IP2vec: An IP node representation model for IP geolocation

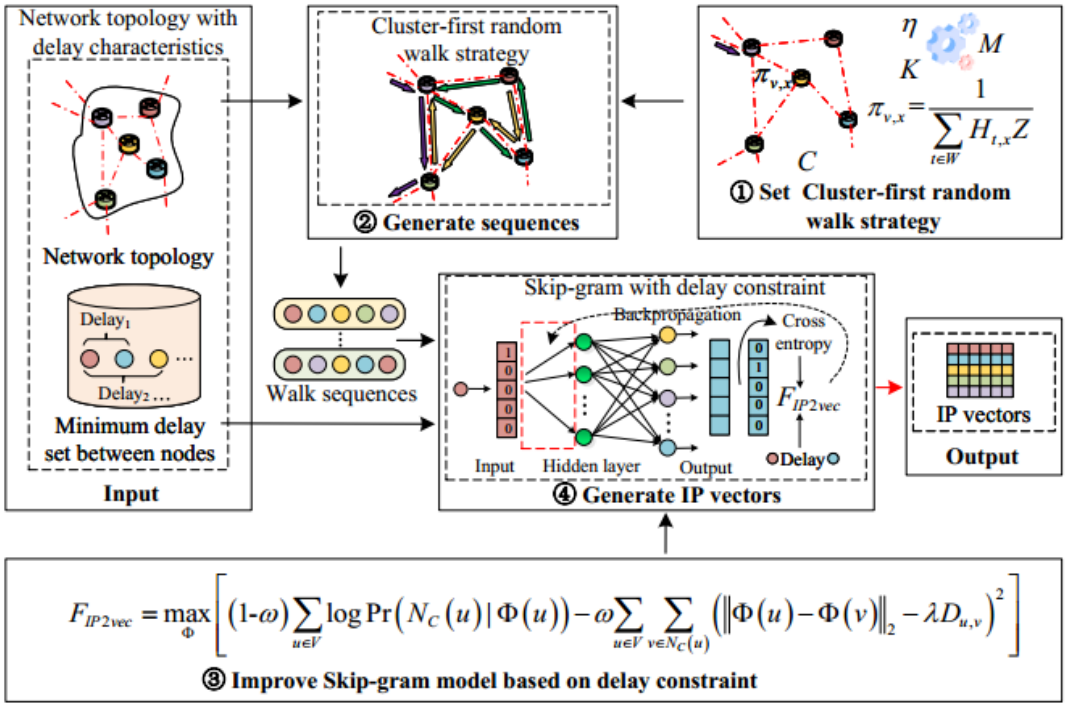
Fan ZHANG, Meijuan YIN, Fenlin LIU, Xiangyang LUO, Shuodi ZU

Frontiers of Computer Science, DOI: 10.1007/s11704-023-2616-9

Problems & Ideas

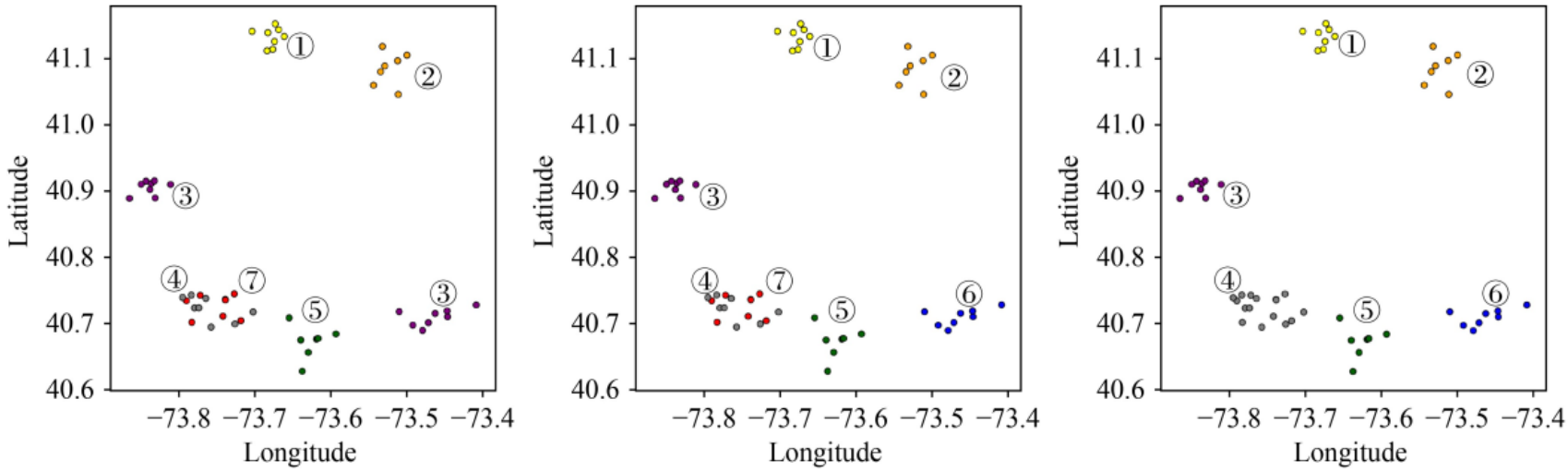
- Problems of conventional IP geolocation algorithms:
  - The accuracy of landmark-based IP geolocation algorithms is limited by the distance between the nearest landmark and the target IP.
  - IP geolocation algorithms based on machine learning cannot accurately represent the distance between any two nodes.
- Ideas: A representation learning model IP2vec that represents IP nodes according to the connection relation and delay between IP nodes.

The model represents IP nodes according to the connection relation and delay between IP nodes so that the similarity between IP vectors can reflect the distance and topological proximity between IP nodes.



Main Contributions

- Contributions:
  - A random walk strategy for IP geolocation is proposed: Cluster-first random walk strategy, which makes the adjacent IP nodes tend to be in the same sequence.
  - A vector-constrained objective function based on network delay is proposed, which enables the IP vector to reflect the distance of nodes in physical space.



(a) Node2vec

(b) IP2vec (Only use Cluster-first random walk strategy)

(c) IP2vec (Cluster-first random walk strategy with delay constraint)

The IP2vec model that adopts the Cluster-first random walk strategy is effective. The vector constraint through delay can also effectively make the landmark vector more related to latitude and longitude.