How much does a galaxy know about its large-scale environment?

Biswajit Pandey

Department of Physics, Visva-Bharati
Santiniketan, India

Theoretical Physics Colloquium, TIFR, August 29, 2017

Galaxies in the Universe

- The Universe is expanding.

Galaxies in the Universe

- The Universe is expanding.
- The Universe is statistically homogeneous and isotropic on large scales.

Galaxies in the Universe

- The Universe is expanding.
- The Universe is statistically homogeneous and isotropic on large scales.
- We live in a flat ΛCDM Universe composed of ordinary matter (5%), dark matter (25%) and dark energy (70%).

Galaxies in the Universe

- The Universe is expanding.
- The Universe is statistically homogeneous and isotropic on large scales.
- We live in a flat ΛCDM Universe composed of ordinary matter (5%), dark matter (25%) and dark energy (70%).
- The present Universe is filled with galaxies.

Galaxies in the Universe

- The Universe is expanding.
- The Universe is statistically homogeneous and isotropic on large scales.
- We live in a flat ΛCDM Universe composed of ordinary matter (5%), dark matter (25%) and dark energy (70%).
- The present Universe is filled with galaxies.
- The galaxies are distributed in a complex filamentary network referred as the cosmic web.

Types of galaxies

- Galaxies are primarily classified by their shapes and come in 3 main types:

Types of galaxies

- Galaxies are primarily classified by their shapes and come in 3 main types:
- Spiral: These galaxies are disk shaped with spiral arms. They rotate and contain interstellar dust, gas and an abundance of young stars.

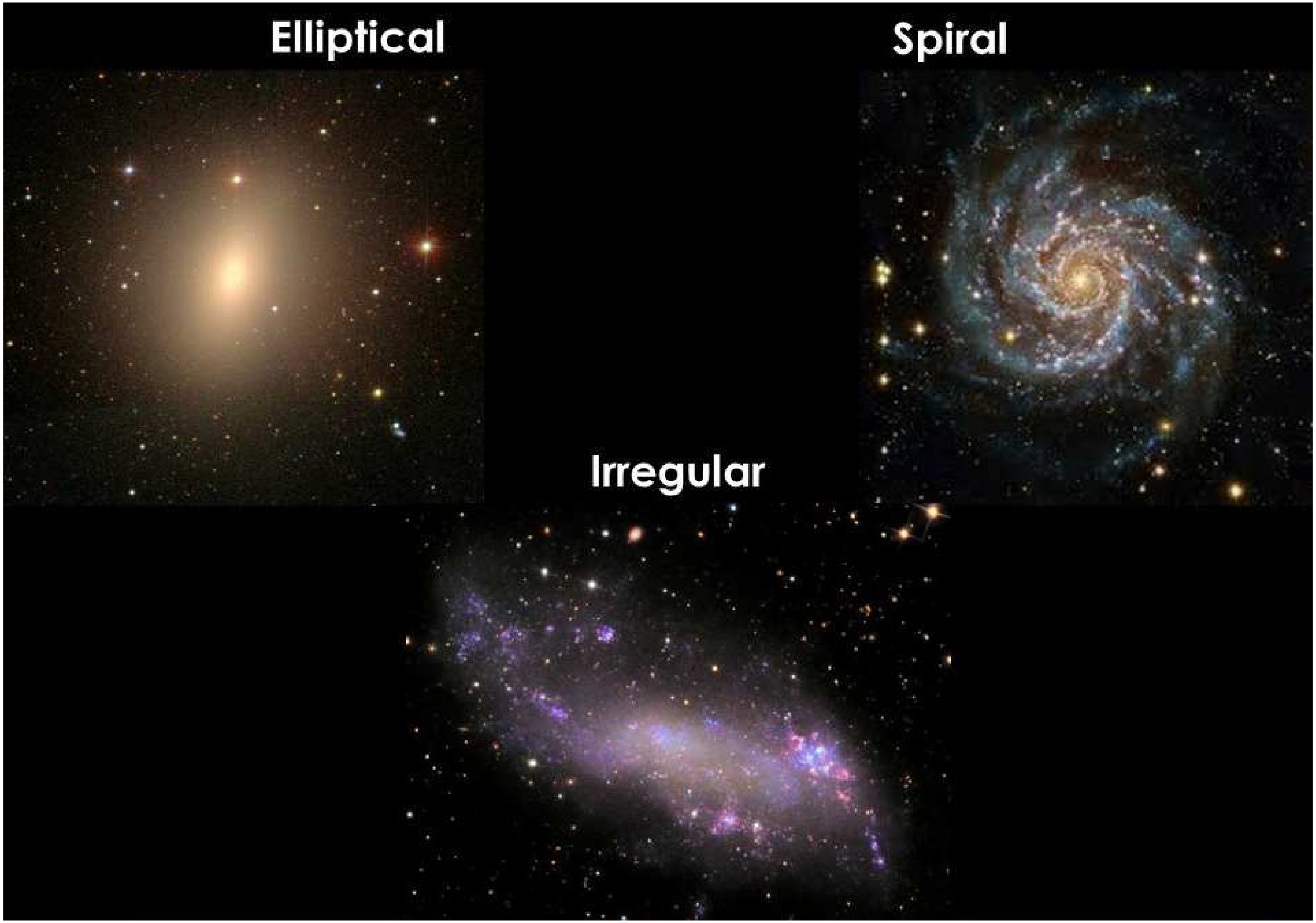
Types of galaxies

- Galaxies are primarily classified by their shapes and come in 3 main types:
- Spiral: These galaxies are disk shaped with spiral arms. They rotate and contain interstellar dust, gas and an abundance of young stars.
- Elliptical: These galaxies are ellipsoid-shaped with little to no structure, rotation and interstellar matter. This results in a minimal star formation and a dominance of the long lived, red stars.

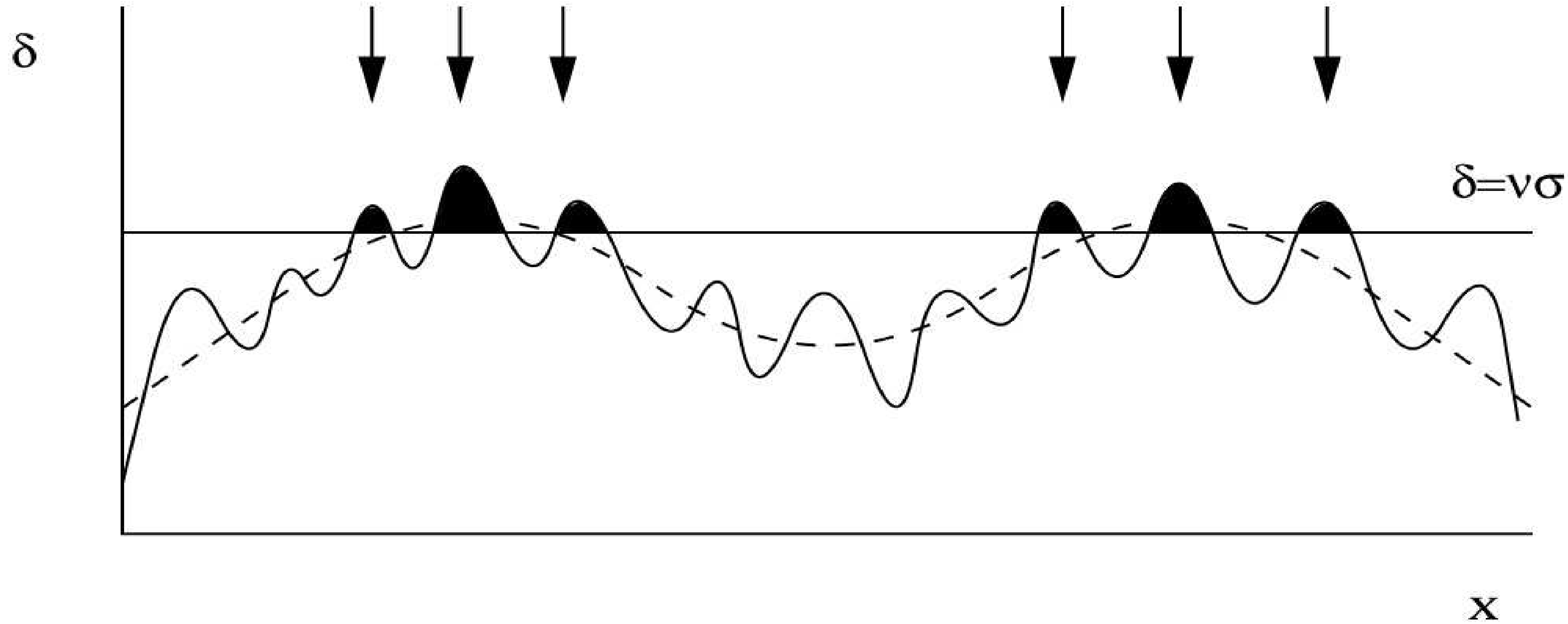
Types of galaxies

- Galaxies are primarily classified by their shapes and come in 3 main types:
- Spiral: These galaxies are disk shaped with spiral arms. They rotate and contain interstellar dust, gas and an abundance of young stars.
- Elliptical: These galaxies are ellipsoid-shaped with little to no structure, rotation and interstellar matter. This results in a minimal star formation and a dominance of the long lived, red stars.
- Irregular: These galaxies have an irregular shape and are considered to be the result of the collision of galaxies. They generally contain a complex mixture of interstellar dust, gas, young stars and old stars.

Types of galaxies

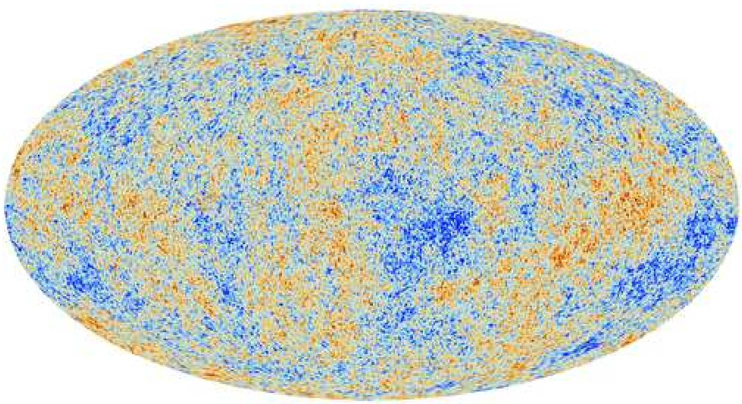


How do galaxies form?



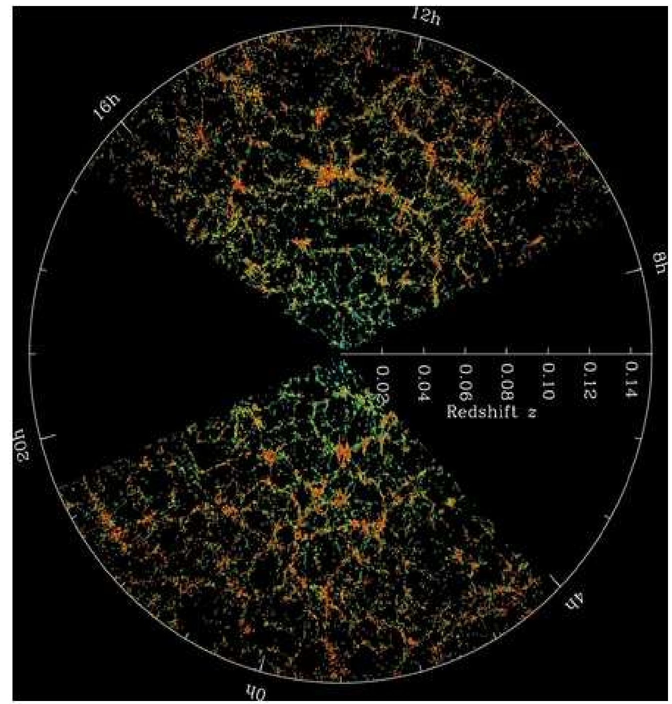
According to the spherical collapse model, the regions in the linear density field with $\delta > \delta_c$ have collapsed to produce virialized dark matter halos.

How do galaxies form?



CMBR map, PLANCK

Universe: ~ 400,000 years after the big bang



Galaxy distribution, SDSS

Universe: ~ 14 billion years after the big bang

Galaxies in the halo model

- In the current paradigm, the dark matter first collapses hierarchically into halos and the baryons settle down later at the centers of these halos to form galaxies by cooling and condensation (White & Rees 1978).

Galaxies in the halo model

- In the current paradigm, the dark matter first collapses hierarchically into halos and the baryons settle down later at the centers of these halos to form galaxies by cooling and condensation (White & Rees 1978).
- In the halo model all galaxies are believed to form and reside in virialized dark matter halos.

Galaxies in the halo model

- In the current paradigm, the dark matter first collapses hierarchically into halos and the baryons settle down later at the centers of these halos to form galaxies by cooling and condensation (White & Rees 1978).
- In the halo model all galaxies are believed to form and reside in virialized dark matter halos.
- The halo model postulates that the number and type of galaxies residing in a dark matter halo are entirely determined by its mass.

Importance of environments on small scales

- If the dark matter halos evolve in isolation, the galaxy properties are largely determined at birth by the initial conditions at the locations where they formed.

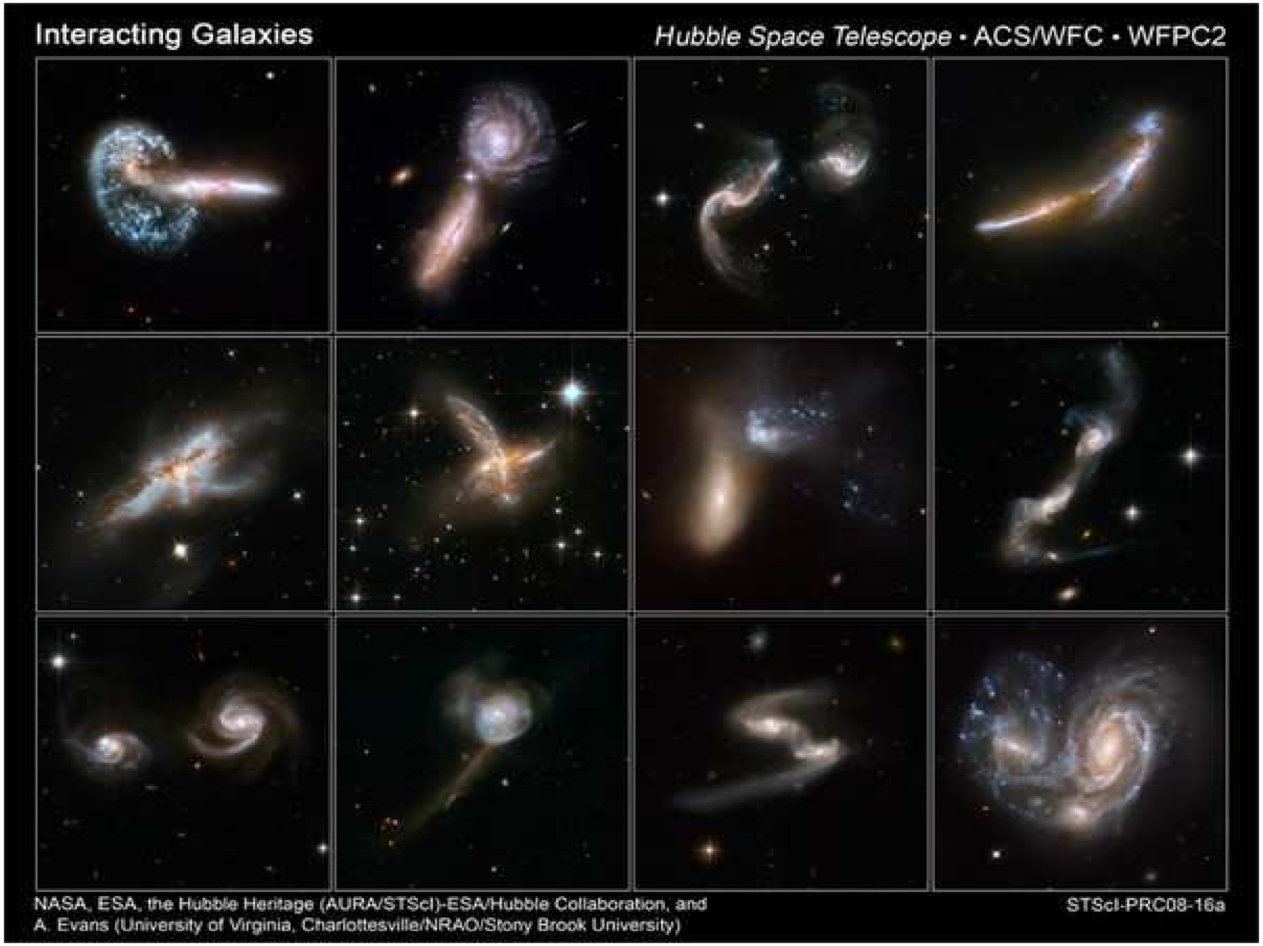
Importance of environments on small scales

- If the dark matter halos evolve in isolation, the galaxy properties are largely determined at birth by the initial conditions at the locations where they formed.
- In the hierarchical model, smaller halos merge to form bigger halos and the galaxy properties evolve according to the nature of their mergers.

Importance of environments on small scales

- If the dark matter halos evolve in isolation, the galaxy properties are largely determined at birth by the initial conditions at the locations where they formed.
- In the hierarchical model, smaller halos merge to form bigger halos and the galaxy properties evolve according to the nature of their mergers.
- The environmental effects such as ram pressure striping and different types of galaxy-galaxy interactions and mergers are thus expected to play a crucial role in the formation and evolution of galaxies.

Galaxy interactions



Local environment and Galaxy properties

- The environment of a galaxy is characterized by the local density.

Local environment and Galaxy properties

- The environment of a galaxy is characterized by the local density.
- The morphology density relation - The elliptical galaxies are known to preferentially reside in rich clusters whereas the spiral galaxies are mostly distributed in the fields (Dressler 1980).

Local environment and Galaxy properties

- The environment of a galaxy is characterized by the local density.
- The morphology density relation - The elliptical galaxies are known to preferentially reside in rich clusters whereas the spiral galaxies are mostly distributed in the fields (Dressler 1980).
- Many other galaxy properties such as luminosity, colour, star formation rate, star formation history, stellar mass, size, metallicity and AGN activity are now known to strongly depend on the environment.

What about the environments on large-scales?

- The local density describes the host halo where the galaxy is embedded.

What about the environments on large-scales?

- The local density describes the host halo where the galaxy is embedded.
- However the host halos themselves may be embedded in filaments, sheets, voids or clusters in the cosmic web and many of the properties of the host halos such as their masses, shapes and spins are determined by their large-scale cosmic environment (Hahn 2007).

What about the environments on large-scales?

- The local density describes the host halo where the galaxy is embedded.
- However the host halos themselves may be embedded in filaments, sheets, voids or clusters in the cosmic web and many of the properties of the host halos such as their masses, shapes and spins are determined by their large-scale cosmic environment (Hahn 2007).
- Do the large scale structures play any important role in the formation and evolution of galaxies?

Large-scale environments and the galaxy properties

- The star formation rates between isolated groups and the groups which are embedded in superstructures are significantly different (Scudder 2012).

Large-scale environments and the galaxy properties

- The star formation rates between isolated groups and the groups which are embedded in superstructures are significantly different (Scudder 2012).
- The presence of filamentary environment elevates the fraction of star forming galaxies (Darvish 2014).

Large-scale environments and the galaxy properties

- The star formation rates between isolated groups and the groups which are embedded in superstructures are significantly different (Scudder 2012).
- The presence of filamentary environment elevates the fraction of star forming galaxies (Darvish 2014).
- $\sim 75\%$ of the extremely metal poor galaxies reside in sheets and voids (Filho 2015).

Large-scale environments and the galaxy properties

- The star formation rates between isolated groups and the groups which are embedded in superstructures are significantly different (Scudder 2012).
- The presence of filamentary environment elevates the fraction of star forming galaxies (Darvish 2014).
- $\sim 75\%$ of the extremely metal poor galaxies reside in sheets and voids (Filho 2015).
- Some studies do not find any evidence for the dependence of galaxy properties on their large-scale environment (Park & Choi 2009, Yan, Fan & White 2013).

SDSS and the Galaxy Zoo:

- The Sloan Digital Sky Survey (SDSS) is the largest galaxy redshift survey todate which has provided the spectroscopic and photometric informations for millions of galaxies and quasars.

SDSS and the Galaxy Zoo:

- The Sloan Digital Sky Survey (SDSS) is the largest galaxy redshift survey todate which has provided the spectroscopic and photometric informations for millions of galaxies and quasars.
- Galaxy Zoo is a citizen science project which invites people to assist in the morphological classification of large number of galaxies observed in the SDSS.

SDSS and the Galaxy Zoo:

- The Sloan Digital Sky Survey (SDSS) is the largest galaxy redshift survey todate which has provided the spectroscopic and photometric informations for millions of galaxies and quasars.
- Galaxy Zoo is a citizen science project which invites people to assist in the morphological classification of large number of galaxies observed in the SDSS.
- More than $200,000$ volunteers in the internet participated in this project and classified ~ 1 million galaxies according to their morphology by visual inspection.

SDSS and the Galaxy Zoo:

- The Sloan Digital Sky Survey (SDSS) is the largest galaxy redshift survey todate which has provided the spectroscopic and photometric informations for millions of galaxies and quasars.
- Galaxy Zoo is a citizen science project which invites people to assist in the morphological classification of large number of galaxies observed in the SDSS.
- More than $200,000$ volunteers in the internet participated in this project and classified ~ 1 million galaxies according to their morphology by visual inspection.
- The morphological classification by direct visual inspection avoids many of the potential biases associated with the proxies for morphology.

The Data:

- Sky Covergae: $135^\circ \leq \alpha \leq 225^\circ$ and $0^\circ \leq \delta \leq 60^\circ$

The Data:

- Sky Covergae: $135^\circ \leq \alpha \leq 225^\circ$ and $0^\circ \leq \delta \leq 60^\circ$
- Primary redshift and magnitude cut: $0 \leq z \leq 0.25$, $m_r < 17.77$

The Data:

- Sky Covergae: $135^\circ \leq \alpha \leq 225^\circ$ and $0^\circ \leq \delta \leq 60^\circ$
- Primary redshift and magnitude cut: $0 \leq z \leq 0.25$, $m_r < 17.77$
- These cuts yield a sample with 343340 galaxies of which 134745 are classified.

The Data:

- Sky Covergae: $135^\circ \le \alpha \le 225^\circ$ and $0^\circ \le \delta \le 60^\circ$
- Primary redshift and magnitude cut: $0 \le z \le 0.25$, $m_r < 17.77$
- These cuts yield a sample with 343340 galaxies of which 134745 are classified.
- We only consider classified galaxies with a debiased vote fraction > 0.8.

The Data:

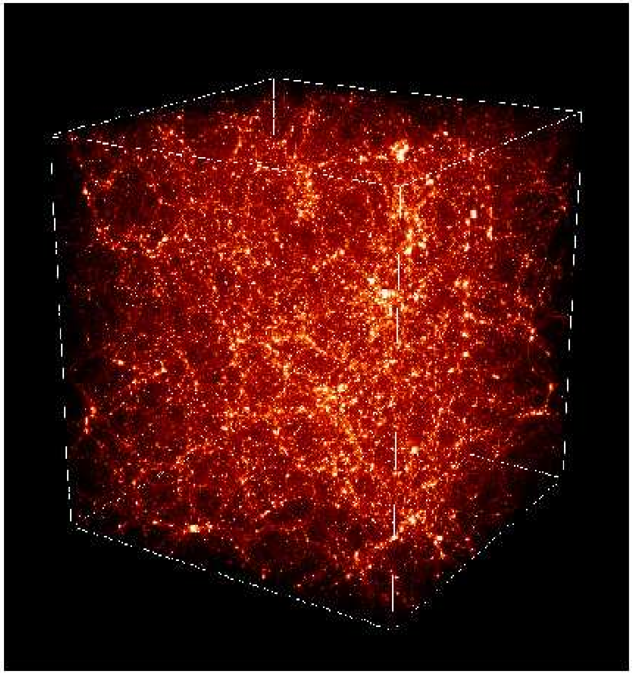
- Sky Covergae: $135^\circ \leq \alpha \leq 225^\circ$ and $0^\circ \leq \delta \leq 60^\circ$
- Primary redshift and magnitude cut: $0 \leq z \leq 0.25$, $m_r < 17.77$
- These cuts yield a sample with 343340 galaxies of which 134745 are classified.
- We only consider classified galaxies with a debiased vote fraction > 0.8.
- We construct a volume limited sample by restricting the extinction corrected and k-corrected r-band absolute magnitude to $M_r < -20$.The resulting volume limited sample extends upto a redshift $z = 0.1067$ and consists of 42334 galaxies with visual morphological classification.

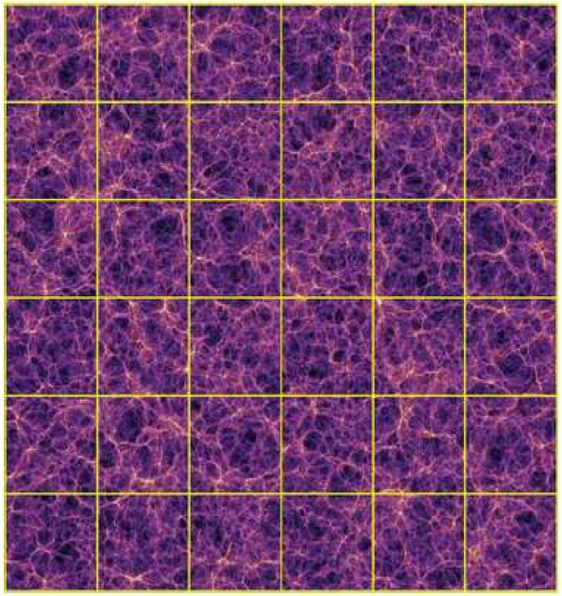
The Method of Analysis

We consider only the galaxies which are visually classified as spiral or elliptical and discarded all the galaxies with uncertain morphology.

We only consider the galaxies with a debiased vote fraction > 0.8

We finally have a cubic region of side 165 Mpc/h which consists of 15860 galaxies out of which 11875 are spirals and 3985 are ellipticals.





We divide the cube into a number of regular cubic voxels and count the number of spiral and elliptical galaxies residing inside each voxel.

We vary the grid size and store the number counts in all voxels for each grid sizes.

Information entropy

- The information contained in any outcome of a probabilistic process is given by $I(p(x_i)) = \log \frac{1}{p(x_i)}$.

Information entropy

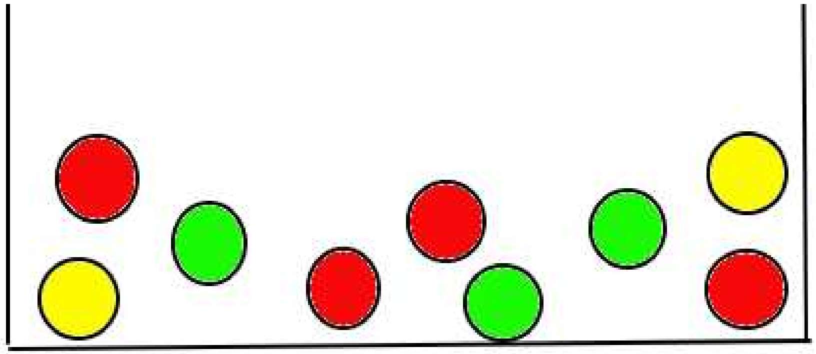
- The information contained in any outcome of a probabilistic process is given by $I(p(x_i)) = \log \frac{1}{p(x_i)}$.
- If there are n outcomes, $\{x_i : i = 1, ...n\}$ then in N trials $N\, p(x_i)$ occurrences are expected for the outcome x_i.

Information entropy

- The information contained in any outcome of a probabilistic process is given by $I(p(x_i)) = \log \frac{1}{p(x_i)}$.
- If there are n outcomes, $\{x_i : i = 1, ...n\}$ then in N trials $N\,p(x_i)$ occurrences are expected for the outcome x_i.
- Information entropy (Shannon 1948) is the average information required to describe the random variable x and defined as,

$$H(x) = \frac{1}{N}\sum_{i=1}^{n} N\,p(x_i)\,\log\frac{1}{p(x_i)} = -\sum_{i=1}^{n} p(x_i)\,\log p(x_i)$$

A simple example



- Pick a ball randomly from the bucket.
- The ball can be either red, green or yellow, total $n = 3$ outcomes.
- Entropy$=(-\frac{4}{9}\log\frac{4}{9})+(-\frac{3}{9}\log\frac{3}{9})+(-\frac{2}{9}\log\frac{2}{9}) = 1.5304755$ implies that you are expected to get 1.5304755 bits information each time you choose a ball from the bucket.
- Minimum entrpy $= 0$ occurs when one of the probabilities is 1 and rest are 0.
- Maximum entropy $= \log(n)$ occurs when all the probabilties have equal values of $\frac{1}{n}$.

Mutual Information

- Mutual information is a measure of statistical dependence between two random variables.

Mutual Information

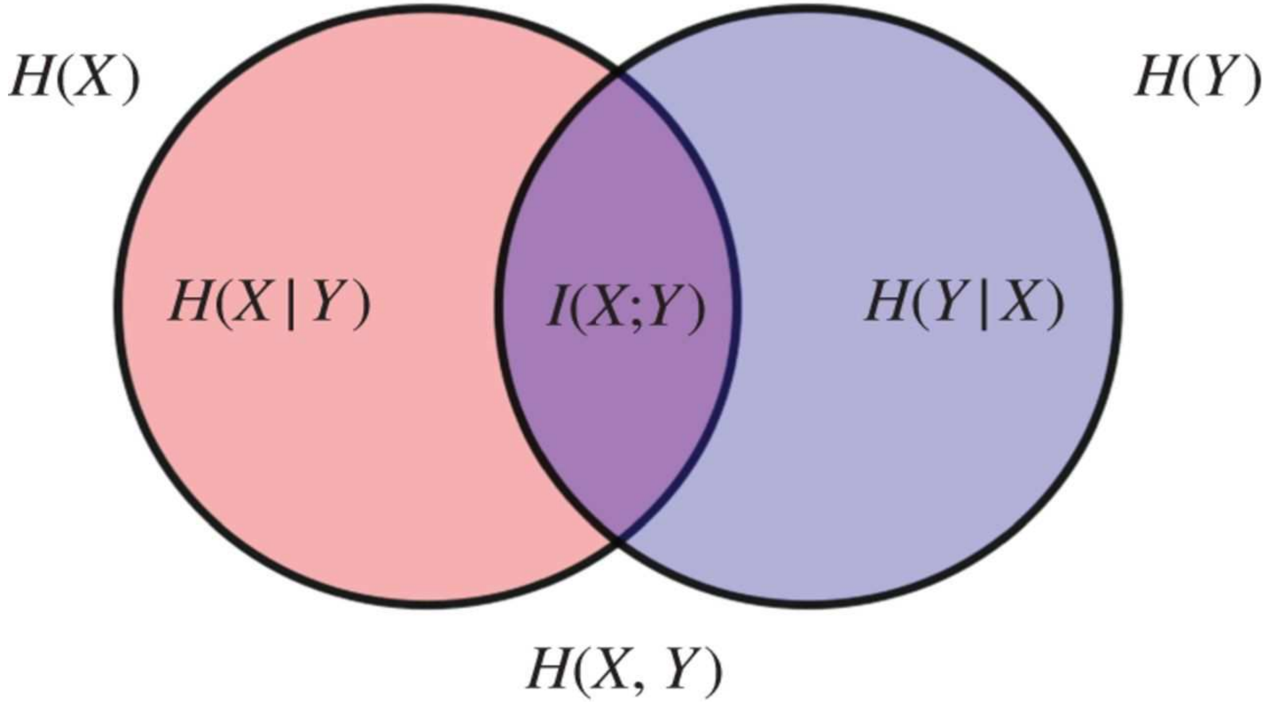
- Mutual information is a measure of statistical dependence between two random variables.
- Mutual information makes no assumptions on the distribution of the variables, or the nature of the relationship between them.

Mutual Information

- Mutual information is a measure of statistical dependence between two random variables.
- Mutual information makes no assumptions on the distribution of the variables, or the nature of the relationship between them.
- The mutual information between two random variables is defined as,

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(X) - H(X|Y) \\
&= H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

Mutual Information



$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(X) - H(X|Y) \\
&= H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

Mutual Information: An example

- Consider a roll of a fair six-sided die.

Mutual Information: An example

- Consider a roll of a fair six-sided die.
- The outcome of any particular roll follows a uniform distribution over the six possible values, with entropy $\log_2 6$.

Mutual Information: An example

- Consider a roll of a fair six-sided die.
- The outcome of any particular roll follows a uniform distribution over the six possible values, with entropy $\log_2 6$.
- If an observer is told that the result of a particular roll is an even number, there are now three possible values, all equally likely, and the entropy of this distribution is $\log_2 3$.

Mutual Information: An example

- Consider a roll of a fair six-sided die.
- The outcome of any particular roll follows a uniform distribution over the six possible values, with entropy $\log_2 6$.
- If an observer is told that the result of a particular roll is an even number, there are now three possible values, all equally likely, and the entropy of this distribution is $\log_2 3$.
- The difference between these entropies is 1 bit:
  $\log_2 6 - \log_2 3 = \log_2 \frac{6}{3} = \log_2 2 = 1$

Mutual Information: An example

- Consider a roll of a fair six-sided die.
- The outcome of any particular roll follows a uniform distribution over the six possible values, with entropy $\log_2 6$.
- If an observer is told that the result of a particular roll is an even number, there are now three possible values, all equally likely, and the entropy of this distribution is $\log_2 3$.
- The difference between these entropies is 1 bit:
  $\log_2 6 - \log_2 3 = \log_2 \frac{6}{3} = \log_2 2 = 1$
- Thus, 1 bit quantifies the amount of information conveyed by the knowledge "this roll is even" and corresponds to a halving of the uncertainty about the outcome.

Mutual Information: environment and morphology

- The entire sample is divided into a number of regular cubic voxels of grid size $d\,h^{-1}\,\mathrm{Mpc}$. Let N_d be the number of resulting voxels for a grid size of d .

Mutual Information: environment and morphology

- The entire sample is divided into a number of regular cubic voxels of grid size $d\,h^{-1}\,\mathrm{Mpc}$. Let N_d be the number of resulting voxels for a grid size of d .
- We define two discrete random variables X and Y with probability distributions $P(X)$ and $P(Y)$ respectively. $P(X_i) = \frac{n_i}{N}$ is the probability that a randomly drawn galaxy resides in the i^{th} voxel where n_i is the number of galaxies in the i^{th} voxel and N is the total number of galaxies. $P(X)$ has a total N_d outcomes. $P(Y_j)$ is the probability that a randomly chosen galaxy is spiral or elliptical and it has only 2 outcomes given by $P(Y_1) = \frac{N_s}{N}$ for spiral and $P(Y_2) = \frac{N_e}{N}$ for elliptical. Here N_s and N_e are the total number of spiral and elliptical galaxies respectively.

Mutual Information: environment and morphology

- For the pair of random variables (X, Y) with joint probability distribution $P(X, Y)$ the joint entropy $H(X, Y)$ is defined as,

$$H(X,Y) \; = \; -\sum_{i=1}^{N_d}\sum_{j=1}^{2} P(X_i, Y_j) \log P(X_i, Y_j)$$

Mutual Information: environment and morphology

- For the pair of random variables (X, Y) with joint probability distribution $P(X, Y)$ the joint entropy $H(X, Y)$ is defined as,

$$H(X,Y) \;=\; -\sum_{i=1}^{N_d}\sum_{j=1}^{2} P(X_i,Y_j)\,\log\,P(X_i,Y_j)$$

- The joint probability that a randomly selected galaxy resides in the i^{th} voxel and is spiral or elliptical is given by $P(X, Y) = P(Y|X)P(X)$ where $P(Y|X)$ is the conditional probability that the randomly selected galaxy is spiral or elliptical given that it resides in the i^{th} voxel. This gives $P(X_i, Y_j) = \frac{(n_s)_i}{N}$ for $j = 1$ (spiral) and $P(X_i, Y_j) = \frac{(n_e)_i}{N}$ for $j = 2$ (elliptical).

Mutual Information: environment and morphology

- The mutual information $I(X;Y)$ between X and Y is then defined as,

$$
\begin{aligned}
I(X;Y) &= \sum_{i=1}^{N_d}\sum_{j=1}^{2} P(X_i,Y_j)\log\frac{P(X_i,Y_j)}{P(X_i)P(Y_j)} \\
&= H(X)+H(Y)-H(X,Y)
\end{aligned}
$$

where $H(X)$ and $H(Y)$ are the individual entropies associated with the variable X and Y respectively.

Mutual Information: environment and morphology

- The mutual information $I(X;Y)$ between X and Y is then defined as,

$$\begin{aligned} I(X;Y) &= \sum_{i=1}^{N_d} \sum_{j=1}^{2} P(X_i, Y_j) \log \frac{P(X_i, Y_j)}{P(X_i)P(Y_j)} \\ &= H(X) + H(Y) - H(X, Y) \end{aligned}$$

  where $H(X)$ and $H(Y)$ are the individual entropies associated with the variable X and Y respectively.

- $I(X;Y) \geq 0$ as $H(X,Y) \leq H(X) + H(Y)$ with equality only if the random variables X and Y are independent.

Mutual Information: environments on different scales

- $$I(X;\bar{X}) = \sum_{i=1}^{N_{d_1}} \sum_{k=1}^{N_{d_2}} P(X_i,\bar{X}_k) \log \frac{P(X_i,\bar{X}_k)}{P(X_i)P(\bar{X}_k)}$$
$$= H(X) + H(\bar{X}) - H(X,\bar{X})$$

  N_{d_1} and N_{d_2} are the number of voxels for grid sizes d_1 and d_2 respectively. The joint probability $P(X_i,\bar{X}_k) = \frac{n_{ik}}{N}$, where n_{ik} is the number of galaxies shared by the i^{th} voxel of grid size d_1 and k^{th} voxel of grid size d_2.

$$H(X,\bar{X}) = -\sum_{i=1}^{N_{d_1}} \sum_{k=1}^{N_{d_2}} P(X_i,\bar{X}_k) \log P(X_i,\bar{X}_k)$$

Conditional mutual information

- The conditional mutual information $I(X;Y|\bar{X})$ between X and Y given the value of $\bar{X}$ can be written as,

$$= \sum_{i=1}^{N_{d_1}} \sum_{j=1}^{2} \sum_{k=1}^{N_{d_2}} P(X_i, Y_j, \bar{X}_k) \log \frac{P(\bar{X}_k) P(X_i, Y_j, \bar{X}_k)}{P(X_i, \bar{X}_k) P(Y_j, \bar{X}_k)}$$

$$= H(X, \bar{X}) + H(Y, \bar{X}) - H(X, Y, \bar{X}) - H(\bar{X})$$

Applying the chain rule of probability one can write the joint probability $P(X_i, Y_j, \bar{X}_k) = P(\bar{X}_k | Y_j, X_i) P(Y_j | X_i) P(X_i)$ which will be $\frac{(n_s)_{ik}}{N}$ for $j = 1$ and $\frac{(n_e)_{ik}}{N}$ for $j = 2$. $(n_s)_{ik}$ and $(n_e)_{ik}$ are respectively the numbers of spirals and ellipticals shared by the i^{th} voxel of size d_1 and k^{th} voxel of size d_2 .

Conditional mutual information

- The conditional mutual information $I(X;Y|\bar{X})$ between X and Y given the value of $\bar{X}$ can be written as,

$$= \sum_{i=1}^{N_{d_1}} \sum_{j=1}^{2} \sum_{k=1}^{N_{d_2}} P(X_i, Y_j, \bar{X}_k) \log \frac{P(\bar{X}_k) P(X_i, Y_j, \bar{X}_k)}{P(X_i, \bar{X}_k) P(Y_j, \bar{X}_k)}$$

$$= H(X, \bar{X}) + H(Y, \bar{X}) - H(X, Y, \bar{X}) - H(\bar{X})$$

$$H(X, Y, \bar{X}) = -\sum_{i=1}^{N_{d_1}} \sum_{j=1}^{2} \sum_{k=1}^{N_{d_2}} P(X_i, Y_j, \bar{X}_k) \log P(X_i, Y_j, \bar{X}_k)$$

Interaction information

- The interaction information (McGill 1954) is defined as,

$$I(X;Y;\bar{X}) = I(X;Y|\bar{X}) - I(X;Y)$$

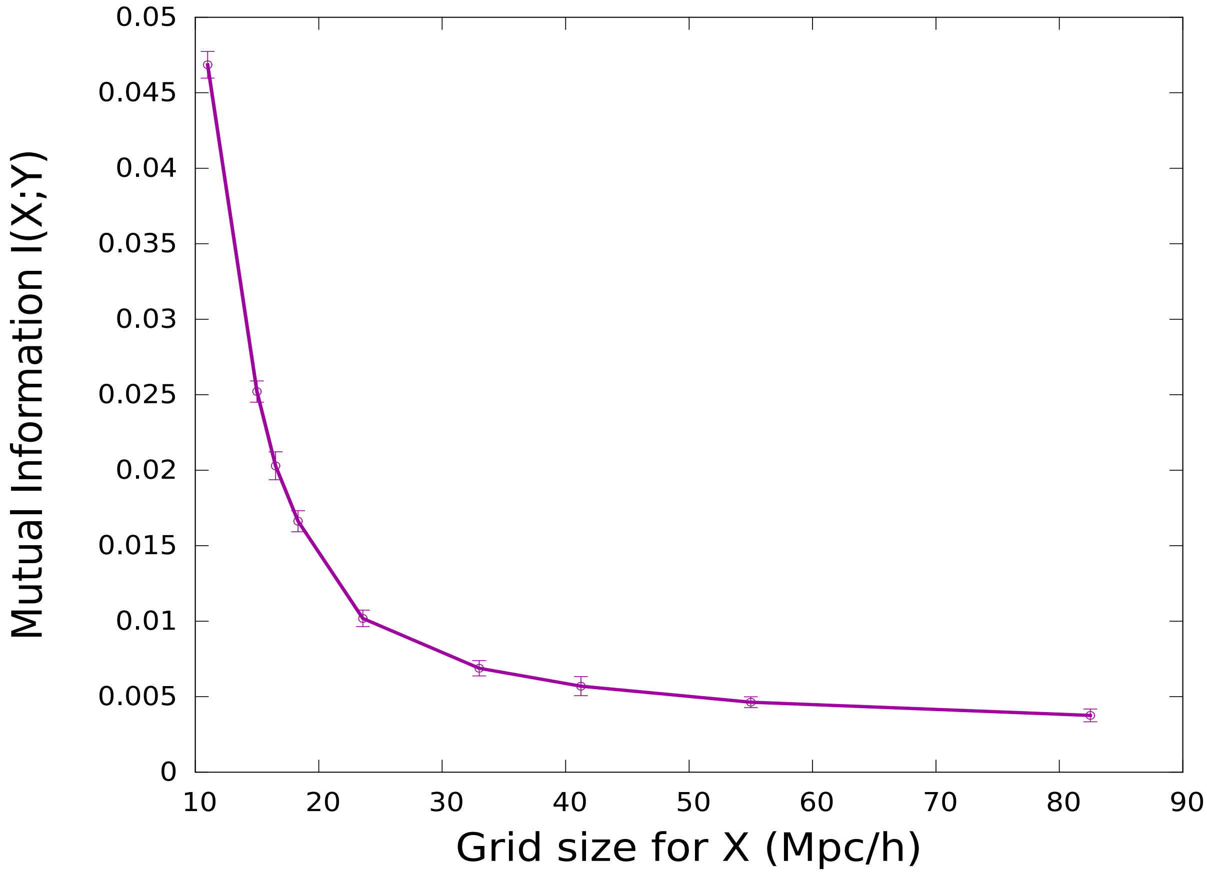
Interaction information

- The interaction information (McGill 1954) is defined as,

$$I(X;Y;\bar{X}) \;=\; I(X;Y|\bar{X}) - I(X;Y)$$

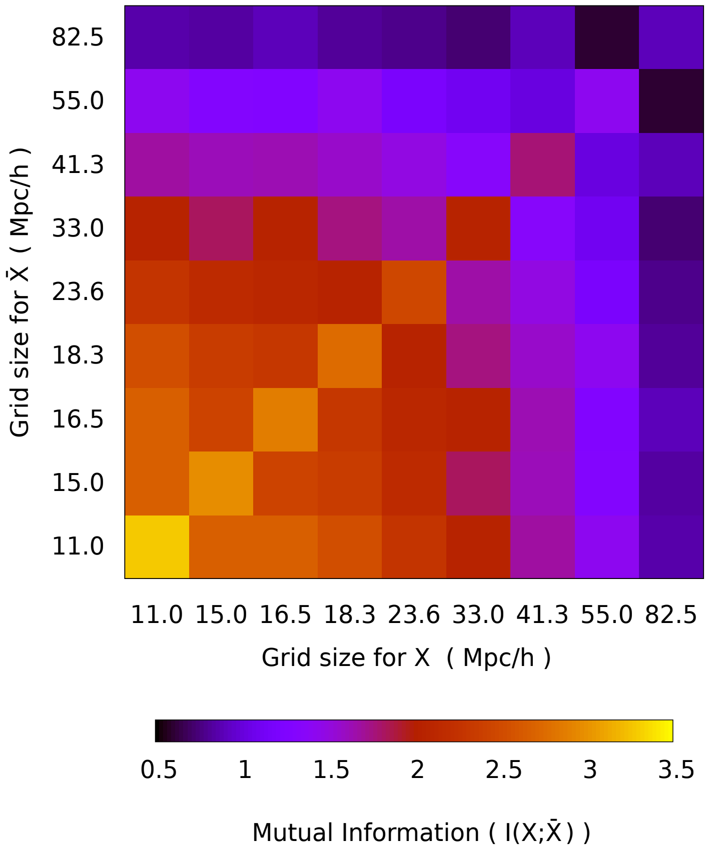
- This measures the loss or gain in information between X and Y due to the additional knowledge of the third random variable $\bar{X}$. Conditioning on $\bar{X}$ may increase, decrease or not change the mutual information between X and Y . Consequently, $I(X;Y;\bar{X})$ may be positive, negative or zero. $I(X;Y;\bar{X}) < 0$ indicates that conditioning on $\bar{X}$ weakens the correlation between X and Y whereas $I(X;Y;\bar{X}) > 0$ indicates that $\bar{X}$ enhances the correlation between X and Y . If conditioning on $\bar{X}$ does not affect the mutual information between X and Y then $I(X;Y;\bar{X}) = 0$.

Results: MI between environment and morphology



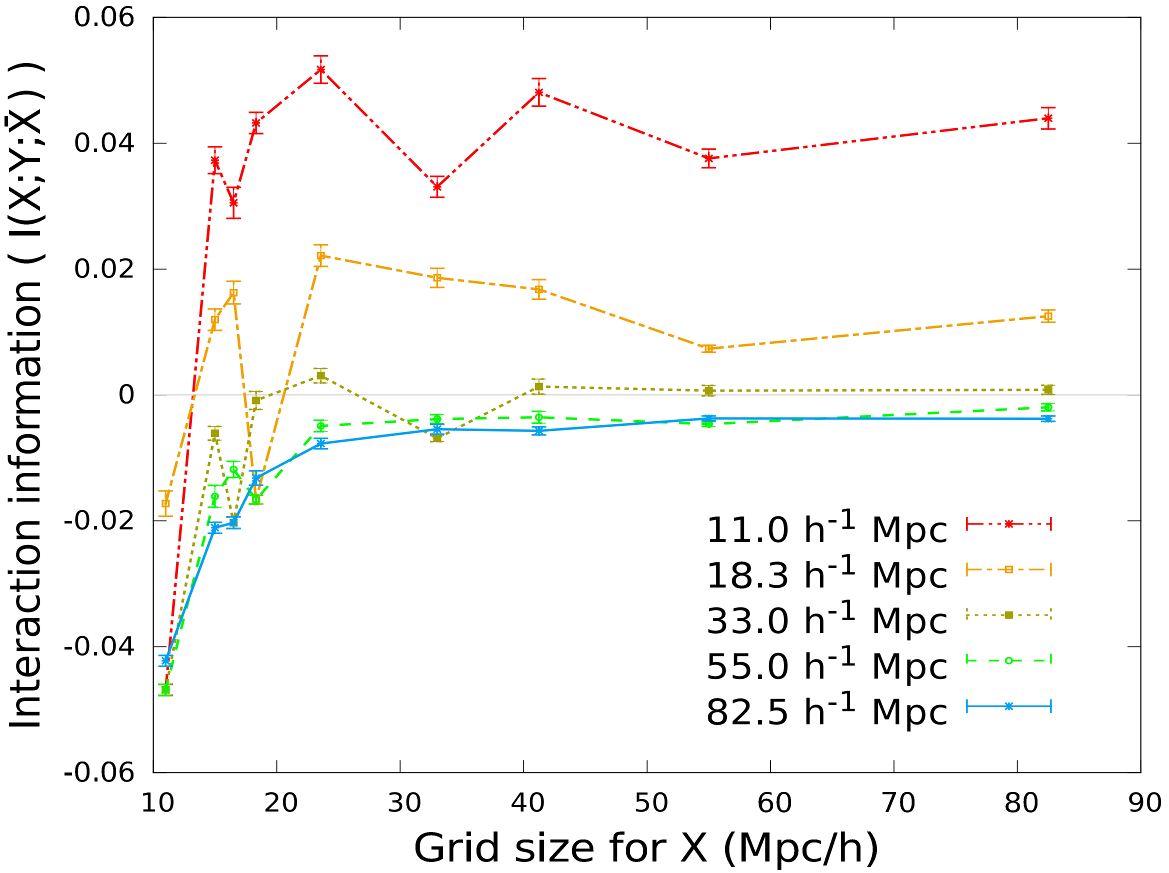
Pandey & Sarkar, 2017, MNRAS Letters

Results: MI between environments on different scales



Pandey & Sarkar, 2017, MNRAS Letters

Results: Interaction information



Pandey & Sarkar, 2017, MNRAS Letters

Conclusions

- We find a non-zero mutual information between morphology and environment which decreases with increasing length scales but persists throughout the entire length scales probed.

Conclusions

- We find a non-zero mutual information between morphology and environment which decreases with increasing length scales but persists throughout the entire length scales probed.
- Conditioning the environments on different scales show that a synergic interaction between morphology and environment exits up to at least a length scales of $\sim 30\,h^{-1}\,\mathrm{Mpc}$.

Conclusions

- We find a non-zero mutual information between morphology and environment which decreases with increasing length scales but persists throughout the entire length scales probed.
- Conditioning the environments on different scales show that a synergic interaction between morphology and environment exits up to at least a length scales of $\sim 30\,h^{-1}\,\mathrm{Mpc}$.
- The mutual information between the environments on different length scales are much larger than the mutual information between the morphology and environment.

Conclusions

- We find a non-zero mutual information between morphology and environment which decreases with increasing length scales but persists throughout the entire length scales probed.
- Conditioning the environments on different scales show that a synergic interaction between morphology and environment exits up to at least a length scales of $\sim 30\, h^{-1}\,\mathrm{Mpc}$.
- The mutual information between the environments on different length scales are much larger than the mutual information between the morphology and environment.
- The observed interaction between morphology and environment largely arise due to the mutual information shared between the environments on different scales.

Collaborators

- Suman Sarkar, Ph.D. student at Visva-Bharati

Acknowledgments

This work has been supported by the SERB DST project EMR/2015/001037.

THANKS