Graph Matching Algorithms: A Spectral Approach

Liam Thomas, working under Professor Abiy Tasissa

Introduction

Graph theory is concerned with sets of lines and points, respectively termed **edges** and **vertices**. Edges and vertices combine to form systems called **graphs**. Social networks and maps of cities can be modeled in applied graph theory, along with other physical, biological, and social systems. The **Graph Matching Problem** is a fundamental topic in the field concerned with training computers to find similarities between graphs.

Our approach to this problem is based in the sub-field of **spectral graph theory**, which combines linear algebra and graph theory to yield unique insights about graphs and their associated matrices.

Methods

In the field of **spectral graph theory**, square matrices are used to describe graphs. Matrices can also be used to describe **random walks**, which can be thought of as a series of random traveling from the nodes of a graph. Our graph matching algorithm uses random walks to generate a novel distance-measuring technique called **diffusion distance**.

Our algorithm measures precisely how different two graphs **G1** and **G2** are, both in terms of their **global structure** and their **local structure**. The formal statement of the optimization problem is written as:

$$\min_{P \in \Pi} ||AP - PB||_F^2 + \beta \, \mathrm{tr}(C^T P)$$

1. Take graphs ***G1*** and ***G2*** as inputs
2. Compute their adjacency matrices, ***A*** and ***B***
3. Use diffusion distance to construct the cost matrix ***C***
4. Use **convex optimization algorithms** to find the optimal **permutation *P***
5. Use this **permutation matrix *P*** to compute the optimal shuffling of the graphs

Approach

We expected to use the technique of **diffusion distance** to construct the cost matrix and inform our graph matching algorithms. This approach has not yet been attempted in the literature.

Results

While our algorithm has not yet been compared side-by-side with the current fastest-recorded solvers of the graph matching problem, initial work on simple graphs and matrices has been promising.

Figures 2 and 3 describe a computation of diffusion distance that was performed on the graph shown in Figure 1, called "AG2105." One of the advantages of diffusion distance is that it takes into account time as a parameter. As is apparent in Figure 3, the distance stabilizes to a steady distribution at an exponential rate as time approaches infinity. The speed of the convergence of the diffusion distance is directly correlated to how strongly connected a graph is; the example shown in figures 2 and 3 is one of rapid convergence.

So far our efforts have been dedicated to the construction of the **cost matrix** (which measures structural differences between two graphs) via diffusion distance. This is a **local, node-by-node** measure of difference. Once this step is completed, we will turn to **global** measures such as taking the Frobenius inner product of the difference between the permuted adjacency matrices of two graphs.

At this point, we will test our algorithm against leading record-holders for the Graph Matching Problem.

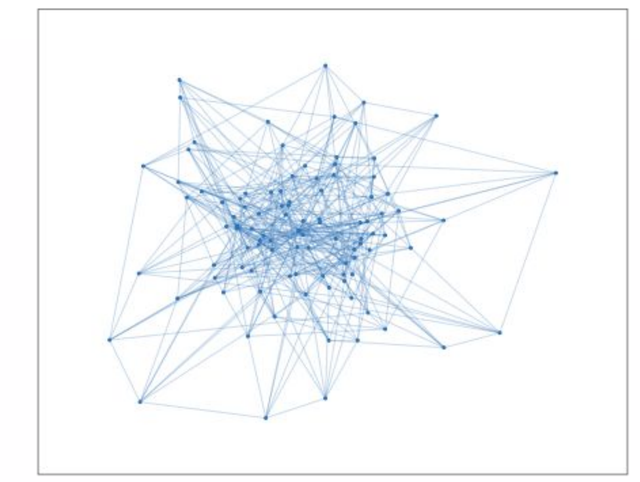

Figure 1: *The Graph AG2105*

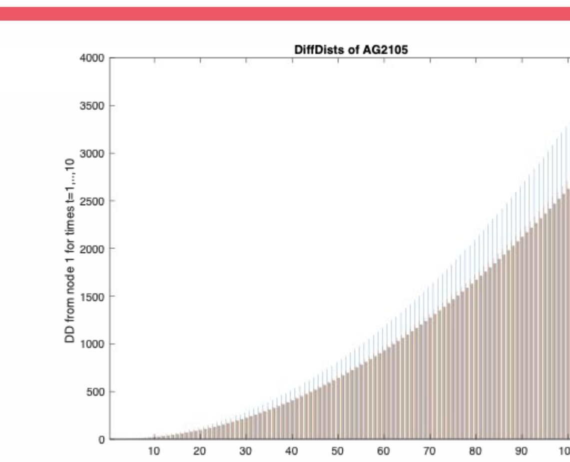

Figure 2: *Diffusion distances between nodes of AG2105*

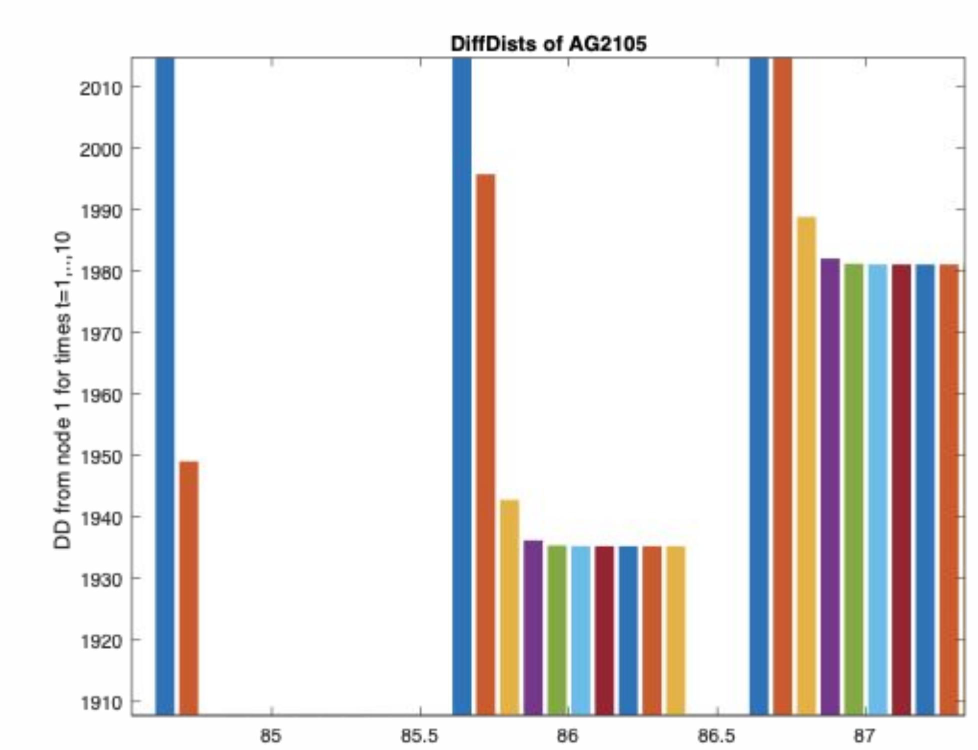

Figure 3: Close-up of node 86 on graph AG2105

- Boaz, Nadler *et al*. "Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators." Neural Information Processing Systems, No. 18. December 2005.
- Chung, Fan. *Spectral Graph Theory*. CBMS No. 92. Providence: American Mathematical Society. 1997.
- Lovásv, L. "Random walks on graphs: a survey." Bolyai Society Mathematical Studies, Vol. 2. 1993.
- Narayanan, Arvind and Shmatikov, Vitaly. "De-anonymizing social networks." 30th IEEE Symposium on Security and Privacy. 2009.
- Nica, Bogdan. *A Brief Introduction to Spectral Graph Theory*. Zürich: European Mathematical Society. 2018.

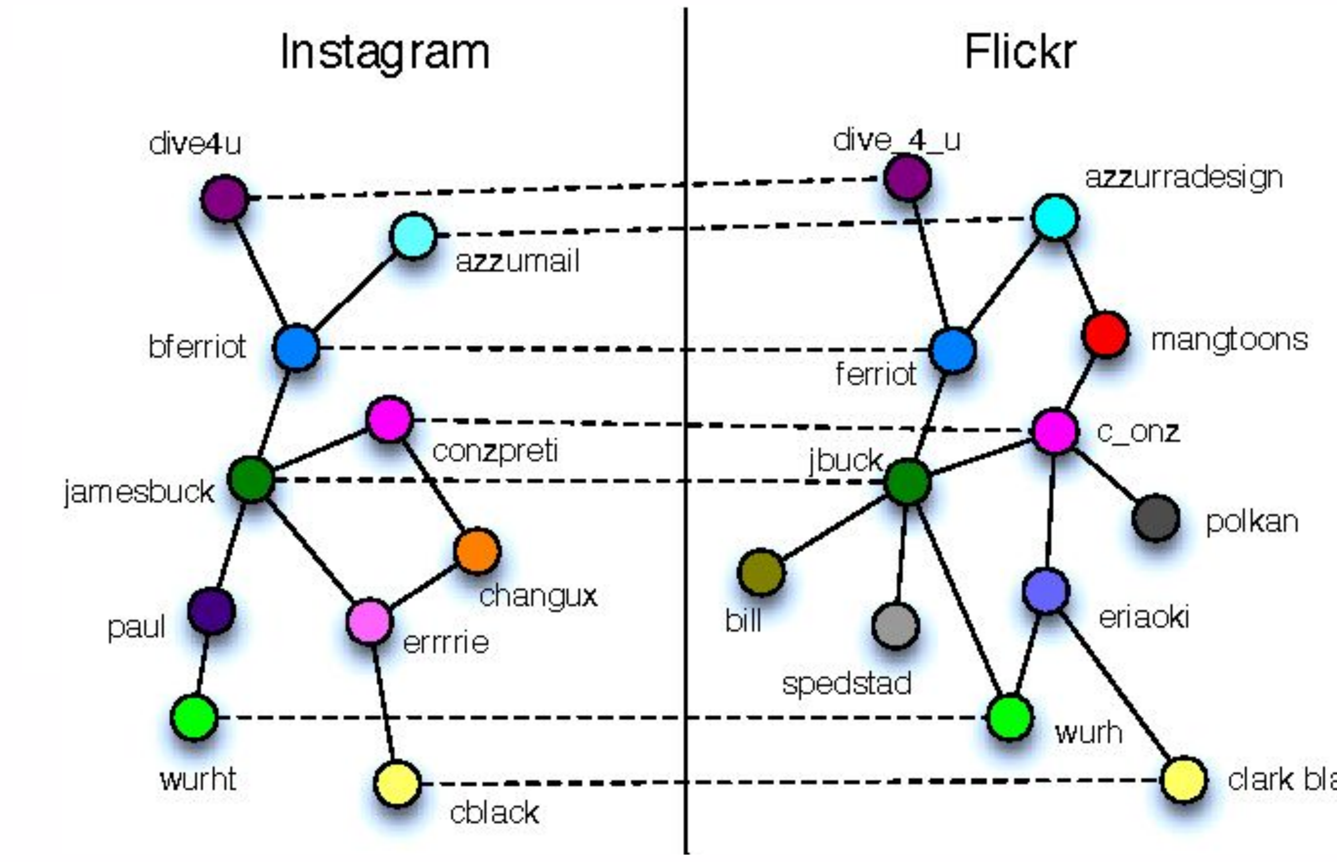

Figure 4: *Modeling social networks*

Future study

After the comparison with current record-holding algorithms is complete, we will modify our algorithm accordingly and write a paper summarizing the results.

One fruitful application of graph matching is in the modeling of social networks, as shown in Figure 4. Graph matching on directed or undirected graphs can quickly compare pairwise connections between different social media accounts across platforms (such as "friends" or "followers") and uniquely identify common accounts.

Following the completion of a technical paper, future endeavours could include more theoretical work on the problem, possibly incorporating other fields of mathematics. In addition, we may apply the algorithm to real-life data sets and graphs from various applications.

Many thanks to the Tufts Summer Scholars program, and to Professor Tasissa.