

Motivation: node2vec

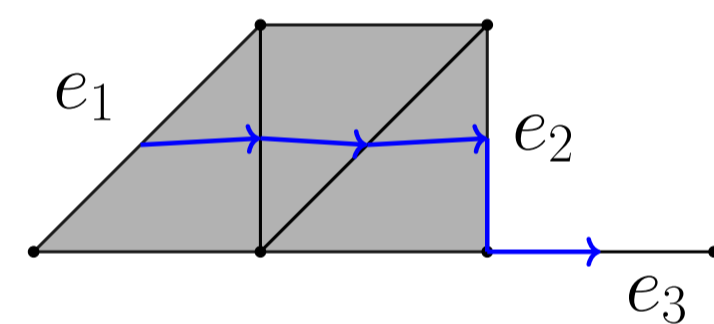
Node2vec [1] is a well known graph embedding method used for dimensionality reduction and preprocessing to obtain machine learning ready input.

- Node2vec used random walks on graphs as a similarity measure between nodes
- Learn a map $f : V \rightarrow \mathbb{R}^n$, with $n \ll |V|$
- For $u, v \in V$ the distance $d(f(u), f(v))$ reflects the probability of u, v co-occurring in a random walk
- Through node2vec gain insight on the structure of the graph

We generalize this method to the higher dimensional simplices of a simplicial complex.

Random walks on simplicial complexes

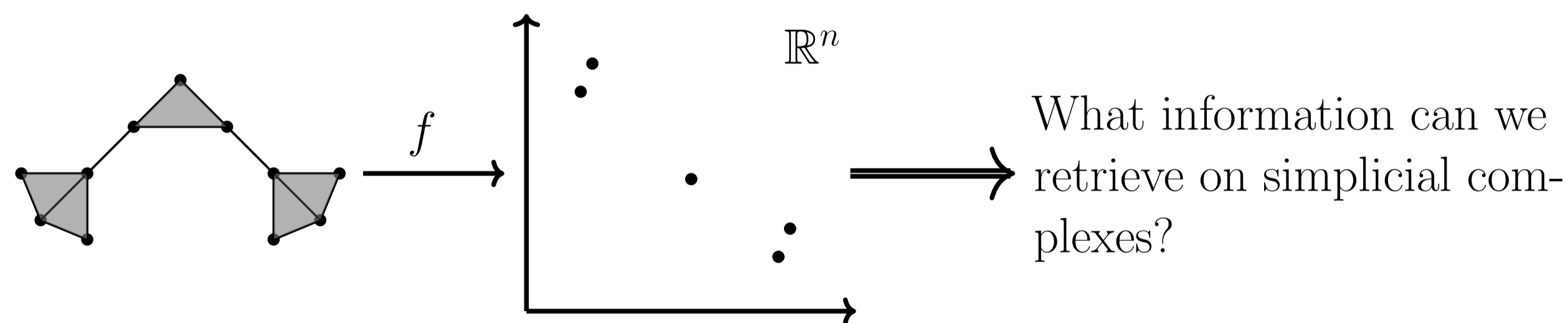
There are several choices for generalizing random walks to complexes. Here we say that k -simplices can be connected through common $(k + 1)$ -simplices and $(k - 1)$ -simplices.



The transition probabilities can be defined using the higher order Laplacians [2].

k -simplex2vec

k -simplex2vec learns a map $f : X_k \rightarrow \mathbb{R}^n$, where X_k is the set of k -simplices of a simplicial complex X .



The main points of the method are:

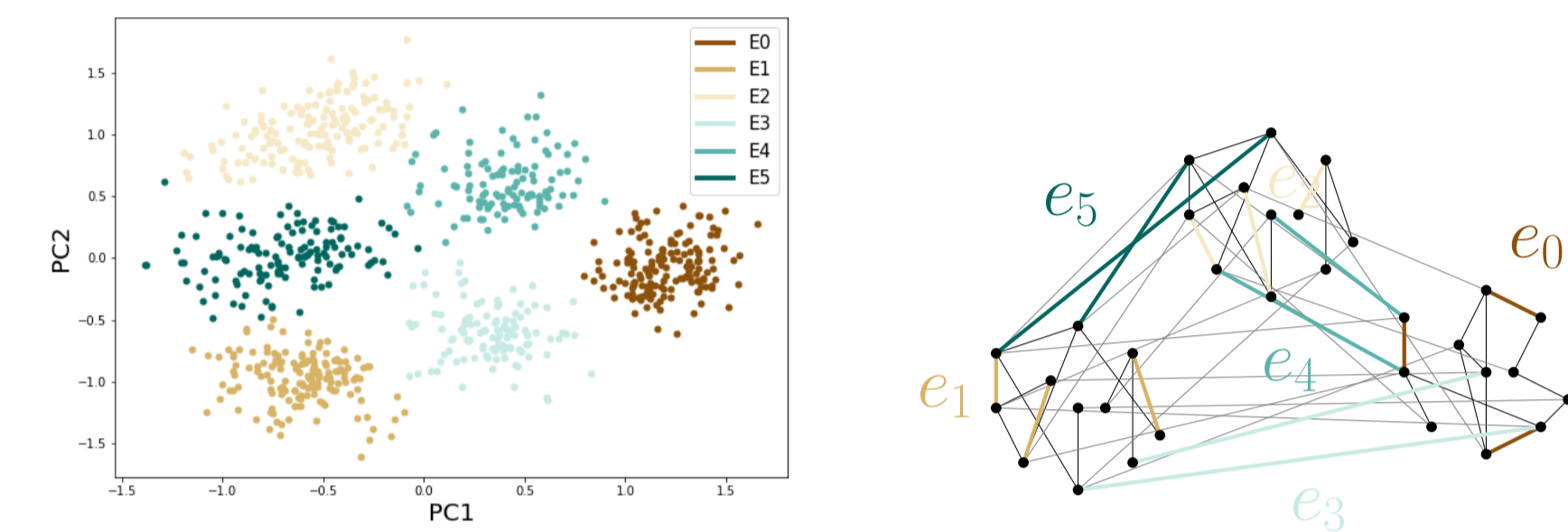
- Fixing the three parameters: walk length, number of walks starting at each simplex and the dimension n
- Using random walks on the k -simplices as a similarity between them
- For $\sigma, \tau \in X_k$, $d(f(\sigma), f(\tau))$ reflects their probability of co-appearing in the random walks

The code is available at: <https://github.com/cehahacker/k-simplex2vec>

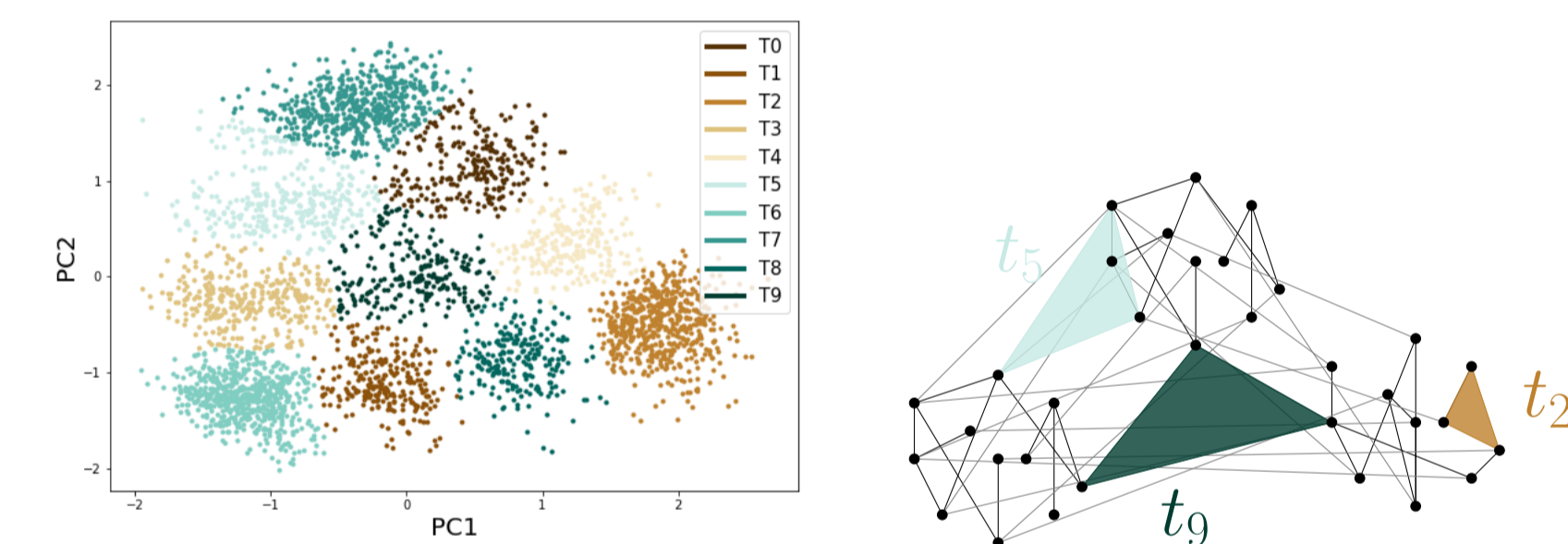
Results

We show the results on the clique complex of a stochastic block model with three blocks of nodes. The point clouds below are the result of k -simplex2vec on the edges and triangles respectively, along with the simplices corresponding to each cluster.

The k -simplex2vec representation of the edges



The k -simplex2vec representation of the 2-simplices



The clustering of the point clouds groups together the points corresponding to similar types of simplices, i.e., those that have nodes in the same blocks.

Conclusion and future directions

- Obtain a meaningful clustering of the simplices according to their structural role in the simplicial complex
- Will further explore the effect of this method on real-world data
- There is a link between the random walks and the homology of the complex [2], it would be interesting to study how this is reflected in k -simplex2vec

References

- [1] Aditya Grover and Jure Leskovec. "Node2vec: Scalable Feature Learning for Networks". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016.
- [2] Sayan Mukherjee and John Steenbergen. "Random walks on simplicial complexes and harmonics". In: *Random Structures & Algorithms* 49.2 (2016), pp. 379–405. DOI: 10.1002/rsa.20645. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20645>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.20645>.