

Making The Local Group In A Computer

James Judd

Level 4 Project, MPhys Physics

Supervisor: Professor A. Jenkins

Department of Physics, Durham University

Submitted: April 20, 2022

I use phase information from the BORG Algorithm to create initial conditions, using IC-GEN, to run dark matter simulations on GADGET-3 code in search of haloes with properties similar to the Local Group (LG). I develop a scoring system for the LG analogues, introducing isolation score to the system used by the SIBELIUS project. I find that simulations whose initial conditions are on nearby iterations of BORG's Markov chain to a LG analogue, have an increased probability of producing an analogue themselves. The Markov chain iteration chosen by SIBELIUS is shown to be unremarkable, however no other iteration stands out as superior. The properties of the best analogues are shown to be less unique than the LG's properties, proving the difficulty in computationally reproducing a LG-like system.

Contents

1. Introduction	2
A. The Local Group	2
B. Galactic Evolution	3
1. Dark Matter	3
2. Top-Down and Bottom-Up Formation	5
3. The Lambda-CDM Model	6
C. This Project and Previous Work	9
1. The SIBELIUS Project and the BORG Algorithm	10
2. Octree Functions	11
3. SIBELIUS' Findings	12
4. This Work	14
2. Methodology	14
1. Simulation Code	14
2. Local Group Criteria	15
3. Scoring Analogues	16
3. Data and Discussion	17
4. Conclusions	26
Acknowledgments	27
References	27

1. INTRODUCTION

In this work I will simulating and analysing Local Group analogues, testing the uniqueness and sensitivity of initial conditions, specifically building upon the work of Sawala T.; McAlpine S.; Jasche J.; et al. (2022)^[1]. This will be done by running simulations on GADGET-3 code^[2] to search for dark matter haloes with properties similar to the Local Group, whose initial conditions were produced with IC-GEN^[3] based on phase information from the BORG algorithm^[4]. To better understand the methods used in these simulations, one must first be acquainted with the Local Group; dark matter; and the processes involved in galaxy formation within the Lambda-CDM (Lambda-Cold Dark Matter) model.

A. The Local Group

Our Milky Way (MW) seems to be a fairly typical galaxy: its 10^{11} stars^[5] across a 10^5 light-year wide^[6] Sbc/SBbc galaxy (Figure 1) enveloped within a dark matter halo for a combined total of 10^{12} solar masses^[7], certainly does not stand out from the 10^{11} other galaxies^[8] in the observable universe. The assumption that there is nothing special about our local universe is important in astronomy as it allows our observational understandings to be applied to the entirety of space.

The Local Group (the Milky Way and it's neighbours) also seems normal. Its other largest constituents, the Andromeda Galaxy (M31) and the Triangulum Galaxy (M33), are amongst at least 100 other galaxies including satellite galaxies like the Large Magellanic Cloud (LMC). The Local Group (LG) is part of the Virgo Supercluster which is itself part of the larger Laniakea Supercluster.

M31 was first described as a "smear" by Persian astronomer Abdal-Rahman al-Sufi in 964^[9]_{A.D.}, nearly a millennium before galaxies were understood to be distinct stellar systems. The nineteenth and twentieth centuries were rife with debate over the nature of the bright nebulous bodies observed in the night sky: small nebulae on the fringes of the Milky Way, then believed

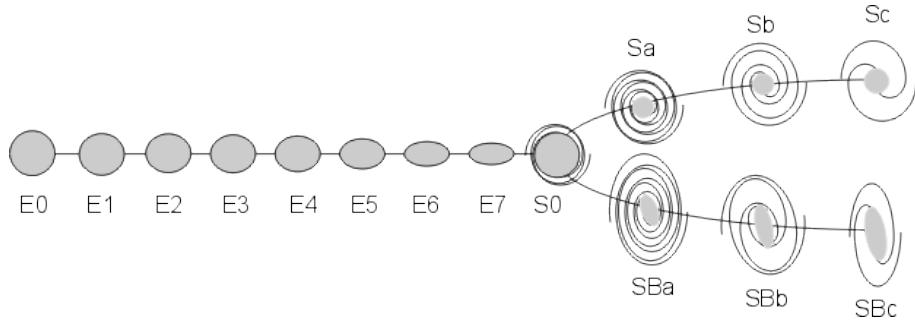


FIG. 1: The Hubble Sequence. Simoes C. Astrnoo. (2013). Elliptical galaxies are denoted by E, spiral galaxies by S, lenticular galaxies by S0, and irregular galaxies by Irr. The degree of elongation of elliptical galaxies is denoted by increasing integers from 0 to 7. Spiral galaxies are split into barred galaxies (denoted by a B) and non-barred (denoted by A or no denotation). The De Vaucouleurs system adds a denotation option of AB for weakly barred galaxies. The prominence of the bulge and degree of tightness of the winding of the spiral arms is then denoted, decreasing through a, ab, b, bc, ...

to be the entire universe; or huge bodies well beyond. This culminated in the "Great Debate of Curtis and Shapely" in 1920, only settled later that decade by Edwin Hubble's measurement of Cepheids in NGC 6822^[10], M33^[11] and M31^[12].

Looking to the future, M31 is set to collide with the Milky Way in approximately 4.5 billion years^[13]. Collisions and many other aspects of galactic evolution will be discussed in the following section.

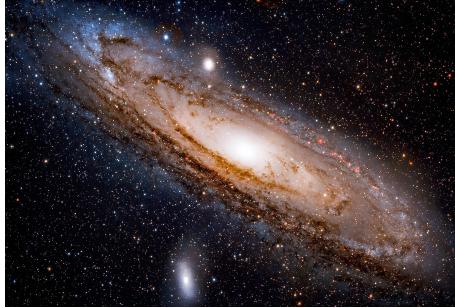


FIG. 2: The Andromeda Galaxy. Dayag, D. (2019). M31 is an SAB galaxy. Satellite galaxies M32 and M110 can also be seen, upper left-centre and lower left-centre respectively.

B. Galactic Evolution

1. Dark Matter

Dark matter has thoroughly encapsulated the public imagination since its conceptualisation in 1884^[14]. The idea that countless imperceptible objects may permeate the universe is certainly a fascinating one, and unavoidable associations with the supernatural may explain why it has garnered such profound interest amongst the non-scientific community. From a more principled position, dark matter perhaps holds even more intrigue; its effects on the universe around us are obvious, yet any definitive explanation of its true nature is as elusive as it appears in the night sky.

Dark matter is simply the name for hypothetical non-electromagnetically-interacting matter that accounts for 85% of the matter in the universe, a value that has been determined through several astronomical observations. One of these is the rate that galaxies rotate about their black hole, first measured by Rubin and Ford in 1970^[15] to remarkably unexpected results^[16]. Kepler's third law of planetary motion gives declining orbital velocities with distance, which holds in star-planet systems, but when we look at disc galaxy rotation profiles this is no longer the case. In Figure 3 we can see the true galaxy rotation curve for NGC 3198 differing greatly from the expected value calculated from its visible matter distribution (denoted by disk).

The observed curve is almost flat, in other words velocity is nearly constant beyond a certain radius; and is typical of other observed galaxies^[17]. Catalogues of evidence supporting our current theories of orbital mechanics leaves the most reasonable explanation to be an invisible halo of mass permeating and surrounding the galaxy, vastly more massive than the galaxy itself. The dark matter density profile to cause this flat rotation curve is well modelled by the Navarro-

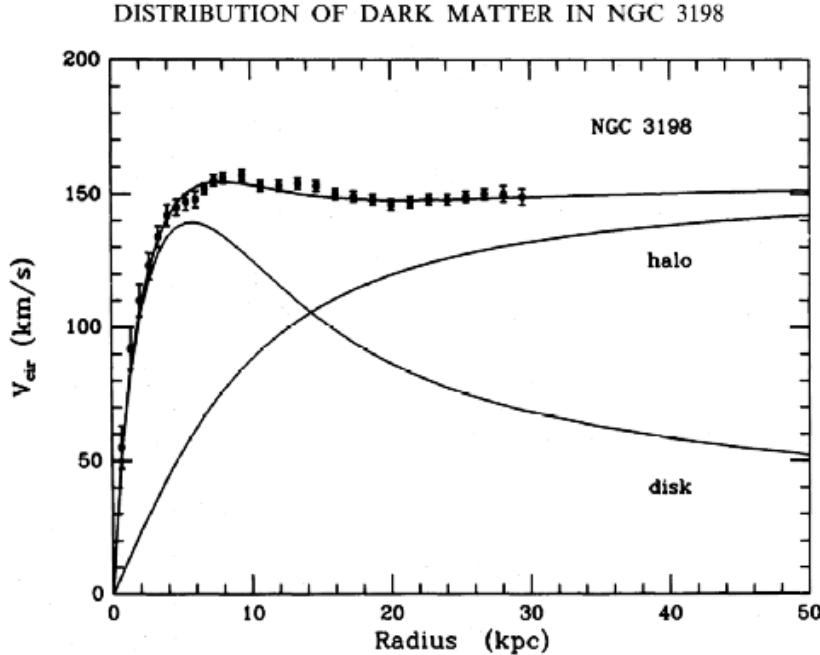


FIG. 3: Rotation Velocity by Radial Distance of NGC 3198. Van Albada, T.S., et al. (1985). This graph shows the discrepancy between the expected rotation velocity curve based solely on NGC 3198's visible matter (labelled disk) and the true curve (labelled NGC 3198), as well as the rotation curve of the interpreted dark matter halo.

Frenk-White (NFW) profile^[18]:

$$\rho(r) = \frac{\rho_0}{r} \left(1 + \frac{r}{R_s}\right)^2 \quad (1)$$

where density as a function of radius, $\rho(r)$, has two parameters that vary between dark matter haloes: inner density, ρ_0 , and scale radius, R_s . Collisionless dark matter simulations have repeatedly shown the NFW to be a good theoretical fit^[19], along with the similar Einasto profile^[20].

Further evidence for the existence of dark matter is extensive, through gravitational lensing (bending of light) and other observations. The Bullet Cluster is one such example, where following a collision of two galaxy clusters, each cluster's respective centre of mass, observed via the geometry of their lensing, is transposed from its baryonic centre^[21]. The reason for this is that when galaxies pass through each other, the relatively low density of interstellar space leaves dark matter and stars largely unperturbed, but gas lags behind due to ram pressure, R_{ram} , exerted upon a body travelling through a fluid medium of density ρ : $P_{ram} = \rho u_i u_j$ where u_z is the z component of the fluid velocity. This effect is thought to have a profound effect on galaxy formation: as galaxies fall towards the centre of their cluster, more and more of the cool, dense gas used to form stars is stripped away. This is shown in the emission spectra of the Virgo and Coma Superclusters^[22].

Despite barely interacting with baryonic matter except through gravity, dark matter is thought to possibly contain some, even if only a small proportion, of baryonic matter. This means that it may interact with light at undetectable levels, but otherwise does not differ from

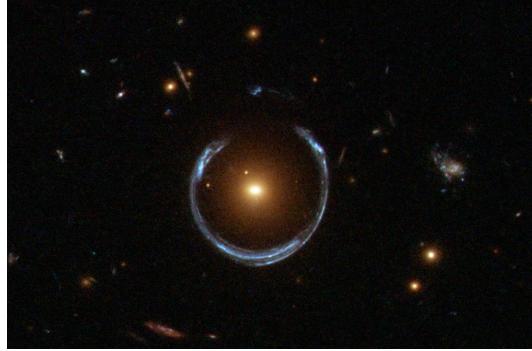


FIG. 4: Gravitational Lensing. ESA/Hubble & NASA. (2011). This image shows an Einstein Ring caused by a galaxy acting as a gravitational lens to an otherwise-obscured galaxy behind it.

regular matter at all. It may be located in massive astrophysical compact halo objects (MACHOs), huge bodies that traverse interstellar space emitting little to no light; in non-luminous gas; in condensed objects such as black holes and neutron stars; or in non-luminous objects such as planets and brown dwarfs. Baryonic dark matter has been shown to only be a small part of the whole via the abundancy of baryonic matter during Big Bang nucleosynthesis^[23]. More baryonic matter in the early universe would lead to more efficient conversion of matter to helium-4 and thus a lower proportion of deuterium in the current universe. Deuterium observations match a lower baryonic-proportioned early universe and thus a minority of baryonic dark matter.

The most likely non-baryonic candidate is an elementary particle that we are yet to discover; weakly interacting massive particles (WIMPs) are believed to be the best candidate for this^[24]. Aside from gravity, WIMPs may interact via the weak nuclear force or forces outside the Standard Model, providing the force is at least as weak as the weak nuclear force. Supersymmetric extensions of the Standard Model predict particles of this type, however the Large Hadron Collider has failed to produce any evidence of supersymmetry and therefore WIMPs remain purely hypothetical^[25]. WIMPs are predicted to have a large mass compared to other fundamental particles and be relatively slow-moving, termed “cold”. A more accurate definition for the “temperature” of dark matter is through its free-streaming length, a key factor in current galaxy formation theory, as we shall discuss next.

2. *Top-Down and Bottom-Up Formation*

In the early universe, small perturbations would only survive if the particles were too slow or massive to escape the heightened local gravitational forces. Take a neutrino of mass m_v ; neutrinos decoupled roughly one second after the Big Bang and travelled at relativistic speeds, $v(t)$, until their temperature, T , dropped to about $3k_B T \sim m_v c^2$ at time t_{NR} ; where k_B is Boltmann’s constant, and c is the speed of light in vacuum. Until present, t_{eq} , neutrinos will have travelled a comoving distance (distance relative to universe scale factor a), λ_{com} , of:

$$\lambda_{com} = \int_1^{t_{eq}} \frac{v(t)}{a(t)} dt. \quad (2)$$

Taking $v \sim c$ prior to t_{NR} , and $a(t) \propto t^{1/2}$ during the radiation era^[26], we get:

$$\lambda_{com} = \frac{2ct_{NR}}{a_{NR}} \left(1 + \ln \left(\frac{a_{eq}}{a_{NR}} \right) \right) \quad (3)$$

which we term the free-streaming length, as it is the distance travelled since the neutrino last interacted with matter.

To give this equation some perspective, in the 1980s neutrinos were a strong dark matter candidate^[27] as an experiment incorrectly measured them to have a mass of $m_\nu c^2 \sim 30$ eV, 300 times too large. This would give them a free-streaming length of 30 Mpc, meaning that no structures smaller than 30 Mpc (the size of the Virgo Supercluster)^[28] would form; clearly inconsistent with our known universe. It was therefore theorised that smaller-scale structures, like galaxies, formed later from the fragmentation of the superstructure. This is an example of top-down structure formation and is fundamental to “hot” dark matter (HDM) theory^[29].

Due to their large masses (~ 100 GeV, if they interact via the weak force)^[30], WIMPs and other cold dark matter candidates would have become non-relativistic before they decoupled, giving them a far shorter free-streaming length. This would imply a significantly lower minimum structure size, leading to the bottom-up hierarchy of cold dark matter (CDM) theory where small objects collapse under mutual gravity into larger structures.

Warm dark matter (WDM) theory has properties in-between the other two. It hypothesises a form of dark matter that causes structure formation to occur bottom-up above its free-streaming scale and top-down below it. The most popular WDM candidates are sterile neutrinos^[31] and gravitinos^{[32][33][34]}.

The Lambda-CDM (Λ CDM) model of cosmology (where Λ is the energy of free space, a cosmological constant associated with dark energy), is the most widely accepted model of our universe as it explains the existence and structure of the cosmic microwave background; the distribution of galaxies; the abundances of light-element isotopes; and the accelerating expansion of the universe.

3. The Lambda-CDM Model

Otherwise known as the standard model, the Lambda-CDM model emerged in the late 1990s to combine the observations mentioned prior which, until then, had seemed mutually inconsistent.

Red shift in galaxy emission and absorption spectra, and the rate of light decay in supernovae, were two problems confronted by astrophysicists at the time. It was known that electrons within elements can only absorb and emit light at very specific frequencies, determined by quantum mechanics. When looking at light from galaxies, it was observed that these specific frequencies of light were shifted towards the “red” end of the spectrum (lower frequency). This can be explained by a Doppler shift.

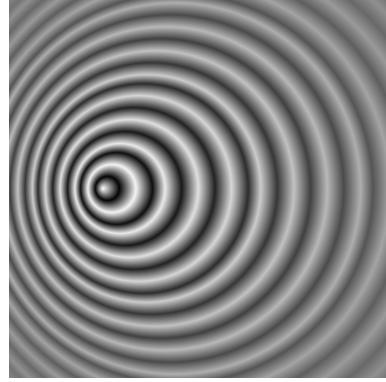


FIG. 5: Doppler Shift. Broks, P. (2008). This image shows a wave source moving towards the left of the image, compressing the waves ahead of it and stretching the waves in its wake.

As can be seen in Figure 5, a moving wave source compresses the waves ahead of it and stretches the waves behind it, relative to a stationary observer, due to the speed of light remaining constant regardless of the velocity of the emitter. The compressed waves therefore have an increased frequency relative to a stationary observer, and the stretched waves a decreased one. Given that redshift was seen in (nearly) all galaxies, this implied that the whole universe was receding away from us. Furthermore, it was seen that more distant galaxies receded faster than closer ones, implying that all galaxies were receding away from a single point, and even more strikingly that the space between galaxies was itself expanding. This corroborated with the observations of time dilation in distant supernovae:

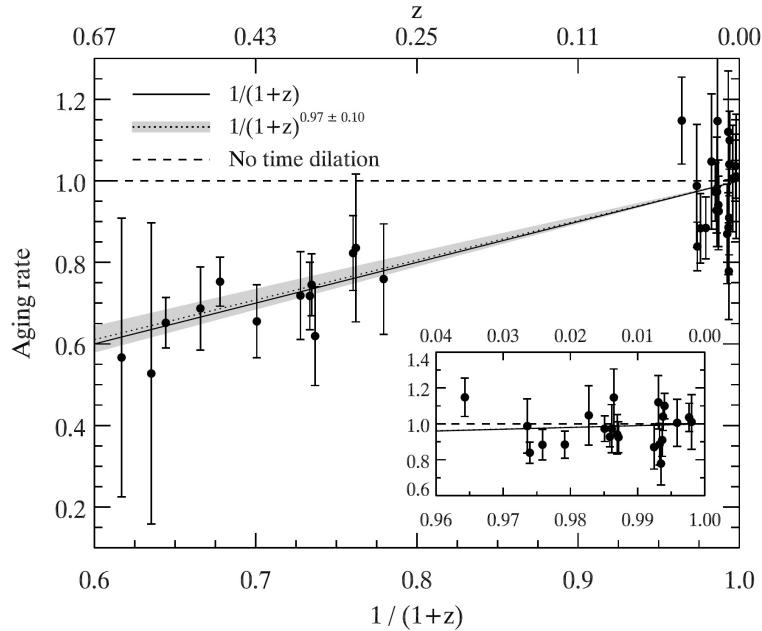


FIG. 6: Time Dilation of Supernova Gamma Ray Bursts. Blondin, S.; et al. (2008). This graph shows the aging of supernovae, specifically the duration of their gamma burst burst relative to nearby supernova. z is the degree of redshift of the light, and clearly shows that the greater the redshift, the slower the supernova seems to have aged.

The only reasonable explanation for the relationship seen in Figure 6 is the time dilation of

distant galaxies caused by mutual recession, with recession velocity increasing linearly with distance. An expanding universe, a pillar of the Lambda-CDM model, is therefore a natural solution to both the redshift and supernovae problems.

Whilst the local speed of expansion is subluminal (slower than the speed of light) at approximately $70 \text{ km s}^{-1} \text{ Mpc}^{-1}$,^[35] over vast distances recession can reach superluminal velocities, causing more and more galaxies to fall out of our observable bubble as the universe expands.

This expansion can be explained by a vacuum energy of empty space, known as dark energy, Λ , creating negative pressure throughout space. Based on 2018 Planck Satellite data^[36], dark matter is believed to constitute over 68% of the mass-energy density of the universe, with dark matter at roughly 27% and ordinary matter at 5%. The Λ CDM model has approximately 0.01% of the energy density stored in the cosmic microwave background and 0.5% in relic neutrinos, both remnants from the birth of the universe.

The Big Bang is modelled to have had temperatures of around 10^{27} Kelvin, followed by cosmic inflation less than 10^{-29} seconds later. As mentioned previously, neutrinos decoupled one second after the Big Bang; however the cosmic microwave background (CMB) is radiation leftover from the following 3.8×10^6 years where the universe remained at 10^5 K. Neutrinos and CMB radiation can be seen to emanate from all directions in the sky, providing more evidence for an expanding universe.

Despite its omnidirectionality, the CMB is not isotropic, as shown in Figure 7. The anisotropies come in two categories: primary and secondary. Secondary, or "late time", anisotropy is the result of gravity and various physical processes since the reionisation phase of the universe, one billion years ago, as there is once again charged matter to interact with light.

More importantly to universe evolution is primary anisotropy, leftover from the early universe. Acoustic oscillations and diffusion damping (photons flowing from hot regions of space to cold) are the principal influences on primary anisotropy. Acoustic oscillations arose from the conflict in the photon-baryon plasma of the early universe. Photonic pressure tended to erase quantum fluctuations whilst gravity tried to collapse them into regions of high density. The conjunction of these affects led to resonances seen now in the CMB, shown in Figure 8. The angular scale of the first peak determines the curvature of the universe^[37]; the third contains information on the dark matter density^[38]; and the temperature ratio of the even peaks to the odd determines baryon density^[39].

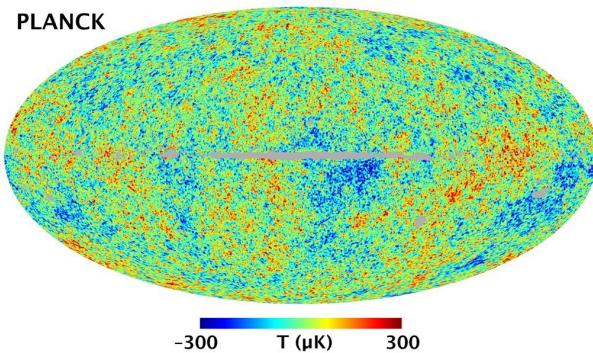


FIG. 7: The Cosmic Microwave Background. Encyclopædia Britannica. (2022). A relative-temperature graph of the CMB highlighting its small-scale anisotropy.

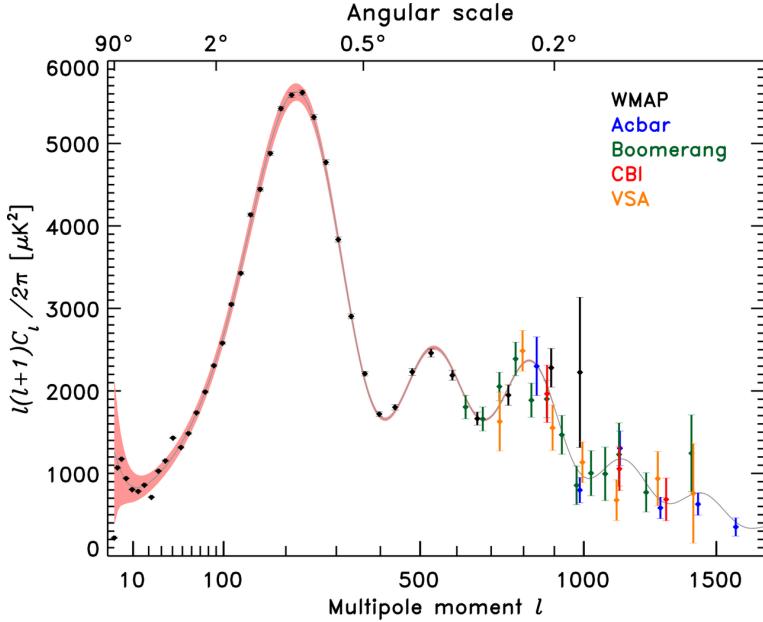


FIG. 8: Power Spectrum of the Cosmic Microwave Background. NASA LAMBDA WMAP Data Product Images. (2006). This graph, showing the theoretical model (bold line) and data from WMAP (2006), Acbar (2004), Boomerang (2005), CBI (2004) and VSA (2004), shows the power spectrum of the CMB temperature anisotropy in terms of angular scale and multipole moment. The data fits the theoretical model exceptionally well, showing that the Lambda-CDM model fits well with this observed characteristic of the CMB.

The ability to model aspects of the universe as a power spectrum is incredibly useful in modern research. Creating the initial conditions for simulations is much simplified and thus commonly used. This work, alongside much of the work I shall be building upon, makes use of power spectra in our simulations of the Local Group.

C. This Project and Previous Work

The capabilities of modern computing have allowed for immense advancements in scientific simulations; 1998 work on dwarf galaxies^[40] within the Local Group contained fewer than 100 simulations, with resolutions below what we would now deem satisfactory. In 1999, Moore, B., et al. examined dark matter substructure within galactic haloes with similar numerical methods, successfully demonstrating the convergence of dark matter on the galactic and galaxy-cluster scales^[41]. Truly hydrodynamic simulations with baryonic physics were not possible for another decade, often working on star formation and evolution^[42]. Galaxy structure has garnered greater scientific focus in recent years with further improvements to computing power and the release of the BORG algorithm^[43], discussed later in the chapter (section 1.C.1.).

I have previously introduced the scientific supposition of the archetypal nature of our region of the universe. One must question the validity of this assumption, and much research into the Local Group subsequently broaches this topic. It has been shown in that, in many ways, the Local Group is quite extraordinary within the standard model: several studies have found that only between 5 and 10 per cent of Milky Way-sized haloes in Λ CDM simulations contain satellites

as massive as the Large and Small Magellanic Clouds^{[44][45][46][47]}; the observed MW-M31 orbit is uncommonly radial^[48]; and very few simulated Λ CDM haloes have satellite distributions as anisotropic as the LG^{[49][50]}. Of course, there are still limitations to modern simulations which may be skewing results, but this is unlikely to be the sole cause of the trends found thus far; extensions to the Λ CDM model (like WDM) may be necessary, or we may have to accept these quirks of the LG as an unlikely coincidence.

1. The SIBELIUS Project and the BORG Algorithm

The SIBELIUS project^[51] was created to directly approach this problem, and is the foundation upon which my work was built. Its aim was to examine the evolution of the local universe using a much wider set of initial conditions than its counterparts like the HESTIA project^[52], revealing the LG's properties' dependence on its initial density field. My work uses its methods to investigate the effect of changing initial conditions on a much larger scale than SIBELIUS.

SIBELIUS and I both used the following cosmological parameters^[53]: the matter density parameter (of the universe), $\Omega_m = 0.307$; the dark energy density parameter, $\Omega_\Lambda = 0.693$; the hubble parameter, $h = 0.6777$; and a parameter to describe the amplitude of the power spectrum of matter distribution in the universe, $\sigma_8 = 0.8288$.

h is the ratio of the current rate of change of the scale factor (the relative size of the universe, see section 1.B.2.), to the value of the current scale factor; in other words how quickly the universe is expanding relative to its current size. σ_8 is essential to describing the structure of primordial fluctuations, thus a small change would have seen drastic differences in the current universe. The density parameters affect the curvature of space, described by the equation $\Omega_0 = \Omega_m + \Omega_\Lambda$ where $\Omega_0 = \frac{\rho}{\rho_c}$ is the density parameter representing the ratio of the universe's density, ρ , to the critical density, ρ_c . The critical density is the density required for the current universe to be "flat", i.e. for self-gravity to slow expansion to the point where it halts after infinite time. If current density is too great ($\Omega_0 > 1$), the universe is geometrically spherical (parallel lines ultimately converge) and the universe will eventually begin shrinking, and if density is too small ($\Omega_0 < 1$), the universe is geometrically hyperbolic (parallel lines diverge) and will expand forever. It is known that $\Omega_0 \approx 1$, meaning an almost perfectly flat universe, yet there is still uncertainty as to whether it exactly equal to one or lies slightly to one side^[54].

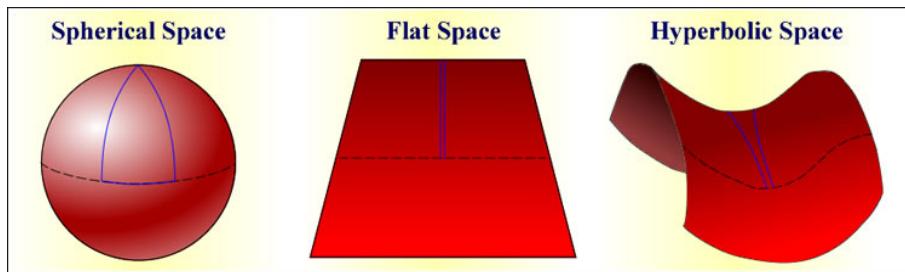


FIG. 9: Potential Geometries of the Universe. Swinburne University of Technology. (2022). This diagram visually represents the three geometric classes of Euclidean space. Travelling through space with each geometry is represented by a line on the shape's surface, showing that locally parallel lines may converge, diverge, or stay parallel depending on the geometry of the space they are contained in.

Using these parameters, we create an initial displacement field using second-order Lagrangian perturbation theory at red-shift, $z = 169$, and compute the displacements (particle locations) using quadratic interpolation, via the IC-GEN algorithm^[55]. The creation of these fields is a multi-step process, beginning with a Gaussian white noise field (GWNF).

A white noise field is a field representing the simultaneous interference of many random waves of equal intensity, giving it a constant power spectral density. If the amplitude of each point in the field has a Gaussian distribution, it is a GWNF, represented in Figure 10.

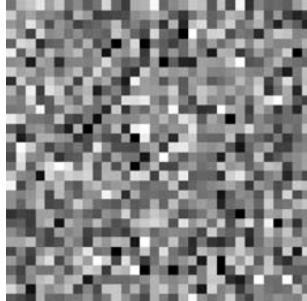


FIG. 10: Gaussian White Noise Field. Grompone von Gioi. (2008). This image of a GWNF shows the Gaussian distribution of colour as greys dominate the image with few whites or blacks, and a clearly random spread.

Our GWNF represents the homogeneous, Gaussian-distributed primordial density fluctuations described previously; and is the product of the BORG algorithm (Bayesian Origin Reconstruction from Galaxies). The BORG algorithm uses a Markov chain to infer the primordial density field from the analysis of over two million galaxies from the 2M++ galaxy redshift catalogue^[56]. Given an arbitrary system, the Markov chain calculates a potential previous state of the system and assigns it a probability of that state evolving into the current one. It does this repeatedly, discarding the least probable states, until it homes in on the most likely initial conditions. Think of this like a tree, with the Markov chain checking each branch, choosing the best branches and checking their sub-branches, and so on, until every sub-sub-...-branch it checks is producing good results. For the purpose of this work, a specific step in the Markov chain will be referred to as the large-scale structure constant (LSS), as the SIBELIUS project found that varying the Markov chain step used affected the formation and evolution of large-scale structure in the local universe.

For most of their work, SIBELIUS took a single value of LSS = 9350, meaning the 9350th step of the Markov chain, as at that time it was the final iteration of the chain and they found it yielded similar results to the true Local Group. My work aims to test the validity of this decision: determining the uniqueness of the initial conditions and their sensitivity to adjustments. To understand the small-scale changes I made to initial conditions we must delve deeper into the methodology of the SIBELIUS project.

2. Octree Functions

Our simulations model the universe as a three dimensional array of cubes, using the S_8 octree orthogonal basis function set laid out in Jenkins (2013)^[57]. Repeated deconstruction of

each cube into eight smaller cubes, half the size in each direction, allows adjustments to be made on incredibly small scales. Each deconstruction is referred to as a layer in the octree. Our simulations are anchored around 15^3 cells at layer 18 of the octree, with a total volume of 1000^3 co-moving Mpc 3 (cMpc 3). This means that the volume of each cell is $\sim 66.67^3$ cMpc 3 , with a minimum cube size of 0.05^3 cMpc 3 at layer 28, the maximum depth of our program.

Octrees are more efficient than the use of Fourier modes when applied to the multi-layer initial conditions required by this project: the octree basis functions are localised, in contrast to Fourier modes, making Fourier modes more computationally expensive when specifying initial conditions at high resolution to only a small region of a large periodic volume.

Reverting back to our GWNF, convolving it (multiplying in Fourier space and taking the Fourier transform) with the power spectrum of the universe, $P(k)$, measured using the CMB (see section 1.B.3.), produces our desired Λ CDM density field of the universe^[58]. To produce an octree representation of this field, we take the Fourier modes of BORG’s 256^3 element GWNF as a set of linear constraints on an unconstrained white noise field produced by octree functions, using methods from the late 20th century^[59]. We now have the template for our initial conditions.

From here, alterations at different layers of the octree can be used to study the formation of the LG using different initial conditions. As demonstrated in Figure 11, the coefficients of the field at layers 1 to 21 are determined by the constraints produced by BORG, with SIBELIUS varying the phase information at layers 22 and above. They initially varied the layers using phase information from different locations in the PANPHASIA Gaussian White Noise Field^[60] (yellow), an extremely large, discrete realisation of a GWNF. From there they used different parts of PANPHASIA for levels 22, 23 and 24 (red), and finally simply randomised subregions of the field at layers 24 and above.

3. SIBELIUS’ Findings

SIBELIUS looked for MW-M31 analogues using varying criteria, and their findings can be summarised in Table 1, representing the number of LG analogues fulfilling the criteria out of 60,000 simulations with variations at the 3.2 cMpc scale.

With no analogues fulfilling the strict criteria, it is clear that my investigation into the validity of LSS = 9350 will be useful.

Their variations at smaller scales led to less pronounced differences between analogues, as one would expect. Interestingly, they found that analogues whose initial conditions only varied on the 0.8 cMpc scale all lay on the same equi-energy (kinetic plus potential) lines, with the value of the sum depending on the specific variation at the 3.2 cMpc scale, as shown in Figure 12.

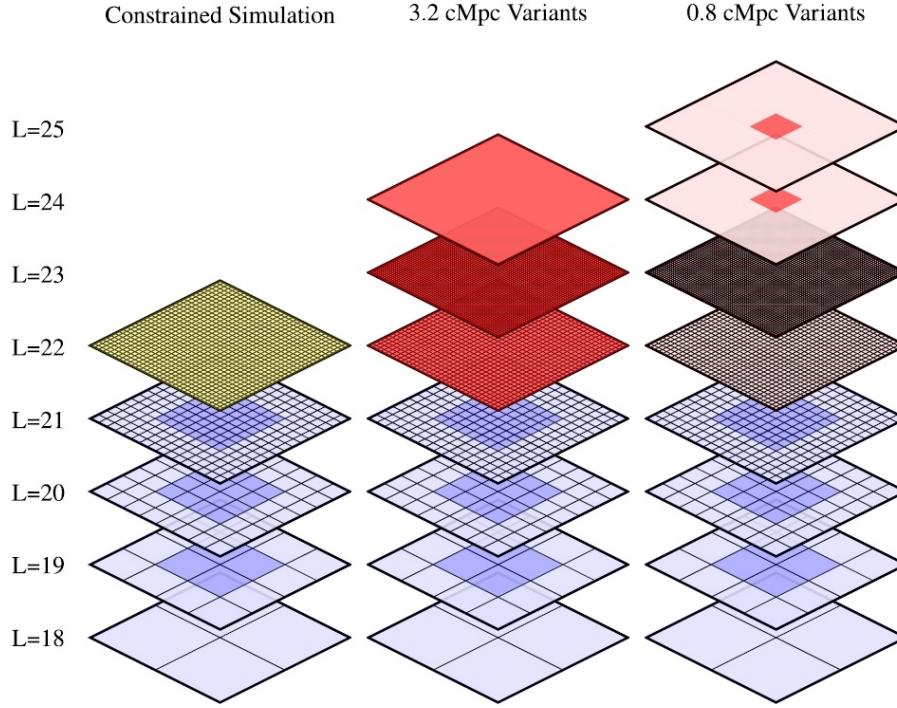


FIG. 11: Octree Representation. Sawala, T. (2022). This diagram represents a two-dimensional view of the octree, demonstrating the bisection of each cube with each layer. The left column represents their initial simulations, varying layer 22 using information from PANPHASIA (Yellow); the middle column represents their simulations using a different region of PANPHASIA for each layer 22 to 24 (red), and the right column represents their randomisation of subregions at layers 24 and above.

	'Loose'	'Intermediate'	'Strict'
$M (10^{12} M_{\odot})$	1.2, ..., 6.0 (46 054)	1.5, ..., 5.0 (39 090)	2.0, ..., 4.0 (24 108)
M31/MW	2/5, ..., 5 (40 490)	2/3, ..., 3 (28 630)	1, ..., 2 (14 162)
$d (\text{Mpc})$	0.5, ..., 1.5 (34 693)	0.6, ..., 1.0 (15 356)	0.74, ..., 0.80 (1665)
$v_r (\text{km s}^{-1})$	-200, ..., 0 (11 414)	-150, ..., -50 (4318)	-109, ..., -99 (597)
$v_t (\text{km s}^{-1})$	< 150 (38 885)	< 100 (23 904)	< 40 (10 792)
$N (M, \text{M31/MW}, d, v_r, v_t)$	6385	489	1
δ	< 45°	< 30°	< 15°
$N (M, \text{M31/MW}, d, v_r, v_t, \delta)$	2309	82	0

TABLE 1: Selection Criteria and Numbers of Successful LG Analogue. Sawala, T. (2022). This shows that a vast majority of the 60,000 simulations produced analogues, however none of them fit the strict criteria. M is total mass; M31/MW is the mass ratio; d is the distance between MW and M31; v_r is the radial velocity; v_t is the tangential velocity; N is number of analogues; and δ is M31's angular separation from its true value.

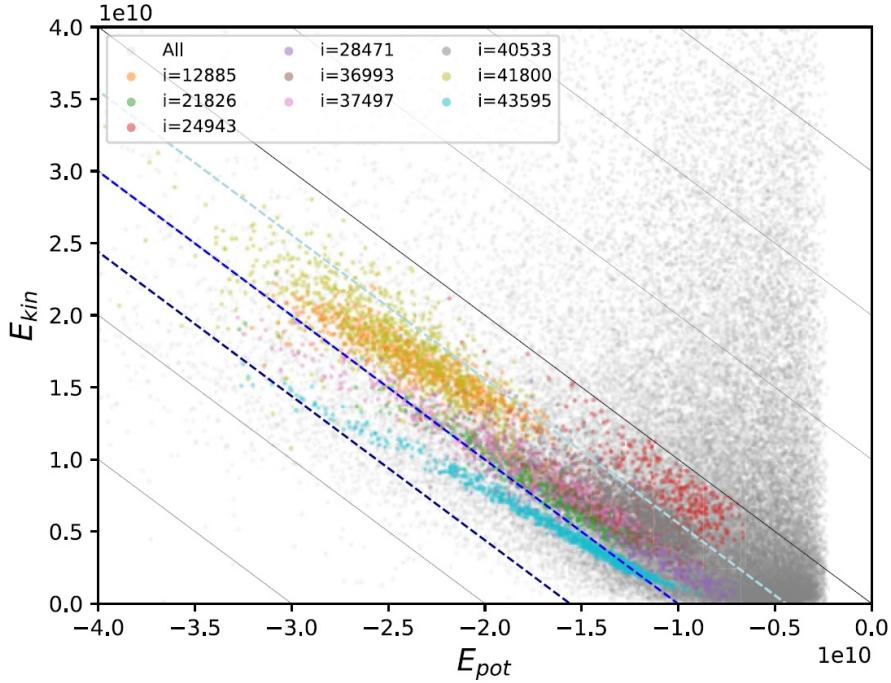


FIG. 12: Energy Distribution of Analogues with Variation at the 0.8 cMpc and 3.2 cMpc Scales. Sawala, T. (2022). Grey dots are the entire set of 0.8 cMpc analogues, with some samples colour co-ordinated by their specific 3.2 cMpc variation, denoted by i . Analogues of identical i can be seen to follow equi-energy lines, with the value of the line varying by i .

4. This Work

As previously explained, I shall be using the methods described so far to investigate whether 9350 is the best value of LSS we currently have, whether it is a different value, or whether no value stands out as particularly special. To do this I shall be randomising the phase information at the 3.2 cMpc scale (layer 22). I shall also search this space for the best LG analogue by scoring them based on their likeness to the true LG, and upon doing so investigate whether this analogue is unique and how sensitive it is to variations in initial conditions.

This is an important area of research as, should a particular LSS stand out as the best option, future simulations will have to cover far less space to achieve reliably good results. If it is found to equal 9350, it also gives added weight to the findings of Sawala, T. et al. (2022). Testing the sensitivity of LG analogues to changes in initial conditions gives us an insight into the delicateness of the true LG's formation, and the probability of it having emerged at all.

2. METHODOLOGY

1. Simulation Code

Much of the initial work of this project was in the development of code to quickly and reliably run the simulations. Unless specifically mentioned, my code is written in Python, C or

Bash. My source code, along with assistance in using it, will happily be provided upon request.

The first step to running a simulation is to adjust the parameter files, provided by A. Jenkins, to the desired values. As the parameter location within the files did not change between runs, I wrote a script to character-search each parameter within the parameter files and return their locations. This is a one-time event, as I then created a code to take new parameters as arguments and place them in these fixed locations; this is the most time-efficient method of editing text files (.inp and others included) at O(n) time complexity, with n being only a few variables.

From here, my parameter files are submitted to the COSMA-5 computer coupled with A. Jenkins' IC-GEN program, returning the initial condition files. These are resubmitted to COSMA-5 attached to GADGET-3 code to run the simulations.

The output of these simulations are sixteen .hdf5 files, containing particle information such as position, mass, and so on, as well as groupings via a nearest-neighbour algorithm. The next step is to test pairs of these groups to find any Local Group analogues. T. Sawala provided a code framework being used for his current research, which I used as inspiration.

My method for searching the groups for analogues was to load them into an array which would remove groups as I applied parameters. The data for each analogue was saved in a dictionary and written to a .pickle file. For groups who only failed on a couple of criteria, the code also output the criteria on which they failed.

From here, efficient automation was a key focus as I could only queue 50 jobs at once on COSMA-5. I therefore wrote a program to automate all of the tasks at the user end, whilst simultaneously checking for completion of the dispatched jobs; this allowed for roughly 80 full simulations per hour, assuming no queue wait times, and for all data collection, plotting and analysis to occur without requiring any user input. Failed simulation runs are also caught by my program and resubmitted at the next opportunity.

File management is not a focus of IC-GEN or GADGET-3, so I wrote a program to store necessary files more efficiently, discarding unused data and folders. This will be useful for future users who have a file-number or memory limit.

2. Local Group Criteria

I ran my program for 138 different LSS values between 7450 and 9800: a spacing of 50 between 7450 and 9100 and between 9600 to 9800; and a spacing of 5 between 9100 and 9600. This was because I wanted to search the area around 9350 more closely, as I believed that if 9350 is indeed better than most, as found by SIBELIUS, good competitors are more likely to be nearby in the Markov chain.

I randomly chose 80 different variants of the 3.2 cMpc phase information to run on each of these LSS values, although I ensured all 80 produced analogues with LSS = 9350. For the purposes of this work, these 80 variants will be known as the small-scale structure constants (SSS) as, compared to LSS, they affect the analogues on a smaller scale. I have indexed the SSS from 1 to 80 but SIBELIUS' full labels for them can be found in appendix A, for those wishing to reproduce this work.

My program initially uses fairly loose criteria to search for analogues, the successors of which are scored based on much stricter values. The loose criteria are:

$3 \times 10^{11} M_{\odot}$	<	masses, M_{MW}, M_{M31}	<	$6 \times 10^{12} M_{\odot}$;
0.5 Mpc	<	separation, d	<	1.3 Mpc;
-150 km s ⁻¹	<	radial velocity, v_r	<	0;
0	<	transverse velocity, v_t	<	150 km s ⁻¹ ;
isolation, s	>	1 Mpc.		

It is also important to note that only groups within 5 Mpc of the centre of the simulation were considered, however this property was not scored as I am more interested in the structure of the LG itself.

These criteria were chosen as they act as broad catchment areas around the true LG values. Isolation is slightly different, however, as it refers to the distance to the nearest other MW-massed galaxy, despite not attempting to reproduce these in the simulations. I set the minimum isolation to 1 Mpc as that is roughly the distance from the the MW-M31 barycentre to the outermost galaxy of the LG, the Sagittarius Dwarf Irregular Galaxy, SagDIG^[61], and thus the edge of the LG. The stricter value for isolation and the other parameters will be discussed later.

When analysing two groups as potential LG analogues, my program allocates them MW and M31 based on their orientation in the sky, implying that the simulated M31's maximum angular separation from the true M31 is $\frac{\pi}{2}$.

3. Scoring Analogues

The successful analogues were then scored on these various criteria. The new values represent a review of current literature, and where literature cannot agree (most notably v_t), I chose a value and uncertainty range to reflect that:

$$\begin{aligned} M_{MW} &= (1.15 \pm 0.35) \times 10^{12} M_{\odot}^{[62][63][64][65]}; \\ M_{M31} &= (1.5 \pm 0.5) \times 10^{12} M_{\odot}^{[66]}; \\ d &= 0.78 \pm 0.13 \text{ Mpc}^{[67]}; \\ v_r &= -109 \pm 5 \text{ km s}^{-1}^{[68]}; \\ v_t &= 40 \pm 40 \text{ km s}^{-1}^{[69][70]}; \\ s &= 3.6 \pm 0.2 \text{ Mpc}^{[71][72]}. \end{aligned}$$

v_t is especially difficult to determine as we have several ways of measuring it, each producing different results. For example, in 2012 R. van der Marel et al. measured it to be 17 ± 17 km s⁻¹ based on Hubble Space Telescope proper-motion measurements; whereas in 2019 he and others used Data Release 2 to measure it as 57_{-31}^{+35} km s⁻¹; having only a small overlap with the first result. For this reason I gave it an exceptionally wide uncertainty region.

s was given a value of 3.6 Mpc as that is the distance to the nearest galaxy (M81) roughly as massive as the MW. The uncertainty was placed at 0.2 Mpc as some measurements of classical Cepheids in the Centaurus A galaxy (NGC 5128) place it as near as 3.4 Mpc. I did not take this to be the isolation as Mira variables typically place it closer to 3.8 Mpc away.

The orientation, o , of each analogue was also scored, using the angle between the real and simulated MW-M31 vectors. M31's true position in the sky is: right ascension, $\alpha \approx \frac{25643\pi}{432000}$ and declination, $\delta \approx \frac{49523\pi}{216000}$. These are values we know with immense precision, forcing me to arbitrarily choose an acceptance region. I chose angular separation, $\theta = \frac{\pi}{12}$, as this would give a slightly larger maximum score than the other criteria, which I deemed necessary as the other

criteria had already been constrained by the initial search whilst orientation had not.

The scoring was done using the following equation:

$$\Sigma_i = \sum_n \left(\max \left(1, \frac{|x_{n,i} - \bar{x}_n|}{\sigma_n} \right) - 1 \right) \quad (4)$$

where Σ_i is the score for the each SSS = i and $x_{n,i}$ is the value of variable x_n who's literature value is \bar{x}_n with uncertainty σ_n . This is except for isolation and orientation. I only gave a non-zero score for isolation values smaller than the literature values, as if a simulation creates a body that far from the LG I do not deem it relevant to my work as it no longer counts as part of the group.

Orientation was scored differently as the angular separation of two random vectors in a sphere is not uniformly distributed (there is a bias towards $\theta = 0, \pi$). This can be easily visualised by considering a vector as a pole of a sphere: take a set of vectors pointing out towards the equator of the sphere and gradually increase their declinations, the vectors will bunch together as the circumference about the pole shrinks. Each angular unit near the poles therefore contains a larger number of vectors than at the equator. This distribution is sinusoidal, therefore it is instead required to score the cosine of the angular separation, making the new equation:

$$\Sigma_o = \max \left(1, \frac{|\cos(\theta) - 1|}{1 - \cos(\sigma_\theta)} \right) \quad (5)$$

This system means that an analogue that falls within all of the strict criteria would receive a total score of zero, a feat not achieved by any of SIBELIUS' analogues, who were scored with a similar system, omitting isolation.

3. DATA AND DISCUSSION

I began by plotting the total number of valid analogues that each LSS produced, shown in Figure 13, to see if there are any trends that may give insight as to the optimal LSS . As you can see, LSS = 9350 gives the largest number of analogues, with a total of 85 of the combined 1195. This was expected as the SSS were chosen to give valid analogues with LSS = 9350. More interestingly is the clear rise in success rate either side of 9350 as well; my idea that Markov chain steps near to a "good" LSS are also more likely to produce valid analogues, seems to be holding true. 85 successful analogues means that some simulations produce two separate LG candidates within the same 5 Mpc-radius sphere, which therefore may receive quite high isolation scores later in this work.

Outside of the peak, the simulations hold a roughly 10% success rate; it would be interesting to test this value against SIBELIUS' full 60,000 simulations with LSS = 9350, using my criteria, as this would give us a definitive idea of whether LSS = 9350 does reliably produce better results than other values. Any perceived periodicity of the success rate between $7500 < \text{LSS} < 9000$ is very likely down to randomness as the variance is too large to imply any true relationship.

I plotted the graphs again with SSS as the independent variable (Figure 14), to look for similar insights, and came across intriguing results. The data follows an approximately Gaussian

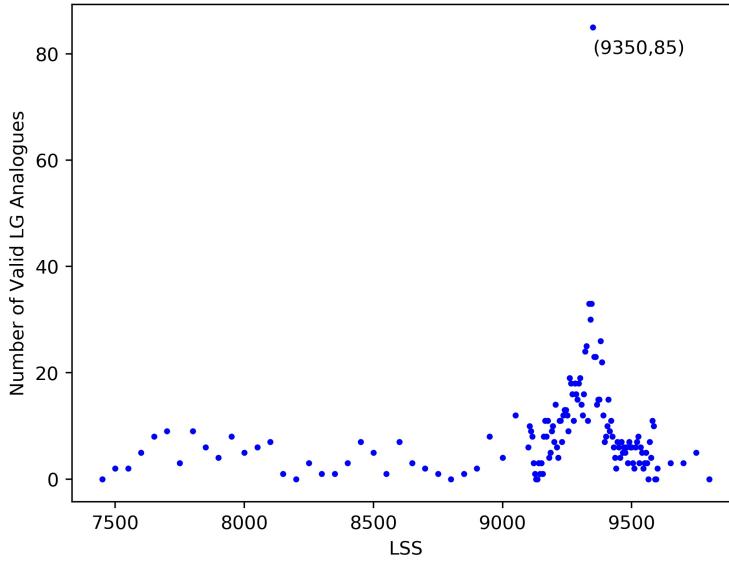


FIG. 13: Success Rate of Each LSS. This graph shows a clear increase in the success rate of simulations around $LSS = 9350$, with a moderately constant rate outside of the central peak.

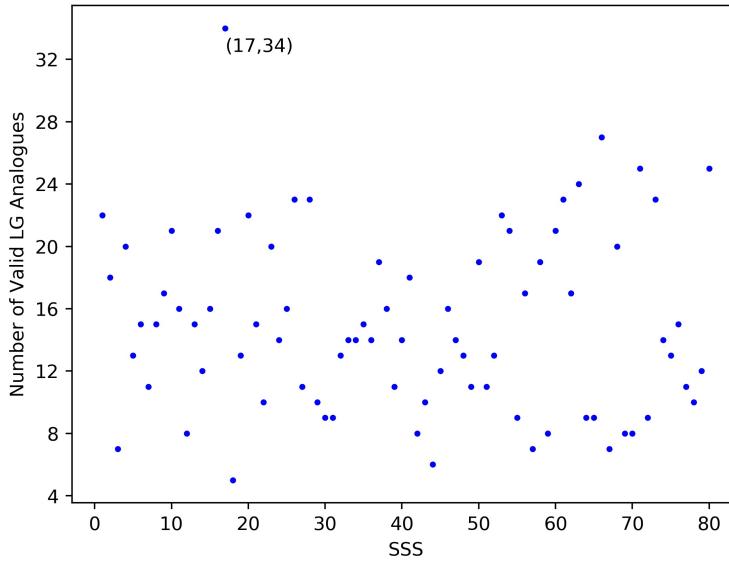


FIG. 14: Success Rate of Each SSS. Analysis deduces that the data follows an approximately Gaussian distribution, with a mean of 14.9 valid LG analogues per SSS. The graph highlights the value of $SSS = 17$, 3.36 standard deviations above the mean.

distribution, as shown in Figure 15, with a mean and standard deviation of 14.9 and 5.68 analogues per SSS respectively. Therefore at 34 analogues, $SSS = 17$ had a far higher success rate than any of the other values: it sits 3.36 standard deviations above the mean giving it a 0.04% probability of having had this many or more analogues. One should not jump to the conclusion that the phase information contained in $SSS = 17$ is particularly good for producing LG analogues, as is not this simple; 29 of the 34 analogues fall within the tall peak, and removing this data region makes it no longer the most successful SSS, let alone the standout value. This means that $SSS = 17$ works immensely well within a narrow band of LSS values, further implying that

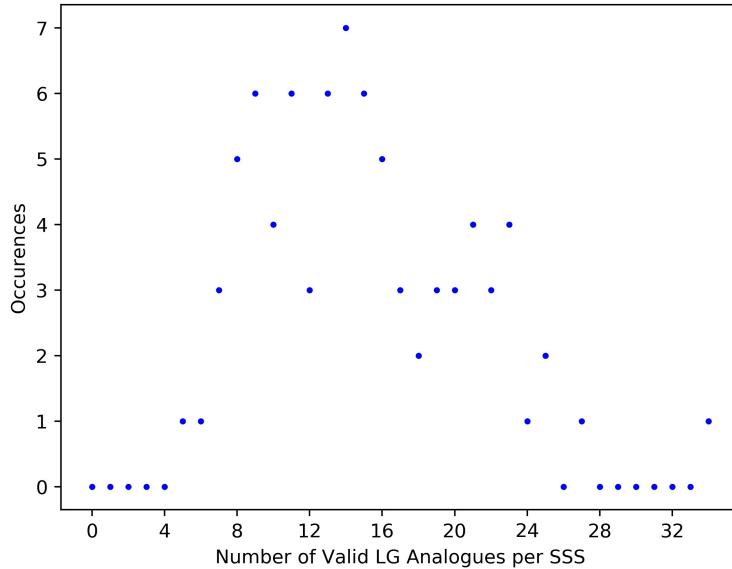


FIG. 15: Distribution of SSS Success Rate. This graph shows the approximately Gaussian distribution, with a mean and standard deviation of 14.9 and 5.68 valid LG analogues per SSS respectively, and again highlights the outlier of SSS = 17 at 34 valid LG analogues.

successful LSS/SSS combinations are not entirely unique: some successful combinations seem to be quite stable, as adjusting the LSS slightly yields very similar results. It is also worth noting that the lack of a trend in the graph is fully expected, as each SSS is random and thus independent of one another.

To search for these regions of stability, I plotted the spread of successful runs as an LSS/SSS map, seen in Figure 16. A stable SSS would appear as a dense horizontal line. As the SSS values are co-independent, stable LSS values will not be shown in the same way, with analogues expected to be randomly scattered along each vertical line.

As you can see there are no distinct horizontal lines; the closest we get are the central portions of SSS = {46, 71, 73} with 10 to 15 near-consecutive analogues. This is not enough evidence to say with certainty that these SSS values are especially stable within this LSS range, however it may be an indication, and repeating this experiment with smaller intervals in the Markov chain may add weight to this finding.

The lack of definitive evidence was unexpected and seems contrary to previous findings regarding SSS = 17. Indeed, the plot shows SSS = 17's analogues to be spread randomly across the peak, rather than on consecutive LSS values as I had previously anticipated. Note that the plot extends down to LSS = 7450, but with no trends in that area I showed only the central peak for visual clarity.

Thus far, solely based on the number of analogues produced there is no clear evidence for or against LSS = 9350 as the optimal value. We have found that some SSS values may work well within a narrow range of LSS values, although the LSS values in this range do not necessarily need to be adjacent (every 5 steps along the Markov chain), for example with SSS = 17. I therefore used the scoring system outlined in section 2.3. to investigate whether LSS = 9350's analogues were superior to those with other LSS values.

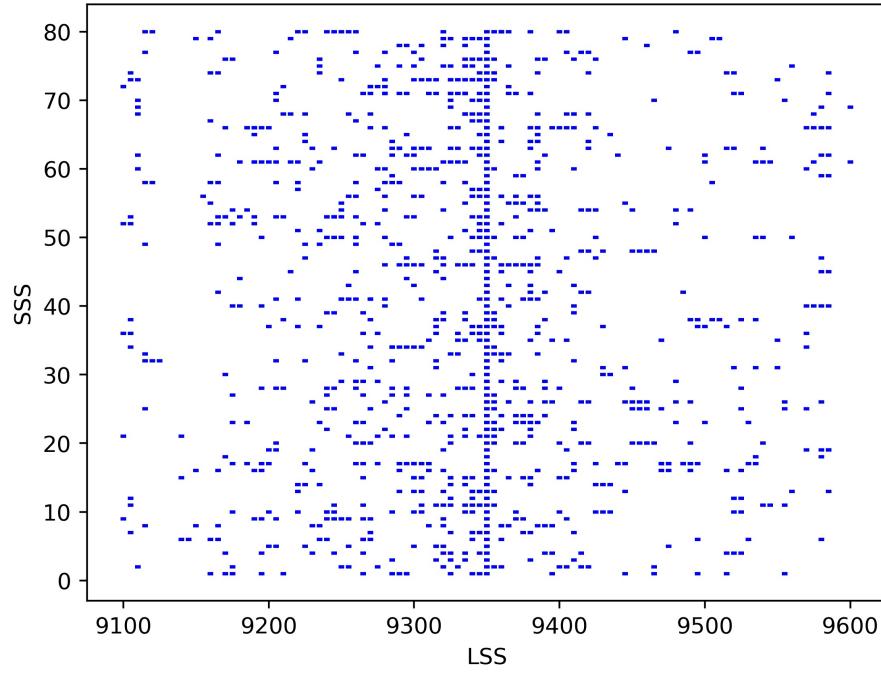


FIG. 16: Analogue Map, displayed as SSS by LSS, for $9100 \leq \text{LSS} \leq 9600$. This graph would show stable SSS values as dense horizontal lines, however this is not seen, implying that all SSS values are sensitive to changes in LSS, contrary to previous evidence.

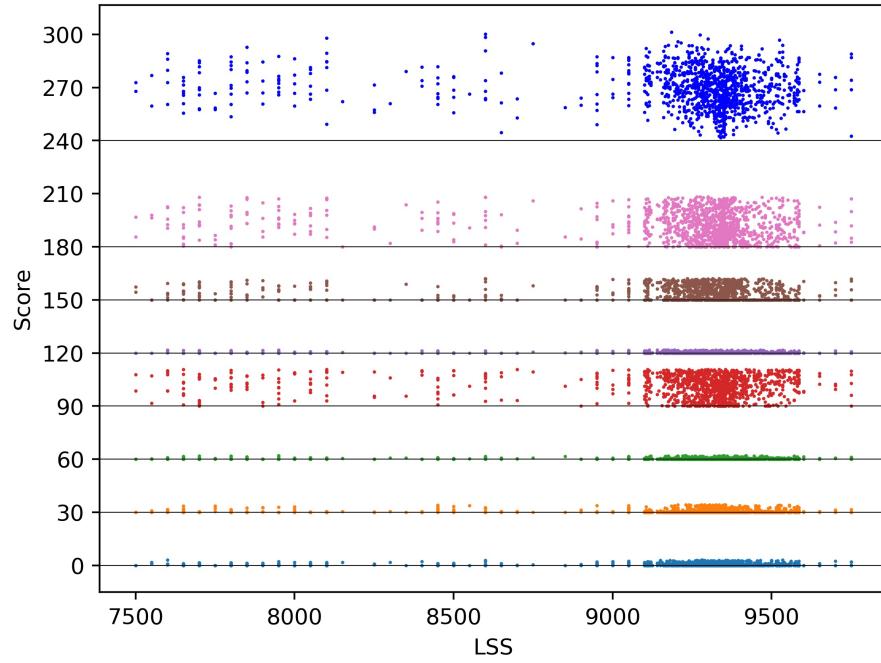


FIG. 17: Full Set of Analogue Scores, Arranged by LSS. Light blue = d ; orange = M_{MW} ; green = M_{M31} ; red = v_r ; purple = v_t ; brown = s ; pink = o ; dark blue = Σ . Large score differences between the criteria are clearly shown, due to different degrees of increased strictness. No clear trends can be seen within the data.

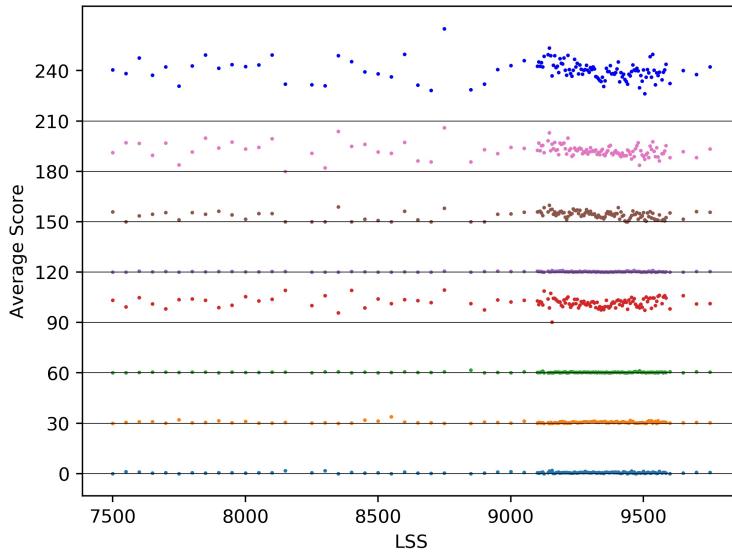


FIG. 18: Full Set of Average Analogue Scores for each LSS. Light blue = d ; orange = M_{MW} ; green = M_{M31} ; red = v_r ; purple = v_t ; brown = s ; pink = o ; dark blue = Σ . No value stands out as having a lower average analogue score - this measure alone in isolation would imply that most LSS are equally good to use in LG simulations.

Figure 17 shows the scores for the entire set of analogues. As you can see, v_r , s and o are significantly higher than the others, due to a greater proportional increase in strictness from their initial screening to their scoring parameters. There are no obvious trends, with high and low scores throughout the set of LSS. I subsequently decided to plot the average scores for each LSS, shown in Figure 18, as this would be a useful measure of how good each LSS value is at producing true-to-life analogues. As the graph shows, there is no standout candidate, again emphasising 9350 as an unremarkable value, with many other values being just as proficient in creating realistic LG analogues.

As before, I reproduced these graphs with SSS as the independent variable (Figure 19), to investigate whether any values of SSS stood out for having especially realistic analogues, subsequently reaching the same conclusion. This was expected as variations on such a small scale are likely to be largely overpowered by the larger variations of the LSS.

Whilst consistently reproducing realistic LG analogues would be incredibly useful for scientific research, so too would the ability to produce a single, true-to-life analogue. I therefore looked into the analogues with the lowest total scores, displayed in Figure 20.

Mirroring the findings of the SIBELIUS project, no analogues achieved a score of zero. LSS = 9340 with SSS = 76, hereby written as (9340, 76), has the lowest score at 1.668, only scoring on M_{MW} and o . This analogue is realised in Figure 21, showing two distinct dark matter haloes of similar masses. M31 clearly has an unusual shape, with an extension of particles in one direction. Shape would be a difficult, but not impossible, property to score and an interesting thing to consider in future research.

Second place (9350, 33) has more spherical haloes and visually looks like a better candidate (Figure 22). LSS = 9350 only appears once in the top ten, however this analogue is incredibly impressive despite trailing (9340, 76) by 0.539 points.

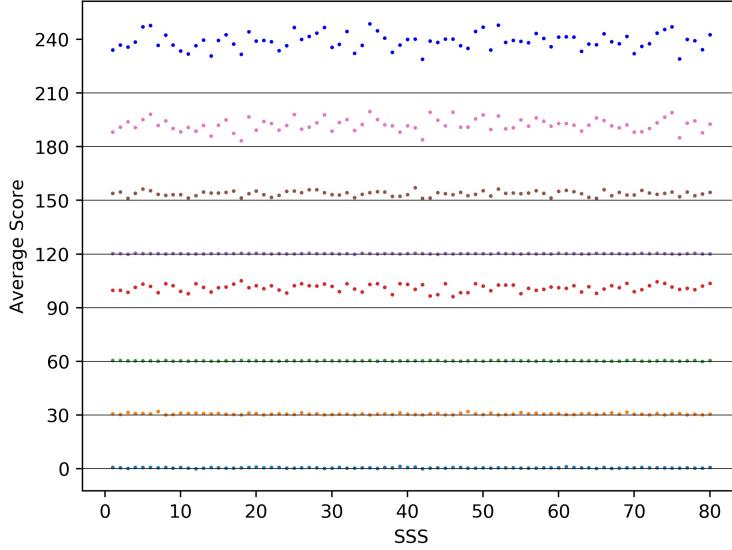


FIG. 19: Full Set of Average Analogue Scores for each SSS. Light blue = d ; orange = M_{MW} ; green = M_{M31} ; red = v_r ; purple = v_t ; brown = s ; pink = o ; dark blue = Σ . No value stands out as having a lower average analogue score - again implying that each SSS is equally good to use in LG simulations.

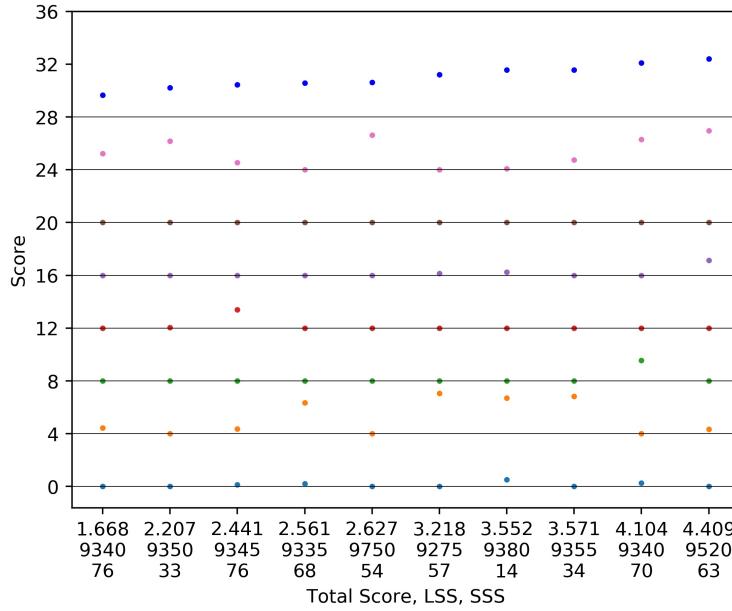


FIG. 20: Lowest 10 Overall Scores. Light blue = d ; orange = M_{MW} ; green = M_{M31} ; red = v_r ; purple = v_t ; brown = s ; pink = o ; dark blue = Σ . LSS = 9340, SSS = 76 has the lowest score, at 1.668, only scoring on M_{MW} and o . LSS = 9350 comes second, but only appears once in the top 10.

One aim of my work was to test whether analogues are unique, or whether they can be reproduced using different initial conditions. To do this, I took these two (best) analogues and readjusted the scoring parameters to exactly match these, whilst keeping the same-sized acceptance region. The score now represents how similar analogues are to these chosen two. For those who wish to use this method, note that the location of the "correct" M31 has changed relative to the "correct" MW. You must therefore reassess the orientation of each analogue,

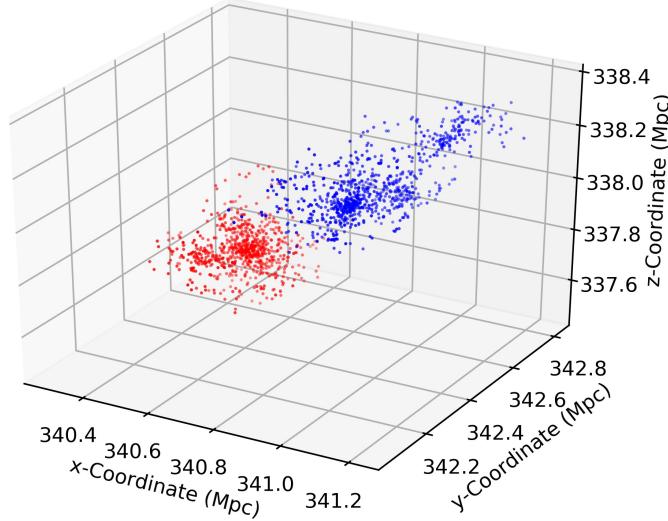


FIG. 21: Three-Dimensional Realisation of LSS = 9340 with SSS = 76. Red = Milky Way; Blue = M31. Each halo is seen to be distinct clusters of particles of similar masses, with M31 appearing to have an large extension of particles into the upper-right corner of the graph.

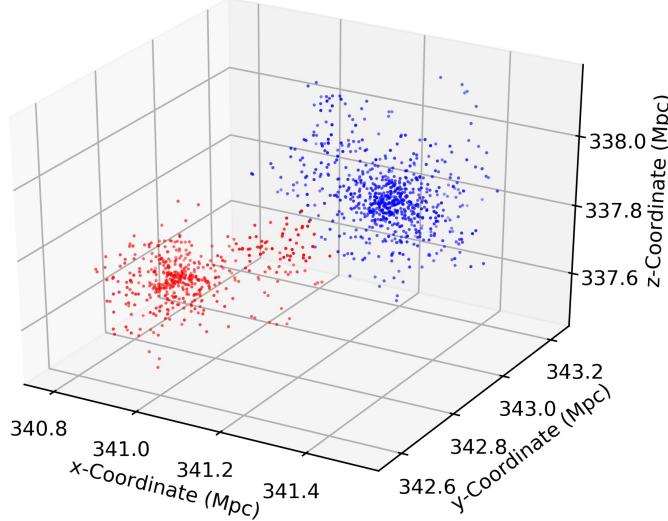


FIG. 22: Three-Dimensional Realisation of LSS = 9350 with SSS = 33. Red = Milky Way; Blue = M31. Each halo is again a distinct cluster of particles, with the MW clearly slightly smaller. On this occasion both haloes have a near spherical shape.

given the new parameters, and reverse the MW-M31 labels if the angular difference is greater than $\frac{\pi}{2}$.

The average score for each LSS did not significantly change for both new sets of parameters, but the set of lowest scores was very interesting (Figures 23 and 24).

These figures show that analogues using these parameters have significantly lower minimum scores than they did using the correct LG parameters (0.744 and 0.617 respectively), however (9340, 76)'s scores are lower across the top 10 whereas (9350, 33)'s are typically higher. This

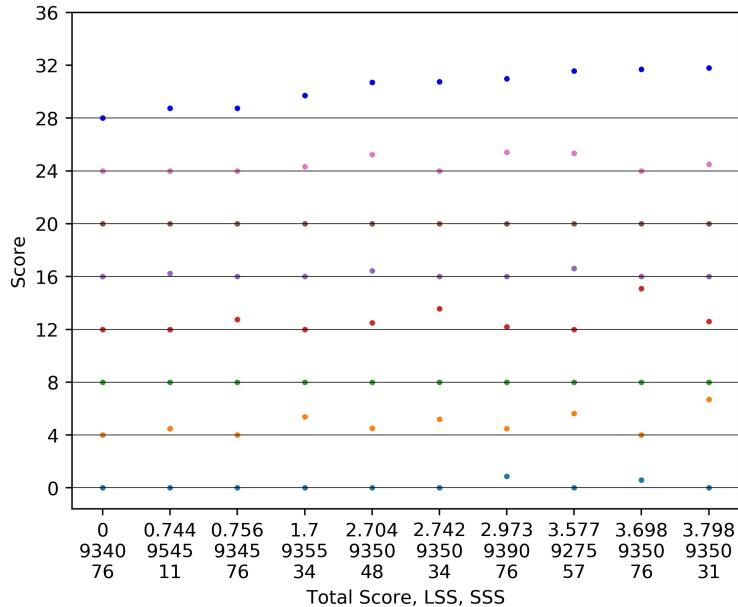


FIG. 23: Lowest 10 Overall (9340, 76) Similarity Scores. Light blue = d ; orange = M_{MW} ; green = M_{M31} ; red = v_r ; purple = v_t ; brown = s ; pink = o ; dark blue = Σ . (9545, 11) and (9345, 76) have significantly lower scores than the other analogues, at ~ 0.75 , however scores across the top 10 are lower than those using the correct LG parameters.

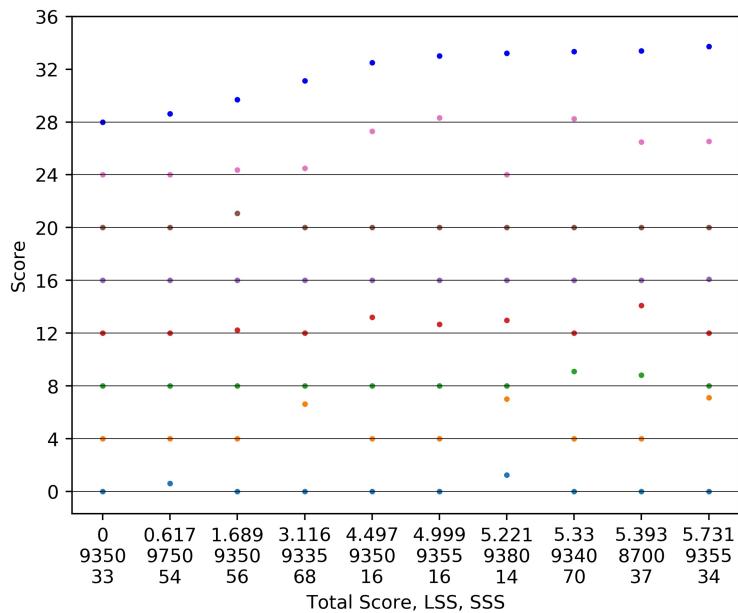


FIG. 24: Lowest 10 Overall (9350, 33) Similarity Scores. Light blue = d ; orange = M_{MW} ; green = M_{M31} ; red = v_r ; purple = v_t ; brown = s ; pink = o ; dark blue = Σ . At a score of 0.617, (9750, 54) has the lowest score seen in any analysis so far, making it immensely similar to (9350, 33), differing only on d . Scores across the top 10 are actually higher than those using the correct LG parameters, however, making this pair fairly unique.

makes both analogues less unique than the real LG, however (9350, 33)'s higher top 10 makes places it in a "niche" set of parameters. We previously found that the average scores across the entire set of LSS did not change and, combined with the vastly lower minimum scores, we must therefore determine (9340, 76) and (9350, 33) to be less unique than the LG.

Finding the LG to be more unique than it's closest analogues allows us to form important conclusions. Simulations using similar large-scale initial conditions (despite added low-level randomness) will always be inherently linked - producing analogues with more in common than those of a truly random output. It also tells us that the LG's properties are indeed fairly difficult to reproduce, hinting at a seeming-uniqueness to our nearby universe, an idea to which astronomers are (rightfully) skeptical.

To culminate my research, I followed in SIBELIUS' footsteps and investigated energy distribution in analogues. They found that analogues sharing the same SSS, with variations at smaller-scales, all lay on the same equi-energy (kinetic + potential) lines. I wanted to know whether this would up-scale to analogues sharing the same LSS with variations at the SSS level. As you can see in Figure 25, analogues were generally potential energy-dominant, with a clear upwards-left trend, implying some degree of uniformity of total energy across the entire set of analogues (as one would expect from the use of mass, distance and velocity parameters). From Figure 26, it appears that analogues sharing an LSS value do not all lie on the same equi-energy line, as the data seems to follow a shallower gradient; this is not certain, however and would greatly benefit from a larger sample size. Should this be the case it is not unexpected, as the initial conditions differ on a much larger scale to the SIBELIUS project's, leading to greater variation as a result.

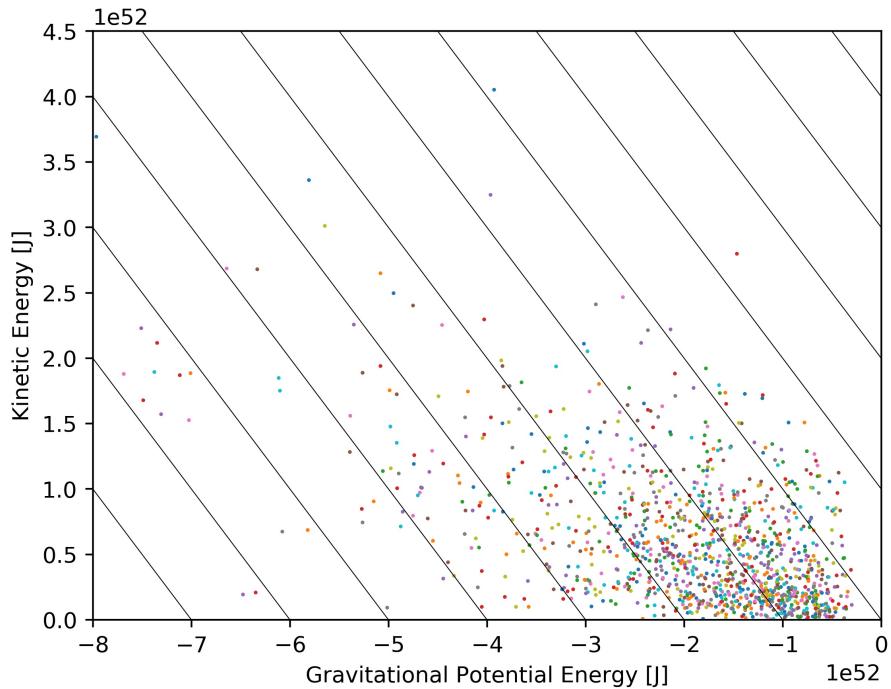


FIG. 25: Energy Distribution for All Analogues. Diagonal lines represent lines of equi-energy. The analogues can be seen to be generally potential energy-dominant, stretching towards the kinetic energy corner of the graph, implying a degree of total-energy uniformity.

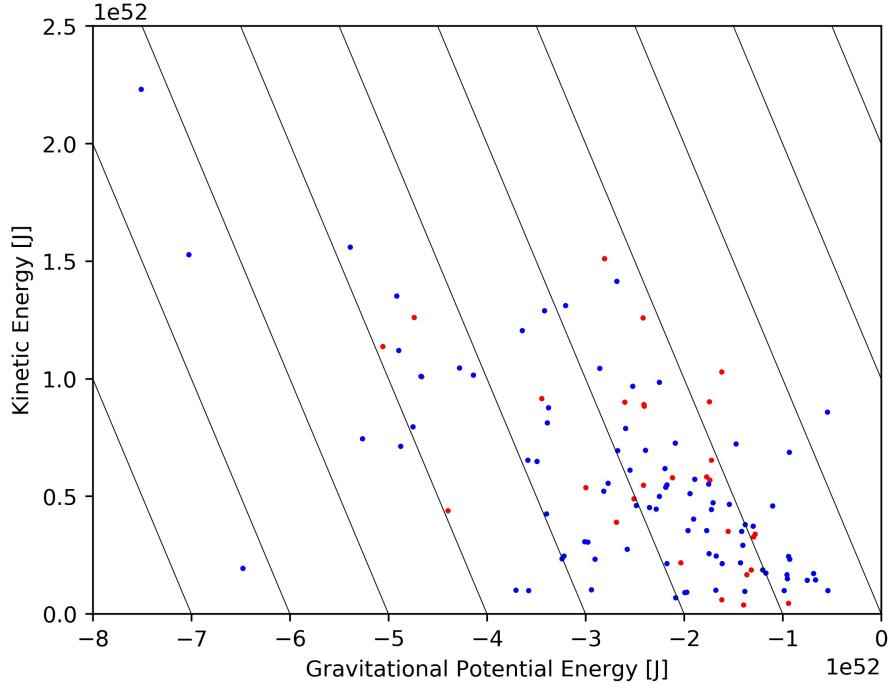


FIG. 26: Energy Distribution for LSS = 9340 (Red) and 9350 (Blue). Diagonal lines represent lines of equi-energy. Both sets of data seem to lie at a shallower gradient than the equi-energy lines, however this is unclear with the large amount of randomness and small sample size.

4. CONCLUSIONS

This project has made significant progress on the work of the SIBELIUS project. The amalgamation of their research into small-scale variations of initial conditions, coupled with my work on broader changes via movement through BORG’s Markov chain, will guide future research aiming to produce Local Group dark matter analogues.

I have built a program to combine A. Jenkins’ IC-GEN code and GADGET-3 into a fluid and storage-efficient process of automated analogue production and analysis. I also developed an analogue-scoring system laid out by SIBELIUS to more closely reflect the LG, introducing an isolation score and narrowing acceptance regions based on current literature.

I found that simulated dark matter structures who have initial conditions close on BORG’s Markov chain (LSS value) to an LG-analogous structure are more likely to be analogous themselves, given the same low-level variations (SSS value). This can lead to specific SSS values forming a large number of LG analogues within a narrow LSS band.

The value of LSS = 9350 used by the SIBELIUS project was shown to be unremarkable, with the average LG-similarity-score being consistent with the other LSS values. One of its analogues did have the second-best score, however, and was visually superior to the first-placed analogue in terms of the expected structure of dark matter haloes. With no LSS value presenting itself to be better in any way, I have no reason to recommend against the use of LSS = 9350 for future simulations.

Taking the two best-scoring analogues, I found them to be less unique than the true LG. This was done by placing each analogue’s parameters within the same scoring system, and

discovering that the average score remains approximately equal, whilst the best score improves drastically; in other words both of these analogues have near-doppelgangers. This tells us that simulations with similar initial conditions will produce non-random results, sharing properties to some degree; and that the Local Group is fairly difficult to reproduce, forcing us to consider whether it perhaps is indeed special.

My work concluded with a look into the energy distribution of my analogues. Most of each analogue's energy was stored in gravitational potential, and analogues did not seem to follow kinetic-gravitational potential equi-energy lines as LSS was varied. This is of note as SIBELIUS found analogues to follow these lines as SSS was varied.

Further research could include repeating my methods with smaller spacing between the LSS values, however I believe that more revealing work can be done. Using the methods laid out by SIBELIUS, one may take the best analogues from my scoring algorithm and vary their initial conditions on smaller and smaller scales, guiding them towards the perfect LG analogue, and indeed into perfect doppelgangers of each other. This will be a true test of the uniqueness of the Local Group and its initial conditions, hoping to providing insight into the galaxies around us, and where they came from.

Acknowledgments

I would like to thank Prof. Adrian Jenkins and Dr Till Sawala for making this project possible. Being given the opportunity to make use of cutting-edge simulations was wonderful, and it is thanks to Adrian's guidance and expertise that I was able to undertake work of this magnitude. Till's contribution to the work was invaluable, providing code and conceptual assistance whenever needed. It was a pleasure to work with them both and an honour to be able to build upon their previous work.

References

- [1] Sawala T.; McAlpine S.; Jasche J.; et al. (2022) "The SIBELIUS Project: E Pluribus Unum". *Monthly Notices of the Royal Astronomical Society*, Vol. 509, Iss. 1, p. 1432–1446.
- [2] Springel, V. (2005). "The cosmological simulation code GADGET-2". *Monthly Notices of the Royal Astronomical Society*, Vol. 364, Iss. 4, p. 1105–1134.
- [3] Jenkins, A. (2010). "Second-order Lagrangian perturbation theory initial conditions for resimulations". *Monthly Notices of the Royal Astronomical Society*, Vol. 403, Iss. 4, p. 1859–1872.
- [4] Jasche, J.; Lavaux, G. (2019). "Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe". *Astronomy and Astrophysics*, Vol. 625, No. A64.
- [5] Masetti, M. (2015). "How Many Stars in the Milky Way?". *NASA Blueshift*.
- [6] Brennan, P. (2019). "Our Milky Way Galaxy: How Big is Space". *NASA Exoplanet Exploration*.
- [7] Peñarrubia, J.; et al. (2014). "A dynamical model of the local cosmic expansion". *Monthly Notices of the Royal Astronomical Society*, Vol. 443, Iss. 3, p. 2204–2222.
- [8] Conselice, C.; Wilkinson, A.; (2016) "The Evolution of Galaxy Number Density at $z > 8$ and Its Implications". *The Astrophysical Journal*, Vol. 830, No. 2.

- [9] Hafez, I. (2010). "Abd al-Rahman al-Sufi and his book of the fixed stars: a journey of re-discovery" (Ph.D. Thesis). *James Cook University*.
- [10] Hubble, E. P. (1925). "NGC 6822, a remote stellar system." *The Astrophysical Journal*, Vol. 62, p. 409–433.
- [11] Hubble, E. P. (1927). "The classification of spiral nebulae" *The Observatory*, Vol. 50, p. 276–281.
- [12] Hubble, E. P. (1929). "A spiral nebula as a stellar system, Messier 31". *The Astrophysical Journal*, Vol. 69, p. 103–158.
- [13] Sohn, S. T.; et al. (2012). "The M31 velocity vector. I. Hubble Space Telescope proper-motion measurements". *The Astrophysical Journal*, Vol. 753, No. 1.
- [14] Kelvin, Lord W. (1904). "Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light". *The Baltimore Lectures*. C.J. Clay and Sons.
- [15] Rubin, V. C.; Ford, W. K. Jr. (1970). "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". *The Astrophysical Journal*, Vol. 159, p. 379.
- [16] Bolles D. (2022). "NASA Science Universe – Dark Energy, Dark Matter". *NASA Science*.
- [17] Bosma, A. (1978). "The Distribution and Kinematics of Neutral Hydrogen in Spiral Galaxies of Various Morphological Types" (PhD Thesis). *Groningen University*.
- [18] Navarro, J.F.; Frenk, C.S.; White, S.D.M. (1996). "The Structure of Cold Dark Matter Halos". *The Astrophysical Journal*, Vol. 462, p. 563.
- [19] Jing, Y. P. (2000). "The Density Profile of Equilibrium and Nonequilibrium Dark Matter Halos". *The Astrophysical Journal*, Vol. 535, p. 30–36.
- [20] Einasto, J. (1965). "Kinematics and dynamics of stellar systems". *Trudy Inst. Astrofiz. Alma-Ata*, Vol. 5.
- [21] Jungman, G.; Kamionkowski, M.; Griest, K. (1996). "Supersymmetric dark matter". *Physics Reports*, Vol. 267, Iss. 5–6, p. 195–373.
- [22] Sparke, L.; Gallagher, III, J. (2007). "Galaxies in The Universe". *Cambridge: University of Cambridge*. pp. 295–296.
- [23] Turner, M. S. (1999). "Cosmological parameters". *AIP Conference Proceedings*, Vol. 478.
- [24] Copi, C.J.; Schramm, D.N.; Turner, M.S. (1995). "Big-Bang Nucleosynthesis and the Baryon Density of the Universe". *Science*, Vol. 267, p. 192–199.
- [25] Craig, N. (2013). "The State of Supersymmetry after Run I of the LHC". *Princeton, Inst. Advanced Study Lectures*.
- [26] van den Bosch, F. (2020). "Theory of Galaxy Formation". *Yale University Lecture Series*.
- [27] Frenk, C. S. (1994). "Neutrinos and the Dark Matter Problem". *The Royal Society*, Vol. 346, No. 1678.
- [28] Einasto, M.; et al. (2007). "The richest superclusters. I. Morphology". *Astronomy and Astrophysics*, Vol. 476, No. 2, p. 697–711.
- [29] Cowen, R. (1996). "Tracing the Architecture of Dark Matter". *Science News*, Vol. 149, p. 87.
- [30] Garrett, K. (2010). "Dark matter: A primer". *Advances in Astronomy*, Vol. 2011, p. 1–22.
- [31] Merle, A. (2013). "keV neutrino model building". *International Journal of Modern Physics D*, Vol. 22, No. 10.
- [32] Moroi, T.; Murayama, H.; Yamaguchi, M. (1993). "Cosmological constraints on the light stable gravitino". *Physics Letters B*, Vol. 303, Iss. 3–4, p. 289–294.

- [33] Okada, N.; Seto, O. (2005). "Brane world cosmological solution to the gravitino problem". *Physical Review D*, Vol. 71, Iss. 2, Code: 023517.
- [34] de Gouvea, A.; Moroi, T.; Murayama, H. (1997). "Cosmology of supersymmetric models with low-energy gauge mediation". *Physical Review D*, Vol. 56, Iss. 2, p. 1281–1299.
- [35] Khetan, N. (2021). "A new measurement of the Hubble constant using Type Ia supernovae calibrated with surface brightness fluctuations". *Astronomy and Astrophysics*, Vol. 647, No. A72.
- [36] Planck Collaboration. (2018). "Planck 2018 results. VI. Cosmological parameters". *Astronomy and Astrophysics*, Vol. 641, No. A6.
- [37] de Bernardis, P.; et al. (2000). "A Flat Universe from High-Resolution Maps of the Cosmic Microwave Background Radiation". *Nature*, Vol. 404, p. 955–959.
- [38] Hu, W. (2022). "Baryons and Inertia". *Department of Astronomy and Astrophysics, University of Chicago*.
- [39] Hu, W. (2022). "Radiation Driving Force". *Department of Astronomy and Astrophysics, University of Chicago*.
- [40] Mateo, M. (1998). "Dwarf Galaxies of the Local Group". *Annual Review of Astronomy and Astrophysics*, Vol. 36, p. 435 – 506.
- [41] Moore, B.; et al. (1999). "Dark Matter Substructure within Galactic Halos". *The Astrophysical Journal*, Vol. 524, No. 1.
- [42] Okamoto, T.; et al. (2008). "Mass loss of galaxies due to an ultraviolet background". *Monthly Notices of the Royal Astronomical Society*, Vol. 390, Iss. 3, p. 920–928.
- [43] Jasche, J.; Lavaux, G. (2019). "Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe". *Astronomy and Astrophysics*, Vol. 625, No. A64.
- [44] Benson, A. J.; et al. (2002). "The effects of photoionization on galaxy formation – I. Model and results at z=0". *Monthly Notices of the Royal Astronomical Society*, Vol. 333, Iss. 1, p. 156–176.
- [45] Boylan-Kolchin, M.; et al. (2011). "Too big to fail? The puzzling darkness of massive Milky Way subhaloes". *Monthly Notices of the Royal Astronomical Society: Letters*, Vol. 415, Iss. 1, p. L40–L44.
- [46] Tollerud, E. J.; et al. (2011). "Small-Scale Structure in the Sloan Digital Sky Survey and Lambda-CDM". *The Astrophysical Journal*, Vol. 738, No. 1.
- [47] Robotham, A. S. G.; et al. (2012). "Galaxy and Mass Assembly (GAMA): in search of Milky Way Magellanic Cloud analogues". *Monthly Notices of the Royal Astronomical Society*, Vol. 424, Iss. 2, p. 1448–1453.
- [48] Fattahi, A.; et al. (2016). "The APOSTLE project: Local Group kinematic mass constraints and simulation candidate selection". *Monthly Notices of the Royal Astronomical Society*, Vol. 457, Iss. 1, p. 844–856.
- [49] Ibata, R. A.; et al. (2013). "A vast, thin plane of corotating dwarf galaxies orbiting the Andromeda galaxy". *Nature*, Vol. 493, p. 62–65.
- [50] Pawłowski, M. S.; et al. (2015). "On the Persistence of Two Small-Scale Problems in Lambda-CDM". *The Astrophysical Journal*, Vol. 815, No. 1.
- [51] Sawala T.; McAlpine S.; Jasche J.; et al. (2022) "The SIBELIUS Project: E Pluribus Unum". *Monthly Notices of the Royal Astronomical Society*, Vol. 509, Iss. 1, p. 1432–1446.
- [52] Libeskind, N.; Carlesi, E.; Grand R.; et al. (2020). "The hestia project: Simulations of the Local

- Group”. *Monthly Notices of the Royal Astronomical Society*, Vol. 498.
- [53] Planck Collaboration. (2014). “Planck 2013 results. XVI. Cosmological parameters.” *Astronomy and Astrophysics*, Vol. 571, No. A16.
- [54] Swinburne University of Technology. (2022). “Density Parameter”. *Swinburne University of Technology*.
- [55] Jenkins, A. (2010). “Second-order Lagrangian perturbation theory initial conditions for resimulations”. *Monthly Notices of the Royal Astronomical Society*, Vol. 403, Iss. 4, p. 1859–1872.
- [56] Lavaux, G.; Hudson, M. J. (2011). “The 2M++ galaxy redshift catalogue”. *Monthly Notices of the Royal Astronomical Society*, Vol. 416, Iss. 4, p. 2840–2856.
- [57] Jenkins, A. (2013). “A new way of setting the phases for cosmological multiscale Gaussian initial conditions”. *Monthly Notices of the Royal Astronomical Society*, Vol. 434, Iss. 3, 21 p. 2094–2120.
- [58] Salmon, J. (1996). “Generation of Correlated and Constrained Gaussian Stochastic Processes for N-Body Simulations”. *The Astrophysical Journal*, Vol. 460, p. 59.
- [59] Salmon, J. (1996). “Generation of Correlated and Constrained Gaussian Stochastic Processes for N-Body Simulations”. *The Astrophysical Journal*, Vol. 460, p. 59.
- [60] Jenkins, A. (2013). “A new way of setting the phases for cosmological multiscale Gaussian initial conditions”. *Monthly Notices of the Royal Astronomical Society*, Vol. 434, Iss. 3, 21 p. 2094–2120.
- [61] Karachentsev, I.D.; Karachentseva, V.E.; Hutchmeier, W.K.; Makarov, D.I. (2004). ”A Catalog of Neighboring Galaxies”. *The Astronomical Journal*, Vol. 127, Iss. 4, p. 2031–2068.
- [62] McMillan, P. J. (2011). “Mass models of the Milky Way”. *Monthly Notices of the Royal Astronomical Society*, Vol. 414, Iss. 3, p. 2446–2457.
- [63] McMillan, P. J. (2016). “The mass distribution and gravitational potential of the Milky Way”. *Monthly Notices of the Royal Astronomical Society*, Vol. 465, Iss. 1, p. 76–94.
- [64] Kafle, P. R.; et al. (2012). “Kinematics of the Stellar Halo and the Mass Distribution of the Milky Way Using Blue Horizontal Branch Stars”. *The Astrophysical Journal*, Vol. 761, No. 2.
- [65] Kafle, P. R.; et al. (2014). “On the Shoulders of Giants: Properties of the Stellar Halo and the Milky Way Mass Distribution”. *The Astrophysical Journal*, Vol. 794, No. 1.
- [66] Penarrubia, J.; et al. (2014). “A dynamical model of the local cosmic expansion”. *Monthly Notices of the Royal Astronomical Society*, Vol. 443, Iss. 3, p. 2204–2222.
- [67] Riess, A. G.; et al. (2012). “Cepheid Period-Luminosity Relations in the Near-Infrared and the Distance to M31 from the Hubble Space Telescope Wide Field Camera 3”. *The Astrophysical Journal*, Vol. 745, No. 2.
- [68] van der Marel, R. P.; et al. (2012). “The M31 Velocity Vector. II. Radial Orbit Toward the Milky Way and Implied Local Group Mass”. *The Astrophysical Journal*, Vol. 753, No. 1.
- [69] van der Marel, R. P.; et al. (2012). “The M31 Velocity Vector. II. Radial Orbit Toward the Milky Way and Implied Local Group Mass”. *The Astrophysical Journal*, Vol. 753, No. 1.
- [70] van der Marel, R. P.; et al. (2019). “First GAIA Dynamics of the Andromeda System: DR2 Proper Motions, Orbits, and Rotation of M31 and M33”. *The Astrophysical Journal*, Vol. 872, Iss. 1, p. 24.
- [71] Karachentsev, I. D. (2005). ”The Local Group and Other Neighboring Galaxy Groups”. *The Astronomical Journal*, Vol. 129, No. 1.
- [72] Ferrarese, L.; et al. (2007). “The Discovery of Cepheids and a Distance to NGC 5128”. *The Astronomical Journal*, Vol. 654, No. 1.

Appendix A

1 166	2 266	3 330	4 349	5 530	6 902	7 3344	8 4678	9 5752	10 6510
11 6812	12 7263	13 8257	14 8760	15 10560	16 11099	17 11768	18 12917	19 13412	20 13746
21 14233	22 14287	23 15308	24 16171	25 17285	26 17309	27 17620	28 18608	29 19439	30 19565
31 20111	32 20578	33 21085	34 21826	35 22715	36 22875	37 23193	38 23653	39 26075	40 26657
41 28091	42 30284	43 30462	44 30562	45 30594	46 30841	47 30969	48 31471	49 32324	50 32604
51 33307	52 33512	53 33656	54 35475	55 35640	56 37539	57 37954	58 39269	59 39574	60 39621
61 40470	62 43564	63 44390	64 45107	65 45799	66 47439	67 47515	68 49289	69 49553	70 50224
71 50690	72 50952	73 51364	74 51825	75 51975	76 52160	77 52511	78 53094	79 53286	80 53647

Table 2: SIBELIUS Labels for SSS Indices.