Self-Adjusting Geometric Structures for Latent-Space Embedding

Eunhui Park and David M. Mount

Problem Statement:

- Analysis of human social networks involves storing and retrieving large dynamic point sets.
- Latent Space Embedding:
  - Given a social network, map the nodes into a geometric space in accordance with a logistic regression model, where the likelihood of an edge increases as the distance between points decreases.
  - Solved by application of MCMC methods, such as Metropolis-Hastings.
  - Efficiency depends on the ability to quickly answer queries regarding point relationships in a dynamic setting.
- Statistical Analysis of Moving Entities:
  - Given the motion sequence for a set of agents, perform statistical analyses of their pattern of motion and their spatial relationships.
  - This involves storing dynamic point sets and performing queries over these sets.

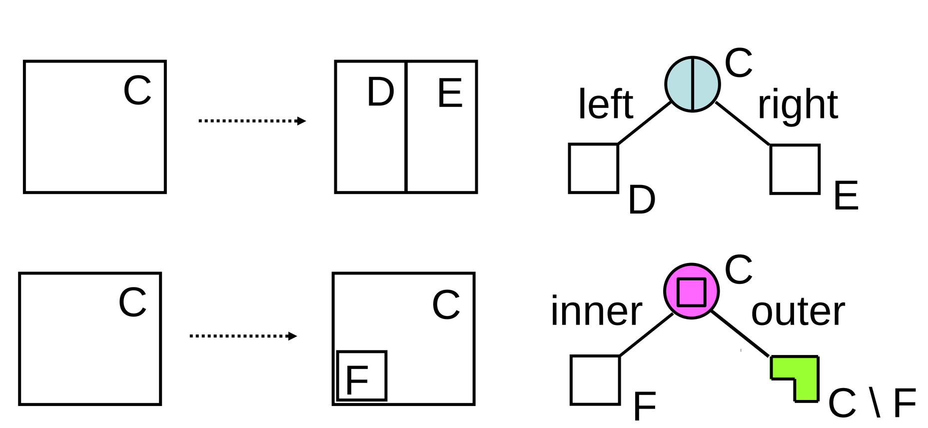
Our Approach:

- Given the unpredictable nature of MCMC algorithms, it is important that data structures adapt to the algorithm's access pattern. This leads to the concept of self-adjusting data structures.
- Sleator & Tarjan (1985) introduced the splay tree, a self-adjusting data structure for 1-dimensional data.
- We developed the quadtreap (SoCG 2010), a dynamic multi-dimensional data structure. It is not self adjusting.
- Splay Quad Tree: A new self-adjusting data structure for multi-dimensional data.

BD-tree:

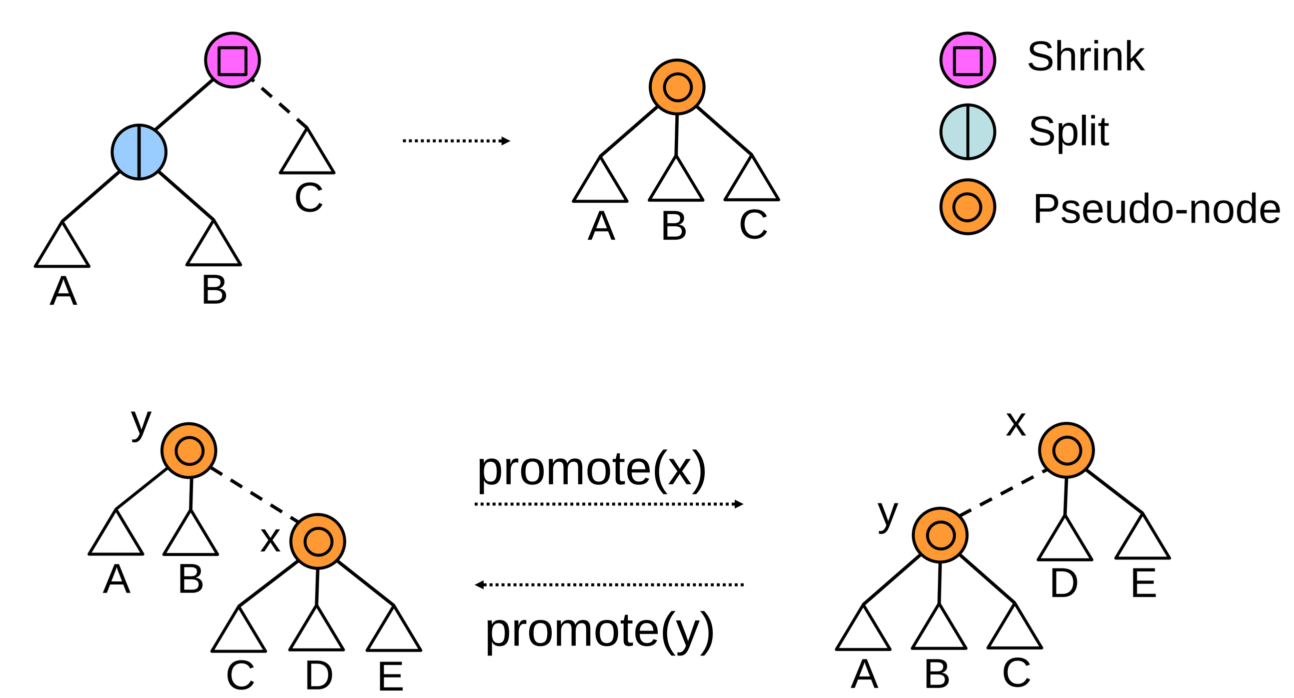
- Spatial decomposition based on:

- Split
- Shrink



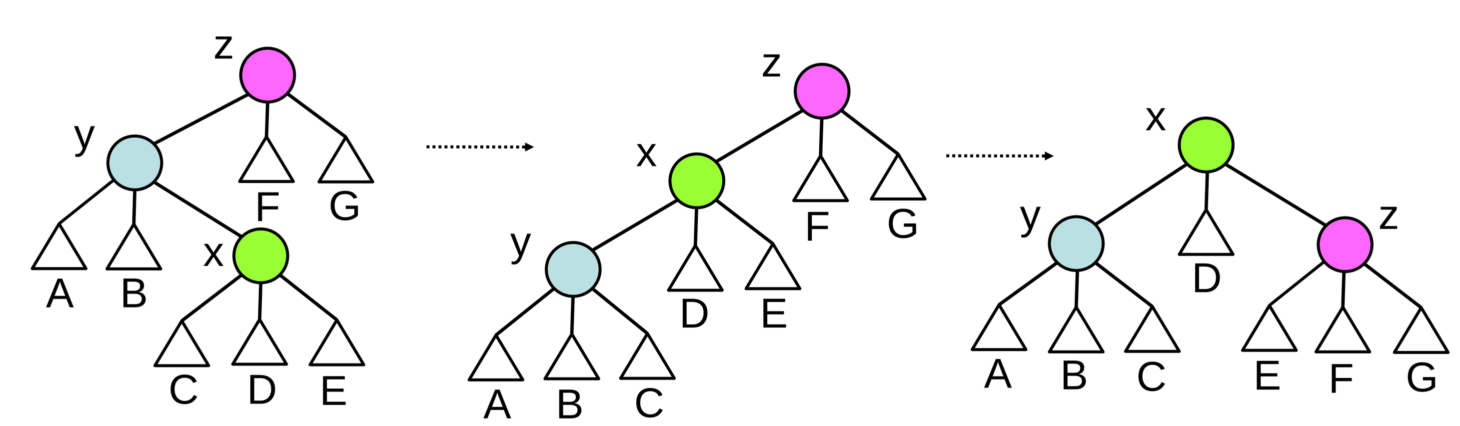
Rotation:

Balanced tree structure is maintained by rotating alternating pairs of shrink-split nodes:

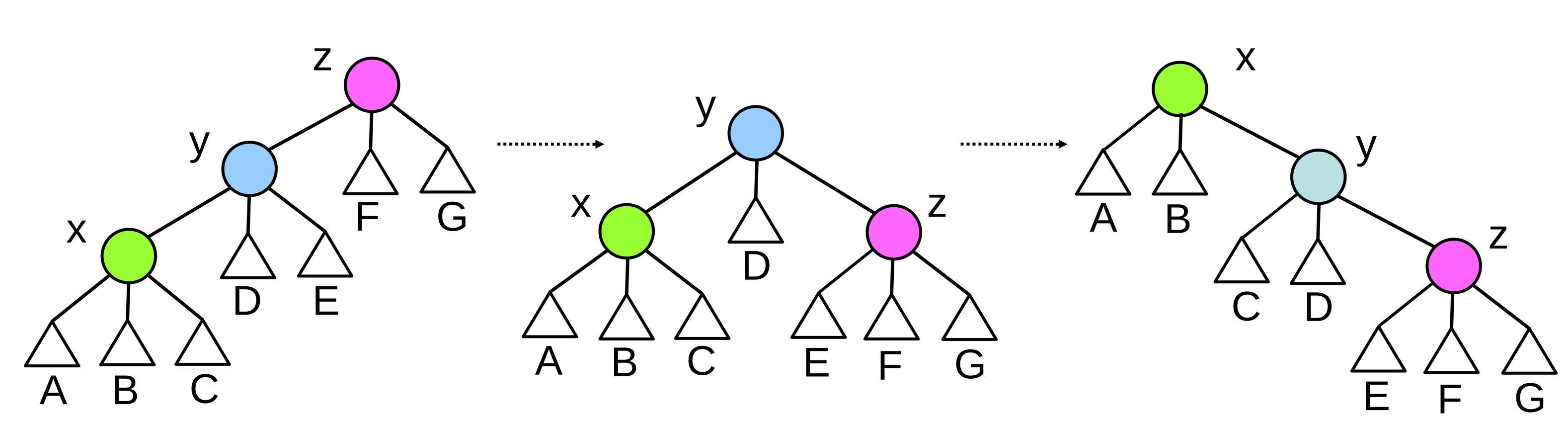


Splaying:

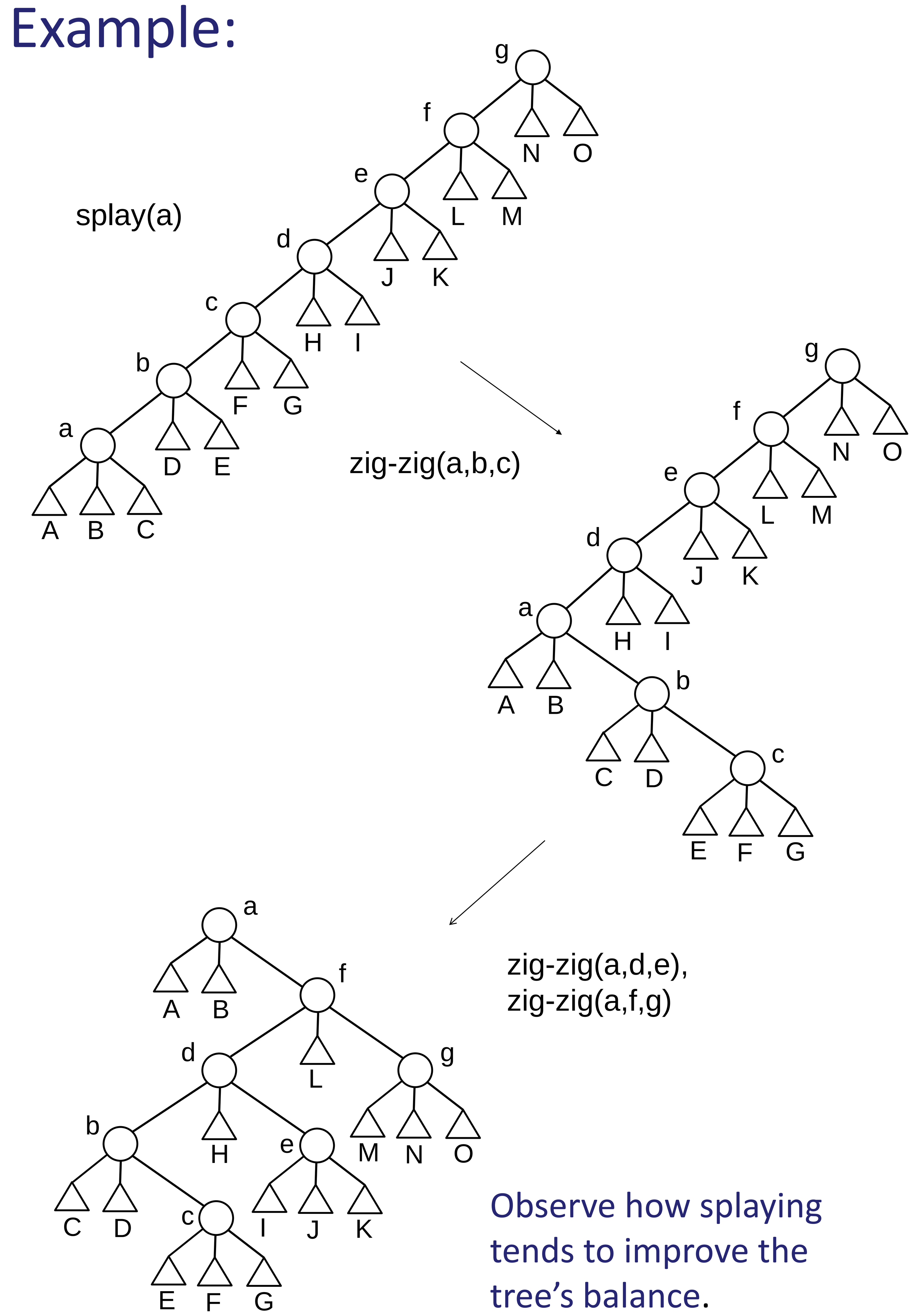
- When a node is accessed, we bring it close to the root through a series of tree rotations.
- Thus, each access makes future accesses to the same node more efficient.
- Efficiency is established through an amortized analysis.
- Primitive rotation operations:
  - Zig-zag



  - Zig-zig



Example:



Observe how splaying tends to improve the tree's balance.

Issues/Progress:

- Care must be taken to assure geometric integrity. (In particular, no multiple inner boxes).
- We have developed a provably correct splaying algorithm.
- Amortized analysis of efficiency:
  - Based on potential function approach from Sleator & Tarjan (1985). (In progress.)

Scalable Methods for the Analysis of Network-Based Data