星系的暗晕占据分布模型

郭宏*

中国科学院上海天文台, 上海 200030

* 联系人, E-mail: guohong@shao.ac.cn

收稿日期: 2016-07-11; 接受日期: 2016-08-05; 网络出版日期: 2017-02-28

国家自然科学基金 (编号: 11543003) 资助项目

摘要 从星系巡天的观测中可以测量宇宙中星系的分布, 而这些我们观测到的可见物质只占据了宇宙组分的很小一部分. 如何从观测的星系分布来联系暗物质分布, 建立星系与暗晕之间的关联, 从而限制宇宙学参数, 一直都是星系宇宙学研究领域的一大课题. 本文具体介绍了描述星系空间分布的暗晕占据分布模型的主要框架和参数形式, 并介绍了模型的一个简单应用. 这个模型被广泛用于解释观测到的星系分布的两点相关函数, 通过模型构建星系和暗晕之间关联, 可以从理论上获得不同星系样本所在暗晕质量分布的信息, 这对于我们理解星系形成与演化模型以及限制宇宙学参数都有着重要的意义.

关键词 宇宙学, 宇宙大尺度结构, 星系, 暗物质晕

PACS: 98.62.Py, 98.62.Ve, 95.35.+d, 98.80.Bp

1 引言

在现代主流星系形成模型里, 星系是在暗物质晕的引力环境中, 通过气体的吸积、冷却和收缩而产生恒星, 从而进一步形成的^[1]. 这其中涉及复杂的物理过程, 也与所在暗物质晕的环境密切相关. 现代星系形成演化模型的一个重要任务就是要理清暗物质晕在星系形成过程中所起到的作用. 从宇宙学的角度来说, 暗物质和暗能量是现代宇宙学的两大难题. 虽然我们仍旧不清楚暗物质的本质, 但是目前较为成功的冷暗物质模型假设暗物质之间只有简单引力相互作用. 在这几十年的研究中, 我们对暗物质

的统计分布已经有了很好的了解, 甚至可以利用高精度的 N 体数值模拟来很好地描述在给定宇宙学条件下的暗物质分布. 如果我们能够建立星系和暗物质晕之间的关联, 就可以利用星系的分布来研究暗物质分布, 从而限制宇宙学参数.

把星系和暗物质晕联系到一起的方法有好几种. 最直接的方案就是流体数值模拟方法, 通过直接考虑星系形成演化中所涉及的复杂物理过程, 同时模拟暗物质、恒星、气体和黑洞等的共同演化, 可以得到星系的各种性质与所在暗晕的关系, 甚至可以用于研究星系形成过程对不同尺度环境的依赖性. 目前国际上最先进的几组高精度流体数值模拟包括 Illustris 项

引用格式: 郭宏. 星系的暗晕占据分布模型. 中国科学: 物理学 力学 天文学, 2017, 47: 049805

Guo H. The halo occupation distribution model for galaxies (in Chinese). Sci Sin-Phys Mech Astron, 2017, 47: 049805, doi: 10.1360/SSPMA2016-00330

目^[2], Eagle 项目^[3]等. 但是这种方法的结果依赖于运行流体数值模拟时所采用的程序算法和星系形成模型方案, 例如活动星系核反馈和超新星反馈的多少直接影响了最终的星系质量大小. 并且我们对目前的星系形成物理过程仍旧所知甚少, 即使是流体数值模拟的结果也很难与观测得到的各种统计性质完全一致. 流体数值模拟的一个缺陷是计算成本较高, 高精度大体积的流体数值模拟在当前的研究中还是一大难题. 目前主流的高精度流体数值模拟的体积也只有 10^6 Mpc^3 左右, 这对于研究星系大尺度结构分布和演化来说还是稍嫌不足的.

另一种方案是星系形成的半解析模型方法. 这种方法是基于 N 体的暗物质数值模拟中的暗晕并合历史, 通过经验性地加入一些星系形成中的可能涉及到的物理过程, 从而利用暗晕的分布来得到模拟的星系分布. 这种方法相比于运行流体数值模拟来说计算成本较低, 也可以用于大尺度结构研究, 它可以快速地生成符合某些观测统计性质的星系目录. 同时这些模型的参数有着富余的可调空间, 可以根据最新的观测结果进行一定的后期优化和修正. 然而, 这样的模型方法也存在和流体数值模拟一样的问题, 对于星系形成物理过程的不同理解会直接影响到最终的结果.

在接下来的篇幅中, 将会主要介绍一种参数化的模型方法, 它不依赖于星系形成的具体物理过程, 而是利用暗物质晕的分布并构建模型描述星系在暗晕里的分布, 就可以从统计上得到星系自身的分布. 这种模型方法的目的是为了经验性地建立星系和暗物质晕之间的关联, 从而避免引入具体的星系形成模型. 所以它不像流体数值模拟和半解析模型一样含有丰富的星系性质和物理过程信息. 但是, 这种统计上的经验性描述仍旧可以很大程度上帮助我们分析星系性质和暗晕之间的关系, 并用于建立宇宙学研究所需要的星系模拟样本. 在这种基于暗晕的模型方法框架下, 目前有几种主流模型, 包括暗晕占据分布模型 (Halo Occupation Distribution, HOD)、条件光度函数模型和子暗晕丰度匹配模型. 本文将侧重于对暗晕占据分布模型的详细介绍.

利用暗晕占据数来描述星系分布的模型思想最早可以追溯到 Jing 等人^[4]于 1998 年发表的文章. 他

们在文章里面利用了星系数目和暗晕质量之间的一个简单函数形式, 成功地解释了 Las Campanas 红移巡天数据所测量到的星系空间成团性分布. 后期人们进一步地发展和完善了 HOD 模型, 包括结合暗晕模型来预言星系成团性分布、提出 HOD 的更具函数形式、区分中心和卫星星系的不同占据数分布等等, 让它可以在更高的精度上和更多的应用方面来解释观测数据.

2 暗晕占据分布模型

在星系成团性研究方面, 人们常用的统计量是星系的空两点相关函数, 它表示了相对于平均分布来说, 在空间中找到一对距离为 r 的星系的额外概率. 在实际的应用中, 人们会测量三维空间中的星系两点相关函数, 即 $\xi(r_p, r_\pi)$, 其中 r_p 和 r_π 分别是两个星系之间空间距离的垂直于和平行于视线方向的分量. 在星系的红移巡天中, 由于星系本动速度的存在, $\xi(r_p, r_\pi)$ 的值与实空间的值有显著不同, 因为星系在视线方向的位置被红移畸变效应改变了. 为了减轻红移畸变效应的影响, 我们可以把 $\xi(r_p, r_\pi)$ 沿着视线方向积分, 从而获得投影的星系两点相关函数, $w_p(r_p)$, 即

$$w_p(r_p) = 2 \int_0^\infty \xi(r_p, r_\pi) dr_\pi, \quad (1)$$

积分前面的系数“2”是因为我们在统计时也需要考虑 r_π 为负数的星系对. 理论上来说, 我们应该沿着视线方向把 $\xi(r_p, r_\pi)$ 积分到无穷远处. 然而在实际的星系观测样本中, r_π 数值很大的星系对有很大一部分来自于噪声的贡献, 它们相互之间可能没有实际的关联, 而且 $\xi(r_p, r_\pi)$ 在 r_π 很大时的值也更接近于 0. 所以把最大的 r_π 值设为一个非常大的数值时, 对 $w_p(r_p)$ 的贡献不大, 反而降低了测量的信噪比, 而且最大的 r_π 值也同时需要考虑所使用的星系样本的体积大小. 我们注意到 $w_p(r_p)$ 只是星系间投影距离 r_p 的函数, r_p 较小时的测量反映了小尺度上的星系成团性, 但是 $w_p(r_p)$ 还有一部分可能来源于 r_π 较大的星系对的贡献.

在实际观测中, 为了测量星系的空两点相关函数, 我们采用的是 Landy-Szalay 估算公式^[5]来计算不同

尺度上星系对的数目,

$$w_p(r_p) = \frac{DD(r_p) - 2DR(r_p) + RR(r_p)}{RR(r_p)}. \quad (2)$$

为了考虑观测星系样本的空间角向和径向的几何分布来计算两点相关函数, 我们需要构建一个满足同样空间几何结构但是相互之间没有相关性的随机样本. 以上公式中的 DD , DR 和 RR 分别是归一化的星系对, 星系 - 随机对和随机 - 随机对的数目.

暗晕模型的基本假设是所有的星系都位于暗晕中, 因此我们可以把不同距离的星系对分成来自于单个暗晕和两个暗晕的项. 单暗晕项中两个星系都在同一个暗晕里面, 而暗晕的大小是有限的, 它反映了星系在小尺度上的成团性, 而双暗晕项则反映了星系在大尺度上的成团性. 通过这样的划分, 我们就把星系的两点相关函数跟暗晕的两点相关函数联系在一起, 我们可以进一步利用星系在暗晕中的分布从暗晕模型来得到星系的成团性.

HOD 模型的根本思想是在给定质量为 M 的暗晕里, 确定星系数目 N 的概率分布函数 $P(N|M)$, 同时结合关于星系在暗晕里的空间和速度分布的假设, 我们就可以利用暗晕的模型框架来预言星系的分布. 通常人们假设星系在暗晕中的分布与暗物质一致, 即星系的径向分布函数近似于满足 NFW 分布^[6], 同时我们假设它们的速度场分布也是一致的. 我们将在后文中讨论星系分布与暗物质分布不一致的情况. 根据上面的结论, 我们可以把实空间的两点相关函数写成如下的形式:

$$1 + \xi(r) = [1 + \xi_{1h}(r)] + [1 + \xi_{2h}(r)], \quad (3)$$

其中, “1h” 和 “2h” 分别代表单暗晕和双暗晕项. 因为总的星系数目正比于 $1 + \xi$, 它可以分解为来自于单暗晕项的星系对 (正比于 $1 + \xi_{1h}$) 和来自于双暗晕项的星系对 (正比于 $1 + \xi_{2h}$).

实空间的单暗晕项可以由如下公式计算^[7-9]:

$$1 + \xi_{1h}(r) = \frac{1}{2\pi r^2 \bar{n}_g^2} \times \int_0^\infty dM \frac{dn}{dM} \frac{\langle N(N-1) \rangle}{2} \frac{1}{2R_{vir}} F' \left(\frac{r}{2R_{vir}} \right), \quad (4)$$

其中, dn/dM 是暗晕的质量函数^[10-12], $\bar{n}_g$ 是星系样本的平均数密度, 以及 R_{vir} 是质量 M 的暗晕的维里

半径. 因为我们假设质量 M 的暗晕里星系数目是 N , 我们可以得到单暗晕里的星系对数目为 $\langle N(N-1) \rangle / 2$. $F(r/2R_{vir})$ 是暗晕内星系对距离小于 r 的累积分布函数, 所以 $F'(r/2R_{vir})$ 是星系对距离为 r 的概率密度函数. 因为星系可以细分为位于暗晕中心的中心星系以及周围的卫星星系, 我们可以进一步把 $F'(r/2R_{vir})$ 分解为中心星系 - 卫星星系以及卫星星系 - 卫星星系对, 即^[8,9,13]

$$\frac{\langle N(N-1) \rangle}{2} F' \left(\frac{r}{2R_{vir}} \right) = \langle N_{cen} N_{sat} \rangle F'_{cs} \left(\frac{r}{2R_{vir}} \right) + \frac{\langle N_{sat}(N_{sat}-1) \rangle}{2} F'_{ss} \left(\frac{r}{2R_{vir}} \right), \quad (5)$$

下标 “c” 和 “s” 分别表示中心和卫星星系, N_{cen} 和 N_{sat} 代表平均的中心和卫星星系数目. 因为我们假设星系径向分布与 NFW 分布一致, $F'_{cs}(r/2R_{vir})$ 实际上正比于 $\rho(r)r^2$, 即

$$F'_{cs}(x) = \frac{4c^2 x}{(1+2cx)^2 [\log(1+c) - c/(1+c)]}, \quad (6)$$

其中, 变量 c 是质量为 M 的暗晕的聚集因子^[14-16], 而 $F'_{ss}(r/2R_{vir})$ 的解析计算可以通过两个 NFW 分布函数的卷积来得到^[17].

对于双暗晕项来说, 它实际上是在暗晕自身的两点相关函数的基础上加入不同 HOD 带来的系数影响. 传统上, 从傅里叶空间计算 ξ_{2h} 会更为方便, 功率谱的双暗晕项可以表述为

$$P_{2h}(k, r) = P_m(k) \times \left[\frac{1}{\bar{n}_g} \int_0^{M_{max}} dM \frac{dn}{dM} N(M) b_h(M, r) y_g(k, M) \right]^2, \quad (7)$$

其中, $P_m(k)$ 是暗物质的功率谱, 在这里我们采用非线性物质功率谱从而更好地描述暗物质在小尺度上的分布^[18]. $y_g(k, M)$ 是归一化的暗物质径向密度分布的傅里叶变换, 对于 NFW 分布来说, 它有一个解析解^[19]. $b_h(M, r)$ 是质量为 M 的暗晕的偏袒因子, 它也与尺度 r 有关, 我们可以采用文献 [9] 的拟合公式. 在上面的积分中, 积分的上限是 M_{max} , 因为我们需要考虑到两个暗晕之间的距离不能小于它们的半径之和, 这被称为暗晕排斥效应 (Halo Exclusion Effect). 所以一个简单的做法是把 M_{max} 设成维里半径为 $r/2$

的暗晕的质量, 即我们只考虑所有维里半径小于 $r/2$ 的暗晕对功率谱的贡献. 然而在实际的情况中, 如果要考虑质量分别为 M_1 和 M_2 的暗晕对功率谱的贡献 (只要它们的半径和等于 r , 其中一个暗晕的半径可以大于 $r/2$), 我们就需要用到更加细致的计算公式, 这可以参见文献 [9]. 同时, $\bar{n}_g$ 是位于质量小于 $M_{\max}$ 暗晕里的星系的平均密度. 在仔细考虑这些效应之后, 我们可以通过傅里叶变换得到 ξ_{2h} , 即

$$\xi_{2h}(r) = \frac{1}{2\pi^2} \int_0^\infty P_{2h}(k, r) k^2 \frac{\sin kr}{kr} dk, \quad (8)$$

从而我们就得到了给定宇宙学条件下的实空间两点相关函数 $\xi(r)$ 的理论值.

根据公式 (1), 投影两点相关函数 $w_p(r_p)$ 和实空间的两点相关函数 $\xi(r)$ 之间可以通过简单的 Abel 积分联系在一起 [20–22],

$$w_p(r_p) = 2 \int_{r_p}^\infty \xi(r) \frac{r dr}{\sqrt{r^2 - r_p^2}}. \quad (9)$$

我们需要注意到, 在实际的计算中, 以上公式的积分只计算到了视线方向有限的距离 $r_{\max}$, 即 ∞ 被替换为 $\sqrt{r_p^2 + r_{\max}^2}$. 因为上述公式的成立条件是积分到无穷远处, 所以为了获得正确的 $w_p(r_p)$, 从而与观测进行对比, 我们可以通过一个额外的校正因子对上述公式进行修正 (详情可以参见文献 [22]).

在上述的暗晕模型中, 我们并没有涉及 HOD 的具体形式, 我们可以把 $\langle N(M) \rangle$ 分为中心星系和卫星星系的部分,

$$\langle N(M) \rangle = \langle N_{\text{cen}}(M) \rangle + \langle N_{\text{sat}}(M) \rangle. \quad (10)$$

对于中心星系来说, 人们最开始使用的是一个从 0 到 1 跳变的函数形式 (Tophat), 即中心星系只位于大于某一质量 $M_{\min}$ 的暗晕里. 但是考虑到星系光度 (或者恒星质量) 与暗晕质量之间存在着一定的涨落 —— 某一给定光度限的星系可能位于不同质量的暗晕中 —— 所以我们采用文献 [23] 中所提出来的一个渐变的函数形式,

$$\langle N_{\text{cen}}(M) \rangle = \frac{1}{2} \left[1 + \text{erf} \left(\frac{\log M - \log M_{\min}}{\sigma_{\log M}} \right) \right], \quad (11)$$

其中 erf 是误差函数. 我们可以看到 $M_{\min}$ 代表了一个特征的暗晕质量指标, 当 $M = M_{\min}$ 时, $\langle N_{\text{cen}}(M_{\min}) \rangle =$

0.5. 参数 $\sigma_{\log M}$ 代表了中心星系 HOD 的渐变分布的宽度, 它并不是给定星系光度条件下的暗晕质量涨落误差. 只有当星系光度和暗晕质量之间满足幂率的关系时, $L(M) \propto M^p$, 我们才可以得到 $\sigma_{\log M}$ 和给定暗晕质量下的星系光度分布涨落 $\sigma_{\log L}$ 的关系, 即 $\sigma_{\log L} = p \sigma_{\log M} / \sqrt{2}$ (参见文献 [24]).

对于卫星星系的 HOD 分布, 我们采用的是如下的形式 [24]:

$$\langle N_{\text{sat}}(M) \rangle = \langle N_{\text{cen}}(M) \rangle \left(\frac{M - M_0}{M'_1} \right)^\alpha, \quad (12)$$

其中 M_0 是暗晕质量下限参数, 即我们只考虑卫星星系位于 $M > M_0$ 的暗晕中. α 是大质量端的斜率. M'_1 是归一化常数, 它的值通常与人们常用的 M_1 (即平均包含一个卫星星系的暗晕的质量) 的值很接近, 但是只有当 $\langle N_{\text{cen}} \rangle = 1$ 并且 $M_0 = 0$ 时, $M = M'_1$ 才能推导出 $\langle N_{\text{sat}} \rangle = 1$. 我们假设卫星星系的数目分布满足平均值为 $\langle N_{\text{sat}} \rangle$ 的泊松分布 [23, 25], 同时我们假设只有当暗晕里有一个中心星系时, 它才会存在卫星星系, 所以我们可以得到 [23]

$$\langle N_{\text{cen}} N_{\text{sat}} \rangle = \langle N_{\text{sat}} \rangle, \quad (13)$$

$$\langle N_{\text{sat}} (N_{\text{sat}} - 1) \rangle = \langle N_{\text{sat}} \rangle^2. \quad (14)$$

在这样的假设下, 我们可以不需要考虑 N_{sat}^2 的概率分布. 当某些星系样本中存在一些暗晕没有中心星系满足星系样本选择条件而卫星星系满足条件时, 我们需要额外考虑公式 (13) 不成立的情况, 加入 $N_{\text{cen}} N_{\text{sat}}$ 的具体概率分布函数的影响. 在这个 HOD 模型里, 我们总共有 5 个自由参数, $M_{\min}$, $\sigma_{\log M}$, M_0 , M_1 和 α .

有了这些基本要素之后, 我们可以理论上计算星系的两点相关函数, 并通过与观测得到的统计值相比, 从而限制模型参数. 由于在暗晕模型里, 我们可以得到理论预言的星系样本数密度,

$$\bar{n}_g = \int_0^\infty dM \frac{dn}{dM} \langle N(M) \rangle. \quad (15)$$

我们同样可以利用样本星系数密度做额外的限制. 在实际应用中, 我们采用马尔科夫链 - 蒙特卡罗模拟的方法来搜索 HOD 的参数空间, 从而找到 χ^2 最小的最佳参数值并得到整个的参数分布统计. 给定 HOD

模型的概率正比于 $\exp(-\chi^2/2)$, 而

$$\chi^2 = (w_p - w_p^*)^T C^{-1} (w_p - w_p^*) + \frac{(\bar{n}_g - \bar{n}_g^*)^2}{\sigma_{\bar{n}_g}^2}, \quad (16)$$

其中 C 是测量误差的协方差矩阵. $\sigma_{\bar{n}_g}$ 是星系样本数密度的误差. 有上标 “*” 的量代表观测得到的统计值.

3 HOD 模型对投影两点相关函数的应用

HOD 模型在实际的观测中有着广泛的应用, 它可以被用于拟合观测得到的两点相关函数, 从而得到星系所在暗晕质量的估计. 它还可以结合弱引力透镜所得到的星系面密度分布函数得到更多宇宙学上的限制 [22, 26] 等. HOD 应用的一个显著的例子是对相关函数的模型解释. 早期人们在星系巡天中所测量得到的投影两点相关函数 $w_p(r_p)$ 大致上可以由一个幂率函数得到较好的拟合. 假设实空间两点相关函数 $\xi(r)$ 是幂率的形式

$$\xi(r) = (r/r_0)^{-\gamma}, \quad (17)$$

其中, r_0 和 γ 是模型的自由参数. 那么我们可以得到 [21]

$$w_p(r_p) = A r_p^{1-\gamma}, \quad (18)$$

$$A = r_0^\gamma \Gamma(0.5) \Gamma[0.5(\gamma-1)] / \Gamma(0.5\gamma),$$

其中, Γ 是 Gamma 函数. 文献 [21] 利用 SDSS 早期数据得到的 γ 最佳拟合值大约是 1.8, 而 r_0 的最佳拟合值在 $6 h^{-1} \text{Mpc}$ 左右. 随着 SDSS 更多星系观测数据的出现, 人们对于 $w_p(r_p)$ 有了更加精确的测量, 发现 $w_p(r_p)$ 在几个 $h^{-1} \text{Mpc}$ 的地方与幂率的函数形式有着很大的偏离, 特别是越亮星系样本的偏离会越严重. 这种偏离一开始让人们很困惑, 但是它在 HOD 模型里面却可以得到十分清晰的解释. 这样的偏离正是发生在单暗晕项与双暗晕项的交界区, 因为越亮的星系位于越大质量的暗晕中, 这些星系样本的交界区就会发生在更大的尺度上, 从而导致越亮的星系样本对幂率形式有越大的偏离.

上述例子只是 HOD 模型应用的很小一方面, 在这一节里将重点讨论 HOD 模型在投影相关函数里

的具体应用. 为了恰当地解释从星系巡天数据中测量的两点相关函数, 人们通常需要测量体积限样本里的星系成团性统计, 以保证在某一红移范围内所有亮于某一光度 (或者大于某一恒星质量) 的星系都被观测到了. 如果需要考虑某一光度区间的星系在暗晕中的分布, 我们可以通过它上下两个光度限的星系样本的成团性来得到. 当然, 如果构造星系样本时还有其他选择条件的话会更为复杂, 因为通常来说复杂的星系样本选择条件是很难在暗晕模型里得到完善的校正.

文献 [27] 在 2005 年测量了 SDSS DR2 低红移 $z \sim 0.1$ 的星系样本的 $w_p(r_p)$, 以及它对于星系的光度和颜色的依赖性. 这个样本有大约 20 万个星系, 占据了 2500 平方度的天区. 他们发现越亮星系的成团性越强, 同时越红的星系也有更高的 w_p 值和更陡的斜率. 在 2011 年, 他们根据 SDSS DR7 的星系样本 (覆盖 8000 平方度, 拥有大约 70 万个星系) 更新了 $w_p(r_p)$ 的统计, 仍旧得到了相同的结论. 他们发现对于亮星系来说, 星系偏袒的值随星系的光度迅速增加. 在 2008 年, 文献 [28] 利用 DEEP2 的 $z \sim 1$ 的星系红移巡天样本, 同样测量了 $w_p(r_p)$ 对光度和颜色的依赖性. 他们得到的结论与文献 [27] 基本一致, 特别是星系成团性对颜色依赖性比光度依赖性要更强, 但是他们在星系颜色光度分布的红序列 (Red Sequence) 里没有发现对颜色的明显依赖性.

随着 SDSS-III 项目的进行, 我们得到了更多中低红移范围的星系样本. 在 2013 年, 文献 [29] 利用了 SDSS-III BOSS 项目中的亮红星系样本, 测量了红移 $0.43 < z < 0.7$ 的 $w_p(r_p)$, 同样发现了与文献 [28, 30] 类似的光度和颜色依赖性. 文献 [31] 利用了 HOD 模型来解释得到的 $w_p(r_p)$ 测量值 (图 1 和 2). 图 1(a) 展示了 BOSS 项目中 CMASS 样本所选择的亮红星系的 $w_p(r_p)$ 对于光度的依赖性, 图中不同颜色的方块分别代表了 i 波段绝对星等为 $M_i < -21.6$, $M_i < -21.8$ 和 $M_i < -22$ 的样本, 相应颜色的实线代表了最佳拟合的 HOD 模型结果. 蓝色的点线是 $M_i < -21.6$ 样本最佳拟合 HOD 模型的单暗晕和双暗晕项. HOD 模型对 3 个光度限样本都拟合地很好, χ^2/dof 的值也在合理范围内.

图 1 (网络版彩图) (a) BOSS 项目中 CMASS 样本所选择的亮红星系的 $w_p(r_p)$ 对于光度的依赖性, 图中不同颜色的方块分别代表了 i -波段绝对星等为 $M_i < -21.6$, $M_i < -21.8$ 和 $M_i < -22$ 的样本, 相应颜色的实线代表了最佳拟合的 HOD 模型结果. 蓝色的点线是 $M_i < -21.6$ 样本最佳拟合 HOD 模型的单暗晕和双暗晕项. (b) 对 3 个星系光度限样本的最佳 HOD 模型拟合. 实线代表 $\langle N(M) \rangle$, 虚线和点线分别代表中心星系和卫星星系的 HOD. (c) 3 个光度限样本的最佳拟合 HOD 模型所预言的卫星星系比例. (d) 星系的寄主暗晕质量的概率密度分布, 实线是所有星系的结果, 虚线和点线分别代表中心和卫星星系的结果. 本图摘自 [31] 的图 2

Figure 1 (Color online) (a) The dependence of $w_p(r_p)$ on the luminosity for the luminous red galaxies of the CMASS sample in BOSS. Squares of different colors display the samples with i -band absolute magnitude of $M_i < -21.6$, $M_i < -21.8$ and $M_i < -22$, respectively. The solid lines of the corresponding colors are for the best-fitting HOD models. The blue dotted lines are the one-halo and two-halo components of the best-fitting HOD model for the $M_i < -21.6$ sample. (b) The best-fitting mean occupation distributions for the three luminosity threshold samples. The dashed and dotted lines are for contribution of central and satellite galaxies, respectively. (c) The predicted satellite fractions for the three best-fitting HOD models. (d) The probability distribution of the galaxy host halo mass, with the solid, dashed, and dotted lines for all, central, and satellite galaxies, respectively. This figure is from Figure 2 of ref. [31].

在小尺度上 ($r_p < 1 h^{-1} \text{Mpc}$), $w_p(r_p)$ 由单暗晕项主导, 而在大尺度上主要贡献来自于双暗晕项. 在小尺度上, 双暗晕项表现近似于一个常数, 这是因为当 r 很小时, 如果不考虑 $P_{2h}(k, r)$ 对 r 的微小依赖性 (来源于暗晕偏袒因子的尺度依赖性), 公式 (8) 积分得到

图 2 (网络版彩图) 类似于图 1, 但是我们显示的是不同颜色星系样本的成团性测量和 HOD 模型结果. 本图摘自 [31] 的图 6

Figure 2 (Color online) Similar to Figure 1, but for the clustering measurements and the HOD modeling results from the galaxy samples of different colors. This figure is from Figure 6 of ref. [31].

的 ξ_{2h} 就近似于一个正比于 $b_g^2 \sigma_8^2$ 的常数, 其中 b_g 是星系的线性偏袒因子. 当然, 严格的说, 双暗晕项贡献的 $w_p(r_p)$ 在 r_p 很小的时候, 并不是常数, 具体数值与 r_p 以及暗晕排斥效应有关. 图 1(b) 是最佳拟合的 HOD 模型, 实线代表 $\langle N(M) \rangle$, 虚线和点线分别代表中心星系和卫星星系的 HOD. 可以看到, 越亮的星系位于越大质量的暗晕中, 这直观表明了亮星系的高成团性是由于大质量暗晕的高成团性导致的, 因为越大质量的暗晕的偏袒因子也越高. 这与文献 [30] 对 SDSS 低红移样本和文献 [24] 对 DEEP2 高红移样本的 HOD 模型拟合结果是一致的.

因为在小尺度上主要是单暗晕项中的中心——卫星星系对主导, 为了解释测量到的小尺度成团性, BOSS 的星系样本中需要很大一部分比例的卫星星系存在. 图 1(c) 展示了对 3 个光度限样本的最佳拟合 HOD 模型所预言的卫星星系比例分布. 卫星星系的的比例是由 $\langle N_{cen} \rangle$ 、 $\langle N_{sat} \rangle$ 和暗晕质量函数共同决定的, 我们在这里展示的是通过蒙特卡罗模拟得到的卫星星系比例分布函数, 它的弥散与 HOD 模型参数的弥散相关. 越亮的星系样本里卫星星系的比例越

低, 但是所有这几个样本的卫星星系比例都在 10% 以下, 与文献 [30] 所得到的 SDSS 低红移样本结果有很大的差别. 在 SDSS DR7 主星系样本里, 最暗的 $M_r < -18$ 的星系样本的卫星星系比例达到了 30% 左右. 这主要是因为 BOSS 所观测的大部分是恒星质量在 $10^{11.5} h^{-1} M_\odot$ 左右的大质量红星系, 这样的亮红星系样本里大多数都是中心星系, 卫星星系只能存在少数非常大质量的暗晕里.

图 1(d) 显示的是星系的寄主暗晕质量的概率密度分布, 实线是所有星系的结果, 虚线和点线分别代表中心和卫星星系的结果. 卫星星系所在的暗晕质量要比中心星系的暗晕质量要高一个量级左右, 这个结果与 SDSS 低红移主星系样本的结果一致. 文献 [30] 发现 M_1 约等于 17 倍的 $M_{\min}$. 当然, $M_1/M_{\min}$ 的具体数值也与 HOD 模型的假设有关, 因为我们假设了暗晕中只有存在中心星系的情况下才能有卫星星系. 如果我们允许卫星星系所在暗晕中没有满足相同样本选择条件的中心星系的话, 中心和卫星星系暗晕质量之间的间隔会有一定的缩小. 同时, 这个比值还跟具体的光度限星系样本的数密度有关, 越亮的星系样本里这个比值通常越小 [31].

除了星系光度与暗晕质量之间的关联之外, 星系的另一个可直接观测量是颜色. 我们知道, 通常来讲越老的星系的颜色也会越红, 星系的颜色蕴含了星系的年龄和形成历史等信息. 所以我们需要研究星系的颜色和暗晕质量之间的关系. 为了剔除星光度的影响, 我们需要选取相同光度范围的星系样本用于研究星系成团性对颜色的依赖性.

我们在图 1 中选取了光度范围为 $-22.2 < M_r < -21.6$ 以及红移范围为 $0.48 < z < 0.55$ 的 BOSS 亮红星系样本, 我们把星系的颜色从红到绿分为 reddest, redseq 和 green 3 种, 其中 redseq 占据了星系颜色光度图的红序列部分. 我们可以看到, 在大尺度上 3 个不同颜色的星系样本的成团性是很接近的, 因为它们的光度范围都相同. 根据上面的结论, 它们所在的暗晕质量也是一致的. 但是在小尺度上, 越红的星系的成团性越高, 这与文献 [30] 用 SDSS 低红移星系样本得到的结果是一致的.

因为在小尺度上, 主要的贡献来自于在同一个暗晕里的星系对, 所以卫星星系的数目对小尺度的

成团性来说是十分重要的. 从图 1 中, 我们发现越红的星系样本中卫星星系的比例越高, 而不同颜色星系样本所在暗晕质量都是一致的 (这一点不完全正确, 因为在这里我们采用了一个简化的模型, 所以得到了一致的暗晕质量分布. 通常来讲, 越亮越红的星系会位于越大质量的暗晕中). 所以, 通过暗晕质量和卫星星系比例, 我们很好地解释了观测到的星系的空间分布对光度和颜色的依赖性.

4 总结

在本文中, 介绍了从理论上解释观测星系分布的暗晕占据分布模型的来源和具体形式, 并把这个模型应用到了测量的星系投影空间两点相关函数中, 从而限制了星系自身的性质, 如光度和颜色, 与所在暗物质晕质量之间的关系. 通过暗晕模型的结果, 我们发现越亮星系的较高成团性是由于它们处于越大质量的暗晕中. 而相同光度的星系中, 越红的星系样本在小尺度上的成团性越高, 这是因为越红的星系中卫星星系的比例越高, 从而增加了小尺度上的星系对数目.

在本文中, 我们通过分别考虑星系成团性对光度和颜色的依赖性, 从而得到这些观测现象的暗晕模型解释. 这样是为了分离不同暗晕模型参数对星系成团性的影响. 而同时分析光度和颜色的依赖性是十分困难的, 因为这就要求对于星系成团性的颜色和光度依赖有一个具体的模型, 例如星系偏袒与星系光度关系等等, 而不同星系样本的成团性可能是非常复杂的, 目前这方面的研究仍旧正在进行.

限于篇幅, 只介绍了 HOD 模型的一个简单应用. 实际上, HOD 模型已经被应用于宇宙学大尺度结构和星系形成研究的许多方面, 例如它可以通过研究红移空间的星系两点相关函数来帮助我们研究星系的本动速度场 [32–34]、它也能结合星系分布的弱引力透镜效应从而更好地限制宇宙学 [22, 26] 等等. 目前暗晕模型已经发展到了一个很成熟的阶段, 但是仍然存在着许多未解决的问题, 例如暗晕模型假设星系的成团性只与暗晕质量相关, 但是有部分研究认为暗晕的形成历史也会对星系的成团性有影响, 这被称为 as-

sembly bias. 还有一些悬而未决的问题包括, 卫星星系在暗晕中分布形式、中心与卫星星系性质关联等等. 未来, 大量星系巡天观测项目的实施将可以提供非常高精度的测量数据, 这对暗晕模型的精度提出

了非常大的挑战, 为了在精确宇宙学时代用暗晕模型来分析星系成团性, 我们还需要在暗晕模型发展上投入更多的努力. 感兴趣的读者可以进一步阅读相关文献以便更好地了解暗晕占据分布模型.

参考文献

- White S D M, Rees M J. Core condensation in heavy halos—A two-stage theory for galaxy formation and clustering. *Mon Not R Astron Soc*, 1978, 183: 341–358
- Vogelsberger M, Genel S, Springel V, et al. Introducing the Illustris Project: Simulating the coevolution of dark and visible matter in the Universe. *Mon Not R Astron Soc*, 2014, 444: 1518–1547
- Schaye J, Crain R A, Bower R G, et al. The EAGLE project: Simulating the evolution and assembly of galaxies and their environments. *Mon Not R Astron Soc*, 2015, 446: 521–554
- Jing Y P, Mo H J, Börner G. Spatial correlation function and pairwise velocity dispersion of galaxies: Cold dark matter models versus the las campanas survey. *Astrophys J*, 1998, 494: 1–12
- Landy S D, Szalay A S. Bias and variance of angular correlation functions. *Astrophys J*, 1993, 412: 64–71
- Navarro J F, Frenk C S, White S D M. A universal density profile from hierarchical clustering. *Astrophys J*, 1997, 490: 493–508
- Berlind A A, Weinberg D H. The halo occupation distribution: Toward an empirical determination of the relation between galaxies and mass. *Astrophys J*, 2002, 575: 587–616
- Zheng Z. Interpreting the observed clustering of red galaxies at $z \sim 3$. *Astrophys J*, 2004, 610: 61–68
- Tinker J L, Weinberg D H, Zheng Z, et al. On the mass-to-light ratio of large-scale structure. *Astrophys J*, 2005, 631: 41–58
- Mo H J, Jing Y P, White S D M. The correlation function of clusters of galaxies and the amplitude of mass fluctuations in the Universe. *Mon Not R Astron Soc*, 1996, 282: 1096–1104
- Sheth R K, Tormen G. Large-scale bias and the peak background split. *Mon Not R Astron Soc*, 1999, 308: 119–126
- Jenkins A, Frenk C S, White S D M, et al. The mass function of dark matter haloes. *Mon Not R Astron Soc*, 2001, 321: 372–384
- Tinker J L. Redshift-space distortions with the halo occupation distribution—II. Analytic model. *Mon Not R Astron Soc*, 2007, 374: 477–492
- Bullock J S, Kolatt T S, Sigad Y, et al. Profiles of dark haloes: Evolution, scatter and environment. *Mon Not R Astron Soc*, 2001, 321: 559–575
- Macciò A V, Dutton A A, van den Bosch F C. Concentration, spin and shape of dark matter haloes as a function of the cosmological model: WMAP1, WMAP3 and WMAP5 results. *Mon Not R Astron Soc*, 2008, 391: 1940–1954
- Zhao D H, Jing Y P, Mo H J, et al. Accurate universal models for the mass accretion histories and concentrations of dark matter halos. *Astrophys J*, 2009, 707: 354–369
- Sheth R K, Hui L, Diaferio A, et al. Linear and non-linear contributions to pairwise peculiar velocities. *Mon Not R Astron Soc*, 2001, 325: 1288–1302
- Smith R E, Peacock J A, Jenkins A, et al. Stable clustering, the halo model and non-linear cosmological power spectra. *Mon Not R Astron Soc*, 2003, 341: 1311–1332
- Cooray A, Sheth R. Halo models of large scale structure. *Phys Rep*, 2002, 372: 1–129
- Davis M, Peebles P J E. A survey of galaxy redshifts. V. The two-point position and velocity correlations. *Astrophys J*, 1983, 267: 465–482
- Zehavi I, Blanton M R, Frieman J A, et al. Galaxy clustering in early sloan digital sky survey redshift data. *Astrophys J*, 2002, 571: 172–190
- van den Bosch F C, More S, Cacciato M, et al. Cosmological constraints from a combination of galaxy clustering and lensing. I. Theoretical framework. *Mon Not R Astron Soc*, 2013, 430: 725–746
- Zheng Z, Berlind A A, Weinberg D H, et al. Theoretical models of the halo occupation distribution: Separating central and satellite galaxies. *Astrophys J*, 2005, 633: 791–809
- Zheng Z, Coil A L, Zehavi I. Galaxy evolution from halo occupation distribution modeling of DEEP2 and SDSS galaxy clustering. *Astrophys J*, 2007, 667: 760–779
- Kravtsov A V, Berlind A A, Wechsler R H, et al. The dark side of the halo occupation distribution. *Astrophys J*, 2004, 609: 35–49
- Zu Y, Mandelbaum R. Mapping stellar content to dark matter haloes using galaxy clustering and galaxy-galaxy lensing in the SDSS DR7. *Mon Not R Astron Soc*, 2015, 454: 1161–1191
- Zehavi I, Zheng Z, Weinberg D H, et al. The luminosity and color dependence of the galaxy correlation function. *Astrophys J*, 2005, 630: 1–27

- 28 Coil A L, Newman J A, Croton D, et al. The DEEP2 galaxy redshift survey: Color and luminosity dependence of galaxy clustering at $z \sim 1$. *Astrophys J*, 2008, 672: 153–176
- 29 Guo H, Zehavi I, Zheng Z, et al. The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Luminosity and color dependence and redshift evolution. *Astrophys J*, 2013, 767: 122
- 30 Zehavi I, Zheng Z, Weinberg D H, et al. Galaxy clustering in the completed sdss redshift survey: The dependence on color and luminosity. *Astrophys J*, 2011, 736: 59
- 31 Guo H, Zheng Z, Zehavi I, et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Modelling of the luminosity and colour dependence in the Data Release 10. *Mon Not R Astron Soc*, 2014, 441: 2398–2413
- 32 Guo H, Zheng Z, Zehavi I, et al. Velocity bias from the small-scale clustering of SDSS-III BOSS galaxies. *Mon Not R Astron Soc*, 2015, 446: 578–594
- 33 Guo H, Zheng Z, Jing Y P, et al. Modelling the redshift-space three-point correlation function in SDSS-III. *Mon Not R Astron Soc*, 2015, 449: L95–L99
- 34 Guo H, Zheng Z, Zehavi I, et al. Redshift-space clustering of SDSS galaxies-luminosity dependence, halo occupation distribution, and velocity bias. *Mon Not R Astron Soc*, 2015, 453: 4368–4383

The halo occupation distribution model for galaxies

GUO Hong*

Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China

We can measure the galaxy distribution in the universe from the observation of the galaxy surveys. However, the baryon component only occupies a small fraction of the universe. An important topic in the research field of galaxy formation and evolution is relating the galaxy distribution to that of the underlying dark matter, building up the connection between galaxies and the dark matter halos, and then constraining the cosmology. In this review, I will introduce the framework and detailed parameterizations of the halo occupation distribution model that is used to describe the galaxy phase-space distribution. This model is commonly used to interpret the observed galaxy two-point correlation function and theoretically obtain the halo mass information for different galaxy samples through modeling the connection between galaxies and halos. Therefore, the application of this model is of great importance in the understanding the galaxy formation and evolution model and constraining cosmology.

cosmology, large-scale structure of the universe, galaxies, dark matter halos

PACS: 98.62.Py, 98.62.Ve, 95.35.+d, 98.80.Bp

doi: 10.1360/SSPMA2016-00330