HENA, Heterogeneous Network-Based Data Set for Alzheimer's Disease

Elena Sügis[1], Ioannis Xenarios[2]

Institute of Computer Science, University of Tartu, Tartu Estonia[1],
UNIL Departement Formation et Recherche, University of Lausanne, Switzerland[2]

Alzheimer's disease and other types of dementia are the top cause for disabilities in later life and various types of experiments have been performed to understand the underlying mechanisms of the disease with the aim of coming up with potential drug targets. These experiments have been carried out by scientists working in different domains such as proteomics, molecular biology, clinical diagnostics and genomics. The results of such experiments are stored in the databases designed for collecting data of similar types. However, in order to get a systematic view of the disease from these independent but complementary data sets, it is necessary to combine them. In this study we describe a heterogeneous network-based data set for Alzheimer's disease (HENA) [1]. It is accessible via the Network Data Exchange (NDEx) repository [2] and via the figshare repository [3]. HENA integrates Alzheimer's disease-related data from well-known public data collections, as well as novel experimental and computational data sets generated by the members of the AgedBrainSYSBIO consortium. It combines 64 distinct computational and experimental data sets of six data types originating from nine data sources. Additionally, we demonstrate the application of state-of-the-art graph convolutional networks [4, 5], i.e. deep learning methods for the analysis of such large heterogeneous graph-structured biological data sets. We expect HENA to allow scientists to explore and analyze their own results in the broader context of Alzheimer's disease research.

References and useful links

[1] Sügis, E., Dauvillier, J., Leontjeva, A., Adler, P., Hindie, V., Moncion, T., Collura, V., Daudin, R., Loe-Mie, Y., Herault, Y., Lambert, J.C., Hermjakob, H., Pupko, T., Rain, J.C., Xenarios, I., Vilo, J., Simonneau, M., Peterson, H., HENA, Heterogeneous network-based data set for Alzheimer's disease. Nature Scientic Data vol 6, no. 1 (2019), https://doi.org/10.1038/s41597-019-0152-0

[2] Sügis, E., HENA ver.2: Heterogeneous network-based data set for Alzheimer's disease (with reduced number of coexpression edges). *The Network Data Exchange* (*NDEx*), https://doi.org/10.18119/N97300 (2019)

[3] Sügis, E. HENA: Heterogeneous network-based data set for Alzheimer's disease. *Figshare*, https://doi.org/10.6084/m9.figshare.c.4469240 (2019).

[4] Hamilton, W., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. *Adv*. *Neur*. *In*. **31**, 1024–1034 (2017).

[5] CSIRO data 61 investigative analytics, Stellar-ml v0.2.0: Machine Learning on graphs, https://github.com/stellargraph (2018).