HIVE PLOTS

LINEAR LAYOUT FOR VISUALIZATION OF NETWORKS

DRAWING NETWORKS RATIONALLY

OR

THE END OF HAIRBALLS

MARTIN KRZYWINSKI

Talks and software – http://mkweb.bcgsc.ca/linnet.
Originally presented at Genome Informatics 2010, Hinxton UK (September 17).

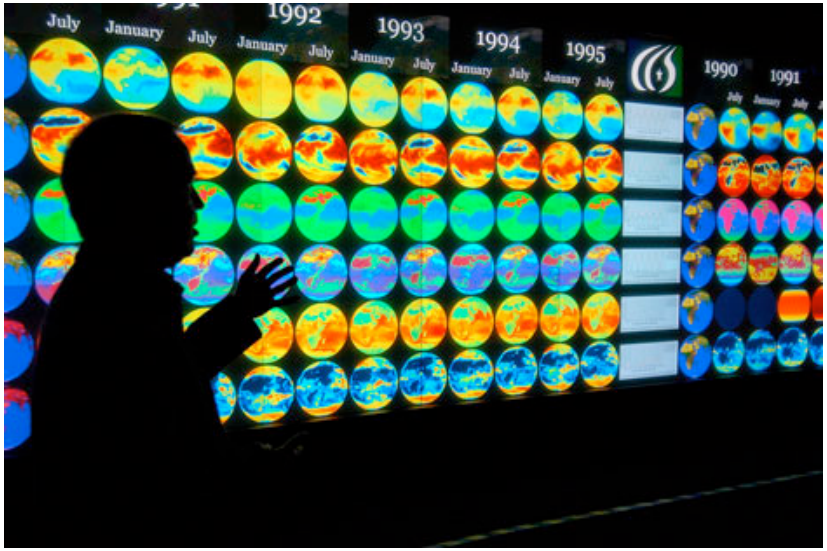
BC Cancer Research Center
Genome Sciences Center

WHAT ARE WE HOPING TO ACHIEVE?

Exploring data sets and **communicating** your findings are two different activities. Typically, the same visualization approach does not suit both.

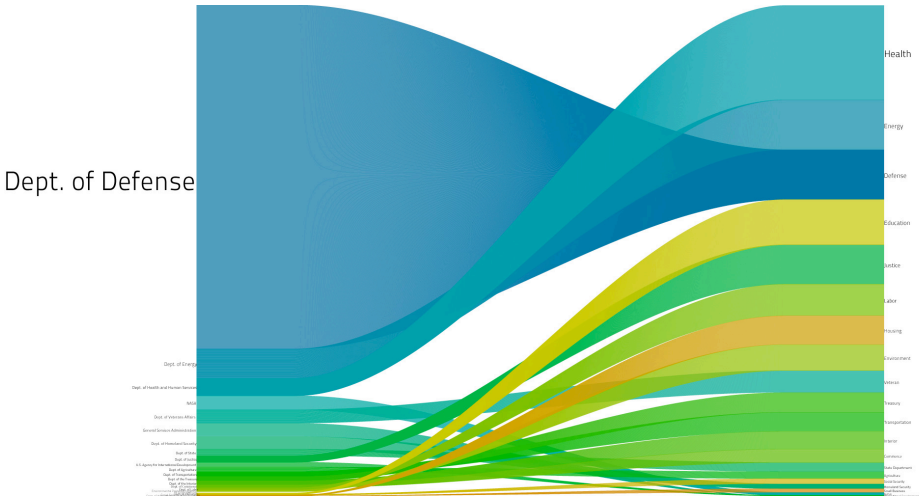
EXPLORATORY VISUALIZATIONS ARE TOO COMPLEX TO COMMUNICATE.

COMMUNICATIVE VISUALIZATIONS CANNOT BE CREATED UNTIL DATA IS EXPLORED.



The EVEREST PowerWall at Oak Ridge National Laboratory, in Tennessee, is a computer visualization facility. EVEREST stands for Exploratory Visualization Environment for Research in Science and Technology. The 9-meter-wide, 2.4-meter-tall screen can display 35 million pixels of information and is now being used as a tool to model climate change.

http://spectrum.ieee.org/energy/nuclear/slideshow-a-nuclear-family-vacation/0

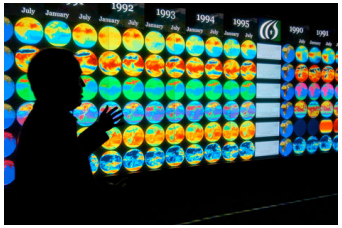


Comparison of US budget spending by department (left) and related media coverage (right).

http://www.pitchinteractive.com/usbudget/

EXPLORING VS COMMUNICATING

EXPLORE



data driven

hypothesis generating

discover patterns and themes

be thorough by applying variety of approaches

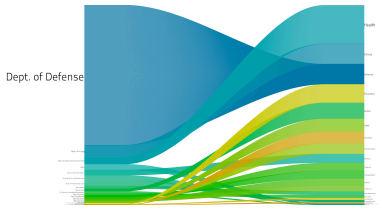
EXPLORATORY VISUALIZATIONS ARE USEFUL TO THE EXPERT, WHO KNOWS THE DATA, CAN DISTINGUISH SIGNAL FROM NOISE, AND IDENTIFY INTERESTING PATTERNS.

PROCESS AND INTEGRATE

identify minimal set of visual elements to communicate your findings

identify appropriate scale at which differences are shown

COMMUNICATE



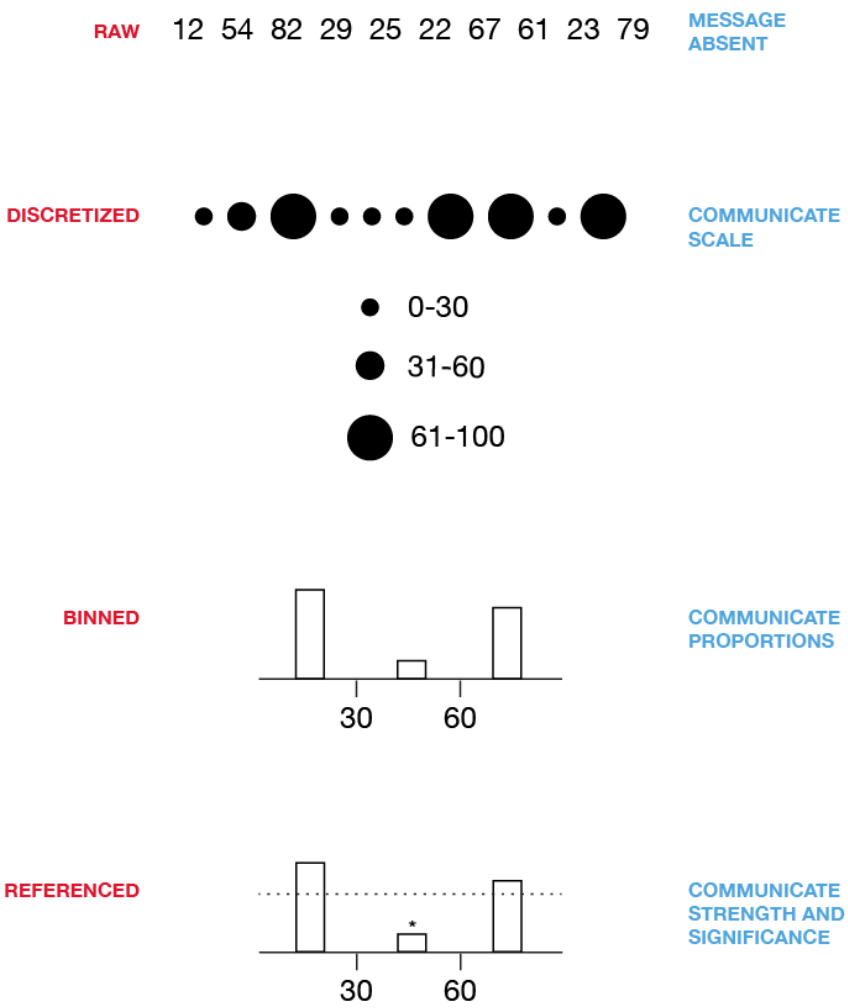
theme driven

communicate patterns and themes

be specific in guiding the reader towards your conclusion – the reader is exploring!

DESIGNED TO EMPHASIZE CONCLUSIONS, THESE VISUALIZATIONS MUST COMMUNICATE THE STRENGTH AND SIGNIFICANCE OF THE EFFECT. OPTIONALLY, OTHER DATA CAN BE INTEGRATED TO PROVIDE CONTEXT, WHILE PRESERVING CLARITY.

CONSTRAIN AMBIGUITY BY IDENTIFYING MEANINGFUL SCALE AND SIGNIFICANCE



EMERGING PATTERNS

Showing the entire data set presents the opportunity to identify emergent patterns – characteristics that are very difficult to express analytically but trivially identified visually.

Nope, no emergent pattern here.

Yup, there's a pattern here.

Consider how long it took you to identify the form and compare this to how long it would take to create a general program to do the same.

CAPTURING IMAGINATION

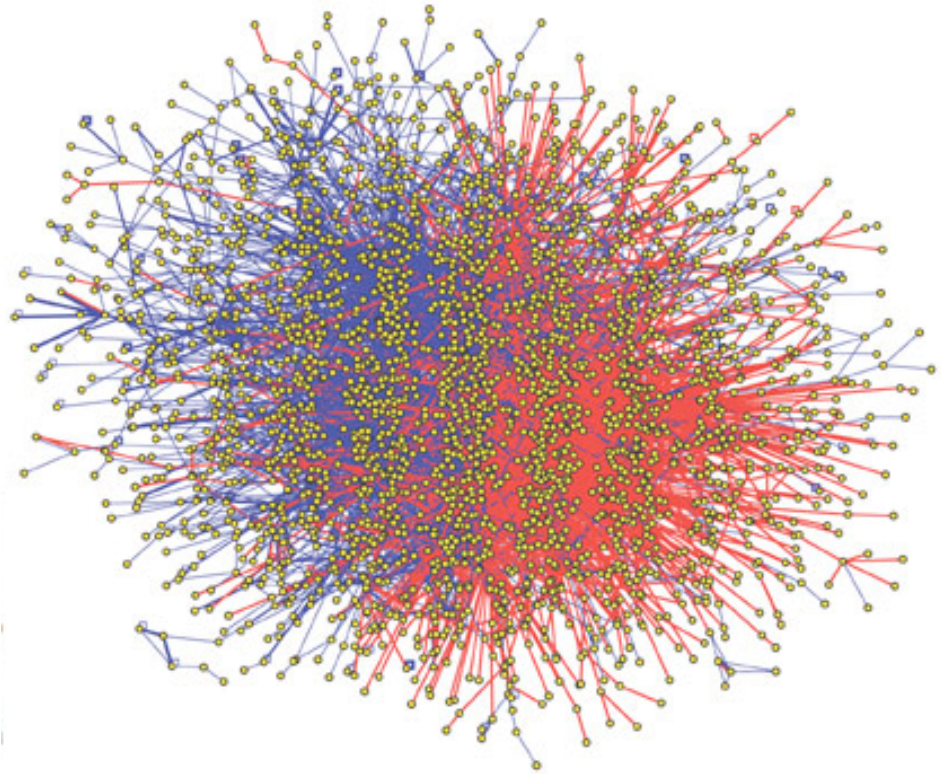
Network data sets are natural inputs for **information art**. What is information art? A visualization which is beautiful, engaging and compels us to look deeper – it does not need to be informative.

Yawn.

Rockin'.

HAIRBALLS

Both are visualizations of a complex system.



Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A et al. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456** (7220): 387-390.

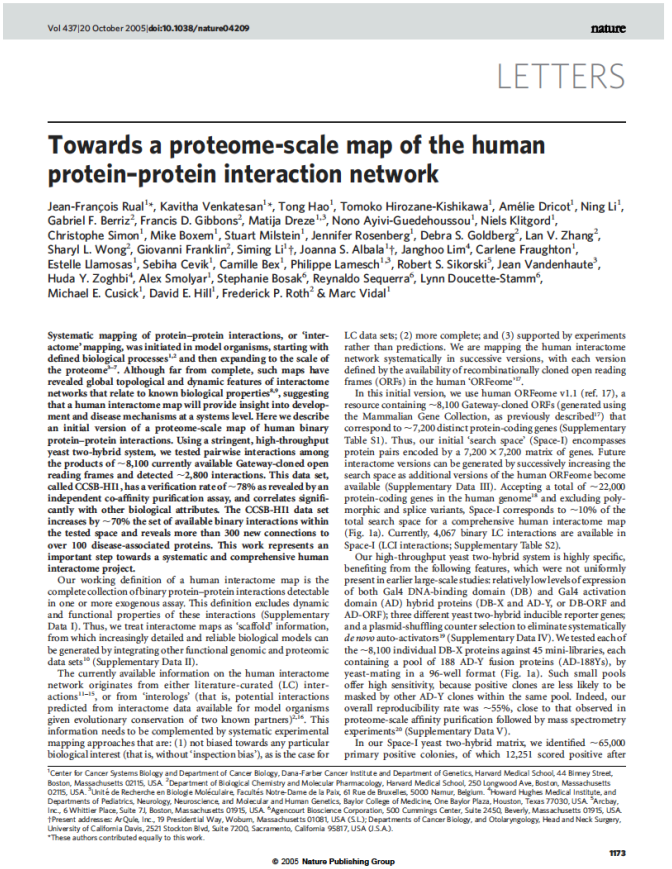
Figure 2 and caption quote from Rual et al., *Nature* **437**(7062):1173-8.

SYSTEM BEHIND THE HAIRBALL

Sometimes a better visualization exists if the question “what does it look like” makes sense (mammoth). If the question does not make sense, a different question must be asked: what is important?

A good visualization of the mammoth.

http://wild-facts.com/2010/09/22/wild-fact-718-the-snow-plow-wooly-mammoth/



Vol 437|20 October 2005|doi:10.1038/nature04209 nature

LETTERS

Towards a proteome-scale map of the human protein–protein interaction network

Jean-François Rual[1]*, Kavitha Venkatesan[1]*, Tong Hao[1], Tomoko Hirozane-Kishikawa[1], Amélie Dricot[1], Ning Li[1], Gabriel F. Berriz[2], Francis D. Gibbons[2], Matija Dreze[1,3], Nono Ayivi-Guedehoussou[1], Niels Klitgord[1], Christophe Simon[1], Mike Boxem[1], Stuart Milstein[1], Jennifer Rosenberg[1], Debra S. Goldberg[2], Lan V. Zhang[2], Sharyl L. Wong[2], Giovanni Franklin[2], Siming Li[1]†, Joanna S. Albala[1]†, Janghoo Lim[4], Carlene Fraughton[1], Estelle Llamosas[1], Sebiha Cevik[1], Camille Bex[1], Philippe Lamesch[1,5], Robert S. Sikorski[5], Jean Vandenhaute[3], Huda Y. Zoghbi[4], Alex Smolyar[1], Stephanie Bosak[6], Reynaldo Sequerra[6], Lynn Doucette-Stamm[6], Michael E. Cusick[1], David E. Hill[1], Frederick P. Roth[2] & Marc Vidal[1]

Systematic mapping of protein–protein interactions, or ‘interactome’ mapping, was initiated in model organisms, starting with defined biological processes[1,2] and then expanding to the scale of the proteome[3–7]. Although far from complete, such maps have revealed global topological and dynamic features of interactome networks that relate to known biological properties[8,9], suggesting that a human interactome map will provide insight into development and disease mechanisms at a systems level. Here we describe an initial version of a proteome-scale map of human binary protein–protein interactions. Using a stringent, high-throughput yeast two-hybrid system, we tested pairwise interactions among the products of ~8,100 currently available Gateway-cloned open reading frames and detected ~2,800 interactions. This data set, called CCSB-HI1, has a verification rate of ~78% as revealed by an independent co-affinity purification assay, and correlates significantly with other biological attributes. The CCSB-HI1 data set increases by ~70% the set of available binary interactions within the tested space and reveals more than 300 new connections to over 100 disease-associated proteins. This work represents an important step towards a systematic and comprehensive human interactome project.

Our working definition of a human interactome map is the complete collection of binary protein–protein interactions detectable in one or more exogenous assay. This definition excludes dynamic and functional properties of these interactions (Supplementary Data I). Thus, we treat interactome maps as ‘scaffold’ information, from which increasingly detailed and reliable biological models can be generated by integrating other functional genomic and proteomic data sets[10] (Supplementary Data II).

The currently available information on the human interactome network originates from either literature-curated (LC) interactions[11–15], or from ‘interologs’ (that is, potential interactions predicted from interactome data available for model organisms given evolutionary conservation of two known partners)[2,16]. This information needs to be complemented by systematic experimental mapping approaches that are: (1) not biased towards any particular biological interest (that is, without ‘inspection bias’), as is the case for LC data sets; (2) more complete; and (3) supported by experiments rather than predictions. We are mapping the human interactome network systematically in successive versions, with each version defined by the availability of recombinationally cloned open reading frames (ORFs) in the human ‘ORFeome’[17].

In this initial version, we use human ORFeome v1.1 (ref. 17), a resource containing ~8,100 Gateway-cloned ORFs (generated using the Mammalian Gene Collection, as previously described[17]) that correspond to ~7,200 distinct protein-coding genes (Supplementary Table S1). Thus, our initial ‘search space’ (Space-I) encompasses protein pairs encoded by a 7,200 × 7,200 matrix of genes. Future interactome versions can be generated by successively increasing the search space as additional versions of the human ORFeome become available (Supplementary Data III). Accepting a total of ~22,000 protein-coding genes in the human genome[18] and excluding polymorphic and splice variants, Space-I corresponds to ~10% of the total search space for a comprehensive human interactome map (Fig. 1a). Currently, 4,067 binary LC interactions are available in Space-I (LCI interactions; Supplementary Table S2).

Our high-throughput yeast two-hybrid system is highly specific, benefiting from the following features, which were not uniformly present in earlier large-scale studies: relatively low levels of expression of both Gal4 DNA-binding domain (DB) and Gal4 activation domain (AD) hybrid proteins (DB-X and AD-Y, or DB-ORF and AD-ORF); three different yeast two-hybrid inducible reporter genes; and a plasmid-shuffling counter selection to eliminate systematically *de novo* auto-activators[19] (Supplementary Data IV). We tested each of the ~8,100 individual DB-X proteins against 45 mini-libraries, each containing a pool of 188 AD-Y fusion proteins (AD-188Ys), by yeast-mating in a 96-well format (Fig. 1a). Such small pools offer high sensitivity, because positive clones are less likely to be masked by other AD-Y clones within the same pool. Indeed, our overall reproducibility rate was ~55%, close to that observed in proteome-scale affinity purification followed by mass spectrometry experiments[20] (Supplementary Data V).

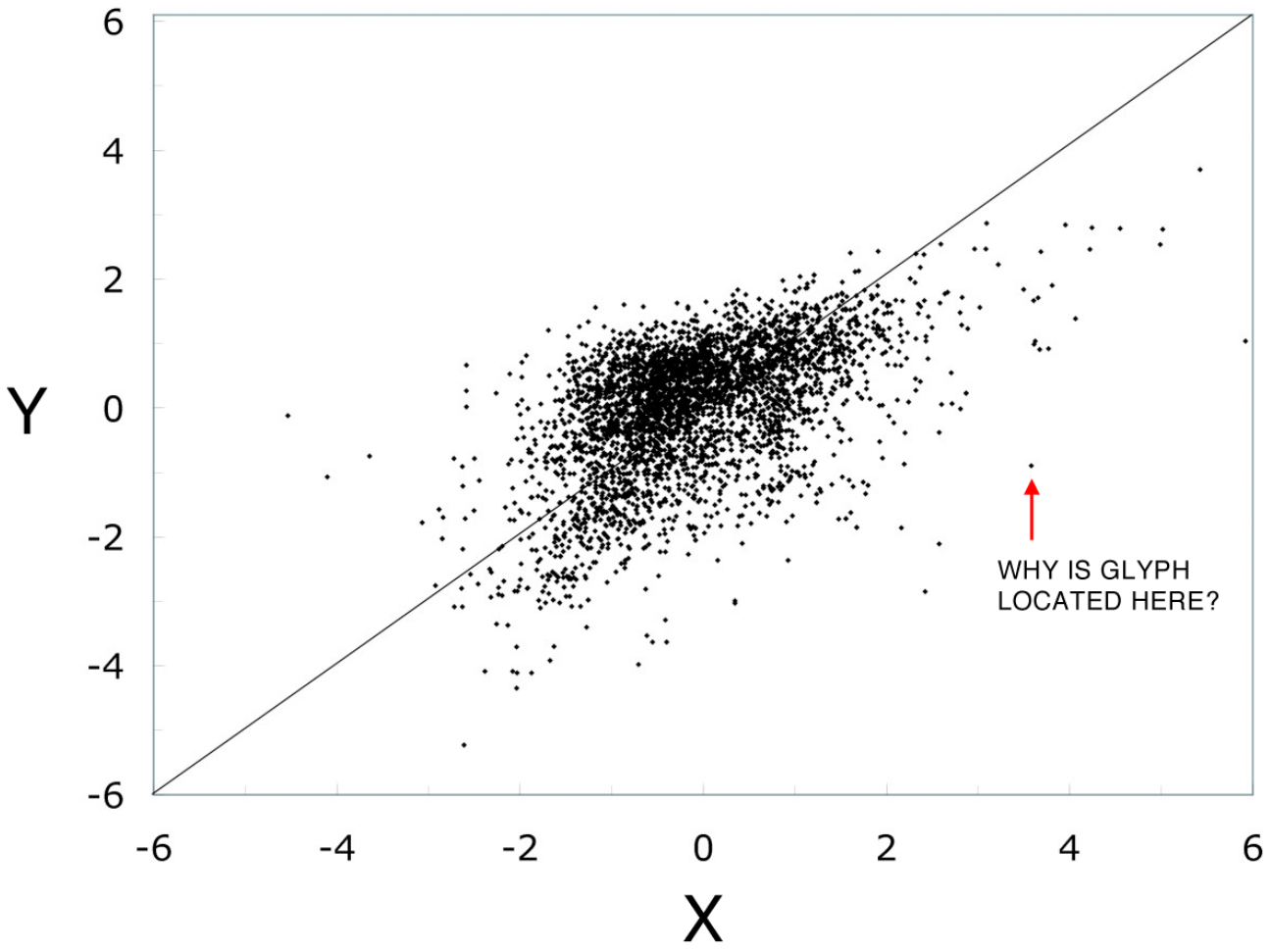
In our Space-I yeast two-hybrid matrix, we identified ~65,000 primary positive colonies, of which 12,251 scored positive after

[1]Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Ave, Boston, Massachusetts 02115, USA. [3]Unité de Recherche en Biologie Moléculaire, Facultés Notre-Dame de la Paix, 61 Rue de Bruxelles, 5000 Namur, Belgium. [4]Howard Hughes Medical Institute, and Departments of Pediatrics, Neurology, Neuroscience, and Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [5]Arcbay, Inc., 6 Whittier Place, Suite 71, Boston, Massachusetts 01915, USA. [6]Agencourt Bioscience Corporation, 500 Cummings Center, Suite 2450, Beverly, Massachusetts 01915, USA. †Present addresses: ArQule, Inc., 19 Presidential Way, Woburn, Massachusetts 01081, USA (S.L.); Departments of Cancer Biology, and Otolaryngology, Head and Neck Surgery, University of California Davis, 2521 Stockton Blvd, Suite 7200, Sacramento, California 95817, USA (J.S.A.). *These authors contributed equally to this work.

An interaction network is not a tangible physical system like the mammoth and its visualization must be approached by first asking: what questions would I like to answer and what properties would I like to communicate?

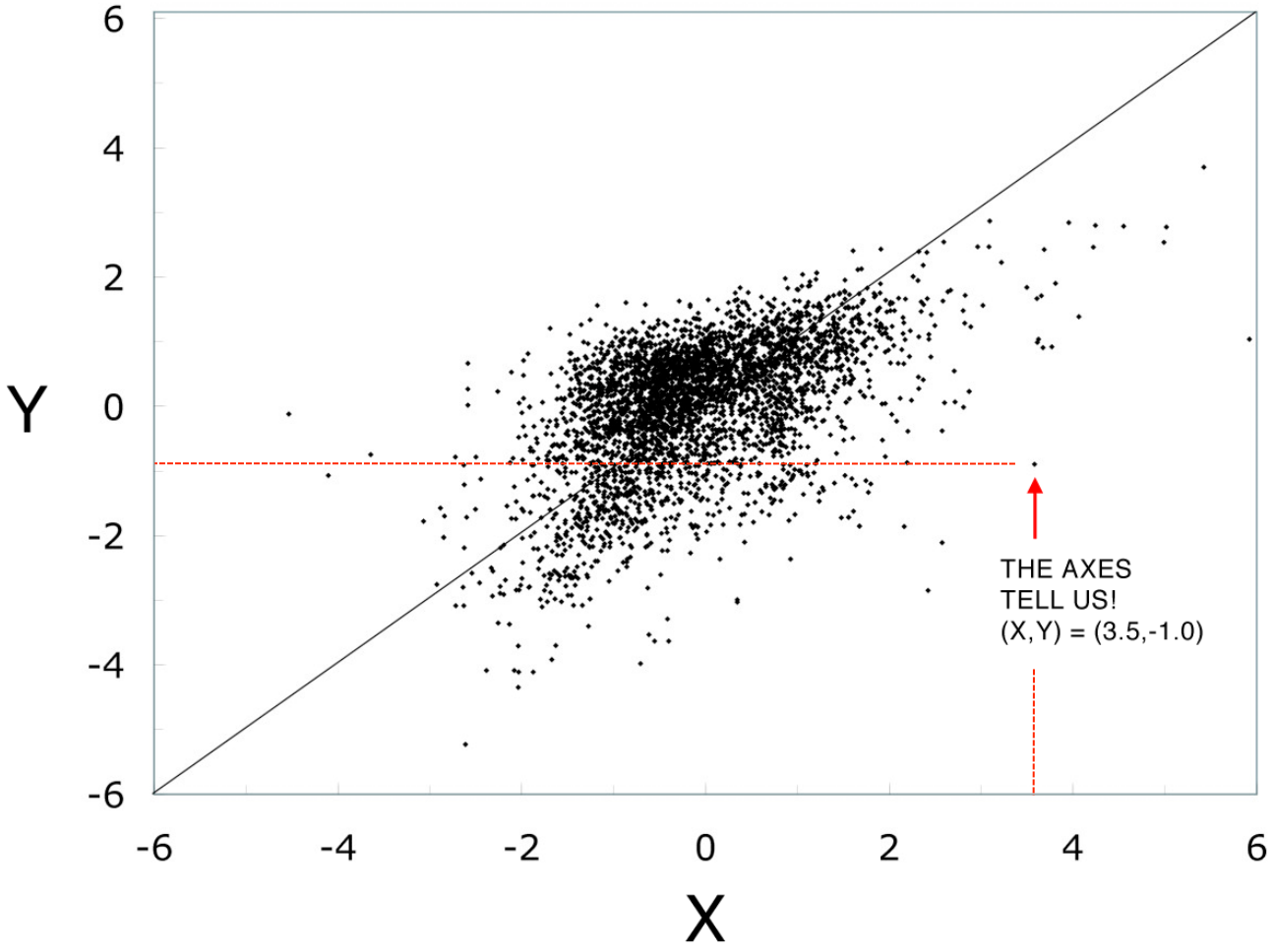
Rual et al., *Nature* **437**(7062):1173-8.

HOW DO WE NAVIGATE A VISUALIZATION?



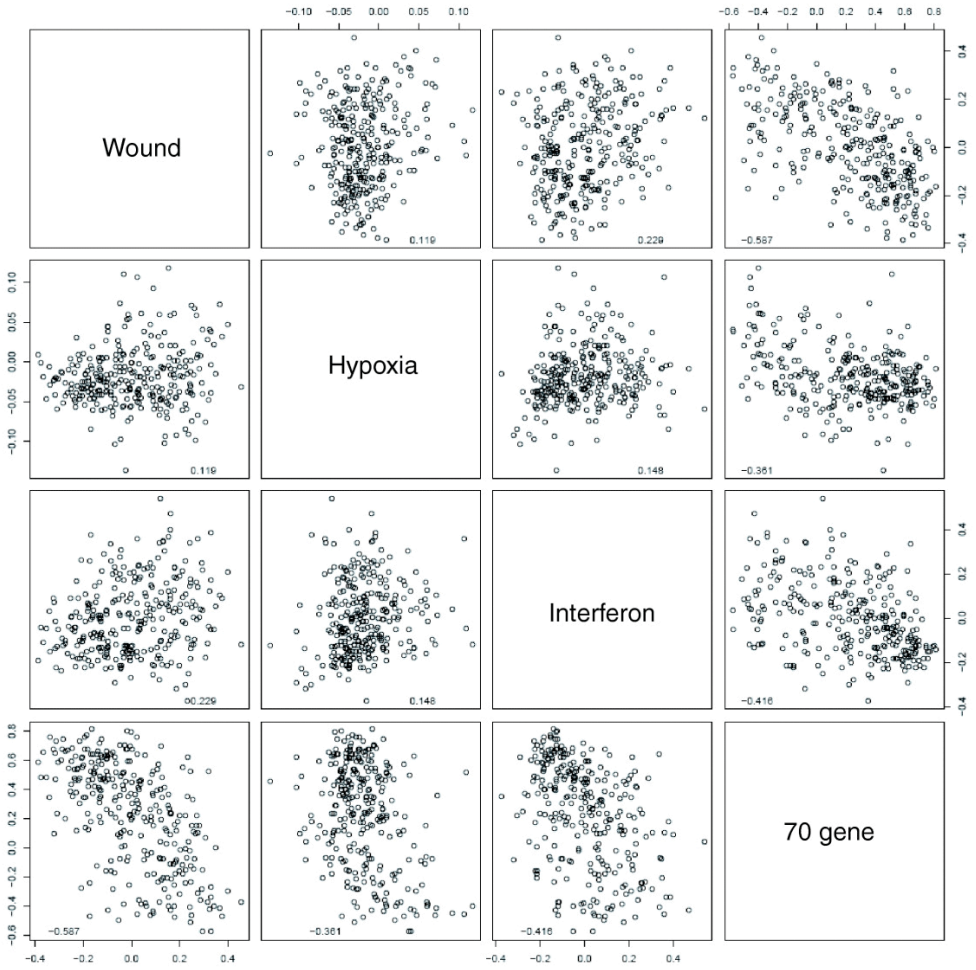
A silly question – bear with me. When looking at a scatter plot you know exactly how to identify the meaning of a single data point.

USE LANDMARKS – AXES



The axes provide a reference system.

A COMMON REFERENCE SYSTEM PERMITS COMPARISONS

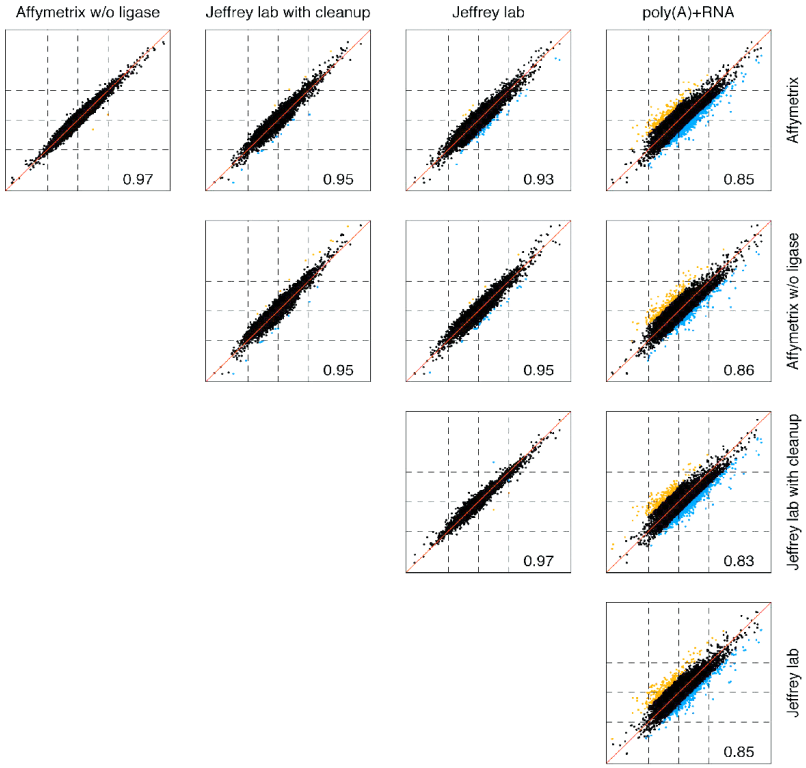


Scatter plots can be easily compared – such as in this matrix – because they have a common coordinate system. Furthermore, the position of any one point on a plot is directly (and solely) decided by meaningful properties (x,y coordinate).

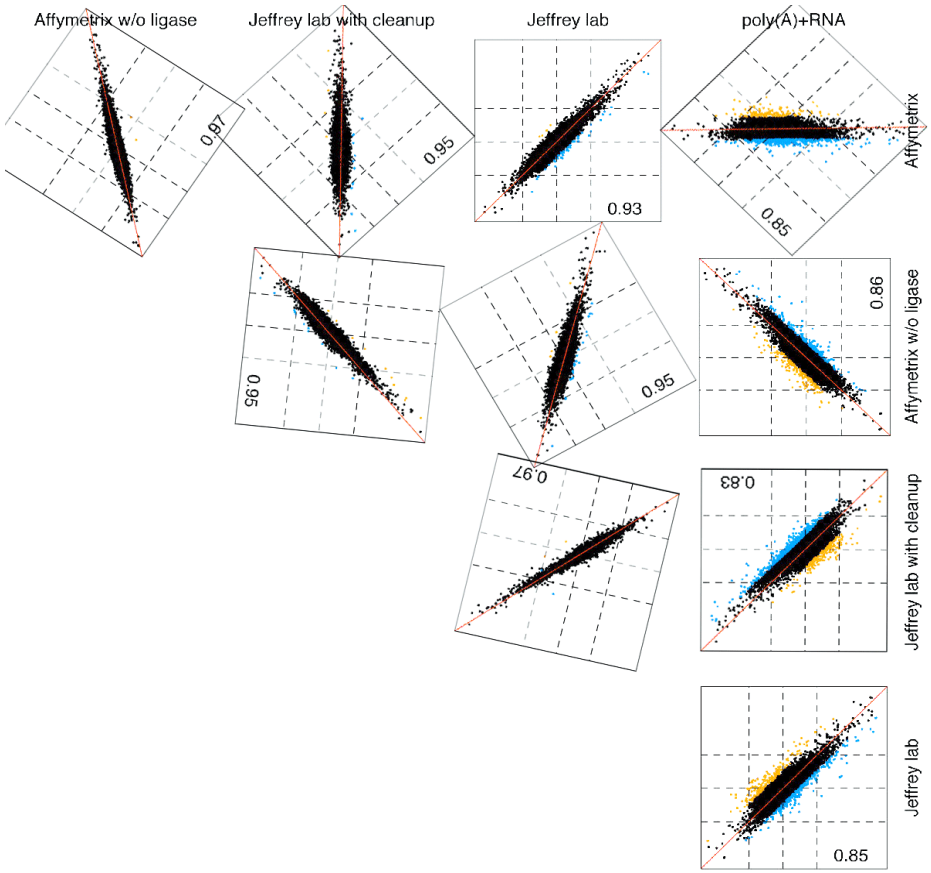
Buess *et al. Genome Biology* 2007 **8**:R191 doi:10.1186/gb-2007-8-9-r191

WHAT IS DIFFERENT? DATA OR COORDINATE SYSTEM?

Our ability to detect linearity is invariant under rotation, but the coordinate system change affects our interpretation of the proportionality in the linear relationship.

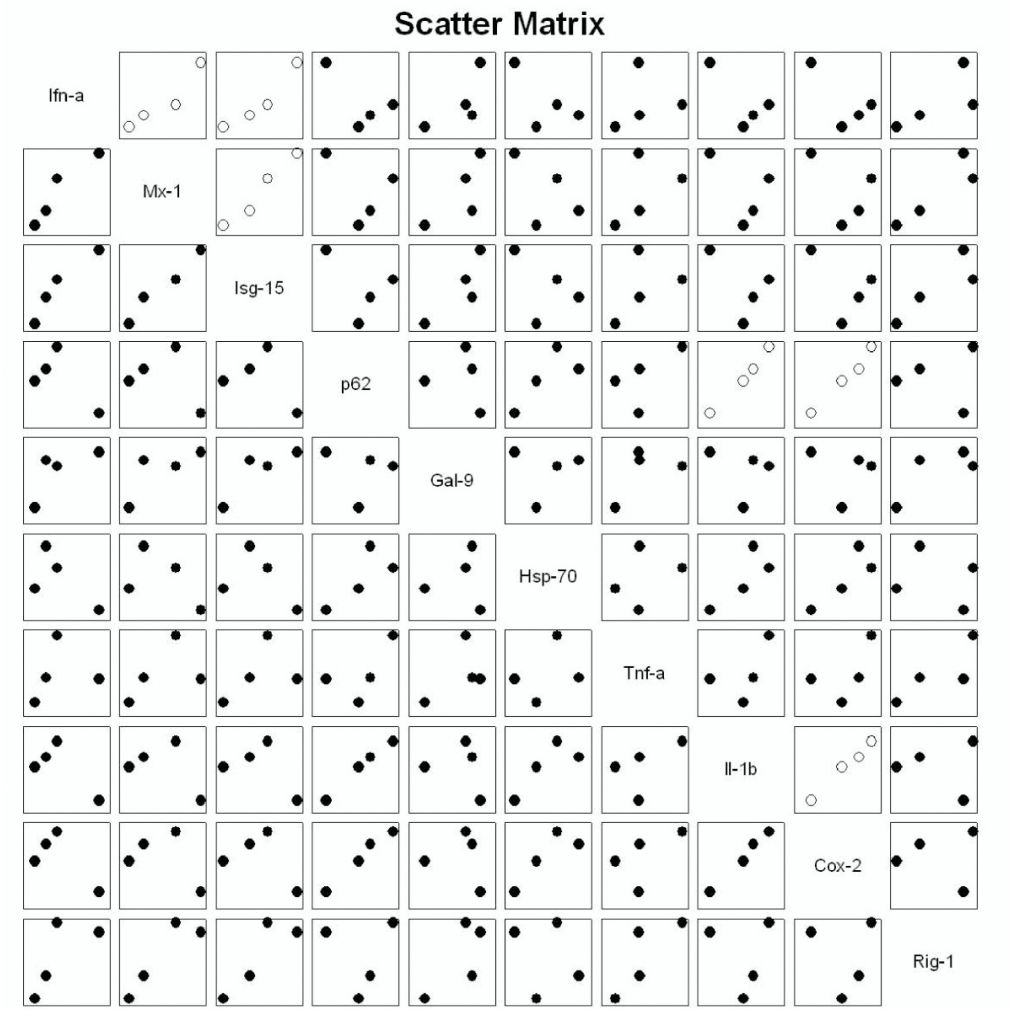


Zhao *et al. BMC Genomics* 2002 **3**:31 doi: 10.1186/1471-2164-3-31



Modified figure. Each scatter plot was arbitrarily rotated.
Zhao *et al. BMC Genomics* 2002 **3**:31 doi:10.1186/1471-2164-3-31

GOOD LUCK!



This figure illustrates the challenge posed by an inadequate coordinate system. The data points are too few and too loosely distributed for us to identify meaningful patterns. We're not very good at identifying loose clusters of patterns against a background of apparent randomness.

Schiøtz *et al. Virology Journal* 2009 **6**:91 doi:10.1186/1743-422X-6-91

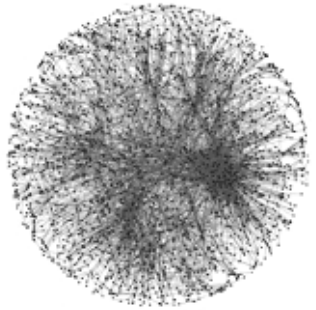
LET'S TALK ABOUT NETWORKS

THIS NETWORK

DOESN'T NEED

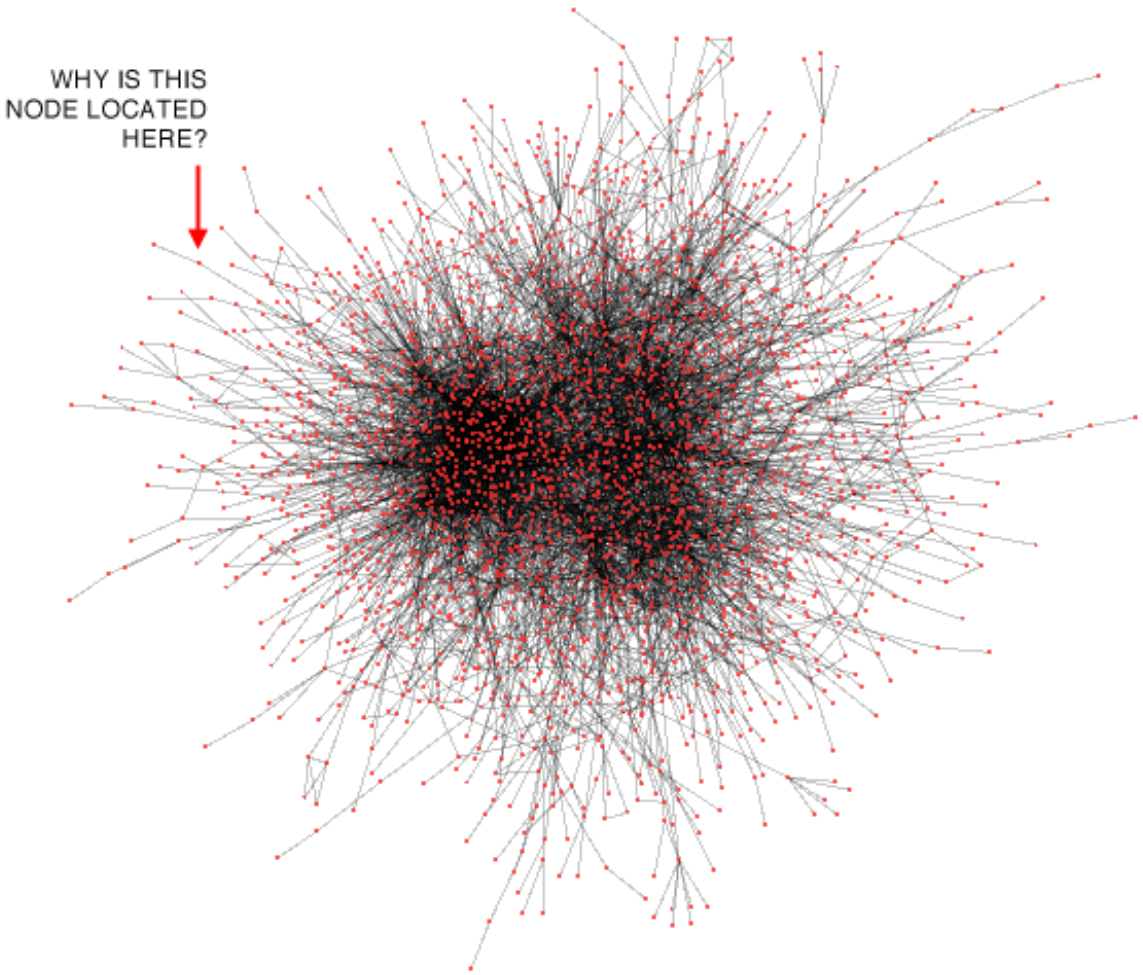
SOPHISTICATED

VISUALIZATION



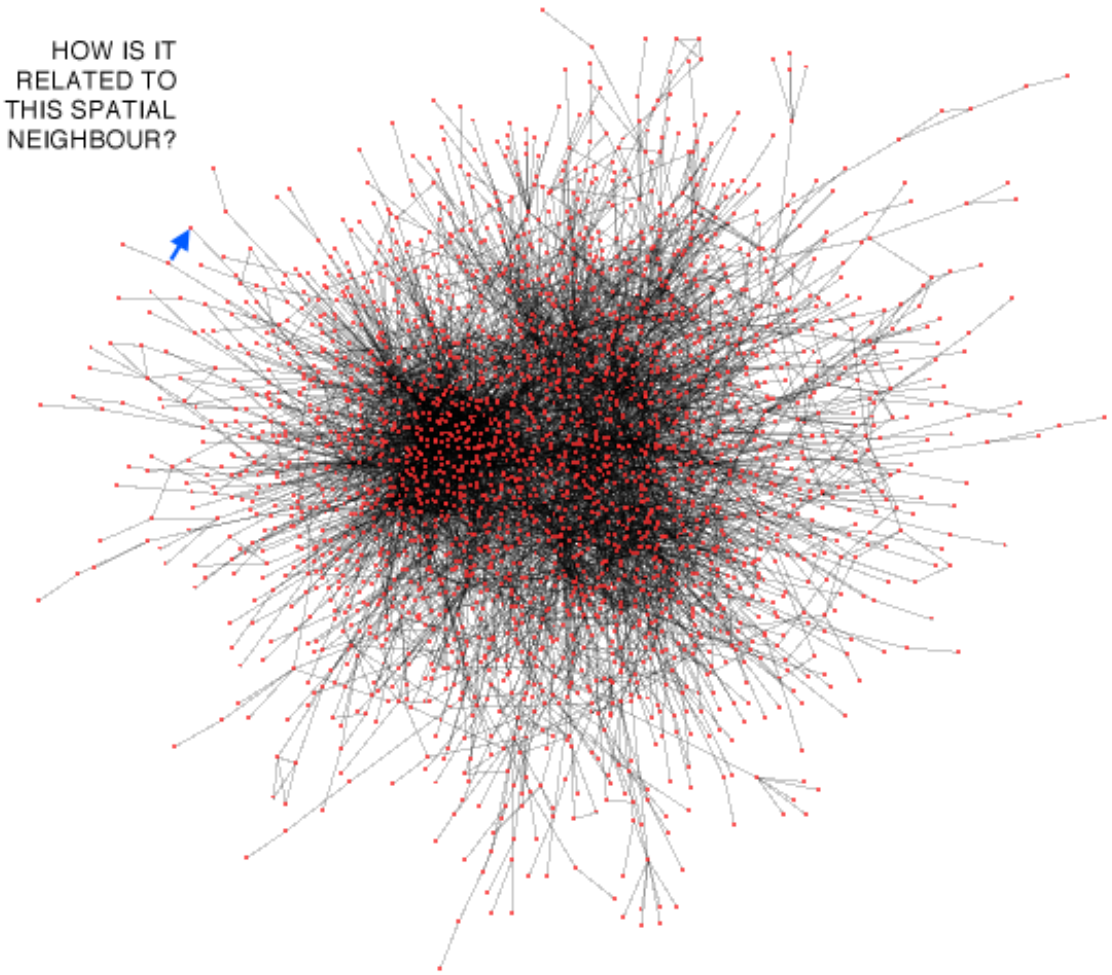
THIS ONE DOES

HAIRBALLS ARE IRRATIONAL – THERE IS NO MEANINGFUL COORDINATE SYSTEM



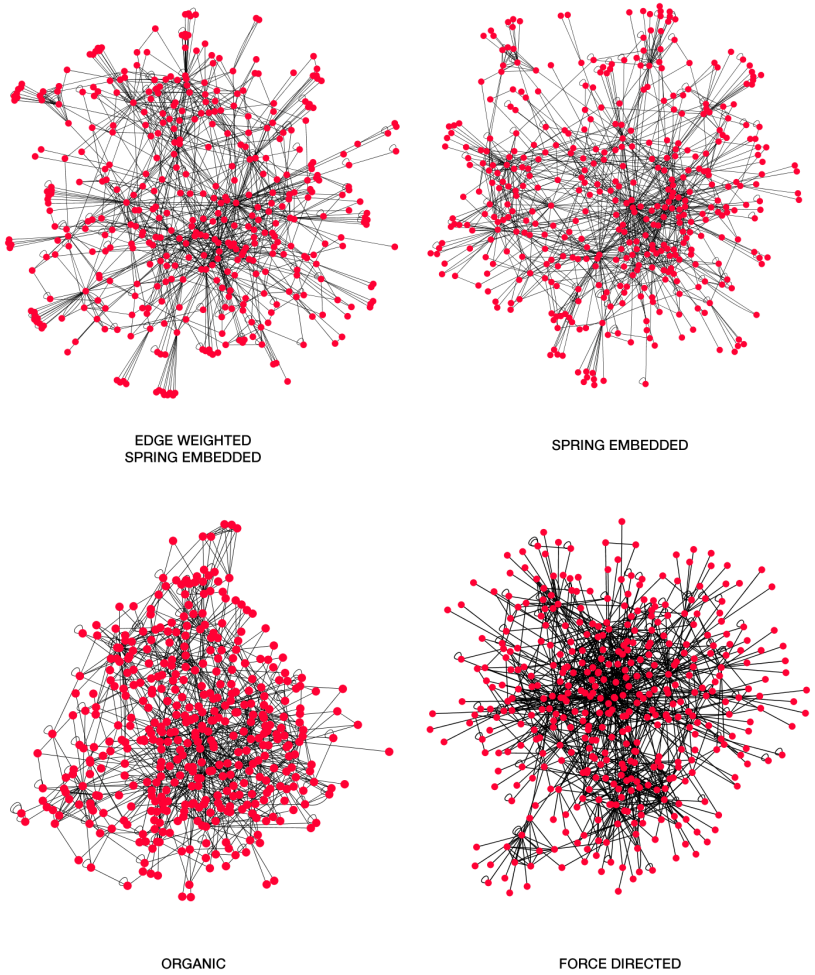
6726 human protein-protein interactions. Rendered with Cytoscape (force directed layout).
Rual et al., *Nature* **437**(7062):1173-8.

SIGNIFICANCE OF SPATIAL PROXIMITY IS ENTANGLED WITH LAYOUT



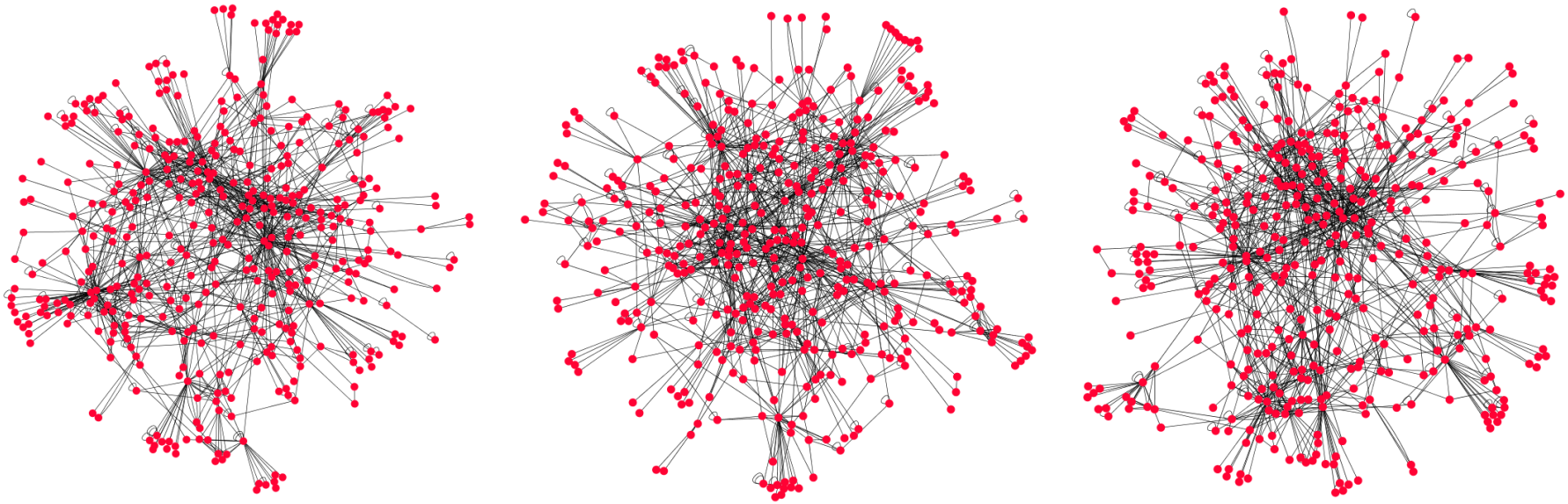
6726 human protein-protein interactions. Rendered with Cytoscape (force directed layout).
Rual et al., *Nature* **437**(7062):1173-8.

DIFFERENT LAYOUTS – SAME NETWORK?



Subset of the human protein-protein interaction network rendered by Cytoscape using a variety of layouts.
Rual et al., *Nature* **437**(7062):1173-8.

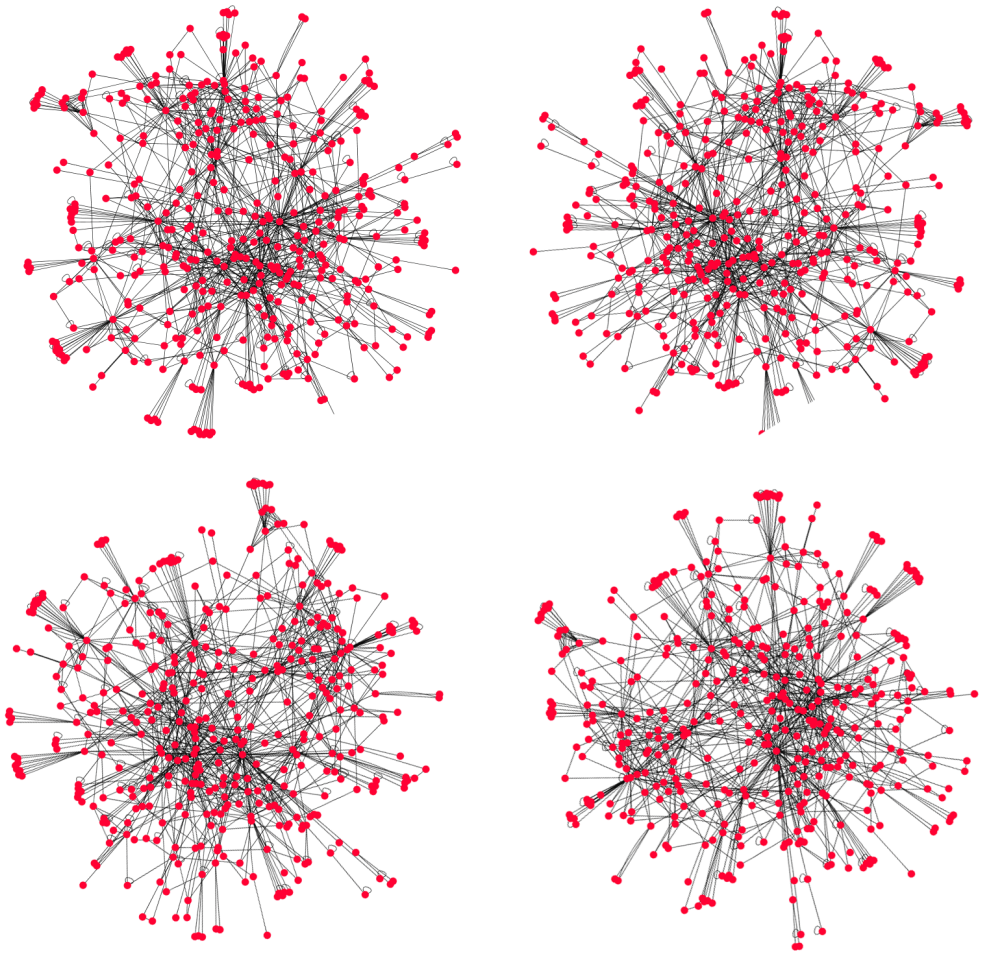
SAME LAYOUT – DIFFERENT NETWORK?



Subset of the human protein-protein interaction network rendered by Cytoscape. Each visualization uses the same layout (spring embedded), using the previous as a starting point.

Rual et al., *Nature* **437**(7062):1173-8.

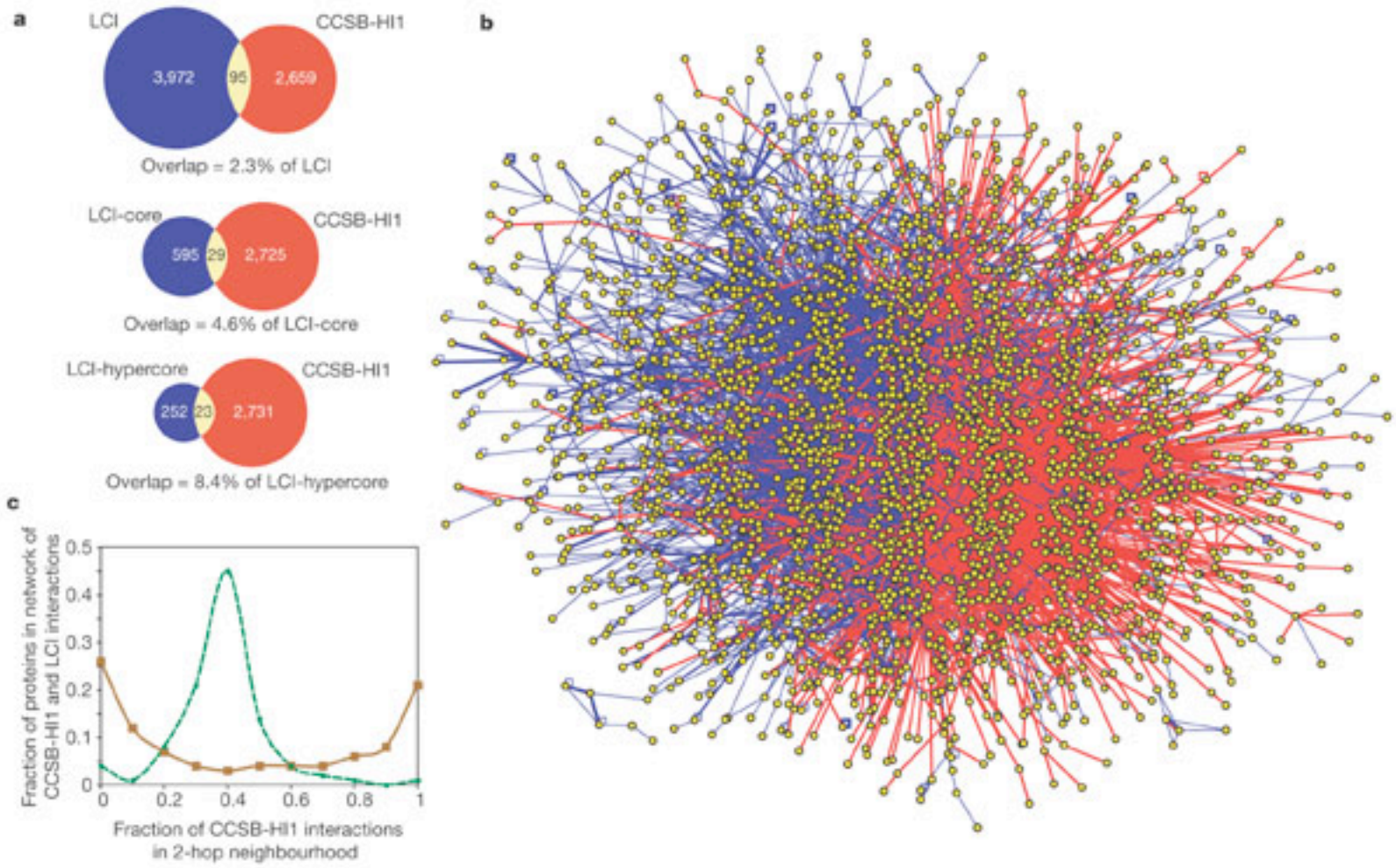
SAME HAIRBALL – CAN YOU TELL?



Subset of the human protein-protein interaction network rendered by Cytoscape (edge weighted spring embedded). Each panel shows the same visualization, but with a random rotation and flip.

Rual et al., *Nature* **437**(7062):1173-8.

VISUALIZATION APOLOGETICS



"The apparent banding pattern of the yellow nodes is an **artefact of the graph layout algorithm** (Supplementary Data). Importantly, the layout algorithm was not informed by type of supporting evidence and therefore does not explain the evident separation of blue and red edges."

Figure 2 and caption quote from Rual et al., *Nature* **437**(7062):1173-8

VISUAL CONFUSION

disambiguating layout algorithm from data is impossible

/ a separation of blue and red edges

/ columns of nodes

Authors acknowledge that both effects are artefacts of the layout algorithm.

This suggests the following

/ what are we seeing?

/ what are we supposed to see?

/ is there a better layout algorithm?

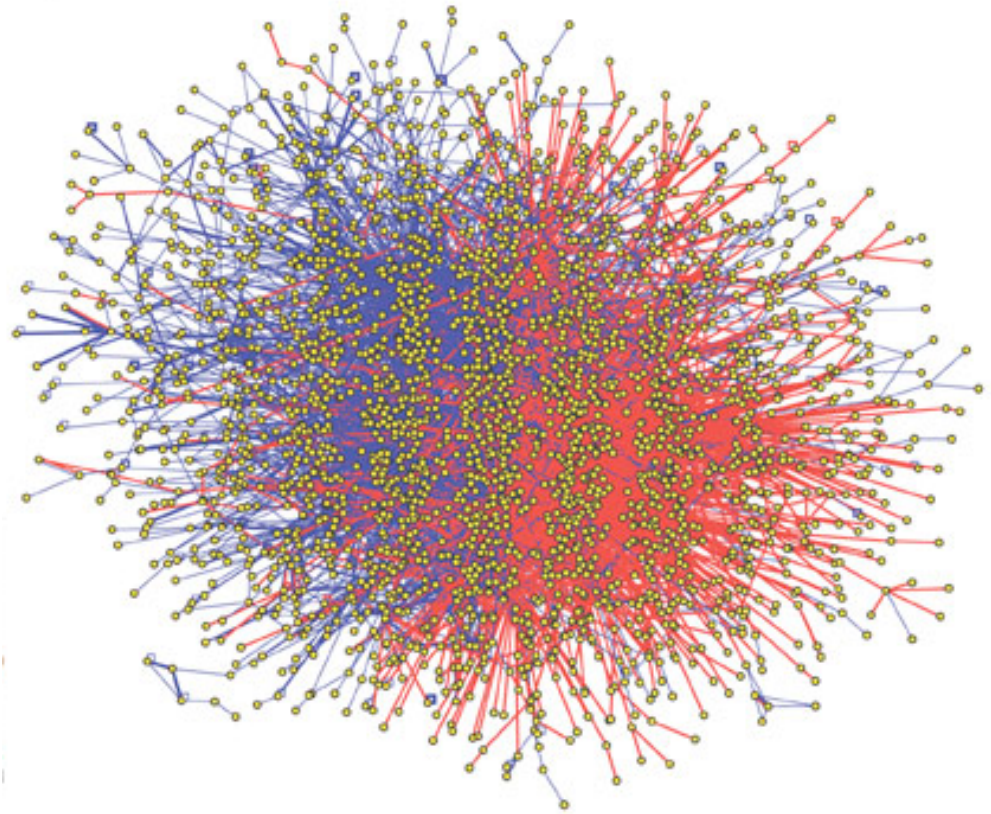
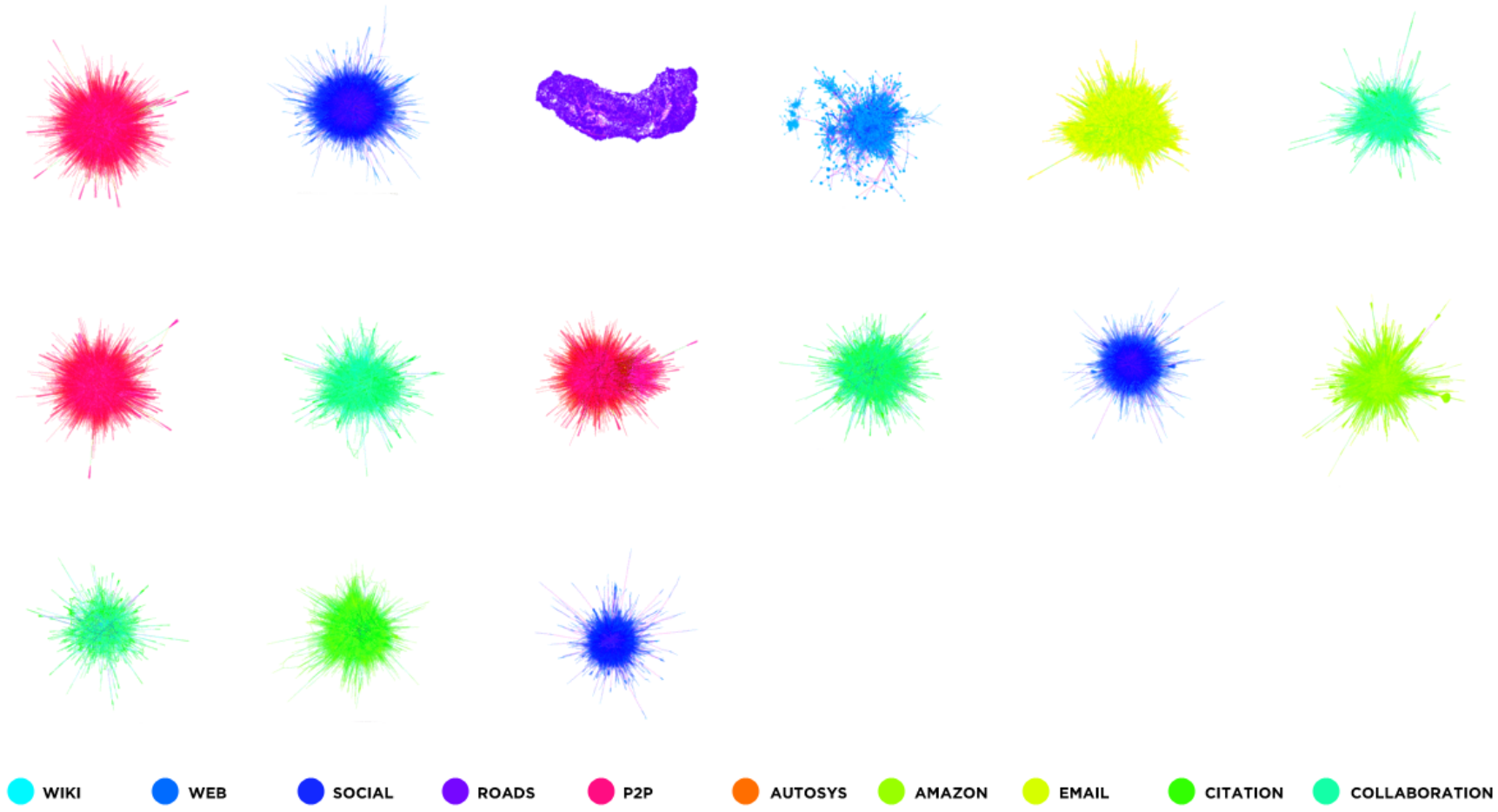


Figure 2 and caption quote from Rual et al., *Nature* **437**(7062):1173-8.

WHAT ARE WE HOPING TO SEE? WHAT IS IMPORTANT?



How are these networks different?

Network visualizations from SNAP (http://snap.stanford.edu/data/index.html). For full size poster see http://mkweb.bcgsc.ca/linnet/img/network-communities.png

CONVENTIONAL NETWORK VISUALIZATION – PROBLEMS

IMPENETRABLE COMPLEXITY

rapidly grow in visual complexity

become visually impenetrable

DEPICTIONS OF LARGE NETWORKS EXCEED RESOLUTION OF OUTPUT AND VISUAL PERCEPTION

DATA SUBORDINATE TO LAYOUT

hairball's form is determined by the layout algorithm

node and edge metadata are subordinate to layout

important characteristics of the network cease to drive the visualization and cannot be evaluated

HAIRBALL VISUALIZATIONS DO NOT CLEARLY REFLECT ASPECTS OF INTEREST

COMPARISON IMPOSSIBLE

layout algorithm is a major influence of the final visualization

similar networks may have different layouts

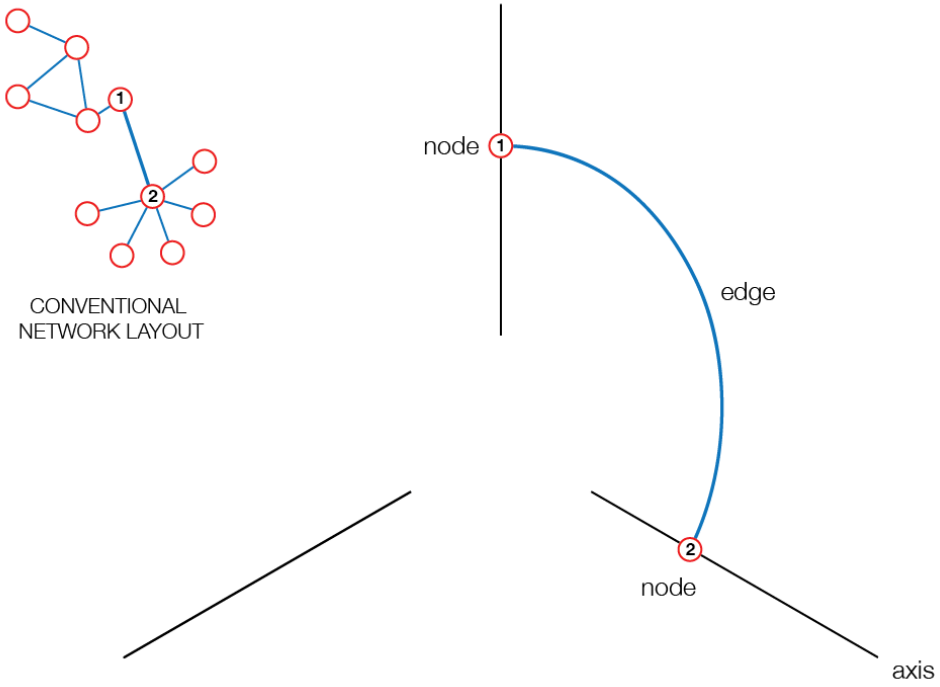
DIFFERENCES BETWEEN TWO HAIRBALLS DO NOT NECESSARILY REFLECT A DIFFERENCE IN THE DATA, NOR CLEARLY CAPTURE THE EXTENT OF THE DIFFERENCE

THE SOLUTION – CONCEPT

the linear network layout addresses the shortcomings of the conventional layout

/ nodes are constrained to linear axes

/ edges are drawn as curves between nodes



In the linear network layout, nodes are constrained to linear axes. Edges are drawn as curves between connected nodes.

THE SOLUTION – MAPPING

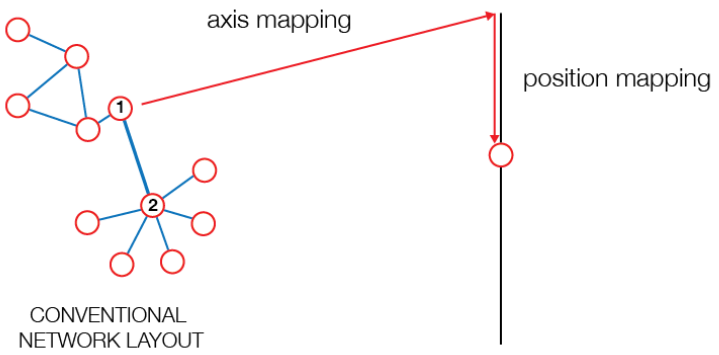
placement of nodes in the linear layout is informed by connectivity and/or annotation

/ layout is controlled solely by meaningful properties

/ interpretation of the visualization is easy, because the layout rules are based on data properties

/ direct comparison between networks is possible

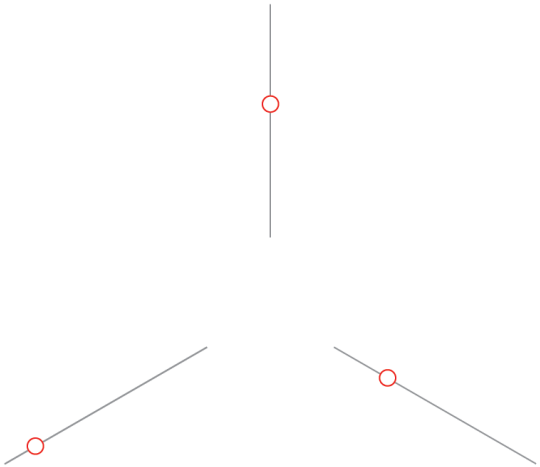
/ describing how the layout was obtained uses meaningful language (*i.e.* based on data properties not aesthetics)



Nodes are mapped and positioned on axes based on structural characteristics and/or annotations. The mappings are meant to be informed by properties of interest, creating a layout that directly illustrates meaningful aspects of the data set.

LAYOUT BASED ON STRUCTURE AND FUNCTION

NODE TO AXIS

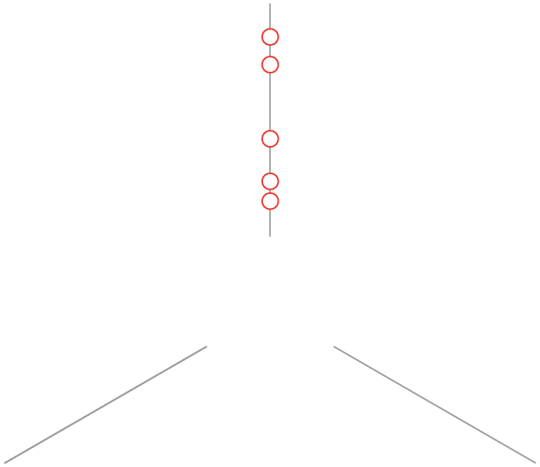


node type (source, sink, both)

node annotation class (*e.g.* gene classification)

AXES CATEGORIZE NODES (NOMINAL SCALE)

AXIS NODE POSITION



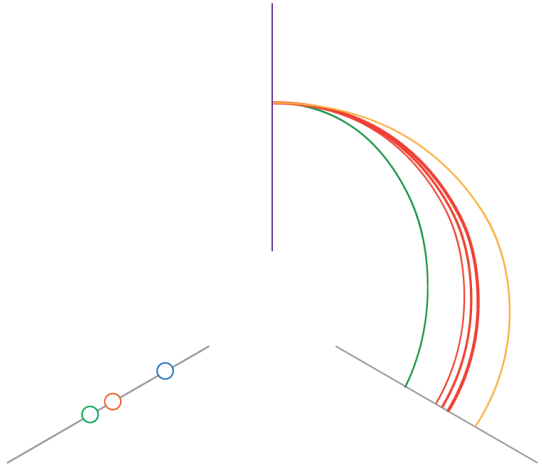
absolute or rank ordered node connectivity

neighbour connectivity

annotation property (*e.g.* expression level)

NODE POSITION ENCODES LOCAL STRUCTURE (ORDINAL OR INTERVAL SCALE)

COLOR AND SHAPE



edge color and transparency controlled by edge weight

glyphs or color codes at node positions classify nodes or layer additional data

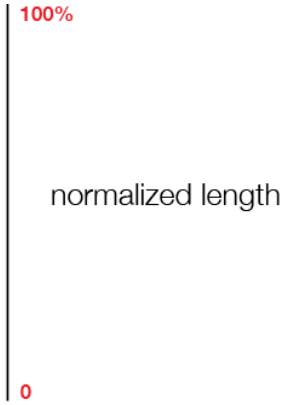
SCALE, ORIENTATION AND SEGMENTATION

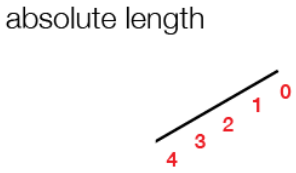
axis subdivision, scale and orientation can be adjusted to add texture and reveal patterns

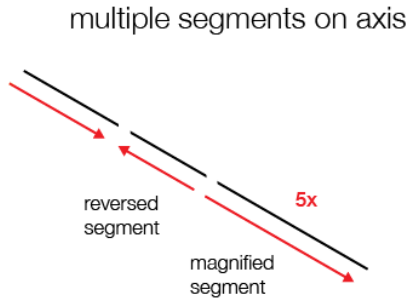
/ axis length can be absolute (e.g. number of nodes on axis), or normalized

/ an axis can be further divided into segments to further classify nodes (e.g. expression state)

/ individual axes or segments can be reversed or scaled







Each axis may have modfied length, orientation, scale and segmentation.

APPLICATION

Yan et al.[1] compare *E. coli* gene regulatory network to the Linux kernel function call network. The linear layout method presented here greatly facilitates in the visual assessment of differences between these networks.

/ the networks are directional (geneA regulates geneB or functionA calls functionB).

/ nodes are classified based on in/out degree

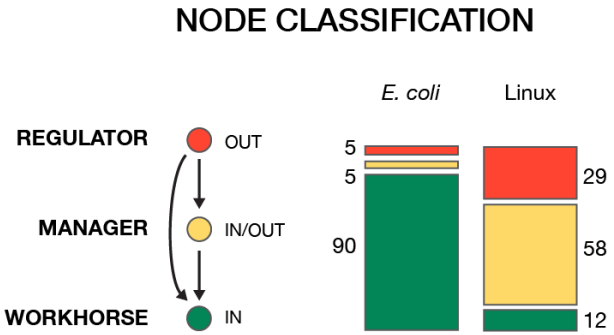
out only (source) – regulator

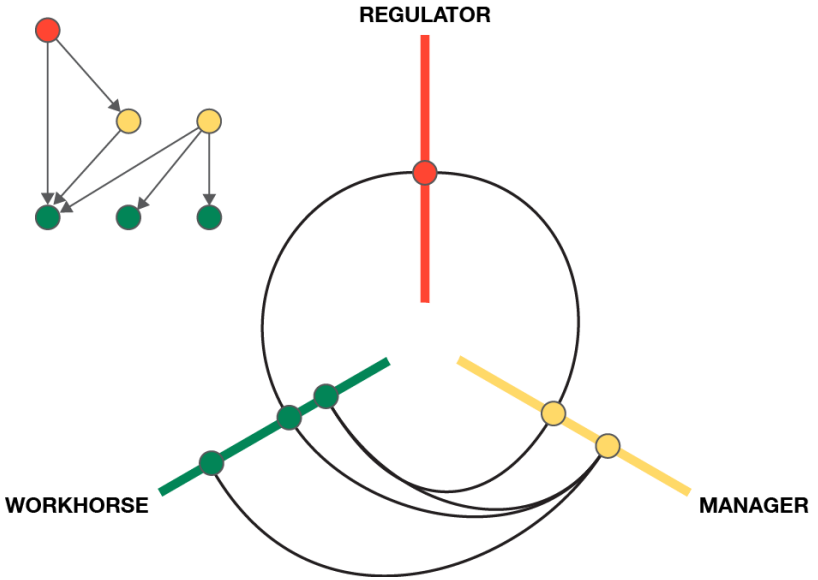
in/out – manager

in only (sink) – workhorse

/ node-to-axis mapping uses this node classification

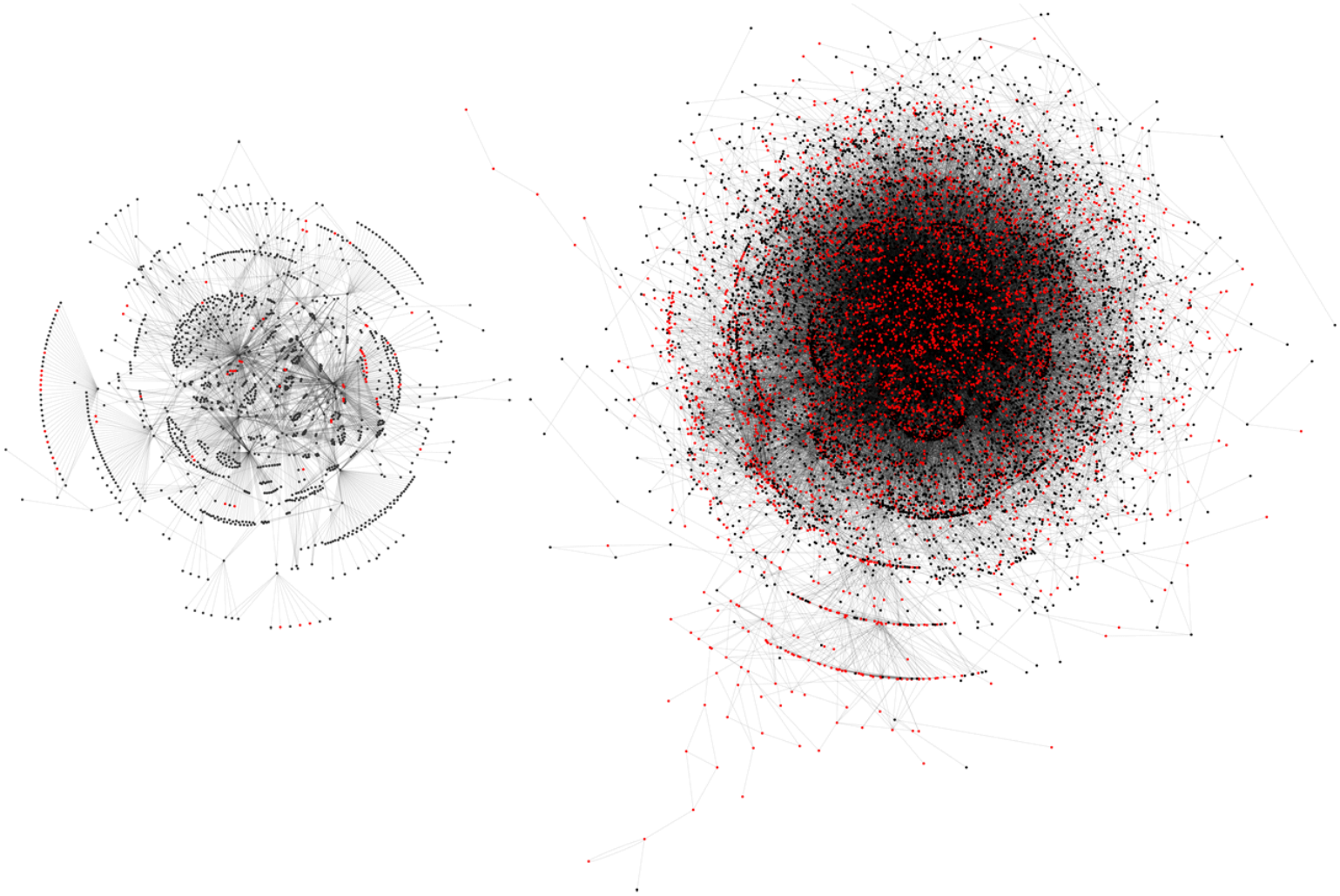
[1] Yan KK, Fang G, Bhardwaj N, Alexander RP, Gerstein M. 2010. Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *Proc Natl Acad Sci U S A* ***107(20):*** 9186-9191.





Nodes in the network are classified as regulators (out only), managers (in/out) and workhorses (in only). The *E. coli* network is bottom-heavy (many workhorses), whereas Linux is top heavy (many regulators and managers).

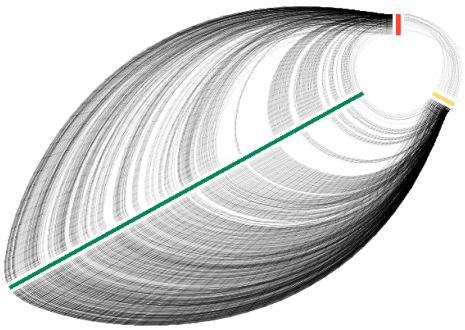
CONVENTIONAL COMPARISON



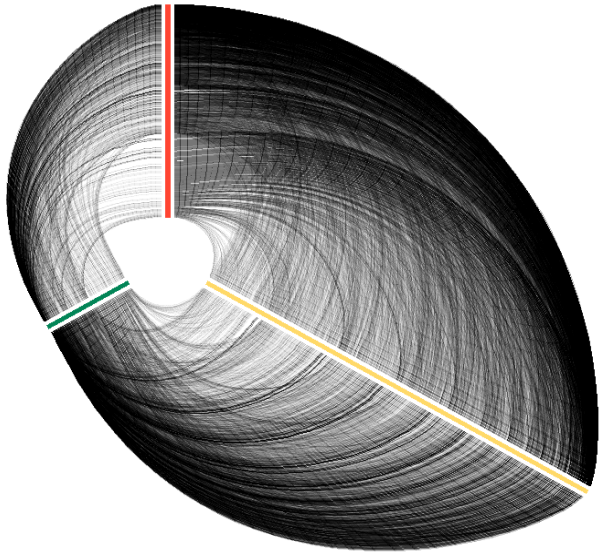
Conventional layouts are not helpful in determining structure of the *E. coli* (left) and Linux (right) networks. Even though the networks are vastly different, except for the network size, all properties are opaque.

LINEAR LAYOUT

Nodes are assigned to axes based on connectivity. Node position is based on rank order of the number of edges at a node (degree). Axis length is proportional to the number of nodes on the axis.



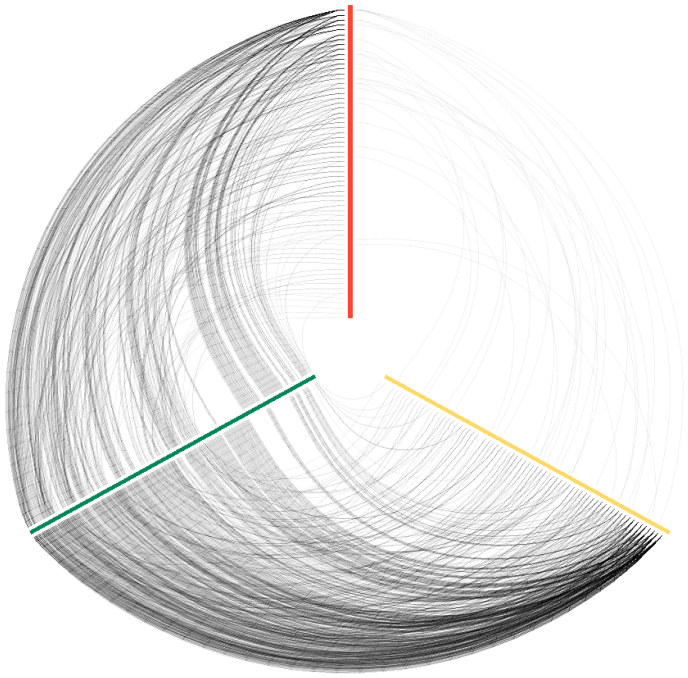
E. coli (6x magnification) / The length of the workhorse (green) axis demonstrates an over-representation in this category. Very few regulator-manager (red-yellow) connections exist. Workhorse connectivity is uniform.



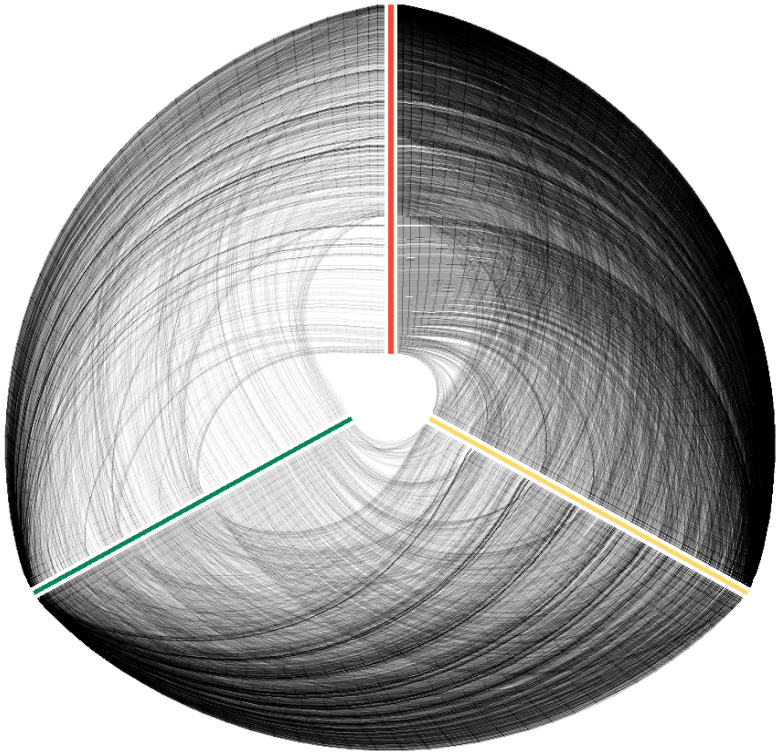
Linux / Large number of regulator-manager connections. Small number of workhorse nodes have very high connectivity (increased density of edges at end of workhorse axis). Approximately 1/3 of the regulators (red) have high connectivity to about 5% of the managers (orange), as evidenced by the converging edge density between the two axes.

NORMALIZED AXIS LENGTH

The layout method is the same as in the previous slide, but here axis length is normalized to decouple node category size from connectivity patterns. This view allows direct comparison based on node category fractions.



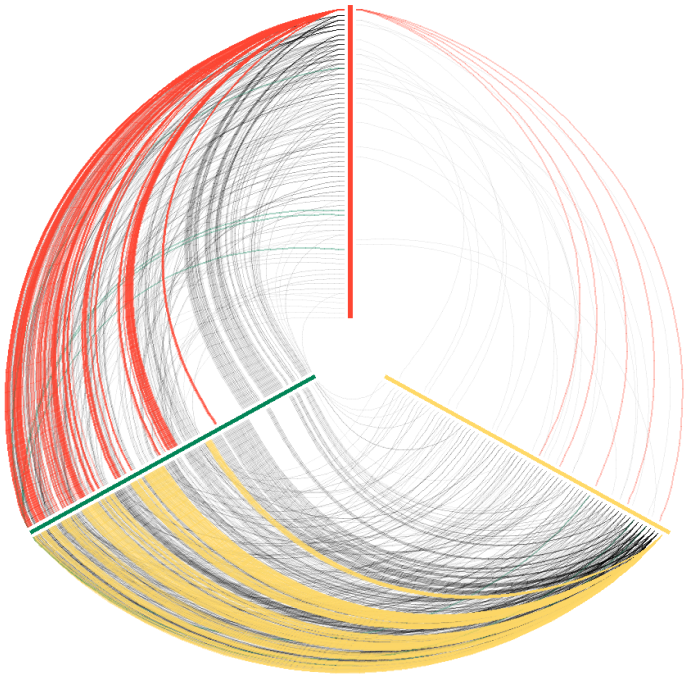
E. coli / Small number of managers are highly connected.



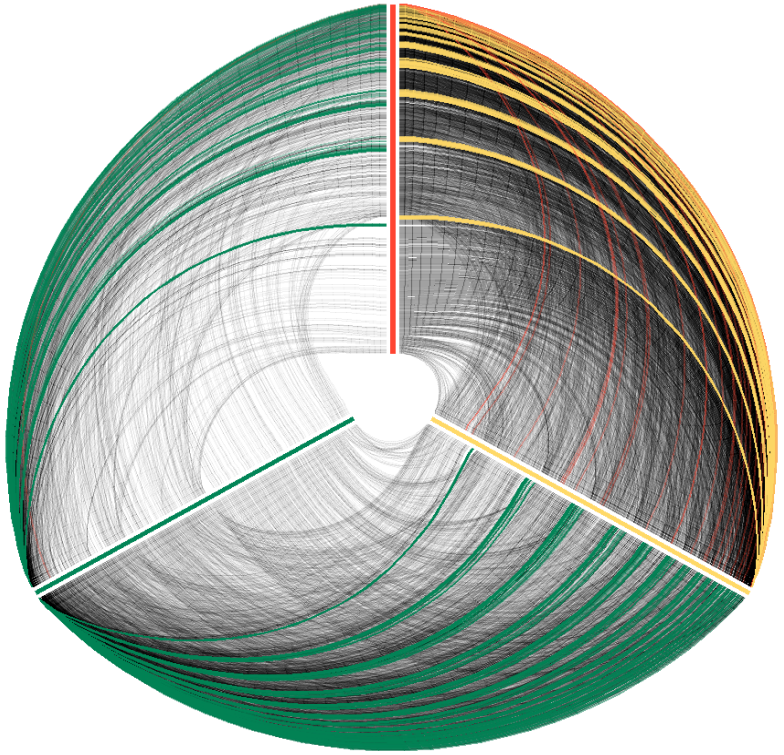
Linux / Heavily connected workhorses are more clearly evidenced when axis length is normalized.

STRUCTURAL ANNOTATION

Layering structural information is easily done using color. Here, links to the most connected node in each group (i.e. most connected regulator, manager, workhorse) are colored by the node's axis color, demonstrating neighbour connectivity around a node category's most connected member.



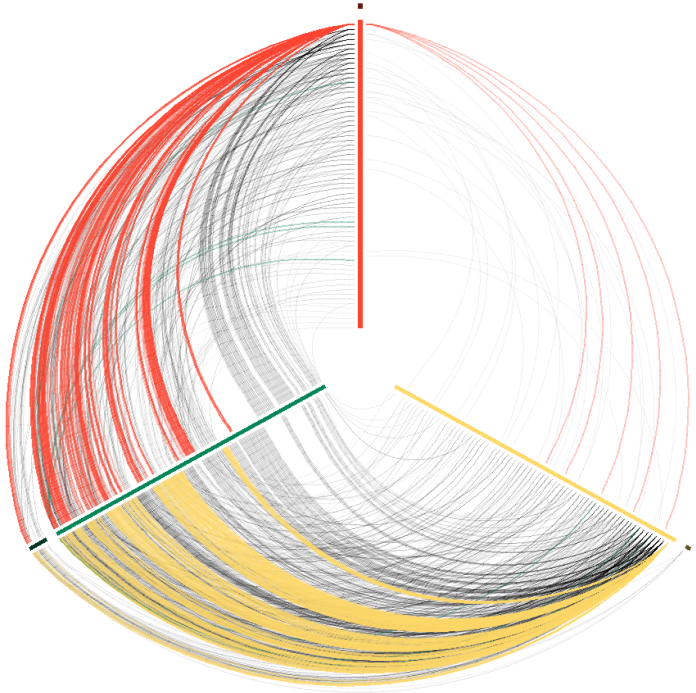
E. coli / The most connected regulator (red) primarily connects to workhorses, but also 5 distinct managers. The most connected manager connects to workhorses.



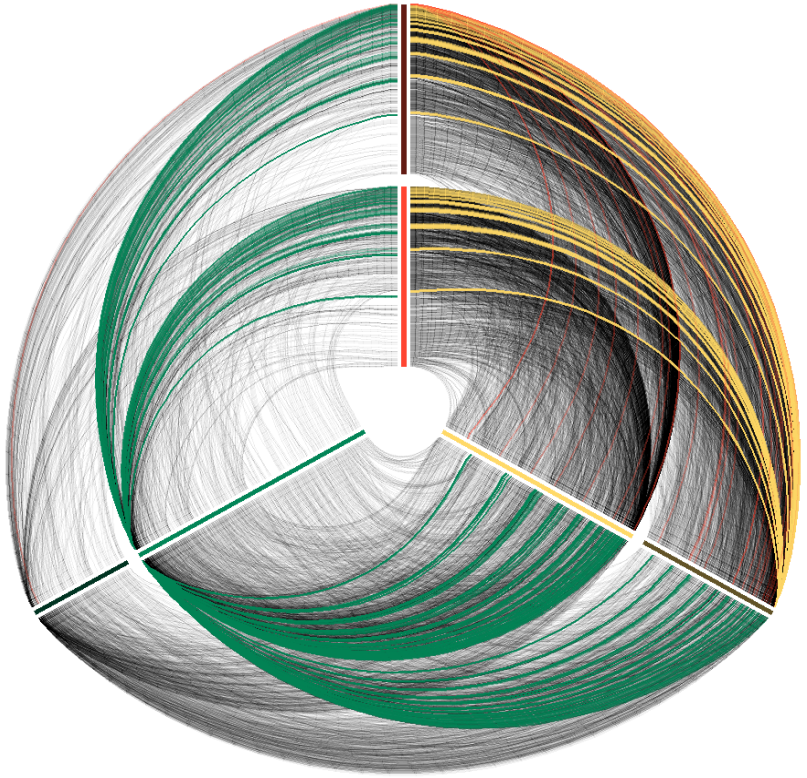
Linux / Unlike *E. coli*, here the most connected manager is connected to regulators. Note the regular banding pattern in the links, suggesting substructure.

SEGMENTATION

Yan et al. further classified each node as either non-persistent or persistent. This is shown here by splitting each axis into two segments that correspond to these two classifications.



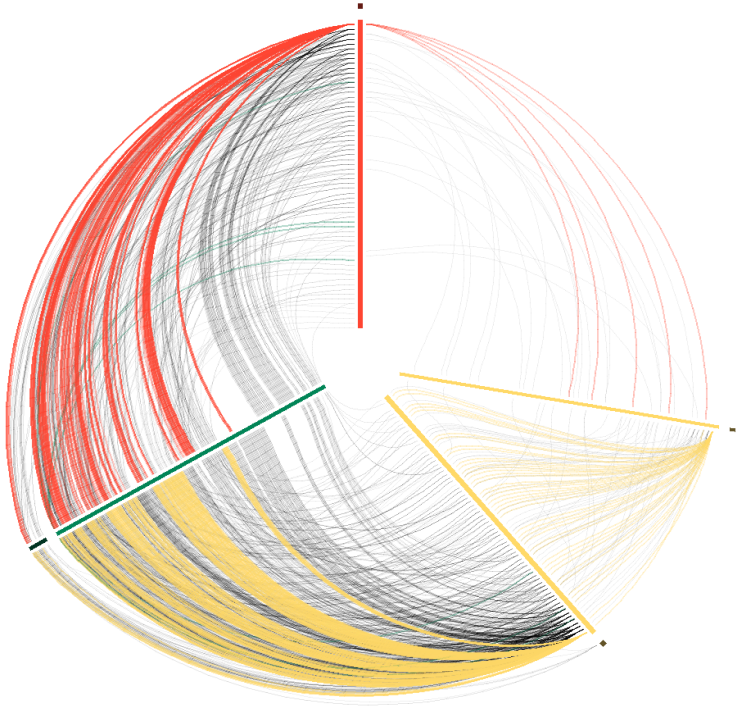
E. coli / Relatively few nodes are classified as persistent (outer segments on each axis).



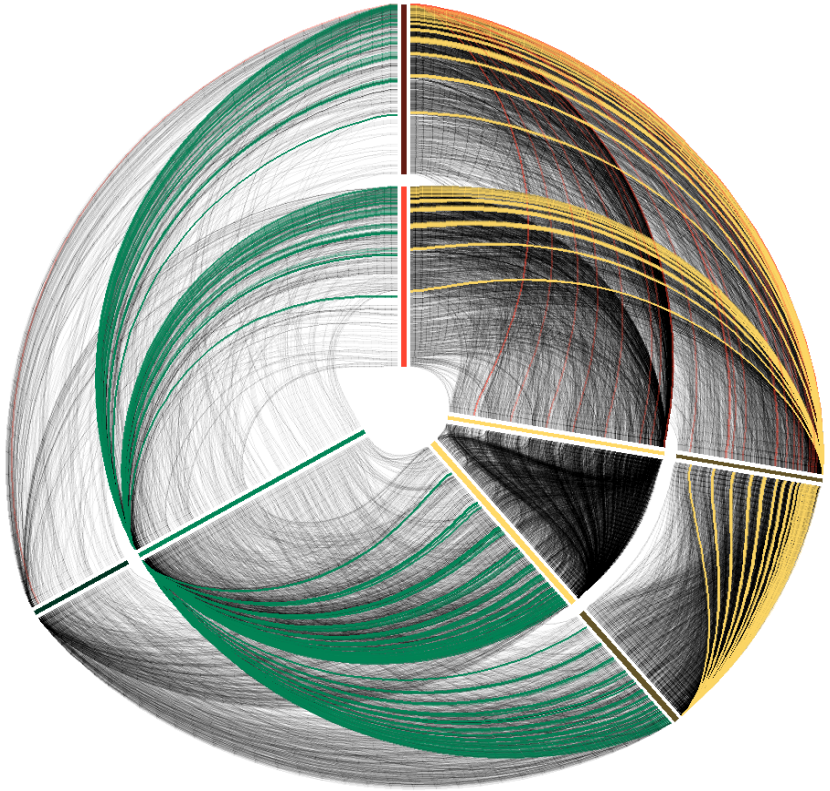
Linux / Each axis contains a near-equal mix of node types. Note that the most connected workhorse is non-persistent (inner segment), whereas the most connected manager is persistent (outer segment).

INTRA-AXIS CONNECTIONS

Managers (in/out nodes) can connect to other managers. These intra-axis links were previously not shown, but can be revealed by cloning the manager axis and displaying manager-manager connections between the cloned axes. The network is directional, with the edge direction clockwise between the two axes.



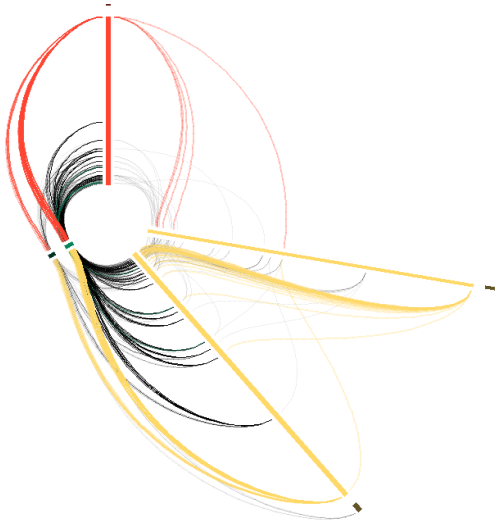
E. coli / The manager-manager connections are largely composed of the most connected manager (its high degree is due to out edges) connecting to other managers. This suggests a cascade.



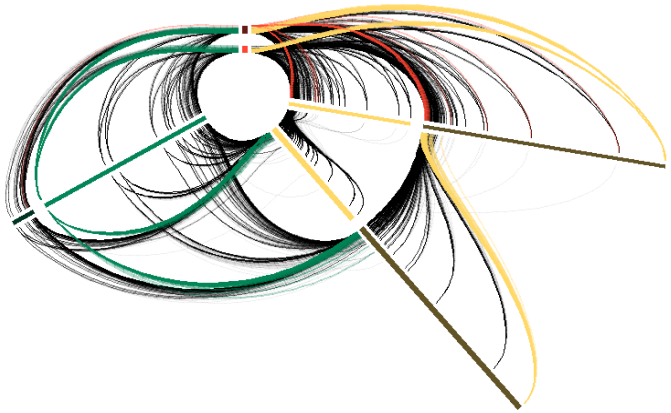
Linux / The most connected manager has a large number of in edges (it's found on the second of the cloned axes, clockwise) and its connectivity to other managers is exclusive to persistent managers.

APPLICATION – ABSOLUTE CONNECTIVITY

Here, node position is based on absolute degree of a node (number of edges). Axis length is therefore proportional to the maximum node degree within a node group.



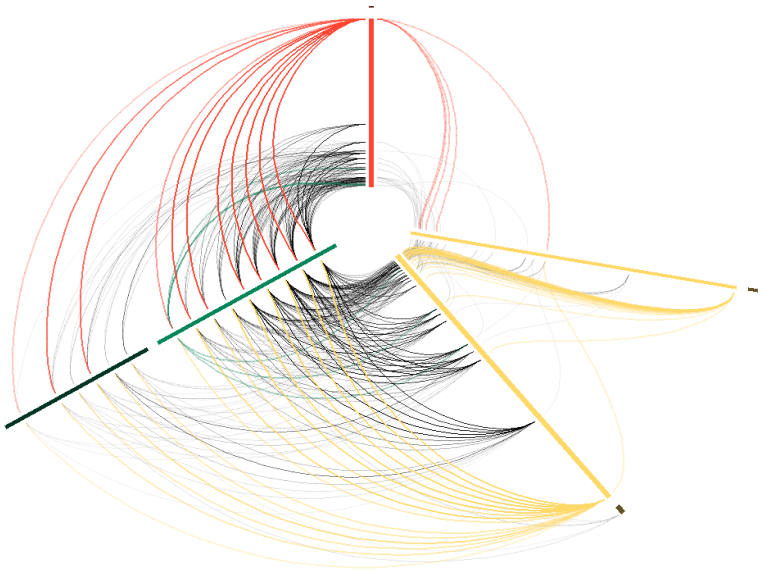
E. coli (3.5x magnification) / The distribution of node degrees becomes evident, with the highest connectivity seen in managers.



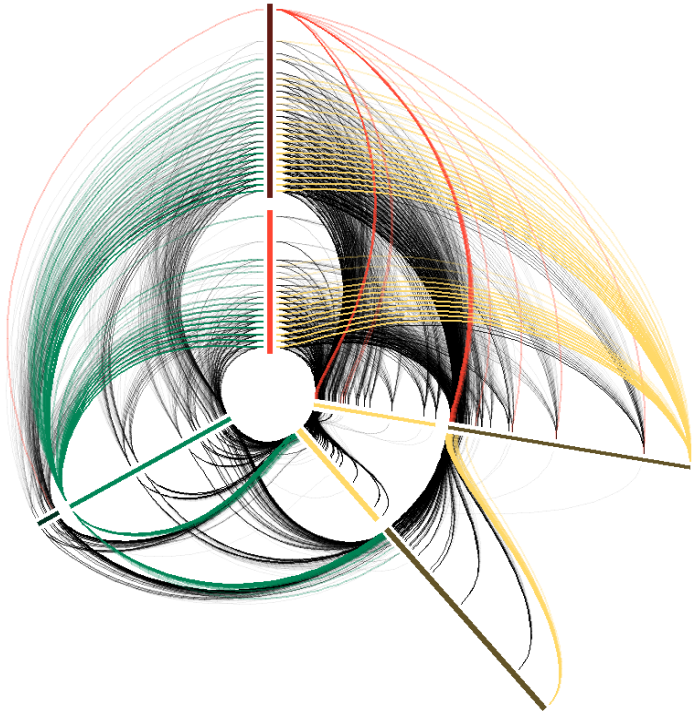
Linux / Intra-manager (yellow) edges reveal that large number of managers with low degree connect to managers with a high degree. From the manager-workhorse links, it is clear that only low degree managers connect to workhorses, whereas high degree managers connect to regulators.

AXIS MAGNIFICATION

Detail can be revealed by magnifying an axis, or individual segments. When the range of node degrees is small, links connect axes at discrete positions.



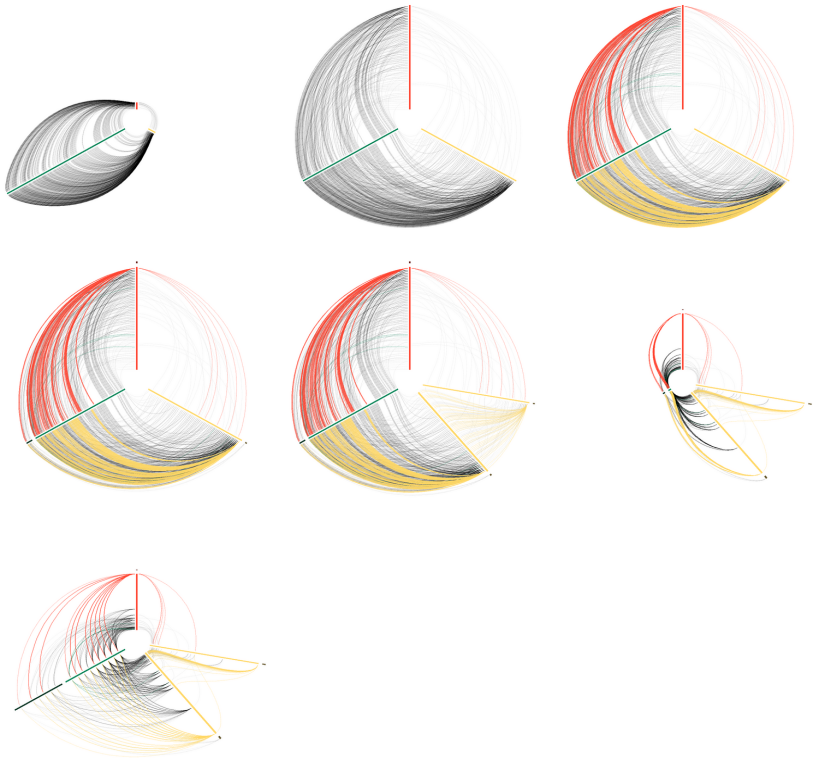
E. coli / Workhorse axis is magnified 25x.



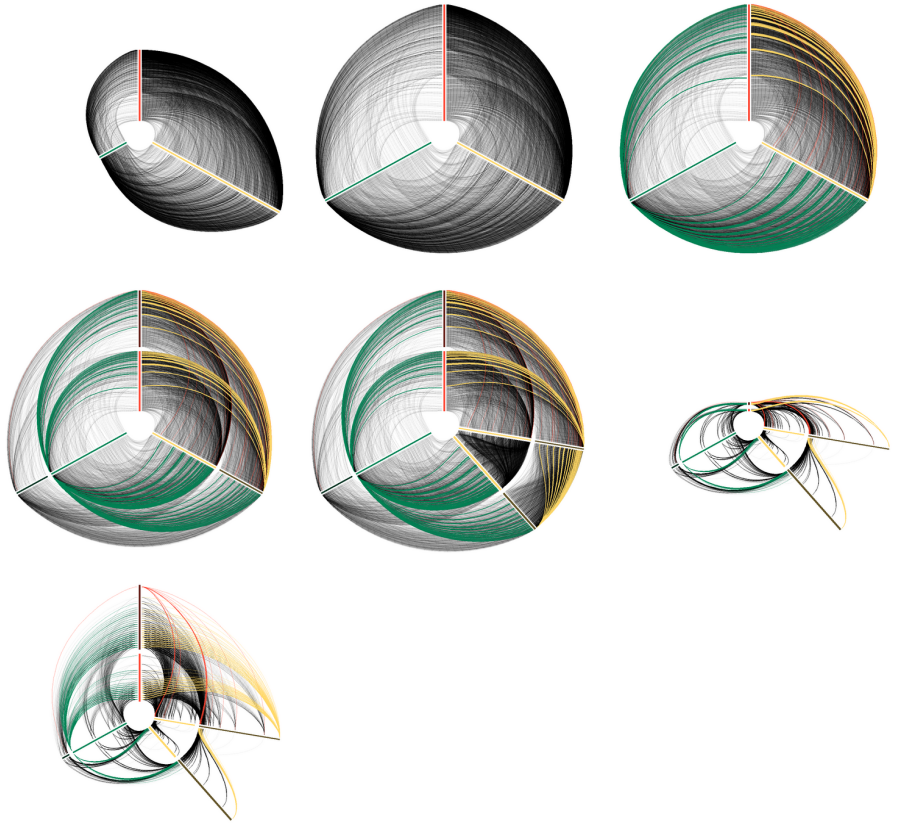
Linux / Regulator axis is magnified 25x.

VISUAL COMPARISON REVEALS DIFFERENCES

Each view reveals different aspects of the two networks, and contrasts distinct differences. Unlike the hairballs, each view is different for the two networks.

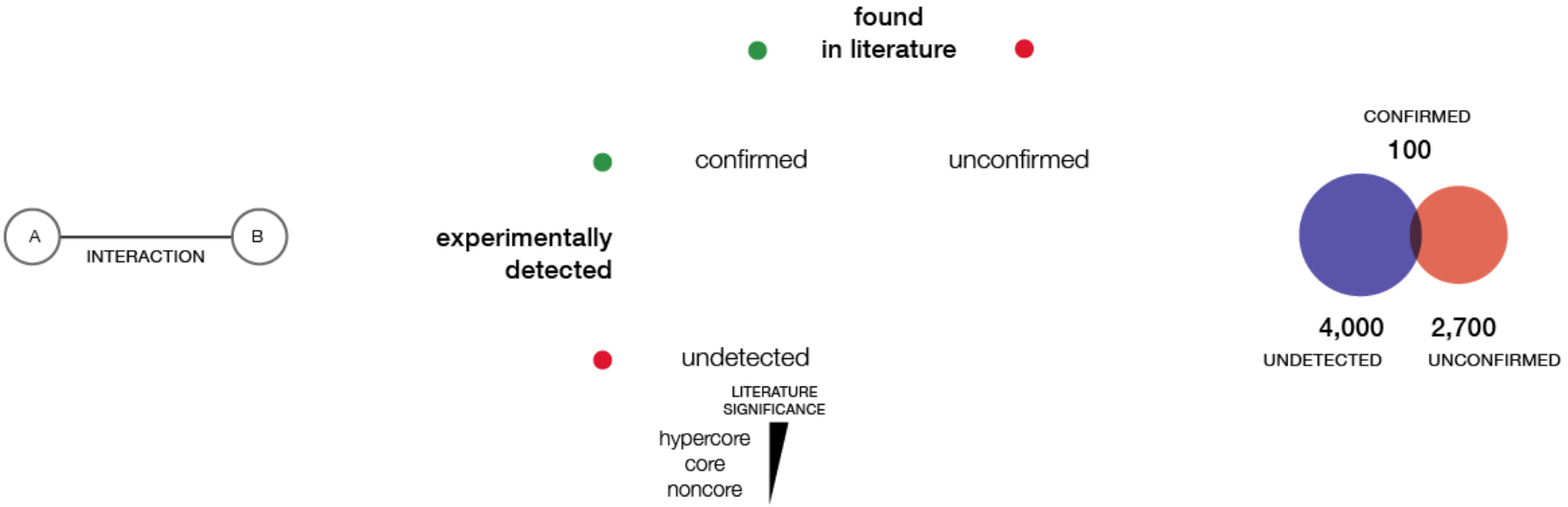


E. coli



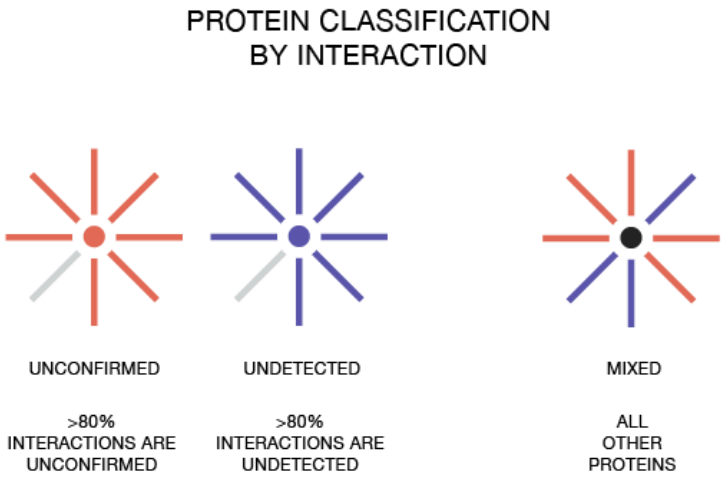
Linux

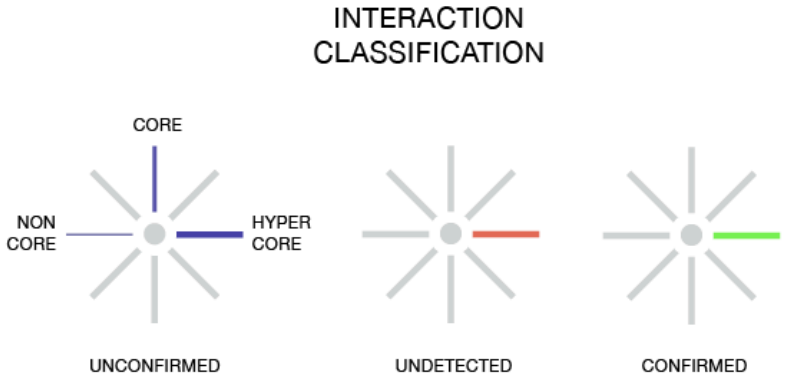
CLASSIFYING PROTEIN INTERACTIONS



Rual et al., *Nature* **437**(7062):1173-8.

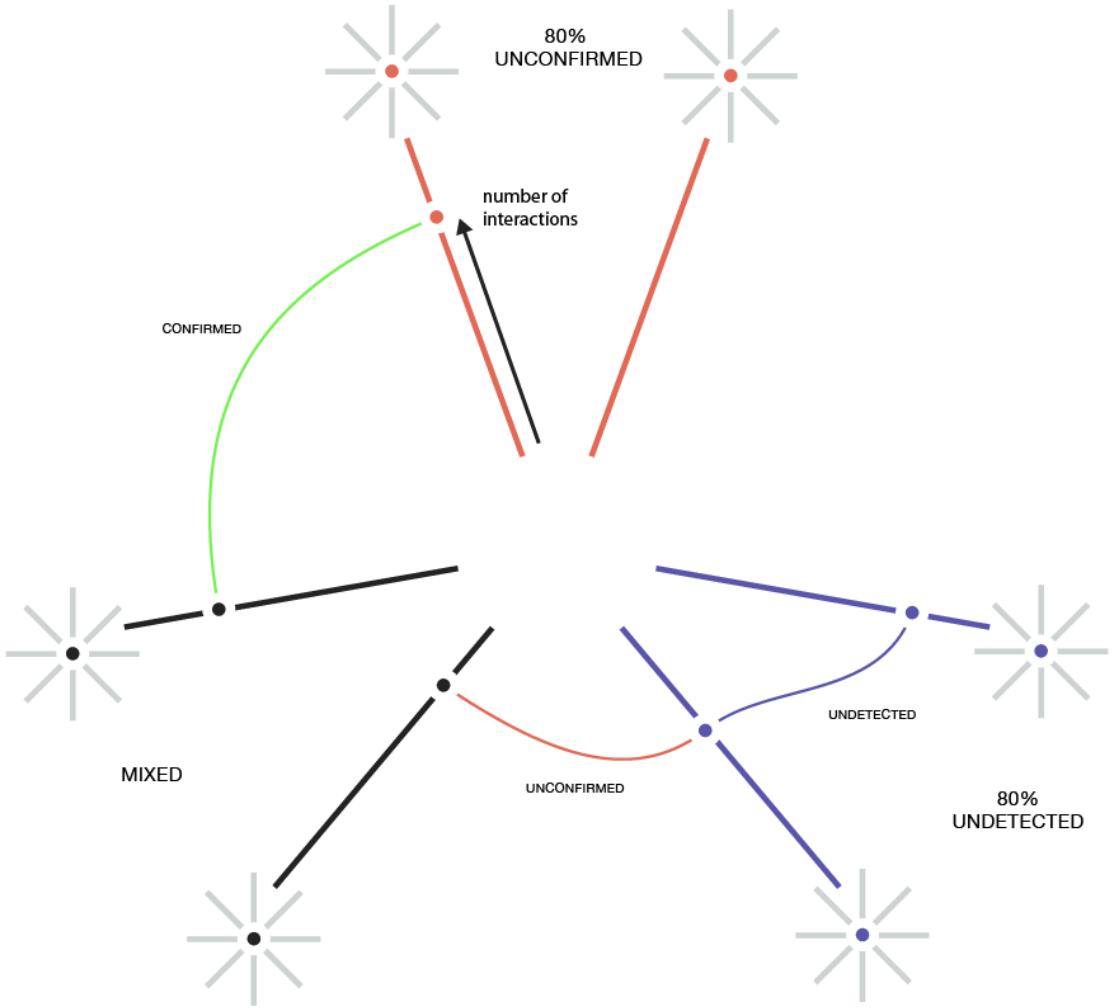
CLASSIFYING PROTEINS AND INTERACTIONS





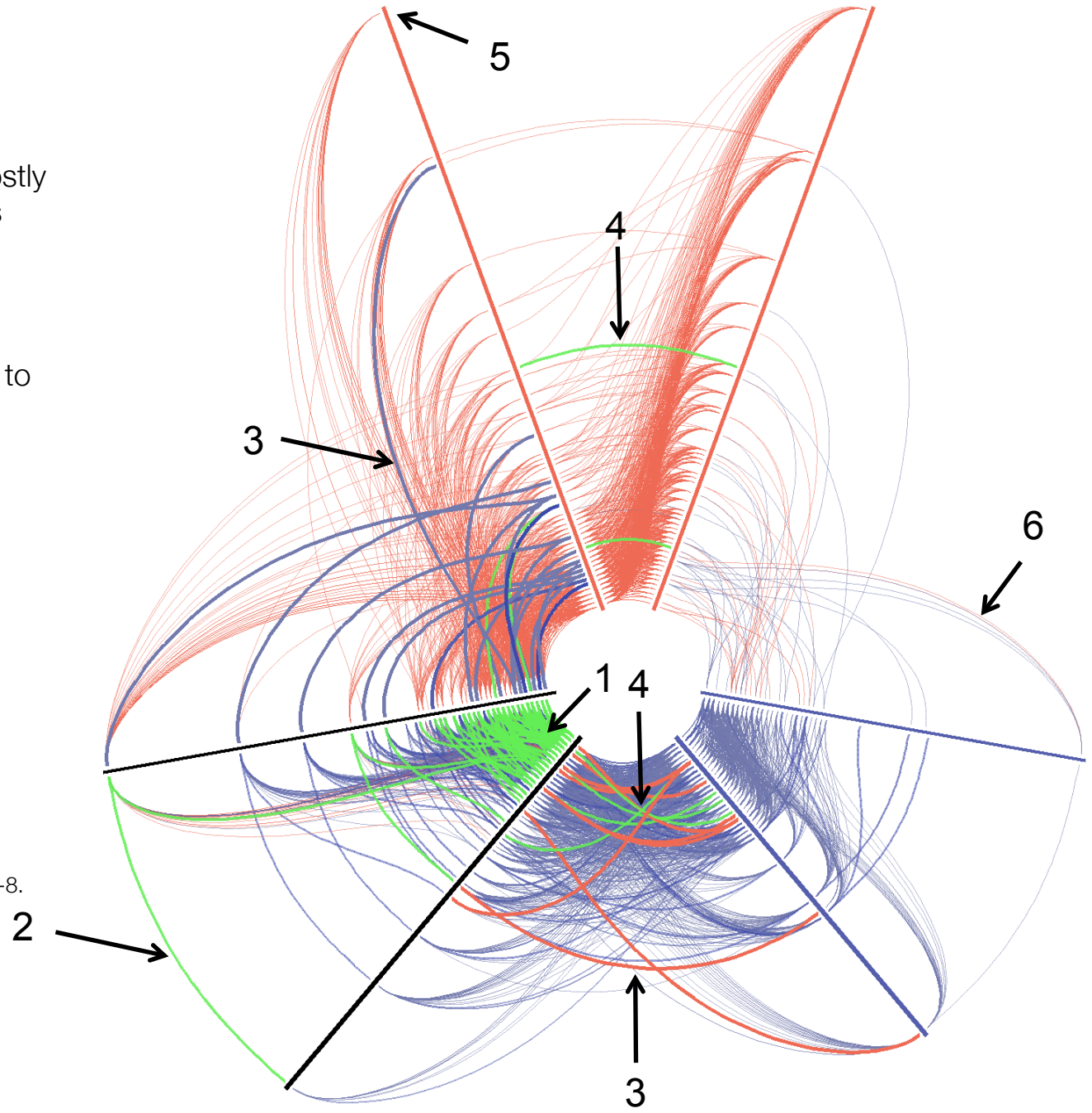
Rual et al., *Nature* **437**(7062):1173-8.

HIVE PLOT SETUP

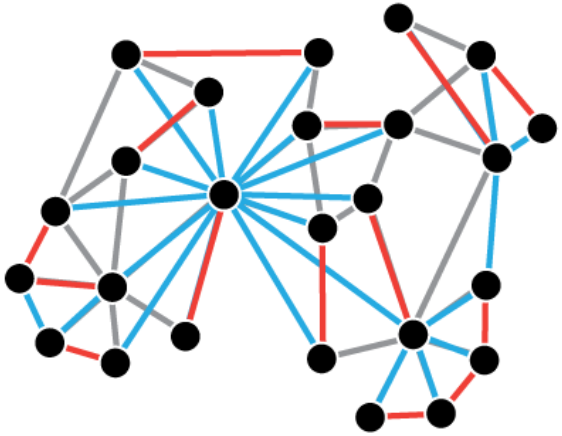


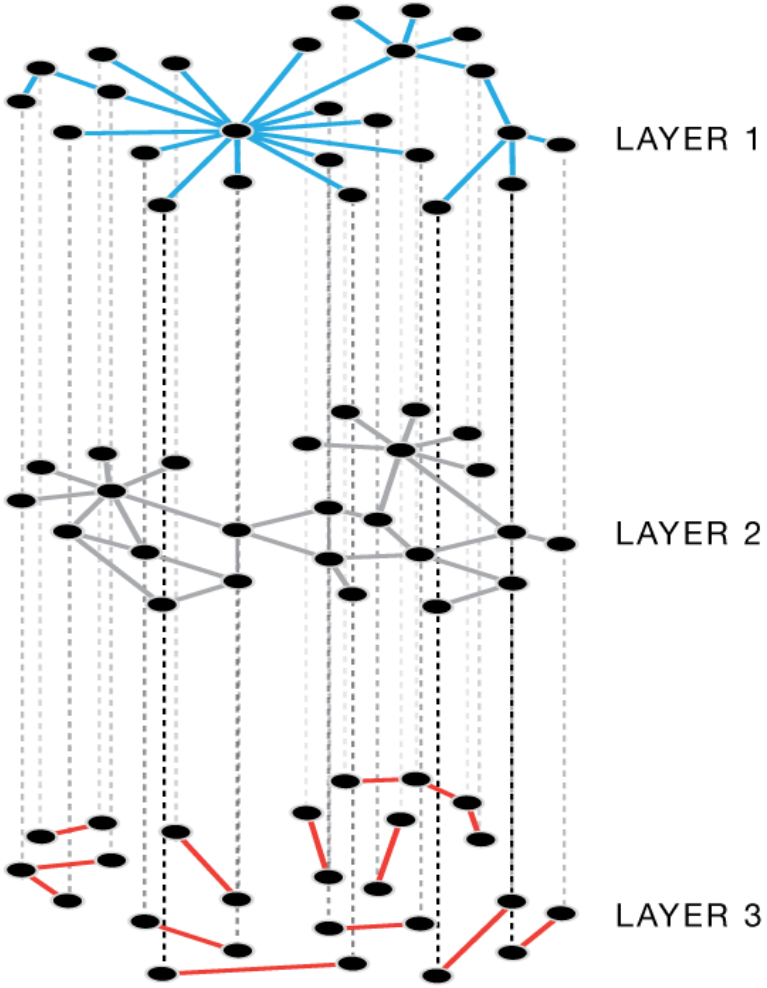
1. confirmed interactions are mostly for proteins with few interactions
2. confirmed self-interaction for most connected mixed protein (TRAF2)
3. more undetected interactions to unconfirmed proteins than unconfirmed interactions to undetected proteins
4. fewer confirmed interactions between unconfirmed than undetected proteins
5. most connected unconfirmed protein has no confirmed or undetected links
6. most connected undetected protein has an unconfirmed link

data from Rual et al., *Nature* **437**(7062):1173-8.



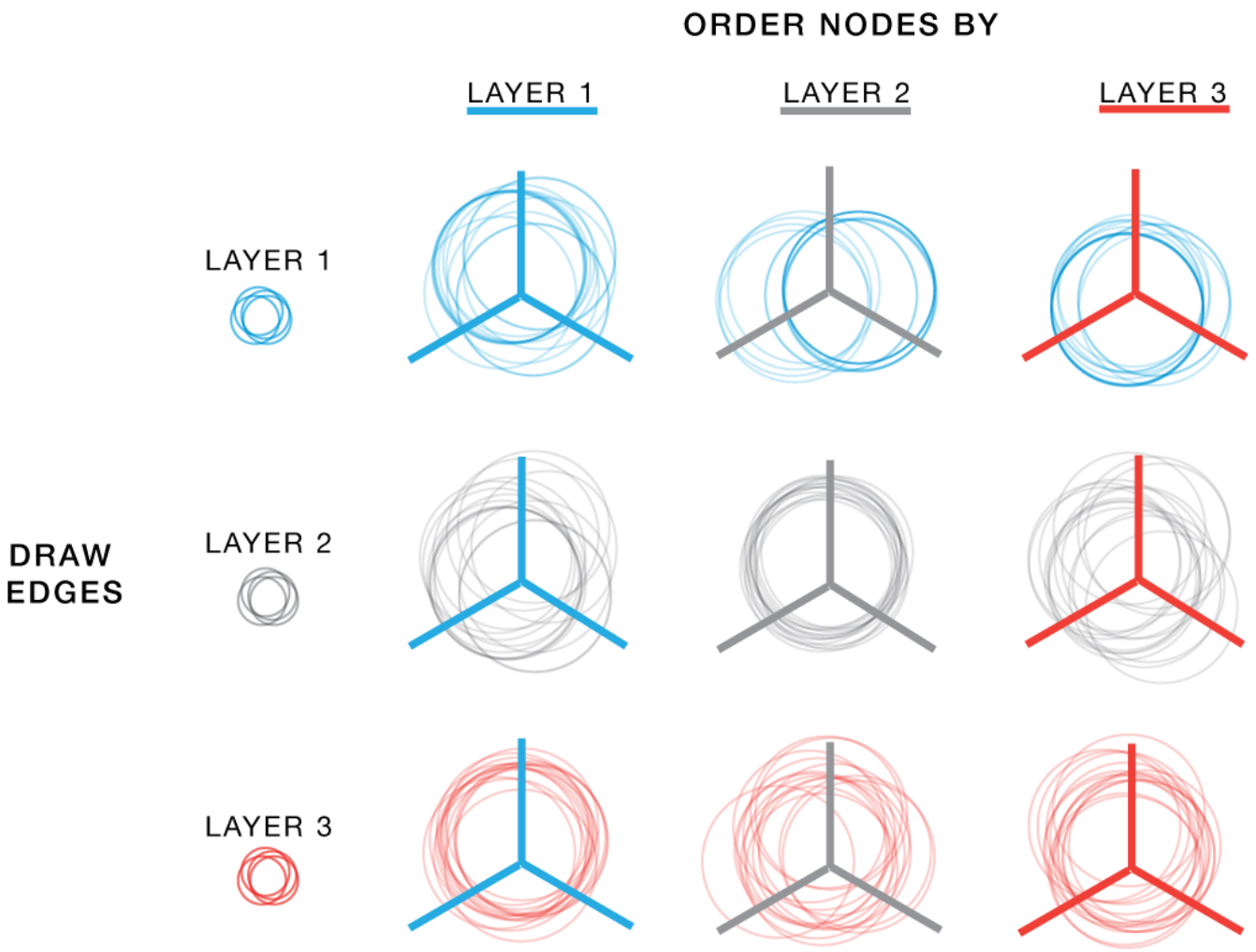
APPLICATION – LAYERED NETWORKS





Suppose you have a network composed of three distinct edge groups. These could be thought of as layers of connectivity, with each layer describing a different type of relationship.

APPLICATION – LAYERED NETWORKS



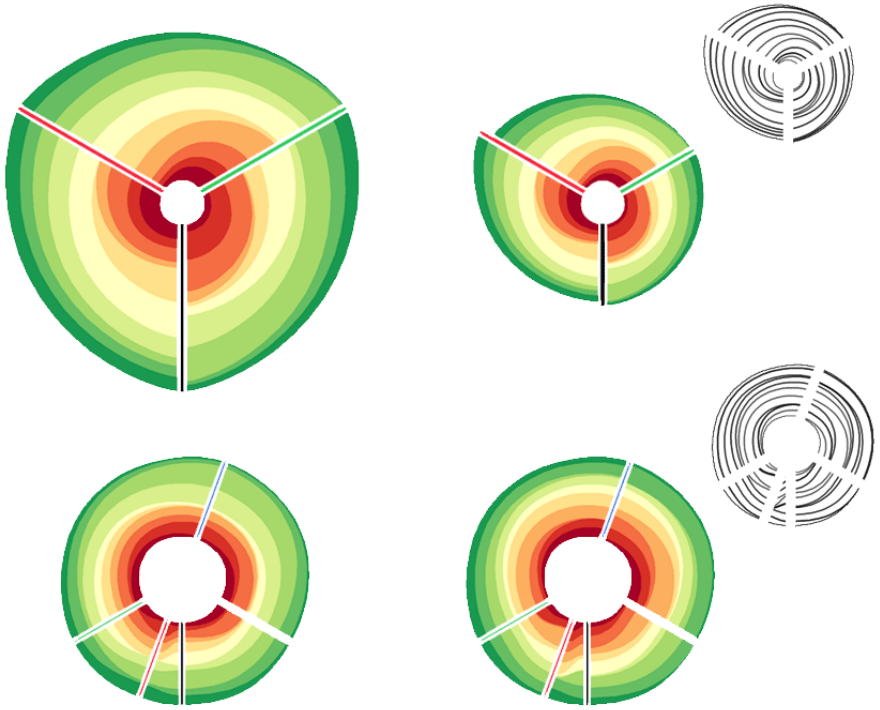
A matrix of linear layouts can reveal how connectivity layers correlate. For each plot the connectivity data that is used to (a) map nodes to axes and determine node position and (b) draw links is *not necessarily the same*.

APPLICATION – STACKED PLOTS

the linear network view can be used to compose stacked bar plots, ideally suitable for comparing multiple ratios

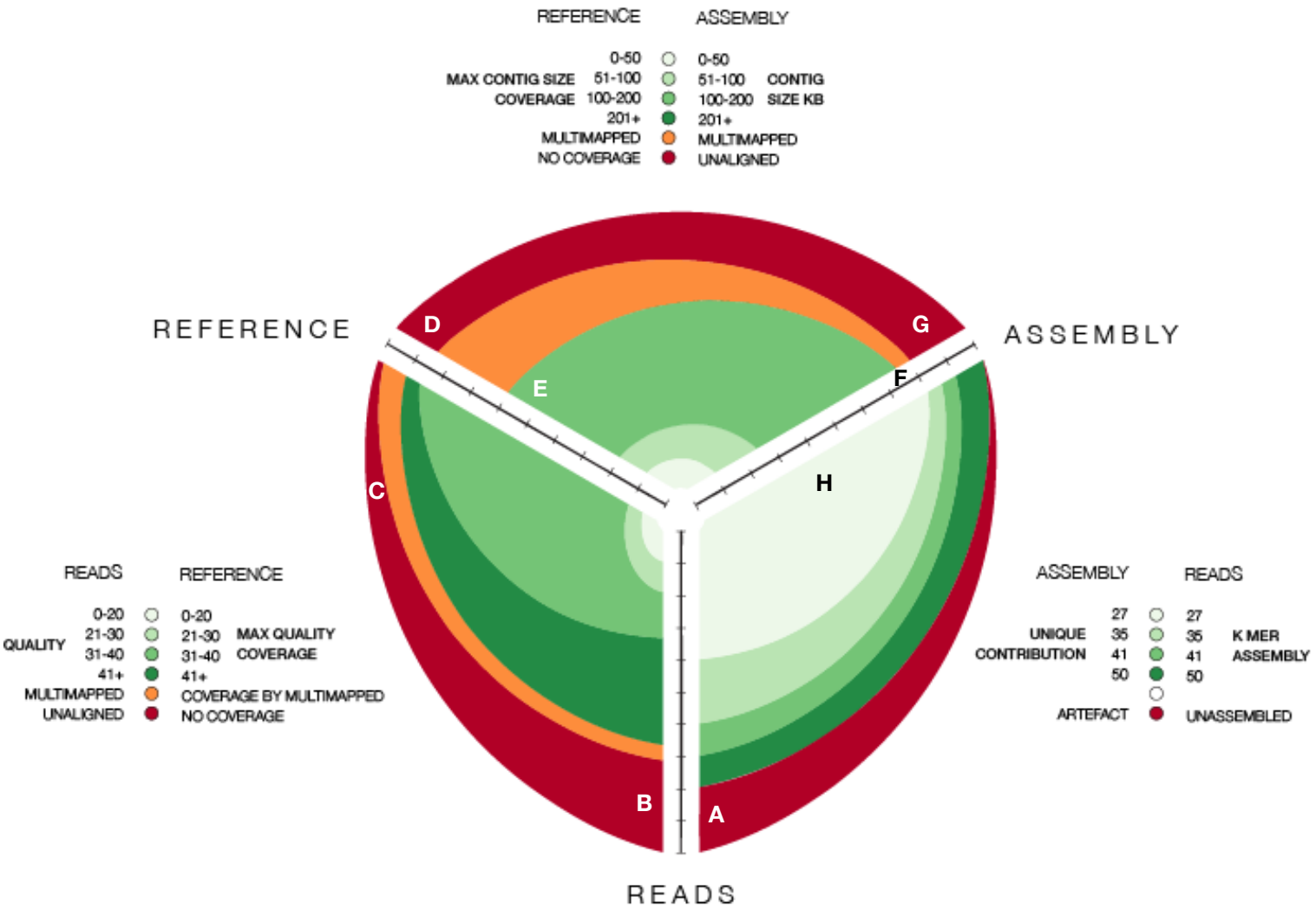
/ edges are drawn as ribbons, with different edge lengths

/ nodes become data values



Application of the network layout to stacked bar plots. The plots are wrapped circularly, creating a comparison loop.

VISUALIZING ASSEMBLY QUALITY



Pairwise comparison between bases in reads, assembly and reference. For example: (A) 20% of reads are unassembled, (B) 30% of reads are unaligned to reference, (C) 2% of reference has no read coverage, (D) 15% of reference has no contig coverage, (E) 60% of reference is constructed by contigs <200kb, (F) there are no contigs >200kb, (G) 20% of contigs are unaligned to reference, (H) 80% of contig bases assembled at k=27.

USING THE SOFTWARE

search GIN for “linnet”

/ use local installation

/ download software from web page

mkweb.bcgsc.ca/linnet

ACKNOWLEDGEMENTS

BC CANCER AGENCY

Cydney Nielsen
Shaun Jackman
Rod Docking
Anthony Fejes
Dan Fornika
Jenny Qian

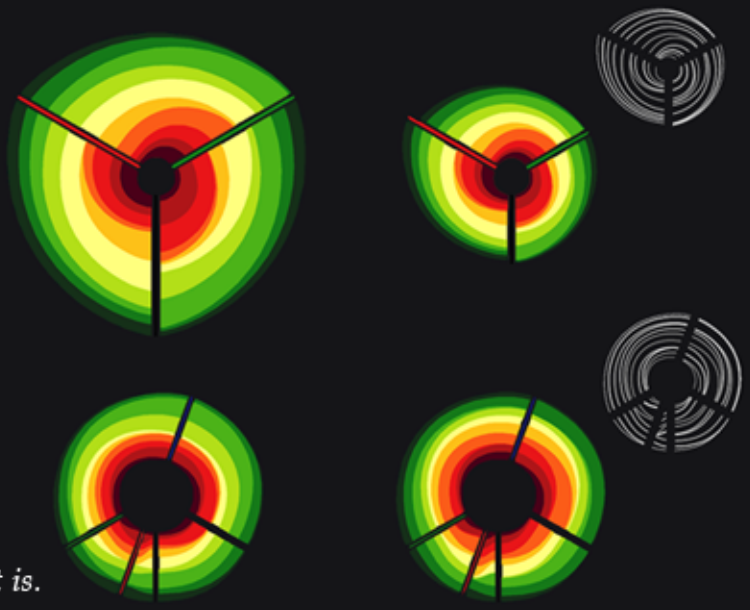
Katayoon Kasaian
Olena Morozova
Inanc Birol
Steven Jones
Marco Marra

MASARYK UNIVERSITY

Martin Lysak

The genius of Gene Rodenberry allowed him to predict a future in which hairballs run amok. In this episode of Star Trek, Trouble with Tribbles, engineer Scott consults with Kirk and Spock about the hairball crisis. Note the tribble in Kirk's cup and those stuck to the walls. It isn't clear how tribbles, which have no legs, can adhere to a vertical surface.

Star Trek Episode 44, 2nd Season, 29 Dec 1967



imagine a world
where something
this pretty
is useful.

we know where it is.