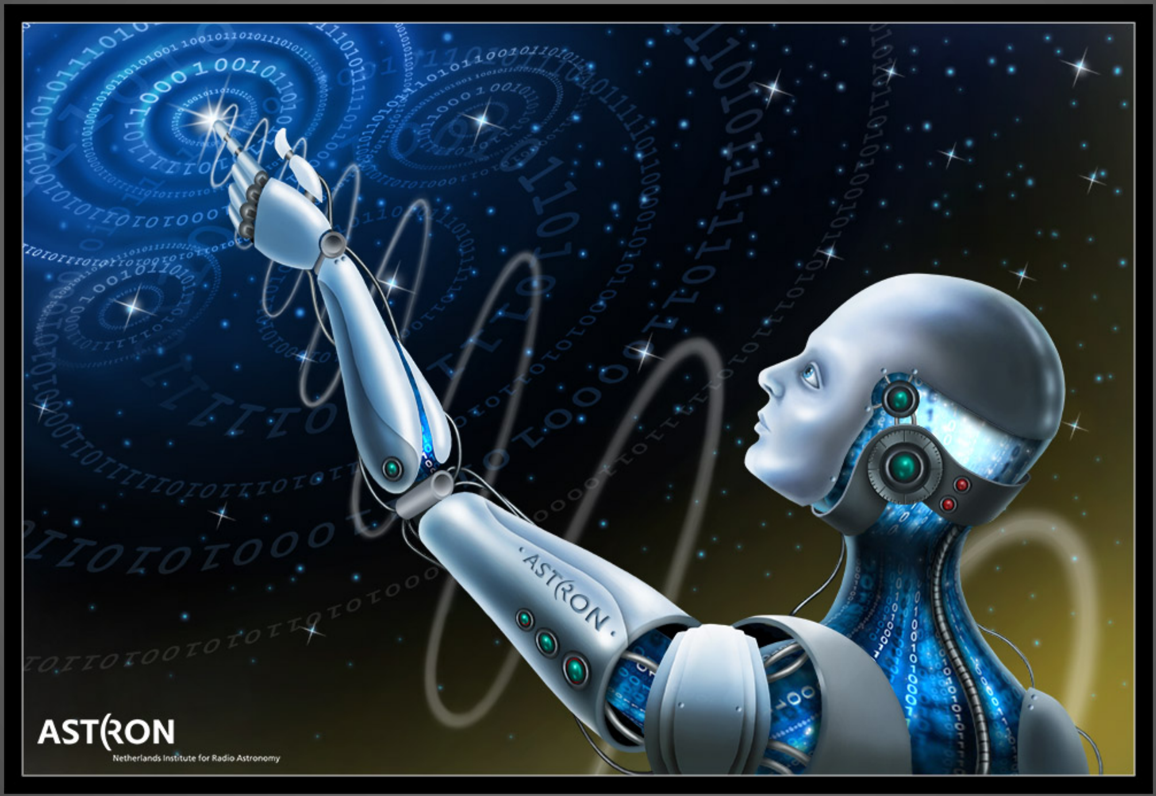
Big data: making the most of surveys with artificial neural networks

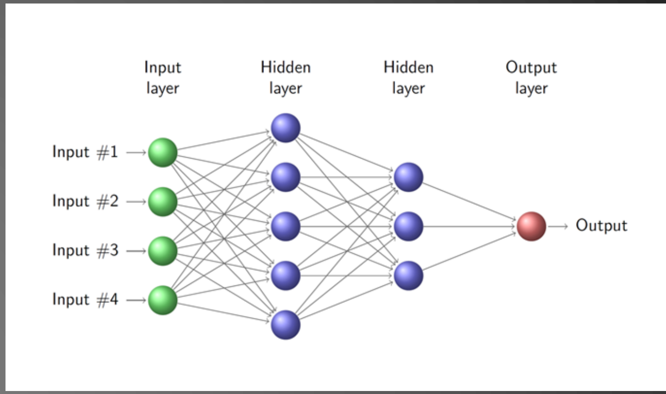
Sara Ellison, University of Victoria



Hossein Teimoorinia (Uvic), Asa Bluck (Uvic/ETH Zurich), David Rosario (MPE), Trevor Mendel (MPE)

Machine learning modes

- Classification.
  - star/galaxy separation
  - AGN classification
  - spectral typing

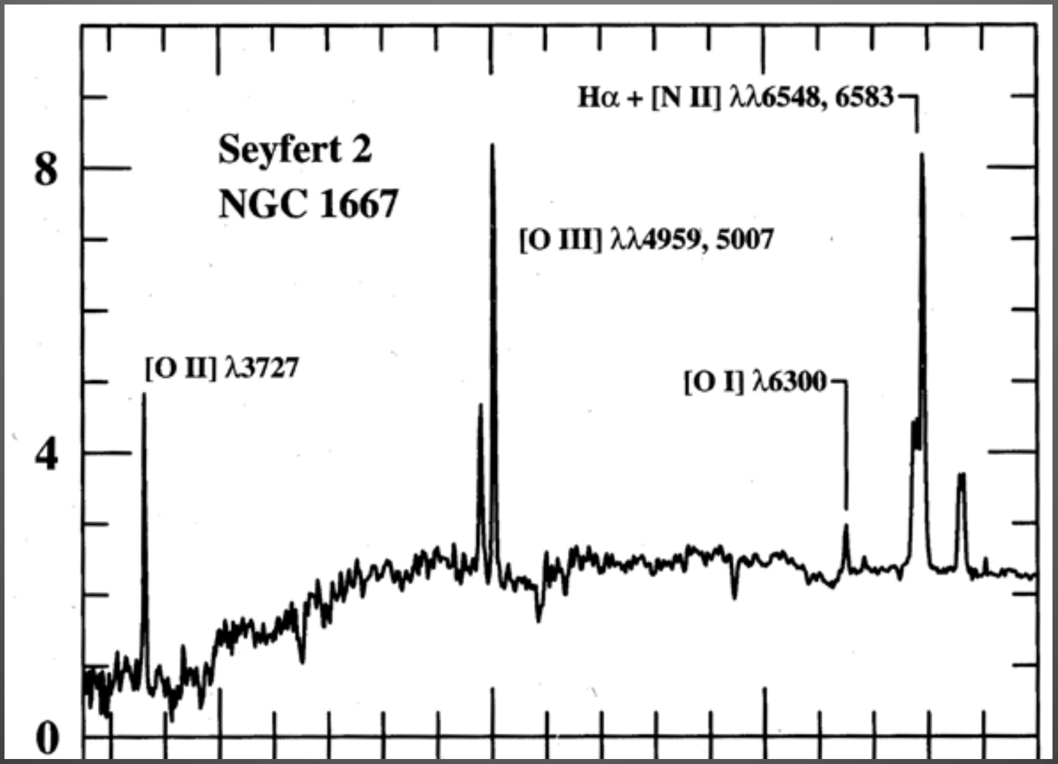


- Parameter ranking.
  - relative importance of physical parameters

- Prediction.
  - photometric redshifts
  - missing data

Predicting physical data

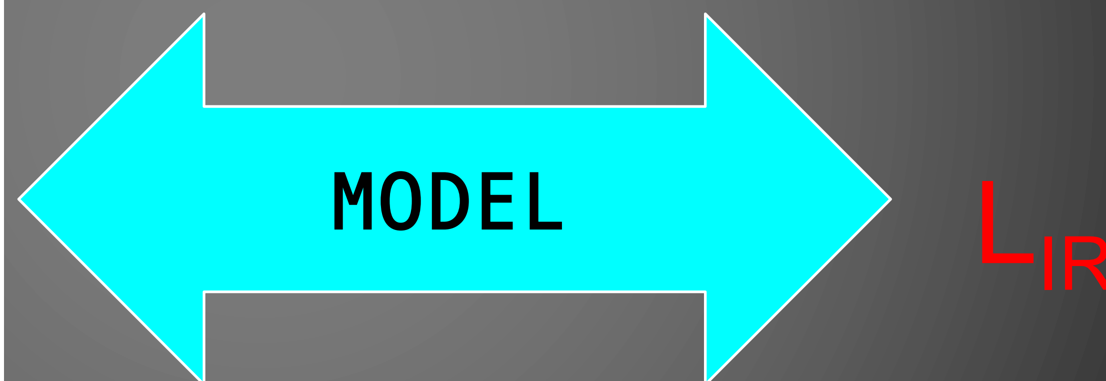
The problem: Optical emission lines are contaminated by AGN. Calibrations against D4000 have fairly large uncertainties. The FIR is great, but current surveys are either small or shallow.



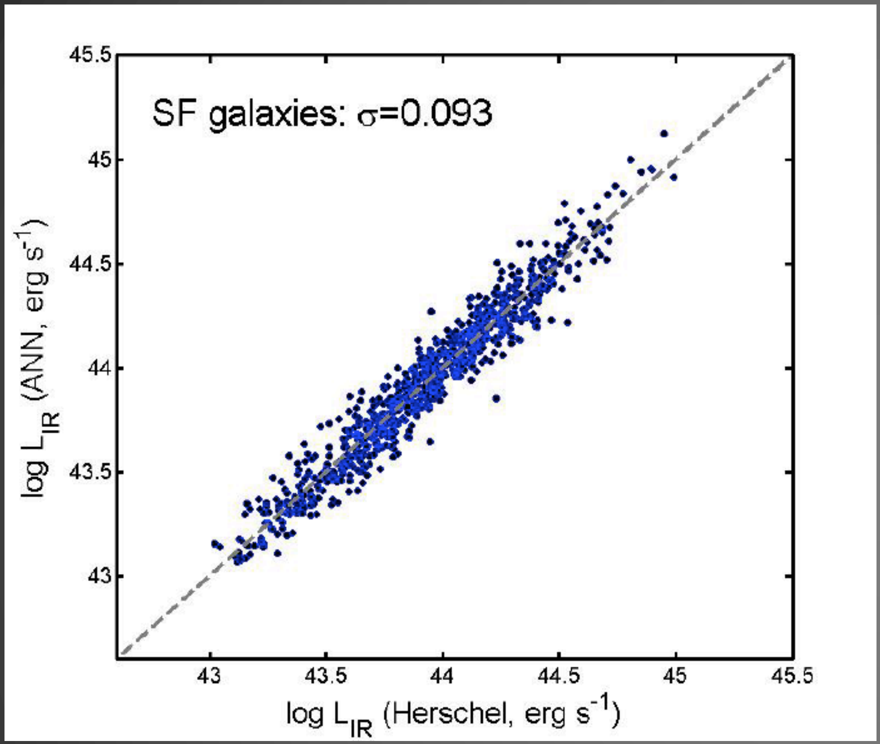
Predicting physical data

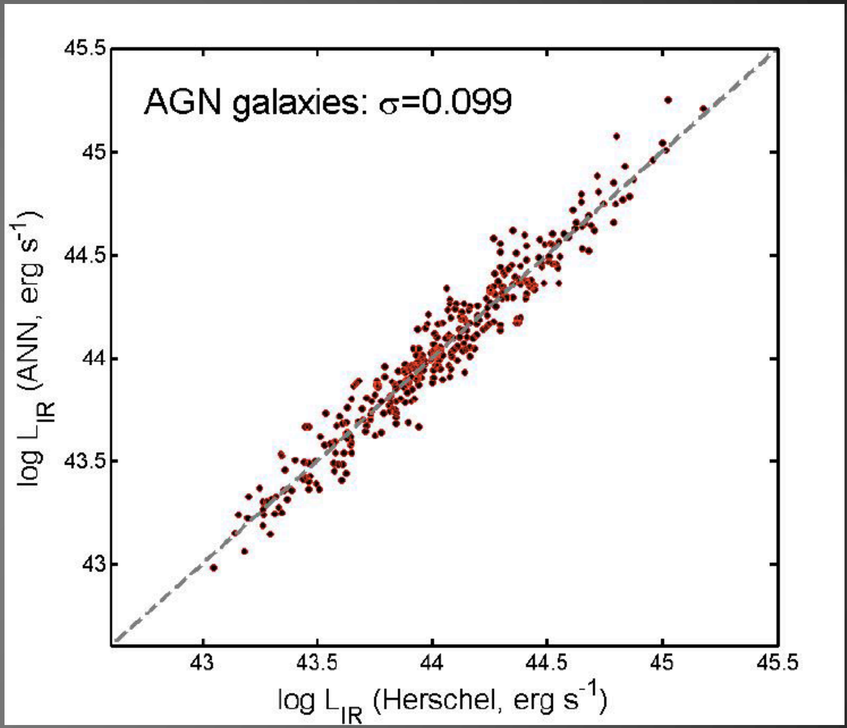
The solution: Train ANN with dataset that combines SDSS optical data with L_{IR} derived from the Herschel Stripe 82 survey (~1200 galaxies)

$M_{\star,fibre}$: Fibre stellar mass
$M_{\star}$: Total stellar mass
Corrected luminosity of [OII] λ 3727 Å emission line
Corrected luminosity of [OII] λ 3729 Å emission line
Corrected luminosity of [OIII] λ 4959 Å emission line
Corrected luminosity of [OIII] λ 5007 Å emission line
Corrected luminosity of Hα λ 6563 Å emission line
Corrected luminosity of Hβ λ 4861 Å emission line
Corrected luminosity of [NII] λ 6582 Å emission line
Corrected luminosity of [SII] λ 6717 Å emission line
Corrected luminosity of [SII] λ 6731 Å emission line
z: Redshift
D_{4000}: 4000 Å break
r-band covering fraction
M_u: absolute u-band magnitude
M_g: absolute g-band magnitude
M_r: absolute r-band magnitude
M_i: absolute i-band magnitude
M_z: absolute z-band magnitude
u-g observed colour
g-r observed colour
r-i observed colour
i-z observed colour



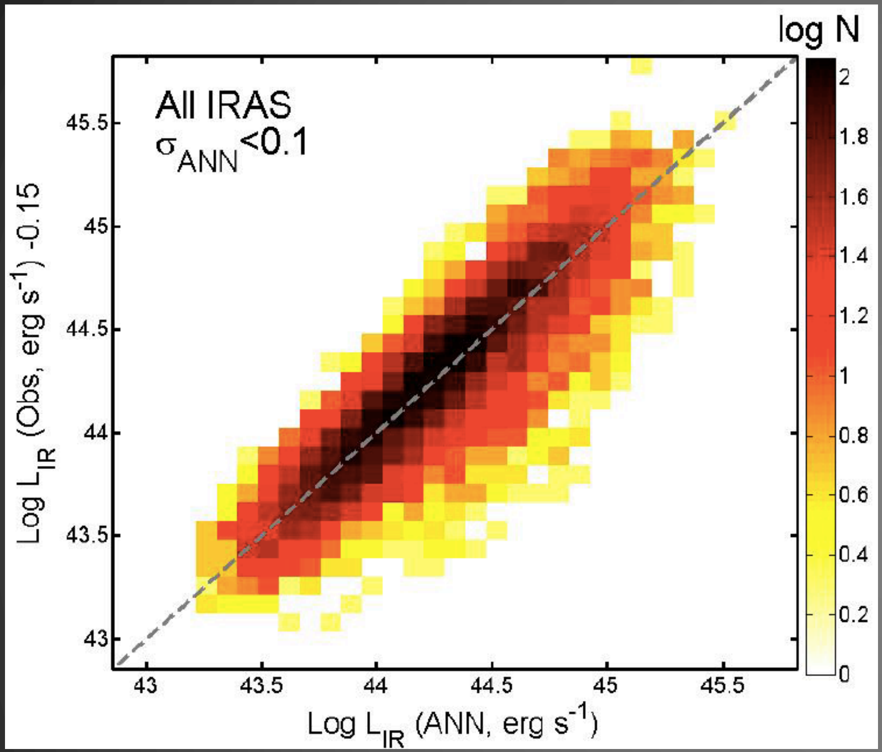
Validating the trained network

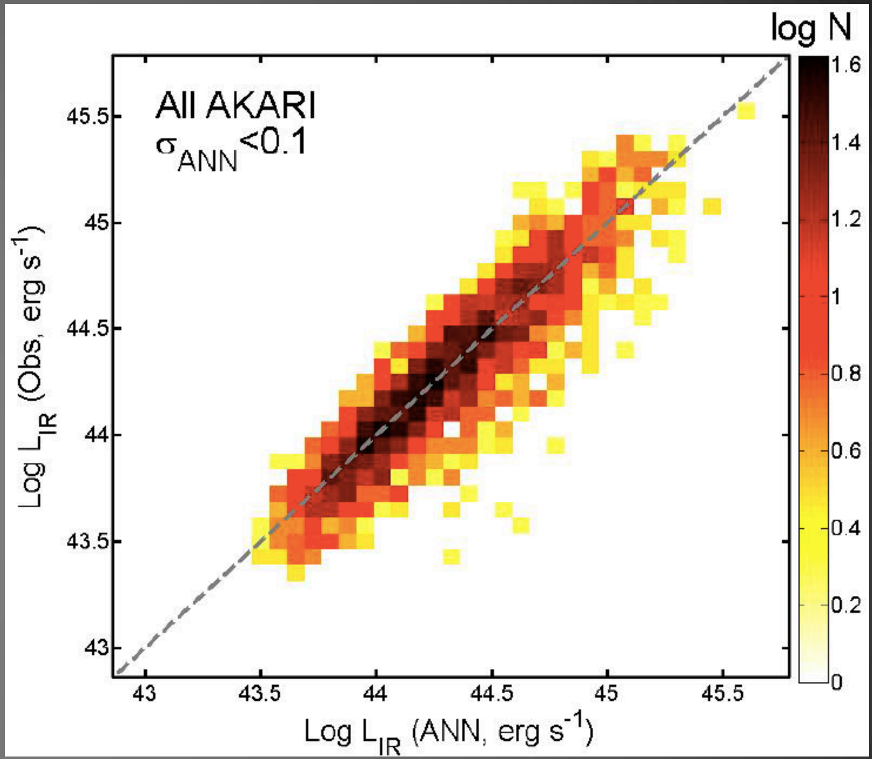




L_{IR} predicted equally well (small scatter, no systematic offset) for both star-forming and AGN dominated galaxies.

Validating the trained network

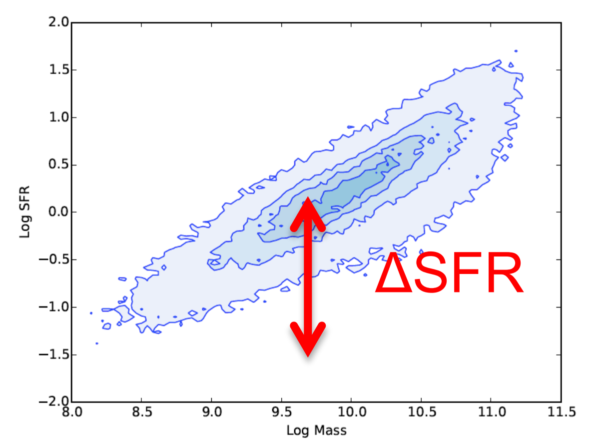


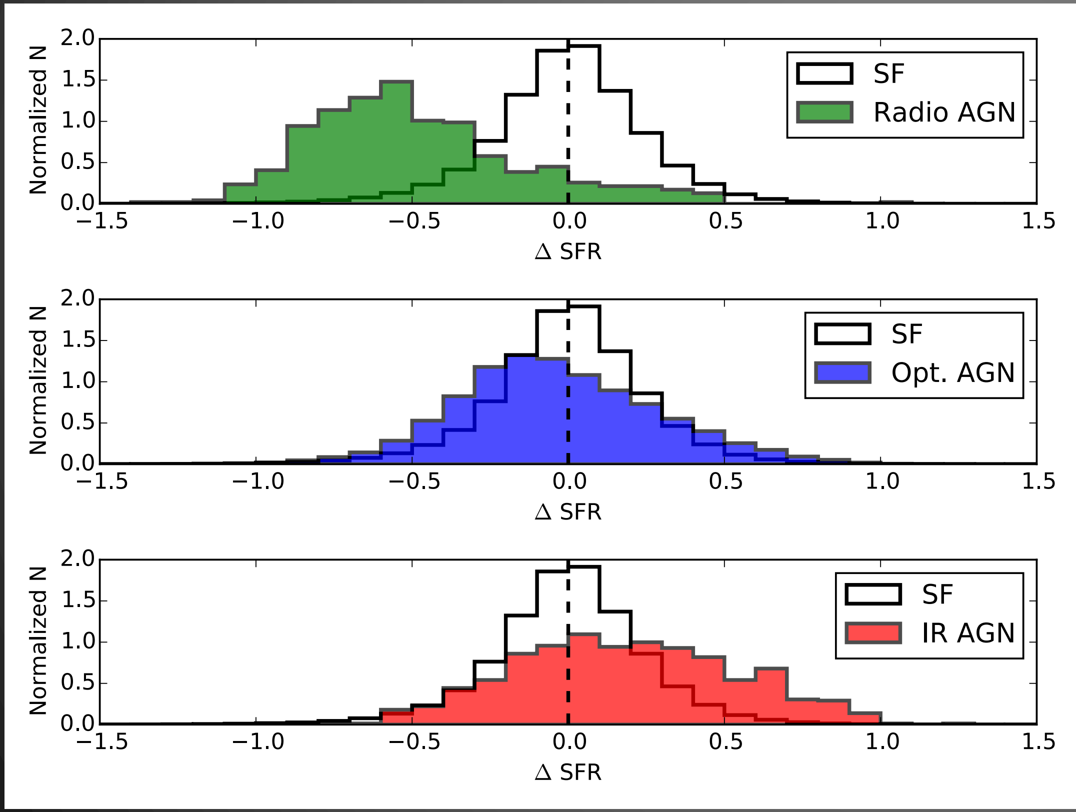


Final product: ~332,000 predicted L_{IR} for SDSS (public catalog).

Ellison, Teimoorinia, Rosario & Mendel (2015)

Application: SFRs in AGN. Are they offset from the main sequence?





Radio-selected AGN (LERGs) are strongly UNDER star forming

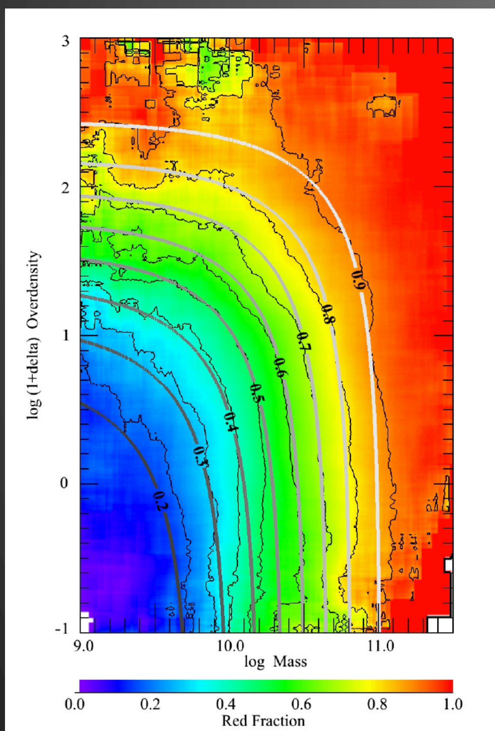
Optically-selected AGN are slightly UNDER star forming

mid-IR-selected AGN are OVER star forming

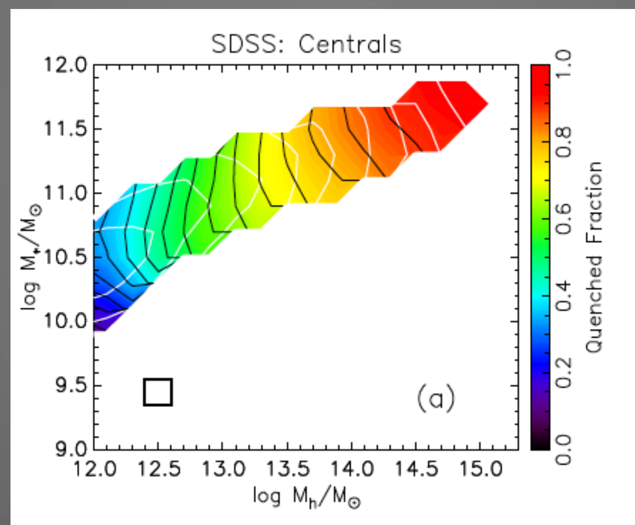
Ellison et al. (in prep)

Ranking physical data

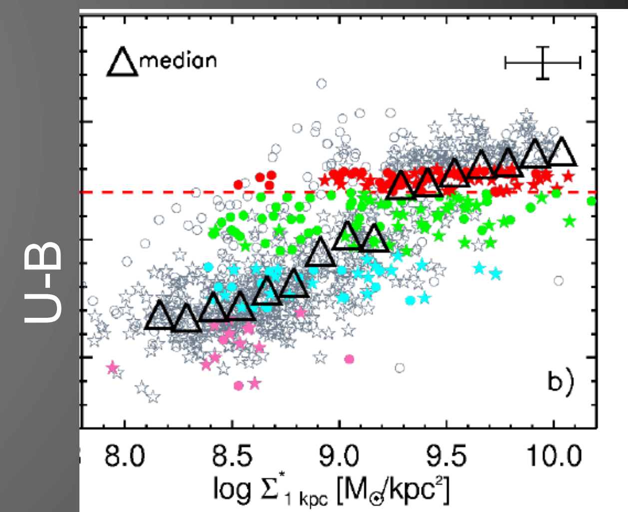
The problem: What drives galaxy quenching? Various physical drivers proposed, such as M_{halo}, M_*, local environment and central mass concentration (or M_{bulge}) . These parameters are inter-related and need to be disentangled.



Stellar mass and environment quenching (e.g. Peng et al. 2010)



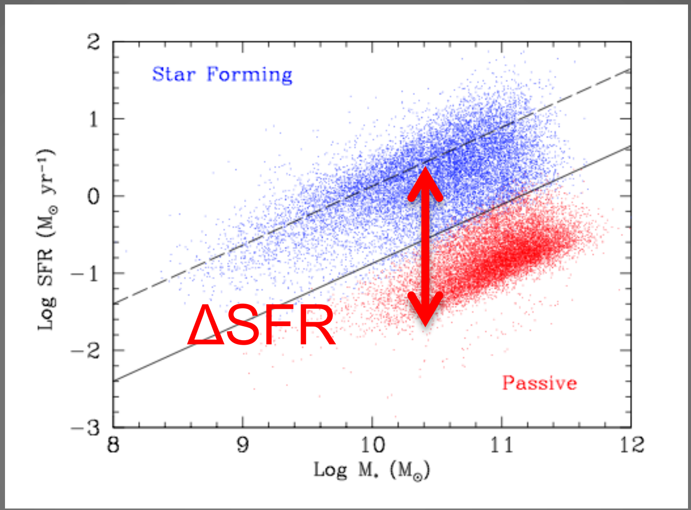
Halo mass quenching (e.g. Woo et al. 2013)



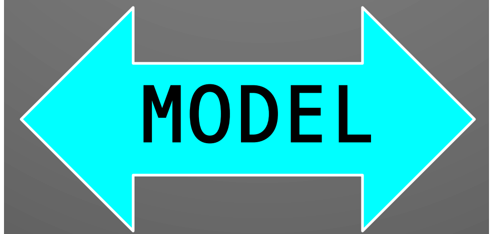
Central (bulge) mass quenching (e.g. Cheung et al. 2012; Bluck et al. 2014, Lang et al. 2014)

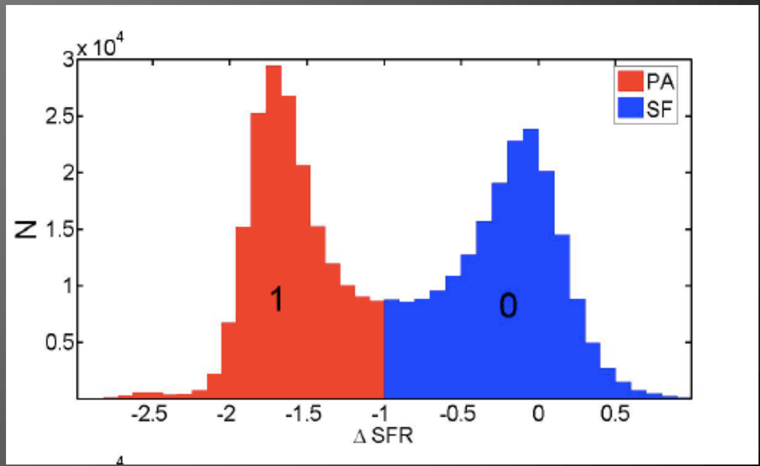
Ranking/classiying physical data

The solution: Use an ANN to rank relative importance of parameters.

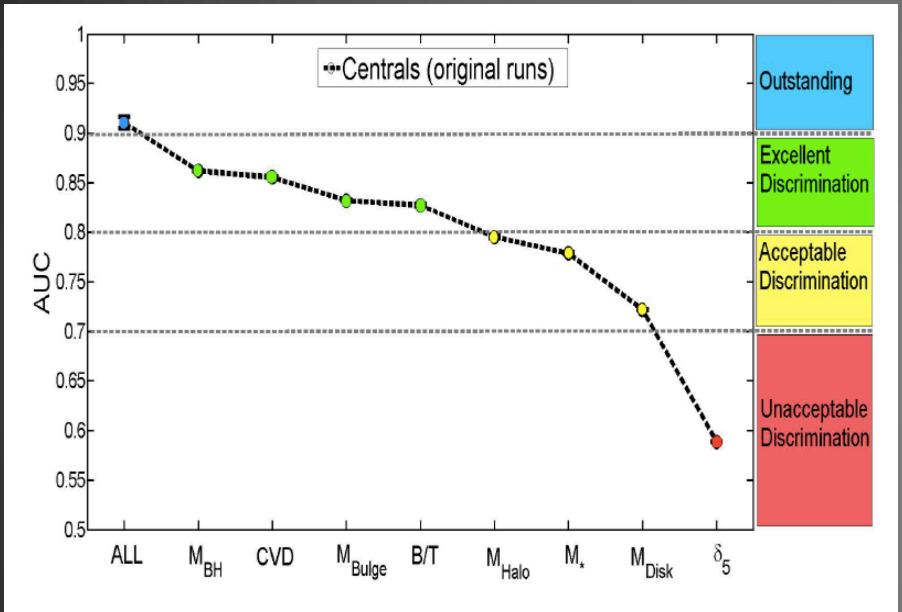


#	Symbol	Description	Scale*
1	M_{BH}	Black Hole Mass	<< 1 kpc
2	CVD	Central Velocity Dispersion	~ 1 kpc
3	M_{Bulge}	Bulge Stellar Mass	0.5 – 4 kpc
4	B/T	Bulge-to-Total Stellar Mass Ratio	0.5 – 8 kpc
5	M_*	Total Stellar Mass	2 – 8 kpc
6	M_{Disk}	Disk Stellar Mass	4 – 10 kpc
7	M_{Halo}	Group Halo Mass	0.1 - 1 Mpc
8	δ_5	Local Density Parameter	0.5 - 3 Mpc





Application: ranking parameters of galaxy quenching



The best single parameters for predicting are those connected with the galaxy centre: M_{BH}, central velocity dispersion and M_{bulge}.

Teimoorinia, Bluck & Ellison (2015, submitted)

Summary

- "Big data" is already a reality – need to exploit it.

- Artificial neural networks can be used in multiple ways, including classification, ranking and data prediction.

- Some example applications:

  - Predictions of L_{IR} for ~332,000 SDSS galaxies (Ellison et al. 2015)
  - L_{IR} catalog can be used for SFRs in AGN (Ellison et al. in prep)
  - Ranking of galaxy quenching parameters (Teimoorinia, Bluck & Ellison, submitted)
  - Prediction of galaxy line fluxes (Teimoorinia & Ellison 2014)
  - HI masses for the SDSS, trained from ALFALFA (Teimoorinia & Ellison, in prep)