glue-ing together the Universe
from Galileo to Gaia

Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics + Radcliffe Insitute for Advanced Study @AlyssaAGoodman
Galileo, Jupiter’s Moons, “3D” thinking

Notes for & re-productions of Siderius Nuncius
Galileo’s 3D thinking, in WorldWide Telescope

January 11, 1610
The Sky at Many Wavelengths in a “WorldWide Telescope”
Dimensions’ many meanings
“Data, Dimensions, Display”

1D: Columns = “Spectra”, “SEDs” or “Time Series”
2D: Faces or Slices = “Images”
3D: Volumes = “3D Renderings”, “2D Movies”
4D: Time Series of Volumes = “3D Movies”
A role for self-gravity at multiple length scales in the process of star formation

Alyssa A. Goodman\textsuperscript{1,2}, Erik W. Rosolowsky\textsuperscript{1,3}, Michelle A. Borkin\textsuperscript{1,4}, Jonathan B. Foster\textsuperscript{2}, Michael Halle\textsuperscript{1,4}, Jens Kauffmann\textsuperscript{1,2} & Jaime E. Pineda\textsuperscript{2}

Self-gravity plays a decisive role in the final stages of star formation, where dense cores (size \(\sim 0.1\) parsecs) inside molecular clouds collapse to form star-plus-disk systems\textsuperscript{4}. But self-gravity’s role at earlier times (and on larger length scales, such as \(\sim 1\) parsec) is unclear; some molecular cloud simulations that do not include self-gravity suggest that ‘turbulent fragmentation’ alone is sufficient to create a mass distribution of dense cores that resembles, and sets, the stellar initial mass function\textsuperscript{4}. Here we report a ‘dendrogram’ (hierarchical tree-diagram) analysis that reveals that self-gravity plays a significant role over the full range of possible scales traced by \(^{13}\)CO observations in the L1448 molecular cloud, but not everywhere in the observed region. In particular, more than 90 per cent of the compact ‘pre-stellar cores’ traced by peaks of dust emission\textsuperscript{6} are projected on the sky within one of the dendrogram’s self-gravitating ‘leaves’. As these peaks mark the locations of already-forming stars, or of those probably about to form, a self-gravitating core seems a critical condition for their existence. Overlapping features as an option, significant emission found between prominent clumps is typically either appended to the nearest clump or turned into a small, usually ‘pathological’, feature needed to encompass all the emission being modelled. When applied to molecular-line
data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND’s two free parameters, the same molecular-line data set can be used to show either that the frequency distribution of clump mass is the same as the mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

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 the advent of 3D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchic structure of a cloud easily visualized and well developed in other data-intensive applications of tree methodologies and almost exclusively within the area of ‘merger trees’ being used with 3D.

Figure 2: Comparison of the ‘dead’ panels and CLUMPFIND feature-identification algorithms as applied to 13CO emission from the L1448 region of Perseus. a, b, 3D-volumetric visualization of the surface indicated by red emission in the dendrogram shown in c. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest structures that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of T_{mb} (main-beam temperature) test-level values for which the total number is less than 2. Blue- and green-colored features of the ‘self-gravitating’ leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D-volumetric visualization shows positions–position–velocity (P–P–V) space.

Figure 3: Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (Shief). The dendrogram (blue) can be constructed by dropping a test constant emission level (purple) from above in tiny steps (g) until all the local maxima and mergers are identified 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 linear curve in two dimensions, and an isosurface in three dimensions. The dendrograms of 3D data shown in Fig. 3 is the direct analogue of the tree shown here, only constructed from ‘isosurfaces’ rather than ‘point’ intersections. It has been sorted and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

© 2009 Macmillan Publishers Limited. All rights reserved
Linked Views of High-dimensional Data

3D

2D

Data Abstraction

Statistics

figure, by M. Borkin, reproduced from Goodman 2012, “Principles of High-Dimensional Data Visualization in Astronomy”
Linked Views of High-dimensional Data (in Python)

glue

video by Tom Robitaille, lead glue developer

glue created by: C. Beaumont, M. Borkin, M. Breddels, P. Qian, T. Robitaille, and A. Goodman, PI
Linked Views of High-dimensional Data (in Python)

**glue**

video by Chris Beaumont, glue developer

**glue** created by: C. Beaumont, M. Borkin, M. Breddels, P. Qian, T. Robitaille, and A. Goodman, PI
Linked views & my FBI file... (glue’s not just for Astronomy)
WorldWide Telescope

multidimensional data exploration

Au

Authorea

glue

Au

Authorea

10Viz
If astronomy had its own Academy Awards, then this part of the Milky Way would have been the Favorite Nebula pick for 2011. Competing against 12,203 other slices of the sky, this got more votes from the 35,000 volunteers searching for cosmic bubbles than any other location.

The volunteers are all citizen scientists working on the Milky Way Project, scanning a vast collection of infrared images from NASA’s Spitzer Space Telescope. Their goal is to identify bubbles that have...
BIG DATA, WIDE data
BIG DATA AND “HUMAN-AIDED COMPUTING”

BIG DATA, WIDE DATA
WIDE DATA

- mm peak (Enoch et al. 2006)
- sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
- $^{13}$CO (Ridge et al. 2006)
- mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)
- Optical image (Barnard 1927)
WIDE DATA, "IN 3D"

- mm peak (Enoch et al. 2006)
- sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
- $^{13}$CO (Ridge et al. 2006)
- mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)
- Optical image (Barnard 1927)
glue
multidimensional data exploration
glue
multidimensional data exploration
Define new variables, import/export insights, and interactive plots for the web, save state, all from GUI.

Link data files' attributes and highlight live or algorithmic selections with Boolean logic.

Custom buttons, features.

Custom plots.

Built-in standard 1D, 2D & 3D plots.

Custom data loaders.

Standard data loaders.

User config.py file (loaders, colors, plot types, +)

Access to all matplotlib functions through built-in IPython terminal.

Run & interact with glue from Jupyter notebook & other tools.

glueviz.org

demo
ALMA
publishing
“No merging of data sets—just glue them.”

An ALMA core
Just drag to visualize, e.g. series of 2D “channel maps.”
Adjust so each tracer is a different color.
Create 3D views...

...see clearly “veil” of wind-blown methanol.
Traditional Rainbow Channel maps
**standard data loaders**
- define new variables, import/export insights, interactive plots
- for the web, save state, all from GUI

**custom data loaders**
- link data files’ attributes
- highlight live or algorithmic selections with Boolean logic
- custom buttons, features

**plug-in**
- standard 1D, 2D & 3D plots
- custom plots

**user config.py file**
- (loaders, colors, plot types, +)

**+options**
- access to all matplotlib functions through built-in IPython terminal
- run & interact with glue from Jupyter notebook & other tools

**glueviz.org**
WorldWide Telescope as a plug-in to glue...
But... Publishing?

The "Paper" of the Future

Alyssa Goodman, Josh Peek, Alberto Accomazzi, Chris Beaumont, Christine L. Borgman, How-Huan Hope Chen, Merce Crosas, Christopher Erdmann, August Muench, Alberto Pepe, Curtis Wong

A 5-minute video demonstration of this paper is available at this YouTube link.

1 Preamble

A variety of research on human cognition demonstrates that humans learn and communicate best when more than one processing system (e.g., visual, auditory, touch) is used. And, related research also shows that, no matter how technical the material, most humans also retain and process information best when they can put a narrative "story" to it. So, when considering the future of scholarly communication, we should be careful not to do blithely away with the linear narrative format that articles and books have followed for centuries: instead, we should enrich it.

Much more than text is used to communicate in Science. Figures, which include images, diagrams, graphs, charts, and more, have enriched scholarly articles since the time of Galileo, and ever-growing volumes of data underpin most scientific papers. When scientists communicate face-to-face, as in talks or small discussions, these figures are often the focus of the conversation. In the best discussions, scientists have the ability to manipulate the figures, and to access underlying data, in real-time, so as to test out various what-if scenarios, and to explain findings more clearly. This short article explains—and shows with demonstrations—how scholarly "papers" can morph into long-lasting rich records of scientific discourse, enriched with deep data and code linkages, interactive figures, audio, video, and commenting.
Gaia & the Milky Way

The Milky Way
(Artist’s Conception)
A Large Catalog of Accurate Distances to Local Molecular Clouds: The Gaia DR2 Edition

Catherine Zucker,1,* Joshua S. Speagle,1, * Edward F. Schlaufly,2 Gregory M. Green,3 Douglas P. Finkbeiner,1 Alyssa A. Goodman,1,4 and João Alves4,5

1 Harvard Astronomy, Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA
2 Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA
3 Kavli Institute for Particle Astrophysics and Cosmology, Physics and Astrophysics Building, 452 Lometa Mall, Stanford, CA 94305, USA
4 Radcliffe Institute for Advanced Study, Harvard University, 10 Garden St, Cambridge, MA 02138
5 University of Vienna, Department of Astrophysics, Türkenkranzstraße 17, 1180 Vienna, Austria

ABSTRACT

We present a uniform catalog of accurate distances to local molecular clouds informed by the Gaia DR2 data release. Our methodology builds on that of Schlaufly et al. (2014). First, we infer the distance and extinction to stars along sightlines towards the clouds using optical and near-infrared photometry. When available, we incorporate knowledge of the stellar distances obtained from Gaia DR2 parallax measurements. We model these per-star distance-extinction estimates as being caused by a dust screen with a 2-D morphology derived from Planck at an unknown distance, which we then fit for using a nested sampling algorithm. We provide updated distances to the Schlaufly et al. (2014) sightlines towards the Dame et al. (2001) and Magnani et al. (1985) clouds, finding good agreement with the earlier work. For a subset of 27 clouds, we construct interactive pixelated distance maps to further study detailed cloud structure, and find several clouds which display clear distance gradients and/or are comprised of multiple components. We use these maps to determine robust average distances to these clouds. The characteristic combined uncertainty on our distances is ≈ 5 – 6%, though this can be higher for clouds at farther distances, due to the limitations of our single-cloud model.

Keywords: ISM: clouds, ISM: dust, extinction, stars: distances, methods: statistical
The 10 Questions

1. **Who** | Who is your audience? How expert will they be about the subject and/or display conventions?
2. **Explore-Explain** | Is your goal to explore, document, or explain your data or ideas, or a combination of these?
3. **Categories** | Do you want to show or explore pre-existing, known, human-interpretable, categories?
4. **Patterns** | Do you want to identify new, previously unknown or undefined patterns?
5. **Predictions & Uncertainty** | Are you making a comparison between data and/or predictions? Is representing uncertainty a concern?
6. **Dimensions** | What is the intrinsic number of dimensions (not necessarily spatial) in your data, and how many do you want to show at once?
7. **Abstraction & Accuracy** | Do you need to show all the data, or is summary or abstraction OK?
8. **Context & Scale** | Can you, and do you want to, put the data into a standard frame of reference, coordinate system, or show scale(s)?
9. **Metadata** | Do you need to display or link to non-quantitative metadata? (Including captions, labels, etc.)
10. **Display Modes** | What display modes might be used in experiencing your display?

---

*To learn more about this site, please visit the About page.*

To read an in-process manuscript giving the scholarship behind the recommendations on this site, see Coltekin & Goodman 2018.
Bonus anyone?

Infoviz

3D selection
Empirical

"Law"

A Priori

Theory

A Priori

Theory

Philosophize
to Explain

Notice

Phenomenon

Collect

Data

Quantify Pattern
Mathematically

Galileo

Copernicus

Ptolemy

Aristotle

Newton

Kepler

Google Maps

Predictive

Systems

Version 1 (Terrible)
Version 2 . . . Simplify!
Predictive Systems

Phenomenon Observation* Data Rule Theory Explanation Prediction

*or, Experiment*
Kepler's Laws of Planetary Motion

Newton's Theory of Gravity
Phenomenon | Observation | Data | Rule | Theory | Explanation | Prediction
---|---|---|---|---|---|---

EMPIRICAL | ANALYTICAL

Ptolemy’s Epicycles | Aristotle’s Natural Philosophy

Wrong
Phenomenon | Observation* | Data | Rule | Theory | Explanation | Prediction
---|---|---|---|---|---|---

*or, Experiment*
Phenomenon Observation Data Rule Prediction

Mendel

Phenomenon Observation Data Theory Explanation Prediction

Darwin

"YOUR" WORLD (BIOLOGY)

Phenomenon Observation Data Rule

Kepler

MY WORLD (PHYSICS)

Phenomenon Observation Data Rule Theory Explanation Prediction

Newton
Mendel

Phenomenon Observation Data Rule Prediction

Darwin

Phenomenon Observation Data Theory Explanation Prediction

Kepler

Phenomenon Observation Data Rule Prediction

Newton

Phenomenon Observation Data Rule Theory Explanation Prediction

NO FULLY PREDICTIVE GENERAL THEORY

FULLY PREDICTIVE GENERAL THEORY
NO FULLY PREDICTIVE GENERAL THEORY
THE FUTURE OF THE FUTURE

20th century

Phenomenon  Observation  Data  Rule  Theory  Explanation  Prediction

21st century?

Phenomenon  Observation  Data  “machine learning”  Prediction
The challenge of 3D Selection
Coming next: glue in the browser
This is all possible, and happening now.

Not for everyone, though.

Critical to *evaluate adoption* & *refine* solutions
Michelle Borkin is doing this for glue
—look for results in Borkin, Goodman, Munzner et al. 2019