M1/M2 Research Internship – Stage M1/M2 2022-2023 : Experimental Investigation of a Stochastic Self-Organizing Overlay

1 Supervision

Advisers	François Taiani, `francois.taiani@irisa.fr`, 02 99 84 75 04
	George Giakkoupis, `george.giakkoupis@inria.fr`
	Davide Frey, `davide.frey@inria.fr`
Lab	Centre Inria Rennes, IRISA (UMR 6074)
Team	WIDE (The World Is DistributEd)
	Équipes-Projet Inria / Département D1 IRISA

2 Context

Self-organizing overlays are a class of epidemic distributed protocols that allow a set of distributed nodes to construct "emergent" topologies in a robust, efficient, and highly scalable manner. Initially proposed 15 years ago [10, 13], they have since been applied to an extensive range of problems, from peer-to-peer recommendation [2, 5, 6, 9, 12], to distributed topology construction [3, 4], and KNN graph computation [7].

Self-organizing overlays (SOOs) assume a distance[1] function $d(-,-)$ on participants (1), and seek to construct for each participant $u \in \Pi$, a neighborhood $N(u)$ (or *view*) of its k closest peers (2).

$$d: \begin{array}{ccc} \Pi \times \Pi & \to & \mathbb{R} \\ (u,v) & & d(u,v). \end{array} \tag{1}$$

$$N(u) \in \operatorname*{argmin}_{S \subseteq \Pi \setminus \{u\} : |S| = k} \sum_{v \in S} d(u,v), \tag{2}$$

Starting from a random configuration, SOOs usually combine two stochastic processes to achieve this goal, that execute in periodic rounds:

- Relying on a Random Peer Sampling Service (RPS) [11], each node periodically probes some random peers and improve its current view accordingly.
- In parallel, each node periodically requests the view of one of its neighbors, chosen at random, and uses these new participants (these 'friends-of-friends') to improve its view.

Experimentally, SOOs have been shown to converge particularly rapidly in many situations, in a number of rounds that is proportional to the logarithm of the size of the system $O(\log(|\Pi|))$ (see for instance Figure 1), but no proof of this result has been proposed yet.

3 Objective

The supervisors and a previous intern, Lucie Guillou, have recently developed a formal analysis of the convergence of an SOO algorithm based on partitions of the search space. The SOO algorithm works in a one-dimensional space of 2^ℓ identifiers and repeatedly partitions this space using limited peer-to-peer exchanges to construct local views of k neighbors. The formal analysis provides a proof that this algorithm converges to top-k neighborhoods with high probability in $\log(n)$ rounds.

This internship aims to perform an in-depth experimental analysis of a generalization of this algorithm to higher dimensions on real and synthetic datasets.

We would first like to consider the procedures described by Dong, Indyk, Razenshteyn, and Wagner in [1] to transform a high-dimensional dataset into a one-dimensional one before applying our decentralized

[1] In some niche cases, d might not even be a distance.

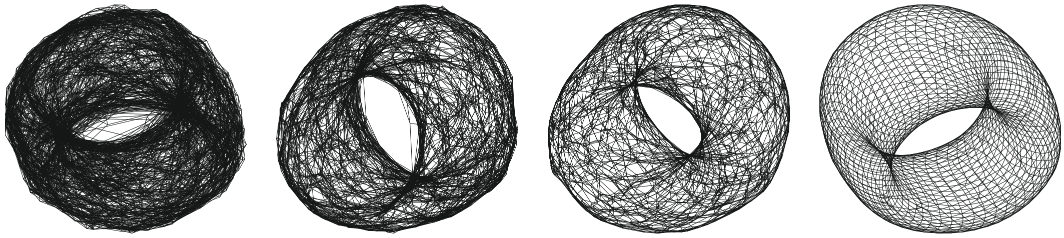

Figure 1: Convergence of a Self-Organizing Overlay with the T-Man algorithm executing on 2500 nodes, taken from [10]

SOO algorithm and measuring key performance metrics. We plan to apply such procedures to the datasets described in [8], and compare the obtained results to that of centralized solutions to approximate k-neighbor search problems.

In a second stage, we plan to perform a sensitivity analysis using synthetic datasets in order to assess to which extent key properties of the datasets (such as the distribution of pair-wise distances against that of the average pair-wise distance) influence the results of our algorithm.

References

[1] Alexandr Andoni, Piotr Indyk, and Ilya P. Razenshteyn. Approximate nearest neighbor search in high dimensions. *CoRR*, abs/1806.09823, 2018.

[2] Marin Bertier, Davide Frey, Rachid Guerraoui, Anne-Marie Kermarrec, and Vincent Leroy. The gossple anonymous social network. In *Middleware'2010*, 2010.

[3] Simon Bouget, Yérom David Bromberg, Adrien Luxey, and François Taïani. Pleiades: Distributed structural invariants at scale. In *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2018)*, pages 542–553, Luxembourg, June 2018.

[4] Simon Bouget, Hoel Kervadec, Anne-Marie Kermarrec, and François Taïani. Polystyrene: the decentralized data shape that never dies. In *2014 IEEE 34th International Conference on Distributed Computing Systems (ICDCS'14)*, pages 288–297, June 2014.

[5] Antoine Boutet, Davide Frey, Rachid Guerraoui, Anne-Marie Kermarrec, and Rhicheek Patra. Hyrec: leveraging browsers for scalable recommenders. In *Proceedings of the 15th International Middleware Conference*, pages 85–96. ACM, 2014.

[6] Antoine Boutet, Davide Frey, Rachid Guerraoui, Anne-Marie Kermarrec, Antoine Rault, François Taïani, and Jingjing Wang. Hide & share: Landmark-based similarity for private knn computation. In *45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'15)*, Rio de Janeiro, Brazil, June 2015.

[7] Wei Dong, Charikar Moses, and Kai Li. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the 20th international conference on World wide web*, pages 577–586. ACM, 2011.

[8] Yihe Dong, Piotr Indyk, Ilya P. Razenshteyn, and Tal Wagner. Learning space partitions for nearest neighbor search. In *ICLR*. OpenReview.net, 2020.

[9] Davide Frey, Anne-Marie Kermarrec, Christopher Maddock, Andreas Mauthe, Pierre-Louis Roman, and François Taïani. Similitude: Decentralised adaptation in large-scale p2p recommenders. In *Distributed Applications and Interoperable Systems*, pages 51–65. Springer, 2015.

[10] Márk Jelasity, Alberto Montresor, and Ozalp Babaoglu. T-man: Gossip-based fast overlay topology construction. *Computer networks*, 53(13):2321–2339, 2009.

[11] Márk Jelasity, Spyros Voulgaris, Rachid Guerraoui, Anne-Marie Kermarrec, and Maarten van Steen. Gossip-based peer sampling. *ACM TOCS*, 25, 2007.

[12] Anne-Marie Kermarrec, Olivier Ruas, and François Taïani. Nobody cares if you liked star wars: Knn graph construction on the cheap. In *Europar 2018*, volume 11014 of *LNCS*, pages 419–431, Turin, Italy, August 2018. Springer.

[13] S. Voulgaris and M. v. Steen. Epidemic-style management of semantic overlays for content-based searching. In *Euro-Par'05*, 2005.