Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding

Carlo Vittorio Cannistraci
University of California San Diego (UCSD), USA & KAUST Saudi Arabia

Most functions within the cell emerge thanks to protein-protein-interactions (PPIs), yet experimental determination of PPIs is both expensive and time-consuming. PPI-networks present signifi-cant levels of noise and incompleteness. Predicting interactions using only PPI-network-topology (topological prediction) is difficult but essential when prior biological knowledge is absent or unreliable.

Methods: Network embedding emphasises the relations between network proteins embedded in a low-dimensional space, in which protein-pairs that are closer to each other represent good candidate-interactions. To achieve network denoising, which boosts prediction performance, we first applied minimum-curvilinear-embedding (MCE), and then adopted shortest-path (SP) in the reduced space to assign likelihood-scores to candidate-interactions. Furthermore, we introduce: (i) a new valid variation of MCE, named non-centred-MCE (ncMCE); (ii) two automatic strategies for selecting the appropriate embedding-dimension; and (iii) two new randomised procedures for evaluating predictions.

Results: We compared our method against several unsupervised and supervised embedding approaches and node-neighbourhood techniques. Despite its computational simplicity, ncMCE-SP was the overall leader, outperforming the current methods in topological link-prediction.

Conclusion: Minimum curvilinearity is a valuable nonlinear framework that we successfully applied to the embedding of protein net-works for the unsupervised prediction of novel PPIs. The rationale for our approach is that biological and evolutionary information is imprinted in the nonlinear patterns hidden behind the protein net-work topology, and can be exploited for predicting new protein-links. The predicted PPIs represent good candidates for testing in high-throughput experiments or for exploitation in systems biology tools such as those used for network-based inference and prediction of disease-related functional-modules.