

Linear and Circular Layouts for Network Visualization

Xiao Nan

Central South University

Mar 29, 2012

Outline

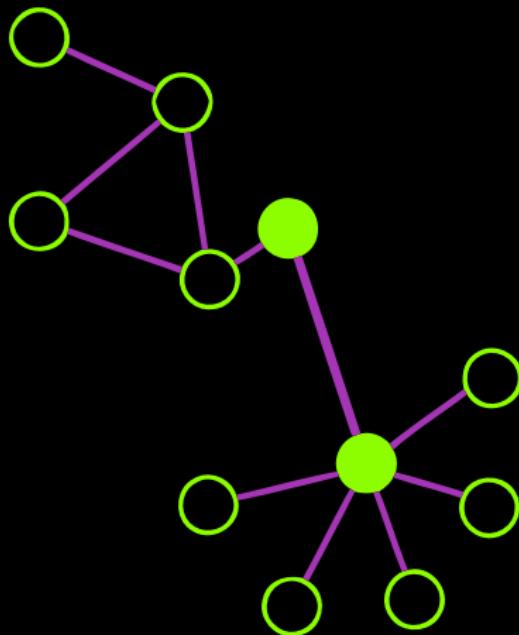
- Traditional Layouts
- Linear Layout
- Circular Layout
- Implementations
- Summary
- References

Totally
instance-based.

Network Data

Online Social Network,
Scientific Collaboration,
Software Dependency,
Genome ...

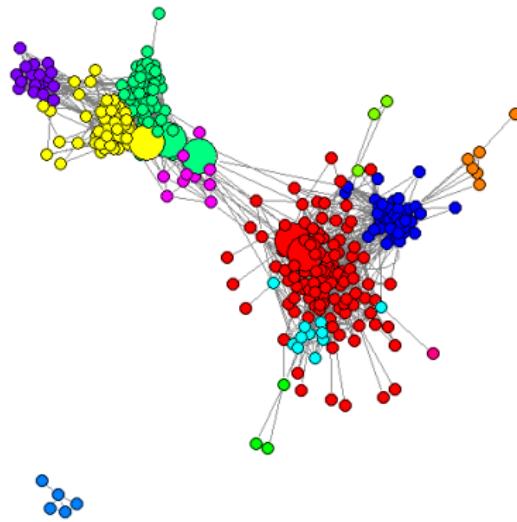
Network data is
Everywhere.



Conventional Network Layout

RAW

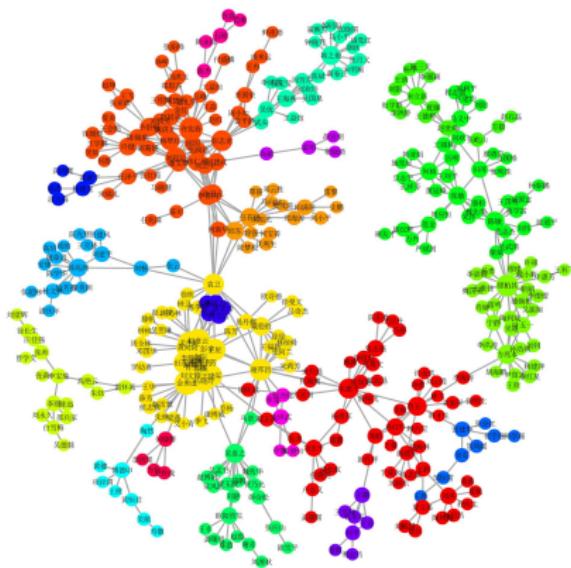
Traditional Layout (RAW)



陈逸波. (2011) 社会网络分析: 探索人人网好友推荐系统.

Property-based

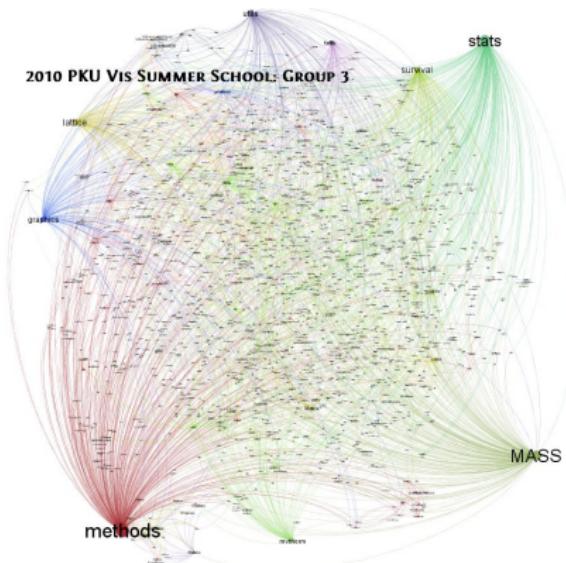
Traditional Layout (Property-based)



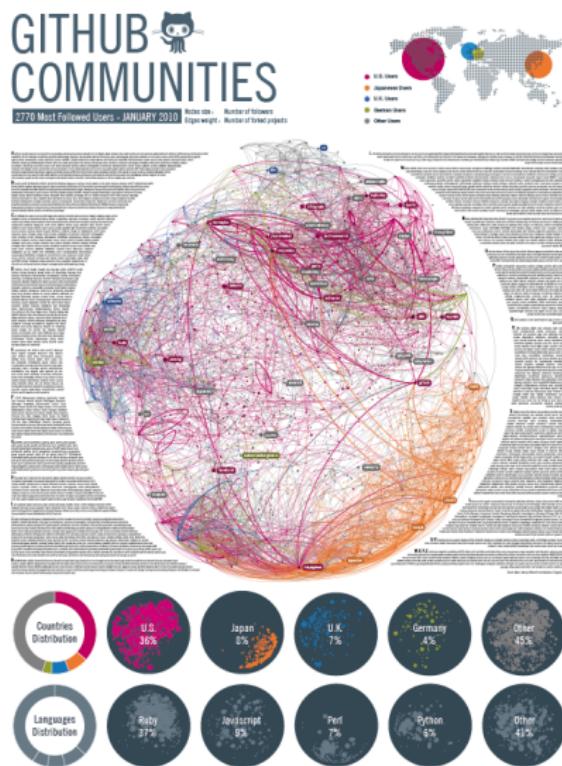
范建宁. (2011) 中国统计之网.

Algorithm-based

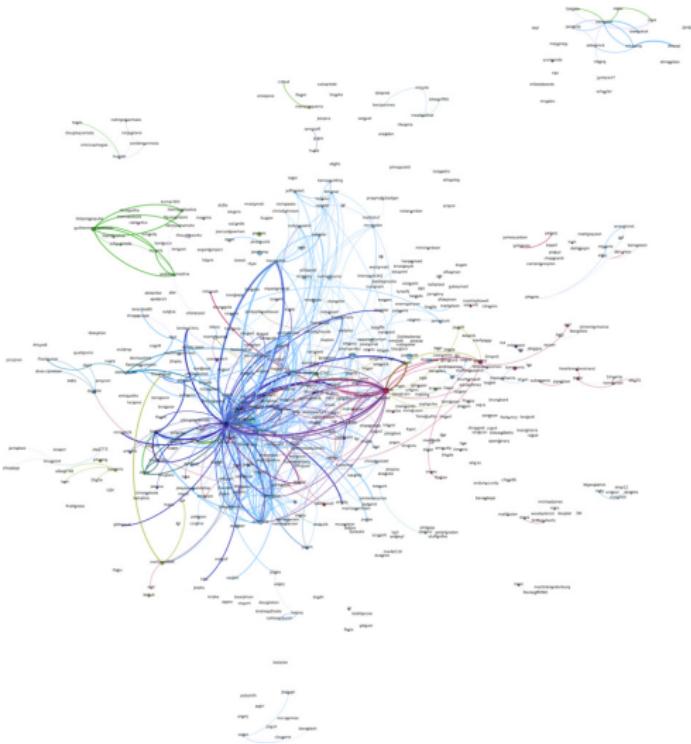
Traditional Layout (Algo-based)



Xiao Nan. (2010) PKU Vis Summer School Final Project.

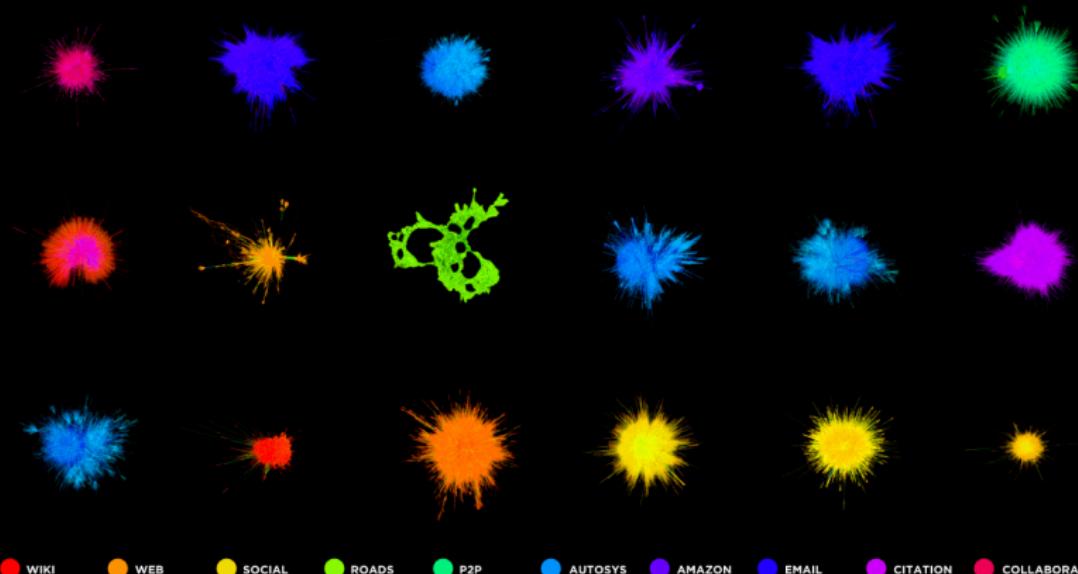


Franck Cuny. (2010) GitHub Explorer.



Franck Cuny. (2010) GitHub Explorer - Python.

Hairball Visualization



● WIKI ● WEB ● SOCIAL ● ROADS ● P2P ● AUTOSYS ● AMAZON ● EMAIL ● CITATION ● COLLABORATION

Hairball Visualization - Beautiful but Useless

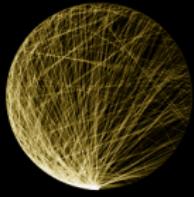
Flaws of Traditional Layout

- Unscalable
- Uninterpretable
- Unreproducible
- Unbeautiful

Unscalable
Leads to overplotting

Uninterpretable
Analysis restricted

Unreproducible
That's fatal



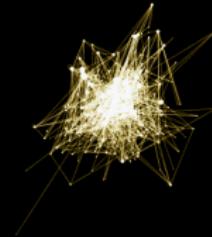
CIRCULAR BY DEGREE



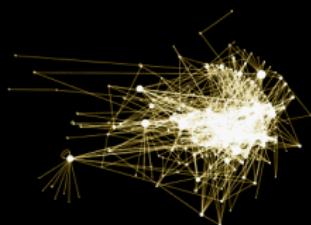
EDGE-WEIGHTED
SPRING EMBEDDED



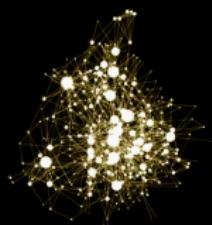
SPRING EMBEDDED



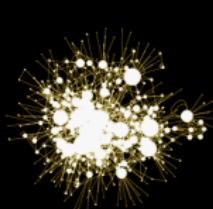
INVERTED SELF
ORGANIZING MAP



RADIAL TREE



ORTHOGONAL



FORCE DIRECTED



FORCE DIRECTED

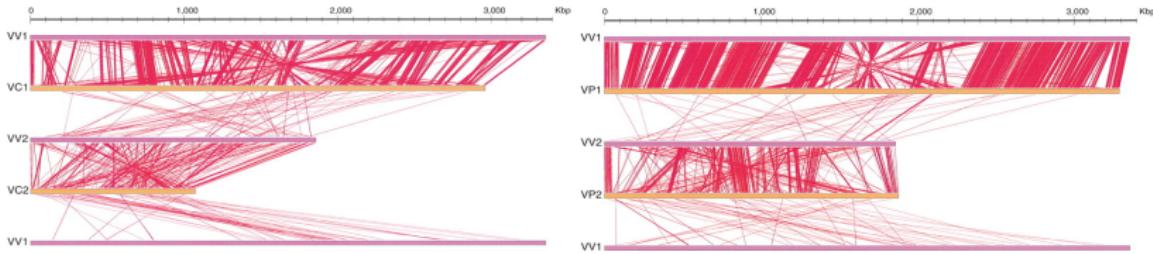
Unreproducible Visualization

Unbeautiful
From aesthetics, to
ugliness

Dumping that layout is
necessary.

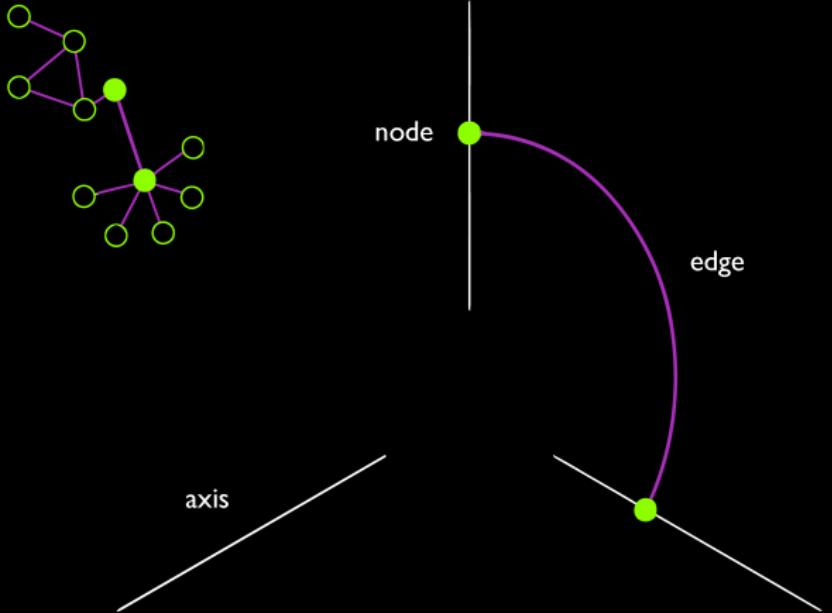
Linear Layout

Linear Layout - Inspiration

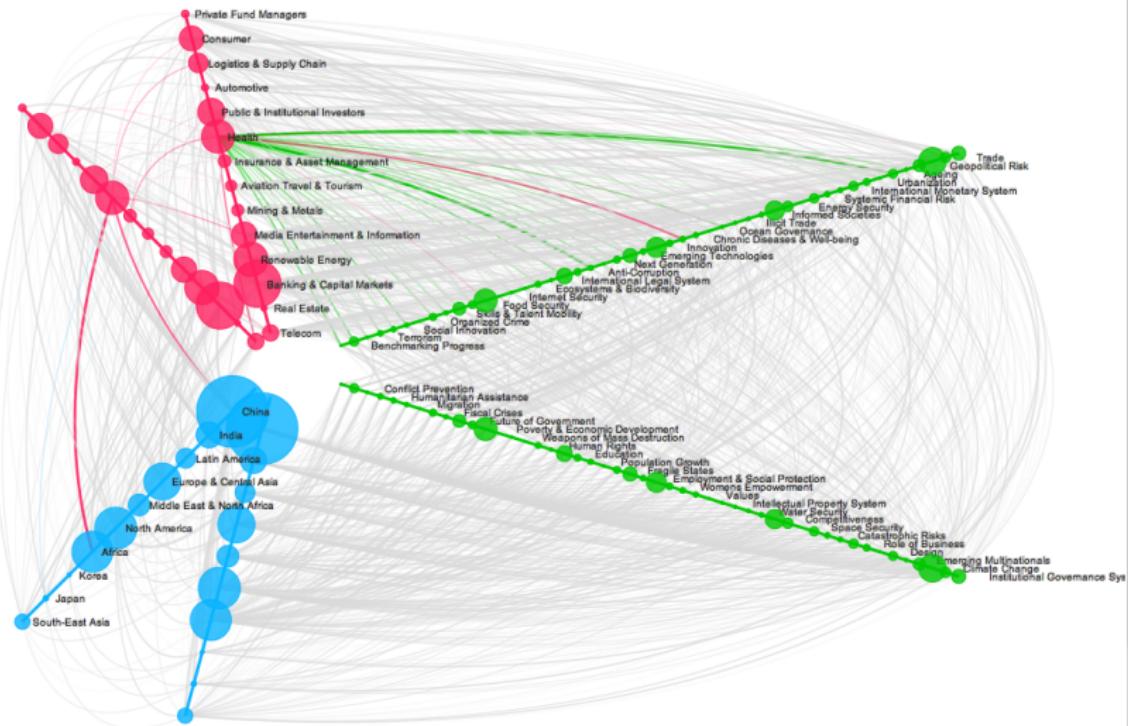


Chen, C.Y., et al., Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome Res.* 2003. 13(12): p. 2577-87.

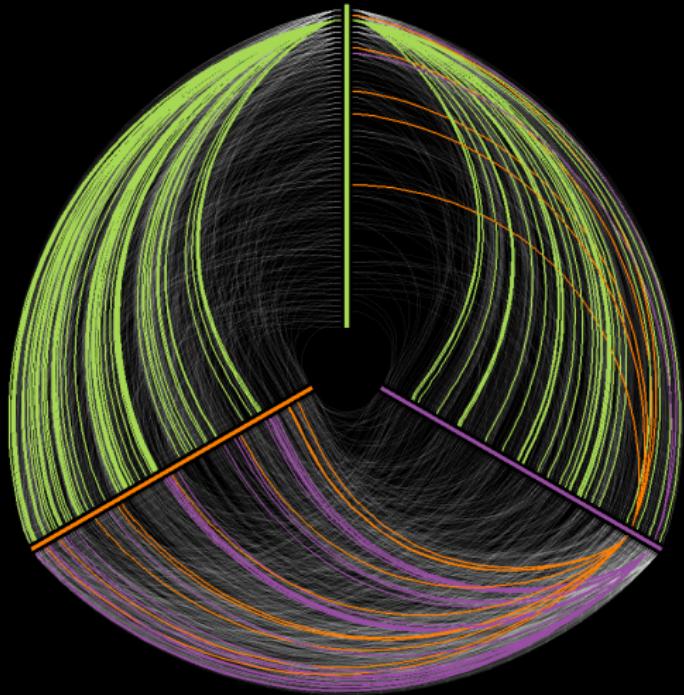
Chen, C.Y., et al. (2003), Comparative genome analysis of
Vibrio vulnificus, a marine pathogen.



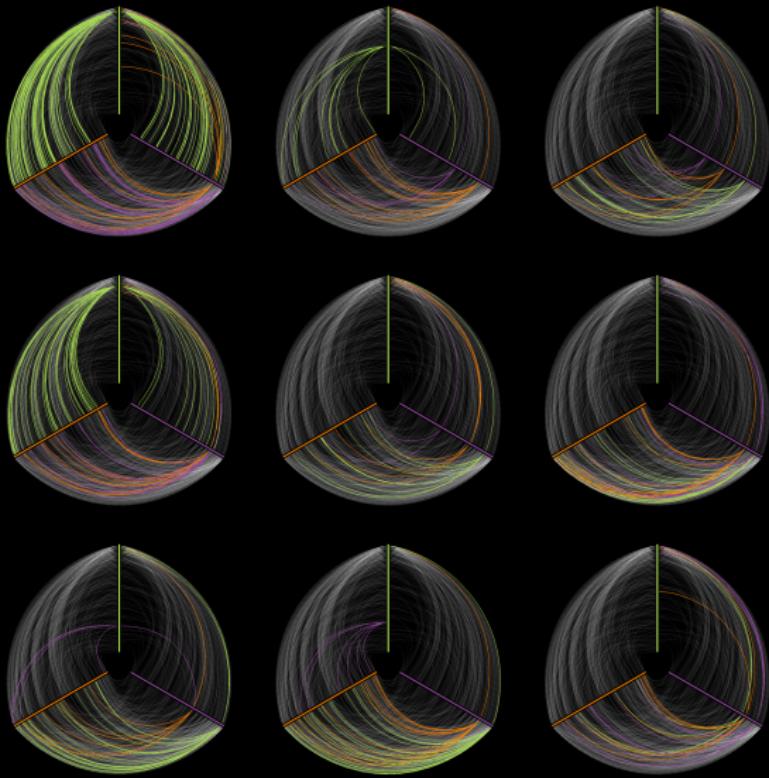
Linear Layout (hiveplot)



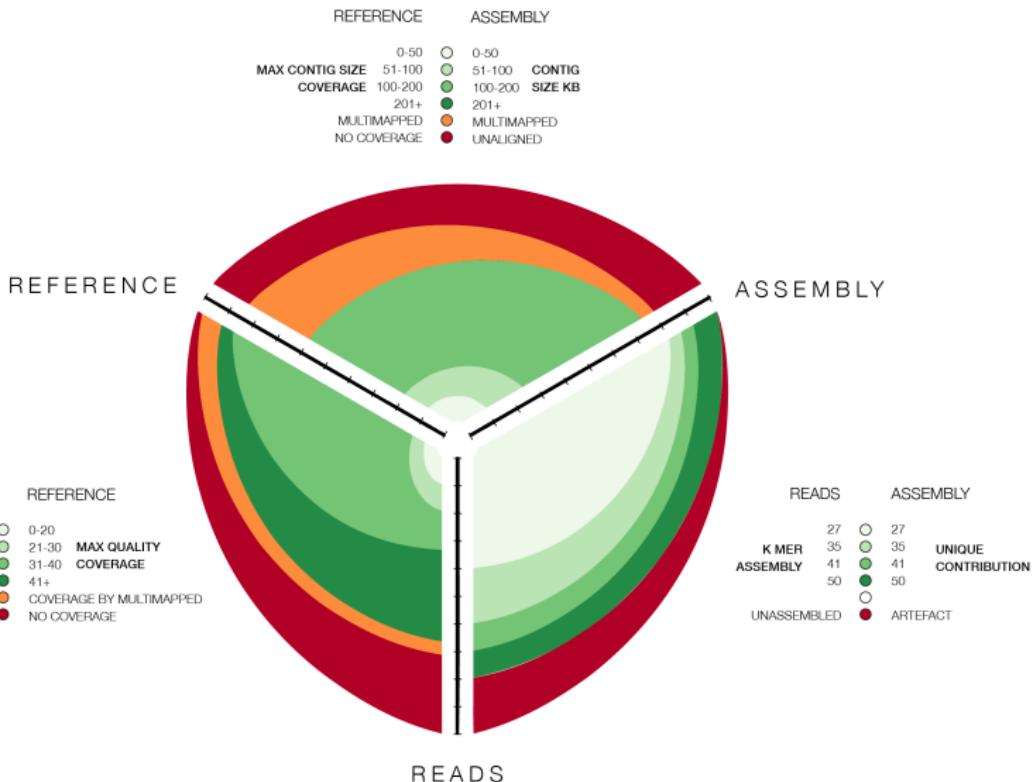
J.R. Heard. (2010) World Economic Forum - Hive Plot



Xiao Nan. (2011) CRAN Package Dependency Network - Hive Plot.



Xiao Nan. (2011) CRAN Package Dependency Network - Hive Panel.



Edge Bundle - Visualizing Ratios

Advantages of Linear Layout

Advantages of Linear Layout

- Scalable

Advantages of Linear Layout

- Scalable
- Interpretable

Advantages of Linear Layout

- Scalable
- Interpretable
- Reproducible

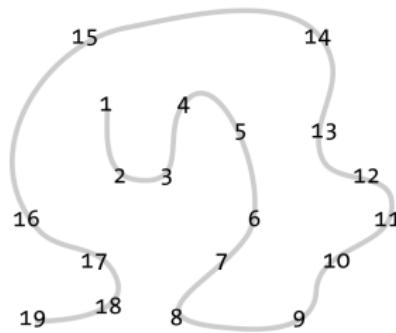
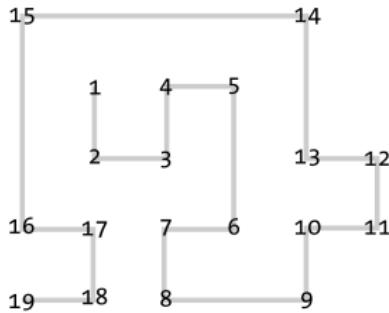
Advantages of Linear Layout

- Scalable
- Interpretable
- Reproducible
- Beautiful

Circular Layout

Graphical Highlights

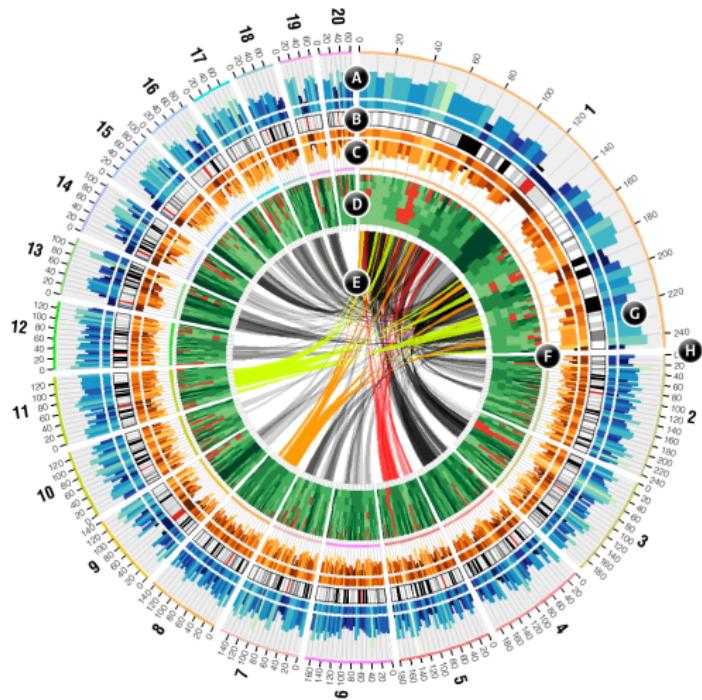
Graphical Highlights



Curved objects are easier to visually follow. Time yourself to see how long it takes you to scan through the numbers in the two shapes. You will find that effort in interpreting the left shape is higher than the right shape.

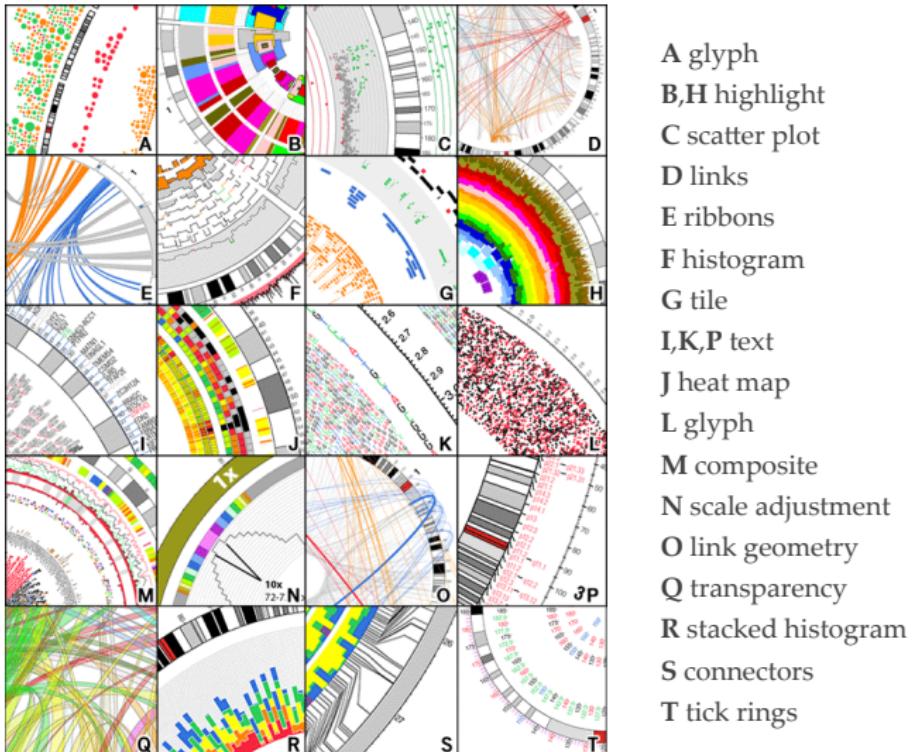
Right angles in the top shape require more energy to traverse – you may find that switching eye movement from vertical immediately to horizontal is uncomfortable.

Curve is Better



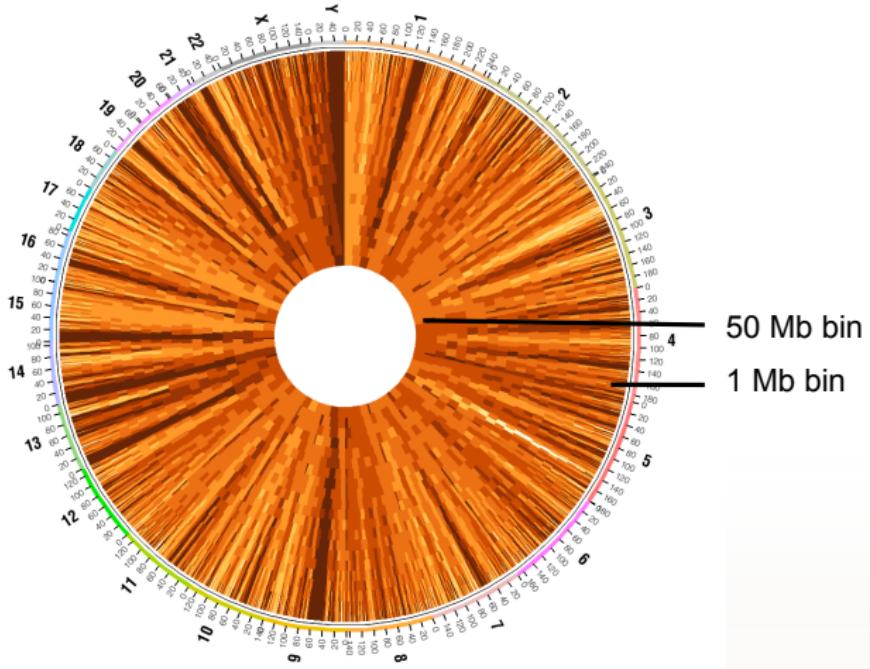
A histogram, B ideograms, C histogram, D heatmap, E links, F highlights, G grid, H ticks. Format of data in tracks A, C, D, E is adjusted by rules based on data values.

Typical Circular Layout (From circos)



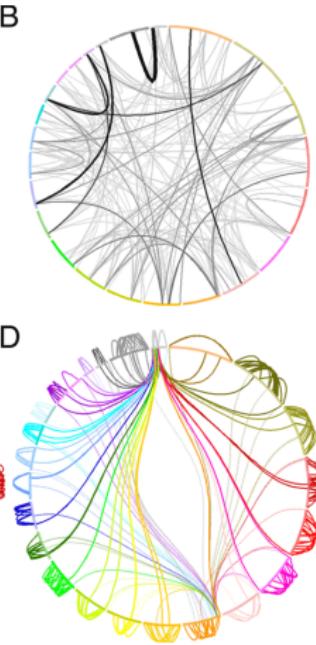
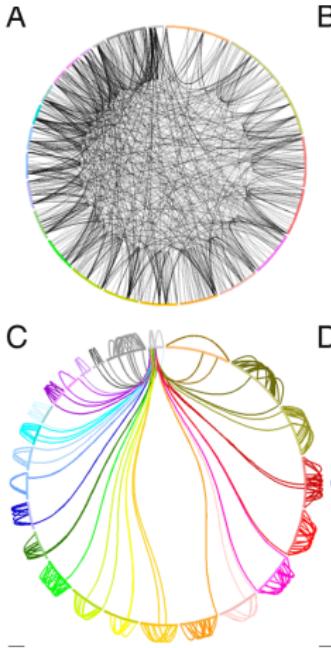
- A** glyph
- B,H** highlight
- C** scatter plot
- D** links
- E** ribbons
- F** histogram
- G** tile
- I,K,P** text
- J** heat map
- L** glyph
- M** composite
- N** scale adjustment
- O** link geometry
- Q** transparency
- R** stacked histogram
- S** connectors
- T** tick rings

Stacking Tracks - Suitable for Publication



Human gene density at resolutions from 50Mb (inner track) to 1Mb (outer track).
The circular form naturally supports a range of resolutions.

Circular Fits More Resolutions



Links and Bundles

The same data set is shown in all panels.

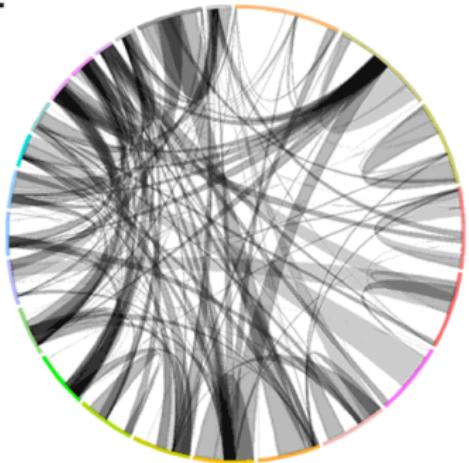
A each link represents one of a subset of 2,500 segmental duplications within the human genome

B rules are used to change link color and thickness

C rules are used to show only links to chrY

D in addition to rules in (C), other rules add a second layer of links from chr8.

E



F



E,F adjacent links are grouped into thicker links (*bundles*) to reduce the complexity of the figure.

Links and Bundles

Examples

The CANCER GENOME challenge

Data bases could soon be flooded with genome sequences from 25,000 tumours. Heidi Ledford looks at the obstacles researchers face as they search for meaning in the data.

When it was first discovered, in 2006, as a study of colon cancer genomes, the method had little consequence. It appeared as only one of the first few papers to come out of the Human Genome Project. The method, called genome-wide association studies (GWAS), had been developed by researchers at the Wellcome Trust Sanger Institute in Cambridge, UK, and the Broad Institute in Cambridge, Massachusetts, and had contributed to the study.

But as the cost of sequencing DNA dropped, so did the cost of analysing the data. In 2008, when the first 100 cancer genomes were sequenced, there were 100 GWAS studies. Now, there are 1,000. Standard studies showed that the mutation rate changed activity of certain genes, such as those involved in repairing mutations or in accelerating cell growth. And then came the copy-number analysis, developed at the Wellcome Trust Sanger Institute in Cambridge, Massachusetts — it didn't stop there.

Four years after the initial discovery, all these analyses have been put together, bringing up the GWAS resolution, the incipience

GENOMES AT A GLANCE

Genome-wide association studies (GWAS) can give a snapshot of the mutations within a population that are associated with a particular disease or other trait type of interest.

Interchromosomal interactions

Point mutations

Deletions

Insertions

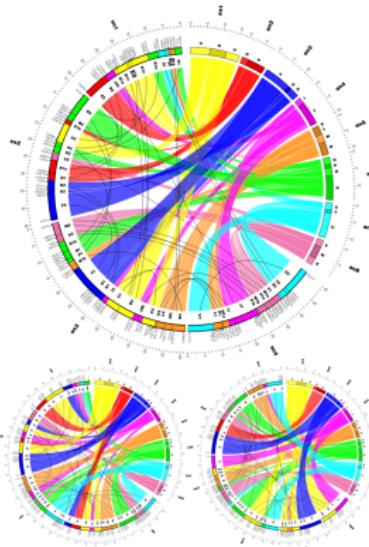
Copy-number change

newly pulled from a sensible bootstrap of cancer-associated mutations thanks to high-throughput genome sequencing. By the time they had finished, the researchers had managed to sequence the DNA from thousands of individuals, mostly cancer patients, from around the world. Roughly 75 cancer genomes have been sequenced so far, and many more which researchers expect to have solved the broadest genetic puzzle.

Therefore, increasingly exciting big questions. Comparing the gene sequences of any two people will reveal millions of differences, of single-letter changes, or point mutations, and millions of additions or deletions, or inverted sequences (see 'Comparing at a genomic level' on page 972). Last year, Vogelstein, a cancer researcher at the Johns Hopkins University School of Medicine and Therapeutics at Johns Hopkins, US, proposed to analyse the mutations in cancer patients rather than in their catalogue lots and lots of healthy people, to see if they might look closely, and perhaps even more closely, at each other. This system offers to disentangle the mutations that cause a cancer from those that are merely associated with it, and those that produce of cancer's growth and threatened its survival, and those that are just along for the ride.

Research can look for mutations that are unique to a specific cancer, or for mutations that are shared at different points. But the real challenge is to find the right answers. Once you take the obvious mutations at the top of the list, how do you make

ANDREW HETHERINGTON/SCIENCE PHOTO LIBRARY



Evolutionary relationship between ancestral and modern crucifer genomes.

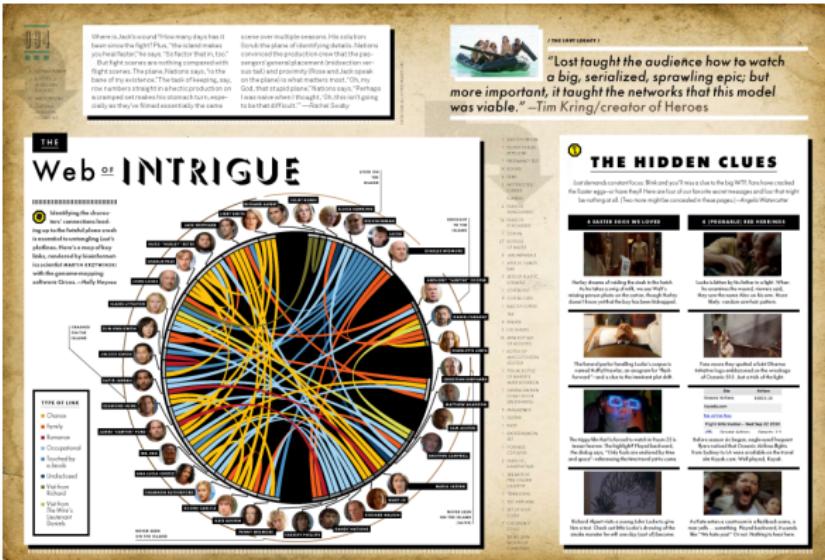
Genomic rearrangements from COSMIC (catalogue of somatic mutations in cancer).

Bedford H 2010 Big science: The cancer genome challenge *Nature* 464 (7291) 972-974.

Lysak M et al 2010 Diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell* (in press)

100+ High Quality Citations

Born for Genome Visualization

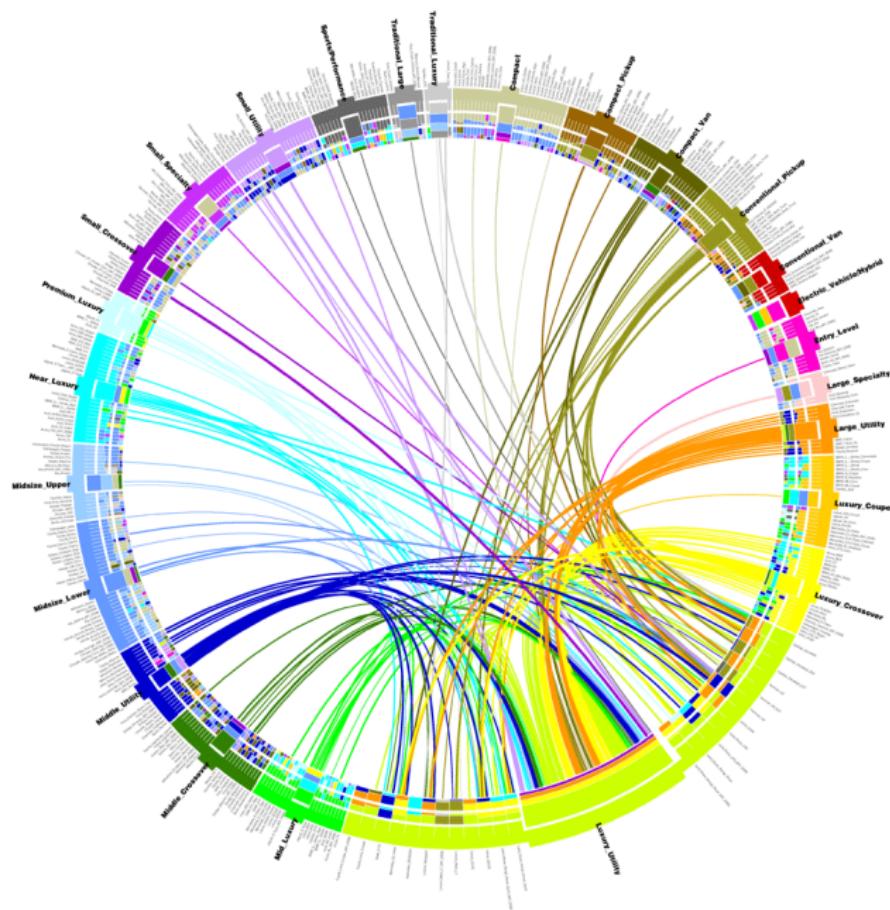


The relationship of characters on *Lost*. Each character is assigned a segment whose size is proportional to the number of relationships. Characters are categorized in five groups. There are 8 types of relationships.

Lost Characters' Relationship



DHL's Commercial



Car Purchase Paradigm Shift

Implementation

Implementation

- Traditional

- igraph (python, R)
- Gephi (Java)
- Cytoscape (Java)
- d3.js (Javascript)

Implementation

- Traditional

- igraph (python, R)
- Gephi (Java)
- Cytoscape (Java)
- d3.js (Javascript)

- Linear

- HiveR (R)
- d3.js (Javascript, experimental)
- linnet (perl, outdated)

Implementation

- Traditional

- igraph (python, R)
- Gephi (Java)
- Cytoscape (Java)
- d3.js (Javascript)

- Linear

- HiveR (R)
- d3.js (Javascript, experimental)
- linnet (perl, outdated)

- Circular

- circos (perl)

Summary

Summary

- 怎么做对是科学, 怎么做好则是艺术.

Summary

- 怎么做对是科学, 怎么做好则是艺术.
- 可视化不能代替模型, 可视化不能代替计算,
可视化不能代替分析. 但我们总可以试图改善图形, 让它更好地辅助建立和解读模型, 指明优化方向.

Summary

- 怎么做对是科学, 怎么做好则是艺术.
- 可视化不能代替模型, 可视化不能代替计算, 可视化不能代替分析. 但我们总可以试图改善图形, 让它更好地辅助建立和解读模型, 指明优化方向.
- 科学有对错, 但艺术没有对错. 对于网络数据, 不妨多尝试一些布局, 选取效果最好的一种. 形式虽然不重要, 但我们总可以把形式做得更好.

References

- Martin Krzywinski. Hiveplots.
<http://www.hiveplots.com>
- HiveR. <http://cran.r-project.org/web/packages/HiveR/index.html>
- d3.js - Hive Plot. <http://bl.ocks.org/2035137>
- Martin Krzywinski. Circos. <http://circos.ca>
- d3.js - Chord Diagram.
<http://mbostock.github.com/d3/ex/chord.html>

References

- Gephi. <http://gephi.org/>
- 陈逸波. (2011) 社会网络分析: 探索人人网好友推荐系统. <http://cos.name/2011/04/exploring-renren-social-network/>
- 范建宁. (2011) 中国统计之网. <http://cos.name/2011/09/coauthorship-network-of-china-stat/>
- Xiao Nan. (2010) PKU Vis Summer School Final Project.
http://vis.pku.edu.cn/wiki/doku.php?id=public_course:summerschool_s10:group:group3
- Xiao Nan. (2011) Visualizing CRAN Package Dependency Network:
Reveal Hidden Patterns with Martin Krzywinski's Hive Panel.
<http://www.road2stat.com/cn/statistics/hivepanel.html>

References

- GitHub Explorer. <http://lumberjaph.net/graph/2010/03/25/github-explorer.html>
- J.R. Heard. (2010) World Economic Forum Hive Plot. <http://www.visualizing.org/visualizations/world-economic-forum-hive-plot>

Q & A

Xiao Nan

road2stat@gmail.com

road2stat.com