Canonical Labelling of Graphs

Aayush Ojha, Arunothia, Yash Singh

Computational Complexity Project
IIT Kanpur

12^{th} April, 2016

Acknowledgements

We would like to thank Prof. Raghunath Tewari for giving us the opportunity to explore this wonderful problem of canonical labelling. We would also like to thank Prof. V. Arvind of IMSc whose notes on the basics of the Graph Isomorphism problem were a great deal of help to us.

Preliminaries: Actions & Blocks

Let us assume that G is a group acting on a set Ω. We call a subset $\Delta \subseteq \Omega$ a block if for any $g \in G$ either,

$$\Delta^g = \Delta$$

or,

$$\Delta^g \cap \Delta = \phi$$

We call the group G transitive on Ω if for any $a, b \in \Omega$, $\exists g \in G$ such that $a^g = b$.

In the event that G is in fact transitive on Ω we see that a block Δ moves around the entire set Ω and covers it with images of itself. This induces a partition of Ω which is clearly G-invariant. Such a partition on Ω is called a *block system*.

Preliminaries: Actions & Blocks

- We note that for any action G a trivial partition exists where the block system is the partition of singletons.
- If G is a transitive action and no non trivial block system of Ω exists then we call G a *primitive* action.
- We call a block system minimal if the number of partitions of Ω created by the block system is minimum.
- Finding the minimal block system is in P (it reduces to finding the largest connected component of a graph).

Preliminaries: Composition Series and Width

- A composition series of a group G is a subnormal series of finite length

$$1 = H_0 \triangleleft H_1 \triangleleft \cdots \triangleleft H_n = G$$

- Composition factors - H_{i+1}/H_i
- For a group G let us define the composition width of G (denoted $cw(G)$) as the smallest positive integer d such that every non-abelian composition factors of G can be embedded in $Sym(d)$.
- A group G is solvable if there are subgroups,

$$\{1\} = G_0 < G_1 < \cdots G_k = G$$

such that G_{j-1} is normal in G_j and G_j/G_{j-1} is an abelian group for $j = 1, \cdots k$. For a solvable group G, $cw(G) = 1$.

Isomorphism and Canonical Forms

Graph Isomorphism

The graphs X_1 and X_2 are G-isomorphic, if

- $X_1 = \sigma(X_2) = X_2^\sigma$
- σ acts on the vertex set, $V(X)$.
- Putting $G = Sym(V(X))$, reduces G-isomorphism to isomorphism.

Let $\mathcal{K}$ be the class of permutations of the vertex set closed under G-isomorphisms. (i.e)

- Let $X \in \mathcal{K}$ and $\sigma \in G$ then, X^σ will also be in $\mathcal{K}$.

Canonical Form

A function $CF : \mathcal{K} \to \mathcal{K}$ is a canonical form if

1. For $X \in \mathcal{K}$, $CF(X) \equiv_G X$
2. For $X, Y \in \mathcal{K}$, $X \equiv_G Y$ iff $CF(X) = CF(Y)$

Canonical Labelling/Placement

- A canonical form corresponds to a set of labellings which consists of the permutations that map X to $CF(X, \sigma G)$.
- We define the function,

$$CL(X, \sigma G) = \{\tau \in \sigma G | X^{\tau} = CF(X, \sigma G)\}$$

We can observe that,

$$CL(X, \sigma G) = \sigma CL(X^{\sigma}, G) \quad (1)$$

$$CL(X, \sigma G) = \tau Aut_G(X^{\tau}) \quad (2)$$

for any $\tau \in CL(X, \sigma G)$.

- By (2) we expect $CL(X, \sigma G)$ to be a sub-coset of G.

Lemma 1

Babai started by proving a very interesting lemma. Let $\mathcal{C}$ the set of sub-cosets of G where $G \subseteq Sym(V)$ and $\mathcal{K}$ is a set of graphs closed under isomorphisms.

Lemma (Babai)

Let $CL : \mathcal{K} \times \mathcal{C} \to \mathcal{C}$ be a function such that, if $X \in \mathcal{K}$ and $\sigma \in \sigma G \in \mathcal{C}$, then $CL(X, \sigma G) \subseteq \sigma G$ and (1) and (2) hold then $CF(X, \sigma G) = X^{\tau}$ for any $\tau \in CL(X, \sigma G)$ defines a canonical form on $\mathcal{K}$ w.r.t σG and CL is the corresponding canonical labelling coset.

Reducing Graphs to Strings

- A graph X with n vertices can be studied through it's adjacency matrix $A(X)$. The group S_n induces an action on $A(X)$ via,

$$A(X)[i \times j]^{\sigma} = A(X)[i^{\sigma^{-1}} \times j^{\sigma^{-1}}]$$

  where $\sigma \in S_n$.
- The canonical form $CF(x, G)$ for a string x is a lexicographically smallest string in the G-orbit of x which we call the lex-leader of x.
- We can see that finding the lex-leader is NP-Hard. (It trivially solves the maximum independent set problem).

A String Canonization Algorithm

We describe the algorithm Babai uses to find a Canonical Placement sub-coset for stings. To use recursion we compute the canonical-placement cosets w.r.t. σG for substrings induced on a G-invariant subset B of A. We denote this sub-coset by $CP_B(x_B, \sigma G)$ where x_B denotes the restriction of x to B.

String Canonization

Input: A string $x \in \Sigma^A$; a coset $\sigma G \subseteq Sym(A)$; a G-stable subset.
Output: $CP_B(x_B, \sigma G)$, a sub-coset of σG.

1. If $|B| = 1$ then $CP_B(x_B, \sigma G) = \sigma G$.
2. If G is intransitive on B let C be the first G orbit. We see $B = C \cup D$. The problem is reduced to calculating,

$$CP_B(x_B, \sigma G) = CP_D(x_D, CP_C(x_C, \sigma G))$$

3. If G is transitive on B and $|B| > 1$ then let H be the stabilizer of the blocks in the first minimal G-block system (recall this is in P) in B. We can write,

$$\sigma G = \bigcup_{i=1}^{r} \sigma_i H$$

Now let us assume that,

$$CP_B(x_B, \sigma_i H) = \rho_i H_i$$

We now distinguish the good sub-cosets (which contain maps to lex-leader) and the bad sub-cosets (which don't map to lex-leader). Without loss of generality assume that the s sub-cosets $\{\rho_i H_i\}_{i=1}^{s}$ are all the good sub-cosets. The canonical placement sub-coset in this case is in fact,

$$\rho_1 < H_1, \{\rho_1^{-1} \rho_i\}_{1 \leq i \leq s} >$$

Using a double induction argument on $|G|$ and $|B|$ we see that our canonical placement sub-coset construction satisfies the conditions of Lemma 1 (subsequently equations (1) and (2)) and hence we end up with a canonical form.
Using (2) we can easily see that in step 3 of our algorithm,

$$H_1 = H_2 = \cdots = H_s = Aut_H(CF(x_B, \sigma G))$$

and hence our construction in step 3 is a very natural one.

Time Complexity Analysis

The group operations (including finding - Orbits, First Minimal Block System and Group Stabilizers) are polynomial and hence, our time analysis reduces to analysing the recursion in Case-2

$$t(|B|) \leq t(|C|) + t(|B| - |C|)$$

which is polynomially bound.

Theorem-1

If G is a primitive permutation group of degree n, and $cw(G) \leq d$ then $|G| \leq n^{\omega(d)}$

For large d, $\omega(d) \leq d * log(d) + c$

In case 3, the problem is reduced from solving (G, B) to solving $[G : H]$ problems of (H, B). Solving (H, B) decomposes into solving problems on disjoint orbit, each of size $(\frac{|B|}{m})$, where m is the number of blocks in the first minimal block decomposition.

Time Complexity Analysis

Hence, using Theorem-1, we get the following bound on $t(|B|)$ in case-3,

$$t(|B|) \leq m^{\omega(d)+1} * t(\frac{|B|}{m})$$

We therefore get,

Theorem-2

The canonical placement algorithm for $\sum^A$ w.r.t G runs in time $O(n^{\omega(d)+c})$, where $n = |A|$ and $d = cw(G)$.

For large d, $\omega(d) \leq d * log(d) + c$
In particular, the algorithm runs in polynomial time if we consider good groups (that have bounded $cw(G)$).

Canonization Strategies

We employ two key strategies in order to calculate the canonical placement coset.

- Refinement : We can classify vertices in a graph X by their degree. Let V_i denote the vertices of degree i. If we assume that the permutation σ reorders vertices by their valence i.e. $i^\sigma \leq j^\sigma$ iff. $deg(i) \leq deg(j)$. Our task now reduces to study the subgroup,

  $$H = Sym(V_0^\sigma) \times \cdots Sym(V_n^\sigma)$$

  Now $CL(X, G) = CL(X, \sigma H)$.

- Individualization: The second strategy is to choose an arbitrary vertex denoted $v \in V$. We look at the stabilizer of v, G_v and study it's cosets and employ an approach similar to that of step 3 of our algorithm. Note that this procedure can be done for any arbitrary edge too.

Applications

We note that using the two strategies we have stated above we can apply the canonization algorithm to Touraments in $n^{c \log n}$ time where $c = 1/2 + o(1)$ and on Bi-Partite graphs $G = B \cup C$ in $O(n^{\omega(d)+c})$ time where $\omega(d) = \max(d_{out}, cw(G))$ where d_{out} is the maximum degree of the vertices of B.

Application: Bounded Valence Graphs

To apply our algorithm for bounded valence graohs we start with a lemma.

Lemma

Let X be a connected graph of degree $\leq d$ and let e be an edge in X. Then the composition factors of $Aut(X(e))$ are subgroups of S_{d-1}. In particular,

$$cw(Aut(X(e)) \leq d - 1$$

The condition that X has degree $\leq d$ can be reduced to saying that the out-valence $(X(e)) \leq d - 1$. Using an inductive argument we see that the total time required is $\mathcal{O}(n^{\omega(d-1)+c})$.

Applications: General Graphs

- Using a Valence Reduction Lemma for the case of general graphs by Zemlyachenko we can see that for general graphs the computation of canonical form can be done in $exp(n^{1/2+o(1)})$ time where $n = |V|$.
- Zemlyachenko's Lemma actually reduces X to a graph with color valence $\leq d$.
- We can extend our previous result to see that for Coloured Graphs the canonical form can be computed in $\mathcal{O}(n^{\omega(d)+c})$ steps where $n = |V|$ if color valence of the graph $\leq d$.

Questions?

References

Babai, Luks
Canonical Labeling of Graphs
Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing
Published by ACM
New York, NY, USA

Eugene M. Luks
Isomorphism of graphs of bounded valence can be tested in polynomial time
Journal of Computer and System Sciences

Arvind, IMSc
Algebra and Computation Lecture Notes
https://www.imsc.res.in/ arvind/notes.pdf

Lang S.
Algebra
Published by Springer