A Galaxy Group Finder Using Machine Learning

Juntao Ma

Collaborators: Jie Wang, Tianxiang Mao, Hongxiang Chen

National Astronomical Observatories, Chinese Academy of Sciences

November 1, 2023

Introduction

Galaxy Group: set of galaxies that reside in the same dark matter halo.

- reveal the underlying density field
- measure halo occupation distribution, conditional luminosity function
- study how galaxies' properties depend on dark halos:
  - galactic conformity (Weinmannn et al. 2006)
  - halo and stellar mass quenching factors (Wang et al. 2018)
  - group-galaxy cross-correlation function (Yang et al. 2005d)

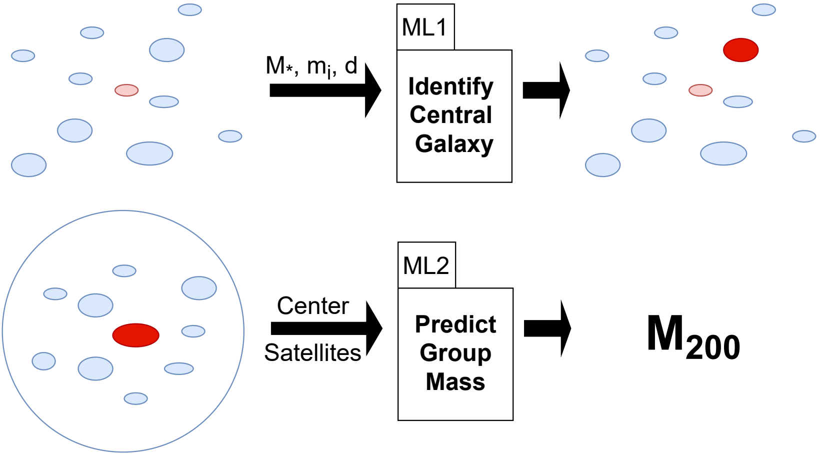
Group Finders

Classic Methods

- Friends of Friends (Huchra & Geller 1982)
- C4 algorithm (Miller et al. 2005)
- Halo-based group finder (Yang et al. 2005b)

We expect machine learning method can achieve

- easy to expand to different redshift surveys
- rely on less physical models
- quicker calculation process



Two separate networks: Central Galaxy Identifier & Group Mass Estimator

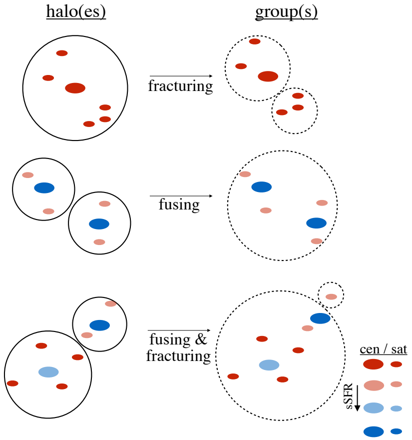


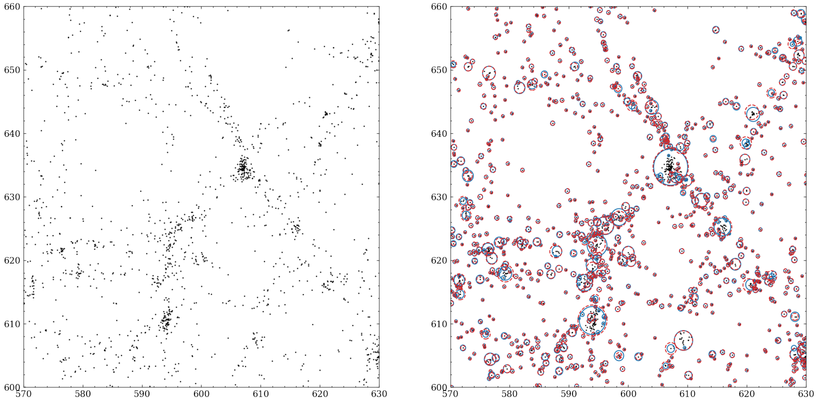
Illustration of group-finder failure modes (Campbell et al. 2015)

- fracturing on high-mass-end halos in direct result of ML models
- add iteration steps to merge sub-groups based on overlap of R_{vir} regions

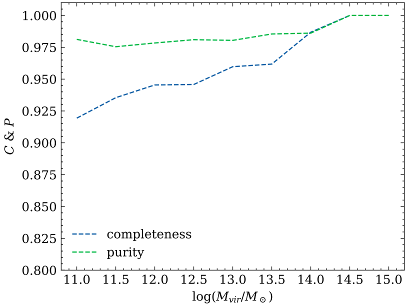
Data

Millennium Simulation: $\Omega_m = 0.25$, $\Omega_b = 0.045$, $h = 0.73$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$, boxsize $(500h^{-1}\mathrm{Mpc})^3$

- $x, y, z < 400\mathrm{Mpc}$ as training box. $x, y, z > 450\mathrm{Mpc}$ as test box.
- > 100 DM particles in the halos
- take one axis as redshift direction
- i-band apparent magnitude ≤ 17.7



Galaxies and groups in a $60 \times 60 \times 20\ \mathrm{Mpc}^3$ slice of Millennium Simulation. Red circles represent DM halos blue circles are predicted galaxy groups and their radii show $2R_{vir}$ range.

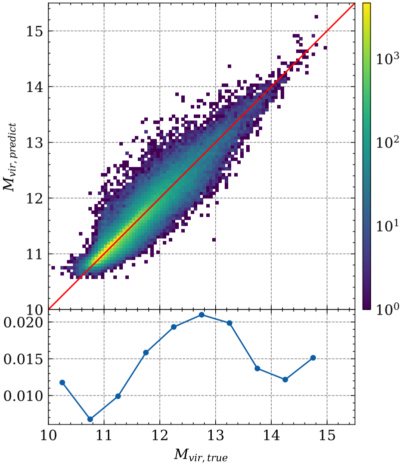


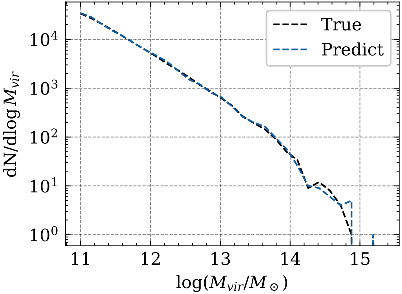
What fraction of true groups found by our model:
Completeness: TIG / TG
What fraction of predicted groups are *real*:
Purity: TIG / IG

Match Identified Group (IG) and True Group (TG)

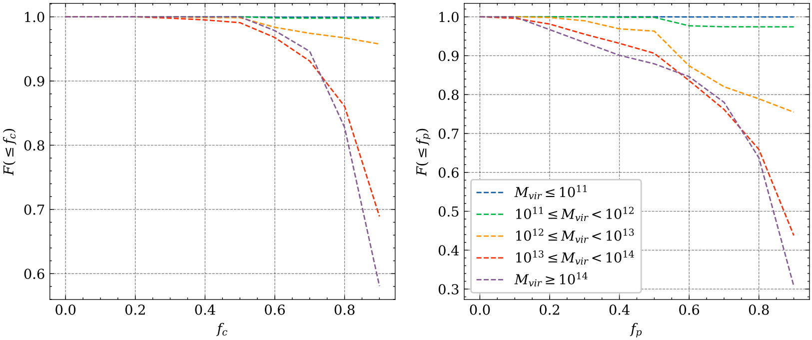
1. Central Matching: common central galaxy
2. Member Matching: $>$ 50% of an IG's members belong to a TG, and vice versa

Results
Halo Mass





Predict halo mass of a group with its central galaxy and 5 most luminous satellites.

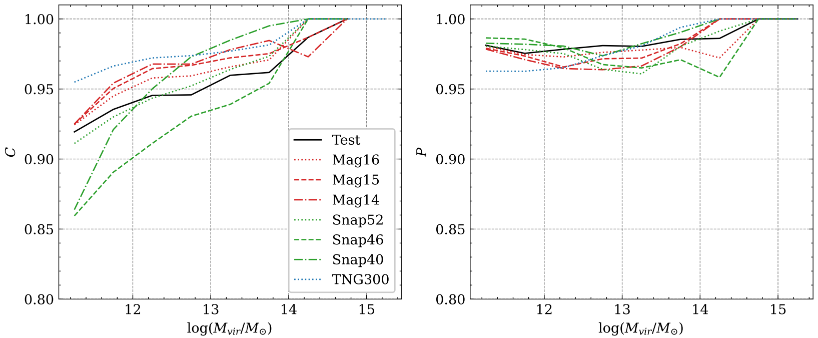


N_t: N of galaxies in TG
N_g: N of galaxies in IG
N_s: N of galaxies both in TG and IG

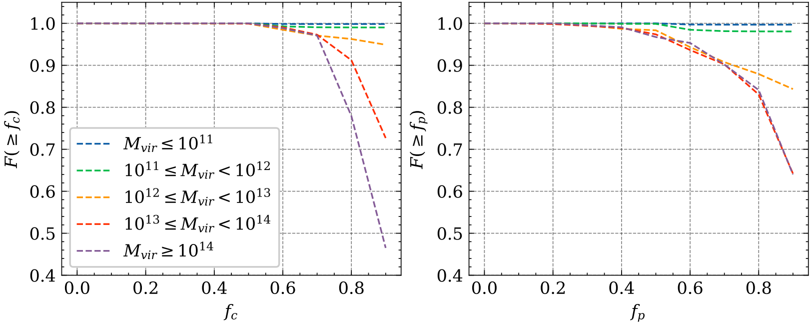
$$f_c = N_s/N_t$$
$$f_p = N_s/N_g$$

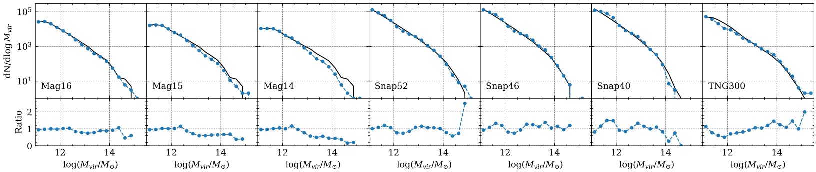
To test the extensibility of our pre-trained model, we apply the model directly on following datasets:

1. Magnitude samples. Same as test data but with different r-band magnitude limitations, $mag_r < 16$, $mag_r < 15$, and $mag_r < 14$.
2. High-redshift samples. Galaxies in high-z snapshots of MSI, z=0.23 (Snap52), z=0.62(Snap46), and z=1.08(Snap40).
3. TNG300 samples. Data from the TNG300 simulation.



Completeness and purity of galaxy groups on the extended test set. Except for the low-mass bins in Snap46 and Snap40, completeness of other samples exceeds 90%, and purity exceeds 95% for all samples.


1.0
0.9
0.8
0.7
0.6
0.5
0.4
$F(\geq f_c)$
0.0
0.2
0.4
0.6
0.8
f_c
$M_{vir} \leq 10^{11}$
$10^{11} \leq M_{vir} < 10^{12}$
$10^{12} \leq M_{vir} < 10^{13}$
$10^{13} \leq M_{vir} < 10^{14}$
$M_{vir} \geq 10^{14}$
1.0
0.9
0.8
0.7
0.6
0.5
0.4
$F(\geq f_p)$
0.0
0.2
0.4
0.6
0.8
f_p

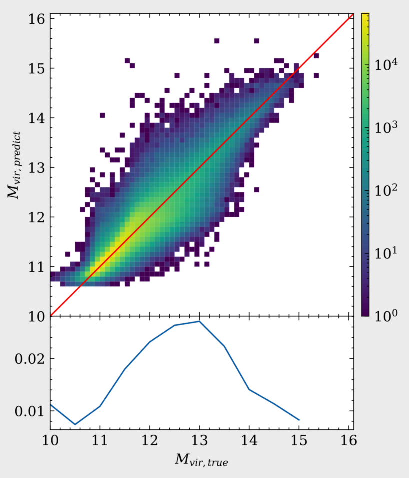


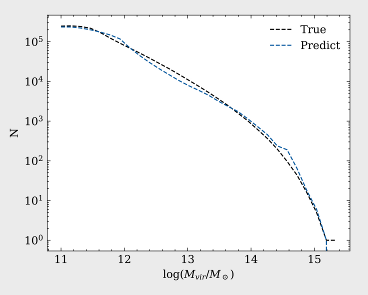
Reasons of mismatch in halo mass function:

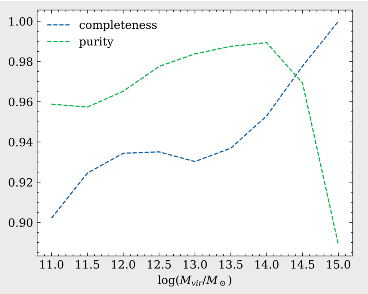
1. Magnitude samples: Reduction on richness of galaxies in group makes the model predict lower mass.
2. Hight-redshift samples: Isolate groups' luminosity - mass relation differs.
3. TNG300 samples: Different stellar mass - halo mass relation.

Conclusion

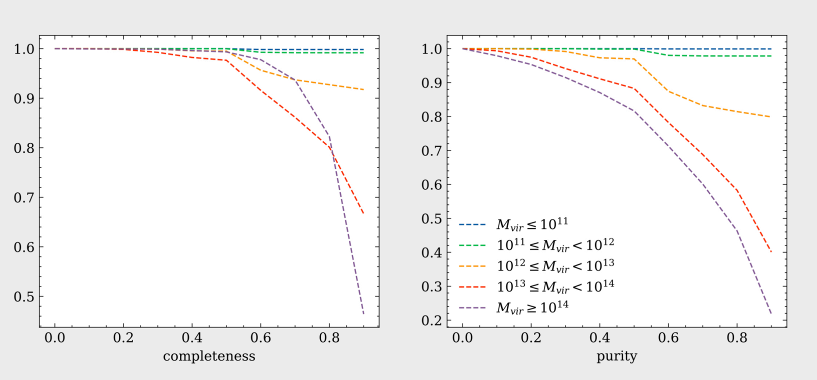
- Galaxies groups can reveal the underlying density field and galaxy-halo connection.
- We developed a new group finder using machine learning. Test result shows >92% completeness and purity down to $10^{11} M_{\odot}$ groups.
- Test on the extended datasets also show remarkable accuracy of groups' members. Mismatches in HMF are mainly due to intrinsic difference of samples.
- We are applying our group finer on SDSS and 2MASS, and their group catalogs will be published later.


16
15
14
13
12
11
10
$M_{vir,predict}$
10^4
10^3
10^2
10^1
10^0
0.02
0.01
10
11
12
13
14
15
16
$M_{vir,true}$


True
Predict
N
10^5
10^4
10^3
10^2
10^1
10^0
11
12
13
14
15
$\log(M_{vir}/M_\odot)$


completeness
purity
1.00
0.98
0.96
0.94
0.92
0.90
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
$\log(M_{vir}/M_\odot)$

Appendix



1.0
0.9
0.8
0.7
0.6
0.5
0.0
0.2
0.4
0.6
0.8
completeness
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.0
0.2
0.4
0.6
0.8
purity
$M_{vir} \leq 10^{11}$
$10^{11} \leq M_{vir} < 10^{12}$
$10^{12} \leq M_{vir} < 10^{13}$
$10^{13} \leq M_{vir} < 10^{14}$
$M_{vir} \geq 10^{14}$