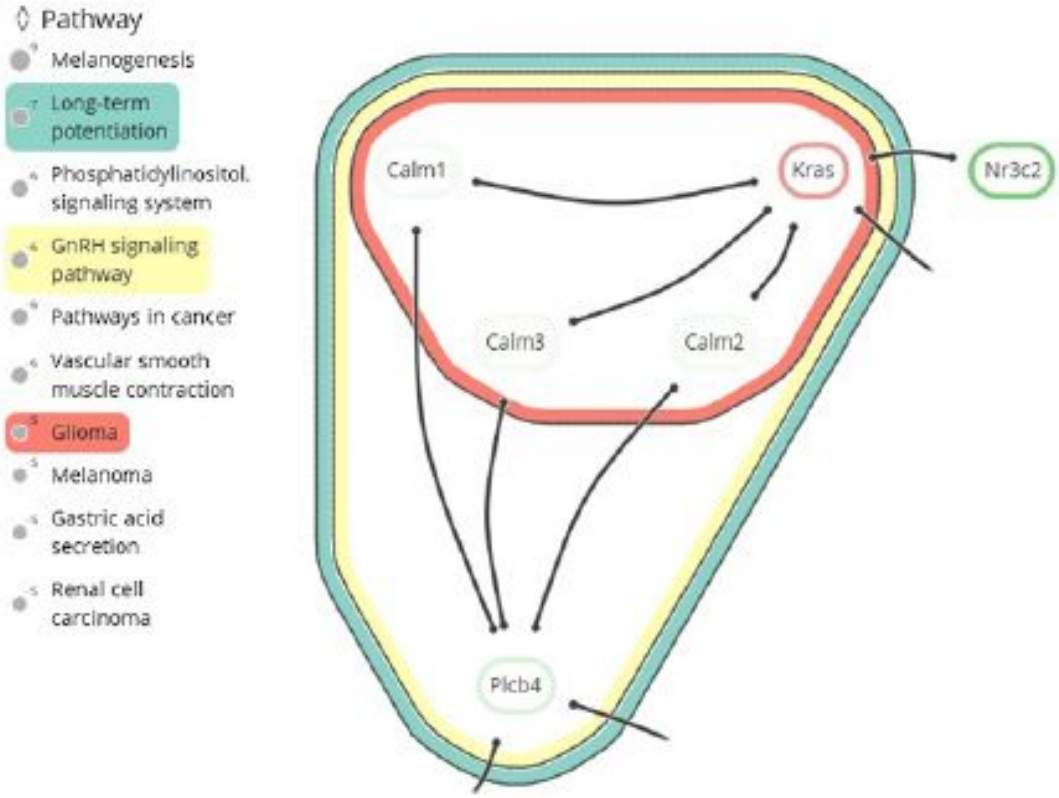
eXamine: Exploring annotated modules in networks

Tim Vigers

Background

- Pathway-based approaches examine overrepresented pathways from databases (e.g. KEGG) rather than focusing on specific genes.
- Network-based approaches produce small subnetwork modules of interest, which may cover several pathways.
- Nodes in these subnetworks are annotated to place the module in a higher-level context (e.g. the node's role in cellular function).
- Data consist of a relatively small network with many annotations. Annotation sets vary in terms of how many nodes each one covers, and often overlap with one another.


Pathway
Melanogenesis
Long-term potentiation
Phosphatidylinositol signaling system
GnRH signaling pathway
Pathways in cancer
Vascular smooth muscle contraction
Glioma
Melanoma
Gastric acid secretion
Renal cell carcinoma
Calm1
Kras
Nr3c2
Calm3
Calm2
Plcb4

Tasks

Gene-based tasks:

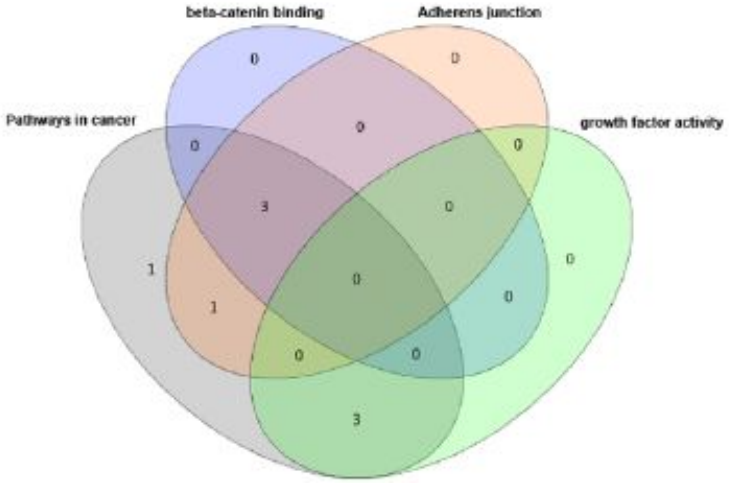
1. Which genes are over or underexpressed? Which genes are not significant?
2. Which genes interact with one another?
3. What annotation set are the genes of interest part of?
4. Which genes share annotations?

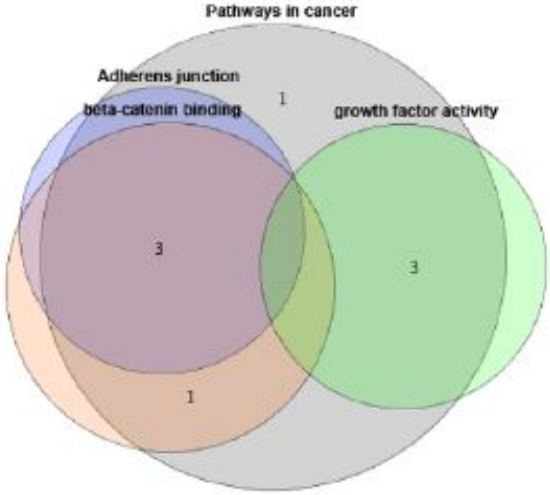
Annotation-based tasks:

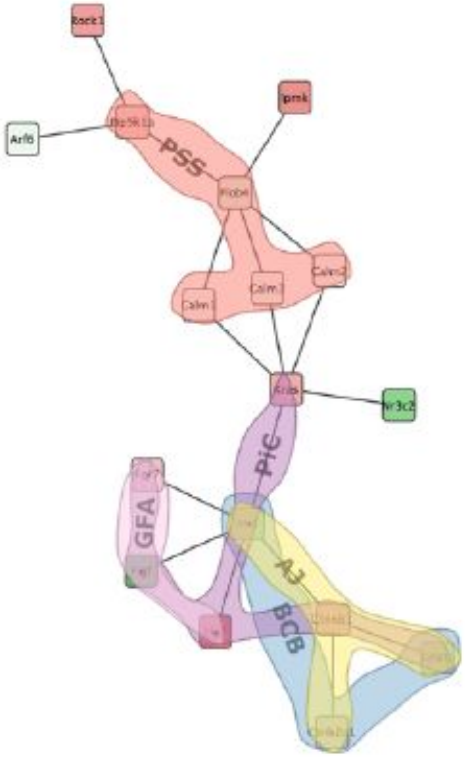
1. How important or relevant is an annotation set to the analysis question (e.g. “pathways in cancer” in a glioma study)?
2. Which genes are part of an annotation set of interest?

Related Tools

- Previous work tended to focus on network visualization with limited support for annotations (usually encoded by node color).
- Set visualization in Cytoscape is limited and separate from network visualization modules.
- Essentially, existing tools focus too much on either network topology or set visualization, at the expense of the other.

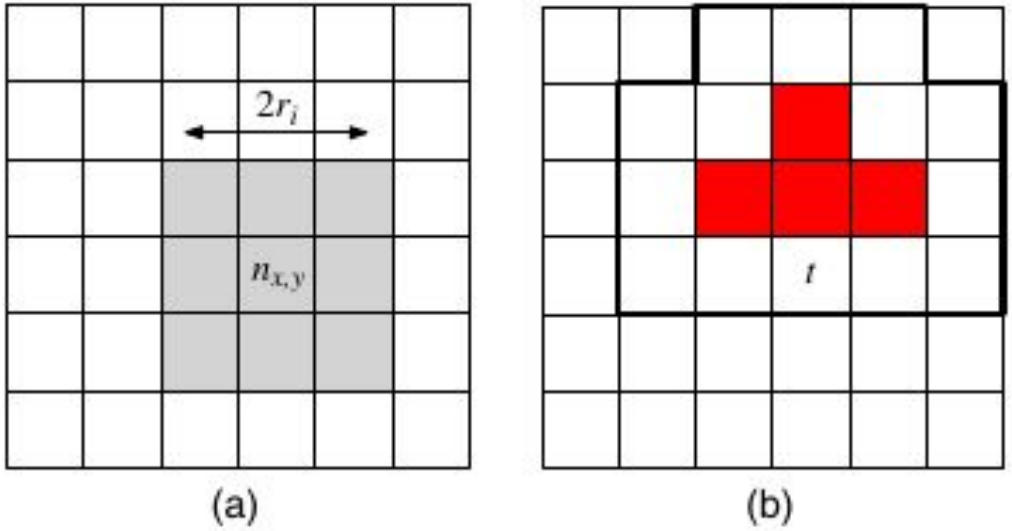

beta-catenin binding
Adherens junction
Pathways in cancer
growth factor activity
0
0
0
0
3
0
1
0
0
1
0
0
0
3


Pathways in cancer
Adherens junction
beta-catenin binding
growth factor activity
1
3
3
1


PSS
PiC
GFA
AJ
BCB

eXamine Methods

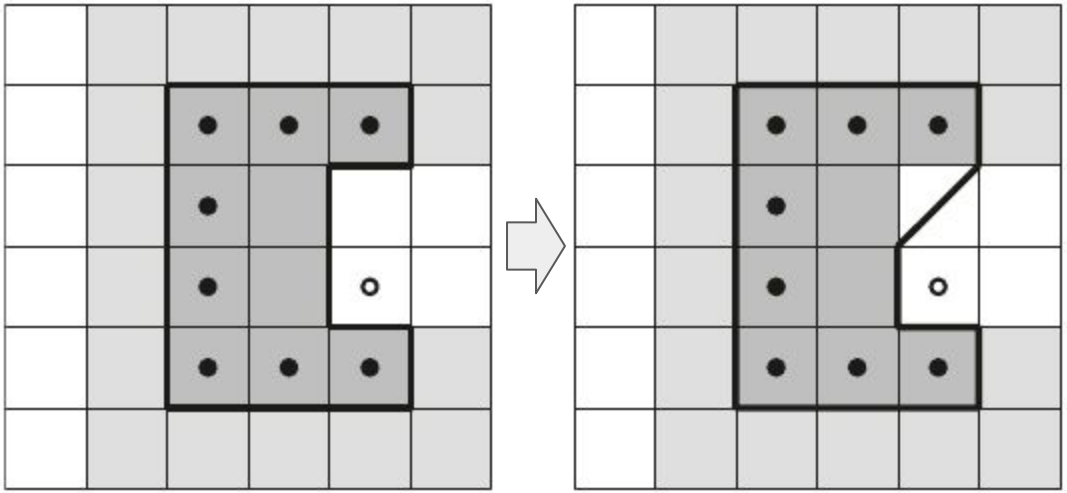
- Annotated modules are treated as a hyper graph with nodes, binary edges (connections between nodes), and “n-ary edges (annotation sets).”
- A unified algorithm treats node connects and set edges with equal importance.
- The reservation-based self organizing maps (RSOM) algorithm causes nodes that differ strongly spread out across the grid quickly, while more similar items separate more slowly (without staying on top of each other).

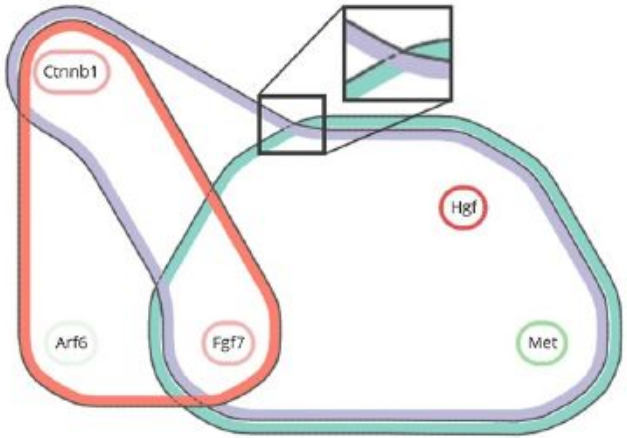


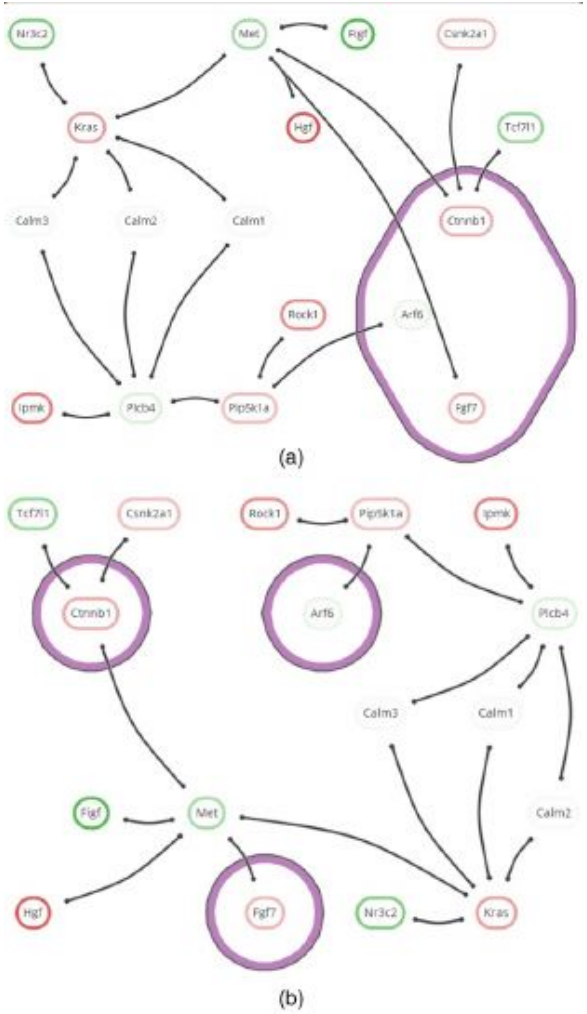
SOMs use a grid of artificial neurons (a) to learn an optimal layout. If a previous item has reserved space on the grid, the next item is forced outwards (b).

Methods

- Set edges are drawn in descending nesting order along the layout grid, and the width is limited to prevent occlusion.
- Corners are tightened using erosion and dilation, and various other refinements are applied.
- New maps are initialized with the previous layout, so that the new map keeps most of the previous structure.




Ctnnb1
Hgf
Arf6
Fgf7
Met



Sets can be interactively weighted by clicking annotations, so that the nodes cluster together more closely (a). Or, the network can completely determine the layout (b).

Demonstration!

Or if Tim can’t get it to work: https://www.youtube.com/watch?v=LFGKek_pgjw

Limitations

- Very slow, at least on my computer.
- Not particularly scalable, although this is partially due to the nature of network visualization.
- Requires “focus and context” prior to analysis. Networks and sets must be narrowed down prior to visualization.
- When generating a new map, there is a tradeoff between conservation of the previous layout and overall quality. The layout sometimes gets stuck after clicking multiple nodes.
- Contour colors are pretty but don’t encode any additional information.

References

1. Dinkla K, El-Kebir M, Bucur C-I, et al. eXamine: Exploring annotated modules in networks. *BMC Bioinformatics*. 2014;15(1):201. doi:10.1186/1471-2105-15-201