

# HiveR: 2D & 3D Hive Plots + Hive Panels

## New Tools for Network Visualization

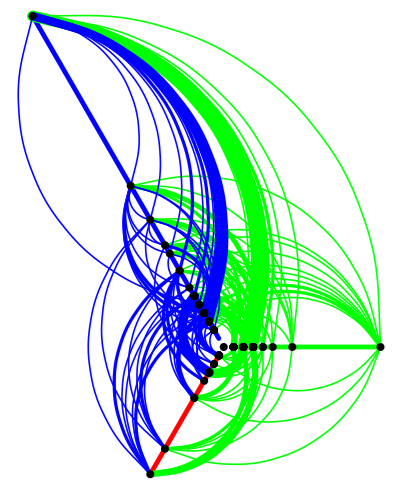
Bryan A. Hanson

Dept. of Chemistry & Biochemistry, DePauw University, Greencastle Indiana USA  
hanson@depauw.edu    github.com/bryanhanson/HiveR    CRAN.R-project.org/package=HiveR

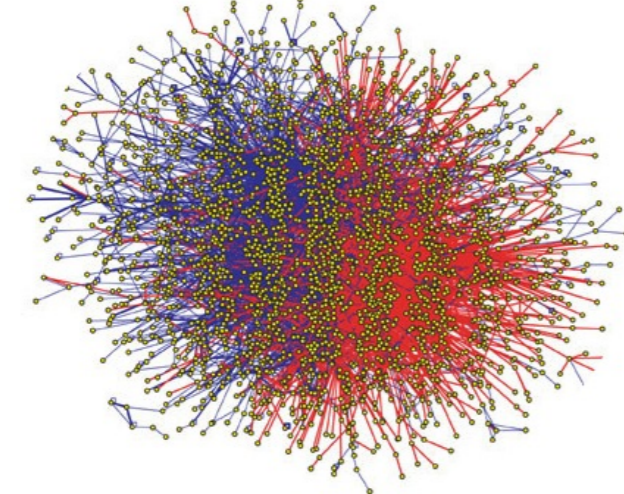
### Inspiration & Motivation

Developed by Martin Krzywinski at the Genome Sciences Center ([www.hiveplot.com](http://www.hiveplot.com))

pretty & useful



artifact-rich infoart



Ruai et al. Nature vol 437 pg 1173 (2005)

- "Hairball" style networks do not meet our standard of reproducible research
- The key innovation in a Hive Plot is that there is a node coordinate system
- Application Areas: ecology, social networks, systems biology, computer science
- HiveR was written from scratch – it is not a port of the original *Perl* prototype. Others have written *Java* & *D3* versions

### Characteristics of Hive Plots

- **Hive Plots are transparent:**
  - Rational: layout determined only by properties of the network (no algorithm)
  - Predictable & Reproducible: network features are mapped to plot features
  - Robust to node/edge loss
- **Hive Plots are practical:**
  - Flexible & can be tuned to show interesting features
  - Complexity scales well – details can be inspected
  - Networks can be directly compared

### Implementation

#### Hive Plot Features Which Can Be Mapped

Axis to which a node is assigned
Radius of a node
Color of a node
Size of a node
Color of an edge
Width of an edge

- Node assignment based upon qualitative or quantitative characteristics
- No jumping or crossing axes allowed
- Mapping limited only by one's creativity & the particular knowledge domain
- Mapping can be readily tuned
- Mapping results in a reproducible plot

### Hive Plots: Axis Units/Scaling Options

method	axis length	center hole	node behavior
native	$f(\text{units})$	asymmetric	nodes may overlap
ranked	$\propto \text{rank}(\text{nodes})$	circular	nodes evenly spaced & don't overlap
normed	all equal	circular	nodes may overlap
ranked & normed	all equal	circular	nodes evenly spaced & don't overlap

### Hive Plot Data Objects (HPD)

\$nodes			
\$id	int	chr	identifier
\$lab	int	chr	label
\$axis	int	chr	axis
\$radius	num	chr	radius
\$size	num	chr	size
\$color	chr	chr	color
\$edges			
\$id1	int	1st node id	
\$id2	int	2nd node id	
\$weight	num	width	
\$color	chr	color	
\$type			
\$desc	chr	description	
\$axis.cols	chr	axis colors	
- attr	chr	"HivePlotData"	

### Nuances of Hive Plots

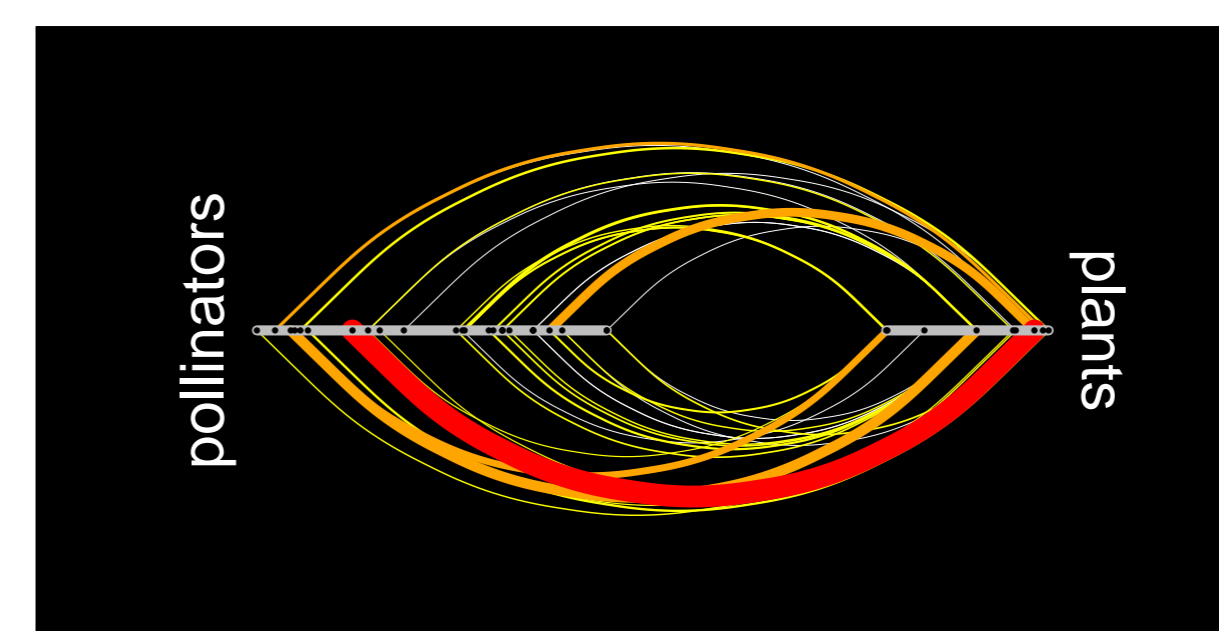
- Hive Plots are radially-arranged parallel coordinate plots
- Assigning the nodes is the toughest part
- Nodes cannot be assigned w/o thinking about the edges
- For 2D Hive Plots with 2 or 3 axes, edges cannot jump axes. For 4+ axes, you must guard against this: Edges should go 1 → 2, 2 → 3, ... 5 → 6, but not 1 → 5
- For 3D Hive Plots, no edges can start & end on the same axis. For 5 or 6 axes, edges may not start on one axis & end on a co-linear axis. For 4 axes, edges cannot jump axes. For 5 or 6 axes, you must guard against this.
- Hive Plots are directionally agnostic. Almost.
- With native or normed coordinates, nodes may overlap.
- The nodes & edges "on top" & showing are the last drawn nodes (sort before drawing)

### HiveR Utilities

- Generate random networks (`ranHiveData`)
- Import data (`dot2HPD`, `adj2HPD`)
- Extract embedded information (`mineHPD`)
- Scale or invert an axis (`manipAxis`)
- Check integrity of the HPD (`chkHPD`)
- Summarize a HPD (`sumHPD`)

### Plant-Pollinator Network

Data set *Safariland* (Vazq. & Simberloff) describes plant-pollinator pairs and the number of visits in grazed and ungrazed habitats. Node radius is  $|d'|$ , an index of specialization. Edge weights are  $\propto \sqrt{\text{no.visits}}$ , divided into 4 groups and colored white to red (redder = more visits)



### E. coli Gene Regulatory Network

Raw data is composed of genes that regulate each other via transcription factors (Yan et al). The initial Dot file is imported with the aid of two files:

#### Sample Dot File Contents

```
[1] "acee [label=nonpersistent];"
[2] "argh [label=persistent];"
[3] "arca -- acea [type=1];"
[4] "fnr -- acef [type=0];"
```

#### Node Mapping Instructions

dot.tag	dot.val	hive.tag	hive.val
label	persistent	color	red
label	nonpersistent	color	black

#### Edge Mapping Instructions

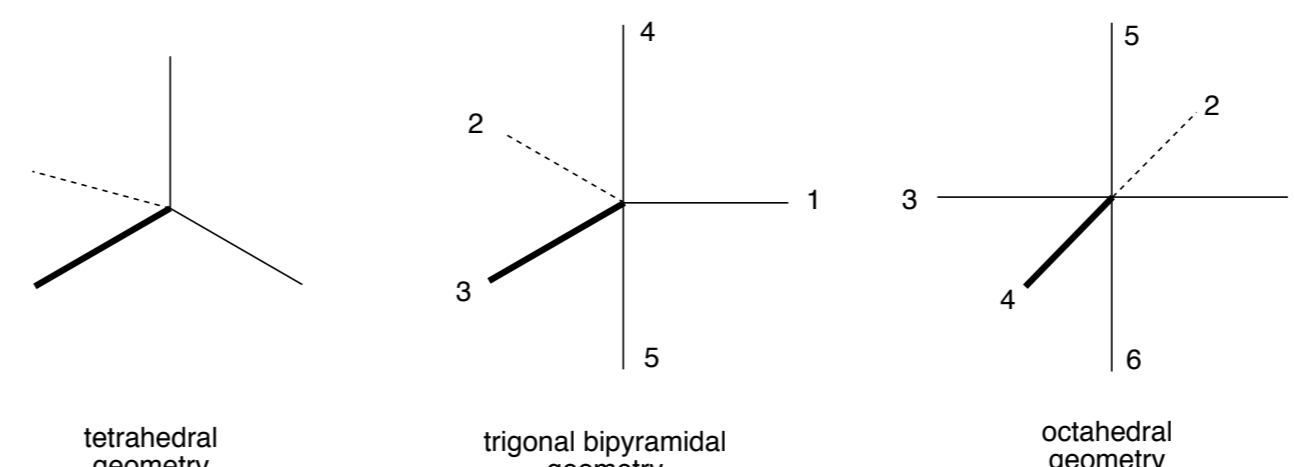
dot.tag	dot.val	hive.tag	hive.val
type	0	color	grey
type	1	color	yellow
type	2	color	orange
type	3	color	red

The following steps were used to create a Hive Plot:

- Nodes to axes according to role (& persistence for 5-axis plots)
- Node radius = edge count/degree
- Orphaned nodes removed, edges sorted
- Gene pair proximity (edges) colored gray → red, red physically closest

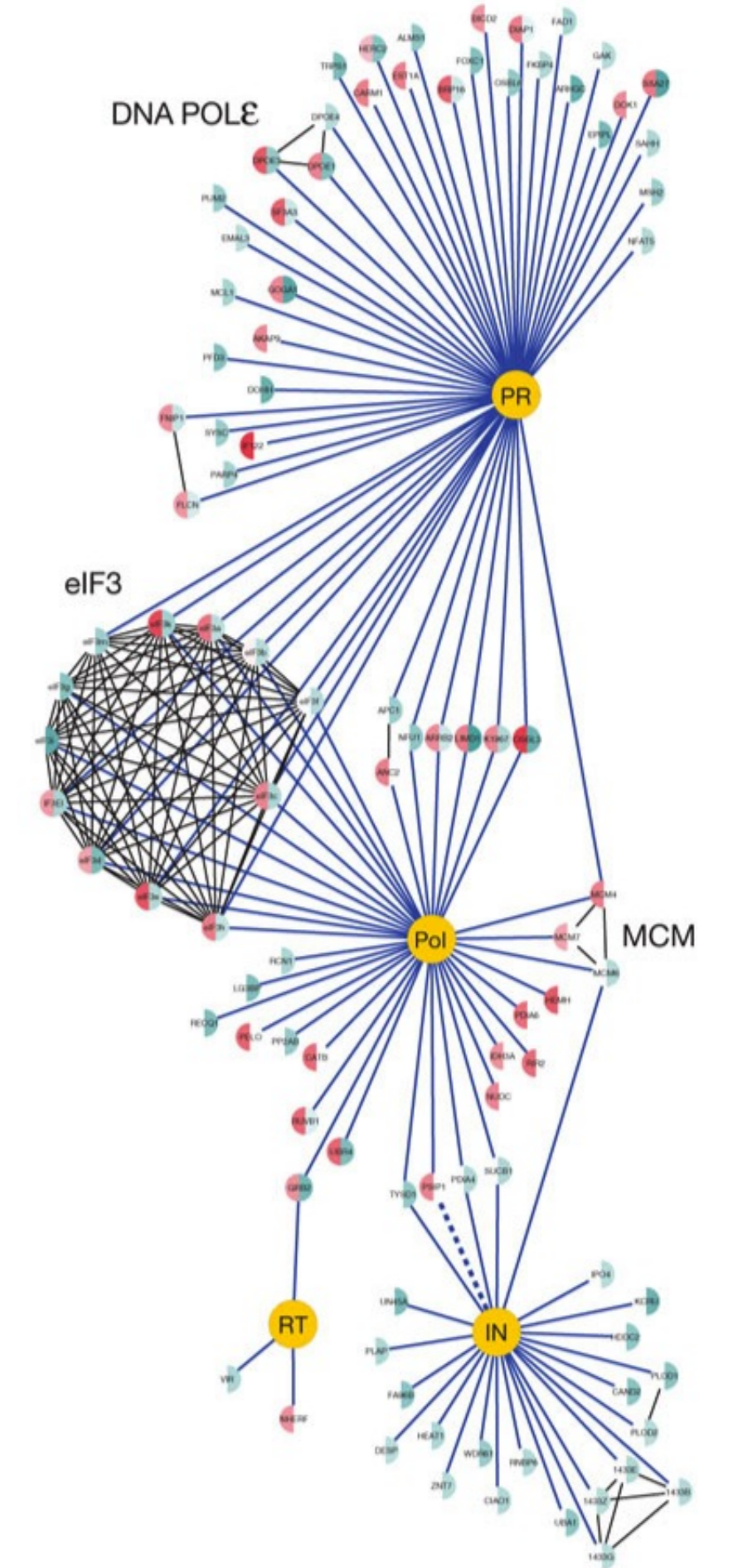
### 3D Hive Plots

- 3D Hive Plots use `rgl` graphics
- More adjacent axes than for 2D Hive Plots
  - Tetrahedron: 8 adjacent axis pairs, *crossings impossible*
  - Trigonal bipyramid: 9 adjacent axis pairs
  - Octahedron: 12 adjacent axis pairs

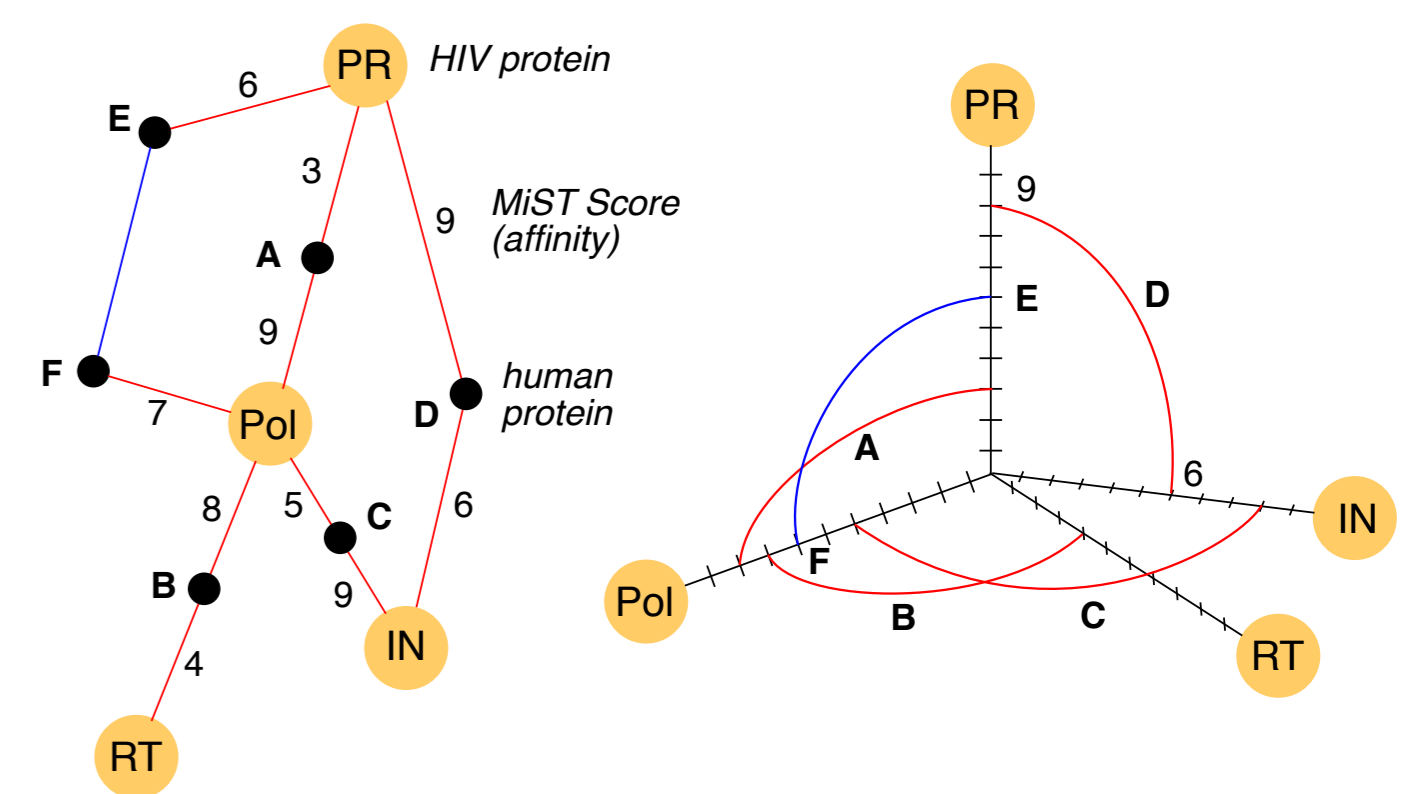


### HIV-Human Protein-Protein Interactions

- Data includes both HIV-human & human-human protein interactions (Jäger et al)
- MiST scores (strength of protein-protein affinity; 2 human cell lines)



#### The Mapping Process

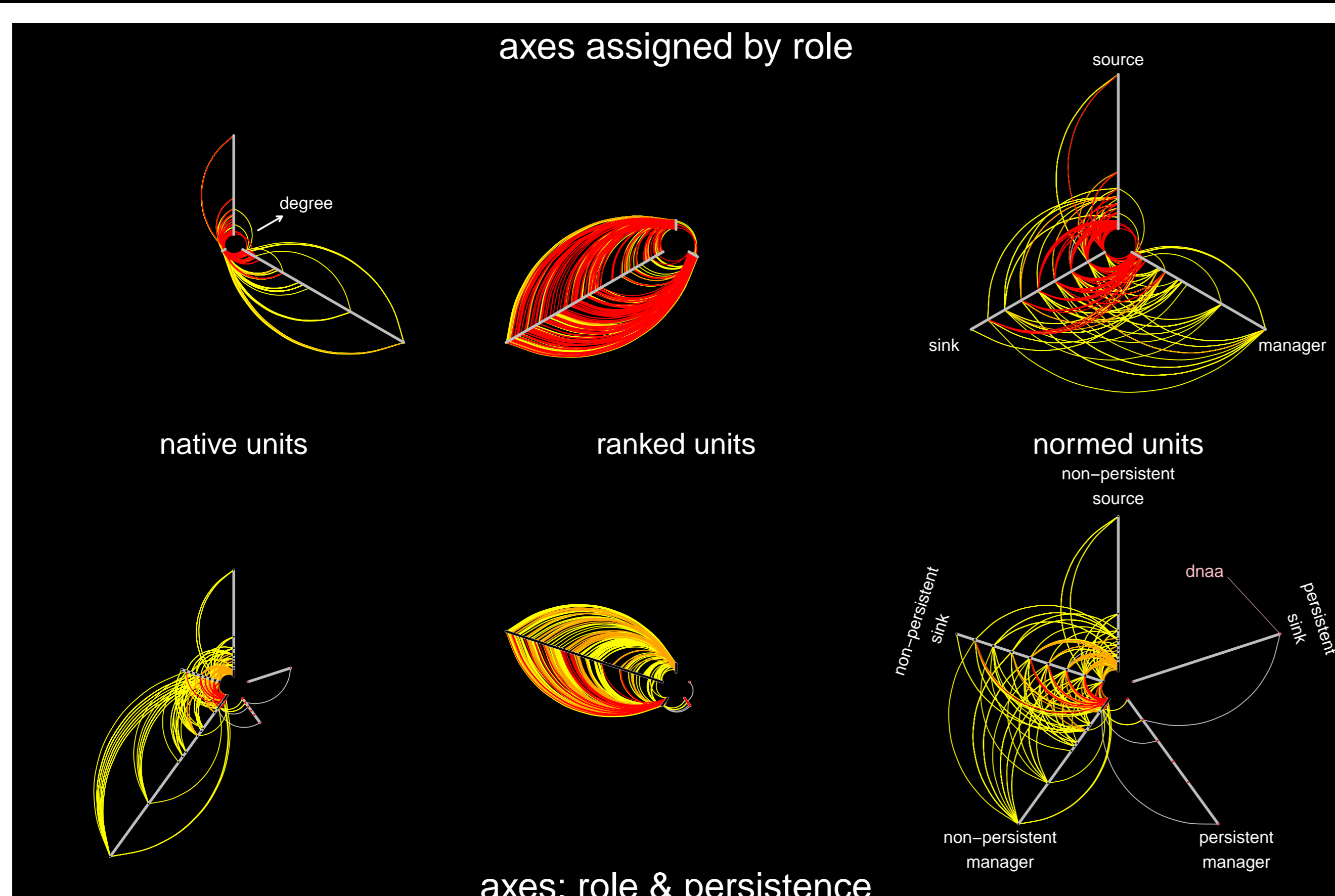


- Each axis is an HIV protein w/human proteins as nodes
- Node radius = MiST score (affinity); yellow nodes have degree  $\geq 2$
- Red edges: a human protein which interacts with 2 HIV proteins
- Blue edges: human-human protein interactions which indirectly link 2 HIV proteins

### References

Krzywinski et al. *Briefings in Bioinformatics* doi:10.1093/bib/bbr069 (2011)  
Vazquez & Simberloff, *Ecology Letters* vol 6 pg 1077 (2003)  
Yan et al. *PNAS* vol 107 pg 9186 (2010)  
Jäger et al. *Nature* vol 481 pg 365 (2012)

### Panel of 2D Hive Plots: E. coli Gene Regulatory Network



### 3D Interactive Hive Plot of HIV-Human Protein-Protein Interactions