Coarse-graining Directed Networks with Ergodic Sets Preserving Diffusive Dynamics

Keywords: Directed Networks, Network Compression, Ergodic Sets, Random Walks

Extended Abstract

Motivation. Directed networks naturally arise in diverse fields such as transportation, food chains, and the World Wide Web, exhibiting complex phenomena like hierarchies and non-normal dynamics [2]. A significant challenge in analyzing these networks is non-ergodicity, where the existence of sources and sinks can make standard algorithms for community detection or node centrality ill-defined. While current approaches often use "tricks" like random teleportation to handle this [1], such methods effectively remove the structural complexity entirely. This work aims to leverage this complexity by identifying *ergodic sets*–subsets of nodes that are dynamically disjoint from the rest of the network–to simplify and compress the network while preserving its essential diffusive dynamics .

Approach and Methodology. We partition network nodes into three groups: forward ergodic sets (wells), backward ergodic sets (sources), and the transient core. We define a forward ergodic set as a strongly connected subgraph where no edges leave the set, essentially acting as an absorbing super-state. Backward ergodic sets are defined symmetrically as strongly connected subgraphs with no incoming edges.

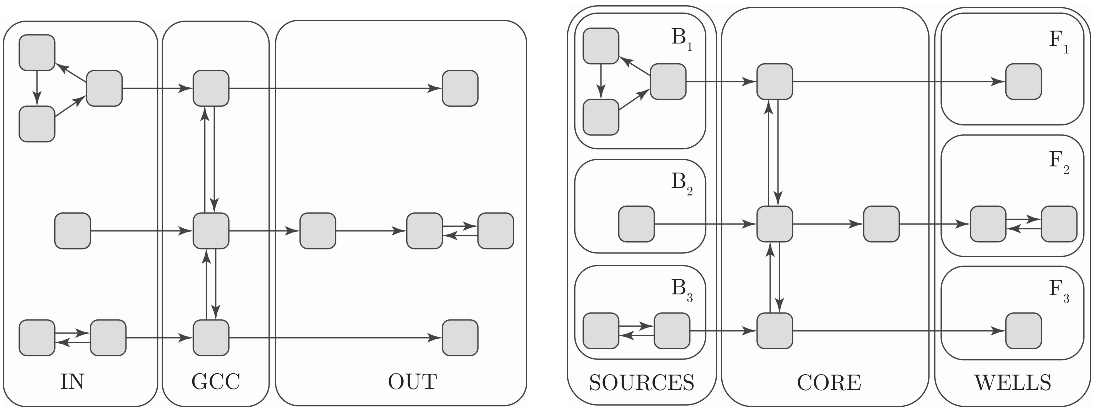

Figure 1: A representation of two different network partitions. **Left:** The classical *bow-tie* structure. **Right:** The proposed division into ergodic sets (Sources, Core, Wells).

The proposed Ergodic Set Compression Algorithm (ESCA) operates in two steps:

1. **Ergodic Set Coarse-graining**: Each ergodic set is deleted and substituted by a single meta-node. Then, we rewire the edges using local properties (in- and out-degrees) to preserve the marginal one-step transition probabilities and random walk flow losslessly. The rewiring weight rule for a backward ergodic set B compressed into a source meta-node n_B is:

$$\omega_{B,v} = \sum_{b \in B} \frac{\mathbb{I}_E(b,v)}{d_{out}(b)} p(b) = \frac{1}{\sum_{b \in B} d_{in}(b)} \sum_{b \in B} \frac{d_{in}(b)}{d_{out}(b)} \mathbb{I}_E(b,v) \tag{1}$$

in which $\mathbb{I}$ is the indicator function, and $p(v)$ is the probability of a random walker to be at node v. The rewiring weight for a forward ergodic set F compressed into a well meta-node v_F is proportional to the number of edges reaching the forward ergodic set before deletion.

2. **Core Compression**: The transient core is treated as a black box that mixes probability between sources and wells. We utilize the Singular Value Decomposition (SVD) on the core mixing matrix to compress the core into latent states, allowing for further reduction of network size.

Results. When applied to real-world networks, the ESCA achieves significant compression. While the first structural step only yields an average compression C_1 of around 1%, it provides am important structural simplification. By leveraging such simplification, the second compression step reaches a compression factor C_2 of approximately 50% on average. Furthermore, we found that the internal structure of real-world cores significantly differs from rewired null models, showing a lower mixing capability than random configurations.

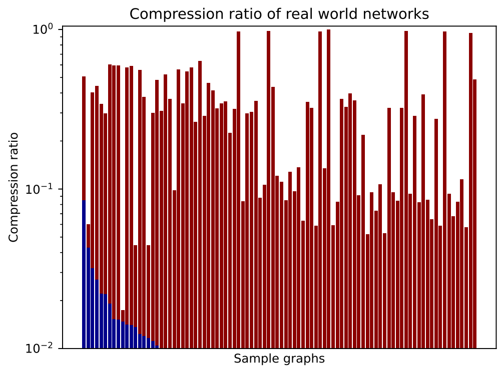

Figure 2: Compression factor the ESCA, applied to a selection of empirical networks from biology (protein interaction networks), information technology (website networks), digital economy (Amazon co-purchase networks), and transportation (airline networks). The blue data is the compression factor C_1, the red data is the total compression factor C_2.

Conclusions and Outlook. The ESCA provides a robust framework for directed network compression that preserves the random walk dynamics. By partitioning nodes based on reachability, we uncover hierarchical structures that traditional "bow-tie" models might miss. Future work will focus on integrating this compression with specific downstream tasks, such as accelerated community detection and identifying dynamical distances within the latent space of the compressed core.

References

[1] David F Gleich. Pagerank beyond the web. *SIAM review*, 57(3):321–363, 2015.

[2] Robert S MacKay, Samuel Johnson, and Benedict Sansom. How directed is a directed network? *Royal Society open science*, 7(9):201138, 2020.