

SGM

<https://arxiv.org/pdf/1209.0367.pdf>

Regular graph matching: minimize adjacency disagreements from 2 graphs

Seeded graph matching: “ “ from 2 partially-matched graphs

Intro

Informally, graph matching is a bijection between 2 vertex sets that minimizes adj. disagreements. Given graphs G_1, G_2 with vertex sets V_1, V_2 , $|V_1|=|V_2|$, and 2 vertices u, v

- u, v adjacent in G_1 but not in $G_2 \rightarrow$ disagreement
- u, v not adjacent in G_1 , but are in $G_2 \rightarrow$ disagreement
- u, v adjacent in both \rightarrow agreement
- u, v not adjacent in both \rightarrow agreement

It turns out that general graph matching is just the quadratic assignment problem! This is NP-hard and its is proven that there is no approximation algorithm in NP unless $P=NP$:(. Even the question asking whether there exists a graph isomorphism is hard and intractable :(.

Pero

What about if we already know about a few of the elements?

For example, let

W_1 in V_1 , W_2 in V_2 s.t. $|W_1|=|W_2|$, and we have a bijection B_1 between the W s.

We can look at the total graph matching bijection B_2 as an extension of B_1 (so all matching from B_2 on the vertices B_1 acts upon are the same). Imagine it as growing B_1 out to B_2 .

The elements of the W s are **seeds** and B_1 is a **seeding**

The paper claims that even a little knowledge about seeds significantly improves our results

Vogelstein et al (2015) developed a Fast Approximate Quadratic Assignment Problem (FAQ), which will be modified to fit the SGM universe with 3 goals

1. Use seeds
2. Work on differently-sized graphs
3. Give each vertex a probability distribution of potential matches

There is a set of quadratic assignment problems from QAPLIB

(<http://anjos.mgi.polymtl.ca/qaplib/>) which FAQ has worked really well on, so it seems like a good thing to try and use from SGM and is faster than PATH. ALSO under mild conditions, the FAQ algorithm asymptotically almost surely provides the optimal graph matching solution when solved exactly.

All graphs covered are undirected and dont have loops or multiple edges. However, this SGM implementation would work on any crazy graph you want.

Notation

The notation in the paper is slightly modified here to fit the typesetting style of google docs.

We have n , m , and n such that $n+m=n$.

We have graphs G_1, G_2 with vertex sets $V_1, V_2 = \{1, 2, \dots, n\}$.

We have seed vertex subsets $W_1, W_2 = \{1, 2, \dots, m\}$ with seeding B_1 .

When $m=0$, this is the unseeded graph matching problem.

We have A, B in $R^{n \times n}$ which are adjacency matrices for G_1, G_2 respectively.

Each entry in these adjacency matrices is either 1 (for adjacency) or 0 (not), indexed as a_{ij} or b_{ij}

A useful partitioning of A is into $A_{11}, A_{12}, A_{21}, A_{22}$, where A_{11} represents the upper-left quadrant which corresponds to the seed. The same holds for B .

We have the following expression which is the objective of SGM

$$\min_{P \in \Pi_n} \|A - (I_m \oplus P)B(I_m \oplus P)^T\|_F^2,$$

Where I_m is the $m \times m$ identity matrix and the plus with circle is direct matrix addition (diagonal concatenation with 0s on the off-diagonal blocks). P one element of Π_n , an $n \times n$ permutation matrix. We can see that $I_m \oplus P$ results in a $n \times n$ matrix. This matrix simply represents the reorganization of B . The post-multiply reorganizes the columns of B and the pre-multiply reorganizes the rows of B . Thus, the permutation that minimizes the frobenius norm of the different between A and permuted B would be considered the “optimal” graph match.

This paper also utilizes the notion of a rho-SBM with number of blocks k and vertices n , vertex set V , block membership function b (that maps from vertex to block), edge probability matrix Λ ($k \times k$), and correlation ρ . We say that random graphs G_1 and G_2 , each having vertex set V with respective adj matrices A and B , have a rho-correlated SBM with parameters k, b , and Λ denoted $(G_1, G_2) \sim \text{rho-SBM}(k, b, \Lambda)$ if:

1. Each graph is marginally distributed as a k -SBM denoted $G_1, G_2 \sim \text{SBM}(k, b, \Lambda)$
2. The random variables A_{ij} and B_{kl} have correlation ρ if $i=k, j=l$ and are otherwise independent

$\rho=0$ implies edge presence is independent across graphs, while $\rho=1$ implies the graphs are almost surely isomorphic. This rho-correlation definition holds for $k=1$ (rho-correlated ER model) and $k=n$ (rho-correlated Bernoulli random graph).

Generation of a rho-SBM(k, b, Λ) is discussed:

1. Sample $G_1 \sim \text{SBM}(k, b, \Lambda)$
2. For G_2 , sample each $B_{ij} \sim \text{Bernoulli}((1-\rho)\Lambda_{\{b(i), b(j)\}} + \rho A_{\{ij\}})$

xd

Seeded-FAQ for approximate seeded graph matching