Data and text mining

Hive Panel Explorer: an interactive network visualization tool

Sarah E. I. Perez1, Aria S. Hahn2, Martin Krzywinski3 and Steven J. Hallam1,3,4,5,*

1Graduate Program in Bioinformatics, University of British Columbia, Genome Sciences Centre, Vancouver, BC V5Z 4S6, Canada, 2Department of Microbiology & Immunology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada, 3Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada, 4Genome Science and Technology Program, University of British Columbia, Vancouver, BC V6T 1Z4, Canada and 5ECOSCOPE Training Program, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received and Revised on May 23, 2020; editorial decision on July 14, 2020; accepted on July 22, 2020

Abstract

Motivation: Networks are used to relate topological structure to system dynamics and function, particularly in ecology systems biology. Network analysis is often guided or complemented by data-driven visualization. Hive one of many network visualizations, distinguish themselves as providing a general, consistent and coherent rule-based representation to motivate hypothesis development and testing.

Results: Here, we present HyPE, Hive Panel Explorer, a software application that creates a panel of interactive hive plots. HyPE enables network exploration based on user-driven layout rules and parameter combinations for simultaneous of multiple network views. We demonstrate HyPE’s features by exploring a microbial co-occurrence network constructed from forest soil microbiomes.

Availability and implementation: HyPE is available under the GNU license: https://github.com/hallamlab/HivePanelExplorer.

Contact: shallam@mail.ubc.ca

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Networks model relationships (edges) between components (nodes) of a system and aid in relating topological structure to dynamics and function (Chumpitazi et al., 2014; Giovannelli et al., 2017; Thompson et al., 2017). For example, nodes within a biological network may represent foodwebs, or pathway and taxonomic relationships in microbial data (Chumpitazi et al., 2014; Giovannelli et al., 2017; Thompson et al., 2017) while the edges may represent their physical or trophic interactions, respectively. Graph theory and measures such as degree distributions, modularity and connectance are commonly used to quantify data associations, inform predictive models or generate hypotheses associated with network topology (Chumpitazi et al., 2014; Giovannelli et al., 2017; Thompson et al., 2017).

HyPE is based on Hive plots (Krzywinski et al., 2012): a rule-based layout that positions nodes using a circular coordinate system (Fig. 1). This coordinate system is based on rules driven by relevant node and edge properties (i.e. degree, clustering coefficient, betweenness, etc.). Nodes are placed onto radially arranged axes and edges are drawn between nodes using curves. The rules determining a node’s assignment and position along an axis can be user defined.

Hive plots are a generalizable, flexible and extensible network visualization modality; however, understanding how to harness their versatility for a particular research question can be challenging. HyPE addresses this issue by providing users with the ability to explore different hive plot constructions in a grid, or ‘hive panel’ (Krzywinski et al., 2012). This design circumvents the need to determine optimal a priori rule sets and enables users to explore multiple combinations of network properties and data dimensions simultaneously. More information related to HyPE’s navigation features are in Supplementary Material.

2 Materials and methods

HyPE was constructed using D3 (Mike Bostock’s hive plot plug-in) (Bostock et al., 2011), JavaScript (Powell et al., 2004) and Python v2.7 (Rossum, 1995). HyPE accepts two comma separated or tab separated (’.tsv’ or ‘.txt’) files as input, a node file and an edge file, making it easy to include node and edge properties and creates a webpage with the interactive interface (Fig. 1).
Environmental metrics (Chumpitazi community structure and network properties can be related back to competition. In this light, a microbial co-occurrence network models interactions between taxa such as trophic exchange, predation or consumption. Correlations capture potential ecological patterns among groups of samples or in this case soil depth and treatment. Based on both weighted mean depth and indicators, the data suggest components A and B are representative of the H and N surface horizons, respectively. For example, Module A was the shallowest (7 cm), and 55% of its nodes were indicators for the H surface horizons, while Module B was deeper (13 cm) and 16% of its nodes were indicators for the N surface horizons. Thus, modules partitioned with harvesting treatment and soil depth (Fig. 1).

To address the second hypothesis, we compared taxonomy of OTU network modules to taxonomic and metabolic data from cognate metagenomes. Detailed information on metagenomes and processing is in Supplementary Material. We observed that differences in taxonomy among the metagenomes were consistent with taxonomic structure of the network modules. For example, Proteobacteria were significantly more abundant within the N LFH microbial community compared to the H LFH microbial community. Similar patterns were observed for Bacteroidetes and Actinobacteria and the annotated open reading frames attributed to these groups. Thus, metabolic function from the metagenomes could be taxonomically linked to network components associated with distinct soil horizons. Together, the data suggest that within LFH horizons, forest harvesting alters the architecture of connectivity, resulting in novel interactions with potential implications for nutrient and energy flow within the ecosystem.

3 Conclusion

HyPE can be used for different network types, sizes and topologies to showcase system properties. The forest soil use case demonstrates HyPE’s versatility in resolving relevant patterns within a complex biological network to support hypothesis development and testing.

Acknowledgements

Special thanks to Mónica Torres Beltrán, Alyse Hawley and Brent Townsend for helpful discussions on development and implementation of HyPE.

Funding

This work was supported by Genome Canada, Genome British Columbia, the Natural Science and Engineering Research Council (NSERC) of Canada and Compute Canada Canada.

Conflict of Interest: A.S.H. and S.J.H. are co-founders of Koonkie Inc., a bioinformatics consulting company that designs and provides scalable algorithmic and data analytics solutions in the cloud.

References

Bostock, M. et al. (2011) D3 data-driven documents. IEEE Trans. Visual. Comput. Graph., 17, 2301–2309.

Chumpitazi, B.P. et al. (2014) Gut microbiota influences low fermentable substrate diet efficacy in children with irritable bowel syndrome. Gut Microbes, 5, 165–175.

Giovannelli, D. et al. (2017) Insight into the evolution of microbial metabolism from the deep-branching bacterium, Thermovibrio ammonificans. Elife, 6, e18890.

Krzyszke, M. et al. (2012) Hive plots – rational approach to visualizing networks. Brief. Bioinf., 13, 627–644.

Powell, T. et al. (2004) JavaScript: The Complete Reference, 2nd edn. McGraw-Hill, Inc., New York, NY, USA.

Rossum, G. (1995) Python tutorial. Technical report, CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands.

Thompson, K. J. et al. (2017) A comprehensive analysis of breast cancer microbiota and host gene expression. PLoS One, 12, e0188873.