“The names and numbers carved on the ring surrounding the installation on the waterfront – Greeting to the Sun – are part of the St. Grisogonus Calendar, developed in Zadar and found in 1964 in the Bodleian Library in Oxford. It dates from 1292 or 1293, and is among the oldest of such documents in the world, and possibly the first to have astronomy data written in Arabic numbers. Besides the calendar with the feast days and names of saints, it also has the astronomy part which shows the sun ephemeride, the coordinates of the heavenly bodies, their angle distances from determined immovable flat surfaces, straight lines or points.”
3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Galileo, 1600

Observatory Tower, Lincolnshire, UK, c. 1300

Reber’s Radio Telescope, 1937

NASA/Explorer 7 (Space-based Observing) 1959

“Virtual Observatories” 21st century

“The Internet”

The “Scientific Revolution”

Long-distance remote-control/“robotic” telescopes 1990s
Stjerneborg
(Tycho Brahe, 1586)

Galileo: 1610

W.H. Keck Observatory
(1995+)

Full-sky virtual astronomy:
c. 2023?
SEAMLESS ASTRONOMY
Alyssa A. Goodman, Harvard-Smithsonian Center for Astrophysics

with
Alberto Accomazzi, Douglas Burke, Raffaele D’Abrusco, Rahul Davé, Christopher Erdmann, Pepi Fabbiano, Edwin Henneken, Jay Luker, Gus Muench, Michael Kurtz, Max Lu, Victoria Mittelbach, Alberto Pepe, Arnold Rots (Harvard-Smithsonian CfA); Mercè Crosas (Harvard Institute for Quantitative Social Science; Christine Borgman (UCLA); Jonathan Fay & Curtis Wong (Microsoft Research); Alberto Conti (Space Telescope Science Institute)

projects.iq.harvard.edu/seamlessastronomy
The (US) Backstory


NVO senior personnel:
Charles Alcock, University of Pennsylvania Kirk Borne, Astronomical Data Center/Raytheon
Tim Cornwell, NSF National Radio Astronomy Observatory
Giuseppina Fabbiano, Smithsonian Astrophysical Observatory
Alyssa Goodman, Harvard University
Jim Gray, Microsoft Research
Robert Hanisch, Space Telescope Science Institute
George Helou, NASA Infrared Processing and Analysis Center
Giuseppe Fabbiano, Smithsonian Astrophysical Observatory
Miron Livny, University of Wisconsin, Madison
Carol Lonsdale, NASA Infrared Processing and Analysis Center
Tom McGlynn, GSFC/HEASARC/USRA
Reagan Moore, San Diego Supercomputer Center
Ray Plante, Naval Observatory, Flagstaff Station
Thomas Prince, California Institute of Technology
STScI Nicholas White, NASA Goddard Space Flight Center

Management and Operation of the Virtual Astronomical Observatory

CONTACTS

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nigel Sharp</td>
<td><a href="mailto:nsharp@nsf.gov">nsharp@nsf.gov</a></td>
</tr>
<tr>
<td>Eileen D. Friel</td>
<td><a href="mailto:efriel@nsf.gov">efriel@nsf.gov</a></td>
</tr>
</tbody>
</table>

Program Guidelines

Solicitation 08-537

Please be advised that the NSF Proposal & Award Policies & Procedures (PAPP) includes revised guidelines to implement the mentoring pro-America COMPETES Act (ACA) (Pub. L. No. 110-69, Aug. 9, 2007.) specified in the ACA, each proposal that requests funding to support postdoctoral researchers must include a description of the mentoring that will be provided for such individuals. Proposals that do not comply with this requirement will be returned without review (see the PAPP Grant Proposal Guide Chapter II for further information about the implementation of this new requirement).
and meanwhile...

Welcome to AstroGrid

AstroGrid is the doorway to the Virtual Observatory (VO). We provide a suite of tools to enable astronomers to explore and bookmark resources from around the world, find data in VOSpace, query databases, plot and manipulate tables, cross-match catalogues, and to automate sequences of tasks. Tools from other Euro-VO projects are integrated.

NVO Virtual Observatory Software for Astronomers

and meanwhile...

The VO
How?
Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
"Seamless Astronomy" (Tools)

Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
"Seamless Astronomy" (Tools)

**Data**

- TOPCAT
- ds9
- WorldWide Telescope
- DataScope
- "Registries"

**Literature**

- arXiv.org
- NASA
- ASTM
- Astrobetter
- Bblogs, Wikis, etc.
- Astrometry.net
- Flickr

**Disclaimer:** This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
"Seamless Astronomy" (Tools)

Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
SAMP

(Simple Application Messaging Protocol)

link to 12/2010 IVOA recommendation
Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
The ADS is operated by the Smithsonian Astrophysical Observatory under NASA Grant NNX09AB39G
Contact: ads at cfa.harvard.edu or through the feedback form.

ADS Labs/Seamless Astronomy Core Collaboration
A. Accomazzi, A. Goodman, M. Kurtz, R. Davé, J. Luker, G. Muench, A. Pepe
3. 2010ApJ...716L...1A The J = 1-0 Transitions of 12CH+, 13CH+, and 12CD+ Amano, T.

4. 2009ApJ...705L.176S Detection of the Zeeman Effect in the 36 GHz Class I CH3OH Maser Line with the EVLA
Sarma, A. P.; Momjian, E.

11. 2003A&A...412..513B The molecular Zeeman effect and diagnostics of solar and stellar magnetic fields. II. Synthetic Stokes profiles in the Zeeman regime
Bordyugina, S. V.; Solanki, S. K.; Frutiger, C.

12. 2000PASP..112..873W Magnetism in Isolated and Binary White Dwarfs
Wickramasinghe, D. T.; Ferrario, Lilja
“shift-click” on object
click
“Research, Information”

...more data
...or more literature
Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
The Seamless Astronomy Group at the Harvard-Smithsonian Center for Astrophysics brings together astronomers, computer scientists, information scientists, librarians and visualization experts involved in the development of tools and systems to study and enable the next generation of online astronomical research.

Current projects include research on the development of systems that seamlessly integrate scientific data and literature, the semantic interlinking and annotation of scientific resources, the study of the impact of social media and networking sites on scientific dissemination, and the analysis and visualization of astronomical research communities. Visit our project page to find out more.
SEAMLESS ASTRONOMY

Projects

- ADS Labs
- High-D Visualization
- ADS All Sky Survey
- Social Networks
- Astronomy Dataverse
- Collaboration Networks
- Data Citation
- WorldWide Telescope
- Semantic Search
Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
Microsoft® Research

WorldWide Telescope

Experience WWT at worldwidetelescope.org

Seamlessly explore imagery from the best ground and space-based telescopes in the world.

Expert led tours of the Universe.

Control time to study how the night sky changes.

View and compare images from across the electromagnetic spectrum.

Much more than “just” the sky at night! 3D features can take you to other planets, stars & galaxies.

Finder Scope links to Wikipedia, publications, and data, so you can learn more.

Context bar shows items of interest in current field of view.

Context globe shows where you’re looking.

Expert led tours of the Universe

Finder Scope links to Wikipedia, publications, and data, so you can learn more

Control time to study how the night sky changes

View and compare images from across the electromagnetic spectrum

Seamlessly explore imagery from the best ground and space-based telescopes in the world

Much more than “just” the sky at night! 3D features can take you to other planets, stars & galaxies.
Experience WWT at worldwidetelescope.org
ADS All Sky Survey

Faceted Heat Map of Articles on the Sky

Historical Image Layer Extracted from ALL ADS holdings (using astrometry.net)

ADS-CDS-Seamless collaboration

ADS-Seamless-astrometry.net collaboration
Astrotagging = Geotagging

Parallel Session 5. (Aula Magna)
Cathal Hoare and Humphrey Sorensen.
On Automatically Geotagging Archived Images

Astroreferencing = Georeferencing

The New York City Historical GIS Project by Matt Knutzen, Stephen A. Schwarzman Building, Map Division June 13, 2012

In 2010, the National Endowment for the Humanities (NEH) awarded the Lionel Pincus and Princess Firyal Map Division of the New York Public Library a three year grant, the New York City Historical Geographic Information Systems project, building digital cartographic resources from our historical paper map and atlas collections.

The project walks a portion of our New York City map collections through a series of workflow steps outlined in a previous blog post Unbinding the Atlas; in a nutshell, maps are scanned (shooting a high resolution digital image), georectified (a.k.a. warped, rubbersheeted, i.e. aligning pixels on an old map to latitude/longitude on a virtual map), cropped (removing extraneous non-map information from the collar area around a map), and finally digitized (think of this as tracing).

In the proposal we committed to scanning 9,000 maps, but were ultimately funded to image 7,200 maps. Work has proceeded much faster than anticipated however, enabling us to scan and mount 7,799 new maps so far. An additional 9,327 metadata records have been created for related collections such as all of New York City’s zoning maps (a bibliography of such maps can be found at the bottom of this great post, or in this .doc file), dating to 1916, most of our public domain fire insurance atlases of areas outside of the city in New York and New Jersey and our entire run of historical and contemporary New York state topographic maps. If the pace of imaging continues as expected, the project will have funded the digitization of 17,126 historical maps, most concentrated on the five boroughs.

http://www.nypl.org/blog/2012/06/13/nyc-historical-gis-project
INVESTIGATING THE COSMIC-RAY IONIZATION RATE NEAR THE SUPERNOVA REMNANT IC 443 THROUGH H$_3^+$ OBSERVATIONS$^{1,2}$

NICK INDIROLO$^3$, GEOFFREY A. BLAKE$^4$, MIWA GOTO$^5$, TOMONORI USUDA$^3$, TAKESHI OKA$^7$, T. R. GEBALLE$^8$, BRIAN D. FIELD$^9$, BENJAMIN J. McCALL$^{10}$

ABSTRACT

Observational and theoretical evidence suggests that high-energy Galactic cosmic rays are primarily accelerated by supernova remnants. If also true for low-energy cosmic rays, the ionization rate near a supernova remnant should be higher than in the general Galactic interstellar medium (ISM). We have searched for H$_3^+$ absorption features in 6 sight lines which pass through molecular material near IC 443—a well-studied case of a supernova remnant interacting with its surrounding molecular material—and the purpose of inferring the cosmic-ray ionization rate in the region. In 2 of the sight lines (toward ALS 882 and HD 254577) we find large H$_3^+$ column densities, N(H$_3^+$) $\sim 10^{19}$ cm$^{-2}$, and deduce ionization rates of $\xi_{\odot} \sim 2 \times 10^{-15}$ s$^{-1}$, about 5 times larger than inferred toward average diffuse molecular cloud sight lines. However, the 3σ upper limits found for the other 4 sight lines are consistent with typical Galactic values. This wide range of ionization rates in the vicinity of particle acceleration and propagation effects, which predict that the cosmic-ray spectrum and thus ionization rate should vary in and around the remnant. While we cannot determine if the H$_3^+$ absorption arises in post-shock (interior) or pre-shock (exterior) gas, the large inferred ionization rates suggest that IC 443 is in fact accelerating a large population of low-energy cosmic rays. Still, it is unclear whether this population can propagate far enough into the ISM to account for the ionization rate inferred in diffuse Galactic sight lines.

Subject headings: astrochemistry – cosmic rays – ISM: supernova remnants

1. INTRODUCTION

As cosmic rays propagate through the interstellar medium (ISM) they interact with the ambient material. These interactions include excitation and ionization of atomic and molecular species, spallation of nuclei, excitation of neutral states, and the production of neutral pions ($\pi^0$) which decay into gamma-rays. Evidence suggests that Galactic cosmic rays are primarily accelerated by supernova remnants (SNRs) through the process of diffusive shock acceleration (e.g. Drury 1983; Blandford & Eichler 1987), so interstellar clouds in close proximity to an SNR should provide a prime “laboratory” for studying these interactions. IC 443 represents such a case, as portions of the SNR shock are known to be interacting with the neighboring molecular clouds. IC 443 is an intermediate age remnant (about 30,000 yr; Chevalier 1999) located in the Galactic anti-center region (l, b $\approx (189^\circ, -3^\circ)$) at a distance of about 1.5 kpc in the G331.20-0.01 OB1 association (Welsh & Sallmen 2003), and is a particularly well-studied SNR. Figure 1 shows the red image of IC 443 taken during the Second Palomar Observatory Sky Survey. The remnant is composed of subshells A and B; shell A is to the NE—the center at $\alpha = 0^h0^m0.3^s$, $\delta = +23\degree26\arcmin50\arcsec$ (2000.0) is marked by the cross—while shell B is to the SW. Adopting a distance of 1.5 kpc, the radii of subshells A and B are about 7 pc and 11 pc, respectively. Between the subshells is a darker lane that runs across the remnant from the NW to SE. This is a molecular cloud which has been mapped in CO emission (Corbell et al. 1977; Dickman et al. 1992; Zhang et al. 2000), and is known to be in the foreground because it absorbs X-rays emitted by the hot remnant interior (Troja et al. 2006). Aside from this quiescent foreground cloud, observations of the J, K, and L traces of CO also show shocked molecular material coincident with IC 443 (DeNoyer 1979; Huang et al. 1986; Dickman et al. 1992; Wang & Scoville 1992). These shocked molecular clouds first identified by DeNoyer (1979) and Huang et al. (1986) in CO have also been observed in several atomic and molecular species (e.g. White et al. 1987; Burton et al. 1988; van Dishoeck et al. 1992; White 1994; Snell et al. 2005), and are thought to be the result of the expanding SNR interacting with the surrounding ISM. While the outermost of the shocked clumps are coincident with the quiescent gas, it...

> 1 Million Articles, like this one

Table 1

<table>
<thead>
<tr>
<th>N &amp; RA</th>
<th>Sp type</th>
<th>#ref</th>
<th>Position</th>
<th>RA</th>
<th>DEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPOC 1</td>
<td>01:24:50</td>
<td>700.1</td>
<td>1</td>
<td>07:50:00</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 2</td>
<td>01:24:54</td>
<td>700.1</td>
<td>1</td>
<td>04:10:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 3</td>
<td>01:25:19</td>
<td>700.1</td>
<td>1</td>
<td>02:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 4</td>
<td>01:27:55</td>
<td>700.1</td>
<td>1</td>
<td>01:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 5</td>
<td>01:28:00</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 6</td>
<td>01:28:32</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 7</td>
<td>01:28:28</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 8</td>
<td>01:28:27</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 9</td>
<td>01:28:40</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 10</td>
<td>01:28:32</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 11</td>
<td>01:28:50</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 12</td>
<td>01:28:57</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 13</td>
<td>01:29:07</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 14</td>
<td>01:29:10</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 15</td>
<td>01:29:41</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 16</td>
<td>01:29:41</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 17</td>
<td>01:29:27</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 18</td>
<td>01:30:06</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
<tr>
<td>CPOC 19</td>
<td>01:30:11</td>
<td>700.1</td>
<td>1</td>
<td>00:30:50</td>
<td>+10:50:50</td>
</tr>
</tbody>
</table>

Figure 3. Abundance map of the core of AWM 4, with GMRT 610-MHz contours overlaid. Rectangular regions were used to examine the variations in abundance ISM. The white cross marks the position of the radio core.
Literature
Optical images
Non Optical images

PRIMARY DATABASE
ADDITIONAL DATABASE

ASTROMETRIC MEASUREMENT

Astrotagged literature
Astro-referenced images

NASA ARCHIVES
DATA VIEWERS

HISTORICAL DATA LAYER
ALL-SKY LITERATURE HEATMAP

astrometry.net
ADS All Sky Survey

[prototype]
Archetypes in a Dataverse

**Asteroid** You have small data sets you’d like to see stay in reliable orbits.

**Protostar** You’re young and eager to become a full-grown star, so you want to share all the data you can, and embed links to it in your publications.

**Main-sequence Star** You’ve been at this for a while, so you have long data history and a good future. You’d like to upload important data to go with “old” papers now, and more in the future.

**Cluster** You collect things in catalogs and lists, and you want to group the catalogs for the greater good.

**Supernova** Your disks are EXPLODING with data, and you don’t know what to do with it. You want to permalink vast data sets directly to papers, and more...

**Pulsar** You really like it when things change. Time-domain astronomy is your thing, and you want online identifiers that understand time.

**Galaxy** You love everything, but you’re organized. You make and collect Surveys you don’t want to lose, and you want people to find them from far away.

**Quasar** Your energy is nearly unlimited, so you suck up (mine) and spit out as much data as you can find. And you like to share in showy ways.

**Black Hole** You suck down any and all data, with unbridled appetite. Dataverse is NOT for you.

Coming soon to **PLoS one... Pepe et al. 2012**

Data handling, archiving, and citing in astronomy
Alberto Pepe¹,²,* August Muench¹, Christopher Erdmann¹, Mercè Crosas², Alyssa Goodman¹
1 Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA
2 Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA
* E-mail: apepe@cfa.harvard.edu
44% of data links from 2001 broken in 2011

Figure 1. Volume of potential data links in astronomy publications. Total volume of external links in all articles published between 1997 and 2008 in the four main astronomy journals, color coded by HTTP status code. Green bars represent accessible links (200), grey bars represent broken links. 

Pepe et al. 2012
Figure 2. Percentage of broken links in astronomy publications according to type of website. Percentages of broken external links in all articles published between 1997 and 2008 in the four main astronomy journals. Black circles represent links to personal websites (link values contain the tilde symbol, ~), while red crosses represent links to curated archives such as governmental and institutional repositories.
Asteroid: You have small data sets you'd like to see stay in reliable orbits.

Supernova: Your disks are exploding with data, and you don’t know what to do with it. You want to permalink vast data sets directly to papers, and more...

Protostar: You’re young and eager to become a full-grown star, so you want to share all the data you can, and embed links to it in your publications.

Pulsar: You really like it when things change. Time-domain astronomy is your thing, and you want online identifiers that understand time.

Main-sequence Star: You’ve been at this for a while, so you have long data history and a good future. You’d like to upload important data to go with “old” papers now, and more in the future.

Galaxy: You love everything, but you’re organized. You make and collect surveys you don’t want to lose, and you want people to find them from far away.

Cluster: You collect things in catalogs and lists, and you want to group the catalogs for the greater good.

Quasar: Your energy is nearly unlimited, so you suck up (mine) and spit out as much data as you can find. And you like to share in showy ways.

Black Hole: You suck down any and all data, with unbridled appetite. Dataverse is NOT for you.

---

Archetypes in a Dataverse

Data Citation & Astronomy Dataverse: theastrodata.org
"Seamless Astronomy" (Tools)

Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.
using 2D maps of column-density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D (p-v) data cube into an easily visualized representation called a ‘dendrogram’ (1). Although well developed in other data-intensive fields (2), it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where ‘merger trees’ are being used with increasing frequency (3).

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we ‘fatten’ the dendrograms of 3D data (see Fig. 5 and its legend), by sorting their ‘branches’ to not cross, which eliminates dimensional information on the x-axis while preserving all information about connectivity and hierarchy. The observer’s blurred half-ball in the figures let the reader match colors between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spatial-line data cube allows for the estimation of key physical properties associated with volumes bounded by iso-surfaces, such as radius (\(R\)), velocity dispersion (\(\sigma\)) and luminosity (\(L\)). The volumes can have any shapes, and in other work (4) we focus on the significance of the especially elongated features seen in L1448 (Fig. 2c). The luminosity is an approximate proxy for mass, such that \(M = \frac{L}{\dot{L}}\), where \(\dot{M}\) is the mass flow rate. As ‘clumps’ are not allowed to belong to larger structures, such as radius (\(R\)), velocity dispersion (\(\sigma\)) and luminosity (\(L\)), the derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an ‘observed’ virial parameter, \(\alpha = \frac{\dot{M}}{GMv}\) (in principle, orientational components of the tree (Fig. 2, yellow highlighting), where \(\alpha = \frac{\dot{M}}{GMv}\) where gravitational energy is comparable to or larger than kinetic energy), or the ratio of self-gravity to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields. Its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

Figure 3: Schematic Illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by ‘dropping’ a test constant emission level (purple) from above in tiny steps until all the local maxima and mergers are found, and connected as shown. The intersection of a test level with the emission profile (black). The dendrogram (blue) can be constructed by

...
Data in Literature

Note: This work came from the “AstroMed” project am.iic.harvard.edu

Four years before the advent of CLUMPFIND, ‘structure trees’ were proposed as a way to characterize clouds’ hierarchical structure using 2D maps of column density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D (p-v) data cube into an easily visualized representation called a ‘dendrogram’. Although we developed in other data-intensive fields (1), it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where ‘merger trees’ are being used with increasing frequency.

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we ‘flatten’ the dendrograms of 3D data (see Fig. 5 and its legend), by sorting their ‘branches’ to not cross, which eliminates dimensional information on the x axis while preserving all information about connectivity and hierarchy. Numbered ‘brilliant ball’ labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a, online) and a sorted dendrogram (Fig. 2d).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by iso-surfaces, such as radius (R), velocity dispersion (σ) and luminosity (L). The volumes can have any shape, and in other work (1) we focus on the significance of the especially elongated features seen in L1448 (Fig. 2b). The luminosity is an approximate proxy for mass, such that \( M_{\text{clump}} \approx \frac{L_{\text{obs}}}{X_{13}\, m_{\text{clump}}} \), where \( X_{13} = 8 \times 10^{13} \text{ cm}^{-3} \text{ K}^{-1} \text{ km}^{-1} \text{ s}^{-1} \) (ref. 13; see Supplementary Methods and Supplementary Fig. 2).

The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an ‘observed’ virial parameter, \( \lambda_v = \frac{2\, \sigma^2}{G\, M} \). As ‘clumps’ are not allowed to belong to larger structures, each pseudo-branch in the ‘dropping’ a test constant emission level (purple) from above in tiny steps, we capture external over-pressure and/or magnetic fields (3), its measured value should only be used as a guide to the longevity (boundlessness) of any particular feature.

**Figure 2**: Comparison of the ‘dendrogram’ and CLUMPFIND features. Inside a 3D map, the dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by iso-surfaces, such as radius (R), velocity dispersion (σ) and luminosity (L). The volumes can have any shape, and in other work (1) we focus on the significance of the especially elongated features seen in L1448 (Fig. 2b). The luminosity is an approximate proxy for mass, such that \( M_{\text{clump}} \approx \frac{L_{\text{obs}}}{X_{13}\, m_{\text{clump}}} \), where \( X_{13} = 8 \times 10^{13} \text{ cm}^{-3} \text{ K}^{-1} \text{ km}^{-1} \text{ s}^{-1} \) (ref. 13; see Supplementary Methods and Supplementary Fig. 2).

The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an ‘observed’ virial parameter, \( \lambda_v = \frac{2\, \sigma^2}{G\, M} \). As ‘clumps’ are not allowed to belong to larger structures, each pseudo-branch in the dendrogram can be rotated to any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.4 or higher). In the printed version, the front face of each 3D cube (the ‘home’ view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1. This choice of orientation allows for a clear and consistent presentation of the standard of best practice from front (\( \sim 0.5 \text{ km s}^{-1} \)) to back (\( \sim 8 \text{ km s}^{-1} \)).

**Figure 3**: Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram ‘trees’ can be constructed by ‘dropping’ a test constant emission level (purple) from above in tiny steps (inverted at some low velocity) until the local maxima and minima are found, and connected as shown. The intersection of a test level with the emission is a set of points (for example the light purple dots) in one dimension, a plane curve in two dimensions, and an isosurface in three dimensions. The dendrogram of 3D data shown in Fig. 2 is the direct analogue of the tree shown here, only constructed from ‘branches’ rather than ‘point’ intersections. It has been sorted and flattened for representation on a flat page, so fully representing dendrograms for 3D data cubes would require four dimensions.


Principles of high-dimensional data visualization in astronomy

A.A. Goodman*

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Received 2012 May 3, accepted 2012 May 4
Published online 2012 Jun 15

Key words cosmology: large-scale structure – ISM: clouds – methods: data analysis – techniques: image processing – techniques: radial velocities

Astronomical researchers often think of analysis and visualization as separate tasks. In the case of high-dimensional data sets, though, interactive exploratory data visualization can give far more insight than an approach where data processing and statistical analysis are followed, rather than accompanied, by visualization. This paper attempts to charts a course toward “linked view” systems, where multiple views of high-dimensional data sets update live as a researcher selects, highlights, or otherwise manipulates, one of several open views. For example, imagine a researcher looking at a 3D volume visualization of simulated or observed data, and simultaneously viewing statistical displays of the data set’s properties (such as an $x$-$y$ plot of temperature vs. velocity, or a histogram of vorticities). Then, imagine that when the researcher selects an interesting group of points in any one of these displays, that the same points become a highlighted subset in all other open displays. Selections can be graphical or algorithmic, and they can be combined, and saved. For tabular (ASCII) data, this kind of analysis has long been possible, even though it has been under-used in astronomy. The bigger issue for astronomy and other “high-dimensional” fields, though, is that no extant system allows for full integration of images and data in a linked, web-like, user-friendly environment. The paper concludes its history and analysis of the present situation by suggesting a model for a new, web-based, cooperatively-developed open-source modular software as a way to create an evolving, linked-view visualization environment useful in astrophysical research.
Glimpse of the Future (that’s here now...)

Results from Tom Rice's Thesis:
Preliminary Hierarchical Catalog of Milky Way Plane Molecular Clouds
Glimpse of the Future (that’s here now...)