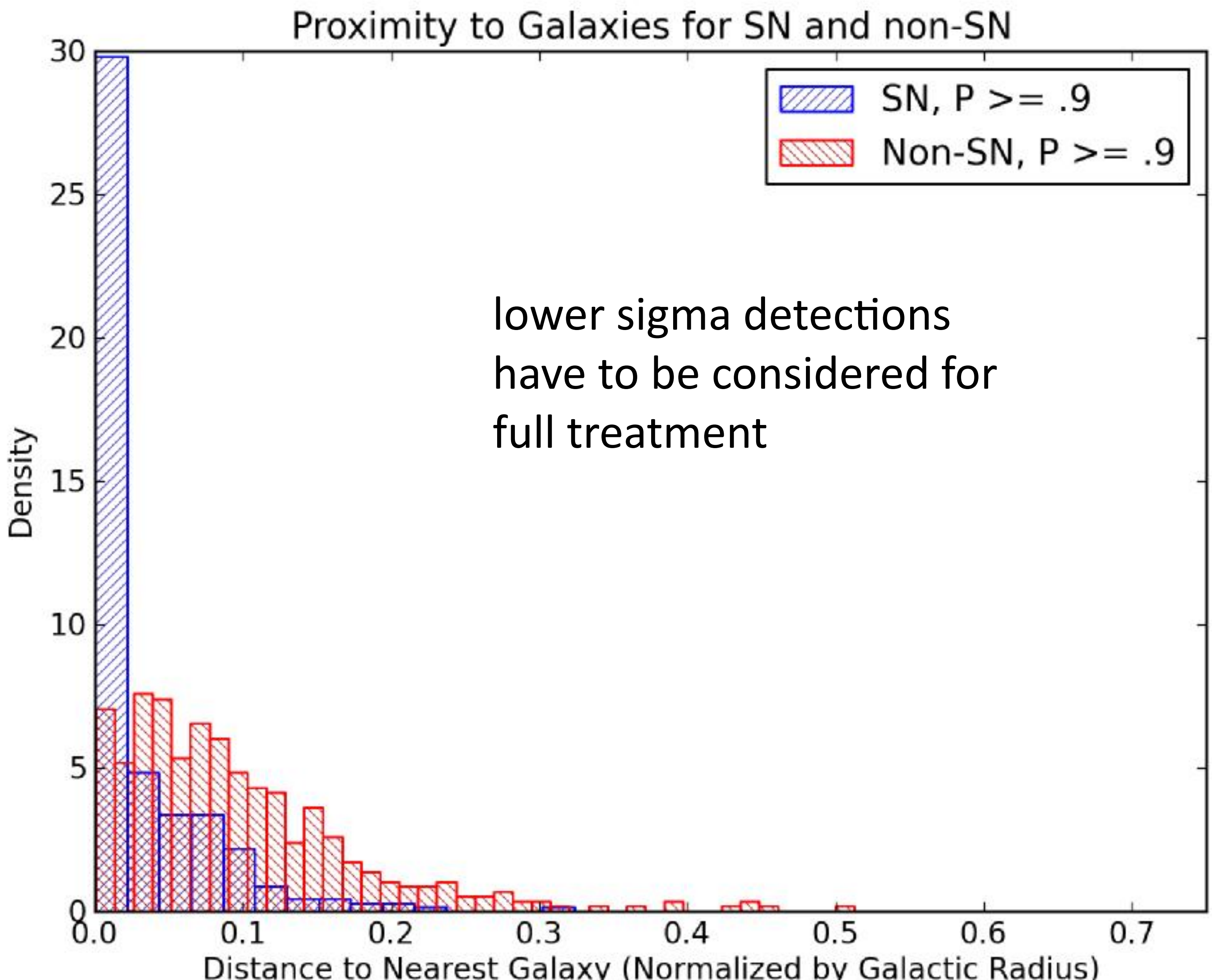
Possible SN in a cluster

No obvious galaxy near the center – can be easily missed



Schematic for the cluster field





Ashish Mahabal

Alex Ball

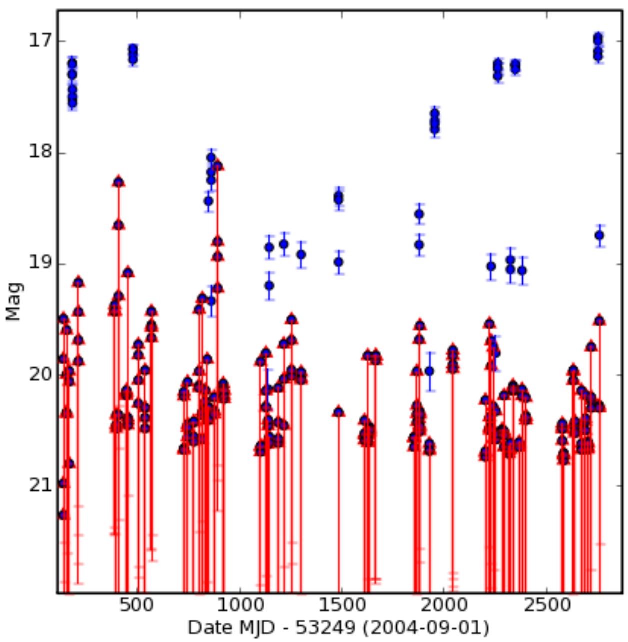
Using archival lightcurves, including upper-limits, one can separate SNe from sources like CVs

(due to specific high thresholds, earlier outbursts could have been missed making one mistake a CV for a supernova)

Exploiting statistics from peaks useful

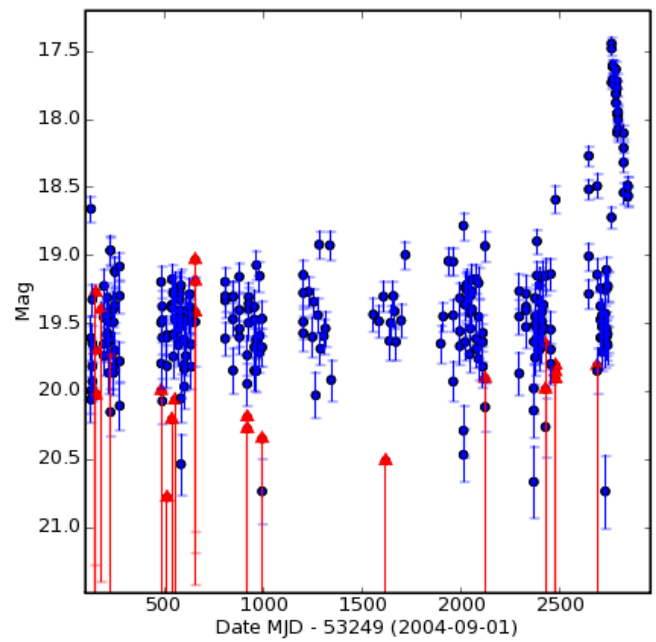
(1) Number of peaks (2) Points in each peak

(3) dt between peaks (4) Ratio of their significance



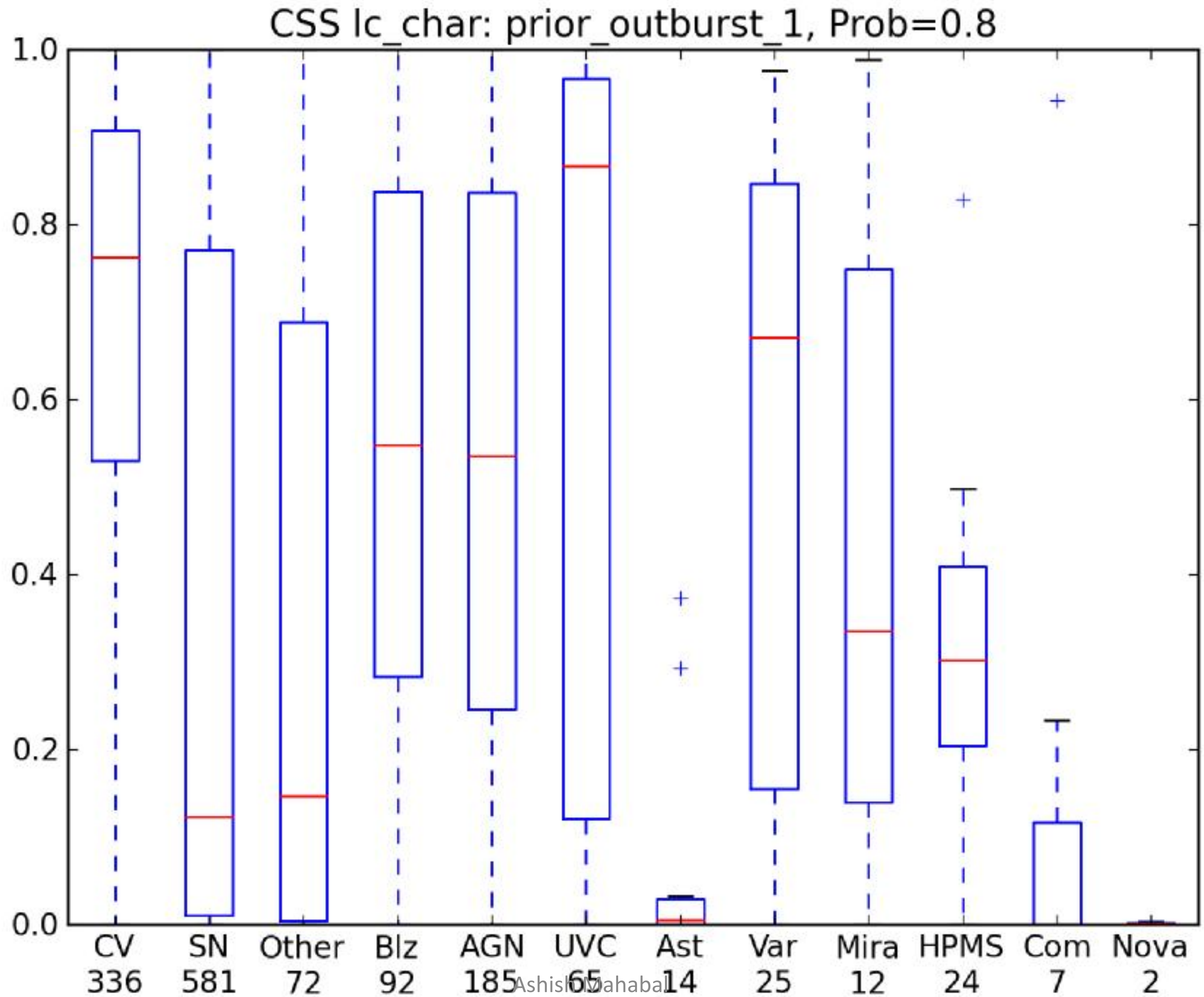
CV

SN



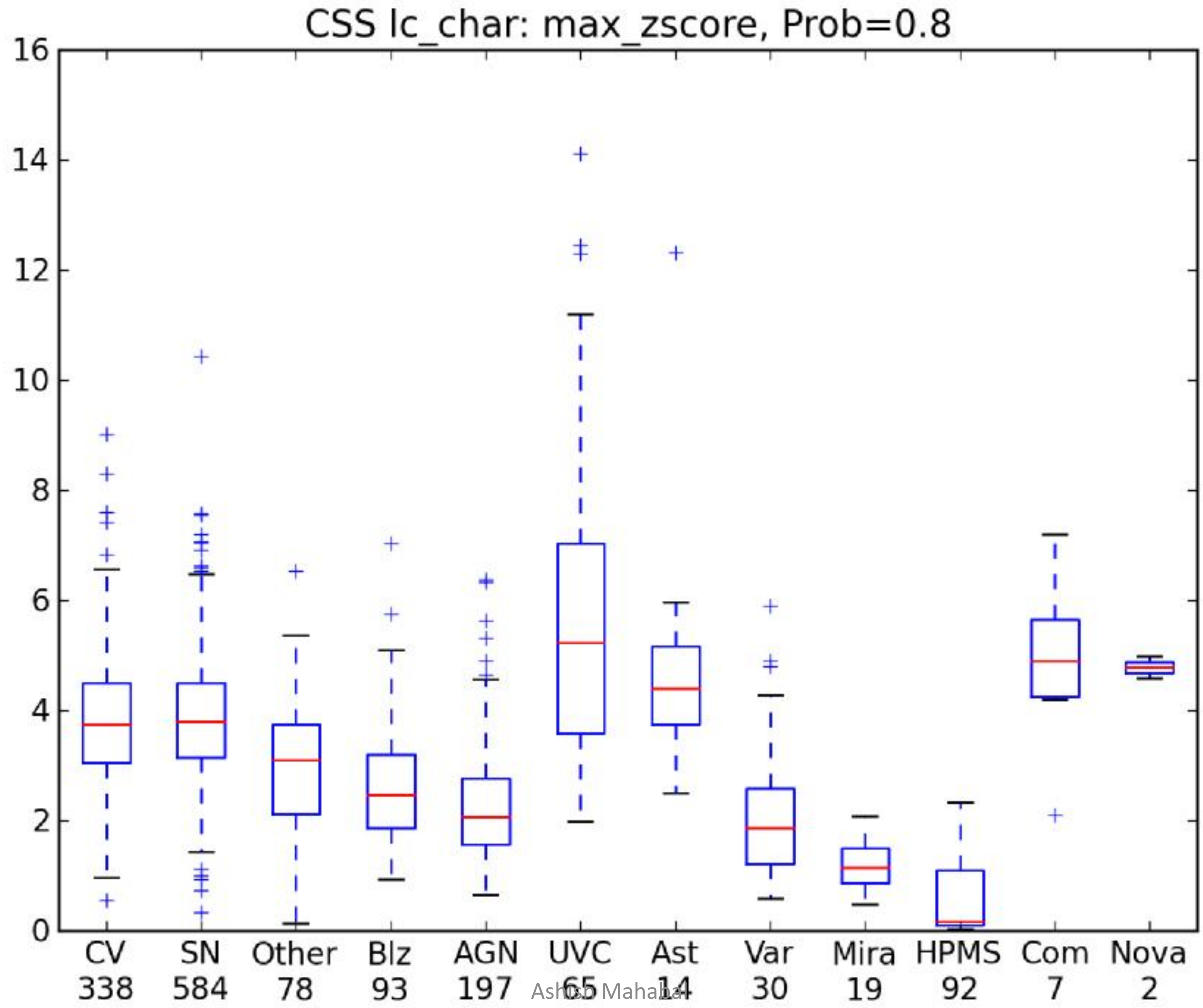
Ashish Mahabal

Boxplot with dt-ratio calculated from 1-sigma peaks

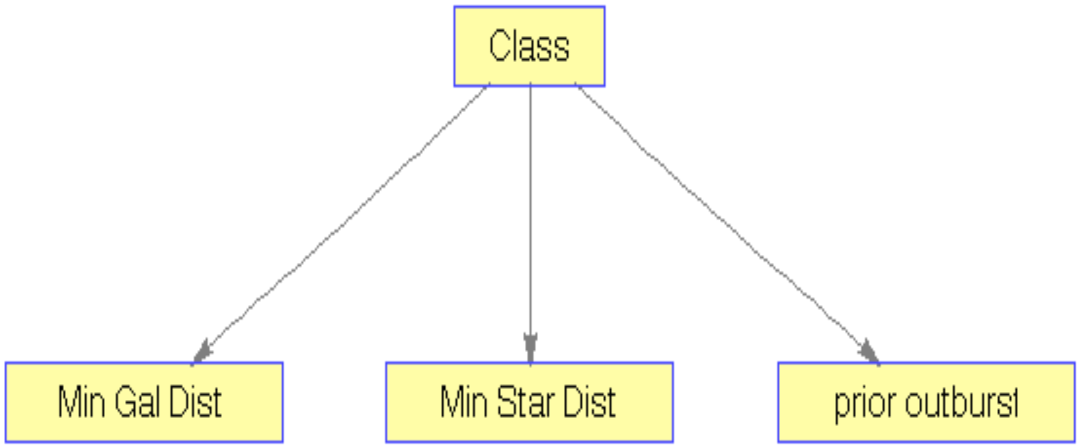


Ashish Mahabal

N-sigma statistics for different types of transients



Ashish Mahabal



normalized

Based on peaks

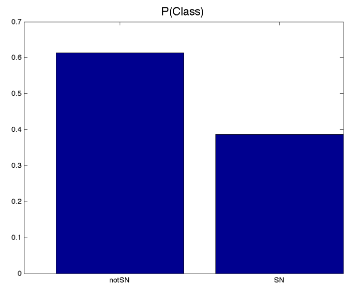
Separation of SNe and non-SNe

80-90% completeness

Classifier	Completeness nonSN	Completeness SN	Contamination nonSN	Contamination SN
3 param incomplete	0.792	0.797	0.139	0.293
3 param complete	0.827	0.917	0.078	0.181
2 param complete	0.807	0.866	0.111	0.228

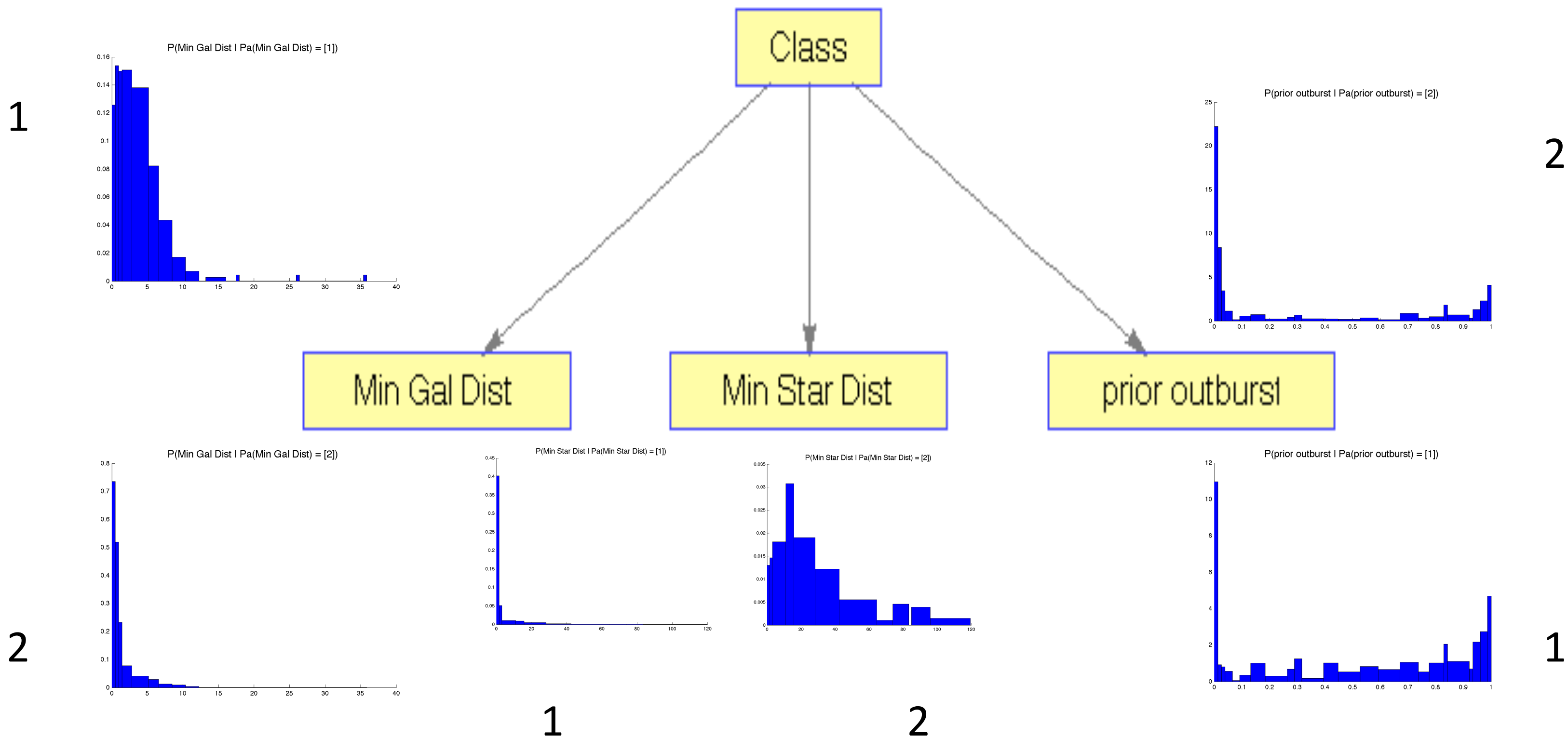
Using 900 non-SNe and 600 SNe

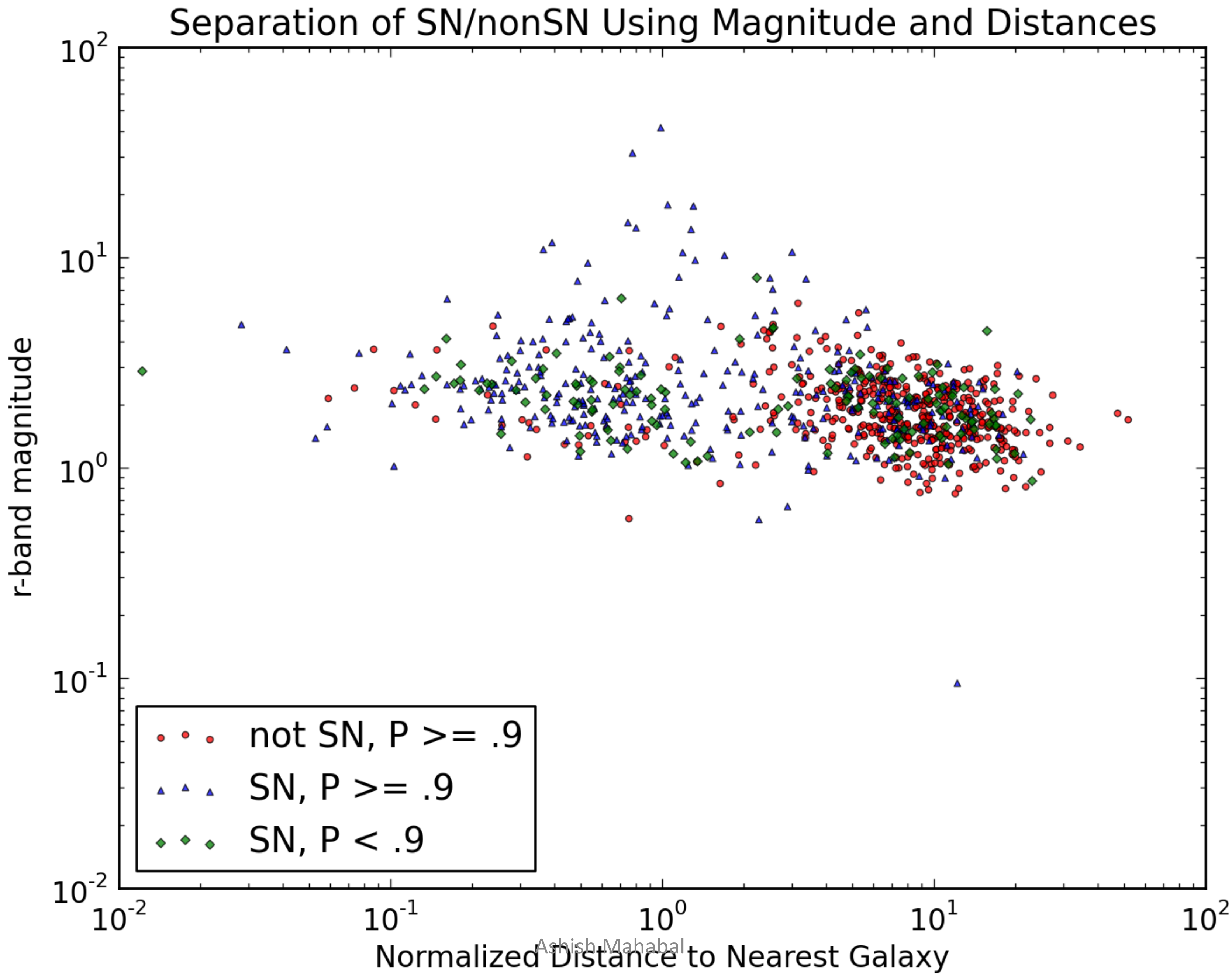
Non-SNe (1)



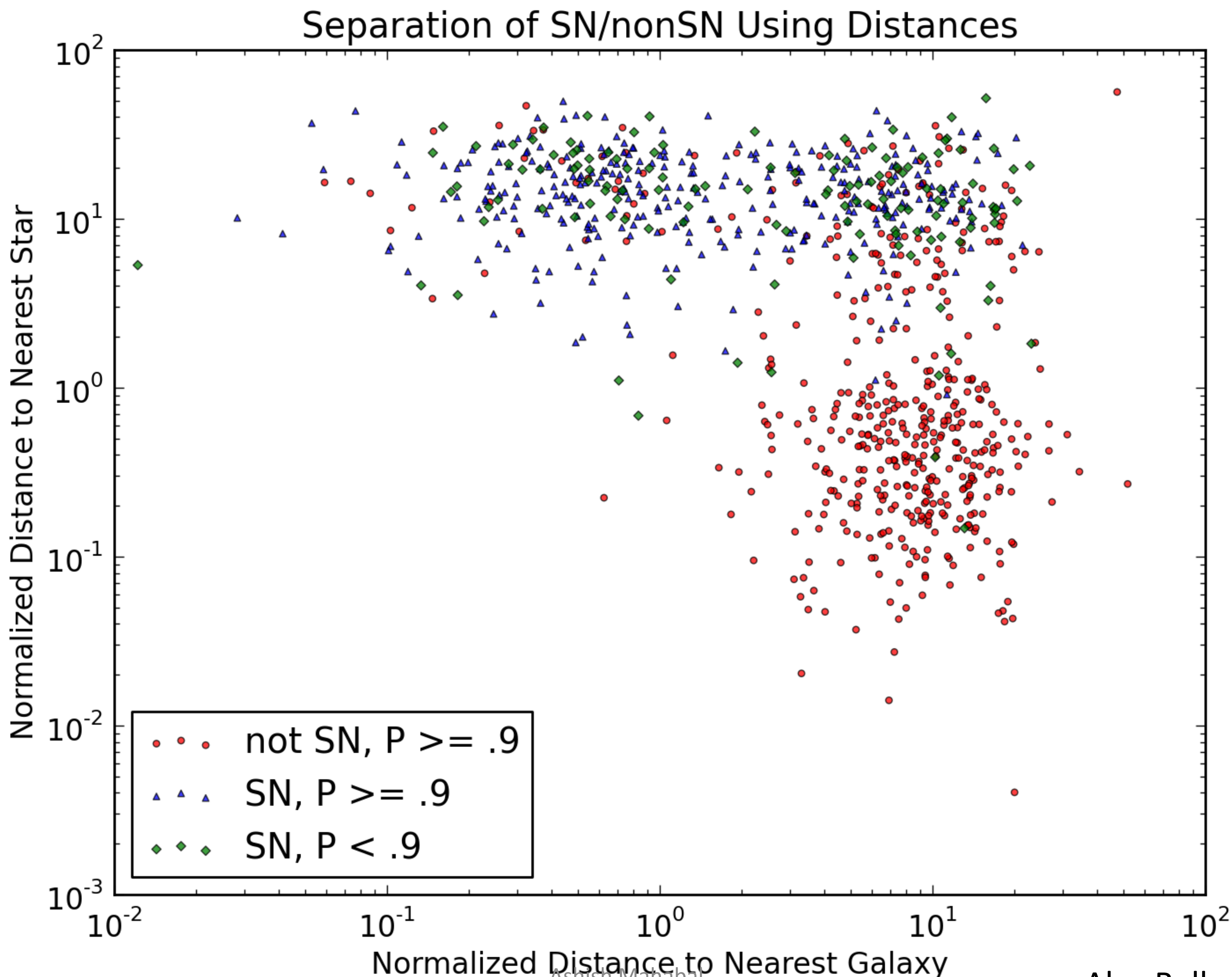
SNe (2)

80-90% completeness using just these parameters





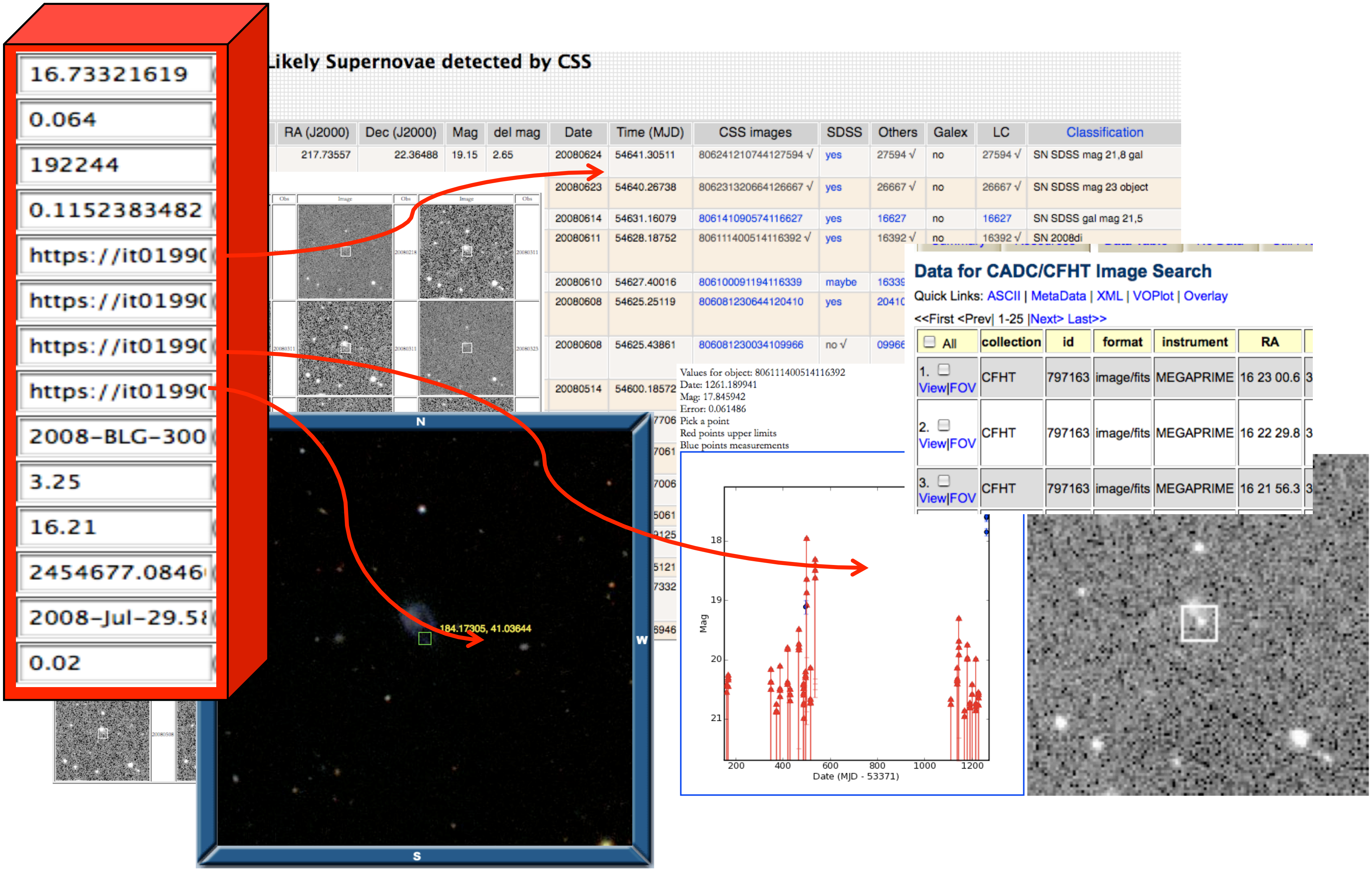
Separation of SN/nonSN Using Magnitude and Distances
r-band magnitude
Normalized Distance to Nearest Galaxy
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
not SN, P >= .9
SN, P >= .9
SN, P < .9
Ashish Mahabal



Ashish Mahabal

Alex Ball

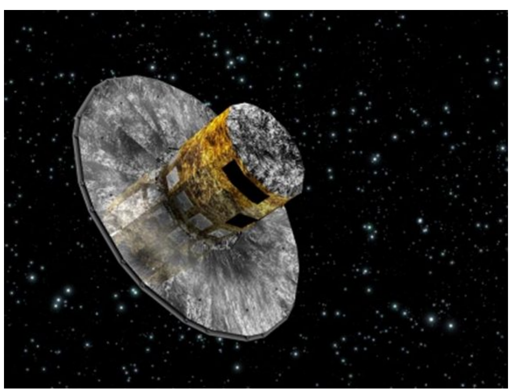
The new classifications feed back in to the portfolio of the object – to be combined with other bits



Ashish Mahabal

CRTS follow-up through various telescopes (Gaia network, IGO, P60, SAAO)

25087	ivo://nvo.caltech/voeventnet/catot#1204240090814131436	2012-05-17 12:31:29.738149	229.05438	-9.91775	6	LC
25086	ivo://nvo.caltech/voeventnet/sssot#1205140310714115953	2012-05-17 12:29:53.886991	164.16582	-31.37003	6	LC
25084	ivo://nvo.caltech/voeventnet/sssot#1204260070624132119	2012-05-14 08:56:37.080339	124.13131	-6.58991	6	LC
25083	ivo://nvo.caltech/voeventnet/catot#1204231150484101073	2012-05-14 08:52:38.205503	138.92333	14.16077	6	LC
25082	ivo://nvo.caltech/voeventnet/catot#1204231090504125181	2012-05-14 08:48:54.636982	141.1142	10.05259	6	LC
25081	ivo://nvo.caltech/voeventnet/catot#1204300010664124998	2012-05-09 14:25:31.463083	184.64554	-1.33181	6	LC
25080	ivo://nvo.caltech/voeventnet/catot#1204251210744104178	2012-05-09 14:24:11.979957	218.66588	19.86692	5	LC
25079	ivo://nvo.caltech/voeventnet/catot#1204111210464120236	2012-05-09 14:22:51.307489	135.61214	20.84623	6	LC
25078	ivo://nvo.caltech/voeventnet/mlsot#1110241160614114623	2012-05-03 14:51:01.711265	67.59319	16.91804	6	LC
25077	ivo://nvo.caltech/voeventnet/catot#1201181150564107728	2012-05-03 14:47:26.647255	162.44317	14.49414	3	LC
25076	ivo://nvo.caltech/voeventnet/catot#1109260040234128189	2012-05-03 14:33:39.574826	62.44411	-4.00095	6	LC
25075	ivo://nvo.caltech/voeventnet/catot#1005220040824107195	2012-05-02 12:51:32.575477	227.95224	-5.22984	4	LC

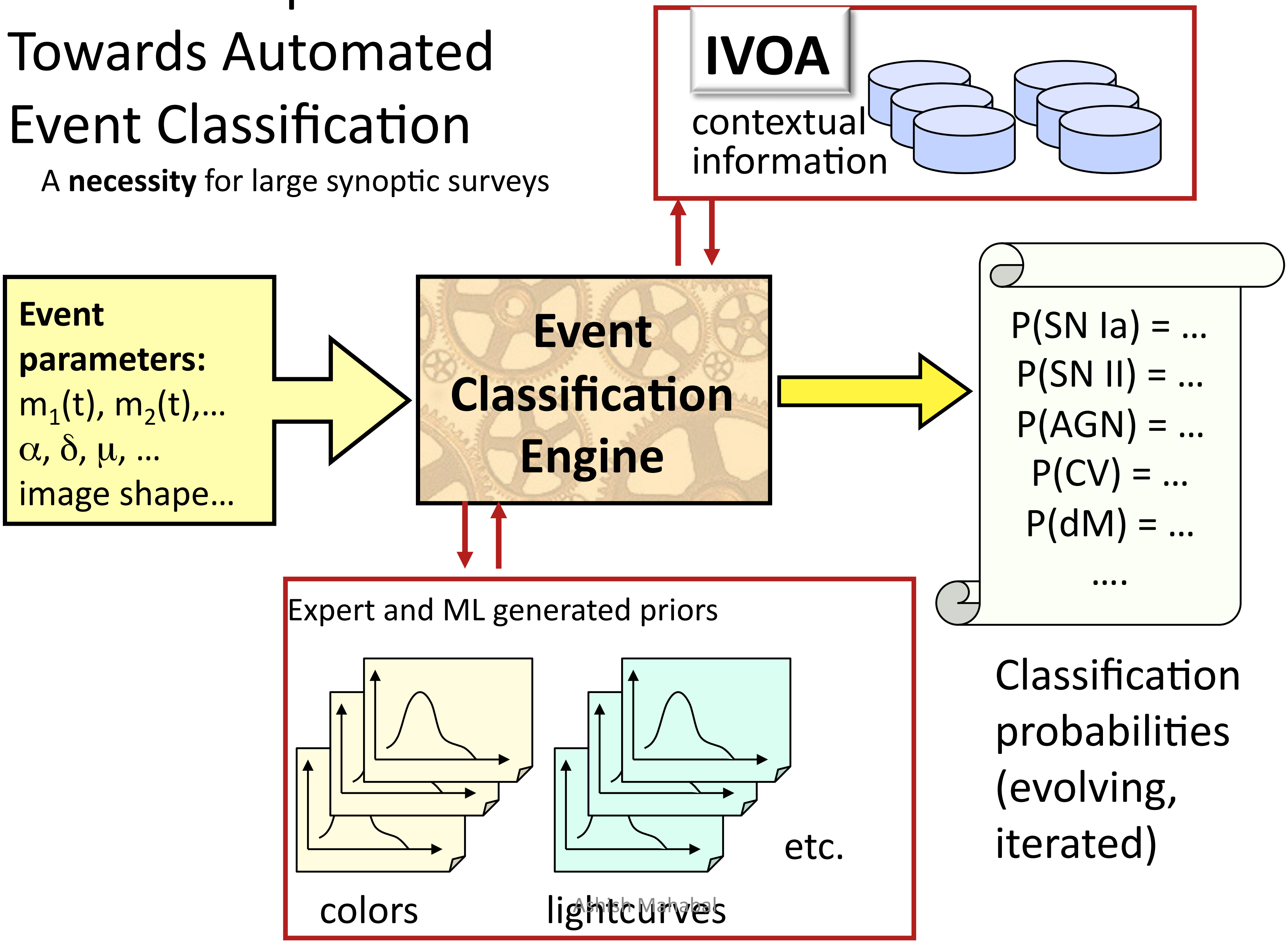


S. Barway, V. Mohan, L Wyrzykowski, S Koposov, S Hodgkin, V Mohan, S Barway, L Eyer, R Anderson, G Altavilla, G Clementini, …

http://gaia020.ast.cam.ac.uk:5000/
http://www.astro.caltech.edu/P60FollowUp/

Definite steps Towards Automated Event Classification

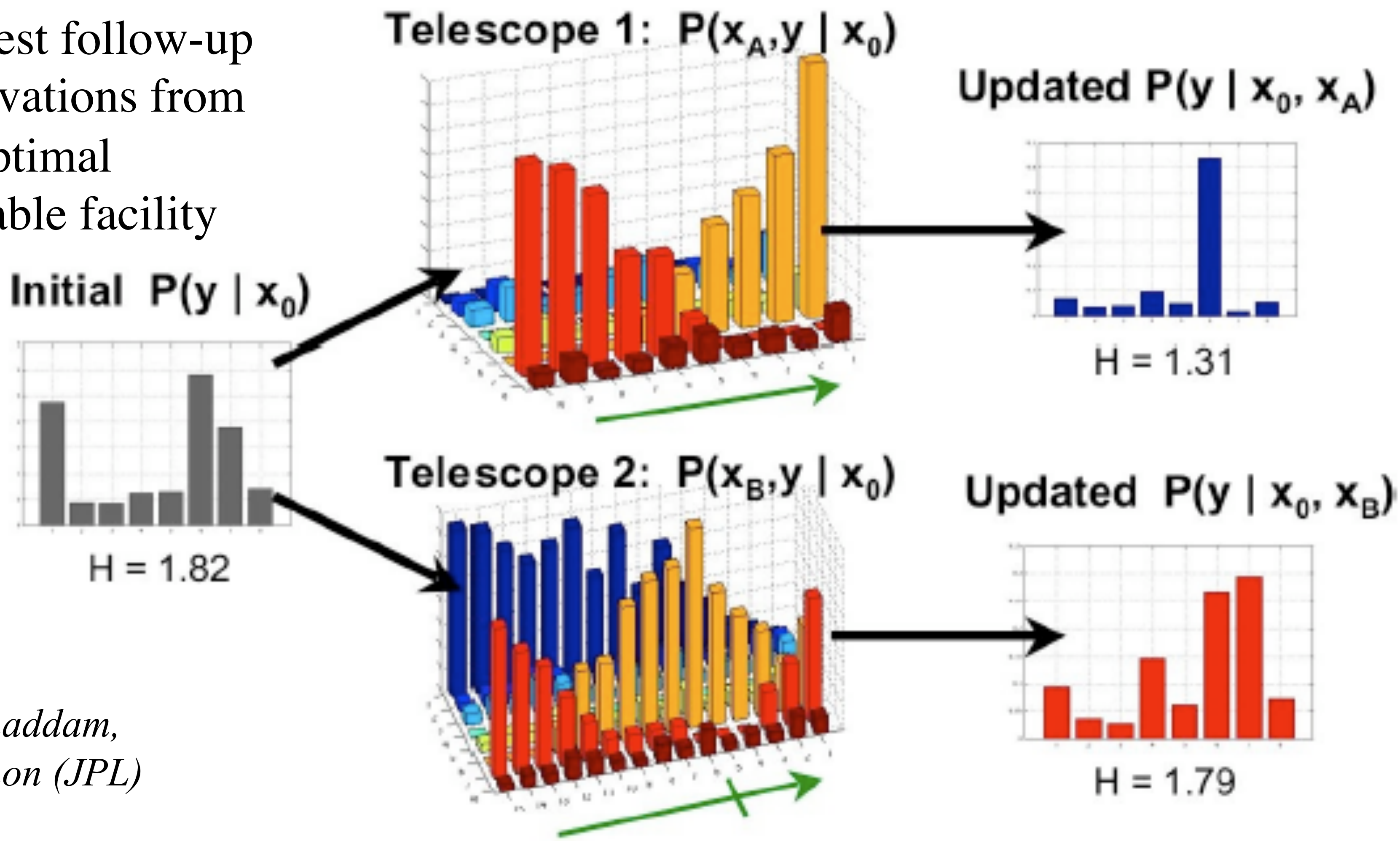
A **necessity** for large synoptic surveys



Automating the Optimal Follow-Up

What type of follow-up data has the greatest potential to discriminate among the competing models (event classes)?

Request follow-up observations from the optimal available facility



B. Moghaddam, M. Turmon (JPL)

Summary

Bayesian networks are powerful tools for classification

- Naïve networks often work best
- Learning from data most principled but also resource intensive (even prohibitive)
- Combining expert knowledge and semantics with combinations of naïve networks likely to be the best compromise

Just a few parameters enough to separate some classes. Such small networks can be combined with others for a comprehensive classification utility that can be run in real-time for surveys like CRTS and more follow-up worthy objects chosen thereof.

Concrete ideas about incorporating more semantics are welcome.